

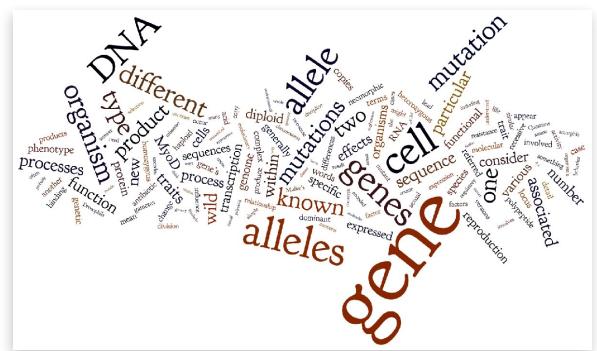
Part II: From molecular biology to genetics and genetic technologies



In which we consider the behavior of genes during the course of asexual and sexual reproduction, how they interact with one another, both along chromosomes and more generally, and how they can be manipulated and studied.

Chapter 11: A short review of concepts with which you should already be familiar

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 again this semester.



Words, terms, and processes we (really) need to understand: As you will have noted, 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. Up through this point, you have probably run into a number of these words. Science is generally referred to as a discipline, and one critical aspect 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 preference – in a sense 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.

So just to review (and if you feel the need, look at chapters 7 & 8), consider what we mean by a **gene**: within a cell, a gene is a stretch of DNA that can be **expressed**. A gene includes the DNA sequences involved in determining where and when 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 the DNA, is generated through the process of **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, is determined by where specific sets of transcription factors (proteins) bind to specific sequences within the DNA; these sites of transcription factor binding are part of the gene’s regulatory sequences. The binding of transcription factors acts to recruit and activate the RNA polymerase. 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. The cell's genome is characteristic, and serves to define the type of organism, the species, the cell is, or is part of. Organisms of the same species have extremely similar genomes, more similar than do organisms of different species – genome differences define a species. A cell can have either one complete genomic copy, in which case it is known as **haploid**, or it can contain two, 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 (3 copies), tetraploid (4 copies), or polyploid (> four copies).

As we will go into greater detail, 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 only a single cell, a single individual. The genome of the cell is duplicated through the process of **DNA replication** (which is mediated primarily by DNA-dependent, DNA polymerases and other factors) and then the cell splits into two. Similarly, at the organismic level, there is no essential need for cooperation between different organisms, or different cells, for reproduction to occur. 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 cooperate with and fuse with one another. Such cooperation requires them to recognize one another 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. Typically, gametes from two different organisms come into proximity through the process of mating and fuse with one another; their initially distinct plasma membranes become one, thereby forming a new diploid cell (organism). Some people might say that this is when life begins, but they would be confused – 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.

The two modes of reproduction have different characteristics. In asexual reproduction, the 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, can modify genomes (something we will discuss). In contrast sexual reproduction (including the process of meiotic recombination, which we will consider in detail) enables alleles to move more or less independently of one another. Sexual reproduction is also associated 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 with a range of evolutionary implications.³⁷⁴

Questions to answer and ponder:

178. How are transcription and translation similar, how are they different?
179. Within a gene, what signals and signal binding proteins are involved in gene expression? make a diagram.
180. How would having two copies of a gene (in a diploid cell) alter the behavior the cell?

Where do genes, alleles, and mutations come from?

When we think about genes, there are 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,

³⁷⁴ here is an interesting book on the topic: [The Mating Mind by Geoffrey Miller](#).

our parents (at least for the sexual among us), through the process of DNA replication. 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), and this organism already had lots of genes. These genes arose even earlier, through processes involving various molecular systems that were active before the appearance of LUCA. New genes have been observed to appear *de novo* out of DNA sequence, in various organisms, in particular the fruit fly *Drosophila*.³⁷⁵ Perhaps even more surprising, many of these *de novo* genes appear to have become essential rather quickly.

³⁷⁶ A number of putative *de novo* genes have been identified in humans.³⁷⁷

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 appear through the process of gene duplication, leading to the production of paralogs, and disappear through gene deletion. A number of studies, beginning with the classic Luria-Delbrück experiment (which we will discuss in detail), indicate that these processes (mutation, recombination, deletion, and duplication) occur largely randomly, based on the molecular nature of DNA, various molecular mechanisms active in the cell, and environmental effects (chemicals and radiation). They appear randomly, and not to meet the adaptive needs of the organism. Once a mutation arises it can, however, effect the organism's **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, 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. 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**). Generally, environmental factors and preexisting adaptations and behaviors determine the selective pressure on a new allele. There are also processes, such as **genetic drift** and various **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

The specific version of a gene, defined by the gene's DNA sequence, is known as an allele. In a diploid organism, the two copies of the gene 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 some genes and heterozygous for others. Different alleles can be expressed differently, due to differences in their regulatory sequences, and they can encode different gene products due to differences in their transcribed and coding (in the case that the gene encodes a

³⁷⁵ 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.

³⁷⁶ 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*.

³⁷⁷ [De novo origin of human protein-coding genes](#)

polypeptide) regions. Within a population of organisms of the same species, there can be multiple versions of a particular gene, multiple alleles. Later we will use the EXAC browser to visualize the alleles of various human genes. 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, a result of the divergent evolutionary histories; they can differ in terms of synteny, with genes deleted, duplicated, or moved to different chromosomal positions within the genome (genomic rearrangements.) 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 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 by its genome, that is all of the genes it contains. 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 the DNA is packaged within the cell, which can 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, essentially anything and everything about an organism that you can observe and measure. 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 (see below). In the case of haploid organisms, genotype (the alleles present in an organism) maps to the organism's traits in a relatively simple manner – an example that we will consider in greater detail is involved in antibiotic resistance. A bacteria that contains a functional copy of a gene that confers resistance to an antibiotic is resistant to that antibiotic. A mutation that inactivates that gene's expression or the gene product it makes can 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 (targeting). Even a simple gene (allele) → phenotype relationship is based on the functioning of a complex biological system, a system composed of thousands of genes and 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 allele and a phenotype is more complex in a diploid organism since there are two copies of most genes (with the possible exception of genes associated with sex determination, which we will return to). The two copies of the gene present can be the same, in which case the organism is referred to as homozygous for that genetic locus (gene), or they can be different, or heterozygous at that genetic locus. These terms always refer to a specific genetic locus. If an organism is homozygous for all genetic loci, it is generally the result of extensive in-breeding. Of course, these terms do not apply to prokaryotes which are generally haploid.

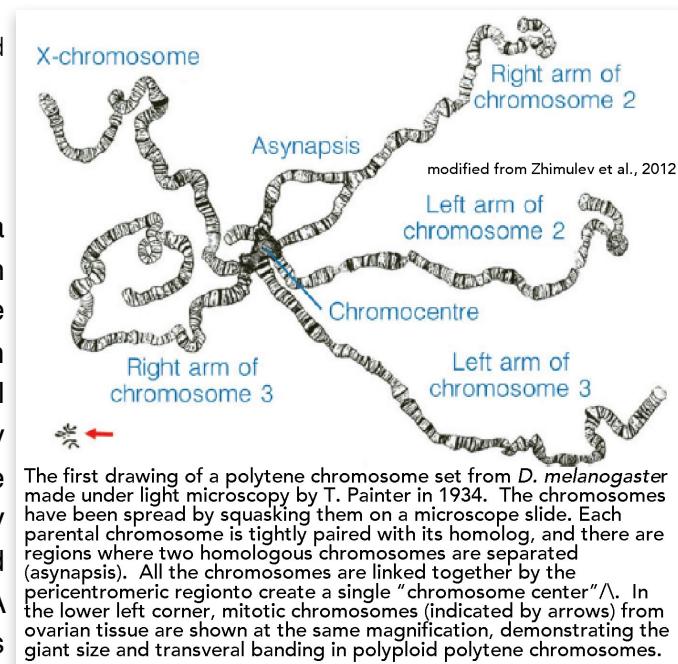
Now consider a trait that is associated with the presence of a particular allele. If the trait is visible when the locus is heterozygous for that allele, the allele is referred to as **dominant** to whatever the other (different) allele might be. On the other hand, if the trait is not apparent when the 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. We will return to, and consider all of these topics in greater detail, when we consider the interactions between combinations of alleles. Similarly, the extent of 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 refer to alleles that are associated with visible traits. Most alleles are neither strictly recessive nor dominant, and contribute in complex ways to a number of traits. Because it is easier to make sense of things we will generally start, at least initially, with strictly dominant and recessive alleles, and then get more complex.

Questions to answer and ponder:

181. Based on your understanding of DNA, draw out (schematically) the relationship between a specific allele and the phenotypic traits it is associated with.
182. Why might the mutation of gene not be associated with any one specific phenotypic trait?

Muller's Morphs

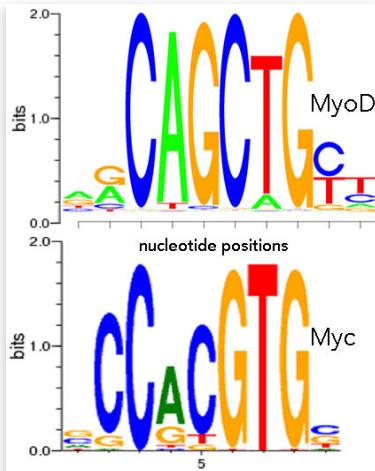
Another way to look at alleles is from a functional perspective. This was the approach taken by Herman J. Muller in the 1920s and 30s. He exploited work done in the fruit fly *Drosophila*, in which a number of gene and regional chromosomal duplications and deletions had been constructed by geneticists, something made possible by unique aspects of chromosome organization in the salivary glands of the fly (→). These cells are polyploid, and each chromosome contains more than 1000 DNA strands.³⁷⁸ Based on the analysis of various



³⁷⁸ [Banding patterns in *Drosophila melanogaster* polytene chromosomes correlate with DNA-binding protein occupancy](#).

mutations he was able to place mutations into distinct functional (with respect to a particular phenotype) classifications: that is amorphic, hypomorphic, hypermorphic, antimorphic, and neomorphic. These classes are compared to the wild type ("normal") version of the allele. At this point we will not consider mutations that have no functional effects on the gene. Compared to the level of 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 (see below). 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 happen, for example the mutation can change the specificity of an enzyme, something that can happen in the course of cancer development.³⁷⁹ To illustrate one such neomorphic mutation, consider the myogenic transcription factor MyoD, a protein that regulates the differentiation of skeletal muscle cells. There are mutations (alleles) associated with an aggressive form of embryonal rhabdomyosarcoma, a cancer of skeletal muscle. One mutant allele changes the DNA sequence so that the leucine found at position 122 of the wild type MyoD protein is replaced by an arginine.³⁸⁰ Such a mutation is known as a **missense** mutation. 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 leads to a change in gene expression. The wild type MyoD protein recognizes a consensus sequence (top panel →); in contrast the mutant allele encodes a protein whose DNA sequence 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 cell differentiation into non-dividing muscle cells (MyoD), acquires a new function, the ability to bind to different binding sites and induce cell division – a key feature of cancer cells. The mutation is neomorphic because the mutated MyoD protein (known as MyoD $\text{Ala}_{122} \rightarrow \text{Arg}$) has a new function, and probably much weaker if any binding to its original target sequence.³⁸¹



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

³⁷⁹ [Neomorphic mutations create therapeutic challenges in cancer](#)

³⁸⁰ from http://crunch.unibas.ch/ENCODE_REPORTS/Myers_HudsonAlpha/BG_5_8/report_BCLAF/JASPAR.Myf.wm.html and [Deep Sequencing of MYC DNA-Binding Sites in Burkitt Lymphoma](#)

³⁸¹ we will return to this topic toward the end of book: see [Neomorphic mutations create therapeutic challenges in cancer](#)
biofundamentals™ Klymkowsky & Cooper - copyright 2010-2018 version: Tuesday, January 16, 2018 230 of 315

behavior known as **haploinsufficiency**, 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 sufficient.

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 do change the DNA sequence – these are known as **synonymous** mutations, and as will see produce what is known as single nucleotide polymorphisms (SNPs), a feature in the DNA that can be detected by various molecular methods and is often used in the analysis of genomic similarities and differences (including human ancestry). There are two other types of generic names for alleles. A **non-sense mutation** is one that 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 depend upon where they occur within the gene. In eukaryotic genes, which often have many exons and introns, there can be mutations that disrupt the sequences involved in recognizing and removing introns following transcription. These are generally referred to as splice-site mutations, since the process of RNA processing 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 polypeptide, both non-sense mutations and mutations that alter an intron-exon junction can result in what is known as a **loss of function** (LoF) mutation, or one of Muller's morphs (although which type of morph is formed is dependent upon the mutation). Similarly, they can produce 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.³⁸² Don't get confused, all biological processes are complex, it is just that some alleles in some genes generate easily recognizable (distinctive) phenotypes.

Questions to answer

183. Draw out the relationship between gene - RNA - polypeptide, and describe the effects of missense, non-sense, and intron-exon junction mutations on gene expression.
184. How does the position within a gene of any of the mutations mentioned above influence their effects on the function of the gene's product?
185. Why is the MyoD mutation (mentioned above) neomorphic? What would you call it, if the mutated MyoD protein blocked the binding of wild type MyoD to its target DNA sequences?
186. Describe how a DNA change (missense, non-sense, junction mutation) produce Muller's morphs.
187. Describe how a neomorphic mutation alters the behavior of transcription factor and 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 *Drosophila* 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?

³⁸² [Pleiotropy: One Gene Can Affect Multiple Traits](#)

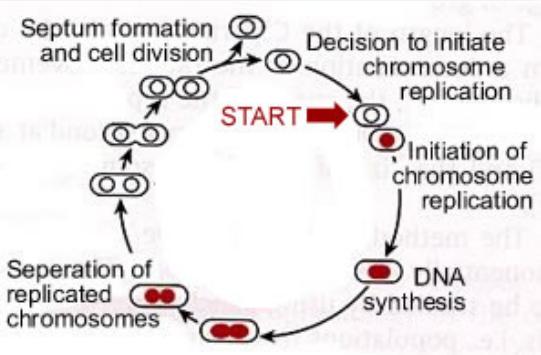
Chapter 12: Reproduction in prokaryotes and horizontal gene transfer

In which we consider how prokaryotic cells replicate asexually, and how they can (under specific conditions) pass genetic information to one another and acquire up such information from their environment.

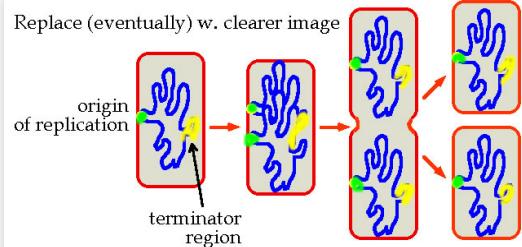
Asexual reproduction in bacteria and archaea



The simplest type of biological (cellular) 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. The chromosome contains two important sequence elements, the origin of replication (ORI) and the terminator (TER). When conditions are appropriate, a cell will pass through a decision point, a molecular switch, known as start (\rightarrow). This switch activates the proteins that bind to the ORI region of the chromosome, and initiates the assembly of the DNA replication complex. A replication bubble 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



Replace (eventually) w. clearer image



around the DNA molecule, and collide in the TER region (\leftarrow). As the DNA replication forks collide, they generate a signal that indicates that DNA replication is complete. During this period the cell itself is also 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 other genes) present. Over time, each organism produces a clone, and various clones interact with the environment and each other independently. These clones can display different levels of reproductive success, some clones can take over the population, while others can become extinct. In the case of studies on the evolution of bacterial antibiotic resistance (see below, page 259), each clone has to develop antibiotic resistance independently of every other clone; a similar

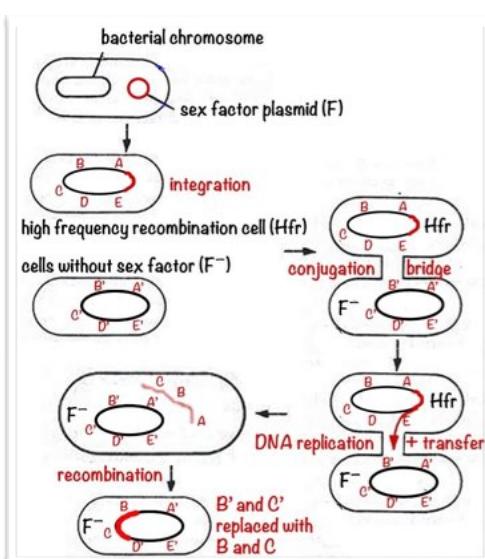
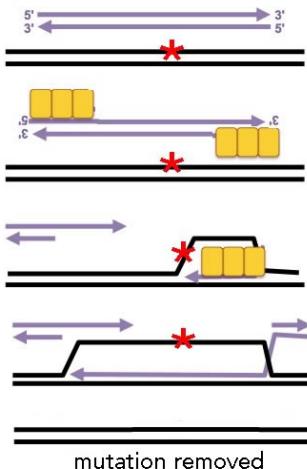
situation was observed in long term bacterial evolution studies.³⁸³ There is no cross talk between lineages in such situations. Of course, if DNA is passed from clone to clone, as occurs within Griffith's transformation experiments (see above, page 157), things can get more complex. The movement of genes between lineages is known as horizontal gene transfer. Here we will review three versions of horizontal gene transfer found in prokaryotes.

Conjugation: what counts as sex in prokaryotes

The process of conjugation in bacteria allows DNA to move from one cell to another, with the moving (donor) DNA replacing the host DNA through the process of homologous recombination, a mechanism that we will consider, only briefly and at seriously over-simplified molecular detail (→). Homologous recombination is used in many systems, and is based on the recognition of a DNA sequence by a similar sequence.

Conjugation is a major pathway for horizontal gene transfer in bacteria.³⁸⁴ In contrast to transformation (see below), conjugation "forces" DNA into what may be a reluctant recipient cell. In the process of conjugation, we can distinguish between two types of bacterial cells (of the same species). One contains a plasmid known as the sex factor (F), the other does not, it is referred to as a F⁻ cell. The F plasmid can exist independently of the host chromosome or it can be integrated into it; cells in which the F-plasmid is integrated into the host chromosome are known as a Hfr (high frequency recombination) cells (↓). The F plasmid contains the

over-simplified homologous recombination



genes needed to transfer a copy of its DNA into a cell that lacks an F-plasmid. In this manner, an F-plasmid can colonize a population. In Hfr cells, the chromosome integrated F-plasmid can transfer host and plasmid genes into a F⁻ cell. To help make things a little simpler, we will refer to the Hfr cell as the DNA donor and F⁻ cells as DNA recipients.

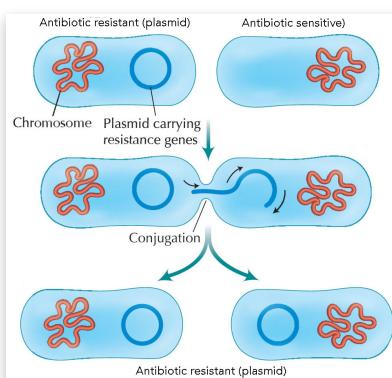
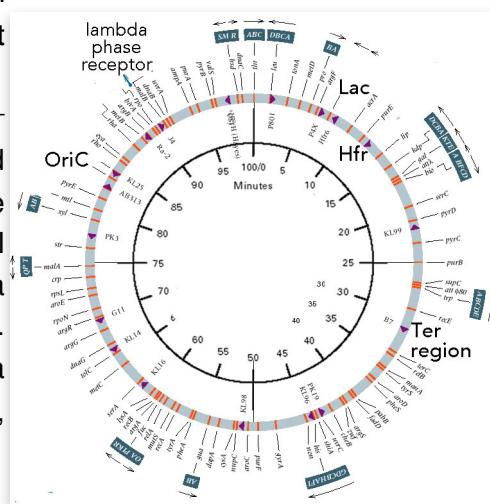
To initiate conjugation, the Hfr/F⁺ cell makes a physical (conjugation) bridge to the F⁻ cell (←). A break in the donor DNA initiates a process by which single stranded DNA is synthesized and moved into the recipient F⁻ cell. The amount of DNA transported is determined largely by how long the bridge between the cells remains intact. It takes ~100 minutes to transfer the entire donor chromosome from an Hfr to an F⁻ cell. Once inside the F⁻ cell, the donor DNA is integrated into the recipient's chromosome, replacing the recipient's versions of the genes transferred, through homologous recombination - which we will return to later. Using Hfr strains carrying different alleles of various genes, and by controlling the duration of conjugation by breaking the conjugation bridge by placing the cells in

³⁸³ see [A cinematic approach to drug resistance](#) and [E. coli Long-term Experimental Evolution Project](#)

³⁸⁴ review of [prokaryotic conjugation](#) (prokaryotes)

a kitchen blender, the experimenters were able to determine the order of genes along the chromosome. The result was the discovery that related organisms had the same genes arranged in the same order, their genes were in syntenic groups (see above).³⁸⁵ 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.



Versions of this process are involved in the transfer of plasmids from cell to cell within a community (←).³⁸⁶ A plasmid contains its own “origin of replication”; some (low copy number) plasmids exist in one to two copies per cell, while others (high copy number plasmids) are present in multiple copies, which is determined in large part by their origin of replication sequences. Some plasmids can exist in as many as 700 copies per cell.³⁸⁷ Plasmids can encode genes responsible for antibiotic resistance and the rapid dispersion of antibiotic resistance phenotype is a cause of increasing concern.³⁸⁸ Many plasmids, also known as mobile genetic elements, are more selfish, that is, their presence in a cell might not directly benefit that cell. Once a plasmid has a Hfr-like element, it can move through (and parasitize) a population. We (that is you) might even be able to generate a plausible mechanism by which viruses could have evolved from such “selfish” plasmids.

Questions to answer & ponder:

188. What factors act to insure that each (prokaryotic) cell generated contains a complete genome?
189. How would mutating the origin or terminator regions of a prokaryotic chromosome influence the cell's reproduction?
190. Describe what you would expect to happen, and why, if a prokaryotic cell received an incomplete genome.
191. Describe (diagram) what happens to the DNA molecule that is introduced to a cell via conjugation?
- How might the regulation of plasmid ORI regions be different in low and high copy number plasmids?

³⁸⁵ Synteny: <http://en.wikipedia.org/wiki/Synteny>

³⁸⁶ Plasmids Spread Very Fast in Heterogeneous Bacterial Communities: <https://www.ncbi.nlm.nih.gov/pubmed/12524329>

³⁸⁷ [Plasmids 101: Origin of Replication](#)

³⁸⁸ 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.³⁸⁹ It is important to note 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 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.³⁹⁰ Alternatively, the nucleotide sequence of a DNA molecule can be integrated into another organism's genome, resulting in the 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 – imagine how inserting a piece of DNA into a genome could be harmful. The study of these natural DNA import systems has identified specific molecular machines that mediate DNA transfer. Some organisms use a system that preferentially imports DNA molecules that are derived from organisms of the same or closely related types as themselves. You can probably even imagine how they do this – they must have receptor systems that can 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 are more common. It turns out that a population of organisms does not have to “invent” all of its own genes, it can adopt 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

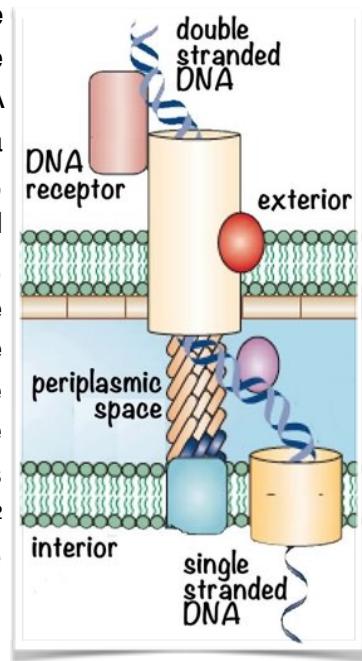
There are well established methods used in genetic engineering to enhance the ability of bacteria to take up plasmids from their environment.³⁹¹ We, however, focus on natural transformation, the process associated with the transfer of DNA molecules from the environment into a cell. Transformation is an active 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 expense to express the genes and synthesize the proteins involved in the DNA uptake and integration machinery. In fact, bacteria can sense the density of organisms in their environment using quorum sensing (see above, pp. 91-95). Bacteria use quorum sensing to generate the DNA uptake system, apparently by activating specific transcription factors. When present in a crowded environment, the quorum sensing system turns on the expression of the DNA uptake system.

³⁸⁹ DNA uptake during bacterial transformation: <http://www.ncbi.nlm.nih.gov/pubmed/15083159>

³⁹⁰ This is of course why genes are rarely if ever transferred from food to the organism doing the eating.

³⁹¹ Making Calcium Competent (bacterial) Cells: http://mcb.berkeley.edu/labs/krantz/protocols/calcium_comp_cells.pdf
biofundamentals™ Klymkowsky & Cooper - copyright 2010-2018 version: Tuesday, January 16, 2018 235 of 315

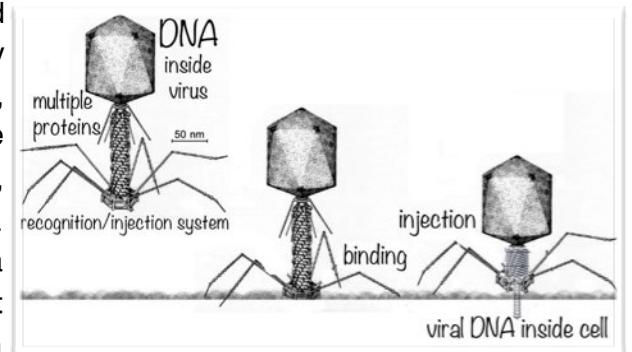
Here we outline the process in one type of bacteria but functionally similar mechanisms are used in other bacterial and archaeal species. Double-stranded DNA binds to the cell's surface through a variety of DNA receptors. In some cases these receptors bind specific DNA sequences, in others they bind 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 periplasmic space in between (→). In an ATP-hydrolysis coupled reaction, DNA bound to the exterior surface of the bacterium is moved, through a protein pore, through the outer membrane and into the periplasmic space, where it is passed to the DNA channel protein. Here one strand 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. 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.³⁹² 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 such 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, 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.

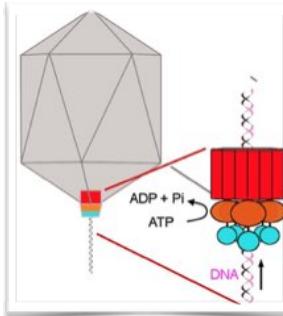
The simplest viruses contain a nucleic acid genome and a protein-based transport and delivery system. We briefly consider a typical bacterial virus, known as a bacteriophage or bacteria eater. The bacterial virus we consider here, the T4 bacteriophage, looks complex and it is, other viruses are simpler (→). 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).³⁹³ 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



³⁹² Bacterial transformation: distribution, shared mechanisms and divergent control & Natural competence and the evolution of DNA uptake specificity

³⁹³ http://en.wikipedia.org/wiki/Bacteriophage_T4

found, based on tail binding to cell surface molecules, the tail domain attaches 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.³⁹⁴ The phage DNA encodes the proteins that are used to assemble new phage heads. DNA is packed into these heads by a protein-based DNA pump, a pump driven by coupling to an ATP hydrolysis reaction complex (\rightarrow).³⁹⁵ In the course of packaging virus 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 injects that cell with a DNA fragment derived from the previous host. Of course, this mis-packaged DNA may not contain all of the genes the virus needs to make a new virus or to kill the host. The transferred DNA can be inserted into the newly infected 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.³⁹⁶ These include the restriction/DNA modification systems used widely for genetic engineering, and the CRISPR-CAS9 system, which enables cells to recognize and destroy foreign (viral) DNA. These systems, evolved as part of prokaryotic immune systems, form the basis of modern genetic engineering methods. They illustrate how studying apparently arcane aspects of the biological world can have dramatic impacts on modern technological, medical, and economic systems.

Questions to answer:

192. What is an asexual clone? How would you recognize it.
193. What is the effect of a amorphic allele / mutation on the behavior of a prokaryotic clone.
194. What are some possible (evolutionary) advantages to the ability to take up and integrate, as opposed to simply eat foreign DNA?
195. Why might the "source" of foreign DNA matter?
196. Present a plausible model that would identify host from foreign DNA.
197. What factors are necessary for homologous recombination?
198. Propose the steps that would be involved in the evolution of a "selfish" plasmid into a virus.

Questions to ponder:

- How might a prokaryotic organism protect itself from invading viruses?
- How might the importation of DNA through transformation be harmful to the host?

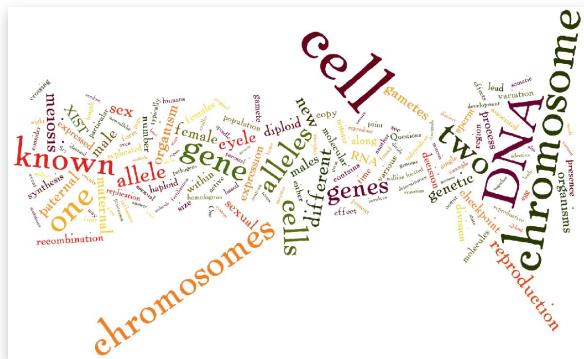
³⁹⁴ An infected bacterial cell can protect its neighbors, often its clonal relatives, if it can kill itself before the virus can replicate. This is an example of a simple altruistic behavior.

³⁹⁵ [The Structure of the Phage T4 DNA Packaging Motor Suggests a Mechanism Dependent on Electrostatic Forces](#)

³⁹⁶ see [The phage-host arms-race: Shaping the evolution of microbes](#)

Chapter 13: Asexual and sexual reproduction in eukaryotes

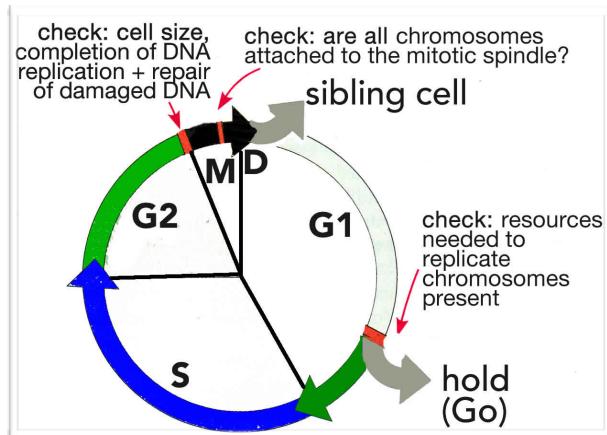
In which we consider the processes of asexual and sexual reproduction in eukaryotes. We note the molecular processes, mitosis & cytokinesis, involved in somatic cell reproduction and how they are modified in meiosis and gamete formation within the germ line. We consider the implications of chromosome pairing, recombination & independent segregation as well as dimorphism of gametes leading to maternal and paternal effects, including mitochondria inheritance and sex determination.



Asexual reproduction in a eukaryote: making a (somatic) clone

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 divide. At that point the genome of the cell is replicated and the cell is then divided into two, each receiving one complete copy of the genome. In addition, all eukaryotes have cytoplasmic organelles (mitochondria, and in algae and plants, chloroplasts) with their own, albeit reduced in size, genomes. 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).³⁹⁷ In the eukaryotes that we will concern ourselves with, most of the cells of the organism are diploid (rest assured, we will let you know when they are not).

Somatic (asexual) reproduction involves what is known as a cell cycle. We think of the cell cycle as beginning with the process of cell division (D in the figure below)(↓). The process of dividing one cell into two, known as cytokinesis, results in two sibling cells, each with identical genomes. Cytokinesis involves cytoskeletal and cytomuscular systems that are generally discussed in detail in a later cell biology course (not here!) Generally, cell division is symmetrical, so that the two sibling cells are half the volume of the parental cell and very similar. Division is followed by a period of cell growth (known as G₁) (→), during which energy and materials are imported from the external environment, or 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 is a decision to be made, will it 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 G₀ (↑). The majority of cells in any particular tissue are in the G₀ state; in G₀ there is no new DNA synthesis, so the possibility of mutation is lower than when DNA is being



³⁹⁷ 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.

replicated. If, however, various external and internal signals act on and within the cell, the cell can reverse the G₀ decision and resume growth and eventually divide (note that it is difficult to talk about these systems without personalizing them, even though they are certainly not conscious).

Once the decision to proceed has been made, that is, the molecular switch has been flipped, the cell will encounter what is known as a checkpoint, discussed in more detail below.³⁹⁸ A checkpoint is a molecular feedback system by which the cell essentially calculates various aspects of its internal state and makes a decision to wait or proceed with DNA synthesis. The decision to start DNA synthesis is based in part on whether the cell has, or will have, sufficient resources to replicate its DNA molecules, which requires (in a human cell) ~12 billion nucleotide addition reactions (both strands of a total of ~6 billion base pairs). The DNA synthesis decision point is known as “start”; once that decision is made the cell will continue to grow and proceed into the part of the cell cycle during which DNA synthesis occurs, known as S (↑). At the end of this phase of the cell cycle, DNA synthesis will be complete and the cell will continue to grow; the cell has entered into what is known as the G₂ phase of the cell cycle. The start decision is particularly critical for the cell (and the organism), since failure to complete DNA replication will likely lead to changes in gene number and increased mutagenesis and inaccurate repair of single stranded DNA molecules.³⁹⁹ There are mutant alleles, originally described in yeast and known as “wee” mutations, in which the molecular switch controlling entry into S is damaged; these mutations lead to a disconnect between growth and division and result in smaller and smaller cells and eventually cell death.⁴⁰⁰

During the asexual reproduction cycle the ploidy (the number of copies of 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 during S-phase of the cell cycle is that there are now two copies of each chromosome. While a cell is diploid during G₁, it is effectively tetraploid during G₂. 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 would expect to see changes in the rates of gene expression in G₂ compared to G₁ cells.

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 sites per chromosome at which DNA synthesis starts. These replication origins are regulated during S phase such that each is activated once and only once, so 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 delays the cell cycle until that DNA has been replicated. 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

³⁹⁸ The quorum sensing systems we discussed previously is a version of a checkpoint system.

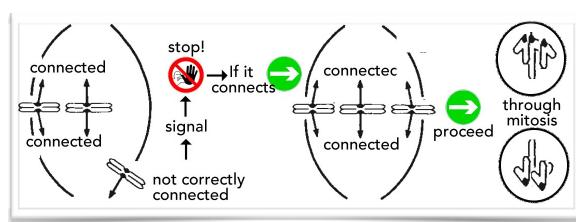
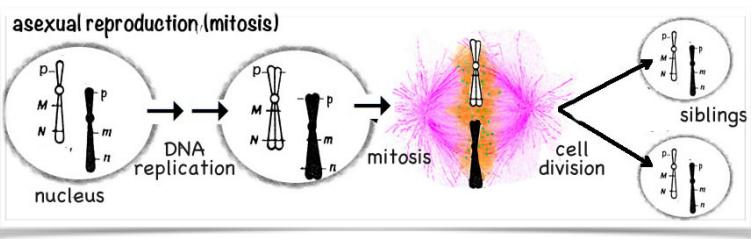
³⁹⁹ DNA replication is complex process, with details considered in subsequent molecular biology courses: this review addresses some of these complexities [Forks on the Run: Can the Stalling of DNA Replication Promote Epigenetic Changes?](#)

⁴⁰⁰ Paul Nurse and Pierre Thuriaux on wee Mutants and Cell Cycle Control: <https://www.ncbi.nlm.nih.gov/pubmed/27927897>
biofundamentals™ Klymkowsky & Cooper - copyright 2010-2018 version: Tuesday, January 16, 2018 239 of 315

Listeria, exploit this DNA damage checkpoint to enhance their own replication.⁴⁰¹

Molecular choices and checkpoints

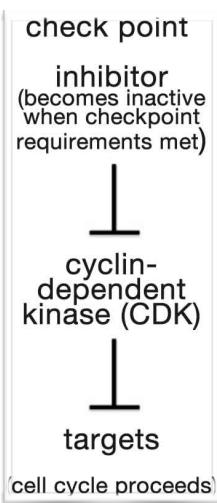
Once the DNA replication/repair checkpoint has been passed, the cell can divide. The first step of this process is known as mitosis (\rightarrow); it involves a molecular machine based on protein polymers ($\alpha\beta$ -tubulin-based microtubules) that are organized into a “mitotic spindle”. There is a molecular checkpoint switch that monitors the assembly of the mitotic spindle, and a second checkpoint that monitors that each replicated chromosome has connected correctly to the spindle (\leftarrow).



Each of the replicated chromosomes interacts independently with the mitotic spindle (this is different from their behavior during meiosis, as we will see below). Together these two checkpoints serve to insure that each of the two future sibling cells gets one and only one copy of each and every chromosome present in the parental cell.⁴⁰² The presence of the second 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.⁴⁰³

Once the chromosomes are segregated to the opposite ends of the parental cell, the parental cell divides, using another protein polymer-based (actin/myosin-based microfilaments) molecular machine, known as the contractile ring, to produce two sibling cells. It is worth noting that while these two cells are genetically 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 (rather briefly) 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. The typical check point is built around a protein kinase, an enzyme that can phosphorylate various targets – such phosphorylation (a post-translational modification) can lead to changes in protein-protein interactions and activities. Checkpoints involve a particular class of kinase, known as cyclin-dependent kinases (CDKs)(\rightarrow). The activity of these CDKs is regulated positively by the binding of a small regulatory protein, known as a cyclin, as



⁴⁰¹ [Listeria monocytogenes induces host DNA damage and delays the host cell cycle to promote infection](#)

⁴⁰² [Kinetochores, microtubules, and spindle assembly checkpoint signaling](#)

⁴⁰³ [Mitotic forces control a cell-cycle checkpoint](#)

well as other interacting proteins and a number of post-translational modifications. Cyclin's themselves are 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 inhibitor is itself inactivated, 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 irreversible until such time as cyclin levels increase again, during the next cell cycle.

Questions to answer:

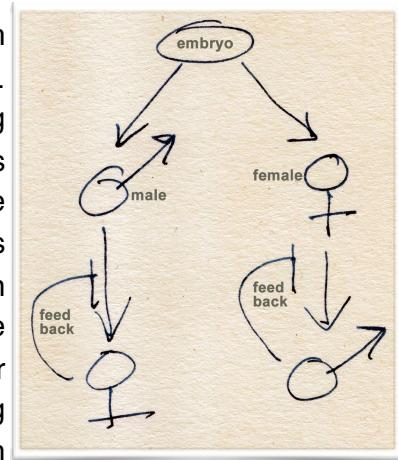
199. How do chromosomes interact with one another during mitosis/cytokinesis?
200. How do checkpoints work and what makes them irreversible?
201. What does it mean that a checkpoint acts to “make a decision based on evidence”?
202. Make a graph of CDK activity and the concentration of the cyclin regulating it, as a function of the cell cycle.
203. What can go wrong if a checkpoint is ignored (start with a cell cycle diagram)?
204. How can a mutation in a checkpoint influence cell behavior during the somatic (mitotic) cell cycle?
205. 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, 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.⁴⁰⁴ There are basically two general mechanisms that determine the sex of an organism: genetic and environmental. 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.⁴⁰⁵ Recently, climate change has been implicated in sea turtle sex ratios.⁴⁰⁶ In other organisms, all organisms originally develop into one sex or the other and, as they mature and



⁴⁰⁴ 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: <https://academic.oup.com/icb/article/46/4/349/634174/Sexual-selection-lessons-from-hermaphrodite-mating>

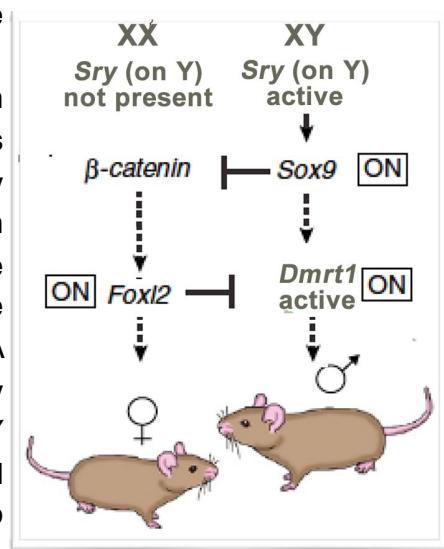
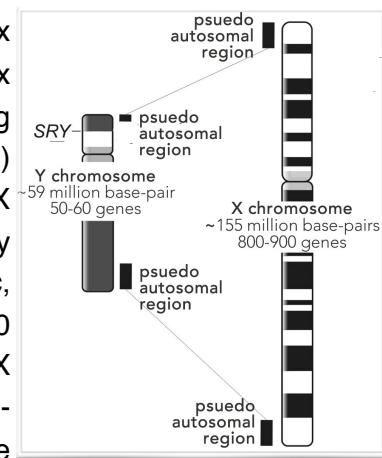
⁴⁰⁵ [Environmental sex determination mechanisms in reptiles](#)

⁴⁰⁶ [Climate change is turning 99 percent of these baby sea turtles female](#)

(often grow larger), transform into the other sex.⁴⁰⁷ 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 smaller females develops into a male.⁴⁰⁸

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.⁴⁰⁹ In humans the sex (23rd) chromosome comes in two forms, known as X and Y (→).⁴¹⁰ 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 has only ~50 genes, while the X-chromosome has 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 will have 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 on the Y (there is no copy of SRY on the X chromosome). The SRY gene is not found in monotremes (egg-laying mammals) and other vertebrates.⁴¹¹ 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, which contains a high-mobility group or HMG box, DNA binding domain. The presence of a Y chromosome, and so an active Sry gene, leads to male sexual development, whereas loss of function mutations in Sry lead to female development, even if the Y chromosome is present (→).⁴¹²



⁴⁰⁷ [Phylogenetic Perspectives on the Evolution of Functional Hermaphroditism](#)

⁴⁰⁸ [Functional hermaphroditism in teleosts](#)

⁴⁰⁹ 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.

⁴¹⁰ X chromosome regulation: diverse patterns in development, tissues and disease: <https://www.ncbi.nlm.nih.gov/pubmed/24733023> and Y-chromosome: <https://ghr.nlm.nih.gov/chromosome/Y>

⁴¹¹ Vertebrates use a number of mechanisms to determine sex, these include "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](#)

⁴¹² see [Molecular Mechanisms of Male Sex Determination: The Enigma of SRY](#) for more details.

Sry encodes a transcription factor that initiates a down-stream cascade, activating some genes and inhibiting others, with the end result being the generation of the various developmental differences associated with male and female anatomy and behavior.⁴¹³ In females other genes are expressed (actively transcribed) 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, and its presence still leads to male determination. The details of these processes are complex, so we refer further details to more advanced classes.⁴¹⁴

At this point we should mention that there are other sex determination strategies that you might come across in your subsequent studies, but which we will ignore here.⁴¹⁵ For example, in some organisms (plants and algae), the haploid (gametic) stage can persist and live independently,⁴¹⁶ but generally the haploid stage of a eukaryote, and particularly animal’s life cycle is short.

So what are the benefits of sexual reproduction, a process that requires social collaboration.⁴¹⁷ As will be noted in our discussion of meiosis (see below), the simple answer is the generation of genetic variation. So why is this variation important. One major reason arises from the presence of rapidly reproducing pathogens. Viruses, bacterial and microbial (eukaryotic) organisms typically reproduce over a period of minutes to hours to days, whereas larger organisms reproduce (generate new organisms) over a period of months, years, and decades. 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, do not kill all of the organisms they infect. And 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.

The level of genetic variation within a population is important as insurance against infectious disease. Similarly, but on somewhat longer time scales, the level of genetic variation within a population enables a population adapt to a changing environment. The larger the population size, the more likely there is some genotypic combination 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 increase in the probability of extinction.⁴¹⁸

In addition to the generation of variation, the process of sexual reproduction offers mechanisms by which to isolate populations reproductively, that is, to create two species from one. Generally males and females have to cooperate to reproduce. 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

⁴¹³ 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](#)

⁴¹⁴ [Sex determination: a primer](#)

⁴¹⁵ [The evolutionary dynamics of haplodiploidy](#)

⁴¹⁶ see wikipedia – gametophyte: <https://en.wikipedia.org/wiki/Gametophyte>

⁴¹⁷ Origins of Eukaryotic Sexual Reproduction: <http://cshperspectives.cshlp.org/content/6/3/a016154.full>

⁴¹⁸ Timing and causes of mid-Holocene [mammoth extinction](#)

diploid cell that forms has to develop normally, which then has to be able to form functional gametes, which involves the pairing of homologous chromosomes, and so on and so forth. Incompatibilities in any of these processes can lead to a reproductive barrier between individuals within populations - that is, speciation.

Questions to answer:

206. If you were design a temperature sensitive form of sex determination, how would you go about it?
207. What might happen if you removed the regions of the Y chromosome that are homologous to the X?

Question to ponder:

- Any thoughts on why different vertebrates would have adopted such different modes of sex-determination, and their evolutionary benefits and drawbacks?

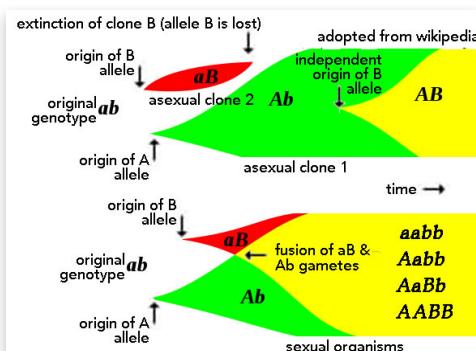
Meiosis, fertilization, and embryogenesis

In contrast to asexual reproduction, which produces clones that are identical (with the exception of newly arising mutations) to their progenitor, the result of sexual reproduction is a genetically distinct organism, different from either parent. There have been a number of explanations for why sexual reproduction is so common, essentially all visible (macroscopic) organisms reproduce (or can reproduce) sexually.⁴¹⁹ One view considers the fact that most parasites and pathogens are small, and reproduce quickly. Populations of such organisms exploit the generation of variations, through mutation, to evolve quickly. In contrast, larger macroscopic organisms typically reproduce much more slowly. So how can they keep up with their parasites and pathogens? Sexual reproduction offers, as we will see, a mechanism 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”

“It takes all the running you can do, to keep in the same place.”
says the Red Queen to Alice

constantly, in terms of generating genetic variation, to keep up with their parasites and pathogens.⁴²⁰ In addition, there is the possibility to eliminate lethal alleles from a lineage, as opposed to having to have the lineage itself go extinct, and the fact that sexual reproduction can speed the appearance of beneficial combinations of alleles (\rightarrow), combinations that would take significantly longer to appear if they had to occur independently in a particular lineage.

One aspect of the haploid state associated with sexual reproduction, is that it can reveal the presence, and lead to the elimination, of highly 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.⁴²¹



⁴¹⁹ C. Zimmer. 2009. [On the Origin of Sexual Reproduction](#)

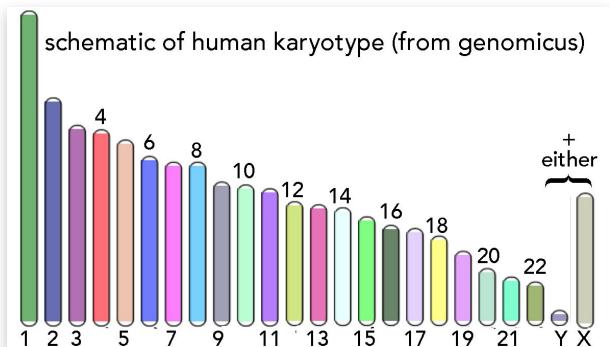
⁴²⁰ see [Sexual reproduction as an adaptation to resist parasites](#)

⁴²¹ see Evolution of haploid selection in predominantly diploid organisms: <http://www.pnas.org/content/112/52/15952.full> and Haplod selection in animals: <http://www.sciencedirect.com/science/article/pii/S0169534704002381>

Steps in meiosis

Sexual reproduction begins with two diploid cells, generally found in two distinct individuals. These two individuals are of different “mating types”, which in macroscopic organisms are referred to as the two sexes, male and female. In species with mating types, the gametes produced appear identical, and both share an equal investment in reproduction. Where gametes are different in size (an example of sexual dimorphism), the two sexes can have discordant investments in reproduction, one can spend more energy generating gametes than the other. The sex with the greater investment in gamete production is known as female, one with less of an investment is referred to as male. 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.⁴²²

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. In females the process of meiosis typically generates a single gamete, known as an egg, and three mini-cells, known as polar bodies. In males, meiosis produces four gametes, known as sperm. 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 stained specimens. In humans, the largest of these chromosomes, chromosome 1, contains ~250 million base pairs of DNA and over 2000 polypeptide-encoding genes, while the smallest, chromosome 22 contains ~52 million base pairs of DNA and around 500 polypeptide encoding genes.⁴²³ Homologous chromosomes are also defined by the order of genes found along their length. Human chromosome #5 contains different genes than are found on chromosome #6. Moreover, the maternal version of each chromosome can contain different alleles of the genes present compared to those that are found in the paternal version. In (therian) 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.⁴²⁴ The fusion event, known as fertilization, is the most discontinuous event in the process of (sexually reproducing) life. Even so,



⁴²² [A peacock's tail: how Darwin arrived at his theory of sexual selection](#) and [Mate choice and sexual selection: What have we learned since Darwin?](#)

⁴²³ We are only discussing polypeptide-encoding genes because it remains unclear whether (and which) other transcribed regions are genes, or physiologically significant.

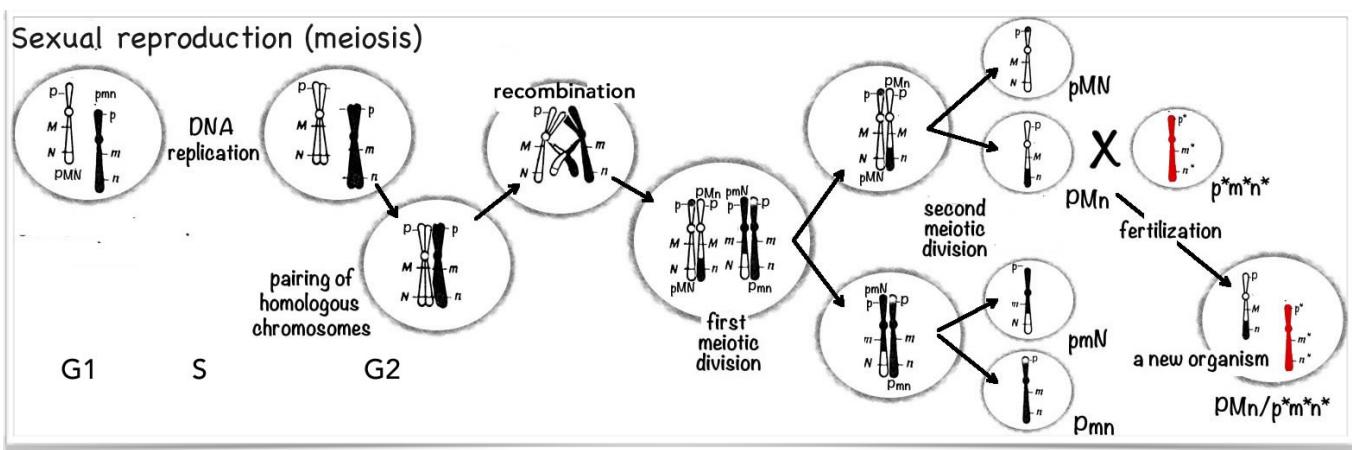
⁴²⁴ While we not deal in detail with this topic, aspects of gender are complex traits: see [Beyond XX and XY: The Extraordinary Complexity of Sex Determination](#)

fertilization does not represent a true discontinuity – both sperm and egg are alive, as is the fertilized egg.⁴²⁵ In a critical sense life (in the post-LUCA world) never begins – it continues and is transformed.

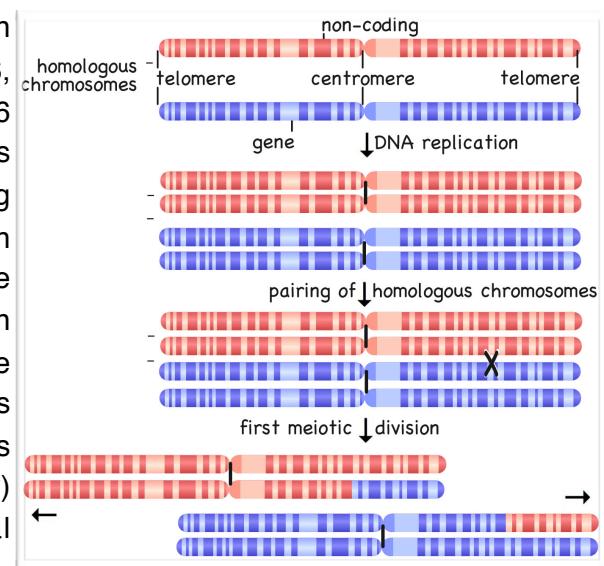
The fused cell that results from fertilization is known as a zygote; through somatic (asexual) cell division (mitosis and cytokinesis) it will develop into a adult, composed of diploid cells. The cells of the adult that produce gametes are known as germ cells, and together are known as the organism's germ line; whereas 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.

Recombination & independent segregation

We will begin our description of meiosis with a germ line cell, a cell that contains two copies of each autosome, and either two X chromosomes in a female and an X and Y chromosome in a male. The chromosome derived from the progenitor female gamete is known as the maternal copy of the chromosome, the one derived from the progenitor male gamete is known as the paternal copy of the chromosome. In order to generate gametes, a diploid germ cell enters meiosis (see video [link](#)). Meiosis (↓) consists of a single round of DNA replication followed by two cell division cycles.



As a cell enters meiosis the first step is the replication of its DNA - the cell goes from a diploid G1 state 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 two resulting replicated (double-stranded) DNA molecules remain attached to one another through a structure known as the centromere. Here is where meiosis diverges from mitosis. In asexual (mitotic) cell division each replicated chromosome remains independent of its homolog and interacts independently with the mitotic spindle (through its centromeric region). In meiosis, during G2 the (now) duplicated homologs (the maternal and paternal



⁴²⁵ 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.

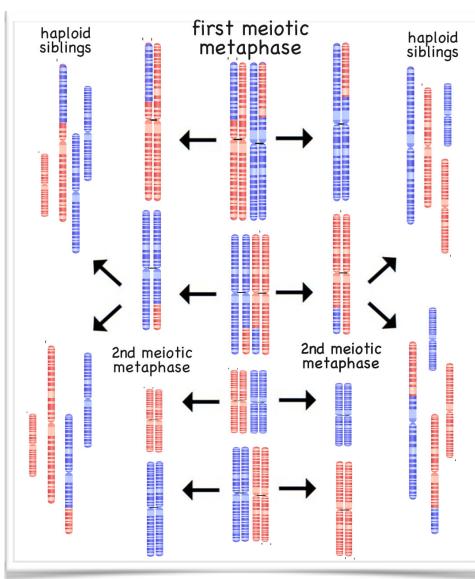
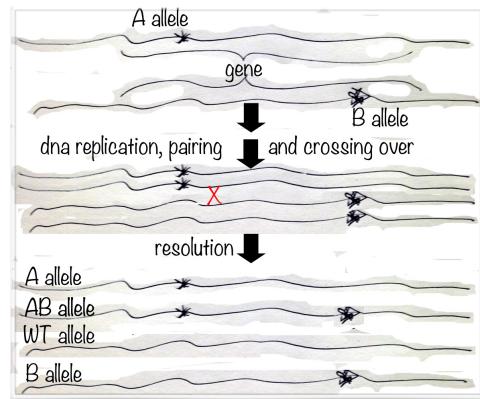
chromosomes) align with one another to form a structure containing four (double stranded) DNA molecules; these four DNA molecules are known historically as a “tetrad”. Homologous chromosome pairing is based on the association of syntenic regions of the chromosome. The DNA sequences along the homologous chromosomes, while not identical, are extremely similar, with the same genes located in the same order on each (if they are not - due to chromosomal rearrangements, things can get messy). After chromosome pairing, and at essentially random positions along the length of the

chromosome, there occurs what are known as cross-over or recombination events (\leftrightarrow). An enzyme (a DNA endonuclease) produces double-strand breaks in two of the four (double-stranded) DNA molecules (at the sites marked by “X” above or by “cross over” to the left).⁴²⁶ 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 (crossing over maternal to maternal or paternal to paternal is 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 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 (\rightarrow). 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 WT). A new allele (AB) has been created, without a new mutation!

In the case of the X and Y chromosomes, the 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 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.

In addition to the shuffling of alleles along chromosomes, and (occasionally) generating new alleles, meiosis leads to a second source of variation (\leftarrow). 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

⁴²⁶ adapted from The Centenary of Janssens's Chiasmatype Theory Koszul et al., 2012. *Genetics* **191**: 309-317.

their centromeres, maternal or paternal) they inherit is determined by chance. For an organism with 23 different chromosomes (such as humans), the first meiotic division produces 2^{23} possible different daughter cells. The process is known as the independent assortment of homologous chromosomes during the first meiotic division, or independent assortment for short.

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. 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 each receives one and only one copy of each chromosome (a double-stranded DNA molecule); again, which molecule they inherit is stochastic. The four haploid cells that are 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, which can be fertilized by fusion with a sperm cell, and the other three form what are known as polar bodies. The result 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.⁴²⁷ 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 generate new phenotypes and phenotypic variation that can impact reproductive success, and so lead to evolutionary effects.

Questions to answer:

208. Consider the odds of an organism obtaining the 3 new mutations necessary for the appearance of a new trait. If you were to predict, which would be faster (in terms of the number of generations required) in achieve this goal, a sexual or an asexual organism. Generate a drawing that illustrates your thinking.
209. You are working with an organism with 5 autosomes and 1 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.
210. Indicate (in a drawing and associated explanation) how a deleterious mutation within a gene could be generated by or eliminated from a gene.
211. How would genetic diversity be altered if meiotic recombination occurred during meiosis II, rather than during 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

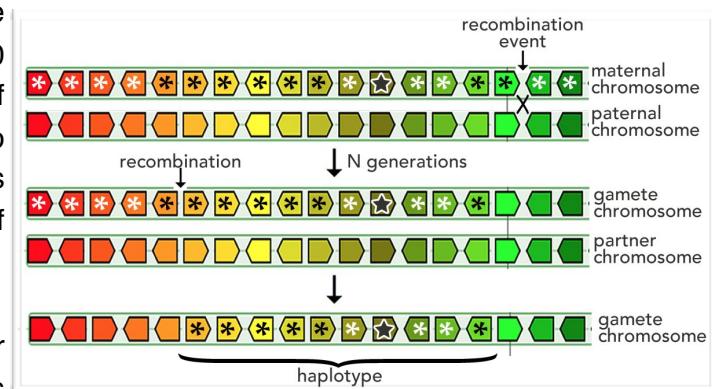
An important feature of meiotic recombination is that it can “disconnect” the alleles of genes located near one another along a chromosome. Consider the situation when a mutation occurs that creates a new allele in gene X; let us call it X^{select} . Now let us assume that this allele is subject to strong 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 selection, positive

⁴²⁷ This even applies to hermaphrodites, in which one organism acts as both mother and father!

effect on reproductive success) or against (negative selection, negative effect on reproductive success) the frequency of the allele will tend to increase or decrease in subsequent generations. The change in the frequency of the X^{select} allele also influences the frequency of alleles near the X gene. If X^{select} is subject to strong positive selection, so will alleles in genes neighboring (linked to) gene X ; similarly, if X^{select} has a negative selective effect, the frequency of the alleles in genes neighboring (linked to) gene X will also decrease over time. The closer the genes are to each other along the chromosome, the longer (over more generations) this “linkage” effect will last. Why? because the probability of recombination between two sites along a chromosome (two genetic loci or positions) is a positive function of distance. 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, 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. Two genetic loci that are 50 centimorgans (or more 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 (\star) here (\rightarrow); 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 asterisks (*). 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 \star 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 days), multiple rounds of crosses (breeding cycles) were required to locate i) on which chromosome the allele (gene) responsible for a particular trait was located, and ii) where, more or less exactly, the allele (gene) was located along the length of the chromosome. 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 (\rightarrow); how could this occur? Two obvious ideas suggest themselves: the allele can influence the packaging of the chromosome region, so that the genes’ accessible to other 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:

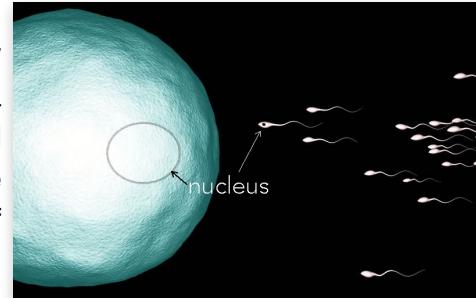
212. Graph, as a function of distance, the likelihood that recombination will disconnect a selected (whether positively or negatively) allele from alleles in surrounding genes.
213. Why might a crossing over event inhibit nearby crossing over events?
214. How can you use the size of a conserved genomic region to estimate time of isolation of a population?
215. 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 (over generational) time, new evolutionary possibilities?

Maternal and paternal effects

One of the implications arising from the dimorphism in gamete size and functional roles is that some alleles preferentially influence oocyte/egg and sperm functions. For example, in a number of organisms, particularly those that develop rapidly and external to the maternal parent after fertilization, most of the gene products and nutrients needed to support early development of the new organism are supplied by the much larger egg (→), accumulated during the course of oogenesis. 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. Similarly, since mitochondria are supplied to the zygote by the egg and not the sperm, defects in the mitochondrial genome cannot be rescued by sperm, even if the sperm is generated by a male with normal mitochondria and a normal mitochondrial genome. Similarly, sperm supply components of the mitotic apparatus to the zygote, fertilization by an aberrant sperm can lead to an early defect in the embryo.



Imprinting:

Another molecular factor that we will consider anon is the fact that for a number of genes, the DNA of sperm and egg are modified differently, in a process known as imprinting.⁴²⁸ Imprinting involves sequence specific modifications of the DNA; they are epigenetic in that they do not alter the gene's 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

⁴²⁸ Genomic Imprinting: <http://learn.genetics.utah.edu/content/epigenetics/imprinting/>

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. As we will see as we come to consider the genetic behaviors of genes (alleles), imprinting complicates things.

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 ~59 million base pairs in length and encodes ~50 genes, while the X chromosome is ~155 million base pairs in length and encodes ~1000 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. 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 the same or very similar). This is the reason for haplo-insufficiency; an amorphic (null) allele can be associated with a dominant phenotype if a single functional copy of the gene is insufficient to produce enough gene product. There would seem to be a need for some form of “dosage compensation”; either genes on the X in males have to be expressed more efficiently or genes on the X in females have to be expressed less efficiently.

The strategy used in humans and other placental mammals is a process known as X-inactivation. At random points during 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). Once the choice of which X chromosome to inactivate is made, it is stable and inherited through mitotic cell division, generating clones of cells with the same X chromosome active and the other inactive. A failure of X-inactivation leads to developmental arrest and embryonic death. While gene expression from the inactivated X is largely 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 chromosome is actively expressed, while the other X chromosome is inactivated (a gene for coat color is located on the X).

The X-chromosomal inactivation system consists of two genes, XIST and TSIX. XIST encodes a functional ~19.3 kilobase long non-coding RNA, that is, not an mRNA, and is expressed only in cells with two X chromosomes – so it is not expressed in males.⁴²⁹ Which of the two X-chromosomes expresses XIST is determined stochastically. When expressed, the XIST RNA associates with regions adjacent to the XIST gene and eventually comes to be 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.

On the other (active) X-chromosome, located on the other strand of the XIST gene is an overlapping gene known as TSIX. As you should be able to explain, the promoter of TSIX is distinct from

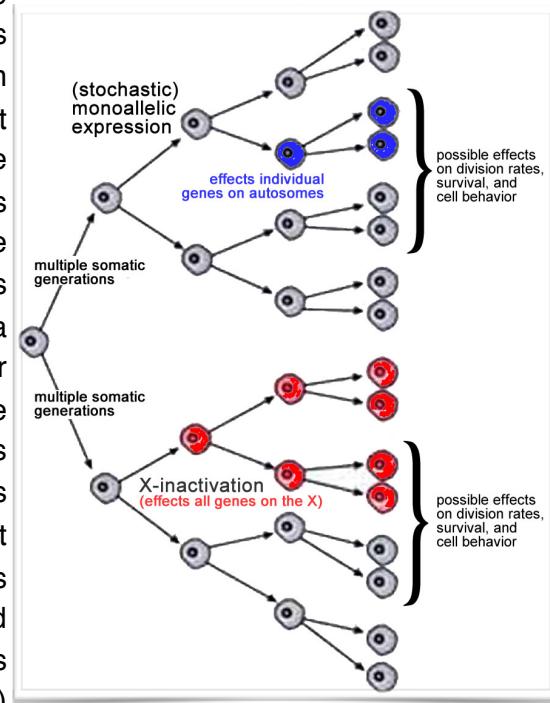


⁴²⁹ X-inactivation-specific transcript ([OMIM](#))

that of XIST. The TSIX gene encodes a ~40 kilobase long non-coding RNA 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 its inactivation. Together the XIST/TSIX system insures that one and only one of the two X chromosomes is active.

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 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 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. In a typical diploid cell, one gene may be active while the other copy of the gene, on the homologous chromosome may be inactive, due to various stochastic events.⁴³⁰ In some cases of stable monoallelic expression there is what is known as somatic selection, which we will return to (→). Given that there are two alleles, when they are different which is expressed may differentially influence cell growth and division, or even cell 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. We will return to this subject in the context of cancer evolution and brain development.



Questions to answer:

216. What does it mean to be mosaic for an allele?
217. Why do males and females differ in the traits they display?
218. Why do males and females differ in the display of phenotypes associated with genes on the X chromosome?
219. Can you provide a plausible mechanism to explain why (autosomal) random monoallelic expression occurs?
220. How can monoallelic expression impact an organism?

Question to ponder:

- Under what conditions might mono-allelic (autosomal) gene expression be beneficial?

⁴³⁰ [Monoallelic Gene Expression in Mammals](#)

Chapter 14: Generating mutations and becoming alleles

In which we consider how mutations appear and become alleles within a population. Distinguishing between randomness and purpose.

We are far enough along to recognize that beginning with a particular genome, any change in that genome, whether due to errors in its replication or environmental (chemical reaction or radiation induced) damage that goes unrepaired, that is, that is not returned to the original state, 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 an allele - a variant within a population. If it lies outside of a gene, it becomes a 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.⁴³¹ Here we distinguish between mutations in the germ line (leading to eggs and sperm) and in somatic cells, the cells of the body. 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 more frequently (~10X) than other dinucleotides. In addition to single nucleotide changes, there are also mutations that involve small insertions and deletions, known collectively as indels; these are defined to be less than 20 base pairs (bps) in length, and are distinguished from larger changes, known as structural variants (> 20 bps).⁴³² It has been estimated that each generation sees the addition of ~3 indels and ~0.16 structural variants in the germ line of each person. In addition, there are copy number variations (CNVs), these can lead to changes in the number of copies of a particular gene or sequence.⁴³³ Many mutations are associated with the process of cell division. Mutations occur more frequently in the soma because there are more cell divisions involved, trillions in the human. Similarly, there are fewer cell divisions involved in the generation of oocytes in females than 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 with the death of the host. The current estimate is that the chance of a de novo germ line mutation in humans is $\sim 1 \times 10^{-8}$ per base pair per generation (remember the human genome contains $\sim 6 \times 10^9$ 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 the first criterion is that it does not have a dominant lethal phenotype; why dominant? In a diploid organism a new mutation will involve only one of the two genes

⁴³¹ see: [The origins determinants, and consequences of human mutations](#)

432 Indels

433 Copy Number Variation

present; for it to have a phenotype, it needs to be dominant over the other allele present. Of course this is not the case in prokaryotes, which are typically 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. Again, this assumes that the allele does not produce a lethal phenotype in the gamete or the early zygote, since where and when a gene is expressed, has a lot to do with the phenotypes it is associated with.

A non-lethal, non-dominant mutation has a chance to become an allele within a population, but first it has to avoid elimination. 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, it is possible that the gamete carrying it fails to fuse with another gamete, if so, this new mutant allele would be lost. This is a version of genetic drift. Similarly, the mutant allele may make it into the next generation even if it is deleterious (although not too deleterious), again just by chance.

If a mutant allele survives these early events, it comes to be referred to as an allele when it is found to occur in >1% of the population. Mutations that occur outside of a gene become what are known as polymorphisms, and are part of the differences between organisms, although unless they lie within an enhancer, are unlikely to influence gene expression or visible phenotype. The difference between allele and polymorphism lies in the ability to recognize what is, and what is not, a gene, something that can be tricky. The genetic variation within a population reflects its past history (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 had no clear understanding of exactly where genetic variation came from, an important question that arose early in the history of evolutionary theory was whether mutations (a prime source of phenotypic and genetic variation) associated with the evolution of new and complex traits – such as the eye – were the result of random (stochastic) events or were they somehow directly and purposefully generated in response to the needs of the organism. In the absence of a clear understanding of how genetic information arose in a population or how it was passed from generation to generation, there was really no way to distinguish between Darwinian (random variation + selection) and Lamarckian (adaptation based on need) evolutionary mechanisms, although Lamarckian mechanisms seemed kinder.⁴³⁴

To understand how this question was resolved, we will consider a classic experiment, known as the Luria-Delbrück experiment after the researchers, Salvador Luria (1912-1991) and Max Delbrück (1906-1981).⁴³⁵ The study was published in 1943, before DNA was recognized as the genetic material and well before anyone understood how genetic information was stored.⁴³⁶ In their study, Luria and Delbrück examined bacterial resistance to viral infection. The bacteria they used could be infected and killed by a specific type of bacteriophage (phage for short). Some of the mutations that arise

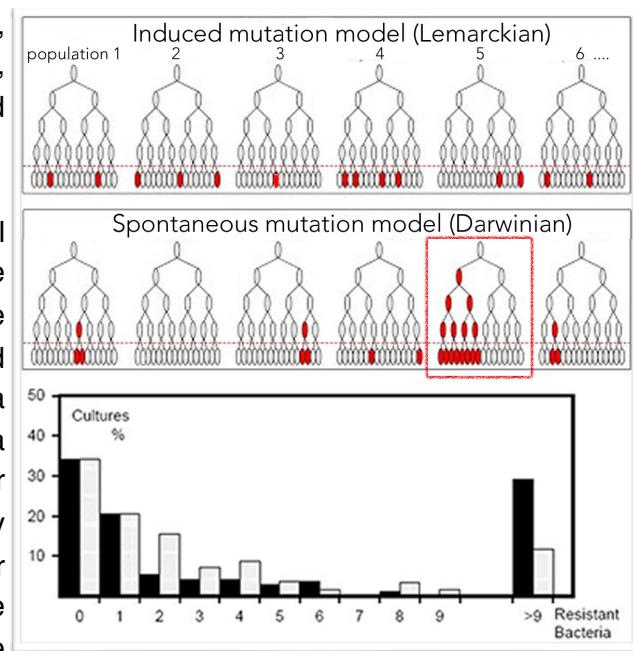
⁴³⁴ This led to what was known as the “[Eclipse of Darwinism](#)”; biology emerged from this “darkness” with the development of an understanding of genes and genetic mechanisms to produce what became known as the “Modern Synthesis”.

⁴³⁵ [Luria–Delbrück experiment](#)

⁴³⁶ Mutations of bacteria from virus sensitivity to virus resistance: <http://www.genetics.org/content/genetics/28/6/491.full.pdf>
biofundamentals™ Klymkowsky & Cooper - copyright 2010-2018 version: Tuesday, January 16, 2018 254 of 315

spontaneously in the bacteria rendered them, and their off-spring, resistant (immune) to phage infection. The question Luria and Delbrück asked was, are phage resistance mutations appearing randomly all of the time in bacteria or is it the presence of the virus that induces their appearance in response to the bacteria's "need" to be immune. Is immunity learned or lucky?⁴³⁷ If the mechanism is random, 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. On the other hand, if the generation of phage resistance mutations is an adaptive process, then we would expect that the frequency of phage 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 (\rightarrow). In a sense, as proposed by Darwin, evolution involves random mutations in individuals, whereas a Lamarckian mechanism involves induced responses by the population as a whole.⁴³⁸

Luria and Delbrück started a number of bacterial cultures to which they then added enough virus (at the time of the horizontal red line) to kill every sensitive bacterium. They then plated out the culture and counted the number of phage-resistant bacteria present, each of which can grow up into a macroscopic (asexual) clone, a colony.⁴³⁹ The number of such cells in a culture will reflect when in the history of the culture the resistance mutation appeared; for example, if it appeared early in the history of the culture, as in the red-boxed culture (\uparrow) in the spontaneous mutation model, it would be common, whereas if it appeared late, it would be rare. The two models (induced/Lamarckian versus spontaneous/Darwinian) make dramatically different predictions. In the induced/Lamarckian model, the variation 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, depending upon whether and when the mutation occurred, a chance event. If the mutation occurs late, most bacteria will be killed (population 2); if the mutation occurs early (as in population 5, boxed in red) there will be many resistant bacteria present. In the lower panel, Luria and Delbrück calculated what they expected from their experiment if the spontaneous/Darwinian model occurred - their observed



⁴³⁷ As we will see later on, there are molecular mechanisms, such as the CRISPR CAS9 system that can learn and lead to acquired immunity.

⁴³⁸ 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/>

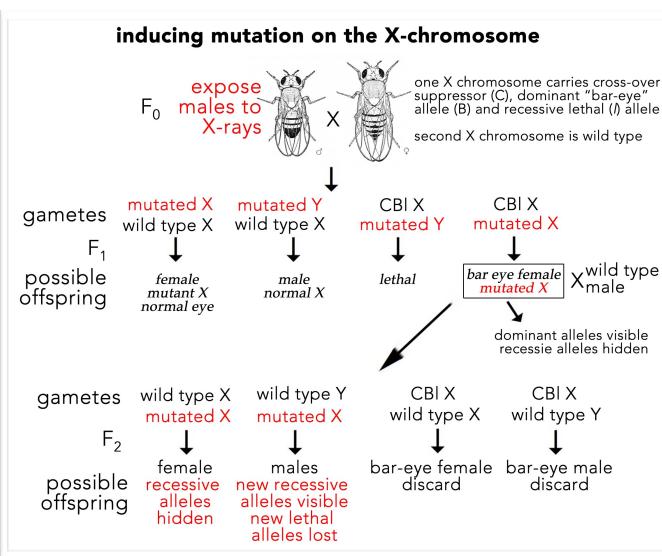
⁴³⁹ The logic and details of their experiment are the subject of this virtual lab lab on the [Luria-Delbrück experiment](#)

results (black bars) matched this prediction, allowing them to conclude that, at least in this system, mutations were occurring 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.⁴⁴⁰ The ability to control mutation rates occurs within the vertebrate immune system, through a process known as somatic hypermutation.⁴⁴¹ This process is involved in the maturation of the 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 to generate uracils, which when repaired generate an A:T base pair, replacing the original C:G base pair. 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”.⁴⁴²

Generating mutations experimentally

When we think about a particular trait or a behavior, a specific phenotype, we often want to know how many different genes are involved in producing that phenotype. One approach to begin to answer that question is to determine how many genes can be mutated so as to disrupt the formation of that phenotype. Such a search for mutations that disrupt a particular phenotype is known as a “forward genetic screen”, and has, historically, been used to identify the molecular components involved in the process. The first step in such a screen is to generate mutations. Waiting for naturally occurring mutations 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. Muller, who we have met before, was the first to create a mutation using X-rays (→).⁴⁴³ He examined the generation of mutations on the X chromosome of the fruit (or more accurately vinegar) fly *Drosophila melanogaster*, an organism chosen in part because of its small size (which



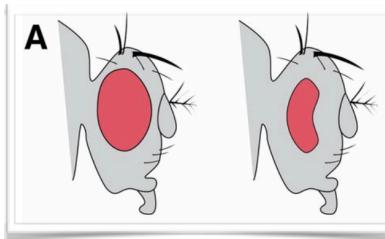
⁴⁴⁰ A trade-off between oxidative stress resistance and DNA repair plays a role in the evolution of [elevated mutation rates in bacteria](#)

⁴⁴¹ Somatic hypermutation: [wikipedia](#)

⁴⁴² Two levels of protection for the B cell genome during [somatic hypermutation](#)

⁴⁴³ [Hermann J. Muller \(1890-1967\) demonstrates that X rays can induce mutations](#)

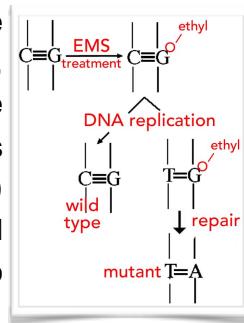
allows lots of animals to be raised in a limited space), rapid life cycle, and the large number (~400) offspring that are produced by a single female after a mating. In previous studies, he had isolated a



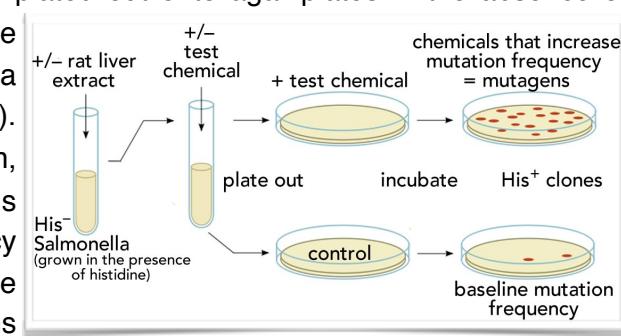
version of the X-chromosome, known as CBI, that carries a dominant allele that produces bar eyes (\leftrightarrow), a recessive lethal mutation, and a large inversion (a flipped region) in the chromosome; the presence of this inversion generates embryonic lethal mutations if recombination occurs within the inverted region (think about how a chromosome with a large inversion will interact with a normal chromosome during meiosis.)

Muller took wild type male flies and irradiated them, so that mutations were induced in their testes, producing sperm with mutations. He then mated females carrying the altered CBI X-chromosome with the irradiated males (you should be able to explain why he could not have used males carrying the CBI chromosome). Based on the markers present, he could identify females that carried the CBI X chromosome and a mutated X chromosome from an irradiated male (\uparrow previous page). When these first filial generation (F_1) females were mated with a wild type male, 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 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)(\rightarrow). EMS reacts, through an esterification reaction, with guanosine residues in DNA, modifying them by 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*.⁴⁴⁴ They began by using a strain of *S. typhimurium* that carries a mutation 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 (\rightarrow). 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



⁴⁴⁴ Ames test (wikipedia)

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 adopt it to organisms such as humans, was the recognition that many chemicals that you might be exposed to are metabolized in the liver. Such reactions generate related chemicals, which 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.⁴⁴⁵

Questions to answer:

221. How would increasing the mutation rate influence the outcome of the Luria-Delbrück experiment.
222. What are the advantages (for a geneticist) for choosing an organism with hundreds of offspring per mating event?
223. What is the advantage of studying traits that alter non-essential structures?
224. Why is it not possible to identify every gene involved in the formation of a complex trait by a simple mutagenesis approach?
225. What is responsible for the baseline mutation frequency (in the Ames test)?
226. 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?

Longer term mutation / evolution studies

We can see the spontaneous mutation model applies throughout the biological world, where ever we look mutations appear to arise randomly. If they can 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 are often lost because initially they are extremely rare within the population, while mildly deleterious mutations can become fixed.

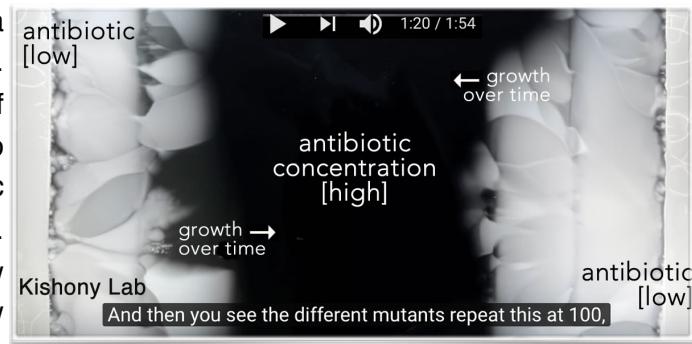
To study such evolutionary processes in a laboratory setting is not easy, but the now classic example of such a study has been carried out by Richard Lenski (p. 1956) and his associates; they have been growing 12 originally identical populations of the bacteria *Escherichia coli* for more than 25 years and > 60,000 generations.⁴⁴⁶ One, of many, characteristics of *E. coli* that distinguish it from other

⁴⁴⁵ "All substances are poisons; there is none which is not a poison. The right dose differentiates a poison..." Paracelsus [[link](#)]

⁴⁴⁶ *E. coli* long-term evolution experiment: [wikipedia](#) and the Lenski lab's *E. coli* [Long-term Experimental Evolution Project site](#)
biofundamentals™ Klymkowsky & Cooper - copyright 2010-2018 version: Tuesday, January 16, 2018 258 of 315

bacteria is that it cannot metabolize citrate in the presence of O₂. 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 O₂ in one of their cultures; a beneficial evolutionary adaptation, since it provided a previously un-utilized energy and carbon source.⁴⁴⁷ 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. Molecular analyses established that it 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 mutant cells’ ability to metabolize citrate.⁴⁴⁸ 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, given 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 new traits via mutation and selection.⁴⁴⁹

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”) with a gradient of antibiotic on it (→). 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.⁴⁵⁰



An important point to recall about the bacterial evolution studies is that these organisms are reproducing asexually, as clones. That means that they have no issue with interbreeding with other organisms in the population, but it also means that (in the absence of horizontal gene transfer) all necessary mutations need to occur 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 they are useful, they will need to appear again, through mutation and selection (or transferred horizontally). In sexually reproducing (diploid) organisms, the various stochastic events involved, including gamete formation and fertilization, can lead to the loss of

⁴⁴⁷ see [Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*](#).

⁴⁴⁸ see [Genomic analysis of a key innovation in an experimental *Escherichia coli* population](#).

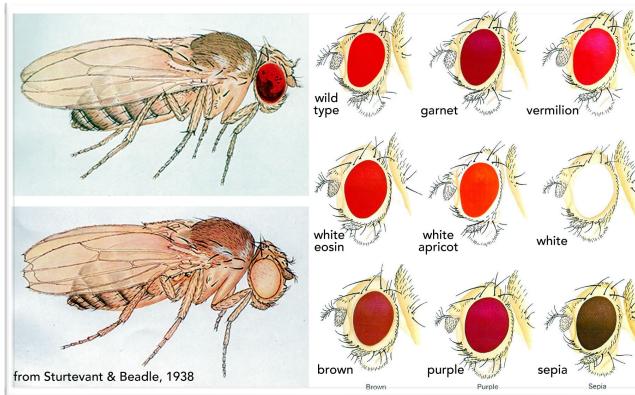
⁴⁴⁹ The evolution of citrate metabolizing *E. coli*: the “[Lenski affair](#)”

⁴⁵⁰ Baym et al., 2016 [Spatiotemporal microbial evolution on antibiotic landscapes](#).

alleles (by genetic drift). Also the importance of reproductive barriers between subgroups within the original population, adapting to different ecological niches, can impact evolutionary processes.

Forward and reverse genetics

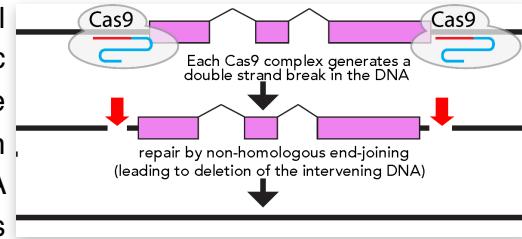
Originally, genetic analyses were carried out through what are known as forward genetics. Forward genetics involves the generation of mutations, essentially at random, and then identifying individuals carrying mutations that disrupt a particular process of interest. As an example, consider eye color in the fruit fly *Drosophila melanogaster* (↓). Eye shape and color are experimentally accessible because a



Drosophila embryo can develop into a fertile adult without an eye; this makes it possible to identify mutations (alleles) that alter the eye but allow other aspects of embryonic development to occur (more or less) normally. 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 eye formation, 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 rarely if ever complete, that is, they do not identify every gene involved in a process.

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 mutate a particular gene, the mechanisms involved are largely beyond the scope of this course. One exception is the recently developed CRISPR CAS9 system, which is one of a number of anti-viral infection systems found in bacteria and archaea.⁴⁵¹ 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) – a sequence that is long enough to (often) occur once and only once within the genome of an organism, even an organism with a genome of more than a billion base pairs, such as humans. This gives an extremely high degree of specificity to the system. After the double-strand break is made, host cell DNA repair systems act to join the two ends of the DNA molecule back together again, but this joining is not accurate – base pairs can be lost or added, generating a mutated form of the original DNA sequence. More and more sophisticated genetic manipulations can be generated using variants of the CRISPR Cas9 system; for example, regions of a gene can be deleted by using pairs of gRNAs (→). If the gRNA sequence is present in both alleles of a gene, both alleles



⁴⁵¹ over-view reference for the Crispr cas9 system: [wikipedia](#).

can be mutated at the same time. Finally, 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. One way to control the effects of a mutation is to use various molecular strategies to express the CRISPR CAS9 system at specific times and in specific cells within a multicellular organism, something that will be expanded upon (we expect) in later classes.

Questions to answer:

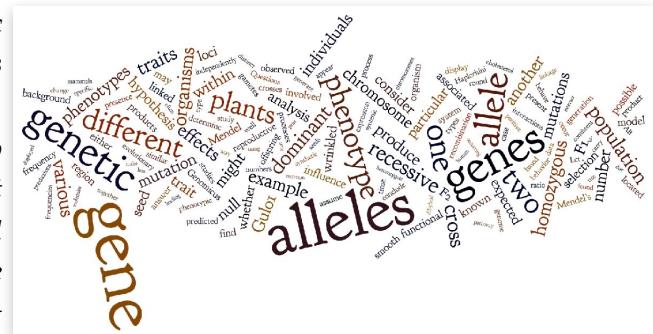
227. How can a “predisposing mutation” influence the possible directions of subsequent evolution?
228. In the antibiotic resistance video (watch!), why is there often (but not always) a delay before the bacteria grow into a region of higher antibiotic resistance?
229. How would the presence of horizontal gene transfer impact the megaplate experiment?
230. How would an evolutionary sweep effect a human population?

Question to ponder:

- How would evolution be altered if the mutations (alleles) were induced rather than selected?

Chapter 15: Becoming Mendelian: analyzing alleles in terms of phenotypes & pathways

In which we consider the contributions of Gregor Mendel, namely the realization that distinct and stable genetic elements influence specific traits. These genetic elements behave in predictable ways during sexual reproduction. The behavior of chromosomes during meiosis leads to Mendel's rules of allele segregation and independent assortment. We come to recognize that the traits Mendel studied reflect a unique 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 origins of genetics, it is worth considering some of the biases imposed by the way that Gregor Mendel did his work, these reflect the realities of science – understanding does not appear fully formed, like religious revelation, rather it is built up by insights, some productive and others a distraction. Subsequent observations and experiments lead to the recognition of the implications and limitations of original ideas (tentative hypotheses and working models), and drive their modification, so as to explain a larger number of observations more and more mechanistically. To make genetic behaviors intelligible, Mendel purposefully grew plants (peas) that displayed highly reproducible and distinguishable traits. These were traits (phenotypes) that were not dramatically influenced by environmental or genetic background effects, but were due to alleles of a single genetic loci, single genes. Moreover, these genes were unlinked to one another – they were located on different chromosomes, and so were segregated independently of one another during meiosis. Finally, he used traits (as we will see) that were well behaved in terms of their interactions with one another (little or no interaction), and an experimental organism that produced high numbers of progeny that could be obtained from a well controlled matings. Moreover, the traits displayed clear dominant-recessive behaviors with respect to one another. It is not that Mendel knew anything about chromosomes and molecular mechanisms, it is just that his choices made the data he obtained intelligible.⁴⁵²

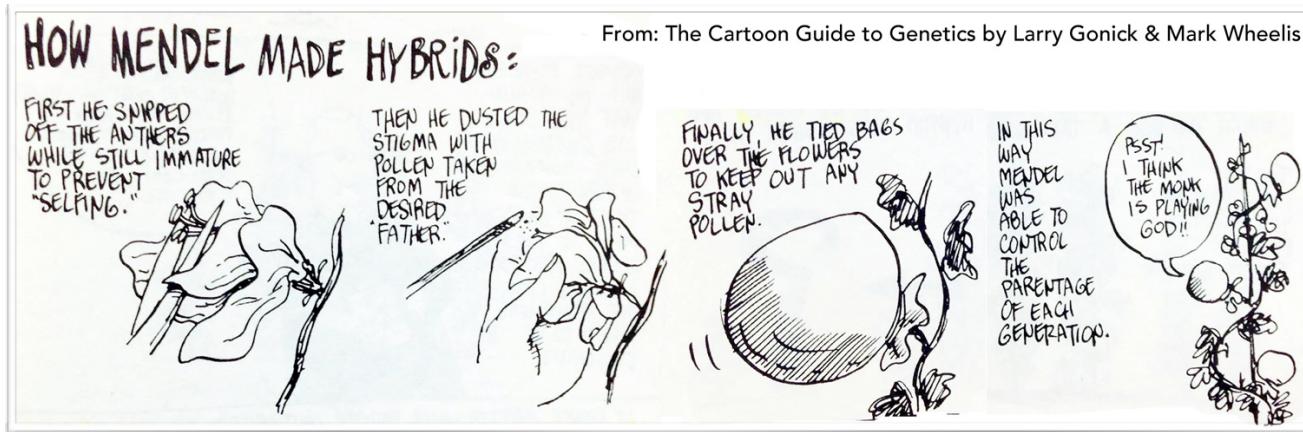
At the same time, it is worth recognizing explicitly that most alleles do not behave in these ways. What phenotypes a particular allele is associated with are often partial, difficult to recognize, and influenced by the genetic background of the organism. In this context it is important to remember that many laboratory studies (including Mendel's) are carried out in in-bred backgrounds, that is, all of the organisms in the study share a common genetic background (overall genotype). Such genotypic homogeneity is an artifact of the way such experiments are conducted; natural populations display much more background genotypic variation - there are many more different alleles present. Such background variation influences the phenotypes associated with a particular allele, whether hetero- or homozygous. Consider a dominant allele; even though the allele is present within the genome of the organism, the associated trait may vary - this variation is characterized in the terms expressivity and penetrance. So, what does that mean exactly? Variable expressivity refers to the observation that even

452 virtuallyGenetics: Mendel lab

in the presence of the associated (dominant) allele, the phenotypic trait can vary. As an example, consider a pea; is each pea really wrinkled to exactly the same extent, or do they vary – are some a little more or less wrinkly, although not smooth? Such behavior indicates variable expressivity. Similarly, it is possible that out of 100 individuals (or peas) that carry a particular dominant allele, all of them display the trait associated with that allele, but only in an in-bred background. If in a wild, out-bred background, only 50% of the organisms carrying the dominant allele express the trait, it is said to be incompletely penetrant, presumably due to various other factors, including different sets of suppressor and enhancer alleles, located in other genetic loci, other genes.⁴⁵³

Hidden alleles within a population

Except for those traits that are fully (100%) penetrant and expressive, for all types of alleles their effects on phenotype can vary dramatically depending on genetic background, that is the rest of the alleles present in the organism – they can disappear and reappear within a family lineage (a family tree or pedigree). 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 not have revealed the simple rules of inheritance that he discovered.



In his studies, Mendel carefully identified dominant and recessive alleles of a number of unlinked genetic loci with completely penetrant and expressive phenotypes, working primarily with the garden pea *Pisum sativum*.⁴⁵⁴ These plants can reproduce by either self-fertilization, that is sperm (pollen) and eggs are derived from the same plant, or by out-breeding, which involves sperm (pollen) and eggs from different plants. He developed or identified pure-breeding lines, that is plants that produce offspring with phenotypes similar (identical) to their own when bred with themselves. One reason to choose the pea was that it is possible to control who breeds with whom (↑); the investigator can block self-fertilization. In such studies, the parents are known as the F₀ generation and their offspring are known as the F₁.

⁴⁵³ here is a particularly relevant recent study: [Genetic background limits generalizability of genotype-phenotype relationships](#)

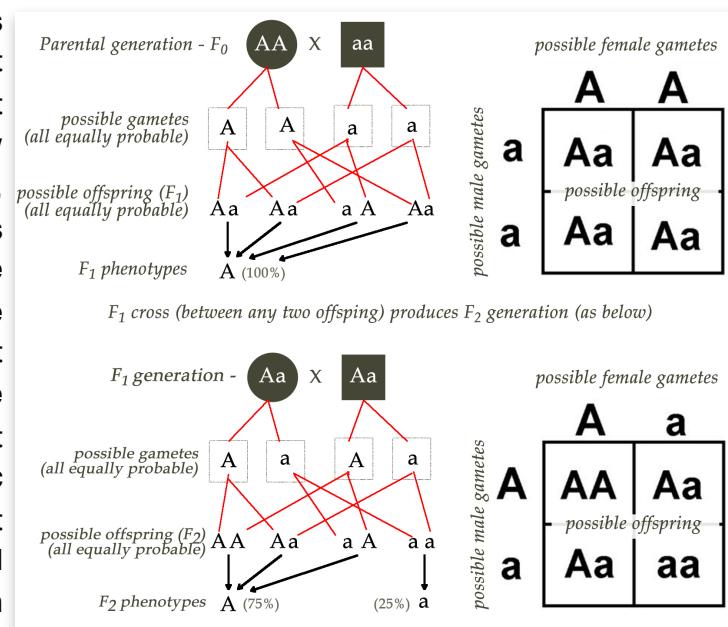
⁴⁵⁴ It might be worth considering the distinction between a study and an experiment. In an experiment, the system is subject to some perturbation, and we examine how the system responds. Typically the experiment begins with a hypothesis, a guess on how a particular perturbation, which we might think we understand, will influence 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 under study might act under different conditions (another observational study) or how a perturbation (an experimental study) might alter the system's behavior. Our prediction of the outcome is known as the null hypothesis - and we generally examine the data collected to determine whether the null hypothesis is supported or not.

generation. The requirement that the lines breed true means, in practice, that they are homozygous for the alleles controlling the trait under consideration. For example, when the F_0 parents are homozygous for either recessive or dominant alleles of a genetic locus, all of the gametes they produce carry the same version of the allele.

Now consider what happens in what is known as a monohybrid cross. In such a cross we are looking at a single trait, presumably controlled by which alleles are present at a particular genetic locus (gene). We assume that we have inbred lines that display either of two different versions of the trait, for example either smooth or wrinkled peas. What do the peas of their offspring look like? Turns out they are round. These F_1 plants must have one allele from each homozygous parent, and these alleles must be different (understand why?); this leads us to assume that the allele associated with the smooth peas' phenotype is dominant to the (recessive) allele associated with the wrinkled pea phenotype. We can then show that all of the F_1 organisms are genetically similar for this genetic locus. How? We can cross each plant to itself and to any of its siblings. Assuming that the number of offspring is large enough, we will find that 75% of the F_2 plants have smooth seeds, while 25% have round seeds, a 3:1 ratio.

An obvious question is are all of the round and wrinkled seeded F_2 plants similar? The answer for the wrinkled seed plants is yes, if we self-cross these plants, or cross them to another wrinkled seed plant, only wrinkled seed plants are produced in the F_3 generation. In contrast, we discover two types of round seed F_2 plants. To reveal these classes, we cross the plants to one another, or to a true breeding wrinkled seed plant. We discover that self-crosses lead to two different outcomes. One third of the smooth seed plants generate only smooth seed plants upon self-crossing, while two-thirds act like F_1 plants, producing both smooth and wrinkled seed plants in a 3 to 1 ratio (75% to 25%). If we cross the round seed F_2 plants to a true breeding wrinkled seed plants, we again find two different outcomes. Again, one third of the F_2 smooth seed plants generate only smooth seed plants upon self-crossing, while two-thirds of the F_2 smooth seed plants produce smooth and wrinkled seed plants in a 1:1 ratio.

We have approached these studies already knowing that peas are diploid, that genes exist in multiple forms (alleles), and that chromosomes segregate independently during meiosis. Mendel deduced all of these "laws" from his studies (rather amazing!). Moreover, he found that the same rules held for a number of distinct traits, including plant height, green/yellow seeds, purple/white flowers, flower position, seed pod shape, and seed pod color. From this type of data, Mendel deduced that: each of these phenotypic traits were controlled by 1) a single genetic element that exists in two different forms, alleles, where one is dominant to the other, recessive allele; 2) that each plant contained two copies of these genetic elements (which we call genes), and 3) that upon breeding, the parent passes one and only one of its alleles to its offspring, but which allele is passed is random (\rightarrow).



Questions to answer:

231. Why was it critical for Mendel's studies to be able to control crosses between individual plants?
232. What led Mendel to be able to discover recessive alleles?
233. Describe, in terms of meiotic behaviors, how the results of a monohybrid cross are produced.
234. 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 answer:

- How are backcross to homozygous recessive individuals informative? Are similar backcrosses to homozygous dominant individuals useful?
- How does not determine, in practice, that a homozygous recessive individual is homozygous recessive?

Chi square analysis, hypothesis testing, and dealing with numbers that are less than infinity

One limitation of Mendel's work is associated with the limited number of plants he could examine. The various ratios he predicted are expected to be true only when the number of individuals examined becomes large, at 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. This is a situation similar to one we considered previously in the case of other stochastic processes, since which gametes contain which alleles and which fuse with one another are both stochastic processes.⁴⁵⁵ Consider the general question, how many rolls of a die would you need to perform to convince yourself that the die is fair? While the stochastic nature of meiosis and fertilization does not effect the F₁ generation of a cross between homozygous dominant and recessive plants, in which all offspring are predicted to be the same (heterozygous), it certainly influences the 3:1 ratio of phenotypically dominant to recessive plants predicted to occur in the F₂ generation. How do we evaluate whether our observations are consistent with our model, or contradict it, and so force us to abandon or substantially revised it?

The answer is a statistical test known as a χ^2 analysis.⁴⁵⁶ Such an analysis uses the equation (↓) together with two other concepts: degrees of freedom and null hypothesis.⁴⁵⁷ If we are testing a model that makes a mathematically precise prediction as to the frequency of the various phenotypic classes observed, our null hypothesis is that our model is correct, that there will be no significant difference between the observed data and the predicted data. We will then try to determine whether our data are consistent with the null hypothesis. Remember, we cannot prove anything, we can only conclude that the data we observe is consistent or inconsistent with our prediction, our null hypothesis.

$$\chi^2 = \sum_{n} \frac{(observed - expected)^2}{expected}$$

To define the degrees of freedom, we need to know how many independent variables there are. In our two phenotype system (wrinkled and round), we assumed that all individuals have either one or the other phenotype, if we know the number of individuals involved and the number of either phenotype, we

⁴⁵⁵ It is similar to the question of which unstable isotope atom will decay next.

⁴⁵⁶ Here is an alternative presentation from [GENETICS AND GENE PROBLEMS](#)

⁴⁵⁷ chi square tutorial: http://www.radford.edu/rsheehy/Gen_flash/Tutorials/Chi-Square_tutorial/x2-tut.htm

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 discussion⁴⁵⁸, we take an observation to be consistent with the null hypothesis if it can be expected to occur by chance at less than 1 time out of 20 (0.05); otherwise we have a good case to reject the “null” hypothesis.

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 (\rightarrow). The prediction of his model is that round : wrinkled seeds in the F₂ will occur in a ratio of 3:1. In his study, he 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

Results of all of Mendel’s monohybrid crosses

Parental phenotype	F ₁	F ₂
1. Round \times wrinkled seeds	All round	5474 round; 1850 wrinkled
2. Yellow \times green seeds	All yellow	6022 yellow; 2001 green
3. Purple \times white petals	All purple	705 purple; 224 white
4. Inflated \times pinched pods	All inflated	882 inflated; 299 pinched
5. Green \times yellow pods	All green	428 green; 152 yellow
6. Axial \times terminal flowers	All axial	651 axial; 207 terminal
7. Long \times short stems	All long	787 long; 277 short

Griffiths et al., 2000

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 (Σ) of these two numbers = 0.0595 . To determine whether these observations are consistent with our null hypothesis, we need to consult a χ^2 probability table (\downarrow). The

higher the χ^2 value the more likely the difference between observed and expected data is

Selected percentile values of the χ^2 distribution						
df*	.99	.95	.50	.10	.05	.01
1	.000157	.00393	.455	2.706	3.841	6.635

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 the observed numbers are consistent with, but by no means proving, that our model is “true.” In fact, there have been suggestions that Mendel’s observed numbers are too good, to close to what would be predicted.⁴⁵⁹ 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.

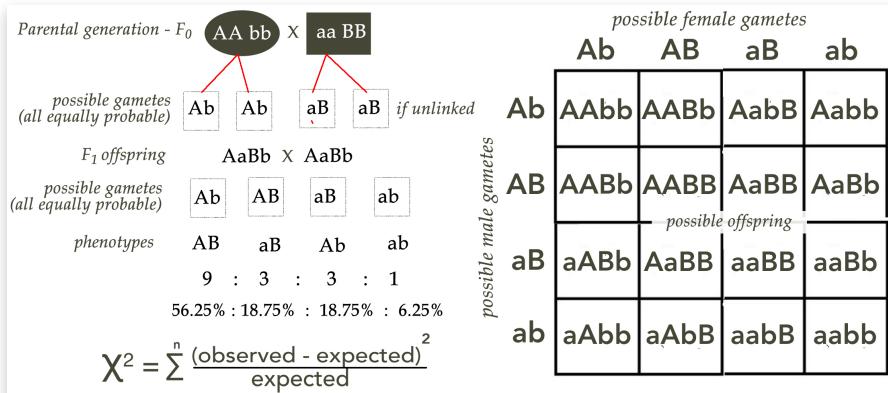
Dihybrid crosses and linkage

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 begin with a null hypothesis, namely that the two traits behave, in meiosis, independently, that is they do not interact with one another. Assume that we begin with two lines that breed true for these traits (\downarrow). As before, each

⁴⁵⁸ [Statistical errors](#) and Colquhoun. 2014. [An investigation of the false discovery rate and the misinterpretation of p-values](#)

⁴⁵⁹ see [On Fisher's Criticism of Mendel's Results With the Garden Pea](#)

parental F₀ organism can produce only one type of gamete, and all F₁ 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 F₁ individuals (F₁ × F₁). Assuming that the two genetic loci behave independently, then each F₁ individual can produce four different types of gametes, and these gametes can fuse (randomly) with gametes from the other F₁ 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.⁴⁶⁰



There are 16 possible combinations of these alleles in the F₂ generation, of these 9 display a dominant:dominant phenotype: AABB (1), AABb (2), AaBb (4), AaBB (2); two 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. This produces F₂ progeny in the 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 numbers observed for the various phenotypic classes of the F₂ offspring do not match our expected values - what can we conclude? The simplest conclusion, and one not made by Mendel (because he excluded such traits), was that the genetic loci involve in these traits are somehow linked together, and only occasionally separated during the process of meiotic recombination.⁴⁶¹ Let us consider one such example, we generate a dihybrid F₂ generation from AB phenotype F₁ offspring (the result of a AB × ab cross), and observed the following outcome :

observed
AB : aB : Ab : aa
981 : 72 : 86 : 964
552 : 394 : 394 : 131
expected

We carry out a χ^2 analysis and obtain a value of 6219. A quick look at the probability table (\downarrow) confirms our suspicion, namely that our null hypothesis, that the genes are

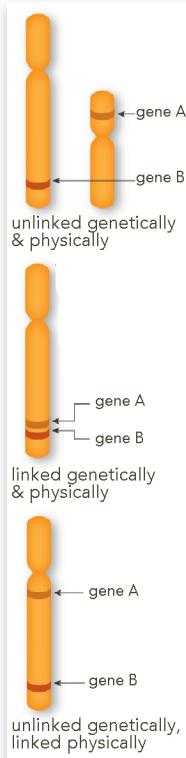
unlinked, is rejected. We are forced to assume that the genes are linked, and we can now generate an estimate of how

closely linked they are on the chromosome.

⁴⁶⁰ Who was this Punnett fellow? see [Reginald Punnett](#)

⁴⁶¹ Why did he miss 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 the actual situation.

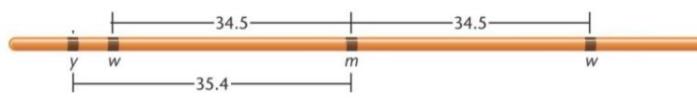
We know from our cross that the parents (F_0) were AB and ab, and that there chromosomes were AB and ab respectively. If the A and B genes are located on the same chromosome, we can assume that, in the absence of recombination, only [AB] and [ab] gametes would be generated and that all F_1 organisms were [AB][ab], with the brackets indicating that the alleles are linked on the same chromosome. Again, in the absence of recombination, we can assume that F_1 organisms can produce only [AB] and [ab] gametes. To produce aB or Ab gametes, we must assume that a recombination event occurred 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%. Recombination frequencies are typically referred to as map units or centimorgans, named in honor of the early geneticist Thomas Hunt Morgan.⁴⁶² A 7.5% recombination frequency equals 20 map units or 50 centimorgans.



We should note that 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, or sets of three loci. 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 centromeric region of the chromosome, centimorgans do not directly or consistently convert into DNA lengths in base pairs. That said, on average (in humans) a 1 centimorgan distance between genetic loci corresponds to ~1 million basepairs of DNA, 1 megabase (abbreviated Mb). From an evolutionary standpoint it is worth remembering that linkage can influence the inheritance of alleles (see above); the closer two genetic loci (and their alleles) are to one another the longer (the more generations) it will take recombination to separate them, so that they are inherited independently.

MUTANT	centimorgan	WILD TYPE
Short aristae	0	Long aristae
Black body	48.5	Gray body
Cinnabar eyes	57.5	Red eyes
Vestigial wings	65.5	Normal wings
Brown eyes	104.5	Red eyes

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. In this example (↑), you should be able to predict what you would expect from a cross if the *w* gene were located upstream or downstream of the *m* gene. In an era (like today) of full genomic sequence data, it



⁴⁶² Thomas Hunt Morgan

is generally easier to use web based tools such as Genomicus [[link](#)], since finding the organisms needed to carry out a multigenerational cross, particularly in humans, can be challenging.

Questions to answer:

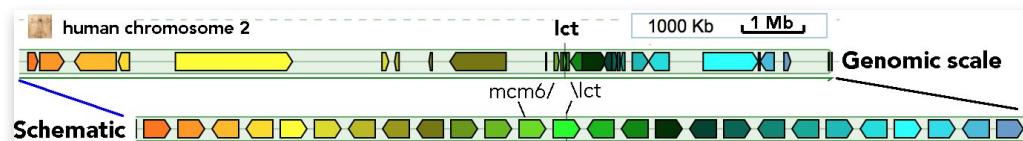
235. What does it mean if the null hypothesis is not supported?
236. A dihybrid cross produces offspring that do not fall into the expected 9:3:3:1 distribution, what kinds of conclusions can we make?
237. In a dihybrid cross, the individuals that are homozygous for both recessive alleles are absent, what might you conclude and why?
238. What might you, at least tentatively, conclude if expected individuals (from a dihybrid cross) that were heterozygous for both dominant alleles, failed to appear?
239. 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?

Using web-based bioinformatic tools: Genomicus

In Genomicus (to be illustrated in class), the user inputs a gene name (rest assured, we will talk more about gene names soon), and the system displays the gene in its genomic context (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”.⁴⁶³ In the example below (↓), 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. 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, we find the trait “adult lactose tolerance”. Adult lactose tolerance is associated with a failure to turn off expression of the *Lct* gene in adults.⁴⁶⁴ 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.⁴⁶⁵ 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 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

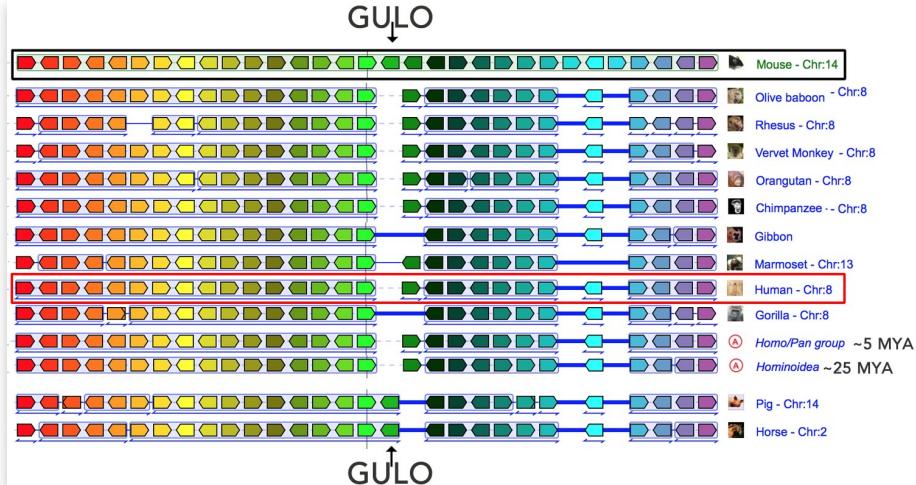
⁴⁶³ [Genomicus update 2015: a genome-wide perspective to multispecies comparative genomics.](#)

⁴⁶⁴ [Lactose digestion and the evolutionary genetics of lactose persistence](#)

⁴⁶⁵ [World-wide distributions of lactase persistence alleles and the complex effects of recombination and selection.](#)

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 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 (we will work with Genomicus in class). 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 chromosome 2), by inputting 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 have already introduced, the *Gulo1* gene. Recall (page 71) that, in contrast to most vertebrates the *Haplorrhini* or dry nose primates are dependent on the presence of vitamin C (ascorbic acid) in their diets. A plausible scenario for this situation is that a functional L-gulonolactone oxidase (*Gulo1*) gene was lost due to mutation in the last common ancestor of the *Haplorrhini*. The remains of the *Gulo1* gene found in humans and other *Haplorrhini* genomes is mutated and 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 phylogenetic (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 that was 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 *Haplorrhini* primates – Whoa!! what gives? The explanation, it turns out, is rather simple (see [link](#)). Because of mutation there is no functional *Gulo1* gene in any *Haplorrhini* primate. But the *Haplorrhini* 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 *Haplorrhini* (↑). 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 *Haplorrhini* primates. Our Genomicus analysis enables us to make a number of readily testable predictions. A newly discovered *Haplorrhini* 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).⁴⁶⁶ Such an analysis also indicates that gene can move around within the genome.

Questions to answer:

240. If you were to add a mouse *Gulo1* gene to a human genome, where would you put it and why?
241. If a gene is missing from a syntenic region, what might have happened to it?

Question to ponder:

- Given what you know about meiosis, how would the deletion of a gene influence the genotypes of the gametes; what about a translocation, in which part of one chromosome was moved to another chromosome?
- What would happen if the homology domains on the Y chromosome were deleted?

Genetic complementation

When we make mutations in various traditional ways (such as X-rays or exposure to mutagenic chemicals – see above), the organisms carrying these mutations can be selected for further study based on their phenotypes, typically chosen because they effect a particular process. The first aspect of such a study is to carry out various “back-crosses” in order to remove unwanted mutations; remember, mutation is random, and generally carried out so as to produce hundreds of mutations within the genome so as to insure that genes of interest are mutated. Organisms that carry mutations that influence a specific process need to have the mutations in other genes removed (through sexual reproduction) before they can be studied. The strategies involved in “cleaning up” a mutation vary between different genetic systems, and we will not consider them in detail here.⁴⁶⁷

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 the 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 (next page ↓). As an example, consider two independently derived alleles that produce the same apparent phenotype. Let us assume that we can generate organisms that are

⁴⁶⁶ [Functional rescue of vitamin C synthesis deficiency in human cells using adenoviral-based expression of murine l-gulono-γ-lactone oxidase](#)

⁴⁶⁷ If you are interested, you can check out: [The art and design of genetic screens: Drosophila melanogaster](#) which has some interesting properties, such as the lack of meiotic recombination in males!

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 organisms, we expect that all of the F₁ generation will be genetically the same, at least at these loci. If the F₁ 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 a a₁/b₁ genotype. We could 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 would conclude that they are alleles of the same gene, if linked to different genetic markers, then it is likely that these are alleles of different genes.

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.	

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) while the a₁/a₂ heterozygote is wild type. 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. We will consider the various other types of genetic interactions, all of which can combine to various extents, through interactions with other allelic variants present throughout the genome, to modify the phenotype displayed by an allele (genetic background effects).⁴⁶⁸ This is one reason that research often examines an allele's phenotype(s) in a number of genetic backgrounds - crossing mutant animals with wild (by which we mean really wild) type animals. Genetic backgrounds can have substantial effects on phenotypes.⁴⁶⁹ Given that different species (such as mice and humans) have dramatically different

⁴⁶⁸ [Genetic Background Limits Generalizability of Genotype-Phenotype Relationships](#) (a paper cited above)

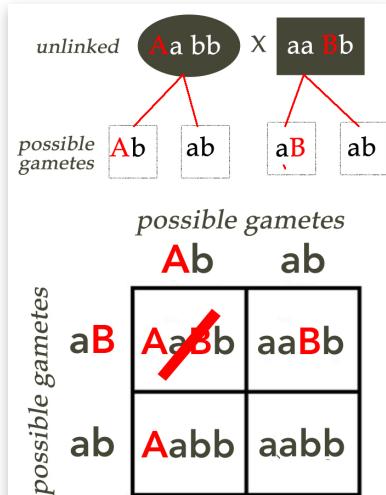
⁴⁶⁹ [Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases](#)

genetic backgrounds, it is not surprising that the same mutation (for example, a null mutation) defined in one organism can produce a different phenotype in another.⁴⁷⁰

Interacting traits: synthetic lethality and co-dominance

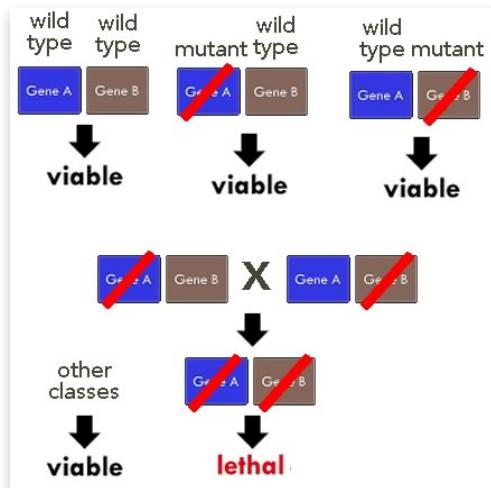
Physical linkage of genetic loci is only one way that genes interact, another involves interactions between gene products and the biological processes they mediate. Perhaps the most dramatic is known as synthetic lethality.⁴⁷¹ In such a situation, often but not necessarily, carried out with dominant alleles of two distinct genes, both heterozygotes, on their own, are viable, while the double heterozygote is dead, the combination is lethal (\rightarrow). Similarly, it can be the case for recessive alleles, that individually homozygous organisms are viable, while double homozygous individuals die, or display a different phenotype. 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 of a

cross (\leftarrow). 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 as we did previously to identify linkage.



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 gene functional product (amorphic or hypomorphic alleles), the product of the second gene may be sufficient, but if both are mutant, not enough of the functional complex is 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



⁴⁷⁰ [Null mutations in human and mouse orthologs frequently result in different phenotypes](#)

⁴⁷¹ [Synthetic lethality and cancer](#)

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:

242. Generate a concept map of all the plausible ways that the products of two distinct genetic loci could interact to produce a synthetic phenotype (we will probably do this in class)
243. If a gene is missing from a syntenic region, what might have happened to it?
244. How might the level of expression of one gene influence the phenotype associated with another?

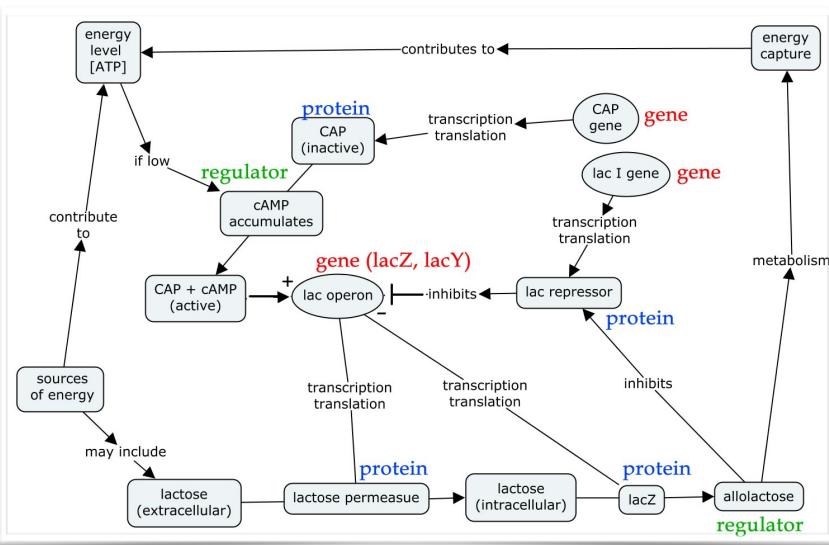
Question to ponder:

- Why did Mendel exclude interacting alleles from his analysis. How did he do this?

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.⁴⁷² In the context of any study in which mutations are generated, it is necessary to remember 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.

Most gene products function within networks in which particular gene products interact with each other and regulatory molecules to produce specific phenotypes. 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 lactose, allolactose, and cyclic AMP (\downarrow). 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. 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



⁴⁷² One point to keep in mind is that normally the process of generating mutants generates lots mutants throughout the genome, which can 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

components of the system. We might have missed a gene/gene product that influences the behavior of the system. At the same time, while DNA-dependent, RNA polymerase is a necessary component of the system, required to express 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 phenotypes.⁴⁷³ 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 often parts of the network that are linear, or perhaps best termed sequential, 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 [On-line Mendelian Inheritance in Man](#) (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 then be transformed into estradiol, a “female” hormone, which is also involved in male reproductive function.⁴⁷⁴ 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.

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 should now 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

⁴⁷³ [Pleiotropy: One Gene Can Affect Multiple Traits](#)

⁴⁷⁴ see [The role of estradiol in male reproductive function](#)

genes (or more accurately gene products) in a particular process.⁴⁷⁵

A complicating aspect of most actual interaction pathways is that there are various forms of feedback 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 bypass 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.

Temperature sensitive alleles

A final type of mutation (allele) is known as a temperature-sensitive mutation / allele. In the case that a gene encodes a polypeptide, changing the amino acid sequence of that polypeptide can influence how the polypeptide chain folds, as well as its stability as a function of temperature. In some cases, the polypeptide (or protein) can be more sensitive to its surroundings. A mutant protein may no longer behave normally when the temperature is reduced (cold-sensitive) or increased (heat-sensitive). This underscores the fact that each 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 denature, or lower temperatures, where they may adopt non-functional configurations.

Questions to answer:

245. Under what conditions might self-fertilization be selected for or selected against?
246. What factors limit the usefulness of genetic crosses to establish epigenetic relationships?
247. How are genetic pathway maps useful, and what are there limitations?
248. Why is a forward genetic screen unlikely to identify all components of a particular process?
249. 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?

Measuring evolution's impact on allele frequencies: Hardy-Weinberg

If we consider a population, each gene is represented by some set of alleles; these occur at various frequencies in different genes. 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 the assumption that evolutionary processes were not occurring within a population, they assumed that: 1) the population was infinite, so that processes such as genetic drift do not occur; 2) the population is isolated, so that no individuals leave or enter; 3) that no new mutations occur; 4) that mating between individuals is random (no sexual selection); and 5) there are no differential reproductive effects, that is,

⁴⁷⁵ [Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems](#)

natural selection is not occurring.⁴⁷⁶ Under these 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 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”⁴⁷⁷, but this is not the case unless the alleles influence reproductive success, that is, unless positive or negative selection is occurring.

So let us consider the situation in which there are two alleles (A and a) of a particular gene; if the frequency of A in the population is p, the frequency of a equals 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. 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. If allele frequencies change over time, one of the assumptions of the model must be wrong - the most obvious is that genotype-based differential reproduction effects (natural selection) are active.

Tracking the fate of mutations (becoming an allele or disappearing from the population)

At this point, you might well ask yourself, given the effectiveness of natural selection, why do alleles that produce severe diseases occur or persist at all? 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, and that these alleles (mutations) disappear from the population with the death of the organism. The prevalence of the disease will then reflect the rate at which such pathogenic mutations arise *de novo* together with the rate at which the individuals carrying them are eliminated (before they have offspring). 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 disease-associated allele might have no discernible effect on the organism’s reproductive success – that is, the allele is recessive. If we remember that whether an allele is recessive or dominant depends upon the phenotypic trait being

⁴⁷⁶ Hardy-Weinberg Equilibrium: <http://www.tiem.utk.edu/~gross/bioed/bealsmodules/hardy-weinberg.html>

⁴⁷⁷ genophagy

considered. As noted above, it is possible that the heterozygotic state conveys a reproductive advantage, that is, the allele has both a dominant (positive) and a recessive (negative) phenotype. In this case, the heterozygote will be subject to positive selection (leading to an increase in allele frequency), while the homozygote will be subject to negative selection (leading to a decrease in allele frequency); this can be sufficient to maintain the allele in the population at a significant levels. Similarly the effects of a dominant allele associated with a pathological condition can be be ameliorated, or even beneficial in the presence of various genetic modifiers (enhancers or suppressors)(we will go into more detail below). 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 effects the final balance. Of course this steady state is sensitive to changes in the environment that influence phenotype and their effects on reproductive success. If we were more mathematical, one could model the system based on such effects.

The pace of selective effects depends upon population size and the strength of selective (both positive and negative) effects. 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 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:

250. Consider conditions in which the deletion of a gene might lead to a selective advantage.
251. 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?)
252. How can combinations of alleles in different genes lead to new traits?

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: Germ line alleles and human pathologies

We consider inherited alleles, their effects on embryonic development and their roles in human disorders and diseases. We consider their patterns of inheritance and techniques used to identify when and where a gene is expressed. We consider how to identify the genes involved in more complex phenotypic traits, traits influenced by allelic variation at tens to hundreds of genes.



Up to this point, we have been considering genes, mutations, and alleles from the perspective of distinct traits. We now focus on how mutations, and the alleles they can become, influence human

"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.

health, leading to genetic dispositions and diseases. We will begin by a short review of embryonic development, although critical details are clearly beyond us here. We focus first on germ line alleles, present in maternal or paternal parents that influence gametes and early developmental

processes. In the next chapter we will consider how these inherited alleles can interact with mutations that occur during the process of development (cell division and differentiation). This is a hugely complex topic, so our intent is to identify core concepts rather than specific molecular and cellular processes.

Developing multicellular organisms: from egg to embryo and more

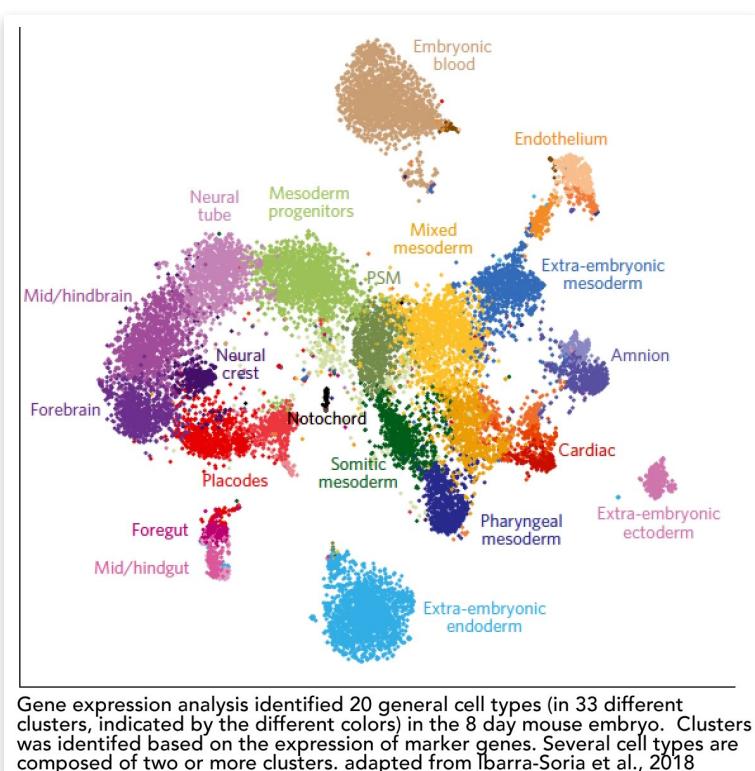
Complex multicellular organisms undergo a process known as embryonic development.⁴⁷⁸ Development begins with the fusion of a haploid sperm and egg, produced through meiosis, to form a diploid zygote that then divides (by mitosis) to produce a multicellular embryo that develops into an adult. Egg and sperm differ in their composition. For example in vertebrates, zygotic mitochondria are supplied in the egg only. Similarly, genomic imprinting (p. 250) means that in the zygote, only the egg-derived or the sperm-derived alleles are expressed. Following the formation of the zygote, cell division leads to a multicellular organism.⁴⁷⁹

While the fertilized egg is totipotent - that is, it can generate all of the cells found in the adult, the cells formed during development become more and more restricted with respect to the types of progeny that they can produce – they become committed to one or another specific fate. In part this fate restriction is due to the fact that as cells divide, different cells come to have different neighbors, and so they experience different environments and different combinations of signals from other cells, leading to the expression of different genes and different cellular behaviors. Differences in gene expression can

478 [Multicellularity: The Evolution of Differentiation](#)

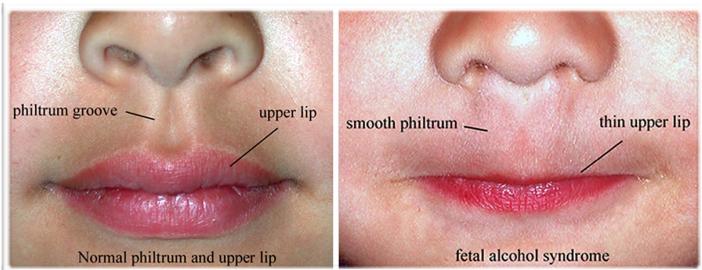
⁴⁷⁹ This process of multicellularity is described in the supplemental chapter.

be used to identify cells, as in this figure in which the cells of an 8 day after fertilization mouse embryo are characterized (in the mouse birth occurs ~21 days post zygote formation, that is fertilization)(→). As development proceeds, cells in various regions of the embryo behave differently from one another, they differentiate into various types of cells, such as neurons, muscle cells, epithelial (surface) cells, etc.⁴⁸⁰ The process of differentiation is associated with, and driven by various internal asymmetries, signals from neighboring cells, and differences in which genes are expressed in the various cells.



Gene expression analysis identified 20 general cell types (in 33 different clusters, indicated by the different colors) in the 8 day mouse embryo. Clusters were identified based on the expression of marker genes. Several cell types are composed of two or more clusters. adapted from Ibarra-Soria et al., 2018

The process of development is complex, and somewhat different in different organisms, leading to the different morphologies of different species and different individuals within a single species. Development is influenced by the gene regulatory systems active, their regulation, together with stochastic effects and environmental influences. As an example, excessive exposure of the human embryo to ethyl alcohol leads to a developmental defect known as fetal alcohol syndrome (FAS), associated with a range of effects and defects, including irreversible brain damage and a number of growth problems, including some minor malformations of facial structures (↓).⁴⁸¹ The extent of the effects of fetal alcohol exposure are also influenced by the genotypes of the mother and the developing embryo, in particular by genes involved in the metabolism of ethanol.⁴⁸² In a similar way, the recent outbreak of Zika virus has been associated with developmental defects, specifically microcephaly, a drastic disruption of brain growth, associated multiple functional (cognitive) defects.⁴⁸³ Again, the severity of these defects is likely to vary based on the timing of infection in terms of embryonic development, together with the genotype of the mother, particularly in terms of her immune response to the virus, and the genotype of embryo, in terms of its susceptibility to perturbation.



⁴⁸⁰ Defining murine organogenesis at single-cell resolution reveals a role for the leukotriene pathway in regulating blood progenitor formation

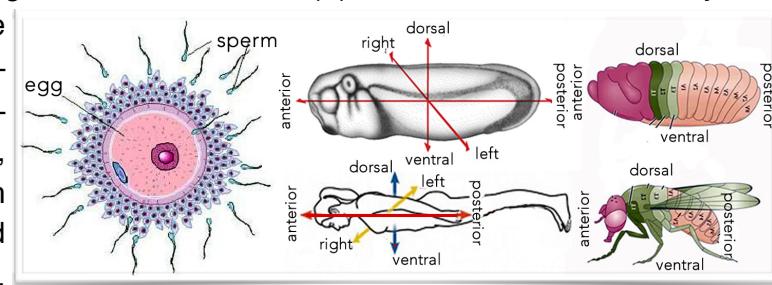
⁴⁸¹ The effects of alcohol on fetal development.

⁴⁸² Genetic and epigenetic insights into fetal alcohol spectrum disorders

⁴⁸³ Zika Virus in the Americas — Yet Another Arbovirus Threat

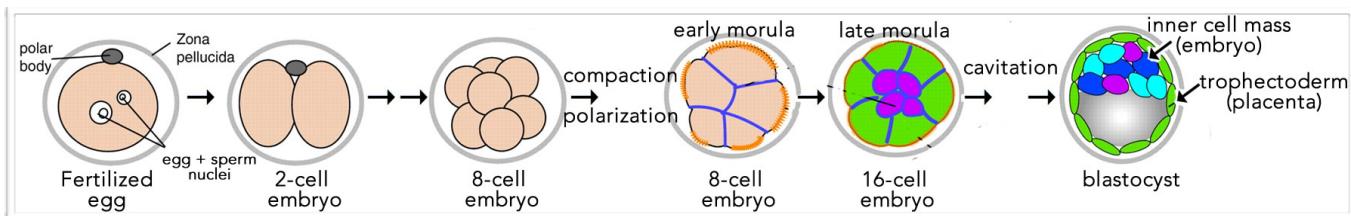
Establishing embryonic axes

During the process of embryonic development, the various axes of the organism, typically anterior-posterior, dorsal-ventral, and left-right in vertebrates form (↓). While the details of embryonic development beyond our scope here, there are two basic, and interacting, processes – asymmetries and inductive interactions – that drive the formation of embryonic axes, the changing shape of the embryo, from fertilized egg to adult, and the associated cellular differentiation and tissue formation.



The mechanisms involved reflect evolutionary pressures, for example, when an egg is fertilized externally (outside of the mother), and without subsequent parental protection, the new organisms are highly vulnerable to predators and their development is normally quite rapid. The eggs are often large and contain all of the nutrients required for development to proceed up to the point where the new organism can feed on its own. To facilitate rapid development, the egg is pre-organized, that is, it is highly asymmetric, with specific factors that can influence gene expression, either directly or indirectly, positioned in various regions of the egg (↓). As an example, in the frog *Xenopus laevis*, the sperm entry point (SEP) establishes a new asymmetry axis that can lead to the reorganization of the cytoplasm before the first embryonic cell division. The oocyte's asymmetrical structure is transformed into an asymmetrical egg through the process of meiosis and the cellular events associated with it. The asymmetric cytoplasmic determinants include various RNAs and proteins, RNAs. The various asymmetries in the fertilized egg are stabilized by the rapid cycles of DNA replication and mitotic cell division, with growth based on the utilization of maternally supplied nutrients. As distinct cells are formed, they become different from one another as they inherit different determinants; the presence of these determinants leads to changes in gene expression and cell differentiation (see above).

On the other hand, in a number of organisms, and specifically mammals, embryonic development occurs within the mother, so there is no compelling need to stockpile nutrients within the egg and the rate of development is (generally) dramatically slower than that seen in externally developing embryos. In such developmental systems, it is often not the asymmetries associated with the oocyte and fertilized egg that are critical, but rather the asymmetries that arise during embryonic development. As the zygote divides, a major factor that drives differentiation is whether a cell comes to lie on the surface of the embryo or within the interior (↓). In mammals, the cells on the exterior form the trophectoderm, which goes on to form

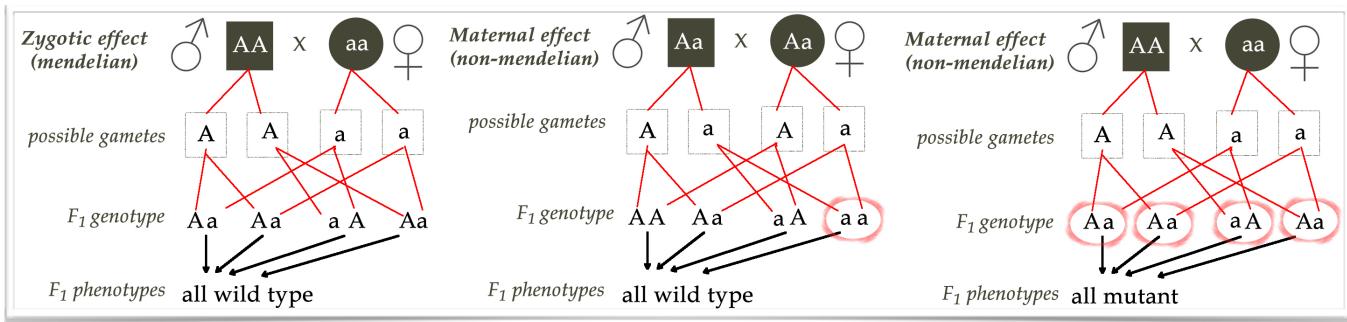


extraembryonic tissues, in particular the membranous tissues that surround the embryo and biofundamentals™ Klymkowsky & Cooper - copyright 2010-2018 version: Tuesday, January 16, 2018 281 of 315

become part of the placenta, the interface between the embryo and the mother. Cells within the interior form the inner cell mass that produces the embryo proper. Changes in gene expression will lead to changes in the ability to produce and respond to inductive signals, which will in turn influence cell behavior and gene expression. Through this process, the cells of the inner cell mass come to form the various tissues and organs of the organism. It is easy to tell a muscle cell from a neuron from a bone cell from a skin cell by the set of genes they express, the proteins they contain, their shapes (morphology), their internal organization, and their behaviors.

Genetic analysis of developmental processes: maternal and zygotic effect mutations

Embryonic development, like any other process or trait, 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. Mutations (alleles) that influence oocyte formation, and then embryonic development are known as "maternal effect mutations". They can be recognized based on their behavior in crosses. Take for example a recessive effect allele "a" - it may be a typical zygotic effect allele or a "maternal effect" allele. Let us consider how they can be distinguished. Let us assume that the mother is homozygous for a. For a typical zygotic effect allele, in cross to a wild type (paternal) homozygote (\downarrow), they behave as expected, with all offspring displaying a wild type phenotype. A cross between two of the resulting heterozygotes (\downarrow) produces all wild-type offspring. But let us consider a maternal effect allele. When we cross the homozygous mother to a male of any genotype we find that all offspring are mutant (\downarrow). We recognize maternal effect mutations by their non-Mendelian behavior (circled genotypes). Similarly, but not shown here, a dominant maternal effect allele will, when crossed to a male of any genotype, produce all mutant offspring.



Mitochondrial inheritance

Eukaryotic cells (such as our cells) have one or more type of intracellular organelle, either mitochondria (all eukaryotes) or mitochondria and chloroplasts (in algae and plants). These organelles have their own genomes, circular DNA molecules known as mtDNAs. A number of genes are encoded by the mtDNA: 37 in human. Also, the mtDNA can, like any DNA molecule, accumulate mutations when it is replicated or in response to free radicals generated during the course of aerobic respiration (something that we will not consider further).

One aspect of typical sexual reproduction is that only the mitochondria of the oocyte are inherited by the fertilized egg; the mitochondria present in the sperm cell, 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 have a number of phenotypes.⁴⁸⁴ The genetics of these mutations are often non-Mendelian and include maternal effects.

One complexity in the study of mitochondrial DNA mutations 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 organisms can contain different numbers of mitochondria and differ in their dependence on mitochondrial function. The result is that we are looking at a population 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, which raises the possibility of somatic selection. In any one cell or tissue, mitochondrial dependent phenotypes will reflect, and be influenced by the multiple 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.⁴⁸⁵

Questions to answer:

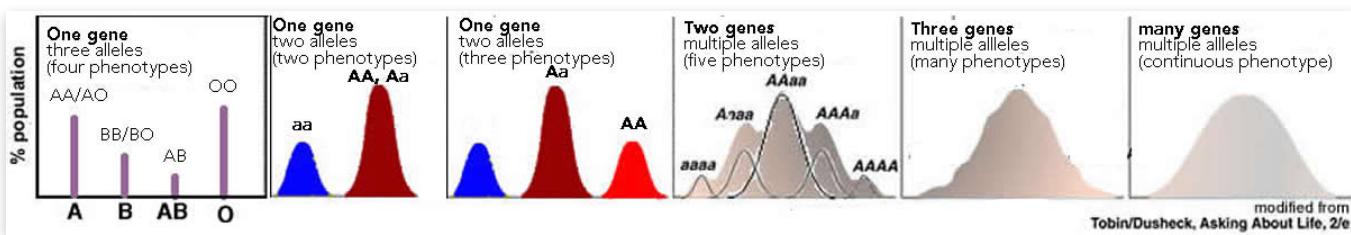
253. Schematically consider mechanisms that would lead to differential gene expression in the various regions of an embryo - how is that even possible?
254. Describe how imprinting can impact Mendelian allele behavior(s)?
255. Most of the genes involved in mitochondrial function are nuclear; how might that influence the phenotypes of mutations in mitochondrial DNA?
256. If you were to predict which tissues would be more severely effected by mutations in mitochondrial DNA, what would you base your predictions 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.

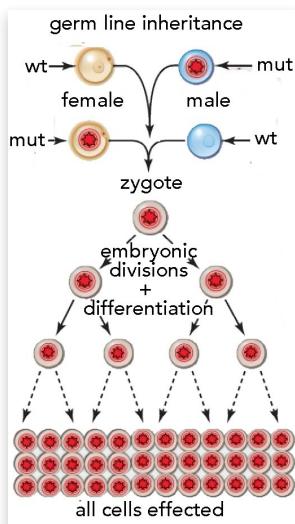
Traits and the number of genes involved

Mutations that become alleles (enter the germ line) 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 pea color or in the case of humans 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, 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 what we 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 mutation - last chapter) will have the same genotype (↓). That said, the phenotypes associated with a particular allele can vary between



different regions of the organism, different tissues, organs, and organ systems. 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 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 will see this effect in our discussion of somatic mutations; a germ line mutation can be inherited but not have a discernible phenotypic effect. However, if a subsequent somatic mutation occurs

that disables the functioning copy of the gene, or compromises the function of a complementary gene, a phenotype can arise. Such events are involved in some heritable cancer susceptibilities (see below).

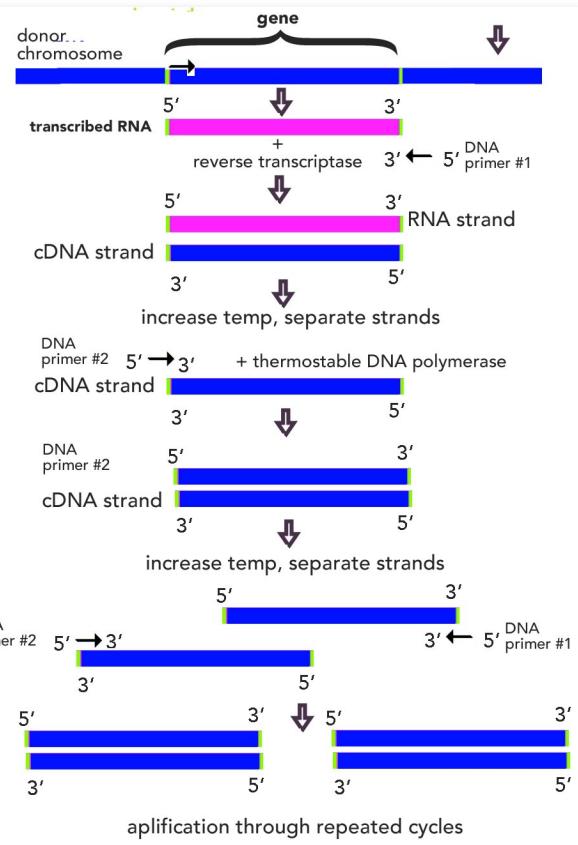
Where is a gene expressed? RT-PCR based systems

The following discussion might encourage you to ask, exactly how do we determine where and when specific genes are expressed within an organism? There are a number of applicable mechanisms 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.

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 in manipulating genes, and visualizing gene expression. 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. This process involves first making a DNA copy of the transcribed RNA, so that we avoid the genomic DNA copies of the genes present in every cell. We isolate RNA from a tissue and then use the virally-derived reverse transcriptase (RT) enzyme; RT uses a DNA primer and makes a DNA copy complementary to the RNA strand, a cDNA (→). The two strands are then separated (typically by increasing the temperature of the system), and then a second DNA primer is added together with a thermostable DNA-dependent, DNA polymerase in order to generate a copy of the cDNA. Now we begin the amplification stage of the reaction. The two strands are separated by increasing system temperature. The 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, a exponential growth in DNA molecule numbers with each cycle. Because the primer sequences, which are synthesized *in vitro*, are complementary to, and specific for, a particular gene sequences, one can 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 compare the numbers of a particular type of RNA molecule (the products of a particular gene) present in the original sample, a measure of the levels of gene expression, at least at the RNA level. With different sets of primers, it is possible to quantify the various splice forms of a gene expressed.

More recently, it has become possible to isolate and sequence the RNAs (or rather cDNAs derived and amplified from mRNAs) in even a single cell and to then sequence those DNA molecules to characterize the genes expressed in that cell.⁴⁸⁶ Because mRNA is used, only exon sequences are 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.⁴⁸⁷



In situ hybridization: A limitation of the RT-PCR approach is that it generally works 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).⁴⁸⁸ 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 digoxigenin) forms of

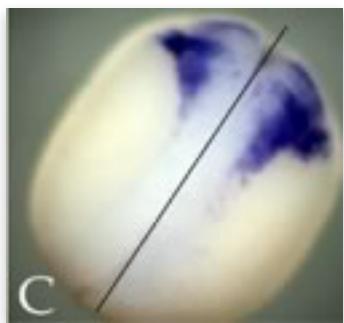
⁴⁸⁶ [A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications](#)

⁴⁸⁷ see [Defining murine organogenesis at single-cell resolution reveals a role for the leukotriene pathway in regulating blood progenitor formation](#)

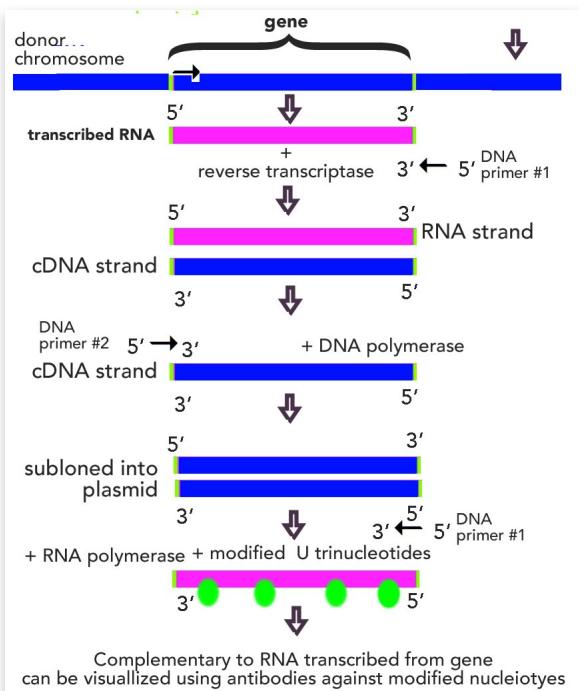
⁴⁸⁸ although things may actually be somewhat more complex: see [Brain Cells Share Information With Virus-Like Capsules](#)

the RNA nucleotide UTP are used; this modified nucleotide can be used by the polymerase and is incorporated into the newly synthesized RNA (\rightarrow).

The overall process is relatively simple. The tissue is chemically stabilized, 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 the gene's 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 mRNA 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 horse radish peroxidase) can then be used to recognize the modified probe RNA:mRNA complex, and color-generating reactions, catalyzed by the enzyme, allow the distribution of probe to be visualized.



The example here (\leftarrow) is a neurula stage *Xenopus* (clawed frog) embryo in which a gene (Snail2/Slug) that is expressed in the neural crest has been visualized.⁴⁸⁹ *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.



Immunocytochemistry: One limitation of RT-PCR and *in situ* hybridization methods 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 level of the accumulated polypeptide. One approach to avoid this disconnect is to use antibodies, proteins generated by the vertebrate immune system that can bind specifically to specific molecular targets. We will ignore how antibodies are generated, but basically they act very much like *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).⁴⁹⁰

⁴⁸⁹ from: [An NF- \$\kappa\$ B and Slug Regulatory Loop Active in Early Vertebrate Mesoderm](#)

⁴⁹⁰ Here is an example of proteomic analysis: [Region and cell-type resolved quantitative proteomic map of the human heart](#)
biofundamentals™ Klymkowsky & Cooper - copyright 2010-2018 version: Tuesday, January 16, 2018 286 of 315

Questions to answer:

257. 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?
258. A gene can be spliced various ways - design primer sets to distinguish the splice variants of a gene.
259. 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?

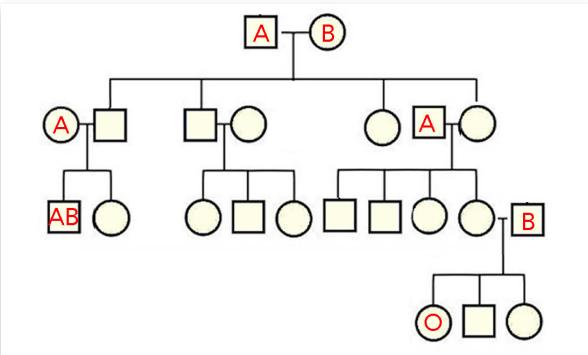
Questions to ponder:

- Why might the number of polypeptides in a cell differ from the number of RNAs that encode it?

Back to Mendelian determinants

Returning to the effects of various alleles, we will begin with discrete traits in humans that behave in a strict Mendelian manner. Perhaps the best known is blood type (see above) which is determined by three different alleles at a single locus. These alleles behave in a dominant A to O, and B to O, and a co-dominant (A with B) manner. The distribution of these alleles in different human populations appears to be due, at least in part to selective advantages associated with specific alleles in specific environments. For example, “*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*”.⁴⁹¹

There are three common alleles that control blood type in humans, A, B, and O. 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 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 may or may not be a realistic assumption. For example, the presence of an AB individual in the second generation (→), indicates that the male parent had to have 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 the lineage.



It is also worth noting that in the modern world, we 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 genetic father of a child (generally, but not always, who the mother of a child is is more unambiguous).

⁴⁹¹ [Beyond immunohaematology: the role of the ABO blood group in human diseases](#)

Disease-associated alleles

There are a number of identified genetic disorders with clear Mendelian inheritance (see [Specific Genetic Disorders](#)). What does this mean? Basically that the alleles associated with the disease act in a simple dominant or recessive manner. In the case of dominant disease-associated alleles, to be inherited means that they are not lethal as homozygotes, and 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 much survive and be able to reproduce. One point to keep in mind is 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. The classic example of such behavior are mutations associated with the hemoglobin B (HBB) gene of humans (→ and below). Alleles of this gene are 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 strong positive selection in environments where malaria is endemic. The same allele is responsible for both traits.

Gene-Phenotype Relationships

Location	Phenotype	Phenotype MIM number	Inheritance	Phenotype mapping key
11p15.4	Delta-beta thalassemia	141749	AD	3
	Erythremias, beta-			3
	Heinz body anemias, beta-	140700	AD	3
	Hereditary persistence of fetal hemoglobin	141749	AD	3
	Methemoglobinemias, beta-			3
	Sickle cell anemia	603903	AR	3
	Thalassemia-beta, dominant inclusion-body	603902		3
	Thalassemias, beta-	613985		3
	{Malaria, resistance to}	611162		3

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, with the two result that two distinct embryos develop within the mother simultaneously, and are generally born at in rapid succession. Fraternal twins are no more or less closely related than are two siblings born years apart. Fraternal twins are also termed dizygotic twins, since they involve two distinct zygotes. In animals that typically have multiple offspring (litters), all of the individuals born 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 then begins development. During development, for one reason or another, the embryo fragments into two distinct embryos, which then develop independently of one another. So, with the exception of (somatic) mutations that may have occurred during embryonic development, the two are genetically (genotypically) identical. This genetic identity enables us to measure the genetic concordance of a trait.⁴⁹² For example, if a trait is totally dependent upon a shared allele, together with a shared genetic (genomic) background, the concordance is 100%. In other cases, while genotype plays a role, it is not completely determinative. As an example, in the auto-immune muscle weakness disease myasthenia

⁴⁹² Does Higher Concordance in Monozygotic Twins Than in Dizygotic Twins Suggest a Genetic Component?

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.⁴⁹³

As we are talking about twins, it is also worth noting another type of outcome, which is known as a chimera.⁴⁹⁴ In a chimeric embryo, two embryos fuse into one - such that a single organism develops, but it has two distinct "sibling" genotypes.⁴⁹⁵ When this dizygotic fusion is complete, a single normal, albeit mosaic, embryo and mature organism is generated. When fusion is incomplete, or occurs at a later developmental stage, incompletely fused embryos are formed - what are known as conjoined twins.

Using web-based bioinformatic tools: Exac Browser

When studying a disease that appears to have a genetic component, it is common to identify the particular allele(s) involved. In the case of recessive alleles, such studies often involve pedigree analysis of more or less in bred 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 the influence of genetic background effects that influence the disease allele's penetrance and expressivity. Over the last decade, there has been an 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 60,000 people from around the globe. This data library can be searched using the [ExAC Browser](#).⁴⁹⁶ To search the ExAC database, the user (you, for example), inputs a gene's official name, as listed in [OMIM](#) or GenBank. ExAC then displays sequence data from 60,706 (as of July 2017) 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 associated with a number of human diseases (see above ↑). 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 ExAC database of "normal", that is

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

disease 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

⁴⁹³ [Immunopathogenesis in myasthenia gravis and neuromyelitic optica](#).

⁴⁹⁴ It is even possible to generate chimeric embryos between different species: [Humanized mice and porcineized people](#).

⁴⁹⁵ 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](#)

⁴⁹⁶ [Genomics, Big Data, and Medicine Seminar Series – Daniel MacArthur](#)

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 (\rightarrow). At the time of this writing (January 2018) 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 ExAC enables us to make informed guesses as to the impact of various genetic differences on the activity of a gene product.⁴⁹⁷ If, for example, a dominant allele has been linked to a disease and yet that allele is detected in the ExAC 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 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)(\downarrow). We see that the Gly70Ser and Glu40Ter alleles are present primarily in non-Finnish Europeans, while the

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

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

Glu122Gln allele is found in South Asians. It is not clear exactly what the effects of such missense mutations will be on 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, but clearly the non-sense mutation will produce a short 39 amino acid polypeptide, compared to the 147 amino acid long full length polypeptide. It is unlikely that such a severely 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

⁴⁹⁷ The [ExAC browser: displaying reference data information from over 60 000 exomes](#).

sequence and search for similar sequences throughout all sequenced genes that have been deposited in a central repository (GenBank). 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) or horizontal gene transfer. 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 ExAC) in such regions would argue for the action of positive selection.

Questions to answer:

260. Outline your strategy to determine whether someone is not telling you the truth about parentage, given a family tree and a simple dominant recessive trait.
261. You find a frequent allele in a population, but no individuals homozygous for that allele - how might you make sense of that observation?
262. Why aren't missense mutations necessarily loss of function mutations?
263. 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:

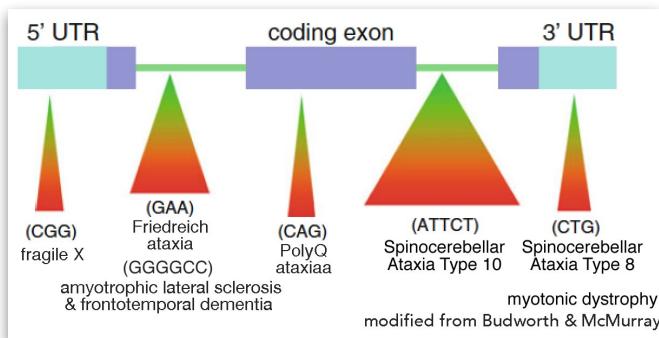
- Provide a model for why an individual homozygous for the Glu7Val allele not have sickle cell disease?

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 (see pp. 175-176). Such alleles are associated with what are known as “trinucleotide repeat” (TNR) 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.⁴⁹⁸ Within the genes involved, there are regions of repeating nucleotides. Because of the slippage of the DNA polymerase during the process of 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.

⁴⁹⁸ [A Brief History of Triplet Repeat Diseases](#)

Mechanisms: Given 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 they 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 clearly 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. Expansion domain effects within a region of the 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 DNA strands (sense and anti-sense) and “that RAN (repeat-associated non-ATG translation) translation⁴⁹⁹ 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-Arg, Gly-Pro). These proteins accumulate in cytoplasmic aggregates in affected brain regions”.⁵⁰⁰ 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 (complex and weird, huh?).⁵⁰¹ Rest assured, the exact mechanisms of these types of alleles and associated phenotypes are complex, but based on the effects of altered transcription on the functional roles of specific cell types.



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 effects of pathogens 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).⁵⁰² The estimates for the heritability of height are not all that accurate (and differs between populations), ranging from between ~60 to ~80% of the variation

⁴⁹⁹ [Non-ATG-initiated translation directed by microsatellite expansions](#)

⁵⁰⁰ [RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia.](#)

⁵⁰¹ [Spt4 selectively regulates the expression of C9orf72 sense and antisense mutant transcripts](#)

⁵⁰² [How much of human height is genetic and how much is due to nutrition?](#)

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 those genes involved in a particular trait?⁵⁰³ 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, it will be near specific genes, and inherited with the gene (allele). Of course meiotic recombination can influence who is linked to whom.

The different SNPs present are identified based on nucleotide complementarity (not unlike the basic process behind *in situ* hybridization (see above). Samples of the person's genome is taken (often from white blood cells, which have nuclei and DNA, in humans - in contrast to enucleated red blood cells). 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 measureable.

We than 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 and 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 simply linked on the chromosome to the actually 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.⁵⁰⁴

Questions to answer:

264. In the case of genetic anticipation, what is the impact if the repeat domain gets shorter?
265. How might the synthesis of small polypeptides influence normal cell behavior?
266. How would a repeat domain influence a coding region?
267. What is critical before one can even consider beginning a GWAS study?

Questions to ponder:

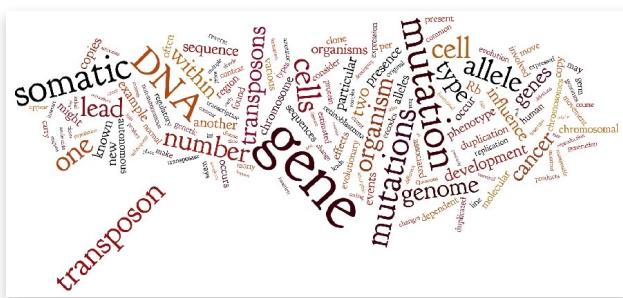
- You discover a gene linked to a particular trait through a GWAS study, how might you go about establishing a significant physiological role for the gene in influencing that trait?

⁵⁰³ [Chapter 11: Genome-Wide Association Studies](#)

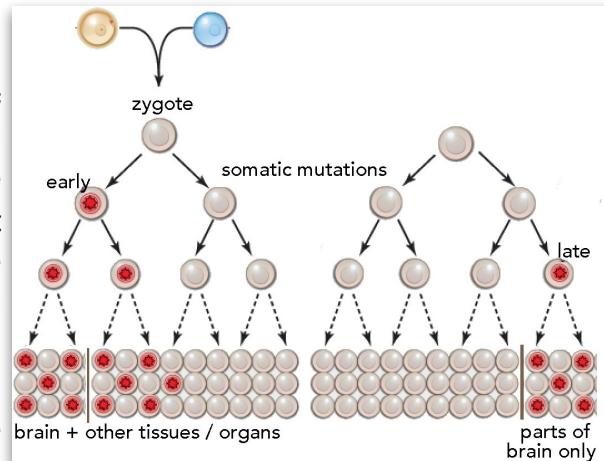
⁵⁰⁴ [The interplay of common, rare variation in autism](#)

Chapter 17: Genome dynamics and pathogenic somatic mutations

In which we consider how genes move around within the genome, through the action of transposable elements of various types, and how that can influence phenotype. We consider how mutations arising in somatic cells can interact with inherited alleles or on their own to impact the development of the nervous and other systems, including the origination and progression of cancer.



Up to this point, we have been considering mutations that have become alleles 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 (p. 252), it can be expected to have its 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 the development of the organism the mutation occurs will determine what percentage of the cells in the organism carry the mutation. Of course, if the mutation leads to a lethal phenotype, the cell that carries it will die, so no cells in the organism will carry the mutation. More often, such mutations are not lethal, but many influence the rate and outcomes of cell divisions, a number have been found to influence the development of the organism, and its various tissues and organs.



Normally, when and where a cell divides is under strict regulatory control, involving both internal regulatory networks, as well as signals from other cells. Another class of somatic mutations underlie the appearance of cancer cells. Cancer itself is a complex process, often involving a number of steps, a number of somatic mutations within a particular clone (cellular lineage); a complete study of cancer is beyond us here, but certain common features of carcinogenesis and progression are worth considering - the most important is to recognize that (as we noted previously) a multicellular organism is a social system. Cells are expected to cooperate in defined ways to keep the society functioning smoothly. Somatic mutations serve to disrupt that coordination. In particular, somatic mutations can lead to cells ignoring various signals meant to control their growth and behavior. As each somatic cell is clonally related to its ancestors and progeny, these clonal populations can (in the absence of appropriate regulation) compete with each other in destructive ways, at least destructive to the goals, survival and reproduction, of the organism as a whole.

Rates and effects of somatic mutation

The rates at which mutations occur within a particular cell type is based on the number of rounds of DNA replication, the error rate associated with that process, and the efficiency of DNA error repair. In mammals, most germ line errors are associated with males, because there are more cell divisions giving rise to sperm than are involved in generating eggs in the female germ line. DNA error rates differ (apparently) between species (you might speculate on why). In the mouse the current estimate is of $\sim 5 \times 10^{-9}$ per base pair per generation. The number is estimated to be higher in humans, closer to 1.2×10^{-8} per base pair per generation. A number of studies have been carried out to determine the mutation rate in somatic cells in the two species (mouse and human); it appears that these rates are higher in the soma than in the germ line.⁵⁰⁵

To think clearly about the effects of a particular somatic mutation, we have to think in some context. How does the new mutation interact with the pre-existing genome. For example, if you inherit a amorphic or hypomorphic allele of a particular gene, that allele may well act in a recessive manner. But what happens if, in a somatic cell, the other (wild type) allele is mutated to an amorphic state, and assume that that cell is capable of, and called upon within the context of the organism, to divide. If the new mutation does not produce a lethal phenotype, the cell will divide to form a somatic clone, and each daughter cell will carry the mutation. If this occurs in the cells that give rise to the brain, it can lead to dysfunction in specific regions, and neurological symptoms. As an example, autism (along a spectrum of severity) is common, occurring in $\sim 1\%$ of the population. Both germ line alleles and somatic mutations have been implicated.⁵⁰⁶ Autism appears to share genetic risk factors with schizophrenia and bipolar disease.⁵⁰⁷ The occurrence of a somatic mutation, in the context of a “sensitized genetic background”, that is in the presence of an allele that influences neural development, can lead to defects in neuronal development. Even as few as 10% of cells that carry the somatic mutation can lead to a neuronal pathology.⁵⁰⁸

In the context of non-neuronal cells, 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. While the steps in the formation of a cancer can be complex, and reflect a number of regulatory pathways - what is clear is that once a somatic mutation has occurred, it can establish a clone that continues to divide. The mutation turns the well behaved somatic cell into a social cheater (see chapter 4). Subsequent mutations can then accumulate that enable the cancer clone to get adequate nutrients and avoid host responses. The evolution of the cancer clone is, however, futile - at best 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.

⁵⁰⁵ see [Differences between germline and somatic mutation rates in humans and mice](#)

⁵⁰⁶ [The interplay of common, rare variation in autism](#)

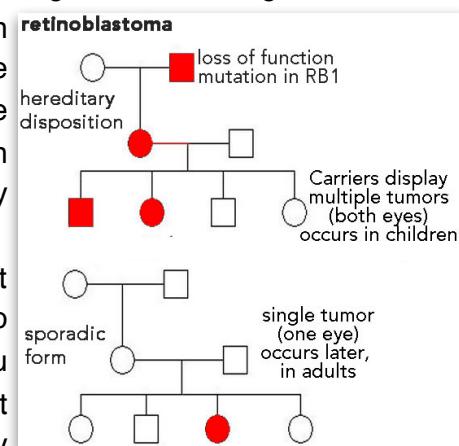
⁵⁰⁷ [Genetic overlap between autism, schizophrenia and bipolar disorder](#)

⁵⁰⁸ see [Somatic Mutation, Genomic Variation, and Neurological Disease](#)

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 complex and beyond our scope here.⁵⁰⁹

Nevertheless, we will consider one type of allele/somatic mutation combination, that is involved in retinoblastoma, a cancer of the retina. There is a trait that we will call “susceptibility to retinoblastoma”; it is conferred by the presence of a dominant, loss of function (amorphic) allele in the RB1 gene, let us call this allele Rb-. In those that inherit the Rb- allele have ~90% chance of developing retinoblastoma early in childhood, ~10% will not – they will be “silent” carriers.⁵¹⁰ Inheriting a single copy of the Rb- allele is not, by itself sufficient to lead retinal cells to become cancerous, a second, somatic mutation needs to occur in order to inactivate the wild type copy of the RB1 gene. But having the Rb- allele dramatically increases the probability that when such a mutation occurs, cancer will result. We see this effect because people (children) homozygous for the Rb- allele typically develop multiple tumors in each of the two eyes; these tumors appear early in childhood. The presence of the Rb- allele leads to a hereditary disposition toward developing retinoblastoma (→).

But people who do not inherit this allele can also get retinoblastoma; the difference is that they have to accumulate two separate mutations, which is a much rarer (improbable) event – you might consider how much rarer. Such rare events do occur, but they tend to occur later in development, so it would be very unlikely that to occur cells whose decedents come to be present in both eyes. When sporadic forms of retinoblastoma appear, they are almost always restricted to one tumor in one eye, and appear in older individuals. A similar pattern of inheritance is associated with breast cancer susceptibility gene 1 (BRCA1).⁵¹¹



Non-disjunction: a disease of aberrant chromosome segregation meiosis

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, contain hundreds of genes, the presence (or absence) of the correct number of chromosomes leads to changes in patterns of gene expression. Generally, when a chromosomal aneuploidy occurs, the resulting embryo fails to complete normal development; recent studies indicate

⁵⁰⁹ [Neomorphic mutations create therapeutic challenges in cancer](#)

⁵¹⁰ [Genetics of Retinoblastoma](#).

⁵¹¹ [BRCA1 and BRCA2: Cancer Risk and Genetic Testing](#)

that chromosomal abnormalities are surprisingly common in humans.⁵¹² 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.⁵¹³ 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!!!!).

Questions to answer:

268. A somatic mutation occurs early in development, what factors will influence the % of cells in the organism over time that carry the mutation?
269. How would you characterize dominant from recessive disease susceptibility alleles?
270. How does exposure to mutagens lead to increased risk of cancer development?
271. What types of molecular defects would lead to chromosomal aneuploidy?
272. How might having three (or one) copy of a chromosome influence normal cell behavior (and gene expression)?
273. In the context of the Rb⁻ allele, how might loss of the chromosome or chromosomal region in which Rb resides influence cellular phenotypes?

Questions to ponder:

- Why doesn't inheriting a cancer associated Rb⁻ or BRCA1 allele not lead to increase risk of cancer in other (all) tissues?
- Under what conditions would a somatic mutation become an inheritable allele?
- How would a mutation in a checkpoint gene influence a somatic cell's clonal evolution?

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, becoming increasingly clear that genomes are more dynamic than previously thought. For example, consider the number of new mutations (single nucleotide polymorphisms, and small insertions and deletions and such) that arise in each generation. The frequency of such events can be estimated based on the number of times a DNA molecule has been replicated in the course of developing from a fertilized egg to the formation of its own gametes – about 400 replication events in a human male, fewer in a female – together with the error rate of DNA replication ($\sim 1 \times 10^{-10}$ per nucleotide per division.) Since each diploid cell contains $\sim 6 \times 10^9$ nucleotides, we can expect ~1 new mutation for every two rounds of DNA replication. It has been estimated that, compared with the chromosomes our parents supplied us, we each have between 60 to 100 new mutations in the chromosomes found in our germ line. Given that less than ~5% of our DNA encodes gene products, that is polypeptides and functional RNAs, only of few of these new mutations are likely to influence a gene coding region or its expression.⁵¹⁴ Even when they occur in a gene’s coding region,

⁵¹² [Chaos in the embryo](#)

⁵¹³ Morris et al. 1999.: Fetal loss in Down syndrome pregnancies. Prenat Diagn. 19: 142-145.

⁵¹⁴ It is, admittedly more difficult to estimate the percentage of the genome involved in the regulation of gene expression, since these regions are harder to recognize than coding regions (can you guess why?) Also, there is an increase recognition that there is regulated transcription of regions that do not encode polypeptides (or at least longer polypeptides). Such long non-coding RNAs (lncRNAs) have been found to regulatory roles. see: [Long noncoding RNAs: past, present, and future](#)

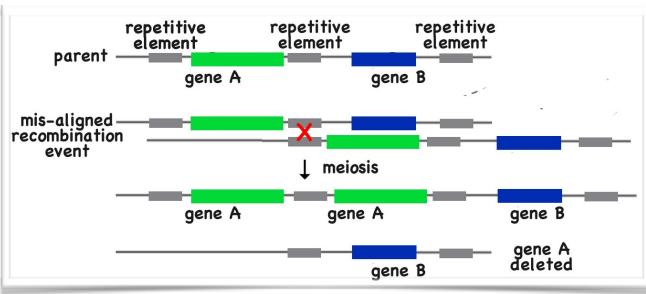
the redundancy of codons means that many SNPs lead to what are termed synonymous mutations, which (generally) do not lead to functionally significant alterations in the gene products. That said, even apparently “neutral” mutations can lead to changes in genotype that can have effects on phenotype, and so evolutionary impacts. For example, they might influence the regulatory region of a gene.⁵¹⁵ As we have already discussed, in small populations genetic drift can influence whether new alleles (with non-lethal effects) are retained in the population.

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.⁵¹⁶ It has been estimated that each person contains about 2000 “structural variants”.⁵¹⁷ Large chromosomal inversions or the movements of regions of DNA molecules between chromosomes can have effects on chromosome pairing during meiosis, and can lead to hybrid sterility and inviability. You can work out for yourself what might happen if recombination events occur in such regions. The mechanisms that lead to these genomic changes can be complex, and largely beyond our scope here.⁵¹⁸

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, 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 cells often actively kill themselves (through apoptosis) or in organisms with an immune system, they can be actively identified and killed.

Gene duplications and deletions

While meiotic alignment generally occurs accurately, there are times where mis-alignment happens. For example, what happens if there are repeated sequences within a chromosomal region. If the homologous chromosomes misalign (\rightarrow), crossing over can lead to haploid cells that emerge from meiosis with either gene duplications or



⁵¹⁵ this appears to have occurred with the human genome: see [Exploring the genesis and functions of Human Accelerated Regions sheds light on their role in human evolution](#)

⁵¹⁶ [Copy number variation in humans:](#)

⁵¹⁷ [Child Development and Structural Variation in the Human Genome](#)

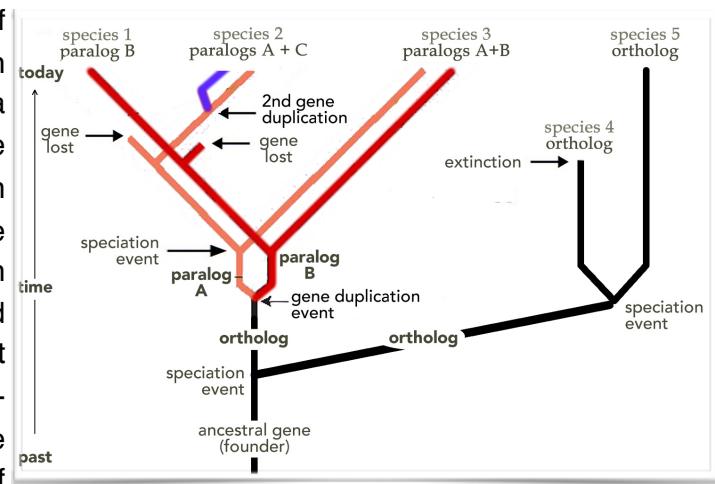
⁵¹⁸ [Mechanisms of Gene Duplication and Amplification](#)

deletions. Such duplication events can have a kind of liberating effect on subsequent evolutionary pathways.⁵¹⁹ 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 a 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 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.⁵²⁰

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. It may catalyze various off-reactions, these are sometimes referred to as off-target or promiscuous activities.⁵²¹ 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, often in ways that 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, particularly in plants, appear to have been duplicated multiple times during the course of their evolution.⁵²² In any gene duplication event, the two duplicated genes can have a number of fates, they can act as a “back-up” for one another, they can be repurposed, or one can be lost. Repeated gene duplication events can generate families of



⁵¹⁹ Ohno's dilemma: evolution of new genes under continuous selection: and [Copy-number changes in evolution: rates, fitness effects and adaptive significance](#)

⁵²⁰ [Dihydrofolate reductase amplification and sensitization to methotrexate of methotrexate-resistant colon cancer cells:](#)

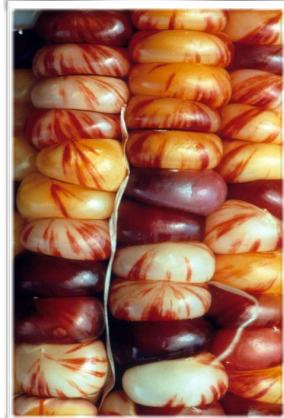
⁵²¹ [Enzyme promiscuity: a mechanistic and evolutionary perspective & Network Context and Selection in the Evolution to Enzyme Specificity](#)

⁵²² Genome and gene duplications and gene expression divergence: [a view from plants](#)

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 gene present in the last common ancestor of those organisms. Paralogous genes are derived from a gene duplication event; they are present together in a particular organism. If one paralog of a pair is subsequently lost, it can be difficult to distinguish the remaining gene from the original ortholog. A particular paralog in one organism can be orthologous to a gene in another organism, or it could have arisen independently in an ancestor, through a gene duplication event.

When both paralogs are present in a species, detailed gene/polypeptide sequences 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 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, meaning that the cell or organism with that mutation dies or fails 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)

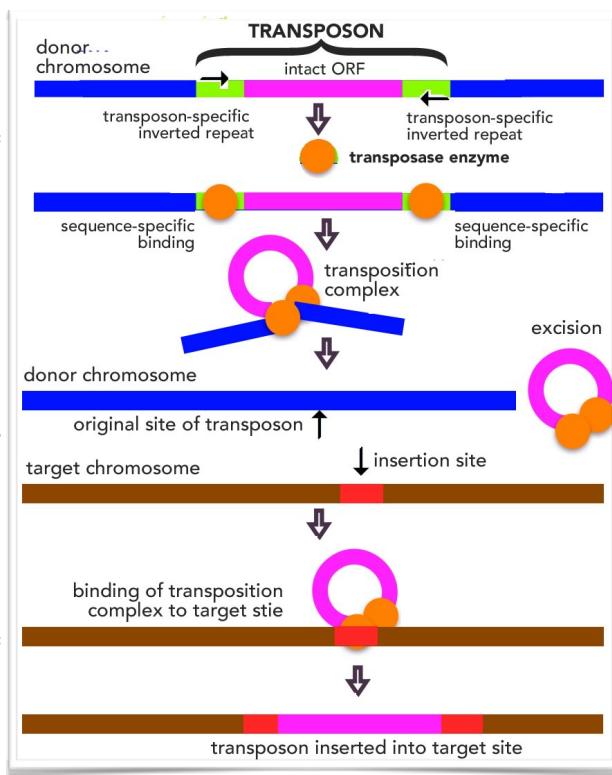
As we are thinking about DNA molecules moving into the genome through horizontal (lateral) gene transfer, and between genomes through conjugation, we can consider another widely important molecular system known as transposons. A transposon is a piece of DNA that can move (jump) from place to place in the genome.⁵²³ The geneticist (and Nobel prize winner) Barbara McClintock (1902–1992)(→) first identified transposons, although she did not know the molecular basis of the effect, while studying maize (*Zea mays*).⁵²⁴ In particular, she studied

⁵²³ Transposons: The Jumping Genes: <http://www.nature.com/scitable/topicpage/transposons-the-jumping-genes-518>

⁵²⁴ Barbara McClintock: http://www.nobelprize.org/nobel_prizes/medicine/laureates/1983/mcclintock-bio.html

the phenomena of variegation in the pigmentation of kernels in maize. The variegation phenotype (\uparrow) 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. 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 organism, due to the presence of the “alternative” allele.⁵²⁵

Transposons can have a number of different effects on the expression of the genes in which they are found.⁵²⁶ 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.⁵²⁷ 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 (\rightarrow). 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, which can be located anywhere in the genome, for example, on another chromosome. 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; if there is an autonomous transposon within the cell, however, then the transposase it encodes can catalyze the excision and insertion of a dependent (non-autonomous) transposon. Why? because when the transposase protein is synthesized (in the cytoplasm) it can move around the cell (and within the nucleus) and interact with multiple transposons (DNA regions).

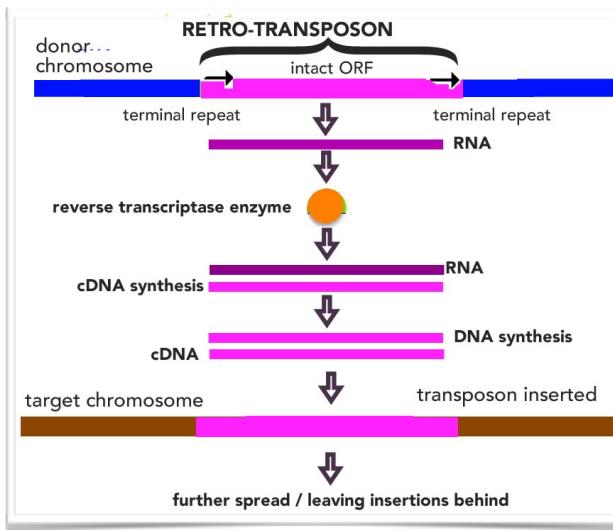


The second type of transposon, known as a type I transposon, is also a DNA sequence, but it uses a different mechanism to move. Again type I transposons come in autonomous and non-autonomous

⁵²⁵ 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.

⁵²⁶ Transposable Elements, Epigenetics, and Genome Evolution: <http://science.sciencemag.org/content/338/6108/758>

⁵²⁷ The Maize Transposable Element Ds Is Spliced from RNA: <https://www.ncbi.nlm.nih.gov/pubmed/3039661>



(dependent) forms (\leftarrow). 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 in that gene 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 an 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); they are dependent, type I transposons that rely on the presence of autonomous transposons to move.⁵²⁸ 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.⁵²⁹ Moreover, since transposon movement is generally stochastic, as populations separate from one another, the patterns of transposons within the genome diverge from the ancestral population.⁵³⁰ 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.⁵³¹

⁵²⁸ Wikipedia: [Alu element](#)

⁵²⁹ [Active transposition in genomes](#)

⁵³⁰ The impact of retrotransposons on human genome evolution: <https://www.ncbi.nlm.nih.gov/pubmed/19763152>

⁵³¹ Stress and transposable elements: co-evolution or useful parasites? <https://www.ncbi.nlm.nih.gov/pubmed/11012710>

Questions to answer:

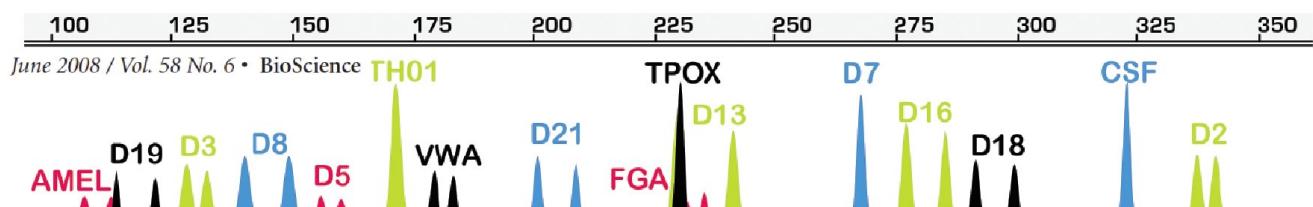
181. How could the movement of a transposon influence gene expression?
182. What are the selective pressures on the maintenance or destruction of active transposons?
183. How could the movement of a transposable element NOT produce a mutation?

Questions 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 dynamics?

Forensics and ancestors (to be completed - stay tuned)

A good paper to read: [Crime Scene Genetics: Transforming Forensic Science through Molecular Technologies](#)

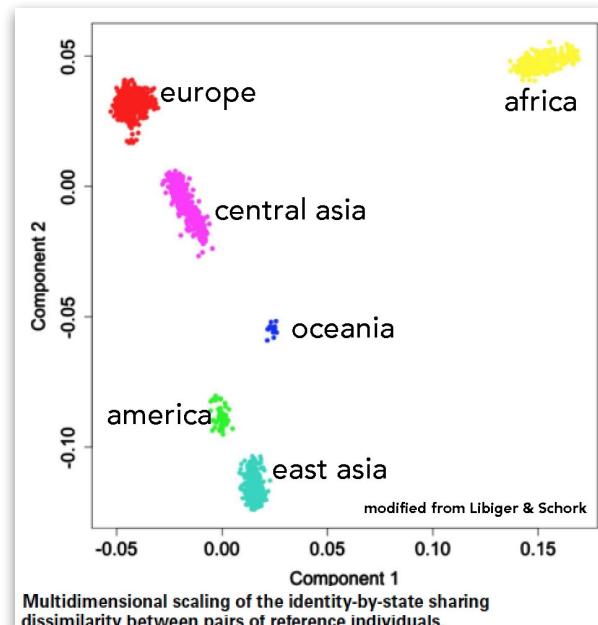


Forensics: How does DNA fingerprinting work?

- Naked Science Scrapbook.⁵³²

forensic studies: <https://youtu.be/ZxWXCT9wVol>

Supplemental materials



⁵³² [A Method for Inferring an Individual's Genetic Ancestry and Degree of Admixture Associated with Six Major Continental Populations.](#)