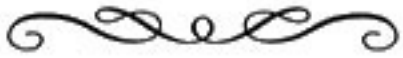
A multi-semester introduction to core concepts and their application in evolutionary, molecular, cellular & developing systems

Michael W. Klymkowsky & Melanie M. Cooper

Molecular, Cellular & Developmental Biology, University of Colorado Boulder
Chemistry, Michigan State University

– *pressbook version* –

with an introduction to genetics & developmental processes (pressbook version 1)



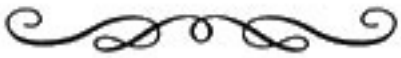
You know how it is.
You pick up a book, flip to the dedication & find that, once again,
the author has dedicated a book to someone else & not to you.

Not this time.

Because we haven't yet met/have only a glancing acquaintance/are just crazy about each other/haven't seen each other in much too long/are in some way related/will never meet, but will, I trust, despite that, always think fondly of each other….

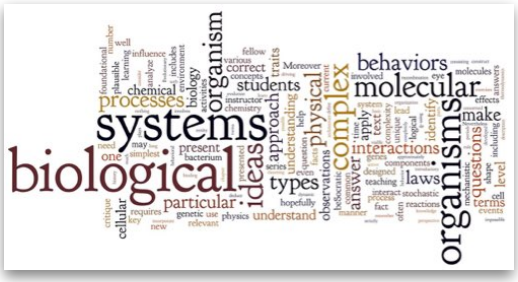
This one's for you.

for the explorer inside all of us



courtesy of Neil Gaiman

Preface: A biofundamentalist's approach to teaching & learning biology



Our overall goal in this project, since the beginning, has been to identify the underlying principles upon which biological systems are based, and to present them in as clear and coherent a narrative as possible. Once understood, the student needs to recognize where and when key concepts and disciplinary principles are relevant, and, as importantly, when they are not, together with how to apply them appropriately. The end goal is to enable you, the student, to approach most any biological system or process, from the origin of disease to how cooperation and kindness arise from a scientific perspective – to be able to generate plausible models and to consider how to test those models in the light of experimental evidence.

To understand biological systems we need to consider them from two complementary perspectives: 1) how they came to be – the historic, that is, the evolutionary, and 2) how they work – the physicochemical, physiological, and the mechanistic. How cellular structures, molecular and macromolecular components arise, and how they behave at the molecular, cellular, organismic, and ecological levels. We also consider what it means to read and answer a question scientifically, how to draw meaningful conclusions from data, and how to recognize when more (or better) data is needed. A recurrent theme is to consider how terms and concepts are to be used; if they are not useful, we will omit them.

We are biological entities, the products of evolutionary and developmental processes acting on inherited information stored in molecules and acting within dynamic (cellular) chemical systems. We live in complex and often unstable social arrangements with other humans and other organisms whose behaviors influence us in both subtle and profound ways.[1] As we alter our environment we inevitably alter ourselves. Science is a communal strategy by which we seek to better understand how the Universe works and how it might be manipulated. Science seeks to reveal how the physical world and its history shapes and constrains what is and what is not possible, and why this is so. That said, science does not provide us with a prescription for how things should be. Science cannot tell us what is morally right or wrong, it can only attempt to explain what is and predict what might be. Science requires a working and useable understanding of the Universe, and, at times, ourselves. Our scientific understanding of almost every topic, and particularly the remarkably complex behaviors of biological systems, is incomplete. It is not even certain that the Universe is coherent and self-consistent. The difficulties in producing a single theory that encompasses the behavior of both the very large and massive (gravity) and the very small (quantum mechanics) raises the possibility that a single theory of everything may not be possible or if possible, may not be comprehensible to us.[2] In a related way, the inherent impossibility of perfect accuracy means that biological systems will never be perfectly predictable.

While science is a powerful strategy to understand and manipulate the world, it is certainly no guide to moral behavior. Nevertheless its power can be seductive. Periodically a perspective (an ideology) known as scientism gains popularity in certain circles. Scientism holds that science provides a complete and exclusively valid description of the Universe, a picture that dictates how we should behave. We caution against this view, in part based on incompleteness of scientific knowledge, the lessons of history, and our deeply held belief that we are each unique individuals

---

[1] The global health and economic effects of the COVID-19 virus come to mind.

[2] Physics's pangolin: Trying to resolve the stubborn paradoxes of their field, physicists craft ever more mind-boggling visions of reality & Scientific method: Defend the integrity of physics

who are inherently valuable, deserving of respect irrespective of current scientific pronouncements. Human beings are not objects to be sacrificed on the altar of abstract ideals, that is, persecuted or harmed based on ideological grounds, whether based on scientific, political, religious, or economic beliefs. A number of serious crimes against humanity as a whole and specific individuals have been justified based on what are claimed to be established "facts" that later turned out to be untrue, incomplete, tragically misapplied, more or less irrelevant, or illusory.[3] Crimes against people in the name of science are as unforgivable as crimes against people in the name of religious beliefs, political ideologies, or simple selfishness, greed, or apathy toward the suffering of others.

That said, scientific thinking is indispensable if we want to distinguish established, empirically supported observations from frauds and fantasies. Such frauds and fantasies can often be harmful, such as the anti-vaccination campaigns that have led to an increase in deaths, birth defects and avoidable diseases.[4] When we want to cure diseases, reduce our impact on the environment, or generate useful tools we are best served by adopting a dispassionate, empirically-based scientific approach to inform, rather than dictate, our decisions. Scientific studies help us decide between the possible and the impossible and to assess the costs and benefits of various interventions. In this context it is worth noting that there are important differences between what has been established scientifically, what those conclusions imply, and how they interact with and influence other social, economic, political, and personal decisions.[5] Particularly important is the fact that all scientific conclusions are tentative, and subject to re-interpretation, although it certain that some are much more likely to be true, or rather more accurately reflect how the world works than others.

Scientific knowledge is a body of knowledge of varying degrees of certainty-some most unsure, some nearly sure, but none absolutely certain … Now we scientists are used to this, and we take it for granted that it is perfectly consistent to be unsure, that it is possible to live and not know. - Richard Feynman.

...it is always advisable to perceive clearly our ignorance. – Charles Darwin.

Montaigne concludes, like Socrates, that ignorance aware of itself is the only true knowledge - Roger Shattuck

How biology differs from physics and chemistry

While it is true that biological systems, that is, cells, organisms, and ecologies, obey the laws and principles of physics and chemistry, they are not deducible simply from a knowledge of physics and chemistry. They are more than just highly complex chemical and physical systems. Why is that? Because each organism is a unique entity, distinguishable from others by the genetic information it carries, the result of mutation and selection, and the stochastic events associated molecular and population level processes. Even identical twins (and quadruplets) can be distinguished in terms of their molecular and behavioral details.[6] Moreover, each organism is the product of a unique history that runs back in time for an unbroken period of more than ~3,500,000,000 years, where the symbol "~" means "approximately". To understand an organism's current shape, internal workings, and behaviors requires an appreciation of the general molecular, cellular, developmental, social, and ecological processes involved in producing these traits. Such mechanistic processes are themselves

---

[3] Walter Gratzer: The Undergrowth of Science

[4] How vaccine denialism in the West is causing measles outbreaks in Brazil & http://www.historyofvaccines.org/content/articles/history-anti-vaccination-movements & The World's Many Measles Conspiracies Are All the Same

[5] What Daniel Sarewitz terms trans-science: Saving science

[6] The impacts of stochastic molecular levels events have been studied in embryos of the nine-banded armadillo, which reproduce by producing four genetically quadruplets: see The transcriptional legacy of developmental stochasticity

the product of what the molecular biologist François Jacob (1920-2013) referred to as "evolutionary tinkering", that is, they reflect each organisms' unique evolutionary history, as well as its current environment.[7]

Looking at the evidence, it is clear that no organism, including ourselves, was designed *de novo* (that is from the Latin meaning, anew). Rather each organism is the product of continuous evolutionary processes that have been in play since the origin of life (~3.8-4.0 billion years ago). A particular individual does not evolve, but populations do. Evolution describes how populations change over time. The reason(s) for these changes involve various evolutionary mechanisms that act together, these have produced distinct populations of individuals adapted to particular life styles (ecological niches) through a combination of random (stochastic) and non-random events. These evolutionary mechanisms, which we will discuss in some detail, include the origin of mutations, that is, changes that alter the genetic material (double-stranded deoxyribonucleic acid, which we refer to as DNA) and the effects of these molecular variations (the organism's genotype) on the shape or behavior of the organism (the organism's phenotype). The genetic material is dynamic and subject to various forms of chemical modification, sequence additions, deletions, and shuffling. The primary driver of the phenotypic changes seen in populations over time is known as "selection" and is due to differences in reproductive success. Various types of selection arise through internal processes and an organism's interactions with other organisms and its environment. Because of the complexity of these processes, one cannot readily deduce the details of a particular organism from physical first principles (or even the sequence of its genome) – and there are many millions of different types (species) of organisms. Take for example the vertebrate eye, which behaves in accord with physical laws, yet displays idiosyncrasies arising from its evolutionary history. Such differences enable us to deduce that the vertebrate eye arose independently from, for example, the eyes of mollusks, that is squid and octopi.[8] Evolutionary processes lead to the emergence of new traits and modified types of organisms while at the same time playing a conservative role, maintaining organisms against the negative effects of molecular noise, that is mutations.[9] The interactions between organisms and their environment can lead to unpredictable evolutionary changes. They can result in the extinction of some lineages and the emergence of new "types" of organisms. Evolutionary processes have produced the millions of different types of organisms currently in existence, in addition to the many more that are now extinct.

Another important difference between biological and physicochemical systems is that even the simplest biological systems are more complex than the most complex non-biological physical system. A bacterium, one of the simplest types of organisms in terms of its molecular components, typically contains more than ~3000 distinct genes, and hundreds to thousands of concurrent and interdependent chemical reactions, whose interactions influence which genes are active (active genes are often said to be "expressed") and which are inactive (not expressed), the range of ecological and environmental interactions that occur between organisms, and how an individual bacterium responds to them. Often these processes are controlled by a small number (one to a few hundreds to thousands) of a particular type of molecule; the small number of molecules involved inevitably results in noisy (stochastic) behaviors that are difficult or impossible to predict on the individual cellular level. We will consider the implications of such stochastic processes repeatedly in various systems.

---

[7] François Jacob: Evolution and Tinkering & Tinkering: a conceptual and historical evaluation

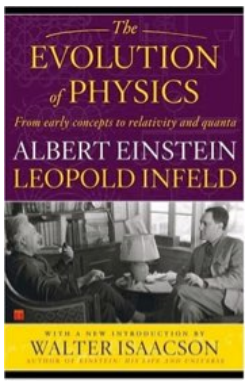
[8] How the Eye Evolved

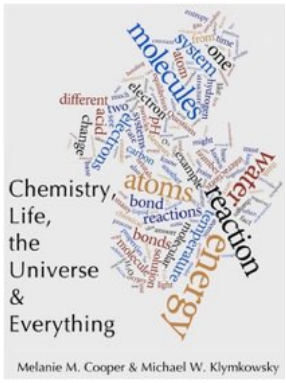
[9] From an evolutionary perspective, a mutation is be considered harmful if it negatively effects on organism's reproductive success; whether a mutation is harmful or beneficial is determined by the context in which it occurs (a point we will return to). There are, for example, cases where removing a gene opens up new possibilities - see When Less Is More: Gene Loss as an Engine of Evolutionary Change.

Not withstanding their complexity, there are common themes within biological systems that we will return to over and over again and that make such systems intelligible. We will rely on the fact that we can understand how molecules interact (through collisions and binding interactions), how chemical reactions interact with one another (through reaction coupling), and how physical laws, in particular the laws of thermodynamics, constrain and shape biological behaviors. The fact that all current (and past) known organisms appear to share a single common ancestor also helps.

Your background and our (Socratic) teaching approach

Biology students are often required to take general introductory physics and chemistry courses. Too often these courses are taught without regard to their relevance to the understanding of biological systems, a situation that seems counter-intuitive and counter-productive. We advocate redesigning introductory chemistry and physics courses so that their relevance to biology is explicit,[10] but recognize that this is rarely the case. We also recognize that many students may not be completely comfortable with the physical and chemical concepts relevant to biology, so we have written biofundamentals presuming very little. Where references to physicochemical concepts are necessary, we have attempted to address them at a level that we believe will be adequate for you to be able to deal productively with the ideas presented. That said, it is your responsibility as a learner to speak up if you do not think (or feel) that you understand an idea or grasp its significance in a particular situation. We suggest that students interested in learning more about the physical and chemical concepts that underlie biological systems read Einstein & Infeld's "The Evolution of Physics"[11] and our own "Chemistry, Life, the Universe, and Everything" (CLUE).[12]





The complexity of biological systems can be daunting and all too often biology is presented as a list of vocabulary terms, with little attention paid to its underlying conceptual (sense-making) foundations. This emphasis on memorization can be off-putting and, in fact, is not particularly valuable in helping you, the learner, develop a working understanding of biological systems. Our driving premise is that while biological systems are complex, both historically and mechanistically, there are a limited set of foundational observations and general principles that apply to all biological systems.[13] The complexity of biological systems, and the incompleteness of our understanding of them, often make an unambiguous (final) answer to biological questions tentative. Nevertheless, it is possible to approach biological questions in an informed, data-based (empirical), and logical manner. In general, we are less concerned with whether you can remember or reproduce the "correct" answer to a particular question and more interested in your ability to identify the observations and over-arching concepts relevant to a question or scenario and to then construct a scientifically plausible, logical, and internally consistent response. More often than not, such a response will be the correct one, or close to it.

Going beyond memorization means that you will need to apply your understanding of key facts, terms, and overarching principles to particular situations; this requires that you develop, through

---

[10] Physics for (molecular) biology students.

[11] Einstein and Infeld's The evolution of physics

[12] CLUE: Chemistry, Life, the Universe & Everything; Organic CLUE may also be useful.

[13] Klymkowsky: Thinking about the **conceptual** foundations of the biological sciences.

practice, the ability to analyze biological situations, to identify what factors are critical, recognize those that are secondary or irrelevant, and then apply your understanding to make predictions or critique conclusions. To give you opportunities to develop these skills, each section of the book includes questions to answer and ponder. As you work with the ideas involved, we expect you will learn to able to generate, explain, and defend plausible, rather than "correct", scenarios, to present them to your instructor and fellow students, and to defend or revision your thinking in response to critical (socratic) questions. When you do not understand how to approach a question you should try to articulate exactly what is confusing you, something that can take serious introspection.

> *We think the way we do because Socrates thought the way he did.*
> *- Bettany Hughes*

As part of this process, we use web-based beSocratic (link) activities to frame in class discussions.[14] These activities are designed to help you develop your ability to analyze problems and to construct models and explanations. In many cases, you will receive feedback within the context of the activity. That said, there is no substitute for engaging in discussions with other students and your instructors. Ideas that you find obscure or that make no sense to you need to be addressed directly, do not let them go unchallenged! Learning to critique or question an explanation will help you identify what is relevant, irrelevant, conceptually correct, or logically absurd in your and your fellow students' thinking. Remember, our goal is that by the time we reach the end of the course you will have learned something substantial about biological systems, and yourself. One mark of an educated person is that they can accurately detect BS in their own thinking, and the thinking of others - this is socratic thinking.[15]

Learning how to explain, critique, and argue scientifically: We have noticed that students often have a difficult time generating scientifically plausible explanations for biological processes, or in explaining the reasoning behind their choices on multiple choice type exams. To this end it is critical that you spend time organizing your thoughts and generating explanations, arguments, or critiques based on explicitly stated assumptions and logic. Practice, feedback, and revision are critical in order to learn how to write (and think) effectively. Learning how to defend (or abandon) ideas in response to questioning is a powerful tool for consolidating your knowledge. This process reflects the fact that "hard thinking" and clear (articulate) speaking and writing are not natural, but need to be learned, nurtured, and mastered.[16]

When you are answering a question we suggest that you write out your answer and then read it back to yourself.[17] Reading your own writing out loud (or having your computer read it) can help you recognize awkwardly phrased or illogical constructions that you might miss when you skim over the words.[18] In part this is due to the fact that different parts of the brain are involved in active listening.[19]

What we are not "covering": An important point is that our aim is to provide an engaging narrative together with a concerted effort to avoid unnecessary distractions. Why? Because it has been found that while experts focus, often unconsciously, on the key aspects of a problem or system, novices, such as students in an introductory biology class, tend to take everything equally seriously – which can be quite distracting. We aim to focus on core terms, concepts, general principles, and key

---

[14] beSocratic is back and we are exploring tools to support useful discussion between students.

[15] Issac Newton and BullSh*t detector A Guide to Being Less Wrong. Also see "On Bullshit" and the book "Calling BS".

[16] Review of "Thinking fast and slow"

[17] NYT: The Benefits of Talking to Yourself

[18] Reading aloud: http://writingcenter.unc.edu/handouts/reading-aloud/

[19] Speech and the Brain: http://webspace.ship.edu/cgboer/speechbrain.html

observations that we will call upon repeatedly. Details will be avoided unless they are critical – as an example, there are many proteins involved in DNA replication, but a key fact is that (most) polymerases work in one direction only, a fact that impacts the behavior of biological systems and one you need to remember, as you will see when we get to it. If you think we have introduced a distraction, please let us know.

Revisions to the text: biofundamentals began as an alternative introductory course in evolutionary and molecular biology. Because the ideas and observations presented are well established, we expect no need for dramatic revisions of content due to new discoveries. At the same time, the advent of inexpensive genomic and single cell RNA sequencing and related techniques, together with high resolution mass spectrometry have led to a flood of observations that illuminate key points and they have incorporated as appropriate.[20] It is, of course, possible that we have missed some important things - if so, let us know and we will consider how they fit into the narrative. We originally thought of biofundamentals as a one semester course, but over the decade it has extended itself and now is more like a three semester course (or the basis of a multicourse curriculum).

That said, we have learned a lot from various studies and personal experiences on how students interact with, and apply (or ignore) the ideas that have been presented to them. In particular our approach to genetic ideas has been influenced by both the complexity of the relationships between genotype and phenotype and the social impacts of how genetic ideas are presented, particularly in terms of the obsolete term "race", a flawed concept that can lead to noxious and scientifically incorrect conclusions. Here our thinking has been influenced by the work of Brian Donovan and colleagues.[21]

At the same time, we have much to learn about how to best help students master and apply complex biological ideas, so we are using student responses from the on-line activities and classroom interactions to identify necessary (and sometimes difficult) ideas and to build more effective learning activities.[22] New “editions” will incorporate these insights. Check the “version date” at the bottom of each page to insure you have the latest version. Observations, criticisms, and suggestions are greatly appreciated, and we welcome your comments on the text and course design.

A note on footnotes: We have an inordinate fondness for footnotes. We do not expect you, the student or the casual reader, to read them or to follow the links within them. Please be careful to avoid getting lost in, or distracted by, the footnotes - although sometimes the world is a labyrinth with treasures (and monsters) along the way.

---

[20] see for example polypeptides and proteins and why genes are getting weirder.

[21] In particular see Donovan B. M. (2014). "Playing with fire? The impact of the hidden curriculum in school genetics on essentialist conceptions of race." Journal of Research in Science Teaching 51: 462-496. And Donovan et al., (2019). "Toward a more humane genetics education: Learning about the social and quantitative complexities of human genetic variation research could reduce racial bias in adolescent and adult populations." Science Education 103: 529-560.

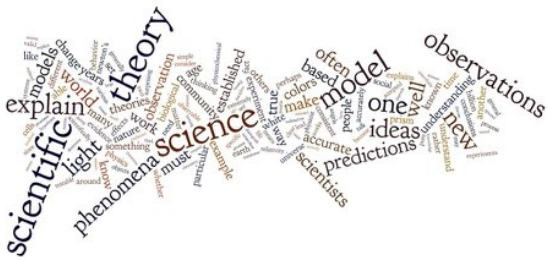
[22] The Design and Transformation of Biofundamentals: A Nonsurvey Introductory Evolutionary and Molecular Biology Course

PART I - Foundations

In which we consider evolutionary mechanisms,
the physicochemical properties of cells,
the capture of energy,
the basic nature of genetic information,
how it accumulates and is encoded, replicated, and read out,
together with how
proteins are assembled, interaction and "work".

Chapter 1: Understanding (biological) science & thinking scientifically

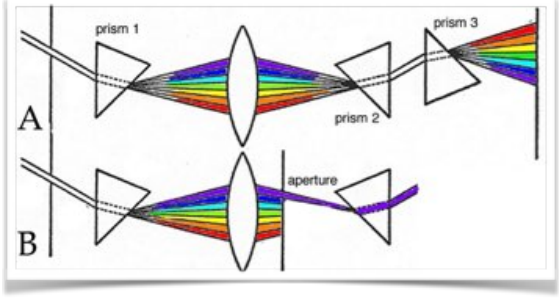
In which we consider what makes science a distinct, productive, and progressive way by which to understand how the universe works. Science enables us to identify what is possible and plausible and what is appears to be impossible or highly implausible. We consider the "rules" that distinguish a scientific approach to a problem from a non-scientific one.



A major feature of science, and one that distinguishes it from many other human activities, is its essential reliance upon shareable experiences rather than personal revelations. Thomas Paine (1737-1809), one of the intellectual parents of the American Revolution, made this point explicitly in his book The Age of Reason (↓).[23] In science, we do not accept that an observation or a conclusion is true simply because another person claims it to be true. We do not accept the validity of revelation or what we might term "personal empiricism." What is critical is that, based on our description of a phenomenon, an observation, or an experiment, others should, if they have the resources, opportunities, and resources, be able to repeat our work. Science is based on social, that is, shared, knowledge rather than revealed (personal) truth.

> *Revelation is necessarily limited to the first communication – after that it is only an account of something which that person says was a revelation made to him; and though he may find himself obliged to believe it, it can not be incumbent on me to believe it in the same manner; for it was not a revelation made to ME, and I have only his word for it that it was made to him.*
> *– Thomas Paine, The Age of Reason.*

As an example consider sunlight. It was originally held that white light was "pure" and that somehow, when light passed through a prism, the various colors of the spectrum, the colors we see in a rainbow, were created *de novo*. In 1665, Isaac Newton (1642–1727) performed a series of experiments that he interpreted as demonstrating that white light was not "pure", but was composed of light of many different colors.[24] This conclusion was based on a number of observations. First, he noted that passing sunlight through a prism generated a spectrum of many colors. He then used a lens to focus the spectrum emerging from one prism so that it passed through a second prism (Part A↓): a beam of white light emerged from the second prism. He went on to show that the light emerging from the prism 1 lens prism 2 combination behaved the same as the original beam of white light; when passed it through a third prism it again produced a spectrum. In a second type of experiment (Part B→), Newton used a screen with a hole in it, an aperture. He found that light of a particular color was not altered when it passed through a second prism - no new colors were emerged. Based on these observations, Newton concluded that white light was not what it appeared to be – that is, a simple "pure" substance – but rather was composed, unexpectedly, of light of many distinct colors. The spectrum was produced because the different colors of light were "bent" or refracted by the prism to different extents. Why this occurred was not clear at the time nor was it clear what, exactly, light is. Newton's experiments left these questions unresolved. This is typical: scientific answers are often extremely specific, elucidating a particular phenomenon, rather than providing a universal explanation.



[23] The Age of Reason: http://www.ushistory.org/paine/reason/singlehtml.htm

[24] Newton's Prism Experiments & http://youtu.be/R8VL4xm_3wk

Two basic features make Newton's approach, observations and conclusions, scientific. The first is its reproducibility. Based on his description of his experiment others could, and did reproduce, confirm, and extend his observations. If you have access to glass prisms and lenses, you can repeat Newton's experiments yourself. You will observe the same phenomena that Newton did.[25] In 1800, William Herschel (1738-1822) did just that. He used Newton's experimental approach and discovered infrared (beyond red) light. While infrared light is invisible to us, other organisms can see it. Its presence can be revealed by the fact that when absorbed by an object, say by a thermometer or a human hand, it leads to an increase in the temperature of the object.[26] In 1801, inspired by Herschel's discovery, Johann Ritter (1776-1810) used the ability of light to initiate the chemical reaction:

silver chloride + light → silver + chlorine

to reveal the existence of another type of light, which Ritter called "chemical light" and that we refer to as ultraviolet light.[27] Subsequent researchers established that visible light accounts for a small portion of a much wider and continuous spectrum of "electromagnetic radiation", ranging from X-rays to radio waves. Studies on how light interacts with matter have led to a wide range of technologies and have helped to construct a coherent understanding of the history of the Universe. All these findings emerge, rather unexpectedly, from attempts to understand the rainbow.

The second scientific aspect of Newton's work was his clear articulation of the meaning and implications of his observations, the logic and limitations of his conclusions. These led to explicit predictions, such as that a particular color will prove to be homogenous, that is, not composed of other types of light, which he then confirmed. His view was that the different types of light, which we see as different colors, differ in the way they interact with matter. One way these differences are revealed is the extent to which the different colors of light are bent when they enter a prism. Newton used some of these ideas when he chose to use mirrors rather than lenses to build his reflecting (Newtonian) telescope. His design avoided the color distortions that arise when light passes through simple lenses.

The features of Newton's approach make science, as a social and progressive enterprise, possible. We can reproduce an observation or experiment, and follow the investigator's explicit thinking. We can identify unappreciated factors that can influence the results observed and identify inconsistencies in logic and explore unappreciated implications that may influence other scientific disciplines. Science rests on the premise that there is a world outside ourselves, that this world is real and constrains what is possible and what is not possible – it rules out "magical thinking", and so can be upsetting to some. It is also the case that science is not about discovering over-arching and immutable truth (aside from the reality of the world), but rather about developing a working understanding of how objects in the world can be expected to behave.

The interconnectedness (self-consistency) of science

It was once thought that there were aspects of biological systems that somehow transcended physics and chemistry, a presumption known as vitalism. If vitalism had proven to be correct, it would have forced a major revision of chemistry and physics. As it turns out, vitalism was wrong. The world described by the sciences is like an extremely complex crossword puzzle (→), where the answer to one question must be compatible with the answers to all

[25] Infrared astronomy

[26] There are some animals that can see infrared light: see link & link

[27] Ritter discovers ultraviolet light

other questions.[28] Alternatively, certain questions, and their answers, once thought of as meaningful can come to be seen as irrelevant or meaningless (not part of the puzzle). For example, how many angels can dance on the head of a pin is no longer considered relevant to a scientific explanation.

What has transpired over the years is that biological processes ranging from the metabolic to the conscious have been found to be consistent with physicochemical principles. What makes biological processes different is their complexity and the fact that they are the product of evolutionary processes, processes influenced by stochastic and historical events that stretch back in an uninterrupted "chain of being" over billions of years. Moreover, biological systems in general are composed of many types of molecules, cells, and organisms that interact in complex ways. All this means is that while biological systems obey physicochemical rules, their behavior often cannot be predicted based on these rules. It may well be that life, as it exists on Earth, is unique in the Universe. The only way we will know for sure is if we discover life on other planets, in other solar systems and galaxies. At present, based on many observations, it appears that all life we know of is related, all organisms are modified (evolved) versions of a "last common universal ancestor", known as LUCA. If other kinds of life are possible, we have no evidence for them - we do not know the "general rules" governing life and its appearance because we only know of one type of life, that found on Earth.

On the other hand, it is possible that studies of biological phenomena could lead to a serious rethinking of physicochemical principles. There are, in fact, research efforts into proving that phenomena such as extrasensory perception, the continuing existence of the mind/soul after death, and the ability to see the future or remember the (long distant) past are real. At present, these all represent various forms of pseudoscience, and most likely, self-delusion and wishful thinking, but they would produce a scientific revolution if they were shown to exist, that is, if they were reproducible and based on discernible mechanisms with explicit implications and testable predictions. These examples emphasize a key feature of scientific explanations: they **must** produce logically consistent, explicit, testable, and potentially falsifiable predictions. Ideas that can explain any possible observation or are based on untestable assumptions, something that some would argue is the case for a number of religions (and aspects of modern physics), are no longer science, whether or not they are "true" in some unprovable sense.[29]

Models, hypotheses, and theories

Scientific models are used in various ways. There are explanatory models that capture a certain approach to a system as well as exploratory and predictive models that are used to test ideas. Predictive, mechanistic models are commonly known as hypotheses. Models are valuable in that they serve as a way to clearly articulate one's assumptions and their implications. They form the logical basis for generating testable predictions about the phenomena they purport to explain. As scientific models become more sophisticated, their predictions can be expected to become more and more accurate or apply to areas that previous forms of the model could not handle. Let us assume that two models are equally good at explaining a particular observation. How might we decide between them? One way is the rule of thumb known as Occam's Razor, named after the medieval philosopher William of Occam (1287–1347). Occam's Razor, also known as the Principle of Parsimony, states that all other things being equal, the simplest explanation is to be preferred. This is not to imply that an accurate scientific explanation will be simple, or that simple explanations are correct, only that to be useful, a scientific model should not be more complex than necessary. Consider two models for a particular phenomenon, one that involves angels and the other that does not. We need not seriously consider the model that invokes angels unless we can accurately monitor the presence of angels and if so, whether they are actively involved in the process to be explained.

---

[28] This analogy is taken from a talk by Alan Sokal:; graphic here

[29] see Farewell to Reality, Not even Wrong, Wronger than Wrong & Lost in Math

Why? Because angels, if they exist, imply more complex factors than does a simple natural explanation. For example, we would need to explain what angels are made of, their origins, and how they intervene in, or interact with the physical world, that is, how they make matter move. Do they obey the laws of thermodynamics? What determines when and where they intervene? Are their interventions consistent, purposeful, or capricious? Assuming that an alternative, angel-free model is as or more accurate at describing the phenomena and making verifiable predictions, the scientific choice would be the angel-free model. Parsimony (an extreme unwillingness to spend money or use resources) has the practical effect that it lets us restrict our thinking to the minimal model that is needed to explain specific phenomena. The surprising result, illustrated by a talk by Murray Gell-Mann[30], is that simple, albeit often counter-intuitive rules can explain much of the Universe with remarkable precision. A model that fails to accurately describe and predict the observable world must be missing something and is either partially or completely wrong (no matter how "beautiful").

Scientific models are continually being modified, expanded, or replaced in order to explain more and more phenomena more and more accurately. It is an implicit assumption of science that the Universe can be understood in scientific terms, and this presumption has been repeatedly confirmed but has by no means been proven. A model that has been repeatedly confirmed and covers many different observations is known as a theory – at least this is how we will use the word.[31] It is worth noting that the word theory is often misused, even by scientists who might be expected to know better. If there are multiple "theories" to explain a particular phenomenon, it is more correct to say that i) these are not actually theories, in the scientific sense, but rather working models or speculations, and that ii) one or more, and perhaps all of these models are incorrect or incomplete. A scientific theory is a very special set of ideas that explains, in a logically consistent, empirically supported, and predictive manner a broad range of phenomena. Moreover, a theory has been tested repeatedly by a number of critical and independent people – that is, people who have no vested interest in the outcome – and it must be found to provide accurate descriptions of the phenomenon it purports to explain. It is not idle speculation. If you are curious, you might count how many times the word theory is misused, at least in the scientific sense, in the course of your day to day experiences.

That said, theories are not static. New or more accurate observations that a theory cannot explain will inevitably drive the theory's revision or replacement. When this occurs, the new theory explains the new observations as well as everything explained by the older theory. Consider for example, gravity. Isaac Newton's law of gravity describes how objects behave; it is possible to make extremely accurate predictions of how objects behave using its rules. However, Newton did not really have a theory of gravity, that is, a naturalistic and mechanistic explanation for why gravity exists and why it behaves the way it does. He relied, in fact, on a supernatural explanation.[32] Later on, it was found that Newton's law of gravity failed in specific situations, such as when an object is in close proximity to a massive object like the sun. New rules were needed. Albert Einstein's Theory of General Relativity not only more accurately predicts the behavior of these systems, but also provides a naturalistic explanation for the origin of gravitational forces.[33] It also makes predictions about future observations, such as gravity waves, that have subsequently been confirmed.[34] So is general

Gravity explains the motions of the planets, but it cannot explain who sets the planets in motion.
- Isaac Newton

[30] Murry Gell-Mann: Beauty, truth and ... physics?

[31] Ideas are cheap, theories are hard

[32] Want to read an interesting biography of Newton, check out "Isaac Newton" by James Gleick

[33] A good video on General Relativity [here]

[34] Physicists find another gravitational wave to suggest that Einstein was right

relativity true? Not necessarily, which is why scientists continue to test its predictions in increasingly extreme situations and to higher and higher degrees of accuracy.

Knowing what you know: constructing models, answers, explanations & critiques

How do we know what we know? This is a central question in philosophy and is equally relevant to teaching and learning. There is plenty of evidence that people consistently over-estimate their own skills, including what they believe they have learned in a class.[35] There is, however, a well-established approach to evaluating one's, and other's, understanding, namely the Socratic dialog. In a Socratic dialog with an engaged and critical person, we can recognize our assumptions and consider the extent to which they are relevant and valid. We use Socratic dialog when we ask you about your answers to questions and when you consider the statements of others: is your application of scientific concepts and relevant observations appropriate and logical? Have you left out important considerations or are unspoken assumptions in play? You should be ready to discuss, Socratically, the answers to the "questions to answer and ponder" found throughout the book.

To answer and explain, it is important to be clear that you understand exactly what it is that the question you are being asked wants to know, or what you need to explain. The ability to read a question, accurately decode what it is asking, and to then compose a coherent and evidence-based response requires basic literacy.[36] While it may be difficult or awkward to ask for clarifications of a question, that is, exactly what you need to do (and what a working scientist would do!) Always feel free to give voice to your confusions and to ask your clarifying questions.[37] It helps to frame your questions in the context of what you think the question is asking and why; what do you find it unclear or confusing. In a testing scenario, this can also be a useful strategy. Restate what you think the question is asking and then answer that question. By asking questions in class or talking with classmates, you can clarify what a question is about, or you can help explain it to others and yourself. If they are equally confused ask the instructor. Typically we will share questions and our responses with the class, since it is very likely that you are not the only person who wants or needs clarification.

Once you understand what a question wants you to explain, you can begin to construct your response. You first need to identify what facts and general principles apply; these will be used in the construction of your answer. As an example, consider the question: "Based on the accumulation of an isotope that is known to be generated only by radioactive decay, a geologist claims a particular rock is ~2 billion years old, while a creationist claims that the rock is ~6000 years old. Why can't both be correct?" To answer the question, we begin by clearly articulating to ourselves what the question and its possible answer is based on. Geologists date rocks, typically igneous (originally molten, often volcano-derived) based on assumptions about the rock's stability and composition. Many observations indicate that the rate and products of the radioactive decay of a particular isotope are constant and universal; they are not influenced by other factors. Assuming that the rock used to assign a date is stable, that is, no atoms enter or leave it, then the ratio of the original isotope and the isotope produced by its decay serves as an atomic clock, providing an estimate of the age of the rock, that is the time since its formation. Fossils are found in sedimentary rocks, but not volcanic ones, since the heat associated with volcanic rocks generally destroys organic remains. Sedimentary rocks are difficult to date accurately, since they are derived, through processes of erosion and deposition from other, older rocks. The geologist dates the fossil containing rock based on the age of the surrounding rock layers. It is less clear what scientific ideas the creationist uses to date rocks and the fossils within them. Since there is no evidence that rates of radioactive decay

---

[35]The Kruger & Dunning effect: Unskilled and Unaware

[36] Norris & Phillips. 2003. How literacy in its fundamental sense is central to scientific literacy

[37] The answers can often be surprising. see McClymers & Knowles.Ersatz Learning, Inauthentic Testing

have changed over the history of the Universe, and assuming no other natural processes are at play (and it is hard to imagine what they might be), the creationist is most likely to be incorrect – their assumptions implicitly contradict well established knowledge from physics, chemistry, and geology.

As you can see, answering a question can be a complex process – constructing an answer can rely on a number of assumptions that need to be recognized and stated explicitly. In the case of dating a fossil, you would consider the observed rate of radioactive decay, the method used to date sedimentary (and igneous) rocks, and the mechanism(s) by which fossils are generated. Our answer needs to identify the assumptions we are making. The complexity of explaining why correct answers are correct is one of the reasons that we may ask you to explain why wrong answers, such as those found in multiple-choice type questions, are wrong or irrelevant. Typically a wrong answer is wrong for a single incorrect assumption or, if correct, is irrelevant to the question at hand.

A similar situation applies when explaining something to someone, you need to identify the various ideas and the observations upon which those ideas are based, what the person you are talking to will need to know to be able to understand your explanation. You should also determine whether they understand what you think they understand. As an example, consider the short video interview [video link →] with the physicist Richard Feynman (1918-1988); in it he describes what it takes to explain magnetic attraction. As you start answering or explaining, you need to be prepared to explain the underlying ideas you are using – the person you are talking with can be expected to ask you to justify your assumptions, clarify your logic, and defend your conclusions. You are taking part in a Socratic dialog. The same applies when you are in class listening to an explanation from an instructor; do their assumptions make sense to you? Are they telling you all you need to know to be able to understand their explanation? Similarly, when you are listening to someone else's explanation, you need to consider whether the evidence they are using is correct, relevant and complete, do their conclusions follow logically? In a scientific discussion, are the methods they are using capable of generating the data upon which their argument rests?

It can be helpful to study with a group of people who are comfortable questioning and explaining to each other, but beware, groups do not always arrive at coherent or reasonable conclusions. It is important to check the group's conclusions by presenting them to a knowledgeable expert (hopefully your instructor). But we often find ourselves called upon to learn materials on our own. One way to cope is to develop your own "inner Socrates", a habit of mind that helps challenge and refine your thinking by asking "am I answering the question I am being asked? have I identified the key ideas and observations needed to answer the question? Are there other observations or concepts that should be considered? Are other, simpler explanations possible?" This is one area in which talking out loud to yourself can be useful!

Questions to answer:

1. How would you use Occam's razor to distinguish between two equally accurate models?
2. What does it mean when there are two explanations for the same phenomena? Can both be correct? How might you resolve this situation?
3. Outline your approach to deciding whether a particular idea, model, or hypothesis is scientific.

Science is social

The social nature of science is something that we want to stress yet again. Science is often portrayed as an activity carried out by isolated (and sometimes crazy or otherwise deranged) individuals, the image of the mad scientist comes to mind (→). The reality is different, science is an extremely social activity. It works only because it involves and depends upon an interactive community who keep each other, in the long run,

honest and anchored in objective reality.[38] Scientists present their observations, hypotheses, and conclusions in the form of scientific papers, where their relevance and accuracy can be evaluated, more or less dispassionately, by others with a working knowledge of the topic under study.

Over the long term, this process of socratic interactions leads to an evidence-based consensus. Certain ideas and observations are so well established that they can be reasonably accepted as universally valid, whereas others are extremely unlikely to be true, such as the possibility of perpetual motion machines and zero-waste processes (a version of the same idea) or "intelligent design creationism." These are ideas that can be safely ignored. As we see it, modern biology is based on a small set of theories: these include the Physicochemical Theory of Life, the Cell Theory, and the Theory of Evolution.[39] That said, as scientists we keep our minds open to exceptions and work to understand them and their implications. The openness of science means that a single person, taking a new observation or idea seriously, can challenge and change accepted scientific understanding. That is not to say that it is easy to change the way scientists think. Most theories are based on large bodies of evidence and have been confirmed on multiple occasions using multiple methods. It turns out that most "revolutionary" observations are either mistaken, misinterpreted, or can be explained within the context of established theories. It is, however, worth keeping in mind that it is not at all clear that all phenomena can be put into a single "theory of everything." It has certainly proven difficult to reconcile quantum mechanics with the general relativity.

A final point, mentioned before, is that the sciences are not independent of one another. Ideas about the behavior of biological systems cannot contradict well established observations and theories in chemistry or physics. If they did, one or the other would have to be modified. For example, there is substantial evidence for the dating of rocks based on the behavior of radioactive isotopes. There are also well established patterns of where rock layers of specific ages are found. When we consider the dating of fossils, we use rules and evidence established by geologists. We cannot change the age we assign to a fossil, making it inconsistent with the rocks that surround it, without challenging our understanding of the atomic nature of matter, the quantum mechanical principles involved in isotope stability, or a range of geological mechanisms. A classic example of this situation arose when the physicist William Thompson (1824-1907), also known as Lord Kelvin, estimated the age of the Earth to be between ~20 to ~100 million years, based on the assumption that the Earth was once completely molten together with the known rate of heat dissipation of such a massive molten object.[40] This was a time-span that seemed too short for a number of geological and evolutionary processes, and greatly troubled Charles Darwin. Somebody was wrong, or better put, their understanding was incomplete or incorrect. The answer in this case was with the assumptions that Kelvin made. His calculations ignored the effects of radioactive decay, not surprising since radioactivity had yet to be discovered. Including the heat released by radioactive decay in such calculations led to an increase in the estimated age of the Earth to ~5 billion years, an age compatible with both evolutionary and geological processes.

Teaching and learning science

An important point to appreciate about science is that because of the communal way that it works, understanding builds by integrating new observations and ideas into a network of previously established ideas and observations. Following this discipline, science often arrives at conclusions that can be strange, counterintuitive, and sometimes disconcerting but that are nevertheless logically

---

[38] A good introduction of how science can be perverted is "The Undergrowth of Science" by Walter Gatzer. You might also want to watch the "The Centrifuge Brain Project" | A Short Film by Till Nowak and consider whether it is scientific or not.

[39] Thinking about the conceptual foundations of the biological sciences

[40] An interesting book on this topic is "Discarded Science: Ideas That Seemed Good at the Time" by Paul Barnett

unavoidable. While it is now accepted that the Earth rotates around its axis and travels around the sun, which is itself moving around the center of the Milky Way galaxy, and that the Universe as a whole is expanding at what appears to be an ever increasing rate, none of these facts are immediately obvious and relatively few people who believe or accept them would be able to explain exactly how we have come to know that these ideas accurately reflect the way the universe works (or at least how it appears to work). At the same time, when these ideas were first being developed they conflicted with the assumption that the Earth was stationary, which, of course it appears to be, and that it is located at the center of a static Universe, which again seems quite reasonable. Scientists' new conclusions about the Earth's actual position in the Universe could be seen as a threat to the sociopolitical order. A number of people were persecuted for holding "heretical" views on the topic. Most famously, the mystic Giordano Bruno (1548-1600) was burnt at the stake for holding these and other ideas, some of which are similar to those proposed by modern physicists. Galileo Galilei (1564-1642) one of the founders of modern physics, was arrested in 1633, tried by the Roman Catholic Inquisition, forced to publicly recant his views on the relative position of the Sun and Earth, and spent the rest of his life under house arrest.[41] In 1616 the Church placed Galileo's book, which held that the sun was the center of the solar system, on the list of forbidden books – it remained there until 1835.

The idea that we are standing on the surface of a planet that is rotating at ~1000 miles an hour and flying through space at ~67,000 miles per hour is difficult to reconcile with our everyday experience, yet science continues to generate (and provide confirmatory evidence for) even weirder ideas. Based on observations and logic, it appears that the Universe arose from "nothing" ~13.8 billion years ago.[42] Current thinking suggests that the Universe will continue to expand forever at an increasingly rapid rate. Einstein's theory of general relativity implies that matter distorts space-time, which is really one rather than two discrete entities, and that this distortion produces the attraction of gravity and leads to black holes. A range of biological observations indicate that all organisms are derived from a single type of ancestral uni-cellular organism (LUCA) that arose from non-living material between 3.5 to 3.8 billion years ago. There appears to be an uninterrupted link between LUCA and every cell in your body, and to the cells within every other living organism, including whales, ants, cats, carrots, and tardigrades, and the various microbes that live in your gut and on your skin. You yourself are a staggeringly complex collection of cells. Your brain and its associated sensory organs, which act together to generate consciousness and self-consciousness, contains ~86 billion (10^9) neurons as well as a similar number of non-neuronal (glial) cells. These cells are connected to one another through ~1.5 x 10^{14} connections, known as synapses.[43] How exactly such a system produces thoughts, ideas, dreams, feelings, and self-awareness remains obscure, but it appears that these are all emergent behaviors that arise from this staggeringly complex natural system. Scientific ideas, however weird, arise from the interactions between the physical world, our brains, and the social system of science that tests ideas based on their ability to explain and predict the behavior of the observable universe.

Understanding scientific ideas

One of the difficulties in understanding scientific ideas and their implications is that these ideas build upon a wide range of observations and are intertwined with one another. One cannot really understand biological systems without understanding the behavior of chemical reaction systems, which in turn requires an understanding of molecules, which rests upon an understanding of how atoms and energy behave and interact. It is our working premise that to understand a topic, or a

---

[41] The History, Philosophy, and Impact of the Index of Prohibited Books

[42] The Origin Of The Universe: From Nothing Everything?

[43] Are There Really as Many Neurons in the Human Brain as Stars in the Milky Way? & Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled–up primate brain

discipline, it is necessary to know the key observations and common principles upon which basic conclusions and working concepts are based. To test one's understanding of a system, you need to be able to construct plausible claims for how, and why the system behaves the way it does, and how various perturbations can be expected to influence it. Your analysis needs to be based on facts, observations, or explicit presumptions that logically support your claim. You also need to be able to present your model to others, knowledgeable in the topic, in a clear way in order to get their feedback, to answer rather than ignore or disparage their questions, and address their criticisms and concerns.[44] Sometimes you will be wrong because your knowledge of the facts is incomplete or inaccurate, your understanding or application of general principles is incorrect, or your logic is faulty. It is important to appreciate that generating coherent scientific explanations and arguments takes time and can be difficult. We hope to help you learn how to understand biological systems and processes through useful coaching and practice. In the context of various questions, we and your fellow students, will attempt to identify when you produce a coherent critique, explanation or prediction, and where you fall short. Our goal is to help you learn how to think accurately and Socratically about biological systems.

Distinguishing the scientific from the trans-scientific

When we consider various personal and public policy decisions, including the ramifications of global warming, and what to do about it, the genetic engineering of human embryos and other organisms, and more generally the use of genetic data in medicine and society, as well as the costs and benefits of various science-informed decisions, we are often told that science has reached a consensus, but what exactly does that mean? By consensus, we mean the common conclusions accepted by scientists working in the field, conclusions supported by available evidence – what we might term "working knowledge". But evidence is rarely complete; for example, measurements can always be more accurate. In addition, when approaching a system scientifically, it is often necessary to make simplifying assumptions. These simplifying assumptions make the system tractable, they make it possible to make the kinds of unambiguous predictions upon which science is based. But when we want to act on scientific conclusions on complex systems such as the human brain and body, Earth's climate, or the response of individuals to specific medical treatments, we find that outcomes are less predictable. How a particular person responds to a particular drug is influenced by many, often interacting, factors, not all of which are perfectly defined in our working model. The limits of our understanding mean that interventions have side-effects, both desirable and undesirable. Only treatments that do nothing, homeopathy comes to mind, have no effects[45] (aside from leaving a serious condition untreated.)[46] There are risks in taking a drug, getting vaccinated, undergoing a surgery, opening or closing nuclear (or coal-based) power plants, but knowing **exactly** what the costs and benefits are may be difficult to predict.

Moreover, such a cost-benefit analysis, when applied to political, social, or economic decisions, often involves non-scientific factors. Consider, for example, the interconnected issues of increasing population, poverty, industrialization, and the ecological impacts of humans. One can argue, rather convincingly, that bringing basic human rights and autonomy, together with access to contraception, to women will help control human population growth – it has already led to reduced populations (fewer children per person) in much of the world.[47] At the same time, the idea of female autonomy

---

[44] This is exact opposite of the alt-fact environment that appears to be all the rage (and depressingly common) these days.

[45] Because homeopathic remedies are in most cases water or other inert chemicals. As we go along, given what we know about the movement of molecules and their constant collisions, you can probably explain why, for homeopathy to work, many laws of physics and chemistry would have to be broken.

[46] The case of Steve Jobs and his pancreatic cancer is a case in point. see link

[47] Hans Rosling: Don't Panic – The Facts About Population

can be deeply troubling (divisive) in certain tradition- and theologically-dominated cultures. There are potential economic effects, such as the extent to which women enter the work-force, and how that might impact cultural dynamics and stability. What, exactly, is the cost of female autonomy in terms of social cohesion and conflict? on personal happiness and political stability? While sensible answers may rely on input from the sciences, they are not scientific questions, they are trans-scientific. Similarly, in the context of evolutionary processes, every adaptation involves an inherent cost-benefit calculation, a design trade-off, opportunity's gained and curtailed, with the final decision based on reproductive success (as we will see).[48] There are no perfect solutions, just compromises that work more or less well. When we think about biological systems and processes, we need to keep this trade-off / cost-benefit calculation in mind.

Questions to answer:

4. A news story reports that spirit forces influence the weather. Produce a set of questions whose answers would enable you to decide whether the report was scientifically plausible.
5. If "science" concludes that free will is an illusion, would you be wise or silly to start behaving like a machine?
6. How would you describe the major differences between scientific thinking in physics and biology?

Questions to ponder

- Is attaining "truth" and developing a theory of everything the goal of science?
- How should we, as a society, deal with the tentative nature of scientific knowledge?
- What distinguishes scientific from trans-scientific conclusions?
- Why are predictions involving the complex phenotype rarely accurate?
- Given that costs and benefits are rarely "fairly apportioned", is it reasonable to think that science can answer social questions?

[48] Weinstein. Evolutionary trade-offs as a central organizing principle in biology

Chapter 2: Life and its origins

In which we consider what biology is all about, namely the study of organisms, their diversity, and how they work. We discover that organisms are built of one or more, sometimes many (millions to billions) cells. Social processes are involved in multicellular organisms and when single-celled organisms act in a coordinated manner. We consider plausible models for the origins of organisms, their basic properties and relationships to one another.

Biology is the science of organisms, how they function, behave, interact, and vary genetically from one another. How they adapt and, as populations, evolve over time. As we will see, organisms are discrete, highly organized, bounded but open, non-equilibrium, physicochemical systems. Now that is a lot of words, so the question is what do they all mean? How is a rock different from a mushroom that looks like a rock? What is genetic variation and how does it influence the properties and behaviors of an organism? What exactly is a bounded, non-equilibrium system? The answers are not simple; they assume a working knowledge of core principles from physics and chemistry and experimental observations. For example, to understand what it means to be a “bounded, non-equilibrium system” you need an understanding of basic thermodynamics, a topic that we will address in some detail in Chapter 5. For the moment, when we talk about a non-equilibrium system, we mean a system that can do various forms of work. Of course we then need to define what we mean by work. For simplicity, we will start by defining work as some outcome that takes the input of energy to achieve. In the context of biological systems, work ranges from generating and maintaining molecular gradients and driving a range of unfavorable, that is energy-requiring reactions, such as the synthesis of biomolecules, including nucleic acids, proteins, lipids, and carbohydrates, required for growth, reproduction, movement, and so on.

We will focus on what is known as Gibbs free energy, which is energy available to make things happen, that is, to do work, work that including assembling unstable molecules. When a system is at equilibrium its free energy is 0, which means that no macroscopic (visible) or net changes are possible. While appearing static at the macroscopic level, at the molecular level there is constant movement and change because, at all temperatures above absolute zero, molecular systems have kinetic energy that manifests itself as movement and vibrations. Organisms maintain their non-equilibrium state by importing "free energy", in various forms (light, chemically unstable molecules) from the external world. Organisms are different from other non-equilibrium systems in that they contain information in a form that can be replicated and passed from parent to offspring. While other types of non-equilibrium systems occur – hurricanes and tornados are non-equilibrium systems – they differ from organisms in that they are transient. They arise *de novo*, they do not have “parents”, and when they dissipate they leave no offspring, no baby hurricanes or tornados. In contrast, each organism alive today arose from one or more pre-existing organisms, its parent(s), and organisms, with some exceptions, have the ability to produce offspring. As we will see, the available evidence indicates that each and every organism, past, present, and future, has, or will have, an uninterrupted history stretching back billions of years. This is a remarkable conclusion, given the all too obvious fragility of life, and makes organisms unique among physicochemical systems.

Biology is based on only a few over arching theories. One of these, the Cell Theory of Life, explains the historic continuity of organisms, while the Theory of Evolution by Natural Selection (and other processes), explains both the diversity of organisms and how populations of organisms change over time. Finally, the Physicochemical Theory of Life explains how it is that organisms can display their remarkable properties without violating the laws that govern all physical and chemical

systems.[49]

What is life, exactly?

Clearly, if we are going to talk about biology, organisms and cells and such, we have to define exactly what we mean by life. This raises a problem peculiar to biology as a science. We cannot define life generically because we know of only one type of life. While you might think that we know of many different types of life, from mushrooms to whales, from humans to the microbial communities growing on the surfaces of your teeth (that is what dental plaque is, after all), we find that the closer we look the these different "types of life" the more we are force to accept the conclusion that they are all, in fact, versions of a single type of life. Based on their common chemistry, molecular composition, cellular structure, and the way that they encode, read, and use hereditary information in the form of molecules of deoxyribonucleic acid (DNA), all topics we will consider in depth as we go on, there is no reasonable doubt that all organisms are descended from a common ancestor, LUCA. We do not know whether this type of life is the only type of life possible or whether radically different forms of life exist elsewhere in the universe or even on Earth, in as yet to be recognized and discovered forms.

We cannot currently answer the question of whether the origin of life is a simple, likely, and predictable event given the conditions that existed on the early Earth when life first arose, or whether the origin and persistence of life is a rare and unlikely event. In the absence of empirical data, one can question whether scientists are acting scientifically, or more as lobbyists for their pet projects, when they talk about doing astrobiology or speculating on when and where we will discover alien life forms. That said, asking seemingly silly questions, provided that empirically-based answers can be generated, is a critical driver of scientific progress. Consider, for example, current searches for life on Earth, almost all of which are dependent upon what we know about life on Earth. Specifically, most of the methods used rely on the fact that all known organisms use DNA to encode their genetic information. If they exist, these methods would not be expected to recognize dramatically different types of life. They would not detect organisms that used a non-DNA-based mechanism to encode genetic information. If we could generate living systems *de novo* in the laboratory we could develop a better understanding of what functions are necessary for life and better methods to look for possible "non-standard" organisms, methods that could reveal whether there are alternative forms of life right here on Earth.[50] That said, until someone manages to create or identify such non-standard forms of life, it seems reasonable to concentrate on the characteristics of life as we know them.

So, let us start again in trying to produce a useful description of what we mean by life. First, the core units of life are organisms, which are individual living objects. From a structural and thermodynamic perspective, each organism is a bounded, non-equilibrium system that persists over time and, from a practical point of view, can produce one or more copies of itself. Even though organisms are composed of one or more cells, it is the organism that is the basic unit of life. It is the organism that produces new organisms.[51] It is the organism that is the real thing. That said, some organisms live in closely integrated mutualistic relationships, and can be difficult to grow in isolation from one another.[52]

Why the requirement for, and emphasis on reproduction? The reasons are pragmatic. Assume that a non-reproducing form of life was possible. Any such system runs the risk of death, or perhaps

---

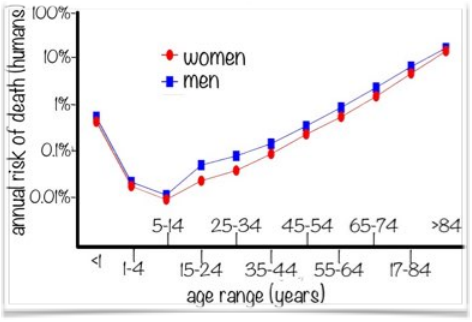
[49] Thinking about the conceptual foundations of the biological sciences

[50]The possibility of alternative microbial life on Earth Signatures of a shadow biosphere Life on Earth but not as we know it

[51] In Chapter 4, we will consider how multicellular and social organisms come to be.

[52] Cultured Asgard archaea shed light on eukaryogenesis by Lopez-Garcıa & Moreir 2020.

better put, accidental extinction. Over time, the probability of death for any individual will approach one – that is, certainty (→).[53] In contrast, a system that can reproduce makes multiple copies of itself and so minimizes, although by no means eliminates, the chance of accidental extinction (that is, the death of all of their descendants). We see the value of this strategy when we consider the history of life. Even though there have been a number of mass extinction events over the course of life's history, descendants of their ancestor (LUCA), an organism that lived billions of years ago, continue to survive and flourish.[54]



Now consider, what does the open nature of biological systems mean? Basically, organisms need to be able to import, in a controlled manner, energy and matter from outside of themselves and to export waste products into their environment.[55] This implies that there is a distinct boundary between the organism and the rest of the world. All organisms have such a barrier (boundary) layer, as we will see. The basic barrier layer of organisms appears to be a homologous structure–that is, it was present in and inherited from their common ancestor. The importation of energy, specifically energy that can be used to drive various cellular processes, is what enables the organism to maintain its non-equilibrium state and its dynamic structure, and to grow and reproduce. The boundary must be able to retain the valuable molecules generated, while at the same time allow waste products to leave. This ability to selectively import matter and export waste enables the organism to grow and to reproduce. While we assume that you have at least a basic understanding of the laws of thermodynamics, we will review the central ideas in Chapter 5.

We find evidence of the non-equilibrium nature of organisms most obviously in their ability of move, but it is important for all aspects of the living state. In particular, organisms use energy captured from their environment to drive a wide range of thermodynamically unfavorable chemical reactions. These unfavorable reactions are driven by coupling them to thermodynamically favorable reactions. An organism that reaches thermodynamic equilibrium is dead.

There are examples of non-living, non-equilibrium systems that can "self-organize" and that can appear *de novo*. Hurricanes and tornados form spontaneously and then disperse. Their formation is dependent upon energy from their environment, energy that is then released back into the environment, a process associated with an increase in the overall entropy of the Universe. These non-living systems differ from organisms in that they do not produce offspring - they are the result of specific atmospheric conditions. They are individual entities, unrelated to one another; they do not and cannot evolve. Tornados and hurricanes that formed billions or millions of years ago would, if we could observe them, be similar to those that form today. Since we understand, more or less, the conditions that produce tornados and hurricanes, we can predict, with some degree of reliability, the conditions that lead to their appearance and how they will behave once formed. In contrast, organisms present in the past were different from those that are alive today. The further into the past we go, the more different they appear. Some ancient organisms became extinct, some gave rise to the ancestors of current organisms. In contrast, each tornado or hurricane originates anew, they are not derived from parental storms.

---

[53] Image modified from "risk of death" graph: http://www.medicine.ox.ac.uk/bandolier/booth/Risk/dyingage.html

[54] Mass extinction events

[55] Cells organize themselves by exporting entropy. So be careful about claims of "zero-waste", they are impossible according to the laws of thermodynamics.

Questions to answer:

7. How might you decide whether a particular object (or system) is alive or not?
8. Using the graph on risk of death as a function of age in humans, provide a plausible explanation for the shape of the graph; what factors influence the various regions of the curve?
9. How does population size influence the risk of extinction?

Questions to ponder:

- Explain whether the points in the risk of death graph (↑) should be connected or whether a smooth "best fit" curve would be a more accurate description of the system.

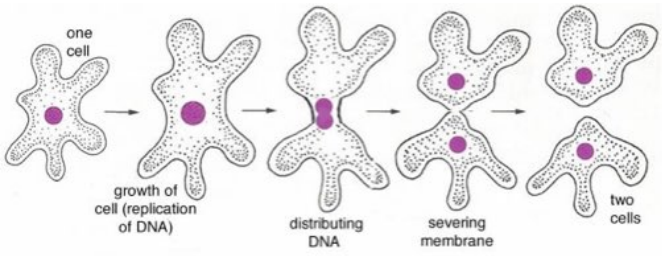
The Cell Theory and the continuity of life

Toward the end of the 1800's, observations using microscopes revealed that all organisms examined contained structurally similar units, termed "cells." Based on such observations, a rather sweeping conclusion, the Cell Theory, was formulated by naturalists. The Cell Theory has two distinct parts. The first is the prediction that every organism is composed of one or more, and in some cases millions to billions, of cells together with their products, such as bone, hair, scales, and slime, produced by cells. The cells that the Cell Theory postulates are membrane-bounded, open, non-equilibrium physicochemical systems, a definition much like that for life itself. Over the course of many observations (up to the present day) there has been no evidence that modern cells can be formed from non-cellular materials. Therefore the second part of the Cell Theory is that cells arise only from pre-existing cells. The implication is that organisms, and the cells that they are composed of, arise in this way and no other. That said, the Cell Theory says nothing as to how the first cell originated or how life on Earth originated.

We now know, and will consider in greater detail as we proceed, that in addition to their basic non-equilibrium nature, cells also contain hereditary information stored in a physical and relatively stable form, namely molecules of double-stranded deoxyribonucleic acid (DNA). Based on a large body of data, the Cell Theory implies that all organisms currently in existence, and the cells that compose them, are related through an unbroken series of DNA replication and cell division (reproductive) events that stretch back in time. Other studies, based on the information present in DNA molecules, as well as careful comparisons of how cells are constructed at the molecular level, suggests that there was a single common ancestor (LUCA) for all life and that this organism lived between ~3.5 to ~3.8 billion years ago. This is a remarkable conclusion, given the fragility of life. It implies that each cell in every currently living organism, including all of the cells that make you up, have an uninterrupted multibillion year old history.

The earliest events in the origin of life, exactly how the first cells were formed and what they looked like, are unknown and essentially unknowable, although there is more than enough speculation about them to go around. Our confusion arises in large measure from the fact that the available evidence indicates that all organisms that have ever lived on Earth share a single common ancestor, and that that ancestor, likely to be a singled-cell organism, was quite complex. Evidence for what living or pre-living systems came before LUCA is lost. We will discuss how we come to these conclusions, and their implications, later on in this chapter.

One point to keep in mind is that the "birth" of a new cell is a continuous process by which one cell becomes two. Each cell is defined, in part, by the presence of a distinct surface barrier, known as the cell or plasma membrane. The new cell is formed when that original membrane pinches off to form two distinct cells (→). The important point here is that there is no discontinuity, the new cell does not spring into existence but rather emerges from the preexisting cell. This continuity, from cell to cell, extends back



in time for billions of years. We often define the start of a new life with the completion of cell division, or in the case of sexually reproducing organisms, including humans, the fusion of an egg cell and a sperm cell. But again there is no discontinuity, both egg cell and sperm cell are derived from other cells and when they fuse, the result is a new hybrid cell. In the modern world, all cells, and the organisms they form, emerge from pre-existing cells and inherit from those cells both their cellular structure, the basis for the non-equilibrium living system, and their genetic material, their DNA. When we talk about cellular or organismic structures, their topologies, we are talking about information present in the living structure, information that is lost if the cell/organism dies. The information stored in DNA molecules, known as an organism's genotype, is more stable than the organism itself; it can survive the death of the organism, at least for a while. In fact, information-containing DNA molecules can move between unrelated cells or from the environment into a cell, a process known as horizontal gene transfer (which we will consider in detail later on). In fact DNA is being explored as a high-density, high-stability data storage system, outside of organisms.[56] That said, DNA means nothing outside of a system that can interpret the information stored within it.

The organization of organisms

Some organisms consist of a single cell, while others are composed of many cells, often many distinct types of cells. Cells vary in a number of ways and can be highly specialized, particularly within the context of multicellular organisms, yet all cells appears related to one another, sharing many molecular and structural details. So why do we consider the organism rather than the cell to be the basic unit of life? The distinction may seem trivial or arbitrary, but it is not. It is a matter of reality versus abstraction. It is organisms, whether single- or multi-cellular, that produce new organisms. As we will discuss in some detail when we consider the origins of multicellular organisms, a cell within a multicellular organism normally cannot survive outside the organism nor can it produce a new organism – it depends upon cooperation with the other cells of the organism. In fact, each multicellular organism is an example of a cooperative, highly integrated social system.

In a typical multicellular organism most cells have given up their ability to reproduce a new organism; their future depends upon the reproductive success of the organism of which they are a part. It is the organism's success in generating new organisms that underlies evolution's selective mechanisms. Within the organism, the cells that give rise to the next generation of organisms are known as germ cells, those that do not, that is, the cells that die when the organism dies, are known as somatic cells.[57] All organisms in the modern world and, apparently for the last ~3.5-3.8 billion years, arose from a pre-existing organism or, in the case of sexually reproducing organisms, from the cooperation of two organisms, an example of social evolution that we will consider in greater detail in Chapter 4. We will also see that breakdowns in such social systems can lead to the death of the organism or the disruption of the social system. Cancer is the most obvious example of an anti-social cellular behavior. In the short term, cancerous behavior maybe "rewarded" (more copies of the cancer cell are produced) but ultimately it leads to the death of the organism and the extinction of the cancer cells.[58] This is because evolutionary mechanisms are not driven by long term outcomes, but only immediate cost-benefit "calculations", revealed in terms of reproductive success.

Spontaneous generation and the origin of life

The ubiquity of organisms raises obvious questions: how did life start and what led to all these different types of organisms? At one point, people believed that these two questions had a single

---

[56] A DNA-Based Archival Storage System

[57] If we use words that we do not define and that you do not understand, look them up or ask your instructor!

[58] Cancer cells as sociopaths: cancer's cheating ways Recently the situation has gotten more complex with the recognition of transmissible cancers and http://www.ncbi.nlm.nih.gov/pubmed/19956175

answer, but we now recognize that they are really two distinct questions and their answers involve distinct mechanisms. An early view, held by those who thought about such things, was that supernatural processes were necessary to produce life in general and human beings in particular. The articulation of the Cell Theory and the Theory of Evolution by Natural Selection, which we will discuss in the next chapter, together with an accumulation of molecular level data enables us to conclude, quite persuasively, that life had a single successful origin and that various natural processes generated the diversity of life.

But how did life itself originate? It was once widely accepted that various types of organisms, such as flies, frogs, and even mice, could arise spontaneously, from non-living matter.[59] Flies, for example, were thought to appear from rotting flesh and mice from wheat. If true, on-going spontaneous generation would have profound implications for our understanding of biological systems. For example, if spontaneous generation based on natural processes was common, there must be a rather simple process at work, a process that presumably can produce remarkably complex outcomes. In contrast, all bets are off if the process is supernatural. If each organism arose independently, we might expect that, at the molecular level, details of each would be unique, since they presumably arose independently from different stuff and under different conditions and for different purposes compared to other organisms. We know, however, that this does not appear to be the case; all organisms use similar molecular mechanisms, are composed of structurally similar cells, and appear to be descended from a single common ancestor.

A key event in the conceptual development of modern biology was the publication in 1668 of Francesco Redi's (1626-1697) paper "Experiments on the Generation of Insects". His hypothesis (informed guess) was that spontaneous generation did not occur.[60] He thought that the organisms that appeared had developed from "seeds" deposited by adults, an idea that led to a number of predictions. One was that if adult flies were kept away from rotting meat maggots, the larval form of flies, would not appear no matter how long one waited. Similarly, the type of organism that appeared would depend not on the type of rotting meat, but rather on the type of adult fly that had access to the meat. To test his hypothesis Redi set up two sets of flasks both of which contained meat. One set of flasks was exposed directly to the air and so to flies, the other was sealed with paper or cloth. Maggots appeared only in the flasks open to the air. Redi concluded that organisms as complex as insects, and too large to pass through the cloth, could arise only from other insects, or rather eggs laid by those insects – that life was continuous, that is, life came from life.

He who experiments increases knowledge. He who merely speculates piles error upon error.
- Arabic epigraph quoted by Francisco Redi.

The invention of the light microscope, and its use to look at biological materials, by Antony van Leeuwenhoek (1632-1723) and Robert Hooke (1635-1703) led to the discovery of a completely new and unexpected world of organisms, known as microbes or microscopic organisms. We now know these as the bacteria, archaea, and a range of unicellular eukaryotes.[61] Although it was relatively easy to generate compelling evidence that macroscopic (that is, big) organisms, such as flies, mice, and people could not arise spontaneously, it seemed plausible that microscopic, and presumably much simpler, organisms could form spontaneously.

The discovery of microbes led a number of scientists to explore their origin and reproduction. Lazzaro Spallazani (1729-1799) showed that after a broth was boiled it remained sterile, that is, without life, as long as it was isolated from contact with fresh air. He concluded that microbes, like larger organisms, could not arise spontaneously but were descended from other microbes, many of

---

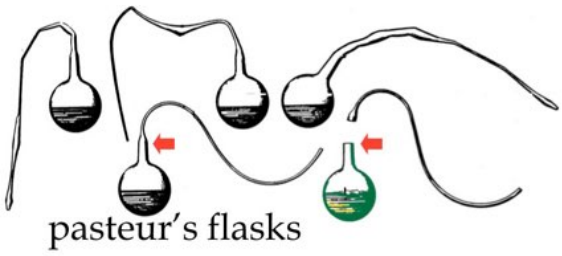
[59] Farley. The spontaneous generation controversy (1700-1860): The origin of parasitic worms. and The spontaneous generation controversy (1859-1880): British and German reactions to the problem of abiogenesis.

[60] see Richard Feynman's description of the role of guessing in the scientific process

[61] see the wikipedia article on protists

which were floating in the air. Think about possible criticisms to this experiment – perhaps you can come up with ones that we do not mention!

One criticism was that perhaps boiling the broth destroyed one or more key components that were necessary for the spontaneous formation of life. Alternatively, perhaps fresh air was the "vital" ingredient. In either case, boiling and isolation would have produced an artifact that obscured rather than revealed the true process. In 1862 (after Charles Darwin had published On the Origin of Species), Louis Pasteur (1822-1895) carried out a particularly convincing set of experiments to address both of these concerns. He sterilized broths by boiling them in special "swan-necked" flasks. What was unique about his experimental design was the shape of the flask neck; it allowed air but not air-borne microorganisms to reach the broth. Microbes in the air were trapped in the bended region of the flask's neck (←). This design enabled Pasteur to address a criticism of previous experiments, namely that access to air was necessary for spontaneous generation to occur. He found that the liquid, even with access to air, remained sterile for months. However, when the neck of the flask (indicated by the red arrows) was broken the broth was quickly overrun with microbial growth. He interpreted this observation to indicate that air, by itself, was not necessary for spontaneous generation, but rather was normally contaminated by microbes. On the other hand, the fact that the broth could support microbial growth after the neck was broken served as what is known as a "positive control" experiment; it indicated that the heating of the broth had not destroyed some vital element needed to support growth. We carry out positive control experiments to test whether specific assumptions are correct. For example, if we are using a drug in a study, we need to establish (rather than take someone's word for it) that the sample of the drug we are using is active. In Pasteur's experiment, if the boiled broth could not support growth (after the flask neck was broken) we would not expect it to support spontaneous generation, and so the experiment would be meaningless. We will return to the description of a "negative control" experiment later.[62]



Of course, not all, in fact, probably not any experiment is perfect, nor does it have to be for science to work. For example, how would one argue against the objection that the process of spontaneous generation normally takes tens to thousands, or millions, of years to occur? If true, this objection would invalidate Pasteur's conclusions. Clearly an experiment to address that particular objection has its own practical issues. Nevertheless, the results of various experiments on spontaneous generation have led to the conclusion that neither microscopic nor macroscopic organisms can arise spontaneously in the modern world. The problem, at least in this form, became uninteresting to working scientists.

So what explains the absence of spontaneous generation in the modern world, or in a world in which life (organisms) already exist? Consider the fact that living systems involve complex chemical reaction networks. In the modern world, there are many organisms around, essentially everywhere, and these organisms are actively eating complex molecules to maintain their non-equilibrium (energy requiring) state, to grow and reproduce. Given the tendency of organisms to eat one another, one might argue (as Darwin did →) that once organisms had appeared in a particular environment they would suppress subsequent events – they would have eaten the molecules needed for spontaneous generation

> *It is often said that all the conditions for the first production of living organisms are now present. But if (and oh! what a big if!) we could conceive in some warm little pond, with all sorts of ammonia and phosphoric salts, light, heat, electricity, etc. present, that a proteine compound was formed, ready to undergo still more complex changes, at the present day such matter would be instantly devoured or absorbed, which would not have been the case before living creatures were formed. - Charles Darwin (1887).*

[62] Wikipedia on control experiments and observations

to occur. But, as we will see, evolutionary processes have led to the presence of organisms essentially everywhere on Earth that life can survive – there are basically no welcoming and sterile, that is, life-less places left within the modern world.

Here we see the importance of history. According to the current scientific view, life could arise *de novo* only in the absence of life. We can put some limits on the minimum time it could take from geological data using the time from when the Earth's surface solidified from its early molten state to the first fossil evidence for life, about 100 to 500 million years. Once life had arisen conditions had changed. The presence of life, that is organisms, would be expected to suppress new spontaneous generation events. Once organisms were present, only their descendants could survive. In such a system, history matters.

The death of vitalism

Naturalists originally thought that life itself was a type of supernatural process, too complex to obey or be understood through the laws of chemistry and physics.[63] In this vitalistic view, organisms were thought to obey different laws from those acting in the non-living world. For example, it was assumed that molecules found only in living organisms, known as organic molecules, could not be synthesized outside of an organism; they had to be made by a living organism. In 1828, Friedrich Wöhler (1800–1882) challenged this view by synthesizing urea in the laboratory. Urea ($O=C(NH_2)_2$) is a simple organic molecule found in the waste derived from living organisms. Urine contains lots of urea. Wöhler's *in vitro* or "in glass", as opposed to *in vivo* or "in life", synthesis of urea was simple. While attempting to synthesize ammonium cyanate (NH_4NCO), he mixed the inorganic compounds ammonium chloride (NH_4Cl) and silver cyanate (AgNCO). Analysis of the products of this reaction revealed the presence of urea. What actually happened was this reaction:

$$AgNCO + NH_4Cl \rightarrow NH_4NCO + AgCl \rightarrow O=C(NH_2)_2 + AgCl.$$

Please do not memorize this reaction! What is important here is to recognize that this is a chemical reaction between two compounds that are not derived from living systems. The point is that the urea synthesized through an "inorganic" reaction is identical to the "natural" urea found in urine.

While simple, Wöhler's *in vitro* synthesis of urea had a profound impact on the way scientists viewed so called organic processes. It suggested that there was nothing supernatural involved in the way organisms worked, the synthesis of urea was a standard chemical process. Based on this and similar observations on the *in vitro* synthesis of other, more complex organic compounds, the scientific consensus is that that all molecules found within cells and organisms can be synthesized in the laboratory using appropriate chemical procedures. This is not to say that all such molecules have been synthesized *in vitro*; it means that we assume that given enough effort (time and resources) they could be. Organic chemistry has been transformed from the study of molecules found in organisms to the study of molecules containing carbon atoms. A huge amount of time and money is devoted to the industrial syntheses of a broad range of organic molecules that are used for purposes as diverse as pharmaceuticals to plastics.

Questions to answer:

10. Why did the discovery of bacteria reopen the debate on spontaneous generation?
11. In Pasteur's experiment would you expect to see microbial growth in the bent loop of the flask? Explain your thinking.
12. What does the result of a positive control experiment tell you?
13. Explain why Wöhler's synthesis of urea transformed thinking about organic molecules.

Questions to ponder:

- Is the assumption of spontaneous generation inherently unscientific? Explain your reasoning.
- Can you imagine an observation that would lead scientists to reject the naturalistic perspective?

---

[63] In a sense this is true since many physicists at least do not seem to understand biology.

- What types of evidence would support the view that the origin of life (or consciousness) requires supernatural intervention?

Thinking about life's origins

There are at least three possible approaches to the study of life's origins. A religious (i.e., non-scientific) approach would likely postulate that life was created by a supernatural being or process. Different religious traditions differ as to the details of this event, but since the process is supernatural it cannot, by definition, be studied scientifically. Nevertheless, intelligent design creationists often claim that we can identify those aspects of life that could not possibly have been produced by natural processes, by which they mean various evolutionary and molecular mechanisms. We will discuss these processes throughout the book, and more specifically in the next chapter. It is important to consider whether these claims would, if true, force us to abandon a scientific approach to the world around us in general, and the origin and evolution of life in particular. Given the previously noted interconnectedness of the sciences, one might well ask whether a supernatural (intelligent design) biology would not also call into question the validity of all scientific disciplines. For example the dating of fossils is based on geological and astrophysical (cosmological) evidence for the age of the Earth and the Universe, which themselves are based on physical and chemical observations and principles. A truly non-scientific biology would be incompatible with a scientific physics and chemistry. The lesson of history, however, is different. Predictions as to what is beyond the ability of science to explain have routinely been found to be wrong, often only a few years after such predictions were made! This speaks to the power of science and science-based technologies. For example, would an intelligent design creationist be tempted to synthesize human proteins in bacteria or plants, something now done routinely to make a range of drugs, such as insulin?[64] Would they predict that genetic modifications could make it possible to transplant pig hearts (and other organs) into the people?[65]

An alternative explanation for the appearance of life on Earth, termed panspermia, assumes that advanced aliens brought (or left) life on Earth. Perhaps we owe our origins to casually discarded litter from these alien visitors. Unfortunately, the principles of general relativity, one of the best confirmed of all scientific theories, limit the speed of travel. Given the size of the Universe, travelers from beyond the solar system seem highly unlikely. More to the point, panspermia does not resolve the question of how life began. Our alien visitors must have come from somewhere and panspermia does not explain their origin. Given our current models for the history of the Universe, understanding the origin of alien life is really no simpler than understanding the origin of life on Earth. On the other hand, if life is discovered on other planets or the moons in our solar system, its structural and molecular details would be extremely informative – it would make "astrobiology" a real scientific discipline.[66]

Experimental studies on the origins of life

One strategy to understanding how life might have arisen naturally involves experiments to generate plausible precursors of living systems. The studies carried out by Stanley Miller (1930-2007) and Harold Urey (1893-1981) were an early and influential example of this approach.[67] These scientists made an educated, although now apparently incorrect, guess as to the composition of Earth's early atmosphere. They assumed the presence of oceans and lightning. They set up an

---

[64] Making human insulin in bacteria

[65] New life for pig-to-human transplants

[66] Top 5 Bets for Extraterrestrial Life in the Solar System

[67] The Miller-Urey experiment & wikipedia: http://en.wikipedia.org/wiki/Miller–Urey_experiment

apparatus to mimic these conditions and then passed electrical sparks through their experimental atmosphere. After a few days they found that a complex mix of compounds had formed. Included in this mix were many of the amino acids found in modern organisms, as well as lots of other organic molecules. Similar experiments have been repeated with other combinations of starting compounds, more likely to represent the environment of the early Earth, with similar results: various biologically important organic molecules accumulate rapidly.[68] Quite complex organic molecules have been detected in interstellar dust clouds, and certain types of meteorites have been found to contain a number of organic molecules. Similarly, the chemistry occurring in deep sea hydrothermal vents can produce complex mixtures of biomolecules abiogenically.[69] Around 4 billion years ago, a time known as the period of the heavy bombardment, meteorite impacts with the Earth could have supplied substantial amounts of organic molecules.[70] It appears likely that early Earth was rich in organic molecules, which are, remember, carbon containing rather than life-derived molecules, the building blocks of life.

Given that the potential building blocks for life were present, the question becomes what set of conditions were necessary and what steps led to the formation of the first living systems? Assuming that these early systems were relatively simple compared to modern organisms, or the precursors to the common ancestor of terrestrial life, we hypothesize that the earliest proto-biotic systems were molecular communities of chemical reactions isolated in some way from the rest of the outside world. This isolation or selective boundary was necessary to keep the system from dissolving away (dissipating). One possible model is that such systems were originally tightly associated with the surface of specific minerals and that these mineral surfaces served as catalysts, speeding up important reactions. We will return to the role of catalysts in biological systems later on. Over time, these pre-living systems acquired more sophisticated boundary structures (membranes) and were able to exist free of the mineral surface, perhaps taking small pieces of the mineral with them.[71]

The generation of an isolated but open system, something we might term a protocell, was a critical step in the origin of life. Such an isolated system has properties that are likely to have facilitated the further development of life. For example, because of the membrane boundary, changes that occur within one such structure will not be shared with neighboring systems. Rather, they would accumulate in, and favor the survival of, one system over its neighbors. Such systems could also reproduce in a crude way by mechanical fragmentation. For example. If changes within one such system improved its stability, its ability to accumulate resources, or its ability to survive, grow, and reproduce, that system, and its progeny, would be likely to become more common. As these changes accumulate and are passed from parent to offspring, the population of organisms will inevitably evolve, as we will see in detail in the next chapter.

As in living systems today, the earliest steps in the formation of the first organisms required a source of energy to maintain the non-equilibrium living (or pre-living) state. There are really only two choices for the source of this energy, light (electromagnetic radiation from the sun) or thermodynamically unstable molecules present in the environment. There have been a number of plausible scenarios, based on various observations, for the steps leading to life. For example, a recent study based on the analysis of the genes, and the proteins that they encode, found in modern organisms, suggests that the last universal common ancestor (LUCA) arose in association with hydrothermal vents and derived energy from thermodynamically favorable chemical reactions.[72] But

---

[68] A reassessment of prebiotic organic synthesis in neutral planetary atmospheres:

[69] The last universal common ancestor between ancient Earth chemistry and the onset of genetics

[70] A time-line of life's evolution: http://exploringorigins.org/timeline.html

[71] Mineral Surfaces, Geochemical Complexities, and the Origins of Life

[72] Meet LUCA, the Ancestor of All Living Things:

whether this reflects LUCA or an ancestor of LUCA that became adapted to living in association with hydrothermal vents is difficult, and perhaps impossible to resolve unambiguously, particularly since LUCA lived ~3.4-3.8 billion years ago and cannot be studied directly.

Mapping the history of life on earth

Assuming, as seems scientifically likely, that life arose spontaneously, we can look at what we know from the fossil record to better understand the diversification of life and life's impact on the Earth. This is probably best done by starting with what we know about where the Universe and Earth came from. The current scientific model for the origin of the universe is known as the "Big Bang", the "primeval atom", or the "cosmic egg" is based on an idea originally proposed by the priest, physicist and astronomer Georges Lemaître (1894-1966).[73] The Big Bang model arose from efforts to answer the question of whether the fuzzy nebulae (patches of light in the night sky) were located within or outside of our galaxy. This required some way to determine how far these nebulae were from Earth. Edwin Hubble (1889-1953) and his co-workers were the first to provide compelling evidence that nebulae were in fact galaxies in their own right, each very much like our own Milky Way and that each is composed of many billions of stars. This was a surprising result. It made Earth, sitting on the edge of one (the Milky Way) among many, many galaxies seem even less important – a change in cosmological perspective similar to that associated with the idea that the Sun, rather than the Earth, was the center of the solar system and the Universe.

To measure the movement of galaxies with respect to the Earth, Hubble and colleagues combined two types of observations. The first of these allowed them to estimate the distance from the Earth to various galaxies. The second measured the Doppler shift of the light from stars within distant galaxies. The Doppler shift is the effect of an object's velocity, relative to an observer, on the wavelength of sound or light it emits. For an object moving toward an observer, the wavelength of emitted light will be shortened, that is, shifted toward the blue end of the spectrum. The wavelength will be lengthened, that is, shifted to the red end of the spectrum, when moving away from the observer. Based on the observed Doppler shifts of light coming from stars in galaxies and the observation that the further a galaxy appears to be from Earth, the greater that shift is toward the red, Hubble concluded that galaxies, outside of our local group, were all moving away from one another. Running time backward, he concluded that at one point in the past, all of the matter and energy in the Universe must have been concentrated in a single point.[74] A prediction of this Big Bang model is that the Universe is ~13.8 +/- 0.2 billion (10^9) years old. This is a length of time well beyond human comprehension; it is sometimes referred to as deep time – you can get some perspective on deep time using the "Here is Today" website (http://hereistoday.com). Other types of data have been used to arrive at an estimated age of the Earth and the other planets in the solar system as ~4.5 to 5 x 10^9 years.

After the Earth formed, it was bombarded by extraterrestrial materials, including comets and asteroids. This bombardment began to subside around ~3.9 billion years ago and reached its current level by ~3.5 billion years ago.[75] It is not clear whether life arose multiple times and was repeatedly destroyed during the early history of the Earth (4.5 to 3.6 billion years ago) or if the origin of life was a one-time event, taking hundreds of millions of years before it succeeded, after which it managed to survive and expand to the present day.

---

[73] Georges Lemaître: http://www.physicsoftheuniverse.com/scientists_lemaitre.html

[74] The origin of the universe and the primeval atom

[75] The violent environment of the origin of life

Fossil evidence for the history of life on earth

The earliest period in Earth's history is known as the Hadean, after Hades, the Greek god of the dead. The Hadean is defined as the period from the origin of the Earth up to the first appearance of life. Fossils provide our only direct evidence for when life appeared on Earth. They are found in sedimentary rock, which is rock formed when fine particles of mud, sand, or dust entomb an organism before it can be eaten by other organisms. Hunters of fossils (paleontologists) do not search for fossils randomly; they use geological information to identify outcroppings of sedimentary rocks of the specific age they are interested in.[76]

Early in the history of geology, before Charles Darwin and Alfred Wallace proposed the modern theory of evolution, geologists recognized that fossils of specific types were associated with rocks of specific ages. This correlation was so robust that rocks could be accurately dated based on the types of fossils they contained. At the same time, particularly in a world that contains young earth creationists who claim that Earth was formed less than ~10,000 years ago, it is worth remembering both the interconnectedness of the sciences and that geologists do not rely solely on fossils to date rocks, in part because many types of rocks do not contain fossils. The non-fossil approach to dating rocks is based on the physics of isotope stability and the chemistry of atomic interactions. It uses the radioactive decay of elements with isotopes with long half-lives, such as ^{235}Ur (uranium) which decays into ^{207}Pb (lead) with a half-life of ~704 million years and into ^{238}Ur which decays into ^{206}Pb with a half-life of ~4.47 billion years. Since these two Pb isotopes appear to be formed exclusively through the decay of uranium isotopes, the ratios of uranium and lead isotopes can be used to estimate the age of a rock, assuming that it originally contained only uranium, and no lead. In order to use isotope abundance to accurately date rocks, it is critical that all of the atoms in a mineral measured originated there and stayed there, that is, that none were washed into or out of the rock. Since uranium and lead have different chemical properties, this can be difficult to establish in some types of minerals. That said, with care, and using rocks that contain chemically inert minerals, like zircons, the isotope ratio method can be used to measure the age of rocks to an accuracy of ~1% or better. Such age estimates, together with other types of evidence, support James Hutton's (1726-1797) dictum that the Earth is ancient, with "no vestige of a beginning, no prospect of an end."[77] We know now, however, that this statement is not true; while very old, Earth had a beginning, it coalesced around ~5 billion years ago, and it will disappear when the sun expands and engulfs it in about ~5.5 billion years from now.[78]

Now, back to fossils. There are many types of fossils. Chemical fossils are molecules that, as far as we know, are naturally produced only through biological processes.[79] Their presence in ancient rock implies that living organisms were present at the time the rock formed. Chemical fossils first appear in rocks that are between ~3.8 to ~3.5 x 10^9 years old. What makes chemical fossils problematic is that there may be non-biological but currently undiscovered or unrecognized mechanisms that could have produced these molecules, so we should be cautious in our conclusions.

Moving from the molecular to the physical, there are what are known as trace fossils. These can be subtle or obvious. Organisms can settle on mud or sand and leave impressions. Burrowing and slithering animals make tunnels or disrupt surface layers. Leaves and immotile organisms can leave impressions. Walking animals can leave footprints in sand, mud, or ash. How does this occur? If the ground is covered, compressed, and converted to rock, these various types of impressions can

---

[76] A process described in some detail by Neil Shubin in The Evolution of Limbs from Fins

[77] Changing Views of the History of the Earth

[78] How the sun will die

[79] Although as Wohler pointed out, they can be generated in the laboratory.

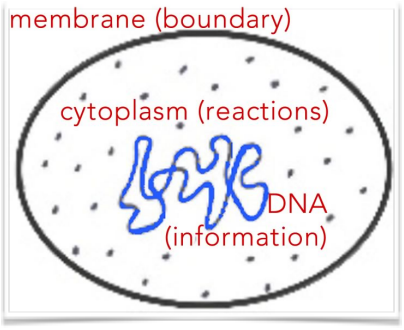
become fossils. Later erosion can then reveal these fossils. For example, if you live near Morrison, Colorado, you can visit the rock outcrop known as Dinosaur Ridge and see trace fossil dinosaur footprints; there may be similar examples near where you live.

We can learn a lot from trace fossils, they can reveal the general shape of an organism, its ability to move, or to move in a particular way. To move, an organism must have some kind of muscle or alternative mobility system and probably some kind of nervous system that can integrate internal and external information and produce coordinated movements. Movement also suggests that the organisms that made the trace had something like a head and a tail. Tunneling organisms are likely to have had a mouth to ingest sediment, much like today's earthworms - they were predators, eating the microbes they found in mud.

In addition to trace fossils, there are also the type of fossils that most people think about, which are known as structural fossils, namely the mineralized remains of the hard parts of organisms such as teeth, scales, shells, or bones. As organisms developed hard parts fossilization, particularly of organisms living in environments where they could be buried within sediment before being dismembered and destroyed by predators or microbes, became more likely.

Unfortunately for us (as scientists), many and perhaps most types of organisms leave no trace when they die. In part this may be because they live in places where fossilization is rare or unlikely. Animals that live in woodlands, for example, rarely leave fossils. The absence of fossils for a particular type of organism does not imply that these types of organisms do not have a long history, rather it means that the conditions where they lived and died or their body structure is not conducive to fossilization. Many types of living organisms have no fossil record at all, even though, as we will see, there is molecular evidence that they arose tens to hundreds of millions of years ago.

Life's impact on the Earth

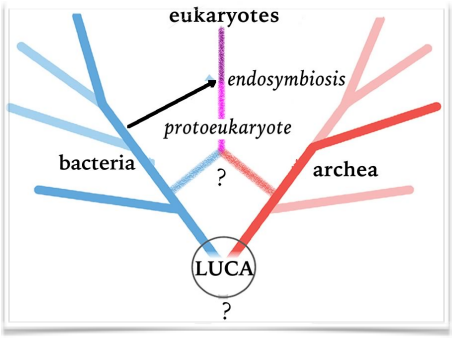


Based on fossil evidence, the current model for life on Earth is that for a period of ~2 x 10^9 (billion) years after the appearance of LUCA, the only forms of life on Earth were microscopic. Today, there are three families of organisms that we describe briefly here and in more detail later: the bacteria, the archaea, and the eukaryotes. While the exact nature of LUCA is unclear, it is likely that it was single celled and relatively simple in general organization (→) consisting of a boundary membrane, controlling the movement of molecules into and out of the cell, a cytoplasm, in which various biosynthetic reactions took place, and molecules of the genetic material, DNA, located within the cytoplasm. Both bacteria and archaea have this same basic type of cellular organization, they differ in a range of molecular details, although not in basic molecular mechanisms.[80] As we will discuss later, eukaryotes are more complex structurally; they contain internal membrane systems and their genetic material is located within a double membrane compartment (the nucleus) located within the cytoplasm. Movement between nuclear interior and cytoplasm is facilitated by molecular machines, known as nuclear pores. How the nucleus came to be remains (not surprisingly) unclear, but it is possible that the proto-eukaryote (that is, with a nucleus) arose through a fusion event that involved both bacterial and archaeal ancestors.[81] Alternatively, it might be directly descended from LUCA. The problem is that we do not have direct evidence as to the details of LUCA's structure, just inferences (informed guesses). It is clear, however, that the formation of eukaryotes involved a symbiotic event (discussed in Chapter 5) in which an α-proteobacterium (a type of bacteria) was engulfed, but not digested, by the proto-eukaryote. This "endogenous bacterium" became the

[80] see the Common Ancestor of Archaea and Eukarya

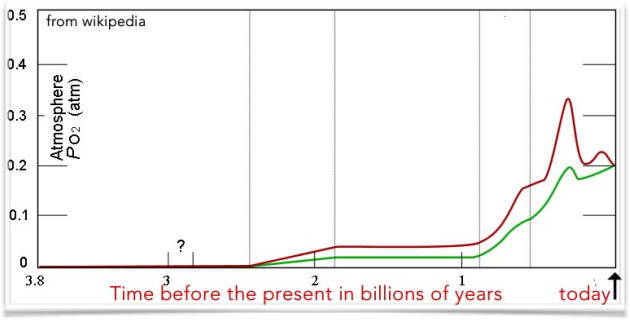
[81] Origin of eukaryotes & The common ancestor of archaea and eukarya was not an archaeon

eukaryotic mitochondrion. Essentially all eukaryotes (the protozoa, fungi, animals, and plants) have mitochondria, apparently descended from this event(→). Later in the history of life, a second endosymbiotic event occurred in which a mitochondria-containing eukaryote engulfed but did not digest a second type of bacteria, a photosynthetic cyanobacterium, leading to the algae and the plants.



While the earliest organisms probably used energy released in the course of chemical reactions to maintain their structural integrity and to grow, relatively soon bacterial-type organisms appeared that could capture the energy in light and use it to drive various thermodynamically unfavorable chemical reactions. A major class of such reactions involves combining CO_2 (carbon dioxide), H_2O (water), and other molecules to form carbohydrates (sugars) and biologically important molecules, such as lipids, proteins, and nucleic acids. At some point during the early history of life on Earth, organisms appeared that released molecular oxygen (O_2) as a waste product of light-driven reactions – a process known generically as oxygenic photosynthesis. These oxygen-releasing organisms became so numerous that they began to change Earth's surface chemistry - they represent the first life-driven ecological catastrophe (or opportunity, depending about your perspective).

The level of atmospheric O_2 represents a balance between its production, primarily by organisms carrying out oxygenic photosynthesis, and its breakdown through various chemical reactions. Early on as O_2 appeared, it reacted with iron to form deposits of water-insoluble Fe(III) oxide (Fe_2O_3) – that is, rust. This rust reaction removed large amounts of O_2 from the atmosphere, keeping levels of free O_2 low. The rusting of iron in the oceans is thought to be largely responsible for the massive banded iron deposits found around the world.[82] O_2 also reacts with organic matter, as in the burning of wood, so when large amounts of organic matter are buried before they can react with O_2, as occurs with the formation of coal, more O_2 accumulates in the atmosphere. Although O_2 was probably being generated and released earlier, by ~2 billion years ago, atmospheric O_2 had appeared in detectable amounts and by ~850 million years ago O_2 had risen to significant levels (→). Atmospheric O_2 levels have changed significantly since then, based on the relative rates of its synthesis and breakdown. Around ~300 million years ago, atmospheric O_2 levels reached ~35%, almost twice the current level. It has been suggested that these high levels of atmospheric O_2 made the evolution of giant insects possible.[83]



Although we tend to think of O_2 as a natural and benign substance, it is in fact highly reactive and potentially toxic; its production and accumulation posed serious challenges and unique opportunities to, organisms. As we will see later on O_2 can be "detoxified" through reactions that lead to the formation of water; this type of thermodynamically favorable reaction appears to have been co-opted for a wide range of biological purposes. For example, through coupled reactions O_2 can be used to capture the maximum amount of energy from the breakdown of complex molecules (food), leading to the generation of CO_2 and H_2O, both of which are stable.

Around the time that O_2 levels were first rising, that is ~10^9 years ago, the first trace fossil burrows appeared in the fossil record. These were likely to have been produced by simple worm-

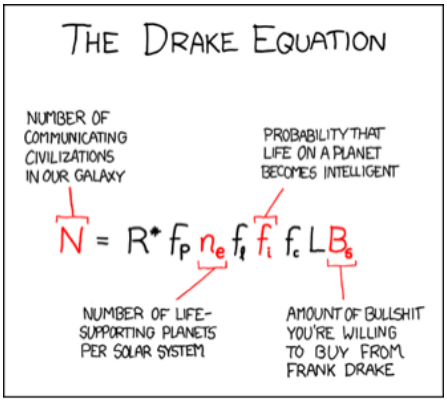
[82] Paleoecological Significance of the Banded Iron-Formation: http://econgeol.geoscienceworld.org/content/68/7/1135.abstract

[83] see Geological history of oxygen & Atmospheric oxygen and giant Paleozoic insects

like, macroscopic multicellular organisms, known as metazoans, that is, multi-cellular animals, capable of moving along and through the mud on the ocean floor. About ~0.6 x 10^9 years ago, new and more complex structural fossils (←) began to appear in the fossil record. The first of these to appear were the so-called Ediacaran organisms, named after the geological formation in which their fossils were first found.[84] Current hypotheses suggest they were immotile, like modern sponges but flatter; it remains unclear how or if they are related to later animals. Since the fossil record does not contain all organisms, we are left to speculate on what earlier metazoans looked like. By the beginning of the Cambrian age (~545 x 10^6 years ago), a wide variety of organisms had appeared within the fossil record, many clearly related to modern animals. Molecular level data suggest that their ancestors originated more than ~30 million years earlier. These Cambrian organisms show a range of body types. Most significantly, many were armored. Since building armor involves expending energy to synthesize these components, the presence of armor suggests the presence of predators, and a need for a defensive response.

Viruses: Before we leave this chapter you might well ask, have we forgotten viruses? Well, no - viruses are often a critical component of an ecosystem and an organism's susceptibility. resistance and response to viral infection can be an important evolutionary factor, but viruses are different from organisms in that they are non-metabolic. That means they do not carry out reactions and cannot replicate on their own, they replicate only within living cells. Basically they are not alive, so even though they are extremely important, we will discuss viruses only occasionally and in quite specific contexts. That said, the recent discovery of giant viruses, such as Mimivirus, suggests that something interesting is going on.[85] Given the recent COVID-19 pandemic and viral illnesses of plants and animals, a understanding of viral-host interactions is of vital scientific, social, and economic importance.

Questions to answer



14. In 1961 Frank Drake, a radio astronomer, proposed an equation to estimate the number of technologically sophisticated civilizations that can be expected to exist within the observable Universe (N).[86]

The equation is $N = R^* \times f_p \times n_e \times f_l \times f_i \times f_c \times L$ where:

R^* = The rate of formation of stars suitable for the development of intelligent life.

f_o = The fraction of those stars with planetary systems.

n_e = The number planets, per solar system, with an environment suitable for life.

f_l = The fraction of suitable planets on which life actually appears.

fi = The fraction of life-bearing planets on which intelligent life emerges.

f_c = The fraction of civilizations that develop a technology that releases detectable signs of their existence into space.

L = The length of time such civilizations release detectable signals into space (that is how long such a civilization persist until it destroys itself or is destroyed by natural disaster).

Identify those parts of the Drake equation that can and those that cannot be established (at present) empirically. Is the Drake equation scientific, or does it just look "sciency"? Explain your reasoning.

15. What factors would influence the probability that a particular type of organism will be fossilized?

16. What factors might drive the appearance of teeth, bones, shells, muscles, nervous systems, and eyes?

17. What factors, biological and geological, determine atmospheric O_2 levels?

---

[84] Ediacarian organomis

[85] http://www.giantvirus.org/intro.html

[86] The Drake equation: http://www.seti.org/drakeequation and cartoon: http://xkcd.com/384/

Questions to ponder

- Is the Drake equation scientific? What factors limit the scientific studies of origin of life?
- If we assume that spontaneous generation occurred in the distant past, why is it not occurring today? How could you tell if it were?

Chapter 3: Evolutionary mechanisms and the diversity of life

In which we consider the rather exuberant diversity of organisms and how they came to be. To understand these processes requires that we introduce core evolutionary mechanisms, both adaptive (natural, sexual, and social selection) and non-adaptive (drift and bottlenecks). As part of our discussion we consider the history of how people considered the diversity (and meaning) of life.

In medieval Europe there was a tradition of books known as bestiaries; these were illustrated catalogs of real and imagined organisms; often each particular organism was associated with a moral lesson. “Male lions were seen as worthy reflections of God the Father, for example, while the dragon was understood as a representative of Satan on earth.”[87] One can see these books as an early version of a natural theology, that is, an attempt to gain an understanding of the supernatural through the study of natural objects.[88] In this case, the presumption was that each type of organism was created for a particular purpose, and that often this purpose was to provide people with a moral lesson. This way of thinking grew more and more problematic as more and more different types of organisms were recognized, many of which had no obvious significance to humans. Currently, scientists have identified approximately ~1,500,000 different species of plants, animals, and microbes. The actual number of different types of organisms, referred to as species, may be much higher.[89] These numbers refer, of course, to the species that currently exist, but we know from the fossil record that many species are not extinct. So the obvious question is, why are there so many different types of organisms?[90] Do they represent multiple independent creation events, and if so, how many such events have occurred? Given how different types of organisms look and behave, it seems possible that trees, mushrooms, spiders, whales, and humans represent distinct lineages and separate creation events.

As the actual diversity of organisms was discovered, a number of observations served to undermine the concept that organisms were created to serve or instruct humanity. The first of these was the fact that a number of organisms had very little obvious importance to the human condition. While (hopefully) obvious in the case of extinct organisms, this extended to a range of newly discovered (by Europeans) organisms; panda bears, potatoes, and maize come to mind. At the same time students of nature, known as naturalists, discovered many different types of upsetting and cruel behaviors within the natural world. Consider the fungus *Ophiocordyceps unilateralis*, which infects the ant *Camponotus leonardi*. The fungus takes control of the ant’s behavior, causing infected ants to migrate to environments that favor fungal growth before killing the infected ant. Similarly, the nematode worm *Myrmeconema neotropicum* infects the ant *Cephalotes atratus*, leading to dramatic changes in the infected ant's morphology and behavior. The infected ant’s abdomen turns red and is held raised up, which makes it resemble a fruit and increases the likelihood of the infected ant being eaten by birds (→). The birds transport the worms, which survive in their digestive systems until

[87] Northumberland Bestiary And as a general note, we focus on the European scientific tradition here, but others are similar.

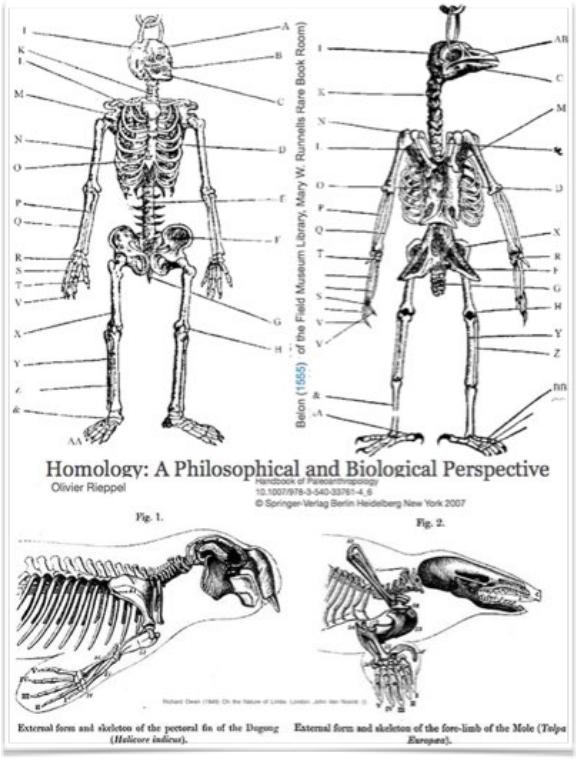
[88] What Is Natural Theology?

[89] How many species are there on Earth and in the ocean?

[90] As a technical point, which we will return to, we will refer to each distinct type of organism as a species.

they are excreted; they are then eaten by, and infect new ants to complete the worm's life cycle.[91] Perhaps the most famous example of this type of apparently cruel behavior involves wasps of the family *Ichneumonidae*. Female wasps deposit their fertilized eggs into the bodies of various types of caterpillars. The wasp's eggs hatch out and produce larvae that feed on the living caterpillar, consuming it from the inside out. Charles Darwin, in a letter to the American naturalist Asa Gray, remarked "There seems to me too much misery in the world. I cannot persuade myself that a beneficent and omnipotent God would have designedly created the *Ichneumonidae* with the express intention of their feeding within the living bodies of caterpillars, or that a cat should play with mice." Rather than presume that a supernatural creator was responsible for such cruel behaviors, Darwin and others sought alternative, morally neutral naturalistic processes that could both generate biological diversity and explain biological behaviors.

As the diversity of organisms became increasingly apparent and difficult to ignore, another broad and inescapable conclusion began to emerge from anatomical studies. Different organisms displayed remarkable structural similarities. For example, as naturalists characterized various types of animals, they found that they either had an internal skeleton (the vertebrates) or did not (the invertebrates). Comparative studies revealed that there were often many similarities between quite different types of organisms. A classic work, published in 1555, compared the skeletons of a human and a bird, both vertebrates.[92] While many bones have different shapes and relative sizes, what is most striking is how many bones are at least superficially similar to one another (→). Studies in "comparative anatomy" revealed many similarities between apparently unrelated organisms. For example, the skeleton of the dugong, a large aquatic mammal, appears quite similar to that of the European mole (→), a small terrestrial mammal that tunnels underground. In fact, there are general skeletal similarities between all vertebrates. The closer we look, the more similarities we find. These similarities run deeper than the anatomical, as we will discover, they extend to the cellular and molecular levels as well and involve both vertebrates and invertebrates. So the scientific question was, what explains such similarities? Why build an organism that walks, runs, and climbs, such as a human, with a skeleton similar to that of a organism that flies (birds), swims (dugongs), or tunnels (moles). Are these anatomical similarities just flukes or do they imply something deeper about how organisms were initially formed?



Organizing organisms, hierarchically

Carl Linnaeus (1707-1778) was a pioneer in taking the similarities between different types of organisms seriously. Based on such similarities (as well as differences), he developed a system to classify organisms in a coherent and hierarchical manner. Each organism had a unique place in this

[91] The Life of a Dead Ant: The Expression of an Adaptive Extended Phenotype

[92] Belon (1555) L'Histoire de la Nature des Oyseaux. Paris, Guillaume Cavellat

scheme, a unique set of coordinates.[93] What was, and occasionally still is, the controversial aspect of such a classification system is in how to decide which traits should be considered significant and which are superficial or unimportant, at least for the purposes of classification. Linnaeus had no real idea for how to explain why organisms be classified in such a hierarchical manner.

This might be a good place to reconsider the importance of guesses, hypotheses, models, and theories in biology, and science in general. Linnaeus noticed the apparent similarities between organisms and used it to generate his classification scheme, but he had no explanation for why such similarities should exist in the first place. Like Newton's law of gravitation, there was no mechanistic explanation for the relationship existed, just how it behaved. So what are the features of a scientific, that is predictive model? Such a model has to suggest observations or predict outcomes that have not yet been observed. It is the validity of these predictions that enables us to identify useful models. A model that makes no empirically testable predictions is not useful scientifically. In this light, Linnaeus's scheme was not scientific, just descriptive. The value of a scientific model, that is, a model that makes explicit predictions, even if they prove to be wrong, is that it enables us to refine, or force us to abandon, our current model. As a scientific model expands what it explains, and its predictions are confirmed, the model becomes a theory (while other "competing" models are abandoned). We assume that the way the model works is the way the world works. This enables us to distinguish between a law and a theory. A law describes what we see but not why we see it. A theory provides the explanation for why the law works.[94]

The Linnaean classification system placed organisms of a particular type together into a species. Similarly, species were grouped into genera, and so on.This, of course, raises a number of interesting questions - how different do two organisms have to be to fall into different species? How do we make such a decision? As we will see, each organism is unique genetically (its genotype) as well as in its various observable traits: its phenotype. If we look at organisms that appear similar, do we place larger individuals (of the same age) into a different species than smaller ones? The situation is even more complex when we think about modes of reproduction. Some organisms can reproduce, that is, produce offspring, by themselves; such organisms can be either asexual or self-fertilizing, often called hermaphroditic - a distinction that we will return to later. Other types of organisms are sexual, individuals need to cooperate with another of the same type to produce offspring. Here we find a reasonably common, but not universal, situation known as sexual dimorphism, in which individuals of the two sexes appear different, often dramatically, from one another.[95] It is often the case that organisms of the same type but different sexes, different developmental stages, and even growing under different conditions can have different phenotypes. It therefore requires careful study to recognize and characterize a particular type of organism.

Of course, what originally counted as a discrete type of organism, a particular species, was based on Linnaeus's or some other naturalists' judgement as an observer and classifier; it depended on which particular traits were assumed to be significant and useful to distinguish organisms of one species from those of another, perhaps quite, similar appearing species. The choice of these key traits is subject to debate. Based on the perceived importance and presence of particular traits, organisms could be split into two or more types (species), or two types originally considered separate could be reclassified into a single species.

As we will see, the individual organisms that make up a species are not identical but share many traits. As noted for organisms that reproduce sexually, there are sometimes dramatic differences

---

[93] Each organism can be identified by a species, within a genus, within a family, within an order, within a class, within a phylum, within a Kingdom.

[94] If we go back, Newton's law of gravity explained how objects behaved gravitationally, but it not why. In contrast, Einstein's theory of general relativity explained why there was gravity, and predicted behaviors that were not predicted by Newton's law.

[95] Sexual dimorphism & sexual dimorphism in spiders

between males and females of the same species (→ left ♂ & right ♀ spiders and ducks). These differences can be so dramatic that without further evidence, it can be difficult to tell whether two animals are members of the same or different species. In this light the primary criteria for determining whether sexually reproducing organisms are members of the same or different species is whether they can and do successfully interbreed with one another in the wild. Reproductive compatibility can be used to determine species distinctions on a more empirical basis, but it is not useful with asexual species, such as most microbes. An asexual organism is essentially a clone and species distinctions have to be based on other criteria, which we will return to later when we discuss genes and genomes. Within a species, there are sometimes regional (geographical) differences that are distinct enough to be recognizable. Where this is the case, these groups are known as populations or subspecies.[96] While distinguishable, the organisms in these groups retain the ability to interbreed and so are considered members of a single species. As an example tigers are *Panthera tigris*, while Siberian tigers are known as *Panthera tigris sumatrae,* sumatrae is the subpecies name.

After defining species, Linnaeus next grouped species that displayed similar traits into more inclusive groups, known as genera. While a species can be considered a natural, interbreeding population, a genus is a more artificial group. Which species are placed together within a particular genus depends on the common traits deemed important or significant by the person doing the classifying. This can lead to conflicts between researchers that are typically resolved by the collection of more comparative data and the building of community consensus. In part this situation arises because of the "flow" of evolution.

In the Linnaean classification scheme, each organism has a unique name, which consists of its genus and species names - this can be consider its primary coordinate within the classification scheme. The accepted usage is to write the name in italics with the genus name capitalized, for example, *Homo sapiens*. Following on this pattern, one or more genera are placed into larger, more inclusive groups (the next larger group is known as a “family”), and these groups, in turn, are placed into even larger groups. The end result of this process is the rather surprising observation that all organisms fall into a small number of “supergroups” or phyla. We will not worry about the traditional group names, because in most cases they really do not help in our understanding of basic biology. Perhaps most surprising of all, all organisms and all phyla – all of the organisms on Earth – can be placed into a single unified phylogenetic “tree” or perhaps better put, bush – they are all connected. That this should be the case is by no means obvious. Such an analysis could have produced multiple, disconnected classification schemes, but it did not. Finally, while forming discrete groups, that is groups with sharp boundaries, can be convenient, don't get confused. There is an inherent continuity through time linking all types of organisms. Where the boundaries between groups are drawn is always, in some important sense arbitrary.

Natural and un-natural groups

While a species, particularly a sexually reproducing species, can be seen as a natural group, the higher classification levels may or may not reflect biologically significant information. Such higher-level classification is an artifact of the human need to make sense of the world; it also has the practical value of organizing information, much like the way books are organized into chapters and

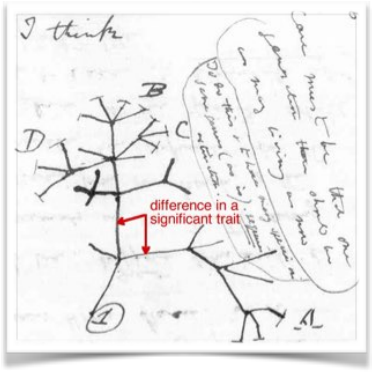
[96] The term race, a social construct, as no real value in biology: see Taking race out of human genetics

placed within in a library. We can be sure that we are referring to the same chapter in the same book or studying the same organism!

Genera and other higher-level classifications are based on a decision to consider one or more traits as more important than others. The assignment of a particular value to a trait can seem arbitrary. Let us consider, for example, the genus *Canis*, which includes wolves and coyotes and the genus *Vulpes*, which includes foxes. The distinction between these two groups is based on smaller size and flatter skulls in *Vulpes* compared to *Canis*. Now let us examine the genus *Felis*, the common house cat, and the genus *Panthera*, which includes tigers, lions, jaguars, and leopards. These two genera are distinguished by cranial features and the fact that *Panthera*, but not *Felix,* have the ability to roar. So what do we make of these distinctions, are they really sufficient to justify distinct groups, or should *Canis* and *Vuples* (and *Felix* and *Panthera*) be merged together? Are the differences between these groups biologically meaningful? They are in the sense that they recognize similarities and differences between organisms, but these similarities and differences may be ambiguous. Such ambiguity is illustrated by the fact that the higher order classification of an organism can change: organisms originally placed in one genus can become a separate genus within a family, the next more inclusive grouping, and vice versa, or a species can be moved from one genera to another. Consider the types of organisms commonly known as bears. There are a number of different types of bear-like organisms, a fact that Linnaeus's classification scheme acknowledged. Looking at all bear-like organisms we currently recognize eight types.[97] Four of these, the brown bear (*Ursus arctos*), the Asiatic black bear (*Ursus thibetanus*), the American bear (*Ursus americanus*), and the polar bear (*Ursus maritimus*) are more similar to one another, based on the presence of various traits, than they are to other types of bears. We therefore placed them in their own genus, *Ursus*. We have placed each of the other types of bear-like organisms, the spectacled bear (*Tremarctos ornatus*), the sloth bear (*Melurus ursinus*), the sun bear (*Helarctos mayalanus*), and the giant panda (*Ailuropoda melanoleuca*) in their own separate genera, because scientists consider these species more different from one another than are the members of the genus *Ursus*. The problem here is how big do these differences have to be to warrant a new genus? Hopefully, it is obvious to you that there are parts of any classification system that are subject to argument and others that are more easily agreed upon.

Evolution: making theoretical sense of Linnaean classification

So where does that leave us? Together with the cell theory (or perhaps better, the theory of biological continuity, we work on the assumption that the more closely related, evolutionarily, two species are, the more traits they will share and that the development of a new, biologically significant traits is what distinguishes one group from another. Traits that underlie a rational classification scheme are known as synapomorphies, a technical term. Basically these are traits that appear in one or the other branch point of a family tree and serve to define that branch point, such that an organism on one branch represent an evolutionary lineage, and so are part of a "natural" group, more closely related to one another and distinct from those on the other branch to which they are less closely related (→). The organisms within each branch are placed in a common Linnaean group. Going back further in time, the two groups, share a common ancestor, and are part of a larger, more inclusive Linnaean group. The continuous (unbroken) ancestral relationships between all organisms provides a reason for why organisms can be arranged into a hierarchical classification scheme.



A remaining question is, how do we determine ancestry when the ancestors lived, thousands, millions, or billions of years in the past. Since we cannot travel back in time, we have to deduce

[97] http://en.wikipedia.org/wiki/List_of_bears

relationships from comparative studies of living and fossilized organisms. Here the biologist Willi Hennig (1913-1976) played a key role.[98] He established rules for using shared, empirically measurable traits to reconstruct ancestral relationships, such that each group should have a single common ancestor, or more realistically, ancestral population. As we will discover later on, one of the traits now commonly used in modern studies are gene (DNA) sequence and genomic organization data, although even here there are plenty of situations where ambiguities remain, due to the very long times that often separate ancestors from present day organisms.

Fossils and family relationships: introducing cladistics (briefly)

As mentioned previously, we continue to discover new fossils, new organisms, and, as we will see, new genes. In most cases, fossils appear to represent organisms that lived many millions to hundreds of millions of years ago but which are now extinct. We can expect that there are dramatic differences between the ability of different types of organisms to become fossilized.[99] Perhaps the easiest organisms to fossilize are those with internal or external skeletons, yet it is estimated that between ~85 to 97% of such organisms are not represented in the fossil record. A number of studies indicate that many types of organisms have left no fossils whatsoever[100] and that the number of organisms at the genus level that have been preserved as fossils may be less, often much less than ~5%.[101] For some categories of modern organisms, such as the wide range of microbes, essentially no informative fossils exist at all.

Once scientists recognized that fossils provide evidence for extinct organisms, the obvious question was, do extinct organisms fit into the same classification scheme as do living organisms or do they form their own groups or even their own separate trees, which could provide evidence for multiple independent origins of life ("creation events") and multiple distinct common ancestors? This can be a difficult question to answer, since many fossils are only fragments of the intact organism. The fragmentary nature of the fossil record can lead to ambiguities. Nevertheless, the most reasonable conclusion that has emerged is that essentially all fossilized organisms fall into the classification scheme developed for modern organisms, although some organisms, such as the Ediacarian organisms, remain ambiguous.[102] The presumption is, however, that if we had samples of Ediacarian organisms for molecular (DNA) analyses, we could quickly resolve this question, and such an analysis would reveal that they fall nicely into the modern classification scheme with all other organisms do (a topic we will return to).[103] A classic example are the dinosaurs which, while extinct, are clearly descended from a specific type of reptile that gave rise to modern birds, while mammals are more closely related to a second, now extinct, group known as the “mammal-like reptiles.”

In rare cases, particularly relevant to human evolution, DNA sequence data can be recovered from bones. For example, it is possible to extract and analyze DNA from the bones of Neanderthals and Denisovan-type humanoids; both types of human-like organisms went extinct ~30,000 years ago. DNA sequence information has been used to clarify the relationship between Neanderthals, Denisovans, and modern humans, *Homo sapiens*.[104] Such data provides compelling evidence for

---

[98] A description of Willi Hennig’s impact on taxonomy

[99] Your inner fish video

[100] The incompleteness of the fossil record

[101] Absolute measures of the completeness of the fossil record

[102] Doser, 2015. The advent of animals: The view from the Ediacaran

[103] On the eve of animal radiation: phylogeny, ecology and evolution of the Ediacara biota

[104]:Paleogenomics of archaic hominins

limited interbreeding between these groups and has led for calls to reclassify Neanderthals and Denisovans as subspecies of *Homo sapiens*.[105]

Questions to answer:

18. Explain how extinct species could fit into the same classification scheme as used for living (observable) organisms.
19. Why are differences between organisms less informative in determining phylogenetic relationships than similarities?
20. What factors would influence your decision as to whether a trait found in two different organisms was present in their common ancestor?
21. You discover life on a planet orbiting another star in another galaxy; would you expect such organisms to fit into the Linnaean classification system?

Questions to ponder:

- What observations would you consider to decide whether Neanderthals and Denisovans were species, distinct from *H. sapiens*?
- Would sex with a Neanderthal be immoral?

The theory of evolution and the organization of life

Why exactly is it that birds, whales, and humans share common features, such as the organization of their skeletons, similarities that led Linnaeus to classify them together as vertebrates? Why are there extinct organisms, known only from their fossils, but which nevertheless share many common features with living organisms? And most importantly, why are there so many different types of organisms? Charles Darwin (1809-1882) and Alfred Wallace (1823-1913) proposed a model, described in great detail in Darwin's book The Theory of Evolution by Natural Selection, originally published in 1858, and more succinctly by Wallace, that answered these and a number of other questions.

> *The main unifying idea in biology is Darwin's theory of evolution through natural selection.*
> *– John Maynard Smith*

As we will see, evolutionary theory is based on a series of direct observations of the natural world and their logical implications. Evolutionary theory explains why similar organisms share similar traits and why we can easily place them into a nested (Linnaean) classification system. Organisms are similar because they are related to one another – they share common ancestors.[106] Moreover, we can infer that the more characters two species share the more recently they shared a common ancestor. We can even begin to make plausible, empirical and testable deductions about what those common ancestors looked like. As an example, we can predict that the common ancestor of all terrestrial vertebrates will resemble a fish with leg-like limbs - and we can predict the number and shape of the bones found in those limbs. Scientists have discovered fossils of such an organism, *Tiktaalik roseae* (←).[107] Its discovery is one more example of the fact that since its original introduction, and well before the mechanisms of heredity and any understanding

Tiktaalik roseae, an extinct organism that lived ~375 million years ago, is likely to be similar to the common ancestor of all terrestrial vertebrates (from "Your inner fish" by Neil Shubin).

[105] Humans mated with Neandertals much earlier and more frequently than thought & The downside of sex with Neanderthals

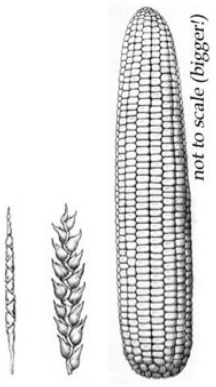
[106] As we will discover, there are organisms can appear similar that are not closely related; this is due to what is known as convergent evolution. That said, such organisms share a common ancestor, although it existed further back in time.

[107] Meet Tiktaalik roseae: An Extraordinary Fossil Fish A similar situation applies to the terrestrial ancestors of whales

of the molecular nature of organisms were resolved, evolutionary theory explained what was observed, made testable predictions about what would be found, and has been supported by what has, in fact, been found. In the case of particularly fast growing organisms, and very strong selection pressures (such as the presence of an antibiotic), we can observe evolutionary processes taking place over the course of days, weeks, and months – that is, in real time.[108]

Evolution theory's core concepts

So what are the facts and inferences upon which the Theory of Evolution is based? Two of its foundational observations are deeply interrelated and based on empirical observations associated with plant and animal breeding and the characteristics of natural populations. The first is the fact that whatever type of organism we examine, if we look carefully enough, making accurate measurements of visible and behavioral traits, which is known as its phenotype, we find that individuals vary with respect to one another. More to the point, plant and animal breeders recognized that the offspring of controlled matings between individuals often displayed phenotypes similar to those of their parents, indicating that the (invisible) factors responsible for phenotypic (observable) traits can be inherited. Over many generations, domestic animal and plant breeders used what is now known as artificial selection to generate the range of domesticated plants and animals with highly exaggerated phenotypes. For example, beginning ~10,000 years ago plant breeders in Mesoamerica developed modern corn (maize) by the selective breeding of variants of the grass teosinte (→).[109] Current evidence supports the idea that all of the various breeds of dogs, from the tiny to the rather gigantic (←), were derived from a common ancestor that lived between ~19,000 to 32,000 years ago. Although it is certainly true that new evidence may emerge that would change our estimates of where and when this common ancestor(s) lived.[110] In all cases, the crafting of domesticated organisms followed the same pattern.



In artificial, that is, human-driven selection, those organisms with desirable (or desired) traits were selected for breeding with one another. Organisms that did not have these traits were not permitted to breed. This process of artificial selection, carried out over hundreds to thousands of generations, led to organisms that display distinct or exaggerated forms of the selected trait. What is crucial to understand is that this strategy could work only if different versions of the trait were present in the original selected population and at least a part of this phenotypic variation was due to genetic, that is stable, heritable, and invisible factors. Originally, the nature of these genetic heritable factors was completely unclear. We refer to them as the organism's genotype, even though early plant and animal breeders would never have used that term.

The power of selection is based on the assumption that different organisms have different genotypes and that different genotypes produce different phenotypes. But the source of genotypic differences was not known to early plant and animal breeders. Were these differences imprinted on the organism in some way based on its experiences or were they the result of environmental factors? Was the genotype stable or could it be modified by experience? How were genotypic factors passed from generation to generation? And how, exactly, did a particular genotype produce or

---

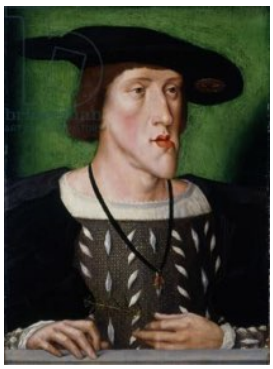
[108] Visualizing evolution as it happens see also Phagotrophy by a flagellate selects of colonial prey: a possible origin of multicellularity - Boraas et al 1998

[109] Molecular Evidence and the Evolution of Maize

[110] From wild animals to domestic pets, an evolutionary view of domestication

influence a specific phenotypic trait. As we will see this last question still remains poorly resolved for many phenotypes.

So what do we mean by genetic factors?



Here the answer is empirical. Traditional plant and animal breeders had come to recognize that offspring tended to display the same or similar traits as their parents. Such observations led them to assume that there was some factor within the parents that was expressed within the offspring and could, in turn, be passed from one generation to the next. A classic example is the Hapsburg lip (→), a trait that was passed through this European ruling family for generations.[111] In the case of artificial selection, an important point to keep in mind is that the various types of domesticated organisms produced are often dependent for their survival on their human creators, much like European royal families. Human protection relieves them of the constraints they would experience in the wild. Because of this dependence, artificial selection can produce quite exaggerated and, in the absence of human intervention, highly deleterious traits. Just look at domesticated chickens and turkeys, which, while not completely flightless, can fly only short distances and so are extremely vulnerable to predators. Neither modern corn (*Zea mays*) or chihuahuas, one of the smallest breeds of dog, developed by Mesoamerican breeders, would be expected to survive for long on their own in the wild.[112]

Limits on populations

It is an empirically demonstrable fact that all types of organisms, as opposed to specific individuals, are capable of producing many more than one copy of themselves. Consider, as an example, a breeding pair of elephants or a single asexually reproducing bacterium. Let us further assume that there are no limits to their reproduction, that is, that once born, the offspring will reproduce periodically over the course of their lifespan. By the end of 500 years, a single pair of elephants could theoretically produce ~15,000,000 living descendants.[113] Clearly if these 15,000,000 elephants paired up to form 7,500,000 breeding pairs, within another 500 years (1000 years altogether) there could be as many as $7.5 \times 10^6 \times 1.5 \times 10^7$ or 1.125×10^{14} elephants. Assuming that each adult elephant weighs ~6000 kilograms, which is the average between larger males and smaller females (an example of sexual dimorphism), the end result would be $\sim 6.75 \times 10^{18}$ kilograms of elephant. Allowed to continue unchecked, within a few thousand years a single pair of elephants could produce a mass of elephants larger than the mass of the Earth, an absurd, that is, impossible outcome. Clearly we must have left something out of our calculations! As another example, let us turn to a solitary, asexual bacterium, which needs no mate to reproduce. Let us assume that this is a photosynthetic bacterium that relies on sunlight and simple compounds, such as water, carbon dioxide, a nitrogen source, and some minerals to grow. A bacterium is much smaller than an elephant but it can produce a new bacterium at a much faster rate. Under optimal conditions our bacterium might divide once every ~20 minutes, or even faster, and would, within

> *A single cell of the bacterium E. coli would, under ideal circumstances, divide every twenty minutes. That is not particularly disturbing until you think about it, but the fact is that bacteria multiply geometrically: one becomes two, two become four, four become eight, and so on. In this way it can be shown that in a single day, one cell of E. coli could produce a super-colony equal in size and weight to the entire planet Earth.*
> *- Michael Crichton (1969) The Andromeda Strain*

---

[111] 'Imperial Stigmata!' The Habsburg Lip, A Grotesque 'Mark' Of Royalty Through The Centuries!: & Genes and Queens

[112] How DNA sequence divides chihuahua and great dane

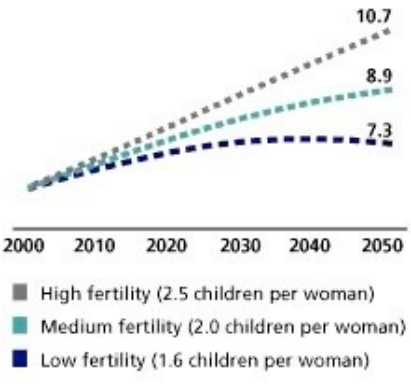
[113] Darwin's elephants

approximately a day, produce a mass of bacteria greater than that of Earth as a whole. Again, we are clearly making at least one mistake in our logic.

Elephants and bacteria are not the only types of organism on the Earth. In fact every known type of organism can produce many more offspring than are needed to replace themselves before they die. This trait is known as superfecundity. But unlimited growth does not and cannot happen for very long - other factors act to constrain it. In fact, if you were to monitor population numbers, you would find that the numbers of most organisms in a particular environment tend to fluctuate around a so-called steady state level. By steady state we mean that, averaging over time, the number of objects added to the system equals the number removed, so that the overall number, over time, remains (on average) constant. As an example, in a steady state population animals are continually being born and are dying, but the total number remains roughly constant.

So what balances the effects of superfecundity, what limits population growth? The obvious answer to this question is the fact that the resources needed for growth are limited and there are limited places for organisms to live. Thomas Malthus (1766-1834) was the first to clearly articulate the role of limited resources as a constraint on population. His was a purely logical argument. Competition between increasing numbers of organisms for a limited supply of resources would necessarily limit the number of organisms. Malthus painted a rather gloomy picture of organisms struggling with one another for access to these resources, with many living in an organismal version of extreme poverty, starving to death because they could not out-compete others for the food or spaces they needed to survive and reproduce. One point that Malthus ignored, or more likely was ignorant of, is that organisms rarely behave in this way. It is common to find various types of behaviors that limit the direct struggle between organisms for resources. For example, in some organisms, an adult has to establish, and defend, a territory before it can successfully reproduce.[114] The end result of this and similar types of behavior is to stabilize the population around a steady state level, which is a function of both environmental and behavioral constraints.

An organism's environment includes all factors that influence the organism. Environmental factors include changes in climate, as well as changes in the presence or absence of other organisms. For example, if one organism depends in important ways upon another, the extinction of the first will necessarily influence the survival of the second.[115] Similarly, the introduction of a new type of organism or a new trait, such as oxygen-generating photosynthesis, into an established environment can disrupt existing interactions and conditions. When the environment changes, existing steady state population levels may be unsustainable or some of the different types of organisms present may not be viable. If the climate gets drier or wetter, colder or hotter, if yearly temperatures reach greater extremes, or if new organisms, including as an example, new disease-causing pathogens, enter an area, the average population density may change or in some cases, if the environmental change is drastic enough, it may drop to zero, in other words certain populations could go extinct. Environmental conditions and changes will influence the sustainable steady-state population level of an organism (something to think about in the context of global warming and the destruction or disruption of natural environments).



An obvious example of this type of behavior involves the human population (→). Once constrained by disease, war, and periodic famine, the introduction of better public health and sanitation measures such as clean water and a more secure food supply, have led to reductions in infant mortality that have resulted in explosive growth in the human population. Now, in many countries, populations appear to be heading to a new steady state

[114] Territorial Defense, Territory Size, and Population Regulation

[115] Why the Avocado Should Have Gone the Way of the Dodo & Neotropical Anachronisms: The Fruits the Gomphotheres Ate

level, although exactly what that final population total level will be is unclear.[116] Various models have been developed based on different levels of average fertility. In a number of countries, the birth rate has already fallen into the low fertility domain, although that is no guarantee that it will stay there![117] In this low fertility domain (ignoring immigration), a country's population will decrease over time, since the number of children born is less than the number of people dying. This itself can generate social stresses. Decreases in birth rate per woman correlate with reductions in infant mortality, generally due to vaccination, improved nutrition, and hygiene, and increases in the educational level and the reproductive self-determination, that is, the emancipation of women. Where women have the right to control their reproductive behavior, the birth rate tends to be lower. Clearly changes in the environment, and here we include the sociopolitical environment, can dramatically influence behavior and impact reproductive rates and population levels.

The conceptual leap made by Darwin and Wallace

Charles Darwin and Alfred Wallace recognized the implications and significance of these key biological facts: the hereditable nature of variation between organisms, the ability of organisms to reproduce many more offspring than are needed to replace themselves, and the constraints on population size due to limited environmental resources. Based on these facts, they drew a logical implication, namely that individuals would differ in their reproductive success – that is, different individuals would leave behind different numbers of viable descendants. Over time, we would expect that the phenotypic variations associated with greater reproductive success, and the genotypes underlying these phenotypic differences, will increase in frequency within the population; over time they will displace those organisms with less reproductively successful phenotypes. Darwin termed this process natural selection, in analogy to the process of artificial selection practiced by plant and animal breeders. As we will see, natural selection is one of the major drivers of biological evolution.

Just to be clear, however, reproductive success is more subtle than the phrase "survival of the fittest" might imply. First and foremost, from the perspective of future generations, surviving alone does not matter much if the organism fails to produce offspring. An organism's impact on future generations will depend not on how long it lives but on how many fertile offspring it generates, a definition of success different from the standard English (American) definition. An organism that can produce many reproductively successful offspring at an early age will have more of an impact on subsequent generations than an organism that lives an extremely long time but has few offspring. Again, there is a subtle point here. It is not simply the number of offspring that matter but the relative number of reproductively successful offspring produced.

If we think about the factors that influence reproductive success, we can classify them into a number of distinct types. For example, organisms that reproduce sexually need access to mates, and must be able to deal successfully with the stresses associated with normal existence and reproduction. This includes the ability to obtain adequate nutrition and to avoid premature death from predators and pathogens. Similarly, organisms can cooperate (help) each other, and through such cooperation increase the odds that their offspring will survive, compare to solitary organisms. Both individual and social traits are part of the organism's phenotype, which is what natural selection acts on. It is worth remembering, however, that not all traits are independent of one another. Often the mechanism (and genotype) involved in producing one trait influences others – traits are often interdependent and sometimes incompatible, after all they are aspects of a single deeply-integrated organism. There are also non-genetic sources of variation. For example, there are molecular fluctuations that occur at the cellular level; these can lead genotypically identical cells to display

[116] Global population growth & The Joy of Stats

[117] Hans Rosling: Religions and babies

different behaviors, that is, different phenotypes.[118] Environmental factors and stresses also influence the growth, health, and behavior of organisms. These are generally termed physiological adaptations. An organism's genotype influences how it responds phenotypically to environmental factors, so the relationship between phenotype, genotype, and the organism's environment is complex.[119]

Mutations and the origins of genotype-based variation

So now the question arises, what is the origin of genetic, that is, inheritable variation? How do genotypes change? As a simple and not completely incorrect analogy, we can think of an organism's genotype as a book of instructions. This book is also known as its genome; do not worry if this seems too simple, we will add needed complexities as we go along. An organism's genome is no ordinary book. For simplicity we can think of it as a single unbroken string of characters. In humans, this string is approximately 3.2 billion (~3,200,000,000) characters or letters long and most types of cells in your body contain two very similar, but not identical copies of this book. A character corresponds to a base pair within a DNA molecule, which we will consider in detail later on. Within this string of characters there are regions that look like words and sentences, that is, regions that appear to have meaning. There are also extensive regions that appear to be meaningless. To continue our analogy, a few critical changes to the words in a sentence can change the meaning of a story, sometimes subtly, sometimes dramatically, and sometimes a change will lead to a story that makes no sense at all.

At this point we will define the meaningful regions, the words and sentences, as corresponding to genes and the other sequences as intragenic regions, that is, spaces between genes. It has been estimated that humans have ~25,000 genes. As we continue to learn more about the molecular biology of organisms, our understanding of both genes and intragenic regions will become more sophisticated. Regions that originally appeared meaningless have been found to influence the meaning of the genome. Many regions of the genome are unique, they occur only once within the string of characters. Others are repeated, sometimes hundreds to thousands of times. When we compare the genotypes of individuals of the same type of organism, we find that they differ at a number of places. For example, over ~55,000,000 variations have been found between all human genomes examined to date, and more are likely to be identified. When present within a population of organisms, these genotypic differences are known as polymorphisms, from the Latin meaning multiple forms. Polymorphisms are the basis for DNA-based forensic identification tests. One thing to note, however, is that only a small number of these variations are present within any one individual, and considering the size of the human genome, most people differ from one another at less than 1 to 4 letters out of every 1000. That amounts to between 3 to 12 million letter differences between two unrelated individuals. Most of these differences are single characters, but there can be changes that involve moving regions from one place to another, or the deletion or duplication of specific regions.

In sexually reproducing organisms, like humans, there are typically two copies of this book in most types of cells of the body, one derived from each of the organism's parents. Organisms (and cells) with two genomic "books" are known as diploid. When a sexual organism reproduces, it produces reproductive cells, known as gametes: sometimes these are the same size. When gametes differ in size, the smaller one is known as a sperm and the larger is known as an egg. Each gamete contains one copy of its own unique version of the genomic book and is said to be haploid. This haploid genome is produced through a complex process known as meiosis (considered in Chapter 11). Meiosis leads to a shuffling of the organism's original parental genomes. When a

[118] Something that has been studied in nine-banded armadillos that produce "identical" quadruplets.

[119] The global influence of genome on traits: An Expanded View of Complex Traits: From Polygenic to Omnigenic

haploid sperm and a haploid egg cell fuse, a new diploid organism is formed with its own unique pair of genomic books. The situation is rather different in asexual organisms.

The origins of polymorphisms: So what produces the genomic variations between individuals found within a population? Are these processes still continuing to produce genotypic and phenotypic variations or have they ended? First, as we have alluded to, and will return to again and again, the sequence of letters in an organism's genome corresponds to the sequence of characters in DNA molecules. A DNA molecule in water (and over ~70% of a typical cell is water) is thermodynamically unstable and can undergo various types of reactions that lead to changes in the sequences of characters within the molecule.[120] In addition, we are continually bombarded by radiation that can damage DNA.[121] Mutagenic radiation, that is, the types of radiation capable of damaging the genome, comes from various sources, including cosmic rays that originate from outside of the solar system, UV light from the sun, the decay of naturally occurring radioactive isotopes found in rocks and soil, including radon, and the ingestion of naturally occurring isotopes, such as potassium-40. DNA molecules can absorb such radiation, which can lead to chemical changes, that is, mutations. Many but not all of these changes can be identified and repaired by cellular repair systems, which we will consider, albeit only briefly, later on.

The second, and major source of change to the genome involves the process of DNA replication itself. DNA replication happens every time a cell divides and while remarkably accurate it is not perfect. Copying creates mistakes. In humans, it appears that replication creates approximately one error for every ~100,000,000 (10^8) characters copied. The cell's proof-reading and error repair systems correct ~99% of these errors, leading to an overall error rate during replication of about 1 in 10^{10} bases replicated. Since a single human cell contains ~6,400,000,000 (> 6 billion) bases of DNA sequence, that means that less than one new mutation is introduced per cell division cycle. Given the number of generations (cell division cycles) from fertilized egg to sexually active adult, that ends up producing ~100-200 new mutations (changes) added to an individual's genome per generation.[122] These mutations can have a wide range of effects, complicated by the fact that essentially all of the various aspects of an organism's phenotype are determined by the action of hundreds to thousands of genes working in a complex network. And here we introduce our last new terms for a while; when a mutation leads to change in a gene, it creates a new version of that gene, which is known as an allele of the gene. When a mutation changes the DNA's sequence, whether or not it is part of a gene, it creates what is known as a sequence polymorphism or simply a polymorphism, a different DNA sequence. Once an allele or polymorphism has been generated, it is as stable as the original molecule - it can be inherited from a parent and passed on to an offspring. Through the various processes associated with reproduction, which we will consider in detail later on, each organism carries its own distinctive set of alleles and its own unique set of polymorphisms. Taken together these genotypic differences, that is, differences in alleles and polymorphisms, produce different phenotypes. The DNA tests used to determine paternity and forensic identity work because they use the unique polymorphisms and alleles present within an individual's genome as a type of bar code for that person.

Two points are worth noting about genomic changes or mutations. First, whether produced by mistakes in replication or chemical or photochemical reactions, it appears that these changes occur randomly within the genome. With a few notable and highly specific exceptions there are no known mechanisms by which the environment (or the organism) can influence where a mutation will occur. The second point is that a mutation may or may not influence an organism's phenotype. The effects

---

[120] Instability and decay of the primary structure of DNA & DNA has a 521-year half-life:

[121] Although not not to worry, the radiation energy associated with cell phones, bluetooth, and various wifi devices is too low to damage DNA. But no matter what you might hear, it is a mistake to swallow a lamp that emits ultraviolet light.

[122] Human mutation rate revealed

of a mutation will depend on a number of factors, including exactly where the mutation is in the genome, its specific nature, the role of the mutated gene, the rest of the genome (the organism's genotype, known as the genetic background), and the environment in which the organism finds itself. We will consider the factors that influence gene and genome dynamics when we return to the behavior of DNA in cells.

Questions to answer:

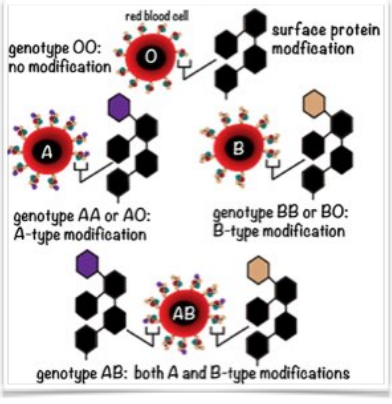
22. Explain why superfecundity is required for evolution to occur.
23. Why is the presence of genetically inheritable variation essential for any evolutionary model?

Questions to ponder:

- What advantages might be associated with self-imposed controls on mating?
- How could behaviors that limit an individual's ability to reproduce arise?

Genotype-phenotype relationships: discrete and continuous traits

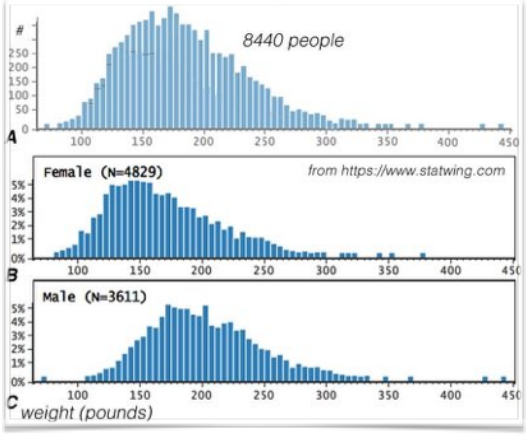
When we think about genetic polymorphisms and alleles, it is tempting to assume simple relationships. In some ways, this is an unfortunate residue from the way you may have been introduced to genetics. Perhaps you remember Gregor Mendel (1822-1884) and his peas. He identified distinct alleles of particular genes that were responsible for distinct phenotypes - yellow versus green peas, wrinkled versus smooth peas, tall versus short plants, etc. Other common examples might be the alleles associated with sickle cell anemia (and increased resistance to malarial infection), cystic fibrosis, and the major blood types. Which alleles of the ABO gene you inherited determines whether you have an O, A, B or AB blood type. We will consider what genes are and how they work in greater detail later on, but for now it is enough to know that the ABO gene encodes for a polypeptide; this polypeptide is a glycotransferase, that is, a protein (an enzyme) that catalyzes the addition of a specific chemical group, a carbohydrate, to a protein. Differences in the DNA sequences of the A, B, and O alleles results in differences in the polypeptides they encode. The polypeptides encoded by the A and B alleles differ in the reactions that they catalyze – different sugar groups are added by the A and B polypeptides. In contrast the polypeptide encoded for by the O allele is inactive, it does not function as a glycotransferase. Remember your cells are diploid; each cell has two copies of each gene (with the exception of the sex chromosomes - in humans, known as X and Y). In the case of the ABO gene, each cell has two copies, one inherited from your mom and one from your dad. The two ABO alleles you inherited may be the same or different.[123] If they are A and B, the proteins on your red blood cells have both the A and B modifications, resulting in an AB blood type. If they are A and O or A and A, your red blood cells have only the A modification, if they are B and O or B and B, your red blood cells have only the B modification, and if you have O and O, no modification (of this type) occurs and you have an O blood type (→). These are examples of what are known as discrete traits; you are either A, B, AB, or O blood type – there are no intermediates. You cannot be 90% A and 10% B.[124] The situation when the presence of a particular allele uniquely determines a particular trait, as in the case of the ABO gene, is rare – most traits are genetically more complex, they are known as polygenic.



[123] There are a number of common alleles of the ABO gene present in the human population, the most common (by far) are the A, B, and O alleles: http://omim.org/entry/110300

[124] Human blood types have deep evolutionary roots (unless of course, there is a mutation that influences the expression of the gene.

Most traits are continuous rather than discrete, they involve hundreds to thousands of genes (and their various alleles). For example, people come in a continuous range of heights, rather than in discrete sizes. If we look at the values of the trait within a population, that is, if we can associate a discrete number to the trait (which is not always possible), we find that each population can be characterized graphically by a distribution. For example, let us consider the distributions of weights in a group of 8440 adults in the USA (←). The top panel (A) presents a graph of the weights, along the horizontal or X-axis, versus the number of people with that weight along the vertical or Y-axis. We can define the "mean" or average of the population ($\bar{x}$) as the sum of the individual values of a trait (in this case each person's weight) divided by the number of individuals measured, as defined by the equation:

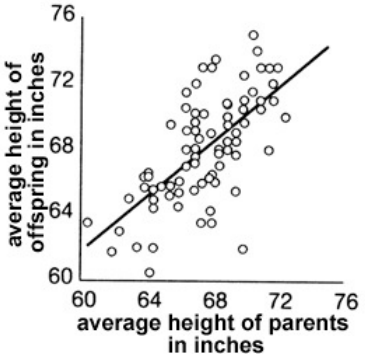


$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

In this particular data set, the mean weight of the population is ~180 pounds. It is common to recognize another characteristic of the population, the median. The median value is the point at which half of the individuals have a smaller value of the trait and half have a larger value. In this case, the median is ~176. Because the mean does not equal the median, we say that the distribution is asymmetric, that is there are more people who are lighter than the mean value compared to those who are heavier. Another way to characterize the shape of the distribution is by what is known as its standard deviation, indicated by the Greek letter sigma (σ). There are different ways to calculate the standard deviation that reflect the shape of the population distribution, but for our purposes we will use a simple one, the so-called uncorrected sample standard deviation (→).[125] To calculate this value subtract the mean value for the population ($\bar{x}$) from the value for each individual (x_i); since x_i can be larger or smaller than the mean, this difference can be a positive or a negative number. We then take the square of the difference, which makes all values positive (hopefully this makes sense to you). We sum these squared differences together, divide that sum by the number of individuals in the population (N), and take the square root, which reverses the effects of our squaring x_i, to arrive at the standard deviation of the population. The smaller the standard deviation, the narrower the distribution - the more organisms in the population have a value near to the mean. The larger σ is, the greater is the extent of the variation in the trait in the population.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}$$

So how do we determine whether a complex (that is, determined by many genes and their allelic variants) trait like weight, or any of a number of other non-discrete, continuously varying traits, is genetically determined? We could imagine, for example, that an organism's weight is simply a matter of how easy it was for it to get food. A standard approach to determine whether a trait has a genetic component is to ask whether there is a correlation between the phenotype in the parents (e.g. their heights) and the phenotypes of the offspring (its height). That such a correlation between parents and offspring exists for height is suggested by this graph (→), but notice we are seeing a trend, parental height is not a perfect predictor of offspring height - other factors must be involved.



One thing that we cannot determine from such data, however, is how many genes are involved in the genetic determination of a trait or how their effects are influenced by the environment and the offspring's specific history. As an example, "human height has been increasing

[125] wikipedia: standard deviation & http://www.mathsisfun.com/data/standard-deviation.html

during the 19th century when comprehensive records began to be kept. The mean height of Dutchmen, for example, increased in height from 165cm in 1860 to a current average height of 184cm, a spectacular increase that probably reflects improvements in health care and diet, rather than changes in genes."[126] Geneticists currently estimate that allelic differences at more than ~50 genetic loci (positions in the genome) make significant contributions to the determination of height, while allelic differences at hundreds of other genes have smaller effects.[127] At the same time, specific alleles of certain genes can lead to extreme shortness or tallness. For example, mutations that inactivate or over-activate genes encoding factors required for growth can lead to dwarfism or gigantism.

On a didaskalogenic note[128], you may remember learning that alleles are often described as if they are either dominant or recessive (a topic we will return to). But the extent to which an allele is dominant or recessive often depends upon how well we define a particular trait and the extent to which it is influenced by other factors and variations. These effects reveal themselves through the fact that people carrying the same alleles of a particular gene can display (or not display) the associated trait, which is known as penetrance, and they can vary in the strength of the trait, which is known as expressivity.[129] Both the penetrance and expressivity of a trait can be influenced by the rest of the genome, that is, the presence or absence of particular alleles of other genes. Environmental factors can also have significant effects on the phenotype associated with a particular allele or genotype.

Variation, selection, and speciation

Combining genetic and associated phenotypic variation, superfecundity, and stable population size, Darwin and Wallace's breakthrough conclusion was that different members of the population would display differences in reproductive success. Some genotypes, and the alleles they contain, would become more common within subsequent generations because the individuals that contained them would reproduce more successfully. Other genotypes would become less common, or disappear altogether. The effects of specific alleles on an organism's reproductive success will, of course, be influenced by the rest of the organism's genotype, its structure and behaviors, both selectable traits (that is traits that influence reproductive success), and its environment. While some alleles can have a strong positive or negative impact on reproductive success, the effects of most alleles are subtle, assuming they produce any noticeable phenotypic effects at all. A strong positive effect will increase the frequency of the allele (and genotype) associated with it in future generations, while a strong negative effect can lead to the allele disappearing altogether. An allele that increases the probability of death before reproductive age is likely to be strongly selected against, whereas an allele that has only modest effects on the number of offspring an organism produces will be selected for, or against, more weakly.

What Darwin and Wallace did not know was that genetic information is stored in molecules of DNA, and that that information can be altered through a variety of mechanisms (mutations) that include sequence duplication, deletion, and recombination (shuffling). Moreover, because DNA molecules are relatively stable they can survive the death of the organism, be released into the environment, and (under certain conditions) be transferred into living organisms and become part of their genetic material. These are all features of the molecular nature of genetic information (genes) and how DNA is manipulated, that is, replicated, repaired, and used to express information within

---

[126] "From Galton to GWAS: quantitative genetics of human height": http://www.ncbi.nlm.nih.gov/pubmed/21429269

[127] Genetics of human height: http://www.ncbi.nlm.nih.gov/pubmed/19818695

[128] We call instruction/instructor-dependent thinking didaskalogenic:

[129] Where genotype is not predictive of phenotype: understanding reduced penetrance in human inherited disease

cells. Recognizing these facts led to what is known as the Modern Synthesis of evolutionary theory.[130] While the basic Darwinian rules are the same, the possible molecular complexities make evolutionary processes even more powerful. We will be considering these various molecular processes as we proceed.

Questions to answer:

23. How would you explain the observation that the products of artificial selection are not generally competitive with "native" organisms?
24. What does the word correlation mean to you? what does it mean mathematically or practically?
25. If an individual's height is determined by the genotypes of their parents, then why don't all height measurements line on a straight line? Where could the scatter come from?
26. Consider a population and generate graphs that display (for a particular trait) the impact of larger and smaller standard deviations as well as median values that are higher or lower than the mean.

Types of (simple) selection

While it is something of an oversimplification, we begin with three basic types of selection: stabilizing (or conservative), directed, and disruptive. We will then introduce the complexities associated with the random aspects of reproduction and the linked nature of genes. We start with a population composed of individuals displaying genetic variation in a particular trait. The ongoing processes of mutation continually introduces new genotypes, and their varying effects on phenotype. The effects of mutations can range from the lethal, the organism that carries the mutation either dies or produces no offspring, to apparently neutral – an organism that carries the mutation displays no obvious change in phenotype or reproductive success. A complicating factor, that we will consider in more detail later, is that the phenotypic effects of a particular mutation, leading to a mutant or alternative allele, often depend upon the rest of the genome - due to so called genetic background effects. At the same time, changes in the population and the general environment influence the predominant types of selection that occur over time, and different types of selection may well (and most certainly are) occurring for different traits.

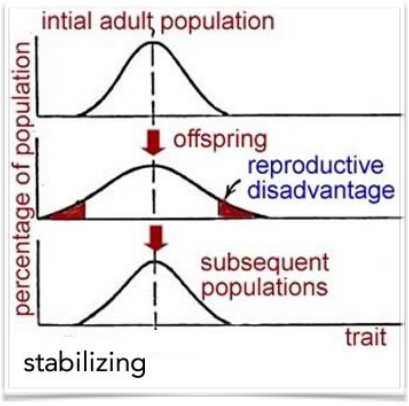
For each type of selection, we will illustrate the effects as if they were acting along a single dimension, for example smaller to larger, stronger to weaker, lighter to darker, or slower to faster. In fact, most traits vary along a number of dimensions. For example, consider the trait of ear, paw, heart, or big toe shape. An appropriate type of graph would be a multi-dimensional surface, but that is harder to draw clearly. It is also possible that a genotype that influences one trait may also influence another, apparently independent, trait. For simplicity's sake, we will start with populations whose distribution for a particular trait can be described by a simple and symmetrical curve, that is the mean and the median are the same. New variants, based on new mutations (new alleles and combinations of alleles), generally fall more or less randomly within this distribution. Under these conditions, for selection NOT to occur we would have to make an seriously unrealistic assumption, namely that an organism (or a pair of organisms, assuming that this is a sexually reproducing species) are all equally successful at surviving and producing offspring, something that is observably not the case. Any time genetic variation influences reproductive success selection occurs, although the strength of selection (the average difference in the number of viable offspring produced) may vary dramatically between traits.

Stabilizing selection: Sometimes a population of organisms appears static for extended periods of time, that is, the mean and standard deviation of a trait are not changing over time. Does that mean that selection has stopped? Obviously we can turn this question around: if we assume that there is a population with a certain stable mean and standard deviation of a trait – what would happen over time if selection disappeared?

---

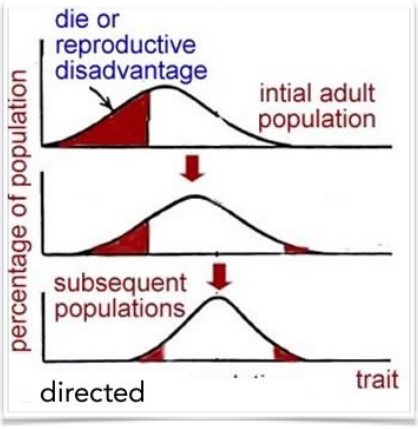
[130] Modern synthesis in evolutionary biology

Let us assume we are dealing with an established population living in a stable environment. This is a real world population, where organisms are superfecund, that is, capable of reproducing more and sometimes, many more organisms than are needed to replace them when they die and that these organisms mate randomly with one another. Now consider the factors that lead to the original population distribution: why is the mean value of the trait the value that it is? What factors influence the observed standard deviation? Assuming that natural selection is active, it must be that organisms that display a value of the trait far from the mean are (on average) at a reproductive disadvantage compare to those with the mean value of the trait (→). We do not know why this is the case and don't really care at the moment. Now if selection, at least for this trait, is inactive what will happen? The organisms far from the mean are no longer at a reproductive disadvantage, so their numbers in the population will increase. The standard deviation will grow larger, until at the extreme, the distribution will be almost flat, characterized only by a maximum and a minimum value, reflecting the limits of what the system can produce and remain viable. New mutations and existing alleles that alter the trait within this range will not be selected against, so they will increase in frequency.



In a real population, the mean and standard deviation associated with the trait remain constant, assuming that the environment is constant. We therefore predict "negative" selection against extreme values of the trait, which means that these individuals tend to produce fewer viable offspring than those with a value of the trait near the mean.[131] We can measure that degree of selection "pressure" by following the reproductive success of individuals with different values of the trait. We might predict that the more extreme the trait, that is, the further from the population mean, the greater its reproductive disadvantage (negative selection) will be, so that with each generation, the contribution of these outliers in the population will be reduced. The distribution's mean will remain constant. The stronger the disadvantage, referred to as negative selective pressure, the outliers face, the narrower the distribution will be – that is, the smaller the standard deviation. In the end, the size of the standard deviation will reflect both the strength of selection against outliers and the rate at which new variations enters the population through mutation. Similarly, we might predict that where a trait's distribution is broad the impact of the trait on reproductive success will be relatively weak.

Directed selection: Imagine that the population's environment changes. It may now be the case that the phenotype of the mean is no longer the optimal phenotype in terms of reproductive success, the only factor that matters, evolutionarily. A different value of the trait may be more favorable. Under these conditions we would expect that, over time, the mean of the distribution would shift toward the phenotypic value associated with maximum reproductive success (→). Once reached, and assuming the environment stays constant, stabilizing selection again becomes the predominant process. One outcome to emerge from a changing environment leading to directed selection is that, as the selected population's mean moves, it may well alter the environment of other organisms.

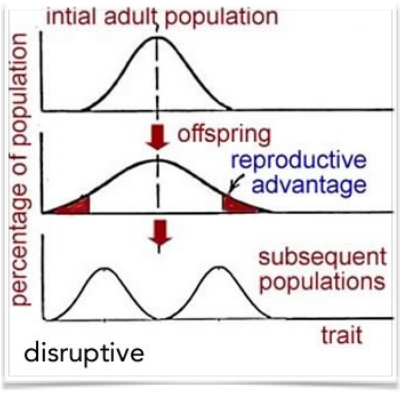


For directed selection to work, the environment must change at a rate and to an extent compatible with the changing mean phenotype of the population. Too big and/or too rapid a change and the reproductive success of all members of the population may be dramatically reduced. The ability of the population to change will depend upon the genetic variation present within the original population and the rate at which new mutations are produced, generally a

[131] By "viable" we mean offspring that live to reproduce, and that themselves reproduce successfully.

relatively slow and constant process.[132] In some cases, the change in the environment may be so fast or so drastic and the associated impact on reproduction so severe that selection will fail to move the population and extinction will occur.

Disruptive selection: A third possibility is that a population of organisms find themselves in an environment in which traits at the extremes of the population's phenotypic distribution have a reproductive advantage over those around the mean. If we think about the trait distribution as a multidimensional surface, it is possible that in a particular environment (which may correspond to multiple geographic regions), there will be multiple distinct strategies that lead to greater reproductive success compared to others. This leads to what is known as disruptive selection (↓). In an asexually reproducing population, various lineages will be subject to selective pressures based on the environments (regions) they come to inhabit, and the likelihood that individuals move from environment to environment, or that the environment changes dramatically. The effect of disruptive selection in a sexually reproducing population will be opposed by the random mating between members of the population, which does not occur in asexual populations. But is random mating a good assumption? It could be that the different environments, which we will refer to as ecological niches, are physically distant from one another and that organisms do not travel far to find a mate. The population may then split into subpopulations in the process of adapting to the two different niches. Over time, two species could emerge, since when and with whom one chooses to mate with and the productivity of such matings, are themselves selectable traits. Disruptive selection will, overtime, lead to the generation of new species, and over long periods of time, the millions of existing species and the even greater number of extinct species. The diversity of life was the observation that Darwin and Wallace originally set out to explain, and evolutionary processes provide a plausible mechanism.



Questions to answer:

27. Why does variation never completely disappear even in the face of strong stabilizing selection?
28. Under what conditions would stabilizing selection be replaced by directed or disruptive selection?
29. By looking at a population, how might one estimate the strength of conservative selection with respect to a particular trait?

Questions to ponder:

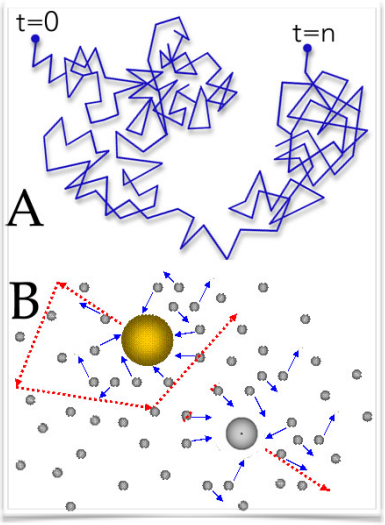
- Why is it difficult to be sure you know why a particular allele or trait was selected?
- How might phenotypic variation influence the choice of a mate (during sexual reproduction)?

Considering stochastic processes

Biological systems are characterized by what are known as stochastic processes. We will find that stochastic processes play an important role in evolutionary mechanisms (population bottlenecks, founder effects, genetic drift, meiotic recombination) as well as molecular processes within cells and tissues (both discussed later on). You may not be familiar with the word stochastic, it is a word whose meaning is often confused with random. So, what exactly distinguishes a stochastic from a random process? A truly random process has no underlying natural cause and so is completely unpredictable. A miracle could be considered a random process. From a scientific perspective, one could argue that there are no truly random natural processes or events, no miracles. Our working hypothesis is that all natural events have identifiable and measurable causes.

[132] As we will consider later when we consider these molecular processes, there are times when physiological stress can lead to increased global mutations rate. Mutation as a Stress Response and the Regulation of Evolvability

That said, that does not mean that every individual event can be predicted. Natural events can be unpredictable for one of two basic reasons: the event may be determined by theoretically unknowable or currently unknown factors, as in the case of the radioactive decay of atoms. Alternatively, the event may be the result of a large number of theoretically knowable events that are, for a variety of practical reasons, impossible to measure accurately. Such events are analogous to, or versions of, Brownian motion, a phenomena named after the Scottish botanist Robert Brown (1773-1858). In Brownian motion, small, but visible particles suspended in a solution (air or water) are found to move in a jerky and irregular manner (A→). Brownian motion arises because the visible particle is colliding with many invisible objects (molecules) present in the environment (air/water: B→).[133] The average energy transferred through these collisions reflects the temperature of the system. At higher temperatures the molecules have a higher average (mean) kinetic energy ($^1/_2\ mv^2$). During a particular time interval, the sum of all collisions can lead to an unbalanced force on the particle that causes it to move. A short time later the vector sum of these collision forces is likely to point in a different direction and the particle will now move in that direction. Collisions between molecules supply the energy to drive the dissociation of molecules from one another and supply the activation energy required for chemical reactions to proceed, topics that we will return to when we consider the thermodynamics of reaction systems (Chapter 5). At the individual event level, the system is unpredictable in practice (but not in theory) because there are so many molecules and collision events involved – for example, in water there are ~3×10^{22} water molecules per cubic centimeter, with the average water molecule traveling ~2.5×10^{-8} centimeters between collisions.[134] The end result is that the speed and direction of visible particle and invisible molecule movements are constantly changing.



In classical (that is, pre-quantum mechanical) physics, it was assumed that if it were possible to know the velocity (speed and direction) of every molecule in the system, as well as the dynamics of the collisions, we could predict the future behavior of the system and the paths of Brownian movements.[135] But it turns out that the world does not behave that way. In fact, we cannot (even theoretically) achieve this level of accurate measurement; we are limited by what is known as the Heisenberg Uncertainty principle, which arises from the fact that matter is composed of objects with both wave- and particle-like properties, rather than simple billiard ball-like particles.[136]

So why, if Brownian motion is a random process is it possible to study it scientifically? The answer is based on the fact that when we look at many objects, the behavior of the population becomes predictable – this predictability implies an underlying cause. For example, consider measurements of a large number of particles undergoing Brownian movement. If we measure the distance between where they start (t=0) and where they end up (t=n) as a function of time (see A↑ above), we find that the average distance travelled (but not the direction of travel or the extent of travel of any particular particle) is predictable and reflects the size of the particle, the nature of the system (water, air, etc), and its temperature. Its predictability indicates that Brownian motion is due

---

[133] Albert Einstein: The Size and Existence of Atoms & Einstein and Brownian Motion

[134] The properties of water: http://galileo.phys.virginia.edu/classes/304/h2o.pdf

[135] see Laplace's demon: https://en.wikipedia.org/wiki/Laplace's_demon

[136] Want to know more? check out: What is the Heisenberg Uncertainty Principle? and How Heisenberg Became Uncertain (https://youtu.be/UFYnsxLuFdQ)

to underlying (calculable) physical processes.

The situation is similar to that of rolling dice. While it is impossible to accurately predict the outcome of a single dice roll, as we increase the number of rolls (the population of rolls), we find that the overall behavior becomes increasingly predictable, each of the six numbers (assuming that this is a fair cube dice) will appear 1/6th of the time. The larger the number of rolls, the more closely the number of each possible outcome will approach 1/6th of the total. While the outcome of any individual roll remains unpredictable, the behavior of a population of rolls is predictable – a behavior known as the law of large numbers. A similar situation occurs with radioactive atoms; while it is impossible to predict when any particular atom will decay, we find that when we consider a large enough population we can accurately predict when any particular percentage of the original population will have decayed. Typically, the time it takes for 50% of the original atoms to decay is known as the "half-life" of the isotope and can be determined to very high precision.

In the case of rolling dice, and other similar (simple) stochastic processes, it is important, but hard to remember, that each individual event is independent, what happened in the past does not influence what will happen next. Forgetting this rule leads to what is known as the Gambler's Fallacy.[137] As an example, you roll a die eight times and get 2, 2, 5, 2, 2, 6, 2, 2. Assuming of course that this is a fair die, what is the probability that the next roll will come up 2? No matter how many times a 2 came up in the past, the chance of rolling a 2 on the next roll remains the same, 1/6.

A complexity that occurs within biological systems is that while a particular event can be stochastic, individually unpredictable but well behaved in a large enough population, in the cell or in an organism, a single event, such as the activation or mutation of a particular gene, can change the system so as to produce different behaviors and outcomes. A mutation can initiate the process by which a cell becomes cancerous. It is therefore possible, and perhaps likely, that if the history of the organism (or life) were to be "rerun" (an impossible situation), outcomes would be different.

Questions to answer:

30. What types of behaviors define a stochastic event; what types of everyday stochastic events are you familiar with. How do you know that they are not random?
31. What types of events are not, in theory, study-able scientifically?

Question to ponder:

- How might you decide whether a pattern in data was due to an underlying process or "just" to chance ?

Population size, founder effects and population bottlenecks

When we think about evolutionary processes from a strictly selection-based perspective, we ignore important factors that can impact the process. For example, what happens when a small number of organisms (derived from a much larger population) colonize a new environment? This is a situation that produces what is known as a founder effect. Something similar happens when a large population is dramatically reduced in size for any of a number of reasons, a situation known as a population bottleneck. In both founder effects and population bottlenecks, the small populations that result can have different allele frequencies than the original "parental" population and are more susceptible to the effects of genetic drift, a stochastic, non-selective process. Together founder effects, bottlenecks, and genetic drift can produce populations with unique traits that are not directly due to the effects of natural selection. Since founder effects and population bottlenecks can occur a number of times during the course of a populations' evolution, it is a mistake to assume that all observed traits have positive effects on reproductive success. If we think of evolutionary change as reflecting the movement of a population through a fitness landscape–the combination of the various

[137] Gambler's Fallacy: https://en.wikipedia.org/wiki/Gambler's_fallacy

factors that influence reproductive success–over time, then the isolation of small populations, and evolutionary changes within them, can cause a jump from one place in the landscape to another. Once in the new position, and as the population grows larger, new adaptations can be possible – selection again becomes the main, but not exclusive, driver of evolutionary change. Deleterious effects, that become frequent due to non-adaptive processes, can be ameliorated. A population invading a new environment will encounter a new set of organisms to compete and/or cooperate with. A catastrophic environmental change will change the selective landscape, removing or introducing competitors, predators, pathogens, and cooperators, favoring new adaptations and selecting against others that might have once been beneficial, in terms of reproductive success. One effect of the major extinction events that have occurred during the evolution of life on Earth is that they provide a new adaptive context, a different and less densely populated playing field with fewer direct competitors.[138] The expansion of various species of birds and mammals that followed the extinction of the dinosaurs is an example of one such opportunity, associated with changes in selective pressures.

Founder effects: What happens when a small subpopulation, a few individuals, becomes isolated, for whatever reason, from its parent population? The original (large) population will contain a number of genotypes and alleles. If this population is in a new environment it will be governed primarily by directed and conservative selection. We can characterize this parental population in terms of the frequencies of the various alleles present within it. For the moment, we will ignore the effects of new mutations, which will continue to arise within the population but generally at a slow rate. Now assume that a small group of organisms comes to colonize a new, geographically separate environment such that it is reproductively isolated from its parental population – no individuals travel between the parent and the colonizing populations.

The classic example of such a situation is the colonization of newly formed islands, but the same process applies more generally during various types of migrations. By chance, the frequency of alleles in a small isolated population is likely to be different from the allele frequencies found in the much larger parent population. Why is that? It is based on the randomness of the sampling of the original population. Consider, as an example, rolling a die (discussed above). Each side will appear 1/6th of the time. But imagine that the number of rolls is small. Would you expect to get each number appearing with equal probability? The answer is decidedly NO!!![139] See how many throws are required to arrive at an equal 1/6th probability distribution; the number is likely larger than you would guess.

Sampling populations: We can apply this “law of large numbers” to populations using the following logic. First, we recognize that if we wanted to determine the exact frequency of each allele of a particular genetic locus or gene in a particular population at a particular time, we would need to determine which allele(s) are present in each individual, BUT that is quite an intensive, expensive, and often impossible task. So we have to use some other method to estimate allele frequencies – we turn to ”sampling”. We examine a random set of individuals, a sample. If the number in the sample is small with respect to the total population size, we can expect significant differences in measured (sampled) and actual (total) population allele frequencies. These differences become smaller as the sample size increases. To provide a concrete example, consider a large population in which each individual carries one (and only one) of six alleles of a particular gene and that the percentage of each type is equal (1/6th). The selection of any one individual from this population is like a throw of a fair die; there is an equal 1/6th chance of selecting an individual with one of the six alleles. Since the parental population is large, the removal of one individual does not appreciably change the distribution of alleles remaining, so the selection of a second individual produces a

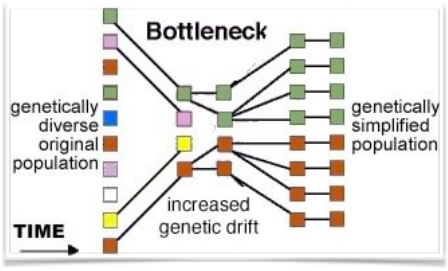
[138] Big Five mass extinction events and How life blossomed after the dinosaurs died

[139] Here is a reasonably good one: http://www.math.uah.edu/stat/apps/DiceExperiment.html

result that is independent of the first, just like individual rolls of the die and are equally likely to result in a 1/6th chance to select any one of the six alleles. But producing a small subpopulation with 1/6th of each allele (or the same percentages of various alleles as are present in the parent population) is, like the die experiment above, unlikely. The more genotypically complex the parent population, the more unlikely it is. Imagine that the smaller colonizing population only has, for example, three members (three rolls of the die) – not all alleles present in the original population can possibly be represented. Similarly, the smaller the subpopulation the more likely that the new population will be genetically distinct from the original population. So when a small group from a parent population invades or migrates into a new environment, it is likely to have a different genotypic (allelic) profile compared to its parent population. This difference is not due to natural selection but rather to chance alone. Nevertheless, it will influence subsequent evolutionary events; the small subpopulation will likely respond in different ways to new mutations and environmental pressures based on which alleles are present. The situation will be further influenced if genetic factors impact migratory behavior or reproductive success in the new environment.

The *Homo sapiens* appears to have emerged out of Africa ~500,000 years ago.[140] Genetic studies reveal that African populations display a much greater overall genotypic (genetic) complexity than do groups derived from it, that is, everyone else. What remains controversial is the extent to which migrating populations of humans in-bred with what are known as archaic humanoids (such as Neanderthals and the Denisovans), which appear to have diverged from the *Homo sapiens* lineage ~1.2 million years ago.[141] Such mating occurred (it appears) outside of Africa, and led to another source of genetic diversity.

Population bottlenecks: A population bottleneck is similar to, but distinct in important ways from a founder effect. Population bottlenecks occur when some environmental change leads to the dramatic reduction in the size of a population. Catastrophic environmental changes, such as asteroid impacts, massive and prolonged volcanic eruptions associated with continental drift, or the introduction of a particularly deadly pathogen that kills a high percentage of the organisms that it infects, can all create population bottlenecks (←). Who survives the bottleneck can be due only to "luck" or may be based on genetic factors, for example, alleles associated with disease resistance.



There is compelling evidence that such drastic environmental events are responsible for population bottlenecks so severe that they led to mass extinctions. The most catastrophic of these extinction events was the Permian extinction that occurred ~250 million years ago; during this event it appears that ~95% of all marine species and ~75% of land species went extinct.[142] If most species were affected, we would not be surprised if the surviving populations experienced serious bottlenecks. The subsequent diversification of the surviving organisms, such as the Dinosauria, which includes the extinct dinosaurs and modern birds, and the Cynodontia, which includes the ancestors of modern mammals, including us, could be due in part to these bottleneck-associated effects, for example, through the removal of competing species or predators. An asteroid impact, known as the Cretaceous-Tertiary event, occurred ~65 million years ago; it contributed to the extinction of the dinosaurs and led to the rapid expansion and diversification of mammals, which had first appeared in the fossil record ~100 million years earlier.

---

[140] Although dating origins depends upon finding fossils: see [140] The great human expansion and Oldest Homo sapiens fossil claim rewrites our species' history

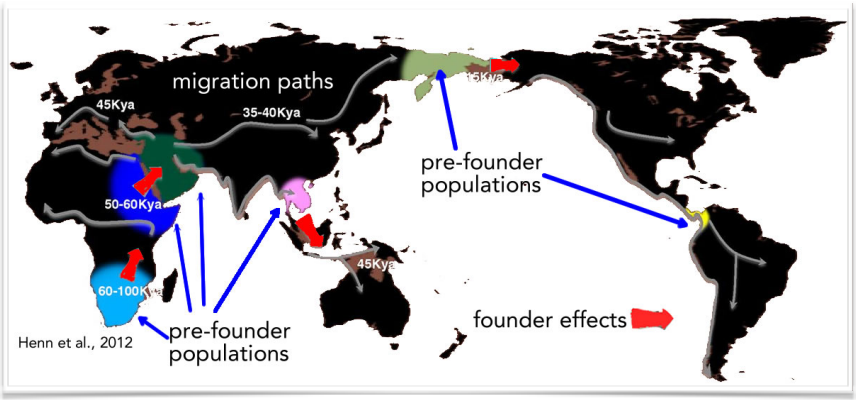
[141] Genetic Data and Fossil Evidence Tell Differing Tales of Human Origins

[142] The Permian extinction and the evolution of endothermy

While surviving an asteroid impact, or other dramatic changes in climate may be random, in other cases who survives a bottleneck is not. Consider the effects of a severe drought or highly virulent bacterial or viral infection; the organisms that survive may have specific phenotypes, and associated genotypes, that influenced their chance of survival. In such a case, the effect of the bottleneck event would produce directed changes in the distribution of genotypes (alleles) in the post-bottleneck population – selective effects that could continue to influence the population in various ways. For example, a trait positively associated with pathogen resistance may also have negative phenotypic effects. After the pathogen-driven bottleneck, mutations that mitigate any negative effects associated with the pathogen resistance trait may have a selective advantage. The end result is that traits that would not be selected in the absence of the pathogen, are selected and become common. In addition, the very occurrence of a rapid and extreme reduction in population size has its own effects. For example, it would be expected to increase the effects of genetic drift (see below) and could make finding a mate more (or less) difficult.

We can identify extreme population reduction events, such as founder effects and bottlenecks, by looking at the variation in genotypes, that is, the sequence of DNA molecules, particularly sequence changes not expected to influence phenotypes, mating preference, or reproductive success. These so-called neutral polymorphisms are expected to accumulate in the regions of the genome between genes (intragenic regions) at a constant rate over time (can you suggest why?) The rate of the accumulation of neutral polymorphisms serves as a type of population-based biological clock. Its rate can be estimated, at least roughly, by comparing the genotypes of individuals of different populations whose time of separation can be accurately estimated, assuming of course that there has been no significant migration between the populations.

Studies using genomic sequence data, the ancestral human population appears to have undergone a bottleneck around ~1.2 million years ago.[143] Once established, groups of modern humans migrated within and out of Africa (→), undergoing a series of founder effect events between ~45,000 to ~60,000 years ago. Groups (small populations) of humans migrated out of southern Africa into the Horn of Africa, then into the Arabian peninsula, and from there into Europe, Asia, Oceania, and finally into North America and throughout central and South America. Comparing genotypes, that is, neutral polymorphisms, between isolated populations enables us to estimate that humans reached Australia ~45,000 years ago and entered the Americas in multiple waves beginning ~16,000 years ago. The arrival of humans has been linked to the extinction of a group of mammals known as the megafauna in those environments.[144] The presence of humans changed the environmental pressures on such organisms around the world.



Genetic drift: Genetic drift is a stochastic process that becomes important in small populations or over long periods of time. It can lead to non-adaptive evolutionary phenomenon that explain a number of observations. Consider the observation that many primates are strictly dependent on the presence of vitamin C (ascorbic acid) in their diet. Primates are divided into two suborders, the Haplorhini, from the Greek meaning “dry noses”, and the Strepsirrhini, meaning “wet noses”. The

[143] Mobile elements reveal small population size in the ancient ancestors of Homo sapiens:

[144] Megafauna extinction effects and an interesting video

Strepsirrhini include the lemurs and lorices, while the Haplorhini include the tarsiers and the anthropoids, monkeys, apes, and humans. The Haplorhini, but not the Strepsirrhini, all share a requirement for vitamin C in their diet. In vertebrates, vitamin C plays an essential role in the synthesis of collagen, a protein involved in the structural integrity of a wide range of tissues. In vitamin C-dependent organisms the absence of dietary vitamin C leads to the disease scurvy, which according to Wikipedia, "often presents itself initially as symptoms of malaise and lethargy, followed by formation of spots on the skin, spongy gums, and bleeding from mucous membranes. Spots are most abundant on the thighs and legs, and a person with the ailment looks pale, feels depressed, and is partially immobilized. As scurvy advances, there can be open, suppurating wounds, loss of teeth, jaundice, fever, neuropathy, and death."[145]

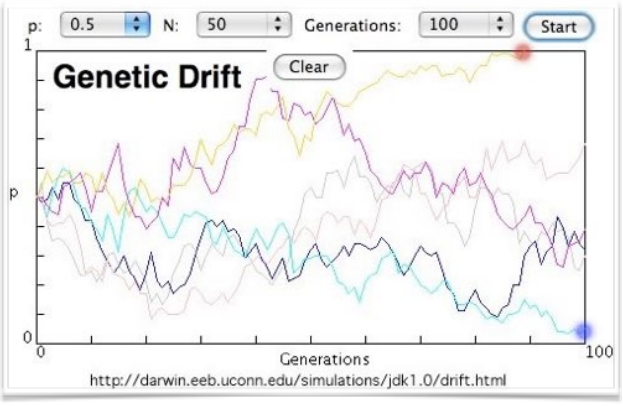
The requirement for dietary vitamin C in the *Haplorhini* is due to a mutation in the *GULO1* gene, which encodes the enzyme 1-gulono-gamma-lactone oxidase (Gulo1) required for the synthesis of vitamin C. One can show that the absence of a functional *GULO1* gene is the root cause of vitamin C dependence in Haplorhini by putting a working copy of the gene, for example derived from a mouse, into human cells. The mouse-derived *GULO1* allele, which encodes a functional form of the Gulo1 enzyme, "cures" the human cells' of their need for exogenous vitamin C. But, no matter how advantageous a working *GULO1* allele might be, particularly for British sailors, who died in large numbers before a preventative treatment for scurvy was discovered[146], no new, functional *GULO1* allele has appeared in the lineage leading to humans or the other Haplorhini, an example of the fact that it is easier to break something than to fix it through random changes. Since mutation is a stochastic process, organisms do not always produce the genes or alleles they "need" or that might be beneficial. Alleles are selected from alleles already present in the population or that appear through *de novo* (new) mutations. In some cases there may be no plausible molecular pathway that can generate such an allele (or such a gene).

The mutant *GULO1* allele appears to have become "fixed", that is the only *GULO1* allele present in the ancestral population that gave rise to the Haplorhini, around ~40 million years ago. So the question is, how did we (that is our ancestors) come to lose a functional version of such an important gene? It seems obvious that when the non-functional allele became fixed in that population, the inability to make vitamin C cannot have been strongly selected against, its loss would appear to have led to little or no effect on reproductive success. We can imagine such an environment and associated behavior; namely, we suspect that these organisms obtained sufficient vitamin C from their diet, so that the loss of their ability to synthesize vitamin C had little if any negative effect on them.

So how was the functional *GULO1* gene lost? We might never know for sure, but we can speculate. In small populations, non-adaptive, that is, non-beneficial and even mildly deleterious genotypic changes can increase in frequency through genetic drift. In such populations, selection continues to be active, but it has significant effects only when a trait and the alleles that produce it strongly influence reproductive success. In asexual populations genetic drift is due to random effects on organismic survival that can, in practice be difficult to distinguish from selective effects. In contrast, drift is unavoidable in small populations of sexually reproducing organisms. This is because cells known as gametes are produced during the process of sexual reproduction (Chapter 4). While the cell that generates these gametes contains two copies of each gene, and each gene reflects one of the alleles present within the population, any particular gamete contains only a single (and possibly new) allele of each gene. Two gametes then fuse to produce a new diploid organism. This process combines a number of chance events: including which allele is present in a particular gamete and which gametes fuse to produce a new organism. Not all gametes produced form a new organism. In a small population, over a reasonably small number of generations, one of multiple

[145] An amazing fact is that it took the deaths of thousands of sailors to understand the nutritional role of vitamin C.

[146] http://mentalfloss.com/article/24149/how-scurvy-was-cured-then-cure-was-lost



alleles at a particular genetic locus may be lost simply by chance. In this figure (→), six distinct experimental outcomes (each line) were analyzed over the course of 100 generations. The population originally contained two different alleles of a particular gene, present in equal numbers, and the population is set to 50 individuals. While we are tracking only one genetic locus, the same type of behavior impacts every gene for which multiple alleles are present. In two of these six populations, one (red dot) or the other allele (blue dot) has been lost or is close to being lost. When a particular allele becomes the only allele within a population, it is said to have been fixed. Assume that the two alleles convey no selective advantage with respect to one another, can you predict what will happen if we let the experiment run through 10,000 generations? For the mathematically inclined, it is possible to estimate the effects of mild to moderate positive or negative selective pressures on allele frequencies and the probability that a particular allele will be lost or fixed through genetic drift.

Since the rest of the organism's genotype can influence the phenotype associated with a particular allele, the presence or absence of various alleles within the population can influence the phenotypes observed (a topic we will return to in chapter 12). If an allele disappears because of genetic drift, future evolutionary changes may be constrained, or perhaps better put, redirected. At each point, the future directions open to evolutionary mechanisms depend in large measure on the alleles currently present in the population. Of course new alleles continue to arise by mutation, but they are originally infrequent, just one of each in the entire population, so unless they are strongly selected for (and even if they are selected for) they may be lost from the population by genetic drift.[147] Drift can lead to some weird outcomes. For example, what happens if drift leads to the fixation of a mildly deleterious allele, let us call this allele BBY. Now the presence of BBY will change the selective landscape: mutations and or alleles that ameliorate the negative effects of BBY will increase reproductive success, selection pressures will favor those alleles. This can lead to evolution changing direction even if only subtly. With similar effects going on across the genome, one quickly begins to understand why evolution is something like a drunken walk across a selective landscape, with genetic drift, founder and bottleneck effects resulting in periodic stochastic staggers in new directions. In fact this can be beneficial, these phenotypic variants enable the population to sample the range of accessible variations, and "select" those that work best (at least in terms of short term reproductive advantage).

The use of pre-existing variation, rather than the idea that an organism invents genetic variations as they are required, was a key point in Darwin's view of evolutionary processes. There is no known mechanism by which organisms can create the alleles they need or "want", generally no simple link between a particular genetic variation (allele) and a specific phenotype. Rather, the allelic variation generated by mutation, selection, and drift are all that evolutionary processes have to work with.[148] Only a very rare mutation that recreates (resurrects or fixes) a lost allele can bring an allele back into the population once it has been lost. Founder and bottleneck effects, together with genetic drift combine to produce what are known as non-adaptive processes and make the history of a population a critical determinant of its future evolution.

---

[147] If the population is small, instead of disappearing, any particular mutation (allele) could become fixed through genetic drift - use the genetic drift applet and look for examples where an allele almost disappears and then becomes fixed; it does happen.

[148] An exception involves the process known as horizontal gene transfer. Viruses also contain genes that they can transfer from organism to organism. We will consider both processes later on.

Questions to answer:

32. How does the extinction of one type of organism influence the evolution of others?
33. What factors make a bottleneck different from a founder effect?
34. How can a founder effect/bottleneck lead to deleterious alleles becoming more frequent in a population? How might the presence of such alleles impact future evolution?
35. How does natural selection influence the effects of genetic drift and vice versa?
36. Describe the relative effects of selection and drift following a bottleneck.
37. How is it that drift (the probability of allele loss) can be accurately quantified, but is unpredictable in any particular population?

Questions to ponder:

- How is determining allele frequency in a population similar to and different from political polling?
- Does passing through a bottleneck improve or hamper a population's chances for evolutionary success?

A reflection on the complexity of phenotypic traits

We can classify traits into three general types: adaptive, non-adaptive, and deleterious. Adaptive traits are those that, when present increase the organism's reproductive success. These are the traits we normally think of when we think about evolutionary processes. Non-adaptive traits are those generated by stochastic processes, like drift, founder effects, and bottlenecks. These traits become established not because they improve reproductive success but simply because they happened to have become fixed within the population. If an allele is deleterious independent of its environment, it will be expected to rapidly disappear from the population, unless other factors are in play. Rare, strongly deleterious alleles are, most likely, the result of new mutations, or they led to a selective advantage in specific situations.

When we consider a deleterious allele we are always referring to its effects on reproductive success. An allele can harm the individual organism carrying it yet persist in the population because it improves reproductive success, that is, it leads to an increased number of viable offspring. Similarly, there are traits that can be seen as actively maladaptive, but which occur within the population because they are linked mechanistically to some other positively selected trait. Many genes are involved in a number of distinct processes and their alleles can lead to multiple phenotypic effects. Such alleles are said to be pleiotropic, meaning they have multiple effects. Not all of the pleiotropic effects of an allele are necessarily of the same type; some can be beneficial, others deleterious. As an example, a trait that dramatically increases the survival of the young, and so increases their potential reproductive success, but leads to senility and sudden death in older adults could well be positively selected for. In this scenario, the senility/death trait is highly maladaptive but is not eliminated by selection because it is mechanistically associated with the highly adaptive juvenile survival trait. What is happening is a form of cost-benefit analysis. If the net evolutionary benefits of an allele exceeds its costs, the allele and the trait associated with it will be subject to positive selection. If the costs exceed the benefits, it will be selected against. It is worth noting that a trait that is advantageous in one environment may be disadvantageous in another, think the effects of diet on the effects of the *GULO1* mutation. All of which is to say that when thinking about evolutionary mechanisms, do not assume that a particular trait exists independently of other traits, that it functions in the same way in all environments, or that the presence of a trait is evidence that it is beneficial.

Gene linkage: one more complication

So far, we have not worried overly much about the organization of genes in an organism. We also have not consider what, exactly a gene is. For now, let us just say that a gene is information encoded within a region of a molecule of DNA (deoxyribonucleic acid) and that multiple genes can be found within a single DNA molecule – we will consider specific aspects of genes below and then again in greater detail in the sections on genetics (Chapter 7).

It could be that each gene behaves like an isolated object, but in fact that is not the case. We bring it up here because the way genes are organized can, in fact, influence evolutionary processes. In his original genetic analyses, Gregor Mendel (1822-1884) spent a fair amount of time looking for "well behaved" genes and alleles, those that displayed simple recessive and dominant behaviors and that acted as if they were independent from one another.[149] In fact, as noted by Kampourakis, "Weldon's (1902) studies of varieties of pea hybrids led him to conclude that there was a continuum of colors from greenish yellow to yellowish green, as well as a continuum of shapes from smooth to wrinkled. It thus appeared that in obtaining purebred plants for his experiments, Mendel had actually eliminated all natural variation in peas, and that characteristics were not as discontinuous as he had assumed". The situation is even more complex for most traits, and the genes that influence them. Traits are rarely dichotomous (one or the other), and often influenced by multiple genes. Genes often act as if they are linked together, because often they are. Gene linkage arises from the organization of genes within chromosomes, that is individual DNA molecules. So what happens to linked genes when a particular allele of a particular gene is strongly selected for or against? That allele, together with alleles found in linked genes, are also selected. We can think of this as a "by-stander" or a "piggy-back" effect, where an allele's frequency in a population increases (or decreases) not because of its direct effects on reproductive success, but because of its location within the genome, its "linkage" to an allele that strongly influences selection.

As we will see later on, linkage between alleles (or between genes) is not a permanent situation; there are processes (meiotic recombination) that can shuffle the alleles on a chromosome. The end result of such recombination events is that the further away two genes are from one another on a DNA molecule (a chromosome), the more likely it is that alleles of those genes will appear to be unlinked, that is, have independent effects on reproductive success. Over time, the effects of linkage will eventually be lost, but not necessarily before particular alleles have been fixed, and other alleles lost, within the population. For example, extremely strong selection for a particular allele of one gene can lead to the fixation of mildly deleterious alleles in closely linked (neighboring) genes.

At this point, let us clarify some terms related to genes. These terms arise from the history of biology in general, and genetics in particular. We now know that genetic information is stored in the sequence of double-stranded DNA molecules. A gene is the region of a DNA molecule that encodes a particular "gene product", either an RNA molecule or a polypeptide, together with regions of the DNA molecule required for the gene product to be "expressed", a term that captures the ability of the gene product to be made and used (that is, to impact the cell/organism within which the gene is located). Where and when a gene is expressed is regulated by networks of interacting molecules. All of the DNA molecules present in a cell are known collectively as the cell's genome. We refer to the position of a particular gene within the genome as a genetic locus (plural, loci). In Latin locus means 'place'; think location – a word derived from the same root. A particular genetic locus (gene) can be occupied by any of a number of distinct alleles (DNA sequences). There are various mechanisms that can duplicate, delete, insert, or move a region of DNA within the genome, creating (or eliminating) new genetic loci. The phenotype associated with an allele is influenced by its position within a genetic locus, as well as the rest of the genome.

It is worth noting that the combination of non-adaptive, non-selective processes can lead to the appearance and maintenance of mildly dis-advantageous (deleterious) traits within a population. Similarly, a trait that increases reproductive success, by increasing the number of surviving offspring, may be associated with other not-so-beneficial, and sometime seriously detrimental (to individuals) effects. The key is to remember that evolutionary mechanisms do not necessarily result in what is best for an individual organism but what in the end enhances net (short term) reproductive success of a population. Evolutionary processes do not select for particular genes or new versions of genes but rather for those combinations of alleles that optimize reproductive success. The situation gets more complicated when evolutionary mechanisms generate organisms, like humans, who think and

[149] Mendelian controversies: a botanical and historical review

feel and can actively object to the outcomes of evolutionary processes. From the point of view of self-conscious organisms, evolution can appear cruel, or at the very least totally uninterested in, and apathetic towards the desires and happiness of individuals. This was one reason that Darwin preferred impersonal (naturalistic) mechanisms over the idea of a God responsible for what can appear to be the gratuitously cruel aspects of their creation.

Questions to answer:

39. How might the linkage of genes along a chromosome influence evolutionary processes?
40. How might interactions between alleles on different chromosomes influence evolutionary processes?
41. What, exactly, is the difference between a gene and an allele? a gene and a chromosome?
42. Consider this quote from Charles Darwin, "Natural selection will never produce in a being any structure more injurious than beneficial to that being, for natural selection acts solely by and for the good of each." How would you modify it in light of our modern understanding of evolutionary mechanisms?

Question to ponder:

- How does evolution's focus on reproductive success, and cost-benefit analysis, rather than individual well-being impact the view that the natural is inherently good (or is it irrelevant)?

Speciation & extinction

As we have noted, an important observation that needs to be explained is why, exactly, are there so many (millions) of different types of organisms. The Theory of Evolution explains this observation through the process of speciation. The basic idea is that populations of organisms can split into distinct groups. Over time evolutionary mechanisms acting on these populations produce distinct types of organisms, that is, different species. At the same time, we know from the fossil record and from modern experiences, that types and groups of organisms can disappear – they can become extinct. What leads to the formation of new species or the disappearance of existing ones?

To answer these questions, we have to consider how populations behave. A population of a particular type of organism will typically inhabit a particular geographical region. The size of these regions can range from over an entire continent or more, to a small limited region, such as a single isolated lake. Moreover, when we consider organisms that reproduce in a sexual manner, which involves a degree of cooperation between individuals, we have to consider how far a particular organism (or its gametes) can travel. The reproductive range of some organisms is quite limited, whereas others can travel significant distances. Another factor to consider is how an organism makes its living - where does it get the matter and energy (that is, food) and space it needs to successfully reproduce? Together these are referred to as a specific specie's (population's) ecological niche.

An organism's ecological niche is the result of its past evolutionary history, past selection pressures acting within a particular environment, and its current behavior. In a stable environment, and a large enough population, reproductive success will reflect how effectively organisms exploit their ecological niche. Over time, stabilizing selection will tend to optimize individual organisms' adaptation to its niche. At the same time, it is possible that different types of organisms will compete for similar resources, for a similar niche. This interspecies competition leads to a new form of selective pressure. If individuals of one population can exploit a different set of resources or the same resources differently, these organisms can minimize competition with other species and become more reproductively successful compared to individuals that continue to compete directly with other species. This can lead to a number of outcomes. In one case, one species becomes much better than others at occupying a particular niche, driving the others to extinction. Alternatively, one species may find a way to occupy a new or

> *So, naturalists observe, a flea has smaller fleas that on him prey; and these have smaller still to bite 'em; and so proceed ad infinitum.*
> *- Jonathan Swift*

related niche, and within that particular niche, it can more effectively compete, so that the two species come to occupy distinct niches. Finally, one of the species may be unable to reproduce successfully in the presence of the other and become (at least locally) extinct.

These scenarios are captured by what is known as the competitive exclusion principle or Gause's Law, which states that two species cannot stably occupy the same ecological niche (something similar to the Pauli exclusion principle in Quantum Mechanics) – over time either one will leave (or rather be forced out) of the niche, or will evolve to fill a different, often subtly different niche.[150] What is sometimes hard to appreciate is how specific a viable ecological niche can be. For example, consider the situations described by the evolutionary biologist Theodosius Dobzhansky (1900-1975): "*Some organisms are amazingly specialized. Perhaps the narrowest ecologic niche of all is that of a species of the fungus family Laboulbeniaceae, which grows exclusively on the rear portion of the elytra (the wing cover) of the beetle Aphenops cronei, which is found only in some limestone caves in southern France. Larvae of the fly Psilopa petrolei develop in seepages of crude oil in California oilfields; as far as is known they occur nowhere else.*"

While it is tempting to think of ecological niches in broad terms, the fact is that subtle environmental differences can favor specific traits and specific organisms. If an organism's range is large enough and each individual's range is limited, distinct traits can be prominent in different regions of the species' range. These different subpopulations[151] reflect local adaptations. For example, it is thought that as human populations migrated out of the equatorial regions of Africa, they were subject to differential selection based on exposure to sunlight, due in part to the role of sunlight in the synthesis of vitamin D and its ability to induce cancer-causing mutations and skin damage (sun burn).[152] In their original ecological niche, the ancestors of humans were thought to hunt in the open savannah (rather than within forests), and so developed adaptations to control body temperature. Our general lack of body hair and ability to sweat compared to other mammals are thought to be such adaptations.

The absence of a thick coat of hair also allowed direct exposure to UV-light from the sun. While UV exposure is critical for the synthesis of vitamin D, too much exposure can lead to skin cancer. Dark skin pigmentation is thought to be an adaptive compromise. As human populations moved away from the equator, the dangers of UV exposure decreased while the need for vitamin D production remained. Under such conditions, allelic variations that favored lighter skin pigmentation, but retained the ability to tan to some extent appears to have been selected (←). Genetic analyses of different populations have begun to reveal exactly which alleles in which genes emerged in different human populations as they migrated out of Africa and across the Earth. Of course, with humans the situation has an added level of complexity. For example, the (relatively recent) trait of wearing clothing directly impacts the pressure of "solar selection." And some pinker folk favor darker (tanned) skin. A number of different phenotypic variations can occur over the geographical range of a species. Differences in climatic conditions, pathogens, predators, and prey can all lead to multiple local adaptations, like those associated with human skin color.

---

[150] Competitive exclusion principle

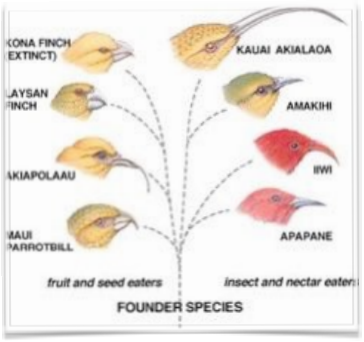
[151] Sometimes sub or local populations are termed subspecies or races. One can (and we will) argue that the term race is obsolete and used to justify group prejudices. Here is a jump point on this topic: Avoiding unrecognized racist implications arising from teaching genetics.

[152] Genetics of skin color: image sources: http://hmg.oxfordjournals.org/content/18/R1/R9.full

Mechanisms of speciation

So now we consider the various mechanisms that can lead a species to give rise to one or more new species. Remembering that species, at least species that reproduce sexually, are defined by the fact that they can and do interbreed to produce fertile offspring, you might already be able to propose a few plausible scenarios. An important point is that the process of speciation is continuous, there is generally no magic moment when one species becomes another, rather a new species emerges over time from a pre-existing species, after which the two populations evolve independently.[153] The origin of species through evolutionary mechanisms is therefore formally analogous to the Cell Theory, where each cell is derived from a pre-existing cell – the difference is that the process of cell division results in a unambiguous benchmark in the history of a cell. The situation is more ambiguous in organisms that reproduce asexually, but we will ignore that for the moment. More generally, species are populations of organisms at a moment in time, they are connected to past species and can produce new species in the future (or go extinct).

Perhaps the simplest way that a new species can form is if the original population is physically divided into isolated subpopulations. This is termed allopatric speciation. By isolated, we mean that individuals of the two subpopulations no longer mingle with one another, they are restricted to specific geographical areas. That also means that they are no longer interact with one another, and so interbreeding does not occur. If we assume that the environments inhabited by the subpopulations are distinct and that they represent distinct sets of occupied and available ecological niches, distinct climate and geographical features, and distinct predators, prey, and pathogens, then these isolated subpopulations will be subject to different selection pressures leading to different phenotypes. Assuming that the physical separation between the populations is stable, and persists over a sufficient period of time, the populations will diverge. Both selective and non-selective processes drive this divergence, which will be influenced by what mutations arise and give rise to the range of alleles present within the populations. The end result will be populations adapted to specific ecological niches, which may well be different from the niche of the parental population. For example, it is possible that while the parental population was a generalist, occupying a broad range of ecological niches, the subpopulations may be specialized to specific niches. Consider the situation with various finches (honeycreepers) found in the Hawai'ian islands.[154] Derived from an ancestral founder population, these organisms have adapted to a number of highly specialized niches. Their specializations give them a competitive edge with respect to one another in feeding off particular types of flowers. As they specialize, however, they become more dependent upon the continued existence of their host flower or flower type (→). It is a little like the fungus that can only grow on one particular place on a particular type of beetle. We begin to understand why the drive to occupy a particular ecological niche also leads to vulnerability, if the niche disappears, a species highly adapted to exploit it may not be able to effectively and competitively exploit other niches, leading to its extinction.[155]



It is a sobering thought that current estimates are that greater that ~98% of all species that have or now live on Earth are extinct, presumably due in large measure in changes in, or the disappearance of, their niches. You might speculate (and provide a plausible argument to support your speculation) as to which of the honeycreepers illustrated above would be most likely to become

---

[153] An interesting exception occurs in some plants (which can self-fertilize), where there are instances new species formed in one generation due to changes in ploidy: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2442920/

[154] Hawaiian honeycreepers and their tangled evolutionary tree

[155] A great video of organisms that have survived (often with human help) the extinction their partners: The Ghosts of Evolution: Nonsensical fruit, missing partners, and other ecological anachronisms

extinct in response to environmental changes.[156] In a complementary way, the migration of organisms into a new environment can produce a range of effects as the competition for existing ecological niches get resolved.[157] If an organism influences its environment, the effects can be complex. As noted earlier, a profound and global example is provided by the appearance, early in the history of life on Earth, of photosynthetic organisms that released molecular oxygen (O_2) into the atmosphere as a waste product. Because of its chemical reactivity, the accumulation of molecular oxygen led to loss of some ecological niches and the creation of new ones. The recent anthropogenic increase in atmospheric CO_2 concentration is such an example. While dramatic, similar events occur on more modest levels all of the time. It turns out that extinction is a fact of life – at the same time, life has continued and diversified in an uninterrupted manner for over ~3,500,000,000 years.

Gradual or sudden environmental changes, ranging from the activity of the sun, to the drift of continents and the impacts of meteors and comets, lead to the disappearance of existing ecological niches and the appearance of new ones. For example, the collision of the continents with one another leads to the formation of mountain ranges and regions of intense volcanic activity, both of which can influence climate and the connectedness of populations. There have been periods when Earth appears to have been completely or almost completely frozen over.[158] These geological processes continue to be active today, with the Atlantic ocean growing wider and the Pacific ocean shrinking, the splitting of Africa along the Great Rift Valley, and the ongoing collision of India with the rest of Asia. As continents move and sea levels change, organisms that evolved on one continent may be able to migrate into another. All of these processes combine to lead to extinctions, which open ecological niches for new organisms, and so it goes.

At this point you should be able to appreciate the fact that evolution never actually stops. Aside from various environmental factors, each species is part of the environment of other species. Changes in one species can have dramatic impacts on others as the selective landscape changes. An obvious example is the interrelationship between predators, pathogens, and prey. Which organisms survive to reproduce will be determined in large part by their ability to avoid predators or recover from infection. Certain traits may make the prey more or less likely to avoid, elude, repulse, discourage, or escape a predator's attack. As the prey population evolves in response to a specific predator or pathogen, these changes will impact the predator or pathogen, which will also have to adapt. This situation is often call the Red Queen hypothesis (→), and it has been invoked as a major driver for the evolution of sexual reproduction, which we will consider in greater detail as we go on.[159]

As the Red Queen said to Alice ... "Here, you see, it takes all the running you can do to keep in the same place"
-Lewis Carroll, Though the Looking Glass

Isolating mechanisms: Think about a population that is on its way to becoming specialized to fill a particular ecological niche. What is the effect of cross breeding with a population that is, perhaps, on an path to another adapting to another ecological niche? Most likely the offspring will be poorly adapted to either niche. This leads to a new selective pressure, selection against cross-breeding between individuals of the two populations. Even small changes in a particular trait or behavior can lead to significant changes in mating preferences and outcomes. Consider Darwin's finches or Hawaiian honeycreepers. A major feature that distinguishes these various types of birds is the size and shapes of their beaks. These adaptations represent both the development of a behavior – that is

---

[156] The Perils of Picky Eating: Dietary Breadth Is Related to Extinction Risk in Insectivorous Bats

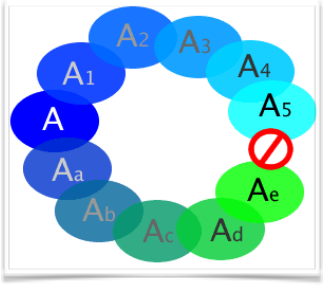
[157] Humans spread through South America like an invasive species

[158] One "snowball Earth" period appears to have been involved in the emergence of macroscopic multicellular life.

[159] Running with the Red Queen: the role of biotic conflicts in evolution

the preference of birds to seek food from particular sources, for example, particular types of flowers or particular size seeds – and the traits needed to successfully harvest that food source, such as bill shape and size. Clearly the organism has to display the behavior, even if it is in a primitive form, that makes selection of the physical trait beneficial. This is a type of loop, where behavioral and physical traits are closely linked. You can ask yourself, thinking about the ancestor of giraffes, could a long neck have evolved if members of the ancestral population did not eat the leaves of trees?

Back to finches and honeycreepers. Mate selection in birds is often mediated by song, generally males sing and females respond (or not). As beak size and shape changes, the song produced also changes.[160] This change is, at least originally, an unselected trait that accompanies the change in beak shape. It can become a selected trait if females recognize and respond to songs more like their own. This would lead to preferential mating between organisms with the same trait (beak shape). Over time, this preference could evolve into a stronger and stronger mating preference, until it becomes a reproductive barrier between organisms adapted to different ecological niches.[161] Similarly, imagine that the flowers that a particular subpopulation feeds on open and close at different times of the day. This could influence when an organism is active and sexually receptive. You can probably generate your own scenarios in which one behavioral trait has an influence on reproductive preferences and success. If a population is isolated from others, such effects may develop but are irrelevant; they become important only when two closely related but phenotypically distinct populations come back into contact. Now matings between individuals in two different populations, sometimes termed hybridization, can lead to offspring poorly adapted to either niche. This can create a selective pressure to minimize hybridization. Again, the reproductive isolation of two populations can arise spontaneously, such as when two populations mate at different times of the day or the year or respond to different behavioral queues, such as mating songs. Traits that enhance reproductive success by reducing the chance of detrimental hybridization will be preferentially selected. The end result is what is known as reproductive isolation.[162] As reproductive isolation occurs, what was one species becomes two. A number of different mechanisms ranging from the behavioral to the structural and the molecular are involved in generating reproductive isolation. Behaviors may not be "attractive," genitalia may not fit together,[163] gametes might not recognize and fuse with one another, or embryos might not be viable - there are many possibilities.



Ring species: Ring species demonstrate a version of allopatric speciation. Imagine populations of the species A. Over the geographic range of A there exist a number of subpopulations. These subpopulations (A_1 to A_5) and (A_a to A_e) have limited regions of overlap with one another but where they overlap they interbreed successfully (→). But populations A_5 and A_e no longer interbreed successfully – are these populations separate species? In this case, there is no unambiguous answer (and sometimes we have to get used to the idea of ambiguity, something that should be more widely appreciated). In part this ambiguity is a basic biological trait, populations are continuous over time, but individuals within a population vary, and it is that variation that leads to evolutionary change. In the real world, ring species are unlikely - it is more likely that that over time the links between the various subpopulations will be broken and one or more species may arise.

---

[160] A good background article on Darin's finches and speciation is here: Sisyphean evolution

[161] Beaks, Adaptation, and Vocal Evolution in Darwin's Finches & Vocal mechanics in Darwin's finches: correlation of beak gape and song frequency

[162] Beak size matters for finches' song: http://news.nationalgeographic.com/news/2004/08/0827_040827_darwins_finch.html

[163] Causes and Consequences of Genital Evolution: http://icb.oxfordjournals.org/content/early/2016/09/13/icb.icw101.abstract

Consider the black bear *Ursus americanus*. Originally distributed across all of North America, its distribution is now much more fragmented. Isolated populations are free to adapt to their own particular environments and migration between populations is limited. Clearly the environment in Florida is different from that in Mexico, Alaska, or Newfoundland. Different environments will favor different adaptations. If, over time, these populations were to come back into contact with one another, they might or might not be able to interbreed successfully - reproductive isolation may occur and one species may become many.

While the logic and mechanisms of allopatric speciation are relatively easy to grasp (we hope), there is a second type of speciation, known as sympatric speciation, that was originally more controversial. It occurs when a single population of organisms splits into two reproductively isolated communities within the same physical region. How could this possibly occur? What stops (or inhibits) the distinct sub-populations from inbreeding; how can these subpopulations become reproductively isolated? Recently a number of plausible mechanisms have been identified. One involves host selection.[164] In host selection, animals (such as insects) that feed off a specific host may find themselves reproducing in distinct zones associated with their hosts. For example, organisms that prefer blueberries may mate in a different place, time of day, or time of year than those that prefer raspberries. There are blueberry- and raspberry-specific niches, and organisms that specialize to one or the other may have a reproductive advantage when they restrict themselves to that food source. Through a process of disruptive selection (see above), organisms that live primarily on one particular plant (or part of a plant) can be subject to different selective pressures. Reproductive isolation will enable the populations to "stay focussed" and so adapt more rapidly. Mutations that reinforce an initial, perhaps weak, mating preference can lead to reproductive isolation - this is a simple form of sexual selection, which we will discuss soon.[165] One population has become two distinct, reproductively independent populations, one species has become two.

Questions to answer:

43. What is involved in establishing reproductive isolation between populations (species formation); what factors favor speciation?
44. How are sympatric and allopatric speciation the same and how do they differ?
45. Describe the (Darwinian) cycle of selection associated with the development of a trait, such as the extended neck of giraffes. Consider the feedback between behavior and anatomy.

Questions to ponder:

- How would you determine whether two species are part of the same genus?
- How might asexual organism be assigned to specific species?
- How might you decide whether an organism, identified through fossil evidence, was part of a extant species?

Signs of evolution: homology and convergence

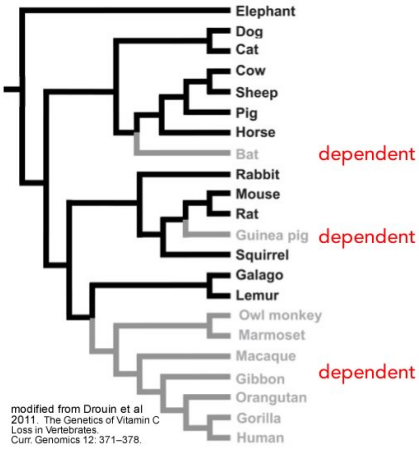
When we compare two different types of organisms we often find traits that are similar. On the basis of evolutionary theory, these traits can arise through either of two processes: the trait could have been present in the ancestral population that gave rise to the two species or the two species could have developed their versions of the trait independently. In this latter case, the trait was not present in the last common ancestor shared by the organism. Where a trait was present in the ancestral species it is said to be a homologous trait. If the trait was not present in the ancestral species but appeared independently within the two lineages, it is known as an analogous trait that arose through convergent evolution.

---

[164] Sympatric speciation by sexual selection & Sympatric speciation in phytophagous insects: moving beyond controversy?

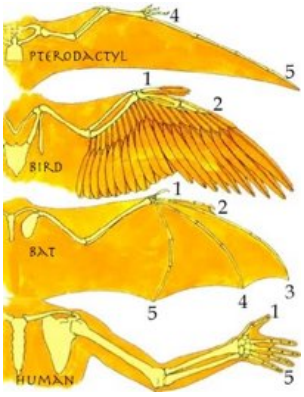
[165] The sexual selection: http://www.youtube.com/watch?v=JakdRczkmNo

For example, consider the trait of vitamin C dependence, found in Haplorrhini primates and discussed above. Based on a number of lines of evidence, we conclude that the ancestor of all Haplorhini primates was vitamin C dependent and that vitamin C dependence in Haplorhini primates is a homologous trait. On the other hand Guinea pigs (*Cavia porcellus*), which are in the order Rodentia, are also vitamin C dependent, but other rodents are not (→).[166] It is estimated that the common ancestor of primates and rodents lived more than ~80 million years ago, that is, well before the common ancestor of the Haplorhini. Given that most rodentia are vitamin C independent, we can assume that the common ancestor of the rodent/primate lineages was itself vitamin C independent. We conclude that vitamin C dependence in Guinea pigs and Halporhini (and bats) are analogous traits, they arose as the result of independent events. If we looked at the molecular details, we would not be surprised to discover different mechanisms (different genomic changes) leading to vitamin C dependence in the two groups.



Question at answer:
46. How would you decide whether vitamin C dependence in Haplorhini and guinea pigs (and bats) were independent events?

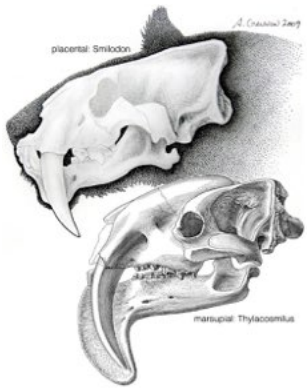
As we consider traits in detail, we have to look carefully, structurally, and more and more frequently, molecularly, that is, directly at the genotype, to determine at least tentatively whether they are homologous or analogous - the result of evolutionary convergence or ancestry. Consider the flying vertebrates. The physics of flight, and many other behaviors that organisms perform, are constant. Organisms of similar size face the same aerodynamic and thermodynamic constraints. In general there are only a limited number of physically workable solutions to deal with these constraints. Under these conditions different populations that are in a position to exploit the benefits of flight will, through the process of variation and selection, end up with structurally similar solutions. This process is known as convergent evolution. Convergent evolution occurs when only certain solutions to a particular problem are evolutionarily accessible.



Consider the wing of a pterodactyl, which is an extinct flying reptile, a bird, and a bat, a flying mammal (←). These organisms are all tetrapod (four legged) vertebrates – their common ancestor had a structurally similar forelimb, so their forelimbs are clearly homologous. Therefore this evolutionary adaptation, using the forelimb for flight, began from a structurally similar starting point. But most tetrapod vertebrates do not fly, and forelimbs have become adapted to many different functions. An analysis of tetrapod vertebrate wings indicates that each took a distinctly different approach to generating wings. In the pterodactyl, the wing membrane is supported by the 5th finger of the forelimb, in the bird by the 2nd finger, and in the bat, by the 3rd, 4th and 5th fingers. The wings of pterodactyls, birds, and bats are analogous structures, while their forelimbs are homologous.

As another example of evolutionary convergence consider teeth. The use of a dagger is an effective solution to the problem of killing another organism. Variations of this solution have been discovered or invented independently many times. Morphologically similar dagger-like teeth have evolved independently, that is, from ancestors without such teeth, in a wide range of distinct lineages. Consider, the placental mammal *Smilodon* and the marsupial mammal *Thyacosmilus*; both

[166] see Drouin et al., 2011. "The genetics of vitamin C loss in vertebrates."



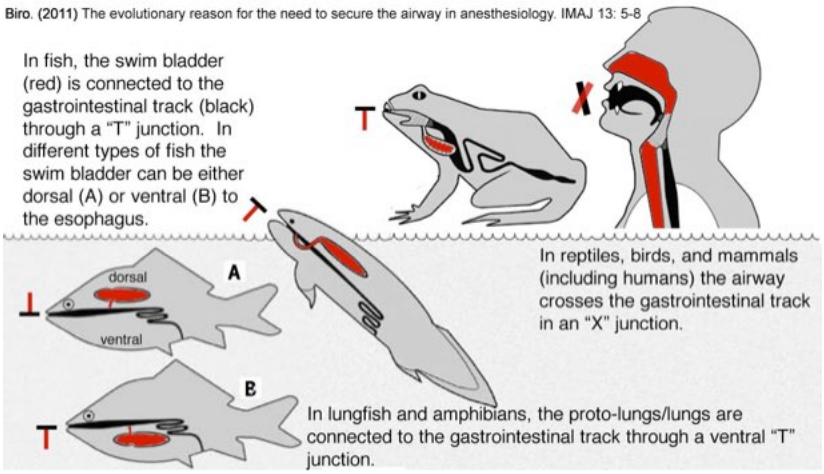
have similarly-shaped highly elongated canine teeth (→). Marsupial and placental mammals diverged from a common ancestor ~160 million years ago and this common ancestor, like most mammals, appears to have lacked such dagger-like teeth. While teeth are a homologous feature of *Smilodon* and *Thyacosmilus*, elongated dagger-like teeth are analogous structures, the result of convergent evolution.

Recognizing phylogenic relationships: A major challenge when trying to determine a plausible relationship between organisms based on anatomy has been to distinguish homologous from convergent (analogous) traits. Homologous traits, known as synapomorphies, are the basis of placing organisms together within a common group. In contrast, convergent traits are independent solutions to a similar problem, and so are irrelevant when it comes to defining evolutionary relationships. It is, however, also true that evolution can lead to the loss of traits; this can confuse or complicate the positioning of an organism in a classification scheme. It is worth noting that very often developing a particular trait, whether it is an enzyme or an eye, requires energy. If the trait does not contribute to an organism's reproductive success it will not be selected for; on the other hand, if it is expensive to build, but has no useful function, its loss may be selected for. As organisms adapt to a specific environment and lifestyle, traits once useful can become irrelevant or distracting, and may be lost. A classic example is the reduction of hind limbs during the evolution of whales [↓]. Another is the common loss of eyes often seen as populations adapt to environments in which light is absent. The most dramatic cases of loss involve organisms that become obligate parasites of other organisms. In many cases, these parasitic organisms are completely dependent on their hosts for many essential functions, this allows them to become quite simplified even though they are in fact highly evolved. For example, they lose many genes as they become dependent upon the host. The loss of traits can itself be an adaptation if it provides an advantage to organisms living in a particular environment. This fact can make it difficult to determine whether an organism is primitive (that is, retains ancestral features) or highly evolved.

Evolution is an ongoing experiment in which random mutations are selected based on the effects of their resulting phenotypes on reproductive success. As we have discussed, various non-adaptive processes are also involved, which can impact evolutionary trajectories. The end result is that adaptations are based on past selective pressures and i) are rarely perfect and ii) may actually have become outdated, if the environment the organisms live in has changed. One wants to keep this in mind when one considers the differences associated with living in small groups in a pre-technological world on the African savannah and living in New York City. In any case, evolution is not a designed process that reflects a predetermined goal but involves responses to current constraints and opportunities - it is a type of tinkering in which selective and non-selective processes interact with pre-existing organismic behaviors and structures and is constrained by those behaviors and structures, as well as by cost and benefits associated with various traits and their effects on reproductive success.[167] What evolution can produce depends on the alleles present in the population, or those that can be generated by mutation, and the current form of the organism. Not all desirable phenotypes (that is, those leading to improved reproductive success) may be accessible from a particular genotype, and even if they are, the cost of attaining a particular adaptation, no matter how desirable to an individual, may not be repaid by the reproductive advantage it provides within a population.

[167] Evolutionary tinkering: Jacob 1977

As an example, our ability to choke on food could be considered a serious design flaw, but it is the result of the evolutionary path that produced us, a path that led to the crossing of our upper airway (leading to the lungs) and our pharynx (leading to our gastrointestinal system). That is why food can lodge in the airway, causing choking or death. It is possible that the costs of a particular "imperfect" evolutionary design are offset by other advantages (→). For example, the small but significant possibility of death by choking may, in an evolutionary sense, be worth the ability to make more complex sounds (speech) involved in social communication.[168]



As a general rule, evolutionary processes generate structures and behaviors that are as good as they need to be for an organism to effectively exploit a specific set of environmental resources and behaviors, and to compete effectively with its neighbors, that is, to successfully occupy its niche. If being better than good enough does not enhance reproductive success, it will not be selected for, and variations in that direction will be lost, particularly if they come at the expense of other important processes or abilities.

In this context it is worth noting that we are always dealing with an organism throughout its life cycle. Different traits can have different reproductive values at different developmental stages. Being cute can have important survival benefits for a baby but be less useful in a corporate board room (although perhaps not). A trait that improves survival during early embryonic development or enhances reproductive success as a young adult can be selected for even, if it produces negative effects on older, post-reproductive individuals. Moreover, since the probability of being dead by accident or disease, and so no longer reproductively active, increases with age, selection for traits that benefit the old will inevitably be weaker than selection for traits that benefit the young, although this trend can be modified in organisms in which the presence of the old, for example, grandparents, positively influences the survival and reproductive success of the young, for example through teaching and babysitting. Of course survival and fertility curves can change in response to changing environmental factors, which alter selective pressures. In fact, lifespan itself is a selected trait, since it is the population not the individual that evolves.[169] In this light, while most large mammals have long lifespans, a number of large and complex invertebrates, such as squid, octopus, and cuttlefish have short lifespans.[170]

We see the evidence for various evolutionary compromises all around us.[171] They explain the limitations of our senses, as well as our tendency to get backaches, need hip-replacements,[172] and our susceptibility to diseases and aging.[173] For example, the design of our eyes leaves a blind spot

---

[168] How the Hyoid Bone Changed History: http://www.livescience.com/7468-hyoid-bone-changed-history.html

[169] Methusaleh's Zoo: clues for extending human health span & Why Men Matter: Mating Patterns & Evolution of Lifespan

[170] As described in Peter Godfrey-Smith's Other Minds: The Octopus, the Sea, and the Deep Origins of Consciousness

[171] Wikipedia: Evidence of common descent

[172] Hip pain may be 'hangover from evolution': http://www.bbc.com/news/health-38251031

[173] How Bipedalism Arose

in the retina. Complex eyes have arisen a number of times during the history of life, apparently independently, and not all have such a blind spot - a blind spot is not a necessary feature of a complex eye. We have adapted to this retinal blind spot through the use of saccadic eye movements because this is an evolutionarily easier fix to the problem than rebuilding the eye from scratch, which is likely to be impossible (evolutionarily). An intelligently designed human eye, that is, an eye designed from scratch would presumably not have such an obvious design flaw, but given the evolutionary path that led to the vertebrate eye, it may simply have been impossible to "back up" and fix this flaw. More to the point, since the vertebrate eye works well, there is no apparent reward in terms in reproductive success associated with removing the blind spot. This is a general rule: current organisms work, at least in the environment that shaped their evolution. Over time, organisms that diverge from the current optimal, however imperfect, solution will be at a selective disadvantage. The current vertebrate eye is maintained by stabilizing selection. The eyes of different vertebrates differ in their acuity, basically how fine a pattern of objects they can resolve at what distance, and sensitivity, what levels and wavelengths of light they can perceive. Each species has eyes, and their connections to the brain, adapted for their specific ecological niche. For example, an eagle sees details at a distance four to five times as far as the typical human; why? because such visual acuity is useful in terms of the eagle's life-style (selection), whereas such visual details might result in non-useful distractions in humans.[174]

Homologies provide evidence for a common ancestor

The more details two structures share, the more likely they are to be homologous. In the 21st century molecular methods, particularly inexpensive genome (DNA) sequencing, have made it possible to treat gene sequences and genomic organization as traits that can be compared quantitatively. Detailed analyses of many different types of organisms reveals the presence of a common molecular signature that strongly suggests that all living organisms share a large numbers of homologies, which implies that they are closely related - that they share a common ancestor. These universal homologies range from the basic structure of cells to the molecular machinery involved in energy capture and transduction, information storage and utilization. All organisms

- use double-stranded DNA as their genetic material;
- use the same molecular systems to access the information stored in DNA;
- express that information initially in the form of RNA molecules;
- use a common genetic code, with a few variations, and messenger RNAs (mRNAs) to specify the sequence of polypeptides (proteins);
- use ribosomes to translate the information stored in messenger RNAs into polypeptides; and
- share common enzymatic (metabolic) pathways and structures (lipid-based boundary membranes).

Questions to answer:

46. How would you decide whether a trait is primitive (ancestral) or specialized (derived)?
47. Describe a scenario in which the loss of a trait or a gene is beneficial?
48. Explain why the loss of a trait or convergent evolution complicates lineage analysis?
49. Describe a scenario in which the simplification of a complex organism would be selected for?
50. Construct a diagram that shows the difference between homologous and analogous traits, and use it to explain the difference.

Anti-evolution arguments

The theory of evolution has been controversial since its inception largely because it deals with issues of human origins and behavior, our place in the Universe, life and its meaning. Its implications can be disconcerting, but many observations support the fact that all organisms on Earth are the

174 What If Humans Had Eagle Vision?

product of evolutionary processes and these processes are consistent with what we know about how matter and energy behave. As we characterize the genomes of diverse organisms, we see evidence for these interrelationships, observations that non-scientific (creationist) models would never have predicted and do not explain. That evolutionary mechanisms have generated the diversity of life and that all organisms found on Earth share a common ancestor is as well-established as the atomic structure of matter, the movement of Earth around the Sun, and the solar system around the Milky Way galaxy. The implications of evolutionary processes remain controversial, but not evolution itself. We would argue that religions and other belief systems that deny the evolutionary relationships between organisms, and the role of evolutionary mechanisms in shaping organisms, including humans, run the risk of making themselves look ridiculous, at least in terms of data-based (scientific) discussions.[175] On the other hand science (and evolution theory) have little to say on how we should behave, what it means to be moral, basically a good person, or why being a selfish unfeeling, narcissist is bad.

Scientific knowledge is a body of knowledge of varying degrees of certainty-some most unsure, some nearly sure, but none absolutely certain … Now we scientists are used to this, and we take it for granted that it is perfectly consistent to be unsure, that it is possible to live and not know. - Richard Feynman.

...it is always advisable to perceive clearly our ignorance. – Charles Darwin.

Questions to ponder:

- Describe testable predictions that emerge from "intelligent design creationism"?
- In what ways might organisms direct (or influence) their own evolution? how about humans specifically?
- If the environment were constant, would extinction or evolution occur?
- Should modern genetic engineering methods be used to fix evolutionary design flaws?

[175] Go ahead and "teach the controversy:" it is the best way to defend science.

Chapter 4: Social evolution, sex & sexual selection

In which we consider how unicellular organisms can cooperate with one another and how cooperation led to the evolution of multicellular organisms composed of distinct cell types. Similar evolutionary mechanisms have produced a range of cooperative (social) behaviors as well as opportunities for cheating to defend against cheaters. One particularly important social behavior is sexual reproduction and we consider its effects on organisms and their evolution.

The naturalist Ernst Mayr (1904-2005) stressed the differences in thinking in biology compared to physics and chemistry. The history of an electron, an atom, or a molecule is irrelevant to its physical and chemical properties. Each carbon isotope atom, for example, is identical to all others - one could be replaced by another and you could never, in practice or in theory, be able to tell the difference. In contrast, each organism, how it is built, how it behaves, how it interacts with other organisms, and the possible futures of its descendants is the result of a continuous evolutionary process involving both adaptive (selective) and non-selective and non-adaptive processes stretching back ~3.5 billion years. This history encompasses an unimaginable number of individually unpredictable events (mutations, noisy gene expression, accidents and environmental disasters, isolated and merging populations). Because of its molecular and cellular complexity and distinct history, each organism is unique and distinguishable from all others.[176]

In biology, we normally talk about organisms, but this is often too simplistic. When does an organism begin? What are its boundaries? The answers can seem obvious, but then again, perhaps not. When a single-celled organism reproduces it goes through some form of cell division, and when division is complete, one of the two organisms present is considered a new organism and the other the old (preexisting) one, but often it is not clear which is which. In fact, both are old, both reflect a continuous history stretching back to the origin of life. When an organism reproduces sexually, the new organism arises from the fusion of two pre-existing cells and it itself produces cells that fuse to form the next generation. But if we trace the steps backward from any modern organism, we find no clear line between the different types (that is, species) of organisms. When, exactly, did humans (*Homo sapiens*) appear from pre-humans, or modern birds from their dinosaurian progenitors? The answer is necessarily arbitrary, since cellular and organismic continuity is never interrupted - life does not start, stop, and start again, it continues until it stops irreversibly in death. Because of superfecundity, selection, and speciation, it also generates branches.

In a similar manner, we typically define the boundaries of an organism in physical terms, but organisms interact with one another, often in remarkably close and complex ways. For example, some unicellular organisms live so closely together that it is impossible for them to live apart.[177] Another, dramatic example of this type of situation are the eusocial organisms. While many of us are familiar with the social structure of ants and bees, fewer (we suspect) are aware of naked (*Heterocephalus glaber*) and Damaraland (*Cryptomys damarensis*) mole rats. In these organisms reproduction occurs at the group level; only select females, termed queens because they are large, produce offspring. Most members of the group are effectively sterile female workers; a few males

[176] While these events obey physical and chemical laws, in practice, the number of variables involved makes them unpredictable. At the same time, because they are based on natural processes, when we consider large numbers of such events, they become predictable. So while the mutation rate is predictable, which mutations occur in which organism is not.

[177] Cultured Asgard Archaea Shed Light on Eukaryogenesis

are present, they inseminate the queen.[178] So what, exactly, is the organism? the social group or the individuals that make it up? From an evolutionary perspective, selection is occurring at a social level as well as the organismic level.

Similarly, consider yourself and other multicellular organisms (animals and plants). Most of the cells in your body, known as somatic cells, do not directly contribute to the next generation, rather they cooperate to insure that a subset of cells, known as germ line cells (sperm and eggs), have a chance to form a new organism. In a real sense, the somatic cells sacrifice themselves so that the germ line cells can produce a new organism. They are the sterile workers to the germ line's queen. The term "sacrifice" in the context of the somatic cells of a multicellular organism may seem weird, and too anthropomorphic, since both germ line and somatic cells are necessary parts of a single organism. We might argue that it is the organism, rather than the cells that compose it, that is the biologically meaningful object. Similarly, in a eusocial organism, it is the social group that matters.

We find examples of social behavior at the level of unicellular organisms as well, and most recently in viruses.[179] For example, think about a unicellular organism that divides but in which the offspring of that division stick together. As this process continues, we get what we might term a colony. Is such a clump of cells one or many organisms? If all of the cells within the group can produce new cells, and so new colonies, we consider it a colony of organisms. So where does a colony of organisms turn into a colonial organism? The distinction can be ambiguous, but we can adopt a set of guidelines or rules of thumb.[180] One criterion would be that a colony becomes an organism when it displays traits that are more than just sticking together or failure to separate, that is, when it acts more like a coordinated group. This involves the differentiation of cells, so that certain cells become specialized to carry out specific roles. Producing the next generation of organisms is one such specialized functional role. Other cells may become specialized for feeding or defense, they support the process of reproduction, in part by enabling the resulting organism to occupy a particular ecological niche. The differentiation of cells from one another within a multicellular aggregate has moved a colony of organisms to a multicellular organism. What is tricky about this process is that originally reproductively competent cells have given up their ability to reproduce, and are now acting, in essence, to defend or support the cells that do reproduce. This is a social event and is similar (analogous) to the behavior of naked mole rats. Given that natural selection acts on reproductive success, one might expect that the evolution of this type of cellular and organismic behavior would be selected against or simply impossible to produce, yet multicellularity and social interactions have arisen independently many times during the history of life on earth.[181] Is this a violation of evolutionary theory or do we have to get a little more sophisticated in our thinking?

Questions to answer:

51. What features (behaviors) are important when defining an organism? Does your definition include both uni- and multi-cellular organisms?
52. How would you characterize humans in terms of sociality?

Selecting social (cooperative) traits

So how does evolution produce multicellularity? To answer this question, we need to approach evolutionary processes more broadly. The first new idea we need to integrate into our theoretical

---

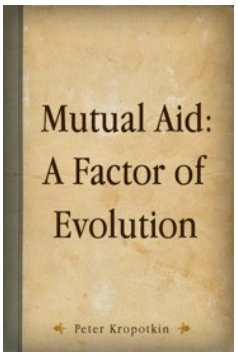
[178] An Introduction to Eusociality: http://www.nature.com/scitable/knowledge/library/an-introduction-to-eusociality-15788128

[179] The secret social lives of viruses

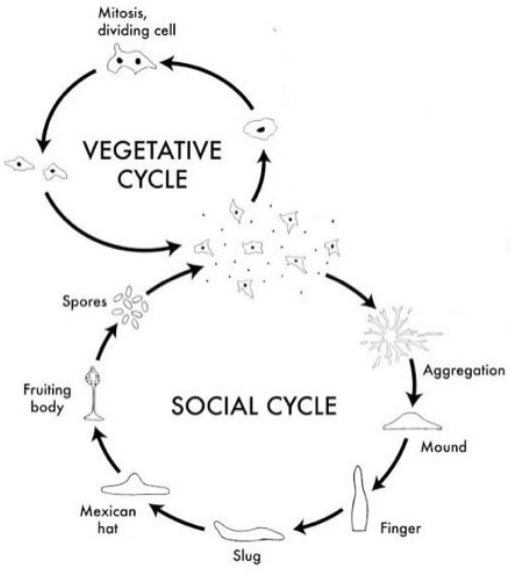
[180] A twelve-step program for evolving multicellularity and a division of labor

[181] The Origins of Multicellularity

framework is that of inclusive fitness, which is sometimes referred to as kin selection. For the moment, let us think about traits that favor the formation of a multicellular organism - later we will consider traits that have a favorable effect on other, related organisms, whether or not they directly benefit the cell or organism that expresses that trait. Finally, we will consider social situations in which behaviors have become fixed to various extents, and are extended to strangers; humans can, but do not always, display such behaviors. The importance of mutual aid in evolutionary thinking, that is the roles of cooperation, empathy, and altruism in social populations, was emphasize by the early evolutionary biologist and anarchist (Prince) Peter Kropotkin (1842–1921)(→).



All traits can be considered from a cost-benefit perspective. There are costs ("c") in terms of energy needed to produce a trait and risks associated with expressing the trait, and benefits ("b") in terms of the trait's effects on reproductive success. To be evolutionarily preferred, that is, "selected for", the benefit b must be greater than the cost c, that is b > c. Previously we had tacitly assumed that both cost and benefit applied to one and the same organism, but when we consider cooperative (social) behaviors and traits, this is not necessarily the case. We can therefore extend our thinking as follows: assume that an organism displays a trait. That trait has a cost to produce and yet may have little or no direct benefit to the organism that produces it; it may even harm it. Now let us assume that this same trait benefits neighboring organisms, a situation similar to the fireman who risks their life to save an unrelated child in a burning building. How is it possible for a biological system (the fireman), the product of evolutionary processes, to display this type of self-sacrificing behavior? The answer is social systems.



As an example of this type of behavior consider the social amoebae *Dictyostelium discoideum*.[182] These organisms have a complex life style that includes a stage in which unicellular amoeba-like organisms crawl around in the soil eating bacteria, growing and dividing. In this phase of their life cycle, known as the vegetative cycle, the cells divide asexually (as if vegetables don't have sex, but we will come back to that!). If, or rather when, the environment turns hostile, the isolated amoeba sense this change and begin to secrete small molecules that influence their own and their neighbor's behaviors. They begin to migrate toward one another, forming aggregates of thousands of cells (←). Now something rather amazing happens: these aggregates begin to act as coordinated entities, they migrate around as multicellular "slugs" for a number of hours. Within the soil they respond to environmental signals, for example moving toward light, and then settle down and undergo a rather spectacular process of differentiation.[183] All through the cellular aggregation and slug migration stages, part of the social cycle, the original amoeboid cells remain distinct. Upon differentiation ~20% of the cells in the slug specialize to form stalk cells that can no longer divide; they go on to die through a process known as programmed cell death or apoptosis. Before they die the stalk cells act together, through changes in

[182] Molecular phylogeny and evolution of morphology in the social amoebas & A Simple Mechanism for Complex Social Behavior. A nice video here: http://youtu.be/bkVhLJLG7ug

[183] Behavior of cellular slime molds in the soil: http://www.mycologia.org/content/97/1/178.full

their composition and shape, to lift the non-stalk cells above the soil, where the non-stalk cells go on to form spores. The stalk cells sacrificed themselves so that non-stalk cells can form spores; specialized cells that can survive harsh conditions. Spores are released and can float in the air and be transported by the wind and other mechanisms into new environments. Once these spores land in a new, and hopefully hospitable environment, they convert back into unicellular amoeba that begin to feed and reproduce vegetatively. The available evidence indicates that within the slug the "decision" on whether a cell will form a stalk or a spore cell is not pre-determined, it arises from molecular level stochastic processes. The decision is not based on genetic (genotypic) differences - two genetically identical cells may both form spores, both stalk cells, or one might become a stalk and one a spore cell.[184]

Community behaviors & quorum sensing

A type of community behavior active at the unicellular level involves what is known as quorum sensing. This is a process by which organisms can sense the density (number of individuals per volume) of organisms in their immediate environment. Each individual secretes specific molecules that they also respond to through specific receptors. The organisms' response to this signaling molecule is dependent on its extracellular concentration. More importantly, the response is non-linear, and displays a "threshold" behavior. Below the system's threshold concentration there is little if any cellular response, above the threshold concentration the cell responds fully. When cells or organisms are present at a low density, the concentration of the signaling molecule never exceeds the threshold concentration. As the density of organisms increases; when the concentration of the signaling molecule exceeds the threshold concentration interesting things can start to happen; there are changes in cellular behavior, often associated with changes in gene expression (we will soon get to what that means).[185] We can think of this type of non-linear response as a strategy to avoid over-reacting to minor fluctuations in the environment. Only when the signal concentration gets high enough (exceeds the threshold concentration) does the system respond. The threshold concentration is a function of the concentration of signaling molecules, their binding affinity to the receptor, and other factors that we will consider in greater detail when we consider molecular interactions and mechanisms.

A classic example of a cooperative and quorum sensing behaviors is provided by the light emitting marine bacteria *Vibrio fischeri*. These bacteria stably colonize a dedicated light organ of the Hawaiian bobtail squid shortly after the squid "hatch".[186] While there are many steps in the colonization process, here we consider just a few to indicate how cooperative behaviors between the bacteria play a critical role. In order to colonize the squid's light organs the *V. fisherei* bacteria must bind to a specific region of the juvenile squid's light emitting organ. Bacteria are small, so you might imagine that very little light would be emitted from a single bacterium. If there were only a small number of bacteria within the light organ, they would be unable to generate a useful level of light, while at the same time, they would be using energy (all costs, no benefit). To increase the numbers (and concentration) of bacteria, the bacteria begin to divide and as they divide, they sense the presence of their neighbors and begin to secrete molecules that form of gooey matrix - this leads to the formation of a specialized aggregate of cells, known as a biofilm. Within the biofilm, the bacteria acquire the ability to follow chemical signals produced by the squid's light organ cells. The bacteria swim, through a process known as chemotaxis, toward the secreted signal and enter and colonize the squid's light organs.

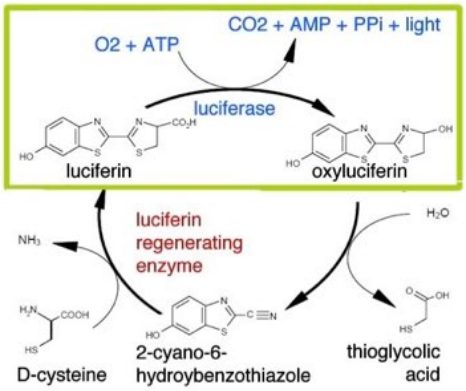
---

[184] This type of behavior occurs in a number of organisms, including the bacteria: see From cell differentiation to cell collectives: *Bacillus subtilis* uses division of labor to migrate: http://www.ncbi.nlm.nih.gov/pubmed/25894589

[185] Quorum sensing in bacteria: http://www.ncbi.nlm.nih.gov/pubmed/11544353

[186] Zink et al (2021). A Small Molecule Coordinates Symbiotic Behaviors in a Host Organ

Within the light organs the bacteria emit light through a reaction system involving the molecules luciferin and O_2 (→): coupled chemical reactions convert chemical energy into the emission of light, electromagnetic energy (the thermodynamics of coupled reactions are considered in chapter 5). The light emitting reaction is catalyzed (that is, sped up) by the protein luciferase, an enzyme (a protein catalyst). The luciferase protein is encoded by a bacterial gene. Its original role in the bacteria has been proposed to be in the "detoxification of deleterious oxygen derivatives".[187] The light emitting reaction is regulated so that it occurs only when the number of bacteria within a light organ is high enough to make the emission of light useful, which decreases the cost to benefit ratio.



So how do the bacteria know that they are in the presence of sufficiently high concentration of neighbors? Here is where quorum sensing comes into play. A molecule secreted by the bacteria regulates the components of the light reaction. At high concentrations of bacteria, the concentration of the secreted molecule rises above a threshold, and the bacteria respond by turning on their light emitting systems - that is, they express the genes encoding the protein luciferase and the proteins involved in the synthesis of luciferin.

Mechanistically similar systems are involved in a range of processes including the generation of toxins (virulence factors), secreted digestive enzymes, and antibiotics directed against other types of organisms. These are produced when the density of bacteria rises above a threshold concentration. This insures that when biologically costly molecules are made (such as luciferase and luciferin), they are effective – that is, they are produced at a level high enough to carry out their intended roles. These high levels can only be attained through cooperative behaviors involving many individuals.

Questions to answer:

53. Why (generally) does a quorum signal need to be secreted (released) from the organism? What other components are necessary for such cooperative behavior to occur.
54. Is a population of bacteria that display quorum sensing behavior a single organism, justify your answer.

Question to ponder:

- How might it impact the social behavior of slime molds if the percentage of spore cells were 1% rather than 80%?
- Why is a non-linear response to a stimulus important in biological systems? How could it be achieved?

Active (altruistic) cell death and survivors

A type of behavior you might think would be impossible for evolutionary processes to produce would be the active and intentional death of a cell or an organism. Yet, such behaviors are surprisingly common in a wide range of systems.[188] The death and release of leaves from deciduous trees in the autumn is an example of a built-in or "programmed" cell death process, also known as apoptosis, from the Greek meaning to fall off. The programmed cell death process amounts to cellular suicide. It plays important roles in the formation of various structures within multicellular organisms, such as the fingers of hands that would develop as paddles without it. Programmed cell death also plays a critical role in the development of the immune and nervous systems, important topics beyond our scope here.[189] Programmed cell death is distinct from accidental cell death, such

---

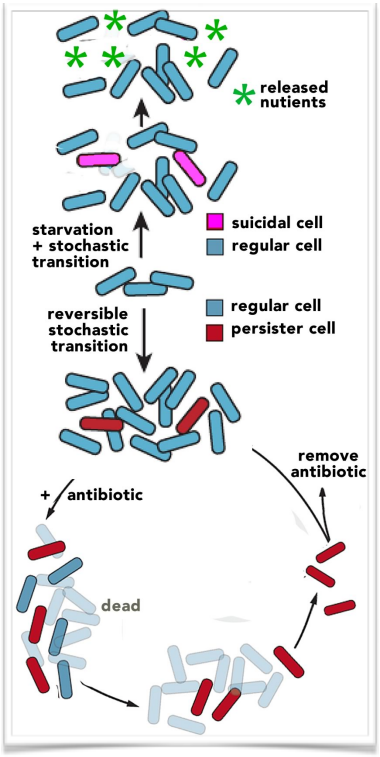
[187] Experimental evidence for the physiological role of bacterial luciferase: http://www.ncbi.nlm.nih.gov/pubmed/14669913

[188] See On the paradigm of altruistic suicide in the unicellular world: http://www.ncbi.nlm.nih.gov/pubmed/20722725

[189] Apoptosis in the nervous system & Apoptosis in the immune system

as occurs when a splinter impales a cell or you burn your skin. Such accidental death leads to what is known as necrosis. In necrosis, cellular contents are spilled out in an uncontrolled manner from the dying cell. The release of cellular debris provokes various organismic defense systems to migrate into the damaged area and (primarily) fight off invading bacteria. The swelling and inflammation associated with injury is an indirect result of necrotic cell death. In contrast, apoptotic cell death occurs using a well-defined pathway that requires energy to carry out. Cell contents are retained during the process; no inflammatory, immune system response is provoked. Surrounding cells actively remove the remains of the apoptotic cells. In programmed cell death/apoptosis appears to play specific and important roles within the context of the organism.

Commitment to active cell death is a tightly controlled process. Here we consider the role programmed cell death in the context of simpler systems, specifically in communities of unicellular organisms. In such systems, programmed cell death is a process triggered by environmental stresses together with quorum sensing. In this situation, a subset of the cells can stochastically "decide" to undergo cell death by activating a cell death pathway. In these systems, when a cell dies, its contents are released and can be used by the living cells that remain (→). These living cells gain a benefit, and we would predict that the increase in nutrients will increase their chances of survival and successful reproduction. This strategy works because as the environment becomes hostile, not all cells die at the same time. It makes no evolutionary sense for an isolated cell to die through programmed cell death, since the release of its nutrients would fail to benefit its (related) neighbors. Instead of dying, better to change into what is known as a "persister". In such a state the bacterium stops growing and minimizes its use of (and need for) energy (→). In the persister state, the bacterium can survive until the stressor (e.g. an antibiotic, a molecule that leads to the death of susceptible bacteria) disappears from the environment. Such behaviors (programmed cell death or the adoption of a persister phenotype) occur in groups of genetically identical cells and involve the action of stochastic processes.



So how do cells kill themselves (on purpose)? Many use a similar strategy. They contain what is known as an addiction module, which consists of two genes - the first encodes a toxic molecule. The toxic molecule, which can kill the cell, is synthesized (expressed) continuously. Many distinct toxin molecules have been identified, so they appear to form analogous rather than homologous systems – meaning that they appear to have evolved independently. Now you may well wonder how such a gene could exist, how does the cell survive in the presence of a gene that encodes and expresses a lethal toxin. The answer is that the cell contains a second gene that encodes an anti-toxin molecule; the anti-toxin typically acts on the toxin and inhibits its activity. Within the cell, the toxin-anti-toxin complex forms but does not harm the cell – the toxin's activity is inhibited by its interactions with the anti-toxin. So far, so good - but you might ask, what is the point - nothing interesting is going on! But the system has one more wrinkle. The toxin and anti-toxin molecules differ in an important way. The toxin molecule is degraded by molecule systems within the cell slowly; once synthesized it has a long "half-life". In contrast, the anti-toxin molecule is degraded rapidly; it has a short half-life. Under normal conditions the steady state concentration of the anti-toxin, a function of its synthesis and degradation rates, is sufficient to inhibit all of the toxin present. The cell has become addicted to the anti-toxin, which must be made continuously in order to inhibit the toxin and avoid cell death.

Now consider what happens if the cell is stressed, either by changes in its environment or perhaps infection by a virus? Generally cellular activity, including gene expression and the synthesis of cellular components, such as the anti-toxin, slows or stops. Can you predict what will happen? The level of the toxin molecule, which has a long half-life, decreases slowly, whereas the level of the

short half-life anti-toxin drops much more rapidly. When the level of the anti-toxin falls below that needed to inhibit the toxin, the now active toxin initiates the process of cell death, leading to the release of the dying cell's components into the environment.

In addition to the dying cell "sharing" its resources with its (presumably related) neighbors, programmed cell death can be used as a population-wide defense mechanism against viral infection. One of the key characteristics of viruses is that they must replicate within a living cell. Once a virus enters a cell, it typically disassembles itself and sets out to reprogram the cell's biosynthetic machinery to generate new copies of the virus. During the period between viral disassembly and the assembly of newly synthesized viruses, the infectious virus disappears - it is said to be latent. If the cell kills itself before new viruses are synthesized, it also "kills" (or rather inactivates or eliminates) the infecting virus. By killing the virus (and itself) the infected cell acts to protect its neighbors from viral infection - this can be seen as a form of the altruistic, self-sacrificing behaviors we have been considering.[190]

Inclusive fitness, kin and group selection, and social evolution

The question that troubled Darwin (and others) was, how can evolutionary processes produce this type of social, self-sacrificing behavior? Consider, for example, the behaviors of bees. Worker bees, who are sterile females, "sacrificed themselves to protect their hives" even though they themselves do not reproduce, they are sterile.[191] Another example, taken from the work of R.A. Fisher (1890-1962), involved the evolution of noxious taste as a defense against predators. We can assume that the organisms eaten by predators do not directly benefit from this trait, after all, they have been eaten. So how can the trait of "distastefulness" arise in the first place? If evolution via natural selection is about an individual's differential reproductive success, how are such traits even possible? W.D. Hamilton (1936-2000) provided the formal answer, expressed in the equation $rb > c$. As before in our consideration of costs and benefits, "b" stands for the trait's benefit to the organism and others, "c" stands for the cost of the trait to the individual, while "r" indicates the extent to which two organisms within the population are related to one another, it is a measure of genetic similarity.

Let us think more about what this means. How might active cell death in bacterial cells be beneficial evolutionarily? In this case, reproduction is asexual; the organism's (cell's) offspring, and its likely neighbors, will be closely related – sharing very similar genomes. They are clonally-related to one another in the same way that the cells of a multicellular organism, such as yourself, are derived from a single cell, the fertilized egg which, once formed, divides in an asexual manner. Aside from occasional mutations (changes in DNA), the cells in a clone and within an organism are genetically identical, that is they have DNA molecules that are identical in sequence.[192] Their genotypic similarity arises from the molecular processes by which the genetic material (DNA) replicates and is delivered to the two daughter cells. We can characterize the degree of relationship, or genotypic similarity, through their r value, the coefficient of relationship. In two genetically identical organisms, r = 1. Two unrelated organisms, with minimum possible genotypic similarity would have an r very close to, but slightly larger than 0 (why is r, very small but not equal to 0?)[193] Now let us return to our cost-benefit analysis of a trait's effect on reproductive success. As we discussed before, each trait has a cost of c to the organism that produces it, as well as a potential benefit of b in terms

---

[190] The evolution of eusociality

[191] Dugatkin, L.A. 2007. Inclusive Fitness Theory from Darwin to Hamilton

[192] There is an exception to this role involving a subset of the cells of the immune system, but it is not important here.

[193] We will consider the complicating effects of sexual reproduction (which is involved in the formation of the fertilized egg) later on. Suffice it to say, that you are not genetically identical to either of your parents or your own siblings (if you have any, and unless you are have an identical twin). As an approximation, you share ~50% of your genetic material with either of your parents and ~25% with your siblings.

of reproductive success. Selection leads to a trait becoming prevalent (frequent or even fixed) within a population if b >> c. But this equation ignores the effects of a trait on other related and neighboring organisms. In this case, we have to consider the benefits accrued by these organisms as well. Let us call the benefits to the individual that result from their cooperative/altruistic behavior b_i and the benefits to others/neighbors b_o. To generate our social equation, known as Hamilton's rule, we need to consider what is known as the inclusive fitness, namely the benefits provided to others as a function of their relationship to the cooperator. So b > c becomes $b_i + r \times b_o > c$. This leads to the conclusion that a trait can evolve if the cost to the cell or organism that displays it, in terms of metabolic, structural, or behavioral impact on its own reproductive ability, is offset by a sufficiently large increase in the reproductive success of individuals related to it. The tendency of an organism to sacrifice itself for others will increase, that is, be selected for, provided that the reproductive success of closely enough related organisms is increased sufficiently. We will see that we can apply this logic to a wide range of situations; it provides an evolutionary mechanism driving the appearance and preservation of various social behaviors. Given the clonal nature of many types of microbes, inclusive fitness can be particularly powerful in these organisms, although it is also significant in small populations of sexually reproducing organisms.

That said, the situation is often more complex. Typically, to have a significant impact, inclusive fitness requires a close relationship to the recipient of the beneficial act. So how can we assess this relationship? How does one individual "know" (that is, how is its behavior influenced by the degree of relationship to others) that it is making a sacrifice for its relatives and not just a bunch of (semi-) complete strangers? As social groups get larger, identifying relatives becomes a more and more difficult task. One approach is to genetically link the social trait, the altruistic behavior, to a physically discernible trait, like smell or a visible structure or behavior. This is sometimes called a "green beard" trait. The likelihood that an organism will behave socially is, one way or the other, linked to the display of a recognizable trait, e.g. a green beard. The presumption is that it is difficult to lose the social cooperation trait without also loosing the green beard trait. The presence of the green beard trait indicates that an organism with the trait will cooperate, it would be "prepared" to "sacrifice" itself for you in the same way you are prepared to sacrifice for it. Assuming a close linkage between the two traits (social and visible), one can expect social behavior from an individual who displays the trait, even if they are only distantly related. In some cases, a trait may evolve to such a degree that it becomes part of an interconnected set of behaviors, a type of biosocial moral system.[194]

Once, for example, humans developed a brain sufficiently complex to do what it was originally selected for (assuming that it was brain complexity that was selected, something we might never know for sure), this complexity may have produced various unintended byproducts. Empathy, self-consciousness, and a tendency to neurosis may not be directly selected for but could be side effects of behavioral processes or tendencies that were. As a completely unsupported (but plausible) example, the development of good memory as an aid to hunting might leave us susceptible to nightmares. Assume, for the moment (since we are speculating here), that empathy and imagination are "unintended" by-products of selective processes. Once present, they themselves can alter future selection pressures and they might not be easy to evolve away from, particularly if they are mechanistically linked to a trait that is highly valued, that is, selected for. The effects of various genetic mutations on personality and behavior strongly supports the idea that such traits have a basis in, or are influenced by, one's genotype. That said, this is a topic well beyond our scope.

Group selection

A proposed alternative to inclusive fitness (sometimes known as kin selection) is the concept of group selection. In this type of evolutionary scenario, small groups of organisms of the same species are effectively acting as single (perhaps colonial) organisms. It is the reproductive success of the group, rather than the individuals within the group, compared to other groups of the organism that is

[194] We might consider organisms that fail to live by these rules as sociopaths or suffering from pernicious narcissism.

the basis of selection. In certain situations, groups that display cooperative and altruistic traits may have a selective advantage over groups that do not. Again, the mathematical analysis is similar, and it has been claimed that group and kin selection are mathematically equivalent, even though one occurs between population groups and the other within a population group.[195] The costs of a trait must be offset by the benefits, but now the key factor is membership in a particular group, and typically, members of a group tend to be more closely related to one another. The life cycle of the bacterium *Myxococcus xanthus* provides an example of this type of behavior. When environmental conditions are harsh, the cells aggregate into dense, 100 μm diameter "fruiting bodies", each containing ~100,000 stress resistant spores. When the environment improves, and nutrients become available, the spores are released en mass and return to active life. They move and feed in a cooperative manner through the release of digestive enzymes that, because they are acting in a quorum mode, can reach high levels.[196] A well-coordinated group is expected to have a significant reproductive advantage over a more anarchic collection of individuals.

While their functional roles are clearly different, analogous types of behavior are seen in flocks of birds, schools (or shoals) of fish, swarms of bees, blooms of algae, and groups of slime mold cells (→).[197] Each of these examples represents a cooperative strategy by which organisms gain a reproductive advantage over those that do not display the behavior. While the original behavior is likely the result of kin selection, in the wild it is possible that different groups (communities) are in competition with one another, and the group(s) that produces the most offspring, that is, the most reproductively successful group will come to dominate.

Defense against social cheaters

Now an interesting question arises: within a social organization, such as a group of cooperating microbes or hunters,[198] we can expect that, through mutation and other behavioral mechanisms, cheaters will arise. What do we mean by a cheater? Imagine a bacterium within a swarm, a cell in an organism, or an animal in a social group that fails to obey the rules - it may benefit from social cooperation without contributing to it.[199] For example when an individual accepts help from others, but fails to help others. In the case of slime mold aggregates, imagine a cell that can avoid becoming a non-reproductive stalk cell, instead it always differentiates into a reproductively competent spore. Let us further assume that this trait has a genetic basis. What happens over time? One plausible scenario would be that this spore cell begins its own clone of migratory amoeba, but when conditions change so that aggregation and fruiting body formation occur, most of the cells avoid forming the stalk. We would predict that the resulting stalk would be short or non-existent and so would not be able to lift the spore forming region above the soil, reducing or eliminating the efficiency of dispersion. Different populations would differ based on the percentage of individuals with the cheater phenotype. If dispersion is important for long term species survival, there would be selection for populations with low levels of cheaters.

---

[195] Mathematics of kin- and group-selection: formally equivalent? http://www.ncbi.nlm.nih.gov/pubmed/19929970

[196] Evolution of sensory complexity recorded in a myxobacterial genome: http://www.ncbi.nlm.nih.gov/pubmed/17015832

[197] How Does Social Behavior Evolve?

[198] An interesting read: The stag hunt and the evolution of social structure.

[199] As an example, consider a person who accepts the protection of police and firefighters, but avoids paying their taxes.

Multicellular organisms are social systems, composed of cells that have given up their ability to reproduce new organisms for the ability to enhance the reproductive success of the organism as a whole. In this context cancers are diseases that arise from mutations that lead to a loss of social control. Cells, whose survival and reproduction is normally strictly controlled, lose that control; they become “anti-social” and begin to divide in an uncontrolled and/or inappropriate manner, disrupting the normal organization of the tissue in which they are located, and they can become malignant, which means that they can breakaway from their original location, migrate, and colonize other areas of the body, a process known as metastasis. The uncontrolled growth of the primary tumor and these metastatic colonies leads eventually to the death of the organism as a whole.

Once a social behavior has evolved, under what conditions can evolutionary mechanisms maintain it, specifically defend it against cheaters (narcissistic sociopaths). One approach is to link the ability to join a social group with various internal and external mechanisms. This makes cooperators recognizable and works to maintain a cooperative or altruistic trait even in the face of individual costs. A complex topic in its own right that we consider only superficially. When we think about maintaining a social behavior, we can think of two general mechanisms: intrinsic and extrinsic policing. For example, assume that a trait associated with the social behavior is also linked to, or required for, cellular survival. In this case, a mutation that leads to the loss of the social trait may lead to cell death (apoptosis). Consider this in the context of cancer. Normal cells can be considered to be addicted to normality. When their normality is disrupted they undergo apoptosis. A cell carrying a mutation that allows it to grow in an uncontrolled and inappropriate manner will likely undergo apoptosis itself before it can produce significant damage.[200] For a tumor to grow and progress, other mutations must somehow disrupt and inactivate the normal (wild-type) apoptotic response. The apoptotic process reflects an intrinsic-mode of social control. It is a little like the guilt experienced by (some) people when they break social rules or transgress social norms. The loss of social guilt is analogous to the inhibition of apoptosis in response to various cues associated with abnormal behavior.[201]

In humans, and in a number of other organisms, there is also an extrinsic social control system. This is analogous to the presence of external policeman. Mutations associated with the loss of social integration – that is, the transformation of a cell to a cancerous state – can lead to changes in the character of the cell. Cells of the immune system can recognize these changes as "non-self" and induce the death of the mutant cell.[202] Of course, given that tumors occur and kill people, we can assume that there are mutations that enable tumor cells to avoid such immune system surveillance. As we will see, one part of the cancerous phenotype is often a loss of normal mutation repair systems. In effect, the mutant cell increases the number of unrepaired mutations, and consequently, the genetic variation in the cancer cell population. While many of these variants are lethal, the overall effect is to increase the rate of cancer cell evolution. This leads to an evolutionary race. If the cancer is killed by intrinsic and extrinsic social control systems, no disease occurs. If, however, the cancer evolves so as to avoid death by these systems, the cancer can progress and spread. As we look at a range of social systems, from cooperating bacteria to complex societies, we see examples of intrinsic and extrinsic control.

Driving the evolutionary appearance of multicellular organisms

Now that we have introduced cooperative behaviors and how evolutionary mechanisms can select and maintain them, we can begin to consider their roles in the evolution of multicellular

---

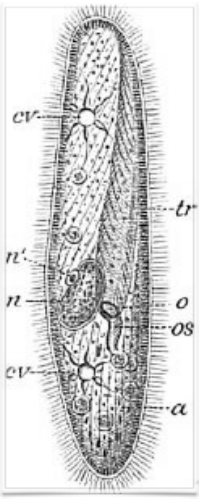
[200] Apoptosis in cancer: http://carcin.oxfordjournals.org/content/21/3/485.full

[201] In an age of rampant narcissism and social cheating – the importance of teaching social evolutionary mechanisms.

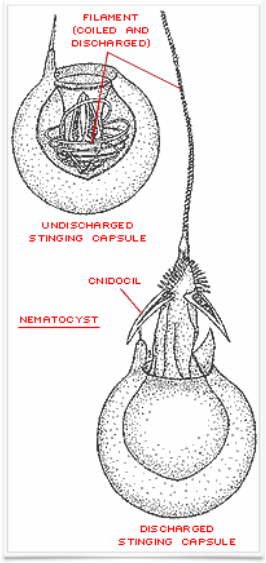
[202] Immune recognition of self in immunity against cancer & Anti-cancer drugs that reactivate the immune survalience

organisms.[203] As we have mentioned there are a number of strategies that organisms take to exploit their environment. Most prokaryotes (bacteria and archaea) are unicellular, but some can grow to substantial (visible) sizes. For example, the bacterium *Epulopiscium fishelsoni* inhabits the gut of the brown surgeonfish *Acanthurus nigrofuscus* and can grow to more than 600 µm in length. As we will see, the unicellular eukaryotic algae of the genus *Acetabularia* can be more than 10 cm in length. Additionally, a number of multicellular prokaryotes exhibit quite complex behaviors. A particularly interesting example is a species of bacteria that form multicellular colonial organisms that sense and migrate in response to magnetic fields.[204] Within the eukaryotes, there are both microscopic unicellular and macroscopic and multicellular species, including the animals, plants, and fungi.

What drove the appearance of multicellular organisms? Scientists have proposed a number of theoretical and empirically supported models. Some have suggested that predation is an important driver, either enabling the organisms to become better (or more specific) predators themselves or to avoid predation. In an experimental study, when the unicellular algae *Chlorella vulgaris* (5 to 6 µm in diameter) was grown together with a unicellular predator *Ochromonas vallescia*, which typically engulfs its prey, it was found that over time, *Chlorella* formed multicellular colonies that *Ochromonas* could not ingest.[205]



At this point what we have is more like a colony of organisms rather than a colonial organism or a true multicellular organism. The change from multi-individual colony to multicellular organism involves cellular specialization, so that different types of cells within the organism come to carry out different functions. The most dramatic specialization being that between the cells that generate the body of the organism, known as somatic cells, and those that give rise to the next generation of organisms, known as germ cells. At the other extreme, instead of producing distinct types of specialized cells to carry out distinct functions, a number of unicellular eukaryotes, known as protists, have complex cells that display a number of highly specialized behaviors such as directed motility, predation, osmotic regulation, and digestion (→). But such specialization can be carried out further in multicellular organisms, where there is a socially based division of labor. The stinging cells of jellyfish provide a classic example; highly specialized cells deliver poison to any organism that touches them through a harpoon-like mechanism (←). The structural specialization of these cells makes processes such as cell division impossible and typically a stinging cell dies after it discharges. Presumably, it is simpler to generate a new stinging cell than it is to reset a discharged cell. The production of these new cells involves both cell division and differentiation, which we will consider later. While we are used to thinking about individual organisms, the same logic can apply to groups of distinct organisms. The presence of cooperation can extends beyond a single species, leading to ecological interactions in which organisms work together to various degrees to achieve that which would be much more difficult or impossible to achieve on their own (while maintaining their ability to reproduce.



Based on the study of a range of organisms and their genetic information, we have begun to clarify the origins of multicellular organisms. Such studies indicate that multicellularity has arisen independently in a number of eukaryotic lineages. This strongly suggests that in a number of contexts, becoming multicellular is a successful way to establish an effective relationship with the environment.

---

[203] The evolutionary-developmental origins of multicellularity: http://www.amjbot.org/content/101/1/6.long

[204] A novel species of ellipsoidal multicellular magnetotactic prokaryotes from Lake Yuehu in China.

[205] Phagotrophy by a flagellate selects for colonial prey: A possible origin of multicellularity

Questions to answer:

55. What type(s) of mutation would enable an organism to escape a cell death module?
56. What types of mechanisms enable organisms (cells) to recognize each other as cooperators?
57. Make a model for the process that could lead to the evolution of social interactions.
58. What factors limit the complexity of a unicellular organism?
59. Is the schooling or herd behavior seen in various types of animals (such as fish and cows) a homologous or an analogous trait?

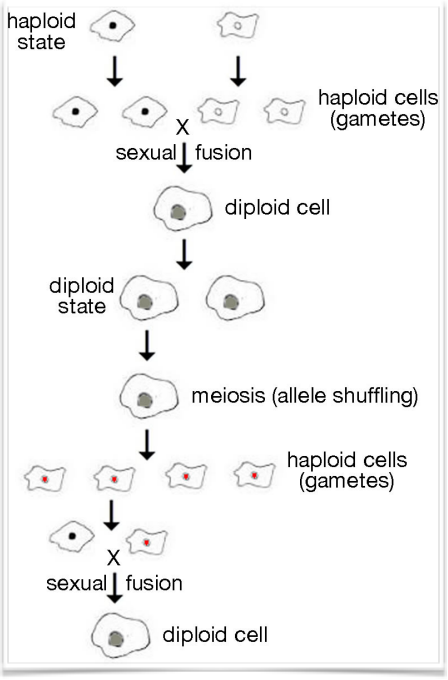
Questions to ponder:

- What strategies can be used to defend against the effects of cheaters in a population?
- Why is r (the relationship between organisms) never 0.
- What are some of the advantages of multicellularity? What are the drawbacks? Why aren't all organisms unicellular or multicellular?

Origins and implications of sexual reproduction

One type of social interaction, mentioned in passing, is sexual reproduction, which involves cooperative interactions between distinctly different organisms. While we are used to two distinct sexes (male and female), this is not universal. Many unicellular eukaryotes are characterized by a number of distinct "mating types". Typically, sexual reproduction involves the fusion of two specialized cells, known as gametes, of different mating types or sexes. Through mechanisms we will consider later, the outcome of sexual reproduction leads to increased genetic diversity among offspring.

So what are the common hallmarks of sexual reproduction? Let us return to the slime mold *Dictyostelium* as an exemplar. We have already considered its asexual life cycle, but *Dictyostelium* also has a sexual life cycle. Under specific conditions, two amoeboid cells of different mating types will fuse together (a version of sex) to form a single cell. The original cells are haploid (↓), meaning that they each have a single copy of their genome. When two haploid cells fuse, the resulting cell has two copies of the genetic material and is referred to as diploid. This diploid cell can then go through a series of events, known collectively as meiosis (a process we will get to). Meiosis results in the shuffling of genetic material and the production of four haploid cells. The critical point is that the genotypes of the haploid cells that emerge from meiosis are different from the haploid cells that originally fused together. Some organisms can spend a significant amount of time in the haploid state, while others spend most of their lives in the diploid state. You, for example, had a reasonably short haploid stage (as both an egg AND a sperm cell), and your diploid stage began when these two cells fused.



The oscillation between haploid and diploid states has some interesting implications. The first is that in the diploid state, there are (generally) two copies of each gene. The different versions of a gene are known as alleles – the two copies of a specific gene can be identical or it can be different. If they are the same, the cell/organism is known as homozygous at that genetic locus (gene); if they are different, it is heterozygous for that gene. Alleles can have a range of effects on phenotype, from cellular lethality to more subtle effects due to differences in the activity, localization, stability, or amount of the gene product. These effects can be influenced by the products of other genes, leading to what are known as genetic background effects. In the diploid phase of the life cycle, the effects of a lethal or deleterious allele can be masked by the presence of the other, functional or wild type allele. Such masked alleles are commonly referred to

as recessive. We will return to these topics later on. Where genes are used, that is, actively expressed and functionally important, in the haploid state, which is not always the case, the presence of a lethal allele can lead to the death of the haploid cell/organism. In this way, the presence of an extended haploid phase of an organisms' life cycle can lead to the elimination of such alleles from the population.

Sexual dimorphism

What, biologically, defines whether an organism is female or male, and why does it matter? The question is meaningless in unicellular organisms with multiple mating types. For example, the microbe *Tetrahymena* has seven different mating types, all of which appear morphologically identical. An individual *Tetrahymena* cell (organism) can mate with another single-celled individual of a different mating type but not with an individual of the same mating type as itself. Mating involves cell fusion and so the identity of the parents is lost; the four cells that are produced by the fused cell (through the process of meiosis) are of one or the other of the original mating types.

In multicellular organisms, the parents do not themselves fuse with one another. Rather they produce cells, known as gametes, that do. Also, instead of multiple mating types, there are usually only two, male and female. This, of course, leads to the question, how do we define male and female? The answer is superficially simple but its implications can be profound. Which sex is which is defined by the relative size of the fusing cells that the organisms produce. The larger fusing cell is termed the egg and an organism that produces eggs is termed a female. The smaller fusing cell, which is often motile (eggs are generally immotile), is termed a sperm and organisms that produce sperm are termed male. At this point, we should note the limits of these definitions. There are organisms that can produce both types of gametes, known as hermaphrodites, after the Greek gods Hermes and Aphrodite. A hermaphroditic organism can self-fertilize. In such cases, males (which produce only sperm) may appear only under certain circumstances. There are organisms that can change their sex, a behavior known as sequential hermaphroditism. For example, in a number of fish it is common for all individuals to originally develop as males; based on environmental cues, the sex of the largest of these males changes to become female.[206]

The size difference between male and female gametes changes the reproductive stakes for the two sexes. Simply because of the larger size of the egg, the female invests more energy in its production (per egg) than a male invests in the production of each sperm cell. It is therefore relatively more important, from the perspective of reproductive success, that each egg produce a viable and fertile offspring. As the cost to the female of generating an egg, and in many organisms, the costs involved in rearing the newly formed offspring increases, the more important the egg's reproductive success becomes. Because sperm are typically small, and so relatively cheap to produce, and because, in many species, males have little investment in rearing their offspring, the selection pressure associated with sperm production and sexual reproduction is often significantly less than that associated with producing an egg and rearing offspring. The end result is that a conflict of interest can emerge between females and males. This conflict of interest increases as the disparity in the relative investment per gamete or offspring increases.

This is an example of evolutionary economics based on cost-benefit analyses. First there is what is known as the two-fold cost of sex, which is associated with the fact that each individual asexual organism can, in theory at least, produce offspring but that two sexually reproducing individuals must cooperate to produce offspring and the resulting offspring are genetically distinct from either parent. Other, more specific factors influence an individual's reproductive costs. For example, the cost to a large female laying a small number of small eggs that develop independently is less than that of a small female laying a large number of large eggs. Similarly, the cost to an organism that feeds and defends its young for some period of time after they are born (that is, leave the body of the female)

[206]Gender-bending fish: http://evolution.berkeley.edu/evolibrary/article/fishtree_07

is larger than the cost to an organism that lays eggs and leaves them to fend for themselves. Similarly, the investment of a female that raises its young on its own is different from that of a male that simply supplies sperm and leaves. As you can imagine, there are many different reproductive strategies (many more than we can consider here), and they all have distinct bio-economic implications, benefits, and constraints. For example, a contributing factor in social evolution is that when raising offspring is particularly biologically expensive, cooperation between the sexes or within groups of organisms in child rearing (protection) can improve reproductive success significantly and increase the return on the investment of the organisms involved. It is important to remember (and be able to apply in specific situations) that the reproductive costs and benefits, and so the evolutionary calculations and conclusions, of the two sexes can diverge dramatically from one another, and that such divergence has behavioral and evolutionary implications.

Consider, for example, the situation in placental mammals, in which fertilization occurs within the female and relatively few new organisms are born from any one female. The female must commit resources to supporting the development and nurturing of the new organisms during the period from fertilization to birth. In addition, female mammals both protect their young and feed them with milk, generated using specialized mammary, that is, milk-secreting glands. Depending on the species, the young are born at various stages of development, from the active and frisky (such as goats (→) to the relatively helpless (humans). During the period when the female feeds and protects its offspring, the female is more stressed and vulnerable than at other times. Under specific conditions, cooperation with other females can occur (as often happens in pack animals) or with a specific male (typically the father) can greatly increase the rate of survival of both mother and offspring, as well as the reproductive success of the male. At the same time, protecting mother and offspring can increase the male's vulnerability. But consider this: how does a cooperating male know that the offspring he is helping to protect and nurture are his? Spending time protecting and gathering food for unrelated offspring is time and energy diverted from the male's search for a new mate and might reduce the male's overall reproductive success, and so could be selected against. Carrying this logic out to its conclusion can lead to behaviors such as males guarding females from interactions with other males.

As we look at the natural world, we see a wide range of sexual behaviors, from males who sexually monopolize multiple females (polygyny) to polyandry, where the female has multiple male "partners." In some situations, no pair bond forms between male and female, whereas in others male and female pairs are stable and (largely) exclusive. In some cases these pairs last for extremely long times; in others there is what has been called serial monogamy, pairs form for a while, break up, and new pairs form. Sometimes females will mate with multiple males, a behavior that is thought to confuse males (they cannot know which offspring are theirs) and so reduces infanticide by males.[207].

It is common that while caring for their young, females are (generally) reproductively inactive. Where a male monopolizes a female, the arrival of a new male who displaces the previous male can lead to behaviors such as infanticide. By killing the young, fathered by another male, the female becomes reproductively active sooner, and so able to produce offspring related to the new male. There are situations, for example in some spiders, in which the male may risk, or even allow itself to be eaten during sexual intercourse as a type of "nuptial gift", which both blocks other males from mating with the female (who is, after all, busy eating and mating) and increases the number of the offspring that result from the mating event. This is an effective reproductive strategy for the male if its odds of mating with a female are low: better (evolutionarily) to mate (reproduce) and die than never

[207] Promiscuous females protect their offspring

to have mated (reproduced) at all. An interesting variation on this behavior is described in a paper by Albo et al.[208] Male *Pisaura mirablis* spiders offer females nuptial gifts, in part perhaps to avoid being eaten during intercourse. Of course where there is a strategy, there are counter strategies. In some cases, instead of an insect wrapped in silk, the males offer a worthless gift, an inedible object (a small stone) wrapped in silk. Females cannot initially tell that the gift is worthless but quickly terminate mating if they discover that it is. This reduces the odds of a male's reproductive success. Over time, as deceptive male strategies become more common, females come to develop counter strategies. For example, a number of female organisms store sperm from a mating and can eject that sperm and replace it with that of another male (or multiple males) obtained from subsequent mating events.[209] Female wild fowl (*Gallus gallus*) can bias the success of a mating event in favor of dominant males; following mating with a more dominant male, they eject the sperm of subdominant males. The result is the production of more robust offspring.[210] This behavior is known as cryptic female choice, cryptic since it is not overtly visible in terms of who the female does or does not mate with. It should be noted that these are not conscious decisions on the part of the female but physiological responses to various cues. And so it goes, each reproductive strategy leads, over time, to counter measures. For example, in species in which a male guards a set of females (its harem), groups of males can work together to distract the guarding male, allowing members of their group to mate with the females. These are only a few of the mating and reproductive strategies that exist.[211] Molecular studies that can distinguish an offspring's parents suggest that "cheating" by both males and females is not unknown even among highly monogamous species. The extent of cheating will, of course, depend on the stakes. The more negative the effects on reproductive success, the more evolutionary processes will select against it.

In humans, a female can have at most one pregnancy a year, while a totally irresponsible male could, in theory at least, make a rather large number of females pregnant during a similar time period. Moreover, the biological cost of generating offspring is substantially greater for the female, compared to the male.[212] There is a low but real danger of the death of the mother during pregnancy, whereas males are not so vulnerable, at least in this context. So, if the female is going to have offspring, it would be in her evolutionary interest that those offspring be as robust as possible, meaning that they are likely to survive and reproduce. How can the female influence that outcome? One approach is to control fertility, that is, the probability that a "reproductive encounter" results in pregnancy. This is accomplished physiologically, so that the odds of pregnancy increase when the female has enough resources to successfully carry the fetus to term. One might argue that the development of various forms of contraception are yet another facet of this type of behavior, but one in which females (and males) consciously control reproductive outcomes.

Sexual selection

As we have already noted, it is not uncommon to see morphological and behavioral differences between the sexes. Sometimes the sexual dimorphism and associated behavioral differences between the sexes are profound; they can even obscure the fact (at least for human observers) that the two sexes are actually members of the same species. In some cases, specific traits associated with one sex can appear to be maladaptive, that is, they might be expected to reduce rather than

---

[208] Worthless donations: male deception and female counter play in a nuptial gift-giving spider

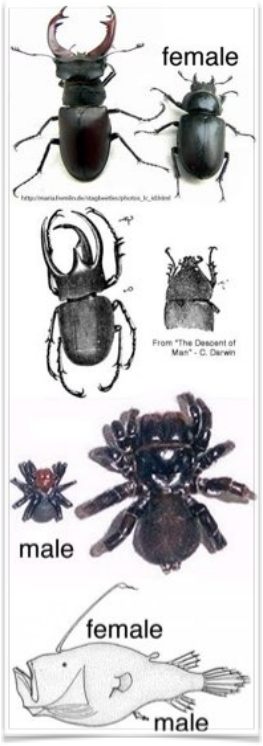
[209] Evolution: Sperm Ejection Near and Far & Sperm Competition and the Evolution of Animal Mating Systems

[210] Female feral fowl eject sperm of subdominant males & Cryptic female choice favors sperm from major histocompatibility complex-dissimilar males

[211] The Evolution of Alternative Reproductive Strategies: Fitness Differential, Heritability, and Genetic Correlations

[212] ' Parental investment

enhance an organism's reproductive potential.[213] The male peacock's tail, the gigantic antlers of male moose, or the bright body colors displayed by some male birds are classic examples (→). Darwin recognized the seriousness of this problem for evolutionary theory and addressed it in his book The Descent of Man and Selection in Relation to Sex (1871). Where the investment of the two sexes in successful reproduction is not the same, as is often the case, the two sexes may have different and potentially antagonistic reproductive strategies. Organisms of different sexes may be "looking" for different traits in their mates. In general, the larger parental investment in the production and rearing of offspring, the less random is mating and the more prominent are the effects of sexual selection, that is, the choice of who to mate with.[214] It is difficult not to place these behaviors in the context of conscious choices, (looking, wanting, etc.), but they appear to be the result of evolved (that is, selected) behaviors and do not imply self-conscious decision making or moral judgements. Presumably, they arise from selection based on costs and benefits. In humans, how consciousness, self-conscious-ness, social organization, ideological and theo-political choices influence sexual behavior (and selection) is even more complex (and way beyond our scope here).



Consider an example in which the female does not require help in raising offspring but in which the cost to the female is high. Selection would be expected to favor a behavior in which females mate preferentially with the most robust, but not necessarily the most cooperative or dependable males available. Females will select their mates based on male phenotype on the (quite reasonable) assumption that the most robust appearing male will be the most likely to produce the most robust offspring. In the context of this behavior, the reproductive success of a male would be enhanced if they could advertise their genetic robustness, generally through visible and unambiguous features. To be a true sign of the male's robustness, this advertisement needs to be difficult to fake and so accurately reflects the true state of the male.[215] For example consider scenarios involving territoriality. Individuals, typically males, establish and defend territories. Since there are a limited number of such territories and females only mate with males that have established and can defend a territory, only the most robust males are reproductively successful. An alternative scenario involves males monopolizing females sexually. Because access to females is central to their reproductive success, males may interact with one another to establish a dominance hierarchy, typically in the form of one or more "alpha" males. Again, the most robust males are likely to emerge as alpha males, which in turn serves the reproductive interests of the females. This type of dominance behavior is difficult to fake. But, cooperation between non-alpha males can be used to thwart the alpha male's monopolization of females.

Now consider how strategies change if the odds of successful reproduction are significantly improved if the male can be counted on to help the female raise their joint offspring. In this situation, there is a significant reproductive advantage if females can accurately identify those males who will, in the future, display this type of reproductive loyalty.[216] Under these conditions (the shared rearing of offspring with a committed male) females will be competing with other females for access to such (perhaps rare) loyal males. Moreover, it is in the male's interest to cooperate with fertile females, and

---

[213] "Flaunting It' - Sexual Selection and the Art of Courtship: http://youtu.be/g3B8hS80k6A

[214] R. Trivers, Parent investment and Sexual selection : http://joelvelasco.net/teaching/3330/trivers72-parentalinvestment.pdf

[215] In Male Rhinoceros Beetle, Horn Size Signals Healthy Mate

[216] From an evolutionary standpoint what is the meaning of romantic love?

often females (but not human females) advertise their state of fertility, that is the probability that mating with them will produce offspring through external signals.

There are of course, alternative strategies. For example, groups of females, including sisters, mothers, daughters, aunts, and grandmothers can cooperate with one another, thereby reducing the importance of male cooperation. At the same time, there may be what could be termed selection conflicts. What happens if the most robust male is not the most committed male? A female could maximize its reproductive success by mating with a robust male and bonding with a committed male, who helps rear another male's offspring. Of course this is not in the committed male's reproductive interest. Selection might favor male's that cooperate with one another to ward off robust but promiscuous and transient males. Since these loyal males already bond and cooperate with females, it may well be a simple matter for them to bond and cooperate with each other. In a semi-counter intuitive manner, the ability to bond with males could be selected for based on its effect on reproductive success with females. On the other hand, a male that commits himself to a cooperative (loyal and exclusive) arrangement with a female necessarily limits his interactions with other females. This implies that he will attempt to insure that the offspring he is raising are genetically related to him. Of course, another possibility is that a loyal male may be attractive to multiple females, who in turn compete for his attention and loyalty. Clearly the outcome of such interactions is influenced by how many females the male can effectively protect (that is, improve their reproductive success) as well as how significant to female reproductive success male cooperation actually is.

The situation quickly gets complex and many competing strategies are possible. Different species make different choices depending upon their evolutionary history and environmental constraints. As we noted above, secondary sexual characteristics, that is, traits that vary dramatically between the two sexes, serve to advertise various traits, including health, loyalty, robustness, and fertility. The size and symmetry of a beetle's or an elk's antlers communicate rather clearly their state of health.[217] The tail of the male peacock is a common example, a male either has a large, colorful and symmetrical tail, all signs of health or it does not – there is little room for ambiguity. These predictions have been confirmed experimentally in a number of systems; the robustness of offspring correlates with the robustness of the male, a win for evolutionary logic.[218]

It is critical that both females and males correctly read and/or respond to various traits, and this ability is likely to be selected for. For example, males that can read the traits of other males can determine whether they are likely to win a fight with that male; an inaccurate determination could result in crippling injuries. A trickier question is how does a one determine whether a potential mate will be loyal? As with advertisements of overall robustness, we might expect that traits that are difficult or expensive to generate will play a key role. So how does one unambiguously signal one's propensity to loyalty and a willingness to cooperate? As noted above, one could use the size and value of nuptial gifts. The more valuable, that is, the more expensive and difficult the gift is to attain, the more loyal the recipient can expect the gift giver to be. On the other hand, once valuable gift-giving is established, one can expect the evolution of traits in which the cost of the gift given is reduced and by which the receiver tests the value of the gift, a behavior we might term rational skepticism, as opposed to naive gullibility.

This points out a general pattern. When it comes to sexual (and social) interactions, organisms have evolved to "know" the rules involved. If the signs an organism must make to another are expensive, there will be selective pressure to cheat. Cheating can be suppressed by making the sign difficult or impossible to fake, or by generating counter-strategies that can be used to identify fakes. These biological realities produce many behaviors, some of which are disconcerting. These include sexual cannibalism, male infanticide, and various forms of infidelity, mentioned above. What we have

---

[217] Attractiveness of grasshopper songs correlates with their robustness against noise

[218] Paternal genetic contribution to offspring condition predicted by size of male secondary sexual character

not considered as yet is the conflict between parents and offspring. Where the female makes a major and potentially debilitating investment in its offspring, there can be situations where continuing a pregnancy can threaten the survival of the mother. In such cases, spontaneous abortion (ending the pregnancy) could save the female, who can go on and mate again. In a number of organisms, spontaneous abortion occurs in response to signs of reproductive distress in the fetus. Of course, spontaneous abortion is not in the interest of the offspring and we can expect that mechanisms will exist to maintain pregnancy, even if it risks the life of the mother, in part because the fetus and the mother, while related are not identical; there can be a conflict of interest between the two.[219]

There are many variations of reproductive behavior to be found in the biological world and a full discussion is beyond our scope here. It is a fascinating subject with often disconcerting moral implications. Part of the complexity arises from the fact that the human brain (and the mind it generates) can respond with a wide range of individualistic behaviors, not all of which seem particularly rational. It may well be that many of these are emergent behaviors; behaviors that were not directly selected for but appeared in the course of the evolution of other traits, and that once present, play important roles in subsequent behavior and evolution. Such emergent traits may be difficult or impossible to remove or modify, evolutionarily, if they are integral to the primary function of the trait.

One of the most robust and reliable findings in the scientific literature on interpersonal attraction is the overwhelming role played by physical attractiveness in defining the ideal romantic partner. Both men and women express marked preference for an attractive partner in a non- committed short-term (casual, one night stand) relationship.

For committed long-term relationships, females appear to be willing to relax their demand for a partner's attractiveness, especially for males with high social status or good financial prospects.

Males also look for various personality qualities (kindness, understanding, good parental skills) in their search for long-term mating partners, but unlike females, they assign disproportionately greater importance to attractiveness compared to other personal qualities.

The paramount importance of attractiveness in males' mate choices has been recently demonstrated by using the distinction between necessities (i.e., essential needs, such as food and shelter) and luxuries (i.e., objects that are sought after essential needs have been satisfied, such as a yacht or expensive car) made by economists.

Using this method, Li et al., reported that males treat female attractiveness as a necessity in romantic relationships; given a limited "mating budget," males allocate the largest proportion of their budget to physical attractiveness rather than to other attributes such as an exciting personality, liveliness, and sense of humor.

- from Mating strategies for young women by Devendra Singh (2004).

Questions to answer

60. How it is possible that individuals of different sexes can be in conflict, reproductively, and how do such differences impact sexual selection?
61. How it is possible that a parent's interests can conflict with the interests of its offspring?
62. Why do the different sexes often display different traits?
63. If the two sexes appear phenotypically identical, what might you conclude (at least tentatively) about their reproductive behaviors?

Curbing "runaway" selection

Sexual selection can lead to what has been termed, but is not really, runaway selection. For example, the more prominent the peacock male's tail the more likely he will find a mate even though larger and larger tails also have significant negative effects. All of which is to say that there will be both positive and negative selection for tail size, which will be influenced by the overall probability that a particular male mates successfully. Selection does not ever really run away, but settles down when the positive benefit, in terms of sexual success, and the negative cost of a trait come to be roughly equal to each other. Sufficient numbers of male peacocks emerge as reproductively successful even if many males are handicapped by their tails and fall prey to predators. In part, this is due to the fact that, in peacocks, there is a reproductive skew for males, that is, a significant number of males in a population will never successfully mate and have offspring. In contrast, almost all females have offspring. For another example, consider the evolution of extremely large antlers

[219] Maternal-Fetal Conflict: https://www2.aap.org/sections/bioethics/PDFs/Curriculum_Session14.pdf

associated with male dominance and mate accessibility, such as occurred in *Megaloceros giganteous* (→). These antlers can be expected to act to constrain the animal's ability to move through heavily wooded areas. In a stable environment, the costs of generating antlers and the benefits of effective sexual advertising would be expected to balance out; selection would produce an optimal solution. But if the environment changes, pre-existing behaviors and phenotypes could act to limit an organism's ability to adapt or to adapt fast enough to avoid extinction. In the end, as with all adaptations, there is a balance between costs and benefits, particularly within a changing environment.

Summary: Social and ecological interactions apply to all organisms, from bacteria to humans. They serve as a counter-balance to the common caricature of evolution as a ruthless and never ceasing competition between organisms. This hyper-competitive view, often known as the struggle for existence or Social Darwinism, may be appealing to ruthless (anti-union / anti-social constraint) capitalists but was not supported by Darwin or by scientifically-established evolutionary mechanisms. It has been promulgated by a number of pundits who used it to justify various political (that is, inherently non-scientific) positions, particularly arguing against social programs that helped the poor (often characterized as the unfit) at the "expense" of the wealthy (who might be viewed as parasites). Assuming that certain organisms were inherently less fit, and that they could be identified, this view of the world gave rise to eugenics, the view that genetically inferior people should be removed from the population or sterilized, before their "bad" traits overwhelmed a particular culture. Eugenics was an influential ideology in the United States during the early part of the 20th century and inspired the genocidal programs of the Nazis in Germany. What is particularly odd about this evolutionary perspective is that it is actually anti-evolutionary, since if the unfit really were unfit, they could not possibly take over a population. In addition, it completely ignores the deeply social (cooperative) aspect of the human species.

Questions to answer

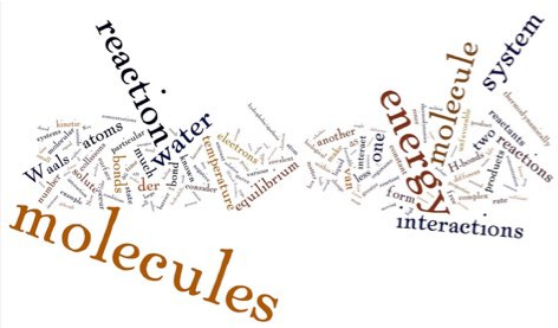
64.What does it mean to cheat, in terms of sexual selection - is a "cheating" organism consciously deceptive?
65. Are there specific types of "cheating" behaviors that females use with males? or males with females?
66. What are the costs involved when a male tries to monopolize multiple females? What are the advantages?
70. What limits runaway selection, or better, why is runaway selection impossible

Questions to ponder

- Should human ethical or ideological beliefs and decisions be more important than evolutionary cost-benefit calculations?

Chapter 5: Getting molecular: interactions, thermodynamics & reaction coupling

In which we change gears, from evolutionary mechanisms to the physicochemical properties of organisms. These physicochemical properties shape and constrain evolutionary possibilities and biological behaviors. We consider how molecules interact and react with one another and how these interactions and reactions determine the properties of substances and systems, particularly the bounded, non-equilibrium system that is life.



Just enough thermodynamics (for now)

While the diversity of organisms and the properties of each individual organism are the products of evolutionary processes initiated billions of years ago, it is equally important to recognize that all biological systems and processes, from cell growth and division, movement, and differentiation to thoughts and feelings, obey the rules of chemistry and physics, The laws of thermodynamics and the ways atoms interact. What makes biological systems unique is that, unlike simpler physicochemical systems that move toward thermodynamic equilibrium, organisms must maintain an uninterrupted non-equilibrium state in order to remain alive. While a chemical reaction system is easy to assemble *de novo*, every current biological system (cells and organisms) has been running continuously for billions of years. So, before we continue we have to be clear about what it means and implies when we say that a system is at equilibrium versus being in a obligate non-equilibrium state, since a biological system at equilibrium is dead, and dead in an (apparently) irreversible state.

To understand the meaning of thermodynamic equilibrium we have to learn to see the world differently, and learn new meanings for a number of words. First we have to make clear the distinction between the macroscopic world that we perceive directly and the sub-microscopic, molecular world that we can understand only through scientific observations and conclusions, and some knowledge of atomic and molecular behavior – it is this molecular world that is particularly important in the context of biological systems. The macroscopic and the molecular worlds behave very differently - in particular, the molecular world often behaves stochastically (that is unpredictably). To illustrate this point we will use a simpler model that displays the basic behaviors that we want to consider but is not as complex as a biological system. In our case let us consider a small, well-insulated air-filled room in which there is a table upon which is resting a bar of gold – we use gold since it is chemically rather inert, that is, un-reactive. Iron bars, for example, could rust, which would complicate things. In our model the room is initially at a cosy 70 ºF (~21 ºC) and the gold bar is at 200ºC. What will happen as a function of time; try and generate a graph that describes how the system behaves.

Our first task is to define the system – that is, the part of the universe we are interested in. We could define the system as the gold bar or the room with the gold bar in it. Notice, we are not really concerned about how the system came to be the way it is - that is, its history. We could, if we wanted to, demonstrate convincingly that (for simple systems like this one) the system's history has no influence on its future behavior – this is a critical difference between biological and simple physicochemical systems. We are, however, concerned as to whether the system is open or closed, that is whether energy and matter can enter or leave the system. For now we will consider the room to be an effectively closed (isolated) system - no energy enters or leaves it.

Common sense tells us that energy will be transferred from the gold bar to the rest of the room and that the temperature of the gold bar will decrease over time, while the final temperature of the room + the gold bar will depend upon relative sizes of both (hope this makes sense). This energy

transfer occurs primarily through molecular collisions between the molecules of the gold bar together with the molecules in the air and the table. The behavior of the system has a temporal direction. Why do you think that is? Why, exactly, doesn't the hot bar get hotter and the rest of the system, the room, get cooler? We will come back to this question shortly. What may not be quite as obvious is that the temperature of the room will increase slightly as the gold bar cools. Eventually the block of gold and the room will reach the same temperature; when that happens, the system will be said to be at thermal equilibrium.

Remember we defined the system as closed; no matter or energy passes into or out of the room. In a closed system, once the system reaches its final temperature no further macroscopic changes occur. The key here is the word macroscopic, which for our purposes means directly observable. This does not mean, however, that nothing is going on. If we could look at the molecular level we would see that molecules of air are moving, constantly colliding with one another and colliding with the particles within the bar, the table, and the walls of the room. The molecules within the bar and the table are also vibrating. The speeds of these molecular movements are a function of temperature, the higher or lower the temperature, the faster or slower these motions, on average, will be. Collisions between molecules can change the velocities of the colliding molecules. What would happen if there was no air in the room or if it were possible to suspend the gold bar in the center of the room, for example if the room were in outer space?

All of the molecules in the system have kinetic energy, the energy of motion, and as a consequence of their interactions (primarily collisions), the kinetic energy of any one particular molecule will change over time. At the molecular level the system is dynamic, even though at the macroscopic level it is static (provided that the system is large enough). And this is what is important about a system at equilibrium: it is macroscopically static, there is no net change possible, even though at the molecular level there is still plenty of movement. The energy of two colliding molecules is the same after a collision as before, even though the energy may be distributed differently between the colliding molecules. In physical terms, the system as a whole cannot do anything; it cannot do work - no macroscopic changes are possible. This is a weird idea, since (at the molecular level) things are still moving. So, as we return to living systems, which are clearly able to do lots of things, including moving macroscopically, growing, thinking, and such, it is clear that they cannot be at equilibrium. We will come back to this insight repeatedly.

We can ask, then, what is necessary to keep a system from reaching equilibrium? The most obvious answer (we believe) is that unlike our imaginary closed system, a non-equilibrium system must be open, that is, energy and matter must be able to enter and leave the system. An open system is no longer isolated from the rest of the universe, it is part of it. Whether the Universe as a whole is open or closed, it is clearly "non-homogenous", that is there are stars emitting tremendous amounts of energy, that maintain non-equilibrium regions. The Earth, and everything on it, is part of a non-equilibrium system, driven by radiation from the Sun (as well as processes such as the radioactive decay of isotopes). If we consider our room with the gold bar, we could maintain a difference in the temperature between the bar and the room by illuminating the bar and removing heat from the room as a whole. A temperature difference between the bar and the room could then (in theory) produce what is known as a heat engine that could do work. As long as we continue to heat the block and remove heat from the rest of the system, it could continue to do work, that is, macroscopically observable changes could happen.

Cryptobiosis: At this point, we have characterized organisms as dynamic, open, non-equilibrium systems. An apparent exception to the dynamic aspect of life are organisms that display a rather special phenotypic adaptation, known generically as cryptobiosis. Organisms, such as the tardigrade or water bear (→), can be freeze-dried and persist in a state of suspended animation for decades. What is critical to note, however, is that when in this cryptobiotic state the organism is not at equilibrium, in much the same way that a battery or piece of wood in

air is not at equilibrium, but capable of reacting. The organism can be reanimated when returned to normal conditions.[220] Cryptobiosis is a genetically-based adaptation that takes energy to produce and energy is needed to emerge from stasis. While the behavior of tardigrades is extreme, many organisms display a range of adaptive behaviors that enable them to survive hostile environmental conditions.

Reactions and energy: favorable and unfavorable, their dynamics and coupling

Biological systems are extremely complex. Both their overall structural elements and many of their molecular components (including DNA and proteins) are the products of thermodynamically unfavorable reactions. How do these reactions take place in living systems? The answer involves the coupling of thermodynamically favorable reactions to thermodynamically unfavorable reactions. This is a type of work, although not in the standard macroscopic physics type of work (w) = force x distance. In the case of (chemical) reaction coupling, the work involved drives thermodynamically unfavorable reactions, typically the synthesis of large and complex molecules and macromolecules (that is, very large molecules). Here we will consider the thermodynamics of these processes.

Thermodynamics is, at its core, about changes in energy. This leads to the non-trivial question, what is energy? Many have struggled to provide an unambiguous answer to this question, and there is no simple satisfactory answer. Perhaps a way around it is to say that for every change to a system, there is an associated change in energy; this implies that such changes can be unambiguously recognized. While it may appear that there are many types of energy (and you may have been taught that this is the case) in fact there are only two forms of energy, kinetic and potential. For example, the energy associated with the movement and vibrations of objects with mass is kinetic energy. Potential energy is associated with an object's position in a field (electrical, magnetic, gravitational) and the particle's nature, its mass, electrical charge, and characteristics, such as "spin". All systems, whether they are macroscopic, microscopic, atomic or sub-atomic can be characterized in terms of the sum of their kinetic and potential energies. But wait, you might say, what about the energy associated with electromagnetic radiation, the most familiar form of which is visible light. Electromagnetic radiation is a form of kinetic energy, energy that is transferred from place to place via photons. Finally, there is the counterintuitive idea that energy and matter, are interconvertible as described by the famous equation:

$$e \text{ (energy)} = m \text{ (mass)} \times c^2 \text{ (c = speed of light)}$$

but not to worry, such interconversion events are not directly relevant to biological systems.

That said, it is clear that kinetic energy can be converted into potential energy and vice versa. To illustrate this principle, we can call on our day-to-day experiences. Forces (which mediate the transfer of energy) can be used to make something move. Imagine a system of a box sitting on a rough floor. You shove the box so that it moves (but do not continue to push it) – the box travels some distance and then stops. By shoving the box you added (kinetic) energy to the system. The first law of thermodynamics states that the total energy in a system is constant. So the question is where has the energy gone when the box slows and stops moving? One answer might be that the energy was destroyed - but the first law of thermodynamics implies that that cannot be the case. Careful observations lead us to conclude that the energy still exists and that it has been transformed and/or transferred somewhere else. Measurements can prove that the mass of the box has not changed. In fact, if we measured the temperature of both the box and the floor we would see that both have increased (by a very small amount). The friction associated with moving the box results in an increase in the movements of the molecules of the box and the floor. Through collisions and vibrations this energy will, over time, be distributed throughout the system–the temperature of the system will increase (if only slightly). The presence of this thermal motion is revealed by what is known as Brownian motion. In 1905, Albert Einstein explained Brownian motion in terms of the

[220] On dormancy strategies in tardigrades & Towards decrypting cryptobiosis

existence, size, and movements of molecules.[221]

In the system we have been considering, the energy that was transferred to the box by pushing it has been spread throughout the system. While one can use a directed push (input of energy) to move something (to do work), the diffuse thermal energy cannot be used to do work. While the total amount of energy is conserved, its ability to do things has decreased (almost abolished). This involves the concept of entropy, which we will turn to next.

Questions to answer:

67. How does energy move from molecule to molecule within a system?
68. What are the common components of a non-equilibrium system; how might you identify such a system.

Questions to ponder

- How is it that a dried out tardigrad can still be alive?

Thinking entropically (and thermodynamically)

We certainly are in no position to teach you (rigorously) the basics of physics, chemistry, and chemical reactions, but we can provide a short refresher that focuses on the key points we will be using over and over again.[222] The first law of thermodynamics is that the total amount of energy within a closed system is constant. The energy may be transformed from kinetic to potential (and vice versa) but in a closed system the total does not change. Again, we need to explicitly recognize the distinction between a particular system and the universe as a whole, although the universe as a whole is itself (as far as we know) a closed system. For any system we must define a system boundary; this can be a real boundary such as a container, or an imaginary boundary. What is inside the boundary is part of the system, and the rest of the universe outside of the boundary layer is not. While we will consider the nature of the boundary of biological systems (cells) in greater molecular detail in the next chapter, we can anticipate that one of the boundary's key features is its selectivity in what it lets pass into and out of the system, the constraints it imposes on those movements.

Assuming that you have been introduced to chemistry, you might recognize the Gibbs free energy equation: $\Delta G = \Delta H - T\Delta S$, where T is the temperature of the system.[223] From our biological perspective, we can think of ΔH as the amount of thermal energy transferred between the system and the surroundings during any change, and ΔS as the change in a system factor known as entropy. Entropy is related to the ways that energy and matter can be arranged, and the more possible ways, the greater the entropy. In the earlier example of the gold bar in the isolated room, energy is transferred between the bar and the room until the two are at equal temperature; over time, the bar and the room come to equilibrium. The process does not run in reverse, the bar does not get hotter while the room cools. This is because transferring energy from hot to cold is very much more probable statistically (See CLUE:Chemistry for a more detailed discussion). The number of arrangements of energy and matter are greater when energy flows from hot to cold, than when it flows from cold to hot. The factor that we use to characterize these probabilities is called entropy (S). Often entropy is used colloquially to describe random or disordered systems, or the "state" of a substance, and it is true that a gas (which is more disordered) has more entropy than a liquid or a solid of the same substance (which is less disordered). The gas has greater entropy because there

---

[221] Albert Einstein: The Size and Existence of Atoms http://youtu.be/nrUBPO6zZ40

[222] Of course, we recommend a chemistry course sequence based on Cooper & Klymkowsky, 2014. Chemistry, Life, the Universe and Everything: here: http://clue.chemistry.msu.edu/

[223] in the real world, the value of ΔG depends upon the concentrations of solute and solvent, but we will ignore that complexity for the moment.

are more possible ways that the gas particles and their associated energies can be arranged, compared to a solid where the particles are fixed in place.

For any change, the entropy of the universe always increases - which is usually stated as the Second Law of Thermodynamics, a behavior that has never been found to be violated. At this point you might be saying wait a minute, aren't there systems in which entropy decreases? For example, it is certainly possible to change a gas (high entropy) into a liquid or a solid (lower entropy), but the critical part here is that this system is not closed. While the system may decrease in entropy, the entropy of the universe as a whole still increases. This is because when gas condenses to a liquid energy must be removed and that energy is transferred to the surroundings, which increases the entropy of the surroundings by making molecules move and vibrate faster. While the entropy of a particular region of the universe (the system) may decrease, the total entropy of the universe always increases.

It turns out that it is difficult to measure energy and entropy changes for the universe. Usually we can only do this for the system we are studying. Fortunately there is a way to account for the total entropy change during a process (or reaction) using the equation $\Delta G = \Delta H - T\Delta S$, which tells us about the change in energy (and therefore entropy) for a process within a system. When $\Delta G < 0$ we say the change is thermodynamically favorable, and can occur. Conversely when $\Delta G > 0$ we say the change is thermodynamically unfavorable, and will not occur. When ΔG for the system = 0 no observable, that is macroscopic changes will occur. The system is at equilibrium.

A reaction is characterized by its equilibrium constant, K_{eq}, that is a function of the reaction itself, the concentrations of the reactants, and system temperature and pressure. In biological systems we generally ignore pressure (and only occasionally consider temperature), although both may be important for organisms that live on the sea floor, mountain tops, or hydrothermal vents.

The equilibrium constant (K_{eq}) for the reaction A + B ⇆ C + D is defined (→) as the product of the concentrations of the products (C and D) divided by the product of the concentrations of the reactants at equilibrium, where nothing macroscopic is happening. At equilibrium the concentrations do not change (that is why K is a constant). For a thermodynamically favorable reaction, that is one that favors the products, K will be greater, often much greater, than one. The larger K_{eq}, the more product and the less reactant there will be when the system reaches equilibrium. If the equilibrium constant is less than 1, then at equilibrium, the concentration of reactants will be greater than the concentration of products.

$$K = \frac{[C][D]}{[A][B]}$$

While the concentration of reactants and products of a reaction at equilibrium remain constant it is not the case that the system is static. If we were to peer into (or imagine) the system at the molecular level we would find that reactants are continuing to form products and products are rearranging to form reactants at equilibrium; the rate of the forward reaction is equal to the rate of the reverse reaction, although both may be very slow.[224] If, at equilibrium, a reaction has gone almost to completion and $K_{eq} >> 1$, there will be very little of the reactants left and lots of the products. Most reactions involve collisions between molecules. The frequency of productive collisions between reactants or products increases as their concentrations increase. Consider the equilibrium state for a highly favorable reaction; the high concentration of products (produced by the reaction) x low probability of effective collisions will equal the low concentration of reactants (remaining) x higher probability of effective collisions.

[224] This, of course, assumes that we have a closed system, that is, that neither the products or the reactants can leave the system, and that the volume of the system also remains constant. If the reactants can "leave the scene" of the reaction, then of course the back reaction, Products ⇆ Reactants, will be much less likely to occur.

Reaction rates

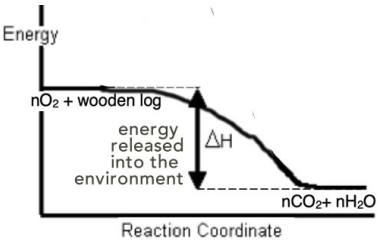
Knowing that a reaction is thermodynamically favorable does not tell us much (or really anything) about whether the reaction occurs to a significant extent under a particular set of conditions. For example, consider a wooden log, composed mainly of the carbohydrate polymer cellulose $(CH_2O)_n$. In the presence of molecular oxygen (O_2) the reaction:

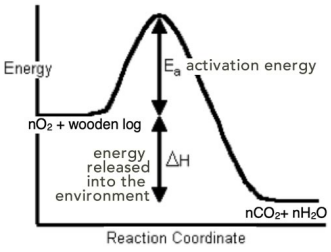
$$nO_2 + \text{wooden log } ((CH_2O)_n) \leftrightarrows nCO_2 + nH_2O + \text{heat}$$

is extremely favorable, thermodynamically, that is. It has a large negative ΔG and a large equilibrium constant (once the reaction starts it goes completely to CO_2 and H_2O). Yet logs are stable - they do not spontaneously burst into flames. The question is, of course, why not? Or more generally why is the world so annoyingly complex?

The answer to this conundrum lies in the fact that both the equilibrium constant and ΔG (or for the more chemically rigorous, ΔG^o) tell us only about whether a reaction is thermodynamically favorable, but they tell us nothing about how fast that reaction will proceed; nothing about whether the reaction will occur under a specific set of conditions. For that we have to turn to the study of reaction rates, also known as reaction kinetics; this requires us to consider the various factors that affect the reaction. In general a reaction will go faster if there are more reactant molecules. For example, in the case of the log and oxygen, oxygen molecules (O_2) must come in contact with the log. Reactant molecules must collide to initiate a reaction. In air (at sea level) O_2 molecules amount to ~20% of the total molecules present. If we increase the O_2 concentration, the log will burn much faster and brighter, because there are more collisions to initiate the reaction.[225] Under normal conditions, however, the log will not start burning spontaneously - added energy is needed. Why? Because the transition between reactants and products requires the breaking of bonds; bond breaking requires the addition of energy and generally the addition of more energy that is available through molecular collisions. The energy required for bonds to break and the reaction to proceed, over and above the energy of the reactants, is known as activation energy. The presence of activation energy explains why chemical systems, such as life, do not quickly move to equilibrium. Why nucleic acids and proteins do not quickly react to produce more stable (but rather more boring) molecules such as CO_2, H_2O, and NH_3 from which they are composed.

To explore the idea of activation energy, let us consider the very simplified model of a log burning in air to produce CO_2 and H_2O, a reaction that is, in fact, quite complex. We could represent this process on a graph of energy (or more accurately Gibbs Free Energy (G)) vs reaction progress like this (→). As the reaction proceeds, a great deal of heat is released into the surroundings; this released energy corresponds to the ΔH between reactants and products. The graph also indicates that the products are more stable (lower energy) than the reactants. But, the reaction energy graph does not give us any indication that energy must be added to start the reaction, or why. If we add in this energy the graph would look like this (↓). The activation energy (E_a) is the energy needed to break the bonds within wood molecules and in O_2. This step, in which pre-existing bonds are broken but new bonds have not yet formed is also known as the transition state. In general the amount of activation energy needed determines the rate of the reaction. If most collisions supply this (or more) energy, the reaction will proceed rapidly, its rate will be fast. If, on the other hand and in the case of a log at room temperature, few if any collisions supply enough energy to break the bonds necessary to start the reaction, the reaction rate will be slow or will, essentially, not occur at all (discussed further below). For the wood burning reactions, the energy needed to start the reaction may involve a downed electrical line, a

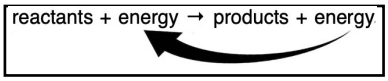




[225] This is one reason why smoking is not a good idea for people who have to use supplementary oxygen to breathe

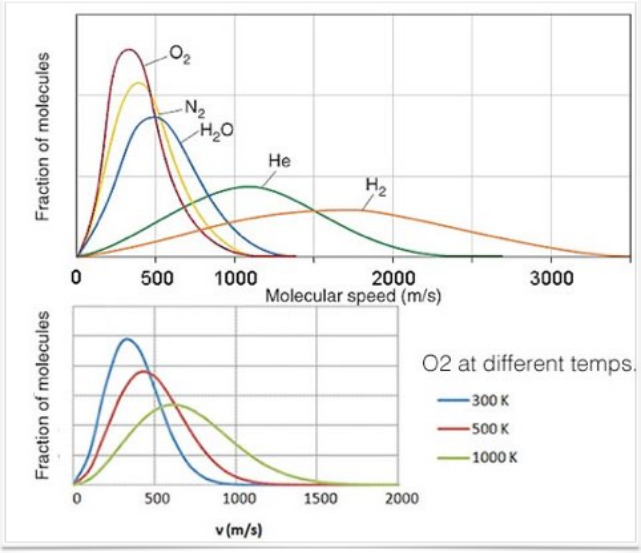
lightening strike, or a burning match.

Once the the reaction starts the energy released when new bonds are formed will be released into the environment. The resulting increase in the temperature (average kinetic energy of molecules) of the reaction system results in more collisions that provide more than the needed amount of activation energy. The result is that the reaction rate will increase and the reaction will become self-sustaining - a form of a positive feedback loop (→). As reactants are used up, however, productive collisions, that is collisions between reactants with sufficient energy, become rarer, the reaction rate slows and less energy is released. At the same time, collisions between products will increase - although as energy is dissipated into the environment, only very rare events will have sufficient energy to break the bonds of the products, the first step in the reverse reaction.

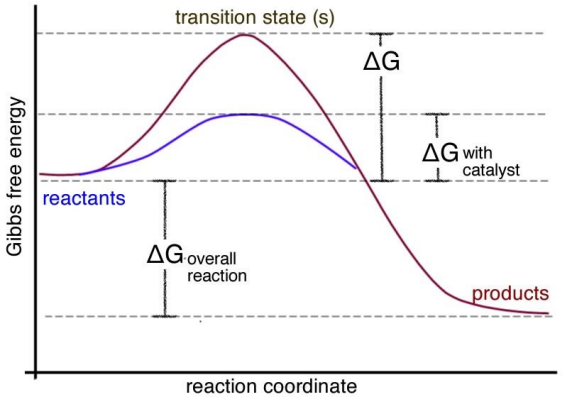


Activation energy and catalysis in biological systems

As noted above, the reason why (most) thermodynamically favorable reactions do not occur immediately when reactants come into contact is that bonds must be broken for the reaction to occur, and breaking bonds, particularly covalent bonds, requires a large amount of energy. In biological systems there are two major sources for this energy: light and collisions with other molecules. A molecule can absorb a photon (a particle of light) or energy can be transferred through collisions with other molecules. In liquid water, molecules are moving; at room temperature they move on average at about 640 meters/second. That is not to say that all molecules are moving with the same speed. If we were to look at the population of molecules, we would find a distribution of speeds known as a Boltzmann (or Maxwell-Boltzmann) distribution (←). As they collide with one another, the molecules exchange kinetic energy, and one molecule can emerge from a collision with much more energy than it entered with. Since reactions occur at temperatures well above absolute zero, there is plenty of energy available in the form of the kinetic energy of molecules.



But, biological systems are constrained in a number of ways. As we will see, the three-dimensional structure of many macromolecules, particularly proteins and nucleic acids, is critical to their normal function, and their 3D structure is basically unstable - even small changes (by the standards of a typical chemistry lab) in temperature can lead to what is known as denaturation and the loss of function. The take home message is that biological systems have to use alternative strategies to control the rates of the reactions they depend upon. Their solution are molecules that act as catalysts. But what exactly does a catalyst do? Basically, it lowers the energy required to reach the transition state (the activation energy) of a reaction by interacting with the reactants (→). The result is that at any particular temperature, the reaction rate will be increased in the presence of an active catalyst. An important feature of biological catalysts, typically proteins - known as enzymes, and nucleic acids - known as ribozymes, is that their activity can be regulated. Their effectiveness as a catalyst for specific reactions can be turned on or off. As we will see, the



regulate-ability of biological catalysts is central to maintaining the dynamic, non-equilibrium state of the cell.

Questions to answer:

69. Where does the energy come from to reach (and pass through) the transition state?
70. A reaction is at equilibrium; we increase the amount of reactant or product. What happens (over time) to the amounts of reactants and products?
71. What does reducing the activation energy of a reaction do to a system at equilibrium? What does it do to a system far from equilibrium?
72. How and why does the feedback system of a burning log change over time?

Question to ponder:

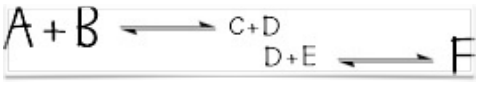
– Propose a model for how (at the molecular level) a catalyst might lower a reaction's activation energy?

Coupling reactions

There are large numbers of different types of reactions that occur within cells. As a rule of thumb, a reaction that produces smaller molecules from larger ones will be thermodynamically favored, while reactions that produce larger molecules from smaller ones will be unfavorable. Similarly a reaction that leads to a molecule moving from a region of higher concentration to a region of lower concentration will be thermodynamically favorable. So how exactly can we build big molecules, such as DNA and proteins, and generate the concentration gradients upon which life depends?

As we noted before reactions can be placed into two groups, those that are thermodynamically favorable (negative ΔG^o, equilibrium constant greater, typically much greater, than 1) and those that are thermodynamically unfavorable (positive ΔG^o, equilibrium constant less, often much less than 1). Thermodynamically favored reactions are typically associated with the breakdown of various forms of food molecules and the release of energy, known generically as catabolism. Reactions that build up biomolecules, known generically as anabolism, are typically thermodynamically unfavorable. An organism's metabolism is the sum total of all of these various reactions. The question is, if a reaction is unfavorable - how can it occur?

The answer to this conundrum lies in the fact that when such a reaction is coupled to a thermodynamically favorable reaction, the unfavorable reaction can be made to occur. The important factor here is that the two reactions share a common intermediate - that is they are "coupled". In this example (↓) there are two reactions occurring at the same time that share the component "D". Let us



assume that the upper reaction is unfavorable while the lower reaction is favorable. Let us further assume that both reactions are occurring at measurable rates and that E is already present within the system. What happens? At the start of our analysis, the concentrations of A and B are high, and C and D are low. We can then use Le Chatelier's principle to make our predictions. Le Chatelier's principle states that if a change is made to a system at equilibrium, then the system will shift to counteract that change, basically because the number of productive collision events associated with one direction of the reaction will increase compared to those associated with the other direction.[226]

Let us illustrate how Le Chatelier's principle works. Assume for the moment that the reaction

$$A + B \leftrightharpoons C + D$$

has reached equilibrium, that is, the rates of the forward and reverse reactions are equal. Now consider what happens to the reaction if, for example, we remove (somehow, do not worry about how) C from the system. Now the rate of the reverse reaction will decrease because there is not as much C to collide with D to initiate the reaction. This means that the rate of the forward reaction will

[226] http://en.wikipedia.org/wiki/Le_Chatelier's_principle

become greater than the reverse reaction: the reaction is no longer at equilibrium. More A and B will react to give C and D, even though that reaction is thermodynamically unfavorable. Similarly if we add B, the rate of the forward reaction will increase and the reaction will move to the right to produce more products, until a new equilibrium position is established. In this case, the addition of B leads to the increased rate of production of C + D until their concentration reaches a point where the rate of the

$$C + D \rightarrow A + B \text{ reaction is equal to the } A + B \rightarrow C + D \text{ reaction.}$$

This type of behavior arises directly from the fact that at equilibrium reaction systems are not static but dynamic at the molecular level – things are still occurring but at the same rate so that there is no net change. When you add or take something away from the system, it becomes unbalanced. Because the reactions are occurring at measurable rates, the system will return to equilibrium over time.

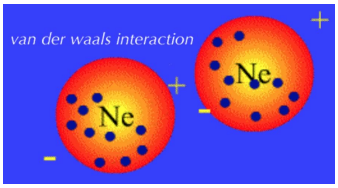
So back to our system of coupled reactions. As the unfavorable A+B reaction occurs and approaches equilibrium it will produce a small amount of C+D. However, the D+E reaction is favorable, and as D is formed it will react with E to produce F, which removes D from the system. As D is removed, it influences the A+B reaction by making the C+D "back reaction" less probable even as the A+B "forward reaction" continues. The result is that more C and D will be produced. Assuming that a sufficient amounts of E is present, more D will be removed. The end result is that, even though it is energetically unfavorable, more and more C and D will be produced, while D will be used up to make F. It is the presence of the common component D and its use as a reactant in the D+E reaction that drives the synthesis of C from A and B, something that would normally not be expected to occur to any great extent. Imagine then, what happens if C is also a reactant in some other favorable reaction(s)? In this way reactions systems are linked together, and the biological system proceeds to use energy and matter from the outside world to produce the complex molecules needed for its maintenance, growth, and reproduction.

Questions to answer:

73. How does adding or removing components of the reaction system change the energy of the system?
74. How is LeChatelier's principle involved in reaction coupling?
75. How would you go about deciding whether the system involved coupled reactions?
76. Assume that the reactions within a reaction system require catalysts to occur at reasonable rates; what happens within reaction systems if the catalysts are missing or inactive?
77. Why are catalysts required for life to be possible?

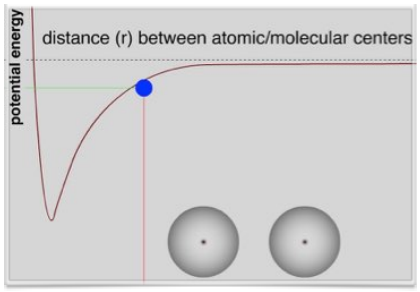
Inter- and Intra-molecular interactions

We have briefly (perhaps too briefly) considered what energy is and have begun to think about how it can be transferred within reaction systems. Now we need to consider what we mean by matter, which implies an understanding of the atomic organization of the molecules that compose matter. As you hopefully know by now, all matter is composed of atoms. The internal structure of atoms is the subject of quantum physics and we will not go into it in any depth. Suffice it to say that each atom consists of a tiny positively charged nucleus and a cloud of negatively charged electrons. Typically atoms and molecules, which after all are collections of atoms, interact with one another through a number of different types of forces. Chemists typically define both as van der Waals interactions, but we will distinguish two types - one common to all molecules, and associated with transient (induced) electrical dipoles and the second associated with permanent dipoles within molecules. The first of these are termed London Dispersion Forces. These forces arise from the fact that the relatively light (in terms of mass) negatively-charged electrons are in continual movement, compared to the relatively massive and stationary positively-charged nuclei (←). Because charges on the protons and electrons are

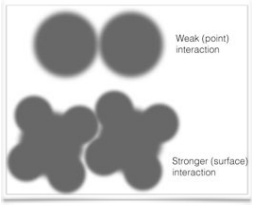


equal in magnitude the atom is electrically neutral, but because the electrons are moving, at any one moment, an observer outside of the atom or molecule will experience a small fluctuating electrical field. At any given instant of time, there may be an unequal distribution of negative charge in a given atom or molecule - an instantaneous dipole.

As two molecules approach one another the distorted electron cloud of one will induce a distortion of the electron cloud of the other (an induced dipole). This results in an attractive force, named after its discoverer Fritz Wolfgang London (1900–1954). This London Dispersion Force (LDF) varies as ~$1/R^6$ where R is the distance between the molecules. As a result LDFs act over very short distances, typically less than 1 nanometer (1 nm = 10^{-9} m). As a frame of reference, a carbon atom has a radius of ~0.07 nm. The magnitude of this attractive force reaches its maximum when the two molecules are separated by what is known as the sum of their van der Waals radii (the van der Waals radius of a carbon atom is ~0.17 nm (→). If they move closer than this distance, the attractive LDF is quickly overwhelmed by the rapidly increasing, and strongly repulsive forces that arise from the electrostatic interactions between the negatively charged electrons of the two molecules. Each atom and molecule has its own characteristic van der Waals radius, although since most molecules are not spherical, it is better to refer to a molecule's van der Waals surface. This surface is the closest distance that two molecules can approach one another before repulsion kicks in and drives them back away from one another. It is common to see molecules displayed in terms of their van der Waals surfaces. Every molecule generates LDFs when it approaches another, so LDF-mediated interactions are universal.



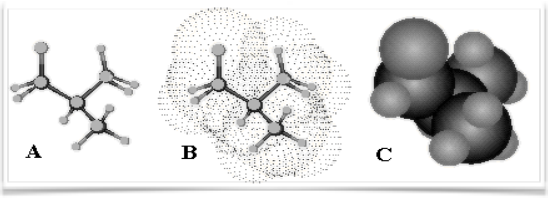
The strength of the LDF-mediated interactions between molecules is determined primarily by their shapes. The greater the surface complementarity between two molecules, the stronger their interaction. Compare the interaction between two monoatomic Noble atoms, such as helium, neon or argon, and two molecules with more complex shapes (→). The two monoatomic particles interact via LDFs at a single point, so the strength of the interaction is minimal. On the other hand, the two more complex molecules interact over extended surfaces, so the LDFs between them are greater, resulting in a stronger van der Waals interaction.



Covalent bonds

In the case of van der Waals interactions, the atoms and molecules involved retain a hold on their electrons, they remain distinct and discrete. There are cases, however, where atoms come to "share" each other's electrons; sharing involves pairs of electrons, one from each atom. When electron pairs are shared, the atoms stop being distinct in that their shared electrons are no longer restricted to one or the other. In fact, since one electron cannot, even in theory, be distinguished from any other electron, they become a part of the molecule's electron system. This sharing of electrons produces what is known as a covalent bond. Covalent bonds are ~20 to 50 times stronger than the interactions based on LDFs. What exactly does that mean? Basically, it takes 20 to 50 times more energy to break a covalent bond compared to the energy needed to break an LDF-mediated interaction. While the bonded form of atoms in a molecule is always more stable than the unbounded form, it may not be stable enough to withstand the energy delivered by collisions with neighboring molecules. Different bonds between different atoms in different molecular contexts differ in terms of bond stability. The bond energy refers to the energy needed to break a particular bond. A molecule is stable if the bond energies associated with bonded atoms within the molecule are high enough to survive the energy delivered to the molecule through collisions with neighboring molecules or the absorption of energy (light).

When atoms form a covalent bond, their van der Waals surfaces merge to produce a new molecular van der Waals surface. There are a number of ways to draw molecules, but the space-filling or van der Waals surface view is the most realistic (at least for our purposes). While realistic it can also be confusing, since it obscures the underlying molecular structure, that is, how the atoms in the molecule are linked together. This can be seen in this set of representations of the simple molecule 2-methylpropane (→). As molecules become larger, as is the case with many biologically important molecules, it rapidly becomes impossible to appreciate their underlying organization based on a van der Waals surface representation.[227]



Because they form a new stable entity, it is not surprising (perhaps) that the properties of a molecule are quite distinct from, although certainly influenced by, the properties of the atoms from which they are composed. Some atoms, common to biological systems, such as hydrogen (H), can form only a single covalent bond. Others can make two (oxygen (O) and sulfur (S)), three (nitrogen (N)), four (carbon (C)), or five (phosphorus (P)) bonds.

In addition to smaller molecules, biological systems contain a number of distinct types of extremely large molecules, composed of many thousands of atoms; these are known as macromolecules. Such macromolecules are not rigid; they can often fold back on themselves leading to **intramolecular** interactions (that is attractions and repulsions within a given molecule). There are also interactions between molecules - which are referred to as **intermolecular** interactions. The strength and specificity of these interactions can vary dramatically and even small changes in a protein's molecular structure, such as caused by mutations and allelic variations, can have dramatic effects on molecular shape and function. Similarly, increasing temperatures can break such weak interactions, leading to changes in molecular shape and function.

Molecules and molecular interactions are dynamic. Collisions with other molecules can lead to parts of a molecule rotating with respect to one another around a single bond. The presence of a double bond restricts these kinds of movements; rotation around a double bond requires what amounts to breaking and then reforming one of the bonds. In addition, and if you have mastered some chemistry you already know that it is often incorrect to consider bonds as distinct entities isolated from one another and their surroundings. In some structures the electrons in bonds are best considered as delocalized (that is not “stuck” between two adjacent atoms). These are often shown as “resonance structures” that behave as mixtures of single and double bonds. Again this restricts free rotation around the bond axis and acts to constrain molecular geometry. As we will come to see, the peptide bond that occurs between a carbon (C) and a nitrogen (N) atom in a polypeptide chain, is an example of such a resonance behavior. Similarly, the ring structures found in the various “bases” present in nucleic acids result in flat structures that pack one on top of another. These various geometric complexities combine to make predicting a molecule’s three dimensional structure increasingly challenging as its size increases.

Bond stability and thermal motion (a non-biological moment)

Molecules do not exist out of context. In the real, or at least the biological world, they do not sit alone in a vacuum. Most biologically-relevant molecular interactions occur in aqueous solution. That means, biological molecules are surrounded by other, mostly water, molecules. As you may already know there is a lowest possible temperature, known as absolute zero (0K, –273.15ºC or –459.67°F).

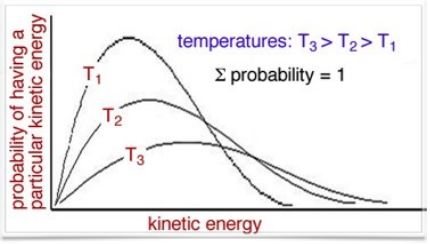
[227] Explicit Concepts of Molecular Topology: http://www.chem.msu.ru/eng/misc/babaev/match/top/top02.htm

At this biologically irrelevant temperature, molecular movements are minimal but not, apparently, absent all together.[228]

When we think about a system, we inevitably think about its temperature. Temperature is a concept that makes sense only at the system level. Individual molecules do not have a temperature, they have kinetic energy. The temperature of a system is a measure of the average kinetic energy of the molecules within it. The average kinetic energy is:

$$E_k = 1/2 \text{ (average mass)} \times (\text{average velocity})^2$$

It does not matter whether the system is composed of only a single type of molecule or many different types of molecules, at a particular temperature the average kinetic energy of all of the different molecules has one value. This is not to say that all molecules have the same kinetic energy, they certainly do not; each forms part of a distribution that is characterized by its average energy, this distribution is known as the Maxwell-Boltzmann distribution (introduced previously →). The higher the temperature, the more molecules will have a higher kinetic energy.



In a gas we can largely overlook the attractive intermolecular interactions between molecules because the average kinetic energies of the molecules is sufficient to disrupt such intermolecular interactions - that is, after all, why they are a gas. As we cool the system, we remove energy from it, and the average kinetic energy of the molecules decreases. When the average kinetic energy gets low enough, the molecules will form a liquid. In a liquid, the movement of molecules is not sufficient to disrupt all of the interactions between them. This is a bit of a simplification, however. Better to think of it more realistically. Consider a closed box partially filled with a substance in a liquid state. What is going on? Assuming there are no changes in temperature over time, the system will be at equilibrium. What we will find, if we think about it, is that there is a phase change going on, that is:

$$\text{Molecule (gas)} \leftrightarrows \text{Molecule (liquid)}.$$

At a particular temperature, the liquid phase is favored, although there will be some molecules in the system's gaseous phase. The point is that at equilibrium, the number of molecules moving from liquid to gas will be equal to the number of molecules moving from the gas to the liquid phase. If we increase or decrease the temperature of the system (that is add or remove energy), we will alter this equilibrium state, that is, the relative amounts of molecules in the gaseous versus the liquid states will change. The equilibrium is dynamic, in that different molecules may be in the gaseous or the liquid states, even though the distribution of molecules between the gaseous and the liquid states will be steady.

In a liquid, while molecules associate with one another, they can still move with respect to one another. That is why liquids can be poured, and why they assume the shape of the (solid) containers into which they are poured. This is in contrast to the container, whose shape is independent of what it contains. In a solid the molecules are tightly associated and so do not translocate with respect to one another, although they can rotate and jiggle in various ways. Solids do not flow. The cell, or more specifically, the cytoplasm, acts primarily as a liquid. Most biological processes take place in the liquid phase: this has a number of implications. First molecules, even very large macromolecules, move with respect to one another. Driven by thermal motion, molecules will move in a Brownian manner, a behavior known as a random walk.

Thermal motion will influence whether and how molecules associate with one another. We can think about this process in the context of an ensemble of molecules, let us call them A and B; A and B interact to form a complex, A:B. Assume that this complex is held together by LDF-mediated interactions. In an aqueous solution, the A:B complex is colliding with water molecules. These water molecules have various energies (from low to high), as described by the Boltzmann distribution.

---

[228] zero point energy (from wikipedia)

There is a probability that in any unit of time, one or more of these collisions will deliver energy greater than the interaction energy that holds A and B together; this will lead to the disassociation of the A:B complex into separate A and B molecules. Assume we start with a population of 100% A:B complexes, the time it takes for 50% of these molecules to dissociate into A and B is considered the "half-life" of the complex. We use the term half-life repeatedly to characterize the stability of a complex or macromolecule. Now here is the tricky part, much like the situation with radioactive decay, but distinctly different. While we can confidently conclude that 50% of the A:B complexes will have disassembled into A and B at the half-life time, we can not predict exactly which A:B complexes will have disassembled and which will remain intact. Why? Because we cannot predict exactly which collisions will provide sufficient energy to disassociate a particular A:B complex.[229] Dissociation is a stochastic process, and like all stochastic processes (such as genetic drift) is best understood in terms of probabilities.

Stochastic processes are particularly important within biological systems because, generally, cells are small and contain relatively small numbers of molecules of a particular type. If, for example, the expression of a gene depends upon a protein binding to a specific site on a DNA molecule, and if there are relatively small numbers of that protein, and usually only one or two copies of the gene, that is, the DNA molecule, present in a cell, we will find that whether or not a copy of the protein is bound to a specific region of the DNA is a stochastic process.[230] If there are enough cells, then the group average may well be predictable, but the behavior of any one cell will not be.[231] In an individual cell, sometimes the protein will be bound and the gene will be expressed and sometimes not, all because of thermal motion and the small numbers of interacting components involved. This stochastic property of cells can play important roles in the control of cell and organismic behaviors.[232] It can even transform a genetically identical population of organisms into subpopulations that display two or more distinct behaviors, a property with important implications, that we will return to.

Questions to answer:

78. How does temperature influence intermolecular interactions? How might changes in temperature influence macromolecular shape?
79. Why is the effect of temperature on covalent bond stability not generally significant in biological systems?
80. Why does population size matter when generating a graph that describes radioactive decay or the dissociation of a complex, like the A:B complex discussed above?

Questions to ponder:

- Why is the Boltzmann distribution asymmetric around the highest point

Bond polarity, inter- and intramolecular interactions

So far, we have been considering covalent bonds in which the sharing of electrons between atoms is more or less equal, but that is not always the case. Because of their atomic structures, based on quantum mechanical principles (not discussed here), different atoms have different affinities for their own electrons. When an electron is removed or added to an atom (or molecule) that atom/molecule becomes an ion. Atoms of different elements differ in the amount of energy it takes to remove an electron from them; this is, in fact, the basis of the photoelectric effect explained

---

[229] It should be noted that, in theory at least, we might be able to make this prediction if we mapped the movement of every water molecule. This is different from radioactive decay, where it is not even theoretically possible to predict the behavior of an individual radioactive atom.

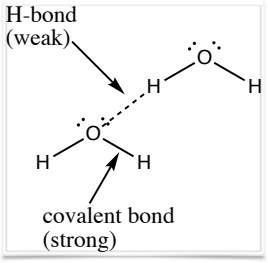
[230] This is illustrated here and we will return to this type of behavior later on.

[231] Biology education in the light of single cell/molecule studies

[232] Single Cells, Multiple Fates, and Biological Non-determinism: https://www.ncbi.nlm.nih.gov/pubmed/27259209

by Albert Einstein in another of his revolutionary 1905 papers.[233] Each type of element has a characteristic electronegativity, a measure of how tightly it holds onto its electrons when it is bonded to another atom, an idea that you may have mastered in general chemistry. If the electronegativities of the two atoms in a bond are equal or similar, then the electrons are shared more or less equally between the two atoms and the bond is said to be non-polar, meaning without direction. There are no stable regions of net negative or positive charge on the surface of the resulting molecule. If the electronegativities of the two bonded atoms are unequal, however, then the electrons will be shared un-equally. On average, there will be more electrons more of the time around the more electronegative atom and fewer around the less electronegative atom. This leads to partially negatively and partially positively-charged regions of the two bonded atoms – the bond has a direction. Charge separation produces an electrical field, known as a dipole. A bond between atoms of differing electronegativities is said to be polar.

Atoms of O and N are more electronegative than C and H, and will sequester electrons when bonded to atoms of H and C. The O and N become partly negative and the C and H become partly positive. Because of the quantum mechanical organization of atoms, these partially negative regions are organized in a non-uniform manner (the atoms have regions with different partial charges). In contrast, there is no significant polarization of charge in bonds between C and H atoms, and such bonds are non-polar. The presence of polar bonds leads to the possibility of electrostatic interactions between molecules (an aspect of van der Waals interactions). Such interactions are stronger than LDF-mediated interactions but weaker than covalent bonds. Like covalent bonds polar bond interactions have a directionality to them – the three atoms involved have to be arranged more or less along a straight line. There is no such geometric constraint on LDF-mediated interactions.



Since the intermolecular forces arising from polarized bonds often involve an H atom interacting with an O or an N atom, these have become known generically and perhaps unfortunately, as hydrogen or H-bonds (→). Why unfortunate? Because H atoms can take part in covalent bonds, but H-bonds are not covalent bonds, they are very much weaker. It takes much less energy to break an H-bond between molecules or between parts of (generally macro-) molecules than it does to break a covalent bond involving a H atom.

The implications of bond polarity

Melting and boiling points are important physical properties of molecules, although this applies primarily to small molecules and not macromolecules. Here we are considering a pure sample that contains extremely large numbers of the molecule in question. Let us start at a temperature at which the sample is liquid. The molecules are moving with respect to one another, there are interactions between the molecules, but they are transient - the molecules are constantly switching neighbors. As we increase the temperature of the system, the energetics of collisions are now such that all interactions between neighboring molecules are broken, and the molecules fly away from one another. If they happen to collide with one another, they (generally) do not adhere; the bond that might form is not strong enough to resist the kinetic energy delivered by collisions with other molecules. The molecules are said to be in a gaseous state and the transition from liquid to gas is the boiling point. Similarly, starting with a liquid, when we reduce the temperature, the interactions between molecules become longer lasting until a temperature is reached at which the energy transferred through collisions is no longer sufficient to disrupt the interactions between molecules.[234] As more and more molecules interact, the positions of neighboring molecules becomes more and

[233]Albert Einstein: Why Light is Quantum: http://youtu.be/LWli7NO1tbk

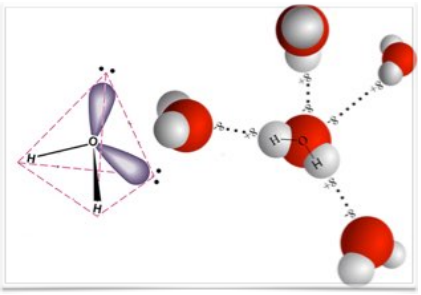
[234] The nature of the geometric constrains on inter-molecular interactions will determine whether the solid is crystalline or amorphous. see: https://en.wikipedia.org/wiki/Crystal

more highly constrained - the liquid is transformed into a solid. While liquids flow and assume the shape of their containers, because neighboring molecules are free to move with respect to one another, solids maintain their shape – neighboring molecules stay put. The temperature at which a liquid changes to a solid is known as the melting point. These temperatures mark what are known as phase transitions: solid to liquid and liquid to gas.

At the macroscopic level, we see the rather dramatic effects of bond polarity on melting and boiling points by comparing molecules of similar size with and without polar bonds and the ability to form H-bonds (↓). For example, neither CH_4 (methane) or Ne (neon) contain polar bonds and so do not form intra-molecular H-bond-type electrostatic interactions. In contrast NH_3 (ammonia), H_2O (water), and FH (hydrogen fluoride) have three, two and one polar bonds, respectively, and can take part in one or more intra-molecular H-bond-type electrostatic interactions. All five compounds have the same number of electrons, ten. When we look at their melting and boiling temperatures, we see how the presence of polar bonds influences these properties. In particular, water stands out as dramatically different from the rest, with significantly higher (> 70ºC) melting and boiling points than its neighbors.

Compounds	CH_4	NH_3	OH_2	FH	Ne
molecular weight	16.04	17.02	18.02	20.01	20.18
bond electronegativity	0.45	0.94	1.34	1.88	N/A
# of electrons	10	10	10	10	10
# of bonds	4	3	2	1	0
melting point	-182°C	-77.7°C	**0°C**	-83°C	-248.6°C
boiling point	-161.5°C	-33.4°C	**100°C**	19.5°C	-246.1°C

So why is water different? Well, in addition to the presence of polar covalent bonds, we have to consider the molecule's shape. Each water molecule has two partially positive Hs and two partially negative sites on its O. These sites of potential H-bond-type electrostatic interactions are arranged in a nearly tetrahedral geometry (→). Because of this arrangement, each water molecule can interact through H-bond-type electrostatic interactions with four neighboring water molecules. To remove a molecule from its neighbors, four H-bond-type electrostatic interactions must be broken, which is relatively easy, energetically, since they are each rather weak. In the liquid state, molecules jostle one another and change their H-bond-type electrostatic interaction partners constantly. Even if one interaction is broken the water molecule is likely to remain linked to multiple neighbors via the remaining H-bond-type electrostatic interactions.



This molecular hand-holding leads to water's high melting and boiling points as well as its high surface tension. We can measure the strength of surface tension in various ways. The most obvious is the weight that the surface can support. Water's surface tension has to be dealt with by those organisms that interact with a liquid-gas interface. Some, like the water strider, use it to cruise along the surface of ponds. (←) As the water strider walks on the surface of the water, the molecules of its feet do not form H-bond-type electrostatic interactions with water molecules, they are said to be hydrophobic, although that is clearly a bad name - they are not afraid of water, rather they are simply apathetic to it. Hydrophobic molecules interact with other molecules, including water molecules, but only through LDF-mediated interactions. Molecules that can make H-bonds or other polar interactions with water are termed hydrophilic. As molecules increase in size they can have regions that are hydrophilic and regions that are hydrophobic. Molecules that have distinct hydrophobic and hydrophilic regions are termed amphipathic and we will consider them in greater detail in the next chapter.

Interacting with water

We can get an idea of the hydrophilic, hydrophobic, and amphipathic nature of molecules through their behaviors when we try to dissolve them in water. Molecules like sugars (carbohydrates), alcohols, and most amino acids are primarily hydrophilic, they dissolve readily in water. Molecules like fats are highly hydrophobic, and they do not dissolve significantly in water. So why the difference? To answer this question we have to be clear what we mean when we say that a molecule is soluble in water. We will consider this from two perspectives. The first is what the solution looks like at the molecular level, the second is how the solution behaves over time. To begin we need to understand what water alone looks like. Because of its ability to make and donate multiple H-bond-type electrostatic interactions in a tetrahedral arrangement, water molecules form a dynamic three-dimensional intermolecular interaction network. In liquid water the H-bond-type electrostatic interactions between the molecules break and form rapidly.

To insert a molecule A, known as a solute, into this network you have to break some of the H-bond-type electrostatic interactions between the water (solvent) molecules. If the A molecules can make H-bond-type electrostatic interactions with water molecules, that is, if they are hydrophilic, then there is little net effect on the free energy of the system. Such a molecule is soluble in water. So what determines how soluble the solute is. As a first order estimate, each solute molecule will need to have at least one layer of water molecules around it, otherwise it will be forced to interact with other solute molecules. If the number of these interacting solute molecules is large enough, the solute will no longer be in solution. In some cases, aggregates of solute molecules can, when small enough, remain suspended in the solution. This is a situation known as a colloid. The cytoplasm of a cell behaves like a colloid in many ways. While a solution consists of individual solute molecules surrounded by solvent molecules, a colloid consists of aggregates of solute molecules in a solvent. We might predict that all other things being equal (an unrealistic assumption), the larger the solute molecule the lower its solubility. You might be able to generate a similar rule for the size of particles in a colloid.

Now we can turn to a conceptually trickier situation, the behavior of a hydrophobic solute molecule in water. Such a molecule cannot make H-bond-type electrostatic interactions with water molecules, so when it is inserted into water the total number of H-bond-type electrostatic interactions in the system decreases - the energy of the system increases (remember, bond forming lowers potential energy). However, it turns out that much of this “enthalpy” change, indicated as ΔH, is compensated for by LDF-mediated interactions between the molecules. Generally, the net enthalpic effect is minimal. Something else must be going on to explain the insolubility of such molecules.

Turning to entropy

In a liquid, water molecules will typically be found in a state that maximizes the number of H-bond-type electrostatic interactions present. Because these interactions have a distinct, roughly tetrahedral geometry, their presence constrains the possible orientations of molecules with respect to one another. This constraint is captured when water freezes; it is the basis for ice crystal formation, why the density of water increases before freezing and decreases with freezing, and why ice floats in liquid water.[235] In the absence of a hydrophobic solute molecule there are many equivalent ways that liquid water molecules can interact to produce these geometrically specified arrangements. But the presence of a solute molecule constrains the number of appropriate orientations of water molecules: a much smaller number of configurations result in maximizing H-bond formation between water molecules. The end result is that the water molecules become arranged in a limited number of ways around each solute molecule; they are in a more ordered, that is, in a more improbable state than they would be in the absence of solute. The end result is that

[235] Why does ice float in water? http://youtu.be/UukRgqzk-KE

there will be a decrease in entropy (indicated as ΔS), the measure of the probability of a state. ΔS will be negative compared to arrangement of water molecules in the absence of the solute.

How does this influence whether dissolving a molecule into water is thermodynamically favorable or unfavorable? Since the change in interaction energy (ΔH) associated with placing most solutes into the solvent is near 0, it is the change in entropy (ΔS) that makes the difference. Keeping in mind that $\Delta G = \Delta H - T\Delta S$, if ΔS is negative, then $-T\Delta S$ will be positive. The ΔG of a thermodynamically favorable reaction is, by definition, negative. This implies that the reaction:

water + solute ⇆ solution (water + solute)

will be thermodynamically unfavorable; the reaction will move to the left. That is, if we start with a solution, it will separate so that the solute is removed from the water. How does this happen? The solute molecules aggregate with one another. This reduces their effects on the organization of water molecules, and so the ΔS for aggregation is positive. If the solute is oil (highly hydrophobic, unable to form H-bonds), and we mix it into water, the oil will separate from the water, driven by the increase in entropy associated with minimizing solute-water interactions. Similar processes can occur at the molecular scale, leading to what known as phase separation - cytoplasmic domains and structure distinct from the bulk cytoplasm. Such liquid-liquid domains occur what are known as emulsions. In the cytoplasm, domain of specific macromolecules can also occur.[236]

Questions to answer:

81. Predict (and explain your prediction), the factors that influence the solubility of a molecule in water
82. Why does the separation of oil and water represent a more disordered state?
83. How would you explain to a "normal" person how it is possible for a water strider to walk on water; or why ice floats – what concepts would you need to introduce them to?
84. Predict (and explain the basis of your prediction) the effects of H-bonding on a molecule's boiling point.

Question to ponder:

What would happen to a water strider if its "feet" were hydrophilic?

[236] McSwiggen et al., 2021. Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences

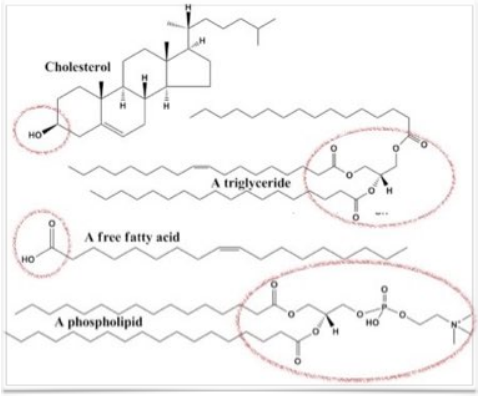
Chapter 6: Membrane boundaries and capturing energy

In which we consider how the aqueous nature of biological systems drives the formation of lipid-based barrier membranes and how such membranes are used to capture and store energy from the environment and chemical reactions. We consider how coupled reactions are used to drive macromolecular syntheses and growth, and how endosymbiotic events, involving the capture of aerobic and photosynthetic bacteria, played a critical role in the evolution of eukaryotic cells

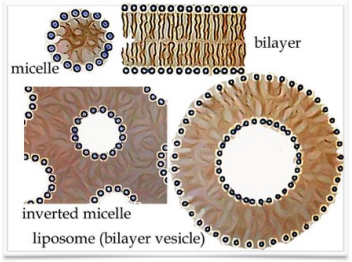
Defining the cell's boundary

A necessary step in the origin of life was the generation of a discrete boundary layer that separates the living non-equilibrium reaction system from the rest of the universe. This original boundary layer, the structural ancestor of the plasma membrane of modern cells, serves to maintain the integrity of the living system and mediates the movement of materials and energy into and out of the cell. The plasma membrane of all cells, whether bacterial, archaeal or eukaryotic, appears to be a homologous structure derived from a precursor present in the last common ancestor of life. So what is the structure of this barrier (plasma) membrane? How is it built and how does it work?

When a new cell is formed its plasma membrane is derived from the plasma membrane of the progenitor cell. As the cell grows, new molecules are added into the membrane to increase its surface area. Biological membranes are composed of two general classes of molecules, proteins (discussed in the next chapter) and lipids. It is worth noting explicitly that, unlike a number of other types of molecules that we will be considering, lipids are not a structurally coherent group, that is they do not have one common structure. Structurally diverse molecules, such as cholesterol and phospholipids, are both considered lipids (→). All lipids have two distinct domains: a hydrophilic domain (circled in red) characterized by polar regions and one or more hydrophobic domains that are usually made up of C and H. While there is a relatively small set of common lipid types, there are many different lipids found in biological systems and the characterization of their structures and functions has led to a new area of analysis known as lipidomics.[237]



Lipids are amphipathic. In aqueous solution, entropic effects will act to drive the hydrophobic parts of the lipid out of an aqueous solution. In contrast to totally non-polar molecules, like oils, the hydrophobic part of the lipid is connected to a hydrophilic domain. Lipid molecules deal with this dichotomy by associating with other lipid molecules in multimolecular structures in which the interactions between the hydrophilic parts of the lipid molecule and water molecules are maximized and the interactions between the lipid's hydrophobic parts and water are minimized. Many such multi-molecular structures that fulfill these constraints can be generated (→). The structures that form depend upon the details of the system, including the shapes of the lipid molecules involved and the

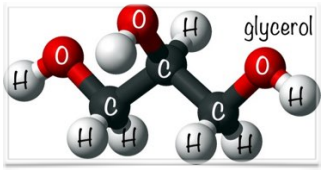


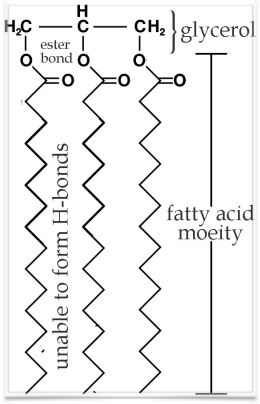
[237] On the future of "omics": lipidomics & Lipidomics: new tools and applications

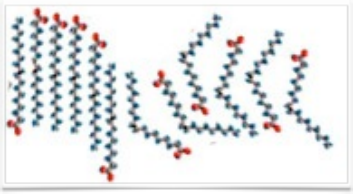
relative amounts of water and lipid present. In every case, the self-assembly of these structures involves an increase in the total overall entropy of the system, a perhaps counterintuitive result. For example, in a micelle the hydrophilic region is in contact with the water, while the hydrophobic regions are inside, away from direct contact with water. This leads to a more complete removal of the lipid's hydrophobic domain from contact with water than can be arrived at by a purely hydrophobic oil molecule, so unlike oil, lipids can form stable structures in solution. The diameter and shape of the micelle is determined by the size of its hydrophobic domain. As this domain gets longer, the center of the micelle becomes more crowded. A type of organization that avoids "lipid-tail crowding" is known as a bilayer vesicle. Here there are two layers of lipid molecules, pointing in opposite directions. The inner layer surrounds a water-filled region, the lumen of the vesicle, while the outer layer interacts with the external environment. In contrast to the situation within a micelle, the geometry of a vesicle means that there is significantly less crowding as a function of lipid tail length. Crowding is further reduced as a vesicle increases in size to become a cellular membrane. Micelles and vesicles can form colloid-like systems with water, that is they exist as distinct structures that can remain suspended in a stable state. We can think of the third type of structure, the planar membrane, as an expansion of the vesicle to a larger and more irregular size. Now the inner layer faces the inner region of the cell (which is mostly water) and the opposite region faces the outside world, which again is often mostly water. For the cell to grow, new lipids need to be inserted into both inner and outer layers of the membrane; how exactly this occurs typically involves interactions with proteins, known as flippases, that can move a lipid from the inner to the outer layer of a bilayer membrane. When we consider proteins, you may consider the energetics of this reaction and how a plausible flipping mechanism might work.

A number of distinct mechanisms are used to insert molecules into membranes, but they all involve a pre-existing membrane – this is another aspect of the continuity of life. Totally new cellular membranes do not form, membranes are built on pre-existing membranes. For example, a vesicle, a spherical lipid bilayer, can fuse into or emerge from a planar (bilayer) membrane. These processes are typically driven by protein-based molecular machines coupled to thermodynamically favorable reactions. When the membrane involved is the plasma (boundary) membrane, these processes are known as endocytosis and exocytosis (into and out of the cell), respectively. These terms refer explicitly to the fate of the material within the vesicle. Exocytosis releases material in the vesicle interior into the outside world, whereas endocytosis captures material from outside of the cell and brings it into the cell. Within a cell, vesicles can fuse with and emerge from one another.

As noted above, there are hundreds of different types of lipids, generated by a variety of biosynthetic pathways catalyzed by proteins encoded in the genetic material. We will not concern ourselves too much about all of these different types of lipids, but we will consider two generic classes, the glycerol-based lipids (←) and cholesterol, because considerations of their structures illustrates general ideas related to membrane behavior. In bacteria and eukaryotes, glycerol-based lipids are typically formed from the highly hydrophilic molecule glycerol combined with two or three fatty acid molecules (a three fatty acid chain molecule is shown →). Fatty acids contain a long chain hydrocarbon with a polar (carboxylic acid) head group. The molecular nature of these fatty acids influences the behavior of the membrane formed. Often these fatty acids have what are known as saturated hydrocarbon tails. A saturated hydrocarbon contains only single bonds between the carbon atoms of its tail domain. While these chains can bend and flex, they tend to adopt a more or less straight configuration. In this straight configuration, they pack closely with one another, which maximizes the lateral (side to side) LDF-mediated interactions between them. Because of the extended surface contact between the chains, lipids with saturated hydrocarbon chains are typically solid around room temperature. Solid means that the molecules rarely exchange positions with one another. On the

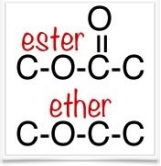






other hand (←), there are cases where the hydrocarbon tails are "unsaturated", that is they contain double bonds (–C=C–). These are typically more fluid and flexible because unsaturated hydrocarbon chains have permanent kinks due to the rigid nature and geometry of C=C bonds; they cannot pack as regularly as saturated hydrocarbon chains. The less regular packing means that there is less interaction area between the molecules, which lowers the strength of the LDF-mediated interactions between them. Lower LDF-mediated interaction energy in turn, lowers the temperature at which these bilayers change from a solid (no movement of the lipids relative to each other within the plane of the membrane) to a liquid with relatively free movements within the plane of the membrane. Recall that the strength of interactions between molecules determines how much energy is needed to overcome a particular type of interaction. Because these LDF-mediated intermolecular interactions are relatively weak, changes in temperature influence the physical state of the membrane. The liquid-like state is often referred to as the fluid state. The membrane's state is important because it can influence the movement, behaviors, and activities of the proteins embedded within it. If the membrane is in a solid state, proteins within the membrane will be relatively immobile. If is in the liquid state, these proteins move rapidly by diffusion, that is, by collision-driven movements within the plane of the membrane. In addition, since lipids and proteins are closely associated with one another in the membrane, the physical state of the membrane can influence the activity of embedded proteins, a topic to which we will return.

Cells can manipulate the solid-to-liquid transition temperature of their membrane by altering the membrane's lipid composition. Increasing the ratio of saturated to unsaturated chains can increase the melting temperature. Controlling chain saturation involves altering the activities of the enzymes involved in various saturation/desaturation reactions. That these enzymes can be regulated implies a feedback mechanism, by which either temperature, membrane fluidity, or protein activity act to regulate metabolic processes and gene expression. This type of feed back mechanism is part of the homeostatic and adaptive systems of the cell (and the organism) and is a topic we will return to in greater depth.



There are a number of differences between the lipids used in bacterial and eukaryotic organisms and archaea.[238] Most dramatically, instead of straight chained hydrocarbons, archaeal lipids are constructed of branched isoprene (CH_2=C(CH_3)CH=CH_2) polymers linked to the glycerol group through an ether, rather than an ester linkage (→). The bumpy and irregular shape of the isoprene groups (compared to the relatively smooth saturated hydrocarbon chains) means that archaeal membranes will tend to melt (go from solid to liquid) at lower temperatures.[239] At the same time the ether linkage is more stable (requires more energy to break) than the ester linkage. It remains unclear why bacteria and eukaryotes use straight chain hydrocarbon lipids, while archaea use isoprene-based lipids. One speculation is that the archaea were originally (or became) adapted to live at higher temperatures, where the greater stability of the ether linkage would provide a critical advantage.

Some archaea and bacteria, known generically as thermophiles and hyper-thermophiles, live (happily, apparently) at temperatures up to 110 ºC.[240] At the highest temperatures, thermal motion might be expected to disrupt the integrity of the membrane, allowing small charged molecules (ions) and other larger hydrophilic molecules to pass through.[241] Given the importance of membrane integrity, you may (perhaps) not be surprised to find "double-headed" lipids in such thermophilic

---

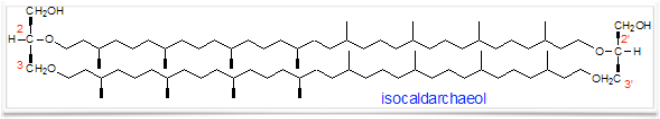
[238] A re-evaluation of the archaeal membrane lipid biosynthetic pathway

[239] The origin and evolution of Archaea: a state of the art

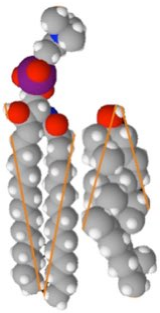
[240] You might consider how this is possible and under want physical conditions you might find these "thermophilic" archaea.

[241] Ion permeability of the cytoplasmic membrane limits the maximum growth temperature of bacteria and archaea

organisms (→). These lipid molecules have two distinct hydrophilic glycerol moieties, one located at each end of the molecule; this enables a single molecule to span the membrane. The presumption is that such lipids act to stabilize the membrane against the disruptive effects of high temperatures.

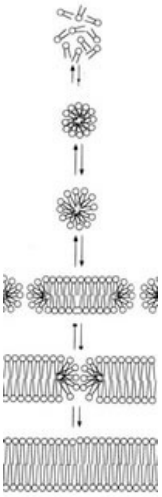


The solid-fluid nature of biological membranes, as a function of temperature, is complicated by the presence of cholesterol and structurally similar lipids. For example, in eukaryotes the plasma membrane can contain as much as 50% cholesterol, in terms of the number of molecules present. Cholesterol has a short bulky hydrophobic domain (→) that does not pack well with other lipids: a hydrocarbon chain lipid (left) and cholesterol (right). The presence of cholesterol dramatically influences the solid-liquid behavior of the membrane. The diverse roles of lipids is a complex subject that goes beyond our scope here.



The origin of biological membranes

The cell membrane is composed of a number of different types of lipids. The hydrophobic “tails” of modern lipids range from 16 to 20 carbons in length. The earliest membranes, however, were likely to have been composed of similar molecules with shorter hydrophobic chains. Based on the properties of lipids, we can map out a plausible scenario for the appearance of membranes. Lipids with very short hydrophobic chains, from 2 to 4 carbons in length, can dissolve in water (can you explain why?) As the lengths of the hydrophobic chains increases, the molecules begin to self-assemble into micelles. By the time the hydrophobic chains reach ~10 carbons in length, it becomes more difficult to fit the hydrocarbon chains into the interior of a micelle without making larger and larger spaces between the hydrophilic heads. Water molecules can begin to move through these spaces and interact with the hydrocarbon tails. At this point, the hydrocarbon-chain lipid molecules begin to associate into semi-stable bilayers (→). One interesting feature of bilayers is that the length of the hydrocarbon chain is no longer structurally limiting, in contrast to the situation in micelles. One problem, though, are the edges of the bilayer, where the hydrocarbon region of the lipid would come in contact with water, a thermodynamically unfavorable situation. This problem is avoided by linking edges of the bilayer to one another, forming a closed balloon-like structure. Such bilayers can capture regions of solvent, that is water and the solutes dissolved within it.



Bilayer stability increases further as hydrophobic chain length increases. At the same time, membrane permeability decreases. It is a reasonable assumption that the earliest biological systems used shorter chain lipids to build their "proto-membranes" and that these membranes were relatively leaky.[242] The appearance of more complex lipids, capable of forming more impermeable membranes, must therefore have depended upon the appearance of mechanisms (presumably protein-based) that enabled hydrophilic molecules to pass through such membranes. The interdependence of change is known as co-evolution. Co-evolutionary processes were apparently common enough to make the establishment of living systems possible.

Questions to answer:

85. Draw diagrams to show how increasing the length of a lipid's hydrocarbon chains affects the structures that it can form and use your diagrams to predict how the effects at the hydrophobic edges of a lipid bilayer are minimized?

[242] Jack Szostak (two videos): The origin of life on Earth & Protocell membranes

86. Some lipids have negatively-charged phosphate groups attached to the glycerol as well as fatty acids - predict how the presence of “phospho-lipids" will impact membrane structure and stability.
87. Make a set of general rules on the effects of size and composition on the ability of a molecule to pass through a membrane.

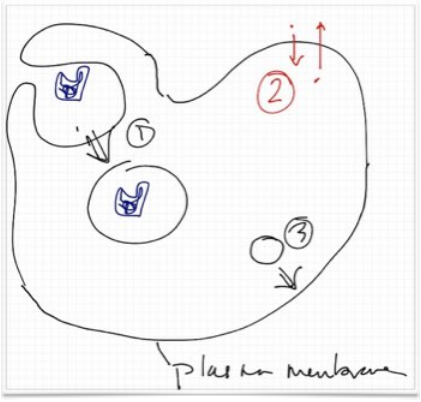
Questions to ponder:

- Why do fatty acid and isoprene lipids form similar bilayer structures?
- Why might early (evolutionarily) membrane be expected to be leaking compared to modern membranes?

Transport across membranes

As we have said before (and will say again), the living cell is a historically continuous non-equilibrium system. To maintain its living state both energy and matter have to move into and out of the cell, which leads us to consider intracellular and extracellular environments and the boundary membrane that separates them. The differences between the regions inside and outside of the plasma membrane are profound. Outside, even for cells within a multicellular organism, the environment is generally mostly water, with relatively few complex molecules. Inside the membrane-defined space is the cytoplasm, a highly concentrated (300 to 400 µg/ml) solution of proteins, nucleic acids, smaller molecules, and thousands of interconnected chemical reactions.[243] Cytoplasm (and the membrane around it) is inherited by each cell when it is formed, and represents an uninterrupted continuous reaction system that first arose more than ~3 billion years ago.

A lipid bilayer membrane poses an interesting barrier to the movement of molecules. First for larger molecules, particles or other organisms, it acts as a physical barrier. Typically when larger molecules, particles (viruses), and other organisms enter a cell, they are first engulfed by the membrane (process 1 known as endocytosis)(→).[244] A superficially similar process, exocytosis, but running in “reverse” (process 3), is involved in moving molecules to the cell surface and releasing them into the extracellular space. Both endocytosis and exocytosis involve membrane vesicles emerging from or fusing into the plasma membrane. These processes leave the topology of the cell unaltered; a molecule within a vesicle is still “outside” of the cell, or at least outside of the cytoplasm. These movements are driven by various protein-based molecular machines that we will consider briefly (they are considered further in more specialized courses on cell biology). We are left with the question of how molecules can enter or leave the cytoplasm, this involves passing directly through a membrane (process 2).



So the question is, how does the membrane “decide” which molecules to allow into and out of the cell. If we think about it, there are three possible general mechanisms (can you think of others?) Molecules can move on their own through the membrane, some move passively across the membrane using specific “carriers” or “channels”, while others are moved actively using a kind of “pump”, an energy dependent process involving coupled reactions. In the majority of cases, these carriers, channels, and pumps are protein-based molecular machines, the structure of which we will consider in greater detail later on. Which types of carriers, channels, and pumps are present will determine what types of molecules move through the cell’s membrane, as well as which directions they move, or rather their net flux into or out of the cell. We can think of this molecular movement as

[243] A model of intracellular organization

[244] These processes, ranging from pinocytosis (cell drinking) to phagocytosis (cell eating) involve different molecular machines.

a reaction, very much in the same way that we consider a conventional chemical reaction reaction generically as:

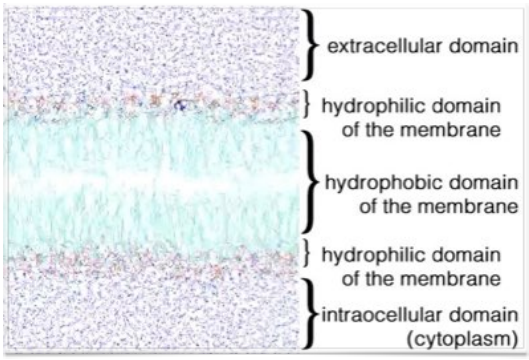
$$Molecule_{outside} \rightleftharpoons Molecule_{inside\ membrane} \rightleftharpoons Molecule_{inside\ cell}.$$

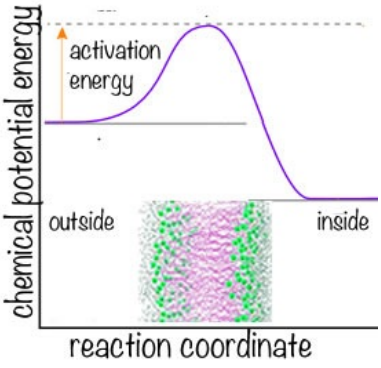
As with standard chemical reactions, movements through a membrane involve an activation energy, that involves the energy needed to remove a water soluble molecule from aqueous solution and then pass the transported molecule through the membrane. So, you might well ask, why does the membrane, particularly the hydrophobic center of the membrane, pose a barrier to the movement of hydrophilic molecules. Here the answer involves the difference in the free energy of the moving molecule within an aqueous solution, including the hydrophilic surface region of the membrane, where H-bond type electrostatic interactions are common between molecules, and the hydrophobic region of the membrane, where only LDF-mediated interactions are present. The situation is exacerbated for charged molecules, since water molecules are typically organized in a dynamic shell around each ion. We are considering molecules of one particular substance moving through the membrane and so the identity of the molecule does not change during the transport reaction. If the concentrations of the molecules are the same on both sides of the membrane, then their Gibbs free energies are also equal, the system will be in equilibrium with respect to this reaction. In this case, as in the case of chemical reactions, there will be no net flux of the molecule across the membrane, but molecules will be moving back and forth at an equal rate. The rate at which they move back and forth will depend on the size of the activation energy associated with moving across the membrane as well as the concentrations of the molecules.

To think about how molecules cross lipid membranes, let us begin with water itself, which is small and uncharged, although polarized. Typically, the concentration of water outside of a cell is greater than the concentration of water inside a cell. This implies that the reaction:

$$H_2O_{\ outside} \rightleftharpoons H_2O_{\ inside}$$

will be favorable, so there will be a net flux of water molecules into the cell. What is happening in this reaction? As a water molecule moves through water, H-bonds are broken and reform - there is no net energetic change. In contrast, when a water molecule begins to leave the aqueous phase the H-bonds between it and its neighbors must be broken but no new H-bonds are formed as the molecule enters the hydrophobic (central) region of the membrane. This asymmetry in H-bonding results in water molecules being "pulled back" into the water phase (←)(video of a water molecule moving through a membrane). In part the $Water_{outside} \rightleftharpoons Water_{inside}$ reaction's activation energy (→) involves breaking these and other H-bonding interactions (with hydrophilic lipid head domains). Thermal movement is generally sufficient for the reaction to occur at a reasonable rate. Once they enter the membrane, water molecules can pass through it rather easily, since they interact with the central region of the membrane solely through weak LDFs.

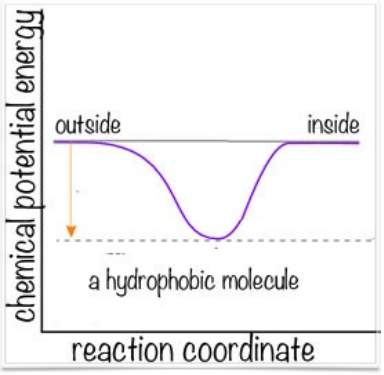




Small non-polar molecules, such as O_2 and CO_2 also pass readily through a biological membrane. There is more than enough energy available through collisions with other molecules (thermal motion) to provide them with the energy needed to overcome the activation energy involved in leaving the aqueous phase and passing through the molecular domains of the membrane. As with

water, there are often differences in the free energies of the molecules on the inside and outside of the cell. For example, in organisms that depend upon O_2 (obligate aerobes), the O_2 outside of the cell is produced by plants that release O_2 as a waste product and carried into the organism's interior by the circulatory system (in animals). When O_2 enters the cell, it can take part in the reactions of respiration (considered soon), leading to an O_2 concentration gradient, $[O_2]_{outside} > [O_2]_{inside}$ leading to a net flux of O_2 into the cell.

Another perspective into membrane behavior is to consider the interactions of different types of molecules within a bilayer membrane. If a molecule is hydrophobic (non-polar) it will be more "soluble" (concentrated) in the membrane's central hydrophobic region than it is in the surrounding aqueous environment (→). A totally hydrophobic molecule will accumulate within the membrane; an activation energy would be associated with its leaving the hydrophobic region, and would involve its entropic effects on water structure (remember, moving a hydrophobic molecule into water will increase water organization (decreasing entropy).



Questions to answer:

88. Consider the reaction diagram for flipping a lipid molecule's orientation by 180º perpendicular to the plane of the membrane: what energy barriers are associated with such a movement?
89. Draw a graph to show how the potential energy changes as an ion moves across a membrane. What is involved when an ion leaves the aqueous phase? How would this differ from a hydrophobic molecule?
90. What do you expect to happen to the O_2 gradient if an aerobic cell's ability to use O_2 is inhibited?

Channels and carriers

Beginning around the turn of the last century, a number of scientists began working to define the nature of the cellular boundary layer. In the 1930's it was noted that small, water soluble molecules entered cells faster than predicted based on the assumption that the membrane acts like a simple hydrophobic barrier. Ernest Overton (1865-1933) and Runar Collander (1894-1973) postulated that membranes were more than simple barriers, specifically that they contained features that enabled them to act as highly selective molecular sieves.[245] Most of these features are proteins (we are getting closer to a discussion of proteins) that can act as channels, carriers, and pores. If we think about crossing the membrane as a reaction, then the activation energy of this reaction can be quite high for highly hydrophilic and larger molecules, we will need a catalyst to reduce the activation energy so that the reaction can proceed at a reasonable rate. There are two generic types of membrane permeability catalysts: carriers and channels.

Carrier proteins are membrane proteins that shuttle back and forth across the membrane. They bind to specific hydrophilic molecules when they are located in the hydrophilic region of the membrane, hold on to the bound molecule as they traverse the membrane's hydrophobic region, and then release their "cargo" when they again reach a hydrophilic region of the membrane. Both the movements of carrier and cargo across the membrane, and the release of transported molecules, are stochastic and are driven by thermal motion (energy transferred as the result of collisions with other molecules), so no other energy source is needed. We can write this class of reactions as:

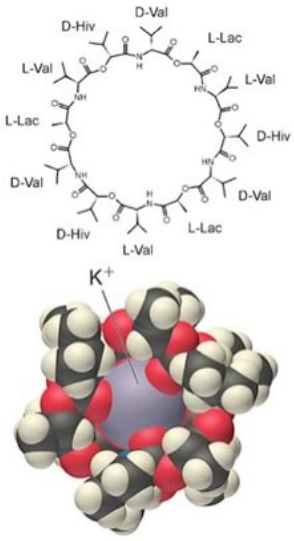
$$Molecule_{outside} + carrier_{empty} \rightleftharpoons carrier\text{–}\ Molecule_{outside} \rightleftharpoons carrier\text{–}\ Molecule_{inside} \rightleftharpoons carrier_{empty} + Molecule_{inside}\ .$$

There are many different types of carrier molecules and each type of carrier has preferred cargo. Related molecules may be bound and transported, but with less specificity and so at a much lower

[245] Does Overton still rule? http://www.nature.com/ncb/journal/v1/n8/full/ncb1299_E201.html

rate. Exactly which molecules a particular cell will allow to enter will be determined in part by which carrier protein genes it expresses. Mutations in a gene encoding a carrier can change (or abolish) the range of molecules that that carrier can transport across a membrane.

Non-protein carriers: An example of a membrane carrier is a class of antibiotics, known generically as ionophores, that carry ions across membranes. They kill cells by disrupting the normal ion balance across the cell's membrane and within the cytoplasm, which in turn disrupts normal metabolic activity.[246] One of these ionophore antibiotics is valinomycin (→), a molecule made by *Streptomyces* type bacteria.[247] The valinomycin molecule has a hydrophobic periphery and a hydrophilic core. It binds K^+ ions ~10^5 times more effectively than it binds Na^+ ions.



In the absence of specific K^+ channels and pumps, K^+ cannot pass through the membrane, the activation energy is too high. The valinomycin molecule continually shuttles back and forth across the membrane. In the presence of a K^+ gradient, that is a higher concentration of K^+ on one side of the membrane compared to the other, K^+ will tend to bind to the valinomycin molecule on the high K^+ concentration side, and be released from valinomycin on the low K^+ concentration side. The result is an increase in the net flux of K^+ from the high to the low concentration sides of the membrane. To be clear, in the absence of a gradient, K^+ ions will move across the membrane (in the presence of valinomycin), but there will be no net movement of K^+, no net flux. There are analogous carrier systems that move hydrophobic molecules within the aqueous phase.

Channels: Channel molecules sit within a membrane and contain an aqueous channel that spans the membrane's hydrophobic region. Hydrophilic molecules of particular sizes and shapes can pass through this aqueous channel and their movement involves a significantly lower activation energy than would be associated with moving through the lipid part of the membrane in the absence of the channel. Channels are generally highly selective in terms of which molecules will pass through them. For example, there are channels which will, on average, pass 10,000 K^+ ions for every one Na^+ ion.

Channel proteins exist in two or more distinct structural states. For example, in one state the channel can be open and allow particles to pass through or it can be closed, that is the channel can be turned on and off. Often the properties of these channels can be regulated. As an example, the binding of small molecules to a channel protein can lead to channel opening. Channels do not, however, determine in which direction an ion will move - net flux is based on the gradients across the membrane.

Another method of channel control depends on the fact that channel proteins are embedded within a membrane and contain charged groups. As we will see, cells can (and generally do) generate ion gradients, that is a separation of charged species across their membranes. For example if the concentration of K^+ is higher on one side of the membrane, there will be an ion gradient where the ions will (if movement is possible) move from the region of higher to lower K^+ concentration.[248] In some cases, the generation of ion gradients can, in turn, produce an electrical field across the plasma membrane. As these fields change, they can produce (induce) changes in channel structure that can switch the channel from open to closed and vice versa. Organisms

---

[246] There is little data in the literature on exactly which cellular processes are disrupted by which ionophore; in mammalian cells (as we will see) these molecules act by disrupting the energy storing ion gradients in mitochondria and chloroplasts, apparently.

[247] Valinomycin: https://en.wikipedia.org/wiki/Valinomycin

[248] In fact this tendency for species to move from high to low concentration until the two concentrations are equal can be explained by the Second Law of Thermodynamics. Check with your chemistry instructor for more details

typically have many genes that encode specific channel proteins that are involved in a range of processes from muscle contraction to thinking. Again, channels do not determine the direction of molecular motion. The net flux of movement is determined by the presence of molecular gradients, with the thermodynamic driver being entropic factors. That said, the actual movement of the molecules through the channel is driven by thermal motion.

Questions to answer:

91. What does it mean to move up (against) a concentration gradient? Is this a favorable or unfavorable event?
92. Where does the energy involved in moving molecules come from?
93. What happens to the movement of molecules through channels and transporters if we reverse the concentration gradients across a membrane?
94. Draw a diagram to show how K^+ ions are transported by an ionophore across a membrane. Draw a graph to show how the potential energy changes as the ion moves. Be sure to include the relative concentrations.

Questions to ponder:

- How might you prove that movements of molecules across a membrane occur in the absence of a gradient.

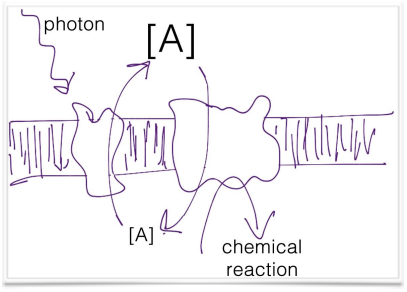
Generating gradients: using coupled reactions and pumps

Both carriers and channels allow the directional movement of molecules across a membrane, but there is a net directional flux only when a concentration gradient is present - that is if the concentration of the molecule is different on each side of the membrane. If a membrane contains active channels and carriers (as all biological membranes do), without the input of energy eventually the concentration gradients across the membrane will disperse. The $[\text{molecule X}]_{\text{outside}}$ will become equal to $[\text{molecule X}]_{\text{inside}}$. Removing a concentration gradient across a cell's plasma membrane is a good way to kill the cell. When we look at cells we find lots of concentration gradients, which raises the question, what produces and maintains these gradients.

The common sense (or rather thermodynamically correct) answer is that there must be molecules (generally proteins) that can transport specific types of molecules across the membrane and against their concentration gradient. We will call these types of molecules pumps and write the reaction they are involved in as:

$$[\text{Molecule}]_{\text{low concentration}} + \text{pump} \longleftrightarrow [\text{Molecule}]_{\text{high concentration}} + \text{pump}$$

As you might suspect moving this reaction to the right is thermodynamically unfavorable; like a familiar macroscopic pump, it will require the input of energy to work. We will have to "plug in" our molecular pump into some source of energy to move a molecule against its concentration gradient. So, what energy sources are available to biological systems? Basically we have two choices: the system can use electromagnetic energy (light) or it can use chemical energy. In a light-driven pump, there is a system that captures (absorbs) light; the absorbance of light (energy) is coupled to the pumping system (→). Where the pump is driven by a chemical reaction, a thermodynamically favorable reaction is often catalyzed by the pump, which also acts to facilitate the movement of one or more molecules against their membrane-associated concentration gradients.



A number of chemical reactions can be used to drive such pumps and these pumps can drive various reactions (remember reactions can move in both directions). One of the most common reactions involves the movement of energetic electrons through a membrane-bound, protein-based "electron transport" system; this, in turn, leads to the creation of an H^+ based electrochemical gradient. The thermodynamically favorable movement of H^+ down such a concentration gradient is

coupled to a reaction that leads to the synthesis of adenosine triphosphate (ATP) through reactions catalyzed by the membrane-bound ATP synthase enzyme:

$$H^+_{\text{(extracellular)}} \rightleftharpoons H^+_{\text{(intracellular)}}$$

ATP synthase (membrane-localized catalyst)

$$H^+ + \text{adenosine diphosphate (ADP)} + \text{phosphate} \rightleftharpoons \text{adenosine triphosphate (ATP)} + H_2O$$

The reaction takes cytoplasmic ADP, phosphate and H^+ and releases ATP and water into the cytoplasm. The thermodynamically favorable movement of H^+ down its concentration gradient is coupled to the thermodynamically unfavorable ATP synthesis reaction. The reaction can run in reverse, so that the thermodynamically favorable ATP hydrolysis reaction:

$$ATP + H_2O \rightleftharpoons ADP + \text{phosphate} + H^+$$

ATPase-driven pump (ATP synthase running "backward")

$$H^+_{\text{(intracellular)}} \rightleftharpoons H^+_{\text{(extracellular)}}$$

a reaction that results in the generation of a H^+ gradient across the membrane. So, we find that the same membrane molecule, the ATP synthase/pump, makes it possible to use energy present in a chemical gradient (across a membrane) to drive ATP synthesis within the cell and can enable ATP hydrolysis to generate a concentration gradient.

Simple Phototrophs

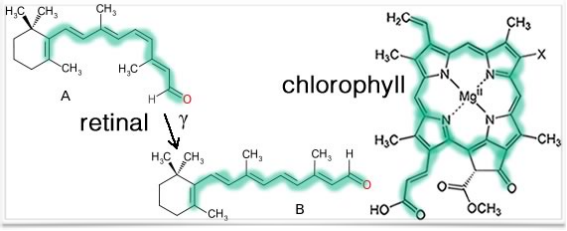
Phototrophs are organisms that capture photons (particles of light) and transform their electromagnetic energy into energy stored in unstable molecules, such as ATP and carbohydrates. Phototrophs “eat” light. Light can be considered as both a wave and a particle (that is quantum physics for you) and the wavelength of a photon reflects its "color" (as perceived by the brain) and the amount of energy it contains. Due to quantum mechanical considerations, a particular molecule will only absorb photons of specific wavelengths (energies). This property makes possible spectroscopic methods, and enables us to identify molecules (even when located at great distances) based on the photons they absorb or emit. Our atmosphere allows mainly visible light from the sun to reach the earth's surface, but most biological molecules do not absorb visible light very effectively if at all. To capture this energy, organisms have evolved the ability to synthesize molecules, known as pigments, that can capture (absorb) visible light, so that organisms can use their energy. The colors we see for a typical pigment are the colors of the light that is not absorbed but has been reflected. For example chlorophyl appears green because light in the red and blue regions of the spectrum is absorbed and green light is reflected. The general question we need to answer then is, how does the organism use this absorbed electromagnetic energy?

One of the simplest examples of a phototrophic system, that is, a system that directly captures the energy of light and transforms it into the energy stored in a chemical system, is provided by the archaea *Halobacterium halobium*.[249] *Halobacteria* are extreme halophiles (salt-loving) organisms. They live in waters that contain up to 5M NaCl. *H. halobium* uses the membrane protein bacteriorhodopsin to capture light. Bacteriorhodopsin consists of two components, a polypeptide, known generically as an opsin, and a non-polypeptide prosthetic group, the pigment retinal, a

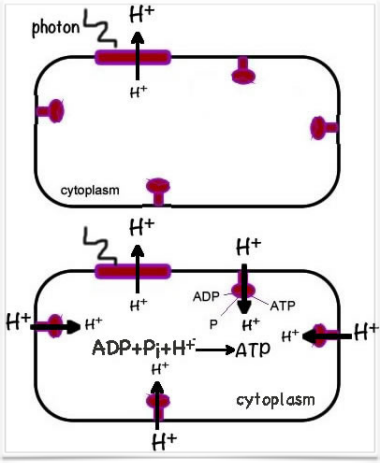
[249] Gradients and reactions (short video)

molecule derived from vitamin A.[250] Together the two, opsin + retinal, form the functional bacteriorhodopsin protein.

Because its electrons are located in extended molecular orbitals with energy gaps between them that are of the same order as the energy of visible light, absorbing a photon of visible light moves an electron from a lower to a higher energy molecular orbital. Such extended molecular orbitals (highlighted here →) are associated with molecular regions that are often drawn as involving alternating single and double bonds between carbons; these are known as conjugated π orbital systems. Conjugated π systems are responsible for the absorption of light by pigments such as chlorophyll and heme (the pigment that makes blood red. Heme includes an iron while chlorophyll includes a magnesium ion). When a photon of light is absorbed by the retinal group, it undergoes a reaction that leads to a change in the pigment molecule's shape and composition, which in turn leads to a change in the structure of the polypeptide to which the retinal group is attached. This is called a photo-isomerization reaction.



The bacteriorhodopsin protein is embedded within the plasma membrane where it associates with other bacteriorhodopsin proteins to form protein patches (→). These patches of membrane protein give the organisms their purple color and are known as purple membrane. When one of these bacteriorhodopsin proteins absorbs light, the change in the associated retinal group produces a light-induced change in protein structure that results in the movement of an H^+ ion from the inside to the outside of the cell. The protein and its associated pigment molecule then returns to its original low energy (ground) state, that is, its state before it absorbed the photon of light. The return of bacteriorhodopsin to the ground state is NOT associated with the movement of a H^+ ion across the membrane. Because all of the bacteriorhodopsin molecules in the membrane have the same orientation, as light is absorbed all of the H^+ ions move in the same direction across the membrane, leading to the formation of an H^+ concentration gradient with $[H^+]_{outside} > [H^+]_{inside.}$ This H^+ gradient is also associated with an electrical gradient because the movement of H^+ leads to more positive charge outside the cell. As light is absorbed the concentration of H^+ outside the cell increases and the concentration of H^+ inside the cell decreases. The question is, where are the moving H^+'s coming from? As you (perhaps) learned in chemistry, water undergoes a dissociation reaction (although this reaction is quite unfavorable):



$$H_2O \rightleftharpoons H^+ + OH^-$$

At pH, 7.0 water contains 10^{-7} moles of H^+ and it is these H^+ s that move.

As H^+s move across the membrane, they leave behind OH^- ions. The result is that the light driven movement of H^+ ions produces an electrical field, with excess + charges outside and excess – charges inside. As you know from your physics, positive and negative charges attract, but the intervening membrane stops them from reuniting. The result is the accumulation of positive charges on the outer surface of the membrane and negative charges on the inner surface. This charge separation produces an electric field across the membrane. Now, an H^+ ion outside of the cell will experience two distinct forces, those associated with the electric field and those arising from the

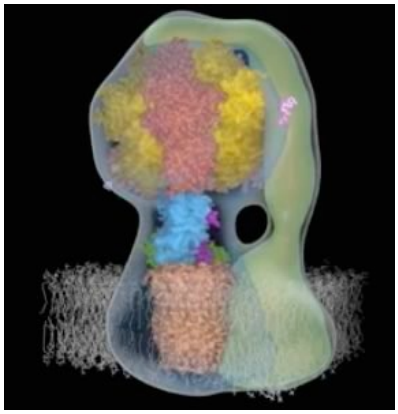
[250] As we will return to later, proteins are functional entities, composed of polypeptides and prosthetic group. The prosthetic group is essential for normal protein function. The protein without the prosthetic group is known as the apoprotein.

concentration gradient. If there is a way across the membrane, such a [H^+] gradient will lead to the movement of H^+ ions back into the cell. Similarly the electrical field will drive the movement of positively charged H^+ back into the cell. The formation of the [H^+] gradient generates a battery, a source of energy that the cell can use.

So how does the cell tap into this battery? The answer is through a second membrane protein, an enzyme known as the H^+–driven ATP synthase (↓). H^+ ions move through the ATP synthase molecule in a thermodynamically favorable sequence of reactions. The ATP synthase couples this favorable movement to an unfavorable chemical reaction, a condensation reaction leading to formation of ATP:

$$H^+_{outside} + ADP + \text{inorganic phosphate (Pi)} + H^+ \overset{\text{ATP synthase}}{\rightleftharpoons} ATP + H_2O + H^+_{inside}$$

ATPase pump (ATP synthase running backward)



This reaction continues as long as light is absorbed and for a short time afterward. In the light, bacteriorhodopsin acts to generate an H^+ gradient. When the light goes off (that is, at night time) the movement of H^+ ions through the ATP synthase continues to drive ATP synthesis until the H^+ gradient no longer has energy sufficient to drive the ATP synthesis reaction. The net result is that the cell uses light to generate ATP, which is stored for later use. ATP acts as a type of chemical battery, in contrast to the electrochemical battery of the H^+ gradient.

An interesting feature of the ATP synthase molecule (→) is that the H^+ ions move through it by hopping from one acidic amino acid to another in a thermodynamically favored sequence (video link). As the protons move, they change the interactions between parts of the ATP synthase, causing changes in shape, which in turn causes a region of the molecule to rotate. It rotates in one direction when it drives the synthesis of ATP and in the opposite direction to couple ATP hydrolysis to the pumping of H^+ ions against their concentration gradient. In this form it is better called an ATPase (or hydrolase) pump, involving the thermodynamically favorable reaction:

$$ATP + H_2O + H^+_{inside} \overset{\text{ATPase pump}}{\rightleftharpoons} H^+_{outside} + ADP + \text{inorganic phosphate (Pi)} + H^+$$

ATP synthase (ATPase pump running backward)

Because the enzyme rotates when it hydrolyzes ATP, it is rather easy to imagine how the energy released through this reaction could be coupled, through the use of an attached paddle-like extension, to drive cellular or fluid movement.

Questions to answer

95. Indicate in a diagram the direction of H^+ movement in a phototroph when exposed to light.
96. Why does the H^+ gradient across the membrane dissipate when the light goes off? What happens to the rate of ATP production? When does ATP production stop and why?
97. Are there limits the "size" of the H^+ gradient that bacteriorhodopsin can produce and why (or why not)?
98. What is photoisomerization? Is this a reversible or an irreversible reaction?

Questions to ponder

- How might ATP hydrolysis lead to cell movement.
- What would happen if bacteriorhodopsin molecules were oriented randomly within the membrane

Chemo-osmosis (an low level overview)

One of the most surprising discoveries in biology was the wide spread, almost universal, use of H^+–based electrochemical gradients to generate ATP. What was originally known as the chemiosmotic hypothesis was produced by the eccentric British scientist, Peter Mitchell (1920–1992).[251] Before the significance of H^+ membrane gradients was widely appreciated, Mitchell proposed that energy captured through the absorption of light (by phototrophs) or the breakdown of molecules into more stable molecules (by various types of chemotrophs) relied on the same basic (homologous, that is, evolutionarily-related) mechanism, namely the generation of H^+ gradients across membranes (the plasma membrane in prokaryotes and the internal membranes of mitochondria and chloroplasts (intracellular organelles, derived from bacteria – see below) in eukaryotes.

What makes us think that these processes might have a similar evolutionary root, that they are homologous? Basically, it is the observation that in both light- and chemical-based processes captured energy is transferred through the movement of electrons through a structurally similar membrane-embedded "electron transport chain" composed of a series of membrane and associated proteins and involving a series of reduction-oxidation (redox) reactions (see below) during which electrons move from a high energy (relatively unstable)donor to a lower energy (more stable) acceptor. Some of the energy difference between the two is used to move H^+ ions across the membrane, generating a H^+ concentration gradient. Subsequently the thermodynamically favorable movement of H^+ down this concentration gradient (across the membrane) is used to drive ATP synthesis, a thermodynamically unfavorable reaction. ATP synthesis itself involves the rotating ATP synthase. The reaction can be written:

$$H^+_{outside} + ADP + Pi + H^+ \rightleftharpoons ATP + H_2O + H^+_{inside,}$$

where "inside" and "outside" refer to compartments defined by the membrane containing the electron transport chain and the ATP synthase, with the ATP synthesis reaction occurring within the membrane-bound compartment. Again, this reaction can run backwards. When this occurs, the ATP synthase acts as an ATPase (ATP hydrolase) that can pump H^+ (or other molecules) against their concentration gradient. Such pumping ATPases establish most of the biologically important ion gradients across membranes. In such a reaction:

ATP+H_2O+molecule in low concentration region $\rightleftharpoons$ ADP+Pi+molecule in high concentration region.

The most important difference between phototrophs and chemotrophs is, essentially, where do the high energy electrons come from - energized by absorption of light or derived from unstable molecules.

Oxygenic photosynthesis

Compared to the salt loving archaea *Halobium,* with its purple bacteriorhodopin-rich membranes, photosynthetic cyanobacteria (which are true or eubacteria), green algae, and higher plants (both eukaryotes) use more complex molecular systems through which to capture and utilize light. The photosynthetic systems of these organisms appear to be homologous, that is, derived from a common ancestor. For simplicity's sake we will describe the photosynthetic system of cyanobacterium; the system in eukaryotic algae and plants, while more complex, follows the same basic logic and appears to derived, evolutionarily, from the cyanobacterial system.[252] We will consider only one aspect of this photosynthetic system, known as the oxygenic or non-cyclic system

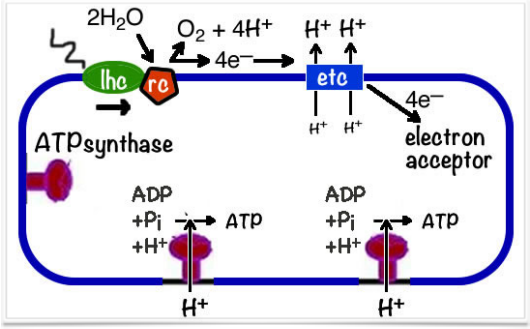
---

[251] Chemo-osmosis and Peter Mitchell (wikipedia)

[252] Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes

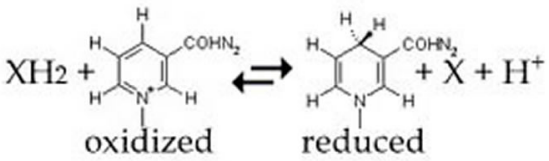
(look to more advanced classes for more details.) The major pigment in this system, chlorophyll, is based on a complex molecule, a porphyrin (see above); it is these pigments that give plants their green color. As in the case of retinal, they absorb visible light due to the presence of a conjugated (resonance) bonding structure (typically drawn as a series of alternating single and double) carbon-carbon bonds. Chlorophyll is synthesized by a conserved biosynthetic pathway. Variants of this scheme are used to synthesize heme, which is found in the hemoglobin of animals and in the cytochromes, within the electron transport chain present in both plants and animals (which we will come to shortly), vitamin B_{12}, and other biologically important prosthetic (that is non-polypeptide) groups associated with proteins and required for their normal function.[253]

Chlorophyll molecules are organized into two distinct membrane-embedded protein complexes. These are known as the light harvesting and reaction center complexes. Light harvesting complexes ("lhc") provide extra surface area to increase the amount of light the organism can capture. When a photon is absorbed, an electron is excited to a higher molecular orbital. An excited electron can be passed between components of the lhc and eventually to the reaction center ("rc") complex (←). Light harvesting complexes are important because photosynthetic organisms often compete with one another for light; increasing the efficiency of the system through which an organism captures light can provide a selective (evolutionary) advantage.



In the oxygenic, that is molecular oxygen (O_2) generating photosynthesis reaction system, high energy (excited) electrons are passed from the reaction center through a set of membrane proteins, the electron transport chain ("etc"). As an excited electron moves through the electron transport chain its energy is used to move H^+s from inside to outside of the cell. This is the same geometry of movement that we saw previously in the case of the purple membrane system. The end result is the generation of an H^+ based electrochemical gradient. As with purple bacteria, the energy stored in this H^+ gradient is used to drive the synthesis of ATP within the cell's cytoplasm, a coupled reaction catalyzed by the ATP synthase.

Now you might wonder, what happens to the originally excited electrons, and the energy that they carry. In what is known as the cyclic form of photosynthesis, low energy electrons from the electron transport chain are returned to the reaction center, where they regenerate the pigment molecules to their original (before they absorbed a photon) state. In contrast, in the non-cyclic process that we have been considering, electrons from the electron transport chain are delivered to an electron acceptor. Generally this involves the absorption of a second photon, a mechanistic detail that need not trouble us here. This is a general type of chemical reaction known as a reduction-oxidation (redox) reaction. Where an electron is within a molecule's electron orbital system influences the amount of energy present in the molecule: adding a negative charge (an electron) to a molecule can increase electron-electron repulsion and raise the molecule's potential energy. When an electron is added to a molecule, that molecule is said to have been "reduced", and yes, it does seem weird that adding an electron "reduces" a molecule (→). Generally, when an electron is removed, the molecule's energy is changed (decreased) and the



[253] Mosaic Origin of the Heme Biosynthesis Pathway in Photosynthetic Eukaryotes:

molecule is said to have been "oxidized".[254] Electrons, like energy, are neither created nor destroyed in biological systems, so the reduction of one molecule is always coupled to the oxidation of another. In a system of redox reactions, electrons removed from the reduced molecule are used to drive various types of thermodynamically unfavorable reactions, including the movement of H^+ across a membrane.

Again, the laws of conservation imply that when electrons leave the photosynthetic system (in the non-cyclic process) they must be replaced. So where do these electrons come from? Here we see what appears to be a major evolutionary breakthrough. During the photosynthetic process, the reaction center couples light absorption to the oxidation (removal of electrons) from water molecules:

$$\text{light} + 2H_2O \rightleftharpoons 4H^+ + 4e^- + O_2.$$

The four electrons, derived from two molecules of water, pass to the reaction center, while the $4H^+$s contribute to the proton gradient across the membrane.[255] O_2 is a waste product of this reaction. Over millions of years, the photosynthesis-driven release of O_2 changed the Earth's atmosphere from containing essentially 0% molecular oxygen to the current ~21% level at sea level. Because O_2 is highly reactive, this transformation is thought to have been a major driver of a number of subsequent evolutionary changes. However, there remain organisms that cannot use O_2 and cannot survive in its presence. They are known as obligate anaerobes, to distinguish them from organisms that normally grow in the absence of O_2 but that can survive in its presence; these are known as facultative anaerobes. In the past the level of atmospheric O_2 has changed dramatically; its level is based (primarily) on how much O_2 is released into the atmosphere by oxygenic photosynthesis and how much is removed by various reactions, such as the decomposition of plant materials. When large amounts of plant materials are buried before they can decay, such as occurred from ~360 to 299 million years ago with the formation of coal beds during the Carboniferous period, the level of atmospheric O_2 increased dramatically, apparently reaching levels of ~35%. It is speculated that such high levels of atmospheric molecular oxygen made it possible for organisms without lungs (like insects) to grow to gigantic sizes.[256]

Chemotrophs

Organisms that are not phototrophic capture energy from other sources, specifically by transforming thermodynamically unstable molecules into more stable species. Such organisms are known generically as chemotrophs. They can be divided into various groups, depending upon the types of food molecules (energy sources) they use: these include organotrophs, which use carbon-containing molecules (you yourself are an organotroph) and lithotrophs or rock eaters, which use various inorganic molecules. In the case of organisms that can "eat" H_2, the electrons that result are delivered, along with accompanying H^+ ions, to CO_2 to form methane (CH_4) following the reaction:

$$CO_2 + 4H_2 \rightleftharpoons CH_4 + 2H_2O.$$

Such organisms are referred to as methanogens (methane-producers).[257] In the modern world methanogens (typically archaea) are found in environments with low levels of O_2, such as your gut. In many cases reactions of this type can occur only in the absence of O_2. In fact O_2 is so reactive, that it can be thought of as a poison for organisms that cannot actively "detoxify" it. When we think

---

[254] you can review redox here or in CLUE

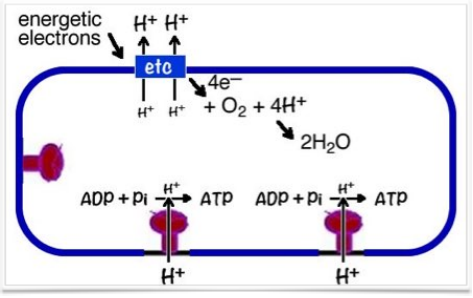
[255] Photosystem II and photosynthetic oxidation of water: an overview

[256] When Giants Had Wings and 6 Legs

[257] Lithotrophic (wikipedia)

about the origins and subsequent evolution of life, we have to consider how organisms that originally arose in the absence of O_2 adapted as significant levels of O_2 began to appear in their environment. It might be that modern obligate anaerobes might still have features common to the earliest organisms.

The amount of energy that an organism can capture is determined by the energy of the electrons that the electron acceptor(s) they employ can accept. If only electrons with high amounts of energy can be captured, which is often the case, then inevitably large amounts of energy are left behind, with the acceptor. On the other hand, the lower the amount of energy that an electron acceptor can accept, the more energy can be extracted and captured from the original "food" molecules and the less energy is left behind. Molecular oxygen is unique in its ability to accept low energy electrons (→). For example, consider an organotroph that eats carbohydrates (carbon plus water); molecules with the general composition $[C_6H_{10}O_5]_n$). This class of molecules includes sugars, starches, and wood. These molecules undergo a process known as glycolysis, from the Greek words meaning sweet (glyco) and splitting (lysis). In the absence of O_2, that is under anaerobic conditions, the end product of the breakdown of a carbohydrate leaves ~94% of the theoretical amount of energy present in the original carbohydrate molecule in molecules that cannot be broken down further, at least by most organisms. These are molecules such as ethanol (C_2H_6O) and lactic acid ($CH_3CH(OH)CO_2H$). However, when O_2 is present, carbohydrates can be broken down more completely into CO_2 and H_2O, a process known as respiration. In such O_2 using (aerobic) organisms, the energy released by the formation of CO_2 and H_2O is transferred to (stored in) energetic electrons and used to generate a membrane-associated H^+ based electrochemical gradient that in turn drives ATP synthesis, through a membrane-based ATP synthase. In an environment that contains molecular oxygen, organisms that can use O_2 as an electron acceptor have a distinct advantage; instead of secreting energy rich molecules, like ethanol, they release the energy poor (stable) molecules CO_2 and H_2O.



No matter how cells (and organisms) capture the energy needed to maintain themselves and to grow, they must make a wide array of complex molecules. Understanding how these molecules are synthesized lies (traditionally) within the purview of biochemistry. That said, in each case, thermodynamically unstable molecules (like lipids, proteins, and nucleic acids) are built through series of coupled reactions that rely on energy captured from light or the break down of food molecules.

Questions to answer

99. How (do you suppose) does an electron move through an electron transport chain? Make a diagram and a graph that describes its energy as it moves through the chain.
100. In non-cyclic photosynthesis, where do electrons end up?
101. What would happen to an aerobic cell's ability to make ATP if it where exposed to an H^+ carrier or channel?
102. Why are oxidation and reduction always coupled?
103. Why are carbohydrates good for storing energy?

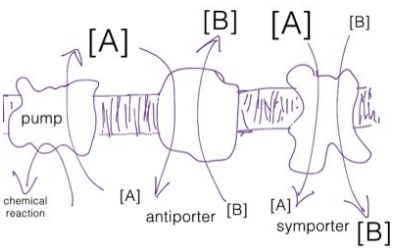
Questions to ponder

- Which do you think would have a greater evolutionary advantage, an organism growing aerobically or anaerobically? What factors influence your answer?

Using the energy stored in membrane gradients

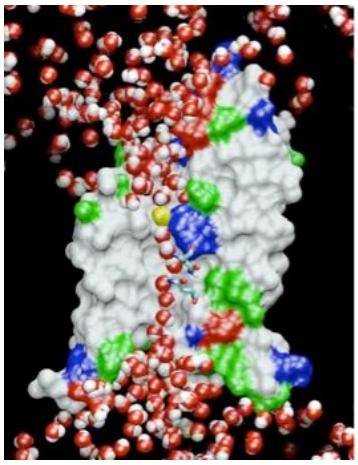
The energy captured by organisms is used to drive a number of processes in addition to synthesis reactions. For example, we have already seen that ATP synthases can act as pumps

(ATP-driven transporters), coupling the favorable ATP hydrolysis reaction to the movement of molecules against their concentration gradients (→). The resulting gradient is a form of stored (potential) energy, energy that can be used to move other molecules, that is molecules that are not moved directly by a ATP-driven transporter.[258] Such processes involve what is known as coupled transport.[259] They rely on membrane-bound proteins that enable a molecule to pass through a membrane, and so allow for a net flux down a concentration gradient. In contrast to simple carriers and channels, however, this thermodynamically favorable net flux down, that is, from high concentration to low concentration, is physically coupled to the movement of a second net flux against a gradient, that is from low to high concentration. When the two transported molecules move in the same direction, the transporter is known as a symporter; when they move in opposite directions, it is known as an antiporter. Which direction(s) the molecules move will be determined by the nature of the transporter and the relative sizes of the concentration gradients of the two types of molecules moved. There is no inherent directionality associated with the transporter itself - the net movement of molecules reflects the relative concentration gradients of the molecules that the transporter can productively bind. What is important here is that energy stored in the concentration gradient of one molecule can be used to drive the movement of a second type of molecule against its concentration gradient. In mammalian systems, it is common to have Na^+, K^+, and Ca^{2+} gradients across the plasma membrane, and these are used to transport molecules into and out of cells. Of course, the presence of these gradients implies that there are ion-specific pumps that couple an energetically favorable reaction, typically ATP hydrolysis, to an energetically unfavorable reaction, the movement of an ion against its concentration gradient. Without these pumps, and the chemical reactions that drive them, the membrane battery would quickly run down. Many of the immediate effects of death are due to the loss of membrane gradients and much of the energy needs of cells (and organisms) involves running pumps maintain the non-equalibrium state of the cell.



Osmosis and living with and without a cell wall

Cells are packed full of molecules. These molecules take up space, space that will not be occupied by water molecules. The concentration of water outside of the cell $[H_2O]_{out}$ will generally be higher than the concentration of water inside the cell $[H_2O]_{in}$. This solvent concentration gradient leads to the net movement of water into the cell.[260] Such a movement of solvent is known generically as osmosis. Much of this movement occurs through the membrane, which is somewhat permeable to water (see above). A surprising finding, which won Peter Agre a share of the 2003 Noble prize in chemistry, was that the membrane also contains water channels, known as aquaporins.[261] Follow the video link (→) to a molecular simulation of a water molecule (yellow) moving across a membrane, through an aquaporin protein. It turns out that the rate of osmotic movement of water is dramatically



[258] Although we will not consider it hear, membrane gradients are also used to send signals throughout the nervous system.

[259] Structural features of the uniporter/symporter/antiporter superfamily

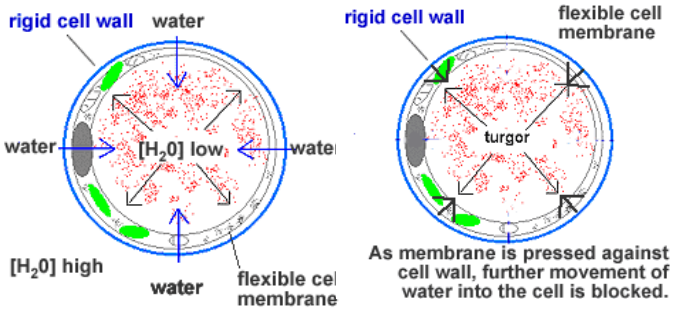
[260] An important note is that in chemistry classes you may be taught that water moves from a region of low to high SOLUTE concentration. These two definitions of osmosis mean the same thing but it is easy to get confused.

[261] Water Homeostasis: Evolutionary Medicine: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3540612/

reduced in the absence of aquaporins. In addition to water, aquaporin-type proteins can facilitate the movement of other small uncharged molecules across cellular membranes.

The difference or gradient in the concentrations of water across the cell membrane, together with the presence of aquaporins, leads to a system that is capable of doing work. The water gradient, can lift a fraction of the solution against the force of gravity, something involved in how plants stand up straight. How is this possible? If we think of a particular molecule in solution, it moves through collisions with its neighbors. These collisions drive the stochastic movement of particles. But if there is a higher concentration of molecules on one side of a membrane compared to the other, then the random movement of molecules will lead to a net flux of molecules from the area of high concentration to that of low concentration, even though each molecule, on its own moves, randomly stochastically, that is, without a preferred direction [this video is a good illustration of this behavior]. At steady state in a biological system, the force generated by the net flux of water moving down its concentration gradient is balanced by forces acting in the other direction.

The water concentration gradient across the plasma membrane of most organisms leads to an influx of water into the cell. As water enters, the plasma membrane expands; you might want to think about how that occurs, in terms of membrane structure. If the influx of water continues unopposed, the membrane would eventually burst like an over-inflated balloon, killing the cell. One strategy to avoid this lethal outcome, adopted by a range of organisms, is to build a semi-rigid "cell wall" external to the plasma membrane (→). The synthesis of this cell wall is based on the controlled assembly of macromolecules secreted by the cell through various processes. As osmosis drives water through the plasma membrane and into the cell, the plasma membrane is pressed up against the cell wall. The force exerted by the rigid cell wall on the membrane balances the force of water entering the cell. When the two forces are equal, the net influx of water into the cell stops. Conversely, if $[H_2O]_{outside}$ decreases, this pressure is reduced, the membrane moves away from the cell wall and, because they are only semi-rigid, the walls flex. It is this behavior that causes plants to wilt when they do not get enough water. These are passive behaviors, based on the structure of the cell wall; they are built into the wall as it is assembled. Once the cell wall has been built, a cell with a cell wall does not need to expend energy to resist osmotic effects. Plants, fungi, bacteria and archaea all have cell walls. A number of antibiotics work by disrupting the assembly of bacterial cell walls. This leaves the bacteria osmotically sensitive, water enters these cells until they burst and die.



Questions to answer:

104. Make a graph of water concentration across a typical cellular membrane for an organism living in fresh water; explain what factors influenced your prediction.
105. How might cell wall-less organisms deal with challenges associated with the absence of a cell wall?
106. Plants and animals are both eukaryotes; how would you decide whether the common ancestor of the eukaryotes had a cell wall.
107. What are potential evolutionary benefits of losing a cell wall?
108. There is a concentration gradient of A across of membrane, but no net flux – what can we conclude?

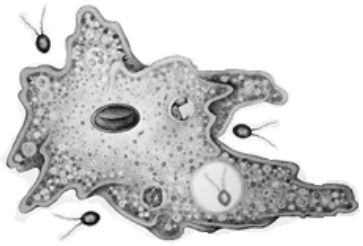
Questions to ponder:

- Why might an aquaporin channel not allow a Na^+ ion to pass through it?

An evolutionary scenario for the origin of eukaryotic cells

When we think about how life arose, and what the first organisms looked like, we are moving into an area where data is fragmentary or unobtainable and speculation is rampant. These are also

events that took place billions of years ago. But there is relevant data present in each organisms' genetic data (its genome), the structure of its cells, and their ecological interactions. It is this type of data that can inform and constrain our various speculations.



Animal cells do not have a rigid cell wall; its absence allows them to be active predators, moving rapidly and engulfing their prey whole or in macroscopic bits through phagocytosis (see above). They use complex "cytoskeletal" and "cytomuscular" systems to drive these thermodynamically unfavorable behaviors (→). Organisms with a rigid cell wall can't perform such functions. Given that bacteria and archaea have cell walls, it is possible that cell walls were present in their common ancestor. This leads us to think more analytically about the nature of the earliest organisms and the path back to the common ancestor. A cell wall is a complex structure that would have had to be assembled through evolutionary processes before it would be useful. If we assume that the original organisms arose in an osmotically friendly, that is, non-challenging environment, then a cell wall could have been generated in steps, and once adequate it could enable the organisms that possessed it to build more complex cytoplasmic spaces and to invade new, more osmotically challenging (dilute) environments, or both. Another plausible scenario is that the ancestors of the bacteria and the archaea originally developed cell walls as a form of protection against predators. So who were these predators? Where they the progenitors of the eukaryotes? If so, it might be that organisms in the eukaryotic lineage never had a cell wall (and that neither did the ancestors of the bacteria and archaea. In this scenario, the development of eukaryotic cell walls by fungi and plants represents an example of convergent evolution and that these structures are analogous (rather than homologous) to the cell walls of prokaryotes (bacteria and archaea).

But now a complexity arises, there are plenty of eukaryotic organisms, including microbes like the amoeba, that live in osmotically challenging environments. How do they deal with the movement of water into their cells? How might they have followed their prey (bacteria and archaea) into the non-salty world? One approach is to actively pump the water that flows into them back out using an organelle known as a contractile vacuole. Water accumulates within the contractile vacuole, a membrane-bounded structure within the cell; as the water accumulates the contractile vacuole inflates. To expel the water, the vacuole connects with the plasma membrane and is squeezed by the contraction of a cytomuscular system, squirting the water out of the cell. The process of vacuole contraction is an active one, it involves work and requires energy.[262] One might speculate that such as cytomuscular system was originally involved in predation in the salty world, that is, enabling the cell to move its membranes, to surround and engulf other organisms (phagocytosis). The resulting vacuole became specialized to aid in killing and digesting the engulfed prey. When digestion is complete, this micro-stomach can fuse with the plasma membrane to discharge the waste, using either a passive or an active contractile system. It turns out that the molecular systems involved in driving active membrane movement are related to the systems involved in dividing the eukaryotic cell into two during cell division; a distinctly different system from that used by prokaryotes.[263] So which came first, distinct cell division mechanisms that led to differences in membrane behavior, with one leading to a predatory active membrane and the other to a passive membrane, perhaps favoring the formation of a cell wall? At this point it is hard (impossible?) to know.

---

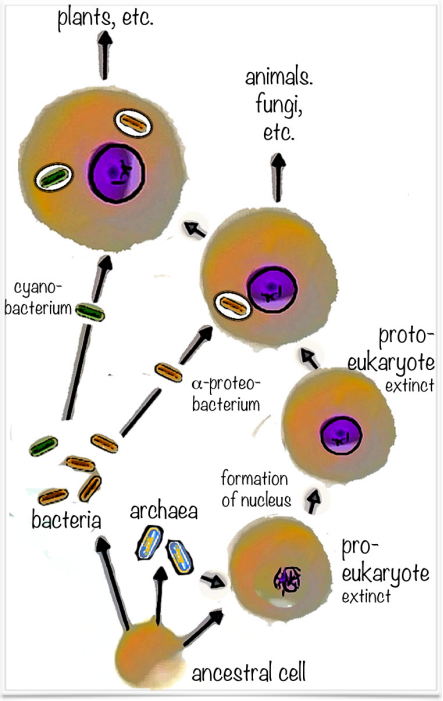
[262] Very cool video of a contractile vacuole in paramecium and explanation

[263] The cell cycle of archaea & Bacterial cell division

Making a complete eukaryote

Up to this point we have touched on only a few of the ways that prokaryotes (bacteria and archaea) differ from eukaryotes. The major differences include the fact that eukaryotes have their genetic material isolated from the cytoplasm by a complex double-layered membrane/pore system known as the nuclear envelope (discussed later on). Exactly how the nucleus came into being in the lineage leading to eukaryotes remains poorly defined, as is often the case in historical processes that occurred billions of years ago.[264] Another difference is the relative locations of chemo-osmotic/ photosynthetic systems in the two types of organisms. In prokaryotes, these systems (light absorbing systems, electron transport chains and ATP synthases) are located within the plasma membrane or within plasma membrane-derived internal membrane vesicles. In contrast, in eukaryotes (plants, animals, fungi, protozoa, and other types of organisms) these structural components are not located on the plasma membrane, but rather within discrete and distinctive intracellular structures. In the case of the system associated with aerobic respiration, these systems are found in the inner membranes of a double-membrane bound cytoplasmic organelles known as a mitochondrion (plural: mitochondria). Photosynthetic eukaryotes (algae and plants) have a second type of membrane-bounded cytoplasmic organelle, known as chloroplasts, in addition to mitochondria. Like mitochondria, chloroplasts are characterized by the presence of a double membrane and an electron transport chain located within the inner membrane and membranes apparently derived from it.

These are just the type of structures one might expect to see if a bacterial cell was engulfed by the ancestral pro-eukaryotic cell, with the host cell's membrane surrounding the engulfed cells plasma membrane (→). A more detailed molecular analysis reveals that the mitochondrial and chloroplast electron transport systems, as well as the ATP synthase proteins, more closely resemble those found in two distinct types of bacteria, rather than in archaea. In fact, detailed analyses of the genes and proteins involved suggest that the electron transport/ATP synthesis systems of eukaryotic mitochondria are homologous to those of a α-proteobacteria while the light harvesting/reaction center complexes, electron transport chains and ATP synthesis proteins of algae and plants appear to be homologous to those of a second type of bacteria, a photosynthetic cyanobacteria.[265] In contrast, many of the nuclear systems found in eukaryotes appear more similar to those systems present in archaea. How do we make sense of these observations?



When a eukaryotic cell divides it must also have replicated its mitochondria and chloroplasts, otherwise they would eventually be lost through dilution. In 1883, Andreas Schimper (1856-1901) noticed that chloroplasts divided independently of their host cells. Building on Schimper's observation, Konstantin Merezhkovsky (1855-1921) proposed that chloroplasts were originally independent organisms and that plant cells were symbionts, essentially two independent organisms living together. In a similar vein, in 1925 Ivan Wallin (1883-1969) proposed that the mitochondria of eukaryotic cells were derived from bacteria. This "endosymbiotic hypothesis" for the origins of eukaryotic mitochondria and chloroplasts fell out of favor, in large part because the molecular methods needed to unambiguously resolve their implications were not available. A breakthrough came with the work of Lynn Margulis (1938-2011)

---

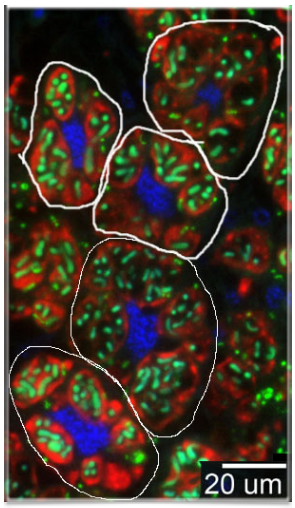
[264] Endosymbiotic theories for eukaryote origin

[265] The origin and early evolution of mitochondria and The Origin and Diversification of Mitochondria

and was further bolstered when it was found that both the mitochondrial and chloroplast protein synthesis machineries were sensitive to drugs that inhibited bacterial but not eukaryotic protein synthesis. In addition, it was discovered that mitochondria and chloroplasts contained circular DNA molecules organized in a manner similar to the DNA molecules found in bacteria (we will consider DNA and its organization soon).

All eukaryotes appear to have mitochondria. Suggestions that some eukaryotes, such as the human anaerobic parasites *Giardia intestinalis*, *Trichomonas vaginalis* and *Entamoeba histolytica* [266] do not failed to recognize cytoplasmic organelles, known as mitosomes, as degenerate (evolutionarily simplified) mitochondria. Based on these and other data it now seems likely that all eukaryotes are derived from a last common (eukaryotic) ancestor (sometime referred to as LECA) that engulfed an aerobic α-proteobacteria-like bacterium. Instead of being killed and digested, these (or even one) of these bacteria survived within the pre-eukaryotic cell, replicated, and were distributed into the progeny cell when the parent cell divided. This process resulted in the engulfed bacterium becoming an endosymbiont, which over time became mitochondria. In the course of time, the original genome of the bacterium has been dramatically reduced in size, with many (but not all) genes transferred to the nucleus (we will consider the implications of this process later on). At the same time the engulfing cell became dependent upon the presence of the endosymbiont, initially to detoxify molecular oxygen, and then to utilize molecular oxygen as an electron acceptor so as to maximize the energy that could be derived from the break down of complex molecules. All eukaryotes, including us, are descended from this mitochondria-containing eukaryotic ancestor, which has been estimated to have appeared ~2 billion years ago. The second endosymbiotic event in eukaryotic evolution occurred when a cyanobacteria-like bacterium formed a relationship with a mitochondria-containing eukaryote. This lineage gave rise to the glaucophytes, the red and the green algae. The green algae, in turn, gave rise to the plants.

As we look through modern organisms there are a number of examples of similar events, that is, one organism becoming inextricably linked to another through symbiotic processes. There are also examples of close couplings between organisms that are more akin to parasitism rather then a mutually beneficial interaction (symbiosis).[267] For example, a number of insects have intracellular bacterial parasites and some pathogens and parasites live inside human cells.[268] In some cases, even these parasites can have parasites. Consider the mealybug *Planococcus citri*, a multicellular eukaryote; this organism contains cells known as bacteriocytes (outlined in white →). Within the bacteriocytes are *Tremblaya princeps* (β-proteobacteria) cells (red). Surprisingly, within these *T. princeps* cells live *Moranella endobia*-type γ-proteobacteria (green).[269] In another example, after the initial endosymbiotic event that formed the proto-algal cell, the ancestor of red and green algae and the plants, there have been other endocytic events in which a eukaryotic cell has engulfed and formed an endosymbiotic relationship with eukaryotic green algal cells, to form a "secondary" endosymbiont, and secondary endosymbionts have been found engulfed by yet another eukaryote, to form a tertiary endosymbiont.[270] The conclusion is that there are combinations of cells that can survive (and more importantly reproduce) better



[266] The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite Entamoeba histolytica

[267] Mechanisms of cellular invasion by intracellular parasites: http://www.ncbi.nlm.nih.gov/pubmed/24221133

[268] Intracellular protozoan parasites of humans: the role of molecular chaperones in development and pathogenesis.

[269] Snug as a Bug in a Bug in a Bug & Mealybugs nested endosymbiosis

[270] Photosynthetic eukaryotes unite: endosymbiosis connects the dots

in a particular ecological niche than either could alone. In these phenomena we see the power of evolutionary processes to populate extremely obscure and limited ecological niches in rather surprising ways.

Questions:

109. How would you define an osmotically friendly environment? what would be its limitations, evolutionarily?
110. Are the mitochondria of plants and animals homologous or analogous? How might you decide?
111. What advantage might a host get from a bacterial symbionts? Was there an advantage for the engulfed bacteria?
112. How would you distinguish a symbiotic from a parasitic relationship? is it always simple?

Questions to ponder:

- Why might a plant cell not notice the loss of its mitochondria? why do you think plants retain mitochondria?
- What evidence would lead you to suggest that there had been multiple symbiotic events that gave rise to the mitochondria of different eukaryotes?

Chapter 7: The molecular nature of the heredity material

In which we discover how the physical basis of genetic inheritance, DNA, was identified and learn about the factors that influence how it is that DNA encodes genetic information, how that information is replicated and read out and often "translated" into useable forms (polypeptides), how mutations occur and may be repaired, and how such extravagantly long molecules are organized within such small cells.

One of the most amazing facts associated with Darwin and Wallace's original evolutionary model was their lack of a coherent understanding of genetic mechanisms. While it was clear, based on the experiences of plant and animal breeders, that organisms varied with respect to one another and that part of that variation could be inherited from the organism's parents, the mechanism(s) by which genetic information is stored and transmitted was unclear and, at the time, essentially unknowable, a situation that promoted much speculation, including a number of hypotheses based on supernatural or metaphysical mechanisms.[271] For example, some proposed that evolutionary variation was generated by an "inner drive" acting at organismic or even at the species level - an idea known as orthogenesis. Orthogenesis had the comforting implication that evolutionary processes reflected some form of purposeful design, that things were going somewhere, that there was a purpose to existence. On the negative side, such an orthogenic model served to support toxic racism, in which different types of organisms or different populations of people represented different levels of perfection.[272] Well before the modern theory of evolution was proposed in 1859, Jean-Baptiste Lamarck (1744–1829) suggested that inheritance somehow reflected the desires and experiences of the parent.[273] Such a model presumes a type of "internally directed" and purposeful form of evolution, the idea that evolutionary change reflects the desires, needs, and experiences of individuals. In contrast Darwin's model, based on random variations in the genetic material, seemed more arbitrary and unsettling, as it implied a lack of an over-arching purpose to life in general, and human existence in particular.

The scientific study of inheritance, which led to the modern disciplines of genetics and molecular biology has its origins in the work of Gregor Mendel (1822–1884). He published his work on sexually reproducing peas in 1865, shortly after the introduction of the modern theory of evolution. Darwin published multiple revised editions of "On the Origin of Species" through 1872, so it is fair to ask why he did not incorporate a Mendelian view of heredity into his theory? The simplest explanation would be that Darwin was unaware of Mendel's work – in fact, the implications of Mendel's work were largely ignored until the early years of the 20th century.

So why was the significance of Mendel's work not immediately recognized? It turns out that Mendel's conclusions were quite specialized and not obviously broadly applicable. Mendel carefully bred pea plants, *Pisum sativum,* to produce discrete traits (phenotypes) that differed from the variable traits found "in the wild" (see above). After this in-breeding, he had plants that displayed what are known as dichotomous traits (one or the other): smooth versus wrinkled seeds, yellow versus green seeds, grey versus white seed coat, tall versus short plants. In contrast, in the wild,

---

[271] The eclipse of Darwin: wikipedia

[272] Evidence for perfection in people, as a species, seems consciously absent.

[273] It is worth reading Evolution in Four Dimensions (reviewed here) which reflects on the factors that influence selection.

these traits occurred along a continuum, with various intermediate phenotypes.[274] Relatively few traits are dichotomous. In addition, the traits he selected were independent, the presence or absence of one trait did not influence any of the other traits he examined. Each trait was controlled, as we know now, by variations at a single genetic locus (gene or position within the genome). Different genes "produced" different traits independently of one another. As we will see, the connection between genetic information and a particular trait is often much more complex.[275] The vast majority of traits do not behave in a simple Mendelian manner; most genes have roles in a number of different traits and a particular trait is generally controlled (and influenced) by variations in many genes. Allelic variations in multiple genes, often referred to as the genetic background, interact in emergent, and not easily predictable, ways. For example, the extent to which a trait is visible, even assuming the underlying genetic factor (allele) is present, can vary dramatically depending upon the rest of the organism's genetic background. Finally, in an attempt to established the general validity of his conclusions Mendel was urged to examine the behavior of a number of other plants, including hawkweed. Unfortunately, hawkweed uses a specialized, asexual reproductive strategy, known as apomixis, which does not follow Mendel's rules.[276] This did not help reassure Mendel or others that his genetic laws were universal or useful. Subsequent work, published in 1900, led to the recognition of the general validity of Mendel's basic conclusions.[277]

Mendel deduced that there are stable hereditary "factors" – which became known as genes – and that genes are present as discrete objects within an organism. Each gene can exist in a number of different forms, known as alleles. In many cases specific alleles (versions of a gene) are associated with specific forms of a trait or the presence or absence of a trait. For example, in mammals, the ability to digest lactose depends upon whether you can make the enzyme lactase. The lactase enzyme is encoded by the *LCT* gene.[278] Lactase is made when the *LCT* gene is expressed. In most mammals, the *LCT* gene stops being expressed with age. In ~65% of human adults the expression of the *LCT* gene, and so lactase production, is off. In various sub-populations *LCT* expression, and so the ability to digest lactose, persists in adults – a trait known as adult lactose tolerance. Adult lactose tolerance has arisen independently in a number of human populations. One version of adult lactose tolerance is based on the allele of the *MCM6* gene you carry. The *MCM6* allele that promotes adult lactose tolerance acts to maintain the expression of the *LCT* gene into adulthood. As we proceed, we will consider the molecular level details involved in processes such as adult lactose tolerance. You have already encountered the terms genes, alleles, genomes, genotypes and phenotypes from our previous discussion of evolutionary mechanisms, and we will consider them again in greater detail as we proceed.

When a cell divides, all of its genes must be replicated so that each daughter cell receives a full set of genes, a genome. The exact set of alleles a cell inherits determines its genotype. Later it was recognized that sets of genes are linked together in a physical way, but that this linkage is not permanent - that is, processes exist that can shuffle the alleles of linked genes. In sexually reproducing organisms, such as the peas that Mendel studied and most multicellular organisms, including humans, two copies of each gene are present in each somatic (body) cell. Such cells are said to be diploid. During sexual reproduction, specialized cells, known as germ cells, are produced; these cells contain only a single copy of each gene and are referred to as haploid, although monoploid might be a better term. Two such haploid cells, known as gametes, fuse to form a new

---

[274] Weldon, W.F.R. (1902). Mendel's laws of alternative inheritance in peas. *Biometrika*, *1*, 228–254..

[275] Actually more complex that we can address here: see An expanded view of complex traits: from polygenic to omnigenic.

[276] Apomixis in hawkweed: Mendel's experimental nemesis: link

[277] Rediscovery of Mendel's work: link

[278] The Co-evolution of Genes and Culture: link

diploid organism. While gametes can be morphologically identical, in animals and plants, they are generally quite different in size and shape. The gametes of animals are known as sperm and egg, while in plants they are known as pollen and ovule. Generally an individual sexually reproducing organism produces only a single type of gamete, with the organism producing the morphologically larger gametes known as the female and the organism producing the smaller gametes are known as male. As we discussed earlier (Chapter 4), this difference in size has evolutionary (selective) implications. In any particular organism there are thousands of genes and within a population there are typically a number of different alleles.[279] An important feature of sexual reproduction is that the new organism carries a unique combination of alleles inherited from its two parents. This increases the genetic variation within the population, which enables the population, as opposed to specific individuals, to deal with a range of environmental factors, including pathogens, predators, prey, and competitors. It leaves unresolved, however, exactly how genetic information is replicated, how new alleles form, and how information is encoded, regulated, and utilized at the molecular, cellular, and organismic levels.

Question to answer

113. Develop a plausible explanation for why adult lactose tolerance is not a universal trait of mammals?

Discovering how nucleic acids store genetic information

To follow the historical pathway that led to our understanding of how heredity works, we have to start back at the cell, the basic living unit. As it became firmly established that all organisms are composed of one or more cells, and that all cells were derived from pre-existing cells, it became more and more likely that inheritance had to be a cellular phenomenon. As part of their studies, cytologists (students of the cell) began to catalog the common components of cells; because of resolution limits associated with available microscopes, these studies were restricted to larger eukaryotic cells. One such component of eukaryotic cells is the nucleus. At this point it is worth remembering that most cells do not contain pigments. Under these early (bright-field) microscopes, they appear clear and transparent, after all they are ~70% water. To discern structural details cytologists had to stabilize the cell. As you might suspect, stabilizing the cell means killing it. Biological samples were killed (known technically as “fixed”) in such a way as to insure that their structure was preserved as close to the living state as possible. Originally, this process involved the use of chemicals, such as formaldehyde or organic solvents that could cross-link or precipitate various molecules together. Fixation stops molecules from moving with respect to one another; it is not unlike boiling an egg. As long as the methods used to view the fixed tissue were of low magnification and resolution, the results obtained using such methods were acceptable. In more modern studies, using higher resolution optical methods[280] and electron microscopes, such crude fixation methods have been replaced by various alternatives, including various forms of cryo-electron microscopy. Even so it can be hard to resolve the different subcomponents of the cell. One approach was to treat fixed cells with various dyes. Some dyes bind preferentially to molecules located within particular parts of the cell. The most dramatic of these cellular sub-regions was the nucleus, which due to its bulk chemical composition, was stained very differently from the surrounding cytoplasm. One common stain consists of a mixture of hematoxylin (actually oxidized hematoxylin and aluminum ions) and eosin; it leaves the cytoplasm pink and the nucleus dark blue.[281] The nucleus was first described by Robert Brown (1773-1858), the person after which Brownian motion was named. The presence of a nucleus was characteristic of eukaryotic (true

[279] You can get an idea of the alleles present in the human population by using the gnomAD browser: link

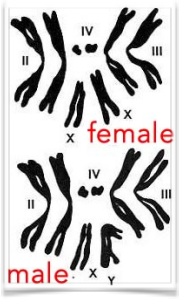
[280] Optical microscopy beyond the diffraction limit: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2645564/

[281] The long history of hematoxylin: http://www.ncbi.nlm.nih.gov/pubmed/16195172

nucleus) organisms.[282] Prokaryotic cells (before a nucleus) are typically much smaller and originally it was technically impossible to determine whether they had a nucleus or not – they do not.

The careful examination of fixed and living cells revealed that the nucleus undergoes a dramatic reorganization during the process of cell division; it loses its roughly spherical shape, which was replaced by discrete stained strands, known as chromosomes (colored bodies). In 1887 Edouard van Beneden (1846-1910) reported that the number of chromosomes in a somatic (diploid) cell was constant for each species and that different species had different numbers of chromosomes (←). Within a particular species the individual chromosomes could be recognized based on their distinctive sizes and shapes. For example, in the somatic cells of the fruit fly *Drosophila melanogaster* there are two copies of each of 4 chromosomes (→). In 1902, Walter Sutton (1877-1916) published his observation that chromosomes obey Mendel's rules of inheritance, that is that during the formation of the cells (gametes) that fuse during sexual reproduction, each cell received one and only one copy of each chromosome. This strongly suggested that Mendel's genetic factors were associated with chromosomes.[283] By this time, it was recognized that there were many more Mendelian factors than chromosomes, which implied that many factors must be present on each chromosome. These observations provided a physical explanation for the observation that many genetic traits did not behave independently but acted as if they were somehow linked together. The behavior of the nucleus, and the chromosomes that appeared to exist within it, mimicked the type of behavior that a genetic material would be expected to display.

species	chromosome #
Ophioglossum reticulatum (a fern)	1260 (630 pairs)
Canis familiaris (dog)	78 (39 pairs)
Cavia cobaya (guinea pig)	60 (30 pairs)
Solanum tuberosom (potato)	48 (24 pairs)
Homo sapiens (humans)	46 (23 pairs)
Macaca mulatta (monkey)	42 (21 pairs)
Mus musculus (mouse)	40 (20 pairs)
Felis domesticus (house cat)	38 (19 pairs)
Saccharamyces cervisae (yeast)	32 (16 pairs)
Drosophila melanogaster (fruit fly)	8 (4 pairs)
Myrmecia pilosula (ant)	2 (1 pair)



Cellular anatomy studies were followed by studies on the composition of the nucleus. As with many scientific studies, progress is often made when one has the right “model system” to work with. It turns out that some of the best systems for the isolation and analysis of the components of the nucleus were sperm and pus, isolated from discarded bandages from infected wounds (yuck). It was therefore assumed, quite reasonably, that components enriched in this material would likely be enriched in nuclear (genetic information containing) components. Using sperm and pus as starting materials Friedrich Miescher (1844-1895) was the first to isolate a phosphorus-rich compound, called nuclein.[284] At the time of its isolation there was no evidence linking nuclein to genetic inheritance. Later nuclein was resolved into an acidic component, deoxyribonucleic acid (DNA), and a basic component, primarily proteins known as histones. Because they have different properties (acidic DNA, basic histones), chemical “stains” that bind or react with specific types of molecules and absorb visible light, could be used to visualize the location of these molecules within cells using a light microscope. The nucleus stained for both highly acidic and basic components - which suggested that both nucleic acids and histones were localized to the nucleus, although what they were doing there was unclear.

Questions to answer

114. How was the nucleus first visualized? What was needed to see it?
115. Is there a correlation between the number of chromosomes and the complexity of an organism. Does chromosome number tell you anything useful about genes?

---

[282] There are some eukaryotic cells, like human red blood cells, that do not have a nucleus, they are unable to divide.

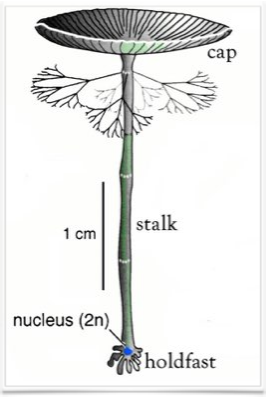
[283] http://www.nature.com/scitable/topicpage/developing-the-chromosome-theory-164

[284] Friedrich Miescher and the discovery of DNA: http://www.sciencedirect.com/science/article/pii/S0012160604008231

Questions to ponder

- How would you define a model system? What is it that makes model systems useful?
- In comparing organisms, what does complexity mean?

Locating hereditary material within the cell



Further evidence suggesting that hereditary information was localized in the nucleus emerged from transplantation experiments carried out in the 1930's by Joachim Hammerling (1901-1980). He used the giant unicellular green alga *Acetabularia acetabulum*, known as the mermaid's wineglass (→). Hammerling's experiments (video link) illustrate two important themes in the biological sciences. The idiosyncrasies of specific organisms can be exploited to carry out useful studies that are simply impossible, difficult, or prohibitively expensive to perform elsewhere. At the same time, the underlying evolutionary homology of organisms makes it possible to draw broadly relevant conclusions from studies on a particular organism, something unlikely to be true if each represented a unique creation event. That said, there are dangers in thinking that complex human traits (such as autism and pathogenic processes) can be studied is evolutionary distinct organisms.[285]

Hammerling exploited three unique features of *Acetabularia*. The first is the fact that each individual is a single cell, with a single nucleus. Through microdissection, it is possible to isolate nuclear and anucleate (without a nucleus) regions of the organism. Second, these cells are very large (1 to 10 cm in height), which makes it possible to remove and transplant regions of one organism (cell) to another. Finally, different species of *Acetabularia* have morphologically distinct "caps" that regrow faithfully following amputation. In his experiments, he removed the head and stalk regions from one individual, leaving a small "holdfast" region that contained the nucleus. He then transplanted large regions of a anuclear stalk, derived from an individual of a different species with a distinctively different cap morphology, onto the smaller nucleus-containing holdfast region. When the cap regrew it had the morphology characteristic of the species that provided the nucleus - no matter that this region was much smaller than the transplanted, anucleate stalk region. The conclusion was that the information needed to determine the cap's morphology was located within the region of the cell that contained the nucleus, rather than dispersed throughout the cytoplasm. It was a short step from these experimental results to the conjecture that all genetic information is located within the nucleus.

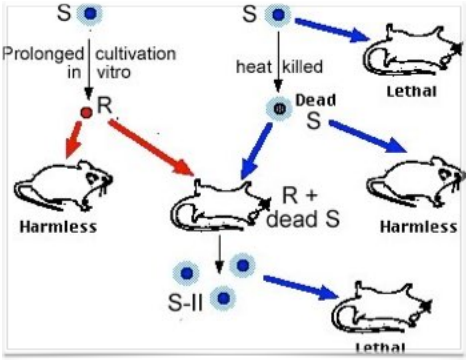
Identifying DNA as the genetic material

The exact location, and the molecular level mechanisms behind the storage and transmission of genetic information, still needed to be determined. Two kinds of experiment led to the realization that genetic information was stored in some chemically stable form. In his studies, H.J. Muller (1890-1967) found that exposing fruit flies to X-rays, a highly energetic form of light, generated a genetic change (a mutation) that could be passed from one generation to the next. Based on this result one conclusion was that genetic information was stored in a chemical form and that that information could be altered through interactions with radiation, which presumably led to a chemical alteration of the molecule(s) storing the information. Moreover, once altered, the information was again stable.

The second piece of experimental evidence supporting the idea that genetic information was encoded in a stable chemical form came from a series of experiments initiated in the 1920s by Fred Griffith (1879-1941). He was studying strains of the bacterium *Streptococcus pneumoniae* that

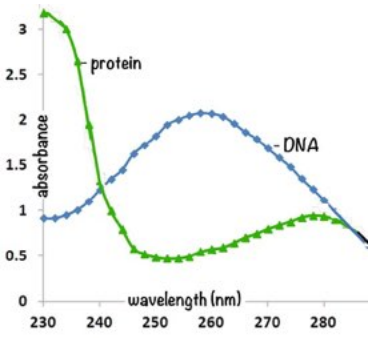
[285] Mice fall short as test subjects - McGlinn 2013 & False analogies & logical fallacies in animal models - Sjoberg 2016

cause bacterial pneumonia. When these bacteria were introduced into mice, the mice got sick and died. Griffith grew these bacteria in the laboratory. Such bacteria are said to be cultured *in vitro* or in glass (although in modern labs they are often grown in plastic), as opposed to growing *in vivo* or within a living animal. Following common methods, he grew the bacteria on plates covered with solidified agar (a jello-like substance derived from sea weed) containing various nutrients. Typically, a liquid culture of bacteria is diluted and spread on the agar surface of the plate. When sufficiently diluted, isolated individual bacteria, separated from one another, come to rest on the agar surface. Bacteria are asexual and so each bacterium can grow up into a colony, a clone of the original bacterium that landed on the plate. The disease-causing strain of *S. pneumoniae* grew up into smooth or S-type colonies, due to the slimy mucus-like substance they secreted. Griffith found that mice injected with S strain *S. pneumoniae* quickly sickened and died. However, if he killed the bacteria with heat before injection the mice did not get sick (→), indicating that it was the living bacteria that produced (or evoked) the disease symptoms rather than some heat-stable chemical toxin.



During extended *in vitro* cultivation the S strain bacteria sometimes gave rise to rough (R) colonies. R colonies are rough rather than smooth and shiny. This appeared to be a genetic change since once isolated, R-type strains produced R-type colonies. More importantly, mice injected with R strain *S. pneumoniae* did not get sick. A confusing complexity emerged however; mice co-injected with the living R strain, which did not get sick, and dead S strain, which also did not get sick, got sick and died! Griffith was able to isolate and culture *S. pneumoniae* from these dying mice and found that, when grown *in vitro*, they produced smooth colonies. He termed these S-II (smooth) strains. His hypothesis was that a stable (that is, non-living) chemical component derived from the dead S bacteria had "transformed" the avirulent (benign) R strain bacteria to produce the new virulent S-II strains.[286] Unfortunately Fred Griffith died in 1941 during the Nazi-bombing of London, which put an abrupt end to his studies.[287]

In 1944 Griffith's studies were continued and extended by Oswald Avery (1877-1955), Colin McLeod (1909-1972), and Maclyn McCarty (1911-2005). They set out to use Griffith's assay to isolate what they termed the “transforming principle” responsible for turning R into S strains. Their approach was to grow up large numbers of cells *in vitro* and to then grind them up and isolate their various components, their proteins, nucleic acids, carbohydrates, and lipids. They then digested these extracts with various enzymes that acted to degrade specific types of molecules and determine whether the transforming principle remained intact. Treating cellular extracts with proteases (that degrade proteins), lipases (that degrade lipids), or RNAases (that degrade RNAs) had no effect on the transforming principle. In contrast, treatment of the extracts with DNAases, enzymes that degrade DNA, destroyed the extracts transforming activity. Further support for the idea that the “transforming substance” was DNA was suggested by the fact that purified transforming substance had the physical properties of DNA; for example it absorbed light like DNA rather than protein (absorption spectra of DNA versus protein →). Subsequent studies confirmed this conclusion. Furthermore DNA isolated from R strain bacteria was not able to produce S-II strains from R strain bacteria, whereas DNA from S strain bacteria could. They concluded that DNA derived from S cells contains the information required for the conversion – it is,



[286] link: Griffith's experiment

[287] And provides yet another good reason (as if we need more) to hold Nazis (and neo-Nazis) in contempt.

or rather contains, a gene required for the S strain phenotype. This information had, presumably, been lost by mutation during the formation of R strains.

The basic phenomena exploited by Griffiths and Avery et al., known as transformation, is an example of horizontal gene transfer, which is discussed in greater detail later on. It is the movement of genetic information from one organism to another. This is a distinctly different process than the movement of genetic information from a parent to an off-spring, which is known as vertical gene transfer. Horizontal gene transfer can occur between unrelated organisms and does not involve cell fusion. Various forms of horizontal gene transfer occur within the microbial world and allow genetic information to move between species. For example horizontal gene transfer is responsible for the rapid expansion of populations of antibiotic-resistant bacteria. Viruses are responsible for a highly specialized form of horizontal gene transfer, known as transduction.[288] An obvious question then is, how is this possible? While we might readily accept that genetic information must be transferred from parent to offspring (we see the evidence for this process with our own eyes in the form of family resemblances), the idea that genetic information can be transferred between different organisms that are not (apparently) related to one another is quite a bit more difficult to swallow. As we will see, horizontal gene transfer is possible primarily because all organisms share the same basic system for encoding, reading, using, and replicating genetic information. The hereditary machinery is homologous among existing organisms.

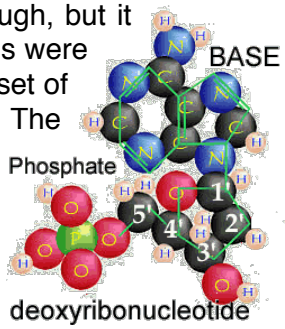
Questions

116. How would Hammerling's observations have been different if hereditary information was localized in the cytoplasm?
117. In Griffith's study, he found that dead smooth *S. pneumoniae* could transform living rough strains of *S. pneumoniae* when co-injected into a mouse. Would you expect that DNA from an unrelated species of bacteria give the same result? Explain your reasoning.
118. What caused the change from S to R strains in culture? Why is DNA from the R strain unable to produce S-II cells?
119. In the spectrometric analysis of DNA and protein, what is plotted on the X- and Y-axes?

Questions to ponder

- What is the difference between a strain and a species?
- How might horizontal gene transfer confuse molecular phylogenies (family trees)?
- How might a creationist explain horizontal gene transfer?

Unraveling Nucleic Acid Structure



Knowing that the genetic material was DNA was a tremendous break through, but it left a mystery - how was genetic information stored and replicated. Nucleic acids were thought of as boring aperiodic polymers, that is, molecules built from a defined set of subunits, known as monomers, but without a simple overall repeating pattern. The basic monomeric units of nucleic acids are known as nucleotides (→). A nucleotide consists of three distinct types of molecules joined together, a five-carbon sugar (ribose or deoxyribose), a nitrogen-rich "base" that is either a purine (guanine (G) or adenine (A)) or a pyrimidine (cytosine (C), or thymine (T)) in DNA or uracil (U) instead of T in RNA, and a phosphate group. The carbon atoms of the sugar are numbered 1' to 5'. The nitrogenous base is attached to the 1' carbon and the phosphate is attached to the 5' carbon. The other functionally important group is a hydroxyl group attached to the 3' carbon of the ribose/deoxyribose moiety.[289] RNA differs from DNA in that there is a hydroxyl group attached to the 2' carbon of the ribose, this

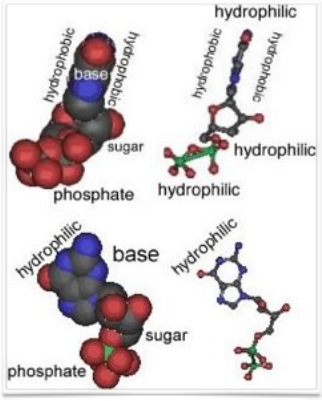
[288] link:: Virus-like particles speed bacterial evolution

[289] "Moiety" defined

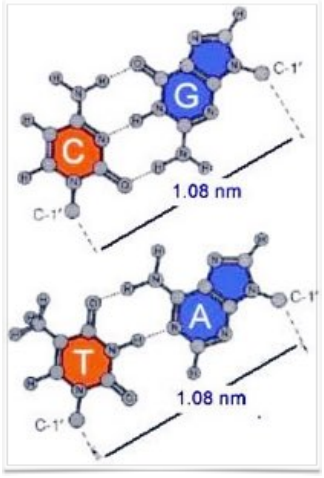
hydroxyl is absent in DNA, which is why it is “deoxy” ribonucleic acid! We take particular note of the 5’ phosphate and 3’ hydroxyl groups of the ribose/deoxyribose because they are directly involved in the linkage of nucleotide monomers together to form nucleic acid polymers.

Discovering the structure of DNA

A critical clue to understanding the structure of nucleic acids came from the work of Erwin Chargaff (1905-2002). When analyzing DNA from various sources, he found that the relative amounts of G, C, T and A nucleotides present varied between organisms but were the same (or very similar) for organisms of the same type or species. On the other hand, the ratios of A to T and of G to C were always equal to 1, no matter where the DNA came from. Knowing these rules, James Watson (1928-) and Francis Crick (1916-2004) built a model of DNA that fit what was known about the structure of nucleotides and structural data from Rosalind Franklin (1920-1958). Franklin got her data by pulling DNA molecules into oriented strands; fibers of many molecules aligned parallel to one another. By passing a beam of X-rays through these fibers she was able to obtain a diffraction pattern; a pattern that defines key parameters that constrain any model of the molecule’s structure.[290] By making a model that was predicted to produce the observed X-ray data, Watson and Crick drew a number of conclusions about the structure of a DNA molecule.[291]



To understand their process, let us consider the chemical nature of a nucleotide and a nucleotide polymer (a nucleic acid) such as DNA. First the nucleotide bases in DNA (A, G, C and T) have a number of similar properties. Each nucleotide (→) has three hydrophilic regions: the negatively charged phosphate group, a sugar which has a number of O–H groups, and the bases' hydrophilic edge, where the N–H and N groups lie. While the phosphate and sugar are three-dimensional moieties, the bases are flat, the atoms in the rings are all in one plane. The upper and lower surfaces of the rings are hydrophobic (non-polar) while the edges have groups that can interact via hydrogen bonds. This means that the amphipathic factors that favor the assembly of lipids into bilayer membranes are also at play in nucleic acid structure. In their model Watson and Crick had the bases stacked on top of one another, hydrophobic surface next to hydrophobic surface, to reduce their interactions with water.

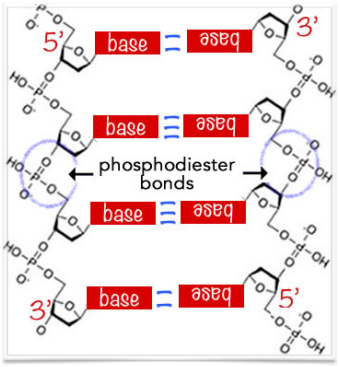


This left each base’s hydrophilic edge, with –C=O and –N-H groups that can act as H-bond acceptors and donors, to be dealt with. How were these hydrophilic groups arranged? With the two polynucleotide strands arranged in opposite orientations, that is, anti-parallel to one another: one from 5’ → 3’ and the other 3’ ← 5’; the bases attached to the sugar-phosphate backbone could interact with one another in a highly specific way (←). An A can form two hydrogen bonding interactions with a T on the opposite (anti-parallel) strand, while a G could form three hydrogen bonding interactions with a C. A key feature of this arrangement is that the lengths of the A::T and G:::C base pairs are almost identical. The hydrophobic surfaces of the bases are stacked on top of each other, while the hydrophilic sugar and phosphate groups are in contact with the surrounding aqueous solution. The repulsion between negatively charged phosphate groups is neutralized (or shielded) by the presence of positively charged ions present in the solution from which the X-ray measurements

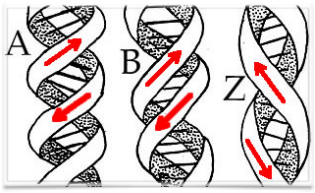
[290] Fiber diffraction

[291] An interesting depiction of this process is provided by the movie “Life Story”

were made. This model also provided a direct explanation for why Chargaff's rules were universal in double stranded DNA.



Each DNA polymer strand has a directionality to it, it runs from the 5' phosphate group of the ribose/deoxyribose at one end to the 3' hydroxyl group of the ribose/deoxyribose at the other end. Each nucleotide monomer is connected to the next through a phosphodiester linkage (→) involving its 5' phosphate group attached to the 3' hydroxyl of the existing strand. In their final model Watson and Crick depicted what is now known as B-form DNA. This is the usual form of DNA in a cell. Under different salt conditions, however, DNA can form two other double helical forms, known as A and Z. While the A and B forms of DNA are "right-handed" helices, the Z-form of DNA is a left-handed helix (←). We will not concern ourselves with these other forms of DNA, leaving that to more advanced courses, but you can imagine that they might well influence the types of intermolecular interactions that occur between DNA and other molecules, particularly proteins.



As soon as the Watson-Crick model of DNA structure was proposed its explanatory power was obvious. Because the A::T and G:::C base pairs are of the same length, the sequence of bases along the length of a DNA molecule (written, by convention in the 5' to 3' direction) has little effect on the overall three-dimensional structure of the molecule. That implies that essentially any sequence can be found, at least theoretically, in a DNA molecule. If information were encoded in the sequence of nucleotides along a DNA strand, information could be placed there and that information would be as stable as the DNA molecule itself. This is similar to the storage of information in various modern computer memory devices, that is, any type of information can be stored, because storage does not involve any dramatic change in the basic structure of the storage material. The structure of a flash memory drive is not dramatically different whether in contains photos of your friends, a song, a video, or a textbook. What matters is how the information is "encoded", most obviously in the specific sequence of nucleotides along a strand.

At the same time, the double-stranded nature of the DNA molecule's structure and the complementary nature of base pairing (A to T and G to C) suggested a simple model for DNA (and information) replication - that is, pull the two strands of the molecule apart and build new (anti-parallel) strands using the two original strands as templates. This model of DNA replication is facilitated by the fact that the two strands of the parental DNA molecule are held together by weak hydrogen bonding interactions; no covalent bonds are broken when the strands are separated from one another. In fact, at physiological temperatures DNA molecules often open up over short stretches and then close again, a process known as DNA breathing.[292] This makes the replication of the information stored in the molecule conceptually straightforward, even though the actual biochemical process is complex, in part because of the importance of accurate replication. The existing strands determine the sequence of nucleotides on the newly synthesized strands. The newly synthesized strand can, in turn, direct the synthesis of a second strand, identical to the original strand. Finally, the double-stranded nature of the DNA molecule means that any information within the molecule is, in fact, stored in a redundant fashion. If one strand is damaged, that is its DNA sequence is lost or altered, the second undamaged strand can be used to repair that damage. A number of mutations in DNA are repaired using this type of mechanism (see below).

Questions to answer

120. How is a DNA molecule structurally analogous to a lipid bilayer? Draw a diagram that reveals the similarities and note the most important differences?
121. Which do you think is stronger (and why), an AT or a GC base pair?
122. Why is the ratio of A to T the same in all organisms?

---

[292] Dynamic approach to DNA breathing: http://www.ncbi.nlm.nih.gov/pubmed/23345902

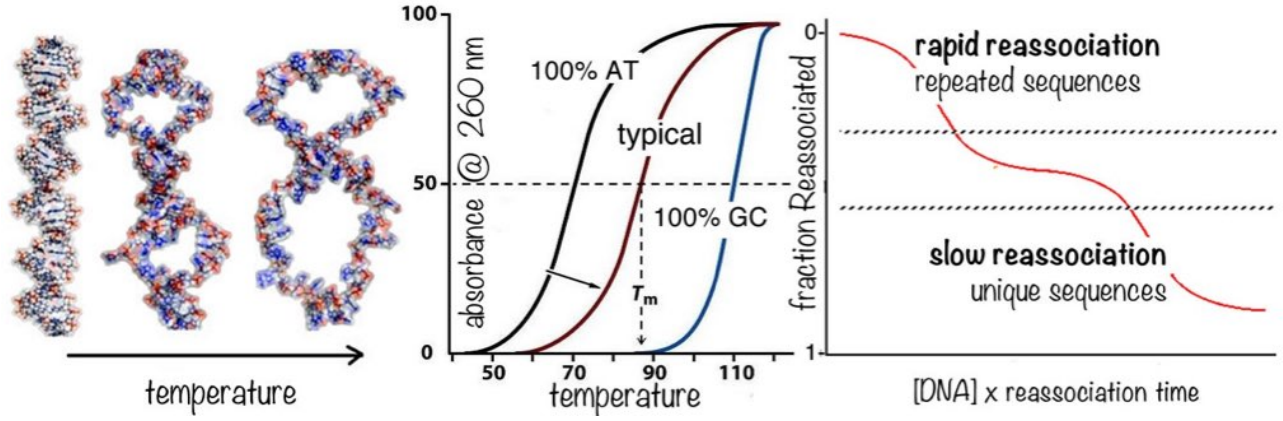
123. Normally DNA exists inside of cells at physiological salt concentration (~140 mM KCl, 10 mM NaCl, 1 mM $MgCl_2$ and some minor ions). Predict what might happen if you placed DNA into pure water.
124. How many general types of mutation can you think of? How would they differ in their impact on the information encoded in a DNA molecule.
125. Generate a model mechanism by which a DNA molecule could be accurately repaired, that is, without the loss of the information originally present within it.

Questions to ponder

- Why does the ratio of A to G differ between organisms?
- You isolated DNA from an organism, and you find it fails to obey Chargaff's rule; what might you predict about the structure of its DNA?

DNA: sequence & information

We can now assume that somehow the sequence of nucleotides in a DNA molecule encodes information but exactly what kinds of information are stored in DNA? Early students of DNA could not read DNA sequences as we can now, so they relied on various measurements to better understand the behavior of DNA molecules. For example, the way a double stranded DNA molecule interacts with light is different from how a single stranded DNA molecule interacts with light. Since the two strands of double stranded DNA molecules, often written dsDNA, are linked only by hydrogen bonds, increasing the temperature of the system will lead to their separation into two single stranded molecules (ssDNA)(left panel ↓). ssDNA absorbs light at 260nm (in the ultraviolet range) more strongly than does dsDNA, so the absorbance of a DNA solution can be used to determine the relative amounts of single and double stranded DNA in a sample. What we find is that



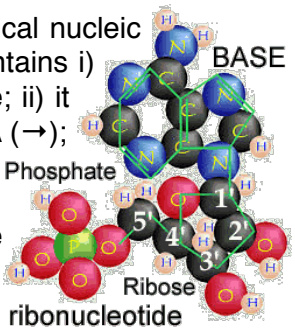
the temperature at which 50% of dsDNA molecules have separated into ssDNA molecules varies between organisms. This is not particularly surprising given Chargaff's observation that the ratio of AT to GC varies between organisms and the fact that GC base pairs, mediated by three H-bonds, are more stable (take more energy to separate) than AT base pairs, which are held together by only two H-bonds. In fact, one can estimate the AT:GC ratio of a DNA molecule based on melting curves (middle pane ↑).

It quickly became clear that things were more complex than previously expected. Here a technical point needs to be introduced. Because of the extreme length of the DNA molecules found in biological systems, it is almost impossible to isolate such molecules intact. In the course of their purification, the molecules are sheared (break) into shorter pieces, typically thousands to tens of thousands of base pairs in length compared to the millions to hundreds of millions of base pairs in intact molecules. In another type of experiment, one can look at how fast ssDNAs (the result of a melting experiment) reform dsDNA. The speed of these "reannealing reactions" depends on DNA concentration. When such experiments were carried out, it was found that there was a fast annealing population of DNA fragments and various slower annealing populations (right panel ↑).

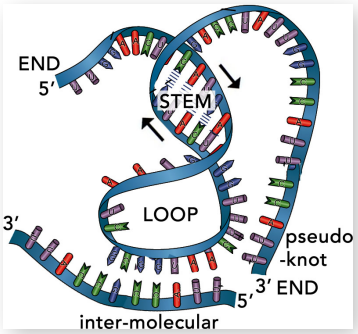
How to explain this observation? Was it a function of AT:GC ratio or was something else going on? Subsequent analyses revealed that it was due to the fact that within the DNA isolated from many organisms, particularly eukaryotes, there were many (hundreds to thousands) of molecular regions that contained very similar nucleotide sequences. Because the single strands of these fragments can associate with one another, these sequences occurred in much higher effective concentrations compared to regions of the DNA with unique sequences. This type of analysis revealed that much of the genome of eukaryotes is composed of various families of repeated sequences and that regions of unique sequence amount to less than ~5% of the total genomic DNA. While a complete discussion of these repeated sequence elements is beyond our scope here, we can make a few points. As we will see, there are mechanisms that can move regions of a DNA molecule from one position to another within the genome, or that can generate a copy of a DNA sequence and insert it into another position of the genome (leaving the original sequence behind). The end result is that the genome (the DNA molecules) of a cell/organism is dynamic, a fact with profound evolutionary implications.

Discovering RNA: structure and some functions

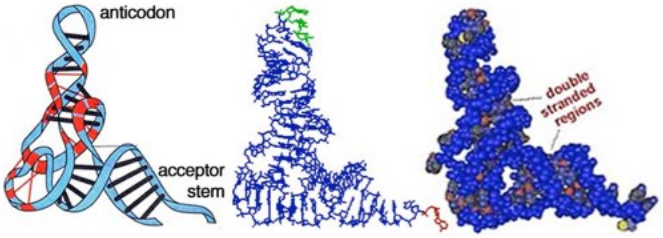
DNA is not the only nucleic acid found in cells. A second class of biological nucleic acid is known as ribonucleic acid (RNA.) RNA differs from DNA in that it contains i) the sugar ribose (with a hydroxyl group on the 2' C) rather than deoxyribose; ii) it contains the pyrimidine uracil instead of the pyrimidine thymine found in DNA (→); and iii) RNA is typically single rather than double stranded.[293] Nevertheless, RNA molecules can associate with an ssDNA molecule with a complementary nucleotide sequence. Instead of the A-T pairing in DNA we find A pairing with U instead. This change does not make any significant difference when the RNA strand interacts with DNA, since the number of hydrogen bonding interactions are the same.



When RNA is isolated from cells, the major population was found to reassociate with unique sequences within the DNA. As we will see later, this class of RNA includes molecules, known as messenger or mRNAs, that carry information from DNA to the molecular machinery that mediates the synthesis of proteins (the ribosome). In addition to mRNAs there are a number of other types of RNAs in cells; in each case, their synthesis is directed by DNA-dependent RNA polymerases. These non-mRNAs include structural, catalytic, and regulatory RNAs. As you may already suspect, the same hydrophobic/hydrophilic/H-bond considerations that were relevant to DNA structure apply to RNA structure, but because RNA is generally single stranded, the structures found in RNA are different and more varied. A single-stranded RNA molecule can fold back on itself, through intra-molecular interactions, to create local double stranded regions (→). Similarly distinct RNA molecules can interact through double-stranded regions (inter-molecular interactions). In both cases, and just as in DNA, these strands are anti-parallel to one another. This results in double-stranded regions ("stems") that end in single-stranded "loops" (or molecular ends). Regions within a stem, that can be as short as 1 base pair, that do not base pair will "bulge out". The end result is that RNA molecules can adopt a wide range of complex three-dimensional structures in solution.



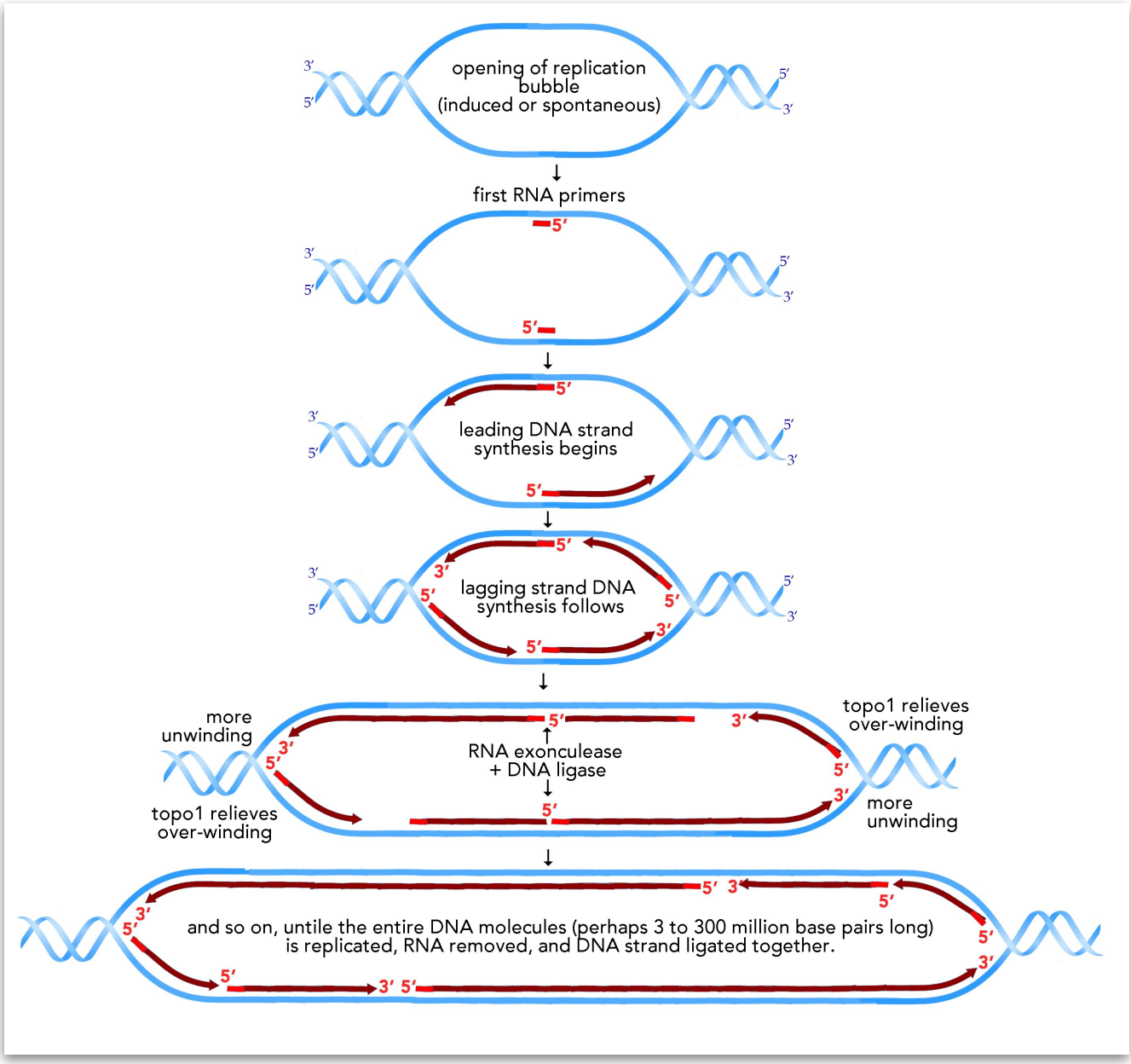
Transfer RNAs (tRNAs)(→), an integral component of the protein synthesis system, are



[293] The exception involves viruses, where double stranded RNA is found as the genetic material

one well studied example of how intermolecular interactions within an RNA molecule can produce complex three-dimensional shapes that carry out specific molecular functions (described in greater detail in the next chapter).

In addition to intra- and inter-molecular interactions involving RNA molecules, RNAs can also interact with proteins to form "riboprotein" complexes. For example, the CRISPR-Cas9 system involves a double-stranded DNA endonuclease (an enzyme that generates the cleavage of both strands of a double-stranded DNA molecule) that is directed to specific DNA sequences through an associated RNA molecule, known as a guide RNA. Other RNA-protein complexes are involved in the control of RNA synthesis and stability, among a number of other functions. The classic example of a riboprotein complex is the ribosome itself, a macromolecular machine that mediates the synthesis of polypeptides. A ribosome is composed of structural and catalytic RNAs (known as ribosomal or rRNAs) and ~50 to 80 proteins (polypeptides), depending upon whether you are prokaryotic or eukaryotic; altogether it has a molecular weight of ~3.2 x 10^6 daltons.



The ability of RNA to both encode information in its base sequence and to mediate catalysis through its three dimensional structure has led to the "RNA world" hypothesis that proposes that early in the evolution of life various proto-organisms relied on RNAs, or more likely simpler RNA-like

molecules, rather than DNA and proteins, to store genetic information and to catalyze at least a subset of metabolic reactions. Some modern day viruses use single or double-stranded RNAs as their genetic material. According to the RNA world hypothesis, it was only later in the history of life that organisms developed the more specialized DNA-based systems for genetic information storage and proteins for most catalytic and structural functions. While this idea is compelling, there is no reason to believe that simple polypeptides and other molecules were not also present and playing a critical role in the early stages of life's origins. At the same time, there are many unsolved issues associated with a simplistic RNA world view, the most important being the complexity of RNA itself, its abiogenic (that is, without life) synthesis, and the survival of nucleotide triphosphates in solution. Nevertheless, it is clear that catalytic and regulatory RNAs play a key role in modern cells and throughout their evolution. The catalytic activity of the ubiquitous ribosome, which is involved in protein synthesis in all known organisms, is based on a ribozyme, a RNA-based catalyst.

Questions to answer:

126. How would you calculate the probability that two DNA sequences (of length N) are identical by chance?
127. Predict how the annealing curve of genomic DNA changes as the number of repeated sequences increases.
128. Propose a plausible model for how a single-stranded RNA molecule could act as a catalyst; consider why double-stranded DNA is unlikely to act catalytically.

Question to ponder:

- What are the possible functions for the unique and repeated sequences of DNA in a genome.

DNA replication

Once it was proposed, the double-helical structure of DNA immediately suggested a simple mechanism for the accurate duplication of the information stored in DNA. Each strand contains all of the information necessary to specify the sequence of the complementary strand. The process begins when a dsDNA molecule opens (next ↓ page) to produce two single-stranded regions. Where DNA is naked, that is, not associated with other molecules (proteins), the opening of the two strands can occur easily, since the two strands are held together only by weak H-bonding interactions. Normally, the single strands simply reassociate with one another. To replicate DNA the open region has to be stabilized and the catalytic machinery involved recruited and organized. We will consider how this is done in general terms, in practice this is a complex and highly regulated process involving a number of components.

The first two issues we have to address in the context of DNA replication may seem arbitrary, but they turn out to be common (conserved) features of DNA synthesis. The enzymes (DNA-dependent, DNA polymerases) that catalyze the synthesis of new DNA strands cannot start the synthesis of a new polynucleotide strand on their own, they must add nucleotides onto the end of a pre-existing nucleic acid polymer, they depend on a "polynucleotide primer". In contrast, the catalysts that synthesize RNA (DNA-dependent, RNA polymerases) do not require a pre-existing nucleic acid strand, they can start the synthesis of a new RNA strand, based on complementary DNA sequence, *de novo*, that is without a polynucleotide primer. Both DNA and RNA polymerases link the 5' end of a nucleotide triphosphate molecule to the pre-existing 3' end of a nucleic acid molecule; the polymerization reaction is said to proceed in the 5' to 3' direction, nucleotides are added sequentially to the 3' end. As we will see later on, the molecules involved in DNA replication and RNA synthesis rely on signals within the DNA that are recognized by proteins; together these determine where and when nucleic acid replication occurs and where synthesis starts and stops. For now let us assume that some process has determined where DNA replication starts.

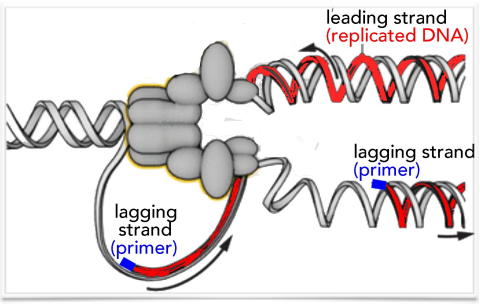
After the dsDNA molecule has locally "opened" (←), a specialized DNA-dependent, RNA polymerase, known as primase collides with, binds to, and synthesizes a short RNA molecule, known as a primer. Because the two strands of the DNA molecule point in opposite directions (they are anti-parallel), one primase complex associates with each of the now separated DNA strands; two

RNA primers are generated, one on each strand. Once these RNA primers are in place, DNA-dependent, DNA polymerases replace the primase enzymes and begin to catalyze the deoxynucleotide-addition reaction; which nucleotide is added is determined by which nucleotide is present next in the existing DNA strand. The nucleotide addition reaction involves various nucleotides colliding with the DNA-primer-polymerase complex; only the appropriate nucleotide, complementary to the nucleotide residue in the existing DNA strand is bound and used in the reaction.

Nucleotides exist in various phosphorylated forms within the cell, including nucleotide monophosphate (NMP), diphosphate (NDP), and triphosphate (NTP) forms. To make the nucleic acid polymerization reaction thermodynamically favorable, the reaction uses the NTP form of the nucleotide monomers, generated through the reaction:

$$(5'P)NTP(3'OH) + (5'P)NTP(3'OH) + H_20 \longleftrightarrow (5'P)NTP\text{-}NMP(3'OH) + diphosphate.$$

During the reaction the terminal diphosphate of the incoming NTP is released (a thermodynamically favorable reaction) and a nucleotide mono-phosphate is added to the existing polymer through the formation of a phosphodiester [-C-O-P-O-C] bond. This reaction creates a new 3' OH end for the polymer that can, in turn, react with another NTP. In theory, this process can continue until the newly synthesized strand reaches the end of the DNA molecule. The strand synthesized from the original primer is known as the "leading" strand. For the process to continue, however, the double stranded region of the original DNA will have to open up further, exposing (generating) more single-stranded DNA. Keep in mind that this process is moving, through independent complexes, in both directions along a DNA molecule. Because the polymerization reaction only proceeds by 3' addition, as new single stranded regions are opened (→) new primers must be created by RNA primase and then extended by DNA polymerase; these are known as the lagging strands. While there are two leading strands leaving a particular DNA replication start site, there are a number of lagging strands involved.



If you try drawing what this looks like, you will realize that i) this process is asymmetric in relation to the start site of replication; ii) the process generates RNA-DNA hybrid molecules; and iii) that eventually an extending DNA polymerase will run into the RNA primer part of an "upstream" molecule. However, keep in mind, RNA regions, derived from the primers, are not found in "mature" DNA molecules, so there must be a mechanism that removes them. As it turns out, the DNA polymerase complex, like a number of other enzyme systems, contains more than one catalytic activity (analogous to the ATP synthase and pump). When the DNA polymerase complex reaches the upstream nucleic acid chain it runs into an RNA containing region; an RNA exonuclease activity associated with the DNA polymerase complex removes the RNA nucleotides and replaces them with DNA nucleotides using the existing DNA strand as the primer. Once the RNA portion is removed, a DNA ligase acts to join (generate a covalent phosphodiester bond between) the two DNA molecules. These reactions, driven by nucleotide hydrolysis, end up producing a continuous DNA strand that runs from one end of the chromosome to the other, or in circular chromosomes, all the way around the circle.

Evolutionary considerations: At this point you might well ask yourself, why (for heavens sake) is the process of DNA replication so complex. Why not use a DNA polymerase that does not need an RNA primer, or any primer for that matter? That should be possible, particularly given that RNA polymerase does not need a primer. Why not have polymerases that can add nucleotides equally well to either end of a polymer? That such a mechanism is possible is suggested by the presence of enzymes in eukaryotic cells that can catalyze the addition of a nucleotide to the 5' end of an RNA

molecule, the 5' capping reaction associated with mRNA synthesis that we will consider (briefly) later on. But while apparently possible, such activities are not known to be used in DNA replication. The real answer to why DNA replication is as complex as it is is that we are not sure. It could be its complexity is an evolutionary relic, based on a process established within the last common ancestor of all organisms and extremely difficult or impossible to change through evolutionary mechanisms, or simply not worth the effort, in terms of its effects on reproductive success. Alternatively, there could be strong selective advantages associated with the system that preclude such changes. What is clear is that this is how the system appears to function in all known organisms. For practical purposes, we need to remember a few key details, these include the direction of polymer synthesis (3' addition) and the need (in the case of DNA synthesis) for an RNA primer.

Replication machines

We have presented DNA replication (an apparently homologous process used in all known organisms) in as conceptually simple terms as we can, but it is important to keep in mind that the actual machinery involved is complex. In part this complexity arises because the process is topologically constrained and needs to be highly accurate. In the bacterium *Escherichia coli* over 100 genes are involved in DNA replication and repair. To insure that replication is controlled and complete, replication begins at specific sequences along the DNA strand, known as origins of replication or origins for short. Origin DNA sequences are recognized by specific DNA binding proteins. The binding of these proteins initiates the assembly of an origin recognition complex, an ORC. Various proteins then bind to the DNA to locally denature (unwind and separate) and block the single strands from re-annealing. This leads to the formation of what is known as a replication bubble. Multiprotein complexes, known as a replication fork, assemble on the two DNA strands. Using a single replication origin and two replication forks, moving in opposite directions, a rapidly growing *E. coli* cell can replicate its ~4,700,000 base pairs of DNA, which are present in the form of a single circular DNA molecule, in ~40 minutes. Each replication fork moves along the DNA adding ~1000 base pairs of DNA per second to the newly formed DNA polymer. While a discussion of the exact mechanisms involved is beyond our scope here, it is critical that DNA is complete before a cell attempts to divide - this implies that there are signaling systems within the cell that can be used to monitor and coordinate the completion of DNA replication which starts of cell division. We will find such "checkpoint" systems in a number of cellular processes. In many bacteria, the signaling system is based on the fact that the chromosome is circular, that DNA replication begins at a single site (the origin), and that replication forks collide with one another in a region of the chromosome known as the terminus.[294]

Questions to answer

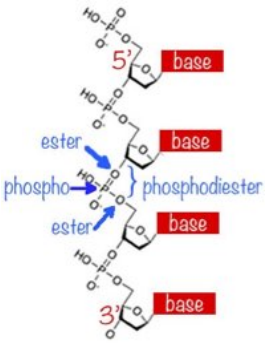
129. Draw a diagram of the key steps in the replication of a circular DNA molecule. How might you adapt this system to replicate much longer linear molecules?
130. What key, non-deducible features of DNA replication do you need to remember (memorize) and why?

Accuracy and error in DNA synthesis

DNA synthesis (replication) is a highly accurate process; the DNA-dependent DNA polymerase makes about one error for every ~10,000 bases it adds. But that level of error would be highly deleterious; in fact most of these errors are quickly recognized as mistakes. To understand how, remember that correct AT and GC base pairs have the same molecular dimensions, that means that incorrect AG, CT, AC, and GT base pairs are either too long or too short. By responding to base pair length, molecular machines can recognize a mistake in base pairing as an abnormal structural feature in the DNA molecule. When a mismatched base pair is formed and recognized, the DNA

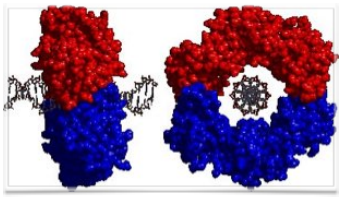
[294] Synchronization of Chromosome Dynamics and Cell Division in Bacteria

polymerase stops forward synthesis, reverses its direction, and removes the region of the DNA containing the mismatched base pair using a “DNA exonuclease” activity. It then resynthesizes the region, (hopefully) correctly. This process is known as proof-reading; the proof-reading activity of the DNA polymerase complex reduces the total DNA synthesis error rate to ~1 error per 1,000,000,000 (10^9) base pairs synthesized.



At this point let us consider nomenclature, which can seem arcane and impossible to understand, but in fact obeys reasonably straightforward rules. An exonuclease is an enzyme that can bind to the free end of a nucleic acid polymer and remove nucleotides through a hydrolysis reaction of the phosphodiester bond (→). A 5' exonuclease cuts off a nucleotide located at the 5' end of the molecule, a 3' exonuclease, cuts off a nucleotide located at the molecule’s 3' end. An intact circular nucleic acid molecule is immune to the effects of an exonuclease. To break the bond between two nucleotides in the interior of a nucleic acid molecule (or in a circular molecule, which has no ends), one needs an endonuclease activity.

As you think about the processes involved, you come to realize that once DNA synthesis begins, it is important that it continues without interruption. But the interactions between nucleic acid chains are based on weak H-bonding interactions, and the enzymes involved in the DNA replication process can be expected to dissociate from the DNA because of the effects of thermal motion, imagine the whole system jiggling and vibrating – held together by relatively weak interactions. We can characterize how well a DNA polymerase molecule remains productively associated with a DNA molecule in terms of the number of nucleotides it adds to a new molecule before it falls off; this is known as its “processivity”. So if you think of the DNA replication complex as a molecular machine, you can design ways to insure that the replication complex has high processivity, basically by keeping it associated with the DNA. One set of such machines is the polymerase sliding clamp - in this system, the DNA polymerase complex is held onto the DNA by a doughnut shaped sliding clamp protein (→), it encircles the DNA double helix and is strongly bound to the DNA polymerase (video link). So the question is, how does a protein come to encircle a DNA molecule? The answer is that the clamp protein is added to DNA by another protein molecular machine known as the clamp loader.[295] Once closed around the DNA the clamp can move freely along the length of the DNA molecule, but it cannot leave the DNA. The clamp’s sliding movement along DNA is diffusive – that is, it is driven by collisions with other molecules, with the average strength of such collisions related to the temperature of the system. Its movement is given a direction because the clamp is attached to the DNA polymerase complex which is adding monomers to the 3’ end of the growing nucleic acid polymer. This moves the replication complex (inhibited from diffusing away from the DNA by the clamp) along the DNA in the direction of synthesis. Processivity is increased since, in order to leave the DNA the polymerase has to disengage from the clamp or the clamp as to be removed by the clamp loader acting in reverse, that is, acting as an unloader.



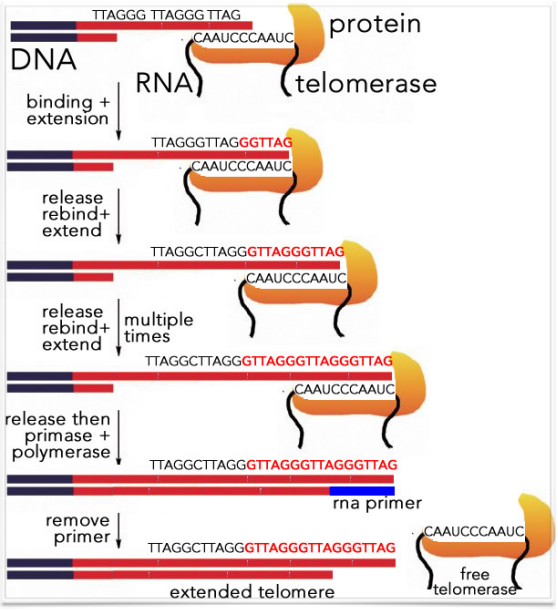
Further replication complexities in eukaryotes: telomeres

The DNA molecules found in bacteria and archaea are circular; they have no free ends.[296] Eukaryotic cells can contain more than 1000 times the DNA found in a typical bacterial cell. Instead of circles, they contain multiple linear molecules that form the structural basis of their chromosomes

[295] see Clamp loader ATPases and the evolution of DNA replication machinery & DNA Clamp & Clamp Loader video

[296] The mitochondria and chloroplasts of eukaryotic cells also contain circular DNA molecules, another homology with their ancestral bacterial parents. ,

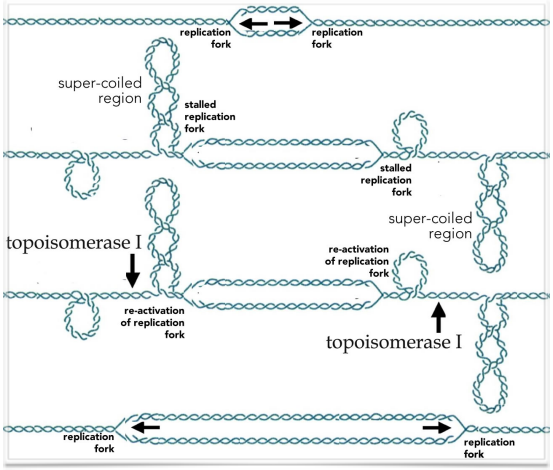
(more details in awhile). The free ends of the chromosomes are known as telomeres. The linearity of eukaryotic chromosomes creates problems replicating the ends of the DNA molecules. Left alone, more and more of the lagging strand end of the chromosome would go unreplicated, the end of the chromosome would begin to disappear with each DNA replication cycle. To address this "design limitation" in the DNA-dependent, DNA polymerase system eukaryotes use another RNA-protein complex, known as telomerase.[297]



Telomeres have a repeated sequence; in the case of human (and all other vertebrates) chromosomes end in repeated copies of the sequence TTAGGG-3' (→). The RNA part of the telomerase enzyme is the product of the TERC gene (OMIM:602322); it combines with the protein product of the TERT gene (OMIM:187270).[298] The TERC RNA contains a sequence complementary to the telomere DNA sequence and serves as the template for the synthesis of GGTTAG from the 3' end of the telomere's lagging strand - this process can occur multiple times, after which the primase and DNA-dependent, DNA polymerase can fill in the telomere end. Follow the footnote for further discussion of telomeres and telomerase.[299]

Topoisomerases

The circular nature of prokaryotic chromosomes creates its own issues, issues based on molecular topology. After replication, the two double-stranded DNA circles are linked together. Long linear DNA molecules can also become knotted together within eukaryotic cells. In addition, the replication of DNA unwinds the DNA, and this unwinding leads to what is known as the supercoiling of the DNA molecule. Left unresolved, supercoiling and knotting will inhibit the separation of replicated strands and DNA synthesis, perhaps you can explain why.[300] These topological issues are resolved by enzymes known as topoisomerases, because they can interconvert topologically distinct versions of the same molecule. There are two generic types of DNA topoisomerases. Type I topoisomerases (←) bind to the DNA, catalyze the breaking of a single bond in one sugar-phosphate-sugar backbone, and allow the release of overwinding through rotation around the bonds in the intact chain. When the tension is released, and the molecule has returned to its "relaxed" form, the enzyme catalyzes the reformation of the broken bond. Both bond breaking and



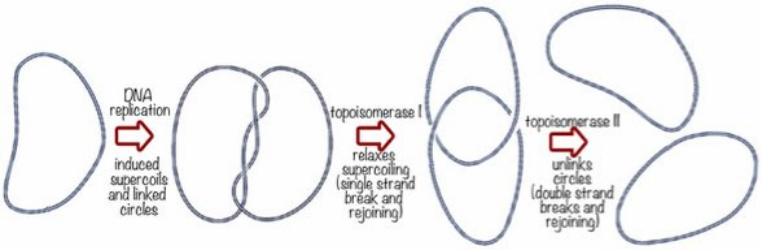
297 http://en.wikipedia.org/wiki/Telomerase

298 You can explore the known genetic diseases by using the web based On-line Mendelian Inheritance in Man (OMIM) database: http://www.ncbi.nlm.nih.gov/omim/

299 more on telomerase: http://blogs.scientificamerican.com/guest-blog/aging-too-much-telomerase-can-be-as-bad-as-too-little/

300 see this video on DNA supercoiling and topoisomerases: http://youtu.be/EYGrElVyHnU

reformation are coupled to ATP hydrolysis. Type II topoisomerases (↓) are involved in "unknotting" DNA molecules. These enzymes bind to the DNA, catalyze the hydrolysis of both backbone chains, but hold on to the now free ends. This allows another strand to "pass through" the broken strand. The enzyme also catalyzes the reverse reaction, reforming the bonds originally broken.



In addition to having typically much more DNA, the eukaryotic DNA replication enzyme complex is much slower, about 1/20th as fast as the prokaryotic system. While a bacterial cell can replicate its circular ~3 x 10^6 base pair chromosome in ~1500 seconds using a single origin of replication, the replication of the billions of base pairs of a typical eukaryote's DNAs involves the use of multiple (many) origins of replication, scattered along the length of each chromosome. So what happens when replication forks collide with one another? In the case of a circular DNA molecule, with its single origin of replication, the replication forks resolve in a specific DNA region known as the terminator. At this point type II topoisomerase allows the two circular DNA molecules to disengage from one another and move to opposite ends of the cell. The cell division machinery forms between the two DNA molecules. The system in eukaryotes, with their multiple linear chromosomes, is much more complex, although topoisomerases are still involved in separating replicated chromosomes, and involves more complex molecular machines that we will return to later, specifically in the complex of sexual reproduction (meiosis).

Questions to answer

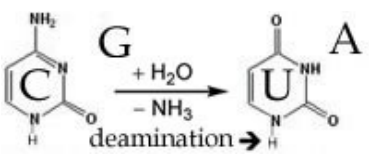
131. During DNA/RNA synthesis what is the average ratio of productive to unproductive interactions between an incoming nucleotide and the polymerase?
132. What are topological isomers?
133. Why do you need to denature (melt) the DNA double-helix to copy it?
134. How would DNA replication change if H-bonds were as strong as covalent bonds?
135. List all of the unrealistic components in this DNA replication video
136. Explain how DNA polymerase might recognize a mistake associated with a mismatched base pair.

Questions to ponder:

- How would evolution be impacted if DNA were totally stable and DNA replication was error-free? What would be the effect if the a mutation inactivated the the proof-reading function of the DNA polymerase complex?
- How might mutations in the genes encoding the clamp/clamp-loader system influence DNA replication?

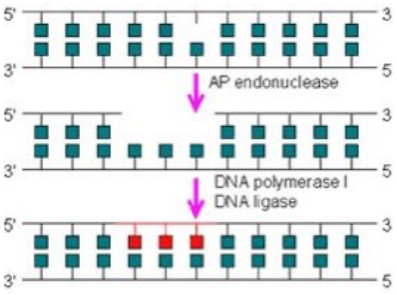
Mutations, deletions, duplications, and repair

While DNA is used as the universal genetic material of organisms, it is worth remembering that DNA is a thermodynamically unstable molecule. Eventually it will breakdown into more stable and dramatically simpler components. As DNA decomposes the information stored within its sequence will be lost. For example, at a temperature of ~13ºC, half of the phosphodiester bonds in a DNA sample will break after ~520 years.[301] But there is more. For example, cytosine groups within the DNA molecule can react with water, which (you might remember) is present at a concentration of ~54M inside a cell. This leads to a deamination reaction that transforms cytosine into uracil (→). If left unrepaired the original CG base pair will be replaced by an AU base pair in one strand during DNA synthesis. But, uracil is not normally found in DNA and its presence will be recognized by an enzyme that severs the bond between the uracil moiety and



[301] Here is the paper from which statement is derived: http://www.nature.com/news/dna-has-a-521-year-half-life-1.11555

the deoxyribose group.[302] The absence of a base, due either to its spontaneous loss or its enzymatic removal, acts as a signal for another enzyme system, the Base Excision Repair complex (←) that removes the section of the DNA strand with the missing base.[303] A DNA-dependent DNA polymerase can then bind to the open DNA and using the existing strand as a primer and the undamaged strand as a template, fill in the gap. Finally, another enzyme (a DNA ligase) joins the newly synthesized segment to the pre-existing strand. In the human genome there are over 130 genes devoted to repairing damaged DNA.[304]



Other hydrolysis reactions including depurination: the loss of a cytosine or thymine group and depyrimidination: the loss of an adenine or guanine group, lead to the removal of a base from the DNA. The rates of these reactions increases at acidic pH, which is probably one reason that the cytoplasm is not acidic. How frequent are such events? A human body contains ~10^{14} cells. Each cell contains about ~10^{9} base pairs of DNA. Each cell, whether it is dividing or not, undergoes ~10,000 base loss events per day or ~10^{18} events per day per person. That's a lot! The basic instability of DNA and the lack of repair after an organism dies means that DNA from dinosaurs, the last of which went extinct ~65,000,000 years ago, has disappeared from the earth, making it impossible to clone (or resurrect) a true dinosaur.[305] In addition DNA can be damaged by environmental factors, such as radiation, ingested chemicals, and reactive compounds made by the cell itself. Many of the most potent mutagens known are natural products, often produced by organisms to defend themselves against being eaten or infected by parasites, predators, or pathogens.[306]

A step back before going forward: what, exactly, is a gene anyway?

Now that we have introduced you to DNA and have casually referred to genes multiple times in various contexts, it is probably well past time that we seriously consider exactly what we mean by a gene.[307] Each organism (cell) carries its genomic DNA, which it replicates when it divides to produce an offspring. The DNA molecules (the genomes) of those organisms that survive and produce offspring become more frequent within a population than the genomes of those organisms that fail to reproduce to the same extent (or at all). As DNA is replicated and maintained within a cell, mutations arise. These mutations can influence the reproductive success of an organism. Over time this process (natural selection) leads to changes in the genomes of a population. When populations split into two (or more), their DNA molecules start changing independently of one another.

From a theoretical perspective there are two types of changes that can occur within a DNA molecule, those that influence the probability of reproductive success and those that do not. Those that influenced reproductive success can have either a positive or negative impact. If over time they become more frequent within the population, they are said to be under positive selection; those that become less frequent are said to be under negative selection. Whether a particular change in the

---

[302] UNG: uracil-DNA-N-glycosidase http://omim.org/entry/191525

[303] absent purine/absent pyrimidine endonuclease http://omim.org/entry/300773

[304] Human DNA Repair Genes – video with lots of misspelled words here: http://youtu.be/g4khROaOO6c

[305] DNA has a 521-year half-life: http://www.nature.com/news/dna-has-a-521-year-half-life-1.11555

[306] Dietary carcinogens, environmental pollution, and cancer: some misconception

[307] Part of the issue here involves the continuity of life and its long history. We always consider living systems that contain a range of molecules and reactive systems derived from their immediate ancestor - there simply is no easy “starting off point”.

DNA is beneficial or detrimental does not necessarily relate to the well being of the individual who carries these changes (mutations) but rather on its reproductive success within a population and in a particular environment. In asexual organisms, without complicating processes like horizontal gene transfer, mutations that have no effect on reproductive success are known as neutral mutations. They can be seen as a kind of molecular clock.[308] If we count the number of neutral changes in the genome sequences of two isolated populations (or organisms) we can use that information to estimate how long ago they shared a common ancestor. Of course this is not a particularly good clock in that there are only three possible changes a mutation that alters a single position in a genomic DNA molecule can make. For example if the original base is an A, it can change to a C, G, or T. Of course, that changed base could itself change; for example, if a A changed to a C, the C could change to an A, T, or G. BUT, if it changes to an A, we could not tell whether it had changed at all. A mutation that changes a single nucleotide at a particular position within the genomic DNA is known as a single nucleotide polymorphism or SNP (pronounced "snip"). Over long periods of time, the ability to date the divergence between organisms using the number of SNPs begins to loose resolution - a situation known as "long branch attraction".

Ah, but how do we know that a genomic change is neutral or subject to positive or negative selection? To begin to answer these questions, we need to know what mutations can do to a gene, and what changing a gene can do to an organism and its reproductive success. The answers to these questions are complex, but the path to such answers begins with recognizing what is stored in genomic DNA - namely information. Mutation, selection, and other evolutionary processes can add and remove information from the genome. Depending upon the circumstances, a mutation can have positive or negative effects on reproductive success.

We can recognize changes (mutations) that give rise to a measurable change in phenotype as influencing what we will call genes. There are many genes in an organism, originally identified by the phenotypes mutations in them produced. In a completely over-simplified view we find that a mutation in a particular region along a DNA molecule produces a similar or related phenotype. In some cases it was clear that a mutation alters the presence or activity of a particular enzyme, which led George Beadle (1903-1989) to put forward the one gene one protein (enzyme) model.[309] After awhile it became clear that many proteins are composed of the products of multiple genes, an example would be telomerase. Some genes encode RNAs that are used directly (e.g. the TERC gene) and some encode RNAs that are used to direct the synthesis of a polypeptide, such as TERT, while others encode RNAs that regulate the expression of genes. Understanding these interactions and their impact on the behavior of biological systems will be considered in detail in the second half of the course.

As we will see, and as you might probably already know, genes can be divided roughly into two domains: these are the regulatory regions and the region that serves to determine the sequence of a newly synthesized RNA molecule (knows as transcribed region). Mutations (changes in DNA sequence) in the regulatory regions influence where RNA synthesis starts and where, when, and how many RNAs are synthesized (per unit time). You will note that we have not mentioned where these two regions are with respect to one another. Defining all of the regulatory regions of a gene can be challenging, particularly since different regulatory regions may be used at different times and in the different cell types present within a multicellular organism. A gene's regulatory regions can span many thousands of kilobases of DNA and be located upstream, downstream, or within the gene's coding region. In addition, because DNA is double stranded, one gene can be located on one strand and another, different gene can be located on the other (anti-parallel) strand. We will return to the mechanisms of gene regulation later on, but as you may have discerned, gene regulation is complex and often the subject of its own course.

---

[308] The Molecular Clock and Estimating Species Divergence

[309] One gene one protein & One gene one enzyme

Transcribed domains can also be complex, particularly in eukaryotic genes: a single gene can produce multiple, functionally distinct gene products through the processes known as alternative promotor usage and RNA splicing.[310] How differences in gene sequence influence the activity and role(s) of a gene is not simple. A critical point to keep in mind is that a gene has meaning only in the context of a cell or an organism. Change the organism and the same, or rather, more accurately put, homologous genes (that is genes that share a common ancestor) can have different roles.

Alleles, their origins and their impact on evolution

Once we understand that a gene corresponds to a specific sequence of DNA, we understand that different versions of a gene, known as alleles, correspond to genes with different sequences. Two alleles of the same gene can differ from one another by as little as one out of thousands of nucleotides, or they can differ a multiple positions. In some cases, the differences between alleles can include deletions and duplications in the sequence. A complicating factor is that a particular gene product may have multiple functional roles, and a particularly trait can be influences by multiple genes. A particular allele of a particular gene may influence different functional roles and traits differently, something to keep in mind in the following discussion which, for simplicity's sake, focusses on a single functional role of a gene product and its influence on a single trait.

An allele can produce a gene product with completely normal function or no remaining functional activity at all, referred to as a null or amorphic allele. It can have less function than the "wild type" allele (hypomorphic), more function than the wild type (hypermorphic), or a new function (neomorphic). Given that many gene products function as part of multimeric complexes that are the products of multiple genes and that many organisms (like us) are diploid, there is one more possibility, the product of one allele can antagonize the activity of the other - this is known as an antimorphic allele. These different types of alleles were defined genetically by Herbert Muller, who won the Nobel prize for showing that X-rays could induce mutations, that is, new alleles.[311] The functional characterization of an allele is typically carried out with respect to how its presence influences a specific trait. Again, remember that most traits are influenced by multiple genes, and a single gene can influence multiple traits.

.The most common version of an allele is often referred to as the wild type allele (← a wild thing), but that is really just because it is the most common. There are often multiple alleles of a particular gene in the population and they all may be equally "normal", although they may influence different traits differently. If there is no significant selective advantage between them, their relative frequencies within a population drift over time. At the same time, the phenotype associated with a particular allele can be influenced by the alleles present at other genetic loci, known collectively as the genetic background. Since most traits are the results of many genes functioning together, and different combinations of alleles can produce different effects, the universe of variation is large. This can make identifying the genetic basis of a disease difficult, particularly when variation at any one locus may make only a minor contribution to the disease phenotype. On top of that, environmental and developmental differences can outweigh genetic influences on phenotype. Genetic background effects can lead to a particular allele producing a disease in one person and not another.[312]

---

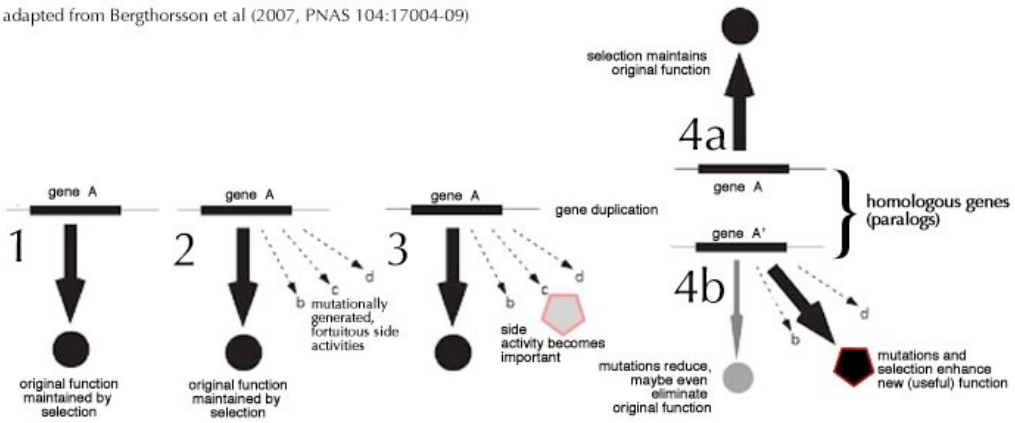
[310] Expansion of the eukaryotic proteome by alternative splicing see also Genes – way weirder than you thought

[311] Muller's morphs: https://en.wikipedia.org/wiki/Muller's_morphs

[312] Genetic background effects: https://www.sciencedaily.com/releases/2015/07/150716135104.htm

Mutations are the ultimate source of genetic variation – without them evolution would not occur. Mutations can lead to a number of effects, in particular, they can create new activities. At the same time most mutations reduce or alter the original (and necessary) activity of a gene, and that gene may encode an essential function. Left unresolved such molecular level conflicts would greatly limit the flexibility of evolutionary mechanisms. For example, it is common to think of a gene (or rather the particular gene product it encodes) as having one and only one function or activity, but in fact, when examined closely many catalytic gene products (typically proteins) can catalyze "off-target" reactions or carry out, even if rather inefficiently, other activities - they interact with other molecules within the cell and the organism. Assume for the moment that a gene encodes a gene product with an essential function as well as a potentially useful (from a reproductive success perspective) activities. Mutations that enhance these "ancillary functions" will survive (that is be passed on to subsequent generations) only to the extent that they do not (overly) negatively influence the gene's primary and essential function. Under these conditions, the evolution of ancillary functions may be severely constrained or blocked altogether.

This problem can be circumvented because the genome is not static. There are molecular level processes through which regions of DNA (and the genes that they contain) can be deleted, duplicated, and moved from place to place within the genome. Such genomic rearrangements, which are mutations because they change genome sequence, may occur during embryonic development. The end result is that not all cells in your body will have exactly the same genome.[313] In the case illustrated here (→), imagine that an essential but multifunctional gene is duplicated and moved elsewhere in the genome. Now one copy can continue to carry out its essential function, while the second copy is free to change as long as it does not interfere with the function of the essential gene. While many mutations will negatively effect the duplicated gene, some may increase and refine its favorable ancillary function. A new gene can emerge freed from the need to continue to perform its original (and essential) function. We see evidence of this type of process throughout the biological world. When a gene is duplicated, the two copies are known as paralogs. Such paralogs often evolve independently.



The origin of new (de novo) genes

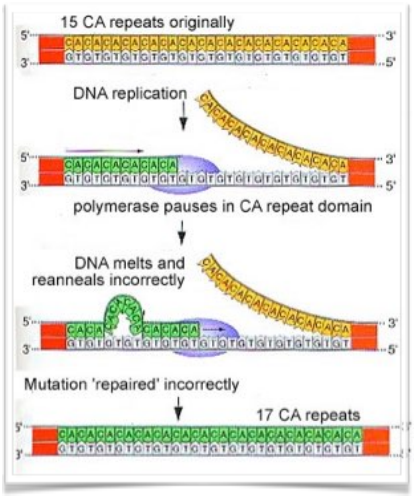
A key question is where, exactly, do brand new (de novo) genes come from?[314] A hint has been found from studies of RNA synthesis. New RNA sequencing and mapping techniques, made possible by the fact that more and more genomes have been sequenced, have revealed that a large percentage of the genome is used to direct RNA synthesis. This includes regions that do not appear to encode polypeptides. While some of these are regulatory RNAs, some do not appear to currently have a function. This opens the possibility that some of these can, because of the presence of

[313] Copy Number Variation in Human Health, Disease, and Evolution and LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes?

[314] Proto-genes and de novo gene birth [**link**] and How evolution builds genes from scratch [**link**]

regions able to encode polypeptides or play useful regulatory roles, will be become "proto-genes". If such a sequence enhances reproductive success, it can be subject to positive selection within a population and may become part of the organisms' genome. There is evidence for such events in fruit flies and humans.[315]

DNA repeat diseases and genetic anticipation



While they are essential for evolution, defects in DNA synthesis and genomic rearrangements more often lead to genetic (that is inherited) diseases than to any benefit to an individual. While we will return to mutational mechanisms and their effects as we continue, here we briefly consider diseases associated with DNA replication, specifically the class of genetic diseases known as the trinucleotide repeat disorders (→). There are a number of such "triplet repeat" diseases, including several forms of mental retardation, Huntington's disease, inherited ataxias, and muscular dystrophy. These diseases are caused by slippage of DNA polymerase and the subsequent duplication of sequences. When these "slipable" repeats occur in a region of DNA encoding a protein, they can lead to regions of repeated amino acids. For example, expansion of a domain of CAGs in the gene encoding the polypeptide Huntingtin (OMIM:613004) leads to the neurological disorder Huntingdon's chorea. OMIM stands for the "On-line Inheritance in Man" website.

A mechanistically related pathogenic syndrome is known as Fragile X (OMIM:300624). The underlying DNA replication defect is the cause of the most common form of autism of known cause (most forms of autism have no known cause). About 6% of autistic individuals have Fragile X syndrome. Fragile X syndrome can also lead to anxiety disorders, attention deficit hyperactivity disorder, psychosis, and obsessive-compulsive disorder. Because the mutation involves the *FMR1* gene (OMIM:309550), which is located on the X chromosome, the disease is sex-linked and effects mainly males, who are XY, compared to females, who have two copies of the X chromosome. In the unaffected population, the *FMR1* gene contains between 6 to 50 copies of a CGG repeat. Individuals with between 6 to 50 repeats are phenotypically normal. Those with 50 to 200 repeats carry what is known as a pre-mutation; these individuals rarely display symptoms but can transmit the disease to their children. Those with more than 200 repeats typically display symptoms and often have what appears to be a broken X chromosome – from which the disease derives its name. The pathogenic sequence in Fragile X is downstream of the *FMR1* gene's coding region. When this region expands, it inhibits the expression of the *FMR1* gene.[316] There are a number of processes that can mediate the pathogenic effects of DNA repeat diseases, some of which we will consider when we discuss the inheritance of these conditions.

Other DNA Defects: Defects in DNA repair can lead to severe diseases and often a susceptibility to cancer. A search of OMIM for DNA repair returns 654 entries! For example, defects in DNA mismatch repair lead to a susceptibility to colon cancer, while defects in translation-coupled DNA repair are associated with Cockayne syndrome. People with Cockayne's syndrome (OMIM:216400 & 133540) are sensitive to light, are of short stature, and appear to age prematurely.[317]

---

[315] Origin and spread of de novo genes in Drosophila melanogaster populations [**link**], Origins of *De Novo* Genes in Human and Chimpanzee [**link**] and De novo mutations across 1,465 diverse genomes reveal mutational insights and reductions in the Amish founder population [**link**]

[316] Molecular mechanisms of fragile X syndrome: a twenty-year perspective.

[317] Cockayne syndrome: http://omim.org/entry/278760

Our introduction to genes has necessarily been quite foundational and we will extend it in the second half of the course. There are lots of variations and associated complexities that occur within the biological world. The key ideas are that genes represent biologically meaningful DNA sequences. To be meaningful, the sequence must play a role within the organism, typically by encoding a gene product (which we will consider next) and/or the information needed to insure its correct expression, that is, where and when the information in the gene is used. A practical problem is that most studies of genes are carried out using organisms grown in the lab or in otherwise artificial or unnatural conditions. It might be possible for an organism to exist with an amorphic allele of a gene in the lab, whereas organisms that carry that allele may well be at a significant reproductive disadvantage in the real world. Moreover, a particular set of alleles, a particular genotype, might have a reproductive advantage in one environment (one ecological/behavioral niche) but not another. Measuring these effects can be difficult. All of which should serve as a warning that would should consider skeptically pronouncements that a gene, or more accurately a specific allele of a gene, is responsible for a certain trait, particularly if the trait is complex, ill-defined, and likely to be significantly influenced by genomic context (the rest of the genotype) and environmental factors. Intelligence is one such complex trait. A dramatic example of the difficulty in defining a gene product's functions is illustrated by the studies of Hutchinson et al; they produced a minimal bacterial genome containing 473 genes.[318] Of these genes, the function(s) of 149 (~32% of the total genome) were unknown, a rather surprising result.

Questions to answer

137. How does a mutation generate a new allele? How is a mutation different from an allele?
138. What would be a reasonable way to determine that you had defined an entire gene?
139. Is it possible to build a system (through evolutionary mechanisms) in which mutations do not occur?

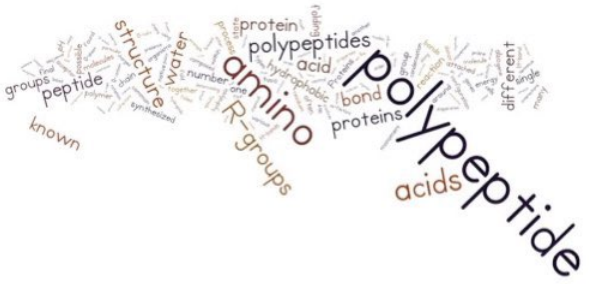
Questions to ponder:

- How could removing information from the genome enhanced reproductive success?
- How might you go about defining the function of a "gene with unknown function"?

[318] Design and synthesis of a minimal bacterial genome. https://www.ncbi.nlm.nih.gov/pubmed/27013737

Chapter 8: Peptide bonds, polypeptides, proteins, and molecular machines

In which we consider the nature of proteins, how they are synthesized and assembled, how they get to where they need to go within the cell and within the organism, how they function, how their activities are regulated, and how mutations can influence their expression, stability, activity, and evolution.



We mentioned proteins many times, since there are few biological processes that do not rely on them. Proteins act as structural elements, signals, regulators, and catalysts in a wide range of molecular machines. Up to this point, however, we have not said much about what they are, how they are made, and how they come to do what they do. The first scientific characterization of what are now known as proteins was published by the Dutch chemist, Gerardus Johannes Mulder (1802–1880).[319] After an analysis of a number of different substances, he proposed that all proteins contain a common chemical core, with the molecular formula $C_{400}H_{620}N_{100}O_{120}P_1S_1$, and that the differences between different proteins were primarily in the numbers of phosphate (P) and sulfur (S) atoms they contained. The name "protein", from the Greek word πρώτα ("prota"), meaning "primary", was suggested by the Swede, Jons Jakob Berzelius (1779–1848) based on the presumed importance of these compounds in biological systems.[320] As you can see, Mulder's molecular formula was not very informative, it tells us little or nothing about protein structure, but suggests that all proteins are fundamentally similar, which while true is confusing since they carry out so many different roles. Subsequent studies revealed that proteins could be dissolved in either water or dilute salt solutions but aggregated and became insoluble when the solution was heated; as we will see this aggregation reaction reflects a change in the structure of the protein. Mulder was able to break down proteins into amino acids through an acid hydrolysis reaction. Amino acids get their name from the fact that they contain both an amino (–NH_2) and a carboxylic acid (–COOH) group. While there are many thousands of possible amino acids, only twenty (or rather twenty two, as we will see) different amino acids could be identified in hydrolyzed samples of proteins. Since their original characterization as a general class of compounds, we now understand that while proteins share a common basic polymer structure, they are remarkably diverse. Proteins are involved in roles from the mechanical strengthening of skin, the building of shells and claws, the regulation of genes, the transport of oxygen, the capture of energy, the release of light, and the catalysis and regulation of essentially all of the chemical reactions that occur within cells and organisms.

While all proteins have a similar bulk composition, this obscures rather than illuminates their structural and functional differences. With the introduction of various chemical methods, it was discovered that different proteins were composed of distinct and specific sets of subunits, and that each subunit is an unbranched polymer with a specific amino acid sequence. Because the amino acids in these polymers are linked by what are known as peptide bonds, the polymers are known generically as polypeptides. At this point, it is important to reiterate that proteins are functional objects, and specific proteins are composed of specific sets of distinct polypeptides; moreover, each distinct polypeptide is encoded by a distinct gene. In addition to polypeptides many proteins also contain other molecular components, known as co-factors or prosthetic groups (we will call them co-

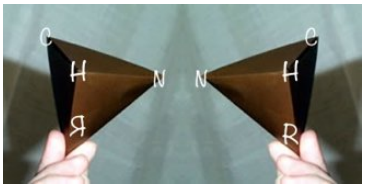
---

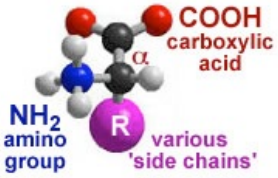
[319] From 'protein' to the beginnings of clinical proteomics: http://www.ncbi.nlm.nih.gov/pubmed/21136729

[320] While historically true, the original claim that proteins get their name from "the ancient Greek sea-god Proteus who, like your typical sea-god, could change shape. The name acknowledges the many different properties and functions of proteins." seems more poetically satisfying to us.

factors for simplicity's sake.) These co-factors can range from metal ions to various small molecules. A protein is a fully assembled and functional entity.

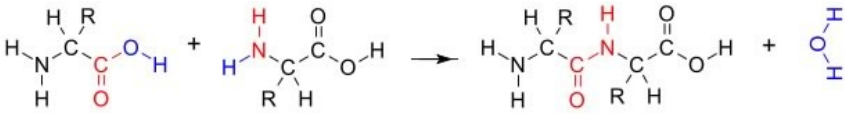
As you might remember from your chemistry courses carbon atoms (C) typically form four bonds. We can think of an amino acid as a (highly) modified form of methane (CH_4), with the C referred to as the alpha carbon (C_α). Instead of four hydrogens, in a biological amino acid there is an H, an amino group ($-NH_2$), a carboxylic acid group (-COOH), and a final, variable (R) group attached to the central C_α atom. The four groups attached to the α-carbon are arranged at the vertices of a tetrahedron (→). If all four groups attached to the α-carbon are different from one another, as they are in all biological amino acids except glycine, the resulting amino acid can exist in two forms, known as enantiomeric stereoisomers. Enantiomers are mirror images of one another and are referred to as the L- and D- forms. Only L-type amino acids are found in proteins, even though there is no obvious chemical reason for why proteins could not have also been made using both types of amino acids or using only D-amino acids for that matter.[321] It appears that the universal use of L-type amino acids in the polypeptides found in biological systems is one more example of the evolutionary relatedness of organisms, it appears to be a homologous trait, presumably established in the last universal common ancestor (LUCA). Similarly, even though there are hundreds of different amino acids known, only 22, the 20 common amino acids and two others, selenocysteine and pyrrolysine, are found in proteins and presumably were present in LUCA.





Amino acids differ from one another by their R-groups, which are often referred to as "side-chains". Some of these R-groups are large, some are small, some are hydrophobic, some are hydrophilic, some of the hydrophilic R-groups contain weak acidic or basic groups. The extent to which these weak acidic or basic groups are positively or negatively charged changes in response to environmental pH. Changes in charge will (as we will see) influence the structure of the polypeptide/protein in which they find themselves. The different R-groups provide proteins with a broad range of chemical properties, which are further extended by the presence of co-factors.[322]

As we noted for nucleic acids, a polymer is a chain of subunits. In the case of a polypeptide, amino acid monomers are linked together by peptide bonds. Under the conditions that exist inside the cell, this is a thermodynamically unfavorable dehydration reaction, and so polypeptide synthesis is coupled to a thermodynamically favorable reaction, a nucleotide triphosphate hydrolysis reaction. A molecule formed from two amino acids, joined together by a peptide bond, is known as a dipeptide. A dipeptide has an N-terminal (amino) end and a C-terminal (carboxylic acid) end. To generate a polypeptide, new amino acids are added sequentially (and only) to the C-terminal end of the polymer – a reaction analogous to the synthesis of a polynucleotide, with addition of monomers to one end of the growing polymer. A peptide bond forms between the amino group of the added amino acid and the carboxylic acid group of the polymer; the formation of a peptide bond is associated with the release of a water molecule (↓). When complete, the addition of an amino acid to the C-terminus of a polypeptide generates a new C-terminal carboxylic acid group. It is important to note that while some amino acids have a



[321] It is not that D-amino acids do not occur in nature, or in organisms, they do. They are found in biomolecules, such as the antibiotic gramicidin, which is composed of alternating L-and D-type amino acids - however gramicidin is synthesized by a different process than that used to synthesize proteins.

[322] Bioengineers are working to go Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon & to incorporation of non-canonical amino acids into proteins in yeast; something made possible due to the redundancy of the genetic code.

carboxylic acid group as part of their R-groups, new amino acids are not added there. Because of this fact, polypeptides are synthesized as unbranched, linear polymers. The process of amino acid addition can continue, theoretically without limit. Biological polypeptides range from the very short (5-10) to very long (many hundreds to thousands) of amino acids in length.[323] For example, the Titin polypeptide (found in muscle cells) can be more than 30,000 amino acids in length.[324] Because there is no theoretical constraint on which amino acid occurs at a particular position within a polypeptide, there is a enormous number of possible polypeptides that can exist. In the case of a 100 amino acid long polypeptide, there are more than 20^{100} possible different polypeptides that could, in theory, be formed.

Questions to answer:

140. How does a polypeptide chain resemble and how does it differ from a nucleic acid molecule?
141. What are the "natural" limits to the structure of an R-group in an amino acid?

Question to ponder:

- Why does it make sense to think that the presence of a common set of amino acids in organisms is a homologous trait?

Specifying a polypeptide's sequence

At this point you might be asking yourself, if there are so many different possible polypeptides, and there is no inherent bias favoring the addition of one amino acid over another, what determines the sequence of amino acids within a polypeptide, presumably it is not random. Here we connect the specification of polypeptide sequence to the information stored in DNA. We begin with a description of the process in bacteria and then extend it to archaea and eukaryotes. We introduce them in this order because, while basically similar (homologous), the system is somewhat simpler in bacteria, although you might find it complex enough for your taste. Even so, we will leave most of the complexities for subsequent courses. One thing that we will do that is not common is that we will consider the network dynamics of these systems. We will even ask you to make plausible predictions about the behavior of these systems, particularly in response to various perturbations, mutations and such. Another important point to keep in mind, one we have made previously, is that the system is continuous. The machinery required for protein synthesis is inherited by the cell, and new copies of it are synthesized as the cell grows; each new polypeptide is synthesized in an environment full of pre-existing proteins and ongoing metabolic processes.

A bacterial cell synthesizes thousands of different polypeptides. The sequence of these polypeptides, the exact amino acids from the N-terminal start to the C-terminal end of the polypeptide, is encoded within the organism's DNA. The bacterial genome is a double-stranded circular DNA molecule that is millions of base pairs in length. Each polypeptide is encoded by a specific region of this DNA molecule. So, our questions are how are specific regions in the DNA recognized and how is the information present in nucleic acid-sequence **translated** into polypeptide sequence.

To address the first question let us think back to the structure of DNA. It was immediately obvious that the one-dimensional sequence of a polypeptide could be encoded in the one-dimensional sequence of the polynucleotide chains in a DNA molecule.[325] The real question was how to translate the language of nucleic acids, which consists of sequences of four different

---

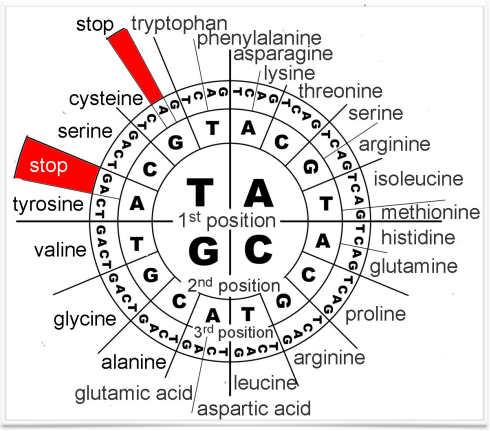
[323] Short polypeptides, or rather the genes that encode them, can be difficult to recognize since short "open reading frames" are difficult to identify unambiguous: see Peptidomic discovery of short open reading frame–encoded peptides in human cells

[324] OMIM entry for TITIN: http://omim.org/entry/188840

[325] Nature of the genetic code finally revealed!: http://www.nature.com/nrmicro/journal/v9/n12/full/nrmicro2707.html

nucleotides, into the language of polypeptides, which consists of sequences of the 20 (or 22) different amino acids. As pointed out by the physicist George Gamow (1904-1968)[326] the minimum set of nucleotides needed to encode all 20-22 amino acids is three; a sequence of one nucleotide (4^1) could encode at most four different amino acids, a two nucleotide sequence could encode (4^2) or 16 different amino acids (not enough), while a three nucleotide sequence (4^3) could encode 64 different amino acids (more than enough).[327] Although the actual coding scheme that Gamow proposed was wrong, his thinking about the coding capacity of DNA influenced those who set out to experimentally determine the actual rules of the "genetic code".

The genetic code is not the information itself, but the algorithm by which nucleotide sequences are "read" to determine polypeptide sequences. A polypeptide is encoded by the sequence of nucleotides. This nucleotide sequence is read in groups of three nucleotides, known as a codon. The codons are read in a non-overlapping manner, with no spaces (that is, non-coding nucleotides) between them. Since there are 64 possible codons but only 20 (or 22) different amino acids used in organisms, the code is redundant, that is, certain amino acids are encoded for by more than one codon. In addition there are three codons, UAA, UAG and UGA, that (in most organisms) do not encode any amino acid but are used to mark the end of a polypeptide, they encode "stops" or periods (→).



The region of the nucleic acid that encodes a polypeptide begins with what is known as the "start" codon and continues until one of the three stop codons is reached.[328] A sequence defined by in-frame start and stop codons, with some number of codons between them, is known as an open reading frame, an ORF. At this point it is important to note that while the information encoding a polypeptide is present in the DNA, the DNA copy of this information is not used directly to specify the polypeptide sequence. Rather, the process is indirect, it involves an intermediate. The information in the DNA is first copied (transcribed) into an RNA molecule, known as a messenger RNA or mRNA; it is the mRNA molecule that directs polypeptide synthesis. The process of copying information within DNA into an RNA molecule is known as transcription because both DNA and RNA use the same nucleotide sequence language. In English, as opposed to molecular biology, transcription is the process of making a written copy of what someone says - the language of both is the same. In contrast polypeptides are written in a different language, amino acid sequences. For this reason the process of RNA-directed polypeptide synthesis is known as translation, which involves changing between languages, from nucleic acid-ese to polypeptide-ese.

The origin of the genetic code

There are a number of hypotheses as to how the genetic code originated. One is the frozen accident model in which the code used in modern cells is the result of an evolutionary accident, a bottleneck event associated with the appearance of LUCA. Early in the evolution of life on Earth, there may have been multiple types of proto-organisms, each using a different genetic code. The common genetic code found in all existing organisms reflects the fact that only one of these proto-organisms gave rise to all modern organisms. Alternatively, the code could reflect specific interactions between RNAs and amino acids that played a role in the initial establishment of the

[326] when he was a professor at UC Boulder

[327] The Big Bang and the genetic code: Gamow, a prankster and physicist, thought of them first

[328] There are situations in which non-start codons occur: see repeat-associated non-ATG translation (RAN translation)

code. It is not clear which model reflects what actually happened, it is likely to be theoretically unknowable, at least until unrelated forms of life are discovered on Earth or elsewhere. What is clear, however, is that the code is not absolutely fixed, there are examples in which certain codons are "repurposed" in various organisms. In fact there are efforts to re-engineer codons to produce proteins made using a range of more than 100 "unnatural" amino acids (uAAs).[329] What these variations in the genetic code illustrate is that evolutionary mechanisms can change the genetic code.[330] Since the genetic code does not appear to be predetermined, the general conservation of the genetic code among organisms is seen as strong evidence that all organisms, even the ones with minor variations in their genetic codes, are derived from a single common ancestor. It appears that the genetic code is a homologous trait shared by all known organisms.

Protein synthesis: transcription (DNA to RNA)

Having introduced you to DNA, mRNA, and the genetic code, however briefly, we now return to the process by which a polypeptide is specified by a DNA sequence. Our first task is to understand how it is that we might be able to find the specific region within a DNA molecule that encodes a specific polypeptide; we are looking for a relatively short region of DNA within millions (in prokaryotes) or billions (in eukaryotes) of base pairs of DNA. So while the double-stranded nature of DNA makes the information stored in it redundant, a fact that makes DNA replication straightforward, the specific nucleotide sequence that will be decoded using the genetic code is present in only one of the two strands. From the point of view of polypeptide sequence the other strand is effectively nonsense. One complexity associated with the double-stranded and anti-parallel nature of DNA is that information containing sequences can, in theory, run along either strand, although in opposite directions. This means that a gene's regulatory sequence must specify where, when and how often RNA synthesis starts and which of the two anti-parallel DNA strands is used to specify the "expressed" RNA's sequence.

If we think about this problem - we recognize one way to "find" a gene involves nucleotide sequences, together with something that can "read" a specific nucleotide sequence. Let us consider a specific form of the problem, say we want to uniquely specify one gene (one sequence) within the ~3,000,000 base pairs of an *E. coli*'s cell's genomic DNA. For simplicity let us assume that the A:T ratio equals the G:C ratio. Clearly a one base pair sequence will not work, since we might expect that half of the base pairs will be recognized, either by directly binding to T or indirectly by binding to an A. To be unique the sequence we want must occur once in 3,000,000 base pairs (1/3,000,000 = 3.33… x 10^{-7} = 0.000000333). If we use a two base sequence, it will occur 1/4 x 1/4 = 1/16 = 0.0625, a four base sequence 0.0039, an eight base sequence 0.00001523, but a 16 base sequence has a probability of occurring purely by chance of ~2.32 x 10^{-10}, which is less than once per genome.[331]

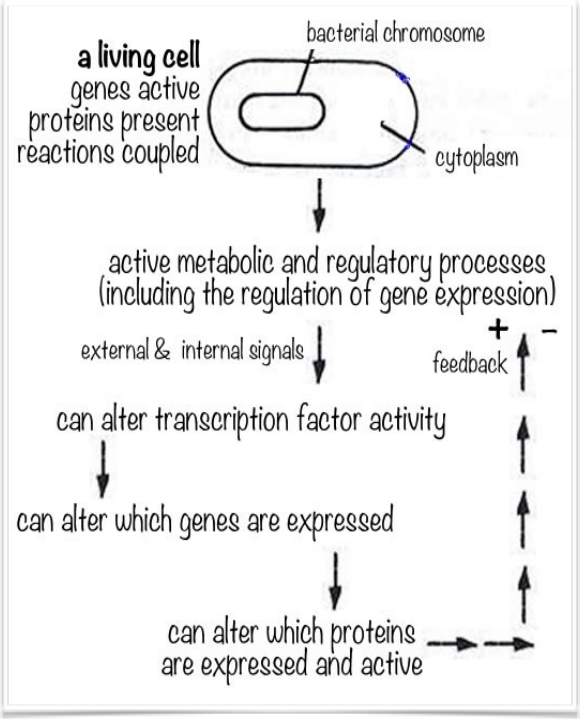
Once a gene's regulatory region is identified (by the binding of a specific type of protein - see below), it can be "expressed". In fact, it is common to say that a gene is expressed only when RNAs are synthesized (transcribed) from it. If a gene is not expressed, that means that no RNAs corresponding to its sequence are being synthesized within the cell. In a sense, it is as if it is not there (at least in a particular cell type or environmental condition). RNA synthesis is mediated by a DNA-dependent, RNA polymerase, which is encoded by genes (→ next page). Where, and in which orientation, the polymerase binds to the gene's DNA is determined by the gene's regulatory sequence(s), inherited from the organism's parent(s), and the protein(s), known as transcription factors, bound to it. Transcription factor proteins are themselves encoded by genes. Polymerase can

---

[329] Designing logical codon reassignment – Expanding the chemistry in biology

[330] The genetic code is nearly optimal for allowing additional information within protein-coding sequences & Stops making sense: translational trade-offs and stop codon reassignment:

[331] As we will return to, the CRISPR CAS9 system for mutagenesis uses a 22-base "guide RNA" to direct an endonuclease; this, in theory at least, would be expected to guarantee one target per genome.



bind to the DNA-transcription factor complex, the first step in the synthesis of a new RNA. Of course, since there are many genes in the genome, the stability of the DNA-Transcription Factor-Polymerase complex, as well as a number of other factors, will impact the number of RNAs from a particular gene that are synthesized per unit time. In addition to mRNAs, a number of other types of RNAs are synthesized, these include structural, catalytic, and regulatory RNAs. We will get to further complexities in a bit.

At this point, it is useful to explicitly recognize some common aspects of biological systems. They are highly regulated, adaptive and homeostatic - that is, they can adjust their behavior to changes in their environment (both internal and external) to maintain the living state. These types of behaviors are based on various forms of feedback regulation. In the case of the bacterial gene expression system, there are genes that encode specific transcription factors. Which of these genes are expressed determines which transcription factor proteins are present and, in turn, which genes are actively expressed. Of course, the gene encoding a specific transcription factor is itself regulated. Transcription factors can act positively or negatively, which means that they can lead to the activation of transcription by recruiting and activating the RNA polymerase or blocking its recruitment and/or its activation. In addition the activity of a particular transcription factor can be regulated (a topic we will return to later on in this chapter).

All organisms are complex. A “simple” bacterium contains thousands of genes and different sets of genes are used in different environments and situations, and in different combinations to produce specific behaviors. In some cases, these behaviors may be mutually antagonistic. For example, a bacterium facing a rapidly drying out environment might turn on specific genes in order to prepare itself to survive in a more hostile environment. Our goal is not to generate perfectly accurate predictions about the behavior of an organism in a particular situation, but rather that you can make plausible predictions about how gene expression will change in response to various perturbations. This requires us to consider, although at a rather elementary level, the regulatory processes active in cells.

For a transcription factor to regulate a specific gene, either positively or negatively, it must be able to bind to specific sequences within the DNA. Whether or not a gene is expressed, whether it is “on” or “off”, depends upon which transcription factors are expressed, are active, and can interact productively with the DNA-dependent, RNA polymerase (commonly referred to as RNA polymerase). You might speculate that groups of genes that are expressed together, under common cellular and environmental conditions, may have similar regulatory sequences, sequences that regulated by the same or related transcription factor proteins, a situation that makes it possible to regulate groups of genes in a coordinated manner. Inactivation of a transcription factor can involve a number of mechanisms, including its destruction, modification, or interactions with other proteins, so that it no longer interacts productively with either its target DNA sequence or the RNA polymerase. Similarly the activity of a transcription factor can be regulated (as we will see). Once a transcription factor is active, it can diffuse through out the cell and (in prokaryotic cells that do not have a barrier to control interactions with DNA) can bind to its target DNA sequences. Now an RNA polymerase can bind to the DNA-transcription factor complex, an interaction that can lead to the activation of the RNA polymerase and the initiation of RNA synthesis, using one DNA strand to direct RNA synthesis. Once RNA polymerase has been activated, it will move away from the transcription factor-DNA complex.

The DNA bound transcription factor can then bind another polymerase or the transcription factor can release from the DNA (in response to molecular level collisions), and can diffuse away, interact with other regulatory factors, or rebind to other sites in the DNA. Clearly the number of copies of a particular transcription factor protein and its interaction partners and DNA binding sites will impact the behavior of the system, as will the number of ancillary factors that must interact with the transcription factor/DNA complex in order to recruit and activate the polymerase.

RNA synthesis is a thermodynamically unfavorable reaction, so for it to occur it must be coupled to a thermodynamically favorable reaction, in particular nucleotide triphosphate hydrolysis reactions. The RNA polymerase moves along the DNA (or the DNA moves through the RNA polymerase, your choice), to generate an RNA molecule (the transcript). Other signals within the DNA, and recognized by proteins associated with the transcription machinery, lead to the termination of transcription and the release of the RNA polymerase. Once released, the RNA polymerase returns to its inactive state. It can act on another gene if the RNA polymerase interacts with transcription factors bound to the gene's promoter. Since multiple type transcription factor proteins are present within the cell and RNA polymerase can interact with all of them, which genes are expressed within a cell will depend upon the relative concentrations and activities of specific transcription factors and their regulatory and associated proteins, together with the binding affinities of particular transcription factors for specific DNA sequences (compared to their general low-affinity binding to DNA in general).

Protein synthesis: translation (RNA to polypeptide)

Translation involves a complex cellular organelle, the ribosome, which together with a number of accessory factors reads the code in an mRNA molecule and produces the appropriate polypeptide.[332] The ribosome is the site of polypeptide synthesis. It holds the various components, the mRNA, tRNAs, and accessory factors, in appropriate juxtaposition to one another to catalyze polypeptide synthesis. But perhaps we are getting ahead of ourselves. For one, what exactly is a tRNA?

The process of transcription is used to generate a number of types of RNAs beside mRNAs; these play structural, catalytic, and regulatory roles within the cell. Of these non-mRNAs, two are particularly important in the context of polypeptide synthesis. The first are molecules known as transfer RNAs (tRNAs). These small single stranded RNA molecules (→) fold back on themselves to generate a compact L-shaped structure. In the bacterium *E. coli*, there are 87 genes that encode tRNAs (there are over 400 such tRNA encoding genes in humans). For each amino acid and each codon there are one or more tRNAs. The only exceptions are the so called stop codons, for which there are no tRNAs. A tRNA specific for the amino acid phenylalanine would be written $tRNA^{Phe}$. Two parts of the tRNA molecule are particularly important and functionally linked: the part that recognizes the codon within the ribosome-bound mRNA complex, and the amino acid acceptor stem, which is where an amino acid is covalently attached to the tRNA. Each specific type of tRNA can recognize a particular codon in an mRNA through base pairing interactions through what is known as its anti-codon. The rest of the tRNA molecule mediates interactions with protein catalysts (enzymes) known as amino acyl tRNA synthetases. There is a distinct amino acyl tRNA synthetase for each amino acid: there is a phenylalanine-tRNA synthetase and a proline-tRNA synthetase, etc. An amino acyl tRNA c binds the appropriate tRNA and the appropriate amino acid and, through a reaction coupled to a thermodynamically favorable nucleotide triphosphate hydrolysis reaction, catalyzes the formation of a covalent bond between the amino acid acceptor stem of the tRNA and the amino acid, to form what is known as a charged or amino acyl-tRNA. The loop containing the anti-codon is located at the other end of the tRNA molecule. As we will see, in the course of polypeptide synthesis, the amino

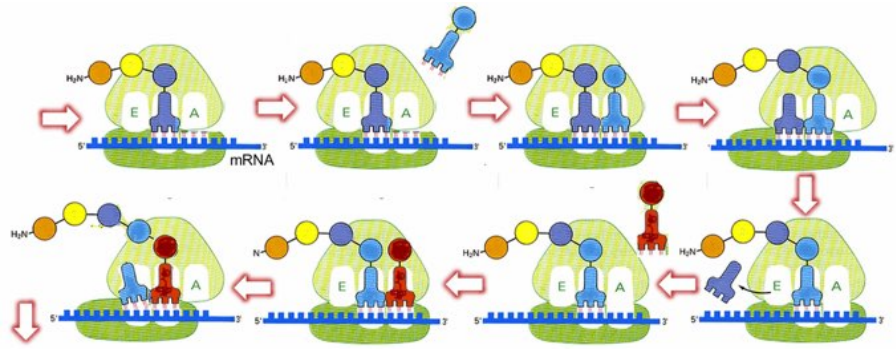
[332] Can't stop yourself? go here for a more detailed description of translation.

acid group attached to the tRNA's acceptor stem will be transferred from the tRNA to the end of a growing polypeptide.

Ribosomes

Ribosomes are composed of roughly equal amounts by mass of ribosomal RNAs (rRNAs) and ribosomal polypeptides. An active ribosome consists of a small and a large ribosomal subunit. In the bacterium *E. coli*, the small subunit is composed of 21 different polypeptides and a 1542 nucleotide long rRNA molecule, while the large subunit is composed of 33 different polypeptides and two rRNAs, one 121 nucleotides long and the other 2904 nucleotides long.[333] Each ribosomal polypeptide and RNA is itself a gene product. The complete ribosome has a molecular weight of ~3 x 10^6 daltons (please note, there is no reason to remember any of these numbers except to appreciate that the ribosome is a complex molecular machine). One of the rRNAs is an evolutionarily conserved catalyst, known as a ribozyme (in analogy to protein based catalysts, which are known as enzymes). This rRNA lies at the heart of the ribosome and catalyzes the transfer of an amino acid bound to a tRNA to the carboxylic acid end of the growing polypeptide chain, also attached to a tRNA. RNA based catalysis is a conserved feature of polypeptide synthesis and appears to represent an evolutionarily homologous trait.

The growing polypeptide chain is bound to a tRNA, known as the peptidyl tRNA. When a new aa-tRNA enters the ribosome's active site (site A), the growing polypeptide is added to it, so that it becomes the peptidyl tRNA, with a newly added amino acid, the amino acid originally associated with the incoming aa-tRNA (↓). This attached polypeptide group is now one amino acid longer.



The cytoplasm of cells is packed with ribosomes. In a rapidly growing bacterial cell, ~25% of the total cell mass consists of ribosomes. Although structurally similar, there are characteristic differences between the ribosomes of bacteria, archaea, and eukaryotes, a point of significance since a number of antibiotics selectively inhibit bacterial but not eukaryotic ribosome-mediated protein synthesis. Both chloroplasts and mitochondria have ribosomes of the bacterial type; another piece of evidence that they are descended from bacterial endosymbionts. Protein synthesis blocking anti-bacterial antibiotics are mostly benign since they do not block most of the protein synthesis that occurs in a eukaryotic cell.

The translation (polypeptide synthesis) cycle

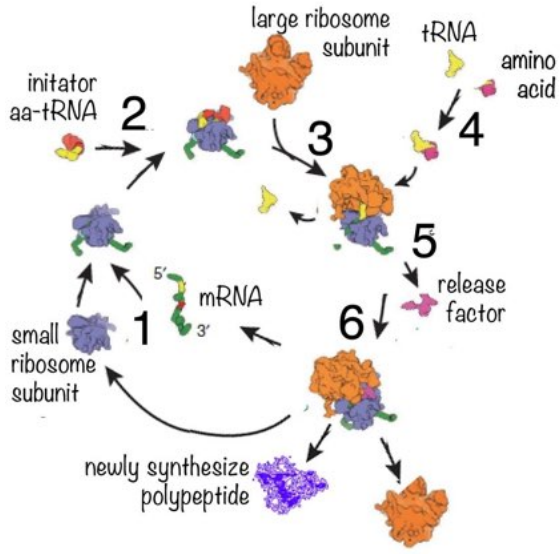
In bacteria and archaea, there is no barrier between the cell's DNA and its cytoplasm, which contains the ribosomal subunits together with the other components involved in polypeptide synthesis. Newly synthesized RNAs emerge from the RNA polymerase directly into the cytoplasm, where they can interact with ribosomes. For this reason, in bacteria and archea, the process of protein synthesis (translation) can begin before mRNA synthesis (transcription) is complete.

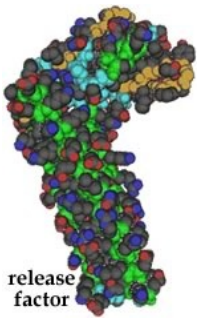
We will walk through the process of protein synthesis, but at each step we will leave out the various accessory factors involved in regulating the process and coupling it to the thermodynamically favorable reactions that make it possible. These are important to consider if you

[333] In the human, the small ribosomal subunit is composed of 33 polypeptides and a 1870 nucleotide rRNA, while the large ribosomal subunit contains 47 polypeptides, and three rRNAs of 121, 156, and 5034 nucleotides in length.

want to re-engineer or manipulate the translation system, but (we think) are unnecessary details that obscure a basic understanding of the underlying processes. Here we will remind you of two recurring themes. The first is to recognize that mRNA-directed polypeptide synthesis (translation) can occur only because all the components needed already exist in the cell. The second is that all of the interactions we will be describing are based on stochastic, thermally driven collisions. For example, consider the addition of an amino acid to a tRNA, the formation of an amino acyl-tRNA or aa-tRNA; random motions bring the correct amino acid and the correct tRNA to their binding sites on the appropriate amino acyl tRNA synthetase. Once the aa-tRNA is formed, only the correct amino acid charged tRNA will bind productively to the ribosome-mRNA-nascent polypeptide complex. Generally, many unproductive collisions occur before a productive (correct) one, since there are more than 20 different amino acid/tRNA molecules bouncing around in the cytoplasm. The stochastic aspects of the peptide synthesis process are rarely illustrated.

The first step in polypeptide synthesis is the synthesis of the specific mRNA that encodes the polypeptide (↓). (1) The mRNA contains a sequence that mediates its binding to the small ribosomal subunit.[334] This sequence is located near the 5' end of the mRNA. (2) The mRNA-small ribosome subunit complex now interacts with and binds to a complex containing an initiator (start) amino acid:tRNA. In bacteria, archaea, and eukaryotes the start codon is generally an AUG codon and encodes the amino acid methionine, although other, non-AUG start codons are possible.[335] This start codon-tRNA complex defines the beginning of the polypeptide as well as the coding region's reading frame. (3) The met-tRNA:mRNA:small ribosome subunit complex can now interact with a large ribosomal subunit to form the functional mRNA:ribosome complex. (4) Catalyzed by amino acid tRNA synthetases, charged amino acyl tRNAs will be present and can interact with the mRNA:ribosome complex to generate a polypeptide. Based on the mRNA sequence and the reading frame defined by the start codon, amino acids will be added sequentially. With each new amino acid added, the ribosome moves along the mRNA (or the mRNA moves through the ribosome). An important point, that we will return to when we consider the folding of polypeptides into their final three-dimensional shapes, is that the newly synthesized polypeptide is threaded through a molecular tunnel within the ribosome. Only after the N-terminal end of the polypeptide begins to emerge from this tunnel can the nascent polypeptide begin to fold. (5) The process of polypeptide polymerization continues until the ribosome reaches a stop codon, that is a UGA, UAA or UAG.[336] Since there are no tRNAs that recognize these codons, the ribosome pauses, waiting for a charged tRNA that will never arrive. Instead, a polypeptide known as release factor, with a shape something like a tRNA (→), binds to the polypeptide:mRNA:ribosome complex instead. (6) This leads to the release of the polypeptide, the disassembly





[334] Known as the Shine-Delgarno sequence for its discovers

[335] Hidden coding potential of eukaryotic genomes: nonAUG started ORFs: http://www.ncbi.nlm.nih.gov/pubmed/22804099

[336] In addition to the common 19 amino and 1 imino (proline) acids, the code can be used to insert two other amino acids selenocysteine and pyrrolysine. In the case of selenocysteine, the amino acid is encoded by a stop codon, UGA, that is in a particular context (surrounding nucleotide sequence) within the mRNA. Pyrrolysine is also encoded by a stop codon. In this case, a gene that encodes a special tRNA that recognizes the normal stop codon UAG is expressed. see Selenocysteine

of the ribosome into small and large subunits, and the release of the mRNA.[337]

When associated with the ribosome, the mRNA is protected against interactions with proteins (ribonucleases) that could catalyze its degradation into nucleotides. Upon its release from the ribosome, an mRNA may interact with a new small ribosome subunit, and begin the process of polypeptide synthesis again or it may interact with a ribonuclease and be degraded. Where it is important to limit the synthesis of particular polypeptides, the relative probabilities of these two events, new translation versus RNA degradation, will be skewed in favor of degradation. Typically an RNA's stability is regulated by the binding of specific proteins to nucleotide sequences within the mRNA. The relationship between mRNA synthesis and degradation will determine the half-life of a population of mRNA molecules, the steady state concentration of the mRNA in the cell, and indirectly, the level of the encoded polypeptide present.

-

Questions to answer:

142. Why so many tRNA genes? How, in basic terms, do different tRNAs differ from one another?
143. How might the concentration of various tRNAs and the frequency of various codons influence the rate of polypeptide synthesis?
144. What is the minimal number of different tRNA-amino acid synthetases in a cell?
145. Would you expect a ribosome to make mistakes in amino acid incorporation or polypeptide termination? How are such mistakes similar to and different from mutations?

Question to ponder:

- How might a ribosome shift its reading frame while translating an mRNA?

Effects of point mutations on polypeptides and proteins

Mutations in a gene's regulatory region can alter the gene's expression by regulating the frequency of transcription. Mutations in a gene's coding region generally do not influence transcription rate (unless of course regulatory regions are located within the coding region) but they can influence the sequence of the encoded polypeptide. We can define three types of mutations that involve changing a single base pair, known as a single nucleotide polymorphism or SNP (pronounced "snip"): synonymous, mis-sense, and non-sense mutations. Because of the semi-redundant nature of the genetic code, it is possible that a single nucleotide change in a coding region can have no effect on the amino acid encoded – this is referred to as a synonymous mutation. That said, different codons for the same amino acid can be recognized by different tRNAs, which are the products of different genes, and may be present at different concentrations in the cell. The efficiency of translation is influenced by the rate of aa-tRNA binding. Different organisms can differ in the codons they use to encode particular amino acids, a fact that leads to what is known as codon bias. Codon bias can influence the efficiency of mRNA translation. It can even lead to ribosome "stalling", if the tRNA needed is absence or present at low concentration. When genetically engineering the synthesis of a mRNA from one organism in another, translational efficiency can be significantly increased by altering the gene that encodes the mRNA so that it reflects the codon bias of the host, rather than the codon bias of the donor.

Another possibility is that the change of a single nucleotide in the coding region will change the amino acid encoded; this is known as a mis-sense mutation. The effect of a mis-sense mutation will depend upon where in the polypeptide it occurs and which amino acid is substituted. We can compare homologous polypeptides found in various organisms; regions that are similar in terms of amino acid sequence and structure are referred to as conserved regions, compared to regions that are more variable, known happily as variable regions.[338] A mis-sense mutation that replaces an amino acid in a conserved region of a polypeptide is likely to have a more drastic effect on the

---

[337] Interested in learning more, check out eukaryotic translation termination factor 1

[338] A polypeptide assumes a 3D-dimensional that shape can be conserved.

polypeptide's function than a similar change in a variable region. Similarly, a mutation that replaces a large hydrophobic amino acid with a acidic or basic, that is, highly hydrophilic amino acid, is more likely to perturb polypeptide structure and function than replacing a large hydrophobic amino acid with a smaller one. The final type of single nucleotide mutation that we consider here leads to the replacement of a codon that specifies an amino acid with a stop codon; it is known as a non-sense mutation. The result of a non-sense mutation is a truncated polypeptide. As a first guess, the effect of a non-sense mutation will be more severe the closer it is to the beginning of the coding region, compared to its effect near the end of the coding region – although other factors are likely to contribute to any particular mutation's effect.

Another type of mutation involves the deletion or addition of nucleotides to a sequence. Such insertions or deletions (known generically as indels) can disrupt or alter the binding of proteins to a gene's regulatory region, influencing gene expression. If they occur within the coding region, they can alter the reading frame that directs polypeptide synthesis. In particular, insertions or deletions that involve non-multiples of three (the lenght of a codon) in the coding region will change the reading frame of the mRNA, so that the sequence of the polypeptide downstream of the insertion site will be changed. In contrast, if the insertion/deletion involves a multiple of three nucleotides, there will be insertion or deletion of amino acids from the final polypeptide, but the normal sequence downstream of the altered region will stay the same.

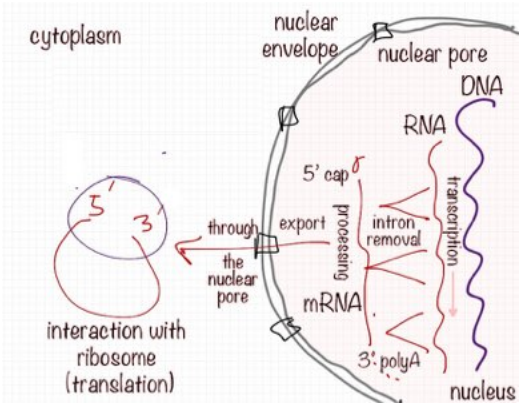
Questions to answer:

146. What do the terms "up-stream" and "down-stream" mean in terms of gene structure.
147. What effects on polypeptide synthesis arise from neglecting codon bias?
148. Why doesn't release factor cause the premature termination of translation at non-stop codons?
149. What might happen if a ribosome starts translating an mRNA at the "wrong" place?
150. When analyzing the effects of a particular non- or mis-sense mutation, what factors would you consider first?

Question to ponder:

- How would you go about reengineering an organism to incorporate non-biological amino acids into its proteins.

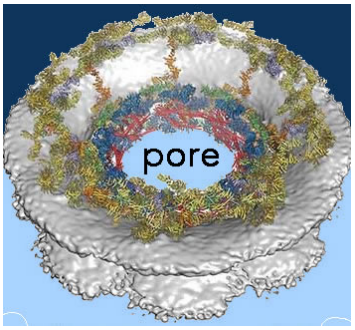
mRNA processing and nuclear export in eukaryotes

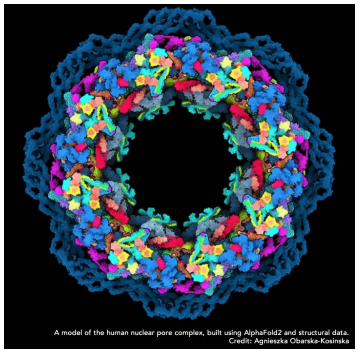


We will briefly reiterate a few points on how gene expression and polypeptide synthesis differ between prokaryotes and eukaryotes. The first and most obvious difference is the presence of a nucleus, a distinct domain within the eukaryotic cell that separates the cell's genetic material, its DNA, from the cytoplasm, where the ribosomes are located (→). Aside from those within mitochondria and chloroplasts, the DNA molecules of eukaryotic cells are located within the nucleus. The barrier between nuclear interior and cytoplasm is known as the nuclear envelope: no similar barrier exists between DNA and ribosomes in prokaryotes. In both bacteria and archaea the DNA is in direct contact with the cytoplasm. In eukaryotes, a newly synthesized mRNA molecule undergoes splicing (see below) and is modified (processed) at both its 5' and 3' ends. Only after RNA processing has occurred will the "mature" mRNA be exported out of the nucleus, through a nuclear pore, into the cytoplasm, where it can interact with ribosomes. Prokaryotic mRNAs are generally not processed.

The nuclear envelope complex (typically considered in greater detail in cell biology courses) consists of two lipid bilayer membranes punctuated by nuclear pores, which are macromolecular complexes (protein machines) of ~125,000,000 daltons. While molecules of molecular weight less than ~40,000 daltons can generally pass through the nuclear pore, larger molecules must be actively transported through a process coupled to a thermodynamically favorable reaction, in this case the

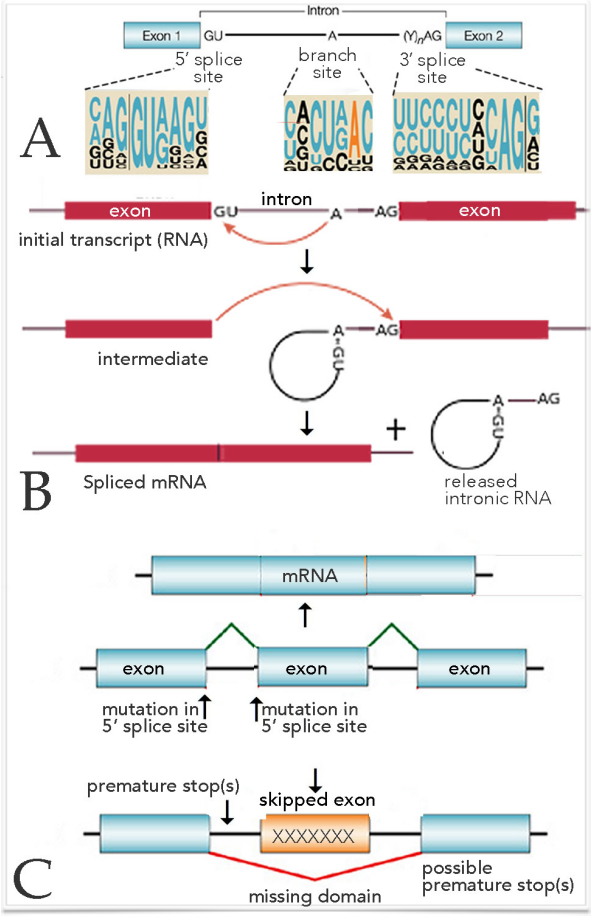
hydrolysis of guanosine triphosphate (GTP). The movement of larger molecules into and out of the nucleus through nuclear pores is regulated by what are known as nuclear localization and nuclear export sequences, located within polypeptides. These are recognized by proteins (receptors) associated with the pore complex (→). A protein with an active nuclear localization sequence (NLS) will be found in the nucleus while a protein with an active nuclear exclusion sequence (NES) will be found in the cytoplasm. By controlling NLS and NES activity a protein can come to accumulate, in a regulated manner, in either the nucleus or the cytoplasm, or can be present in both cellular regions. As we will see later on, the nuclear envelope breaks down during cell division (mitosis) in many but not all eukaryotes. Tears in the nuclear envelope have also been found to occur when migrating cells try to squeeze through small openings.[339] Once the integrity of the nuclear envelop is re-established, proteins with NLS and NES sequences are moved back to their appropriate location within the cell through active, that is energy driven, coupled reaction-based processes.





Mutations influencing splicing

While we ignore many details, a final class of point mutations are worth noting explicitly; these influence the "splicing" of a newly synthesized RNA molecule. Eukaryotic genes are generally broken up into coding regions, known as exons, and the non-coding regions between exons, known as intervening regions or introns. When a polypeptide-encoding gene is expressed, the RNA made, the initial transcript, contains both introns and exons. But ribosomes cannot distinguish between exon and intron sequences (probably one reason that prokaryotes do not have introns). In eukaryotes, introns are removed before the mature mRNA is exported across the nuclear envelope and into the cytoplasm, where the ribosomes are located. So the obvious question is, how exactly are introns recognized and removed, what mechanisms (molecular machines) are used? As you might already have guessed, there must be information, present in the sequence of the newly synthesized RNA that identifies the intronic sequences to be removed. There are nucleotide sequences that indicate the end of an exon and the start of an intron, known as the 5' splice site, and the end of an intron and the start of the next exon, known as the 3' splice site. Finally, there is information within the intron known as the branch site (A→). We can visualize this information through what are known as a "sequence logo" plot.[340] Such a plot indicates the information associated within a sequence; where there is no preference, that is, where any of the four nucleotides



[339] Tearing the nuclear envelope: http://www.sciencemag.org/news/2016/03/cells-can-do-twist-sometimes-their-nuclei-burst

[340] Sequence logos: a new way to display consensus sequences: http://www.ncbi.nlm.nih.gov/pubmed/2172928

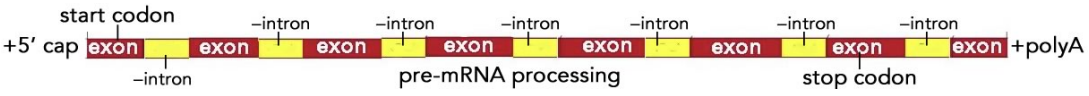
are acceptable, the information present at that site is 0. Where either of two nucleotides are acceptable, the information is 1, and where only one particular nucleotide is acceptable, the information content is 2.

Splicing is carried out by polypeptide-RNA complex known as the spliceosome. The spliceosome can recognize intron-exon boundary sequences and, using endonuclease and ligase activities, cut out the intron and join the 3' end of one exon to the 5' of the next (B→), releasing the intervening intron sequence in a looped form. A point mutations that disrupt the normal intron-exon boundary sequences (C→) can inhibit splicing, so that the intron remains in the final mRNA. Since introns do not encode polypeptides, there is no selection against the presence of stop codons in their sequence. A ribosome reading along a non-spliced RNA will likely add a series of inappropriate amino acids to the growing polypeptide, and is likely to encounter a stop codon, leading to premature termination of polypeptide synthesis. Alternatively if, for example a 3' splice site is disabled, a "down-stream" exon may be used for splicing; the result is that an exon normally included is lost from the spliced mRNA, the polypeptide sequence it encodes will be missing from the synthesized polypeptide, and it is possible that the down-stream reading frame will be wrong, leading to the synthesis of incorrect amino acid sequences and the creation of stop codons. The result is that mutations that disrupt splicing can have dramatic hypomorphic, anti-morphic, and possible neo-morphic effects, and such mutations (alleles) have been associated with a number of human diseases.[341]

The complexity of eukaryotic genomes is greatly increased by the fact that most genes contain multiple exons and introns; different sets of exons can be spliced together, a regulate-able process known as alternative splicing, in different cells and within a single cell to produce mRNA molecules that encode variants of the same polypeptide. These processes can lead to a range of complex behaviors that can muddy the interpretation of experimental manipulations.[342]

Non-sense mediated RNA decay

The truncated polypeptide generated by a non-sense mutation can produce phenotypic effects that are more severe than those associated with the failure to produce any polypeptide at all. To protect against the negative effects of non-sense mutations, particularly those that occur well "up-stream" of the normal stop codon, eukaryotic organisms have developed a defense mechanism known as non-sense mediated decay (NMD). In a typical gene, the "normal" stop codon is generally located within an exon located near the 3' end of newly synthesized "pre-mRNA". During mRNA processing, introns are recognized and removed by the splicing system (↓); the 5' end is "capped" and the 3' end processed and (generally) a stretch of A nucleotides, a "polyA tail", is added. Typically, all of these modifications are completed before the processed transcript, now an mRNA, is transported through the nuclear pore complex into the cytoplasm. The removal of an intron leads to the formation of an exon-exon junction (eej)(↓) and the association of an exon-exon junction protein complex (EJC) immediately "upstream" of each exon-exon junction.[343] When a ribosome engages with the 5' end of the mRNA and begins to move along mRNA during translation it displaces the EJCs, so what when the first ribosome reaches the end of the mRNA's coding region all of the EJCs have been removed (→). The stability of the EJ complex-free mRNA is regulated by signals located primarily in its 5' and 3' untranslated regions.

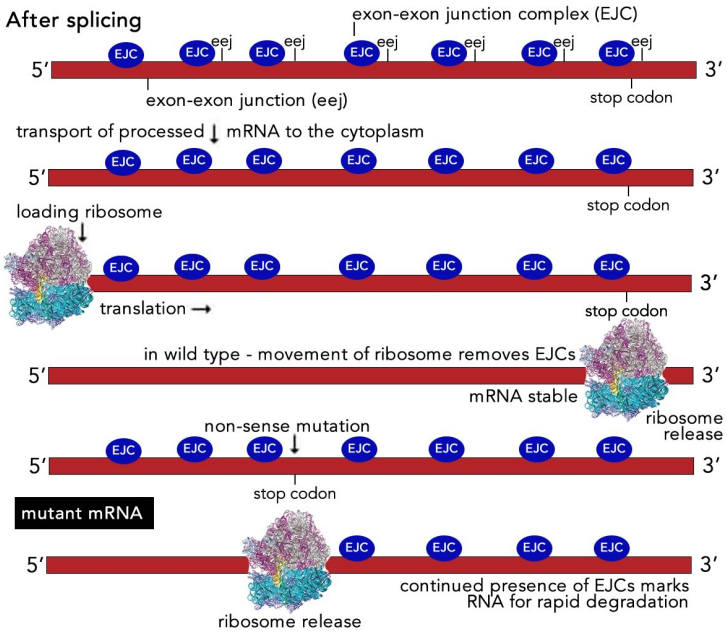


[341] The pathobiology of splicing: https://www.ncbi.nlm.nih.gov/pubmed/19918805

[342] See Biological plasticity rescues target activity in CRISPR knock outs

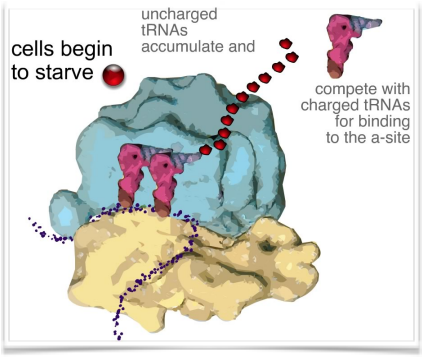
[343] The exon junction complex as a node of post-transcriptional networks

The situation is different when a non-sense mutation generates a stop codon within an upstream exon (→). The ribosome engages with the mRNA and continues until it reaches this stop codon, upon which release factor binds and ribosome disengages. All of the EJCs downstream of the mutation-generated stop codon remain associated with the mRNA. The failure to remove the EJCs marks the mRNA as aberrant and triggers the non-sense mediated decay (NMD) response.[344] NMD leads to the degradation of mRNAs containing out-of-context non-sense codon and dramatically reduces the synthesis of potentially toxic polypeptides. In a further weird twist, it has recently been reported that RNA fragments generated from the degraded mRNA re-enter the nucleus and regulate other genes - which can further complicate the already complicated relationship between mutation, genotype, and phenotype.[345]



Alarm generation

The translation system is a major consumer of energy within the cell.[346] When a cell is starving, it does not have the energy to generate amino acid charged tRNAs (→). The result is that uncharged tRNAs accumulate. Since uncharged tRNAs fit into the amino-acyl-tRNA binding sites on the ribosome, their presence increases the probability of unproductive tRNA interactions with the mRNA-ribosome complex, a situation that can lead to the premature termination of translation. When this occurs the stalled ribosome generates a signal (illustrated here: link) that can lead to adaptive changes in the cell that enable the cell to survive for long periods in a "dormant" state.[347]



Another response that can occur is a more social one. Some cells in the population can "sacrifice" themselves for their closely related neighbors (remember kin selection and inclusive fitness.) By shutting down mRNA synthesis (transcription) and RNA-dependent polypeptide synthesis (translation), a cell containing an addiction module can undergo what is known as programmed cell death. The mechanism is based on the fact that proteins (a toxin and an anti-toxin) can differ in the rates at which they are degraded within the cell. Just as ribonucleases can degrade mRNAs, proteases degrade proteins and polypeptides. How stable a protein/polypeptide is depends upon its structure, which we will be turning to soon, and more importantly the presence of proteases that degrade it. As discussed previously, interrupting protein synthesis leads to the rapid

---

[344] Mechanism and regulation of the nonsense-mediated decay pathway

[345] Wilkinson, M. F. (2019). Genetic paradox explained by nonsense,

[346] Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources

[347] Characterization of the Starvation-Survival Response of Staphylococcus aureus:

disappearance (turn-over) of the anti-toxin while the toxin persists, leading to cell death, which in turn leads to the release of the cell's nutrients, nutrients that can be used by its neighbors, in part to maintain active gene expression and protein synthesis. Of course, sacrificing for ones neighbors makes evolutionary sense only if one has neighbors and those neighbors are close relatives.

Questions to answer:

151. A gene has many introns - provide a model for how it might encode functionally distinct polypeptides.
152. How can a mutation in splice site sequence influence gene expression and protein function?
153. How does non-sense mediated decay (NMD) protect against potentially deleterious mutations (alleles)?
154. Why would a cell want to stop (rather than continue) polypeptide synthesis when it is starving?

Question to ponder:

– How might the presence on uncharged tRNA lead to the termination of translation?

Turning polypeptides into proteins

A protein is a functional entity, typically composed of one or more polypeptides.[348] These polypeptides can be the same or different, that is encoded by different genes. Polypeptides are synthesize in a linear manner. In contrast to a linear DNA molecule, a polypeptide is a three dimensional object, and it comes to fold into its three dimensional shape as it is synthesized. In a protein composed of multiple polypeptides, these polypeptides must interact with one another and assume a functional conformation, the protein's structure. When we think about how a polypeptide folds, we have to think about the directionality of synthesis, the environment that the newly synthesized polypeptide comes to inhabit, and how it interacts with itself and with other polypeptides. In the case of a protein composed of multiple polypeptides (subunits), each is synthesized independently, so we have to consider how these polypeptides come to interact with one another, and avoid "inappropriate" interactions.

As we think about protein structure it is common to see the terms primary, secondary, tertiary, and quaternary structure (video link). The primary structure of a polypeptide is the sequence of amino acids along the polypeptide chain, written from its N- or amino terminus to its C- or carboxyl terminus. The secondary structure of a polypeptide consists of local folding motifs: the α-helix, the β-sheet, and connecting domains. The tertiary structure of a polypeptide is the overall three dimensional shape a polypeptide takes in space, as well as how its R-chains are oriented. Quaternary structure refers to how the various polypeptides and co-factors combine and are arranged to form a functional protein (remember the distinction between polypeptide and protein). In a protein that consists of a single polypeptide and no co-factors, tertiary and quaternary structures are the same. As a final complexity, a particular polypeptide can be part of a number of different proteins – the universe of proteins that a polypeptide is a part of could be considered another level of structure. Some of these interactions are relatively stable, others more ephemeral and regulative. This is one way in which a gene can play a role in a number of different processes and be involved in the generation of a number of different phenotypes.

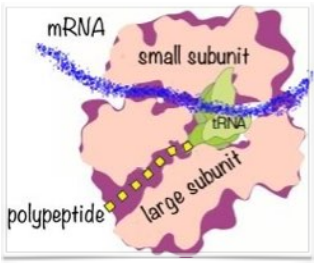
Polypeptide synthesis (translation), like most all processes that occur within cells, is a stochastic process, meaning that it is based on random collisions between molecules. In the specific case of translation, the association of the mRNA with ribosomal components occurs stochastically. Given that a human cell contains ~24,000 genes that can generate mRNAs and millions of ribosomes, most RNAs find a ribosome. Similarly, the addition of a new amino acid to the end of a growing polypeptide depends on a productive collision between the appropriate amino acid-charged tRNA and the RNA-ribosome complex. Since there are many different amino-acid charged tRNAs in the cytoplasm, the ribosomal complex will bind only the amino-acyl-tRNA that the mRNA specifies, that is the tRNA with the right anticodon. This enables its attached amino acid to interact productively, leading to the addition of the amino acid to C-terminus of the growing polypeptide chain. You rarely

[348] see also: When is a gene product a protein when is it a polypeptide?

see this fact illustrated in most presentations of polypeptide synthesis. In bacterial cells from 10 to 20 amino acids are added to the end of a growing polypeptide chain per second, the rate is about half that in mammalian cells.[349]

Now you might wonder whether there are errors in polypeptide synthesis as there are in nucleic acid synthesis. In fact there are! Such translation errors can lead to an in-frame stop codon that terminates translation and the release of an aberrant polypeptide that is (generally) rapidly degraded.[350] There are also cases that are "programmed" such that at certain positions along an mRNA the ribosome can "slip back" one nucleotide (a –1 frameshift) or skip one nucleotide (a +1 frameshift), leading to a different sequence of amino acids added from the point of the frameshift to the end of the polypeptide.[351] Similarly, if the wrong amino acid is inserted at a particular position and it disrupts normal folding, the polypeptide may disrupt normal cellular functions. There are molecular machines that recognize mis-folded proteins and mark them for degradation. What limits the effects of mistakes made during translation is that most proteins (unlike DNA molecules) have finite and relatively short half-lives; that is, the time an average polypeptide exists before it is degraded by various enzymes. Normally this limits the damage that a mis-translated polypeptide can do to the cell and organism.

Factors influencing polypeptide folding and structure



Polypeptides are synthesized, and they fold, in a vectorial, that is, directional manner. Synthesis occurs in an N- to C- terminal direction and the newly synthesized polypeptide exits the ribosome through a ~10 nm long and ~1.5 nm diameter tunnel (→). This tunnel is narrow enough to block the folding of the newly synthesized polypeptide chain. As the polypeptide emerges from the tunnel it begins to fold (video link). At the same time it encounters the crowded cytoplasmic environment; the newly synthesized polypeptide needs to avoid low affinity, non-specific, and non-functional interactions with other cellular components.[352] If the polypeptide is part of a multi-subunit protein, once synthesis is complete it must "find" its correct partner polypeptides, which again is a stochastic process. If the polypeptide does not fold correctly, it will not function correctly and may even damage the cell or the organism. A number of degenerative neurological disorders appear to be due, at least in part, to the accumulation of mis-folded polypeptides (see below).

We can think of the folding process as a “drunken” walk across an energy landscape, with movements driven by intermolecular interactions and collisions with other molecules. The goal of this process is to find the lowest point in the landscape, the energy minimum of the system. This is generally assumed to be the native or functional state of the polypeptide. That said, this native state is not necessarily static, since the folded polypeptide (and the final protein) will be subject to thermal fluctuations (collisions with neighboring molecules). It is possible that it will move between various states with similar, but not identical stabilities.[353] The challenge to calculating the final folded state of a polypeptide is that it is a complex computational problem. Generally two approaches are taken to characterize the structure of a functional protein. In the first the structure of the protein is determined directly by X-ray crystallography, cryo-electron microscopy, or Nuclear Magnetic Resonance (NMR) spectroscopy (which, as you will notice, we are not going to explain here, but which you may

---

[349] see http://bionumbers.hms.harvard.edu/default.aspx

[350] Quality control by the ribosome following peptide bond formation

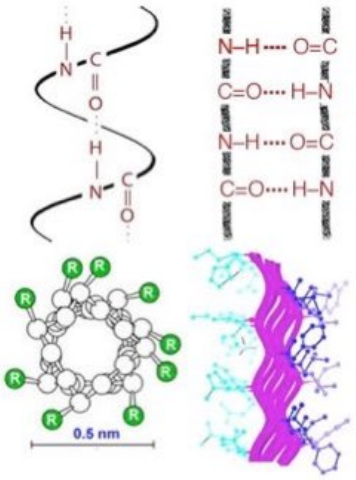
[351] Ketteler 2012. On programmed ribosomal frameshifting: the alternative proteomes

[352] Remember, all molecules interact with each other via LDF-mediate interactions.

[353] folding video: from YOUTUBE - Stoneybrook: https://youtu.be/YANAso8Jxrk

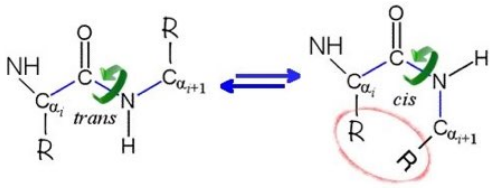
encounter in a chemistry or a biophysics class). In the second, if the structure of a homologous (evolutionarily-related) protein is known, it can be used as a framework to model the structure of a previously unsolved protein. There are a number of on-line tools to generate such structural models.

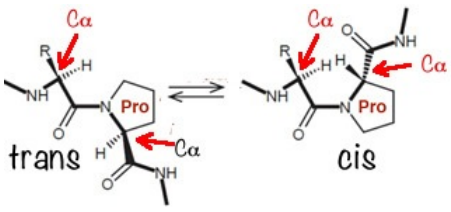
A number of constraints influence the folding of a polypeptide. The first is the peptide bond itself. All polypeptides consist of a string of peptide bonds. It is therefore not surprising that there are common patterns in polypeptide folding. The first of these common patterns to be recognized, the α-helix (left →), was discovered by Linus Pauling (1901-1994) and Robert Corey (1897-1971) in 1951. This was followed shortly thereafter by their description of the β-sheet (right →). The forces that drive the formation of the α-helix and the β-sheet will be familiar, they are the same forces that underlie water structure, namely H-bonding interactions.



In an α-helix and a β-sheet, all of the possible H-bonds involving the peptide bond's donor and acceptor groups (–N–H and O=C– , with "..." indicating a H-bond) are formed within the polypeptide. In an α-helix these H-bond interactions run parallel to the polypeptide chain. In the β-sheet, these H-bonding interactions occur between polypeptide chains. The interacting strands within a β-sheet can run parallel or anti-parallel to one another, and can occur within a single polypeptide chain, folded back on itself in various ways, or between different polypeptide chains. In an α-helix, the R-groups point outward from the helix axis. In β-sheets the R-groups point in an alternating manner either above or below the plane of the sheet. While all amino acids can take part in either α-helix or β-sheet structures, the imino acid proline cannot - the N-group coming off the α-carbon has no H, so its presence in a polypeptide chain leads to a break in the pattern of intrachain H-bonds. It is worth noting that some polypeptides can adopt functionally different structures: for example in one form (PrPC) the prion protein contains a high level of α-helix (~42%) and essentially no β-sheet (~3%), while an alternative form, (PrPSc) associated with the disease scrapie, contains high levels of β-sheet (~43%) and ~30% α-helix.[354] The result is two very different 3-dimensional protein structures, even though the primary sequences of the two are identical.

Peptide bond rotation and proline: Although typically drawn as a single bond, the peptide bond behaves more like a double bond, or rather like a bond and a half. In the case of a single bond, there is free rotation around the bond axis in response to molecular collisions. In contrast, rotation around a peptide bond requires more energy to move from the trans to the cis configuration and back again (→). It is more difficult to rotate around the peptide bond because it involves the partial breakage of the bond. In addition, in the cis configuration the R groups of adjacent amino acids are on the same side of the polypeptide chain. If both R groups are large they can bump into each other. If they get too close they will repel each other. The result is that usually the polypeptide chain will be in the trans arrangement. In both α-helix and β-sheet configurations, the peptide bonds are in the trans configuration because the cis configuration disrupts their regular organization.





Peptide bonds involving a proline residue have a different problem. The amino group is "locked" into a particular shape by the ring and therefore inherently destabilizes both α-helix and β-sheet structures (see above). In addition, peptide bonds involving prolines (←) are found in the cis configuration ~100 times as often as those between other amino acids. This cis configuration leads to a bend or kink in the polypeptide chain. The energy

[354] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC47901/ and prion disease: https://en.wikipedia.org/wiki/Prion

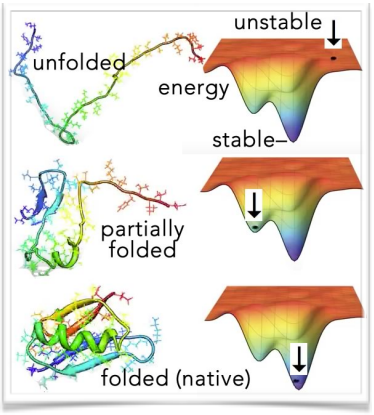
involved in the rotation around a peptide bond involving a proline is much higher than that of a standard peptide bond; so high, in fact, that there are protein catalysts, peptidyl proline isomerases such as PIN1 (OMIM:601052), that facilitate the cis-trans rotation.

Hydrophobic R-groups: Many polypeptides and proteins exist primarily in an aqueous (water-based) environment. Yet, a number of their amino acid R-groups are hydrophobic. That means that their interactions with water will decrease the entropy of the system by leading to the organization of water molecules around the hydrophobic group, a thermodynamically unfavorable situation. This is very much like the process that drives the assembly of lipids into micelles and bilayers. A typical polypeptide, with large hydrophobic R groups along its length will, in aqueous solution, tend to collapse onto itself so as to minimize, although not always completely eliminate, the interactions of its hydrophobic residues with water. In practice this means that the first step in the folding of a newly synthesized polypeptide after it leaves the ribosomal tunnel is its collapse onto itself so that the majority of its hydrophobic R groups are located internally, out of contact with water. In contrast, where there are no (or few) hydrophobic R groups in the polypeptide, the polypeptide will tend to adopt an extended configuration. On the other hand, if a protein comes to be embedded within a membrane (considered later on), then the hydrophobic R-groups will tend to be located on the surface of the folded polypeptide that interacts with the hydrophobic interior of the lipid bilayer. Hopefully this makes sense to you, thermodynamically.

Acidic and basic R-groups: Some amino acid R-groups contain carboxylic acid or amino groups and so act as weak acids or bases, respectively. Depending on the pH of their environment these groups may be uncharged, positively charged, or negatively charged. Whether a group is charged or uncharged can have a dramatic effect on the structure, and therefore the activity, of a protein. By regulating pH in specific cellular compartments, an organism can modulate the activity of specific proteins. There are, in fact, compartments within eukaryotic cells that are maintained at low pH in part to influence protein structure and activity. As an example, the internal regions of the vesicles associated with endocytosis become acidic (through the ATP-dependent pumping of H^+ ions across their membranes), which in turn activates a number of enzymes located within the vesicle, these enzymes mediate the hydrolytic breakdown of proteins, nucleic acids, and other compounds.

Subunits and prosthetic groups: Many proteins contain non-amino acid-based components, known generically as co-factors. A protein minus its cofactors is known as an apoprotein. Together with its cofactors, it is known as a holoprotein. Generally, without its cofactors, a protein is inactive and often unstable. Cofactors can range in complexity from a single metal ion to complex molecules, such as vitamin B12. The retinal group of bacteriorhodopsin and the heme group (with its central iron ion) are co-factors. In general, co-factors are synthesized by various anabolic pathways, and so they depend on the activities of a number of genes. A functional protein can therefor be the direct product of a single gene, many genes, or (indirectly) entire metabolic pathways.

Chaperones



The path to the native, that is, stable, functional state is not necessarily a smooth or predetermined one. The folding polypeptide can get "stuck" in a local energy minimum; there may not be enough energy, derived from thermal collisions, for it to get out again. If a polypeptide gets stuck, structurally, there are active mechanisms to unfold it and let the process leading to the native state proceed again (→). The process of unfolding misfolded polypeptides is carried out by proteins known as chaperones; we will call them folding/re-folding chaperones to distinguish them from other types of chaperones. Chaperones are protein-based molecular machines that are encoded

by other genes. The unfolding of a misfolded protein by a chaperone requires energy, and so is coupled to a thermodynamically favorable reaction, such as ATP hydrolysis.

An important point to recognize is that chaperones do not determine the native state of a polypeptide–that is a function of the polypeptide's primary amino acid sequence. Rather, they suppress the probability of misfolded alternative structures. Consider, for example, the effect of a mis-sense mutation. Such a mutation can change the pattern of folding of a polypeptide; it may get caught more frequently in a mis-folded form. A folding/refolding chaperone can recognize such a mis-folded polypeptide, unfold it, either totally or partially, and release it to refold again, enabling the polypeptide to reach a functional structure, even in the presence of a destabilizing mutation.

There are many types of protein chaperones; some interact with specific polypeptides as they are synthesized and attempt to keep them from getting into trouble, that is, folding in an unproductive way. Others can recognize inappropriately folded polypeptides and, through coupling to ATP hydrolysis, catalyze the unfolding of the polypeptide, allowing the polypeptide a second (or third or …) chance to fold correctly. In the "simple" eukaryote, the yeast *Saccharomyces cerevisiae*, at least 63 distinct molecular chaperones have been recognized.[355]

Now you may ask yourself, if most proteins are composed of multiple polypeptides but polypeptides are synthesized individually, how do polypeptides come to be correctly assembled into functional proteins in a cytoplasm crowded with other proteins and molecules? Protein assembly often involves specific "assembly" chaperones, that bind to a newly synthesized polypeptide and either stabilize their folding, or hold them until they interact with other polypeptides to form the final, functional protein.[356] When proteins are synthesized *in vitro*, the absence of appropriate chaperones can make it difficult to assemble multi-subunit proteins into functional proteins.

Another class of chaperones are known as "heat shock proteins." The genes that encode these proteins are expressed in response to increased temperature (and other stressors), assuming that the temperature increase does not kill the cell or organism immediately. At these higher temperatures collisions with surrounding molecules can lead a protein to unfold and misfold, the protein can become "denatured". Once expressed, heat shock proteins recognize denatured polypeptides, couple ATP hydrolysis reactions to unfold them, and then release the unfolded protein, giving them another chance to refold correctly.

Heat shock proteins help an organism adapt.[357] In classic experiments, when bacteria were grown at temperatures sufficient to activate the expression of the genes that encode heat shock proteins, the bacteria had a higher survival rate when re-exposed to elevated temperatures compared to bacteria that had been grown continuously at lower temperature. Heat shock response-mediated survival at higher temperatures is an example of the ability of an organism to adapt to its environment - it is a physiological response. The presence of the heat shock system itself, however, is a selectable trait, encouraged by temperature variation in the environment. It is the result of evolutionary factors.

By now you might be asking yourself, how do chaperones recognize unfolded or abnormally folded proteins? In the case of a water soluble protein, most of their hydrophobic R-groups will be found within the interior of the correctly folded protein. In contrast, an unfolded protein will tend to have hydrophobic amino acid side chains exposed on its surface. The presence of these surface hydrophobic residues will lead to a tendency to aggregate; interacting hydrophobic regions will minimize hydrophobic-water interactions. Chaperones for water-soluble proteins recognize and interact with surface hydrophobic regions. For assembly chaperones, we can expect that specific sequences or structures in the target protein are recognized, which presumably is one reason that there are so many chaperone-like proteins, and specific chaperones for specific polypeptides and

---

[355] An atlas of chaperone–protein interactions in Saccharomyces cerevisiae: implications to protein folding pathways

[356] Assembly chaperones: a perspective: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3638391/

[357] The heat shock response: life on the verge of death

proteins.

Questions to answer

155. Why does it matter that rotation around a peptide bond is constrained?
156. How can changing the pH of a solution alter a protein's structure and activity?
157. Make models of polypeptides all of whose R-groups are hydrophilic or hydrophobic?
158. How might the presence of a folding/refolding-chaperone mitigate the effects of a mis-sense mutation?
159. How do assembly-chaperones facilitate the assembly of multi-polypeptide proteins?
160. Under what conditions might you expect heat shock proteins to be unnecessary for an organism?

Questions to ponder

- How does entropy drive protein folding and assembly?
- How might surface hydrophobic R-groups facilitate protein-protein interactions.
- How many ways can you imagine that the absence of a polypeptide/protein will influence the phenotype of an organism, consider a polypeptide that interacts with a number of other polypeptides (proteins).
- Develop a plausible model for how the expression of heat shock genes is regulated in response to temperature.

Regulating protein activity, concentrations and stability (half-life)

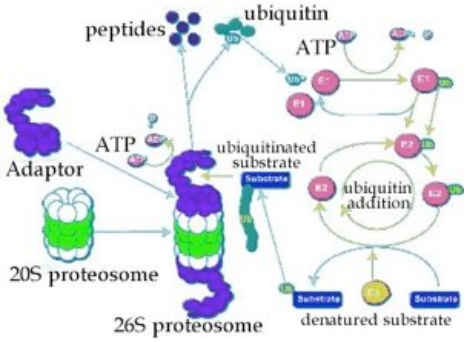
Proteins act through their interactions with other molecules. Catalytic proteins (enzymes) interact with substrate molecules; these interactions lower the activation energy of the reaction's rate limiting step, leading to an increase in the overall reaction rate. At the same time, cells and organisms are not static. They must regulate which proteins they produce, the final concentrations of those proteins within the cell or organism, how active those proteins are, and where those proteins are located. It is primarily by altering proteins, which in turn influences gene expression, that cells and organisms adapt to changes in their environment.

A protein's activity can be regulated in a number of ways. The first and most obvious is to control the total number of protein molecules present within the system. Let us assume that once synthesized a protein is fully active. With this simplifying assumption, the total concentration of a protein, and the total protein activity in a system [P_{sys}] is proportional to the rate of that protein's synthesis (dSynthesis/dt) minus the rate of that protein's degradation (dDegradation/dt), with dt indicating per unit time. The combination of these two processes, synthesis and degradation, determines the protein's concentration in the cell. Both the rate of protein's synthesis and degradation can be regulated. These processes can influence the rate at which a cell (or organism) can respond to various perturbations.

The degradation of proteins is mediated by a special class of enzymes known as proteases. Proteases cleave peptide bonds via hydrolysis (adding water) reactions. Proteases that cleave a polypeptide chain internally are known as endoproteases - they generate two polypeptides. Those that hydrolyze polypeptides from one end or the other, generally release one or two amino acids at a time, and are known as exoproteases. Proteases can also act more specifically, recognizing and removing specific parts of a protein in order to activate or inactivate it, or to control where it is found in a cell. For example, nuclear proteins become localized to the nucleus (typically) because they contain a NLS or they can be excluded because they contain an NES (see above). For these sequences to work they have to be able to interact with the transport machinery associated with nuclear pores; but the protein may be folded so that the NLS/NES sequences are hidden. Changes in a protein's structure can reveal or hide such sequences, thereby altering the protein's distribution within the cell and therefore its activity. As an example, a transcription factor located in the cytoplasm is, in terms of its effects on gene expression, inactive; it can become active if it enters the nucleus. Similarly, many proteins are originally synthesized in a longer and inactive "pro-form". When the pro-peptide is removed, cut away by an endoprotease, the processed protein becomes active. Proteolytic processing is itself often regulated.

The amount of a protein within a cell or organism is a function of the number of mRNAs encoding the protein, the rate that these mRNAs are recognized and translated, the rate at which functional protein is formed, which in turn depends upon folding rates and their efficiency. Generally once translation begins it continues at a more or less constant rate until a stop codon is reached. In the bacterium *E. coli*, the rate of translation at 37ºC is ~15 amino acids per second.[358] The translation of a polypeptide of 1500 amino acids therefore takes about 100 seconds. After translation, folding and, in multi-subunit proteins, assembly, the protein will function, assuming that it is active, until it is degraded.

In the case of both mRNAs and proteins, the breakdown process is stochastic, based on collisions with the degradative machinery. While the probability that a molecule is degraded can be measured, how long any particular molecule persists (that is, the time from its synthesis to its degradation) can not be predicted accurately. Degradation can be regulated, signals within or added to a molecule can influence whether a collision with a degrading complex will be productive, that is, whether the molecule is broken down. Protein degradation is particularly important for controlling the levels of "regulated" proteins, whose presence (or concentration) within the cell may lead to unwanted effects. The rate of molecular degradation can be regulated, generally through the presence or addition of a signal that serves to influence the outcome of collisions with the degradative machinery (→). Degradation is an active and highly regulated process, involving ATP hydrolysis and multi-subunit complexes. One of these, involved in proteins degradation, is known as the proteasome. The proteasome degrades the polypeptide into small peptides and amino acids that can be reused. As a mechanism for regulating protein activity, however, degradation has a serious drawback, it is irreversible.



Allosteric and post-translational regulation

Allosteric regulation is a reversible way to control a protein's activity; a regulatory molecule binds reversibly to the protein altering the protein's structure, its activity, its location within the cell, and/or its stability. When an allosteric effector binds to a protein, it interact through van der Waals interactions - it is not covalently bound to the protein. Such interactions are reversible, influenced by thermal factors. Allosteric regulators can act either positively or negatively. The nature of such factors is broad, they can be a small molecule or another protein. What is important is that the allosteric binding site is distinct from the enzyme's catalytic site. In fact allosteric means "other site". Because allosteric regulators do not bind to the same site on the protein as the substrate, changing substrate concentration generally does not alter their effects.

Of course there are other types of regulation as well. A molecule may bind to and block the active site of an enzyme. If this binding is reversible, then increasing the amount of substrate can over-come the inhibition. An inhibitor of this type is known as a competitive inhibitor. Increasing the substrate concentration can overcome inhibition. In other cases, the inhibitor reacts with the enzyme, forming a covalent bond. This type of inhibitor is essentially irreversible; so increasing substrate concentration does not overcome inhibition. These are therefore known as non-competitive inhibitors. Allosteric effectors are also non-competitive, since they do not compete with substrate for binding to the active site. That said, binding of substrate could, in theory, change the affinity of the protein for its allosteric effectors, just as binding of the allosteric effector changes the binding affinity of the protein for the substrate.

---

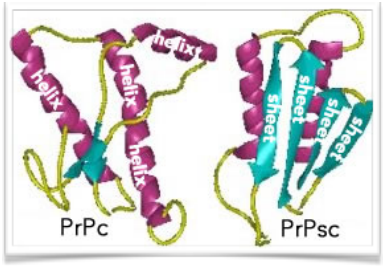
[358] We are going to totally ignore the fact that different tRNAs are present at difference concentrations, which gives rise to what is known as codon bias. The presence of codons recognized by rare tRNAs slows down translation. To learn more look at Codon Bias as a Means to Fine-Tune Gene Expression: https://www.ncbi.nlm.nih.gov/pubmed/26186290

Proteins may be modified, through various covalent-modifications, after their synthesis, folding, and assembly - this process is known as post-translational modification. A number of different types of post-translational modifications have been found to occur within cells. Here we consider post-translational modification only generically. In general they involve the formation of a covalent bond linking a specific chemical group to specific amino acid side chains in the protein - these groups can range from a phosphate group (phosphorylation), an acetate group (acetylation), the attachment of lipid/hydrophobic groups (lipid modification), carbohydrates (glycosylation) and others. In general where a protein can be modified that modification can be reversed, except, of course, when the modification involves protein degradation or proteolytic processing. One type of enzyme catalyzes the addition of the modifying group while another type of enzyme catalyzes its removal. For example, proteins are phosphorylated by enzymes known as protein kinases, while protein phosphotases remove phosphate groups from proteins. Post-translational modifications act in much the same way as do allosteric effectors, they modify the structure and, in turn, the activity of the polypeptide or protein modified. They can also modify a protein's interactions with other proteins, the protein's localization within the cell, and its stability.

Diseases of folding and misfolding

If a functional protein is in its native (or natural) state, a dysfunctional mis-folded protein is said to be denatured. It does not take much of a perturbation to unfold or denature many proteins. In fact, under normal conditions, proteins often become partially denatured spontaneously, normally these are either refolded, often with the help of chaperones or degraded through the action of proteases. A number of diseases, however, arise from irreversible protein mis-folding.

Kuru was among the first of these protein mis-folding diseases to be identified. Beginning in the 1950s, D. Carleton Gadjusek (1923–2008)[359] studied a neurological disorder common among the Fore people of New Guinea. The symptoms of kuru, which means "trembling with fear", are similar to those of scrapie, a disease of sheep, and variant Creutzfeld-Jakob disease (vCJD) in humans. Among the Fore people, Kuru was linked to the ritual eating of the dead. Since this practice has ended (we are told), the disease has disappeared. The cause of kuru, scrapie, and vCJD appears to be the presence of an abnormal form of a normal protein, known as a prion (mentioned above). We can think of prions as a type of anti-chaperone. The idea of proteins as infectious agents was championed by Stan Prusiner (b. 1942), who was awarded the Nobel Prize in Medicine in 1997.[360]



The protein (PrPc) responsible for Kuru and Scrapie is encoded by the PRP gene (OMIM:176640). It normally exists in a largely α-helical form. There is a second, abnormal form of the protein, PrPsc (the "sc" indicates scrapie); its structure contains a high level of β-sheet (→). The two polypeptides have the same primary sequence. PrPsc acts to catalyze the transformation of PrPc into PrPsc. Once initiated, this leads to a chain reaction and the accumulation of PrPsc. As it accumulates PrPsc assembles into rod-shaped aggregates that appear to damage cells. When this process occurs within the cells of the central nervous system it leads to neuronal cell death, dysfunction, and severe neurological defects. There is no natural defense, since the protein responsible is a normal protein.

When the Fore ate the brains of their beloved ancestors, they inadvertently introduced PrPsc protein into their bodies. Genetic studies indicate that early humans evolved resistance to prion diseases, suggesting that cannibalism might have been an important selective factor during human evolution. Since cannibalism is reasonably uncommon today, how does one get such diseases in the

---

[359] Carleton Gajdusek: http://www.theguardian.com/science/2009/feb/25/carleton-gajdusek-obituary

[360] Stanley Prusiner: 'A Nobel prize doesn't wipe the skepticism away' & http://youtu.be/yzDQ8WgFB_U

modern world? There are rare cases of iatrogenic transmission, that is, where the disease is caused by faulty medical practice, for example through the use of contaminated surgical instruments or when diseased tissue is used for transplantation.

But where did people get the disease originally? Since the disease is caused by the formation of PrPsc, any event that leads to PrPsc formation could cause the disease. Normally, the formation of PrPsc from PrPc occurs only rarely. We all have PrPc but very few of us spontaneously develop Kuru-like symptoms. There are, however, mutations in *PRP* gene that greatly increase the frequency of the PrPc → PrPsc conversion event. Such mutations may be inherited (genetic) or may occur during the life of an organism (sporadic). Fatal familial insomnia (FFI)(OMIM:600072) is due to the inheritance of a mutation in the *PRP* gene, a mutation that replaces the aspartic acid normally found at position 178 of the PrPc protein with an asparagine. When combined with a second mutation in the PRP gene at position 129, the FFI mutation leads to Creutzfeld-Jacob disease (CJD).[361] If one were to eat the brain of a person with FFI or CJD, one might well develop a prion disease.

So why do PrPsc aggregates accumulate? To cut a peptide bond, a protease (an enzyme that cuts peptide bonds) must position the target peptide bond within its catalytically active site. If the target protein's peptide bonds do not fit into the active site, they cannot be cut. Because of their structure, PrPsc aggregates are highly resistant to proteolysis. They gradually accumulate over many years, a fact that may explain the late onset of PrP-based diseases.

Questions to answer

161. A protein binds an allosteric regulator - what might happen to the protein?
162. How is the post-translational modification of a protein analogous to allosteric regulation? how is it different?
163. Assuming that synthesis rate decreases by 50% what happens to steady state polypeptide concentration? What happens if degradation rate increases by 50%? Generate predictive graphs of these (and other) possibilities.
164. How is the proteolytic processing of a polypeptide like and unlike an allosteric effector or a post-translational modification.
165. Why do post-translational modifications (and their reversals) require energy?
166. How might a mutation that alters a signal sequence influence the translation, assembly, localization, and function of a polypeptide (protein)?

Questions to ponder

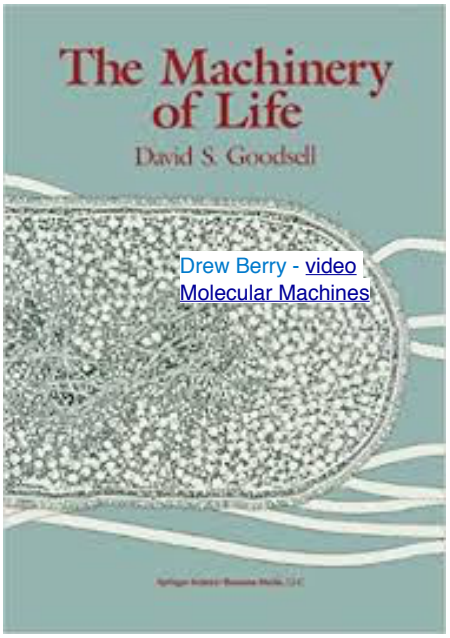
- Why is a negative allosteric regulator not considered a "competitive" inhibitor?
- How might the concentration of an allosteric effector influence the activity of the target protein?
- How would a cell recover from the effects of exposure to an irreversible, non-competitive inhibitor?
- In terms of energy used, explain the advantages of allosteric and post-translational modification based regulation compared to protein degradation.

Molecular machines

Polypeptides and the proteins and macromolecular complexes they form are what we might reasonably refer to as molecular machines. Essentially every process within a cell or an organism is mediated by some sort of molecular machine. When we think about these molecular machines it is important to consider how they find their site of action, and how they carry out their function(s) - their molecular mechanism(s) of action. Molecules cannot see, they can only "feel" - that is, they can bind to specific targets with various levels of specificity and stability through inter-molecular interactions. We see this type of interaction in the ability of chaperone proteins to recognize and unfold misfolded proteins, the binding of proteins involved in the replication of DNA and transcription of genes, and the binding and post-translational modification of proteins by various enzymes. Other types molecular machines (which we only briefly mention) are involved in various cellular movements (cellular swimming driven by flagella and cilia, cellular contractions based on the actin-

[361] OMIM entry for Creutzfeld-Jacob diease: http://omim.org/entry/123400

myosin system, and the movements of chromosomes based on motor molecules walking along cytoplasmic polymers - microtubules). Because machines, even molecular machines, have to "do" things, make things happen (repair damaged DNA, move chromosomes, form ATP), they require energy, energy that is supplied by coupling to thermodynamically favorable chemical reactions (or the absorption of light). Also, much like macroscopic machines, molecular machines often need to be turned on and off. The DNA replication and transcription machines have to work were and when they are needed. Both post-translational modifications, allosteric effectors, and target-recognition binding interactions play a role in when and where molecular machines act and are not active. At the same time, and something rarely illustrated in fancy video animations, the stochastic nature of molecular machines (driven by thermal interactions) is often ignored but since we have stressed it, you may consider how it will influence such animations. Remembering the machine nature of proteins and other macromolecular complexes (e.g. the ribosome and the nuclear pore) can be useful when considering the effects of mutations and allelic variants.

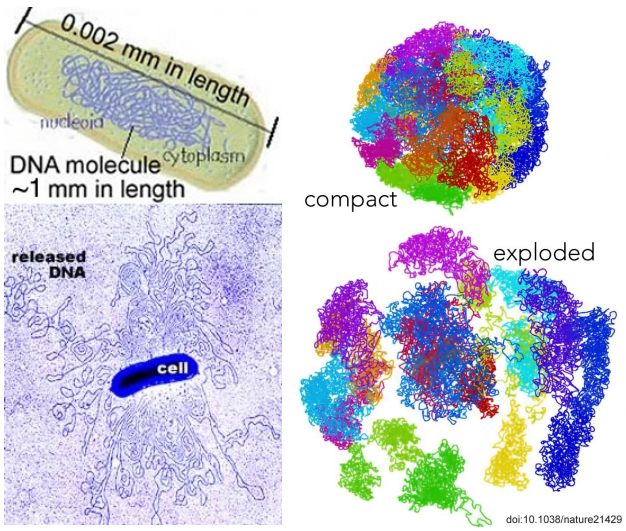


Drew Berry - video
Molecular Machines

Chapter 9: Organizing and expressing genes in regulatory networks

In which we consider how DNA molecules, and the genes that reside within them, are organized, how genes are recognized, and how their expression is controlled and organized into regulatory networks.

An important part of our approach to the study of biology is to think concretely about the molecules we are considering. Nowhere is this more important than with DNA. DNA molecules are very long and cells, even the largest cells, are (generally) small. For example, a typical bacterium is roughly cylindrical and ~2 μm in length and ~1 μm in circumference. Based on the structure of DNA, each base pair is ~0.34 nm in length. A region of DNA that is 1000 (10^3) base pairs long is therefore ~0.34 μm in length. A bacterium, like *E. coli*, has ~3 x 10^6 base pairs of DNA – that's a DNA molecule almost a millimeter in length or about 500 times the length of the cell in which it finds itself. That implies that at the very least the DNA has to be folded back on itself many times (←). A human cell has ~6000 times more DNA, resulting in a total length of greater than 2 meters of DNA per cell; these DNA molecules have to fit into a nucleus that is typically ~10 μm in diameter. In both cases, the DNA has to be folded and packaged in ways that allow it to fit within the cell and yet still be accessible to the various proteins involved in the regulation of gene expression and DNA replication. To accomplish this, the DNA molecule is associated with specific proteins; the resulting DNA:protein complex is known as chromatin.

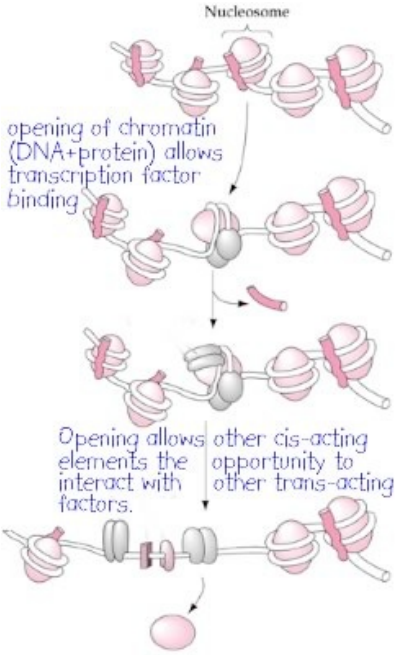


Left: A diagram of a bacterial cell showing its DNA molecule; disrupting the cell membrane (below) allows the DNA molecule within the cell to unfold.
Right: The (color coded) chromosomes within the nucleus of a human cell in compact & "exploded" views.

The study of how DNA structure is regulated and the information stored in DNA is used is the general topic of epigenetics (on top of genetics). Genetics refers to the genetic information itself, encoded in the sequence of DNA molecules. A mutation will effect the sequence of DNA, it may or may not effect a gene, what a gene encodes and/or how the gene is regulated. When considering a particular gene, you may suspect that to be "expressed", transcription factor proteins must be able to find (by diffusion) and bind, through various intermolecular interactions, to specific regions within the gene's regulatory region(s). These interactions are based on nucleotide sequences within the DNA. But the way the DNA is organized into chromatin, particularly in eukaryotic cells, can dramatically influence the ability of transcription factors to interact with and bind to their regulatory sequences. For example, if a gene's regulatory regions are inaccessible to protein binding because of the structure of the chromatin, the gene will be "off" (unexpressed) even if the transcription factors that would normally turn it on are present and active. As with essentially all biological systems, the interactions between DNA and various proteins can be regulated. At this point it is also worth remembering that there are typically only one to two copies of any particular gene within a cell, so we also have to consider the stochastic aspects of these molecular recognition processes.

Different types of cells can often have their DNA organized differently through the differential expression and activities of genes encoding proteins and non-coding RNAs involved in opening up

(making accessible) or closing down (making inaccessible) regions of DNA. You might wonder what accessible means; it means that proteins, and various molecular machines, can bump into and directly interact with specific regions of the DNA. Accessible, transcriptionally active regions of DNA are known as euchromatin while DNA packaged so that the DNA is inaccessible to the regulatory protein binding is known as heterochromatin (→). A particularly dramatic example of this process occurs in female mammals. The human X chromosome contains ~1100 polypeptide-encoding genes that play important roles in both males and females.[362] But the level of gene expression is influenced by the number of copies of a particular gene present within a cell. Only so many RNA polymerase complexes can move along a DNA molecule at a time, and each assembles a single RNA molecule as it moves; each ribosome assembles a single polypeptide as it moves along an mRNA molecule.



While various mechanisms can compensate for differences in gene copy number, this is not always the case. For example, there are genes in which the mutational inactivation of one of the two copies leads to a distinct dominant phenotype, a situation known as haploinsufficiency. This raises issues for genes located on the X chromosome, since XX organisms (females) have two copies of these genes, while XY organisms (males) have only one.[363] While one could imagine a mechanism that increased expression of genes on the male's single X chromosome, the actual mechanism used is to inhibit the expression of genes on one of the female's two X chromosomes (we return to what it means to be "dominant" in Chapter 13). In each XX cell, one of the two X chromosomes is packed into a heterochromatic state, known as a Barr body, more or less permanently. The "decision" as to which of the two X chromosomes is to be packed away ("inactivated") is made in the early embryo and appears to be stochastic - that means that it is equally likely that in any particular cell, either the X chromosome inherited from the mother or the X chromosome inherited from the father may be inactivated, that is, made heterochromatic. Importantly, once made this choice is inherited, the offspring of a cell will maintain the active/inactivated states of the X chromosomes of its parental cell – the inactivation event is inherited vertically.[364] The result is that XX females are epigenetic mosaics, they are made of clones of cells in which either one or the other of their X chromosomes have been inactivated. Many epigenetic events can persist through DNA replication and cell division, so these states can be inherited through the soma. There is even the possibility of evolutionary selection, for example, if the expression of one X chromosome leads to a reproductive advantage (more frequent cell division or survival) than that associated with the expression of the the other X chromosome. The result can be that cells of one "type" out reproduce the other. A particular tissue may end up preferentially expressing genes on the maternal or the paternal X chromosome. A question remains whether epigenetic states can be transmitted through the generation of sperm and egg and into the next generation.[365] Most epigenetic information appears to be reset during the process of embryonic development.

---

[362] Human Genome Project: Chromosome X: http://www.sanger.ac.uk/about/history/hgp/chrx.html

[363] The Y chromosome is not that serious an issue, since its ~50 genes are primarily involved in producing the male phenotype.

[364] X Chromosome: X Inactivation: http://www.nature.com/scitable/topicpage/x-chromosome-x-inactivation-323

[365] Identification of genes preventing transgenerational transmission of stress-induced epigenetic states

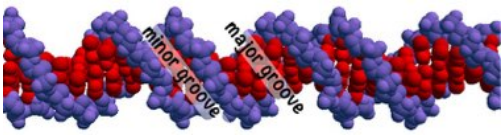
Locating information within DNA

For genes to be useful there needs to be mechanisms by which specific genes can be recognized and expressed (RNA synthesized) at specific times, at specific levels, and in multicellular organisms, in specific types of cells.[366] Recognizing genes involves a two-component system. The first part involves nucleotide sequences that provide a molecular address; this molecular address (a type of bar code) identifies a specific region of a DNA molecule as well as which strand of the DNA should be transcribed, that is, used to direct RNA synthesis. The second component of the system are the proteins that recognize and specifically bind to such "regulatory" DNA sequences. The regulatory region of a gene can be simple and relatively short or long and complex. In some human genes, the regulatory region is spread over thousands of base-pairs of DNA, located “up-stream” and/or "down-stream", within introns or within the coding region.[367] The DNA within a chromosome can fold back on itself, allowing widely separated regions to interact.

The proteins that bind to regulatory sequences are known as transcription factors.[368] Many different transcription factors and transcription factor binding sites can be involved in the regulation of a gene’s expression. In early genetic studies, two general types of mutations were identified that could influence the expression of a gene. “cis” mutations are located within a gene’s regulatory region, often near the gene’s coding (transcribed) region. In contrast “trans” mutations mapped to other, more distant sites, within the genome – often sites located on different chromosomes. Such mutations turned out to alter genes that encode transcription factors and other molecular components involved in gene expression. A transcription factor protein binds specifically (with high affinity) to sequences within the target gene’s regulatory region. A particular transcription factor can influence the expression of many hundreds of genes. Transcription factors can act either positively to recruit and activate DNA-dependent, RNA polymerase or negatively, to block polymerase binding and activation. Post-translational modifications and the binding of allosteric factors can alter the activity of transcription factors, while interactions with other proteins can alter binding specificity and down-stream effects on gene expression.

Genes that efficiently recruit and activate RNA polymerase will make many copies of the transcribed RNA and are said to be highly expressed. Generally (but not always), high levels of an mRNA will lead to high levels of the encoded polypeptide. A mutation in a gene encoding a transcription factor protein (a trans mutation) can influence the expression of many genes, while mutations in a gene’s regulatory sequence (a cis mutation) will directly effect only its own expression, unless of course the gene encodes a transcription factor or its activity influences the regulatory circuitry of the cell. Genes are organized in interacting systems, with associated feed-back mechanisms involved in homeostatic, adaptive, and developmental processes. An experimental point is often to determine whether the expression of a particular gene is directly or indirectly influence by a mutation or an environmental factor.

Transcription regulatory proteins recognize specific DNA sequences by interacting with the edges of base pairs accessible through the major and/or minor grooves of the DNA helix (→). There are a number of different types of transcription factors, with structurally distinct DNA binding domains; transcription factor proteins can be grouped in various structurally, and presumably



[366] As an aside, are many transcribed DNA sequences that do not appear to encode a polypeptide or regulatory RNAs. It is not clear whether this transcription is an error, due to molecular level noise or whether such RNAs play a physiological role..

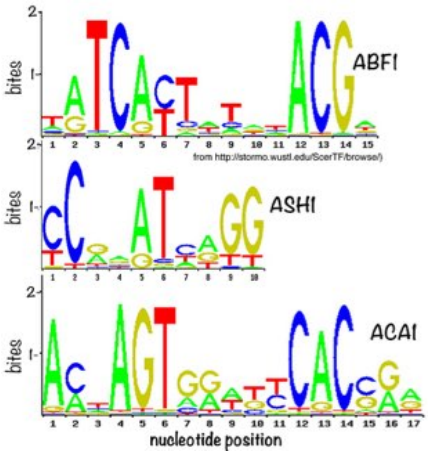
[367] Regulatory regions located far from the gene’s transcribed region are known as enhancer elements.

[368] In prokaryotes transcription factors are often referred to as sigma (σ) factors.

evolutionarily related, families.[369] The binding affinity of a particular transcription factor to a particular regulatory sequence will be influenced by the DNA sequence as well as the binding of other proteins in the molecular neighborhood. We can compare affinities of different proteins for different binding sites by using an assay in which short DNA molecules containing a particular nucleotide sequence are mixed in a 1:1 molar ratio, that is, equal numbers of protein and DNA molecules:

$$DNA_{sequence} + protein \leftrightarrows DNA{:}protein.$$

After the binding reaction has reached equilibrium we can measure the percentage of the DNA bound to the protein. If the protein is in its native (functional) form, binds with high affinity and on its own, that is, with no needed accessory factors, the value will be close to 100%. The ratio of bound to unbound protein will be close to 0% if the transcription factor protein binds with low affinity to the target sequence. In this way we can empirically determine the relative binding specificities (binding affinities) for particular sequences) of various proteins, assuming that we can generate DNA molecules of specific length and sequence (simple) and that we can purify proteins that remain properly folded in a native rather than in a denatured or inactive configuration, which may or may not be simple.[370] What we discover is that transcription factors (very much like the factors that mediate RNA splicing) do not recognize a single, unique nucleotide sequence, but rather have a range of affinities for related sequences. This binding preference is a characteristic of each transcription factor protein; it involves both the length of the DNA sequence recognized and the pattern of nucleotides within that sequence. A simple approach to this problem considers the binding information present at each nucleotide position as independent of all others in the binding sequence, which is not accurate but close enough for most situations. As noted before, the data is presented as a "sequence logo".[371] In such a plot, we indicate the amount of binding information at each position along the length of the binding site (→). Where there is no preference any of the four nucleotides is acceptable. The fewer the number of nucleotides that are acceptable the more information is present. Different transcription factor proteins produce different preference plots.



As you might predict, mutations that influence the transcription factor's DNA binding site can have dramatically different effects; they can abolish site-specific DNA binding altogether or they may alter the DNA sequences bound, leading to changes in patterns of gene expression (addressed later on). Similarly, changes in the sequence recognized by a transcription factor can range from little effect on the binding of a particularly transcription factor to completely abolishing its binding, depending on the nature of the change and the information content of the position altered.

This is not to say that proteins cannot be perfectly specific in their binding to nucleic acid sequences. There are classes of proteins, known as restriction endonucleases and site specific DNA modification enzymes (methylases and acetylases) that bind to unique nucleotide sequences. For example the restriction endonuclease EcoR1 binds to (and cleaves) the nucleotide sequence GAATTC; change any one of these bases and there is no significant binding and no cleavage of the sequence. The CRISPR CAS9 system for genetic manipulation is also highly specific, using a 22

---

[369] Determining the specificity of protein-DNA interactions: http://www.ncbi.nlm.nih.gov/pubmed/20877328

[370] Of course we are assuming that physiologically significant aspect of protein binding involves only the DNA, rather than DNA in the context of chromatin, and ignores the effects of other proteins, but it is a good initial assumption.

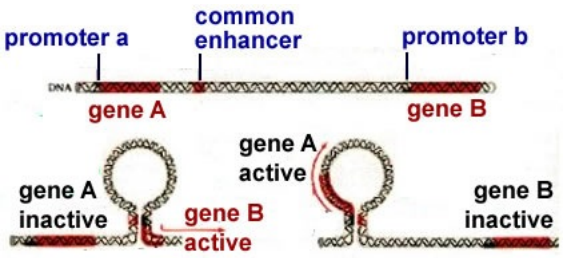
[371] Sequence logos: a new way to display consensus sequences: http://www.ncbi.nlm.nih.gov/pubmed/2172928

nucleotide RNA to target an endonuclease to a specific site in the genome.[372] So the fact that the binding specificities of transcription factors are more flexible suggests that there is a reason for such flexibility, although exactly what that reason is remains conjectural.

A point worth making is that most transcription factor proteins also bind weakly (with low affinity) to generic DNA sequences. Such non-sequence specific binding is transient and rapidly broken by thermal motion. That said, since there are huge numbers of such non-sequence specific binding sites within a cell's DNA, much of the time transcription factors are found transiently associated with DNA. To be effective in recruiting a functional RNA polymerase complex to a specific sites along a DNA molecule, the binding of a protein to a specific DNA sequence must be relatively long lasting. A common approach to achieving this outcome is for the transcription factor to be multivalent, that is, so that it can bind to multiple (typically two) sequence elements at the same time. This has the effect that if the transcription factor dissociates from one binding site, it can remains tethered to the other. Since the molecule is held, by this binding, close to the DNA it is more likely to rebind to its original site. In contrast, a protein with a single binding site is more likely to diffuse away before rebinding can occur. A related behavior involving the low affinity binding of proteins to DNA is that it leads to one-dimensional diffusion along the length of the bound DNA molecule.[373] Collisions are more likely to move the protein along rather than away from the DNA molecule. This enables a transcription factor protein to bind weakly to DNA and then move back and forth along the DNA molecule until it interacts with, and binds to, a high affinity site or until it dissociates completely. This type of "facilitated target search" behavior can greatly reduce the time it takes for a protein to find a high affinity binding site among the millions of low affinity sites present in the genome.[374]

As the conditions in which an organism lives get more complex, the more dynamic gene expression needs to be. This is particularly the case in multicellular eukaryotes, where different cell types need to express different genes, or different versions (splice variants) of genes. One approach is to have different gene regulatory regions, that bind different sets of transcription factors. Such regulatory factors not only bind to DNA, they interact with one another. We can imagine that the binding affinity of a particular transcription factor will be influenced by the presence of transcription factors already bound to a neighboring or overlapping site on the DNA. Similarly the structure of a protein can change when it is bound to DNA, and such a change can lead to interactions with DNA:protein complexes located at more distant sites, known as enhancers. Such regulatory elements, can be part of multiple regulatory systems.

For example, consider the following situation. Two genes share a common enhancer, depending upon which interaction occurs, gene A or gene B but not both could be active (→). The end result is that combinations of transcription factors are involved in turning on and off gene expression. In some cases, the same protein can act either positively or negatively, depending upon molecular context, that is, the specific gene regulatory sequences accessible, the other transcription factors present and their various post-translational modifications. Here it is worth noting (again) that the organization of regulatory and coding sequences in DNA imposes directionality on the system. A transcription factor bound to DNA in one orientation or at one position may block the binding of other proteins (or RNA polymerase),



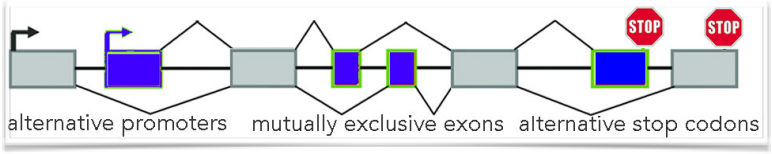
[372] The CRISPR-CAS9 system involves targeting a double-stranded DNA exonuclease to a specific site in a DNA sequence; it uses a RNA molecule to achieve very high levels of specificity. see CRISPR/Cas9 and Targeted Genome Editing

[373] As illustrated in the PhET applet:http://phet.colorado.edu/en/simulation/gene-expression-basics

[374] Physics of protein-DNA interactions: mechanisms of facilitated target search

while bound to another site it may stabilize protein (RNA polymerase) binding. Similarly, DNA binding proteins can interact with other proteins to control chromatin configurations that can facilitate or block accessibility to regulatory sequences. While it is common to see a particular transcription factor protein labelled as either a transcriptional activator or repressor, in reality the activity of a protein often reflects the specific gene under consideration, and its interactions with various accessory factors, all of which can influence gene expression outcomes.

The exact position on the DNA where RNA polymerase starts transcribing an RNA molecule is known as the transcription start site. Different regulatory sequences can lead to different transcription start sites. Similarly, in genes with introns, where transcription starts can determine which exons are included in the final transcript (mRNA molecule). Other factors influence splicing, and so determine which exons are included and which are excluded from the final RNA (→). Where the RNA polymerase falls off the DNA, and so stops transcribing RNA, is known as the transcription termination site.



Once transcription initiates, the RNA polymerase moves down the DNA; as it clears the transcription start site there is now room for another polymerase complex to associate with the DNA. Assuming that the factors associated with the regulatory region remains intact and active, the time to load a new polymerase on an existing regulatory complex will be much faster than the time it takes to build up a new regulatory complex from scratch; the result is that transcription is often found to occur in bursts, a number of RNAs are synthesized from a particular gene in a short period of time followed by a period of transcriptional silence associated with the disassembly and reassembly of the transcription start complex. A similar bursting behavior is observed in polypeptide synthesis (translation). The onset of translation begins with the small ribosomal subunit interacting with the 5' end of the mRNA; the assembly of this initial complex involves a number of components, and takes time but once formed persists for awhile. While this complex exists multiple ribosomes can interact with the mRNA, each synthesizing a polypeptide, leading to bursts (multiple rounds) of translation. Once the translation initiation complex dissociates, it takes time, more time than just colliding with another small ribosomal subunit, for a new complex to form. The combination of transcriptional and translational bursting contributes to noisy protein synthesis. Since cellular behavior can be influenced by changes in gene expression, these processes can lead to phenotypic differences between genetically identical cells. An analogous process involving differential DNA (chromatin) accessibility can lead to monoallelic gene expression, in which only one or the other of the two genes present in a diploid cell is expressed. Monoallelic expression can lead to phenotypic differences between cells.[375]

Questions to answer:

167. How might a transcription factor determine which DNA strand will be transcribed?
168. How could one increase the specificity of a particular transcription factor protein?
169. A mutation inhibits the expression of a gene, how might determine whether the mutation altered a transcription factor or the DNA sequences that regulate gene expression.
170. What factors are likely to influence the length of a gene's regulatory region?
171. How might you tell which X chromosome was inactivated in a particular cell of a female person?

Questions to ponder:

– What factors might drive the evolution of overlapping genes?
– How can over-lapping genes, or genes on different DNA strands influence each others' expression?
– How might you determine which allele is expressed in a cell displaying monoallelic gene expression?

[375] Monoallelic Gene Expression in Mammals - Chess, 2016

Interaction networks and model systems

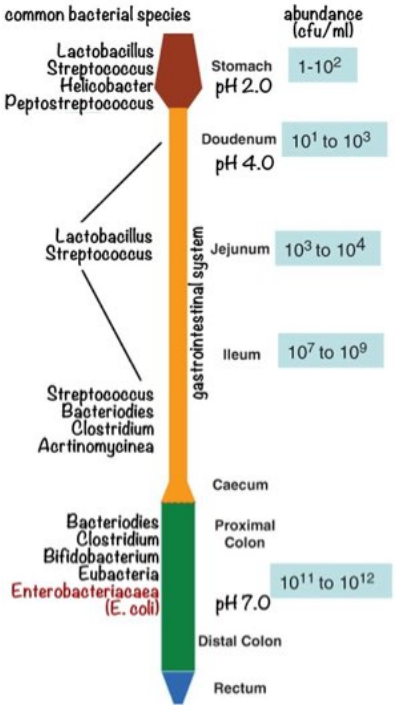
Interaction networks are a universal feature of biological systems, from the molecular to the social. These are generally organized in a hierarchical and bidirectional manner, involving various forms of "feedback". So what exactly does that mean? Most obviously, at the macroscopic level, the behavior of ecosystems depends upon the interactions between of organisms. As we move down the size scale the behavior of individual organisms is based on the interactions between the cells and tissues formed during the process of embryonic development. Gene expression also involves interaction networks; genes express proteins that regulate the expression of other genes (including the genes that encode them) and multiple gene products are involved in the regulation of a particular gene. Since many of these interactions have a stochastic nature, chance plays a role. At the same time there are regulatory interactions and feedback loops that can act to control stochastic effects and serve to make biological behaviors more robust. All of these interactions, and the processes that underlie particular biological systems, are the result of evolutionary processes and historical situations, including past adaptations and non-adaptive events in ancestral populations.

Scientific studies of biological systems are driven by the desire to understand how it is that such systems came to be and how they behave the way they do. Such knowledge is helpful, particularly in the age of genetic engineering, in order to treat or avoid a disease. But there are a number of reasons that some questions cannot be answered directly; it may not be possible (or ethical) to carry out the necessary experiments. But here the evolutionary relationships between organisms come to our aid; we can choose organisms that are easier to study, develop faster, or are “simpler” in a way. By studying various “model” organisms, we can come to identify what can be common and relevant mechanisms. At the same time, it is important to recognize that the various “types” of organism that have been useful experimentally are each adapted to a specific environmental niche, generally evolving independently of others for millions to hundreds of millions of years. Even the most closely related of organisms, such as the great apes, a group that includes humans, display functionally significant differences. Once isolated, and maintained in the laboratory, we put organisms in an unnatural situation, a situation that subjects them to different selection pressures. At the same time, isolated organisms are often maintained under conditions that reduce genetic variation - they become inbred. Such inbreeding can be desirable (for science), since it reduces variability and makes experiments more interpretable, while at the same time making them less realistic or relevant to “real” organisms.

Not withstanding the complexities of biological systems, we can approach them at various levels of resolution through a systems perspective, using specific organisms to study specific processes and behaviors. At each level, there are objects that interact with one another in various ways to produce specific behaviors. Many of these systems are conserved, related to one another evolutionarily. To analyze a system we need to define, identify, and appreciate the nature of the objects involved, how they interact, and the behaviors and that emerge from such interactions, in particular how such interactions influence the system. Does the system move to a new state or does it return, after a perturbation, to its original state? There are many ways to illustrate this way of thinking but we will get concrete by looking at a (relatively) simple system and consider how it behaves at the molecular, cellular, and social levels. The model system we consider here is the bacterium *Escherichia coli* in particular how it behaves in isolation, in social groups, and how it metabolizes the milk sugar lactose.[376] Together these illustrate a number of common regulatory principles that apply more or less universally to biological systems at all levels of organization.

[376] The Lac Operon: A Short History of a Genetic Paradigm

E. coli as a model system



Every surface of your body harbors a flourishing microbial ecosystem. This is particularly true of the gastrointestinal system, which runs from your mouth and esophagus (with a branch leading to your nose), through the stomach, into the small and large intestine and the colon (→).[377] Each of region supports its own unique microbial community, known as a microbiome. These environments differ in terms of a number properties, including differences in pH and O_2 levels. Near the mouth and esophagus O_2 levels are high and microbes can use aerobic (O_2 dependent) respiration to maximize the extraction of energy from food. Moving through the system O_2 levels decrease until anaerobic (without O_2) mechanisms are necessary. At different positions along the length of the gastrointestinal track microbes with different ecological preferences and adaptabilities are found.[378]

One challenge associated with characterizing the complexity of the microbiome present at various locations is that often the organisms present are dependent upon one another for growth and survival. When isolated from one another (and their normal environment) they do not grow. The standard way to count bacteria is to grow them in the lab. Samples are diluted so that single bacteria land in isolation from one another on an agar plate surface. When they grow and divide, they form macroscopic (visible) colonies; we count the number of "colony forming units" (CFUs) per original sample volume; this number provides a measure of the number of individual viable bacteria present, or rather the number of bacteria capable of growing and dividing. If an organism cannot form a colony under the assay conditions, it will appear to be absent from the population. Many bacteria are dependent on others and could not be grown in isolation. Recent studies, however, have found ways to culture more of such organisms.[379] To avoid this issue, molecular methods use DNA sequence analyses to identify which organisms are present without having to grow them.[380] The result of these types of analysis has revealed the true complexity of the microbial ecosystems living on and within us.[381]

Much early work in molecular biology was carried out using a relatively minor member of this microbial community, *Escherichia coli*. *E. coli* is a member of the Enterobacteriaceae family of bacteria and is found in the colons of birds and mammals.[382] *E. coli* is what is known as a facultative aerobe, it can survive in both anaerobic and an aerobic environments. This flexibility, as well as *E. coli*'s generally non-fastidious nutrient requirements make it easy to grow in the laboratory. Moreover, the commonly used laboratory strain of *E. coli,* known as K12, does not cause disease in humans. That said, there are strains of *E. coli*, such as *E. coli* O157:H7, that are pathogenic (disease-causing). *E. coli* O157:H7 contains 1,387 genes that are not found in the *E. coli* K12 strain and it is estimated that the two strains diverged from a common ancestor ~4 million years ago. The

---

377 The gut microbiome: scourge, sentinel or spectator?

378 The Gut Microbiome: Connecting Spatial Organization to Function and Gut biogeography of the bacterial microbiota

379 See Lopez-Garcia & Moreira, (2020) Cultured Asgard Archaea Shed Light on Eukaryogenesis

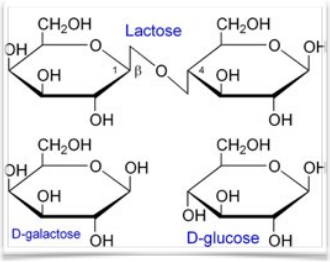
380 Application of sequence-based methods in human microbial ecology: http://www.ncbi.nlm.nih.gov/pubmed/16461883

381 The human microbiome: our second genome

382 Evolutionary ecology of E.coli

details of what makes *E. coli* O157:H7 pathogenic is a fascinating topic, but beyond our scope here.[383]

Adaptive behavior and gene networks: the lac response



Lactose is a disaccharide (a sugar) composed of D-galactose and D-glucose (←). It is synthesized, biologically, exclusively by female mammals. Mammals use lactose in milk as a source of calories (energy) for infants. One reason, it is thought, is that lactose is not easily digested by most microbes. The lactose synthesis system is derived from an evolutionary modification of an ancestral gene that encodes the enzyme lysozyme. Through a gene duplication event and mutations, a gene encoding the protein α-lactoalbumin was generated. α-lactoalbumin is expressed in mammary glands, where it forms a macromolecular complex with a ubiquitously expressed protein, galactosyltransferase, to form the protein lactose synthase.[384]

E. coli is capable of metabolizing lactose, but only when there are no better (easier) sugars to eat. If glucose or other compounds are present in the environment, the genes required to metabolize lactose are turned off, they are not expressed. Two genes are required for *E. coli* to metabolize lactose. The first encodes lactose permease. Lactose, being large and highly hydrophilic cannot pass through the *E. coli* cell's membrane. Lactose permease is a membrane protein that allows lactose to enter the cell, moving down its concentration gradient. The second gene involved in lactose utilization encodes the enzyme β-galactosidase, which catalyzes the reaction that splits lactose into D-galactose and D-glucose, both of which can be metabolized by proteins expressed constitutively, that is, all of the time. So how exactly does this system work? How are the lactose utilization genes turned off in the absence of lactose and how are they turned on when lactose is present and energy is needed? How does the cell sense when lactose is present in the environment? The answers illustrate general principles of the interaction networks controlling gene expression.

In *E. coli*, like many bacteria, multiple genes are organized into what are known as operons. In an operon, a single regulatory region controls the expression of multiple genes, often genes involved in the same metabolic pathway. A powerful approach to the study of genes is to look for mutations that abolish a specific process, and so produce a discernible phenotype. As we said, wild type (that is, normal) *E. coli* can grow on lactose as their sole energy source. So to understand lactose utilization, we can look for mutant *E. coli* that fail to grow on lactose.[385] To make the screen for such mutations more relevant, we first check to make sure that the mutant can grow on glucose. Why? Because we are not really interested (in this case) in mutations in genes that disrupt standard metabolism, such as the ability to use glucose. We seek to identify the genes (and gene products) involved in a specific process, lactose metabolism. Such an analysis revealed a number of distinct classes of mutations: some led to an inability to respond to the presence of lactose in the medium, others led to the de-repression, that is the constant expression of the two genes involved in the ability to metabolize lactose, lactose permease and β-galactosidase. In such mutant strains both genes were expressed whether or not lactose is present.

By mapping where these mutations are in the genome of *E. coli* (using the Hfr horizontal gene transfer system described in chapter 12) and a number of other experiments, the following model was generated (↓). The genes encoding lactose permease (lacY) and β-galactosidase (lacZ) are part of an operon, known as the lac operon. The lac operon is regulated by two distinct factors. The

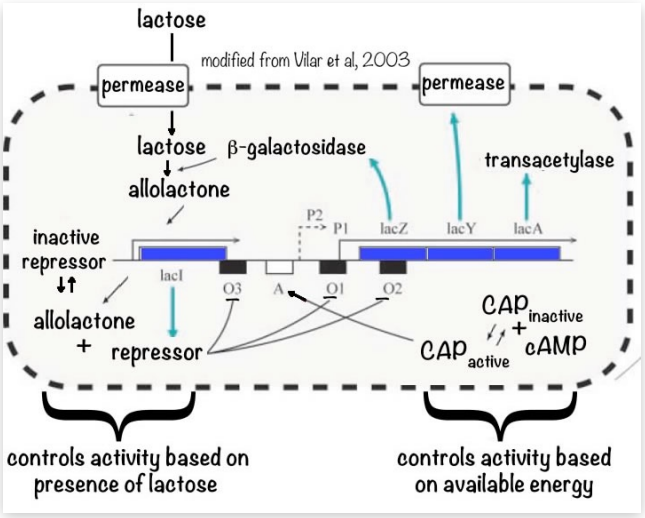
---

[383] Enterohemorrhagic E. coli (EHEC) pathogenesis: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3417627/

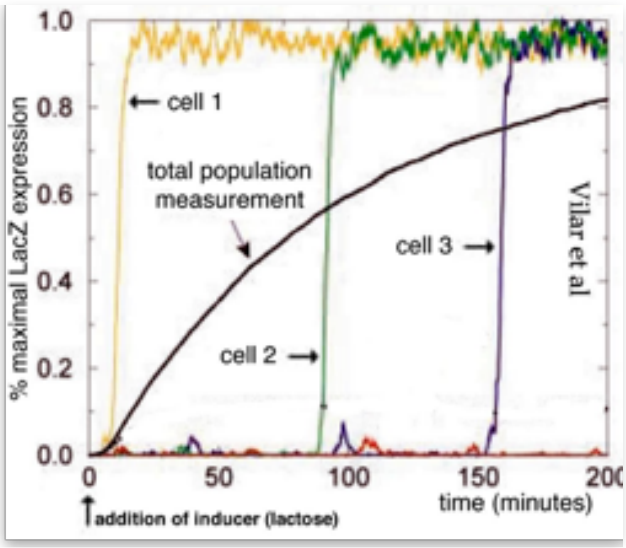
[384] Molecular divergence of lysozymes and alpha-lactalbumin: http://www.ncbi.nlm.nih.gov/pubmed/9307874

[385] The basic experimental approach involves a technique known as replica plating

first is the product of a constitutively active (that is, expressed) gene, lacI; the lacI-encoded polypeptide assembles into a tetrameric protein that acts as a transcriptional repressor. A typical cell contains ~10 lac repressor proteins and generally one or two copies of the lac operon. The lac repressor protein binds to sites in the promoter of the lac operon; the binding of the repressor blocks the expression (transcription) of the lac operon. The repressor's binding sites within the lac operon promoter appear to be its' only functionally significant binding sites in the *E. coli* genome. The second regulatory element in the system is known as the activator site. It can bind the catabolite activator protein (CAP). CAP is encoded by a gene located outside of the lac operon. CAP is a homodimer, that is, it is composed of two identical polypeptides. The DNA binding activity of CAP is regulated by the binding of an allosteric co-factor, cyclic adenosine monophosphate (cAMP). cAMP accumulates in the cell when nutrients, specifically free energy delivering nutrients (like glucose), are low. An increase in cAMP concentration [cAMP] acts as a signal that the cell needs energy. In the absence of cAMP, CAP does not bind to or activate expression of the lac operon, but in its presence (that is, when energy is needed), CAP-cAMP is active, binds to a site in the lac operon promoter, and recruits and activates RNA polymerase, leading to the synthesis of lactose permease and β-galactosidase RNAs and proteins. However, even if energy levels are low and [cAMP] is high, the lac operon will be inactive (not expressed) if lactose is absent because binding of the lac repressor protein to sites (labeled O1, O2, and O3) in the lac operon's regulatory region blocks polymerase recruitment.



So what happens when lactose appears in the cell's environment? Well, obviously nothing, since the cells are expressing the lac repressor, so the lac operon is not expressed and the lactose permease is not present. Lactose cannot enter the cell without it. A simple prediction might assume the system works perfectly and deterministically, but this is not the case. The system is stochastic, that is, it is noisy and probabilistic. Given the small number of lac repressor molecules per cell (~10), there is a small but significant (non-zero) chance that, at random, the lac operon will be free of bound repressor. If this occurs under conditions in which CAP is active, β-galactosidase and lactose permease will be expressed independently of the presence of lactose. If, however, lactose is present, there is a positive feedback loop (←).[386] Those few cells that have, by chance, expressed both lacY (lactose permease) and lacZ (β-galactosidase) genes will respond. The permease will enable lactose to enter these cells. This lactose will be converted to allolactone, in a reaction catalyzed by β-galactosidase. Allolactone binds to, and inhibits the lac repressor protein. Unrepressed, there is a further increase (~1000 fold) in the rate of expression of the lacZ and lacY genes. In addition to generating allolactone from lactose, β-galactosidase catalyzes the hydrolysis of lactose into D-galactosidase and D-glucose, which are used to drive cellular metabolism. Through this process,



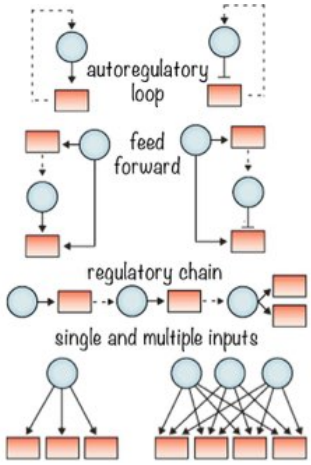
[386] Modeling network dynamics: the lac operon, a case study: http://www.ncbi.nlm.nih.gov/pubmed/12743100

the cell goes from essentially no expression to the full expression of the lac operon, which enables the cell to metabolize lactose. At the same time, those cells that did not (by chance) express lactose operon will be unable to metabolize lactose, even though lactose is present outside of those cells. So even though all of the *E. coli* cells present in a culture may be genetically identical, they can express different phenotypes due to the stochastic nature of gene expression.[387] In the case of the lac system, over time the noisy nature of gene expression leads to more and more cells activating their copy of the lac operon. Also cells that can metabolize lactose have energy for growth. The offspring of such a cell will inherent lactose permease and β-galactosidase, so will be able to use lactose. Once "on", the operon will be expressed as long as lactose is present, since allolactone, derived from lactose, binds to and inactivates the lac repressor protein.

What happens if (and when) lactose disappears from the environment, what determines how long it takes for the cells to return to the state in which they no longer express the lac operon? The answer is determined by the effects of cell division and regulatory processes. In the absence of lactose, the [allolactone] falls and the lac repressor protein will return to its active (repressive) state, inhibiting lac operon expression. No new lactose permease and β-galactosidase will be synthesized and their concentrations will fall based on the rate of their dilution by growth and cell division and their degradation (proteolysis). In the absence of lactose, each cell division will reduce the concentration of the lactose permease and β-galactosidase by ~50%. As the proteins are diluted or degraded, the cells return to their initial state, that is, with the lac operon off and no copies of either lactose permease or β-galactosidase present.

Types of regulatory interactions

A comprehensive analysis of the interactions between 106 transcription factors and (many more) regulatory sequences in the baker's yeast *Saccharomyces cerevisiae* revealed the presence of a number of common regulatory motifs.[388] These include (→):



- **Auto-regulatory loops**: A transcription factor binds to sequences that regulate its own transcription. Such interactions can be positive (amplifying) or negative (squelching).
- **Feed forward interactions**: A transcription factor regulates the expression of a second transcription factor; the two transcription factors then cooperate to regulate the expression of a third gene.
- **Regulatory chains**: A transcription factor binds to the regulatory sequences in another gene and induces expression of a second transcription factor, which in turn binds to regulatory sequences in a third gene, etc. The chain ends with the production of some non-transcription factor products.
- **Single and multiple input modules**: A transcription factor binds to sequences in a number of genes, regulating their coordinated expression. In most cases, sets of target genes are regulated by sets of transcription factors that bind in concert.

In each case the activity of a protein involved in an interaction network can, like the lac repressor, be regulated through interactions with other proteins, allosteric factors, and post-translational modifications. It is through such interactions that signals from inside and outside the cell can control patterns of gene expression leading to maintenance of the homeostatic state or various adaptations.

[387] An example of such behavior here: http://www.elowitz.caltech.edu/publications/Noise.pdf

[388] Transcriptional regulatory networks in *Saccharomyces cerevisiae*: http://www.ncbi.nlm.nih.gov/pubmed/12399584

Final thoughts on (molecular) noise, for now

When we think about the stochastic behaviors of cells, we can identify a few reasonably obvious sources of molecular and cellular level noise. First, there are generally only one or two copies of a particular gene within a cell. The probability that those genes are accessible and able to recruit transcription factors, associated proteins, and RNA polymerase molecules is determined by the frequency of productive collisions between regulatory sequences and relevant transcription factors together with their dissociation rates. Cells are small, and the numbers of different transcription factors can vary quite dramatically. Some transcription factors are present in high numbers (~250,000 per cell) while others (like the lac repressor) may be present in less than 10 copies per cell. The probability that particular molecules interact will be controlled by their relative concentrations, diffusion, binding, and kinetic energies. This will influence the probability that a particular gene regulated by a particular transcription factor is active or not. Once on, transcriptional and translational bursting will produce gene products that can alter the state of the cell so that secondary, down-stream changes occur in gene expression and other cellular processes. These changes may (like the lac operon system) be reversible once the stimulus (lactose) is removed or they may be more or less irreversible, as occurs during cellular differentiation and embryonic development.[389]

Questions to answer:

172. How would you design a regulatory network to produce a steady level of product?
173. How would you design a regulatory network that oscillates like a clock?

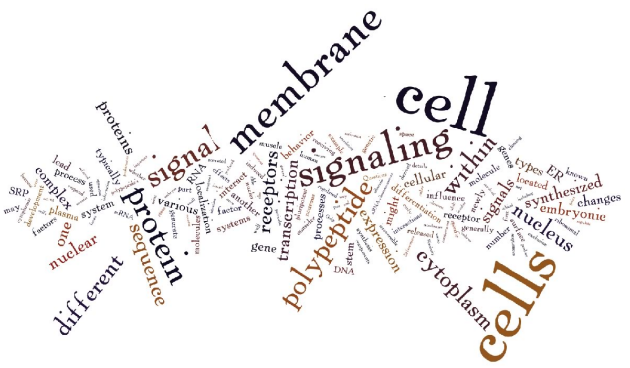
Question to ponder:

– Design a gene regulatory system that acts as an irreversible switch between states?

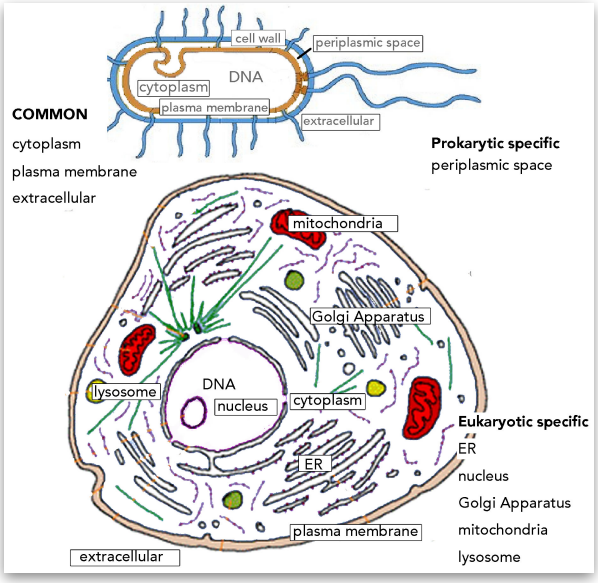
[389] A single molecule view of gene expression: http://www.ncbi.nlm.nih.gov/pubmed/19819144

Chapter 10: Cellular topology and intercellular signaling

In which we consider the signals, receptors, and machinery that control how proteins come to be where they are needed within cells and organisms, and how cells interact with one another through various signaling systems.



As noted earlier, each cell is a bounded non-equilibrium system. The plasma membrane forms an unambiguous boundary between the rest of the universe and the cell. In prokaryotes, the cell is typically surrounded by a cell wall, a semi-rigid structure that protects the cell from osmotic effects among other things. As we have discussed, the cell's metabolic activities occur primarily within the space defined by its cell membrane, the cytoplasm.[390] A polypeptide synthesized within the cytoplasm has a number of places it might end up, and like other biological processes these outcomes are controlled by signals and receptors. In a prokaryote (upper image →), a newly synthesized polypeptide can remain in the cytoplasm where it can interact with the organism's genetic material, its DNA, since it is also located directly in the cytoplasm. Alternatively, a newly synthesized polypeptide can end up embedded within the plasma membrane (an integral membrane protein) or it can pass through the membrane and be secreted. Secreted proteins (in a prokaryote) can remain within the periplasmic space, can become part of the cell wall, or can pass through the cell wall and into the external environment.



Eukaryotic cells (→) are more topologically complex and contain a distinctive double membrane structure, the nucleus. The cell's genetic material, its DNA, organized into chromosomes (linear molecules compared to circular molecules found in prokaryotes), is located within the nucleus. The synthesis of RNA molecules, occurs within the nucleus. The membranes of the nucleus are elaborated within the cytoplasm into a network known as the endoplasmic reticulum or ER. There are a number of other intracellular membranes, including the Golgi apparatus and various types of small vesicles, involved in moving molecules to and from the plasma membrane and between the ER and Golgi apparatus. Finally, there are the mitochondria and (in plants) chloroplasts, double-membrane structures with their own genomic DNAs, derived from apparent endosymbiotic events early in the history of eukaryotes (discussed earlier). Complex signal-receptor interaction mechanisms serve to maintain the topological details of cells as they give rise to new cells. Our focus here are the general rules by which specific proteins end up are "targeted" to specific cellular compartments.

Targeting proteins to where they need to be: membrane proteins

So the question is, what determines where a polypeptide (protein) ends up? As you might suspect, there are signals and receptors involved. Signals are typically part of the polypeptide's primary (amino acid) sequence and receptors are proteins encoded for by a number of other genes.

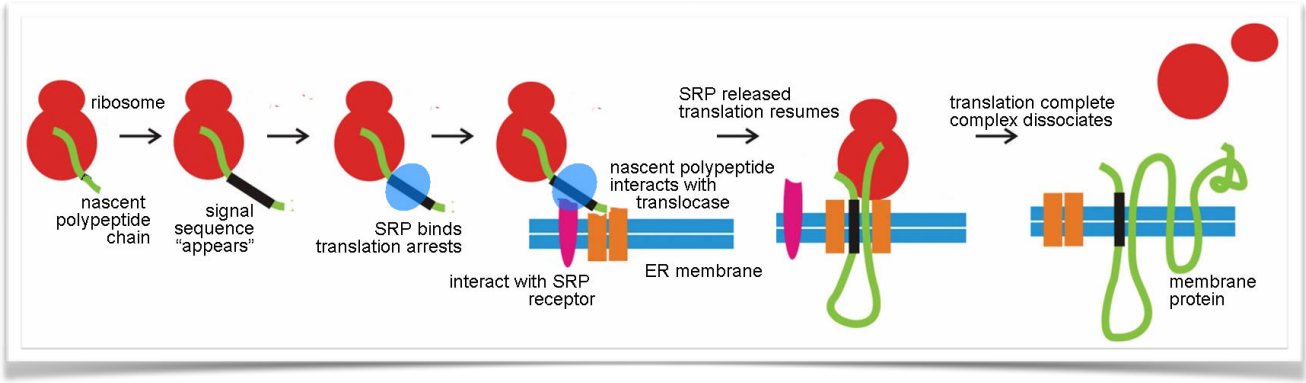
---

[390] In prokaryotes, there is a distinct space between the cell membrane and the cell wall known as the periplasmic space; a number of reactions occur within this region. A similar space exists in plants and fungi that, unlike animal cells, have evolutionarily distinct cell walls.

The receptors are already present in the living cell, they are part of what is inherited from a cell's progenitor, part of the continuity of life captured in the cell theory. We begin our description of polypeptide targeting with prokaryotes, because they are simpler. We will consider how a newly synthesized polypeptide comes to end up in the cytoplasm, the plasma membrane, or outside of the plasma membrane.

In prokaryotes, the genomic DNA is located in the cytoplasm; there is no barrier between a newly synthesized RNA molecule and the ribosomes, tRNAs, and the other components involved in RNA-dependent polypeptide synthesis. The newly synthesized mRNA molecule can interact with the small and large ribosomal subunits, assemble with them to form a functional ribosome and direct polypeptide synthesis. For a water-soluble cytoplasmic polypeptide, as opposed to a polypeptide that resides in, or passes through the membrane, no further "signals" are necessary. The ribosomal complex moves along the mRNA, the polypeptide is synthesized, passes through the ribosomal channel, and emerges into the cytoplasm. When the ribosome reaches a stop codon, release factor binds, leading to the disassembly of the ribosomal-mRNA-polypeptide complex. The ribosomal components, as well as the mRNA can then initiate a new mRNA-ribosome complex, to produce another polypeptide. The released (newly synthesized) polypeptide may fold on its own or associate with other polypeptides to form a functional protein. Some of these folding steps may involve interactions with chaperones.

So what is going on with a polypeptide destined for insertion into a membrane? Clearly it has a different structure than a water-soluble protein; differences you should be able to predict. The first step in delivering a membrane protein to or through a membrane is to recognize a newly synthesized polypeptide as a membrane protein, or one that needs to pass through a membrane. The general mechanism (and the only one we will consider) involves what is known as a signal sequence (↓). A signal sequence is composed primarily of hydrophobic amino acids; the typical



signal sequence is between 8 to 12 amino acids in length and generally located near the polypeptide's N-terminus, the first part of the polypeptide to be synthesized. The presence of such a signal sequence marks the polypeptide as a membrane protein. As a new synthesized polypeptide emerges from the ribosomal tunnel, the signal sequence is recognized through its binding of a cytoplasmic receptor, the signal recognition particle (SRP). SRP is composed of polypeptides and a structural RNA. The binding of a SRP to a signal sequence causes translation to halt, although the mRNA-ribosome-nascent polypeptide-SRP complex remains intact. The mRNA-ribosome-nascent polypeptide-SRP complex diffuses within the cell until it engages an SRP-receptor located on the cytoplasmic surface of the plasma membrane; the SRP receptor is associated with a transmembrane polypeptide translocator (↑). When the mRNA-ribosome-nascent polypeptide-SRP+SRP Receptor complex forms, SRP disassociates from the ribosome-nascent polypeptide complex, translation resumes and the nascent polypeptide interacts with the translocon and either folds to become embedded within the membrane, or passes through the membrane, and is released (secreted) on the other side. Typically, if the polypeptide is secreted, the signal sequence is removed by proteolytic processing.

Now let us consider the situation in eukaryotic cells. Although more topologically complex the same basic process applies. The difference is that the SRP receptor is not located in the plasma membrane, rather it is located in the ER membrane. A protein with a signal sequence will be delivered to the ER membrane or released into the lumen of the ER. From there other signals will determine whether the protein stays in the ER, moves to the Golgi apparatus, where it is post-translationally modified, and may then move to the plasma membrane, or to some other membrane compartment within the cell. A protein in the lumen of the ER is effectively outside of the cytoplasm, and can be retained within a membrane compartment (such as the ER) or secreted from the cell. At this point, we will not concern ourselves with further details, except to say that whenever a protein is targeted to a specific cellular compartment, we can assume that the protein contains signals that are recognized by receptors that lead to its localization.

Nuclear targeting and nuclear exclusion

All polypeptides are synthesized in the cytoplasm, but can be assembled in any of the cell's topologically distinct compartments. So, what happens if the protein needs to be assembled and functions in the nucleus or within the endoplasmic reticulum, say as part of the DNA replication, DNA repair, RNA transcription, or RNA processing machinery? And what about a cytoplasmic protein that might interfere with such processes if it were to find its way into the nucleus? Again we find the same pattern, there must be signals, typically amino acid sequences that indicate the protein should be located to or excluded from the nucleus. Such signals exist, and are referred to as nuclear localization (NLS) or nuclear exclusion (NES) sequences. Such sequences interact with receptors, that is, molecular machines associated with the nuclear pore complex that mediate the polypeptide's (protein's) translocation into or out of the nucleus.

It is worth noting that a protein can contain both NLS and NES sequences. Their "activities" can be regulated by allosteric effector binding or post-translational modifications. NLS and NES sequences may be accessible or inaccessible, that is unable to interact with the nuclear pore machinery. Where a protein is within a cell, that is, the percent of the protein in a cell located in the nucleus, the cytoplasm, or both, can be controlled. The extent to which a protein, such as a transcription factor or kinase (for example), is within the nucleus will influence its functional impact on the cell. Nuclear localization of a positively acting transcription factor can lead to the activation of a gene, as can the nuclear exclusion of a negatively acting transcription factor. Changing the intracellular distribution of a transcription factor, whether positively or negatively acting, can influence the expression of the genes the transcription factor regulates. The situation is different from that found in membrane targeting (the signal sequence-SRP system), which is essentially irreversible - once a protein is inserted into a membrane or excreted from the cell, and its signal sequence removed, the protein cannot return to the cytoplasm. Many proteins can shuttle back and forth between nucleus and cytoplasm.

Questions to answer:

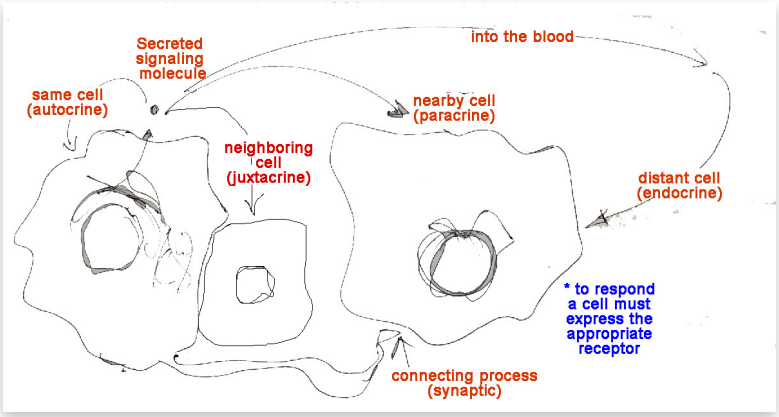
174. How is a water soluble protein different from a protein that resides in a membrane?
175. What are the components needed to insert of polypeptide/protein into or through a membrane? How might mutations in these proteins influence a polypeptide's localization within a cell?
176. Predict what would happen if a signal sequence were mutated.
177. How might you activate a NLS or NES sequence within a protein? How might such a sequence be rendered inactive?

Question to ponder:

- How might a cytoplasmic protein be inserted into a membrane?

Intercellular signaling: signals, receptors & responses

The ability of cells to place proteins on their surface and to secrete proteins into the extracellular space, opens up the possibility of various forms of signaling between cells. Intercellular signaling enables cells to influence each other in various ways.[391] Here we consider only the basics of such processes, more details will be added later on. Intercellular signaling system involves the synthesis of a signaling molecule. This depends turning on the expression of the gene(s) encoding the signaling molecule or the metabolic machinery needed for its synthesis, followed by its processing, and secretion or localization to the cell surface (↓). Similarly, for a cell to respond to a signal, whether from another cell or from itself, a cell has to express a receptor for the signal molecule. Such receptors are proteins and are generally located on the responding cell's surface. When the signal binds to the receptor it acts as an allosteric effector, changing the behavior of the receptor. Different signal-receptor combinations produce different types of changes in the receptor, changes that initiate a cascade of events leading to changes in cell behavior, gene expression, or (often) both.



When signaling molecules are released from a cell into the extracellular space they are (generally) free to diffuse. They can interact with receptors present on the surface of cells within the immediate neighborhood of the signal secreting cell. If the signal is high enough, cells that have the appropriate receptors on their surface will respond. In autocrine signaling (↑), the cell that released the signal also has receptors for the signal; in a sense the cell can talk to itself.[392] If the signal interacts with receptors on neighboring cells, it is referred to as paracrine signaling. A third form of signaling occurs when the signal is released from one type of cell (or cells in one region) and is transported throughout the body of the (multicellular) organism, typically through the blood stream, which is referred to as endocrine signaling. Juxtacrine signaling occurs when the signaling and receiving cells need to touch one another molecularly, through surface membrane proteins. Altogether such interactions underlie the coordination of the behavior of neighboring cells; they are the basis for multicellularity, cellular differentiation, organ formation and coordination, and the formation and function of the immune and nervous systems. The effects of intercellular signaling can be largely transient, for example, as in muscle contraction, or can lead to irreversible changes in gene expression, cell morphology, and behavior. Signaling induced cascades in changing gene expression and cellular behaviors underlie embryonic development and disease progression.

Signaling molecules and receptors

Molecules that provoke a signaling responses are typically called agonists. Different agonists interact with agonist-specific receptors, typically composed of one or more integral membrane proteins. Their interactions produces distinct "down-stream" molecular cascades that exploit post-translational modification or allosteric effects to activate or inactivate various enzymes and transcription factors. In general for each component of a signaling system, there are molecules (generally proteins) that act antagonistically; they inhibit the signaling process - these are known as

---

[391] Antebi et al. 2017. An operational view of intercellular signaling pathways

[392] as an example, see Glucagon regulates its own synthesis by autocrine signaling

antagonists. Antagonists (or inhibitors) may bind agonists, receptors, or "downstream" effectors and so block signaling. Moreover, any one particular cell may express a number of different signaling pathway components; cells of different types will express different combinations of signaling systems, so they will be responsive to different incoming signals. Different combination of signaling factors can produce different effects.

In cases where signaling leads to changes in gene expression, these changes can modify the behavior of the cell, and lead to changes in cellular phenotype. As a general rule, any particular signaling input will generate both direct and indirect effects. For example, activation of a signaling system may lead to the activation (or repression) of a specific set of transcription factors. These can directly regulate the expression of a set of target genes. Some of these genes may themselves encode transcription factors, or polypeptides that regulate transcription factor activity and gene accessibility. The expression of these genes will, in turn, regulate other genes – these are considered indirect or secondary targets of the signaling system. Since which genes will be turned on or off will be influenced by the total set of transcription factors and associated proteins that are expressed and active in a cell, the response of different types of cells to the same signal can be different, and characteristic of the cell type. For example, a muscle cell might respond differently from a kidney cell to the same signal. Similarly, once a cell has been signaled to, the changes in the patterns of gene expression can lead to subsequent changes in cell morphology and behavior, including evolving changes in patterns of gene expression, it can differentiate, that is become different from what it was originally. The process of embryonic development consists of a series of signals and cellular responses that lead to the specialization of cells, the development of tissues, and organ systems. Normally, this process of signal-driven differentiation is irreversible. It proceeds in one and only one direction. The processes result in what is known as terminal differentiation. Only recently have strategies been developed that can reverse these effects.

Cellular reprogramming: embryonic and induced pluripotent stem cells

An important question, asked by early developmental biologists, was is cellular differentiation due to the loss of genetic information? Is the genetic complement of a neuron different from a skin cell or a muscle cell? This question was first approached by Briggs and King in the 1950s through nuclear transfer experiments in frogs. These experiments were extended by Gurdon and McKinnell in the early 1960s; they were able to generate adult frogs via nuclear transfer using embryonic (differentiated) cells.[393] The process was inefficient however - only a small percentage of the nuclei taken from differentiated cells supported normal embryonic development. The ability of somatic cells to be "reprogrammed" by the egg so that they could support embryonic development differs between different types of cells. In part this seems to be due to effectively irreversible changes associated with DNA/chromatin modification.[394] Finally, stochastic processes can influence the patterns of gene expression, so that even cells of the "same type" can differ in their patterns of gene expression, and in their ability to support embryonic development. Nevertheless, these experiments suggested that it was the regulation rather than the loss of genetic information that was important in embryonic differentiation.

In 1996 Wilmut et al used somatic cell nuclear transplantation to clone the first mammal, the sheep Dolly. Since then many different species of mammal have been cloned, and there is serious debate about the cloning of humans. In 2004, cloned mice were derived from the nuclei of olfactory neurons using a method similar to that used by Gurdon. These neurons came from a genetically engineered mouse that expressed the fluorescent protein GFP in most cell types. After the nuclei of a mature (haploid) oocyte was removed, a neuronal nucleus derived from the GFP-mouse was introduced. Blastula derived from these cells were then used to generate totipotent embryonic stem cells from cells of the inner cell mass. A totipotent cell is capable of producing, through cell division

[393] The egg and the nucleus: a battle for supremacy: http://www.nobelprize.org/mediaplayer/?id=1864

[394] see: Individual neurons may carry over 1,000 mutations

and differentiation, all of the different types of cells in the adult. It was the nuclei from these cells that were then transplanted into enucleated eggs. The resulting embryos were able to develop into fully grown fluorescent mice, proving that neuronal nuclei retained all of the information required to generate a complete adult animal.

The process of cloning from somatic cells is inefficient – many attempts had to be performed, each using an egg, to generate an embryo that is apparently normal (most embryos produced this way were abnormal). There are serious ethical issues associated with the entire process of reproductive cloning, particularly given the persistent inequalities in modern society.[395] For example the types of cells used, embryonic stem cells, are derived from the inner cell mass of mouse or human embryos - their isolation involves destroying the original embryo.

In a breakthrough series of studies, Takahashi and Yamanaka (2006) determined that introducing a set of four transcription factors (Oct3/4, Sox2, c-Myc, and Klf4) into terminally differentiated cells led some of the transfected cells to reverse their differentiation, and return to a more pluripotent state, that is a state that can subsequently differentiate into many other cell types.[396] This process of dedifferentiation has been found to be robust, and the dedifferentiated cells produced are known as “induced pluripotent stem cells” or iPSCs. iPSCs behave much like embryonic stem cells. The hope is that patient-derived iPSCs can be used to generate tissues or even organs that could be transplanted back into the patient, and so reverse and repair disease-associated damage.

Questions to answer:

178. What cellular factors determine how (or whether) a cell responds to a particular signaling molecule?
179. What is necessary for cells to become different from one another?
180. Based on your understanding of the control of gene expression, outline the steps required to reprogram a nucleus so that it might be able to support embryonic development.

Questions to ponder:

- Why, if differentiation is normally uni-directional and irreversible, is it possible to artificially reprogram somatic cells to an “earlier” state? Why doesn’t this happen all the time in your body?
- What are the main ethical objections to human cloning? What if the clone were designed to lack a brain, and destined to be used for "spare parts”? Does that change anything, or does it make things worse?

[395] J. Gray. 2017. A History of the Future: how writers envisioned tomorrow’s world

[396] Takahashi & Yamanaka. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.

Part II: From molecular biology to the behavior of genes in organisms

In which we consider the behavior of genes (genetics) during the course of asexual and sexual reproduction, how they interact with one another, both along a chromosome and more generally, within a genome, and how they can be manipulated and studied.

Looking back: Concepts you should be familiar with & need to understand

In which we reflect back on what we have learned and review the meanings and implications of various terms and processes associated with molecular mechanisms, genes, gene products, and cellular reproduction, together with a few key terms that we will be using over and over and that you will encounter repeatedly in the biological sciences.

As you may have noticed by now, biology is full of words. Some of these words are familiar and others are strange or have biological meanings that differ from their common usage. You have probably run into a number of these words. Science is generally referred to as a discipline, and a common feature of a discipline is discipline, meaning that there are strict rules involved, there are ways to act, behave, and speak. The discipline of science in general, and biology in particular, is that what words mean is unambiguously defined, and not subject to personal preferences – we are not allowed to be creative in the meaning of scientific words; this is probably one reason people like to invent new words to describe new phenomena and new ideas. Therefore, we will begin by considering a number of words and what they mean scientifically, as opposed to colloquially. To understand the meaning of a word, you have to be able to use it correctly, apply it when appropriate, and understand its implications. Don't be afraid to ask a question if you are unsure whether a particular word is appropriate to a particular situation.

To review let us consider what we mean by a gene**.** Within a cell, a gene is a stretch of DNA that contains information that can be expressed**.** A gene includes the DNA sequences involved in determining where, when, and (in part) how strongly the gene is expressed. So what does it mean to be "expressed"? To say that a gene is expressed, we mean that an RNA molecule, complementary to the sequence of one of the two DNA strands in a region, is generated through the process of DNA-dependent and directed RNA synthesis, a process also termed transcription. Transcription is mediated through the action of DNA-dependent, RNA polymerases. Where such enzymes bind to the DNA and act, that is, where RNA synthesis starts and in which direction it proceeds, is determined by where specific sets of transcription factors (proteins) bind to specific sequences within the DNA. Transcription factor binding sites are part of the gene's regulatory sequences. The binding of transcription and accessory factors act to recruit and activate an RNA polymerase molecule. A gene's regulatory sequences can be located near to the gene's transcribed region (generally referred to as the gene's promoter) or at more distant sites, known as enhancer elements.

The sum of all the DNA molecules within a cell constitutes the cell's genome. Generally, each cell in an organism contains a full copy of the genome.[397] The cell's genome is characteristic of the type of organism (the species) and unique to the individual. Genomic differences impact and in some cases determine the phenotypic variations observed between individuals. Organisms of the same species have extremely similar, but not identical, genomes, more similar than do organisms of different species. As an example, when comparing different human beings, they share more that 99.9% of the same DNA sequences (although given that each person has ~6,000,000,000 base pairs of DNA, there are still lots of differences).

Within an organism, a cell can have either one complete genomic copy, in which case it is known as haploid, or it can contain two copies, in which case it is known as diploid. In certain cases, a cell can have more than two copies of its genome, in which case it is termed triploid (three copies), tetraploid (four copies), or polyploid (> four copies). The cells within an organism can differ from one

[397] In mammals the exception are the red blood cells, which have lost their DNA during the process of their formation.

another due to sequence differences that arise during the course of DNA replication or from the failure to accurately repair spontaneous mutations.[398]

As we will consider further, a haploid or a diploid cell can divide asexually to produce two haploid or diploid cells, respectively. So what does sexual and asexual mean, exactly? Asexual reproduction involves a single cell, a single individual. The genome of the cell is duplicated through the process of DNA replication (which is mediated by DNA-dependent, DNA polymerases and other factors) and then the cell splits into two. Similarly, at the organismic level, asexual reproduction requires no need for cooperation between different organisms, or different cells. Of course, for such a process to continue there has to be growth of the cell between cell division events. Generally the cell doubles in volume and mass between one division and the next. The growth of the cell involves the import of energy and other materials into the cell, and their metabolic transformation into various cellular parts, proteins, nucleic acid polymers, lipids, etc. There is a continuity, one cell becomes two; this is the simplest version of the cell theory of life.

The process of sexual reproduction is more complex. Two different cells, generally but not always from two different organisms, have to find and fuse with one another. Such cooperation requires them to recognize each other as appropriate fusion partners. Sexual reproduction involves a diploid cell that first generates a number of haploid cells, known as gametes, through a process known as meiosis in eukaryotes - other processes mediate related processes in prokaryotes (bacteria and archaea). Typically, gametes from two different organisms come into proximity through the process of mating. Their initially distinct plasma membranes become one, they fuse, thereby forming a new diploid cell, a new organism. Some people might say that this is when life begins, but they would be confused, or perhaps better put, inaccurate – life began ~3.4 billion years ago. Both gametes are alive, as is the zygote, the cell formed by their fusion. That said, the fusion of gametes generates a genetically distinct (and so new) organism and is an unambiguous event.

The two modes of reproduction have different characteristics. In a purely asexual organism the various versions of genes, known as alleles***,*** within a cell evolve together, as a group - there is no simple way to remove deleterious alleles from future progeny, although the processes of horizontal gene transfer, that is, transformation, conjugation, or transduction, common in prokaryotes, can modify genomes. In contrast during sexual reproduction, the process of meiotic recombination enables alleles to be "disconnected" from one another. Sexual reproduction is also associated with a number of features, particularly in multicellular organisms. Sexual dimorphism means that the two gametes, and the organisms that produce them, can be different in morphology and behavior. Such differences can lead to sexual selection, a distinctive process associated with the evolution of a range of traits and evolutionary implications.[399]

Questions to answer and ponder:

– Make a list of all the bio-words you can think of, can you define what each one means?

181. How are transcription and translation (RNA directed polypeptide synthesis) similar, how are they different?
181. Within a gene, what signals and signal binding proteins are involved in gene expression? make a diagram.
182. How might having two copies of a gene (in a diploid cell) alter the effects of a mutation or the cell's behavior?

[398] We will ignore the directed mutational events that can occur within the vertebrate immune system.

[399] here is an interesting book on the topic: The Mating Mind by Geoffrey Miller.

Where do genes, alleles, and mutations come from?

When we think about genes, there are, to start, two issues to consider. The first is where do genes come from? The most obvious (and perhaps unsatisfying) answer is that our genes come from our ancestors, our parents through the processes of DNA replication, cell division, and for sexual organisms, cell fusion. Unfortunately, this leaves the ultimate origin of genes shrouded in mystery. As discussed earlier, all life on Earth appears to be descended from a last universal common ancestor (LUCA). LUCA had lots of genes, genes that arose even earlier, through processes involving molecular systems active before the appearance of LUCA. New genes have been observed to appear *de novo* in various organisms, in particular the fruit fly *Drosophila*.[400] Perhaps even more surprising, many of these *de novo* (new) genes appear to have become essential rather quickly.[401] A number of putative *de novo* genes have been identified in humans.[402]

Once DNA (nucleic acid) molecules and genes existed, new versions of genes (alleles) can appear through processes of mutation and recombination, which lead to alterations in DNA sequence. Moreover, an existing gene can give rise to new copies of itself through the process of gene duplication, leading to the production of what are known as paralogs. Genes can also disappear through gene deletion or loss. A number of studies, beginning with the classic Luria-Delbrück experiment (which we will discuss in detail), indicate that these processes, that is, mutation, recombination, deletion, and duplication occur stochastically, based on the molecular nature of DNA, various molecular mechanisms active in cells, and environmental effects (chemicals and radiation). Mutations appear by chance and not to meet the adaptive needs of the organism. Once a mutation arises it can, however, effect phenotype, that is the traits displayed by an organism. These phenotypic effects can include effects on reproductive success. The most severe of such effects is lethality or sterility, generally arising because the mutation inactivates an essential gene, a gene whose activity (gene product) is necessary for the organism's survival, that is the maintenance of life, or its ability to produce offspring that are themselves viable and fertile. Evolutionary processes act to "select" against mutant alleles that reduce reproductive success (negative selection) and increase the frequency of mutant alleles that improve it (positive selection). Of course, the rest of the genome influences the extent to which an allele has positive or negative selective effects. Generally, environmental factors and preexisting adaptations and behaviors determine the selective pressures on a new allele. There are also processes, such as genetic drift together with founder and bottleneck effects, that can influence which alleles are found within a population. These principles apply both to the cells within a multicellular organism (somatic selection) as well as organisms within a population.

Alleles

Each gene is characterized by a specific region of DNA, a "locus", a position or place within the genome, with a specific nucleotide sequence.[403] Versions of a gene with different DNA sequences are known as alleles; determining which of these differences "matter" biologically, that is are associated with recognizable phenotypes can be non-trivial. In a diploid organism, the two copies of the gene present can have different sequences, they can be different alleles. If the two alleles in a diploid organism are the same, the organism is said to be homozygous for that gene, if they are different it is said to be heterozygous for that gene. An organism can be homozygous for

---

[400] see: Schlotter, 2015. Genes from scratch – the evolutionary fate of de novo genes and Fact or fiction: updates on how protein-coding genes might emerge de novo from previously non-coding DNA.

[401] see New genes in Drosophila quickly become essential and The Goddard and Saturn Genes Are Essential for Drosophila Male Fertility and May Have Arisen De Novo.

[402] De novo origin of human protein-coding genes

[403] Although exactly what is a gene can get complicated - see Portin & A. Wilkins (2017). The evolving definition of the term "gene". Genetics 205: 1353-1364.

some genes and heterozygous for others. If an organism is homozygous for all genetic loci, it is generally the result of extensive in-breeding. Different alleles can be expressed differently, due to differences in their regulatory sequences, and they can encode different gene products due to differences in where transcription starts and differences in their coding (in cases where the gene encodes a polypeptide) regions, as well as differences in RNA splicing (in eukaryotes).

Within a particular population, there may be only a few or many different alleles present at a particular genetic locus (gene). Some alleles are predicted to lead to a "loss of function" of the gene, a failure to produce a functioning gene product. Within a population, the absence of such loss of function"alleles is often taken as evidence that the gene's normal function(s) is essential for the survival or reproduction of the organism. Later on we will learn to use the "On-Line Inheritance in Man" (OMIM) and other public genomic data sites to get information on genetic variations and their effects. Closely related species often share many genes, organized along chromosomes in similar patterns, a situation known as synteny, something that can be visualized using the Genomicus web tool. Different species are likely to have different alleles (and some different genes), a result of their divergent evolutionary histories; and these genes can be located in different molecular neighborhoods in the genome. Genes can be deleted, duplicated, or moved to different chromosomal positions within the genome (genomic rearrangements). New genes can appear and conserved genes can disappear. Some of the differences between alleles have little or no impact on the function of a gene or the gene product that it encodes, these allelic variants can all be considered normal or wild type. In contrast, other alleles are associated with or contribute to specific traits, or versions of a trait – in some cases these are traits associated with disease, disease susceptibility, developmental defects, or cellular and organismic lethality. In other cases, they are associated with evolutionary novelties, the traits that distinguish one species from another. A mutation in a wild type allele is much more likely to lead to a defect than an improvement in the gene product's function or a useful new trait, but such beneficial mutations do occur; they appear, together with other environmental and selective factors, to drive evolutionary processes.

Phenotypes

The traits of an organism, including how it develops and responds to its environment, are determined, constrained, or influenced by its genome, that is all of the genes it contains, and how the genome interacts with the cellular state. The various regulatory interactions that occur between genes, gene products, and the cells' metabolic processes are known as its epigenome. The epigenome includes non-DNA sequence components, including how DNA is modified and packaged within the cell through interactions with various molecules. Epigenetic factors often influence which genes are, or can be, expressed in a particular cell type, or in response to particular signals. As we will explore in detail, all of the observable or measurable aspects of an organism constitute its phenotype. Phenotypes can range from blood type, allergic reactions, susceptibility or resistance to disease, height, skin color, eye color, the speed of reflexes, or various behaviors in various situations – essentially anything and everything about an organism that you can observe and measure objectively. In some, relatively rare cases there is a 1 to 1 correspondence between which allele of a gene an organism carries and the specific trait(s) it displays. This type of allele:trait association was used by Gregor Mendel to establish his rules of inheritance. More often, however, many alleles together with stochastic factors and environmental influence, combine to produce the organism's phenotype.

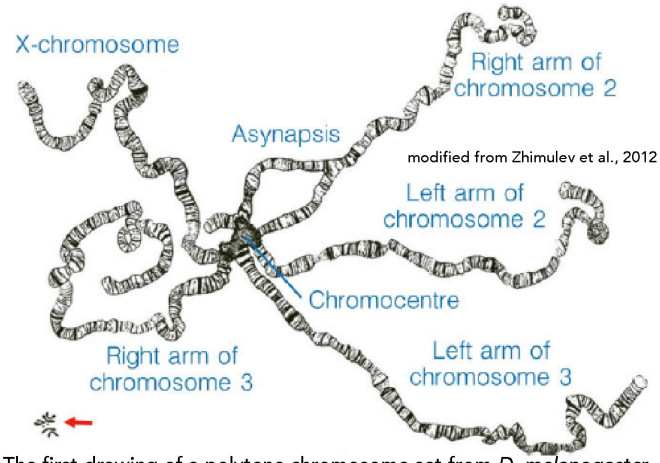
An example that we will consider in detail is antibiotic resistance in bacteria. A bacteria that contains a functional copy of a gene that confers resistance to an antibiotic is resistant to that antibiotic. A mutation that inhibits the antibiotic resistance gene's expression or leads to an inactive gene product will leave the bacteria susceptible to the antibiotic. Of course it is a mistake to think that the gene and the product that it encodes are the only components needed for antibiotic resistance; no gene acts alone – for a gene to influence a phenotype (such as antibiotic resistance) the gene needs to be recognized and expressed (transcribed), the encoded protein synthesized

(translated), and delivered to the right location (targeted). Even a simple gene (allele)→phenotype relationship is based on the functioning of the complex underlying biological system, a system composed of hundreds to thousands of gene products. Most traits are based on many gene products, and often the impact of a particular allele of a particular gene is subtle, something that can be identified through complex molecular genetic studies, which we will consider anon. The relationship between an geno type and a phenotype is more complex in a diploid organism since there are two copies of most genes. The two copies of a particular gene can be the same or different, and in some cases only one allele may be expressed in a particular cell. Different versions of the same gene/gene product can interact in various ways.

Now consider a trait that is associated with the presence of a particular allele. If the trait is visible when the locus is heterozygous for that allele, the allele is referred to as dominant to whatever the other (different) allele might be. On the other hand, if the trait is not apparent when the locus is heterozygous, but is visible when the locus is homozygous for the allele, it is referred to as recessive. Finally, if the trait displayed by an organism that is heterozygous for a particular locus is different from either of the homozygous versions, the alleles are referred to as co-dominant or semi-dominant. In such cases, the nature of the phenotype observed will depend on exactly which alleles are involved. One point to keep in mind is that an allele can be dominant for one trait and recessive or semi-dominant for others. In addition, the extent and the appearance of a phenotype, known as its penetrance and its expressivity, can be influenced by the other alleles within the genome, the organism's genetic background. Remember however, the terms recessive and dominant generally refer to alleles that are associated with simple and visible traits. Most alleles are neither strictly recessive nor dominant, and contribute in complex ways to a number of measurable traits. Because it is easier to make sense of things we will generally start, at least initially, with strict dominant and recessive alleles, and then get more complex in order to consider the molecular mechanisms that connect genotype to phenotypes.

Questions to answer and ponder:

182. What types of mutation might you predict would lead to a "loss of function" of a gene?
184. Draw out (schematically) the relationship between a specific allele and its molecular effects. Why might the relationship between mutation (allele) of gene not be associated in a straightforward way with a specific phenotypic change?



The first drawing of a polytene chromosome set from *D. melanogaster* made under light microscopy by T. Painter in 1934. The chromosomes have been spread by squasking them on a microscope slide. Each parental chromosome is tightly paired with its homolog, and there are regions where two homologous chromosomes are separated (asynapsis). All the chromosomes are linked together by the pericentromeric regionto create a single "chromosome center"/\. In the lower left corner, mitotic chromosomes (indicated by arrows) from ovarian tissue are shown at the same magnification, demonstrating the giant size and transveral banding in polyploid polytene chromosomes.

Muller's Morphs

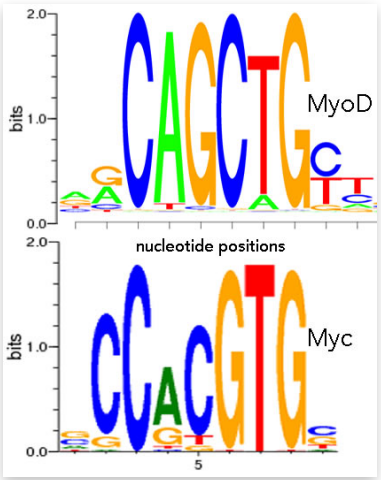
Another way to look at alleles is from a functional perspective. This was the approach taken by Herman J. Muller (1890-1967) in the 1920s and 30s. He exploited work done in the fruit fly *Drosophila*. Geneticists had isolated a number of chromosomal duplications and deletions, something made possible by unique aspects of chromosome organization in the salivary glands of the fly (←). These cells are polyploid; each chromosome contains more than 1000 double-stranded DNA molecules lined up from end to end.[404] Based on the analysis of various mutations he was able to place mutations into distinct functional (with respect to a particular phenotype) groups: that is amorphic, hypomorphic, hypermorphic,

[404] Banding patterns in Drosophila melanogaster polytene chromosomes correlate with DNA-binding protein occupancy.

antimorphic, and neomorphic. These classes are compared to the wild type ("normal") version of the gene. It is, however, worth keeping in the back of your mind that a particular gene (and gene product) may have more than one functional role, and a particular mutation may influence these different functions differently, it may be associated with different phenotypic effects As an example, an allele could be hypomorphic for one trait and antimorphic for another. At this point we will not consider mutations that have no phenotypic effects.

Compared to the functional gene product produced by a wild type allele, an amorphic allele has no function - it might not be expressed, or if expressed the gene product may not carry out the trait-specific functions of a wild type gene product. Importantly, an amorphic allele does not interfere in any way with the expression or functioning of the wild type gene product encoded by the other allele in a diploid cell. Amorphic alleles are also known as null or loss of function (LoF) alleles. In a similar manner, a hypomorphic allele has less functional activity, whatever that might be, compared to a wild type allele, whereas a hypermorphic allele has more, but the same, functional activity as the wild type allele. Again, for both hypo- and hypermorphic alleles, the mutant gene product does not interact with the wild type gene product. In contrast, an antimorphic allele is not only non-functional with respect to a trait-specific function, but it interacts with and inhibits the activity of the wild type gene product.

The final class of mutation (allele) is known as neomorphic; it changes the activity of the gene product, producing a new (neo-) function. There are a number of ways a new function can be generated by a mutation. As an example the mutation can change the specificity of an enzyme, something that can happen in the course of the development of cancer.[405] To illustrate one such neomorphic mutation, consider the transcription factor MyoD, a protein that regulates the formation (differentiation) of skeletal muscle cells. There are mutations (alleles) of the *MyoD* gene associated with an aggressive form of embryonal rhabdomyosarcoma, a cancer of skeletal muscle. One missense mutant allele changes the DNA sequence so that the leucine present at position 122 of the wild type MyoD protein is replaced by an arginine.[406] So what is the effect of this change in the MyoD protein? To understand, you need to remember that MyoD is a transcription factor, a protein that recognizes specific sequences in DNA and, when bound to such sites, leads to a change in gene expression. The wild type MyoD protein recognizes and binds to a consensus sequence (top panel →); in contrast the mutant allele encodes a protein whose DNA sequence binding specificity is altered (bottom panel →); it now binds better to a sequence that is also recognized by the transcription factor Myc. Myc regulates genes associated with active cell division. The result is that a gene product that normally inhibits cell division and encourages the formation of non-dividing muscle cells (MyoD), acquires a new function, the ability to bind to different DNA sequences, turning on different sets of genes, and inducing (aberrant) cell division – a key feature of cancer cells. The mutation is neomorphic because the mutated MyoD protein (known as MyoDAla$_{122}$→Arg) has a new function, and (probably) weaker binding to its original target sequence.[407]



It is worth noting explicitly, that the relationship between the type of mutation (in Muller's terminology) and recessivity or dominance is not simple. An amorphic allele could be dominant, a behavior known as halpoinsufficiency, arising because one copy of the gene does not produce the necessary amount of the gene product, or it can be recessive, if one functional copy of the gene is

---

[405] Neomorphic mutations create therapeutic challenges in cancer

[406] from Myc and MyoD and Deep Sequencing of MYC DNA-Binding Sites in Burkitt Lymphoma

[407] We will return to this topic toward the end of book: see Neomorphic mutations create therapeutic challenges in cancer

sufficient to produce the phenotype.

Before we move on, let us consider (again) the effects of mutations in a coding region of a gene. We have already mentioned missense mutations, mutations that lead to the replacement of one amino acid by another, different amino acid. There are mutations that do not change the amino acid sequence of the encoded polypeptide, but change the DNA sequence – these are known as synonymous mutations, and as will see such mutations produce what are known as single nucleotide polymorphisms (SNPs), a feature in the DNA that can be detected by various molecular methods. SNPs are often used in the analysis of genomic similarities and differences, including human ancestry. There are two other types of generic terms for alleles. A non-sense mutation leads to a stop codon replacing a sequence encoding an amino acid in a polypeptide. Non-sense mutations lead to the premature truncation of the encoded polypeptide; their effects on gene function often depend upon where they occur within the gene. Another type of mutation leads to the insertion or deletion of one or more nucleotides from the gene sequence (known generically as indels for "insertion/deletion"); these can lead to a range of effects. In eukaryotic genes, which can have many exons and introns, there can be mutations that disrupt the sequences involved in recognizing and removing introns from newly synthesized RNAs. These are generally referred to as splice-site mutations; the processing of a newly synthesized RNA to generate an mRNA involves splicing out (removing) of the introns before the RNA is transported from the nucleus to the cytoplasm. Depending upon their effects on the final encoded polypeptide, indels, non-sense mutations, and mutations that alter an intron-exon junction can result in frame-shift mutations, mutations that alter the wild type reading frame and lead to multiple changes in polypeptide sequence and pre-mature termination. These can lead to any one of Muller's morphs depending upon the exact nature of the mutation and gene. Similarly, such mutations can produce either recessive or dominant alleles. Finally, it is worth remembering that essentially all traits are dependent upon a number of gene products, and so are polygenic, whereas a particular gene product may have a functional role in a number of processes; its mutational alteration can influence some or all of these processes, in which case it is considered pleiotrophic.[408] Don't get confused, all biological processes are complex, it is just that (occasionally) some alleles in some genes generate easily recognizable (distinctive) phenotypes.

Questions to answer

185. Draw out the relationship between gene→RNA→polypeptide→protein, and describe the effects of mis-sense, non-sense, frame-shift, and intron-exon junction mutations on gene expression.
186. Can you produce some "rules of thumb" relating the position of a mutation within a gene to their effects on the gene product's function?
187. Why is the MyoD mutation neomorphic? What would you call it, if the mutated MyoD protein blocked the binding of wild type MyoD to its target DNA sequences but failed to activate transcription?
188. Describe how a DNA change could produce the various Muller's morphs.
189. Describe how a neomorphic mutation might alter the behavior of transcription factor or an enzyme.

Questions to ponder

- A *Drosophila* polytene chromosome can have over 1000 DNA molecules (strands). How, do you imagine, does the banding pattern observed in these polytene chromosomes relate to the genes on the chromosome?
- How does the polyploid nature of these chromosomes make visualizing chromosomal duplications and deletions possible? What are its limits, do you think?

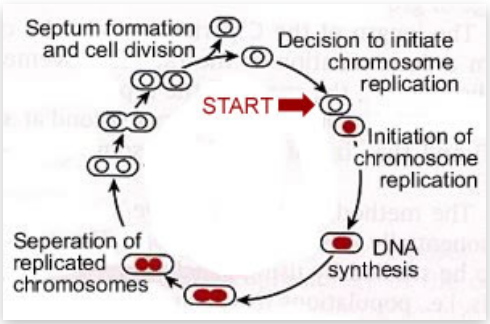
[408] Pleiotropy: One Gene Can Affect Multiple Traits

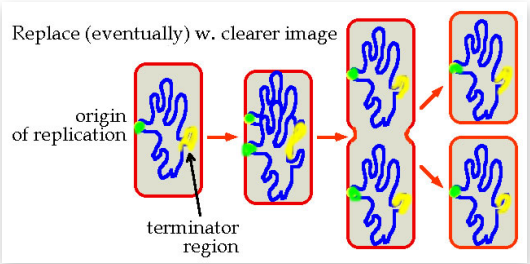
Chapter 11: Reproduction in prokaryotes & horizontal gene transfer

In which we consider how prokaryotic, that is bacterial and archaeal, cells replicate asexually, and how they can (under specific conditions) pass genetic information to one another and acquire such information from their environment.

Asexual reproduction in bacteria and archaea

The simplest type of biological (cellular as opposed to viral) reproduction is probably the asexual process found in prokaryotes. In both bacteria and archaea, the genome typically consists of a single large circular DNA molecule, known as the bacterial chromosome. In some cases, the cell also contains smaller circular DNA molecules, known as plasmids. For the moment we will ignore plasmids and focus on the chromosome.[409] The chromosome contains two important DNA sequence elements, the origin of replication (ORI) and the terminator region (TER). When conditions are appropriate, a cell will pass through a decision point, a molecular switch, known as "start" (→). This switch activates the proteins that bind to the ORI region of the chromosome, and initiates the assembly of the DNA replication complex, a molecular machine known as the replisome. A replication bubble (a region of the DNA in which the two strands have separated) forms, and the replication forks begin to move around the DNA molecule, making a copy. As the ORI sequence is replicated, the two ORI sites remain associated with the plasma membrane. The replication forks move around the DNA molecule, and collide in the TER region (←). The collision of the DNA replication forks generates a signal that indicates that DNA replication is complete. During this period the cell is growing, adding mass and volume. The division of one cell into two is mediated by the formation of a septum, an extension of the plasma membrane and the cell wall. Septum growth initiates between the two membrane-bound ORI sequences, which insures that each daughter cell receives one complete chromosome, one total genome.





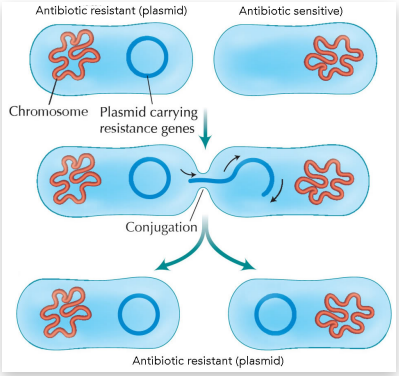
If we consider the chromosome itself, it is worth noting that the order of genes around the circular molecule is conserved between organisms of the same species. The genes along the chromosome constitute a syntenic linkage group, the same genes in the same order along a chromosome (discussed further below). In the standard asexual mode of replication, all of the alleles are inherited together, the result is that a mutation in any particular gene (generating a new allele) acts in concert with the other alleles (in the other genes) present. Over time, each organism produces a clone, and various clones interact with the environment and each other. These clones can display different levels of reproductive success; some clones can take over the population, while others may become extinct. In the case of studies on the evolution of bacterial antibiotic resistance (see below), each clone has to develop antibiotic resistance independently of every other clone; a

[409] Noirot-Gros et al., 2002. An expanded view of bacterial DNA replication

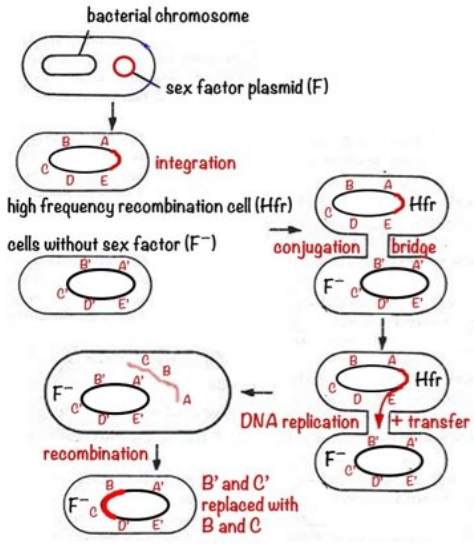
similar situation was observed in long term bacterial evolution studies.[410] There is no cross talk between lineages in such situations. Of course, if DNA is passed from clone to clone, as occurs within Griffith's (previously considered) transformation experiments, things get more complex. The movement of genes between lineages is known as horizontal gene transfer. We will consider the three versions of horizontal gene transfer found in prokaryotes.

Conjugation: what counts as sex in prokaryotes

Conjugation is a major pathway for horizontal gene transfer in bacteria.[411] In contrast to transformation, conjugation "forces" DNA into what may be a reluctant recipient cell. In the process of conjugation, we start by distinguishing between two types of bacterial cells (of the same species). One contains a DNA sequence known as the fertility (or sex) factor (F), the other does not and is referred to as a F⁻ cell. The F factor can exist independently of the host chromosome as a plasmid. The F plasmid, discovered by Esther Lederberg (1922–2006), was the first plasmid discovered. Cells in which the F-plasmid is integrated into the host chromosome are known as a high frequency recombination (Hfr) cells. We start by considering the situation in which a cell contains a free F plasmid. The ~100 kilobase F plasmid contains ~100 genes that encode the proteins needed to transfer a single-stranded copy of its DNA into a cell that lacks an F-plasmid.[412] In this manner, an F-plasmid can colonize a population of F⁻ cells. F-type plasmids often include genes that encode an addiction system. Such systems encode a stable toxin and an unstable (rapidly degraded) anti-toxin. Once the plasmid enters a cell, both toxin and anti-toxin proteins are synthesized. If the plasmid is lost, the cell dies because of the anti-toxin disappears before the toxin, leading to toxin activation and cell death.



The F-plasmid contains two distinct origins of replication - one, known as oriV, is involved in normal replication during cell growth and division. The second, known as oriT, is involved in generating the single stranded DNA molecule that is transferred into the recipient cell. To initiate conjugation, the F⁺ cell makes a physical (conjugation) bridge, known as a pillus, to the F⁻ cell (→). A single stranded copy of the F plasmid is synthesized and transferred through the pillus into the recipient F⁻ cell. Subsequent DNA synthesis generates a double-stranded copy of the F-plasmid in the recipient cell, while the donor cell retains the original plasmid.



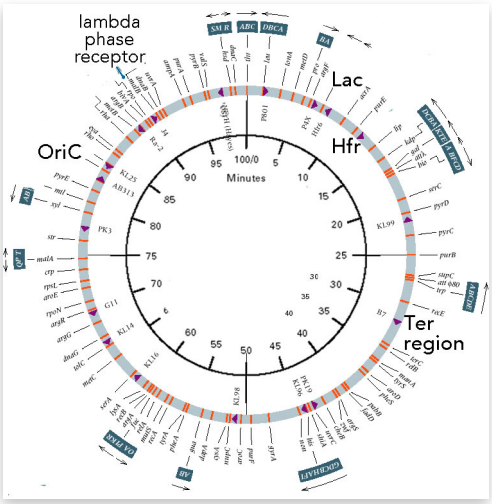
In Hfr cells (←), integration of the F-plasmid can occur at various points along the host chromosome. As with the free plasmids, the integrated F-plasmid can initiate (at its oriT site) the transfer of its own as well as linked host genes into a F⁻ cell. The amount of DNA transferred will be determined largely by how long the bridge between the cells remains intact. In *E. coli* it takes ~100 minutes to transfer the entire donor genome (chromosome) from an Hfr to an F⁻ cell. Once inside the F⁻ cell, the transferred donor DNA will be integrated, via homologous recombination, into the recipient's chromosome, replacing the recipient's versions of the genes transferred (a process to which we will return). Using Hfr strains carrying

[410] see A cinematic approach to drug resistance and E. coli Long-term Experimental Evolution Project

[411] review of prokaryotic conjugation and Pull in and Push Out: Mechanisms of Horizontal Gene Transfer in Bacteria

[412] fertility factor review by S.M. Rosenberg & P.J. Hastings 2001.

different alleles of various genes (associated with recognizable phenotypes), and by controlling the duration of conjugation by breaking the conjugation bridge by shearing the cells in a kitchen blender, experimenters were able to determined the order of genes along the bacterial chromosome. The result was the discovery that related organisms had the same genes arranged in the same order.[413] The typical drawing of the circular bacterial chromosome is like a clock going from 0 to 100 (→), with the genes placed in their respective positions, based on the time it takes to transfer them in minutes.



If the entire F-plasmid sequence is transferred, the original F⁻ cell becomes an Hfr cell. If the Hfr cell loses the F-plasmid sequence, it reverts to a F⁻ state. The end result of the conjugation process is similar to that obtained in sexual reproduction in eukaryotes, namely the original F⁻ cell now has a genome derived in part from itself and from the "donor" Hfr cell. Because the outcome of an Hfr/F⁻ cell interaction can lead to a cell with a different set of alleles than either of the "parental" cells, this process is often referred to as bacterial (prokaryotic) sex, although it is really quite distinct from sexual reproduction in eukaryotes.

Versions of this process are involved in the transfer of plasmids from cell to cell within a community.[414] All plasmids contain an "origin of replication"; some (low copy number) plasmids exist in one to two copies per cell, while high copy number plasmids may be present in as many as ~700 copies per cell.[415] Which is which is determined in large part by their origin of replication sequences. Plasmids can encode genes responsible for antibiotic resistance and the rapid dispersion of the antibiotic resistance phenotype is a cause of increasing concern.[416] Many plasmids, also known as mobile genetic elements, are selfish, that is, their presence in a cell may not directly benefit that cell and the loss of the plasmid may result in the death of the host cell, due to the presence of an addiction module.

Questions to answer:

190. What factors act to insure that each (prokaryotic) cell generated contains a complete genome?
191. How would mutating the origin or terminator regions influence the cell's reproduction?
192. How would the progeny of an incomplete F-factor mediated recombination event differ from its "parents"?
193. Generate a mechanism that could produce a high copy plasmid, and how might a mutation lead to a low copy plasmid.

Questions to ponder:

- How might a "selfish" plasmid evolve into a virus.

---

[413] Synteny: http://en.wikipedia.org/wiki/Synteny

[414] Plasmids Spread Very Fast in Heterogeneous Bacterial Communities: https://www.ncbi.nlm.nih.gov/pubmed/12524329

[415] Plasmids 101: Origin of Replication

[416] Addgene: Mechanisms of Antibiotic Resistance

Other naturally occurring horizontal gene transfer mechanisms

Many horizontal transfer mechanisms are regulated by social and/or ecological interactions between organisms.[417] It is worth noting that the mechanisms involved can be complex; one could easily imagine an entire course focused on this topic alone. We introduce only the broad features of these systems. Also, we want to be clear about the various mechanisms of DNA uptake. First recognize that when an organism dies its DNA can be eaten by others as a source of energy, as well as carbon, nitrogen, and phosphorus. When eaten, any information in the DNA, the result of mutation and selection, is lost.[418] Alternatively, the nucleotide sequence of a DNA molecule can be integrated into another organism's genome, resulting in the possible acquisition of whatever information developed (evolved) within that lineage. This is information that might be useful, harmful, or irrelevant to the organism that acquires it. The study of these natural DNA import (as distinct from direct conjugation-mediated transfer) systems has identified specific molecular machines that mediate DNA transfer. Some organisms use a system that preferentially imports DNA molecules derived from organisms of the same or closely related types as themselves. You can probably even imagine how they do this – one way could be that they have receptor systems that recognize species-specific "DNA uptake sequences." The various mechanisms of horizontal gene transfer, unsuspected until relatively recently, have had profound influences on evolutionary processes, particularly among microbial communities, where they appear to be more common than in eukaryotes. It turns out that, in many cases, a population of organisms does not have to "invent" all of its own genes, it can adopt (import) genes generated by evolutionary mechanisms in other organisms in other environments for other purposes. So the question is, what advantages might such information uptake systems convey, and (on the darker side), what dangers do they make possible?

Transformation

There are well established methods, used in genetic engineering, to enhance the ability of bacteria to take up DNA from their environment.[419] We, however, will focus on natural transformation, the process associated with the transfer of DNA molecules from the environment into a cell. Natural transformation is an active (energy-requiring) process that involves a number of components, encoded by genes that can be expressed or not depending upon environmental conditions. Consider a type of bacteria that can import DNA from its environment. If the density of bacteria is low, there will be little DNA to import, and it may not be worth the (energetic) expense associated with expressing the genes and synthesizing and assembling the proteins involved in the DNA uptake and integration machinery. Bacteria use quorum sensing systems (considered earlier) to monitor cell density and to control the expression of genes involved in synthesis of the DNA uptake system. When present in a crowded environment, the quorum sensing system can turn on the expression of the genes involved in the assembly of the DNA uptake system.

Here we outline the process in one type of bacteria but functionally similar mechanisms are used in other bacterial and archaeal species. Double-stranded DNA binds to the cell's surface through a variety of DNA receptor proteins (themselves the products of genes). In some cases these receptors bind specific DNA sequences, in others they bind DNA generically, that is, any DNA sequence. As shown, Gram negative bacteria have two lipid membranes, an outer one and an inner (plasma) membrane, with a space, known as the periplasmic space, between them. In an ATP-hydrolysis coupled reaction, DNA bound to the exterior surface of the bacterium is moved, through a protein

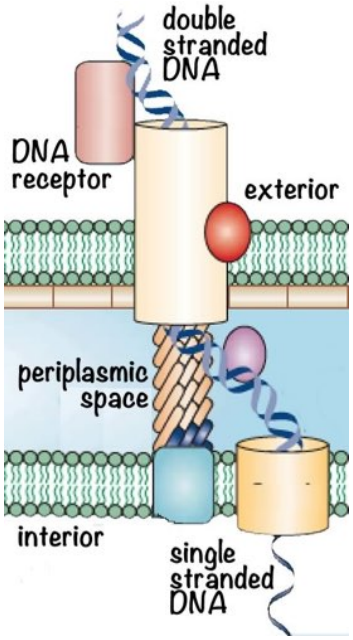
---

[417] DNA uptake during bacterial transformation: http://www.ncbi.nlm.nih.gov/pubmed/15083159

[418] This is of course why genes are rarely if ever transferred from food to the organism doing the eating.

[419] Making Calcium Competent (bacterial) Cells: http://mcb.berkeley.edu/labs/krantz/protocols/calcium_comp_cells.pdf

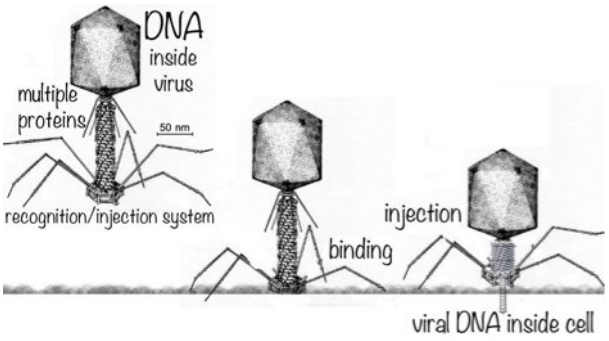
pore across the outer membrane and into the periplasmic space, where it is passed to the DNA channel protein (→). Here one strand of the DNA is degraded by a nuclease while the other moves intact through the channel into the cytoplasm of the cell in a 5' to 3' direction (similar to the one-strand transfer seen in bacterial conjugation). Once inside the cell, the DNA associates with specific single-stranded DNA binding proteins and, by homologous recombination, it is inserted into the host genome (or degraded, depending on the system).[420] While the molecular details of this and functionally similar processes are best addressed elsewhere, what is key is that transformation enables a cell to decide whether or not to take up foreign DNA and whether to add the imported DNA sequences to its own genome.

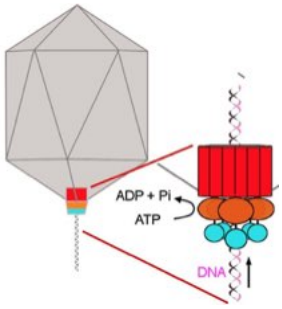


Viruses moving genes: transduction

The final form of horizontal gene transfer that we will consider involves viruses. The structure and behavior of viruses is a complex topic, the details of which are largely beyond us here, but it is not unreasonable to consider viruses as nucleic acid transport machines. Viruses are completely dependent for their replication on the infected host cell, they have no active metabolic processes and so are not alive in any meaningful sense of the word, although they can certainly be infectious, that is they can spread through a population. Viruses cannot be killed, because they are not alive, but they can be inactivated by various treatments.

The simplest viruses contain a nucleic acid genome and a protein-based transport and delivery system. We briefly consider a typical bacterial virus, known as a bacteriophage or bacteria eater. The bacterial virus we consider here, the T4 bacteriophage, looks complex and it is (→), other viruses are simpler. The T4 phage (short for bacteriophage) has a ~169,000 base pair double-stranded DNA genome that encodes 289 polypeptides, almost as many as a minimal cell (see above).[421] The assembled virus has an icosahedral protein head that contains a DNA molecule attached to a tail assembly that recognizes and binds to target cells. Once a suitable host is found, based on tail binding to cell surface molecules, the tail domain attaches to the cell's surface and contracts, like a syringe, punching a hole through the cell's external wall and plasma membrane. The DNA emerges from the bacteriophage and enters the cytoplasm, infecting the cell. Genes within the phage genome are expressed, leading to the replication of the phage DNA molecule and the fragmentation of the host cell's genome.[422] The phage DNA encodes the proteins that are used to assembled new phage heads. DNA is packed into these heads by a protein-based DNA pump (←), a pump driven





[420] Bacterial transformation: distribution, shared mechanisms and divergent control & Natural competence and the evolution of DNA uptake specificity

[421] http://en.wikipedia.org/wiki/Bacteriophage_T4

[422] An infected bacterial cell can protect is neighbors, often its clonal relatives, if it can kill itself before the virus can replicate. This is an example of a simple altruistic behavior.

by coupling to an ATP hydrolysis reaction complex.[423] In the course of packaging viral DNA, the system will, occasionally, make a mistake and package a fragment of the host cell's DNA. When such a phage particle infects another cell, it can inject that cell with a DNA fragment derived from the previous host. The mis-packaged DNA may not contain all of the genes the virus needs to make a new virus or to kill the host. If this is the case, the host cell may have to be co-infected by a wild type virus for the mutant virus to replicate. The DNA transferred by the virus to the host can be inserted into the host cell genome, with the end result being similar to that discussed previously for transformation and conjugation. DNA from one organism is delivered to another, horizontally rather than vertically.

Because the horizontal movement of DNA is so common in the microbial world, a number of defense mechanisms have evolved to control it.[424] These include the restriction endonuclease / DNA modification systems used widely for genetic engineering, and the CRISPR-CAS9 system, which enables cells to recognize and destroy foreign (viral) DNAs. These systems, evolved as part of prokaryotic immune systems, together with various plasmids, form the tools used in modern molecular biology and genetic engineering methods. They illustrate how studying apparently arcane aspects of the biological world, bacterial viral defense mechanisms, can have dramatic impacts on modern technological, medical, and economic systems.

Questions to answer:

194. What is an asexual clone? How would you recognize it.
195. What is the effect of an amorphic allele / mutation on the behavior of a prokaryotic clone.
196. What are some possible (evolutionary) advantages to the ability to take up and integrate, as opposed to simply eat foreign DNA?
197. Why might the "source" of foreign DNA matter?
198. Present a plausible model that would identify host from foreign DNA
199. Propose a model by which a "selfish" plasmid might evolve into a virus.
200. How can co-infection of a cell with wild type virus "rescue" a virus that has lost some of its essential genes?
201. How might inserting a piece of DNA into a bacterium's genome be harmful

Questions to ponder:

- Describe a mechanism by which a prokaryotic organism might protect itself from invading viruses?
- How is it that "punching a hole" in a membrane (during DNA uptake or phage infection) does not kill the cell?
- How does vertical differ from horizontal inheritance?

Possible extension:

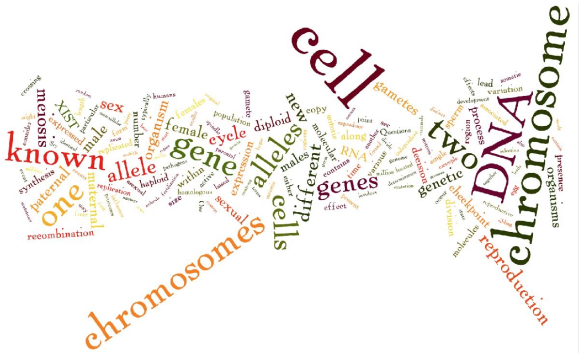
- Introduce and consider the role of the lysogenic / lytic switch in bacteriophage / bacterial interactions.
- Extend discussion to mobile genetic elements

[423] The Structure of the Phage T4 DNA Packaging Motor Suggests a Mechanism Dependent on Electrostatic Forces

[424] see The phage-host arms-race: Shaping the evolution of microbes

Chapter 12: Asexual & sexual reproduction in eukaryotes

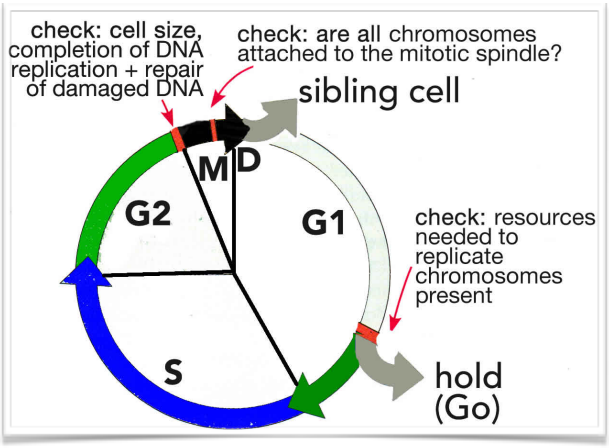
In which we consider asexual and sexual reproduction, including the processes of chromosome segregation (mitosis) & cell division (cytokinesis) and how they are modified to produce the haploid gametes that fuse to form a diploid organism. We consider the implications of chromosome pairing, recombination & independent segregation as well as sex determination, the dimorphism of gametes, maternal inheritance of mitochondria, and maternal and paternal effects.



Asexual reproduction in a eukaryote: making a (somatic) clone

In general terms, asexual reproduction in a eukaryote is similar to that in a prokaryote. The cell grows and at some point there is a molecular decision to replicate its genomic DNA and to divide the cell into two. At that point the genome of the cell is replicated and the cell divides to form two cells, each receiving one complete copy of the genome. In addition, all eukaryotic cells have cytoplasmic organelles (mitochondria, and in algae and plants, chloroplasts as well) which have their own small but essential genomes. As you might guess, since they appear to be derived from prokaryotes, these organellar genomes are circular double stranded DNA molecules. In the course of asexual, what is termed somatic, reproduction, each of the sibling cells also receive a number of mitochondria (and in plants, chloroplasts).[425] In the eukaryotes that we will concern ourselves with, most of the cells of the organism are diploid – we will let you know when they are not.

Somatic (asexual) reproduction involves what is known as the cell cycle. We can think of the cell cycle as beginning with cell division (D in the figure)(↓). The process of dividing one cell into two, known as cytokinesis, results in two sibling cells, each with (usually) identical genomes. Cytokinesis involves cytoskeletal and cytomuscular systems that are discussed in detail in a later cell biology course (not here!) Generally, but not necessarily, cell division is symmetrical, so that the two sibling cells are half the volume of the parental cell and very similar. Asymmetric divisions can occur, and generally result in cells that behave differently.[426] Cytokinesis is followed by a period of cell growth, known as G_1 (→), during which energy and materials that are imported from the external environment, or have been previously stored within the parental cell, are converted into lipids, nucleic acids, proteins, and other molecules leading to an increase in cell volume, the growth of the cell. As the cell grows, there are a number of decisions to be made: will the cell continue to grow (and perhaps divide) or will it stop growing and enter a steady state where it maintains itself (building and disassembling molecules, repairing DNA, etc) – a state known as G_o (↑). Generally, the majority of cells in any particular tissue are in the G_o state. In G_o there is no new DNA synthesis, so the possibility of mutation is lower than when DNA is being replicated. If, however, various external and internal signals act on and within the cell, many (but not all) cells can reverse the G_o decision and resume



[425] Plants and algae, which we will not be discussing in any detail, contain a second type of intracellular, DNA-containing organelle, known as chloroplasts. Their inheritance is similar to that of mitochondria.

[426] These differences are discussed in detail in the section on developmental biology.

growth and eventually divide (note that it is difficult to talk about these systems without personalizing them, even though these are not conscious "decisions" but the outcomes of molecular switches).

The decision to start DNA synthesis is based in part on whether the cell has, or can expect to have, sufficient resources to completely replicate its DNA molecules which, in a human cell, requires ~12 billion nucleotide addition reactions (both strands of a total of ~6 billion base pairs). The DNA synthesis decision point is known as “start”. There are mutant alleles, originally described through genetic studies in yeast, that result in a malfunctioning molecular switch controlling the start switch (the entry into S); such mutations, known as "wee" mutants by their Scottish discoverer, lead to a disconnect between growth and division and result in smaller and smaller cells and eventually cell death.[427]

Once a cell passes through the start checkpoint, the cell will enter the part of the cell cycle during which DNA synthesis occurs, known as S. As it begins genomic DNA synthesis the cell will encounter various checkpoints.[428] Checkpoints are molecular feedback systems and switches by which the cell monitors various aspects of its internal state and makes a decision to pause or proceed with a process, in this case DNA synthesis and later cell division.

During S the cell continues to grow and to replicate its DNA. In contrast to circular prokaryotic genomes, which typically have a single origin of replication (the site where DNA synthesis begins), the much larger size of eukaryotic genomes and the presence of multiple linear chromosomes requires multiple DNA synthesis start sites per chromosome. These replication origins are regulated during S phase such that each is activated once and only once per cell cycle in order to insure that each region of the genomic DNA is replicated once and only once. Before cell division (cytokinesis), a checkpoint monitors the presence of unreplicated DNA and will delay the cell cycle until that DNA has been replicated.[429] The process of DNA replication can lead to mutations, so this checkpoint also monitors the completion of various DNA repair processes. The presence of such a DNA repair checkpoint explains the observation that damaging DNA, for example by radiation, or inhibiting DNA synthesis enzymes using drugs, leads to delays in the cell cycle. Pathogens, such as the bacteria *Listeria monocytogenes*, exploit this DNA damage checkpoint to enhance their own replication.[430]

Questions to answer:

202. How many ways can you think up by which a cell could detect, and attempt to repair, damaged DNA or errors in DNA synthesis?
203. What factors limit the efficiency of DNA repair mechanisms? Why are mutations possible?
204. Why, do you suppose, does a wee mutant cell eventually die?
205. What effects could arise from the local over- or under-replication of DNA during S phase?

Ploidy during the cell cycle

By the end of S phase DNA synthesis is complete; the cell's genome has been replicated - the cell now has two complete copies of each chromosome. At this point the cell has entered into what is known as the G_2 phase of the cell cycle. Cells can continue to grow in G_2. During the asexual reproduction cycle the ploidy, the number of copies of the genome and each chromosome, is conserved. A haploid cell gives rise to a haploid cell, while a diploid cell gives rise to a diploid cell. The one detail that is altered is that by the end of S-phase of the cell cycle and during G_2 there are now twice the number of copies of the genome, and of each chromosome. While a diploid cell is

---

[427] Paul Nurse and Pierre Thuriaux on wee Mutants and Cell Cycle Control: https://www.ncbi.nlm.nih.gov/pubmed/27927897

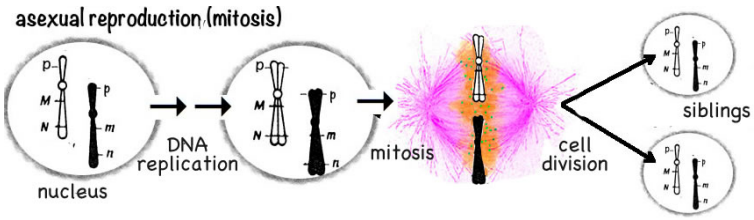
[428] The quorum sensing systems we discussed previously is a version of a checkpoint system.

[429] DNA replication is complex process, see Can the Stalling of DNA Replication Promote Epigenetic Changes?

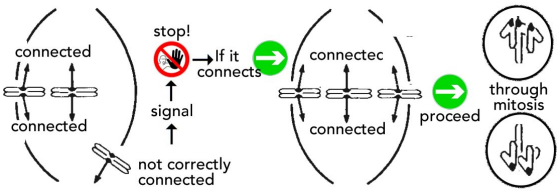
[430] Listeria monocytogenes induces host DNA damage and delays the host cell cycle to promote infection

diploid during G_1, it is effectively tetraploid after S and during G_2. This can have physiological effects because two copies of a gene can, in theory and generally in practice, support the synthesis of more RNA molecules per unit time than one copy of a gene. Based on this logic, we might expect to see changes in the rates of gene expression in G2 compared to G1 cells.

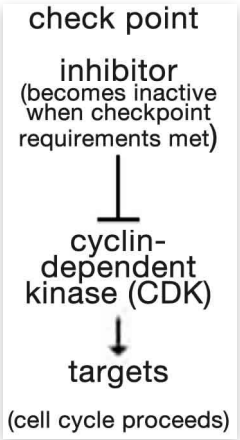
Molecular choices and checkpoints



Once the DNA replication/repair checkpoint has been passed, the cell can divide. The first step of this process (in eukaryotes) is known as mitosis (→). Mitosis involves a molecular machine, the mitotic spindle, based on protein polymers (αβ-tubulin-based microtubules). There is a molecular checkpoint that monitors the assembly of the mitotic spindle, and a second checkpoint that monitors that each replicated chromosome has connected correctly to the spindle (←). Each replicated chromosome consists of two linear double stranded DNA molecules. The pair of replicated chromosomes interacts with the mitotic spindle through a specific protein structure known as the kinetocore. Kinetocores are assembled in association with specific DNA regions known as centromeric sequences. Each replicated chromosome will have its own kinetocore and each interacts independently with the mitotic spindle (this is different from their behavior during meiosis, as we will see). The presence of the chromosome attachment mitotic checkpoint was recognized in experiments in which chromosomes were manipulated so that they could not connect correctly to the mitotic spindle; such a manipulation caused a delay or halt in mitosis.[431] The mitotic checkpoints serve to insure that each sibling cell gets one and only one copy of each and every chromosome present in the parental cell.[432]



Once activated, links between replicated chromosomes are severed, and the mitotic spindle moves chromosomes to opposites sides (poles) of the parental cell. The parental cell then divides using another protein (actin/myosin) polymer-based molecular machine, known as the contractile ring, to produce two sibling cells. It is worth noting that while these two cells are genotypically identical, as they inherit the same set of alleles as were present in the parental cell, they may behave differently due to differences in their environment and differences in internal components - factors that we will return to when we consider developmental processes.



The cell cycle decision check points are composed of multicomponent interaction networks. While we consider check point mechanisms only briefly here, they play a number of important roles in development and disease. A typical check point is commonly built around a protein kinase, an enzyme that can phosphorylate various targets – such phosphorylation (a post-translational modification) can lead to changes in protein structure, protein-protein interactions, protein activities, and a protein's stability and intracellular localization. Cell cycle checkpoints often involve a particular class of kinases, known as cyclin-dependent kinases (CDKs)(→). The activity of these CDKs is regulated positively by the binding of a small regulatory protein, known as a cyclin, as well as other interacting proteins and post-translational modifications. Cyclin's themselves are

[431] Mitotic forces control a cell-cycle checkpoint

[432] Kinetochores, microtubules, and spindle assembly checkpoint signaling

the target of various forms of regulation, including proteolytic degradation, triggered by their post-translational modification. Typically the activity of the cyclin-CDK complex is inhibited by various factors (proteins). When the conditions involved in the checkpoint are met, this inhibition is removed, allowing the cyclin-CDK complex to become active; the active kinase then phosphorylates and regulates the activity (and stability) of its targets, allowing the cell to pass through the check point and proceed along the cell cycle. One effect of activating the CDK is the rapid degradation (removal) of the cyclin, this makes the switch effectively irreversible until such time as cyclin levels increase again, during the next cell cycle.

Questions to answer:

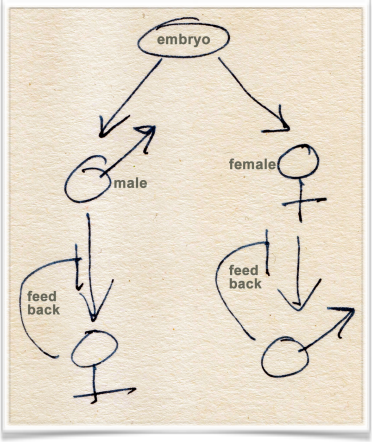
206. How do chromosomes interact with one another during mitosis/cytokinesis?
207. What does it mean that a checkpoint acts to "make a decision based on evidence"?
208. How does cyclin degradation make a checkpoint decision effectively irreversible?
209. Make a graph of CDK activity and the concentration of the cyclin regulating it, as a function of the cell cycle.
210. Predict what might go wrong if a checkpoint is ignored? (start with a cell cycle diagram)
211. How can a mutation in a checkpoint influence cell behavior during the somatic (mitotic) cell cycle?
212. How does gene expression change over the course of the somatic cell cycle?

Questions to ponder:

- Why is the decision to start a new cell cycle critical?
- When is the decision to start a new cycle made?

Sex-determination and its chromosomal basis

In eukaryotes, the generation of a new organism, distinct from previous organisms, generally involves the process of sexual reproduction. Different types of organisms determine an individual's sex using different mechanisms, and in some cases, a single individual, known as a hermaphrodite, can display traits of both sexes at either the same time or sequentially.[433] There are basically two general mechanisms that determine the sex of an organism: genetic and environmental, although do not be confused, environmental processes are based on molecular and cellular switches encoded genetically. In environmental sex determination various external signals influence the sex of the organism. For example in a number of reptiles (and other organisms), the sex of the adult is determined by temperature during key developmental periods, with different temperatures associated with male and female outcomes.[434] Recently, climate change (global warming) has been implicated in altering sea turtle sex ratios.[435] In other organisms, all individuals originally develop into one or the other sex and, as they mature (often growing larger) transform into the other sex.[436] In some cases the presence of a mature animal of one sex can inhibit the sex change in smaller individuals (→). As an example, the largest clownfish in a group is typically female; if that female is removed, one of the smaller males will develop into a female (think about the impact on Nemo). In other species, the situation is reversed, the largest animal is a male, and if this male is removed, one of the



[433] We will not go into any great detail about hermaphroditic models of reproduction, but this is an interesting paper related to the subject: Sexual selection: lessons from hermaphrodite mating systems.

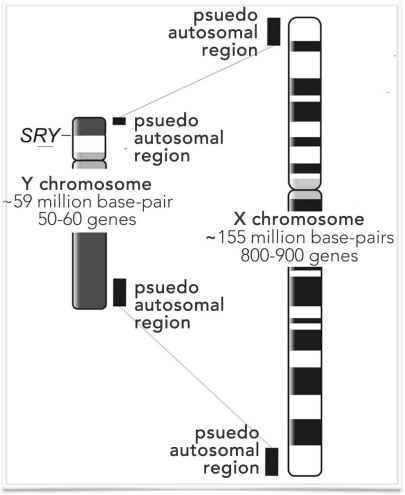
[434] Environmental sex determination mechanisms in reptiles

[435] Climate change is turning 99 percent of these baby sea turtles female

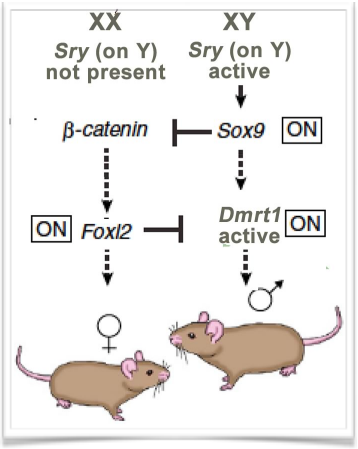
[436] Phylogenetic Perspectives on the Evolution of Functional Hermaphroditism

(smaller) females develops into a male.[437]

In humans, and most mammals, birds, and reptiles the phenotypic sex of an individual is determined chromosomally, that is, by which sex chromosomes their cells contain. The other, non-sex determining chromosomes are known as autosomes.[438] In humans the sex (23rd) chromosome comes in two forms, known as X and Y (→).[439] An XX individual typically develops as a female, while an XY individual typically develops as a male. Most of the X and Y chromosomes are non-syntenic, as you might have suspected given that the Y chromosome contains only ~50 genes, while the X-chromosome contains between 800 and 900 genes. The X and Y chromosomes are syntenic in what are known as their pseudo-autosomal regions. As we will see below, the organization of these chromosomes has effects on how they behave during the course of meiosis (sexual reproduction).



One key difference between X and Y chromosomes in therian mammals (marsupials and placental mammals, which includes humans), is the presence of the *SRY* gene in the Y chromosome. There is no copy of *SRY* on the X chromosome. The *SRY* gene is not found in monotremes (egg-laying mammals) and other vertebrates.[440] The *SRY* gene appears to have originated in the therian mammal lineage ~150 million years ago, derived by duplication of a Sox-type DNA binding protein/transcription factor that contains a high-mobility group (HMG) DNA binding domain. The presence of a Y chromosome, and so (presumably) an active *SRY* gene, leads to male sexual development, whereas the absence of *SRY* or loss of function mutations in *SRY* lead to female development, even if the Y chromosome is present (→).[441]



SRY encodes a transcription factor that initiates a down-stream gene regulatory cascade, activating some genes and inhibiting others, with the end result being the generation of the various developmental difference associated with male and female anatomy and behavior.[442] In females other genes are expressed and they act to inhibit the male differentiation system, just as Sry and its "downstream" targets act to inhibit female differentiation. In molecular studies, it is possible to show the importance of *SRY*, since the *SRY* gene can be transferred to one of the other chromosomes (an autosome), and its presence still leads to male determination. The details of these processes are complex, so we refer further details to more advanced classes.[443] That said, as you can imagine, defects in any of the genes in the pathway can influence outcomes.

---

[437] Functional hermaphroditism in teleosts

[438] In other species (e.g. birds, some reptiles, and some insects) the system is based on Z and W sex chromosomes. In contrast to the XY system, males are ZZ while females are ZW.

[439] X chromosome regulation: diverse patterns in development, tissues and disease and Y-chromosome

[440] "Environmental sex determination is widely employed in fish, where a range of stimuli from social cues to temperature establishes sex. Temperature sex determination is also extensively utilized in reptiles." see Sex determination in mammals--before and after the evolution of SRY

[441] see Molecular Mechanisms of Male Sex Determination: The Enigma of SRY for more details.

[442] In a recent study, the primary sex determination event in humans has been found to be associated with changes in ~6500 genes: see 6,500 Genes That Are Expressed Differently in Men and Women

[443] Sex determination: a primer

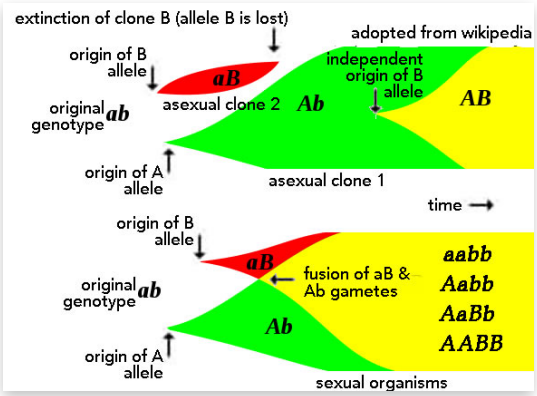
At this point please note that there are other sex (mating-type) determination strategies that you might come across in your subsequent studies, but which we ignore here.[444]

In contrast to asexual reproduction, which produces largely identical clones, the result of sexual reproduction is the generation of genetically distinct organisms, different from either parent. So what are the benefits of sexual reproduction, a process that involves collaboration between male and female organisms.[445] There have been a number of explanations for why sexual reproduction is so common, essentially all visible (macroscopic) organisms, with the possible exception of bdelloid rotifers,[446] reproduce (or can reproduce) sexually.[447]

A simple answer is the generation of genetic variation. So why is this variation important. One plausible reason involves the presence of rapidly reproducing pathogens. Viruses, bacterial and microbial (eukaryotic) organisms typically reproduce over periods of minutes to hours to days, whereas larger, multicellular organisms reproduce over periods of months, years, and decades. Similarly, but on somewhat longer time scales, the level of genetic variation within a population enables a population adapt to changing environmental conditions (of which pathogens are a part). Susceptibility to infection by pathogens is itself a phenotype, one with a genetic component. The genetic variability within a population can serve as insurance against pathogens; even the most lethal pathogens known, viruses like smallpox and bacteria such as those that cause plague, generally do not kill all of the organisms they infect. Those organisms that survive infection are often immune to subsequent infections, a phenomena that is the basis of vaccination and various other processes, including the CRISPR-CAS9 system of prokaryotes.

Sexual reproduction, specifically the processes of meiosis and fertilization offers a mechanism by which to generate huge amounts of genetic variation within a population. This view of the selective advantage of sex is often referred to as the Red Queen Hypothesis, since organisms have to "run" constantly, in terms of generating genetic variation, to keep up with their parasites and pathogens.[448]

"It takes all the running you can do, to keep in the same place." says the Red Queen to Alice



In addition, sexual production inserts a genetic bottleneck through which a multicellular organism must pass to generate the next generation; this bottleneck can remove deleterious alleles from a population.[449] In addition, sexual reproduction can speed the appearance of beneficial combinations of alleles, combinations that would take significantly longer to appear if they had to occur independently in a particular lineage (→).

The larger the population size, the more likely there is some genotypic combination already present that will make adaptation to a changing environment possible. The reduction in genetic variation is one of the reasons that reductions in population size have been linked to an increased

---

[444] The evolutionary dynamics of haplodiploidy

[445] Origins of Eukaryotic Sexual Reproduction: http://cshperspectives.cshlp.org/content/6/3/a016154.ful

[446] Uptake and Genomic Incorporation of Environmental DNA in the "Ancient Asexual" Bdelloid Rotifer Philodina roseola

[447] C. Zimmer. 2009. On the Origin of Sexual Reproduction

[448] see Sexual reproduction as an adaptation to resist parasites

[449] Add the sex as genetic bottleneck.

probability of extinction.[450]

In addition to the generation of variation, the process of sexual reproduction offers mechanisms by which populations can become reproductively isolated from one another, that is, to create two species from one. Generally males and females have to cooperate to reproduce; sexual reproduction is a social process. They have to be producing functional gametes at the same time, these gametes have to be able to meet each other, recognize each other, and fuse together, the diploid cell that forms has to develop normally, and the organism formed has to be able to form functional gametes, and so on. Incompatibilities in any of these processes can produce a reproductive barrier between the individuals within different populations - that is, speciation.

Questions to answer:

213. If you were to design a temperature sensitive form of sex determination, how might you go about it?
214. What might happen during meiosis if you were to remove the regions of the Y chromosome that are homologous to the X?

Question to ponder:

- How might variations in sexual behavior come about, molecularly?

Steps in meiosis: from diploid to haploid

Sexual reproduction begins with diploid cells, generally found in two distinct individuals. The basic process of sexual reproduction can be summarized as follows: a diploid cell generates, through the process of meiosis, one or more haploid "gametes". Haploid gametes (from two distinct "parents") fuse to form a new diploid individual. In some organisms, the haploid (gametic) stage can persist and live independently,[451] but generally the haploid stage of a eukaryote, and particularly animals, life cycle is short. In some, primarily unicellular, species there are multiple "mating types", and only gametes of different types can fuse. One aspect of the haploid state is that it can reveal the presence, and lead to the elimination, of deleterious recessive alleles. Haploid cells that contain, and are dependent upon the expression of such alleles will be eliminated, removing the allele from the population, which can have a strong evolutionary effect on the population.[452]

While the gametes of different mating types differ molecularly, they are similar morphologically, they both share an equal investment in reproductive outcomes. In multicellular organisms there are generally only two “mating types”. Moreover the gametes they produce differ in size: the mating type that produces the larger gamete (the oocyte) is known as female (♀) and the smaller (sperm or spermatozoa) as male (♂). The difference in the size of the gametes, an example of sexual dimorphism, can mean that the two sexes can have discordant investments in reproduction, one can spend more energy generating gametes than the other. This difference can become even more pronounced in terms of parental investment, a fact that underlies sexual selection, one of the key aspects of modern (Darwinian) evolutionary theory.[453]

In females the process of meiosis typically generates a single gamete, known as an egg, and three non-viable mini-cells, known as polar bodies. In males, meiosis produces four gametes. Each gamete will contain one and only one copy of each autosomal chromosome present in the original diploid cell. Historically, chromosomes were numbered based on their apparent size in histologically

---

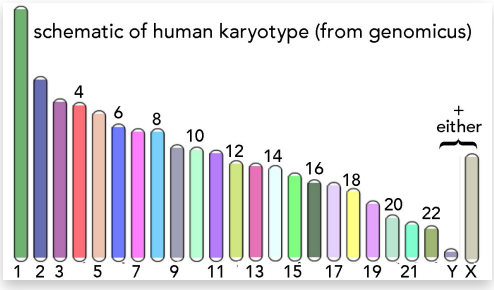
[450] Timing and causes of mid-Holocene mammoth extinction

[451] see wikipedia – gametophyte: https://en.wikipedia.org/wiki/Gametophyte

[452] see: Evolution of haploid selection in predominantly diploid organisms and Haploid selection in animals

[453] How Darwin arrived at his theory of sexual selection and Mate choice and sexual selection since Darwin?

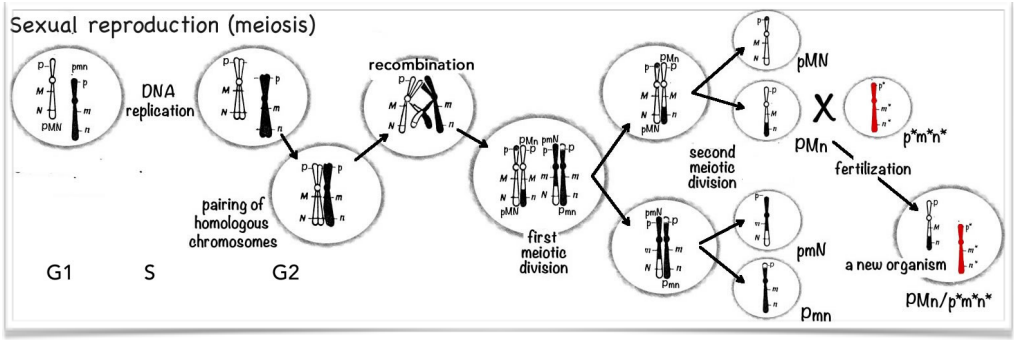
stained specimens. In humans, the largest chromosome, chromosome 1, contains ~250 million base pairs of DNA and over 2000 polypeptide-encoding genes, while the smallest, chromosome 22 contains ~52 million based pairs of DNA and ~500 polypeptide encoding genes (→).[454] Homologous chromosomes are also defined by the order of genes found along their length. Human chromosome #5 contains different genes from those found on chromosome #6. Moreover, the maternal (from the mother) version of each chromosome can contain different alleles of the genes present compared to those found in the paternal (from the father) version. The maternally and paternally derived chromosomes are known as homologs.



In mammals males have both an X and a Y chromosome; meiosis generates four gametes that contain one copy of each of the autosomes and either an X or a Y chromosome. Females have two X chromosomes, so all gametes they produce contain an X chromosome. A male gamete (a sperm) fuses with a female gamete (an egg) to form a new diploid cell, a new organism. If the male gamete contains a Y chromosome, the new (diploid) organism is chromosomally male, if the male gamete contains an X chromosome, the new organism is chromosomally female.[455] The fusion event, known as fertilization, is the most discontinuous event in the process of (sexually reproducing) life. Even so, fertilization does not represent a true discontinuity, at least with respect to life – both sperm and egg are alive, as is the fertilized egg.[456] In a critical sense life (in the post-LUCA world) never begins – it continues and is transformed. That said, fertilization is the start of a new, genetically distinct organism. The fused cell (new organism) that results from fertilization is known as a zygote. Through somatic (asexual) cell division (mitosis and cytokinesis) the zygote (fertilized egg) will develop into an adult, composed of diploid cells. The cells of the adult that produce gametes are known as germ cells, and together are known as the organism's germ line. The rest of the adult is composed of somatic cells, cells that divide (if they divide) by mitosis. Meiosis is restricted to germ line cells and gamete formation.

Recombination & independent segregation

We begin our description of meiosis (↓) with a diploid germ line cell that contains two copies of each autosome and, in mammals, either two X chromosomes in a female and an X and a Y chromosome in a male. The chromosomes derived from the female gamete are known as the maternal copy of the chromosome, while the chromosomes derived from the male gamete are known as the paternal copy of the chromosome. The maternal and paternal chromosomes are known as homologs. To generate gametes, a diploid germ cell enters



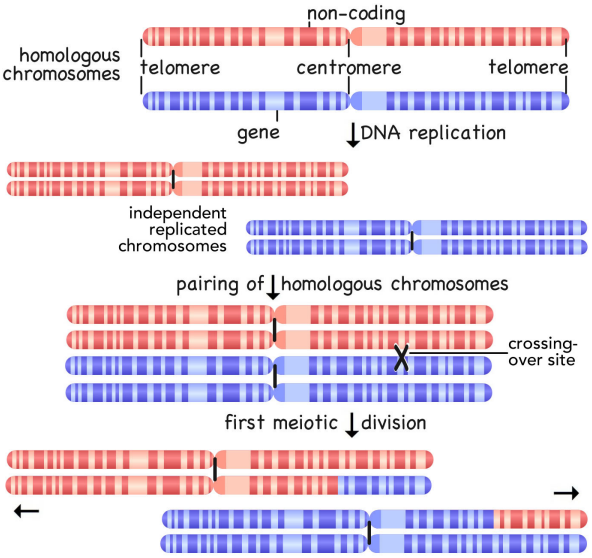
[454] We are only discussing polypeptide-encoding genes because it remains unclear whether (and which) other transcribed regions are genes, or physiologically significant.

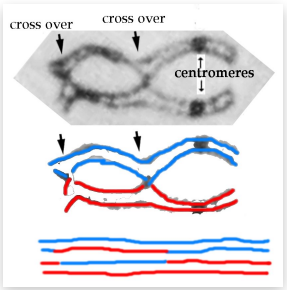
[455] While we not deal in detail with this topic, aspects of gender are complex traits: see Beyond XX and XY: The Extraordinary Complexity of Sex Determination

[456] In fact, there are examples of cell fusion within organisms - as an example, during the development of skeletal muscle, muscle precursor cells fused to generate large multi-nuclear cells, known as myotubes.

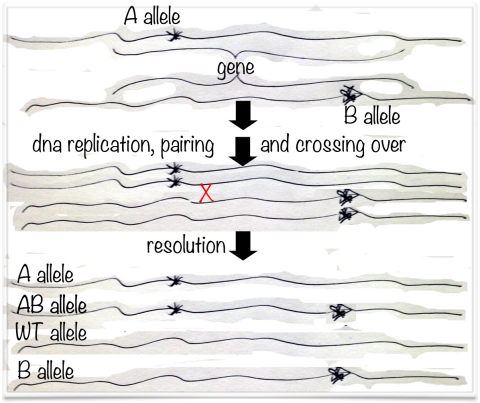
meiosis (see video link). Meiosis consists of a single round of DNA replication followed by two rounds of cell division.

As a diploid cell enters meiosis it moves from G1 into S, just as in mitosis. Each of its individual chromosomes (46 in humans, 2 copies each of the 23 homologous chromosomes) is duplicated. The resulting replicated (double-stranded) DNA molecules remain attached to one another through a structural complex known as the centromere. Here is where meiosis diverges from mitosis. In an asexual (mitotic) cell division each replicated chromosome remains independent of its homolog and each replicated chromosome interacts independently with the mitotic spindle through its centromere, and associated kinetochore complex. In meiosis, during G2 the (now) duplicated homologs (the maternal and paternal chromosomes) align with one another to form a structure containing four (double-stranded) DNA molecules (→). These four DNA molecules are known historically as a "tetrad"; each consists of four double-stranded DNA molecules. The pairing of the homologous chromosomes is based on the association of syntenic chromosomal regions.[457] The DNA sequences along the homologous chromosomes, while not identical, are extremely similar, with the same genes located in the same order on each. When they are not, due to chromosomal rearrangements, things can get messy - as we will see. After chromosome pairing, and at essentially random positions along the length of the chromosomes, "crossing-over" or recombination events can occur. An enzyme, a DNA endonuclease, produces double-strand breaks in two of the four (double-stranded) DNA molecules at the site marked by "X" above (↑) or by "cross over" to the left (←).[458] The DNA molecules are then rejoined, either back to themselves (maternal to maternal, paternal to paternal) or to the other DNA molecule (maternal to paternal or paternal to maternal), leading to a visible "crossing-over" event – maternal to maternal or paternal to paternal crossing over events are generally invisible. Typically, multiple "cross-over" events occur along the length of each set of paired (replicated) homologous chromosomes. Whenever maternal-paternal crossing over occurs the resulting recombinant chromosome contains a different set of alleles than either the original paternal or maternal chromosomes. You can convince yourself by following any one DNA molecule from beginning to end.





In addition to shuffling alleles, crossing over can create new alleles. Consider the situation in which two alleles of a particular gene are different from one another (→). Let us assume that each allele contains a distinct sequence difference (as marked). If, during meiosis, a crossing over event takes place between these sites, it results in one allele that contains both molecular sequences (AB), and another allele with neither (indicated as wild type "WT"). A new allele (AB) has been created, without a new mutation!

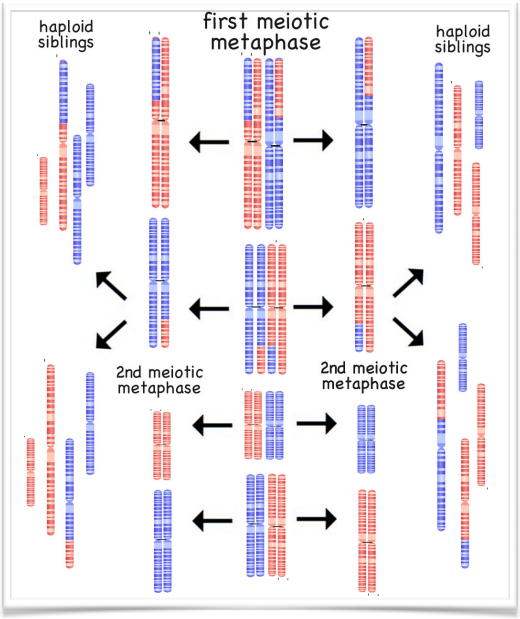


In the case of the X and Y chromosomes, the

[457] Synaptonemal complex formation: where does it start?

[458] adapted from The Centenary of Janssens's Chiasmatype Theory Koszul et al., 2012. Genetics **191:** 309-317.

chromosomes pair with one another through their common pseudo-autosomal regions (see above), which are syntenic. Outside of these regions there is no significant synteny between the X and Y chromosomes, leading to the suppression of crossing over much of the X and Y chromosomes' length in males. In contrast, crossing over can occur normally (that is, just like for autosomes) between the two X chromosomes in a female.



Meiosis leads to yet another source of variation. At the first meiotic division, the duplicated (and recombined) chromosomes remain attached at their centromeres, so that each of the two resulting daughter cells receives either the duplicated maternal or paternal chromosome centromere region. However, what set of chromosomes (defined by their centromeres, maternal or paternal) they inherit is determined by chance.The process is known as the independent assortment of homologous chromosomes during the first meiotic division, or independent assortment for short. For an organism with 23 different chromosomes (such as humans), the first meiotic division can produce 2^{23} different daughter cells (←).

There is no DNA replication between the first (M1) and the second (M2) meiotic divisions. During the second meiotic division the replicated chromosomes, held together at their centromeres, attach to the spindle, very much as in mitosis. Because of recombination, the two chromosomes are not necessarily identical, which further increases (to rather astronomical levels) the number of different chromosome sets a particular haploid cell can inherit. When they separate, the two resulting sibling cells normally each receives one and only one copy of each chromosome (a double-stranded DNA molecule). Again, which particular molecules they inherit is stochastic. The four haploid cells generated by meiosis are known as gametes (or at least are potential gametes). In males, all four haploid cells differentiate to form sperm cells, whereas in females, typically one of the four haploid cells differentiates to form an oocyte, which becomes an egg that can fuse with a sperm cell (fertilization); the other three cells are known as polar bodies. Polar bodies do not fuse with sperm. In essence, the polar bodies donate their cytoplasm to the oocyte - supporting the development of the fertilized egg, the new organism.

The result, and basically the point, of meiosis is to generate gametes in which the alleles present in the maternal and paternal chromosomes have been shuffled in various ways, so that the resultant offspring has a genome related to, but distinct from that of either of its parents.[459] Fertilization (the fusion of gametes) combines two such genomes, one maternal and one paternal, to form a new organism, with a novel combination of alleles. Most phenotypes are influenced, to a greater or lesser degree, by the set of alleles within a genotype, and new combinations of alleles will lead to new phenotypes and phenotypic variations that can impact reproductive success, and so lead to evolutionary effects.

Questions to answer:

215. Consider the odds of an organism obtaining the three new mutations necessary for the appearance of a new trait. Predict which would be faster (in terms of the number of generations required) in achieving this goal, sexual or asexual reproduction and why.
216. You are working with an organism with five autosomes and one sex chromosome. Considering only the effects of independent assortment during meiosis, how many different types of gametes could be generated? A drawing of the process could help.

[459] This even applies to hermaphrodites, in which one organism acts as both mother and father!

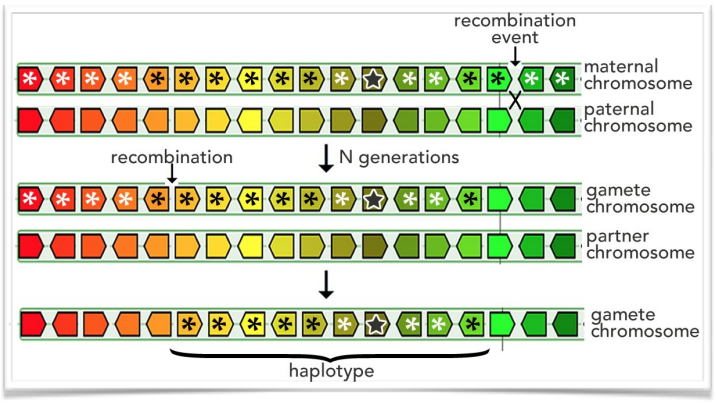
217. Indicate (in a drawing and associated explanation) how a deleterious mutation within a gene could be generated by or eliminated from a gene through recombination.
218. Would genetic diversity be altered if meiotic recombination occurred during meiosis II, rather than meiosis I?

Questions to ponder

- Under what conditions might you expect the evolution of sexual reproduction to be selected against.
- Why are parents and their siblings not necessarily good donors for organ transplantation?

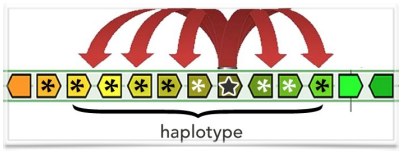
Linkage & haplotypes

An important feature of meiotic recombination is that it can "disconnect" the alleles of genes located near one another along a chromosome. Consider the situation when a mutation occurs that creates a new allele in gene X; let us call it X^{select}. Now let us assume that this allele is subject to strong positive or negative selection. That means that the presence of the X^{select} allele in an organism has a strong effect on reproductive success. Because it is either strongly selected for (positive effect on reproductive success) or against (negative effect on reproductive success) the frequency of the allele will tend to increase or decrease in subsequent generations, unless it is lost through the effects of genetic drift. The change in the frequency of the X^{select} allele also influences the frequency of alleles of genes located near the X gene on the chromosome. If X^{select} is subject to strong positive selection, such selection will also increase the frequency of the alleles in these neighboring "linked" genes. Similarly, if X^{select} has a negative selective effect, the frequency of the alleles in genes neighboring (linked to) gene X will decrease over time, even if these alleles are, on their own, beneficial. These effects will depend upon the relative selective effects of the various alleles. The closer the genes are to each other along the chromosome, the longer (over more generations) such linkage effects will persist. Why? because the probability of recombination between two sites along a chromosome (two genetic loci or positions) is a function of their distance from on another. As the distance between two genetic loci increases, the probability that the original alleles at these positions will be separated by recombination increases. When the probability of a recombination event between two genes reaches 50% or greater (per meiotic division), the genes behave as if they are on different chromosomes – they become "unlinked." Linkage distances are calculated in terms of centimorgans, named after the geneticist Thomas Hunt Morgan (1866-1945). A centimorgan corresponds to a 1% chance of a crossing over event between two specific sites along a chromosome. In humans, a centimorgan corresponds to ~1 million base pairs of DNA, although this value varies somewhat in different regions of different chromosomes. Two genetic loci that are 50 or more centimorgans apart are separated by ~50 million or more base pairs. In the context of meiosis, two genetic loci on the same chromosome, but separated by >50 centimorgans, have the same probability of being inherited together as if they were on two different chromosomes. We will return to this again, when we consider the interpretation of genetic crosses.



Consider a particular allele of a particular gene, marked by the star (★) here (→); let us assume that this allele is associated with a visible trait. We will mark the alleles found in neighboring genes on this chromosome with asterixs (*). For the sake of clarity assume that different alleles (un-marked) are found on the homologous chromosome. During meiosis, recombination events will occur randomly across these chromosomes. Over time independent recombination events occur that will increasingly reduce the size of the region of the original chromosome (containing the ★ allele). This original region is known as a haplotype; it is a group of alleles that are inherited together from a single parent. From a formal point

of view, it is not clear which variation within the haplotype region is responsible for the trait observed. In the era of genetic (pre-molecular biological methods) days, multiple rounds of crosses (breeding cycles) are required to identify on which region of which chromosome the allele (gene) responsible for a particular trait was located, With more and more generations, the size of haplotype regions becomes smaller.



Now consider how the alleles within a particular region can be maintained together. Let us assume that the original allelic variant has effects on the expression of neighboring genes (→); how might this occur? Two obvious mechanisms suggest themselves: the allele could influence the packaging of the chromosomal region, so that the genes' accessibility to regulatory factors is modified or the allele can itself effect or be in an gene regulatory element (an enhancer) that plays an important role in the regulation of multiple genes in this molecular neighborhood. Both options could lead to selective effects based on the maintenance of the integrity of the chromosomal region (a haplotype) - that is, recombination events within the region can occur, but because they have a negative effect on reproductive outcomes they would be selected against.

Questions to answer:

219. Graph, as a function of distance, the likelihood that recombination will disconnect a selected (whether positively or negatively) allele from alleles in surrounding genes.
220. Why might a crossing over event inhibit nearby crossing over events?
221. How can you use the size of a conserved genomic region to estimate time of isolation of a population?
222. What are the benefits of recombination in terms of environmental adaptation?

Questions to ponder:

- How does the size of haplotype regions reflect the reproductive history of a population?
- How does the presence of a deleterious allele influence the selective pressures on an organism? How might it open up time, new evolutionary possibilities?

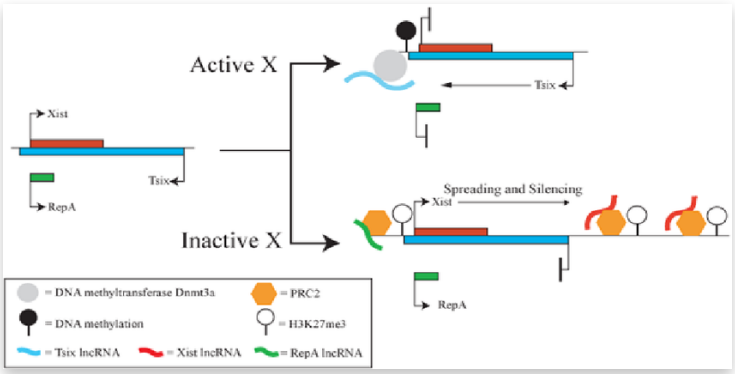
X-inactivation and sex-linked traits

One aspect of the XY chromosome-based system of sex determination is that the two sexes have different genotypes, at least with respect to these chromosomes. As mentioned above, the Y chromosome is short and encodes relatively few genes, while the X chromosome is much longer and encodes many more genes. This creates a genetic imbalance between the two sexes in terms of gene copy numbers. A single gene can direct the synthesis of only so many RNA molecules per unit time, based on the rate of RNA polymerase binding, activation, and RNA synthesis along a DNA molecule. This is the reason for haplo-insufficiency, a phenomena associated with genes on autosomes, where a null allele leads to a dominant phenotype due to the fact that a single functional copy of the gene does not produce sufficient gene product. Without some "balancing" mechanism, we would predict that female cells would have about twice as many RNAs for genes on the X as do similar cells in a male (and most cells in males and females are, in fact, similar). There therefore seems to be a need for a form of "dosage compensation"; either genes on the X in males have to be expressed more efficiently or genes on the X in females should be expressed less efficiently. The strategy used in humans and other placental mammals is a process known as X-inactivation. Early in embryonic development, one or the other of a female's X chromosomes becomes associated with specific RNAs and proteins, and is packed into a compact structure that can no longer support gene expression (RNA transcription).[460] Once the choice of which X chromosome to inactivate is made, it is stable and inherited through subsequent mitotic cell divisions, generating clones of cells with the one or the other X chromosome active (and the other inactive). A failure of X-inactivation generally leads to developmental arrest and embryonic death in female embryos. While gene expression from

[460] X Chromosome Inactivation Is Initiated in Human Preimplantation Embryos

the inactivated X is inhibited, the replication of the inactivated chromosome continues with each cell cycle. We can see the effect of this choice in female calico cats (→), in which the different coat colors reflect domains in which one or the other X chromosomes is actively expressed, while the other X chromosome is inactivated. As you may have already deduced, a gene involved in the generation of coat color is located on the X chromosome.

The X-chromosome inactivation system consists of two genes, *XIST* and *TSIX*. *XIST* encodes a functional ~19.3 kilobase long non-coding RNA, known as an lncRNA; such an RNA does not (as far as is currently known) encode any polypeptides - it is not (apparently) an mRNA (↓). XIST is expressed only in cells with two X chromosomes – so it is not expressed in males.[461] Which of the two X-chromosomes expresses *XIST* is initially determined (during embryonic development) stochastically. When expressed, the XIST RNA associates with regions adjacent to the *XIST* gene and eventually comes to localized along the entire length of the X-chromosome on which the active *XIST* gene is located. The XIST RNA comes to associate with a number of protein complexes involved in inhibiting gene expression and producing the compact state of the inactivated X, also known as a Barr body, named after its co-discoverer Murray Barr (1908 – 1995).



On the DNA strand opposite to the *XIST* gene is an over-lapping gene known as *TSIX* (↑). The *TSIX* gene on the active X-chromosome is expressed. The *TSIX* promoter is distinct from that of *XIST*; expression of *TSIX* is expected to interfere with *XIST* expression. The *TSIX* gene encodes a ~40 kilobase lncRNA that is partially complementary to the XIST RNA. The TSIX RNA acts to inhibit XIST activity, and so blocks the action of XIST on the active X chromosome, blocking that chromosomes's inactivation. Together the XIST/TSIX system insures that one and only one of the two X chromosomes is active in a particular cell.

X-linked diseases and mono-allelic gene expression

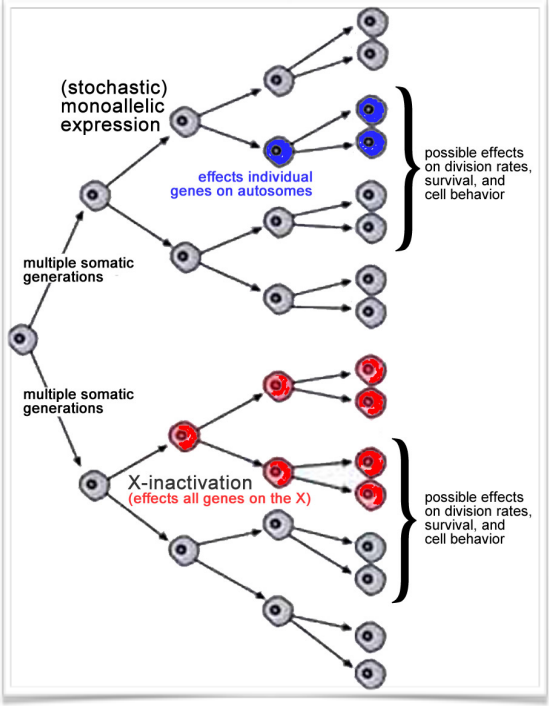
While calico spots occur only in female cats, there are a number of genetic susceptibilities that are more commonly seen in males; these arise because males have only a single X chromosome. The result is that, in contrast to the rest of the genome, genes on the X are effectively haploid in males. The result is that the phenotypes associated with recessive alleles of genes located on the X chromosome are visible in males. In contrast, in females that are formally heterozygotic for that gene, some cells express one allele while others express the other. This situation (in females) leads to what is known as random monoallelic expression. Recent studies have revealed that random monoallelic expression occurs throughout the genome, even in autosomal genes, but it is essentially universal for genes presence on the X chromosome, in females. In a typical diploid cell, it is sometimes the case that one gene is active while the other copy of the gene, on the homologous chromosome is inactive, due to stochastic "transcriptional silencing" events.[462] In some cases of stable monoallelic expression there is what is known as somatic selection, which we will return to.

---

[461] X-inactivation-specific transcript (OMIM)

[462] Monoallelic Gene Expression in Mammals

Given that there are two alleles, when they are different which is expressed may influence cell growth, division, and even survival, so that over time, cells expressing one allele may come to dominate (in numbers) those that express the other (→). The extent to which random monoallelic expression influences human development and disease is just now being recognized and examined carefully.



Questions to answer:

223. What does it mean to be mosaic for an allele?
224. Why do males and females differ in the traits they display?
225. Why do males and females differ in the display of phenotypes associated with genes on the X chromosome?
226. Can you provide a plausible mechanism to explain why (autosomal) random monoallelic expression occurs?
227. How might monoallelic expression impact an organism?

Question to ponder:

- Under what conditions might monoallelic (autosomal) gene expression be beneficial?

Chapter 13: Generating mutations & becoming alleles

In which we consider how mutations appear and become alleles within a population and distinguish between the effects of chance and selection on allele frequencies.

We are far enough along to recognize that beginning with a particular genome, any change in that genome, such as those that arise due to errors in its replication or un-repaired or mis-repaired environmentally induced damage (through chemical reaction or radiation) results in a mutation. If the mutated cell/organism survives and gives rise to offspring, and if the mutation lies within a gene, it becomes what is known as an allele - a genetic variant of a gene within a population. If it lies outside of a gene, it becomes known as a sequence polymorphism. With the advent of genomic sequencing, and related technologies, it is possible to estimate the rates of mutation in a particular organism or a particular cell type.[463] Here we distinguish between mutations in the germ line (leading to eggs and sperm) that can be inherited by offspring and those that occur in somatic cells, the cells of the body. Mutations not inherited from one's parents are said to occur spontaneously. A spontaneous mutation may be passed on to offspring if it occurs in a germ line cell.

As a first approximation, mutations occur randomly within genomes, but in fact there are what are known as mutational hotspots - for example, CpG dinucleotides are mutated ~10X more frequently than other dinucleotides. In addition to single nucleotide changes, there are also mutations that involve small insertions and deletions in the DNA, known as indels. Indels are defined to be less than 20 base pairs (bps) in length to distinguish them from larger changes, known as structural variants that can be much larger (kilo or megabases). It has been estimated that each generation sees the addition of ~3 indels and ~0.16 structural variants in the germ line of a person. Another class of structural variant, known as a copy number variation (CNV), that lead to changes in the number of copies of a particular genomic region, leading to multiple copies of the gene(s) localized in the region and their regulatory elements.[464] There are also mutations associated with the process of cell division, these can lead to the loss a chromosome or the duplication of the entire genome.

Mutations occur more frequently in the soma of an organism because there are more cells (trillions) and so more cycles of DNA replication and cell division. Similarly, there are fewer cell divisions involved in the generation of oocytes in females than in the generation of sperm in males, and the number of mutations, particularly in the male germ line, increases with age. As you can probably predict, germ line mutations can be passed from generation to generation, while somatic mutations are lost on the death of the body. The current estimate is that the chance of a *de novo* germ line mutation in humans is ~$1x10^{-8}$ per base pair per generation (remember the human genome contains ~6×10^9 bps). Somatic mutations appear to be the prime driver of cancer, and we will discuss both germ line and somatic mutations and their effects in awhile.

Mutations into alleles

For a mutation to become an allele within a population the first criterion is that it does not have a early (pre-reproductive age) dominant lethal phenotype - that is a phenotype that results in the death of the organism before it can produce offspring. Why? In a diploid organism a new mutation will involve only one of the two genes present; for it to have a phenotype, it needs to be "dominant" over

---

[463] see: The origins determinants, and consequences of human mutations

[464] Copy Number Variation & Indels

the other allele present. Of course this is not the case in prokaryotes, which are effectively haploid. If the mutation is not dominant lethal, and if it occurs in the germ line, it can be passed to a gamete and from there into the next generation, it has a chance to persist within the population. Again, this assumes that the presence of the allele does not result in a lethal phenotype in gametes or the early embryo, since where and when a gene is expressed has a lot to do with the phenotypes it is associated with.

A non-lethal dominant or a recessive mutation has to avoid elimination through the stochastic effects of genetic drift. Remember that when it first appears in the germ line of a sexually reproducing organism (we will ignore somatic mutations for the moment, since they are "trapped" within a particular organism), there is only one copy of the mutated allele in the population; it is possible that gametes carrying this allele with fail to find and fuse with another gamete to form a new organism – if so, the mutant allele will be lost. Similarly, the mutant allele may make it into the next generation if it is not too deleterious, just by chance.

If a mutant allele survives these early events, it comes to be referred to as an allele, particularly when it is found in >1% of the population. Mutations that occur outside of a gene become what are known as polymorphisms; such polymorphisms generally do not have effects on phenotype since they do not influence gene expression. The difference between an allele and a polymorphism lies in the ability to recognize what is, and what is not, part of a gene, something that can be tricky. The total genetic variation within a population, the sum of alleles and polymorphisms reflects the population's past history, that is, the combination of selective pressures and non-adaptive events, such as founder effects, bottlenecks, and genetic drift, and serves as the basis for subsequent evolutionary change.

Luria & Delbrück: Discovering the origin of mutations

Keeping in mind that Darwin and Wallace lacked a clear understanding of where genetic variation came from, how it is stored, or replicated from one generation to the next, an important question that arose early in the history of evolutionary theory was whether the mutations (a prime source of phenotypic and genetic variation) associated with the evolution of new species and complex traits – such as the eye – were the result of chance (stochastic) events or whether they were somehow purposefully generated in response to the needs of the organism. As proposed by Darwin, evolution involves random variations that arise in individuals; a Lamarckian mechanism involves induced responses by individuals.[465] In the absence of a clear understanding of how genetic information and variants in that information arise in a population or how they are passed from generation to generation, there was really no way to distinguish between Darwinian (random variation + selection) and Lamarckian (adaptation based on the organism's "needs") evolutionary mechanisms, although Lamarckian mechanisms seemed more direct.[466]

To understand how this question was resolved, consider a classic experiment, known as the Luria-Delbrück experiment after the two researchers, Salvador Luria (1912-1991) and Max Delbrück (1906-1981) who carried it out.[467] Their study was published in 1943, before DNA was recognized as the genetic material and well before anyone understood how genetic information was stored.[468] Luria and Delbrück examined the resistance of bacteria to viral infection. They used bacteria that could be infected and killed by a specific type of bacteriophage. Mutations arose spontaneously in the

---

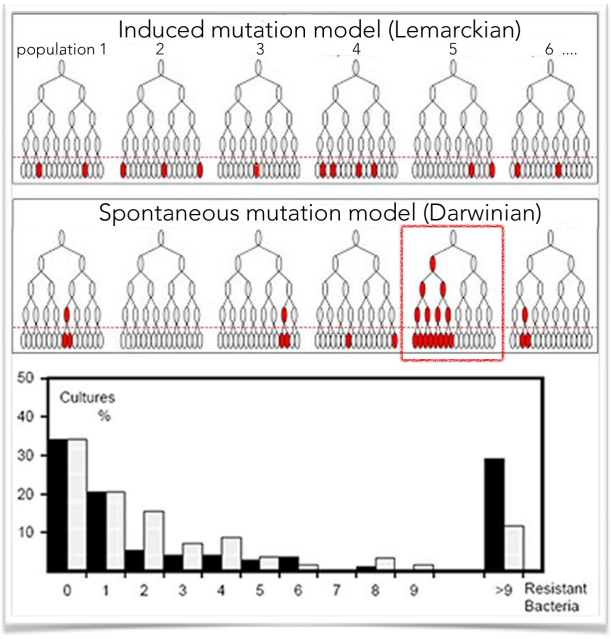
[465] This is perhaps one reason that collectivist ideologies, such as the Soviet Union under Stalin, so disliked Darwinian evolution (and harshly prosecuted geneticists). see http://blogs.plos.org/scied/2017/04/10/science-politics-marches/

[466] This led to what was known as the "Eclipse of Darwinism"; biology emerged from this "darkness" with the development of an understanding of genes and genetic mechanisms to produce what became know as the "Modern Synthesis".

[467] Luria–Delbrück experiment

[468] Mutations of bacteria from virus sensitivity to virus resistance: http://www.genetics.org/content/genetics/28/6/491.full.pdf

bacteria rendered them, and their off-spring, avoid or survive phage infection. The question Luria and Delbrück asked was, are phage resistance mutations appearing randomly all of the time or is it that the presence of the virus 'induces" the appearance of mutations in response to the bacteria's "need" to be immune. Is immunity learned or lucky?[469] If the generation of phage resistance mutations is an adaptive process, then we would expect that the frequency of resistance (mutations) will be more or less uniform from one population to the next – repeating experiments on different cultures should produce resistant bacteria at approximately the same rate in each (top panel →). If, on the other hand, the mechanism occurs by chance (middle panel →), then we can expect that the number of mutational events will vary dramatically from one population (culture) to the next - the variation in the frequency of phage resistance (and the mutations that produce it) between independent populations will be large.



Luria and Delbrück started a number of bacterial cultures to which they then added enough virus (at the time of the horizontal red line in the top two panels) to kill every sensitive bacterium. They then plated out the cultures and counted the number of phage-resistant bacteria present, each of which could grow up into a macroscopic (asexual) clone, a colony. The number of such phage resistant cells in a culture reflects when, in the history of the culture, the resistance mutation appeared; for example, if the resistance mutation appeared early in the history of the culture, as in the red-boxed culture (↑) it would be common, whereas if it appeared late, it would be rare. The two models (induced/Lemarckian versus spontaneous/Darwinian) make dramatically different predictions. In the induced/Lemarckian model, the variation in the numbers of resistant bacteria between cultures is expected to be low, since resistance arises through a common "inductive", physiological process, even though we do not know how that process works. In contrast, in the spontaneous/Darwinian model we expect large variations, with many cultures having no resistant bacteria and some having many. When the mutation occurs late, or not at all, as in lower panel, population 2, there will be few phage resistant cells. If the mutation occurs early there will be many resistant bacteria. Luria and Delbrück calculated what the two models predicted. The observed results (black bars) matched the prediction for the spontaneous/Darwinian mechanism, leading them to conclude that, at least in this system, mutations occurred independently of the presence of the virus.

To date there is no evidence that environmental factors can specifically induce the generation of beneficial or useful mutations. What can happen, however, is that the general (non-specific) mutation rate can increase in response to various stress conditions, arising from internal or environmental effects. Typically an increased mutation rate involves effects on the efficiency of DNA error repair systems, which leads to increased levels of genetic variation upon which selection can act.[470] The ability to control mutation rates occurs within the vertebrate immune system, through a process known as somatic hypermutation.[471] This process is involved in the maturation of the

---

[469] As we will see later on, there are molecular mechanisms, such as the CRISPR CAS9 system that can learn and lead to acquired immunity.

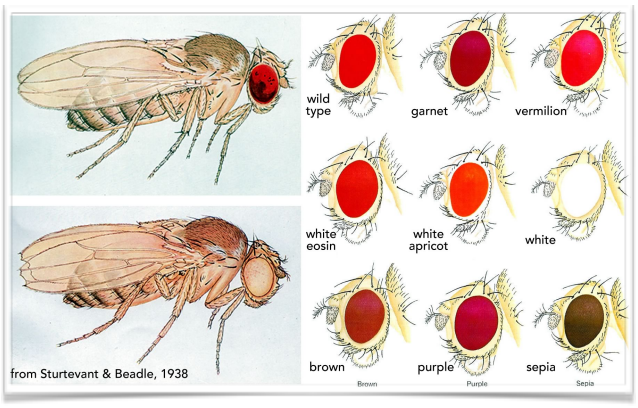
[470] A trade-off between oxidative stress resistance and DNA repair plays a role in the evolution of elevated mutation rates in bacteria

[471] Somatic hypermutation: wikipedia

immune response and the generation of increasingly specific antibodies, a topic well beyond our scope here. That said, the mechanism is known; these cells activate a gene that encodes an "activation-induced deaminase" or AID (OMIM:605257). AID acts on cytosine residues in DNA to generate uracils that, when repaired, replace the original C:G base pair with an A:T base. The other genes in these cells appear to be at least partially protected by "selective targeting of AID and gene-specific, high-fidelity repair of AID-generated uracils".[472]

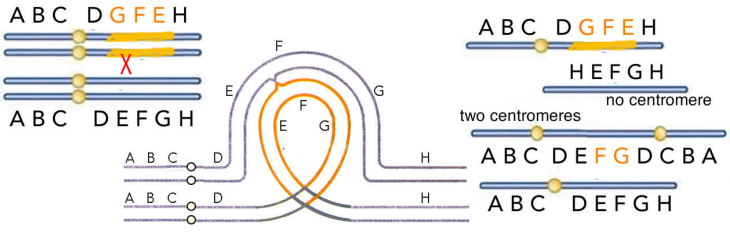
Forward and reverse genetics

Originally, genetic analyses were carried out through what is now known as forward genetics. Forward genetics involves the generation of mutations by chance and then identifying individuals carrying mutations that disrupt a particular process or structure of interest. As an example, consider eye shape or color in the fruit fly *Drosophila melanogaster* (↓); these are traits that are experimentally accessible because a *Drosophila* embryo can develop into a fertile adult without an eye. It is therefore possible to identify mutant alleles that alter the eye but allow other aspects of embryonic development to occur (more or less) normally, at least in the context of the laboratory. When we think about a particular trait or behavior, a specific phenotype, we want to know how many different genes are involve in producing that phenotype. On the other hand, if the product of the mutated gene plays multiple roles in the developing organism, perhaps in processes distinct from those involved in the formation of the eye, the embryo may die before eyes form, and no mutations in that gene will be recovered, even though the gene's product plays a key role in eye development or pigmentation. It is for this reason that forward genetic screens for mutations that influence a particular process are never complete, that is, they do not identify every gene/gene product involved in a process.



The classical approach to identifying genes involved in producing a particular phenotype is known as a "forward genetic screen"; it involves a search for mutations that disrupt that phenotype. Waiting for naturally occurring mutations to appear is too slow for the ambitious (and mortal) researcher, so steps are taken to induce large numbers of mutations. Among the first of these mutagenesis methods was irradiation using X-rays. In 1927, H.J. (Joe) Muller, who we have met before, was the first to create a mutation using X-rays.[473] It earned him a Nobel prize.

A brief aside on inversions: Before we go on, let us consider how the presence of a chromosomal inversion in one of the two homologous chromosomes can influence meiotic outcomes. If the inverted region is large enough, the region of one chromosome can loop around to maximize pairing with the other during meiosis (homologous chromosomes do not align during mitosis). During the process of chromosome pairing, there is a significant chance that a crossing over event will occur between the inverted and non-inverted regions(→); different effects will occur depending upon exactly where the inversion is located

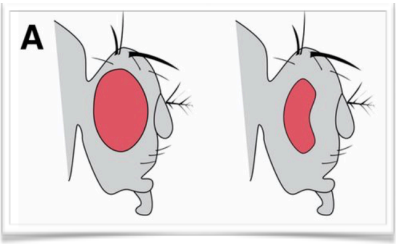


[472] Two levels of protection for the B cell genome during somatic hypermutation

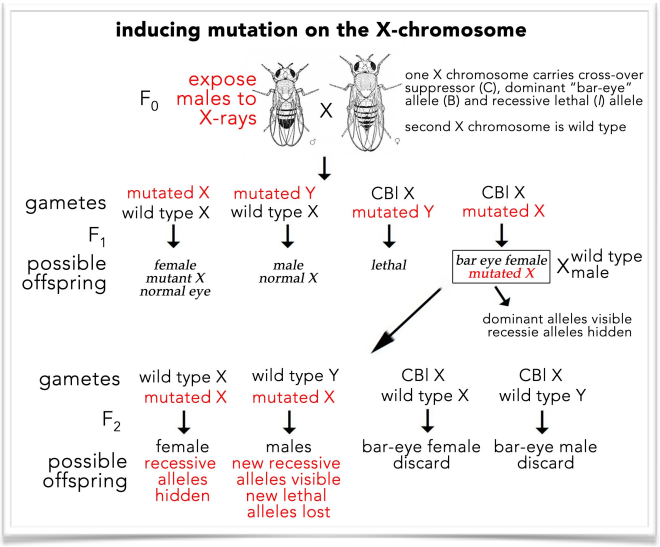
[473] Hermann J. Muller (1890-1967) demonstrates that X rays can induce mutations

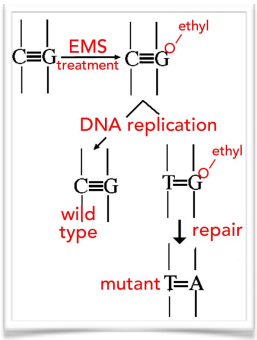
along the chromosome. Here we consider an inversion that does not include the region of the centromere. A crossing over event in this region will result in a duplication of DNA sequence (and genes) in one chromosome and DNA sequence (and gene) deletion in the other. One recombinant chromosome will have two centromeres (it is "di-centric") while the other has none, it is "acentric". During the first meiotic division, the acentric chromosome will fail to interact with the meiotic spindle and will not be accurately segregated to daughter cells. The dicentric chromosome can associated with both spindle poles; if it does it can be "ripped" apart during the first meiotic division leading to mutations. These effects, together with the effects of the duplications and deletions can lead to lethality during embryonic development.

Back to Muller: He examined the generation of mutations on the X chromosome of *D. melanogaster*, an organism chosen in part because of its small size (which allows lots of animals to be raised in a limited space), rapid life cycle, and the large number (~400) of offspring produced by a single female after a mating. In previous studies, he had isolated a version of the X-chromosome, known as CBI, that carries a dominant allele that produces bar eyes (←), a recessive lethal mutation in a different gene, and a large chromosomal inversion (a flipped region of DNA) in the chromosome. If meiotic crossing over (recombination) event occurs within the inverted region, embryonic lethal mutations are generated. The result is to effectively suppress recombination, since individuals that inherent recombinant chromosomes do not survive, and so do not effect subsequent conclusions.



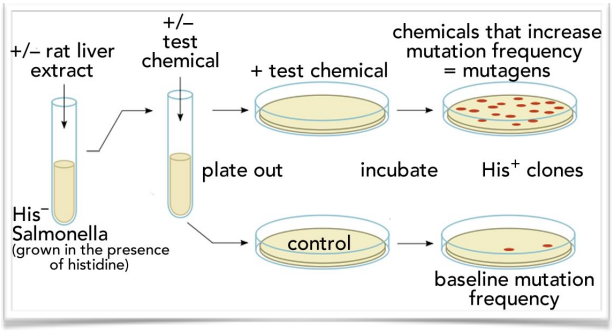
Muller took wild type male flies and irradiated them, which induced mutations in their testes resulting in sperm carrying those mutations. He then mated females carrying the altered CBI X-chromosome with the irradiated males. Based on the markers present, he could identify females that carried the CBI X chromosome and a mutated X chromosome from an irradiated male (→). When these first filial generation (F_1) females were mated with wild type males, the offspring that carried a mutated X chromosome could be identified and analyzed. Males displayed phenotypes associated with recessive alleles (mutations) on the X, while dominant mutations were visible in females. Through this analysis, Muller identified hundreds of new mutations (alleles) and, more importantly, showed that the genetic material could be damaged, or rather altered, by radiation.





Since these studies, a number of other methods have been found to induce mutations, all act by damaging the DNA in one way or the other. For example, animals can be fed potent mutagenic chemicals, such as ethyl methane sulfonate (EMS)(←). EMS reacts, through an esterification reaction, with guanosine residues in DNA, modifying them through the addition of an ethyl group. The modified G base (G*) pairs with T rather than C; when the modified DNA is replicated, one copy is wild type while the other generates an aberrant AG* base pair, which is then repaired to produce a mutation, replacing the original CG base pair with an TA base pair.

To identify chemicals that can induce mutations, Bruce Ames (b.1928) and colleagues developed a test using the bacterium *Salmonella typhimurium*.[474] They began by using a strain of *S. typhimurium* that carries a mutant allele that rendered it unable to grow in the absence of the amino acid histidine; they termed this strain His^-. The His^- strain can be reverted to a his^+ strain by mutation. To test whether a chemical is mutagenic in *S. typhimurium*, His^- cells were grown up in the presence of histidine (to allow for growth) together with the chemical to be tested. Typically, a number of different concentrations of the chemical are tested. After some time the cultures are plated out onto agar plates in the absence of histidine. The result is that only those bacteria that have acquired a mutation that converts them from a His^- to a His^+ phenotype can grow into macroscopic colonies (→). There is, of course, a low rate of spontaneous mutation, that is mutation in the absence of test chemical; this enables us to estimate the baseline mutation frequency for the *S. typhimurium* strain used. If the chemical to be tested is mutagenic, then the frequency of mutations should increase above this baseline rate; we also expect that the mutation rate will increase as a function of the concentration of the chemical tested. Hopefully you appreciate (but we will remind you) that while we are assaying for the appearance of His^- to His^+ mutations, mutations are occurring randomly throughout the genome of the organism - most fail to produce a discernible phenotype.



An important variation of this assay, needed to adapt it to organisms such as humans, was based on the recognition that many chemicals that you might be exposed to are metabolized in the liver. Such reactions generate related chemicals that may well be significantly more (or less) mutagenic than the original compound. To mimic such metabolic effects, it is possible to add liver extracts to the original culture. Because cancer arises due to somatic mutations, it is clear that we would like to minimize our exposure to mutagenic chemicals. But often a particular chemical is significantly mutagenic only at high concentrations, much higher than you would ever be exposed to. So while many chemicals can induce mutagenesis many fewer are carcinogenic, in part because most mutations are repaired and exposure levels are low enough to have little effect on the baseline mutation frequency.[475]

Questions to answer:

228. How would increasing the mutation rate influence the outcome of the Luria-Delbrück experiment.
229. What are the advantages (for a geneticist) for choosing an organism with hundreds of offspring per mating event?
230. What is the advantage of studying traits that alter non-essential structures?
231. Why does simple mutagenesis fail to identify every gene involved in the formation of a complex trait?
232. What is responsible for the baseline mutation frequency (for example, in the Ames test)?
233. A compound produces mutations in the Ames test; what factors would influence your decision about whether to worry about exposure to that compound?

Questions to ponder:

- Given the frequency at which phage resistance arises, can you provide a plausible reason for why resistance to bacteriophage is not already a universal trait in prokaryotes?
- How would it change your perspective if mutations occurred because organisms need them, rather than randomly?
- How does the apparent fact that evolution depends upon random mutations to generate new genes and new "types" of organisms, new species, influence your view of the meaning of existence?

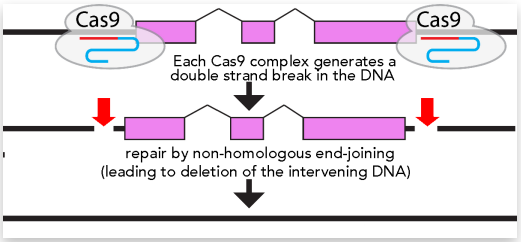
---

[474] Ames test (wikipedia)

[475] "All substances are poisons; there is none which is not a poison. The right dose differentiates a poison…" Paracelsus [link]

Generating mutations rationally - CRISPR CAS9 and related technologies

While early geneticists worked with forward genetics, often known as classical genetics, there are reasons that this approach generally fails to generate a complete map of the genes involved in a particular process. An alternative approach is to determine whether a specific gene is involved in a particular process. While there are a number of ways to identify and then mutate the genes involved in a particular developmental process, the strategies used are largely beyond the scope of this course. Two methods we will consider are single cell RNA sequencing (later in the developmental biology section) and CRISPR-CAS9 mediated mutagenesis, which is one of a number of anti-viral infection systems found in bacteria and archaea.[476] In 2020, Emmanuelle Charpentier (b. 1968) and Jennifer Doudna (b. 1964) won the Nobel prize in Chemistry "for the development of a method for genome editing". The Cas9 enzyme is an endonuclease that creates double-stranded breaks in DNA. What makes the system distinctly different, and extremely powerful, is that the site at which the endonuclease cuts the DNA is determined by a ~23 base pair RNA sequence, a guide RNA (gRNA) – this sequence is long enough to (often) occur once and only once within the genome of an organism, even an organism with a genome of more that a billion base pairs, such as humans. This gives an extremely high degree of specificity to the system. Versions of the system have been engineered to catalyze base changes at the target site, rather than cutting the DNA.[477] In the DNA cleavage system, the cell's DNA repair systems act to join the two ends of the cleaved DNA molecule back together again, but this joining is rarely accurate – base pairs can be lost or added, generating a mutated form of the original DNA sequence. If the gRNA sequence is present in both alleles of a gene, both alleles can be mutated at the same time. One variation, to insure that a region is removed, is to use pairs of gRNAs (→). If the CRISPR-CAS9 system is activated (or introduced) early in the development of an organism all or most cells can be mutated, which can lead to multiple phenotypes. Alternatively, it is possible to activate the system only in specific types of cells, or at specific times of development, allowing for finer experimental control.



Longer term mutation and evolution studies

We can see the spontaneous mutation model applies throughout the biological world, where ever we look mutations appear to arise by chance. If they persist within the population (see above), they become alleles. It is worth reiterating that because of non-adaptive processes such as genetic drift, new neutral or beneficial mutations may be lost because initially they are extremely rare within the population, while mildly deleterious mutations can become fixed by chance.

To study such evolutionary processes in a laboratory setting is not easy, but the now classic example of such a study has been carried out by Richard Lenski (p. 1956) and his associates. They have been growing twelve originally identical populations of the bacteria *E. coli* for more than 25 years and 60,000 generations.[478] One, of many, characteristics of *E. coli* that distinguish it from other bacteria is that it is unable to metabolize citrate in the presence of O_2. In the course of their studies, Blount et al observed the appearance of variants of *E. coli* that could metabolize citrate in the presence of O_2 in one of their cultures; a beneficial evolutionary adaptation, since it provided those

---

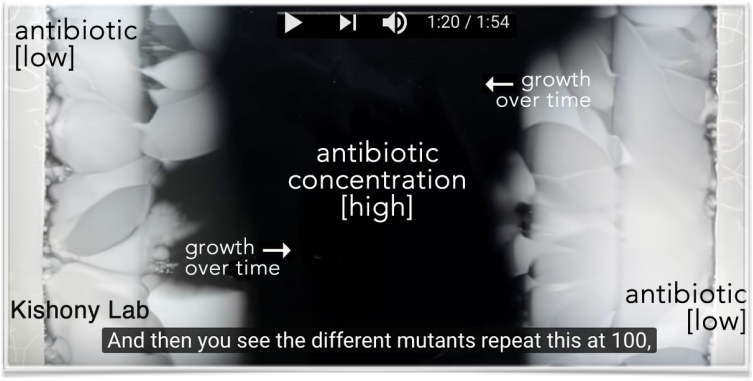
[476] over-view reference for the Crispr cas9 system: wikipedia. The ADDGENE CRISPR website is useful - link.

[477] The next generation of CRISPR–Cas technologies and applications

[478] *E. coli* long-term evolution experiment: wikipedia and the Lenski lab's *E. coli* Long-term Experimental Evolution Project site

cells with a previously un-utilized energy and carbon source.[479] By tracking backward, the investigators identified a "pre-disposing" mutation that occurred in this lineage around generation 20,000. The presence of this mutation made it more likely that subsequent mutations would enable cells to grow on citrate, the Cit+ phenotype. Molecular analyses indicated that the initial Cit+ phenotype, which appeared around generation ~31,500, was weak and involved a ~3000 bp genomic duplication that led to increased expression of the *citT* gene, which encodes a protein involved in the import of citrate into the cell. Subsequent studies identified mutations in other genes in the Cit+ strain that further improved the cells' ability to metabolize citrate.[480] One of these mutations led to increased expression of *DctA*, a gene that encodes a membrane transport protein that increases the cell's ability to import various nutrients normally released into the media, giving the cell a reproductive advantage when grown on citrate. An interesting aspect of these studies was the backlash from some creationists, who reject the possibility of the evolution of new traits via mutation and selection.[481]

A second more recent study on bacterial evolution, this time looking at the evolution of resistance to an antibiotic, used a giant agar plate (a "megaplate") and a gradient of antibiotic (→). Bacterial cells were placed in the regions free of antibiotic, and over time their ability to grow into regions of higher and higher antibiotic concentrations was visualized directly (video link). It is possible to watch the emergence of new variants at the boundary regions, as new mutations arise.[482]



An important point to recall about such bacterial evolution studies is that these organisms are reproducing asexually, as clones. That means that they do not interbreed with other organisms in the population, but it also means that (in the absence of horizontal gene transfer) that all mutations necessary for a phenotype need to occur independently in a single clonal population. As we discussed in the evolution section, if such mutations lead to a reproductive advantage they can, barring accidental death, take over the population – a process known as a reproductive sweep. This can lead to the loss of alleles present in other clones within the population. If these lost alleles were useful (that is enhanced reproductive success), they would need to appear again, independently, through mutation and selection (or be transferred horizontally, something that is not occurring in this system). In sexually reproducing organisms, alleles from different individuals can be mixed to more rapidly produce beneficial phenotypes.

Questions to answer:

234. How can a "predisposing mutation" influence the possible directions of subsequent evolution?
235. In the antibiotic resistance video (watch!), why is there often (but not always) a delay before the bacteria grow into a region of higher antibiotic resistance?
236. How might the presence of horizontal gene transfer impact the megaplate experiment?
237. How might an evolutionary sweep effect a human population?

---

[479] see Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli.

[480] see Genomic analysis of a key innovation in an experimental Escherichia coli population.

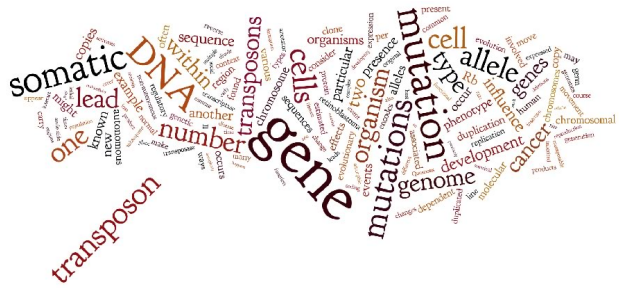
[481] The evolution of citrate metabolizing *E. coli*: the "Lenski affair"

[482] Baym et al., 2016 Spatiotemporal microbial evolution on antibiotic landscapes.

Question to ponder:

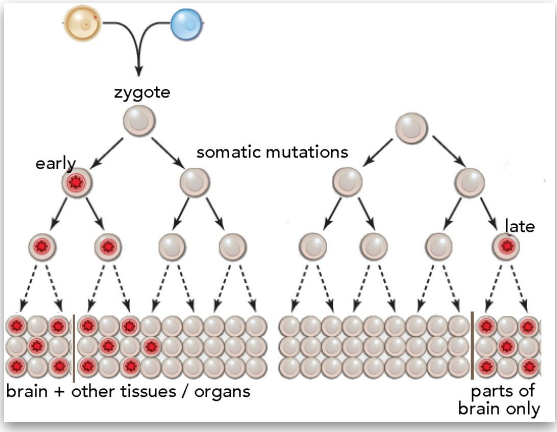
- How would evolution be altered if the mutations (alleles) were induced rather than selected?34

Chapter 14: Somatic mutations & genome dynamics



In which we consider how somatic mutations can influence cellular behaviors and disease and how genes can move around within the genome, often involving the actions of transposable elements.

Up to this point, we have been considering mutations that have become alleles within a population and that are inherited from one's parents. We have considered the shuffling of alleles through meiosis and the formation of a new diploid organism from haploid gametes. Now we introduce the reality of mutations that occur during the development of the organism. First let us reiterate, an inherited allele is present in all cells of the developing and adult organism. With the exception of processes such as X-inactivation and monoallelic expression, it can be expected to have effects on all tissues in which it is expressed. In contrast, when a mutation occurs within a somatic cell, it is passed on as part of a clone (↓), through asexual reproduction. When during development the mutation occurs will determine what percentage of the cells in the organism carry the mutation. If the mutation leads to a lethal phenotype, cells that carry it will die, so no cells in the organism will carry the mutation. More often, such mutations are not lethal but may influence the rate and outcomes of cell divisions.



As noted, a multicellular organism is a social system. Cells cooperate in defined ways to keep the system functioning smoothly. In particular, when and where a cell divides is under strict regulatory control, involving both internal regulatory networks, as well as signals from other cells. Some somatic mutations disrupt this regulatory network, leading to inappropriate cell division, a behavior that underlies the appearance of tumors and metastatic cancer. Carcinogenesis itself is a complex process, involving a number of steps, a number of distinct mutations within a particular clone (cellular lineage). While a complete study of cancer is well beyond us here, certain common features are worth considering. In particular, somatic mutations can lead to cells ignoring signals meant to control their growth and behavior. A mutant somatic cell can generate a clone that will compete with wild type clones in various ways.

Rates and effects of somatic mutation

The rates at which mutations are found within a particular cell type is based on the number of rounds of DNA replication leading to that cell type, the error rate associated with DNA synthesis, the rate of non-replication associated mutations, and the efficiency of DNA error repair. DNA error rates differ between species. In the mouse the current estimate is of one mutation per ~5 x 10^{-9} per base pairs per generation. The number is estimated to be higher in humans, closer to ~1.2 x 10^{-8} per base pair per generation. Mutations rates in somatic cells appear to be higher than in germ line cells.[483] If we think about cumulative effects, that is from fertilized egg to the production of gametes, there are about 400 replication events in a human male, fewer in a female. It has been estimated that, compared with the chromosomes our parents supplied us, we each have ~100 new mutations in our germ line chromosomes.

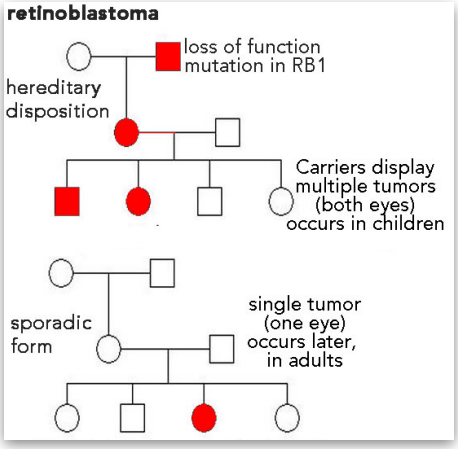
Now consider how a new somatic mutation interacts with the pre-existing genome. For example, consider the differences in outcomes between an organism that is homozygous for a wild type allele

[483] see Differences between germline and somatic mutation rates in humans and mice and from 2019 here

compared to one that is heterozygous, with one allele being wild type and the other being non-functional and recessive. Assuming that a single wild type allele is necessary and sufficient to produce a normal phenotype; a somatic mutation that inactivates a wild type allele in a homozygous cell will have little or no effect, whereas a similar mutation in a cell heterozygous for a non-functional recessive allele will produce a pathogenic phenotype. Two independent mutations will be required to produce a pathogenic phenotype in cells that are originally homozygous wild type. Since the probability of producing two such mutations in the same cell is proportional to the mutation rate squared, this will be a highly improbable event. A mutation induced phenotype is more likely in the heterozygote. The effects of the mutation will depend upon when and where the mutation occurs during the development of the organism. If the mutation occurs early, many tissues may be affected, if late, few and the effects may be restricted to specific organs. In the cells that give rise to the brain, as few as 10% of cells that carry a somatic mutation can lead to a neuronal pathology.[484] As an example, autism (a trait associated with a wide range or "spectrum" of effects) is common, and estimated to occur (to some extent) in ~1% of the population. Both germ line alleles and somatic mutations (rather than vaccinations) have been implicated in such disorders.[485]

The effects of somatic mutations can lead to the loss of growth control, and subsequent over-proliferation - the formation of a tumor, both benign (non-malignant) and malignant. The steps in the formation of a cancer are complex, and reflect a number of regulatory pathways. The first step is often a mutation that leads to a clone that divides when it should not. The mutation turns the well behaved somatic cell into a social cheater (chapter 4). Subsequent mutations can accumulate that enable the cancer clone to get better at competing with its normal neighbors and avoid the host's various defensive responses. The evolution of the cancer clone is, however, ultimately futile. From the clone's perspective, it will continue to divide and grow, but in the end such growth is incompatible with the survival of the host, both the clone and the host will die of the disease, the cancer.[486] There are a number of ways that genes can be mutated to lead to cancer, and a number of ways such somatic mutations can interact with inherited alleles, the details are beyond our scope here.[487]

We will consider just one type of "predisposing" genetic interactions that leads to susceptibility to retinoblastoma, a cancer of the retina. Typically retinoblastoma is rare, but there is a form associated with an inherited dominant, loss of function allele (within incomplete penetrance) in the *RB1* gene, let us call this allele Rb^- (→). Those that inherit a copy of the Rb^- allele have an ~90% chance of developing retinoblastoma early in childhood.[488] Inheriting a single copy of the Rb^- allele is not, however, sufficient to lead retinal cells to become cancerous. A second, somatic, mutation is necessary; this mutation inactivates the wild type *RB1* gene and leads to a dramatic increase in the probability of cancer. People (children) homozygous for the Rb^- allele typically develop multiple tumors in each of their eyes; these tumors appear early in childhood. The presence of the Rb^- allele leads to a hereditary disposition toward developing retinoblastoma.



---

[484] see Somatic Mutation, Genomic Variation, and Neurological Disease

[485] Discovery of autism/intellectual disability somatic mutations in Alzheimer's brains

[486] The exception is the occurrence of cellularly transmissible cancers, described in Tasmanian devils (*Sarcophilus harisii*) and a small number of other species- see Some Cancers Become Contagious

[487] Neomorphic mutations create therapeutic challenges in cancer

[488] Genetics of Retinoblastoma.

People who do not inherit the Rb^- allele can get retinoblastoma; the difference is that they have to accumulate two separate somatic mutations, a much rarer (more unlikely) event. Such rare events do occur, but they tend to occur later in development, so it is unlikely that decedents of the newly (somatic) mutant cell will be present in both eyes. When sporadic forms of retinoblastoma do appear, they are almost always restricted to one eye, and they appear in older individuals. Such somatic mutation are unlikely to affect the germ line, and so will not be inherited. A similar pattern of inheritance is associated with breast cancer susceptibility gene 1 (BRCA1).[489]

Non-disjunction: aberrant chromosome segregation

There is one more genetic disorder that we will consider, but only briefly, namely non-disjunction. Non-disjunction refers to the situation where there is a failure of normal chromosome segregation. In the case of somatic (mitotic) cell division, one daughter cell may receive two copies of a chromosome, while the other daughter receives none; this can lead to lethality or differential reproduction (somatic evolution) within the two resulting clones.

In the germ line, non-disjunction can lead to a gamete containing extra copies of one or more chromosomes, a situation known as chromosomal aneuploidy. Given that each chromosome, even the smallest ones, contains hundreds of genes. The presence (or absence) of the correct number of chromosomes leads to many changes in patterns of gene expression. Generally, when a chromosomal aneuploidy occurs, the resulting embryo fails to complete normal development; recent studies indicate that chromosomal abnormalities are surprisingly common in early humans embryos.[490] For example, when a human embryo carries three copies of one of the smaller human chromosomes, chromosome 21 (the basis for Down Syndrome), it is estimated that ~80% of such embryos perish *in utero* or in the neonatal period.[491] In cases where the early embryo is mosaic for chromosomal abnormalities, somatic evolution in which euploid blastomeres (embryonic cells) replace aneuploid cells appears to lead to normal embryos (and people!)[492]

Questions to answer:

238. A somatic mutation occurs early in development, what factors will influence the % of cells in the organism over time that carry the mutation?
239. How does exposure to mutagens lead to increased risk of cancer development?
240. What types of molecular defects would lead to chromosomal aneuploidy?
241. How might having three (or one) copy of a chromosome influence normal cell behavior (and gene expression)?
242. In the context of the Rb^- allele, how might loss of the chromosome or chromosomal region in which Rb resides influence cellular phenotypes?
243. Propose a model that explains why inheriting a cancer Rb^- or BRCA1 allele lead to increase risk of cancer in some but not all tissues?

Questions to ponder:

- Can you imagine a situation in which a somatic mutation became an inheritable allele in the next generaton?
- How would a mutation in a checkpoint gene influence a somatic cell's clonal evolution?

---

[489] BRCA1 and BRCA2: Cancer Risk and Genetic Testing

[490] Chaos in the embryo

[491] Morris et al. 1999.: Fetal loss in Down syndrome pregnancies. Prenat Diagn. **19**: 142-145.

[492] Mosaicism in preimplantation human embryos: when chromosomal abnormalities are the norm

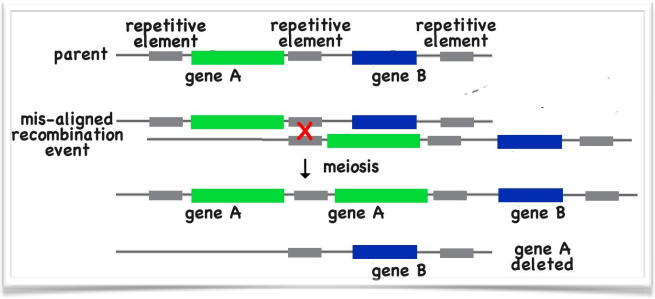
Genome dynamics

Aside from the insertion of “external” DNA through horizontal gene transfer, something that is rare in eukaryotes, and abnormal meiotic recombination events (see below), we might assume that the genome itself, is static. It is, however, clear that genomes are more dynamic than previously thought. In addition to the point mutations that arise from mistakes in DNA replication, a whole other type of genomic variation has been uncovered in the course of genome sequencing studies, these include the movements of transposable elements, discussed below. These are known as “structural variants.” They include flipping of the orientation of a DNA region (an inversion) and sequence insertions or deletions, known as copy number variations.[493] It has been estimated that each person contains about 2000 “structural variants”.[494] Large chromosomal inversions or the movements of regions of DNA molecules between chromosomes can have effects on chromosome pairing during meiosis (described above), and can lead to hybrid sterility and inviability. The mechanisms that lead to these genomic changes are largely beyond our scope here.[495]

An important point with all types of new genetic variants is that if they occur in the soma, that is in cells that do not give rise to the haploid cells (gametes) involved in reproduction, they will be lost when the host organism dies. Moreover, if a mutation disrupts an essential function, the affected cell will die and is likely to be replaced by surrounding normal cells, a version of somatic selection (see above). Finally, as we have discussed before, multicellular organisms are social systems. Mutations, such as those that give rise to cancer, can be seen as cheating the evolutionary (cooperative) bargain that multicellular organisms are based on. It is often the case that organisms have both internal (cellular) and social (organismic) policing systems. Mutant or "eccentric" (that is, misbehaving) cells often actively kill themselves (through apoptosis), can be induced to die by their normal neighbors, or in organisms with an immune system, they can be actively identified and killed.[496]

Gene duplications and deletions

While meiotic alignment generally occurs accurately, there are times were mis-alignment happens. For example, what happens if there are repeated sequences within a chromosomal region. If the homologous chromosomes misalign, crossing over can lead to haploid cells that emerge from meiosis with either gene duplications or deletions (→). Such duplication events can have a kind of liberating effect on subsequent evolutionary pathways.[497] Most obviously, having two copies of a previously single copy gene means that it is possible for the cell/organism to make twice as many transcripts per unit time. This extra activity can be useful. For example, imagine that the original gene product was involved in inactivating an environmental toxin; one copy of the gene might not make enough polypeptide/protein to allow the cell/organism to grow or survive, whereas two copies might. When



[493] Copy number variation in humans:

[494] Child Development and Structural Variation in the Human Genome

[495] Mechanisms of Gene Duplication and Amplification

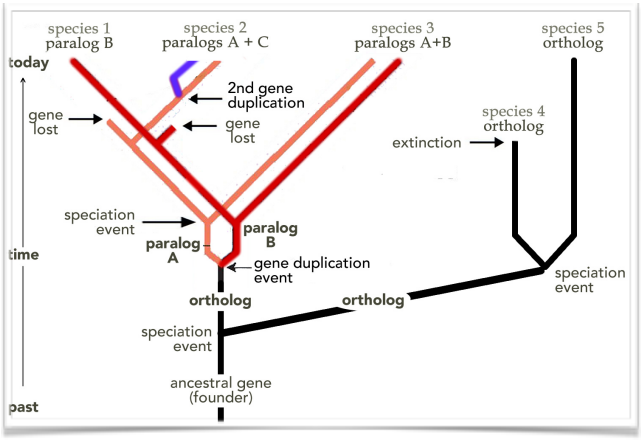
[496] Conceptual simplicity and mechanistic complexity: the implications of un-intelligent design

[497] Ohno's dilemma: evolution of new genes under continuous selection: and Copy-number changes in evolution: rates, fitness effects and adaptive significance

one analyzes bacterial (or cancer) cells that can grow in the presence of a toxic compound, it is not uncommon to find that a gene that encodes a polypeptide/protein involved in the degradation or export of the toxin from the cell has been duplicated one or more times.[498]

Another adaptive mechanism depends upon the fact (noted above) that while a particular gene product may have a clear “primary” activity, it may also have weaker (often much weaker) secondary activities. For example, an enzyme may catalyze "off-target" reactions.[499] Assuming that a gene product’s primary function is essential for survival or reproductive success, changes that negatively influence survival or reproductive success will be strongly selected against, even if they improve valuable secondary activities. In this context, the duplication of the gene allows the original activity to be preserved, while the duplicated gene can evolve freely, and may improve its various, and useful, off-target activities or alter when and where the gene is expressed.

Orthologs and paralogs



When a gene with similar sequence properties is found in distinct organisms, our general assumption is that an ancestor of that gene was present in the organisms’ common ancestor and that the two genes are homologs, or orthologs, of one another. Because of gene duplication events, a gene in an organism (and eventually a population) can be duplicated (→). Even more dramatically, entire genomes appear to have been duplicated multiple times during the course of their evolution.[500] In any gene duplication event, the duplicated genes can have a number of fates, they can act as a “back-up” for one another, they can be re-purposed, or one can be lost. Repeated gene duplication events can generate families of evolutionarily-related genes that are recognized by the presence of similar nucleotide and amino acid sequences and structural motifs in the encoded polypeptides. In the analysis of gene families, we make a distinction between paralogs and orthologs. Orthologs are homologous genes found in different organisms; they are presumed to be derived from a single gene present in the last common ancestor of those organisms. Paralogous genes are derived from a gene duplication event; they are present together in the ancestral organism. If one paralog of a pair is subsequently lost, it can be difficult to distinguish the remaining gene from the original ortholog.

When both paralogs are present in a species, detailed gene/polypeptide sequence comparisons can often be used to distinguish the evolutionary family tree of a gene. That said, the further in the past that a gene duplication event occurred, the more mutational noise can obscure the relationship between the duplicated genes. For example, when looking at a DNA sequence there are only four possible bases at each position. A mutation can change a base from an A to a G, and a subsequent mutation can change the G back to A. With time, this becomes more and more frequent, making it difficult to accurately calculate the number of mutational events that separate two genes, since it could be 0, 1, 2 or a greater number. We can only generate estimates of probable relationships. Since many multigene families appear to have their origins in organisms that lived hundreds of millions of years ago, the older the common ancestor, the more obscure the relationship can be. The

---

498 Dihydrofolate reductase amplification and sensitization to methotrexate of methotrexate-resistant colon cancer cells:

499 Enzyme promiscuity: a mechanistic and evolutionary perspective & Network Context and Selection in the Evolution to Enzyme Specificity

500 Genome and gene duplications and gene expression divergence: a view from plants

exceptions involve genes that are very highly conserved, which basically means that their sequences are constrained by the sequence of their gene product and natural selection. In this case most mutations produce a lethal or highly disadvantageous phenotype, so that cells or organisms with the mutation die or fail to reproduce. These genes evolve (change sequence) very slowly. In contrast, gene/gene products with less rigid constraints, and this includes many genes/gene products, evolve more rapidly, which can make determining the relationships between genes found in distantly related organisms more tentative and speculative. Also, while functional similarities are often seen as evidence for evolutionary homology, it is worth considering the possibility, particularly in highly divergent genes and gene products, of convergent evolution. As with wings, the number of ways to carry out a particular molecular level function may be limited.

Transposons: moving DNA within a genome (and weird genetics)

As we are thinking about DNA molecules moving into the genome through horizontal (lateral) gene transfer, and between genomes through conjugation, we can consider another widely important molecular system known as transposable elements or transposons. A transposon is a piece of DNA that can move (jump) from place to place in the genome.[501] The geneticist and Nobel prize winner Barbara McClintock (1902–1992)(→) first identified transposons while studying maize (*Zea mays*).[502] In particular, she studied the phenomena known as variegation in the pigmentation of kernels (→). The variegation phenotype is due to what are known as unstable alleles; these are pairs of alleles in which one allele is associated with one phenotype (e.g. dark pigment) and the other allele is associated with another phenotype (e.g. lighter pigmentation or a different color). During development an allele can change from one state to another (which is reasonably weird). Since tissues are built from (asexual) clones of somatic cells, the earlier in development an allele change occurs, the larger the region associated with the phenotype in the adult organism, due to the presence of the "alternative" allele.[503]

Transposons can have a number of different effects on the expression of the genes in which they are found.[504] For example, some transposons are found in the coding region of a gene, and are then spliced out of the RNA, resulting in the synthesis of a normally functioning gene product.[505] In other cases, the movement of a transposon can inactivate the gene into which it inserts. Transposons are classified into two general types - those that move a DNA sequence from one place in the genome to another with no increase in total transposon copy number – these are known, for historical reasons, as type II transposons (↓). Type II transposons come in two types, known as autonomous and non-autonomous (dependent). Autonomous transposons encode a protein known as transposase. The transposon is characterized by the presence of repeat nucleotide sequences at each end. The transposase protein recognizes these sequences and catalyzes the removal of the intervening sequence from the original site on the DNA and its subsequent insertion into another site, a site that can be located anywhere in the genome where the chromatin is in an "open" state. In fact, this property has been used to map the regions of the genome that are open, a method

---

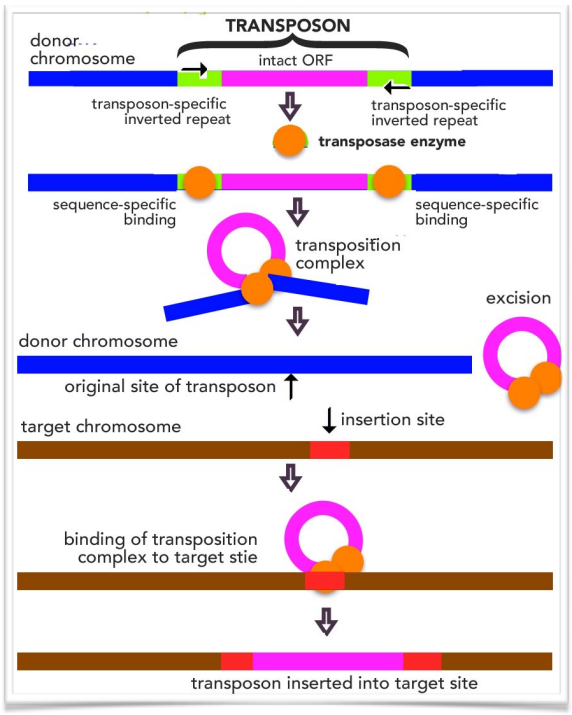
[501] Transposons: The Jumping Genes: http://www.nature.com/scitable/topicpage/transposons-the-jumping-genes-518

[502] Barbara McClintock: http://www.nobelprize.org/nobel_prizes/medicine/laureates/1983/mcclintock-bio.html

[503] In you can't stop yourself, check out: Controlling elements in maize – https://www.ncbi.nlm.nih.gov/books/NBK21808/. We will not go into the genetics of corn, that is something to look forward to in an advanced class in plant genetics.

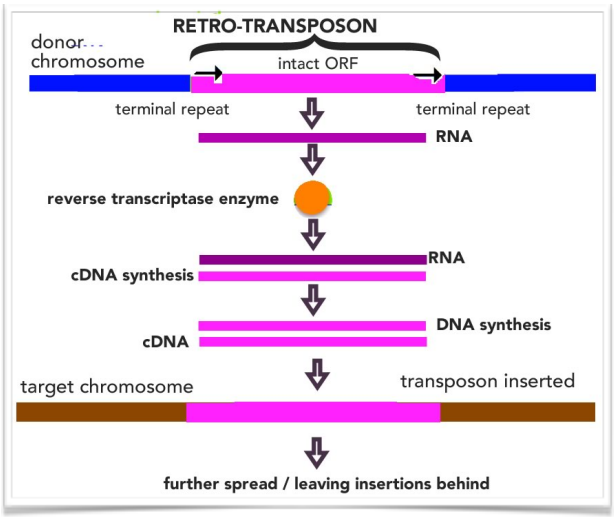
[504] Transposable Elements, Epigenetics, and Genome Evolution: http://science.sciencemag.org/content/338/6108/758

[505] The Maize Transposable Element Ds Is Spliced from RNA: https://www.ncbi.nlm.nih.gov/pubmed/3039661



known as "Assay for Transposase Accessible Chromatin with high-throughput sequencing" (ATAC-seq).[506] In non-autonomous (dependent) type II transposons, mutations have led to the loss of a functional transposase gene within the transposon. By itself, such a dependent transposon cannot move, but if there is an autonomous transposon active within the cell then the transposase it encodes can catalyze the excision and insertion of a dependent transposon. Why? because when the transposase protein is synthesized (in the cytoplasm) it can move into the nucleus and interact with multiple transposons (DNA regions).

The second type of transposon, known as a type I transposon, is also a DNA sequence, but it uses a different mechanism to move. Again type I transposons come in autonomous and non-autonomous (dependent) forms (↓). The autonomous form encodes a protein known as reverse transcriptase. When expressed, the type I transposon leads to the generation of an mRNA that encodes the reverse transcriptase (or RNA-directed, DNA polymerase) protein. The reverse transcriptase can recognize and make a complementary DNA (cDNA) copy of the transposon-encoded RNA. The cDNA can, in turn, be used as the template to generate a double-stranded DNA molecule that can then be inserted, more or less randomly, into the genome. In contrast to a type II transposon, the original transposon's DNA sequence remains in place, and a new transposable element is created and inserted into the genome. If the transposon sequence is inserted into a gene, it can create a null or amorphic mutation by disrupting the gene's regulatory or coding sequences. It can also act as a regulatory element, leading to changes in when and where the gene is expressed. In contrast to an autonomous type II transposon, an autonomous type I transposon encodes a functional reverse transcriptase protein, copies itself, and leads to an increase in the number of copies of the transposon in the genome. In dependent (non-autonomous) type I transposons, mutations in the transposon sequence render the reverse transcriptase non-functional; it can only make copies of itself if another, separate autonomous type I transposon is present and actively expressed within the genome.



Because transposons do not normally encode essential functions, random mutations can inhibit the various molecular components involved in their recognition, excision, replication, and insertion within a genome. They can be inactivated ("killed") by random mutation. If you remember back to our discussion of DNA, human and many other types of genomes contain multiple copies of specific sequences - these are clearly derived from once active transposons, but most are now "dead" – they are the remains of molecular parasites. It is estimated that the human genome contains ~1,000,000 copies of the Alu type transposon (~11% of the total genome); these are dependent, type I

[506] ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide

transposons that rely on the presence of autonomous transposons to move.[507] About ~50% or more of the human genome consists of various dead transposons. It is probably not too surprising then that there is movement within genomes during the course of an organism's life time, since some transposons are still active.[508] Moreover, since transposon movement is generally stochastic, as populations separate from one another, the patterns of transposons within the genome diverge from that of the ancestral population.[509] In addition, various stresses within an organism can enhance transposon movement, which may play a role in the generation of genetic variation - a primary driver of evolutionary diversity and adaptation.[510]

Questions to answer:

244. How many ways can you imagine that the movement of a transposon could influence gene expression?
245. What are the selective pressures on the maintenance or destruction of active transposons?
246. How could the movement of a transposable element NOT produce a mutation?

Question to ponder:

Does the presence of molecular parasites represent an evolutionary design feature or an unintended consequence of molecular machines involved in"normal" DNA dynamics and mutational repair?

[507] Wikipedia: Alu element

[508] Active transposition in genomes

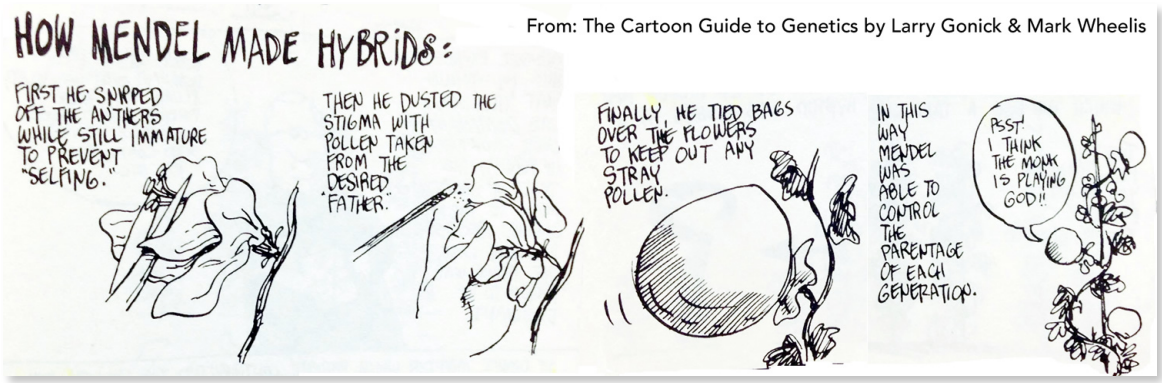
[509] The impact of retrotransposons on human genome evolution: https://www.ncbi.nlm.nih.gov/pubmed/19763152

[510] Stress and transposable elements: co-evolution or useful parasites? https://www.ncbi.nlm.nih.gov/pubmed/11012710

Chapter 15: Becoming Mendelian & recognizing non-mendelian genetic behaviors

In which we (finally) consider the contributions of Gregor Mendel, namely the realization that distinct and semi-stable genetic elements behave in predictable ways during sexual reproduction and influence specific traits. The behavior of chromosomes during meiosis leads to Mendel's rules of allele segregation and independent assortment. We recognize that the traits Mendel studied reflect a unique, and in his case artificially constructed, subclass of genetic elements. We consider how exceptions to Mendel's rules, including linkage, synthetic phenotypes, and epistatic behaviors, arise.

As we think about the historical origins of genetics, it is worth considering the biases imposed by the way that Gregor Mendel (1822-1884) did his work, these reflect the realities of science – understanding does not appear fully formed, like religious revelation, rather it is build up through often experimentally constrained insights, some of which are productive and others that turn out to be distractions. Subsequent observations and experiments lead to the recognition of the implications and limitations of original ideas (tentative hypotheses and working models), and drive their refinement or abandonment. It is worth noting, particularly for the student, that the path from an idea to new discoveries and concrete conclusions is rarely as linear as they are often made to appear to be when the results are presented in a scientific paper. In Mendel's case, he began his work around 1854 and published it 11 years later in 1865; it took 35 years from the time he published his work until it was recognized (1900) as establishing the fundamentals of genetic mechanisms – its significance was not immediately obvious.[511]



How Mendel did what he did

To make genetic behaviors intelligible, Mendel purposefully selected (and bred) plants whose mating partners he could control, that produced high numbers of progeny, and that displayed easily characterized and uniform (from plant to plant) traits. In addition, the traits he chose were independent of one another and were not dramatically influenced by environmental effects (growth conditions). His most famous work involved the garden pea *Pisum sativum,* which displays all of

[511] It is not as though people did not know of his work, "The methodical monk sent reprints of the article to 40 leading biologists around Europe, including Charles Darwin. Darwin's copy was found later, with its double pages still uncut: It had not been read." and "Mendel's work received little notice elsewhere and was cited a mere three times over the next 35 years."

these features.[512] Mating in peas involves male pollen (the plant equivalent of animal sperm). During fertilization a pollen cell fuses with an ovule cell, the plant equivalent of an animal egg. Pea plants can self fertilize, but this can be prevented and the experimenter can control the source of the pollen.

Over a number of years, Mendel identified or developed lines of peas that displayed one or the other of various pairs of traits (↓). In many cases, this involved "breeding out natural variation. A case in point is pea color. The type of pea plants that Mendel worked with normally display a continuous range of seed colors, from green to yellow. Over a number of generations, Mendel selected the greenest and yellowest plants for in-breeding, leading to strains that produced seeds of a uniform green or yellow color - the plants "bred true" for seed color.

Results of all of Mendel's monohybrid crosses

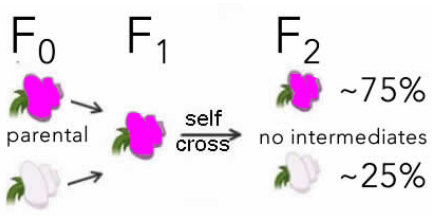
	Parental phenotype	F_1	F_2
1.	Round×wrinkled seeds	All round	5474 round; 1850 wrinkled
2.	Yellow×green seeds	All yellow	6022 yellow; 2001 green
3.	Purple×white petals	All purple	705 purple; 224 white
4.	Inflated×pinched pods	All inflated	882 inflated; 299 pinched
5.	Green×yellow pods	All green	428 green; 152 yellow
6.	Axial×terminal flowers	All axial	651 axial; 207 terminal
7.	Long×short stems	All long	787 long; 277 short

Griffiths et al., 2000

An individual plant is derived from a single pollen grain fertilizing an ovule. To say that a plant line breeds true means that when a plant is allowed to self-fertilize all of the offspring produced display the same form of the trait. These offspring will, if allowed to self-fertilize or to fertilize each other, again produce offspring that display the same form of the trait as the parent.

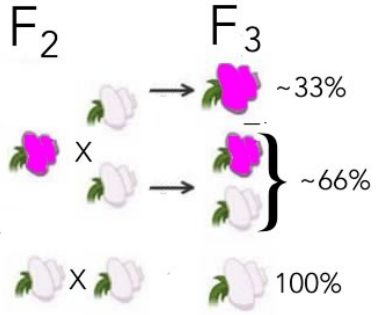
Next he crossed (fertilized) one plant with the gametes of another. For example he fertilized a plant with white flowers with pollen from a plant with purple flowers, and examined the traits expressed in the offspring, known as the F_1 generation. On analyzing the traits of a large number of F_1 offspring he found that among this set of traits, only one of the pair of traits was displayed or expressed. When the parents (the F_0 generation) had purple or white petals, all of the offspring (F_1) individuals had purple flowers. It did not matter if the purple plant was the maternal or the paternal parent. In such a cross the parental trait displayed in the F_1 generation was said to be "dominant" to the "recessive" parental trait, that is the trait that was not displayed. The traits he worked with all behaved in this way. Moreover, when two or three of these traits were displayed in the same individual, they did not influence each other - they behaved independently. The result was not surprising in that Mendel did not start with random traits, he selected traits that followed these rules, they were "well behaved", a fact we will consider further later on.

Mendel continued his experiments by crossing true breeding F_0 individuals expressing one or the other of these traits, to produce F_1 individuals. He then crossed such F_1 individuals to themselves or other F_1 individuals (←). Here was the surprise, from such an F_1 x F_1 cross there emerged F_2 individuals that displayed the recessive form of the trait. As he collected more and more such F_2 individuals, he discerned a pattern - approximately 25% of the F_2 individuals displayed the recessive form of the trait - that is, when a large enough number of individuals were collected there was a clear 1 to 3 ratio of individuals expressing the recessive traits compared to those that expressed the dominant trait. When the F_2 individuals that display the recessive trait were crossed to one another (or themselves), the resulting F_3 individuals all (100%)



[512] Considering the distinction between a study and an experiment. In an experiment, the system is subject to some perturbation and we examine how the system responds. A typical experiment begins with a hypothesis, a guess on how a particular perturbation, which we think we understand, influences the system. A study is more about observing and collecting data about a system. From such observations, we can make hypotheses about how the system will act under various conditions (an observational study) or how a perturbation (an experimental study) will alter the system's behavior. Our prediction of the outcome is known as the null hypothesis - we examine the data collected to determine whether the predication's null hypothesis is supported or not, or whether the data produced could have arising by chance (stochastic fluctuations).

expressed the recessive trait (→). The F_2's that expressed the recessive trait were like the in-bred, recessive F_0 parent. However, the F_2 individuals that displayed the dominant trait were not all the same. When Mendel crossed the F_2 individuals that expressed the dominant trait to recessive F_0 individuals the results fell into two classes (→). In one third (~33%) of cases, all the offspring displayed the dominant trait, while in two thirds (~66%) of the cases approximately half of the offspring expressed the dominant trait and half expressed the recessive trait.



Mendel used such data to come up with a model for trait behavior. He assumed that each trait was controlled by two factors (alleles) at a particular genetic position (locus), what we refer to as a gene. In each of the parental lines these two factors were the same, they are homozygous for either the dominant or the recessive allele of the gene. All of the gametes produced by an F_0 individual therefore carry the same allele of the gene associated with the trait. In a cross between parents homozygous for different alleles of a particular locus, the model predicts that all F_1 individuals are heterozygous and display the same "dominant" phenotype. A heterozygous (for a particular gene) individual will (normally) produce equal number of two different types of gametes; a particular gamete will carry either the dominant or the recessive allele. When an F_1 individual mate, there are four possibilities for the offspring. A gamete carrying the dominant allele can fuse with a gamete carrying either the dominant or the recessive allele. Similarly, a gamete carrying the recessive allele can fuse with a gamete carrying either the dominant or the recessive allele. If we assume that these events are all equally probable, we expect to find two phenotypic outcomes in the F_2 generation in the following proportions (if the number of offspring are large enough) 3 dominant to 1 recessive phenotype. As we might suspect, all of the recessive phenotype individuals are necessarily homozygous for the recessive allele. On the other hand, the individuals displaying the dominant phenotype are not necessarily the same. We discover, using what is known as a backcross to a homozygous recessive individual, that one-third of the F_2 individuals produce offspring with the dominant phenotype only. We conclude that these were homozygous for the dominant allele. On the other hand, two-thirds of the dominant phenotype F_2 offspring, backcrossed to a recessive homozygous individual, produce equal numbers of dominant and recessive phenotype individuals. Based on such studies, we conclude that (given large enough numbers) that the F_2 generation will be 25% homozygous dominant, 50% heterozygous, and 25% homozygous recessive, a 1 to 2 to 1 ratio. Observations were consistent with these ratios for the traits Mendel considered, thereby providing experimental support for his model.

Key to Mendel's model were factors unique to genetic control of the traits he used for his studies. First the variants of a specific trait were unambiguously distinguishable and determined by the genotype (alleles) present at a single genetic locus - these are what are known as monoallelic traits. In addition, the traits had to display clear dominant-recessive behaviors with respect to one another; not all traits behave this way. In some cases individuals heterozygous for a particular gene display a phenotype distinct and often "intermediate" between a "mixture" of the homozygous forms of the trait. Finally, none of the genes associated with the traits he examined were located near each other on any chromosome. They behaved (segregated) independently during meiosis. But remember, Mendel knew nothing about chromosomes and molecular mechanisms, it was just that his choices of genes and alleles that made the data he obtained intelligible and enabled him to build a relatively simple predictive model.

It is worth recognizing explicitly that most traits are controlled in more complex ways than by simple dominant or recessive alleles at a single genetic locus. A particular allele might influence only a limited aspect of one or many phenotypes and may be influenced by the genetic background (alleles present at other genetic loci) within the organism - they may be influenced by allelic variation at hundreds of different genes, and that different alleles of the same gene can produce unique phenotypes, dependent on their interactions with allelic variants at other genetic loci. It is important

to note that many laboratory studies (including Mendel's) are carried out in in-bred backgrounds; all of the organisms in the study may share a common genetic background (similar overall genotypes). Such genotypic homogeneity is an artifact of the way such experiments are conducted; natural populations display much more genotypic variation - there are many different alleles present.

"Background" genetic variation can influence the phenotypes associated with a particular allele, whether hetero- or homozygous. Consider a dominant allele; the phenotypic trait associated with that allele may vary - the extent of such variation is characterized through the terms expressivity and penetrance. So, what do those terms mean exactly? Variable expressivity refers to the observation that even in the presence of the associated (dominant or homozygous recessive) allele, the phenotype may vary. As an example, consider a hypothetical pea; is each pea really wrinkled to exactly the same extent, or do they vary – are some a little more or a little less wrinkly? Such variation in wrinkliness indicates variable expressivity. Similarly, it is possible that out of 100 individuals that carry a particular dominant or homozygous recessive allele, not all display the trait associated with that allele. The extent to which a trait is present is known as its penetrance, the percentage of individuals with the allele that display the trait associated with the allele. Genetic background can influence both the expressivity and penetrance of an allele. If found in a wild, out-bred background, often only a proportion of the organisms carrying a dominant allele (or homozygous for the recessive allele) will display the trait. That allele will be said to be incompletely or variably penetrant; such phenotypic outcomes can be due to various factors, including different combinations of alleles that act to "suppress" and "enhancer" another allele's phenotype.[513] It was for that reason that Mendel restricted his studies to only fully penetrant dominant and recessive alleles; otherwise, the results of his studies would have failed to revealed the simple rules of inheritance that he discovered. Similarly only large numbers of offspring could provide the statistical power needed to come to clear conclusions.

Questions to answer:

247. Why was it critical for Mendel's studies to be able to control crosses between individual plants?
248. What led Mendel to be able to discover recessive alleles?
249. Describe, in terms of meiotic behaviors, how the results of a monohybrid cross are produced.
250. Explain why, when small numbers of offspring are generated, the ratio of phenotypes in a F2 cross can differ from the expected 3:1 ratio.

Questions to ponder:

- Why are backcrosses to homozygous recessive individuals informative? Are backcrosses to homozygous dominant individuals useful?
- How does one determine, in practice, that a homozygous recessive individual is homozygous recessive?

Chi square analysis, hypothesis testing, and numbers that are less than infinity

One limitation of Mendel's work involved the limited number of plants he could examine. The various ratios he predicted are expected to be true and reproducibly observed only when the number of individuals examined becomes large. With smaller numbers of individuals, there can be serious divergences between what is observed and what is (according to the hypothesis or model being tested) predicted, a situation common to stochastic processes. Which gametes contain which alleles and which fuse with one another are both stochastic processes.[514] Consider the general question, how many rolls of a die would you need to perform to convince yourself, with high confidence, that a particular die is fair? or perhaps better put, not unfair. While the stochastic nature of meiosis and fertilization does not effect the (F_1) offspring of a cross between homozygous dominant and

---

[513] here is a particularly relevant recent study: Genetic background limits generalizability of genotype-phenotype relationships

[514] It is similar to the question of which unstable isotope atom will decay next.

recessive plants, in which all offspring are expected to have the same (heterozygous) phenotype, it will influence the 3:1 ratio of phenotypically dominant to recessive plants predicted to occur when F_1 individuals are crossed (the F_2 generation). How do we evaluate whether what we observe is consistent with our model or contradicts it? A model that does not produce the observed results will need to be abandoned or revised.

The answer is a statistical test known as a χ^2 (chi square) analysis.[515] Such an analysis uses this equation (↓) together with two other concepts: degrees of freedom and null hypothesis.[516] If we are testing a model that makes a mathematically precise prediction, such as the frequency of the phenotypic classes observed in a genetic cross, our null hypothesis is that the data are unlikely to be generated simply by chance. Remember, we are not trying to prove that our specific hypotheses are correct; we are trying estimate the probability that the values observed could have occurred by chance.

$$\chi^2 = \sum^{n} \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

To define the degrees of freedom, we need to know how many independent variables there are. In our two phenotype system (wrinkled or round, purple or white, etc.), we assumed that all individuals have either one or the other (unambiguously characterizable) phenotype, if we know the number of individuals involved and the number of either phenotype, we automatically know the number of the other. In the case of two phenotypic classes, the degree of freedom is 1 (if there are four classes, the degree of freedom is 3, and so on). What is the degree of freedom for a six-sided die? By convention, which is currently under some discussion[517], we take an observation to be consistent with the null hypothesis if it can be expected to occur by chance at less that 1 time out of 20 (0.05) or one time out of one hundred (0.01); otherwise we have a good case to reject our hypothesis - that is, the data observed could well be due to a chance occurrence.

For any particular experiment, we make observations to test our null hypothesis, are our predictions supported or rejected? Just for fun, let us consider here (and as a classroom assignment) Mendel's monohybrid crosses. The prediction of his model is that the ratio of round to wrinkled seeds in the F_2 will be 3:1. Mendel reported that he examined 7324 plants. Given his model, he would have predicted that 5492 of these plants would have round seeds, while 1849 plants would have wrinkled seeds. We can now do our χ^2 calculation. We have (5474 (observed) – 5492 (expected))^2 = (-18)^2 = 324/5492 (expected) equals 0.059 and (1850 (observed) – 1849 (expected))^2 = 1^2 = 1/1849 (expected) = 0.00054. The sum (∑) of these two numbers is 0.0595. To determine whether these observations are consistent with our null hypothesis, we consult a χ^2 probability table (↓). The higher the χ^2 value the more likely the difference between observed and expected data is due to chance, rather than because our assumption, our null hypothesis, is correct. Our value of 0.059 lies well below the 0.05 probability value of 3.841, suggesting that the observed numbers are consistent with our model and unlikely to be generated by chance. But keep in mind, consistency does not imply "truth." In fact, there have been suggestions that Mendel's observed numbers are too good, too close to what would be predicted from his model.[518] Be that as it may, Mendel's conclusions for the behavior of the types of traits he chose to study have been repeatedly verified - we can trust his general conclusions given his assumptions.

Chi Square Values and Probability

Degrees of Freedom	$P = 0.99$	0.95	0.80	0.50	0.20	0.05	0.01
3	0.115	0.352	1.005	2.366	4.642	7.815	11.345

---

[515] Here is an alternative presentation from GENETICS AND GENE PROBLEMS

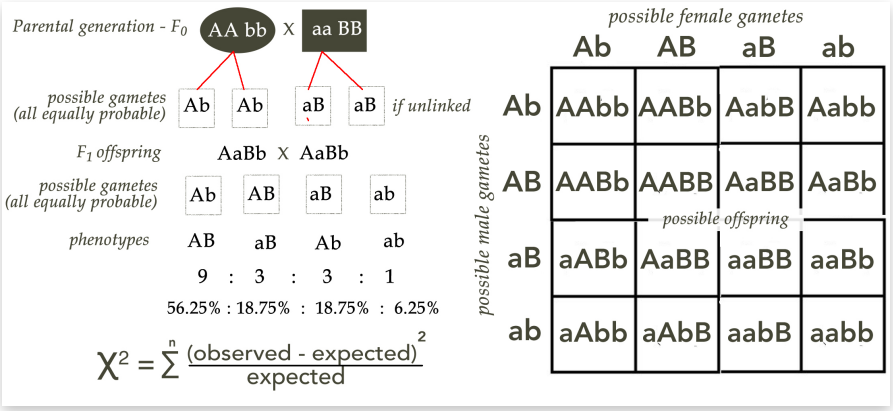
[516] chi square tutorial: http://www.radford.edu/rsheehy/Gen_flash/Tutorials/Chi-Square_tutorial/x2-tut.htm

[517] Statistical errors and Colquhoun. 2014. An investigation of the false discovery rate and the misinterpretation of p-values

[518] see On Fisher's Criticism of Mendel's Results With the Garden Pea

Dihybrid crosses: linkage & recombination

Now we can move to more complex questions. As an example, let us consider two distinct traits (smooth/wrinkled and yellow/green seeds), we can ask, do the alleles involved behave independently of one another or do they interact in some way? We begin, based on a monohybrid analysis, knowing which traits are determined by recessive and dominant alleles. We can assume a null hypothesis, namely that the two traits behave independently; that is they do not interact with one another and that they are not linked to one another. Assume that we begin with two lines that breed true for these traits. As before, each parental F_0 organism can produce only one type of gamete, and all F_1 organisms will have the same AaBb genotype (which is independent of which parent was AA and which was BB). We can then predict the outcome of a cross between F_1 individuals. Assuming that the two genetic loci are independent, we predict that each F_1 individual will produce four different types of gametes in equal numbers and that these gametes will fuse (randomly) with gametes from the other F_1 individual. We can visualize this behavior, and the outcome of the cross, using what is known as a Punnett square (→), which enables us to determine the various possible phenotypically distinct outcomes and their relative frequencies given our assumptions (→).[519] There are 16 possible combinations of these alleles in the F_2 generation, of these 9 display a dominant:dominant phenotype: AABB (1), AABb (2), AaBb (4), AaBB (2); three display a dominant:recessive phenotype: AAbb (1), Aabb (2) or a recessive:dominant phenotype: aaBB (1), aaBb (2); and one (aabb) displays a recessive:recessive phenotype. If we examine enough F_2 progeny we expect to find these phenotypic classes in a ratio of 9:3:3:1. Test crosses to recessive:recessive organisms can be used to identify the genotypes (allele composition) of these various classes of organisms. We can, again, use a χ^2 analysis to determine whether the outcome of a particular dihybrid (two trait) cross is consistent with the hypothesis that the alleles involved do not interact with one another, that they are unlinked.



But what happens if we find that the cross produces the same phenotypic combinations but that the numbers observed do not match our predicted (expected) values - what can we conclude? The simplest conclusion, and one not made by Mendel (because he excluded such traits), was that the i) the genetic loci involved are somehow "linked together", and ii) there are processes that can, on occasion, separate the linked gene - the process of meiotic recombination.[520] Let us consider one such example, we generate a dihybrid F_2 generation from AB phenotype F_1 offspring (the result of a AB X ab cross), and observed the following outcome (→).

expected			
AB	aB	Ab	aa
981	72	86	964
552	394	394	131
observed			

We carry out a χ^2 analysis and obtain a value of 3492. A quick look at the probability table (next page) confirms our suspicion, namely that our null hypothesis, that the genes are unlinked, is rejected. An alternative hypothesis is that the genes are linked to one another and separated by a certain distance; we can now generate an estimate of how

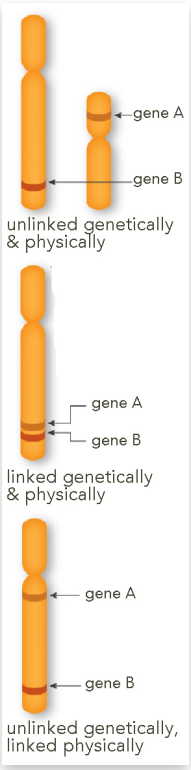
[519] Who was this Punnett fellow? see Reginald Punnett

[520] Why did he missing this type of genetic behavior, because i) he did not have linked traits in his analysis or ii) because he excluded traits that behaved in this way from his analysis - I have not checked with was the actual situation.

closely to one another they line on the chromosome.

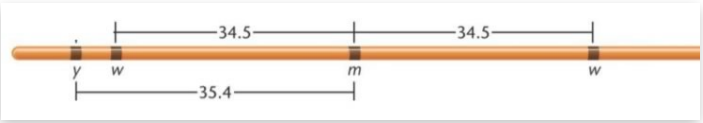
Selected percentile values of the X^2 distribution						
df*	.99	.95	.50	.10	.05	.01
1	.000157	.00393	.455	2.706	3.841	6.635

We know from our cross that the parents (F_o) were AB and ab, and that the chromosomes were AB and ab respectively. If the A and B genes are located on the same chromosome, we can assume that, in the absence of recombination, only [AB] and [ab] gametes will be generated and that all F_1 organisms were [AB][ab], with the brackets indicating that the alleles are linked on the same chromosome. Again, in the absence of meiotic recombination, we can assume that F_1 organisms can produce only [AB] and [ab] gametes. To produce aB or Ab gametes, there must have been a recombination event between the A and B loci. To calculate the frequency at which such recombination (cross-over) events occurred, we add the number of aB and Ab organisms and divide by the total number of organisms, in our case this results in 72 + 86 / 2103 = 0.0751. This indicates a recombination frequency of ~7.5%, significantly less than the 50% recombination frequency we would predict if the genes were unlinked. Recombination frequencies are typically referred to as map units or centimorgans, named in honor of the geneticist Thomas Hunt Morgan (1866 – 1945).[521] A 7.5% recombination frequency equals 7.5 centimorgans.



When the linkage distance exceeds 50 centimorgans (cM), the two genetic loci behave as if they are unlinked, that is, located on different chromosomes, even if they are actually located on the same chromosome (←). It is, of course, possible to walk along a chromosome using pairs of loci located near one another. In this way, we find that a typical chromosome is more than 50 cM in length. Because recombination (crossing-over) can be influenced by the physical state of the chromosome, for example crossing over is often inhibited within the chromosome's centromeric region. Centimorgans do not directly or consistently convert into DNA lengths in base pairs. That said, on average (in humans) a 1 centimorgan recombination distance between genetic loci corresponds to a physical distance of ~1 million base pairs of DNA, 1 megabase (abbreviated Mb). From an evolutionary standpoint it is worth remembering that linkage can influence the inheritance of alleles; the closer two genetic loci (and their alleles) are to one another the longer (the more generations) it will take for recombination to separate them, so that they are inherited independently.

Using conventional genetic methods, we can extend our analysis of linkage from two to three or more genes, in order to identify the order of genes along a chromosome. If two different genes are linked to the same gene, for example, the m gene is linked to the w and the y genes (→), they can be in various orientations with respect to one another.



Genetic crosses using organisms that are originally homozygous for all three alleles, assuming that at least two forms of the alleles at each locus can be identified and that these homozygous organisms are viable, can be used to map genes with respect to one another. This enables one to determine if the w gene is located upstream or downstream, along the length of the chromosome, of the m gene. In an era (like today) of full genomic sequence data, it is easier to use web based tools such as Genomicus [link](see below).

[521] Thomas Hunt Morgan

Questions to answer:

251. What does it mean if the null hypothesis is not supported?
252. A dihybrid cross produces offspring that do not fall into the expected 9:3:3:1 distribution, what kinds of conclusions can we make?
253. In a dihybrid cross, the individuals that are homozygous for both recessive alleles are absent, what might you conclude and why?
254. Alleles in two different genes appear linked to an allele in a third gene, but they do not appear to be linked to each other. What can you conclude and why?

Question to ponder:

- Do genes on opposite sides of the centromeric region of a chromosome appear closer or further away (genetically) than they are molecularly? (assume that recombination is suppressed in the region of the centromere)

Genetic complementation

When we make mutations in various traditional ways, such as by exposure to X-rays or mutagenic chemicals, the organisms carrying these mutations are initially identified for further study based on their phenotypes, typically on how the mutation influences a particular process. The first aspect of such a study is the need to carry out a number of "back-crosses" in order to remove unwanted mutations. Why? Because mutation occurs by chance and generally carried out so as to produce many mutations within each genome so as to insure that genes of interest are mutated. Organisms that carry mutations that influence a specific process need to have such "background" mutations in other genes removed (through sexual reproduction) before they can be studied, and meaningful conclusions reached. The strategies involved in "cleaning up" a mutation vary between different genetic systems, and we will not considered them in detail here.[522]

A priori we do not know whether mutations (alleles) producing similar or related phenotypes, generated following mutagenesis, are in the same or different genes. One way to answer this question is through genetic complementation tests. Let us assume that two (newly defined) mutant alleles influence molecular processes leading to clearly discernible traits. We can use dihybrid crosses to carry out a preliminary examination of the various types of interactions between these alleles. These are outlined in this table (→). As an example, consider two independently derived alleles that produce the same apparent phenotype. Let us assume that we can generate organisms that are homozygous for these alleles, which implies that they are not homozygous lethal. If we cross these, let us call them a1/a1 and b1/b1,

Allelic interactions

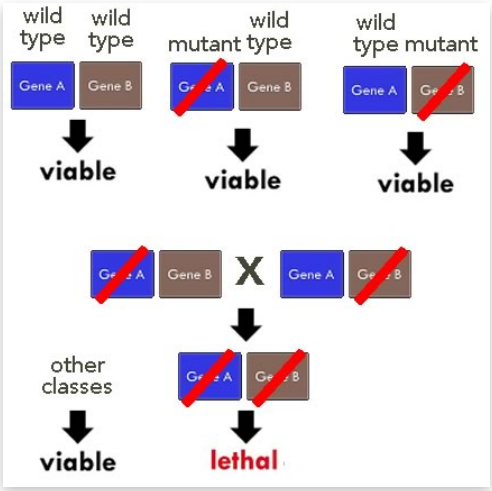
Independent	allele a in a gene is associated with a particular phenotype allele b in a different gene is associated with different phenotype	a/b organism displays both phenotypes.
synthetic	allele a in a gene is associated with a particular phenotype allele b in a different gene is associated with different phenotype	a/b organism displays a new phenotype (such as lethality)
complementary	allele a in a gene is associated with a particular phenotype allele b in the same or a different gene is associated with the same or a different phenotype	a/b organism displays wild type phenotype
enhancement	allele a in a gene is associated with a particular phenotype allele b is in a different gene	phenotype of a/b organism is more severe than a/+
suppression	allele a in a gene is associated with a particular phenotype allele b is in a different gene	phenotype of a/b organism is less severe than a/+
epistasis	allele a in a gene is associated with a particular phenotype allele b in a different gene is associated with different phenotype	a/b organism expresses only one of the two phenotypes.

[522] If interested, check out: The art and design of genetic screens

organisms, we expect that all of the F_1 generation will be genetically the same, at least at these loci. If the F_1 organisms exhibit a wild type phenotype, we can tentatively conclude that these alleles are located in different genetic loci (genes), and have an a1/+ b1/+ genotype. If they display a mutant phenotype, we could tentatively conclude that these are alleles of the same gene, with an a1/b1 genotype. We might seek to confirm these conclusions by asking whether the alleles are linked, although this can be difficult (or impossible) if a1/a1 and b1/b1 have similar phenotypes. We could avoid this problem if we had enough phenotypically distinct genetic markers; that would enable us to determine whether the two genes are linked to the same or different genes. If they were found to be linked to the same markers (allelic versions of other genes), we might conclude that they are alleles of the same gene. If they are linked to different genetic markers, then it is likely that these are alleles of different genes.

Another formal possibility is that these two alleles are in the same gene, but display what is known as intragenic complementation, that is, while the a1 and a2 alleles are both recessive, leading to a mutant phenotype as homozygotes (that is, as either a1/a1 or a2/a2) the a1/a2 heterozygote displays a wild type phenotype. This type of intragenic complementation is relatively rare, since generally both allelic versions of the gene product are inactive (amorphic/null, or hypomorphic), but there are cases, particularly involving proteins composed of multiple copies of the same gene product, in which the combination of allelic polypeptides retains sufficient activity to produce a wild type phenotype. Various other types of allele-specific interactions are possible.[523] This is one reason that researchers often examine multiple alleles of a gene, as well as allelic phenotypes in a number of genetic backgrounds. Genetic backgrounds can have substantial effects on phenotype.[524] Given that different species (such as mice and humans) have dramatically different genetic backgrounds (and evolutionary histories and ecological adaptations), it is not surprising that the same mutation (for example, a null mutation) defined in one organism can produce a different phenotype in another.[525]

Interacting traits: synthetic lethality and co-dominance



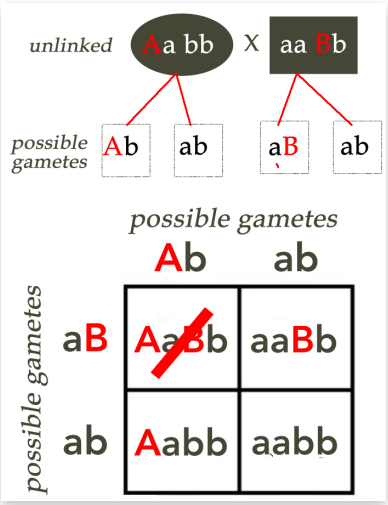
Physical linkage of genetic loci is only one of the ways that genes interact, another involves interactions between gene products and the biological processes they mediate, there are also interactions between proteins encoded an other proteins within the concentrated confines of the cell, which we will consider later in this chapter. Perhaps the most dramatic type of interaction, from the perspective of phenotype (as opposed to molecular mechanism) is known as synthetic lethality.[526] In such a situation, often but not necessarily, carried out with dominant alleles of two distinct genes, both heterozygotes, on their own, are viable, while the double heterozygote is dead, the combination is lethal (→). Similarly, it can be the case for recessive alleles, that individually are viable in homozygous organisms, are lethal or display a different phenotype in double homozygous individuals. We can detect the presence of synthetic lethality through various crosses in which individuals with specific combinations of alleles (such as the dominant A and B alleles) fail to appear in the progeny

[523] Genetic Background Limits Generalizability of Genotype-Phenotype Relationships (a paper cited above)

[524] Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases

[525] Null mutations in human and mouse orthologs frequently result in different phenotypes

[526] Synthetic lethality and cancer



of a cross (← next page). Again, as long as we can identify expected progeny phenotypes, and so count their presence in a population, such deviations from expected outcomes can be detected using a χ^2 analysis similar to our approach to identify linkage.

The presence of synthetic lethality suggests that the two gene products are involved in a common, essential process. Less extreme interaction outcomes are associated with other types of synthetic interactions between alleles of different genetic loci; these are recognized because the phenotype produced by the presence of both alleles is different from the phenotype of either allele on its own. This is different from the behavior of Mendel's genetic factors whose phenotypes are (because of Mendel's choices) independent of one another.

Synthetic phenotypes can arise in a number of different ways. As an example, a process may depend upon multiple gene products interacting to form a functional complex, necessary to produce a trait. Two, often paralogous, genes may produce functionally similar gene products. If one is mutated so as to produce little or no functional gene product, the product of the second gene may be sufficient, but if both are mutant, not enough of the functional complex may be present, resulting in a new version of the trait or lethality. In some cases, alleles of both genes may be recessive, but when present together, they may appear dominant. Such a situation can be generated using various molecular methods, generating what is known as a "sensitized background" that reveals the roles of gene products in specific tissues.

Questions to answer:

255. What types of plausible scenarios can you imagine by which the products of two distinct genetic loci interact to produce a synthetic lethal phenotype.
256. If a gene is missing from a syntenic region, what might have happened to it?
257. How might the level of expression of one gene influence the phenotype associated with another?

Question to ponder:

- Why (and how) did Mendel exclude interacting alleles from his analysis?

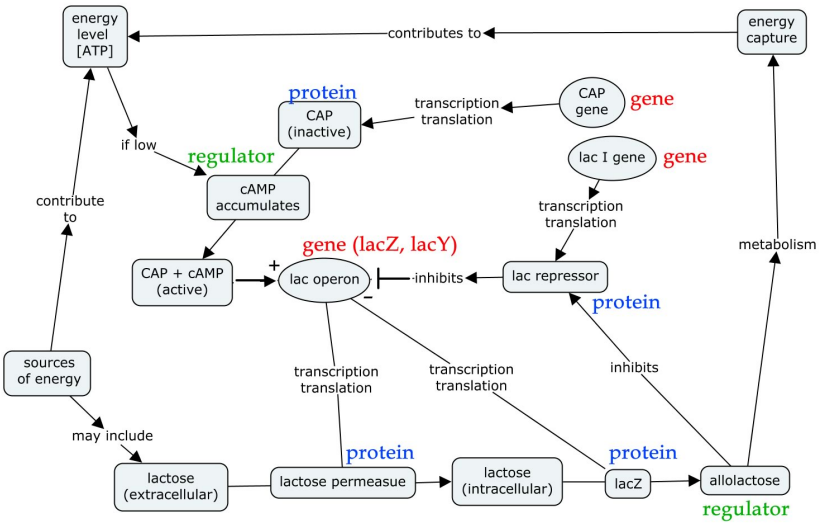
Interacting traits: epistasis

Once mutations (alleles) that alter a particular phenotype, such as eye shape or color, limb formation, or a specific behavior have been identified, they can be used to study the underlying cellular and molecular processes involved. Our first task is to determine whether the mutations are in the same gene or different genes. Different genes are recognized by the fact that they are (generally) unlinked or genetically separable.[527] In the context of any study in which mutations are generated, it is necessary to remember that there are number of possible effects on the gene product, as well as the phenotype, that can arise from a mutation – it is important to characterize the nature of the mutation, an amorphic mutation will behave differently from an anti-morphic or neomorphic mutation.

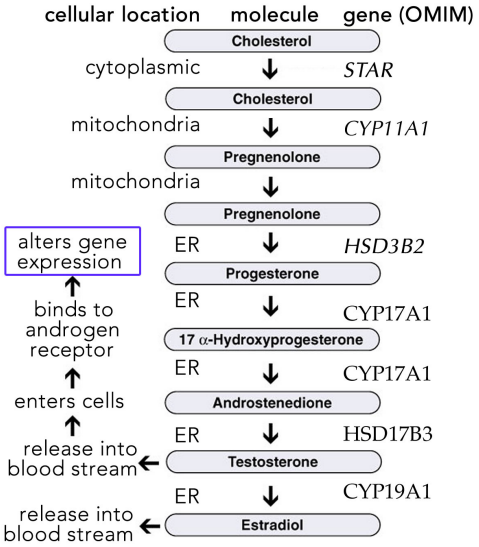
The molecular systems that produce biological behaviors (phenotypes) involve multiple gene products that often act within macromolecular complexes and interaction networks and regulatory feedback. Within such a network, we can consider the types of effects that a particular mutation will have on the phenotype. As an example, let us return to the lac operon. We can generate a schematic of the interactions between genes, gene products, and regulatory molecules - in this case

[527] Traditional processes of generating mutations generate lots mutants throughout the genome; these complicate the analysis. To remove these "background" mutations, mutated organisms that display the trait under study are crossed to wild-type animals, this is known as a backcross. Those organisms that display the trait in subsequent generations selected for further study

lactose, allolactone, and cyclic AMP (→). Based on such a scheme, we could, if we were so motivated, generate a mathematical (graphical) model to serve as the basis for making predictions about the effects of mutations in the various genes involved in the process. These models would include descriptions of the concentrations of various components, binding affinities between molecules, and such. If those predictions are confirmed experimentally, we have increased faith that our understanding of the system is complete; if the predictions are not confirmed, it is possible (likely) that we have missed important components of the system. At the same time, while DNA-dependent, RNA polymerase is a necessary component of the system, required to expressed the genes involved, it is not explicitly included in our model because mutations that alter polymerase function would be expected to disrupt many (essentially all) systems within a cell or an organism, and produce complicating phenotypes. These are known as pleiotropic effects arising from a mutation (allele).[528] Similarly, if any of the components of the system we include are involved in other processes, the model may be influenced by effects on those systems and processes.



In a number of systems, there are parts of the network that act in a linear, or perhaps best termed sequential manner, with one gene product acting on another, "down-stream" aspect of the system. An example is the testosterone/estradiol system; both testosterone and estradiol are derived from cholesterol and both play key roles in the generation of male and female sexual characteristics in mammals. If we begin with cholesterol (ignoring the pathway of reactions involved in cholesterol synthesis), we find a number of gene products, identified by their OMIM designations, that catalyze the various steps in this pathway (→), reactions that occur in both the cytoplasmic and mitochondrial compartments of the cell. Entry of cytoplasmic cholesterol into mitochondria is facilitated by the STAR gene product; within mitochondria, an enzyme (a gene product) catalyzes the chemical reaction that transforms cholesterol into pregnenolone, which then leaves the mitochondria and accumulates in the endoplasmic reticulum (ER). A series of reactions then leads to the formation of testosterone, the "male" hormone, which can be transformed into estradiol, a "female" hormone; estradiol is also involved in male reproductive function.[529] Both testosterone and estradiol are released into the blood stream, allowing them to interact with cytoplasmic receptor proteins (androgen/estrogen receptors) in various cell types. Testosterone and estradiol act as allosteric effectors of these transcription factor proteins, activating them to enter the nucleus and regulate the expression of specific target genes.



[528] Pleiotropy: One Gene Can Affect Multiple Traits

[529] see The role of estradiol in male reproductive function

In the context of such a pathway analysis, we find that the effects of mutations/alleles of genes can be ordered. For example, assume that there is a mutation in the CYP17A1 gene which leads to a non-functional (amorphic or null) version of the encoded protein. In an individual homozygous for this CYP17A1 mutation, we would expect to see the accumulation of progesterone in the ER. Now consider a second null mutation in the CYP11A1 gene, an individual that is homozygous for this mutation would be expected to accumulate cholesterol in mitochondria. So, you may be able to predict the phenotype, in molecular terms, of an organism homozygous for null alleles in both CYP17A1 and CYP11A1 genes, as well as predicting the phenotype resulting from a genetic cross between CYP17A1 and CYP11A1 homozygous individuals, assuming of course that both are viable and fertile. The result of such a genetic analysis allows us to establish what is known as the epistatic relationship between genes (or more accurately gene products) in a particular process.[530]

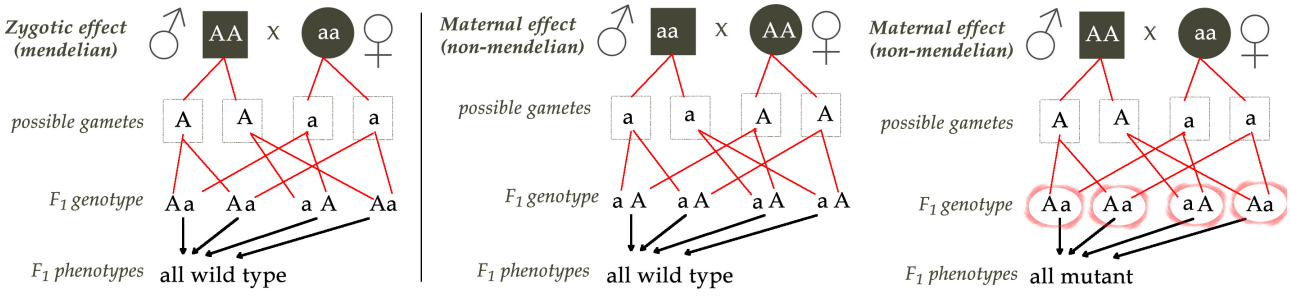
A complicating aspect of most actual interaction pathways is that there are various forms of feed-back and feed-forward interactions that can influence the behavior of a pathway when its normal functioning is inhibited or perturbed. As an example, the accumulation of one compound might influence the expression of other genes, or the activity of other enzymes. In some cases, this can result in a by-pass of the block, so that phenotypic effects are minimized. Consider the cholesterol to testosterone/estradiol pathway - both testosterone and estradiol influence gene expression by serving as allosteric effectors of transcription factors; just as their presence can activate or inhibit the expression of genes, their absence can activate or inhibit the expression of a range of genes. At this point, what is important is to consider what the phenotypes of various genetic crosses might tell you about underlying molecular and cellular systems, while recognizing the limitations of such predictions.

Questions to answer:

258. What factors limit the usefulness of genetic crosses to establish epigenetic relationships?
259. How are genetic pathway maps useful, and what are their limitations?
260. Why is a forward genetic screen unlikely to identify all components of a particular process?
261. Consider a dominant allele in which the associated phenotype is lost on a particular genetic background. How might you reveal the presence of such an allele through a genetic analysis?

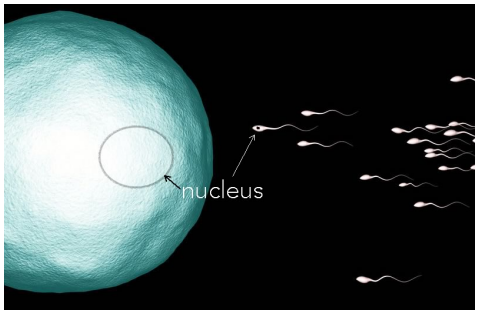
Maternal and paternal effects

Like any other process or trait, embryonic development can be studied and underlying mechanisms identified through the generation and analysis of mutations in the genes that influence the processes involved. From a genetic perspective, there are two general types of mutations (alleles) - there are those that effect the formation of gametes, particularly the egg, and those that effect the process of embryonic development directly or indirectly. Mutations (alleles) that influence oocyte formation and the maternally-constructed developmental environment, are known as “maternal effect mutations”. Take for example a recessive allele “a” - it may be a typical zygotic effect allele or a “maternal effect” allele. Let us consider how they can be distinguished. Begin with a standard cross between homozygous individuals, the outcome will be the same whether the male or



[530] Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems

the female is homozygous for the "a" allele (↑). The traits Mendel use all behave in this way. In contrast, the outcome of the cross will be dramatically different for a maternal effect allele if the female is homozygous for the wild type "A" or mutant "a" allele (↑). In this case, the genotype of the female parent (aa) rather than the genotype of the offspring (Aa) determines the phenotype, a decidedly non-Mendelian behavior. A similar situation will arise if the maternal effect allele is dominant, assuming it effects female reproduction, rather the phenotype of the female itself.



Gamete dimorphism (that is the difference in gamete size and composition)(→) implies that some genes preferentially influence oocyte/egg and sperm behaviors and functions. For example, in a number of organisms, particularly those that develop rapidly and outside of the maternal parent, most of the gene products and nutrients needed to support early development of the new organism are supplied by the much larger egg. Defects in the oocyte, due for example to recessive alleles in a homozygous mother, may lead to defects in the behavior of the fertilized egg and embryo that cannot be rescued by a sperm cell carrying a wild type (dominant) allele - they are dependent upon the maternal genotype and independent of the offspring's genotype. As you might well expect, paternal effects have also been identified.[531]

Mitochondrial inheritance

A obvious example of a maternal effect involves the inheritance of mitochondria. Essentially all eukaryotic cells have intracellular organelles known as mitochondria. Mitochondria have their own genomes, circular DNA molecules known as mtDNAs. A number of genes are encoded by the mtDNA: 37 in human. mtDNAs can, like any DNA molecule, accumulate mutations, whether during replication or in response to free radicals generated during the course of aerobic respiration (something that we will not consider further). Mitochondria are supplied to the zygote by the egg and not the sperm. While mitochondria are present in the sperm cell, they either do not enter the egg or if they do, they and their DNA is destroyed – degraded in various ways, by activated endonucleases and other processes. Mutations in mitochondrial DNA can lead to dysfunctional mitochondria, which can lead to a number of phenotypes.[532] Defects in the mitochondrial genomes present in the egg cannot be rescued by sperm, and so produce a maternal effect on the zygote.

A complexity in the study of mutations in mitochondrial DNA is that each mitochondrion contains a DNA molecule, and the cell contains many mitochondria (hundreds to a few thousand). Different cell types within the same organism can contain different numbers of mitochondria and differ in their dependence on mitochondrial function. The result is that we are looking at populations of mitochondria, with a number of different mitochondrial genotypes. Moreover, the numbers of mitochondria can change, raising the possibility of population bottlenecks and associated changes in genotype. There is the possibility of somatic selection - the differential replication of somatic cells based on mitochondrial genotype and function. In any one cell or tissue, mitochondrial dependent phenotypes will reflect, and be influenced by, the mitochondrial DNA genotypes present – that is, the percentage of mutant (dysfunctional) to wild type (functional) genotypes. A detailed consideration of mitochondrial influences on disease phenotypes in humans and other organisms is beyond us here, but the interested can find a database of mitochondrial DNA mutations at the MitoMap web site.

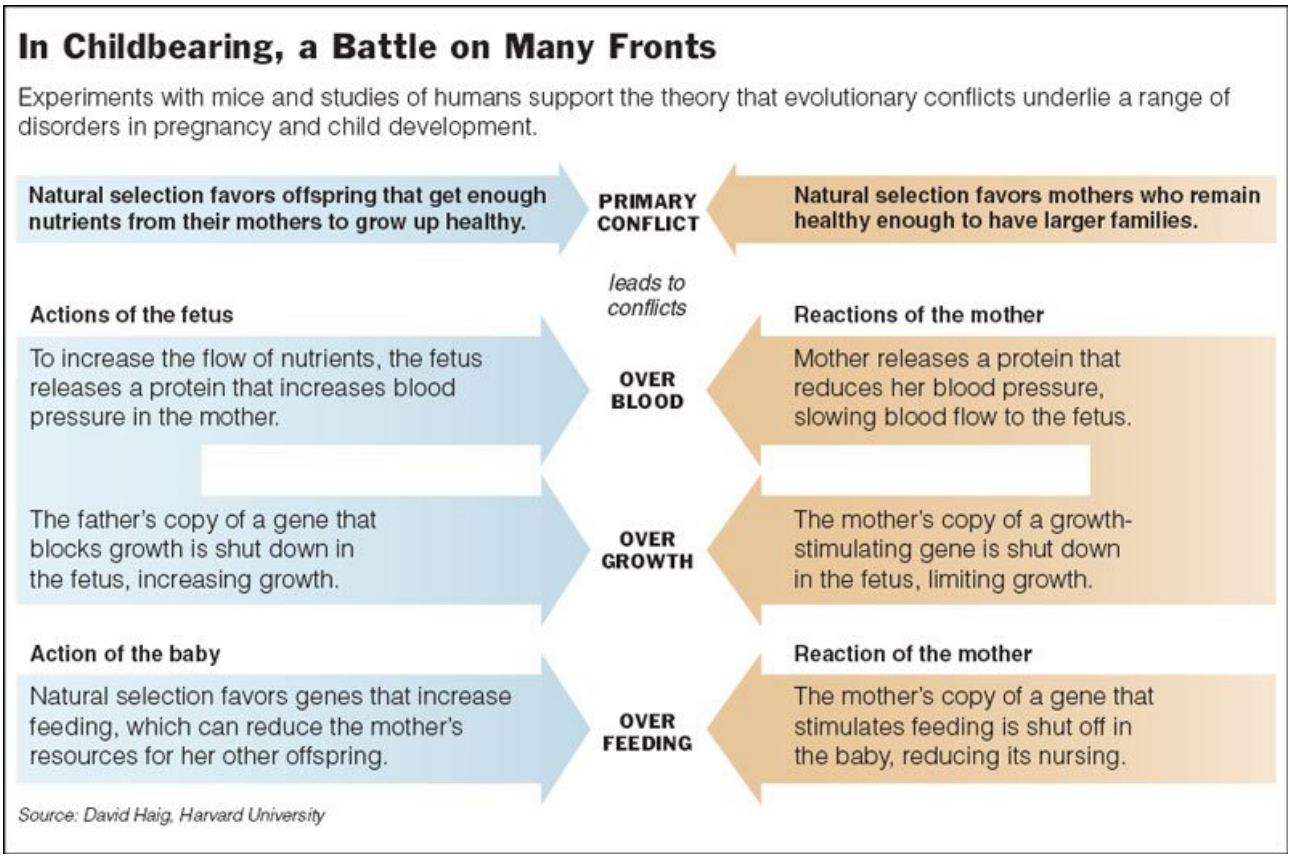
---

[531] What is a paternal effect?

[532] Mitochondrial DNA mutations and human disease

Imprinting: conflicts between mother, father, and fetus

While we have considered sexual selection and the various conflicts between the reproductive interests of the two sexes (particularly in sexually dimorphic species), another conflict that can occur is particularly important in a subset of placental mammals, such as humans. In these organisms, the risks to, and costs on, the mother in raising an embryo are substantial. Under such a condition, carrying the pregnancy to term has the potential to harm the mother, and there may be situations in which it is to the mother's benefit to terminate the pregnancy. In contrast, the embryo's (and in many cases the father's) overriding interest is to be born. Under these conditions, the embryo can benefit from suppressing or modulating the mother's "self-defense" responses. In turn these embryonic defense strategies can be countered by maternal effects on zygotic gene expression (↓). Both



strategies involve a process known as imprinting, in which the DNA of sperm and egg are modified differently.[533] Imprinting involves sequence specific modifications of the DNA; these changes are epigenetic in that they do not alter the gene's nucleotide sequence but rather influence when and where a gene is expressed. Because patterns of imprinting are different in males and females, the maternal and paternal alleles present in a new diploid organism may be expressed differently, that is in some cells only the maternal allele of a gene will be expressed, whereas in other cells only the paternal allele will be expressed.[534]

In a typical scenario the paternal (sperm-supplied) copy of a gene that promotes embryo growth (which if excessive can threaten the survival of the mother) is over-expressed. In response, the maternal (egg-supplied) copy of the gene is turn off. This balances the behavior of the paternal copy, leading to normal development. A similar situation can occur if a maternal gene is expressed,

[533] Genomic Imprinting: http://learn.genetics.utah.edu/content/epigenetics/imprinting/

[534] The origin and evolution of genomic imprinting and viviparity in mammals.

leading to the suppression of expression of the paternal copy. Developmental problem can arise, however, if (for example) the paternal (expressed) copy of the gene is defective or visa versa.[535] Imprinting often involves (it appears) the modification a gene's promoter region. Imprinting complicates things.

Questions to answer:

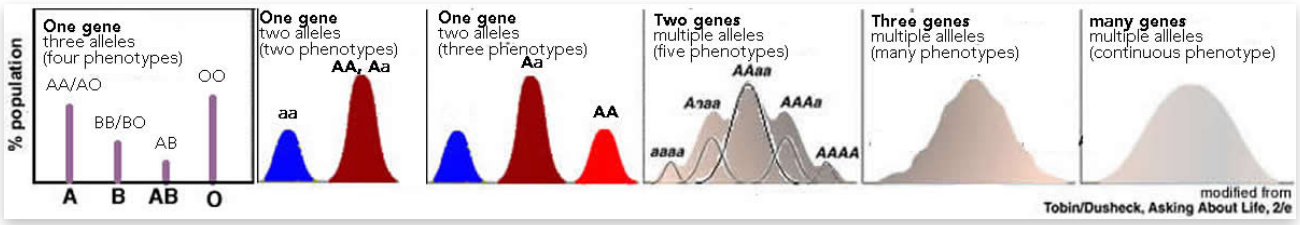
262. How many mechanisms can you imagine that would lead to the expression of different genes in different regions of an embryo?.
263. Describe how imprinting can impact Mendelian allele behavior(s)?
264. Most of the genes involved in mitochondrial function are nuclear; how might that influence the phenotypes of mutations in mitochondrial DNA?
265. If you were to predict which tissues would be more severely effected by mutations in mitochondrial DNA, what would you base your predications on?

Questions to ponder:

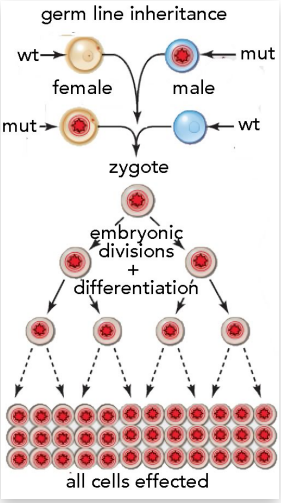
- What has to happen to change the events or timing of early developmental events?
- Explain the evolutionary pressures egg and sperm behavior and the speed of early development.

Estimating the number of genes involved in a particular traits

Mutations that become alleles (enter the germ line and the population) can be seen as lying along a continuum. At one end of this continuum are alleles that behave as do the alleles that Mendel used; these are alleles of a gene that control what we might term discrete features of a particular trait, such as human (ABO) blood type, or a number of genetic diseases that you either have or you do not have (↓ left side). As the number of genes (and the alleles) that influence a



particular trait increases, the distribution of versions of the trait, say for example, height, approaches a smooth curve, a curve often termed a bell curve (right side ↓). Such a distribution is characterized by a mean, a median (which is the same as the mean when the curve is symmetrical), and a standard deviation, which reflects the width of the distribution. The alleles in the various genes involved in a trait can display dominant, recessive, or synergistic (interactive) behaviors.



An important feature of germ line alleles is that all cells of the resulting organism (with the exception of the gametes produced by that organism and any new somatic mutations) will have the same genotype (←). That said, for heterozygous loci, single cell RNA sequence has revealed what is known as monoallelic gene expression, where one or the other allele is expressed. The result can be differences (and selection) between genetically identical cells.[536] Phenotypes associated with a particular allele can vary between cell types. Genes that encode common, often termed house-keeping functions, generally have global effects, while those expressed in only one or a few cell types may have effects in only these cells. The fact that many genes have been

[535] genomic imprinting

[536] Monoallelic Gene Expression on Mammals.

duplicated during evolution, to form paralogous genes, which often have similar although rarely identical functions can also influence the phenotypes associated with various alleles. A gene may be expressed in a particular cell type, but the behavior of the gene product may be more or less critical in those cells because of the presence of functionally complementary gene products (both due to expression of a paralogous gene, or genes in various compensatory or parallel molecular processes and pathways. We saw this effect in our discussion of somatic mutations (see above); a germ line mutation can be inherited but not have a discernible phenotypic effect until a subsequent somatic mutation occurs that disables or alters the functioning copy of the gene, or compromises the function of a complementary gene, a phenotype can arise.

On the nature of mutations (again)

A mutation that changes a single nucleotide position within a gene is known as a point mutation. To produce a phenotypic effect, a point mutation needs to alter a regulatory region, a coding region, or sequences involved in splicing. A point mutation that alters a codon without changing the encoded amino acid is referred to as a neutral or synonymous mutation; such a mutation can have effects if it changes a codon that is recognized by a highly expressed tRNA to a infrequently expressed tRNA, an effect associated with codon bias. tRNAs with different codon-anti-codon interactions, can bind with different affinities. The result is that some codons are misread more frequently than others, leading to an increase probability of a frameshift or even translation termination.[537] When a single nucleotide change alters the amino acid encoded it is often referred to as missense mutation; such mutations can influence the behavior of the encoded polypeptide. If, for example, the altered amino acid forms part of the active site of an enzyme, or its three-dimensional structure, sites of post-translational modification or processing, or influences interactions with water or other polypeptides in the cell of active site, it can alter the polypeptide's assembly, activity, stability, and cellular localization. For example, a single amino acid change can alter the energetics of polypeptide folding; it can misfold and may be unstable as lower (cold-sensitive) or increased (heat-sensitive) temperatures. This underscores the fact that organism typically has an optimal growth temperature. As part of its evolutionary adaptation, its polypeptides/proteins are optimally functional at that temperature, and are relatively less functional at higher temperatures, where they may unfold, or lower temperatures, where they may adopt non-functional configurations. Abnormal protein folding can lead function-disrupting interactions with other molecules in the crowded cytoplasm.

A third type of "point" mutation, known as a non-sense mutation, introduces a non-coding codon, often referred to as a stop codon, upstream of the normal translation termination site. Such a mutation leads to a truncated polypeptide, which can fail to fold or function correctly, and may inappropriately interact with and disrupt the function of other proteins. Because such mutations can be generally disruptive there are mechanisms in eukaryotic cells in which such mutations, when they occur early in the coding region of an mRNA, can trigger the nuclease mediated degradation of the mRNA, a process known as non-sense mediated decay (discussed previously). Degradation of the mRNA suppresses the synthesis of the mutant polypeptide and so mitigates the effects of the aberrant (truncated) gene product.

Alleles, traits, and genetic diseases in humans.

"In human genetics, we try to avoid referring to patients as "mutants," even when it is fully justified scientifically; the word carries unfortunate cultural connotations." D. Botstein. Decoding the language of genetics. 2016. CSH press.

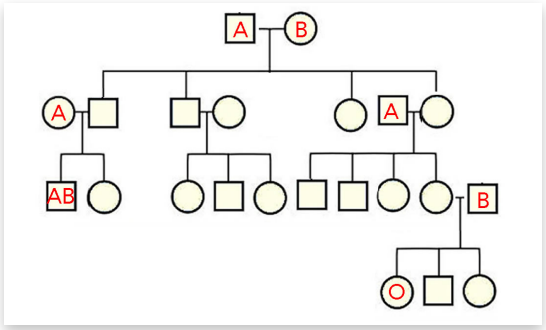
Often mutations lead to alleles; the range of alleles present within a population influence the various phenotypes observed - ranging from body size and shape to disease susceptibility. Some (rare) alleles produce discrete traits that

[537] Different organisms vary in their use of different codons, which form the basis of what is known as "codon bias". Optimal expression of gene from organisms in another (e.g. a bacterium) often involves optimizing the codons used.

behave in a modified Mendelian manner. Perhaps the best known is ABO blood type, which is determined by three distinct alleles of the *ABO* gene, which encodes the protein ABO glycosyltransferase. Both A and B alleles behave in a dominant manner with respect to O, which acts in a recessive manner. A and B behave in a co-dominant manner with respect to one another, that is, when both are present they generate a new phenotype, the AB phenotype. The distribution of these alleles in different human populations appears to be due, at least in part, to selective advantages associated of specific alleles in specific environments.[538]

> *"Mourant suggested that the major differences in the geographical distribution of ABO blood groups may be the consequence of epidemics that occurred in the past. The concept of evolutionary selection based on pathogen-driven blood group changes is currently supported by studies on the genetic characterization of the ABO blood group in Neanderthals and ancient Egyptian mummies. These studies suggest a potential selective advantage of the O allele influencing the susceptibility to several different pathogens responsible for diseases such as severe malaria, H. pylori infections and severe forms of cholera".*

Because blood type can be determined unambiguously, the mode of interaction of these alleles is well defined, it is possible to trace their inheritance across multiple generations. If we know an individual's blood type, we have an initial (although extremely incomplete) model of their genotype. As we examine the phenotypes of their progeny, we can further constrain their genotypes. In such studies, we assume that we know with certainty who mated with whom, something that is often not true. In this family free the presence of an AB individual in the second generation (←), indicates that the male parent must (if they are the father) have had an AB or BO genotype, other genotypes could not have been produced by the parental (A X B) cross. Similarly in the lineage giving rise to the O individual, we can conclude that its male parent had to be BO, while its female parent had to be OO. The more of the individual phenotypes we know in a pedigree, the more we can constrain the genotypes of members of their lineage.



In the modern world we can use molecular markers to identify the alleles present in specific individuals. One issue with such pedigree analysis is that it can lead to potentially embarrassing or disruptive conclusions; for example revealing that a father cannot be the biological father of a child. Generally, but not always, who the mother of a child is is more unambiguous.[539] Molecular details can influence these conclusions. For example of the A and B alleles encode enzymes that catalyze distinct reactions (giving rise to the A and B phenotypes), while the O allele encodes a non-functional enzyme. The reactions catalyzed by the A and B enzymes are dependent upon another "upstream" enzyme - a fucosyltransferase, the product of another gene, necessary to create the substrate upon which the A and B enzymes act. If this enzyme is not present (due to a non-functional allele of that gene) a person with an A or B allele can display an O type phenotype even though genetically that are not homozygous of the O allele.

> *"When people think of skin color in Africa most would think of darker skin, but we show that within Africa there is a huge amount of variation, ranging from skin as light as some Asians to the darkest skin on a global level and everything in between. We identify genetic variants affecting these traits and show that mutations influencing light and dark skin have been around for a long time, since before the origin of modern humans." – Sarah Tishkoff*

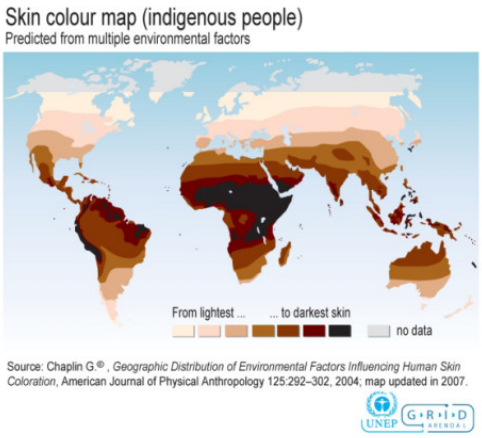
A different type of trait that differs between populations, as well as between individuals within a population, is skin color a trait linked to skin

---

[538] Beyond immunohaematology: the role of the ABO blood group in human diseases

[539] That said there are strange situations, often involving embryological events, that can lead to unexpected results [link to add]

exposure to solar UV radiation and the role of UV light in the synthesis of vitamin D.[540] The extent of exposure of skin to sunlight depends on a number of factors. As genomic studies include more people from geographically diverse groups, DNA sequence analyses have revealed that a number of genes are involved in the determination of skin color. As one might predict, given that people originated in Africa, African populations are expected to display the most dramatic genetic diversity in skin color, a prediction confirmed by direct observation. Genomic studies indicate that four genomic regions (genes) are responsible for ~30% of the variation in skin pigmentation (the rest is due to allelic variation in a number of other genes.[541] Based on modern primates, it appears that our primate ancestor had largely unpigmented skin, but were protected from sun damage by fur. Skin pigmentation is expected to have increased as fur was lost, an adaptation a to more active (heat-generating) life style, dependent on more effective cooling of the body. As human populations migrated away from their site of origin within Africa different levels of UV exposure impacted their adaptation to the antagonist pressures of skin damage and vitamin D production, leading to selection pressures based on skin pigmentation.[542] As populations migrated away from the equator, reduced levels of skin pigmentation were selected (←).



Concordance between monozygotic twins and genetic influence on a trait

An interesting phenomenon that can be used to characterize the genetic contribution to a trait involves twins. There are two generic types of twins. Fraternal twins involve two eggs, and two sperm, leading to two distinct embryos developing together within the mother, and generally both born in rapid succession. Fraternal twins are no more or less closely related than are two siblings born years apart, except that the uterine environment is distinctly different. Fraternal twins are also termed dizygotic twins, since they involve two distinct pairs of zygotes. In animals that typically have multiple offspring (litters), the individuals born generally arise from distinct zygotes. In contrast, identical twins are known as monozygotic twins. Identical twins occur because a single sperm fertilizes a single egg, and generates a single zygote, which begins development. During development, for one reason or another, the embryo fragments and produces two embryos, rather than one, which then develop independently of one another. So, with the exception of (somatic) mutations that occurred independently during embryonic development, the two individuals are genetically identical. This genetic identity enables us to measure the genetic concordance of a trait.[543] For example, if a trait is determined solely by the individual's genetics, then the concordance between identical twins should be 100% (blood type is one example). In other cases, while genotype plays a role it is not completely determinative. As an example, in the auto-immune muscle weakness disease myasthenia gravis, the genetic concordance is ~35%, a level of genetic concordance that implies other factors play an important role in the appearance and progression of the disease.[544] These can include stochastic effects on gene expression and cell behavior.

---

[540] Evolution, Prehistory and Vitamin D

[541] Genes responsible for diversity of human skin colors identified: (paper) Loci associated with skin pigmentation identified in African populations

[542] Low levels of vitamin D can lead to the skeletal malformations; in women this can affect the pelvis and lead to higher levels of fetal and maternal death.

[543] Does Higher Concordance in Monozygotic Twins Than in Dizygotic Twins Suggest a Genetic Component?

[544] Immunopathogenesis in myathenia gravis and neuromyelitic optica.

As we are talking about twins, it is worth noting (for completeness) another type of outcome, which is known as a chimera.[545] In a chimeric embryo, two embryos fuse into one - such that a single organism develops, but it has two distinct "sibling" genotypes.[546] When dizygotic fusion is complete, a single normal, albeit mosaic, embryo and mature organism is generated, a situation that can lead to genotypic confusion. When fusion is incomplete, or occurs at a later developmental stage, incompletely fused embryos are formed - what are known as conjoined twins.

Measuring evolution's impact on allele frequencies: Hardy-Weinberg

If we consider a population, each gene is represented by some set of alleles. In a particular population, different alleles are present in different frequencies. These differences reflect the history of the population and evolutionary pressures. To determine whether evolution is occurring within a population, we use what is known as the Hardy-Weinberg (H-W) equation, based on the work of G.H. Hardy (1877-1947) and Wilhelm Weinberg (1862-1937) – published independently in 1908. Their analysis was based on a set of five assumptions: 1) the population is infinite, so that processes such as genetic drift do not occur; 2) the population is isolated, so that no individuals leave or enter; 3) no new mutations occur; 4) mating between individuals is random (no sexual selection); and 5) there are no differential reproductive effects, that is, natural selection is not occurring.[547] Under these (completely unreal) conditions, the allele frequencies found in the initial population do not change over time. If, on the other hand, allele frequencies are found to change, selection (or some other process) must be occurring.

Before Hardy-Weinberg's analysis there was a belief that dominant alleles were somehow "stronger" than recessive alleles, that "dominant alleles must, over time, inevitably swamp recessive alleles out of existence. This incorrect assumption was called "genophagy", literally "gene eating""[548], but this is not the case unless the alleles influence reproductive success, that is, unless positive or negative selection are occurring.

So let us consider the situation in which there are only two alleles (A and a) of a particular gene. If the frequency of A in the population is p, the frequency of a is q. It is clear (hopefully) that $p + q = 1$. We can then calculate the frequency of homozygotes and heterozygotes by expanding the term $(p+q)^2$; simple mathematical considerations indicate that within this population, the probability of an AA homozygote is p^2, the probability of an aa homozygote is q^2, and the probability of an Aa heterozygote is 2pq, such that:

$$p^2 + 2pq + q^2 = 1.$$

How is this possible? remember, both p and q are less than 1. Our null hypothesis is that these alleles are NOT subject to natural selection, which means that they have no effect on reproductive success within the population. Now we can look at the frequency of recessive homozygotes in a population and calculate the χ^2 value and use it to estimate whether the population is at equilibrium, that is, no evolutionary changes are occurring, or whether there is active selection for or against certain alleles. For example, it might be that homozygous recessive individuals are either not viable, they die, or they are not fertile, or that their offspring die more often that the offspring of others. Alternatively, the heterozygote might have a reproductive advantage compared to the recessive homozygote; such a heterozygote reproductive advantage can maintain significant levels of an allele that is deleterious as a homozygote within a population. The classic example of such behavior are mutations associated with the hemoglobin B (*HBB*) gene of humans. Alleles of this gene are

---

[545] It is even possible to generate chimeric embryos between different species: Humanized mice and porcinized people.

[546] Such human chimeras have been identified: see 3 Human Chimeras That Already Exist and One Person, Two Sets of DNA: The Strange Case of the Human Chimera

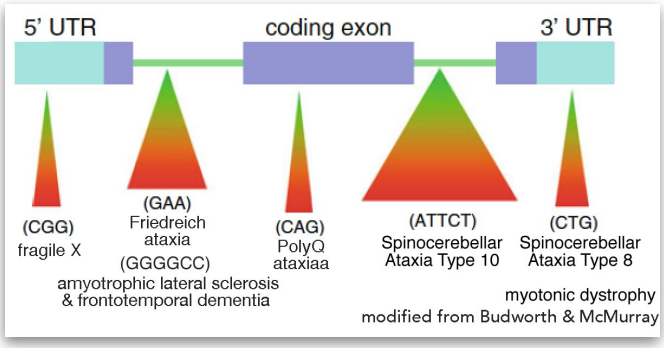
[547] Hardy-Weinberg Equilibrium: http://www.tiem.utk.edu/~gross/bioed/bealsmodules/hardy-weinberg.html

[548] genophagy

associated with a dominant trait, resistance to malarial infection, as well as a homozygous (often lethal) trait, sickle cell anemia. While the recessive trait is subject to strong negative selection, the dominant trait is subject to positive selection in environments where malaria is endemic. The same allele is responsible for both traits.

Genetic anticipation

There is a type of inherited allele that differs in interesting ways from conventional alleles, these are alleles that change from generation to generation, a behavior that has been termed genetic anticipation (discussed previously). Such alleles are associated with what are known as "trinucleotide repeat" expansion diseases, although some involve sequences longer than repeating triplets, and are known as microsatellite expansion mutations. Such repeated microsatellite sequences (3 to 6 repeating units) account for ~30% of human genome sequence. Nucleotide repeat expansion diseases include several forms of mental retardation, Huntington's disease, inherited ataxias, and muscular dystrophies.[549] Within the genes involved, there are regions of repeating nucleotides. Because of the slippage of the DNA polymerase during DNA replication, the number of such repeats can grow bigger or smaller. The result? the allele delivered to an offspring can be more deleterious than the allele present in the parent - over generations, the symptoms of such an allele grow more and more severe. The length of the repeat correlates with the age of disease onset, but the age of onset is variable between individuals with the same repeat length, suggesting the impact of various genetic modifiers. In addition to standard inheritance, many of these genes play roles in the function of nervous tissue, and it is possible that somatic (as opposed to germ line mutations) can influence the allele associated phenotype. As an example, there is evidence that genetic anticipation is important in the context of schizophrenia and bipolar disorder, which together occur in ~1% of the population and have an estimated ~80% heritability risk, which means that on average, about 80% of the differences between individual organisms is due to genetic factors. Of course such estimates depend critically on how accurately various phenotypes can be recognized and quantitated.



Mechanisms: The number of sites in which nucleotide repeats are found, and where their expansion can lead to disease (→) implies a number of possible mechanisms behind the pathogenic state. First, all of the pathology-associated nucleotide expansion regions appear to occur within the transcribed region of the gene, and that includes the 5' and 3' untranslated regions, as well as within introns and exons. For example, if such a domain occurs in a coding region it can lead to increased stretches of repeating amino acids in a polypeptide. Alternatively, they may reflect toxic interactions between the transcribed RNA and other cellular components. To illustrate the potential complexity (a full exploration is beyond our scope here), consider recent work on the role of a nucleotide expansion domain in the gene *C9ORF72* (OMIM: 614620), which encodes a polypeptide implicated in vesicle trafficking within the cell. The expansion domain effects within *C9ORF72* gene have been linked to both amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). Studies indicate that the expanded nucleotide region is targeted for inappropriate transcription; RNAs are synthesized bidirectionally from both sense and anti-sense DNA strands and "that RAN (repeat-associated non-ATG translation) translation occurs from both sense and antisense expansion transcripts, resulting in the expression of six RAN proteins (antisense: Pro-Arg, Pro-Ala, Gly-Pro; and sense: Gly-Ala, Gly-

[549] A Brief History of Triplet Repeat Diseases

Arg, Gly-Pro)".[550] These proteins accumulate in cytoplasmic aggregates in affected brain regions.[551] Interestingly, another gene product, encoded for by the *Supt4H1* gene (OMIM: 603555) appears to play a role in the inappropriate transcription of the C9ORF72 gene; reducing the levels of the *Supt4H1* gene product ameliorates the phenotypic effects of nucleotide expansion in C9ORF72.[552] The exact mechanisms of these types of alleles and associated phenotypes are complex, based likely on the effects of altered transcription on the functional roles of specific cell types.[553]

The persistence of deleterious alleles

A number of genetic disorders display clear Mendelian inheritance (see Specific Genetic Disorders). What does this mean? Basically that inheriting specific alleles leads to the disease, and that these alleles act in a simple dominant or recessive manner, although variations in expressivity and penetrance factors may be involved. In the case of dominant disease-associated alleles, to be inherited means that they are not lethal as heterozygotes, and so result in fertile individuals, otherwise they could not pass the allele on to the next generation. Recessive alleles can be lethal in the homozygous state (as might be dominant alleles), but heterozygotes must survive and be able to reproduce. Keep in mind that the terms recessive or dominant are always in reference to specific phenotypic traits. An allele can be recessive with respect to one phenotype and dominant with respect to another.

You might well ask yourself, given the effectiveness of natural selection, why do alleles that produce severe diseases persist? There are a number of possible scenarios that the previous discussion should help you consider. One is that new mutations are continuously arising, either in the germ line of the organism's parents or early in the development of the organism itself. The prevalence of the disease will reflect the rate at which pathogenic mutations arise together with the rate at which the individuals carrying them are eliminated (before they have off-spring). The second, more complex reason involves the fact that in diploid organisms there are two copies of each gene and that carrying a single functional copy of a recessive disease-associated allele might have no discernible effect, or may even enhance the heterozygous organism's reproductive success. In this case, the recessive pathogenic allele has a dominant positive effect leading to an increase in allele frequency (as in the case of malarial resistance associated with the sickle cell allele). Such a heterozygous effect can be sufficient to maintain the allele in the population at a significant level. Similarly the effects of a dominant allele associated with a pathological condition can be ameliorated, or even beneficial in the presence of various genetic modifiers (enhancers or suppressors). Eventually the population will reach a point where negative and positive effects balance. This is better considered a "steady state" than an equilibrium, since selection is active, but positive and negative, that together effect the final balance (allele frequencies). Of course this steady state is sensitive to changes in the environment that influence phenotype and their effects on reproductive success. If we were being more mathematical, one could model the system based on such effects.

The pace of selective effects depends upon population size and the strength of the selection pressures. As selection acts, and the population's allele frequencies change, the degree to which a particular trait influences reproductive success can also change. The effects of selection are not static, but evolve over time. For example, a trait that is beneficial when rare may be less beneficial when common, and competition between individuals that express the trait increases. New mutations that appear in the same or different genes can influence the trait and selective effects, leading to

---

[550] Non-ATG–initiated translation directed by microsatellite expansions

[551] RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia.

[552] Spt4 selectively regulates the expression of C9orf72 sense and antisense mutant transcripts

[553] C9orf72-mediated ALS and FTD: multiple pathways to disease

changes in the population over time. The example of the evolution of the ability to utilize citrate (described above) appeared in a population pre-disposed to such a change.

Questions to answer:

266. Consider conditions in which the deletion of a gene might lead to a selective advantage.
267. How might you determine whether the appearance of an allele in a population is due to a new mutation, as opposed to some other mechanism (or is there no other way?)
268. How can combinations of alleles in different genes lead to new traits?
269. In the case of genetic anticipation, what is the impact if the repeat domain gets shorter?
270. How might the synthesis of small polypeptides influence normal cell behavior?
271. How would a repeat domain influence a coding region?

Questions to ponder:

- Do genomes always become more complex over (evolutionary) time? Why might they become simpler?
- Are there broader implications arising from the maintenance of deleterious alleles within a population?

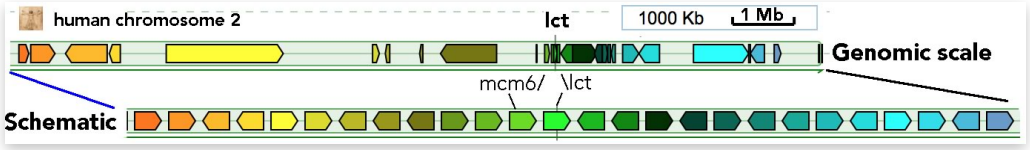
Chapter 16: Tools to study genes & genomes

In which we consider a few tools available to examine the organization and evolution of genes and genomes and the role of genes and gene products in

As we think about genes and the functional roles they play within biological systems it becomes increasingly useful to understand available methods, their power and limitations. We now know a lot about biological system at the single cell (and single nucleotide) level, and can begin to piece together how they function. Here we will introduce a few useful tools.

Synteny examined using Genomicus

In Genomicus, the user inputs a gene name and the system displays the gene in its genomic context, that is within a chromosome, as well as the genomic positions "of all its orthologous and paralogous copies in all the other sequenced metazoan genomes" together with "predicted ancestral genome structure".[554] In this example (↓), we inputed the gene name *LCT* (OMIM: 603202), a gene than encodes the enzyme lactase, the enzyme that enables mammals to digest lactose, and so survive on their mother's milk, one of the defining traits of mammals. In most mammals, the *LCT* gene is expressed in infants and then turned off as they mature into adults. The trait "adult lactose tolerance" is found in populations of humans known to raise domesticated animals from which milk can be harvested, and so provide a significant source of energy and nutrients. Adult lactose tolerance is associated with a ***failure*** to turn off expression of the *LCT* gene in adults.[555] Molecular studies indicate that expression of the *Lct* gene in adults is negatively regulated by an enhancer element ~14 kbs upstream of the *LCT* gene, located within an intron of the *MCM6* gene. Mutations within this enhancer element are found in populations in which adult lactase tolerance is common, apparently due to positive selection.[556] Genomicus enables us to analyze the region around the *LCT* gene. Two views are possible, in the genomic scale view, the genes are displayed based on their actual size in base pairs), relative locations, and the direction of transcription, indicated by a pointed box (↑). Different genes get different colors and the direction of the box indicates the direction of RNA synthesis; here are two genes that are transcribed in opposing directions  , hopefully you can explain how such a thing is possible. While each pointed box indicates the region of the gene, it does not show the positions of introns and exons. Intergenic regions (the regions between genes) are indicated, with their relative lengths accurately displayed. In the schematic view, each gene is again indicated by a pointed box, but now all genes, no matter their actual length, are indicated by the same size box. It can be easier to recognize genes in the schematic view. On the web, holding your cursor on a gene (in either view) will display the gene name and more information about it. Note that the *MCM6* gene is located adjacent to the *LCT* gene. We could, if we wanted to, walk along the chromosome (the *Lct* gene is located on human



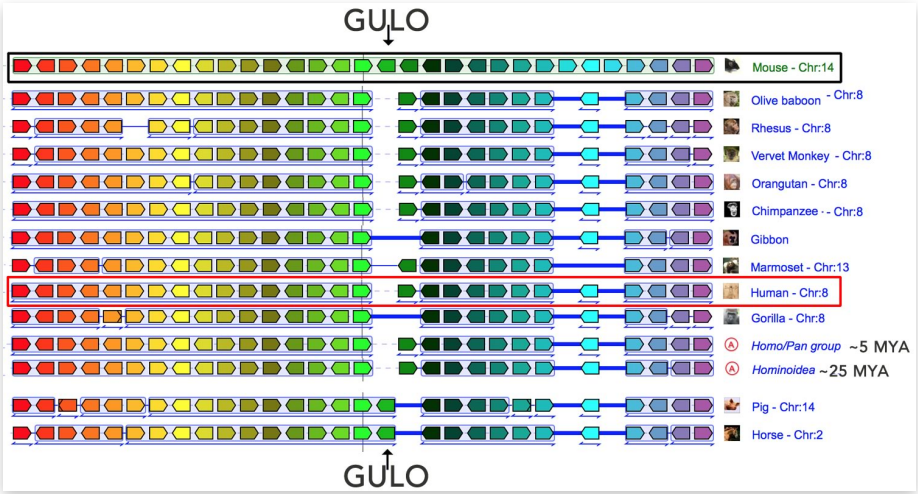
[554] Genomicus update 2015: a genome-wide perspective to multispecies comparative genomics.

[555] Lactose digestion and the evolutionary genetics of lactose persistence

[556] World-wide distributions of lactase persistence alleles and the complex effects of recombination and selection.

chromosome 2), by inputing genes at each end of the region displayed. Genomicus also presents syntenic regions in other organisms, and provides predictions of the genomic organization of evolutionary ancestors.

To use Genomicus to study evolutionary change, let us consider a gene we introduced previously, the *GULO1* gene. Recall that, and in contrast to most vertebrates, the *Haplorhini* or dry nose primates are dependent on the presence of vitamin C (ascorbic acid) in their diets. One plausible scenario for how this situation came to be is that a functional L-gulonolactone oxidase (*GULO1*) gene was lost due to mutation in the last common ancestor of the *Haplorhini*. The remains of the *GULO1* gene found in humans and other *Haplorhini* genomes is non-functional, leading to our requirement for dietary vitamin C. If we use the human genome as a reference, Genomicus fails to find the non-functional *GULO1* gene. In contrast, if we enter *GULO1* using the mouse or a *Strepsirrhini* (wet nose primate) genome, Genomicus finds the gene (↓). Each horizontal line in the diagram represents a segment of a chromosome from a particular species selected, together with predicted phylogenic (evolutionary) relationships based on synteny between species. We find a *GULO1* gene in the mouse together with orthologs in a wide range of eukaryotes, including single-celled eukaryotes such as baker's yeast (which appears to have diverged from other eukaryotes about ~1,500,000,000 years ago). Moreover, we find that the genes surrounding the *GULO1* locus in mammals are also (largely) the same; mammals are estimated to have shared a common ancestor ~184 Mya. The syntenic region around the *GULO1* gene, and the presence of a *GULO1* gene in yeast and other distantly related organisms, suggests that the ability to synthesize vitamin C is a trait present in the ancestor of all eukaryotes.



Humans are eukaryotes, but an examination of the resulting map reveals the absence of humans (*Homo sapiens*) and other Haplorhini primates – Whoa!!! what gives? The explanation, it turns out, is rather simple.[557] There is (apparently) no functional *GULO1* gene in any *Haplorhini* primate. But the *Haplorhini* are related to the rest of the mammals, aren't they? We can test this assumption, and circumvent the absence of a functional *GULO1* gene, by exploiting synteny – when we search for genes in the neighboring region, we find that this region, with the exception of *GULO1*, is present and conserved in the *Haplorhini* (↑). The *Gulo1* syntenic region (without *GULO1*) lies on human chromosome 8 (highlighted by the red box) and similar syntenic regions are found in the homologous chromosomes of other *Haplorhini* primates. Our Genomicus analysis enables us to make a number of readily testable predictions. A newly discovered *Haplorhini* primate would be predicted to share the same syntenic region and to be missing a functional *GULO1* gene, whereas a newly discovered *Strepsirrhini* primate, or any mammal that does not require dietary ascorbic acid, should have a functional *GULO1* gene within this syntenic region. We might also predict that adding a functional *GULO1* gene, for example from a mouse, would make a human cell (or a human) vitamin C independent (perhaps something a future genetic engineer with do).[558] Such an analysis

[557] see Visualizing and teaching evolution through synteny

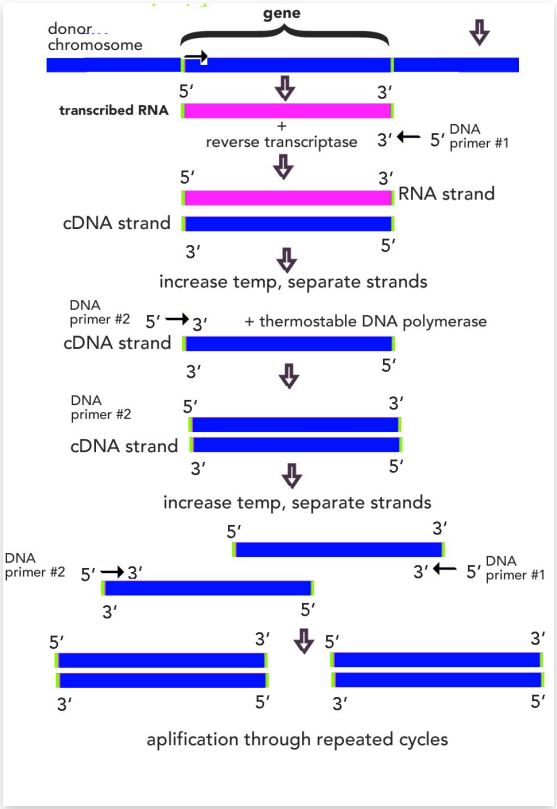
[558] Functional rescue of vitamin C synthesis deficiency in human cells by expression of murine l-gulono-γ-lactone oxidase

also reveals that genes and chromosomal regions can and often do move around within the genome.

Questions to answer:
271. If you were to add a mouse *Gulo1* gene to a a human genome, where would you put it and why?
273. If a gene is missing from a syntenic region, what might have happened to it?
Questions to ponder:
- Would growers citric fruits be right in working to ban the genetic engineering of vitamin C independent people?

Where is a gene expressed?

When we consider the role of a particular gene in generating a particular phenotype, an important question is whether the effect is direct or indirect - is the gene expressed in the cells/ tissues/organ that produces the phenotype or does it influence an earlier event? How, exactly do we know where and when a specific gene is expressed within an organism? There are a number of applicable methods that fall into two basic types - there are those that detect transcribed gene products (RNAs) and those that detect the polypeptide encoded by an RNA. We consider them briefly here.

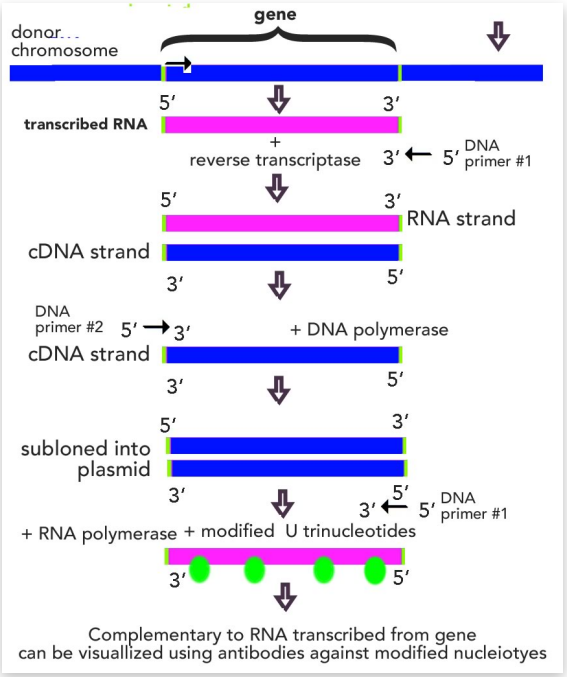


RT-PCR: A transformative technology, made feasible by the discovery of heat stable DNA-dependent, DNA polymerases, isolated from archaea that live in very high temperature environments (thermophiles and hyperthermophiles), polymerase chain reaction (PCR) has been a powerful technique for isolating and manipulating genes, as well as for visualizing gene expression and genome sequencing. In the context of gene expression analysis, we can use PCR to quantify the amount of a particular transcribed (expressed) RNA within a particular tissue, cell type, or together with single cell isolation technology, a single cell (→). The first step in this process involves making a DNA copy of the transcribed RNA - this enables us to avoid the genomic DNA copies of genes which are present in every cell. We isolate RNA from a tissue and then use a "reverse transcriptase" enzyme. The reverse transcriptase (RT) enzyme is derived from viruses and transposable elements that convert RNA into DNA as part of their replication cycle.[559] The RT enzyme uses a DNA primer and makes a DNA copy complementary to the RNA strand, a cDNA. The RNA-DNA strands are then separated (in laboratory by increasing the temperature of the system), and then a second DNA primer acts together with a thermostable DNA-dependent, DNA polymerase to generate a copy of the cDNA, leading to a doubled stranded DNA molecule with primer sequences at each end. Now we begin the amplification stage of the reaction. The two strands are separated by increasing temperature. The original two DNA primers are present in excess, so that when the temperature is reduced, they bind back to the DNA strands, and initiate a new round of DNA-dependent, DNA synthesis. With each cycle the number of DNA strands doubles, so that there is exponential growth in the number of specific DNA molecules with each cycle. Because the primer sequences, which are designed by the investigator and synthesized in vitro, are complementary to, and specific for, a particular gene sequence (the

[559] insert reference to reverse transcriptase.

RNA of interest), one can expect to amplify one and only one of the RNAs (gene products) present in the tissue under analysis. If the gene is not expressed, no amplified DNA will be synthesized. By using various tricks (beyond us here, but relatively simple to employ with the right equipment) the process can be made quantitative, so that it is possible to accurately compare the numbers of different types of RNA molecules (the products of a particular gene) present in the original sample, a measure of the level of gene expression, at least at the RNA level. With different sets of primers, it is possible to quantify the expression of various splice forms of a gene.

More recently, it has become possible to isolate and sequence the RNAs (or rather cDNAs derived and amplified from mRNAs) in a single cell and to then sequence those DNA molecules to characterize the genes expressed in that cell.[560] Because mRNA is used, only exon sequences are (generally) included - and the result is known as an exome sequence. This is a method that can be particularly useful in characterizing the genes expressed in a particular cell type, or in a cancer.[561]



In situ hybridization: A limitation of the RT-PCR approach is that it is generally used on tissue samples, which contain multiple different types of cells. To achieve spatial resolution, we need to use other methods. Perhaps the most common is known as in situ hybridization. When a gene is expressed, an RNA molecule complementary to one strand of the gene is synthesized, and these "sense" RNAs accumulate in the cells that express the gene (there is little evidence for significant transport of RNA from cell to cell, across the plasma membrane.)[562] To identify cells that express a gene, we generate modified "anti-sense" RNA molecules (→). Typically, we first isolate and subclone a DNA molecule that encodes the sense (mRNA) and antisense RNA of a gene's expressed (exonic) region – this can be based on a cDNA generated from an mRNA or a genomic exon. Using specific primers, recognized by different bacteriophage-derived DNA-dependent, RNA polymerases, we can generate either sense or anti-sense RNA molecules. In these reactions modified (with either fluorescein or digoxygenin) forms of the RNA nucleotide UTP are used; this modified nucleotide can be used by the polymerase and is incorporated into the newly synthesized RNA.

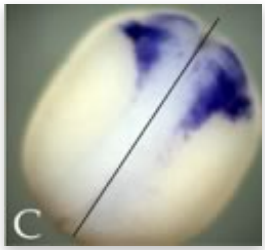
The overall process is relatively simple. The tissue is chemically stabilized and permeabilized (so that molecules can diffuse into and out of it) and then incubated with either sense or anti-sense probe. Because of the complementary nature of nucleic acids, the anti-sense probe RNA will bind to RNA transcripts, generated during gene expression. In contrast, the sense probe is the same sequence as the RNA transcript, and so does not bind - it is used as a control, since (generally) such a sense RNA probe is not complementary to any of the other mRNAs (or other RNAs) present. By controlling the hybridization temperature, we can remove low affinity, non-specific interactions, leaving only the high affinity sense (transcript)-anti-sense complexes. The probe will be retained in regions that express the gene, and washed away from regions where the gene is not expressed (the level of binding to genomic sequence is too low to be visible). Antibodies, conjugated with various enzymes (typically alkaline phosphatase or horseradish peroxidase) can then be used to recognize

---

[560] A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications

[561] see Defining murine organogenesis at single-cell resolution

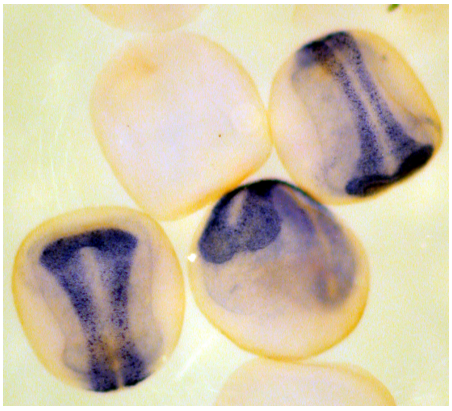
[562] although things may actually be somewhat more complex: see Brain Cells Share Information With Virus-Like Capsules



the modified probe RNA:mRNA complex, and color-generating reactions, catalyzed by the enzymes, allow the distribution of probe to be visualized. The example here (←) is a neurula stage *Xenopus laevis* (clawed frog) embryo in which a gene (Snai2/Slug) expressed in the neural crest has been visualized by in situ hybridization.[563] In situ hybridization can provide single cell resolution, distinguishing cells that do, from those that do not, express a particular gene. The specificity of the technique is influenced by the length of the probe and the hybridization temperatures used.

Single cell RNA Sequencing: The advent of more efficient DNA sequencing methods, together with PCR-based amplification, has made it possible to isolate and sequence the RNA molecules within a single cell. Once sequenced, the number of molecules of each RNA (each gene product) can be counted to provide a catalogue of the genes expressed within a cell. In previously approaches, the genes expressed in a tissue, composed of many different cells, could be identified - but variations between the cells was lost. Single cell RNA sequencing (known as ssRNA SEQ) reveals not just the genes expressed, but (in heterozygotes) whether one or both alleles are expressed. The result is that cells that were once considered identical have been shown to vary in terms of gene expression. These variations can give rise to cell to cell variations that can influence cell behavior and organismic phenotype.

Immunocytochemistry: One limitation of RT-PCR, in situ hybridization, and ssRNA SEQ is that they monitor RNA levels. In cases where the ultimate gene product is a polypeptide, it can be the case that RNA levels are not strictly correlated with the level of the accumulated polypeptide. One approach to avoid this disconnect is to use antibodies, proteins generated by the vertebrate immune system that bind specifically to particular molecular targets. We will ignore how antibodies are generated (since it involves understanding of the immune system, a complex cellular system), but basically antibodies act very much like anti-sense RNA in situ probes, binding to specific molecular (protein) targets. A full characterization of the proteins present in a cell or tissue relies on physicochemical approaches, such as mass spectrometry, to define the proteome (a subject beyond us here).[564] The example here (↑) is a neurula stage *Xenopus laevis* (clawed frog) embryo stained for the transcription factor Sox3.



Questions to answer:

274. How can observed variation in a trait be used to develop a model for the number of genes involved in determining the trait. How might you test your model? (move up! I think)
275. A gene can be spliced various ways - design primer sets to distinguish the splice variants of a gene.
276. Explain why a sense strand RNA probe serves as a useful control for in situ hybridization studies; what does it control for, and why does it work?

Questions to ponder:

- Why might the number of polypeptides in a cell differ from the number of RNAs that encode it?

[563] from: An NF-κB and Slug Regulatory Loop Active in Early Vertebrate Mesoderm

[564] Here is an example of proteomic analysis: Region and cell-type resolved quantitative proteomic map of the human heart

Using web-based bioinformatic tools: gnomAD

When studying a disease that appears to have a genetic component, it is common to identify the causative allele(s) involved. In the case of recessive alleles, such studies often involve pedigree analysis of more or less inbred families. Once a disease-associated allele is identified, it can be important to determine whether that allele is found in individuals who do not display the disease trait. Particularly for dominant alleles, the presence of an allele without the disease phenotype indicates genetic background effects that influence the disease allele's penetrance and expressivity. Over the last decade, there has been in increasing number of human genome or exon sequences. The exome is all of the DNA sequences, the exons, that make it into mature RNA, and even more specifically into mRNA. Most genomic DNA is not transcribed into RNA, which makes generating exomic sequences easier and less expensive - less DNA to sequence.

The accumulating library of exomic sequence data now includes more than 120,000 people from around the globe (and continues to increase and will become more diverse - more non-European people analyzed over time). This data library can be searched using the gnomAD.[565] To search the database, the user (you, for example), inputs a gene's official name, as listed in OMIM or GenBank. gnomAD then displays sequence data from unrelated individuals; this allows for the identification of alleles and mutations present in a range of human populations. Let us try using the gene associated with sickle cell anemia, the *HBB* gene (hemoglobin, beta, OMIM: 141900). Mutations (disease-associated alleles) in HBB have been implicated in a number of human diseases. The allele associated with the sickle cell phenotype involves a missense mutation from GLU to VAL, now known as GLU7VAL (↓). We discover that within the gnomAD database of "normal", that is disease-

Variant	Chrom	Position	Consequence	Filter	Annotation	Flags	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
11:5248232 T / A (rs77121243)	11	5248232	p.Glu7Val	PASS	missense		532	121340	1	0.004384
11:5248233 C / T	11	5248233	p.Glu7Lys	PASS	missense		149	121340	0	0.001228

free individuals, this allele occurs with a frequency of ~0.0044 (with a single homozygous individual identified). The heterozygotic individuals would not be expected to display any overt phenotype under most conditions, while the homozygous individual would be expected to have sickle cell disease. The vast majority of the people with the HBB Glu7Val allele are of African descent, as is the one homozygous individual. When this was originally written (June 2019) there was only one other homozygous individual within the library (Glu122Gln). 71 out of 85 of the people carrying this allele are of African descent, as is the homozygous individual.

Data from gnomAD enables us to make informed guesses as to the impact of various genetic differences on the activity of a gene product.[566] If, for example, a dominant allele has been linked to a disease and yet that allele is detected in the gnomAD database, we might suggest either that that allele is not the cause of the disease, or that the effects of the allele are influenced by variation (alleles) in other genes, leading to reduced penetrance and/or expressivity. If an allele is present in a heterozygous condition, but not a homozygous one, we can tentatively assume that negative selection is acting on the allele. If, on the other hand, alleles are present at different frequencies

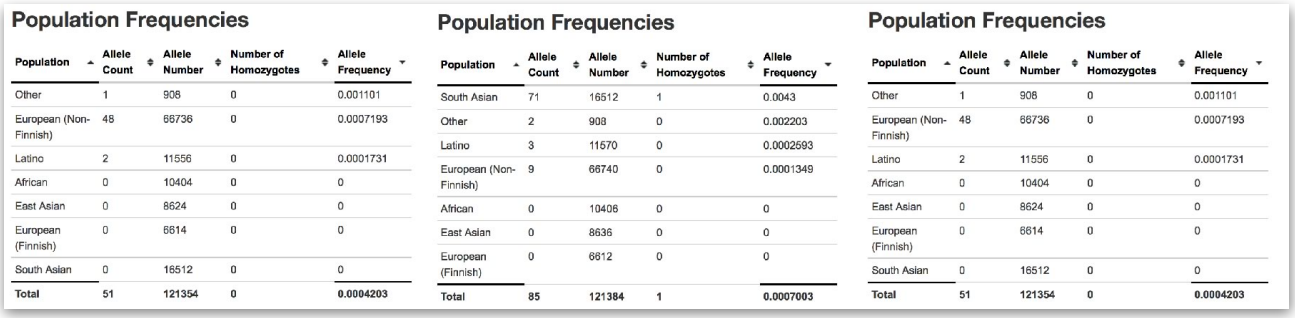
Population Frequencies

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
African	505	10404	1	0.04854
Latino	12	11548	0	0.001039
South Asian	9	16512	0	0.0005451
European (Non-Finnish)	6	66734	0	8.991e-05
East Asian	0	8620	0	0
European (Finnish)	0	6614	0	0
Other	0	908	0	0
Total	**532**	**121340**	**1**	**0.004384**

[565] Genomics, Big Data, and Medicine Seminar Series – Daniel MacArthur

[566] The ExAC browser: displaying reference data information from over 60 000 exomes.

in different populations, that may be evidence for the action of positive selection dependent on environmental factors. In addition, the frequency of alleles in different populations often reflects the effects of founder effects, bottlenecks, and drift. Take for example three other HBB alleles, p.Gly70Ser, p.Glu122Gln, and p.Gln40Ter (Ter=stop)(↓). We see that the Gly70Ser and Glu40Ter



Population Frequencies

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
Other	1	908	0	0.001101
European (Non-Finnish)	48	66736	0	0.0007193
Latino	2	11556	0	0.0001731
African	0	10404	0	0
East Asian	0	8624	0	0
European (Finnish)	0	6614	0	0
South Asian	0	16512	0	0
Total	**51**	**121354**	**0**	**0.0004203**

Population Frequencies

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
South Asian	71	16512	1	0.0043
Other	2	908	0	0.002203
Latino	3	11570	0	0.0002593
European (Non-Finnish)	9	66740	0	0.0001349
African	0	10406	0	0
East Asian	0	8636	0	0
European (Finnish)	0	6612	0	0
Total	**85**	**121384**	**1**	**0.0007003**

Population Frequencies

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
Other	1	908	0	0.001101
European (Non-Finnish)	48	66736	0	0.0007193
Latino	2	11556	0	0.0001731
African	0	10404	0	0
East Asian	0	8624	0	0
European (Finnish)	0	6614	0	0
South Asian	0	16512	0	0
Total	**51**	**121354**	**0**	**0.0004203**

alleles are present primarily in non-Finnish Europeans, while the Glu122Gln allele is found in South Asians. It is not clear exactly what the effects of such missense mutations will be on the functions of the polypeptide – it could change folding, change interactions with other polypeptides and molecules, add or remove sites of post-translational modification, or change catalytic activity, if the polypeptide has such an activity. It is likely that the Glu40Ter mutation will produce a short. likely non-functional 39 amino acid polypeptide (compared to the 147 amino acid long wild type polypeptide). It is unlikely that the truncated protein is functional, but if it accumulates it could interfere with the function or molecular interactions of the full length polypeptide.

Using web-based bioinformatic tools: BLAST

There are other web based tools to identify evolutionarily conserved regions in related gene products. Perhaps the most useful is BLAST. It enables you to take either a nucleotide or a polypeptide sequence and search for similar sequences in all sequenced genes (deposited in GenBack, a central repository). The program returns similar sequences in other organisms. The presence of such sequences can be best explained through either evolutionary relationships (inherited from a common ancestor), horizontal gene transfer, or convergent evolution towards a similar function from different starting points or via different pathways (think wings). The BLAST tool is also useful for identifying those parts of nucleic acid or polypeptide sequences that are conserved, that is, that vary the least from organism to organism – we might well expect such regions to be particularly sensitive to mutational change. The absence of allelic (missense/non-sense) variants (in gnomAD) in such regions would argue for the action of positive selection.

Questions to answer:

277. You find a frequent allele in a population but no individuals homozygous for that allele - how might you make sense of that observation?
278. Why aren't missense mutations necessarily loss of function mutations?
279. Looking at two populations, you find a particular allele to be much more common in one than the other - what processes and historic events could explain such an observation?

Questions to ponder:

- Provide a model for why an individual homozygous for the Glu7Val allele not have sickle cell disease?

Genome-wide Association Studies (GWAS)

The majority of phenotypic traits are not associated with simple Mendelian inheritance, rather a number of different genetic loci (genes) and the combination of alleles present determines the genetic aspect of the trait. In addition, there are non-genetic, that is environmental factors involved. How much nutrition an organism gets when developing, the presence of toxins or absence of vital

nutrients, the effects of pathogens and other stressors and such, combine to influence the final phenotype. A classic example of a trait influenced by both genetics and environment is height, because it is what is known as a quantitative trait – we characterize it by a simple number (although in fact, posture can influence our measurement).[567] The estimates for the heritability of height are not all that accurate and differ between populations, ranging from between ~60 to ~80% of the variation attributed to genetic differences and ~20 to ~40% environmental (nutritional) factors. In addition, height (in humans) is a sexually dimorphic trait - on average males are taller than females.

So how, if many genes are involved, do we identify the genes involved in a particular trait?[568] We begin with a trait that can be accurately measured. In this regard, height is better than friendliness, for example. Then we need a method to identify the various differences found between different organisms (people in this case). Typically between 500,000 to 1 million single nucleotide polymorphisms (SNPs) are used. A useful SNP occurs at high frequency (>10 to 30%) in the population - it does not need to be located within a particular gene, but with a high enough density of SNPs, a some SNP will be near essentially every genes and inherited with the gene (allele). Of course meiotic recombination can influence who is linked to whom.

The different SNPs present in a particular genome are identified based on nucleotide complementarity. Samples of a person's genome are taken, often from white blood cells, which have nuclei and DNA (in contrast to enucleated red blood cells in humans). Since alleles and SNPs differ in their nucleotide sequences, two perfectly complementary (single-stranded) DNA molecules bind more strongly to one another than two mis-matched molecules. We can use this difference in binding stability to identify which SNP or allele is present at a particular position. Finally, we ask how the presence of particular SNPs/alleles relates to the level of the trait, for example the height of the person or the levels of LDL and HDL (low and high density lipoproteins) in their blood. Of course you see some of the issues right away. People are different heights at different times of their lives, and different levels of LDL and HDL depending on their diet, and when they last ate. So the trait we are trying to study has to be accurately and reproducibly measurable.

We then ask which markers (SNPs or alleles) are found in correlation with the trait phenotype (height, LDL/HDL levels, etc.). With a large enough population of people (genotypes and phenotypes) we can identify those markers (alleles and SNPs) that are in or near specific genes that are associated with the phenotype in question. However correlation does not imply (or better put prove) causation. It may be that the allele/SNP is linked on to functionally significant allele. This is one reason that it is important that there has been time (generations) to separate, by meiotic recombination, one allele from another. To prove that a particular allele plays a functionally significant role in producing or modifying a trait, further experimental studies are necessary.[569]

Questions to answer:

280. What is critical before one can even consider beginning a GWAS study?

Questions to ponder:

- You discover a gene linked to a particular trait through a GWAS study, how might you go about establishing a significant physiological role for the gene in influencing that trait?

[567] How much of human height is genetic and how much is due to nutrition?

[568] Chapter 11: Genome-Wide Association Studies

[569] The interplay of common, rare variation in autism

A few conclusions before we move on ...

At this point, you will have completed what is meant to be a two semester introductory course on modern biology. Of course it is limited in scope, primarily because what it aims to teach is important to master confidently. As noted by Oscar Whitney (per. comm.), who served as a learning assistant for the course (awhile ago), the goal of any such course should be to help you build effective and productive intuitions regarding biological systems. That does not mean memorizing large numbers of facts, but rather developing a reasonable feeling for how a system could work. What molecular level processes are likely to be involved. So what comes next? Typically that might be courses in cell and more advanced molecular biology - looking at common mechanisms regulating the behaviors of biological systems. More and more details, but all anchored in the core concepts introduced in biofundamentals. In the next section we consider how these processes are applied in the context of developing systems.

To be added (here or to the appendix): Brief descriptions of forward and reverse genetic screens and the generation of targeted and conditional mutations

Appendix: Fundamental concepts & their application to developing systems.

In which we consider the basic molecular and cellular processes involved in the behavior of groups of cells, including the processes behind the transformation of a single cell, the fertilized egg, into a complex multicellular organism composed of multiple and integrated cell types.

By this point, we have introduced many and perhaps most of the core molecular and cellular ideas that are needed to understand the mechanisms involved in the behavior of biological systems.[570] These are the ideas we will call upon to build plausible models of specific processes. With some modifications, these ideas serve as the basic toolkit that working scientists, and hopefully students, learn to call on to design experiments, interpret observations, and construct models that help test their understanding of the system they are considering. The ability to generate plausible models is a high level skill and involves recognizing and applying appropriate ideas and ignoring irrelevant ideas. In an important way, it enables us to simplify our thinking, we can focus on what is important and avoid distractions. It starts by reflecting on what ideas apply to specific situations; a skill that takes practice and informed feedback - one reason that manuscripts written by the most experienced scientists can benefit dramatically from what is known as peer review and revision. As a focus, we will consider how fundamental ideas are applied to understanding the behaviors of developing systems.[571]

By developing systems we mean a cell or group of cells and how they change over time in response to various perturbations, including their interactions with one another. Cells monitor their external and internal environments, and regulatory networks are critical to maintaining the living state (a process known as homeostasis), adapting to changing condition as well as decisions on whether to grow and divide, to change, or to remain in a steady state. These involve an interconnected network of networks that control cellular machinery, their organization and activity, and associated changes in gene expression - together they produce the emergent behaviors we referred to as the living state. Emergent behaviors are those that "cannot be predicted through analysis at any level simpler than that of the system as a whole".[572]

When we think about emergent behaviors, an obvious example is the development of multicellular organisms. In these systems a fertilized egg, formed by the fusion of haploid gametes, goes on to form a multicellular embryo composed of multiple cell types in specific juxtapositions, and controlled by a combination of intracellular and extracellular asymmetries and cellular responses to those asymmetries. Similar developmental processes are also found among unicellular organisms. Microbes of various types sense their neighborhood, including the number (concentration) of related and unrelated organisms and alter their cooperative and competitive behaviors in various ways through processes, such as quorum sensing, positive and negative feedback interactions, and molecular cascades that produce changes in gene expression and phenotype.[573]

---

[570] adapted from the blog post: on teaching developmental biology from a biofundamentalist perspective

[571] We (almost) completely ignore plants, not because they are unimportant (clearly they are important), but to stay focussed.

[572] review of Dyson, 1997. Darwin among the Machines: The Evolution of Global Intelligence

[573] see: Community behaviors & quorum sensing (page 81)

How do systems change at the molecular level?

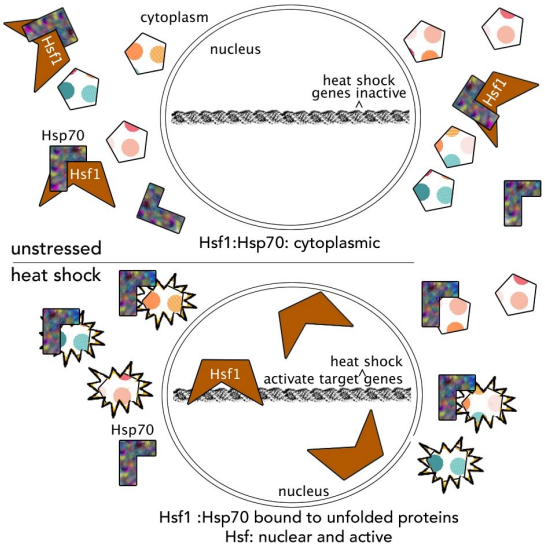
The fundamental biological system is the cell. A cell grows (captures energy from its environment, increases in mass, builds proteins, lipids, and other molecules, replicates its genetic material) and divides producing two cells similar to the original cell. The various molecular processes involved are essential to the living state, they are often referred to as "housekeeping" functions - they underlie every process carried out by the cell. As an example, the mRNA-directed synthesis of polypeptides (translation) is a housekeeping function as is the maintenance of the ionic state of the cellular interior, mediated by ion pumps located in various cellular membranes. At the same time, cells have the ability to respond in different ways to different external and internal factors. As we consider multicellular organisms, various factors combine to produce the patterns of cell division and differentiation that underlie the formation of specific cell types, tissues, organs, and organismic behaviors. How a cell responds to various external and internal signals is a product of the organism's evolutionary history as well as the genes it has been, or is currently expressing, together with the proteins (and other molecules) present. Typically when cells change, when they take on different shapes, express different genes and gene products, and display different behaviors, these changes are driven by the interactions between internal systems and external factors. Again these are emergent behaviors, behaviors that can be explained only at the level of the interacting systems that produce them. This is one reason why the link between a specific allele and a specific phenotypic trait can be complex and difficult to predict.

So how, exactly, do cell's change their behavior? Cells can respond to physical changes in their environment, changes in temperature, the availability of nutrients (food), the presence of toxins or damaging radiation, and such. They can respond to changes in specific signaling or adhesion molecules. Let us start by considering the effects of physical changes that directly effect cellular components. Radiation, such as UV light, can provide the energy to initiate a chemical reaction. This type of process is involved when light exposure leads to the tanning of skin, the conversion of chemicals (as in the synthesis of vitamin D), the capture of energy (photosynthesis), or the generation of mutations. Starvation, the lack of necessary nutrients, can lead to various stress responses associated with the interruption of on-going processes dependent upon (driven by) coupling to thermodynamically favorable reactions. Without ATP and other molecules involved in coupled reactions, the thermodynamically unfavorable reactions associated with maintaining the living state, DNA, RNA, and polypeptide synthesis and DNA repair, as well as most metabolic reactions will cease. Synthesis reactions can stall, and aberrant molecules can accumulate.

A classic example of a physical effector involves changes in temperature and is known as the heat shock response. At the temperatures that a cell normally experiences, many of their proteins are semi-stable, folding and partially unfolding. A class of evolutionary conserved proteins known as chaperones play a key role in allowing unfolded proteins to refold, or to target unfolded or abnormally folded proteins for degradation, removing potentially toxic molecules.[574] When temperature goes up, whether for a bacterial cell or a human being, the extent of unfolding and misfolding of many proteins increases. Because the number of chaperone molecules present in a cell is limited, there will be a competition - proteins normally associated with a chaperone may lose that interaction as chaperones come to interact with the increased numbers of unfolded proteins generated in response to higher temperature. In many systems, there are evolutionarily conserved cellular responses to heat shock (and related stresses) that increase the expression of genes that encode "heat shock proteins", chaperones and other "defense" factors. The transcription factor Hsf1 is constitutively expressed but normally sequestered in the cytoplasm through interactions with the heat shock protein Hsp70. The Hsf1:Hsp70 complex cannot enter the nucleus. In response to a temperature increase there is protein unfolding/misfolding so there is a sudden jump in the concentration of Hsp70 binding proteins, which (following Le Chatelier's principle) will lead to the

[574] Rosenzweig et al., 2019. The Hsp70 Chaperone Network

movement of Hsp70 out of Hsp70:Hsf1 complexes, and an increase in "free" (unbound) Hsf1 (↓). Binding to Hsp70 normally blocks the nuclear import of Hsf1. The increase in free cytoplasmic Hsf1 leads to its transport into the nucleus where it activates the expression of various genes; the expression of these genes further protects the cell from the potentially toxic effects of unfolded proteins. When the system temperature returns to normal, and unfolded proteins are refolded or degraded (broken down to amino acids by various proteolytic systems), the concentration of available Hsp70 increases. As Hsf1 moves into the cytoplasm, it interacts with Hsp70 and again becomes sequestered. Genes, normally dependent upon Hsf1 for their expression "turn off".



Thinking about this process, we recognize a number of common conceptual themes. First, binding interactions are based on molecular structure and the numbers of chaperone molecules present. There will always be a competition between all possible "target" molecules for chaperone binding. Different proteins will differ in the stability of their functional state(s), so changing temperature will change the pattern of chaperone binding proteins and will influence the degree to which various chaperone:target complexes exist. This is a general rule; it also applies to transcription and associated factors and the genes they regulate. The combination of binding site affinity and transcription factor concentration will determine the extent to which specific DNA binding sites are occupied, and will influence the extent to which the genes they regulate are expressed. Changes in molecular shape, such as are associated with unfolding, post-translational modifications, interactions with other proteins, or the binding of allosteric effectors can influence molecular behaviors and properties.

Question to consider: What defines a chaperone target and how can it be recognized?

Steady state and changing molecular concentrations: synthesis and degradation

A key factor involved in the interaction between molecular components of biological systems is the concentration of these components. The concentration of a molecular component is determined dynamically, it is a function of the rates of its synthesis and its degradation, both active (energy-dependent) and regulate-able processes. As an example, the synthesis rate of a gene product is determined by a number of factors - including the number of mRNA molecules synthesized, processed (introns removed, 5' cap and 3' polyA tail added, and transport to the cytoplasm in eukaryotes), the efficiency of their interactions with ribosomes and various associated proteins involved in polypeptide synthesis. The length of the transcribed and translated regions will determine the time it takes to synthesize RNA and polypeptides. Both processes, transcription and translation, can reflect stochastic effects - resulting in what is known as "bursting", namely oscillations between periods when there are multiple RNAs or polypeptides synthesized, and periods when none are.[575] Particularly when time-averaged levels of a gene product are low, bursting (stochastic) expression can lead to functionally significant variations in the concentration of a gene product, one more reason that the behavior of biological systems, particularly at the single cell level, can be difficult to predict. Given the cascade effects that we will discuss further, a transient increase in a protein, particularly if it influences the pattern of gene expression, can lead to long lasting effects on cellular behaviors – such stochastic effects can generate phenotypic variations between cells

[575] What is a transcriptional burst? & Beyond initiation-limited translational bursting

within a homogenous environment.

Turnover/degradation rate: Another key factor controlling intracellular concentration of a specific molecule is the rate of its degradation - in many situations the rate of degradation is often referred to as a molecule's half-life. Unlike the situation with the half-life of a radioactive isotope, in biological systems the degradation rate of a molecule is not intrinsic to the molecule but determined by active (that is, energy-dependent) and regulateable processes. Polypeptides can contain sequences that mark them for rapid degradation by proteolytic enzymes. Alternatively, they can be marked by post-translational modifications, particularly the covalent addition of ubiquitin, a small (76 amino acid long) protein. Degradation is a population behavior, so that the smaller the population size, the greater the statistical fluctuations - the more noise, the more variation. The effect is similar to that seen in genetic drift (allele behavior in populations) and the case of the lac operon in bacteria.[576] When regulator concentrations are low, stochastic variations in regulator concentration leads to noisy gene expression that can generate significant phenotypic variation between genetically identical cells and their progeny.

The concentration of any molecule within a cell, or within a biological system more generally, will reflect both the rates of its synthesis and degradation. At the same time, these rates and their regulation determine the speed at which the system can readjust molecular concentrations in response to changes in external and internal factors. For example, in a system in which the degradation rate of a specific molecule is slow, even if synthesis stops, the molecule will persist for some time.[577] Alternatively, if the degradation rate is rapid, the concentration of the molecule will change quickly in response to changes in the synthesis rate. Of course, changes in the synthesis rate are not immediate, since (in the case of a polypeptide) the times involved are influenced by the length of the transcribed region (the length of the synthesize RNA) and the time involved for polypeptide synthesis. The longer the RNA and the polypeptide (not always correlated), the longer the delay between the signal to increase gene expression and the appearance of newly synthesized polypeptide. In cases where rapid changes in molecular activity are involved, synthesis and degradation rates can stay unchanged, the binding of allosteric effectors or post-translational modifications can act more quickly to alter activity. As an example, the rate of degradation of a stable protein can be quickly accelerated by a post-translational modification, such as the covalent addition of ubiquitin groups.

Direct and indirect cellular responses to signaling molecules

A typical biological signaling system uses both fast acting responses (allosteric effectors and post-translational modifications, including proteolytic processing) and slower acting changes in gene expression (synthesis and degradation rates). Each signaling system can be characterized by common features, these include i) the signal itself - generally molecules synthesized and released by other cells (although in some cases, a cell can signal to itself - a process known as autocrine signaling). ii) A receptor for the signal, generally receptors are proteins synthesized by the responding cell. Finally, iii) the effect(s) that occurs when the signaling molecule interacts with (binds to) the receptor. The heat shock system behaves similarly; the signal is unfolded proteins, the receptor is Hsp70, and the response is the release of Hsf-1, its nuclear localization and its effect(s) on gene expression. Many (most?) cellular signaling systems result in changes to molecular networks and patterns of gene expression. Perturbations that target one cellular system generally also influence gene expression, which in turn can influence the behavior of the targeted system.[578]

---

[576] lac operon: page 191

[577] You may recognize the toxin-antiToxin system associated with programmed cells death, discussed earlier.

[578] an example: Cytoskeletal control of gene expression: depolymerization of microtubules activates NF-kappa B

In biological systems, the behaviors produced and their regulatory dynamics are based on interacting molecules (the products of their evolutionary history). We can begin to model these behaviors by considering the factors that influence them. Interactions between molecules are based on the thermodynamics of surface-surface and surface-solvent interactions. Such surface features, for example, determine the relative binding specificity of transcription factors for specific versus generic DNA sequences. The binding energy will determine the stability of the interaction, that is the average time an interaction, once formed, persists before it is knocked apart by collisions with other molecules, an inherently stochastic process. Low affinity interactions will likely be transient, they will persist for shorter periods of time than higher affinity interactions. The assembly of multi-molecular components can be expected to be more stable than simpler ones. Of course, given their stochastic nature, we can predict the average duration that two interacting molecules will remain bound to one another, but not the duration of any particular interaction. This matters at the cellular level, since there are only two copies of most genes and limited numbers of the other molecules involved. Noise in gene expression associated with low transcription factor levels is to be expected.

Often what were originally thought to be independent molecular interaction networks can themselves interact, producing systems of systems that lead to emergent behaviors. Moreover, various experimental and genetic manipulations perturb the system of multiple ways. For example, the removal of a gene can be expected (naively) to lead to the "simple" absence of gene product, but the effects of a gene's removal may be more complex. For example, if the gene product normally interacts with other gene products, then the behavior of these interacting gene products may be altered, often in unexpected ways. As an example, polypeptides "orphaned" by the absence of their normal interaction partner may interact with molecules they would not usually interact with, disrupting their normal function(s), or they may fail to fold normally and so form toxic aggregates. In part these effects can be modulated by the levels of various chaperones, proteins that can (in some cases) reverse the effects of protein aggregation and misfolding.[579]

Question to consider: How can the position of a mutation in a gene influence the strength and duration of an interaction between two molecules, or populations of molecules? What types of information would help you with your predictions?

Modeling gene expression

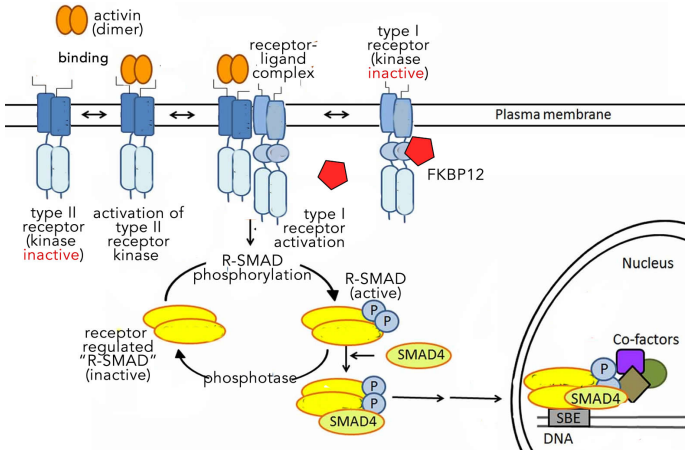
Let us get more specific and consider a model of a gene regulatory system. We will use the model proposed by Saka & Smith[580] to illustrate a number of points (but not to memorize). Their model aims to understand how an extracellular signaling molecule can regulate the mutually exclusive expression of one or another target gene in a system. They consider the case of cellular responses to the secreted signaling molecule activin, a member of the Transforming Growth Factor (TGF) family of proteins. The activin protein is synthesized and secreted by cells during embryonic development in the frog *Xenopus laevis* (and lots of other systems).[581] So what makes a protein, or other type of molecule, a signaling molecule? As noted above, cells contain and express genes that encode polypeptides that assemble into receptors; receptors that bind the signaling molecule, leading to a change in the receptor's three dimensional shape and its catalytic activity and/or its interactions with other molecules, which in turn alters their activities. The signaling molecule is an allosteric effector of the receptor. In the case of the activin system, the receptor is a membrane protein with a protein kinase activity. The receptor is a surface membrane protein; its activin binding site is extracellular while its kinase domain is intracellular. The binding of activin to its (type II)

---

[579] Such behaviors are discussed here: *Filaments & phenotypes: cellular roles and orphan effects associated with mutations in cytoplasmic intermediate filament proteins*.

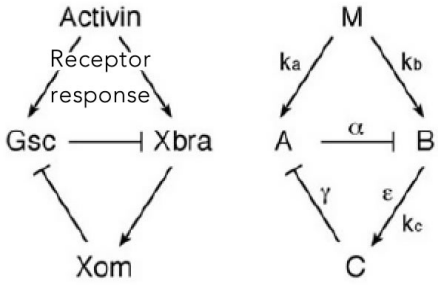
[580] Saka & Smith 2007. A mechanins for the shapr transition of morophogen gradient interpretation in Xenopus

[581] Activin is a member of the TGFb family of signaling molecules. see Chaikuad & Bullock 2016. Structural Basis of Intracellular TGF-b Signaling: Receptors and Smads



receptor leads to the activin:receptor complex's binding to a "type I" co-receptor, another membrane protein with protein kinase activity. (←) In this activin-binding regulated type I/II receptor complex, the type II receptor kinase is active and phosphorylates the co-receptor. This phosphorylation alters the co-receptor's structure leading to i) the dissociation of a cytoplasmic inhibitor (FKBP12) from the co-receptor, ii) the activation of the co-receptor's protein kinase domain, and iii) the phosphorylation of cytoplasmic receptor-regulated SMAD (R-SMAD) proteins. The phosphorylation of the R-SMAD protein changes its shape so that two phosphorylated R-SMAD polypeptides associate with a common "co-SMAD" polypeptide, SMAD4. The SMAD4 polypeptide, normally localized to the cytoplasm (excluded from the nucleus), contains a transcription-activating domain. The R-SMAD:SMAD4 complex is transported from the cytoplasm into the nucleus through nuclear pores. In the nucleus the R-SMAD:SMAD4 complex interacts with specific DNA sequences and associated proteins, and regulates the expression of target genes. There are a number of different R-SMAD proteins; different combinations of R-SMADs in trimeric R-SMAD/SMAD4 complexes lead to the activation of different target genes. In addition, there are other proteins that can interact with the SMAD complex and inhibit its activity, turning it into a transcriptional repressor. At each point along the pathway there are inhibitors that can modulate the effects of extracellular activin: there are activin-binding proteins that block its binding to receptors, proteins that bind to the receptor and block its activation, and cytoplasmic proteins that block R-SMAD phosphorylation. The system is dynamic and, importantly, all of the events associated with activin signaling are reversible - including co-receptor and R-SMAD phosphorylation. R-SMAD dephosphorylation leads to the disassembly of the SMAD complex, the export of SMAD4 from the nucleus, and the inactivation of activin-regulated genes.

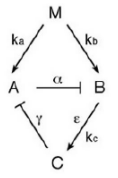
In the Saka and Smith model, the level of activin leads to SMAD regulated expression of two genes, *Gsc* and *Xbra* (↓)[582] – at this point, what these gene names "mean" and where they come from is not important, what is important is that both genes encode sequence specific DNA binding proteins and act as regulators of transcription. In this scenario, both *Gsc* and *Xbra* are directly regulated by the Activin signaling pathway; there are no intervening genes whose transcription and translation are necessary for *Xbra* and *Gsc* gene expression – the system is poised to respond to activin binding to activin receptors. There are, however, downstream effects based on the ability of Gsc to inhibit *Xbra* expression and Xbra's ability to induce *Xom* expression. The product of the *Xom* gene is a transcriptional repressor that inhibits expression of *Gsc*. While *Xbra* and *Gsc* are direct targets of activin signaling, *Xom* is an indirect (downstream) target. Generally, there are a limited number of direct regulatory targets of a signaling system; these act to control a regulatory cascade of downstream targets. In this case, while *Gsc*, *Xbra*, and *Xom* (and the polypeptides that they encode) are the focus of the analysis, it is reasonable to assume that the Gcs, Xbra, and Xom proteins directly regulate, perhaps tens to hundreds, other genes - they might positively regulate some genes, and negatively regulate others, depending upon promoter binding and context. It is their interactions with each other that are the primary determinants of system behaviors.



[582] As a reminder, gene names are in italics while the polypeptides encoded for by a gene is in standard font.

What makes analyzing, predicting, and understanding signaling effects complex is that the regulatory cascade usually includes what are known as feedback interactions. In the activin-system we have three such feedback interactions. First the Gsc and Xbra gene products negatively regulate each other's expression, so that at a high enough concentration of Gsc (for example), *Xbra* expression is inhibited, and visa versa, even in the presence of active activin-based activation. There is also a secondary, indirect negative feedback interaction mediated by the Xom gene product's effect on *Gsc* expression. There can also be negative negative feedback interactions that involve the degradation of receptors or other essential components of the signaling system; these act to turn down or turn off signaling after a period of activation, even if the signal is still present. There are also (but not here) positive feedback interactions, in which a gene product further activates the expression of the gene that encodes it. We will consider what limits such positive feedback loops, and the amount of gene product within a cell shortly

Predicting the behavior of the Activin-Gsc-Xbra system is not simple, we need to generate a quantitative model. We can abstract and generalize the system, replacing protein and gene names with symbols (←). In such a model, many of the molecular mechanisms involved are "collapsed" into more general variables and used to generate systems of (solvable) differential equations (→). These enable us to make predictions as to how the system will behave in response to various perturbations. In this case we characterize the relationship between the strength of the original signal (M), the relative effects on the direct (A and B) and indirect (C) genes, the concentrations of various proteins, and their affinities for their regulatory targets. The variables that apply to the system can take on a number of values, variations in such values reflect the situations in different cells, since cells can vary in terms of the concentration and activity levels of various system components. In addition, while the same activin (M) signaling system directly regulates transcription of the A and B genes, the rate of A and B synthesis can be quite different; for example, differences in the length of the RNA molecule and its coding region, as well as RNA and polypeptide degradation rates, folding and assembly rates (in the case of polypeptides that are part of a multimeric complex) will lead to different time delays for the appearance of functionally significant levels of the various encoded proteins. The functionally significant level of a particular protein will depend on their binding affinities for various target DNA sequences and interaction partners,, and their roles in generating a functional response.

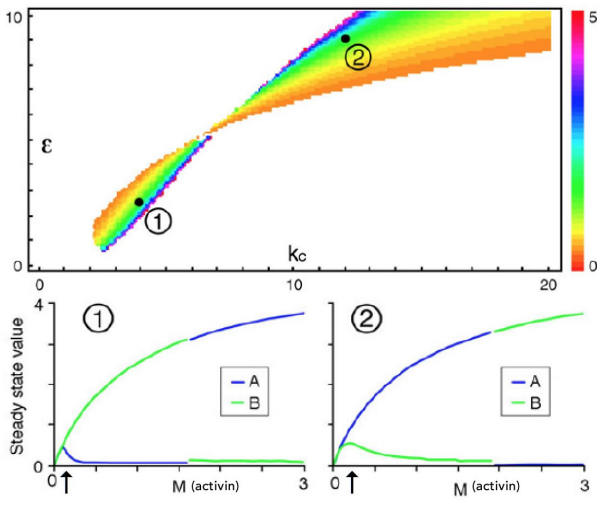


$$\frac{dA}{dt} = \frac{k_a}{1+C^{\gamma}} \cdot \frac{M^{\mu}}{1+M^{\mu}} - kd_a \cdot A$$

$$\frac{dB}{dt} = \frac{k_b}{1+A^{\alpha}} \cdot \frac{M^{\mu}}{1+M^{\mu}} - kd_b \cdot B$$

$$\frac{dC}{dt} = k_c \cdot \frac{B^{\varepsilon}}{1+B^{\varepsilon}} - kd_c \cdot C$$

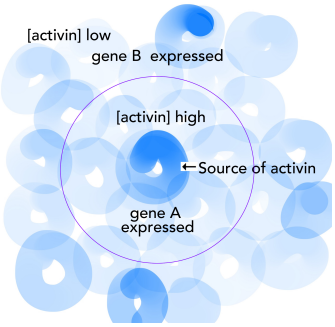
k_a, k_b and k_c are the synthesis rates of A, B and C. α and γ reflect the cooperativities of repression by A and C, ε and μ are the cooperativities of induction by B and M. kd_a, kd_b and kd_c are degradation rates of A,B, and C proteins.



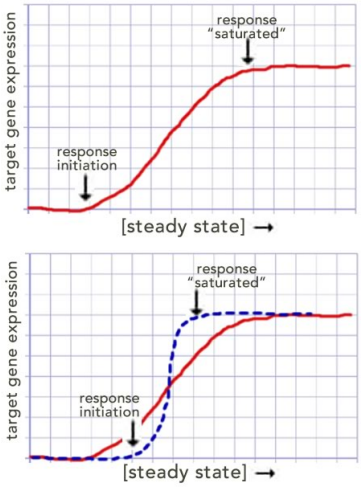
How the system behaves depends on these parameters, which may or may not be easily determined experimentally. Saka and Smith modeled the system's behavior at two parameter positions (marked 1 and 2 in the top graph (←). In both, behavior is similar at low concentrations of activin (bottom graphs). Both gene A and B (*Gsc* and *Xbra*) are expressed at low levels of activin signaling (the "↑s" in the lower panels). Expression behavior changes dramatically as activin concentration increases. In the two domains, expression of one or the other of the target genes increases, while the other drops to near zero. Expression of the active gene continues to increase until activin concentration crosses a threshold, at which point the system

Modified from Saka & Smith: Simulations were performed assuming ka = kb = 5, α = γ = 3, μ = 1, kda = kdb = kdc = 1. Threshold values of M, which are color-coded, are plotted in the parameter plane (kc, ε). kc and ε determine how the value of C changes over time. Consider two sets of variables (top panel). At smaller values of kc and ε (area 1), B is on and A is off at low M, and vice versa with high M (once the system reaches) steady state. With larger values of kc and ε (area 2), A is on and B is off with low M, and vice versa with high M at steady state.

flips, the expression of the previously expressed gene drops to near zero while the expression of the unexpressed gene jumps to high levels. If we were to think of a plane of cells, in which there is a localized source of activin that decreases with diffusion from that source, resulting in an activin concentration gradient, we might predict that, assuming that the cells are similar, that we would see a circular domain of cells expressing gene A surrounded by a domain of cells expressing gene B. The two domains would be separated by a distinct boundary (→). The expression of A or B would be expected to lead to different cellular behaviors, different "downstream" effects.



Another type of threshold effect: Often when the level of signal (or a transcription factor) increases, the effects on the target genes it directly regulates is not a linear one – the relationship between signal and response is not accurately described by a straight line. Generally the dose-response relationship is best described by a sigmoidal curve, a smooth curve with a characteristic shape – it looks like a flattened S (←). Often there is little or no response to low levels of signal, after which there begins a smooth increase, until at higher signal levels, the response flattens again. When response onset and saturation concentrations are close, the response curve looks more like a step function, basically an off-on (or on-off) switch. There can be many reasons for why low levels of a signal fail to activate a response, these can involve the need to assemble a stable multicomponent complex before a response can occur. For example, if the synthesis/activation rate of a necessary response component is a function of signal concentration, while the degradation/inactivation rate is constant, sufficient active activator may only appear above a certain signal concentration, and then increase more or less linearly after that, essentially after the degradation machinery has been saturated (that is, reaching its maximum rate). The saturation level is determined by limiting components. As an example, there are only two copies of a particular gene in a diploid cell, the number of RNA molecules that can be synthesized per unit time is limited by the rate at which RNA polymerase molecules can load onto one or the other of these genes. And, of course, as in the case with the gene regulatory system described above, there can be both positive and negative interactions between components, including positive and negative feedback loops, wherein one component effects its own synthesis, activity, or stability (degradation).



Reversible, irreversible, and cascade effects

A final consideration is whether, when a cell receives a signal, its response is transient - that is, does it return to its original state when the signal is removed or does it adapt to the presence of the signal, for example, through a negative feedback interaction that leads to decreased levels of receptor or critical response components, or does it move on into a new cellular state, characterized by the expression of different genes and different cellular behaviors. As an example, if the signal up-regulates expression of a transcription factor that in turn regulates expression of down-stream transcription factors, that results in altered receptors and regulatory molecules, the cell can become

physiologically and phenotypically different in significant ways. It can become a new cell type. It may well no longer respond to the original signal even if the original signal has been removed for an extended period of time. The first type of response can be considered an adaptive response. The cell responds to a signal, but then "resets" back to its original state. The cell may even adapt (get used to) the presence of one level of signal, and require a higher level to continue to respond. The second type of response can be irreversible, the cell has changed in terms of the genes it expresses, the proteins and molecules it contains. Chromatin organization may be altered, so that the same signaling molecule produces either no response or a different response. This second type of response is common in embryonic development, cells move from an originally totipotent state to an increasingly restricted one. While an early embryonic cell may be induced, in response to a specific combination of signals, to differentiate into a range of different cell types; at a later stage, the same signals may have no significant effect on a differentiated cell. The differentiated state is only irreversible. A neuron, once formed, remains a neuron - it is a terminally differentiated cell type.

A recent technical breakthrough has been the discovery of protocols that can reverse terminal differentiation in some cell types, to reprogram a cell, producing what are known as induced pluripotent stem cells or iPS cells. But these protocols do not work equally well with all differentiated cell types, which is one reason (among many) that the cells that go on to form gametes, the cells of the germ line, are maintained in a distinctive state compared to the cells that go on to form the body, the somatic cells, which are differentiated to various extents. The process of reprogramming a somatic cell is itself associated with stochastic effects, effects that can be best observed through single cell analyzes of gene expression. When a culture of supposedly identical cells are expose to the factors use to generate iPSCs, analysis of individual cells indicates that most cells fail to "reset", and that those that do can different in significant ways from one another.[583]

Questions to answer

- Make (and describe) a model for how a cell move adapt to level of signaling molecule, and then required a higher level of signaling molecule to produce the same response.
- Make (and describe) a model for an irreversible response to a pulse of signaling molecule; what factors will determine the behavior of the system.
- Make (and describe) a model by would a point source of a signaling molecule could produce patterns (such as the "eye spots" in a butterflies wing.

Social interactions between cells

In which we consider the social behaviors displayed by both uni-cellular populations of cells within multicellular organisms.

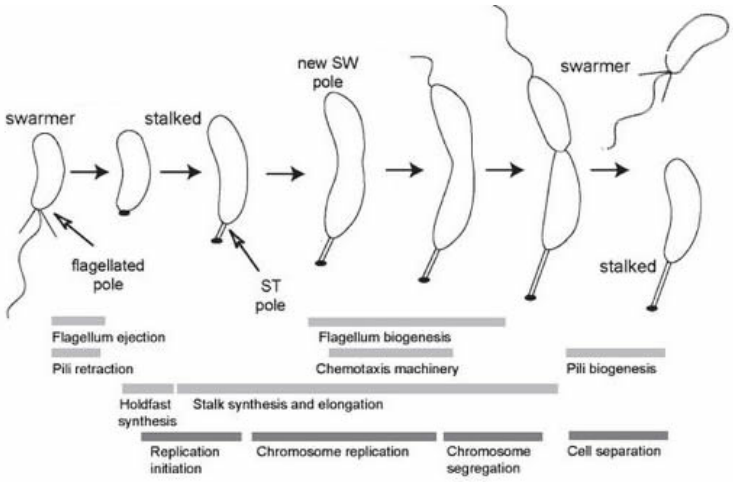
Biology is often presented as a fragmented discipline. There can be multiple biology departments on a single college campus. Yet, underneath the diversity of organisms, systems of organisms (micro and macro ecologies), and idiosyncratic molecular mechanisms, there are evolutionary (family) resemblances that go deep. This is the main reason we can use studies of dramatically diverse organisms to reveal common mechanisms. As the result of evolutionary adaptations, different organisms can display behaviors in an exaggerated form, or can be more accessible (convenient and economical, or both) to scientific studies. At the same time it is important to remember that a molecular/cellular mechanism characterized in one type of organism may be different, often in subtle, but important ways. Mice are not people, and there are mechanistically important differences between even the

[583] see Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming

most closely related species, as well as between individuals of the same species due to genetic variation and life histories. Related (homologous) molecule may play different roles in different species.[584]

An important feature of many organisms are the social aspects of their behavior. How is it that unicellular organisms can cooperate with one another under specific circumstances to generate behaviors that simply would not work if attempted at the single cell level? Based on quorum sensing and the ability to produce multiple phenotypes from a single genotype, these behaviors range from self-sacrifice to the construction of complex molecular machines and communal feeding strategies. A particularly dramatic example occurs when normally unicellular organisms come together and coordinate their behaviors to form what we might term a temporary metazoan. In addition to self-sacrifice, we see examples of cellular differentiation in response to environmental and internal factors. Similar mechanisms are used in a wide range of responses, including those involved in producing a human from a fertilized egg. Network behavior and integration underlie the emergent behaviors of a range of systems, from the immune system to the brain. Now we will go on to consider what we can learn about general processes from studies of specific types of animals (we will largely ignore plants).

How do unicellular organisms generate phenotypic diversity?



In most unicellular organisms, the cell division process is reasonably uneventful, the cells produced are similar to the original cell – but not always. A well studied example is the bacterium *Caulobacter crescentus* (and related species)[→].[585] In cases such as these, the process of growth leads to phenotypically distinct daughters. While it makes no sense to talk about a beginning (given the continuity of life after the appearance of LUCA), we can start with a "swarmer" cell, characterized by the presence of a motile flagellum, a molecular machine[586] driving cellular motility.[587]

A swarmer cell will eventually settle down, lose its flagellum and replace it with a specialized structure, a holdfast, that anchors the cell to a solid substrate. As the organism grows, the holdfast develops a stalk that lifts the cell away from the substrate. As growth continues, the end of the cell opposite the holdfast begins to differentiate – it begins the process leading to the assembly of a new flagellar apparatus. When reproduction (cell growth, DNA replication, and cell division) occurs, a swarmer cell is released; it can swim away and colonize another area or settle nearby. The holdfast-anchored cell continues to grow, producing new swarmers. This process is based on the inherent asymmetry of the system – the holdfast end of the cell is molecularly distinct from the flagellar end. As we will see, this type of behavior is similar to that displayed by what is known as a stem cell in multicellular organisms.

---

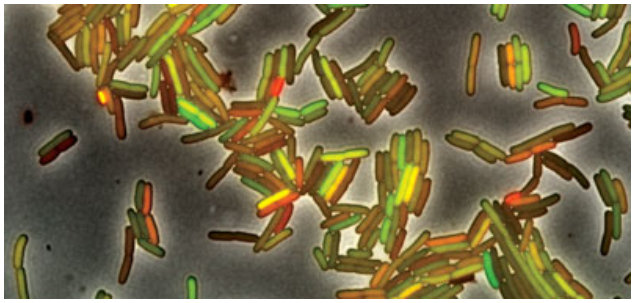
[584] here is an example: Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells

[585] further reading: Caulobacter microbewiki, C. crecentus and Hughes et al 20112. C. crescentus. Current biology.

[586] Molecular machines and the place of physics in the biology curriculum

[587] from Jacobs-Wagner (2004). Regulatory proteins with a sense of direction: cell cycle signaling network in Caulobacter.

The process of swarmer cell formation in *Caulobacter* is an example of what we will term deterministic phenotypic switching. Cells can also exploit molecular level noise (stochastic processes) that influence gene expression to generate phenotypic heterogeneity, different behaviors expressed by genetically identical cells within the same environment. This process enables members of a population to sample phenotypic space.[588] Molecular noise arises from the stochastic nature of molecular movements and the rather small (compared to macroscopic systems) numbers of (most) molecules within a cell.[589] Most cells contain one or two copies of any particular gene, and a small number of molecular sequences involved in their regulation. Which molecules are bound to which regulatory sequences, and for how long, is governed by inter-molecular surface interactions and thermally driven collisions, as well as their physical accessibility, and is inherently noisy. How the chromatin is folded, what other proteins may be bound may influence expression. There are strategies that can suppress but not eliminate such noise.[590] As dramatically illustrated by Elowitz et al (↓) and others, molecular level noise can produce cells with different phenotypes. Similar



processes are active in eukaryotes (including humans), and can lead to the expression of one of the two copies of a gene. If the two alleles at a particular locus are not the same, monoallelic expression can lead to phenotypic differences between different lineages.[591] Recent studies suggest the presence of competitive interactions between such clones.[592] Such stochastic phenotypic heterogeneity between what are genetically identical cells is rarely considered in most biology courses, but is becoming increasingly easy to identify using techniques such as single cell RNA sequencing and is found in essentially all cellular systems.[593] Control of such variation has been reported based on various social / community responses.[594]

The ability to sample different phenotypes can be a valuable trait if an organism's environment is subject to significant changes. As an example, when the environment gets hostile, some bacterial cells transition from a rapidly dividing to a slow or non-dividing state - they are known as "persisters" since they are resistant to antibiotics and other drugs. Some cells in the population can survive until the environment becomes hospitable again.[595] In some cases, cells differentiate to form "spores", which are resistant to killing by dehydration, radiation, and other stresses. If changes in environment are rapid, a population can protect itself by continually having some cells (stochastically) differentiating into spores, while others continue to divide rapidly. Only a few individuals need to survive a catastrophic environmental change to re-establish the population.

---

[588] Elowitz et al 2002. *Stochastic gene expression in a single cell.* Science **297**:1183-6 & Balázsi et al., 2011. Cellular decision making and biological noise: from microbes to mammals. Cell **144**: 910-925.

[589] Fedoroff, N. and W. Fontana 2002. *Small numbers of big molecules.* Science **297**:1129-1131.

[590] Lestas et al., 2010. Fundamental limits on the suppression of molecular fluctuations. Nature **467**:174-178.

[591] Zakharova et al., 2009. Monoallelic gene expression in mammals. Chromosoma, **118**:279-290 & Deng et al., 2014. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science. **343**: 193-196.

[592] Ellis et al., 2019. "Distinct modes of cell competition shape mammalian tissue morphogenesis." Nature **569**: 497.

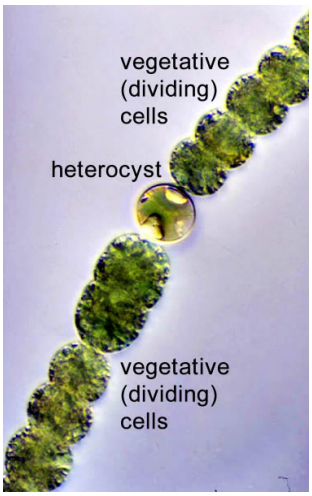
[593] Biology education in the light of single cell/molecule studies

[594] see Cell competition corrects noisy Wnt morphogen gradients to achieve robust patterning in the zebrafish embryo (2019)

[595] Fisher et al., 2017. Persistent bacterial infections and persister cells. Nature Reviews Microbiology **15:**453.

While some of these responses are transient, re-setting quickly as conditions change, others are the result of regulatory cascades, and lead to the establishment of new and stable phenotypes.

Dying for others – social interactions between "unicellular" organisms



Many might conclude that self-sacrificing behaviors are contrary to evolutionary mechanisms, and would be surprised to learn that one bacterial cell (organism) can sacrifice itself (die) to benefit another, but there are a number of examples of this type of self-sacrificing behavior, known as programmed cell death. One interesting example is provided by the cellular specialization decisions associated with photosynthesis or nitrogen fixation in cyanobacteria. These two processes require mutually exclusive cellular environments; specifically molecular oxygen (O_2) released by photosynthesis inhibits the process of nitrogen fixation. Nevertheless, both are required for optimal growth. The solution? Some cells differentiate into what are known as heterocysts (→), cells committed to nitrogen fixation, while most "vegetative" cells continue with photosynthesis. Heterocysts cannot divide, and eventually die – they sacrifice themselves for the benefit of their neighbors, the vegetative cells, cells that can reproduce.

The process by which the death of an individual can release resources that can be used by its neighbors to insure or enhance their survival and reproduction is an inherently social process, and it is subject to control by social mechanisms.[596] Social behaviors can be selected for because the organism's neighbors, the beneficiaries of the self-sacrificial behavior, are likely to be closely (clonally) related to themselves. One result of social behavior, mediated by "inclusive fitness" is, at the population level, an increase in one aspect of evolutionary fitness. This can lead to an increase in the frequency of the genes, alleles, and regulatory networks that produce the behavior.

Such social behaviors can enable a subset of the population to survive various forms of environmental stress (see spore formation above). An obvious environmental stress involves the impact of viral infection. Recall that viruses are completely dependent upon the metabolic machinery of infected cells to replicate. While there are a number of viral reproductive strategies, a common one is bacterial lysis – the virus replicates explosively, kills the infected cell leading to the release of virus into the environment to infect others. But, what if the infected cell kills itself BEFORE the virus replicates – the dying (self-sacrificing, altruistic) cell "kills" the virus (although viruses are not really alive) and stops the spread of the infection.[597] Often such genetically programmed cell death responses are based on a simple two-part system, involving a long lived toxin and a short-lived anti-toxin. When the cell is stressed, for example early during viral infection, protein synthesis rates fall leading to a reduction in the level of the anti-toxin, the activation of the toxin, and cell death.

Quorum effects

Some types of behaviors only make sense when the density of organisms rises above a certain critical level. For example, it makes no sense evolutionarily (or practically) for a single *Anabaena* cell to differentiate into a heterocyst (see above) if there are no vegetative cells nearby. Similarly, there are processes in which a behavior of a single bacterial cell, such as the synthesis and secretion of a specific enzyme, a specific import or export machine, or the construction of a

[596] In an age of rampant narcissism and social cheating – the importance of teaching social evolutionary mechanisms

[597] One can imagine a similar process in the context of COVID-19. If an infected individual self-isolates themselves (a sacrificial behavior for most people) until their immune system eliminates the virus, they effective kill the virus and spare others from infection

complex, such as a DNA uptake machine (discussed earlier), makes no sense in isolation – the secreted molecule will just diffuse away, and so be ineffective, the molecule to be imported (e.g. lactose) or exported (an antibiotic) may not be present, or there may be no free DNA to import.[598] As the concentration (organisms per volume) of bacteria increases, however, these behaviors begin to make biological sense – there is DNA to eat or incorporate and the concentration of secreted enzyme can be high enough to degrade the target molecules (so they are inactivated or can be imported as food).

How exactly does a bacterium determine whether it has neighbors or whether it wants to join a community of similar organisms? After all, it does not have eyes to see. The process used is known as quorum sensing, a process that relies on threshold (non-linear) responses to signaling systems. Each individual synthesizes and secretes a signaling molecule and a receptor protein whose activity is regulated by the binding of the signaling molecule. Species specificity in signaling molecules and receptors insures that organisms of the same kind are "talking to one another" and not to other, types of organisms that may be present in the environment. At low signaling molecule concentrations, below the activation point, such as those produced by a small number of bacteria in isolation, the receptor is not activated, and the cell's behavior remains unchanged. However, as the concentration of bacteria increases, the concentration of the signal increases, leading to the activation of the receptor. Such a "threshold" effect, no response to the presence of the signaling molecule below a set concentration, and essentially a full response above it, is similar to behaviors observed in a number of developmental systems. It may involve the assembly of active receptors, or various feed back interactions. Activation of the receptor can have a number of effects, including increased synthesis of the signal (a positive feedback effect) and other changes. In unicellular organisms, it can lead to expression of genes involved in various behaviors, including directed movement, aggregation and differentiation. In multicellular organisms, it can lead to the formation of different cell types and different cellular behaviors at different signal molecule concentrations.

In addition to driving the synthesis of a common good (such as a useful extracellular enzyme), social interactions can control processes such as programmed cell death. When the concentration of related neighbors is high, the programmed death of an individual can be beneficial, it can lead to the release of nutrients, common goods, including DNA molecules, that can be used by neighbors (relatives).[599] A quorum regulated increase in the probability of cell death can enhance survival of relatives, and so be selected through inclusive fitness. On the other hand, if there are few related individuals in the neighborhood, programmed cell death "wastes" these resources, and so is likely to be suppressed.

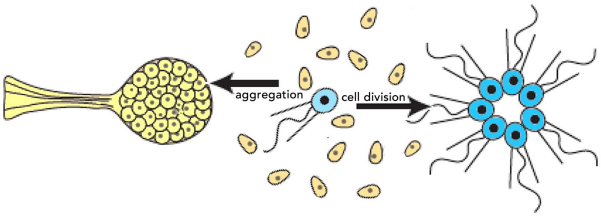
Of course, as in any social system, such "altruistic" (self-sacrificing and cooperative) behaviors are vulnerable to cheaters. A cheater might avoid programmed cell death (for example due to a mutation that inactivates the cell killing system) and could come to take over the population over time. On the other hand, if such cheaters take over, the population will be less likely to survive the types of hostile environmental events that the social (altruistic) behavior was evolve to address. In response to the realities of cheating, social organisms have evolved various strategies that enforce the commitment to social cooperation.

Questions to answer:
How might cheaters be recognized by non-cheaters? What other ways might a cheater cheat?
Describe a situation in which the ability to produce multiple phenotypes from a single genotype is beneficial.

[598] page 208

[599] Durand & Ramsey, 2018. The Nature of Programmed Cell Death. Biological Theory, 1-12.

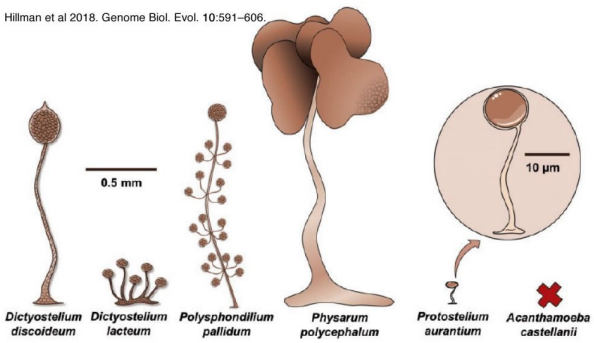
Transient and clonal ("true") metazoans



Although we often think about developmental processes as restricted to multicellular organisms, there are versions that involve organisms that exist in both unicellular and multicellular forms - the multicellular form is transient. Because they are simpler, we can learn important and relevant lessons from these transient metazoans.[600] Forming a transient multicellular organism requires that single celled organisms cooperate with one another, they get social.

The ability of individuals to cooperate, through processes such as quorum sensing, enables them to tune their responses so that they are appropriate and useful. Social interactions also make it possible for them to produce behaviors impossible for isolated individuals. Once individual organisms develop, evolutionarily, the ability to cooperate, new opportunities and challenges (cheaters) emerge. There are strategies that can enable an organism to adapt to a wider range of environments, or to become highly specialized to a specific environment, through the production of increasingly complex behaviors. Many cooperative strategies can be adopted by single celled organisms, but others require a level of multicellularity. Multicellularity can be transient – a pragmatic response to specific conditions, or it can be (if we ignore the short time that gametes exist as single cells) permanent, allowing the organism to develop the range of specialized cell types needed to build large, macroscopic organisms with complex and coordinated behaviors. We can divide multicellularity into two distinct types, aggregative and clonal. These appear to have arisen independently in a number of lineages .[601]

Transient multicellularity: Quorum and environmental/internal sensing systems enable single celled organisms to monitor the density of related organisms in their environment, as well as the supply of nutrients, and to turn on or off specific sets of genes necessary to produce specific and complex cooperative behaviors. The classic example is the cellular slime mold *Dicty*ostelium *discoideum*.[602] Under the appropriate conditions such signaling systems provoke the directional migration of single celled amoeba to associate and form multicellular aggregates that coordinate their behavior to form transient multicellular “slugs" that can migrate and undergo a process of differentiation, forming multiple cell types. Such behaviors have been observed in a range of normally unicellular organisms (→).[603] Under normal conditions, these unicellular amoeboid eukaryotes migrate, eating bacteria and such. In this state, the range of an individual’s movement is restricted to short distances. However when conditions turn hostile (or perhaps better put, unsupportive), specifically due to a lack of necessary nitrogen compounds, there is a compelling reason to abandon one environment and migrate to another, a journey impossible for a single-celled organism. This is a behavior that depends upon the presence of a sufficient density (cells/unit volume) of cells that enables those cells to: 1) recognize one



[600] We will be restricting our considerations to animals, so metazoans makes sense. Behavioral systems in multicellular plants (metaphyta) are beyond us.

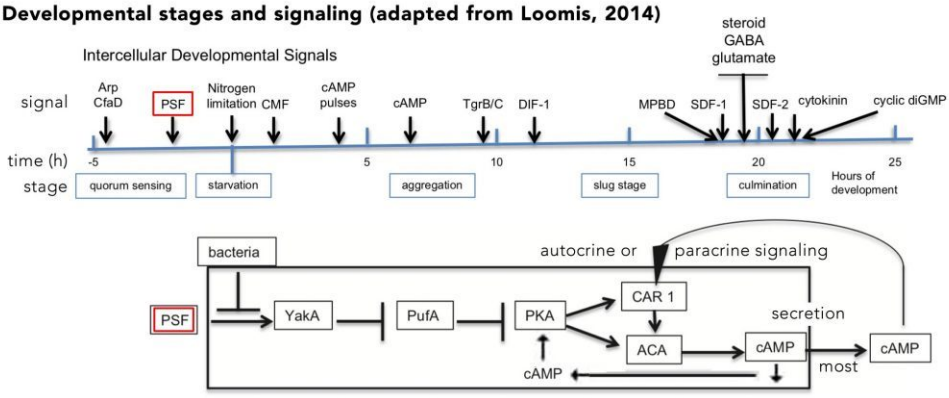
[601] Bonner. 1998. The origins of multicellularity and Knoll. 2011. The multiple origins of complex multicellularity.

[602] Loomis. 2014. Cell signaling during development of Dictyostelium.

[603] Hillmann et al., 2018. Multiple roots of fruiting body formation in Amoebozoa.

another's presence (through quorum sensing), 2) find each other through directed (chemotactic) migration, and 3) form a multicellular slug that goes on to differentiate. Upon differentiation about 20% of the cells differentiate (and die), the process of differentiation produces a stalk that lifts the other ~80% of the cells into the air. These non-stalk cells (the survivors) differentiate into spore cells that are resistant to drying out and essentially inert. The spore cells are released into the air where they can be carried to new locations, establishing new populations.

The process of cellular differentiation in *D. discoideum* has been worked out in molecular detail and involves two distinct signaling systems: the secreted pre-starvation factor (PSF) protein and cyclic AMP (cAMP)(↓). PSF is a quorum signaling the inactivation of PufA and increased PKA activity. Active PKA induces the synthesis of two downstream proteins, adenylate cyclase (ACA) and the cAMP receptor (CAR1). ACA catalyzes cAMP synthesis, much of which is secreted from the cell as a signaling molecule. The membrane-bound CAR1 protein acts as a receptor for autocrine (on the cAMP secreting cell) and paracrine (on neighboring cells) signaling. The binding of cAMP to CAR1 leads to further activation of PKA, further increasing cAMP synthesis and secretion – a positive feed-back loop. As cAMP levels increase, downstream genes are activated (and inhibited) leading cells to migrate toward, and to adhere to on another to form a slug. Once the slug forms it begins to migrate to an appropriate site; the processes of cellular differentiation, morphogenesis, and death lead to stalk and spore formation. The fates of the aggregated cells are determined stochastically. Social cheaters can arise. Mutations can lead to individuals that avoid becoming stalk cells. In the long run, if all individuals became cheaters, it would be impossible to form a stalk, so the purpose of social cooperation (to form a structure that disperses spores) would fail. In the face of environmental variation, populations invaded by cheaters are more likely to become extinct. The various defenses against cheaters are best left to other, more advanced courses.[604]



Growing amoebae constituatively synthesize and secret the protein PSF; the extracellular concentration of PSF reflects cell density. Above a threshold [PSF] level, PSF activates the protein kinase YakA. YakA activation is inhibited by the presence of bacteria (that is , food). When active, YakA inhibits the inhibitor PufA, which inhibits protein kinase PKA. PKA activity is dependent on cylic AMP (cAMP). Activation of PKA leads to increase expression of adenyla cyclase (ACA) with catalyzes the synthesis of cAMP. Most cAMP is secreted and can bind to cell surface cAMP (CAR1) receptors on the secreting cell (autocrine signaling) or neighboring cells (paracrine signaling). Activated CAR1 activates ACA leading to increased cAMP levels that, in turn, lead to cell migration, slug formation, and differentiation.

Evolutionary origins of clonal (permanent) multicellularity

An interesting aspect of the unicellular-multicellular-unicellular behaviors of social slime molds, is that evolutionary selection acts on both stages, the uni- and the multi-cellular. A major evolutionary transition, leading to the appearance of permanently multicellular plants, animals, and fungi is estimated to have occurred some time in the Cryogenian period (834–780 Ma).[605] Exactly how this transition occurred, on how many occasions, and exactly why remains unclear - presumably it involved selection for organisms that could exploit a new range of ecological niches.

---

[604] Strassmann et al., 2000. Altruism and social cheating in the social amoeba Dictyostelium discoideum.

[605] Snowball Earth climate dynamics and Cryogenian geology-geobiology and Uncertainty in the Timing of Origin of Animalsand the Limits of Precision in Molecular Timescales

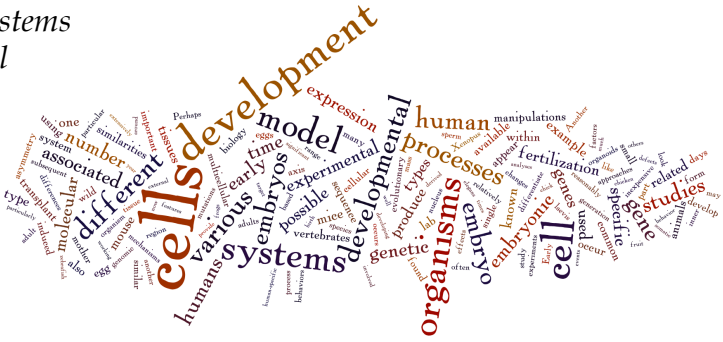
Aggregative multicellularity involves an extension of quorum sensing and social cooperation between genetically distinct, but related individuals. We can speculate on the drivers of true multicellularity, in which all cells of the organism are clonally related and one generation is related to the next through meiosis, gamete formation, and fusion.[606] Of imaginable adaptive (evolutionary) drivers two spring to mind: a way to avoid or discourage predators by getting bigger and as a way to produce varied structures needed to more efficiently exploit ecological niches and life styles, opportunities not available or as efficiently exploited by unicellular organisms.

An example of the first type of driver of multicellularity is offered by the studies of Boraas et al; they cultured the unicellular green alga *Chlorella vulgaris*, together with a unicellular predator, the phagotrophic flagellated protist *Ochromonas vallescia*.[607] After less than 100 generations (cell divisions), they observed the appearance of multicellular, and presumably inedible (or at least less easily edible), forms of *Chlorella*. Once selected, this trait appeared to be stable, such that "colonies retained the eight-celled form indefinitely in continuous culture". The genetic basis for this multicellularity remains to be determined.

Another critical step in the evolution of true multicellularity involves the appearance of specialized cells involved in reproduction. Another evolution of multicellularity experiment that seems relevant are the studies by Ratcliff et al; they selected yeast that failed to separate after mitosis – the resulting clumps of cells were selected because they fell to the bottom of culture tubes, forming multicellular aggregates known as "snowflakes". Within each snowflake cluster, the cells were linked mother to daughter in chains. A snowflake would "divide" into two when one of the cells in a chain died by programmed cell death (apoptosis).[608] This type of division is quite distinct from the process of sexual reproduction, involving meiosis followed by gamete formation and fusion to generate a genetically distinct individual.

The role of model systems in studying metazoan development

In which we consider some common model systems (organisms) use in the study of developmental processes, their strengths and limitations.



A range of inter-cellular communication and cooperative mechanism exist at the unicellular level and can lead to complex multicellular behaviors. Now, we move to explore how animals (multicellular metazoans) develop. Similar processes are involved, although the outcomes are generally more complex, brains versus fruiting bodies. Also because of the evolutionary factors and times involved, each system displays its own idiosyncratic details. The scientific discipline of developmental biology has its roots in the comparative study of embryos, but embryology as a discipline is beyond the scope of any introductory course in developmental biology, in part because there are 35 (assuming no more are discovered) different "types" (phyla) of animals – nicely illustrated at this website [BBC: 35 types of animals, most of whom are really weird]. What has transformed classical embryology into developmental biology has been the increased availability of genomic sequence data from more

---

[606] J.T. Bonner 1998. The Origins of Multicellularity

[607] Boraas et al., 1998. Phagotrophy by a flagellate selects for colonial prey: a possible origin of multicellularity see also Herron et al., 2019. De novo origins of multicellularity in response to predation

[608] Ratcliff et al 2015 Origins of multicellular evolvability in snowflake yeast & Pentz et al., 2016. Apoptosis in snowflake yeast: novel trait, or side effect of toxic waste?

and more different organisms and the more detailed understanding of cellular and molecular processes, which has transformed embryology into the field of evo-devo.[609] But for many biologists the principle driver of studying developing systems is to gain a better (practical and working) understanding of the origins of human birth defects and pathogenic processes. While humans are connected to the rest of the tree of life, and specifically mammals, we are a (now) a distinct species, derived from a population that separated from other mammals sometime around 6,000,000 years ago. In response to the various evolutionary pressures and events associated with this speciation event and subsequent human evolution there have been a number of functionally significant molecular changes.[610] Some lead to therapeutically significant differences in the response of humans to treatments that have proven effective in other organisms, such as the mouse.[611] At the same time, experimentation with humans is constrained by a number of considerations, both ethical and practical. These constraints are clearly appropriate and necessary given the depressing history of medical atrocities. To circumvent these limitations, at least in the early and exploratory stages of biomedical research, it is common to turn to model systems. So what do we mean by a model system and what have we learned from studying such model systems about development in general, and human development in particular?

Model Systems: As our focus is on human development, we consider developmental processes in animals (and ignore plants). "All members of Animalia are multicellular, and all are heterotrophs, that is, they rely directly or indirectly on other organisms for their nourishment). Most ingest food and digest it in an internal cavity."[612] From a macroscopic perspective, most animals have (or had at one time during their development) an axis of asymmetry. This asymmetry may pre-exist within the unfertilized egg or it may appear in response to external factors, such as sperm entry or early events in development. This axis of asymmetry underlies the development of embryonic axes: anterior-head to posterior-tail (or oral-aboral). Animals that can crawl, swim, walk, or fly typically have a dorsal-ventral (back to belly) axis and a left-right axis. When seeking model organisms that can be studied profitably in terms of insights into developmental processes also found in humans, we look for some common features. First we need to be able to cultivate the organism in captivity (in the lab). We would also like organisms that are small and can be fed non-esoteric foods, that maintaining individuals and colonies is reasonably inexpensive. A rapid replication time would also be desirable, we would like to get experiments done in a timely manner. At the same time we would like the stages of early development to be experimentally accessible - external fertilization is one example, in which development occurs outside the mother. Processes that occur within the mother are more technically challenging. At the same time, we might want to avoid organisms that display unique behaviors. An example would be the nematode *Ascaris suum*, in which ~13% of the genome is discarded in somatic cells. While this process may be of interest, since it occurs in a human parasite, it is unlikely to provide direct insights into processes associated with human development.[613]

On the other hand, there are deep molecular level similarities between organisms that appear to be completely different. Perhaps the most dramatic is the HOX cluster(s) of genes associated with anterior-posterior and proximal-distal (in limbs) axes specification. These genes encode DNA binding regulators of gene expression and their genomic organization and patterns of

---

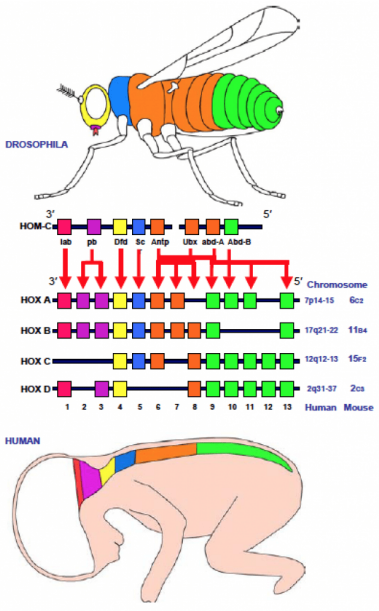
[609] see Arthur, W. (2002) The emerging conceptual framework of evolutionary developmental biology and Wilson, E.B. (1940) The cell in development and heredity.

[610] Guo et al 2020. Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells

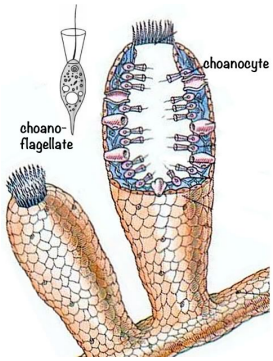
[611] NYTs 2013. Mice Fall Short as Test Subjects for Some of Humans' Deadly Ills

[612] Phil Myers. Animals

[613] The Occurrence, Role and Evolution of Chromatin Diminution in Nematodes and Silencing of Germline-Expressed Genes by DNA Elimination in Somatic Cells



expression are similar throughout the metazoans, from fruit flies to mice and humans (→).[614] It should be noted, however, Hox gene organization is often presented in textbooks in a distorted manner.[615] The Hox clusters of vertebrates are compact, but they are split, disorganized, and even "atomized" in other types of organisms - another illustration of how what might seem to be the most conserved features of organisms can, through evolutionary processes, be altered.[616] Such molecular similarities extend to cell-cell and cell-matrix adhesion systems and the systems that release and respond to various signaling molecules, controlling cell behavior and gene expression. These similarities reflect the evolutionary conservation and the common ancestry of all animals.[617] Differences often reflect adaptions.



Where do these similarities come from? Presumably they were present in the common ancestor of all metazoans. Early in the history of comparative cellular anatomy, it was noted that there are striking structural similarities between the feeding system of choanoflagellate protozoans, a motile (microtubule-based) flagellum surrounded by a "collar" of microfilament-based microvilli, and a structurally similar organelle (←) found in choanocytes, cells present in a number of multicellular organisms, such as sponges. The implication is that the Choanozoan ancestor was predisposed to exploit some of the evolutionary opportunities offered by clonal multicellularity. These pre-existing affordances, together with newly arising genes and proteins were exploited in multiple lineages in the generation of multicellular organisms.[618]

Model Systems

Here we briefly consider a number of the most commonly used model organisms, focussing in particular on what types of experimental analyses and developmental processes they are best suited for. While developmental processes have been studied in many organisms, over time scientists have narrowed their attention to just a few. These range throughout the animal kingdom, and generally have been chosen based on a few practical considerations.[619] Perhaps the most important is the availability of embryos throughout the year, experiments can be carried out as the are imagined by researchers. Since one experiment is inspired or necessitated by results and observations from the last, it is important not to have to wait until next year to do the follow on experiment. At the same time, the maintenance of organisms in the lab needs to be reasonably inexpensive. This tends to favor smaller organisms that can be housed in compact quarters. Other factors that influence choice of experimental organisms are the ease of their experimental manipulation; clearly such manipulations are easier when fertilization and subsequent development occur outside of the mother. The ease with which organisms survive and heal from surgical

---

[614] Figure from Lappin et al, 2006. HOX genes: seductive science, mysterious mechanisms.

[615] Duboule 2007. The rise and fall of Hox gene clusters.

[616] Similar to the limited repurposing of codons in some organisms (link?)

[617] Brunet & King. 2017. The origin of animal multicellularity and cell differentiation.

[618] Long et al., 2013. New gene evolution: little did we know.

[619] Hopwood 2019. Inclusion and exclusion in the history of developmental biology

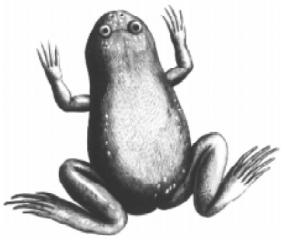
manipulations can also be a factor. As we will see, different model systems offer specific benefits for answering questions about specific processes.

Early on the experimental manipulations available to researchers were limited. Regions of a developing embryo could be moved or removed. Alternatively, one could generate, select, and analyze mutations that influenced developmental processes. More recently, a much wider range of molecular interventions have become available. Embryonic cells can be injected with various inert dyes that can be used to trace cellular lineages (what types of cells a particular cell in the early embryo differentiates into). Molecular biology tools make it possible to construct plasmids that encode RNAs that encode wild type or mutant gene products; chimeric polypeptides that contain regions derived from fluorescent proteins can be used to visualize the intracellular localization of the encoded polypeptides. DNA-based promotor reporters that reveal where different signaling systems are active. Monoclonal antibodies can be injected into cells, where they bind to and disrupt intracellular protein function(s). The expression of gene products can be suppressed by reagents that block the translation of mRNAs (morpholinos) or act to destabilize or block the translation of target mRNAs (based on microRNAs). Most recently CRISPR CAS9-based approaches have been developed that can mutate target genes in various ways.

An equally important aspect of experimental studies involves the techniques available to analyze the effects of various manipulations on developmental processes. Early on, analyses were primarily based on microscopy-based examinations, often associated with the preparation of thin sections of the organism or tissue. Such sections could be stained with dyes to reveal various subcellular components, such as the nucleus, the nucleoli, or connective tissues. Over the last few decades the tools available for analyzing experimental effects and mutant phenotypes have grown dramatically more sensitive and sophisticated. Microscopy, together with various fluorescent reagents has been extended to three dimensions and higher resolution using whole-mount confocal and light sheet microscopy. Single cells and subcellular organelles and their normal or abnormal morphologies can be characterized. Similarly, it is now possible to dissociate embryos or tissues into single cells and to sequence the mRNAs present (single cell RNA SEQ) providing a read-out of the genes expressed as well as the variation between superficially similar cells. Analogous methods exist (affinity-isolation and mass spectrometry-based proteomics) to examine the polypeptides present in a cell, as well as their interaction partners.

The following is meant to be but a short introduction to key model systems.

Frogs & fish



As a model system, the frog *Xenopus laevis* has a number of advantages, and some limitations.[620] Adults are remarkably disease resistant with a wholly aquatic lifestyle. Its lifecycle (from fertilization to sexual maturity) is relatively short, and that of the related species *X. tropicalis* is even shorter. It can be induced to lay eggs and produce functional sperm year round through the injection of commercially available hormones into females. Fertilization and subsequent development occur externally and rapidly, resulting in swimming tadpole within a day or so. A single female produces a large number (hundreds) of eggs of a size that make injection of individual blastomeres (up the the 16-32 cell stage) and microsurgical manipulations possible with limited training.

Xenopus and other frogs have been particularly useful in identifying and in some cases resolving a number of key questions about developmental mechanisms. For example, studies in frog

[620] Gurdon & Hopwood. 2000. The introduction of Xenopus laevis into developmental biology: of empire, pregnancy testing and ribosomal genes

embryos identified the "organizer", a region of the early embryo that acts to induce the formation of the embryonic anterior-posterior axis. Nuclear transplant experiments in *Xenopus* were used to illustrate that genetic information is (generally) not lost during development, an observation that laid the groundwork for somatic cell reprogramming (the generation of induced pluripotent stem (iPS) cells. Nuclear transplant experiments were facilitated by the identification of a dominant mutation in the gene encoding rRNA (0-nu); heterozygous 0-nu cells have a single nucleolus (the site of ribosomal gene expression), whereas wild type cells have two. Transplanting a one-nucleoli nucleus into a wild type cell enabled experimenters to confirm that the transplanted nucleus was driving development. Finally, because early development is supported by maternal components, isolated cells continue to grow and behave. The surrounding vitelline membrane / fertilization envelop of the early embryo can be easily removed, making microsurgical approaches (often using eyebrow hairs as scalpels) reasonably straightforward with a little practice. Various types of embryonic explants have been used extensively to study cellular behaviors, morphogenic movements, and inductive interactions that drive developmental processes.

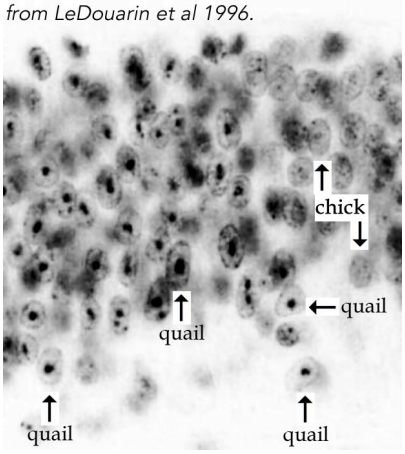
A type of analysis that is rare in *Xenopus laevis* are genetic studies. While there are experimental approaches to the manipulation gene expression, these are one off, involving the manipulation of a single embryo. In part this is because the generation time of *X. laevis* is much longer than that of most of the organisms used for genetic studies, and in part because *X. laevis* is effectively tetraploid. There has been some interest in genetic studies using the related species *X. tropicalis*, which reaches sexual maturity faster and is diploid.

A vertebrate that has been used extensively for genetic studies is the zebrafish, *Danio rerio*. As with frogs, fertilization and embryonic development are external, and so experimentally accessible. Moreover, unlike frogs eggs and early embryos, which are pigmented and opaque, zebrafish embryos are nearly transparent, so high resolution optical microscopy is possible. Zebrafish are easy and (relatively) inexpensive to maintain in the lab, which facilitates classical mutagenesis and analysis, although with the advent of genome sequence data and directed (CRISPR-CAS9 mediated) mutagenesis the process has become increasingly efficient. It is now reasonably straightforward to "knock-in" various alleles, for example alleles associated with diseases in humans, and examine the effects of related processes in the fish. There are companies that will edit genomes in various ways for you![621]

Chick and Quail

Another classic system in which to vertebrate development has been studied extensively is the chick embryo. Fertilization occurs internally; the egg is laid after ~24 hours and hatches ~21 days later. It is possible to open the egg without disturbing embryo development, which allows for various tissue removal (extirpation) and transplant type studies. Fertilized chicken eggs are relatively inexpensive and the tools involved are fairly standard.[622] Another important factor is that it is possible to transplant tissues between quail (*Coturnix coturnix japonica*) and chicken (*Gallus gallus*) embryos. While these birds and their eggs are of different sizes, their developmental rates are similar (quail 17 days). Importantly, cells from transplant and host can be distinguished based on chromatin organization (→): in both embryonic and adult quail cells heterochromatin is condensed into a small number (1 to 3)

from LeDouarin et al 1996.



[621] Zebrafish Genome-Editing Services

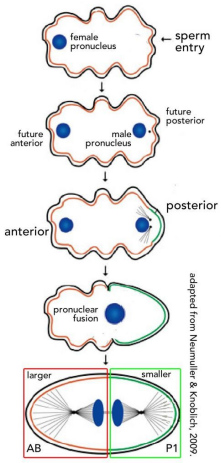
[622] Le Douarin et al. 1996 Quail-Chick Transplantations

aggregates in the central region of the nucleus, and can be visualized using a histological staining (Feulgen-Rossenbeck) reaction. In more modern studies, it is possible to transplant chick cell transgenic for GFP expression; such chimeric embryos often develop normally and may hatch.

The fruit fly *Drosophila melanogaster*

The fruit fly has many of the features we look for in a model organism, it is easy and economical to maintain in the lab. Mating (fertilization is internal) produces many offspring, resulting in the laying of embryos that develop quickly in culture, produce motile larvae that undergo metamorphosis to produce sexually mature adults in 10-12 days. Adults can be anesthetized and easily sorted under a dissecting microscope while virgin flies can be distinguished, making controlled crosses of phenotypically and genotypically characterized males and females possible for various genetic analyses. A number of chromosomal rearrangements are available to control for meiotic recombination effects, and recombination does not occur in males. The characterization of genetic mutations influencing early, and highly stereotyped events in early embryonic development, as well as the identification of what are known as homeotic mutations, in which a body part or region is transformed into another, set the stage for the application of molecular techniques that revealed the distribution of gene products, their binding partners (and in the case of transcription factors, the genes they regulate), and defined many of the basic mechanisms underlying embryonic development, such as, the establishment of molecular gradients, and the responses of cells to such gradients.

The nematode *Caenorhabditus elegans*



Another primarily genetic organism, at least originally, is the soil nematode *C. elegans,* in part because most adults are self-fertilizing hermaphrodites.[623] Some of the attractive aspects of *C. elegans* are that it is easy to grow in the lab, and embryos and adults can be frozen.[624] It is small (adults are ~1 mm in length). The embryo (and adult) are, like zebrafish, transparent. Its life cycle is about 3 days from fertilized egg to sexually mature adult. The embryo hatches to produce the first larval stage with 558 nuclei (some cells are multinucleate). The cell divisions that produce these cells occur in an invariant pattern, based on an early asymmetry within the egg and the site of sperm entry (→). The pattern of cell division and differentiation enables investigators to identify (and so study) cells that undergo programmed cell death (apoptosis), and to look at how mutations change patterns of cell division and differentiation. Another aspect has been centered around the ability of dsRNA to silence target gene expression for multiple generations, a phenomena known as RNAi and a form of transgenerational epigenetic gene regulation.[625] Studies of RNAi have elucidated the molecular mechanisms involved in related processes associated with small RNAs.

---

[623] C. elegans outside the Petri dish

[624] Corsi et al. 2015. A Transparent window into biology: A primer on Caenorhabditis elegans

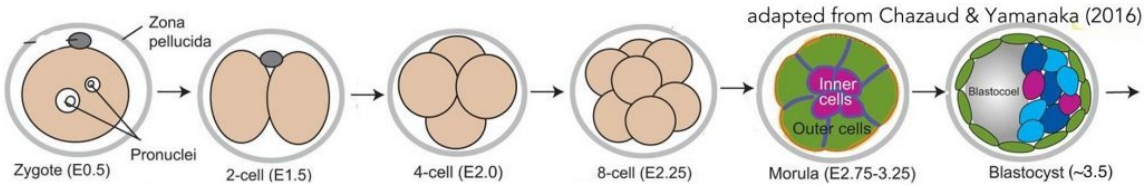
[625] Spraklin et al., 2017. The RNAi Inheritance Machinery of Caenorhabditis elegans

The Mouse

For studies of development in mammals, in which both fertilization and subsequent embryonic development occur internally (within the mother), the mouse has been the model system of choice.[626] While the costs associated with working with mice are significantly higher than the other model systems considered so far, they remain reasonable (much lower, for example, compared to working with pigs or primates), and provide experimental access, particularly through the generation of various genetically manipulated mouse lines, to carry out quite sophisticated studies. Perhaps the technique that had the most dramatic impact has been the Cre-Lox (and related) systems in which genetic manipulations (gene deletions and such) can be activated in specific cell types and at specific times during embryonic development. These have now been supplemented and extended using CRISPR-Cas9-based systems.

So why mouse, *Mus musculus*? Mice and humans shared a common ancestor ~80 million years ago, and while different share a number of physiological similarities. Pet mice have been kept for centuries, and most lab strains are derived from such mice, and so are relatively docile and, to be sure, different from wild (as opposed to wild type, i.e. non-mutant) mice. Mice have a gestation period of 19–20 days (from fertilization to birth), reach sexual maturity in 6 to 8 weeks after birth, and produce litters of 5–8 offspring. At the same time, humans are, on average ~roughly 2500 times larger than mice.

In contrast to the other model systems introduced so far, the mouse (mammalian) egg appears grossly symmetric, and sperm entry itself does not appear to impose any long lasting asymmetries. As the zygote divides, the first cells formed appear to be similar to one another. As cell division continues, however, some cells find themselves on the surface while others are located within the interior of the forming ball of cells, or morula (↓). These two cell populations are



exposed to different environments, particularly when the embryo implants into the wall of the uterus.

The surface cells differentiate to form the trophectoderm, which in turn differentiates into extra-embryonic placental tissues, the interface between mother and developing embryo. The internal cells become the inner cell mass, which differentiate to form the embryo proper, the future mouse (or human). Early on inner cell mass cells appear similar to one another, but they experience different environments, leading to emerging asymmetries associated with the activation of different signaling systems, the expression of different sets of genes, and differences in behavior – they begin the process of differentiating into distinct cell lineages and cell types forming, as embryogenesis continues, different tissues and organs. It is possible to establish "embryonic stem cell" (ES) lines from inner cell mass cells retain the totipotency displayed by inner cell mass cells, they can differentiate to form essentially any cell type found in the adult.

ESC and iPSC derived organoids

While model systems have provided a wide range of insights into the processes involved in development, and humans are clearly related to other mammals, it is immediately obvious that there are important differences – after all people are instantly distinguishable from members of closely

[626] Perlman 2016. Mouse models of human disease: An evolutionary perspective

related species and certainly look and behave differently from mice. For example, the surface layer of our brains are extensively folded (they are known as gyrencephalic) while the brain of a mouse is smooth as a baby's bottom (and referred to as lissencephalic). The failure of the human brain cortex to fold is known as lissencephaly, a disorder associated with several severe neurological defects.[627] With the advent of more and more genomic sequence data, we can identify human specific molecular (genomic) differences. Many of these sequence differences occur in regions of our DNA that regulate when and where specific genes are expressed. Sholtis & Noonan provide an example: the HACNS1 locus is an 81 basepair region that is highly conserved in various vertebrates from birds to chimpanzees; there are 13 human specific changes in this sequence that appear to alter its activity, leading to human-specific changes in the expression of nearby genes.[628] At this point ~1000 genetic elements that are different in humans compared to other vertebrates have been identified and more are likely to emerge.[629] Such human-specific changes can make modeling human-specific behaviors, at the cellular, tissue, organ, and organismic level, in non-human model systems difficult and problematic. It is for this reason that scientists have attempted to generate better human specific systems.

The Nobel prize winning work of Kazutoshi Takahashi and Shinya Yamanaka, who devised the methods to take differentiated (somatic) human cells and reprogram them into ESC/PSC-like cells, cells known as induced pluripotent stem cells (iPSCs), represented a technical breakthrough that jump-started this field.[630] Since then progress as been rapid. In particular, Madeline Lancaster, Jurgen Knōblich, Yoshiki Sasai, and a growing community of others have devised approaches by which such cells can be induced to form tissue specific organoids. Cerebral organoids, which produce brain-like tissues, have been used to examine developmental defects associated with microencephaly and Zika-virus infection-induced microencephaly, lissencephaly, Down's syndrome and others. Both ES and iPS cells can be induced to differentiate into what are known as gastruloids. Gastruloids can develop anterior-posterior (head-tail), dorsal-ventral (back-belly), and left-right axes analogous to those found in human embryos.[631] Perhaps surprisingly (and perhaps not) human organoids develop along a time-line to that observed in intact human embryos, which means that these studies can take significant amounts of time.

Done for now (revisions envisioned)

---

[627] lissencephaly

[628] Sholtis & Noonan. 2010. Gene regulation and the origins of human biological uniqueness

[629] McLean et al. 2011. Human-specific loss of regulatory DNA and the evolution of human-specific traits

[630] Takahashi & Yamanaka 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors and How iPS cells changed the world

[631] Turner et al 2017. Anteroposterior polarity and elongation in the absence of extra-embryonic tissues and of spatially localised signalling in gastruloids: mammalian embryonic organoids

Acknowledgements

biofundamentals began more than a decade ago when I found myself teaching the introductory course in molecular, cellular and developmental biology. Dissatisfied with the books available, I decided to try to present the foundations of modern (mostly molecular) biology in what I hoped was a clearer, more logical, and more interactive way through a web site. On realizing that evolutionary biology had been omitted from the curriculum, I introduced a section on that topic as part of the over-arching framework for the course (and perhaps even the subsequent curriculum). The inspiration to turn it into a book came from the success of the CLUE (chemistry, life, the universe and everything) project, a collaboration with Melanie Cooper whose aim was to improve the design of introductory courses in general and organic chemistry. biofundamentals has since been extended to consider topics in molecular genetics and developmental biology. Throughout its evolution I have been grateful to the students involved together with Jeremy Rentsch and Emina Begovic, who helped me gain some perspective on what is and is not important, and to my children for providing escape, meaning, and a recognition of the importance of inclusive teaching in an increasingly weird world.

I greatly appreciated the support of Spencer I. and Lynn Browne early in the development of the virtuallaboratories project, and Hillary Browne for giving us space in her building! Tom Lundy was a great partner in the virtuallaboratory project, transforming my appreciation of what might be done through his amazing FLASH applets (which has become particularly poignant with the demise of FLASH). Similarly my involvement in the Dynamic Cell project (Springer) got me thinking about what was and was not useful to present to students. Looking back, I recognize that Bruce Alberts and Harvey Lodish were an inspiration, prestigious scientists who took education seriously enough to think about it when (rather surprisingly) all too many in academia see thinking about education as a distraction. When Harvey Lodish asked me to contribute a “Working with the Literature” section for Molecular Cell Biology, it helped me focus my thinking on underlying biological processes.

As I began building the first web-based version of biofundamentals I was inspired by a great collaboration with Kathy Garvin-Doxas and Isidoros Doxas, who cared about revealing what students think. I greatly appreciated the benign neglect of my academic department and college for not generating too many obstacles to my following my educational passions, interests, and obsessions, although I would have welcomed their more active engagement in the project. I am particularly grateful for the fantastic collaboration I have had with Melanie Cooper, who opened my eyes to many educational and chemical ideas - our many discussions (and a few disagreements) have been transformative. Over the years interactions with many students in the lab and in various classes, have made all the stresses associated with this project totally worthwhile and deeply rewarding, thanks!

I particularly appreciate my colleague Jon Van Blerkom for his supportive comments on the text and his general encouragement, such things really matter and are often too rare. I appreciate all those who have looked, read, and commented on the materials presented - there is nothing more useful than an engaged and critical reader. Now if only the powers that be would make educational engagement, effectiveness, and outcomes the institutional priority it needs to be.

–

We end here! Please excuse (and let us know) about any errors you find – this is clearly a work in progress, particularly this last section).