MRU Open

Genetics Lectures

Note that this is organized to set up online exercises at `http://opengenetics.net`

This project was funded by the Alberta Open Educational Resources (ABOER) Initiative, which is made possible through an investment from the Alberta government

OPEN GENETICS History

The first edition of this textbook, called OPEN GENETICS, was produced in January, 2009 as instructional material for students in Biology 207 at the University of Alberta, and is released to the public for noncommercial use under the Creative Commons License (See below). Users are encouraged to make modifications and improvements to the book. All the text in the original 2009 edition was written by Michael Deyholos, Ph.D. In subsequent editions (2010-2014), additional chapters were written by Mike Harrington, Ph.D., at the University of Alberta. Additional content and editing by John Locke, Ph.D. and Mark Wolansky, M.Sc., at the University of Alberta. Photos and some diagrams were obtained from various, non-copyrighted sources, including Flickr, Wikipedia, Public Library of Science, and Wikimedia Commons. Photo attributions are listed in the legend with each image.

Open Genetic Lectures (OGL) Origin: 2015

OGL is an alternative approach to an open source textbook. Much of its content is derived from the OG textbook. The 13 chapters in OG were cut up and distributed into 41 shorter chapters that parallel the current lecture topics in BIOL 207 (Molecular Genetics and Heredity) at the University of Alberta. More text content, figures, and chapter-end questions were added in this revision.

This reorganization of OG content into OGL was accomplished during the summer of 2015 by John Locke, Mike Harrington, Lindsay Canham, and Min Ku Kang. Lindsay Canham was supported by a grant from the Alberta Open Education Initiative (OEI) through the University of Alberta. Min Ku Kang was supported by a Summer Student Scholarship from the Centre for Teaching and Learning (CTL), University of Alberta. Without these sources of financial aid this project would not have been possible. John Locke and Michael Harrington appreciate their help, as well as that of Michael Deyholos, who initiated this endeavor.

Todd Nickle re-ordered some chapters to correspond with the sequence he uses in class and to relate to the online activity site http://opengenetics.net.

In the future, molecular chapters will be incorporated into the text. For now, they are individual files not included in this booklet.

Access to OGL (MRU) text files

The final versions of this work and the website are available as .zip files at http://opengenetics.net/Files

They include all the .docx, .indd (InDesign), css, .html, and image files for each chapter and other relevant files. They are available for anyone to use, adapt, or improve for educational purposes. If you have edits, improvements or additions that you wish to share under the same license terms, please contact John Locke, University of Alberta or Todd Nickle, Mount Royal University (tnickle@mtroyal.ca).

Table of Contents

Chapter 1: Nomenclature

Chapter 2: Alleles at a Single Locus

Chapter 3: Mendel's Laws

Chapter 4: Sex Chromosomes

Chapter 5: Gene Interactions

Chapter 6: Pedigree Analysis

Chapter 7: Cell Cycles

Chapter 8: Human Chromosomes

Chapter 9: Gene Interactions

Chapter 10: Eukaryotic Mapping

Chapter 11: DNA is the Genetic Material

Chapter 12A: DNA Structure and Replication

Chapter 12B: PCR

Chapter 13: DNA Sequencing

Chapter 14: Prokaryotic Operon Regulation

Chapter 15: Recombinant DNA

Chapter 16: Molecular Basis of Recombination

Chapter 17: Chromosome Rearrangements

Chapter 1

Part 1: Simple Nomenclature

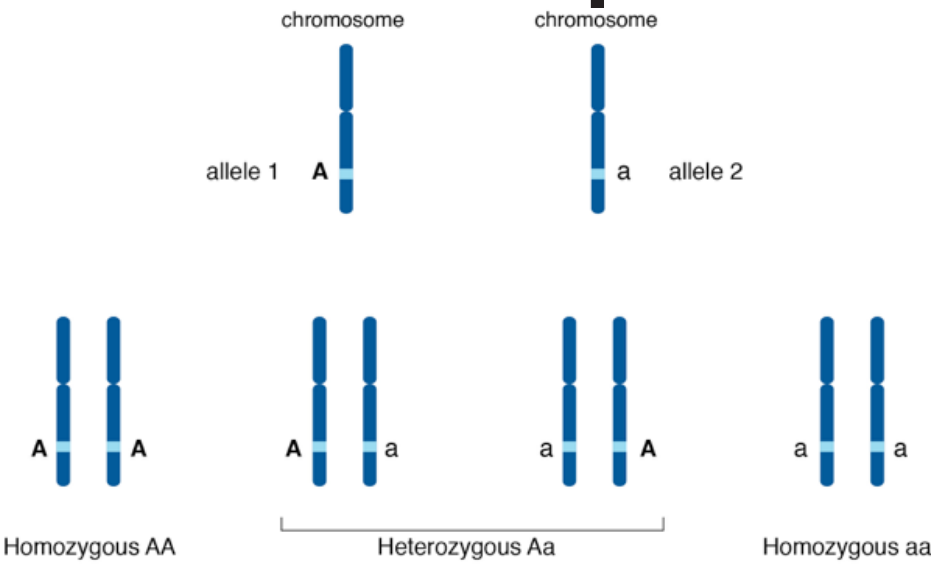


Figure 1-1
Two chromosomes with the same gene but different alleles, A and a.

(Image used with permission of www.genome.gov provided it includes the link http://xoax.net/me

diaGallery/Allele-dat-342)

INTRODUCTION

Classical genetics requires higher-level thinking and an organized strategy. You have probably encountered symbols in high school genetics and perhaps in previous chapters assigned to you in this course. Chapter 13 points out that a position on a chromosome is a gene locus, which simply means the location at which a gene is found. At that position, the gene could be one of several variants that we call alleles. The set of alleles comprise an organism's genotype, which might affect the organism's phenotype through the action of the genotype.

When you are solving genetics problems you have to keep track of different alleles of one or more genes. This can become confusing, so understanding how a gene works, how alleles vary, how they are expressed, and the influence they have on phenotype is important. A consistent nomenclature system – a set of rules dictating how you name alleles – becomes immensely helpful. Online Open Genetics has an activity that demonstrates why a good nomenclature system is helpful. The icon to the right indicates an online module.

A WHY USE NOMENCLATURE?

http://tinyurl.com/oog-name

Word and practical problems in biology can get confusing in a hurry, particularly if you're distracted by something like exam stress! When you are manipulating several ideas, it is good practice to be thoughtful and follow rules that keep you consistent in your interpretation. As an instructor, I have seen work in which the student clearly got flustered and forgot that the mutation he or she was working on was dominant. This often leads to an answer that is inconsistent with the data.

There are a few simple rules we can use for nomenclature. Essentially, we use capital letters to indicate an allele that, in a heterozygote, expresses the proteins from that particular allele. This first appendix uses a simplified system to communicate allele characteristics. The next appendix will show a more complicated system that carries even more information in the gene symbols.

B THE BIOCHEMISTRY OF GENE BEHAVIOUR

B.1 THE MECHANICS OF GENE EXPRESSION

The biochemical action of gene expression is in Chapter 3. We often think of genes as "made of DNA" and they reside in the nucleus and endosymbiotic organelles of eukaryotes or the nucleoid region of prokaryotes. They are transcribed into an RNA message by RNA polymerase then interpreted by ribosomes that assemble particular amino acids into a polypeptide strand (also known as a protein) based on the sequence of nucleotides. In a cell, proteins can act as enzymes, structural features, pigments, and a host of other functions, including regulating the expression of other genes. This expression of genes leads to how an organism looks – its phenotype.

Chapter 2 points out that organisms usually fall into the classes of being diploid or haploid. Humans and eukaryotic genetic systems usually assume the organism is diploid, which means that most chromosomes are represented as pairs. Each pair has homologous loci: the term homologous means "the same information" and refers to the gene at each locus. Note that although the loci, or genes, are the same, the alleles that comprise them may be different!

B.2 DOMINANT AND RECESSIVE; HOMOZYGOUS, HETEROZYGOUS, AND HEMIZYGOUS

Alleles themselves do not directly exert an effect on the phenotype of an organism. You will recall that genes are instructions, often for proteins. It is the effect of the protein that causes an organism to take on particular traits. In this chapter we will look at how to design symbols appropriate to communicate the characteristics of alleles you want to investigate, but keep in mind that the allele instructs what kind of protein to make. If an organism is diploid and homozygous for an allele (**Figure 1-2**, left and right), the gene at the same position of the homologous chromosomes is the same allele. Only one type of protein is made. If an organism is diploid and heterozygous (Figure 1-2, middle), and the protein from one allele influences the phenotype more than the protein from the other allele, we use the terms dominant and recessive, respectively. Note that we don't use the terms homozygous or heterozygous in haploid organisms. If they are haploid, their phenotype will reflect the genotype of the only allele present. The proper term for their genotype is hemizygous to reflect only one copy of each gene.

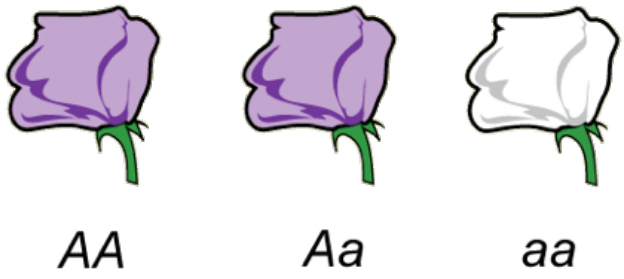


Figure 1-2
Relationship between genotype and phenotype for an allele that is completely dominant to another allele. (Original-M. Deyholos -CC:AN)

A mutation is a change in nucleotide sequence. Chapter 11 goes into more detail of what this means. What's important now is that you understand that the amino acids of a protein can be different if we compare different alleles of a gene and they may behave differently – often one protein will "work better" than the other. If the promoter of a "functional" allele of a gene is damaged, the allele that is created might not even create an mRNA so no protein will be encoded by that allele. This is called an amorphic or "null" mutant (See Chapter 2).

If one copy of an allele makes enough protein to compensate for the absence of protein from the other allele in a heterozygote, it will influence the phenotype. If this phenotype looks identical to that of an organism homozygous for the "functional" allele, we consider the "functional" allele to be dominant to the "null". We could also say the "null" is recessive to the "functional" allele. Keep in mind that the alleles themselves aren't doing anything, but it's common practice to label the alleles as dominant or recessive, although in reality we are talking about the expression of those alleles.

http://tinyurl.com/oog-biochem

Why this is important is how the proteins from two alleles interact. If both proteins are identical (from a homozygous genotype) the phenotype that results will be that of the action of one "type" of protein, even if though there are two copies of the gene – they are both the same allele. There's no real interaction because the proteins do the same thing. It's in heterozygotes that we can see whether a particular allele is dominant over another or otherwise influences the phenotype in an interesting way.

Here's another point about the "normal" allele: it was

honed by natural selection over a long period of time. Genes are instructions for the protein tools of an organism's cells. For this reason we often call the "normal" allele of a gene the "wild type" allele. This would be the allele most common in the wild, presumably because it provides a benefit to the organism. Thus, most mutations are likely to reduce the effectiveness of the wild type allele, although the process of evolution allows (and, in fact, requires) an occasional beneficial allele to permeate a population if it provides a selective advantage.

B.3 THERE ARE MANY KINDS OF ALLELES FOR A GENE

Keep also in mind that a gene can be mutated in different parts of its sequence to create different alleles. A diploid organism can have a maximum of two alleles (aside from gene duplication or abnormal chromosome structure, but ignore that for now). In a population, though, there can be many, many different alleles. Perhaps allele one decreases protein function and allele 2 is even less effective. This describes an allelic series as follows: wild-type > allele 1 > allele 2 > any null allele. Wild-type alleles in this case encode the most effective protein. Null alleles represent catastrophic mutations that eliminate transcription or produces proteins that can't function at all. See section 3.2 on page 4 below for another example of an allelic series.

B.4 GENDER AND GENE INTERACTIONS

Finally, here are a couple points about other interactions before we move on to gene symbols. Some organisms have a pair of sex chromosomes that dictate gender (see Chapter 4). We'll ignore how we indicate this as a symbol for now; the second part of this chapter will show a trick to indicate this.

Incomplete dominance, codominance, and epistasis are not accommodated in any of the nomenclature systems we present.

C BASIC NOMENCLATURE

C.1 SINGLE LETTER SYSTEM

Sometimes what you want to do is a little rough work for investigating your genetic model. A genetic model is a diagram of the logic that you propose for inheritance. For instance, if you cross a true-breeding purple plant with a true-breeding white plant (*e.g.* see Figure 1-2 on page 8; cross the outer two plants) you will get a heterozygote (the middle plant in Figure 1-2 on page 8, also shown at left). If we name the gene after the recessive trait (*a* is the first letter in *albino*), we know that the heterozygote will have one capital letter "*A*" and a lower case "*a*". The heterozygote is the F_1 generation ("first filial", which means it's the first child from parents that are crossed). The F_1 is purple, which means the "*a*" allele is recessive; only one copy of the "*A*" allele is needed for enough purple pigment to make it identical to one true-breeding parent. This is complete dominance.

We can't know from the information given which allele is wild type or mutant. One hypothesis, though, is that purple pigments are required to attract pollinators and therefore would help the plant in the wild. Albino plants could be a mutant and might not generate as many seeds for lack of pollinators. If the context of your genetics problem doesn't indicate which allele is wild-type, it's good practice to name your allele based on the recessive trait. Often the recessive allele is the mutation. We might consider that the "*a*" allele is null and makes no pigment at all. One or two "*A*" alleles make enough protein to cause the plant to be purple.

Figure 1-2 on page 8 already assumes that the capital letter (*A*) stands for an allele that encodes a protein for purple pigment and the recessive allele (*a*) doesn't make pigment. Thus the *Aa* heterozygote is sufficient evidence to adopt upper- and lower-case letter "*A*"s to communicate the characteristics of purple and white alleles.

A note of caution: When you're writing down gene symbols for homework or on an exam, be sure to make the characters distinct. A typewritten "*y*" is easy to distinguish from the upper case "*Y*" but not as easy when writing it down. Instructors who ask you to show your work need to be able to follow your logic train. More important than that is YOU have to be able to follow your own reasoning. Students often switch symbols and come up with an answer that is inconsistent with the data given because of this. Consider underlining your capitals or putting a line through one of them to make it distinct (e.g. $\not{Y}$ for the dominant allele; *y* for the recessive).

C.2 NAME THE GENE AFTER THE MUTANT PHENOTYPE

Some instructors would accept "*P*" for "*purple*" for the previous cross. However, the better answer is to follow an established system. During "exam fog", it's easy to get lost if you are inconsistent with how you devel-

op your symbols. During your study period and when you're practicing genetics problems, be thoughtful about the gene names you choose.

Let's always choose a letter based on the mutant phenotype for our gene symbol. If we are presented with a ladybug mutant that is small, we might choose "*d*" for "*dwarf*". Geneticists sometimes set up a research program based on unusual phenotypes of the organism they are studying. The fact that a mutant phenotype that is heritable exists tells us that there is a genetic control for the trait and that it might be isolated in the lab. When you look at your classmates, you don't necessarily note that none of them has an arm growing from the tops of their heads. If one student had this trait, however, you couldn't help but notice it. If you found out it showed up in that student's ancestry in a predictable fashion, you might reasonably suggest that there is a genetic basis for that. If it happened to be controlled by a single gene, you might call the gene "*extra arm*" or "*arm head*". If it happened to be a dominant trait, you might use the letter "*X*" (for "*Xtra*") or "*A*" (for "*Arm*") for the mutant allele. The wild type allele would be "*x*" or "*a*", respectively.

C.3 If you don't know which is mutant, use the recessive trait for the gene name

What if you don't know which allele is mutant? What if you're presented with two true-breeding frogs: one that is gold and one that is yellow. If you don't know what the predominant colour in nature is you can't know which one is mutant. If you crossed them and all the progeny are gold, then you know the dominant allele encodes a protein to make it gold. The recessive, therefore, is "*yellow*" and you should name the gene "*y*" after the recessive phenotype. This means the dominant allele would be "*Y*". Your offspring would therefore be *Yy* and the gold parent would be *YY*. The yellow parent would be *yy*.

Apply these ideas at the Online Open Genetics exercises.

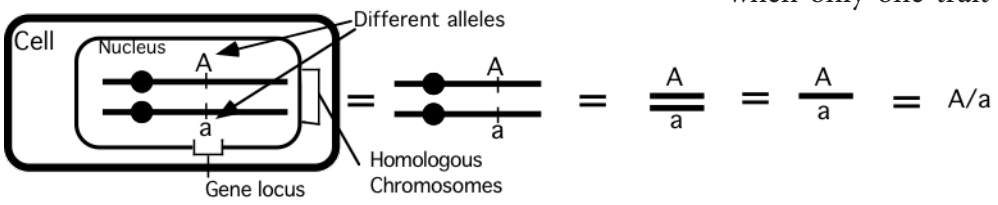


Figure 1-3

A diagram of how chromosomes, loci and alleles look in the cell, and how we depict them in written form. Note that only one gene is represented in this figure and chromosomes have not replicated. (Original-J.Locke- CC BY-NC 3.0)

D LINKED GENES

http://tinyurl.com/oog-basic

Mendel was lucky. He studied a variety of traits in pea plants and his data were consistent with his idea of traits being encoded by pairs of discrete heritable units. He didn't call these "genes" and had no idea about their chromosomal origins or chemical makeup. It turns out that the genes he studied were either on different chromosomes and so assorted independently (See Chapter 9), or so far apart on the same chromosome that linkage could not be detected (See Chapter 10).

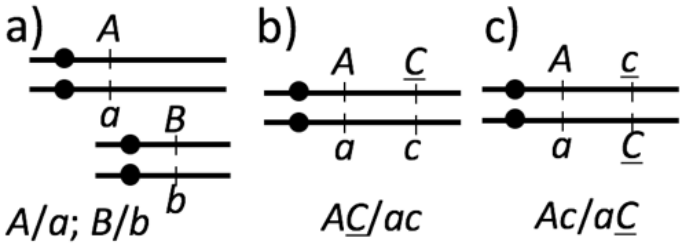


Figure 1-4

Three gene arrangements for cells of genotype *AaBb*. Chromosomes are replicated (shown with sister chromatids). a) demonstrates unlinked genes: *A/a; B/b*. Linked genes are shown in b) coupling *AC/ac* and c) repulsion *Ac/aC*. (Modified by T. Nickle from an original by J.Locke- CC BY-NC 3.0

Symbols for a gene can be drawn on a page to communicate their position on a chromosome (**Figure 1-3**). To do this, we use a forward slash (/) to demonstrate what is on each chromosome. Figure 1-4 shows how we might conceptualize the position of a gene on two chromosomes by collapsing the chromosomes into a single line.

There's no question about where the gene is located when only one trait is under investigation: it will be at the same position on each homolog. Two genes, however, can be one of three possibilities (**Figure 1-4**). Each possibility has implications for gene mapping and predicting ratios from a dihybrid cross. Figure 1-4 shows the positions of genes for an unlinked situation as well as linked genes in coupling and

repulsion configurations. If genes are unlinked, put the allele symbols for one gene on either side of one slash followed by a semicolon (indicating that it's unlinked) and the other gene with the alleles separated by a second slash (*A/a; B/b*). When genes are linked, only one slash is used: remember, the slash stands for a pair of homologous chromosomes. Genes in coupling would have the dominant genes together on one side of the slash and recessives on the other side (*AB/ab*). Repulsion would represent the other arrangement (*Ab/aB*).

Practice your skills with identifying linked and unlinked genes online. Some examples of different forms of gene symbols are shown in Table 1-2 on page 14. Keep in mind that sometimes you have flexibility in which system of nomenclature you use, but sometimes it is dictated to you, for example in publications or other formal submissions. You are discouraged from inventing your own system or mixing up different systems because it will confuse your readers (or graders!).

A more advanced system of nomenclature is outlined in the next section. New rules are introduced to help you identify sex linked genes and predict phenotypes from gene symbols alone.

http://tinyurl.com/oog-linked

Part 2: Advanced Nomenclature

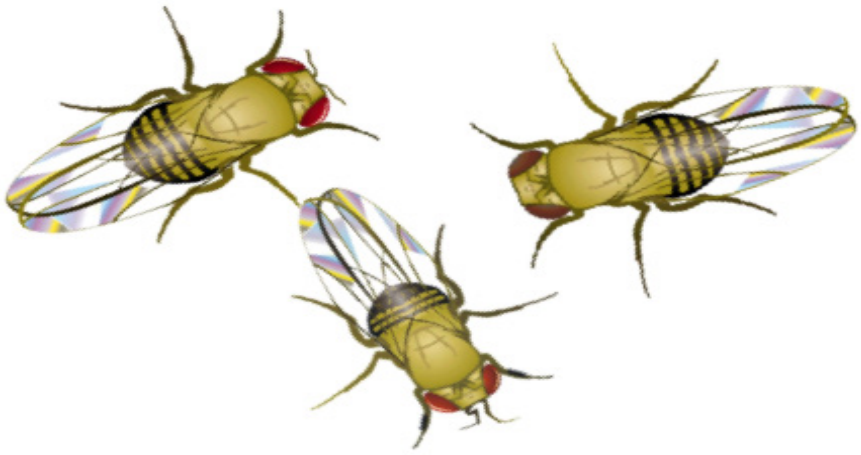


Figure 1-5
Wildtype and mutant "flies" of *Sumino-spauci combibo*, an imaginary model system. These will be used in this chapter to show how to apply advanced nomenclature.

(Image created by D. Bird based on templates generously provided by The Imaginarium. Used with permission)

In the first part of this chapter, you saw how several nomenclature systems can be used to communicate the nature of the mutant allele such as the mutant phenotype and whether the mutant allele is dominant or recessive to the wild-type. This appendix extends the toolbox of symbols by introducing a powerful system that will help you practice thinking about genetics like a professional. It's a strictly-defined method of naming alleles and is designed for you to be able to look at a single gene symbol and determine its inheritance automatically. We will call it the Sumnospauci combibo convention (*Sóo-me-no spów-see*, a hypothetical organism used for genetics practice questions). We'll build up to this by working through extensions to the rules.

Many students find this formalized nomenclature scheme cumbersome at first. The rules are straightforward, but practice makes this system simple. I've had many students report that they hated the it at first, but when using it they found it easier to track the alleles during complex problem solving.

A SUPERSCRIPTS

Sometimes a letter is used as the name of a gene, and superscripts can modify it to indicate the different alleles (**Figure 1-6**). One common single letter code for an allelic series is "*I*". Red blood cells can have their cell membranes modified by sugar tags that give rise to our blood type. One allele of *I* gives rise to blood type A and is therefore called I^A. An enzyme encoded by I^B modifies sugars to create blood type B. A heterozygote I^AI^B demonstrates both sugar tags because those alleles are expressed – they are codominant. People with blood type O only possess alleles for the *I* gene that don't work and are therefore recessive – they don't modify the extracellular sugar tags. Because it is recessive, individuals are homozygous for *i*: they are *ii*.

Sometimes a superscript "plus sign" is used to denote the wild type allele. One might use the symbol W^+ to indicate a wild-type allele that promotes wing growth. Note that the generic "*wing*" gene name isn't a best practice – name the gene after the mutant phenotype! A wingless mutant would be W^-. You should never use a "+" and shift the case of the letter unless you are dealing with a special case such as the codominance in the blood type example above. The capital "*I*" letter indicates it is dominant to "*i*". The superscript A and B for the codominant alleles indicate the dominant alleles are different from each other.

Superscripts can be symbols, a single letter, or many letters (a few examples are in **Table 1-1**). They modify the gene name only in the superscripted symbols: the regular-sized letters are identical between them (see Table 1-1). This means that Abc^+ and *abc* would be different genes (*i.e.* not allelic); Abc^+ and *Abc* are alleles, as are abc^+ and *abc*. Note that a superscript is not mandatory for all alleles of that gene. It's use depends on the convention being used.

Alleles of bacterial genes are typically indicated with a superscript + or -. For example, a bacterial allele that creates an enzyme that makes methionine would be met^+, and a defective allele of that gene is met^-.

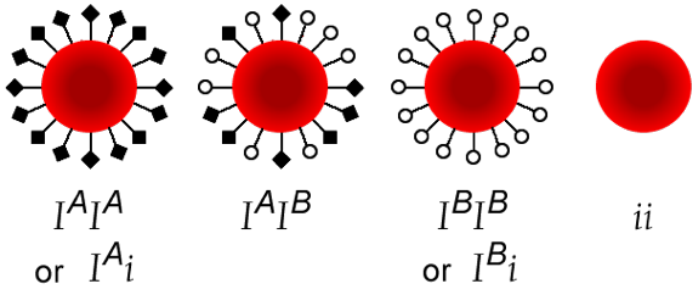


Figure 1-6
Relationship between genotype and phenotype for three alleles of the human ABO gene. The I^A and I^B alleles show co-dominance. The I^A allele is completely dominant to the *i* allele. The I^B allele is completely dominant to the *i* allele. (Original-Deholos -CC:AN)

B THREE-LETTER SYMBOLS

It's perfectly acceptable to use a single letter or even two letters. Sometimes, though, multiple traits spelled with the same first letter can get confusing. Using three letters for a gene symbol can make it easier to remember what the letters stand for. In fact, for some model systems those who study them adopt a defined nomenclature system. A plant often used for genetic studies, called *Arabidopsis thaliana*, has a three-letter code (`https://www.arabidopsis.org/portals/ nomenclature/namerule.jsp`).

Just as we saw for the one-letter symbols, the dominant allele has the first letter capitalized and the last two letters are lower-case. Recessive alleles are all lower-case. With three letters, you can make gene names that are easier to keep track of. For example, you might see a fly with an extra set of wings. Instead of calling it "*w*" for "*wings*" (which is a poor choice because it doesn't represent the mutant phenotype), you can instead call it "*exw*" for "*extra wings*". Then, when you see it, you can sound out the abbreviation and remember that it stands for the mutation. If the wild-type allele for this is dominant, then you would write that one "*Exw*". The first letter indicates that it is the dominant allele. What you can't tell just from these examples is

Table 1-1
Examples of genes using a superscript modifier.

$white^{apr}$ or $white^{apricot}$	An allele of the *white* gene which has an "apricot" phenotype
Abc^+ *Abc*	Two alleles for the *Abc* gene (wild-type and mutant, respectively). Note the mutant allele is dominant.
w^+; w^a; *w*	Three alleles in a series for the *w* gene. The first is wild type; the second two are different mutant alleles.
bio^+; bio^-	A wild-type allele of a *biotin* gene and its recessive counterpart. This is likely a bacterial gene due to the convention.

that wild-type allele is dominant! *Hint:* remember in the previous section that the "+" superscript indicates the wild-type allele.

http://tinyurl.com/oog-advanced

WHY USE THE *SUMINOSPAUCI COMBIBO* SYSTEM?

In the first part of the chapter you saw different systems of expressing gene inheritance. Some rules we established were:

- Alleles are instructions for a protein. It is the protein that determines the phenotype of a trait.
- Traditionally, we call an allele "dominant" or "recessive", but these terms actually refer to the action of the protein the allele encodes. The alleles are merely "instructions".
- Symbols should be chosen based on the phenotype of the mutant. If the mutant allele cannot be identified, the symbol should reflect the recessive phenotype.
- Symbols for "recessive" alleles are in lower case.
- Symbols for "dominant" alleles have a capitalized first letter (if the symbol has more than one letter, all remaining letters are lower case).
- Some mutant alleles act as recessive to wild-type, and some mutant alleles act as dominant to their wild-type counterpart.
- A "+" superscript indicates that an allele is wild-type. If this style of nomenclature is chosen, then all alleles are identical (including the type of letter: *e.g.* a/a^+ reflect a recessive mutation and the corresponding wild type in a heterozygous individual; a/A^+ is incorrect).

We noted that if a mutant allele is recessive to the wild-type, and the wild-type allele is dominant to the mutant allele. Conversely, if the mutant allele is "dominant", the wild-type allele is "recessive". When writing these things out fully, the descriptions get complex; and the more complex, the more likely you are to make a mistake. Likewise, if you're writing out a description and start talking about the mutant alleles and switch to the wild-type allele later on, chances are good you'll confuse the reader – and likely yourself!

There is a simple solution: always name gene after the mutant phenotype. If you can't tell which phenotype is mutant, name the gene after the recessive. The mutant allele will have a capitalized first letter if it's dominant. If it is recessive, it will be all in lower case. Be sure to italicize or underline the gene symbol. Once the mutant allele is set, use exactly the same letters (including the upper- or lower-case first letter) and put a superscript "+" behind it.

The "+" means wild type. Always. But it also means "not". Bear with me.

If you see a gene symbol and you know it follows the *S. combibo* convention, you can immediately know if it is dominant or recessive. If you see *wht* as a gene symbol, you know, without any extra information, that:

1. The mutant allele:
   a) Is recessive (because the symbol is in all lower case)
   b) Probably looks like something that wht would describe (perhaps "white"?)
2. the wild type allele:
   a) Has the symbol wht^+
   b) Is "not" mutant (because of the "+", but it should be obvious because we name the allele after the mutation)
   c) Is "not" recessive (therefore it is dominant to *wht*)
   d) Is "not" white (or whatever *wht* stands for)

That's a lot of information in just three letters! Note that we can get the same information from the wild type symbol alone. Consider Pnk^+. It seems like it refers to something that's pink when it displays the mutant phenotype (but this allele is NOT the mutant; it has the "+" so it is NOT pink). The capital "P" tells us the mutant allele is dominant, and so the wild-type allele is NOT dominant (it is recessive).

Finally, note that we're using the term "gene name". The gene is named after the mutation. Whether it is *wht* or wht^+, the "gene name" is *wht*. For *Pnk* or Pnk^+, the gene name is *Pnk*. We use the superscript when it is required. Try to get into the habit of asking exactly what the mutant symbol is trying to communicate.

C MORE POWER: X-LINKAGE CHANGES THE GENE SYMBOL

So far we've dealt with symbols that involve autosomal genes. In earlier courses, you might have encountered sex-linkage (also known as X-linkage, because most of these genes are on the X chromosome). More information about sex-linked genes is in the next chapter (Chapter 4).

Sex chromosomes, by definition, are represented differently between males and females. In mammalian systems - the human system being most familiar to you - females have two X-chromosomes and males have one X- and one Y-chromosome. Be aware that this isn't universal for all organisms!

Because males in the X/Y system do not have homologous chromosomes, it's obvious when a gene is X-linked: the male has only one allele compared to having two alleles for his autosomes. But how would you indicate that a female's genotype contains X-linked genes? The system so far doesn't contain enough information to illustrate this.

We'll conform to the three letter system but put these letters as a superscript after a large"X" for X-linked genes. Since there are very few Y-linked genes, the Y chromosome is represented as just a "Y" with no superscript. Wild type alleles will still have a "+" after the three letters. For example, a male with an X-linked gene for a wild-type gene encoding black eyes would be represented as X^{bey+}/Y. A female heterozygous for this gene would be X^{bey+}/X^{bey}.

More examples are in **Table 1-2**.

Practice writing out genotypes and making gene symbols at

`http://opengenetics.net`.

http://tinyurl.com/oog-sex

Table 1-2
Examples of symbols used to represent genes and alleles.

Examples	Interpretation
A and *a*	Uppercase letters represent dominant alleles and lowercase letters indicate recessive alleles. Mendel invented this system but it is not commonly used in publications because not all alleles show complete dominance and many genes have more than two alleles. It's quick and easy for you to use when working out genetics problems when you are sure each gene involves only two alleles.
a^+ and *a*	Superscripts are used to indicate alleles. For wild type alleles the symbol is a superscript +. The mutant allele of gene *a* would be recessive.
met^+ and met^-	This is typical of a prokaryote gene symbol. It could be referring to wild-type (functional) and mutant (nonfunctional) alleles of a gene that makes a protein in the methionine synthesis pathway.
AA or *A/A*	Sometimes a forward slash is used to indicate that the two symbols are alleles of the same gene, but on homologous chromosomes. Both representations in this row are identical: it represents a homozygous dominant.
Aa/Aa or *Aa/aa*	Note that this example shows two alleles of the gene *Aa*. We know that the gene symbol is two letters because the slash separates the allele found on each of the homologous chromosomes. We cannot tell if the mutant phenotype is recessive because there's no indication which is wild type.
Grn^+ *shr/Grn shr*	The three-letter system is used here. "*Grn*" might mean that the phenotype is "green", but we can't be sure. What we do know is that the mutant allele codes for a protein leading to a dominant phenotype. The wild-type allele must be recessive to the mutant allele. Maybe "*shr*" means "shrunken" or "short", but we know that the mutant phenotype can only be seen in the homozygous recessive configuration. The phenotype for this organism is mutant for both *Grn* and *shr* traits. Final note: the genes are on the same chromosome based on the position of the slash.
bob^+*/bob; mia/mia*	This also uses the three-letter system. The organism is heterozygous for *bob* but shows the wild-type trait in its phenotype. It is homozygous recessive for *mia* and therefore shows that mutant phenotype. The genes are unlinked.

SUMMARY:

- Symbols are used to denote the alleles, or genotype, of a locus.
- Phenotype depends on the alleles that are present, their dominance relationships, and sometimes also interactions with the environment and other factors.
- In a diploid organism, the alleles can be homozygous, heterozygous or hemizygous.
- Allelic interactions at a locus can be described as dominant vs. recessive, incompletely dominant, or co-dominant.
- Two different genes can be on the same chromosome (linked) or on different ones (unlinked)
- Steps:
  - Can you identify the mutant? Name the gene after its phenotype
  - If you cannot tell which allele is mutant, name the gene after the recessive allele
  - Capitalize the first letter for the dominant allele; use lower-case for the recessive allele
  - If genes are linked, write the gene symbols together on each side of the slash. If genes are unlinked, they appear on different sides of a semicolon

Chapter 2
Alleles at a Single Locus

Figure 2-1
A flower called Camellia showing co-dominance of the red and white alleles of flower colour.

(Flickr- darwin cruz-CC BY 2.0)

Introduction

A mutation is any change in the nucleic acid code for an organism. If the change of "letter" does not alter a gene in a fashion of changing a characteristic, it is called a silent mutation. However, sometimes the change alters a trait. These variations in traits first lead Gregor Mendel to formulate his theories of inheritance. More background about this is available in the prerequisite reading and Chapter 1.

A TERMINOLOGY

A specific position along a chromosome is called a locus. Because each gene occupies a specific locus along a chromosome, the terms locus and gene are often used interchangeably. However, the term "gene" is a much more general term, while "locus" usually is limited to defining the position along a chromosome. Each locus will have an allelic form (allele); a specific DNA sequence. In a population of individuals there will be sequence variation so there will be different alleles. Some may be defined as wild type, some as variants, others as mutant.

The complete set of alleles at all loci in an individual is its genotype. Typically, when writing out a genotype, only the alleles at the locus (or loci) of interest are considered and written down – all the others are still present and assumed to be wild type. So, typically only the alleles at the few mutant loci appear in the written genotype. All the many, many others that are wild type are not.

The visible or detectable effect of alleles on the structure or function of that individual is called its phenotype – what it looks like. The phenotype studied in any particular genetic experiment may range from simple, visible traits such as hair color, to more complex phenotypes including disease susceptibility or behavior. If two alleles are present in an individual, as is the case with diploid organisms, then various interactions between them may influence their expression in the phenotype.

B SOMATIC VS. GERMLINE MUTATIONS

A mutation occurs in the DNA of a single cell. It can be passed on to the cell's direct descendents through mitosis, and it also may be passed on to the organism's descendants in the process of meiosis. In single-cell organisms a mutation is passed on directly to its descendants. In multicellular organisms, there is a partitioning into somatic cells and germline cells that occurs very early in development of animals. In plants, this separation occurs later, in the cells that form the flower.

B.1 SOMATIC MUTATIONS

Somatic cells form the tissues of the organism and are not passed on in reproduction. Any mutations in somatic cells will only affect the individual in which they occur, not its progeny. If mutations occur in somatic cells, its mutant descendants will exist along side other non-mutant (wild type) cells. If the mutation occurs at a very early stage of development the impact of the mutation will involve more cells. This gives rise to an individual composed of two or more types of cells that differ in their genetic composition. Such an individual is said to be a mosaic. An example is shown in **Figure 3-2** and cancer cells are another example.

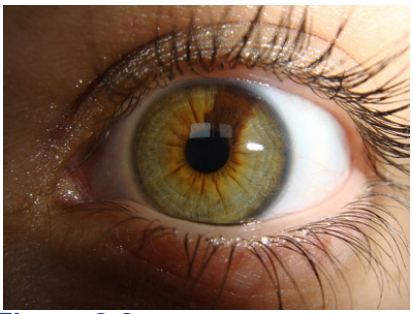


Figure 2-2
Patch of brown eye colour in a green eye. (Wikipedia-Sheila. lorquiana-CC BY-SA 3.0)

B.2 GERMLINE MUTATIONS

Germline cells are those that form the eggs or sperm cells (ovum or pollen in plants), and are passed on to form the next generation. Therefore, mutations in germline cells will be passed on to the next generation but won't affect the individual in which they occur.

In animals, somatic cells are segregated from germ line cells. In plants, somatic cells become germline cells; so somatic mutations can become germline mutations.

B.3 HAPLOID VS. DIPLOID ORGANISM

In haploid organisms, there is only one copy of a gene so a mutation will directly affect the organism's phenotype. Therefore, one can infer the genotype of the organism just by looking at its phenotype. However, in diploid organisms, there are two copies of each gene. This can result in an interaction between two alleles of a gene so that a mutation may not have a direct impact on the organism's phenotype. The mutated allele can demonstrate complete dominance, incomplete dominance, co-dominance, or it can be recessive. Therefore, inferring one's genotype based upon its phenotype is not always simple in diploids.

C ALLELES: HETERO-, HOMO-, HEMIZYGOSITY

Mendel's First Law (segregation of alleles) is especially remarkable because he made his observations and conclusions (1865) without knowing about the relationships between genes, chromosomes, and DNA. We now know the reason why more than one allele of a gene can be present in an individual: most eukaryotic organisms are diploid and have at least two sets of homologous chromosomes. For organisms that are predominantly diploid, such as humans or Mendel's peas, chromosomes exist as pairs, with one homolog inherited from each parent. Diploid cells therefore can contain two different alleles of each gene, with one allele on each member of a pair of homologous chromosomes. If both alleles of a particular gene are the same (indistinguishable), the individual is said to be homozygous at that gene or locus. On the other hand, if the alleles are different from each other, the genotype is heterozygous.

In cases where there is only one copy of a gene present, for example if there is a deletion of the locus on the homologous chromosome, we use the term hemizygous. Another example is the single X- and Y- chromosomes in X/Y males in which almost all the loci on those chromosomes are hemizygous. (The exception is the pseudo-autosomal region which is an evolutionary holdover before X/Y sex determination – see the chapter on sex chromosomes.)

Although a single diploid individual can have at most two different alleles of a particular gene, many more alleles can exist in a population of individuals. In a natural population the most common allelic form is usually called the wild-type allele. However, in many populations there can be multiple variants at the DNA sequence level that are visibly indistinguishable as all exhibit a normal, wild type appearance. There can also be various mutant alleles (in wild populations and in lab strains) that vary from wild type in their appearance, each with a different change at the DNA sequence level. The many different mutations (alleles) at the same locus are called an allelic series for a locus.

D PLEIOTROPY AND POLYGENIC INHERITANCE

There is usually not a one-to-one correspondence between a gene and a physical characteristic. Often a gene is responsible for several phenotypic traits and it is said to be pleiotropic. For example, mutations in the vestigial gene (vg) in Drosophila results in an easily visible short wing phenotype. However, mutations in this gene also affect the number of egg strings, position of the bristles on scutellum, and lifespan in Drosophila. Therefore, vg gene is said to be pleiotropic in that it affects many different phenotypic characteristics.

The opposite is also found. Single characteristics can be affected by mutations in multiple, different genes. This implies that many genes are needed to make each characteristic. For example, if we return to the Drosophila wing, there are dozens of genes that when mutant alter the normal shape of the wing, not just the vg locus. Thus there are many genes that are needed to make a normal wing; the mutation of any one causes an abnormal, mutant, phenotype. This type of arrangement is called polygenic inheritance.

E COMPLETE DOMINANCE AND RECESSIVE

Let us return to an example of a simple phenotype: flower color in Mendel's peas. We have already said that one allele as a homozygote produces purple flowers, while the other allele as a homozygote produces white flowers (Figure 2). But what about an individual that has one purple allele and one white allele; what is the phenotype of an individual whose genotype is heterozygous? This can only be determined by experiment, such as observation. We know from observation that individuals heterozygous for the purple and white alleles of the flower color gene have purple flowers. Thus, the allele associated with purple color is therefore said to be dominant to the allele that produces the white color. The white allele, whose phenotype is masked by the purple allele in a heterozygote, is recessive to the purple allele. The dominant/recessive character is a relationship between proteins formed by each allele and must be determined by observation of the heterozygote phenotype.

Sometimes, to represent this relationship, a dominant allele will be written as a capital letter (e.g. A) while a recessive allele will be written in lower case (e.g. *a*). However, this is not the only system. Many different systems of genetic symbols are in use. The most common are shown in Table 3-1. Also note that genotypes (alleles) are usually written in italics and chromosomes and proteins are not. For example, the white gene in Drosophila melanogaster on the X chromosome encodes a protein called WHITE, which is a pigment precursor transmembrane transporter enzyme.

Table 2-1
Examples of symbols used to represent genes and alleles.

Examples	Interpretation
A and *a*	Uppercase letters represent dominant alleles and lowercase letters indicate recessive alleles. Mendel invented this system but it is not commonly used because not all alleles show complete dominance and many genes have more than two alleles.
a^+ and a^1	Superscripts or subscripts are used to indicate alleles. For wild type alleles the symbol is a superscript +.
AA or *A/A*	Sometimes a forward slash is used to indicate that the two symbols are alleles of the same gene locus, but on homologous chromosomes.

F INCOMPLETE DOMINANCE

Besides the complete dominant and recessive relationship, other relationships can exist between alleles. In incomplete dominance (also called semi-dominance), both alleles affect the trait additively, and the phenotype of the heterozygote shows a typically intermediate between the homozygotes, which is often referred to as blended phenotype. For example, alleles for color in carnation flowers (and many other species) exhibit incomplete dominance. Plants with an allele for red petals (A^1) and an allele for white petals (A^2) have pink petals. We say that the A^1 and the A^2 alleles show incomplete dominance because neither allele is completely dominant over the other (Figure 3-3 on page 17).

G CODOMINANCE

Codominance is another type of allelic relationship in which a heterozygous individual expresses the phenotype of both alleles simultaneously (**Figure 3-3**). An example of codominance is found within the ABO blood

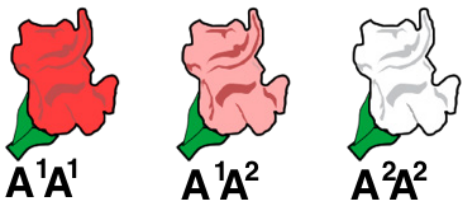


Figure 2-3
Relationship between genotype and phenotype for incompletely dominant alleles affecting petal colour in carnations. (Original-Deyholos- CC BY-NC 3.0)

group of humans. The ABO gene has three common alleles that were named (for historical reasons) I^A, I^B, and *i*. People homozygous for I^A or I^B display only A or B type antigens, respectively, on the surface of their blood cells, and therefore have either type A or type B blood (**Figure 3-4**). Heterozygous I^AI^B individuals have both A and B antigens on their cells, and so have type AB blood. Note that the heterozygote expresses both alleles simultaneously, and is not some kind of novel intermediate between A and B. Codominance is therefore distinct from incomplete dominance, although they are sometimes confused.

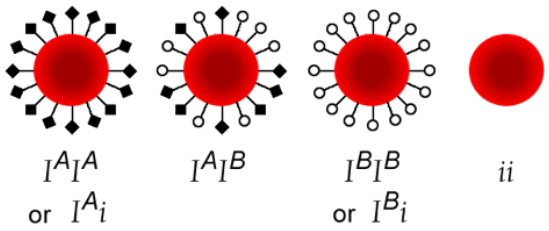


Figure 2-4

Relationship between genotype and phenotype for three alleles of the human ABO gene. The I^A and I^B alleles show co-dominance. The IA allele is completely dominant to the *i* allele. The I^B allele is completely dominant to the i allele.

It is also important to note that the third allele, *i*, does not make either antigen and thus is recessive to the other alleles. I^A/i or I^B/i individuals display only A or B antigens, respectively. People homozygous for the i allele have type O blood.

This is a useful reminder that different types of dominance relationships can exist, even for alleles of the same gene. Many types of molecular markers, which we will discuss in a later chapter, display a co-dominant relationship among alleles. Another example of co-dominance is shown in the first figure of this chapter – flower colour in *Camellia* sp.

H BIOCHEMICAL BASIS OF DOMINANCE

Given that a heterozygote's phenotype cannot simply be predicted from the phenotype of homozygotes, what does the type of dominance tell us about the biochemical nature of the gene product? How does dominance work at the biochemical level? There are several different biochemical mechanisms that may make one allele dominant to another.

For the majority of genes studied, the normal (i.e. wild-type) alleles are haplosufficient. So in diploids, even with a mutation that causes a complete loss of function in one allele, the other allele, a wild-type allele, will provide sufficient normal biochemical activity to yield a wild type phenotype and thus be dominant and dictate the heterozygote phenotype.

On the other hand, in some biochemical pathways, a single wild-type allele is not enough protein and may be haploinsufficient to produce enough biochemical activity to result in a normal phenotype, when heterozygous with a non-functioning mutant allele. In this case, the non-functional mutant allele will be dominant (or semidominant) to a wild-type allele.

Mutant alleles may also encode products that have new and/or different biochemical activities instead of, or in addition to, the normal ones. These novel activities could cause a new phenotype that would be dominantly expressed.

I MUTANT CLASSIFICATION

I.1 MORPHOLOGICAL

Morphological mutations cause changes in the visible form of the organism. An example could be a change in size, shape, colour, number etc.

I.2 LETHAL

A lethal mutation causes the premature death to an organism. For example, in Drosophila lethal mutations can result in the death during the embryonic, larval, or pupal period of development. Lethal mutations are usually recessive, so both copies of a gene have to be lost for the premature death to occur (homozygous lethal alleles will not be viable). Heterozygotes which have one lethal allele and one wild type allele are typically viable..

I.3 BIOCHEMICAL

Biochemical mutations produce auxotrophic mutants from prototrophic parents. This type of mutation blocks a step in a biochemical pathway as discussed for the arg- mutants of Beadle and Tatum in the chapter on biochemical pathways. Biochemical mutations are a specific type of the conditional mutation class (next).

I.4 CONDITIONAL

This type of mutation relies on the concept of: phenotype = genotype + environment + interaction. Organisms with this kind of mutation express mutant phenotype, but only under specific environmental conditions. Under restrictive conditions, they express

the mutant phenotype while under permissive conditions, they show a wild type phenotype. One example of a conditional mutation is the temperature sensitive pigmentation of Siamese cats. Siamese cats have temperature sensitive fur colour; their fur appears white (unpigmented) when grown in a normal, warm temperature environment and appears black (pigmented) when under a cooler temperature, such as the peripheral regions of the feet, snout, and ears (**Figure 3-5**). This is because in warm temperature, the enzyme that is needed for melanin pigment synthesis becomes nonfunctional. However, in cooler temperature, the enzyme needed for melanin synthesis is functional and the deposition of melanin makes the fur look dark.

Figure 2-5

Siamese cats have temperature sensitive pigmentation due to genetic mutation.

(Wikimedia-Telekokopelli-CC BY-SA 3.0)

J MULLER'S MORPHS

Exposure of an organism to a mutagen causes mutations in essentially random positions along the chromosomes. Consequently, most of the mutant phenotypes recovered from a genetic screen are caused by loss-of-function mutations. These alleles are due to random changes in the DNA sequence that cause it to no longer produce the same level of active protein as the wild-type allele (**Figure 3-6**). Loss-of-function alleles tend to be recessive because the wild-type allele is haplosufficient. A loss-of-function allele that produces no active protein is called an amorph, or null. On the other hand, alleles with only a partial loss-of-function are called hypomorphic. More rarely, a mutant allele may have a gain-of-function, producing either more of the active protein (hypermorph) or producing an active protein with a new and different function (neomorph). Finally, antimorph alleles have an activity that is dominant and opposite to the wild-type function; antimorphs are also known as dominant negative mutations.

Thus, mutations (changes in a gene sequence) can result in mutant alleles that no longer produce the same level or type of active product as the wild-type allele (**Figure 3-6**). Any mutant allele can be classified into one of five types: (1) amorph, (2) hypomorph, (3) hypermorph, (4) neomorph, and (5) antimorph.

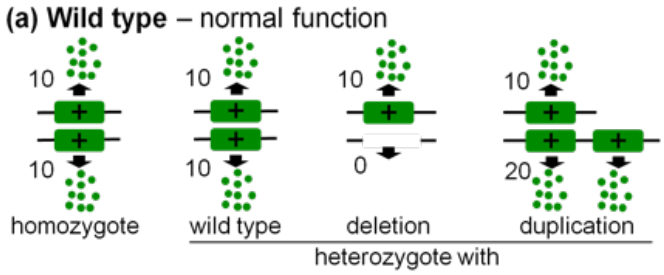


Figure 2-6

"Normal" function of a gene. This is the situation to which Muller's morphs are compared to explain their mechanism of expression.

J.1 AMORPH

Amorphic alleles have a complete loss-of-function. They make no active product – zero function. They are known as a "Null" mutation or a "loss-of-function" mutation (**Figure 3-7**).

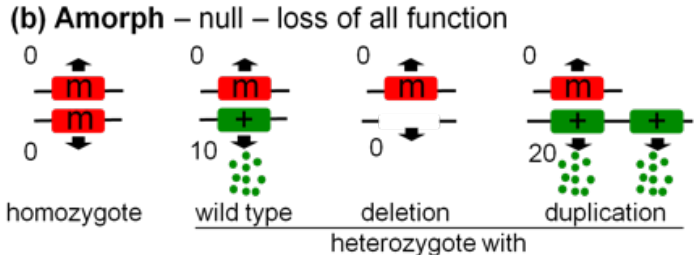


Figure 2-7

A mutant allele that produces no product or afunctional product is considered an amorph.

(Original-Locke- CC BY-NC 3.0)

Molecular explanation - Changes in the DNA base pair sequence of an amorphic allele may cause one or more of the following:

(1) Gene deletion - The DNA sequence is removed from the chromosome.

(2) Gene is present, but is not transcribed due to gene regulation mutation, such as mutation in the promoter or enhancer/regulatory elements.

(3) Gene is present but the transcript is not processed properly. There is normal transcription but base pair changes cause the mature mRNA to incorrectly splice introns, therefore the translated amino acid se-

quence would be altered and nonfunctional.

(4) Gene is present and a transcript is produced but no translation occurs – changes in the base pair sequences would preclude the mRNA from binding to the ribosome for translation.

(5) Gene is present and a transcript is produced and translated but a nonfunctional protein product is produced – the mutation alters a key amino acid in the polypeptide sequence producing a completely non-functional polypeptide.

Genetic/phenotypic explanation - Amorphic mutations of most genes usually act as recessive to wild type (**Figure 3-8**). However, with some genes the amorphic mutations are dominant to wild type (**Table 2-2**). For the *Minute* gene, we concluded that the organism needs both copies to have a wild type phenotype. Loss of one copy (an amorphic mutation) produces a dominant visible mutant phenotype. Deletion of the gene is an example of classical amorphic mutation.

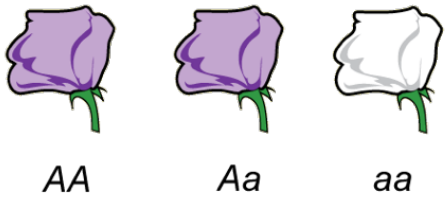


Figure 2-8

An amorph lacking the gene to produce pigment results in an allele that is expressed as a recessive pattern.

Table 2-2

Phenotypes that arise from cases #1 and #2 in the text.

case #1: *white* gene in Drosophila	
w^+/w^+	wildtype (red eyed phenotype)
w^+/w^-	wildtype (red eyed phenotype)
w^-/w^-	mutant (white eyed phenotype)
case #2: *Minute* locus in Drosophila	
M^+/M^+	wildtype (long bristled phenotype)
M^+/M^-	Mutant (short bristled phenotype)
M^-/M^-	Phenotype = dead; recessive lethal

J.2 HYPOMORPH

Hypomorphic alleles are only a partial loss-of-function. These alleles are sometimes referred to as "leaky" mutations, because they provide some function, but not complete, normal function (**Figure 3-9**).

Molecular explanation - Changes in the DNA base pair sequence of the hypomorphic allele may cause one or more of the following, with gene still being present:

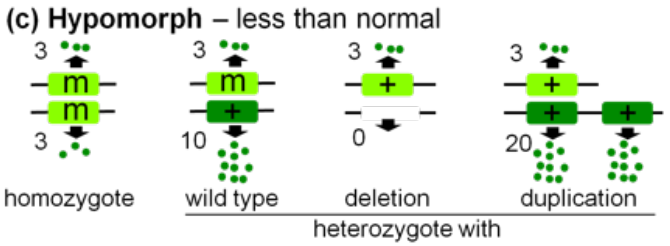


Figure 2-9

A mutant allele that produces fewer proteins or proteins that have lower effectivity than wild-type is considered a hypomorph.

(Original-Locke- CC BY-NC 3.0)

(1) reduced transcription – changed DNA sequence in the promoter or enhancer/regulatory elements can reduce the level of transcription.

(2) aberrant processing of the transcript – normal transcription but base pair changes cause the mature mRNA to incorrectly splice introns, therefore the translated protein sequence would be altered and function at a reduced level.

(3) reduced translation – changes in the base pair sequences would reduce the efficiency of the mRNA binding to the ribosome for translation.

(4) reduced-function protein product – normal transcription, processing, and translation but mutation changes certain amino acid in the polypeptide sequence so its function is reduced.

Genetic/phenotypic explanation - Hypomorphic mutations of most genes usually act as recessive to wild type, though hypomorphic mutations theoretically could be dominant to wildtype.

Both amorphs and hypomorphs tend to be recessive to wild type in diploids because the wild type allele is usually able to supply sufficient product to produce a wild type phenotype (called haplo-sufficient). If the mutant allele is not able to produce a wild type phenotype, then it is haplo-insufficient, and it will be dominant to the wild type allele. Here -/+ heterozygotes produce a mutant phenotype.

While the first two classes involve a loss-of-function, the next two involve a gain-of-function – quantity or quality. Gain-of-function alleles are almost always dominant to the wild type allele.

J.3 HYPERMORPH

Hypermorphic alleles produce quantitatively more of the same, active product (**Figure 3-10**).

Molecular explanation - Changes in the DNA base pair sequence of the hypermorphic allele may cause one or more of the following, with gene still being present:

(1) increased transcription – changed DNA sequence in the promoter or enhancer/regulatory elements that increase the level of transcription.

(2) increased translation – changes in the base pair sequences would increase the efficiency of the mRNA binding to the ribosome for translation.

(3) increased function protein product – normal transcription, processing, translation but base pair changes alter certain amino acid in the polypeptide sequence so its function is normal but increased in amount.

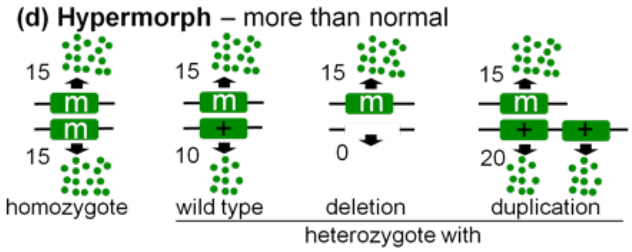


Figure 2-10

A mutant allele that produces more proteins or proteins that have increased effectivity than wild-type is considered a hypermorph.

(Original-Locke- CC BY-NC 3.0)

Genetic/phenotypic explanation - Hypermorphic mutations of most genes usually act as dominant to wild type since they are a gain of function, The classic hypermorph is a gene duplication.

J.4 NEOMORPH

Neomorphic alleles produce a product with a new, different function, something that the wild type allele doesn't do (**Figure 3-11**).

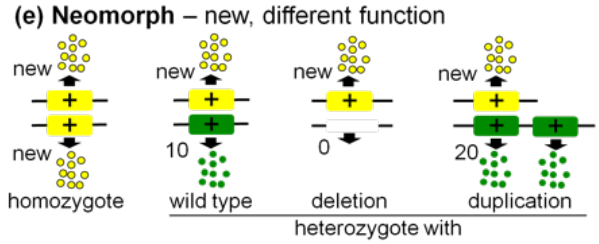


Figure 2-11

Neomorphic alleles encode proteins that have a function very different from those from the wild-type allele.

(Original-Locke- CC BY-NC 3.0)

Molecular explanation - Changes in the nucleotide sequence of the neomorphic allele may cause one or more of the following (gene is still present):

(1) transcription – new DNA sequence in the regulatory elements alters transcription either temporally or in a tissue specific manner.

(2) protein product with new function – normal transcription, processing, translation, but encode a different amino acid(s) so the protein acquires a new function (e.g. additional substrate or new binding site).

Genetic/phenotypic explanation - Neomorphic mutations of most genes usually act as a dominant to wild type since they are a gain-of-function. The classical neomorphic mutation is usually a translocation that moves a new regulatory element next to a gene promoter so it is expressed in a new tissue or at a new time during development. Such mutations are often produced when chromosome breaks are rejoined and the regulatory sequences of one gene are put adjacent to the transcriptional unit of another, creating a novel, chimeric gene.

J.5 ANTIMORPH

Antimorphic alleles are relatively rare, and have a new activity that is dominant and opposes the wild-type function. These alleles usually have lost their normal function and they interfere with the function from the wild type allele (**Figure 3-12**). This can happen at the transcriptional, translational, or later level of expression. Thus, when an antimorph allele is heterozygous with wild type, the wild type allele function is reduced or prevented. While at the molecular level there are many ways this can happen, the simplest model to explain an antimorph effect is that the product acts as a dimer (or any multimer) and the inclusion of one mutant subunit poisons the whole complex, thereby preventing it from functioning. Antimorphs are also known as dominant negative mutations because they are usually dominant and act negatively against the wild type function.

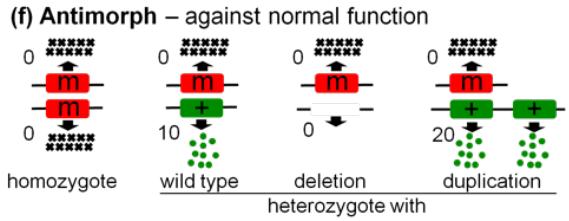


Figure 2-12

In an antimorph, the protein from the mutant allele binds the wild-type protein, inactivating it.

(Original-Locke- CC BY-NC 3.0)

J.6 IDENTIFYING MULLER'S MORPHS

All mutations can be sorted into one of the five morphs base on how they behave when heterozygous with three other standard alleles (**Figure 7**): (1) null alleles (zero function), (2) wild type alleles (normal function), and (3) duplication alleles (double normal function).

SUMMARY:

- Symbols are used to denote the alleles, or genotype, of a locus. See Appendix 1 for rules about symbols.
- Phenotype depends on the alleles that are present, their dominance relationships, and sometimes also interactions with the environment and other factors.
- In a diploid organism, the alleles can be homozygous, heterozygous or hemizygous.
- Somatic mutations affect the individual but not the progeny, whereas germline mutation affect the progeny in the next generation but not the individual in which they occur.
- Allelic interactions at a locus can be described as dominant vs. recessive, incomplete dominance, or co-dominance.
- Muller's morphs classify all types of mutations including: amorph, hypomorph, hypermorph, neomorph, and antimorph.

KEY TERMS:

silent mutation
homozygous
heterozygous
hemizygous
wild-type
variant
locus
genotype
phenotype
dominant
recessive
complete dominance
incomplete (semi) dominance
co-dominance
ABO blood group
haplosufficiency
haploinsufficiency
loss-of-function
gain-of-function
amorph
null
hypomorph
hypermorph
neomorph

STUDY QUESTIONS:

1. If your blood type is B, what are the possible genotypes of your parents at the locus that controls the ABO blood types?
2. In **Table 3-3**, match the mouse hair color phenotypes with the term from the list that best explains the observed phenotype, given the genotypes shown. In this case, the allele symbols do not imply anything about the dominance relationships between the alleles. List of terms: haplosufficiency, haploinsufficiency, pleiotropy, incomplete dominance, co-dominance, incomplete penetrance, broad (variable) expressivity.

Table 2-3
Data for Question 2.

	A_1A_1	A_1A_2	A_2A_2
1	all hairs black	on the same individual: 50% of hairs are all black and 50% of hairs are all white	all hairs white
2	all hairs black	all hairs are the same shade of grey	all hairs white
3	all hairs black	all hairs black	50% of individuals have all white hairs and 50% of individuals have all black hairs
4	all hairs black	all hairs black	mice have no hair
5	all hairs black	all hairs white	all hairs white
6	all hairs black	all hairs black	all hairs white
7	all hairs black	all hairs black	hairs are a wide range of shades of grey

Chapter 3

Part 1: Mendel's First Law Segregation of Alleles

Figure 3-1
Pea plants were used by Gregor Mendel to discover fundamental laws of genetics.

His first law, the segregation of alleles, is covered in this chapter. His second law, independent assortment, is covered in the next chapter. (Wikimedia commons-B. Ebbesen-CC BY-SA 3.0)

Introduction

The once prevalent concept of blended inheritance proposed that some undefined essence, in its entirety, contained all of the heritable information for an individual. It was thought that mating combined the essences from each parent, much like the mixing of two colors of paint. Once blended together, the individual characteristics of the parents could not be separated again. It was just "common sense".

However, Gregor Mendel (**Figure 3-2 on page 24**) was one of the first to take a quantitative, scientific approach to the study of heredity. He started with well-characterized strains, repeated his experiments many times, and kept careful records of his observations. Working with peas, Mendel showed that white-flowered plants could be produced by crossing two purple-flowered plants, but only if the purple-flowered plants themselves had at least one white-flowered parent (**Figure 3-3 on page 24**). This was evidence that a discrete genetic factor that produced white-flowers had not blended irreversibly with the factor for purple-flowers. Mendel's observations disprove blending inheritance and favor an alternative concept, called particulate inheritance, in which heredity is the product of discrete factors that control independent traits.

Because of this work, blended inheritance was discredited as a theory of inheritance.

Through careful study of patterns of inheritance, Mendel recognized that a single trait could exist in different versions, or alleles, even within an individual plant or animal. For example, he found two allelic forms of a gene for seed color: one allele gave green seeds, and the other gave yellow seeds. Mendel also observed that although different alleles could influence a single trait, they remained indivisible and could be inherited separately. This is the basis of Mendel's First Law, also called The Law of Equal Segregation, which states: during gamete formation, the two alleles at a gene locus segregate from each other; each gamete has an equal probability of containing either allele.

A Overview

Mendel first made his discoveries of inheritance in the 1850's. In his 1866 publication he didn't use the word "gene" as the fundamental unit of heredity because it wasn't coined until 1909 by Danish botanist Wilhelm Johannsen. Thomas Hunt Morgan proposed that genes resided on chromosomes in 1910, and occupied distinct regions on those chromosomes. DNA as a substance was discovered in the 1860's, but it took until the 1940s to realize that DNA was the molecule that contained the genetic information. Then in the 1950's Watson and Crick discovered the structure of DNA.

Figure 3-2

Gregor Johann Mendel (1822-1884), an Augustinian Friar, who lived in Moravia (now part of the Czech Republic), published his work in 1866 on what has become known as the laws of Mendelian Inheritance. (Wikipedia-Hugo Iltis- CC BY 4.0)

Nevertheless, Mendel made his discoveries without any of this information. Today we have overwhelming knowledge from research allowing us to understand the molecular mechanism behind Mendel's laws. To explain Mendel's First Law, segregation, we will explain it through the concept of meiosis.

B Dominant and Recessive Alleles

The concepts of dominant and recessive alleles were introduced in Chapter 1. Remember, alleles are different versions of a gene. The traits Mendel studied with his peas were all completely dominant traits, and so will only be described briefly here.

In a diploid organism, if an allele is dominant only one copy of that allele is necessary to see the dominant phenotype. If an allele is recessive, then the gene needs to have two copies (or be homozygous) to see the recessive phenotype. If an organism is a heterozygote, or has one copy of each allele type, then it will show the dominant phenotype. When representing these in written form, a dominant allele is written as a capital letter (e.g. *A*), while a recessive allele will be written in lower case (e.g. *a*). If these are alleles of the same gene, they should be written with the same letter. This is the most common way of writing genotypes (See Chapter 1), but there are many different systems that often have deviations from these general rules. Note that genes and alleles are usually written in italics and chromosomes and proteins are not, proteins often written in all capitals. For example, the white gene in Drosophila melanogaster on the X chromosome encodes a protein called WHITE.

C One Locus on a Chromosome - Segregation - Monohybrid

Not only did Mendel solve the mystery of inheritance as units (genes), he also invented several testing and analysis techniques still used today. **Classical genetics** is the science of solving biological questions using controlled matings of model organisms. It began with Mendel in 1865 but did not attain wide spread usage until Mendel's work was rediscovered in 1903 by four researchers (E. von Tschermak, H. de Vries, C. Correns, and W. J. Spillman). Then Thomas Morgan began working with fruit flies in 1908 and used this work. Later, starting with Watson and Crick's structure of DNA in 1953, classical genetics was joined by **molecular ge-**

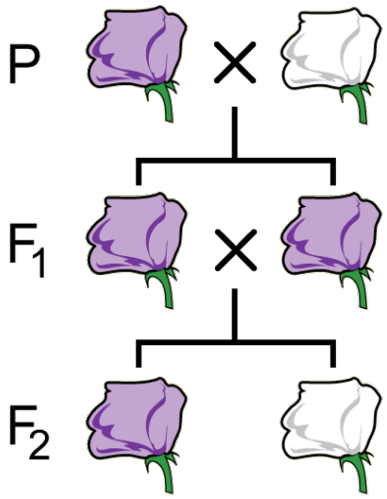


Figure 3-3

Inheritance of flower color in peas. Mendel observed that a cross between pure breeding, white and purple peas (generation P) produced only progeny (generation F1) with purple flowers. However, white flowered plant reappeared among the F2 generation progeny of a mating between two F1 plants. The symbols P, F1 and F2 are abbreviations for parental, first filial, and second filial generations, respectively. (Original-Deyholos- CC BY-NC 3.0)

netics, the science of solving biological questions using DNA, RNA, and proteins isolated from organisms. The genetics of **DNA cloning** began in 1970 with the discovery of restriction enzymes and plasmids as cloning vectors.

Knowing what we now know of the process of meiosis, we can further understand what Mendel's First Law means. The Law of Segregation states that every individual contains a pair of alleles for each gene, which segregate during the creation of gametes, and so for every gene pair each parent passes on a random allele to its offspring. The way Mendel decided on his First Law was through the process of **Monohybrid crosses**, which will be described below.

C.1 Terminology

A specific position along a chromosome is called a **locus**. Each gene occupies a specific locus (so the terms locus and gene are often used interchangeably). Each locus will have an allelic form (allele). The complete set of alleles (at all loci of interest) in an individual is its **genotype**. Typically, when writing out a genotype, only the alleles at the locus (loci) of interest are considered – all the others are present and assumed to be wild type but are normally not written in the genotype. The visible or detectable effect of these alleles on the structure or function of that individual is called its **phenotype** – what the individual looks like. The phenotype studied in any particular genetic experiment may range from simple, visible traits such as hair color, to more complex phenotypes including disease susceptibility or behavior. If two alleles are present in an individual, then various interactions between them may influence their expression in the phenotype.

C.2 True breeding lines

Geneticists make use of true breeding lines just as Mendel did (**Figure 3-4a**). These are in-bred populations of plants or animals in which all parents and their offspring (over many generations) have the same phenotypes with respect to a particular trait. True breeding lines are useful, because they are typically assumed to be homozygous for the alleles that affect the trait of interest. When two individuals that are homozygous for the same alleles are crossed, all offspring will all also be homozygous. The continuation of such crosses constitutes a true breeding line or strain. A large variety of different strains, each with a different, true breeding character, can be collected and maintained for genetic research.

C.3 Monohybrid crosses

A monohybrid cross is one in which both parents are heterozygous (or a hybrid) for a single (mono) trait. The trait might be petal colour in pea plants (**Figure 3-4b**). Recall from **Figure 3-2 on page 24** that the generations in a cross are named P (parental), F_1 (first filial), F_2 (second filial), and so on.

The monohybrid cross is how Mendel discovered genes were discrete units that separated in the creation of offspring. Previous ideas of blending inheritance would mean that a cross between a white flower and a purple flower would create a 'blended' phenotype. Instead what Mendel saw was distinct parental colours in the hybrids, that when crossed would produce in specific ratios the purple and white seen in the parents. These traits weren't blended when the true-breeding lines were crossed, but instead those parental alleles were carried on through the offspring. Through the monohybrid cross he was able to discern the **dominant** and **recessive** alleles of each gene he studied in the pea plants. In further crosses (F_3, F_4, etc.), these traits were continuously transmitted and not lost, though they may be hidden as seen in the F_1 generation.

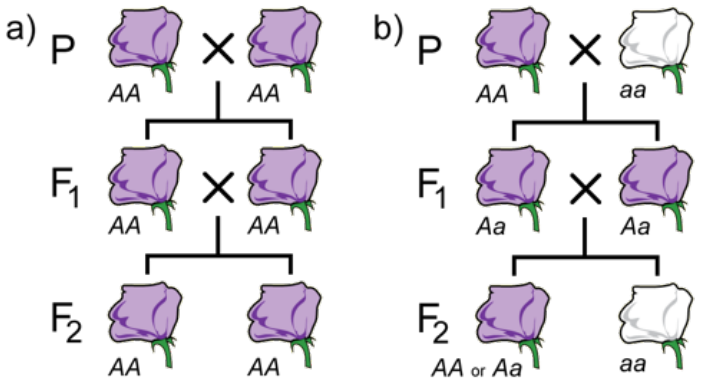


Figure 3-4
(**a**) A true-breeding line (**b**) A monohybrid cross produced by mating two different pure-breeding lines.
(Original-Deyholos-CC BY-NC 3.0)

D Punnett Squares - 3:1 Ratio

The specific ratios seen in the monohybrid cross can be described using a **Punnett square**, named after R.C. Punnett who devised this approach.

Given the genotypes of any two parents, we can predict all of the possible genotypes of the offspring. Furthermore, if we also know the dominance relationships for all of the alleles, we can predict the phenotypes of

the offspring. This provides a convenient method for calculating the expected genotypic and phenotypic ratios from a cross.

A Punnett square is a matrix in which all of the possible gametes produced by one parent are listed along one axis, and the gametes from the other parent are listed along the other axis. Each possible combination of gametes is listed at the intersection of each row and column, since we know through the process of meiosis that the alleles on each chromosome separate to form the gametes.

The F_1 cross from **Figure 3-4 on page 25** would be drawn as in **Figure 3-5**. As you can see, in a Monohybrid cross, the offspring ratios will be 3:1 of dominant phenotype (purple): recessive phenotype (white). Punnett squares can also be used to calculate the frequency of offspring. The frequency of each offspring is the frequency of the male gametes multiplied by the frequency of the female gamete.

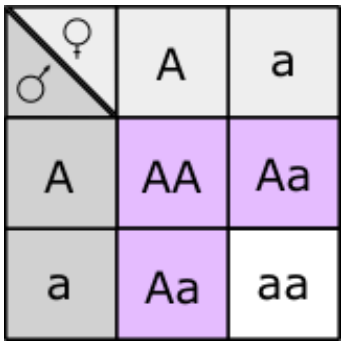


Figure 3-5
A Punnett square showing a monohybrid cross. The purple boxes represent the purple colour of the dominant (A) allele, while the white box represents the recessive (aa) allele homozygote.

(Original-L. Canham- CC BY-NC 3.0)

Go online to practice working with monohybrid ratios.

http://tinyurl.com/oog-monohybrid

E SINGLE LOCUS TEST CROSSES

Knowing the genotypes of an individual is an important part of a genetic experiment. However, genotypes cannot be observed directly; they must be inferred based on phenotypes. Because of dominance, it is often not possible to distinguish between a heterozygote and a homozygote based on phenotype alone (*e.g.* see the purple-flowered F_2 plants in **Figure 3-4 on page 25**). To determine the genotype of a specific individual, a **test cross** can be performed, in which the individual with an uncertain genotype is crossed with an individual that is homozygous recessive for all of the loci being tested.

For example, if you were given a pea plant with purple flowers it might be a homozygote (*A*/*A*) or a heterozygote (*A*/*a*). You could cross this purple-flowered plant to a white-flowered plant as a **tester**, since you know the genotype of the tester is *a*/*a*. Depending on the genotype of the purple-flowered parent (**Figure 3-6**), you will observe one of two phenotypic ratios in the F_1 generation. If the purple-flowered parent was a homozygote *AA*, all of the F_1 progeny will be purple. If the purple-flowered parent was a heterozygote *A*/*a*, the F_1 progeny should segregate purple-flowered and white-flowered plants in a 1:1 ratio.

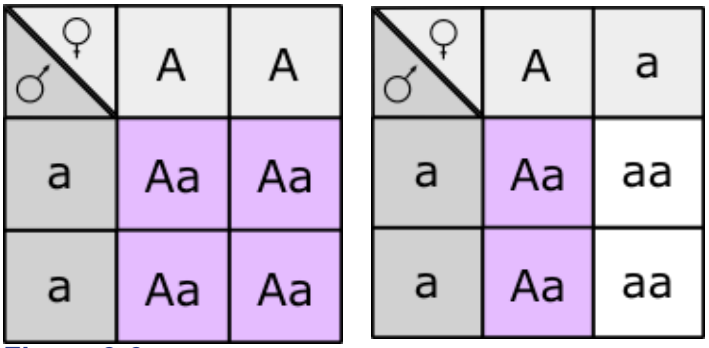


Figure 3-6
Punnett squares showing the two possible outcomes of a single locus test cross. (Original-L. Canham- CC BY-NC 3.0)

SUMMARY:

- Mendel demonstrated that heredity involved discrete, heritable factors that affected specific traits.
- A gene can be defined abstractly as a unit of inheritance.
- Phenotype depends on the alleles that are present, their dominance relationships, and sometimes also interactions with the environment and other factors.
- Classical geneticists make use of true breeding lines, monohybrid crosses, Punnett squares, test crosses, reciprocal crosses, and the chi-square test.

Part 2: Mendel's Second Law Independent Assortment

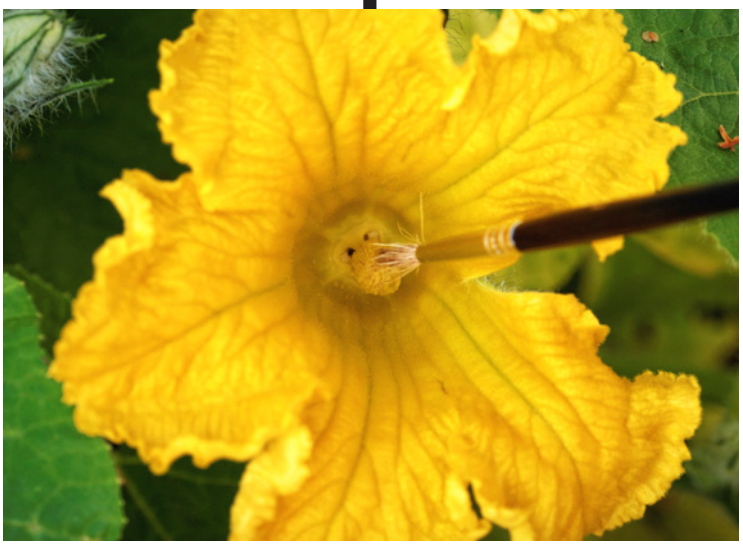


Figure 3-7
Hand pollination of a pumpkin flower. When Mendel was doing his crosses with pea plants, he pollinated each flower by hand in a similar way in order to be sure he knew the parents of each cross. (Flickr-S. Hurmerinta- CC BY-NC 2.0

Introduction

The principles of genetic analysis that we have described for a single locus in Part 1 will be extended to the study of alleles at two loci in Part 2. The analysis of two loci in the same cross provides information for testing gene interactions (Chapter 9) and genetic mapping (Chapter 10). These techniques are very useful for both basic and applied research. Before discussing these techniques, we will first revisit Mendel's classical experiments.

Before Mendel's 1865 publication, **blended inheritance** was the accepted model to explain the transmission of traits. It was Mendel's work that established that **heritable traits** were controlled by discrete factors, which we now call alleles, in a **particulate inheritance** model. At the time it was an important question as to whether heritable traits, controlled by discrete factors, were inherited independently of each other? To answer this, Mendel took two apparently unrelated traits, such as seed shape and seed color, and studied their inheritance together in one individual. For example, he studied two variants of each trait: seed color was either green or yellow, and seed shape was either round or wrinkled. (He studied seven traits in all, each on a different chromosome.) When either of these traits was studied individually, the phenotypes segregated in the classical 3:1 ratio among the progeny of a monohybrid cross (**Figure 3-8**), with ¾ of the seeds green and ¼ yellow in one cross, and ¾ round and ¼ wrinkled in the other cross. Would this be true when both hybrids were in the same individual?

Like in the previous chapter, we will first walk through how a dihybrid cross works on at the DNA level, and then we will explain the results that Mendel saw that led him to his law, the Law of Independent Assortment.

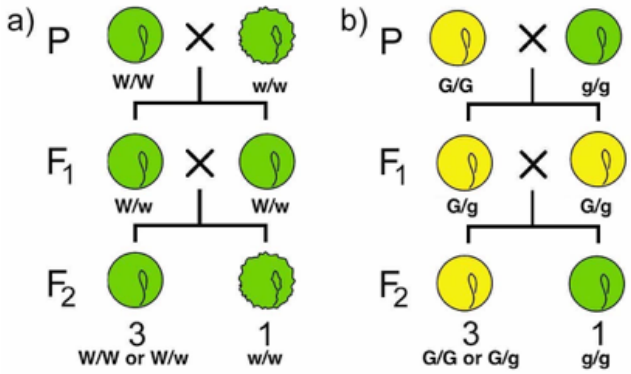


Figure 3-8
Monohybrid crosses involving two distinct traits in peas. a) is *W/w* and b) is *G/g*. Note that seeds in a) also contain a "g" gene that would be homozygous recessive, but this isn't written out. Seeds in b) are have a "*w*" gene, but they would be heterozygous or homozygous dominant. By convention, unless a gene is specifically referenced in the problem, we don't write it down – but it's still there! Monohybrid crosses are detailed in Part 1. (Original-Deyholos-CC BY-NC 3.0-Modified-Nickle)

A Two Loci on Different Chromosomes

The separation of gametes through the process of meiosis has already been introduced. But what does that mean when you are taking multiple different genes (or loci) into account?

Remember the main stages of meiosis. The homologous pairs align during metaphase I, and complete one round of cell division. Then during metaphase II in those two cells the replicated chromosomes align individually and the sister chromatid separate, so when complete you have two daughter cells. Let's say one chromosome has Gene *A* on it, and another chromosome has Gene *B* on it, and the individual is heterozygous at each gene (a.k.a. has the genotype *A/a ; B/b*). There are a variety of ways that the homologous pairs can align themselves during metaphase I. The orientation of that alignment will affect the alleles each gamete receives at the end of telophase II (**Figure 3-9**).

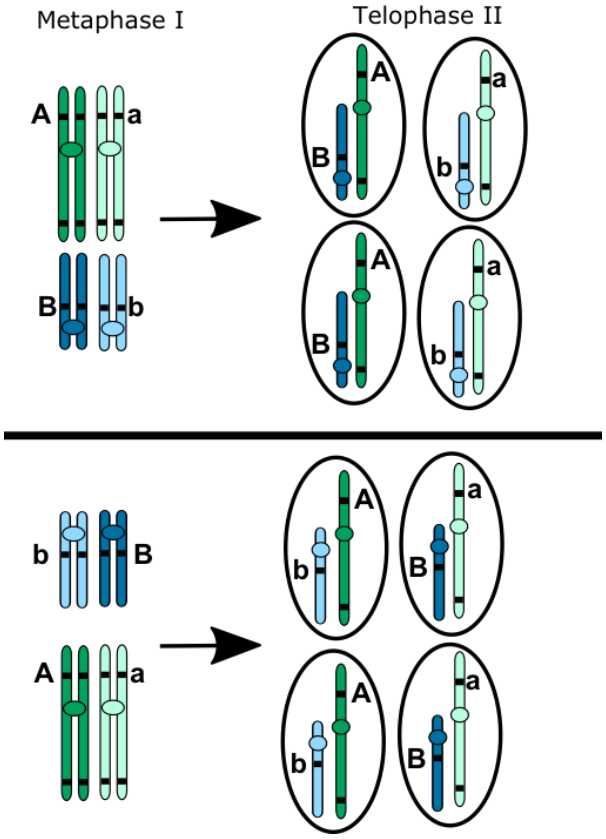


Figure 3-9

Independent assortment as seen on two different chromosomes. Gene A is found on the short chromosome and Gene B is found on the long chromosome, and both genes are heterozygotes for the dominant (*A* and *B*) and recessive (*a* and *b*) alleles. The orientation that the chromosomes align themselves during metaphase I affect the alleles found in the 4 gametes produced after telophase II.

These are just two of many orientations the chromosomes can arrange themselves in at metaphase I. The full stages of meiosis were removed for simplicity. See Chapter 7 for details. (Original-L.Canham-CC BY-NC 3.0)

Because the alignment at metaphase I is always random you will see a random, equal distribution of alleles in all the gametes produced. This means that one allele doesn't affect the distribution of another allele, or in other words, each allele assorts independently (**Independent Assortment**).

B Two Loci on One Chromosome

Based on the description in the last section, it would be expected that if the genes were on the same chromosome the alleles would travel together through meiosis (**Figure 3-10 top**). However, when tested this is not always the case. The recombination of alleles can be explained through the phenomenon of crossing over, which occurs during prophase I. Details are described in Chapter 7. Crossing over is an exchange between non-sister chromatids that can occur at any position along the entire chromosome. If the two loci being observed are sufficiently far enough apart on the chromosome, crossover events can occur between the two loci.

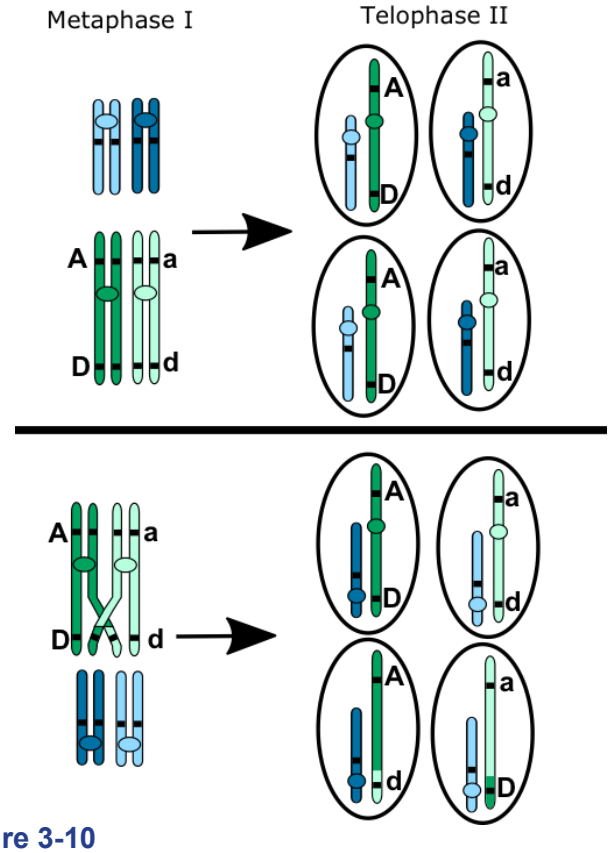


Figure 3-10

Independent assortment as seen on the same chromosome. **Top**: no crossover. The dominant alleles of Gene A and Gene D travel together, not leading to independent assortment. **Bottom**: if a crossover occurs between the two genes, then the alleles will transfer to the other non-sister chromatid, shuffling the alleles. This allows for independent assortment, despite being on the same chromosome.

(Original-L.Canham-CC BY-NC 3.0)

This, coupled with the random orientation that the chromosomes align during metaphase I, will allow the other combination of alleles in the gametes (**Figure 3-10 bottom**).

While not shown in **Figure 3-10**, if the two loci are very far apart, multiple crossover events can also take place, further increasing the shuffling of alleles.

The farther apart on the chromosome the more crossover events take place between the two loci. Ultimately, this will result in similar allele combinations to that observed in independent assortment shown above, even if they are on the same chromosome.

If the loci are very close together on the same chromosome, fewer or no crossovers will occur between them. We will not take this situation into account in this chapter, but will be discussed later in **Chapter 9**.

C A Dihybrid Cross showing Mendel's Second Law (Independent Assortment)

Mendel found for each locus that there were two alleles, which segregated from each other during the creation of gametes. He wondered if dealing with multiple traits at a time would affect this segregation, so he created a **dihybrid cross**. The distribution of offspring from his experiments led him to formulate **Mendel's Second Law**, the **Law of Independent Assortment**, which states that the segregation of alleles as one locus will not influence the segregation of alleles at another locus during gamete formation – the alleles segregate independently. Next we will walk through how he came to this understanding, now that it's already been explained how independent assortment happens.

C.1 Mendel's second law

To analyze the segregation of both traits at the same time in the same individual, he crossed a pure breeding line of green, wrinkled peas with a pure breeding line of yellow, round peas to produce F_1 progeny that were all green and round. They were called **dihybrids** in that they carried two alleles at each of two loci (**Figure 3-11**)

If the inheritance of seed color was truly independent of seed shape, then when the F_1 dihybrids were crossed to each other, a 3:1 ratio of one trait should be observed within each phenotypic class of the other

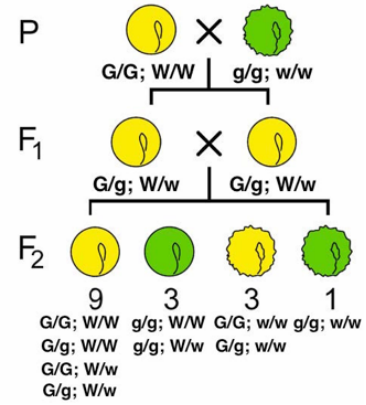


Figure 3-11

Two pure-breeding lines are crossed to produce dihybrids in the F_1 generation. These F_1 are crossed to produce four phenotypic classes, which appear in a 9:3:3:1 ratio. (Original-Deyholos-CC BY-NC 3.0 modified by Nickle)

trait (**Figure 3-11**). From **Figure 3-8 on page 27** we know that yellow and round are dominant, and green and wrinkled are recessive. Using the product law, we would therefore predict that if ¾ of the progeny were yellow, and ¾ of the progeny were round, then ¾ × ¾ = 9/16 of the progeny would be both round and yellow (**Table 3-1**).

Likewise, ¾ × ¼ = 3/16 of the progeny would be both round and green. And ¾ × ¼ = 3/16 of the progeny would be both wrinkled and yellow. And ¼ × ¼ = 1/16 of the progeny would be both wrinkled and green. So by applying the product rule to all of these combinations of phenotypes, we can predict that if the two loci assort independently in a **9:3:3:1** phenotypic ratio among the progeny of this dihybrid cross, if certain conditions are met (see section below). Indeed, 9:3:3:1 is very close to the ratio Mendel observed in his studies of dihybrid crosses, leading him to state his Second Law, the **Law of Independent Assortment**.

The 9:3:3:1 phenotypic ratio that we calculated using the product rule could also be obtained using Punnett Square (**Figure 3-12 on page 30**). First, we list the

Table 3-1

Phenotypic classes expected in monohybrid and dihybrid crosses for two seed traits in pea.

Frequency of phenotypic crosses within separate monohybrid crosses:				
seed shape:		¾ round		¼ wrinkled
seed color:		¾ yellow		¼ green
Frequency of phenotypic crosses within a dihybrid cross:				
¾ round	×	¾ yellow	=	9/16 round & yellow
¾ round	×	¼ green	=	3/16 round & green
¼ wrinkled	×	¾ yellow	=	3/16 wrinkled & yellow
¼ wrinkled	×	¼ green	=	1/16 wrinkled & green

	W ; G	W ; g	w ; G	w ; g
W ; G	W/W; G/G	W/W; G/g	W/w; G/G	W/w; G/g
W ; g	W/W; G/g	W/W; g/g	W/w; G/g	W/w; g/g
w ; G	W/w; G/G	W/w; G/g	w/w; G/G	w/w; G/g
w ; g	W/w; G/g	W/w; g/g	w/w; G/g	w/w; g/g

Figure 3-12

A Punnett Square showing the results of the dihybrid cross from **Figure 3-11 on page 29**. Each of the four phenotypic classes is represented by a different color of shading. (Original-Nickle-CC BY-NC 3.0)

genotypes of the possible gametes along each axis of the Punnett Square. In a diploid with two heterozygous genes of interest, there are up to four combinations of alleles in the gametes of each parent. The gametes from the respective rows and column are then combined in the each cell of the array. When working with two unlinked loci, genotypes are written with the symbols for both alleles of one locus, followed by both alleles of the next locus (e.g. *AaBb*, not *ABab*). Note that the order in which the loci are written does not imply anything about the actual position of the loci on the chromosomes.

This is an encore presentation of the online exercise for writing out linked and unlinked genes.

http://tinyurl.com/oog-linked

To calculate the expected phenotypic ratios, we assign a phenotype to each of the 16 genotypes in the Punnett Square, based on our knowledge of the alleles and their dominance relationships.

In the case of Mendel's seeds, any genotype with at least one *W* allele and one *G* allele will be round and yellow; these genotypes are shown in the nine, green-shaded cells in **Figure 3-12**.

We can represent all of four of the different genotypes shown in these cells with the notation (*W/-;G/-*), where the dash line (-), means "any allele". The semicolon means that the genes assort independently (Perhaps they are on different chromosomes). The three offspring that have at least one R allele and are homozygous recessive for *y* (*i.e. R/-;y/y*) will have a round, green phenotype. Conversely the three progeny that are homozygous recessive *r*, but have at least one *Y* allele (*r/r;Y/-*) will have wrinkled, yellow seeds. Finally, the rarest phenotypic class of wrinkled, yellow seeds is produced by the doubly homozygous recessive genotype, *r/r;y/y*, which is expected to occur in only one of the sixteen possible offspring represented in the square.

C.2 Assumptions of the 9:3:3:1 Ratio

Both the product rule and the Punnett Square approaches showed that a 9:3:3:1 phenotypic ratio is expected among the progeny of a dihybrid cross such as Mendel's *R/r;Y/y* × *R/r;Y/y*. In making these calculations, we assumed that:

(1) alleles at each locus segregate independently of the alleles at the other;

(2) one allele at each locus is completely dominant (the other recessive); and

(3) each of four possible phenotypes can be distinguished unambiguously, with no interactions between the two genes that would interfere with determining the genotype correctly.

For simplicity, most student examples involve easily scored phenotypes, such as pigmentation or changes in visible structures. However, keep in mind that the analysis of segregation ratios of any two markers can provide insight into a wide range of biological processes they may represent.

C.3 Deviations from the 9:3:3:1 Phenotypic Ratio

There can be deviations from the 9:3:3:1 phenotypic ratio. These situations may indicate that one or more of the above conditions has not been met. Modified ratios in the progeny of a dihybrid cross can therefore reveal useful information about the genes involved. One such example is linkage.

Linkage is one of the most important reasons for distortion of the ratios expected from independent assortment. Two loci show linkage if they are located close together on the same chromosome. This close proximity alters the frequency of allele combinations in the gametes. We will return to the concept of linkage in Chapter 9. Deviations from 9:3:3:1 ratios can also be due to interactions between genes, such as epistasis, duplicate gene action and complementary gene action. These interactions are discussed in Chapter 9.

D The Dihybrid Test Cross

While the cross of an F_1 x F_1 gives a ratio of 9:3:3:1, there is a better, easier cross to test for independent assortment: the dihybrid test cross. In a dihybrid test cross, independent assortment is seen as a ratio of 1:1:1:1, which is easier to score than the 9:3:3:1 ratio. This test cross will also be easier to use when testing for linkage (Chapters 9 and 10).

Like in monohybrid crosses (Part1), you can do test crosses with dihybrids to determine the genotype of an individual with dominant phenotypes, to see if they are heterozygous or homozygous dominant. This type of cross is set up in the same fashion: an individual with an unknown genotype in two loci is crossed to an individual that is homozygous recessive for both loci.

Punnett squares should be done ahead of the crosses, so you know what to expect for any of the possible outcomes. Using the example from the rest of this chapter, you cross a double homozygous recessive pea plant (r/r ; y/y. green and wrinkled) to an unknown individual that has two dominant phenotypes (R/- ; Y/-. yellow and round). There are four possible genotypes the unknown individual could be: R/R ; Y/Y or R/R ; Y/y or R/r ; Y/Y or R/r; Y/y. The Punnett squares for the first two are listed below (**Figure 3-13**). Notice on the left you only get the dominant phenotype for both, so you know both genes in the unknown are homozygous recessive. On the right you get only the dominant phenotype for round peas, but you get 50% yellow and 50% green peas, showing that the unknown is homozygous for round, but heterozygous for colour of the peas. **Figure 3-14** is blank for you to fill in the two other gamete and genotype possibilities.

♂ \ ♀	R;Y	R;Y	R;Y	R;Y
r;y	R/r Y/y	R/r Y/y	R/r Y/y	R/r Y/y
r;y	R/r Y/y	R/r Y/y	R/r Y/y	R/r Y/y
r;y	R/r Y/y	R/r Y/y	R/r Y/y	R/r Y/y
r;y	R/r Y/y	R/r Y/y	R/r Y/y	R/r Y/y

♂ \ ♀	R;Y	R;y	R;Y	R;y
r;y	R/r Y/y	R/r y/y	R/r Y/y	R/r y/y
r;y	R/r Y/y	R/r y/y	R/r Y/y	R/r y/y
r;y	R/r Y/y	R/r y/y	R/r Y/y	R/r y/y
r;y	R/r Y/y	R/r y/y	R/r Y/y	R/r y/y

Figure 3-13

Punnett square for two test crosses. The tester in both cases is the male with the genotype *r/r ; y/y* . On the left, the unknown has a genotype of *R/R; Y/Y*. On the right, the unknown has the genotype *R/R ; Y/y*. (Original-L.Canham-CC BY-NC 3.0)

♂ \ ♀				
r;y				
r;y				
r;y				
r;y				

♂ \ ♀				
r;y				
r;y				
r;y				
r;y				

Figure 3-14

Blank Punnett squares to fill in the other two possibilities of the test cross.

Go online to practice working with dihybrid ratios.

http://tinyurl.com/oog-dihybrid

Summary:

- The alleles of loci in different chromosomes are inherited independently of each other.
- The expected phenotypic ratio of a dihybrid cross is 9:3:3:1.
- The 9:3:3:1 ratio can be modified if the loci are not simple Dominant/recessive to each other, or if there are gene interactions, or if the two loci are linked.
- A test cross gives a ratio of 1:1:1:1 for loci that assort independently.

Key terms:

alleles
blending inheritance
classical genetics
crossing over
dihybrid
dihybrid cross
DNA cloning
dominant
genotype
Gregor Mendel
heritable traits
independent assortment
Law of Independent Assortment
linkage
locus
Mendel's First Law
Mendel's Second Law
molecular genetics
monohybrid cross
particulate inheritance
phenotype
polar bodies
Punnett square
recessive
test cross
tester
The Law of Equal Segregation
true breeding lines

Study Questions:

1. How would the results of the cross in **Figure 3-3 on page 24** have been different if heredity worked through blending inheritance rather than particulate inheritance? If the F_1 was selfed, how would successive generations appear? What in particular did Mendel's work demonstrate regarding the nature of whatever particle determined inheritance?

2. Imagine that astronauts provide you with living samples of multicellular organisms discovered on another planet. These organisms reproduce with a short generation time, but nothing else is known about their genetics.

   a) How could you define laws of heredity for these organisms?
   b) How could you determine what molecules within these organisms contained genetic information?
   c) Would the mechanisms of genetic inheritance likely be similar for all organisms from this planet?
   d) Would the mechanisms of genetic inheritance likely be similar to organisms from Earth?

3. What is the maximum number of alleles for a given locus in a normal gamete of a diploid species?

4. Does equal segregation of alleles into daughter cells happen during mitosis, meiosis, or both?

5. Wiry hair (*W*) is dominant to smooth hair (*w*) in dogs.

   a) If you cross a homozygous, wiry-haired dog with a smooth-haired dog, what will be the genotype and phenotype of the F1 generation?
   b) If two dogs from the F1 generation mated, what would be the most likely ratio of hair phenotypes among their progeny?
   c) When two wiry-haired *W/w* dogs actually mated, they had a litter of three puppies, which all had smooth hair. How do you explain this observation?
   d) Someone left a wiry-haired dog on your doorstep. Without extracting DNA, what would be the easiest way to determine the genotype of this dog?
   e) Based on the information provided in question 1, can you tell which, if either, of the alleles is wild-type?

6. An important part of Mendel's experiments was the use of homozygous lines as parents for his crosses. How did he know they were homozygous, and why was the use of the lines important?

7. **Figure 3-13 on page 31** shows Punnett squares for two of the four possible test crosses. Fill in the Punnett squares in **Figure 3-14 on page 31** for the other two possible genotypes of the unknown that aren't shown.

8. Based on meiosis, when dealing with two loci there will always be four distinct gamete types. But if the organism is homozygous, like the tester, all those gametes will look the same. In this situation, when writing a Punnett square, is it necessary to write out the four similar gametes? How would you redraw the Punnett Square on the right in **Figure 3-13 on page 31**?

Chapter 4
Part 1 - Sex Linkage

Figure 4-1
The *E/e* gene in turkeys is responsible for bronze or brown feather colour, and is located on the Z-chromosome (Flickr-stevevoght- CC BY-SA 2.0).

Introduction

Previously, Mendel, working with plants, showed patterns of inheritance derived from gene loci on autosomal chromosomes. One complication to this model of inheritance in animals is that loci present on sex chromosomes, called sex-linked loci, don't follow this pattern. This chapter covers the various patterns of inheritance for various sex-linked loci.

A Autosomes and Sex Chromosomes

In diploids, most chromosomes exist in pairs (same length, centromere location, and banding pattern) with one set coming from each parent. These chromosomes are called **autosomes**. However many species have an additional pair of chromosomes that do not look alike. These are **sex chromosomes** because they differ between the sexes. In humans, males have one of each while females have two X chromosomes. Autosomes are those chromosomes present in the same number in males and females, while sex chromosomes are those that are not. When sex chromosomes were first discovered their function was unknown and the name X was used to indicate this mystery. The next ones were named Y, then Z, and then W.

The combination of sex chromosomes within a species is associated with either male or female individuals. In mammals, fruit flies, and some dioecious plants, those with two *X* chromosomes are females while those with an *X* and a *Y* are males. In birds, moths, and butterflies males are *Z/Z* and females are *Z/W*. Because sex chromosomes have arisen multiple times during evolution the molecular mechanism(s) through which they determine sex differs among those organisms. For example, although humans and Drosophila both have X and Y sex chromosomes, they have different mechanisms for determining sex (see the next chapter).

How do the sex chromosomes behave during meiosis? Well, in those individuals with two of the same chromosome (i.e. **homogametic** sexes: *X/X* females and *Z/Z* males) the chromosomes pair and segregate during

meiosis I the same as autosomes do. During meiosis in *X/Y* males or *Z/W* females (**heterogametic** sexes) the sex chromosomes pair with each other.

In mammals (*X/X*, *X/Y*) the consequence of this is that all egg cells will carry an *X* chromosome, while the sperm cells will carry either an *X* or a *Y* chromosome. Half of the offspring will receive two

X chromosomes and become female while half will receive an *X* and a *Y* and become male (**Figure 4-2**). In species with *Z/Z* males, all sperm carry a Z chromosome, while in females, *Z/W*, half will have a *Z* and half a *W*. Sex Linkage = "An Exception to Mendel's First Law".

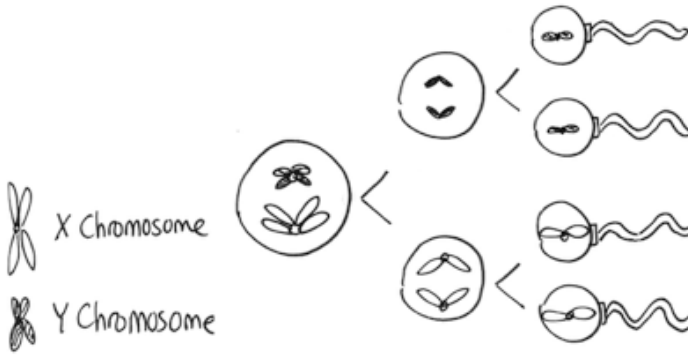


Figure 4-2

Meiosis in an X/Y mammal. The stages shown are anaphase I, anaphase II, and mature sperm. Note how half of the sperm contain Y chromosomes and half contain X chromosomes (Original-Harrington-CC BY-NC 3.0).

Above we introduced sex chromosomes and autosomes (non-sex-linked chromosomes). For loci on autosomes, the alleles follow the classic Mendelian pattern of inheritance. However, for loci on the sex chromosomes this doesn't follow because most (not all) of the loci on the typical X-chromosome are absent from the Y-chromosome, even though they act as a homologous pair during meiosis. Instead, they will follow a **sex-linked** pattern of inheritance. An X-linked allele in the father will always be passed on to his daughters only, but an X-linked allele in the mother will be passed on to both daughters and sons equally.

A.1 X-LINKED GENES: THE WHITE GENE IN DROSOPHILA MELANOGASTER

A well-studied sex-linked gene is the *white* gene on the *X* chromosome of *Drosophila melanogaster*. Normally flies have red eyes but flies with a mutant allele of this gene called *white* (X^{whi}) have white eyes because the red pigments are absent. Because this mutation is recessive to the wild type X^{whi+} allele, females that are heterozygous have normal red eyes. Female flies that are homozygous for the mutant allele have white eyes. Because there is no *white* gene on the *Y* chromosome, male flies can only be hemizygous for the wild type allele or the mutant allele, and consequently express that phenotype.

Note that the symbols X^{whi} *and* X^{whi+} are explained in **Appendix 2**. The letter "X" states that this is a gene on the X chromosome, and the superscript identifies the gene as one in which the mutant phenotype is "*whi*" ("short for white"). **Figure 4-3** and **Figure 4-4 on page 35** use the symbols "w^-" and "w^+" to state the same thing. These symbols don't inform you that the gene is on the X-chromosome; you infer it by the presence of a Y-chromosome in the male.

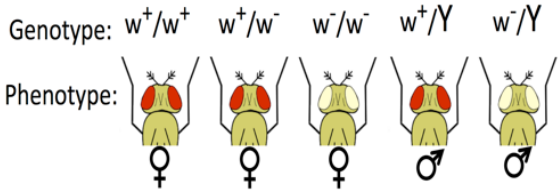


Figure 4-3

Relationship between genotype and phenotype for the white gene on the X-linked gene in Drosophila melanogaster. The Y chromosome is indicated with a capital Y. There is not white gene on the Y.

(Original-Deyholos/Harrington/Locke-CC BY-NC 3.0)

A researcher may not know beforehand whether a novel mutation is sex-linked. The definitive method to test for sex-linkage is **reciprocal crosses** (**Figure 4-4 on page 35**the next page). This means to cross a male and a female that have different phenotypes, and then conduct a second set of crosses, in which the phenotypes are reversed relative to the sex of the parents in the first cross. For example, if you were to set up reciprocal crosses with flies from pure-breeding w^- and w^+ strains, the results would be as shown in **Figure 4-4 on page 35** on the next page. Whenever reciprocal crosses give different results in the F1 and F2 and whenever the male and female offspring have different phenotypes the usual explanation is sex-linkage. Remember, if the locus were autosomal the F1 and F2 progeny would be different from either of these crosses.

A similar pattern of sex-linked inheritance is seen for X-chromosome loci in other species with an *X/X-X/Y* sex chromosome system, including mammals and humans. The *Z/Z-Z/W* system is similar, but reversed.

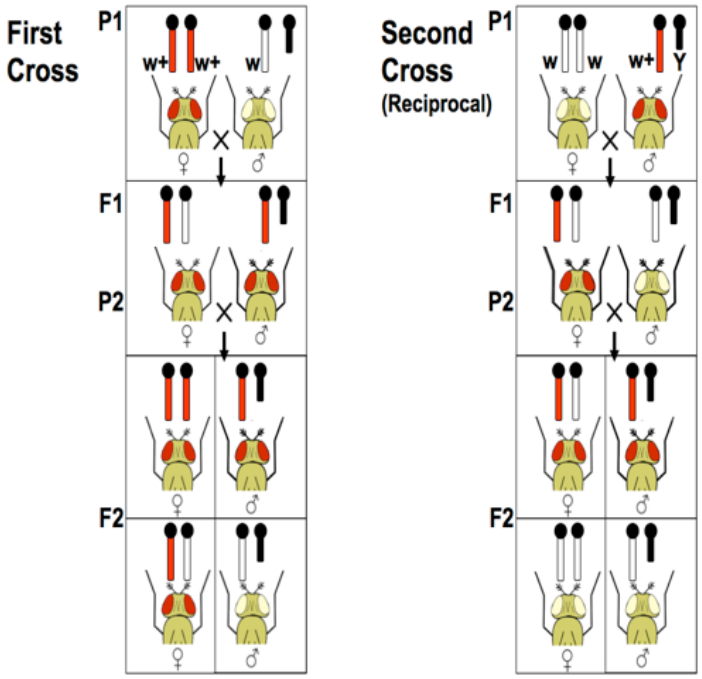


Figure 4-4

Reciprocal crosses involving an X-linked gene in Drosophila melanogaster. In the first cross (left) all of the offspring have red eyes. In second (reciprocal) cross (right) all of the female offspring have red eyes and the male offspring all have white eyes. If the F1 progeny are crossed (to make the P2), the F2 progeny will be different in each cross. The first cross has all red-eyed females and half red-eyed males. The reciprocal cross has half red-eyed males and females.

Thomas Morgan was awarded the Nobel Prize, in part, for using these crosses to demonstrate that genes (such as white) were on chromosomes (in this case the X-chromosome).

(Wikipedia- Deyholos/Harrington/Locke -PD)

A.2 Y-linked gene

Genes located on the Y-chromosome exhibit Y-linkage. For examples, the TDF gene that is responsible for sex determination and hairy ear rim phenotype show only father-to-son inheritance patterns.

A.3 Z-linked Genes in Birds

One last example is a **Z-linked gene** that influences feather colour in turkeys. Turkeys are birds, which use the ZZ-ZW sex chromosome system. The *E* allele makes the feathers bronze and the *e* allele makes the feathers brown (**Figure 5**). Only male turkeys can be heterozygous for this locus, because they have two Z chromosomes. They are also uniformly bronze because the *E* allele is completely dominant to the *e* allele and birds use a dosage compensation system similar to *Drosophila* and not mammals. Reciprocal crosses between turkeys from pure-breeding bronze and brown breeds would reveal that this gene is in fact Z-linked.

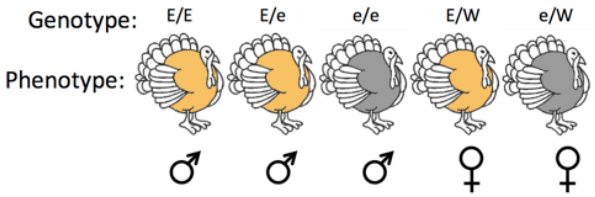


Figure 4-5

Relationship between genotype and phenotype for a Z-linked gene in turkeys. The W chromosome does not have an E/e-gene so it is just indicated with a capital W. (Original-Harrington/Locke-CC BY-NC 3.0)

More on these in the next part!

PART 2: Sex Chromosomes: Sex Determination

Figure 4-6
Not all species determine sex using the same mechanism. There are many factors that can determine a species' sex and one of them is growth temperature. For alligators, the temperature of the eggs in their nest determines sex. (Flickr-Florida Fish and Wildlife-CC BY-ND 2.0)

Introduction

In the previous Part, sex chromosome were described and their inheritance was compared to that of the autosomes. The linkage of sex chromosomes to the sex of individuals was presumed. In this chapter we will cover the mechanisms of sex determination by chromosomes (genes) as well as other, environmental, mechanisms. In the diversity of animal life, sex is not always determined by genetics (sex chromosomes).

B Sex Determination Mechanisms in Animals

There are various mechanisms for sex determination in animals. These include sex chromosomes, chromosome dosage, and environmental cues.

B.1 Sex chromosome systems:

X/Y system

Different combinations of the X and Y sex chromosomes can determine the sex of an organism. For example in humans and other mammals *X/Y* embryos develop as males while *X/X* embryos become females. This difference in development is due to the presence of only a **single gene**, the ***Testis-Determining Factor*** (***TDF***), also known as ***Sex-determining Region*** **Y** (***SRY***) gene, on the Y-chromosome. Its presence in the genome and expression in gonad tissues dictates that the sex of that individual will be male. Its absence or lack of correct expression results in a female phenotype for that individual.

In mammals, the *X/Y* sex determining system evolved just after the divergence of the monotreme lineage (mammals that lay eggs) from the lineage that led to marsupial mammals (young are carried in a pouch) and placental mammals. Thus nearly every mammal species uses the same sex determination system. In this system, during embryogenesis, the gonads will develop into either ovaries or testes (**Figure 4-7**).

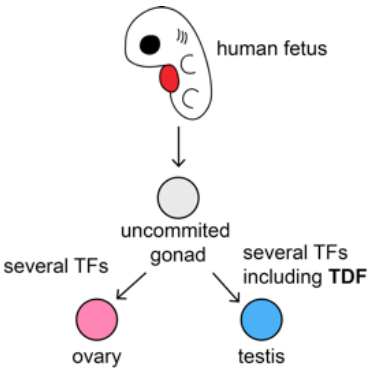


Figure 4-7
Gonad differentiation is under the control of several genes including Testis-determining factor (TDF, SRY) at Yp11.3. (y chromosome, p arm, region 1, band 1, sub-band 3). (Original-Harrington/Kang-CC BY-NC 3.0)

Fun fact: monotremes have five pairs of sex chromosomes and all the "X" chromosomes segregate together, as to all the "Y" chromosomes during sperm production! So much for independent assortment!

A gene present only on the Y chromosome called *TDF* encodes a protein that directs the gonads to mature into testes. *X/X* embryos do not have this gene and their gonads mature into ovaries instead, a default (**Figure 4-7 on page 36**).

TDF in **therians** (placental mammals and marsupials) encodes a DNA-binding transcription factor that, when combined with other factors, turns on genes that encode male-specific transcription factors. This begins a cascade of gene expression that leads to the differentiation of the gonad into testes. Mutations in the TDF-Y gene lead to a range of sex-related disorders with varying effects on an individual's phenotype. In some cases, the individual will morphologically develop as a female although both X and Y chromosomes are present.

Once formed the testes produce sex hormones that direct the rest of the developing embryo to become male, while the ovaries make different sex hormones that promote female development. The testes and ovaries are also the organs where gametes (sperm or eggs) are produced.

Z/W system

In birds, some fish, some insects (butterflies and moths) and reptiles, they use different chromosome for sex determination, Z- and W-chromosomes. Z-chromosome is larger and has more genes than the W-chromosome. *Z/Z* embryos become male and *Z/W* embryo become females. This sex linkage pattern is backwards of the X and Y sex linkage pattern. It is currently unknown if the presence of W chromosome induces female features or two copies of Z chromosome induces male features. In birds, researches have not yet found a *Z/Z/W* or *Z/0* individual.

X/O system

The X/O system (XX-female, X/O male), where O is an absence of a chromosome, is found in insects (e.g. grasshoppers). The absence of a chromosome means that there is not a specific gene that determines the sex of an individual, instead it is usually determined via chromosome dosage.

B.2 Chromosome dosage

X-Autosome Ratio

This mechanism involves ratios of autosome to sex chromosomes. This can occur even in species that have two sex chromosomes For example, although *Drosophila melanogaster* has *X/X-X/Y* sex chromosomes, its sex determination system uses a chromosome ration method, that of **X:Autosome (X:A) ratio**. In this system it is the ratio of autosome chromosome sets (A) relative to the number of X-chromosomes (X) that determines the sex. Individuals with two autosome sets and two X-chromosomes (2A:2X) develop as females, while those with only one X-chromosome (2A:1X) develop as males. The presence/absence of the Y-chromosome and its genes are not significant for determining sex, however there are genes on the Y-chromosome that are needed for male fertility. An *X/O* fly is phenotypically male but not fertile. By comparison, *X/O* mammals are phenotypically female because they lack the TDF gene.

Ploidy Level

In other species of animals the number of chromosome sets can determine sex. For example the **haploid-diploid system** is used in bees, ants, and wasps. Typically haploids are male and diploids are female.

C Environmental Factors

C.1 Growth temperature

Alligators – Sex is determined by the temperature during development in the egg and individuals are fully determined by the time of hatching. Developmental temperatures of 30°C produce all females (nests constructed on levees). Developmental temperatures of 34°C yield all males (wet marsh nests). The natural sex ratio at hatching is five females to 1 male. Note that such a mechanism is sensitive to warming environmental temperatures.

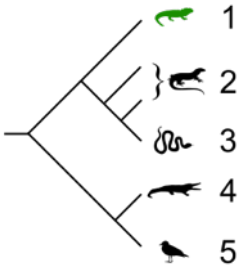


Figure 4-8
The tuatara (left) is a reptile, but not a lizard, although it is related to lizards (right).
Cladogram:
1=Tuatara
2=lizards
3=snakes
4=crocodiles
5=birds
(Left: Flickr-PhillipC- CC BY 2.0) (Right: Wikipedia-Benchill-CC BY 3.0)

Tuatara (**Figure 4-8**) – These reptiles look like lizards but are a distinctly separate Order, which has survived for over 200 million years. There are currently only two extant species. Embryo's development temperature determines the animal's sex; low temperatures (below a threshold) develop into females. High temperatures (above a threshold) develop into males. Global warming will affect the sex ratio in the population. By 2080 there will be conditions that produce 100% male

C.2 Social Organization

Sex-ratio in a population determines the sex of a population. For example, most Reef fish can change their sex during their lifetime. For example, the Wrasse family (**Figure 4-9**) includes many different species of various sizes and colours. In this family, sex change is typically female-to-male (male-to-female sex change has been seen in experimental conditions). The individual to change sex is generally the largest female in a group.

Figure 4-9
Moon Wrasse (Thalassoma lunare) can change sex.
(Flickr- Nick Hobgood- CC BY-NC 2.0)

C.3 Parthenogenetic Species

In **parthenogenetic** species, females can lay fertile eggs without requiring males. Examples include walking stick insects, some fish and lizards, and sharks in captivity.

Table 4-1
A summary table outlining various factors that affect sex determination and its genetic and cell response mechanism.

Determining Factors	Genetic Mechanism	Cell Response Mechanism
Chromosomal: • *XX/XY* • *ZW/ZZ* • *XX/XO* • Haploid/Diploid	• Single gene • X-Autosome Ratio (gene dosage)	• Hormonal: directs cells to sex phenotype • Cell-autonomous (each cell "knows" what sex it is)
Environment: • Rearing temp. • Social interactions • Parthenogenesis	Not genetic	Hormonal?

Part 3: Sex Chromosomes: Dosage Compensation

Figure 4-10
A calico cat showing the random inactivation (X-inactivation) of one or the other X-chromosome giving either an orange or black fur colour. The inactivation is a mechanism of dosage compensation. (Note: the white colour pattern is due to another gene.) (Original-J. Locke-CC:AS)

Introduction

The previous chapters on sex chromosomes dealt with sex linkage and sex determination. Now, there is one last issue dealing with sex chromosomes, that of **dosage compensation**. Because the number of X chromosomes (and Z chromosomes) differs between the sexes, there is a difference in the number of copies for each locus on the chromosome: females have two, while males only have one (opposite for the *Z/Z-Z/W* system).

D Gene Dosage Problem

For many loci, the different number of chromosomes is inconsequential. That is, the phenotype is unaffected whether there are one or two alleles present. However, for some loci, it is significant and can affect the phenotype. These loci need to have the correct gene dosage to generate a wild type phenotype. The dosage difference between the sexes is reconciled in one of two ways. Either the single X chromosome in males is up-regulated to produce the expression equivalent of two doses. Or, one of the two doses in females is inactivated so as to only have one active dose.

Mammals and *Drosophila* both have *X/X – X/Y* sex determination systems. However, because these systems evolved independently, and very early in evolution, they work differently with regard to compensating for the difference in gene dosage. Remember, in most cases the sex chromosomes act as a homologous pair even though the Y-chromosome has lost most of the loci when compared to the X-chromosome. Typically, the X and the Y chromosomes were once similar but, for unclear reasons, the Y chromosomes have degenerated, slowly mutating and losing its loci. In modern day mammals the Y chromosomes have very few genes left while the X chromosomes remain as they were. This is a general feature of all organisms that use chromosome based sex determination systems. Chromosomes found in both sexes (the X or the Z) have retained their genes while the chromosome found in only one sex (the Y or the W) have lost most of their genes. In either case there is a gene dosage difference between the sexes: *e.g. X/X* females have two doses of X-chromosome genes while XY males only have one. This gene dosage needs to be compensated in a process called **dosage compensation**. There are two major mechanisms.

D.1 Dosage Compensation in Drosophila

In Drosophila and many other insects, dosage compensation takes place in males. To make up for having only a single X chromosome, the genes on it are transcribed at twice the normal rate. This increased gene expression restores a balance between proteins encoded by **X-linked genes** and those made by **autosomal genes**.

D.2 X-chromosome Inactivation in Mammals

Basics

In mammals a different mechanism is used, called **X-chromosome inactivation** and it operates in females, not males. In XX embryos one X in each cell is randomly marked and inactivated. From that point forward most of the genes on this chromosome will be unexpressed or "inactive", hence its name $X_{inactive}$ (X_i). The other X chromosome, the X_{active} (X_a), is unaffected and genes are expressed as they normally would be. The inactivation process is under the control of the X-inactivation centre (XIC), located at Xq13 on the X-chromosome, which contains several genes including *XIST* gene. XIST gene is transcribed (but not translated into a protein) and is responsible for the initiation and propagation of inactivation of one X-chromosome in an XX cell. These *XIST* RNA transcripts coat the X chromosome so that the transcription from that X chromosome is prevented (inactivated).

The X_i chromosome is replicated during S phase and transmitted during mitosis the same as any other chromosome, but most of its genes are not transcribed (**Figure 4-11**). The chromosome appears as a condensed mass within interphase nuclei and is called the **Barr body** (**Figure 4-12**) and does not decondense to be expressed. (The Barr body is named after Canadian researcher, Murray Barr, who along with his graduate student Ewart Bertram at Western University in London Ontario discovered it in 1948.) With the inactivation of genes on one X-chromosome, females have the same number of functioning X-linked genes as males. However, some genes and particularly those in the pseudoautosomal regions escape inactivation and the alleles are expressed from both active and inactive X chromosomes. These genes may explain clinical features in sex chromosome aneuploidy as gene products may be either under or over expressed in relation to normal females and males.

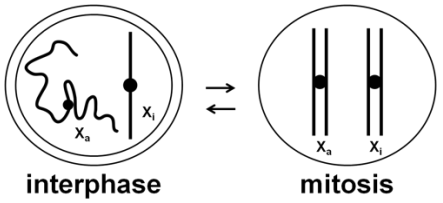


Figure 4-11
X chromosome inactivation during mitosis and after mitosis. (Original-Harrington/Kang-CC BY-NC 3.0)

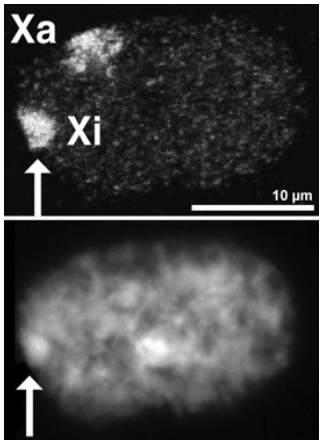


Figure 4-12
X chromosomes detected by FISH method in a female cell's nucleus. (Wikipedia-Steffen Dietzel-CC BY-SA 3.0)

E X-linked Genes – Orange Gene in Cats

A classic X-linked gene that shows X-inactivation is the *Orange* gene (*O*) in cats. The O^O allele encodes an enzyme that results in orange pigment in the fur hairs. The O^B allele results in the hairs being black. The phenotypes of various genotypes of cats are shown in **Figure 13.** Note that the heterozygous females have an orange and black mottled phenotype known as tortoiseshell. This is due to patches of skin cells having different X-chromosomes inactivated. In each orange hair the X_i chromosome carrying the O^B allele is inactivated. The O^O allele on the X_a is functional and orange pigments are made. In black hairs the reverse is true, the X_i chromosome with the O^O allele is inactive and the X_a chromosome with the O^B allele is active. Because the inactivation decision happens early during embryogenesis, the cells continue to divide to make large patches on the adult cat skin where one or the other X is inactivated.

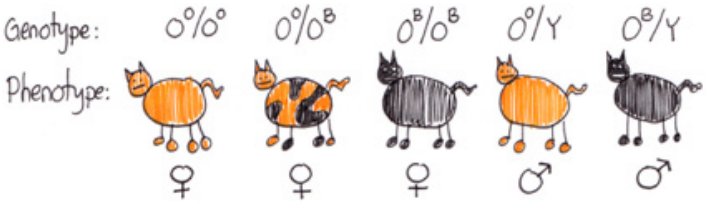


Figure 4-13
Relationship between genotype and phenotype for an X-linked gene in cats. The O^O allele = orange while the O^B allele = black. (Original-Harrington-CC BY-NC 3.0)

The *Orange* gene in cats is also a good demonstration of how the mammalian dosage compensation system affects gene expression. However, most X-linked genes do not produce such dramatic, easy to see, mosaic phenotypes in heterozygous females.

E.1 A typical X-linked gene – F8 gene in humans

A more typical example of an X-linked gene is the *F8* gene in humans. It makes Factor VIII blood clotting proteins in liver cells. If a male is hemizygous for a mutant allele ($F8^-/Y$) the result is hemophilia type A. Females homozygous for mutant alleles ($F8^-/F8^-$) will also

have hemophilia. However, heterozygous females, those people who are $F8^{+}/F8^{-}$, do not have hemophilia because even though half of their liver cells do not make Factor VIII (because the X with the $F8^{+}$ allele is inactive) the other 50% can (**Figure 4-14**). Because some of their liver cells are producing and exporting Factor VIII proteins into the blood stream they have the ability to form blood clots throughout their bodies. The genetic mosaicism in the liver cells of their bodies does not result in a visible mosaic phenotype.

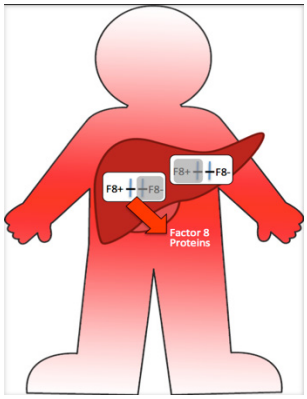


Figure 4-14
This figure shows the two types of liver cells in females heterozygous for an *F8* mutation. Because people with the $F8^{+}/F8^{-}$ genotype have the same phenotype, normal blood clotting, as $F8^{+}/F8^{+}$ people the $F8^{-}$ mutation is classified as recessive. (Original-Harrington/ Locke-CC BY-NC 3.0)

F Mechanisms of Sex Determination Systems

Sex is a phenotype. Typically, in most species, there are multiple characteristics, in addition to sex organs, that distinguish male from female individuals (although some species are normally **hermaphrodites** where both sex organs are present in the same individual; e.g. worms). The morphology and physiology of male and females is a phenotype just like hair or eye colour or wing shape. The sex of an organism is part of its phenotype and can be genetically (or environmentally) determined.

For each species, the genetic determination relies on one of several gene or chromosome based mechanisms. See **Figure** 15 for a summary. There are, for other species, also a variety of environmental mechanisms, too (rearing temperature, social interactions, **parthenogenesis**).

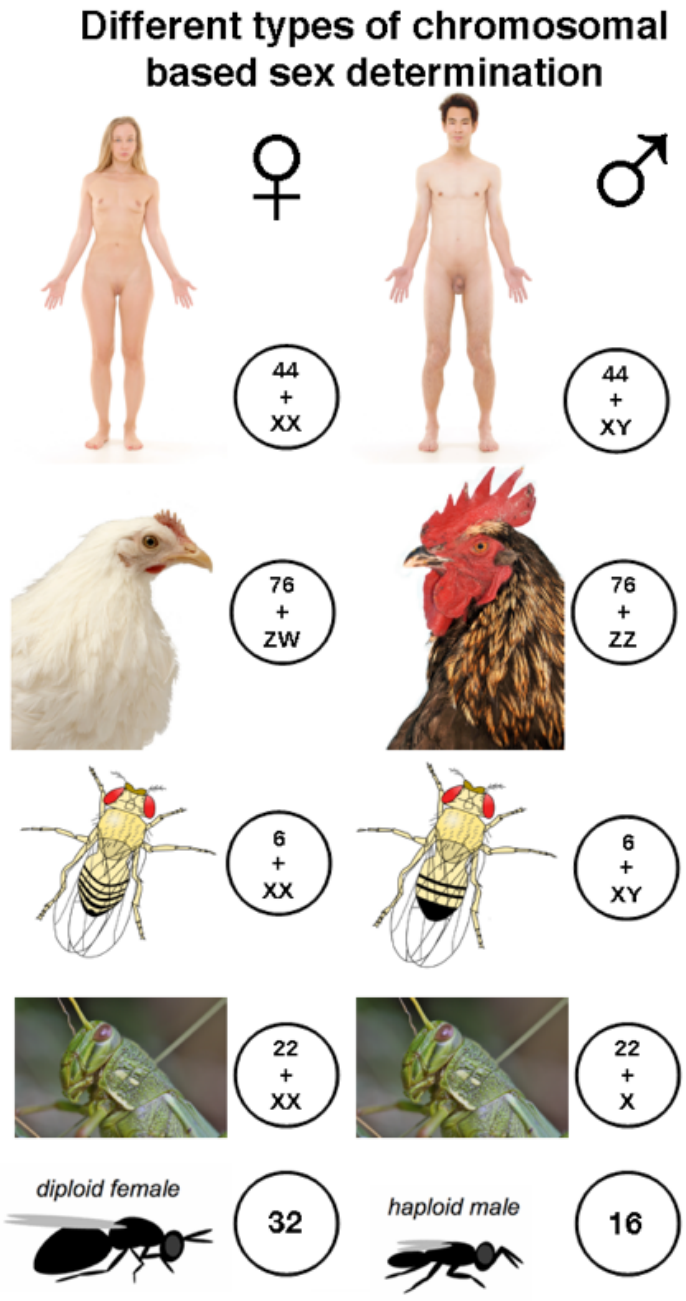


Figure 4-15
Different types of chromosomal (or gene) based sex determination. From top to bottom, there is the archetypal *X/X-X/Y* system found in humans (and most mammals) with the TDF-Y gene leading to a male phenotype; the *Z/W-Z/Z* system found in chickens (birds, moths, and butterflies); the same *X/X-X/Y* system in Drosophila (sex is determined by the X-chromosome:autosome ratio); the *X/X-X/O* system as found in grasshoppers; and the diploid/haploid system as found in bees (and ants, and wasps). Also, the hormonal mechanism is used in humans, while all the other examples use the cell-autonomous mechanism for development of the male or female sex phenotype.

(Wikipedia-original - CFCF with additions and corrections by J. Locke- CC BY-SA 3.0)

This is an encore presentation of the online exercise in assigning gene names to X-linked loci.

http://tinyurl.com/oog-sex

Summary:

- Autosomes and sex chromosomes differ in that the former exist in pairs but the latter depends on the sex of the chromosome.
- Pseudo-autosomal regions are regions on X and Y chromosome that can pair up and recombine.
- Sex-linked genes are an exception to standard Mendelian inheritance. The type of sex chromosome system and the type of dosage compensation system found in the species influence their phenotypes.
- Some of the examples of sex-linked genes are: *white* gene on the Drosophila's X chromosome, *TDF* gene on Y chromosome, *E/e* gene on Z chromosome.
- The sex of an individual can be determined by sex chromosomes
- This includes the X/Y, Z/W, and X/O system
- Also, differences in the ploidy level (haploid vs diploid) determine sex in some species
- Lastly, environmental factors such as rearing temperature or social organization (male vs female ratio) can determine sex. In order to compensate for under or over dosage of gene products, organisms use various methods such as expressing genes twice the normal rate or inactiving X chromosome.
- X-chromosome inactivation occurs randomly (except for special circumstances), and during interphase the inactivated chromosome appears as a condensed mass in the nucleus called the Barr body.
- Orange gene in cats and F8 gene in humans are examples of X-linked genes.

Key terms:

autosome
autosomal genes
Barr body
dosage compensation
F8 gene
haploid-diploid system
heterogametic
homogametic
Orange gene
parthenogenesis
parthenogenetic
pseudoautosomal regions heteromorphic
reciprocal cross
sex chromosome
Sex-determining Region Y (SRY)
sex-linked
Sex-ratio
sexually dimorphic
single gene
Testis-Determining Factor (TDF)
therians
Tuatara
X-chromosome inactivation
X-linked genes
X-linked genes
X:Autosome (X:A) ratio
X/O system
X/Y system
Z-linked genes
Z/W system

Study Questions:

1. Draw reciprocal crosses that would demonstrate that the turkey *E*-gene is on the Z chromosome.
2. Mendel's First Law (as stated in class) does not apply to alleles of most genes located on sex chromosomes. Does the law apply to the chromosomes themselves?
3. A rare dominant mutation causes a neurological disease that appears late in life in all people that carry the mutation. If a father has this disease, what is the probability that his daughter will also have the disease?
4. Make Punnett Squares to accompany the crosses shown in **Figure 4-4 on page 35**.
5. Another cat hair colour gene is called *White Spotting*. This gene is autosomal. Cats that have the dominant *S* allele have white spots. What are the possible genotypes of cats that are:
   a) entirely black
   b) entirely orange
   c) black and white
   d) orange and white
   e) orange and black (tortoiseshell)
   f) orange, black, and white (calico)
6. What is the relationship between the O^0 and O^B alleles of the *Orange* gene in cats?
7. Make a diagram similar to **Figure 4-14 on page 41** that shows the relationship between genotype and phenotype in females and males for the *F8* gene.

Chapter 5
Gene Interactions

Figure 5-1
Coat color in mammals is an example of a phenotypic trait that is controlled by more than one locus and the alleles at these loci can interact to alter the expected Mendelian ratios. (Flickr-David Blaikie- CC BY 2.0)

Introduction

The principles of genetic analysis that we have described for a single locus (dominance/ recessiveness) can be extended to the study of alleles at two different loci. While the analysis of two loci concurrently is required for genetic mapping, it can also reveal interactions between genes that affect the phenotype. Understanding these interactions is very useful for both basic and applied research. Before discussing these interactions, we will first revisit Mendelian inheritance for two loci.

A Mendelian Dihybrid Crosses

A.1 Mendel's second law (a quick review)

To analyze the segregation of two traits (e.g. colour, wrinkle) at the same time, in the same individual, Mendel crossed a pure breeding line of green, wrinkled peas with a pure breeding line of yellow, round peas to produce F_1 progeny that were all green and round, and which were also **dihybrids**; they carried two alleles at each of two loci (**Figure 5-2**).

If the inheritance of seed color was truly independent of seed shape, then when the F_1 dihybrids were crossed to each other, a 3:1 ratio of one trait should be observed within each phenotypic class of the other trait (Figure 5-2). Using the product law, we would therefore predict that if ¾ of the progeny were green, and ¾ of the progeny were round, then ¾ × ¾ = 9/16 of the progeny would be both round and green. Likewise, ¾ × ¼ = 3/16 of the progeny would be both round and yellow, and so on. By applying the product

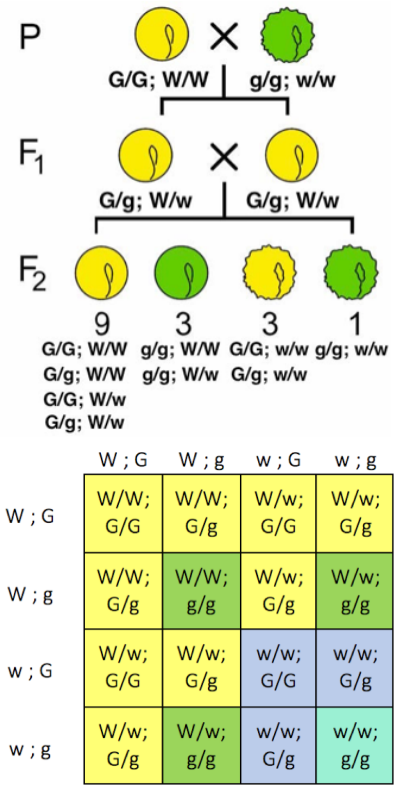


Figure 5-2
Pure-breeding lines are crossed to produce dihybrids in the F_1 generation. The cross of these particular dihybrids produces four phenotypic classes.

(Original-Deyholos-CC BY-NC 3.0)

rule to all of these combinations of phenotypes, we can predict a **9:3:3:1** phenotypic ratio among the progeny of a dihybrid cross, if certain conditions are met, including the independent segregation of the alleles at each locus. Indeed, 9:3:3:1 is very close to the ratio Mendel observed in his studies of dihybrid crosses, leading him to state his Second Law, the **Law of Independent Assortment**, which we now express as follows: two loci assort independently of each other during gamete formation.

A.2 Assumptions of the 9:3:3:1 ratio

Both the product rule and the Punnett Square approaches showed that a 9:3:3:1 phenotypic ratio is expected among the progeny of a dihybrid cross such as Mendel's *R/r;Y/y* × *R/r;Y/y*. In making these expectations, we assumed that:

(1) both loci assort independently (not the semicolon);

(2) one allele at each locus is completely dominant; and

(3) each of four possible phenotypes can be distinguished unambiguously, with no interactions between the two genes that would alter the phenotypes.

Deviations from the 9:3:3:1 phenotypic ratio may indicate that one or more of the above conditions has not been met. For example, Linkage of the two loci results in a distortion of the ratios expected from independent assortment. Also, if complete dominance is lacking (e.g. co-dominance or incomplete dominance) then the ratios will also be distorted. Finally, it there is an interaction between the two loci such that the four classes cannot be distinguished (which is the topic under consideration in this chapter) the ratio will also deviate from 9:3:3:1.

Modified ratios in the progeny of a dihybrid cross can therefore reveal useful information about the genes being investigated. Such interactions lead to **Modified Mendelian Ratios**.

B Epistasis and Other Gene Interactions

Some dihybrid crosses produce a phenotypic ratio that differs from the typical 9:3:3:1. These include 9:3:4, 12:3:1, 9:7, or 15:1. Note that each of these modified ratios can be obtained by summing one or more of the 9:3:3:1 classes expected from our original dihybrid cross. In the following sections, we will look at some modified phenotypic ratios obtained from dihybrid crosses and what they might tell us about the interactions between the genes involved.

B.1 Epistasis

Epistasis (which means "standing upon") occurs when the phenotype of one locus masks, or prevents, the phenotypic expression of another locus. Thus, following a dihybrid cross fewer than the typical four phenotypic classes will be observed with epistasis. As we have already discussed, in the absence of epistasis, there are four phenotypic classes among the progeny of a dihybrid cross. The four phenotypic classes correspond to the genotypes: *A/-;B/-, A/-;b/b, a/a;B/-,* and *a/a;b/b*. If either of the singly homozygous recessive genotypes (i.e. *A/-;b/b* or *a/a;B/-*) has the same phenotype as the double homozygous recessive (*a/a;b/b*), then a **9:3:4** phenotypic ratio will be obtained.

For example, in the Labrador Retriever breed of dogs (**Figure 5-3**), the B locus encodes a gene for an important step in the production of melanin. The dominant allele, *B* is more efficient at pigment production than the recessive *b* allele, thus *B*/- hair appears black, and *b/b* hair appears brown. A second locus, which we will call *Y*, controls the deposition of melanin in the hairs. At least one functional *Y* allele is required to deposit any

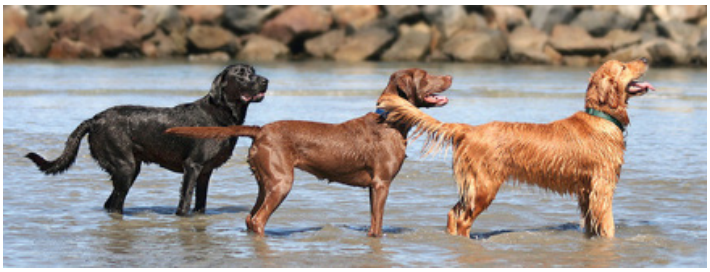


Figure 5-3

Retrievers with different coat colors: (from left to right) black, chocolate, yellow: an example of recessive epistasis phenotypes.

(Flickr- Pirate Scott - CC BY-NC 2.0)

	Y/B	Y/b	y/B	y/b
Y/B	Y/Y;B/B	Y/Y;B/b	Y/y;B/B	Y/y;B/b
Y/b	Y/Y;B/b	Y/Y;b/b	Y/y;B/b	Y/y;b/b
y/B	Y/y;B/B	Y/y;B/b	y/y;B/B	y/y;B/b
y/b	Y/y;B/b	Y/y;b/b	y/y;B/b	y/y;b/b

Figure 5-4

Genotypes and phenotypes among the progeny of a dihybrid cross of Labrador Retrievers heterozygous for two loci affecting coat color. The phenotypes of the progeny are indicated by the shading of the cells in the table: black coat (black, *Y/-;B/-*); chocolate coat (brown, *Y/-;b/b*); yellow coat (yellow, *y/y;B/-* or *y/y;b/b*).

(Original-Nickle-CC BY-SA 3.0)

pigment, whether it is black or brown. Thus, all retrievers that are *y/y* fail to deposit any melanin (and so appear pale yellow-white), regardless of the genotype at the *B* locus (**Figure 5-3**, right side).

The *y/y* genotype is therefore said to be **epistatic** to both the B and b alleles, since the homozygous *y/y* phenotype masks the phenotype of the *B* locus. The *B/b* locus is said to be **hypostatic** to the *y/y* genotype. A graphic showing all the possible progeny genotypes and their phenotypes is shown in **Figure 5-4**.

In some cases, a dominant allele at one locus may mask the phenotype of a second locus. This produces a segregation ratio of **12:3:1**, which can be viewed as a modification of the 9:3:3:1 ratio in which the *A/-;B/-* class is combined with one of the other genotypic classes (9+3) that contains a dominant allele. One of the best known examples of a 12:3:1 segregation ratio is fruit color in some types of squash (**Figure 5-5**). Alleles of a locus that we will call *B* produce either yellow (*B*/-) or green (*b/b*) fruit. However, in the presence of a dominant allele at a second locus that we call *A*, no pigment is produced at all, and fruit are white. The dominant *A* allele is therefore epistatic to both *B* and *b/b* combinations (**Figure 5-6**). One possible biological interpretation of this segregation pattern is that the function of the *A* allele somehow blocks an early stage of pigment synthesis, before either yellow or green pigments are produced.

Figure 5-5
Green, yellow, and white fruits of squash.

(Flickr-Unknown-CC BY-NC 3.0)

	A/B	A/b	a/B	a/b
A/B	A/A;B/B	A/A;B/b	A/a;B/B	A/a;B/b
A/b	A/A;B/b	A/A;b/b	A/a;B/b	A/a;b/b
a/B	A/a;B/B	A/a;B/b	a/a;B/B	a/a;B/b
a/b	A/a;B/b	A/a;b/b	a/a;B/b	a/a;b/b

Figure 5-6
Genotypes and phenotypes among the progeny of a dihybrid cross of squash plants heterozygous for two loci affecting fruit color.

(Original-Nickle-CC BY-SA 3.0)

B.2 Duplicate Gene Action

When a dihybrid cross produces progeny in two phenotypic classes in a 15:1 ratio, this can be because the proteins from each different gene have the same (redundant) functions within the same biological pathway. With yet another pigmentation pathway example, wheat shows this form of epistasis. The biosynthesis of red pigment near the surface of wheat seeds (**Figure 5-7**) involves many genes, two of which we will label *A* and *B*. Normal, red coloration of the wheat seeds is maintained if function of either of these genes is lost in a homozygous mutant (e.g. in either *a/a;B/-* or *A/-;b/b*). Only the doubly recessive mutant (*a/a;b/b*), which lacks function of both genes, shows a phenotype that differs from that produced by any of the other genotypes (**Figure 5-8**). A reasonable interpretation of this result is that both genes encode the same biological function, and either one alone is sufficient for the normal activity of that pathway.

Figure 5-7
Red (left) and white (right) wheat seeds.

(cropwatch.unl.edu)

	A/B	A/b	a/B	a/b
A/B	A/A;B/B	A/A;B/b	A/a;B/B	A/a;B/b
A/b	A/A;B/b	A/A;b/b	A/a;B/b	A/a;b/b
a/B	A/a;B/B	A/a;B/b	a/a;B/B	a/a;B/b
a/b	A/a;B/b	A/a;b/b	a/a;B/b	a/a;b/b

Figure 5-8
Genotypes and phenotypes among the progeny of a dihybrid cross of a wheat plants heterozygous for two loci affecting seed color.

(Original-Nickle-CC BY-SA 3.0)

B.3 Complementary Gene Action

The progeny of a dihybrid cross may produce just two phenotypic classes, in an approximately **9:7** ratio. An interpretation of this ratio is that the loss of function of either A or B gene function has the same phenotype as the loss of function of both genes. For example, consider a simple biochemical pathway in which a colorless substrate is converted by the action of gene A to another colorless product, which is then converted by the action of gene B to a visible pigment (**Figure 5-9** next page).

Loss of function of either A or B, or both, will have the same result: no pigment production. Thus *A/-;b/b*,

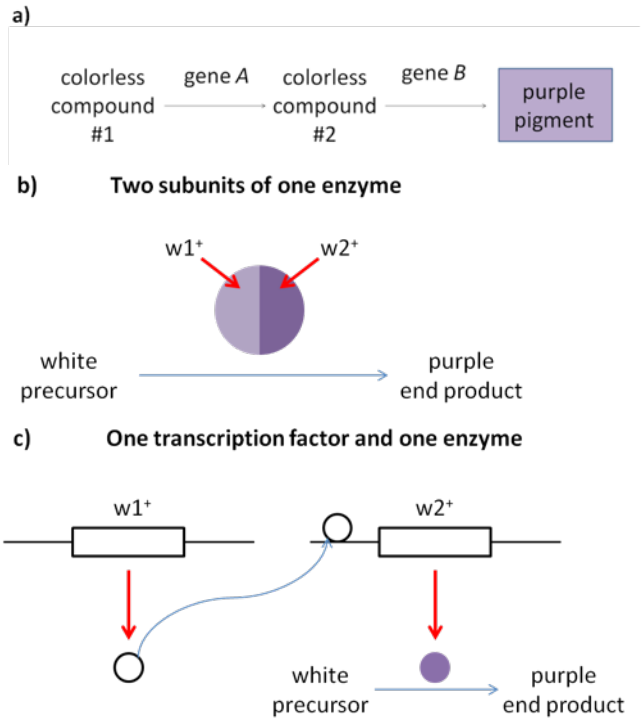


Figure 5-9

a) A simplified biochemical pathway showing complementary gene action of A and B. Note that in this case, the same phenotypic ratios would be obtained if gene B acted before gene A in the pathway.

b) biochemical pathway showing two subunits of one enzyme

c) biochemical pathway showing one transcription factor and one enzyme

(Original-Deyholos/KangCC BY-NC 3.0)

a/a;B/-, and *a/a;b/b* will all be colorless, while only *A/-;B/-* genotypes will produce pigmented product (**Figure 5-10**). The modified 9:7 ratio may therefore be obtained when two genes act together in the same biochemical pathway, and when their loss of function phenotypes are indistinguishable from each other or from the loss of both genes. There are also other possible biochemical explanations for complementary gene action.

	A/B	A/b	a/B	a/b
A/B	A/A;B/B	A/A;B/b	A/a;B/B	A/a;B/b
A/b	A/A;B/b	A/A;b/b	A/a;B/b	A/a;b/b
a/B	A/a;B/B	A/a;B/b	a/a;B/B	a/a;B/b
a/b	A/a;B/b	A/a;b/b	a/a;B/b	a/a;b/b

Figure 5-10

Genotypes and phenotypes among the progeny of a dihybrid cross of a hypothetical plant heterozygous for two loci affecting flower color.

(Original-Nickle-CC BY-SA 3.0)

C GENETIC SUPPRESSION: RECESSIVE AND DOMINANT SUPPRESSION

A **Suppressor mutation** is a type of mutation that suppresses the phenotypic expression of another mutation that already exists, which results in a more wild type (less mutant) phenotype. On the other hand, **enhancer mutations** have the opposite effect of suppressor mutations as they make the phenotype more mutant (enhance the mutant phenotype).

For example, if a fly has a mottled (whi^m) phenotype, it can be suppressed to look more like whi^+ phenotype by a dominant suppressor mutation, or enhanced to look more like whi^- by a dominant enhancer mutation (whi^-/whi^- whi^m/whi^m $whi^+/-$ or whi^m/whi^-; see **Figure 5-11**).

Note that whi^m is recessive to whi^+ but dominant to whi^-. This is an example of an **allelic series** (more than one allele of a gene can be in play. Another example is the A/B/O blood type series of alleles).

Figure 5-11

Mutation in the *white* gene impacts the pigmentation in *Drosophila* eyes. Note that whi^m (for $white^{mottled}$) is recessive to whi^+ and dominant to whi^-.

(Original-Locke-CC BY-NC 3.0)

The suppressor mutation can happen within the original gene itself (**intragenic**) or outside the gene, at some other gene elsewhere in the genome (**extragenic**). For example, a frameshift mutation caused by a deletion in gene A can be reverted by an insertion in the same gene to regain the reading frame (intragenic suppressor mutation). On the other hand in extragenic suppressor mutation, a defect caused by mutation in gene A can be suppressed by a mutation in gene B. In extragenic suppressor mutation, there are two types of suppressor mutations: (1) **dominant suppression** and (2) **recessive suppression.**

C.1 Dominant Suppression

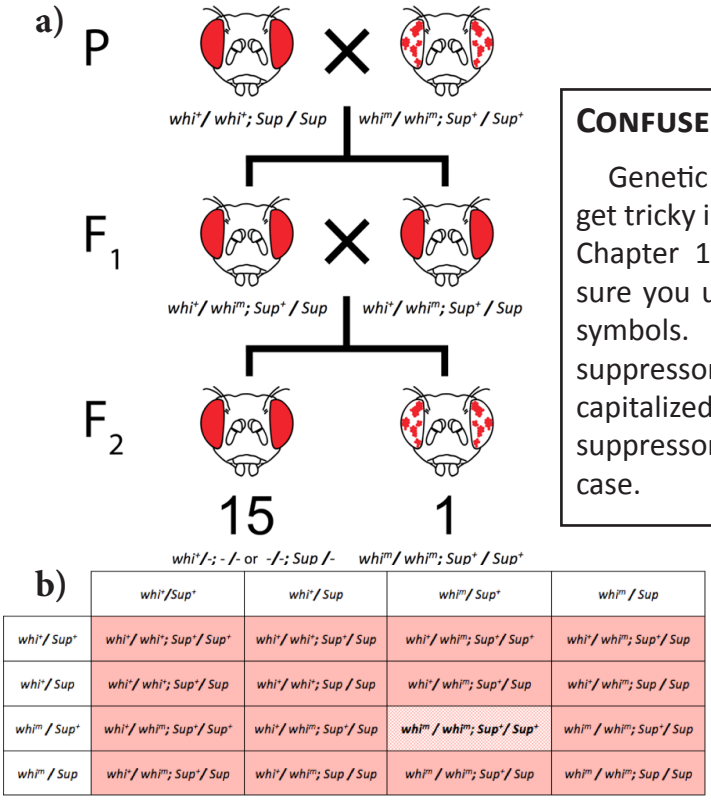
In **dominant suppression**, the mutant suppressor allele (*Sup*) is dominant to the wild type suppressor allele (*Sup*$^+$). Therefore, one mutant suppressor allele is sufficient to suppress the mutant phenotype. For example, in **Figure 5-12**, the *Sup* gene represents the suppressor gene. Flies that have at least one *Sup* allele, even though they have a homozygous recessive *whi*^{m}/*whi*^{m} genotype, will show a wild-type (*whi*$^+$) phenotype. A fly will have *whi*^{m} phenotype only if it has homozygous recessive *Sup*$^+$/*Sup*$^+$ genotype. *If whi*$^+$/*whi*^{m};*Sup*$^+$/*Sup* flies are crossed together, the ratio of *whi*$^+$/-(wild type) to *whi*^{m} /*whi*^{m} (mutant) would be 15:1.

C.2 Recessive Suppression

On the other hand, in **recessive suppression**, the mutant suppressor allele (*rsp*) is recessive to the wild type suppressor allele (*rsp*$^+$). Therefore, two of the mutant alleles are needed to suppress the *whi*^{m} (mottled) phenotype. For example, in **Figure 5-13**, flies that have at least one *whi*$^+$ allele will show a wild-type phenotype. Also, flies that homozygous for *rsp* will have wildtype phenotype since only two mutant alleles can suppress the *white* gene mutation. On the other hand, flies that have the *whi*^{m}/*whi*^{m} alleles will have mottled phenotype unless they have homozygous *rsp* alleles. If *whi*$^+$/*whi*^{m};*rsp*$^+$/*rsp* flies are crossed, the ratio of *whi*$^+$/- (wild type) to *whi*^{m}/ *whi*^{m} (mutant) would be 13:3.

Confused?

Genetic problems can get tricky in a hurry. Check Chapter 1 again to make sure you understand gene symbols. The dominant suppressor, *Sup* is capitalized; the recessive suppressor *rsp* is all lower case.



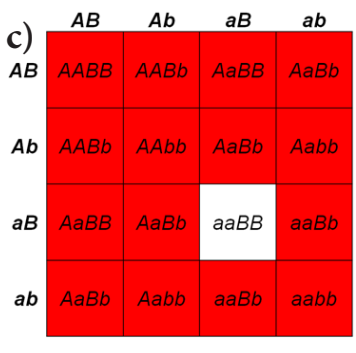
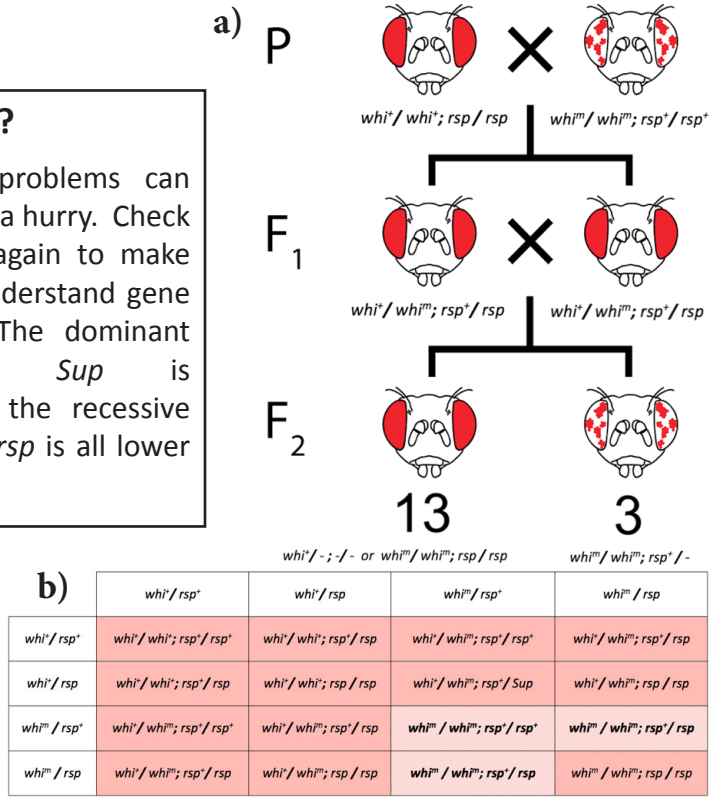


Figure 5-12

Drosophila cross and its Punnett square showing the effects of dominant suppression of the *Sup* gene on the *whi* gene. Note that a dash(-) = indicates any allele for that locus.

a) crosses and genotypes; b) Punnett square with proper gene names; c) simplified Punnett square where A=*whi*$^+$, a=*whi*, B=*Sup*$^+$, and b=*Sup*

(Image-Kang, modified by Nickle-CC BY-NC 3.0; Table – Nickle – CC BY-SA 3.0)



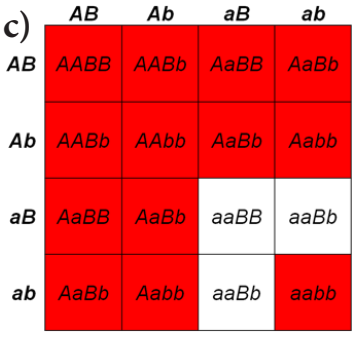


Figure 5-13

Drosophila cross and its Punnett square showing the effects of dominant suppression of the *rsp* gene on the *whi* gene. Note that a dash(-) = indicates any allele for that locus.

a) crosses and genotypes; b) Punnett square with proper gene names; c) simplified Punnett square where A=*whi*□, a=*whi*, B=*rsp*□, and b=*rsp*

(Image-Kang, modified by Nickle-CC BY-NC 3.0; Table – Nickle – CC BY-SA 3.0)

You've had an awful lot of interactions to consider! So far you've seen how the 9:3:3:1 ratios are altered by having a unique phenotype for each genotype. With no gene interaction, you should see 2^n phenotypes where n=# genes involved.

Also critical for this section is that the genes must be independently assorting. Genes that are close enough together on the same chromosome are "linked", and linkage always shows a shift in the ratios that differs from unlinked genes. See Chapters 3 (page 28), 9, and 10 for details.

Table 5-1 summarizes the different forms of epistasis and the ratios they produce for interacting, unlinked loci.

Table 5-1

Summary showing gene interactions and their genotypic (italic) and phenotypic (bold) ratios. Shading represents combined classes. Letters in italics represent alleles. Bolded letters refer to phenotypes. Where possible, the phenotype is given the same letter as the allele which is responsible for it. The names of the various forms of epistasis are provided, but for this course you need not memorize them. Just recognize how the 9:3:3:1 ratios are altered and be able to provide possible biochemical reasons for the alteration in ratio.

Fraction:	9	3	3	1	Ratio
Genotype:	*A/-;B/-*	*A/-;b/b*	*a/a;B/-*	*a/a;b/b*	
None	9 **A;B**	3 **A;b**	3 **a;B**	1 **a;b**	9:3:3:1
"Recessive" epistasis *a/a* influences *B* and *b* alleles	9 **A;B**	3 **A;b**	4 **a**		9:3:4
"Dominant" epistasis *A* influences *B* and *b* alleles	12 **A**		3 **a;B**	1 **a;b**	12:3:1
Duplicate Genes *Dominant alleles either gene A or B creates phenotype* **C**, otherwise **c**	15 **C**			1 **c**	15:1
Complementary Genes *a/a* and *b/b* are identical (phenotype **c**) but distinct from *A&B which gives phenotype* **C**	9 **C**	7 **c**			9:7
Recessive Suppression *a/a* influences *b/b*	9 **B**	3 **A;b**	4 **B**		13:3
Dominant Suppression *A* disables expression of *b/b*	15 **B**			1 **a;b**	15:1

D Example of Multiple Genes Affecting One Character (Polygenic Inheritance)

D.1 Continuous Variation

Most of the phenotypic traits commonly used in introductory genetics are qualitative, meaning that the phenotype exists in only two (or possibly a few more) discrete, alternative forms, such as either purple or white flowers, or red or white eyes. These qualitative traits are therefore said to exhibit **discrete variation**. On the other hand, many interesting and important traits exhibit **continuous variation**; these exhibit a continuous range of phenotypes that are usually measured quantitatively, such as intelligence, body mass, blood pressure in animals (including humans), and yield, water use, or vitamin content in crops. Traits with continuous variation are often complex, and do not show the simple Mendelian segregation ratios (e.g. 3:1) observed with some qualitative traits. The environment also influences many complex traits. Nevertheless, complex traits can

often be shown to have a component that is heritable, and which must therefore involve one or more genes.

How can genes, which are inherited (in the case of a diploid) as at most two variants each, explain the wide range of continuous variation observed for many traits? The lack of an immediately obvious explanation to this question was one of the early objections to Mendel's explanation of the mechanisms of heredity. However, upon further consideration, it becomes clear that the more loci that contribute to trait, the more phenotypic classes may be observed for that trait (**Figure 5-14**).

If the number of phenotypic classes is sufficiently large (as with three or more loci), individual classes may become indistinguishable from each other (particularly

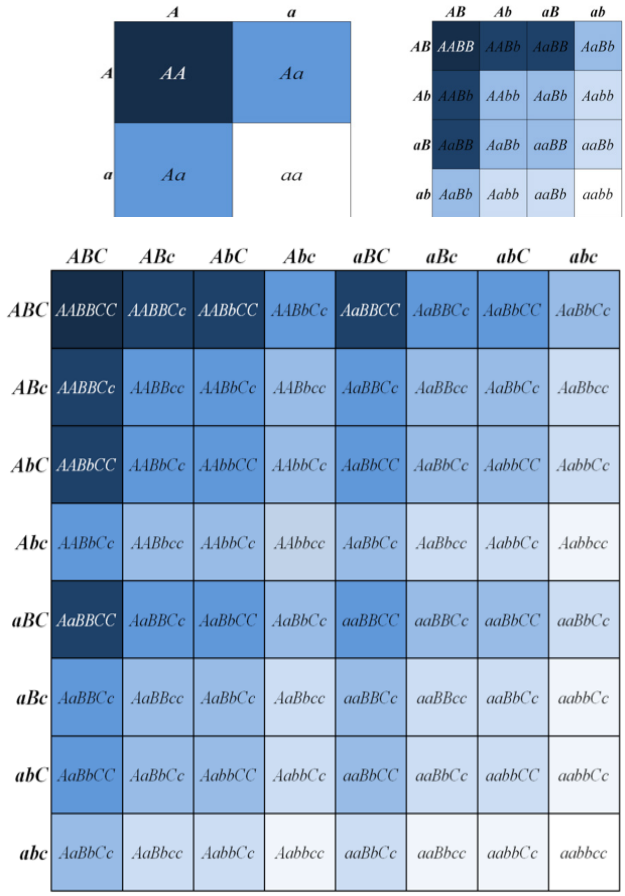


Figure 5-14

Punnett Squares for one, two, and three loci. We are using a simplified example of up to three semi-dominant genes, and in each case the effect on the phenotype is additive, meaning the more "upper case" alleles present, the stronger the phenotype. Comparison of the Punnett Squares and the associated phenotypes shows that under these conditions, the larger the number of genes that affect a trait, the more intermediate phenotypic classes that will be expected.

(Original-Deyholos-CC BY-NC 3.0)

when environmental effects are included), and the phenotype appears as a continuous variation (**Figure 5-15**). Thus, quantitative traits are sometimes called **polygenic traits**, because it is assumed that their phenotypes are controlled by the combined activity of many genes. Note that this does not imply that each of the individual genes has an equal influence on a polygenic trait – some may have major effect, while others only minor. Furthermore, any single gene may influence more than one trait, whether these traits are quantitative or qualitative traits.

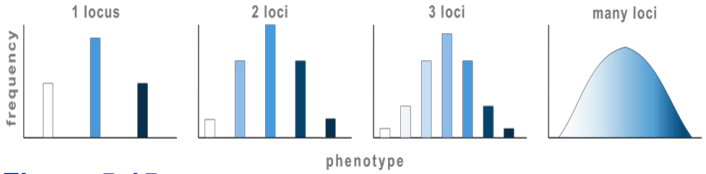


Figure 5-15

The more loci that affect a trait, the larger the number of phenotypic classes that can be expected. For some traits, the number of contributing loci is so large that the phenotypic classes blend together in apparently continuous variation.

(Original-Deyholos-CC BY-NC 3.0)

D.2 Cat Fur Genetics – (adapted from Christensen (2000) Genetics 155:999-1004)

Most aspects of the fur phenotypes of common cats can be explained by the action of just a few genes (Table 2). Other genes, not described here, may further modify these traits and account for the phenotypes seen in tabby cats and in more exotic breeds, such as Siamese.

For example, the X-linked ***Orange*** gene has two allelic forms. The O^O allele produces orange fur, while the O^B alleles produce non-orange (often black) fur. Note however, that because of X-chromosome inactivation the result is mosaicism in expression. In O^O / O^B female heterozygotes patches of black and orange are seen, which produces the tortoise shell pattern (**Figure 5-16 on page 50** A,B). This is a rare example of **codominance** since the phenotype of both alleles can be seen. Note that the cat in part A has short fur compared to the cat in part B; recessive alleles at an independent locus (L/l) produce long (*l/l*) rather than short (*L/-*) fur.

Alleles of the ***dilute*** gene affect the intensity of pigmentation, regardless of whether that pigmentation is due to black or orange pigment. Part C shows a black cat with at least one dominant allele of *dilute* (*D/-*), in contrast to the cat in D, which is grey rather than black, because it has the *d/d* genotype.

Epistasis is demonstrated by an allele of only one of

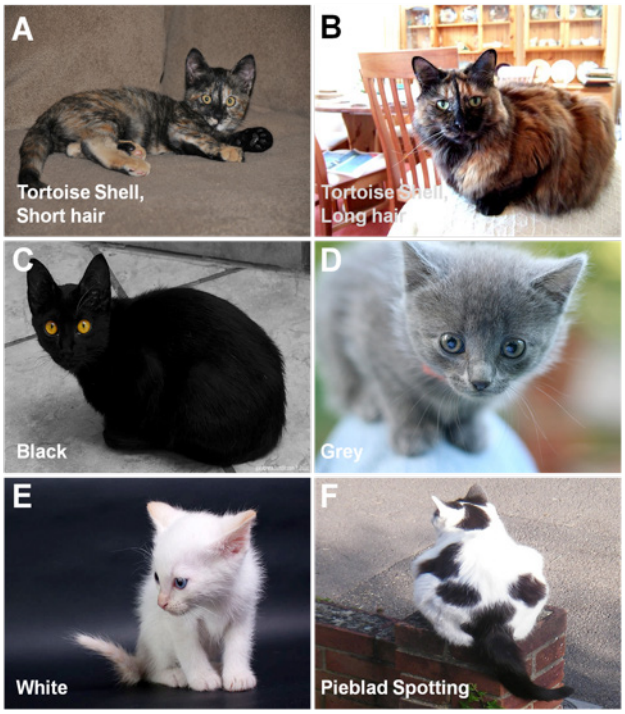


Figure 5-16

Representatives of various fur phenotypes in cats. Tortoise shell (A,B) pigmentation in cats with short (A) and long (B) fur; black (C) and grey (D) cats that differ in genotype at the dilute locus. The pure white pattern (E) is distinct from piebald spotting (F). A: (Flickr-Bill Kuffrey-CC BY 2.0), B: (Wikipedia-Dieter Simon-PD), C: (Flickr-atilave-lo-CC BY 2.0), D: (Flickr-Waldo Jaquith-CC BY-SA 2.0), E: (Wikipedia-Valerius Geng-CC BY-SA 3.0), F: (Flickr-Denni Schnapp-CC BY-NC-SA 2.0) *Changes: Letters and descriptions were added to the pictures.

Table 5-2

Summary of simplified cat fur phenotypes and genotypes.

Trait	Phenotype	Genotype	Comments
fur length	short	*L/L* or *L/l*	*L* is completely dominant
	long	*l/l*	
all white fur (non-albino)	100% white fur	*W/W* or *W/w*	If the cat has red eyes it is albino, not *W/-*. *W* is epistatic to all other fur color genes; if cat is *W/-*, can't infer genotypes for any other fur color genes.
	<100% white fur	*w/w*	
piebald spotting	> 50% white patches (but not 100%)	*S/S*	*S* is incompletely dominant and shows variable expressivity
	< 50% white patches	*S/s*	
	no white patches	*s/s*	
orange fur	all orange fur	X^O/X^O or X^O/Y	*O* is X-linked
	tortoise shell variegation	X^O/X^B	
	no orange fur (often black)	X^B/X^B or X^B/Y	
dilute pigmentation	pigmentation is intense	*D/D* or *D/d*	*D* is completely dominant
	pigmentation is dilute (e.g. gray rather than black; cream rather than orange; light brown rather than brown)	*d/d*	
tabby	tabby pattern	*A/A* or *A/a*	This is a simplification of the tabby phenotype, which involves multiple genes
	solid coloration	*a/a*	
sex	female	X/X	
	male	X/Y	

the genes in **Table 5-2**. One dominant allele of ***white masking*** (*W*) prevents normal development of melanocytes (pigment producing cells). Therefore, cats with genotype (*W/-*) will have entirely white fur regardless of the genotype at the *Orange* or *dilute* loci (part E). Although this locus produces a white colour, *W/-* is not the same as albinism, which is a much rarer phenotype caused by mutations in other genes. Albino cats can be distinguished by having red eyes, while *W/-* cats have eyes that are not red.

Piebald spotting is the occurrence of patches of white fur. These patches vary in size due to many reasons, including genotype. Homozygous cats with genotype *s/s* do not have any patches of white, while cats of genotype *S/s* and *S/S* do have patches of white, and the homozygotes tend to have a larger proportion of white fur than heterozygotes (part F). The combination of piebald spotting and tortoise shell patterning produce a **calico cat**, which has separate patches of orange, black, and white fur.

E Environmental Factors

The phenotypes described thus far have a nearly perfect correlation with their associated genotypes; in other words an individual with a particular genotype always has the expected phenotype. However, many (most?) phenotypes are not determined entirely by genotype alone. Instead, they are determined by an interaction between genotype and environmental factors and can be conceptualized in the following relationship:

Genotype + Environment

$$\Rightarrow \textbf{Phenotype (G + E} \Rightarrow \textbf{P)}$$

Or:

$$\textbf{Genotype + Environment + Interaction}_{GE}$$

$$\Rightarrow \textbf{Phenotype (G + E + I}_{GE} \Rightarrow \textbf{P)}$$

***GE = Genetics and Environment**

This interaction is especially relevant in the study of economically important phenotypes, such as human diseases or agricultural productivity. For example, a particular genotype may predispose an individual to cancer, but cancer may only develop if the individual is exposed to certain DNA-damaging chemicals or carcinogens. Therefore, not all individuals with the particular genotype will develop the cancer phenotype, only those who experience a particular environment.

F Penetrance and Expressivity

The terms penetrance and expressivity are also useful to describe the relationship between certain genotypes and their phenotypes.

F.1 Penetrance

Penetrance is the proportion of individuals with a particular genotype that display a corresponding phenotype (**Figure 18**). It is usually expressed as a percentage of the population. Because all pea plants that are homozygous for the allele for white flowers (e.g. *a/a in Chapter* 2 (page 17)) actually have white flowers, this genotype is completely (100%) penetrant. In contrast, many human genetic diseases are incompletely penetrant, since not all individuals with the disease genotype actually develop symptoms associated with the disease (less than 100%).

F.2 Expressivity

Expressivity describes the variability in mutant phenotypes observed in individuals with a particular phenotype (**Figure 5-17** and **Figure 5-18**). Many human genetic diseases provide examples of broad expressivity, since individuals with the same genotypes may vary greatly in the severity of their symptoms. Incomplete penetrance and broad expressivity are due to random chance, non-genetic (environmental), and genetic factors (mutations in other genes).

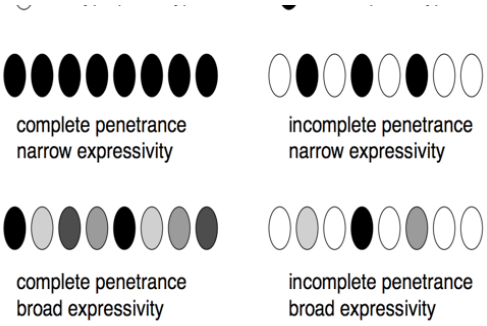


Figure 5-17

Relationship between penetrance and expressivity in eight individuals that all have a mutant genotype. Penetrance can be complete (all eight have the mutant phenotype) or incomplete (only some have the mutant phenotype). Amongst those individuals with the mutant phenotype the expressivity can be narrow (very little variation) to broad (lots of variation).

(Original-Locke-CC BY-NC 3.0)

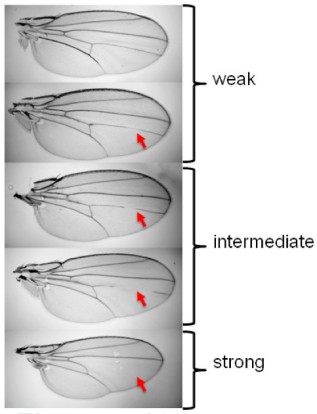


Figure 5-18

Mutations in wings of *Drosophila melanogaster* showing weak to strong expressivity.

(Original-J. Locke-CC; AN)

G Mendelian Phenotypic Ratios May Not Be As Expected

G.1 Other factors

There are other factors that that affects organism's phenotype and thus appear to alter Mendelian inheritance.

(1) Genetic heterogeneity: There is more than one gene or genetic mechanism that can produce the same phenotype.

(2) Polygenic determination: One phenotypic trait is controlled by multiple genes.

(3) Phenocopy: Organisms that do not have the genotype for trait A can also express trait A due to environmental conditions; they do not have the same genotype but the environment simply "copies" the genetic phenotype.

(4) Incomplete penetrance: even though an organism possesses the genotype for trait A, it might not be expressed with 100% effect.

(5) Certain genotypes show a survival rate that is less than 100%. For example, genotypes that cause death, **recessive lethal mutations**, at the embryo or larval stage will be under-represented when adult flies are counted.

G.2 The X^2 test for goodness-of-fit

For a variety of reasons, the phenotypic ratios observed from real crosses rarely match the exact ratios expected based on a Punnett Square or other prediction techniques. There are many possible explanations for deviations from expected ratios. Sometimes these deviations are due to **sampling effects**, in other words, the random selection of a non-representative subset of individuals for observation.

A statistical procedure called the **chi-square** (χ^2) test can be used to help a geneticist decide whether the deviation between observed and expected ratios is due to sampling effects, or whether the difference is so large that some other explanation must be sought by re-examining the assumptions used to calculate the expected ratio. The procedure for performing a chi-square test is shown at http://tinyurl.com/chi2 .

Summary:

- Phenotype depends on the alleles that are present, their dominance relationships, and sometimes also interactions with the environment and other factors.
- The alleles of different loci are inherited independently of each other, unless they are genetically linked.
- Many important traits show continuous, rather than discrete variation. These are called quantitative traits.
- Many quantitative traits are influenced by a combination of environment and genetics.
- The expected phenotypic ratio of a dihybrid cross is 9:3:3:1, except in cases of linkage or gene interactions that modify this ratio.
- Modified ratios from 9:3:3:1 are seen in the case of recessive and dominant epistasis, duplicate genes, and complementary gene action. This usually indicates that the two genes interact within the same biological pathway.
- There are other factors that alter the expected Mendelian ratios.

Key terms:

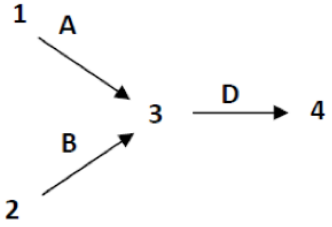
calico
continuous variation
dihybrid
dilute
discrete variation
duplicate gene action
expressivity
independent assortment
linkage
masking
Mendel's Second Law
modified Mendelian Ratios
penetrance
piebald spotting
polygenic traits
recessive epistasis
recessive lethal mutations
redundancy

Study Questions:

1. In the table to the right, match the mouse hair color phenotypes with the term from the list that best explains the observed phenotype, given the genotypes shown. In this case, the allele symbols do not imply anything about the dominance relationships between the alleles. List of terms: haplosufficiency, haplo-insufficiency, pleiotropy, incomplete dominance, co-dominance, incomplete penetrance, broad (variable) expressivity.

	A_1A_1	A_1A_2	A_2A_2
1	all hairs black	on the same individual: 50% of hairs are all black and 50% of hairs are all white	all hairs white
2	all hairs black	all hairs are the same shade of grey	all hairs white
3	all hairs black	all hairs black	50% of individuals have all white hairs and 50% of individuals have all black hairs
4	all hairs black	all hairs black	mice have no hair
5	all hairs black	all hairs white	all hairs white
6	all hairs black	all hairs black	all hairs white
7	all hairs black	all hairs black	hairs are a wide range of shades of grey

2. Answer questions 2-4 using the following biochemical pathway for fruit color. Assume all mutations (lower case allele symbols) are recessive, and that *either* precursor 1 or precursor 2 can be used to produce precursor 3. If the alleles for a particular gene are not listed in a genotype, you can assume that they are wild-type.



3. If 1 and 2 and 3 are all colorless, and 4 is red, what will be the phenotypes associated with the following genotypes?

a) *a/a*
b) *b/b*
c) *d/d*
d) *a/a;b/b*
e) *a/a;d/d*
f) *b/b;d/d*
g) *a/a;b/b;d/d*
h) What will be the phenotypic ratios among the offspring of a cross *A/a;B/b* × *A/a;B/b*?
i) What will be the phenotypic ratios among the offspring of a cross *B/b;D/d* × *B/b;D/d*?
j) What will be the phenotypic ratios among the offspring of a cross *A/a;D/d* × *A/a;D/d*?

4. If 1 and 2 are both colorless, and 3 is blue and 4 is red, what will be the phenotypes associated with the following genotypes?

   a) a/a
   b) b/b
   c) d/d
   d) a/a;b/b
   e) a/a;d/d
   f) b/b;d/d
   g) a/a;b/b;d/d
   h) What will be the phenotypic ratios among the offspring of a cross *A/a;B/b* × *A/a;B/b*?
   i) What will be the phenotypic ratios among the offspring of a cross *B/b;D/d* × *B/b;D/d*?
   j) What will be the phenotypic ratios among the offspring of a cross *A/a;D/d* × *A/a;D/d*?

5. If 1 is colorless, 2 is yellow and 3 is blue and 4 is red, what will be the phenotypes associated with the following genotypes?

   a) *a/a*
   b) b/b
   c) d/d
   d) a/a;b/b
   e) a/a;d/d
   f) b/b;d/d
   g) a/a;b/b;d/d
   h) What will be the phenotypic ratios among the offspring of a cross A/a;B/b × A/a;B/b?
   i) What will be the phenotypic ratios among the offspring of a cross B/b;D/d × B/b;D/d?
   j) What will be the phenotypic ratios among the offspring of a cross A/a;D/d × A/a;D/d?

6. Which of the situations in questions 2 – 4 demonstrate epistasis?

7. If the genotypes written within the Punnett Square are from the F_2 generation, what would be the phenotypes and genotypes of the F_1 and P generations for:

   a) Figure 5-4
   b) Figure 5-6
   c) Figure 5-8
   d) Figure 5-10

8. To better understand how genes control the development of three-dimensional structures, you conducted a mutant screen in *Arabidopsis* and identified a recessive point mutation allele of a single gene (*g*) that causes leaves to develop as narrow tubes rather than the broad flat surfaces that develop in wild-type (*G*). Allele *g* causes a complete loss of function. Now you want to identify more genes involved in the same process. Diagram a process you could use to identify other genes that interact with gene *g*. Show all of the possible genotypes that could arise in the F_1 generation.

9. With reference to question 7, if the recessive allele, *g* is mutated again to make allele *g**, what are the possible phenotypes of a homozygous *g* g** individual?

10. Again, in reference to question 8, what are the possible phenotypes of a homozygous *a/a;g/g* individual, where *a* is a recessive allele of a second gene? In each case, also specify the phenotypic ratios that would be observed among the F_1 progeny of a cross of *A/a;G/g* x *A/a;G/g*

11. Use the product rule to calculate the phenotypic ratios expected from a trihybrid cross. Assume independent assortment and no epistasis/gene interactions.

Chapter 6
Pedigree Analysis

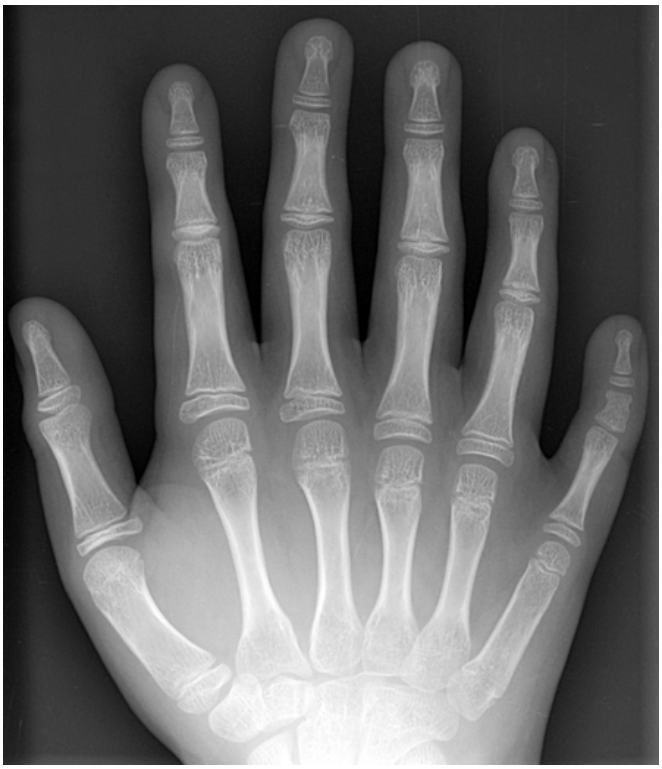


Figure 6-1
Polydactyly (six fingers in this case) is an example of a human trait that can be studied by pedigree analysis. (Wikipedia- Drg-nu23- CC BY-SA 3.0)

Introduction

The basic concepts of genetics described in the preceding chapters can be applied to almost any eukaryotic organism. However, some techniques, such as **testcrosses**, can only be performed with model organisms or other species that can be experimentally manipulated. To study the inheritance patterns of genes in humans and other species for which controlled matings are not possible, geneticists use the analysis of pedigrees and populations.

A. Pedigree Analysis

A.1. Pedigree charts

Pedigree charts are diagrams that show the phenotypes and/or genotypes for a particular organism, its ancestors, and descendants. While commonly used in human families to track genetic diseases, they can be used for any species and any inherited trait. Geneticists use a standardized set of symbols to represent an individual's sex, family relationships and phenotype. These diagrams are used to determine the **mode of inheritance** of a particular disease or trait, and to predict the probability of its appearance among offspring. Pedigree analysis is therefore an important tool in basic research, agriculture, and **genetic counseling**.

Each pedigree chart represents all of the available information about the inheritance of a single trait (most often a disease) within a family. The pedigree chart is therefore drawn using factual information, but there is always some possibility of errors in this information, especially when relying on family members' recollections or even clinical diagnoses. In real pedigrees, further complications can arise due to **incomplete penetrance**

(including age of onset) and **variable expressivity** of disease alleles, but for the examples presented in this book, we will presume complete accuracy of the pedigrees – that is, the phenotype accurately reflects the genotype. A pedigree may be drawn when trying to determine the nature of a newly discovered disease, or when an individual with a family history of a disease wants to know the probability of passing the disease on to their children. In either case, a tree is drawn, as shown in **Figure 6-2**, with circles to represent females, and squares to represent males. Matings are drawn as a line joining a male and female, while a consanguineous mating (closely related, such as siblings or first cousins) is two lines.

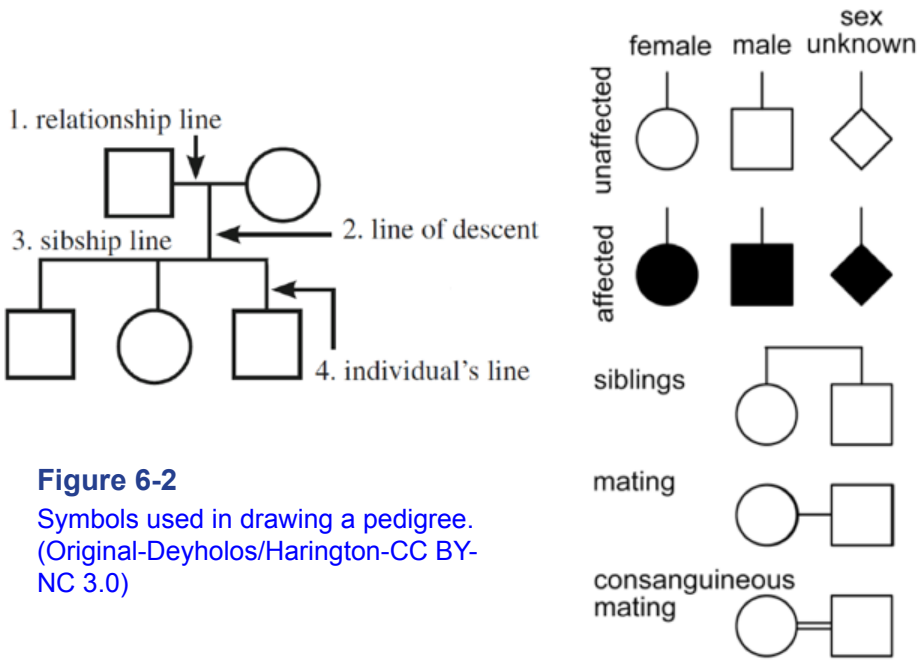


Figure 6-2
Symbols used in drawing a pedigree. (Original-Deyholos/Harington-CC BY-NC 3.0)

The affected individual that brings the family to the attention of a geneticist is called the **proband** (or propositus). If the individual is unaffected, they are called the consultand. If an individual is known to have symptoms of the disease (**affected**), the symbol is filled in. Sometimes a half-filled in symbol is used to indicate a known **carrier** of a disease; this is someone who does not have any symptoms of the disease, but who passed the disease on to subsequent generations because they are a heterozygote. A circle with a dot in the centre indicates female carriers of X-linked traits. Note that when a pedigree is constructed, we often don't know whether a particular individual is a carrier or not. Not all carriers are always explicitly indicated in a pedigree. If you don't know the exact genotype, don't guess. If someone can be a homozygote or heterozygote, indicate the second allele a slash (*e.g. A/-*). For simplicity, in this course we will assume that the pedigrees presented are accurate, and represent fully penetrant traits.

A.2. Pedigree Chart Convention Symbols

In pedigree analysis, standardized human pedigree nomenclature is used. If possible, the male partner should be left of female partner on relationship line. Siblings should be listed from left to right in birth order, oldest to youngest.

B. Modes of Inheritance

Given a pedigree of an uncharacterized disease or trait, one of the first tasks is to determine which modes of inheritance are possible and then which mode of inheritance is most likely. This information is essential in calculating the probability that the trait will be inherited in any future offspring. We will mostly consider five major types of inheritance: autosomal dominant (AD), autosomal recessive (AR), X-linked dominant (XD), X-linked recessive (XR), and Y-linked (Y).

Warning: do not use ratios to determine which allele is dominant or recessive. There are not enough data points in a pedigree to validate the assumptions upon which the ratios are based. Try out gene symbols and find out which are consistent and which are inconsistent. Most small pedigrees are consistent with more than one inheritance pattern.

B.1. Autosomal Dominant (AD)

When a disease is caused by a dominant allele of a gene, every person with that allele will show symptoms of the disease (assuming complete penetrance), and only one disease allele needs to be inherited for an individual to be affected (**Figure 6-3**). Thus, every affected individual must have an affected parent. A pedigree with affected individuals in every generation is typical of **AD** (*autosomal dominant*) diseases. However, beware that other modes of inheritance can also show the disease in every generation, as described below. It is also possible for an affected individual with an AD disease

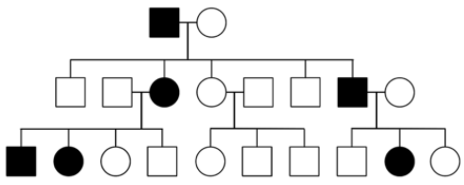


Figure 6-3
A pedigree consistent with AD inheritance.
(Original-Deyholos-CC BY-NC 3.0)

Table 6-1
Genotype nomenclature consistent with AD inheritance.

(Original-Nickle-CC BY-NC 3.0)

AD Nomenclature	
A	mutant allele
a	WT allele
A / a	affected
A / A	affected to the same or greater extent (may be recessive lethal)

to have a family without any affected children, if the affected parent is a heterozygote. This is particularly true in small families, where the probability of every child inheriting the normal, rather than disease allele is not extremely small. Note that AD diseases are usually rare in populations, therefore affected individuals with AD diseases tend to be heterozygotes (otherwise, both parents would have had to been affected with the same rare disease). Achondroplastic dwarfism, and polydactyly are both examples of human conditions that may follow an AD mode of inheritance. **Table 6-1** shows how symbols can be used for AD expression.

AD EXAMPLE: ACHONDROPLASIA

Achondroplasia (**Figure 6-4**) is a common form of dwarfism. *FGFR3* gene at 4p16 (chromosome 4, p arm, region 1, band 6) encodes a receptor protein that negatively regulates bone development. A specific bp substitution in the gene makes an over-active protein and this results in shortened bones (**Figure 6-5**). Acondroplasia is considered autosomal dominant because the defective proteins made in *A/a* embryos halt bone growth prematurely. *A/A* embryos do not make enough limb bones to survive. Most, but not all dominant mutations are also recessive lethal. In achondroplasia, the *A* allele shows dominant visible phenotype (shortness) and recessive lethal phenotype.

B.2. X-LINKED DOMINANT (XD)

In X-linked dominant inheritance, the gene responsible for the disease is located on the X-chromosome, and the allele that causes the disease is dominant to the normal allele in females. Because females have twice as many X-chromosomes as males, females tend to be more frequently affected than males in the population. However, not all pedigrees provide sufficient information to distinguish XD and AD. One definitive indication that a trait is inherited as AD, and not XD, is that an affected father passes the disease to a son; this type of transmission is not possible with XD, since males inherit their X chromosome from their mothers.

Figure 6-4
Portrait of Sebastián de Morra by Diego Velázquez, a court dwarf and was painted ~1645. He likely had achondroplasia, a condition that has autosomal dominant inheritance.

(Wikimedia Commons-Diego Velázquez-PD)

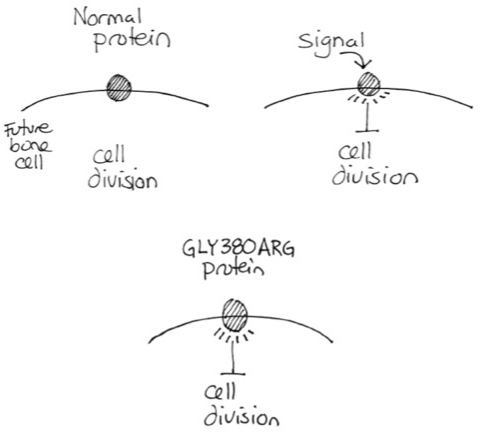


Figure 6-5
Diagram showing the mechanism of achondroplasia.

(Original-Harrington-CC BY-NC 3.0)

Table 6-2 shows how gene symbols work with XD; **Figure 6-6** and **Figure 6-7** demonstrate XD syndromes.

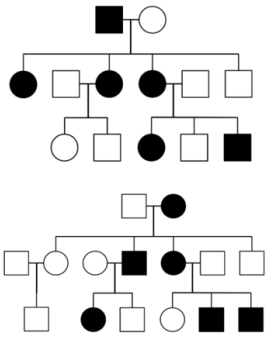


Figure 6-6
Two pedigrees consistent with XD inheritance.
(Original-Deyholos-CC BY-NC 3.0)

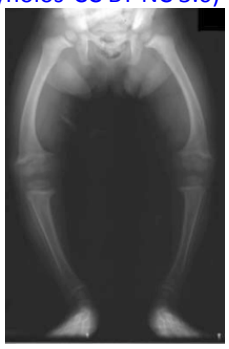


Figure 6-7
Some types of rickets may follow an XD mode of inheritance.
(Wikipedia-Mrish-CC BY-SA 1.0)

XD EXAMPLE: FRAGILE X SYNDROME

The *FMR1* gene at Xq21 (X chromosome, q arm, region 2, band 1) encodes a protein needed for neuron development. There is a $(CGG)_n$ repeat array in the 5'UTR (untranslated region). If there is expansion of the repeat in the germline cell the child will inherit a non-functional allele. X^A / Y males have fragile X mental retardation

Table 6-2
Genotype nomenclature consistent with XD inheritance.

(Original-Nickle-CC BY-NC 3.0)

XD Nomenclature	
X^A	mutant allele
X^a	WT allele
X^A/ Y	affected male
X^A / X^A	affected females
X^A / X^a	

(IQ < 50) because none of their neurons can make FMR1 proteins. Fragile X syndrome is considered X-linked dominant because only some neurons in X^A / X^a females can make FMR1 proteins. The severity (IQ 50 – 70) in these females depends upon the number and location of these cells within in the brain.

B.3. Autosomal recessive (AR)

Diseases that are inherited in an autosomal recessive pattern require that both parents of an affected individual carry at least one copy of the disease allele. With AR traits, many individuals in a pedigree can be carriers, probably without knowing it. Compared to pedigrees of dominant traits, AR pedigrees tend to show fewer affected individuals and are more likely than AD or XD to "skip a generation". Thus, the major feature that distinguishes AR from AD or XD is that unaffected individuals can have affected offspring. Attached earlobes is a human condition that may follow an AR mode of inheritance.

Table 6-3 shows AR gene symbols and **Figure 6-8** shows an AR pedigree.

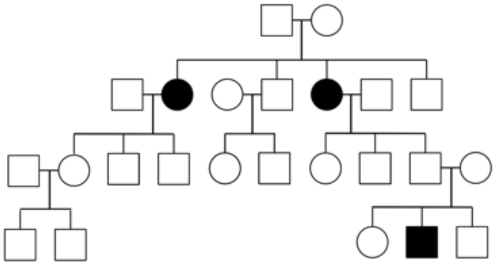


Figure 6-8
A pedigree consistent with AR inheritance.
(Original-Deyholos-CC BY-NC 3.0)

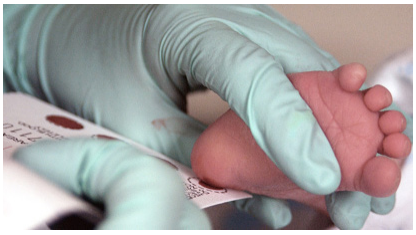
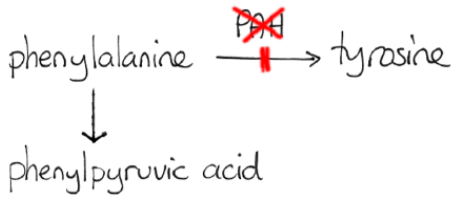


Figure 6-9
Many inborn errors of metabolism, such as phenylketonuria (PKU) are inherited as AR. Newborns are often tested for a few of the most common metabolic diseases.

(Wikipedia-U.S. Air Force photo/Staff Sgt. Eric T. Sheler-PD)

Figure 6-10
Mutation in the *PAH* gene cannot catalyze the breakdown of phenylalanine into tyrosine. This causes build up of phenylpyruvic acid, which would damage the central nervous system.



(Original-Harrington-CC BY-NC 3.0)

Table 6-3
Genotype nomenclature consistent with AR inheritance.
(Original-Nickle-CC BY-NC 3.0)

AR Nomenclature	
A	WT allele
a	mutant allele
a / a	affected
A / a	carrier
A / -	status unknown

AR example: phenylketonuria (PKU)

Individuals with phenylketonuria (PKU) have a mutation in the *PAH* gene at 12q24 (chromosome 12, q arm, region 2, band 4), which encodes an enzyme that breaks down phenylalanine into tyrosine called phenylalanine hydrolase (PAH). Without PAH, the accumulation of phenylalanine and other metabolites, such as phenylpyruvic acid (**Figure 6-10**), disrupts brain development, typically within a year after birth, and can lead to intellectual disability. Fortunately, this condition is both easy to diagnose (**Figure 6-9**) and can be successfully treated with a low phenylalanine diet. There are over 450 different mutant alleles of the PAH gene, so most people with PKU are **compound heterozygotes**. Compound heterozygotes have two different mutant alleles (different base pair changes) at a given locus, in this case the *PAH* gene.

B.4. X-linked recessive (XR)

Because males have only one X-chromosome, any male that inherits an X-linked recessive disease allele will be affected by it (assuming complete penetrance). Therefore, in XR modes of inheritance, males tend to be affected more frequently than females in a population (*e.g.* **Figure 6-11, 12**). This is in contrast to AR and AD, where both sexes tend to be affected equally, and XD, in which females are affected more frequently. Note, however, in the small sample sizes typical of human families, it is usually not possible to accurately determine whether one sex is affected more frequently than others.

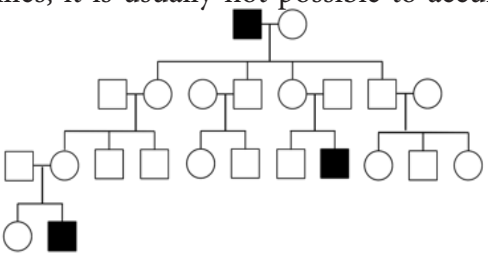


Figure 6-11
A pedigree consistent with XR inheritance.
(Original-Deyholos-CC BY-NC 3.0)

On the other hand, pedigree that can be used to definitively es-

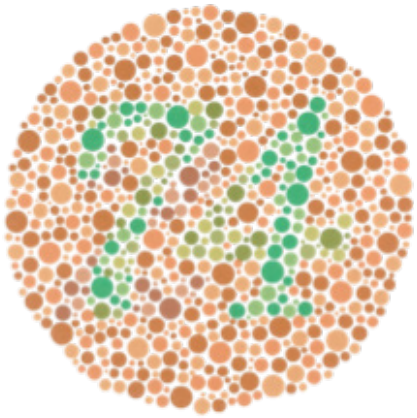


Figure 6-12
Some forms of colour blindness are inherited as XR-traits. Colour blindness is diagnosed using tests such as this Ishihara Test, which is shown above.

(Wikipedia-unknown-PD)

tablish that an inheritance pattern is not XR is the presence of an affected daughter from unaffected parents; because she would have had to inherit one X-chromosome from her father, he would also have been affected in XR. **Table 6-4** show XR symbols.

Table 6-4
Genotype nomenclature consistent with XR inheritance. (Original-Harrington-CC BY-NC 3.0)

XR Nomenclature	
X^A	WT allele
X^a	mutant allele
X^a / Y	affected
X^A / X^a	carrier female
X^a / X^a	affected female

XR EXAMPLE: HEMOPHILIA A

F8 gene at Xq28 (X chromosome, q arm, region 2, band 8) encodes blood clotting factor VIIIc (**Figure 6-13**). Without Factor VIIIc, internal and external bleeding can't be stopped. Back in the 1900s, X^a / Y male's average life expectancy was 1.4 years, but in the 2000s it has increased to 65 years with the advent of Recombinant Human Factor VIIIc. Hemophilia A is recessive because X^A / X^a females have normal blood coagulation, while X^a / X^a females have hemophilia.

Figure 6-13
F8 gene encodes for blood clotting factor VIIIC which is responsible for blood coagulation.

(Original-Harrington-CC BY-NC 3.0)

B.5. Y-LINKED

Only males are affected in human Y-linked inheritance (and other species with the X/Y sex determining system). There is only father-to-son transmission. This is the easiest mode of inheritance to identify, but it is one of the rarest because there are so few genes located only on the Y-chromosome.

An example of Y-linked inheritance is the **hairy-ear-rim** phenotype seen in some Indian families. As expected this trait is passed on from father to all sons and no daughters. Y-chromosome DNA polymorphisms can be used to follow the male lineage in large families or through ancient ancestral lineages. For example, the Y-chromosome of Mongolian ruler Genghis Khan (1162-1227 CE), and his male relatives, accounts for ~8% of the Y-chromosome lineage of men in Asia (0.5% world wide).

B.6. ORGANELLAR GENOMES

In eukaryotes, DNA and genes also exist outside of the chromosomes found in the nucleus. Both the **chloroplast** and **mitochondrion** have circular chromosomes .These organellar genomes are often present in multiple copies within each organelle. In most sexually reproducing species, **organellar chromosomes** are inherited from only one parent, usually the one that produces the largest gamete. Thus, in mammals, angiosperms, and many other organisms, mitochondria and chloroplasts are inherited only through the mother (maternally). Therefore, mutations in mitochondrial DNA (mtDNA) are inherited through the maternal line. There are some human diseases associated with mutations in mitochondria genes. These mutations can affect both males and females, but males cannot pass them on as all mitochondria are inherited via the egg, not the sperm. Mitochondrial DNA polymorphisms are also used to investigate evolutionary and historical lineages, both ancient and recent. Because of the relative similarity of sequence mtDNA is also used in species identification in ecology studies. An example of Mitochondrial inheritance is the Leber hereditary optical neuropathy (LHON). Mitochondria are very important in retinal cells for ATP and/or a specialized function. Mutations in several mtDNA genes result in blindness during early childhood.

C. SPORADIC AND NON-HERITABLE DISEASES

Not all of the characterized human traits and diseases are attributed to mutant alleles at a single gene locus. Many diseases that have a heritable component, have more complex inheritance patterns due to (1) the involvement of multiple genes, and/or (2) environmental factors.

On the other hand, some non-genetic diseases may appear to be heritable because they affect multiple

members of the same family, but this is due to the family members being exposed to the same toxins or other environmental factors (e.g. in their homes).

Finally, diseases with similar symptoms may have different causes, some of which may be genetic while others are not. One example of this is ALS (amyotrophic lateral sclerosis); approximately 5-10% of cases are inherited in an AD pattern, while the majority of the remaining cases appear to be **sporadic**, in other words, not caused by a mutation inherited from a parent. We now know that different genes or proteins are affected in the inherited and sporadic forms of ALS. The physicist Stephen Hawking (**Figure 6-14**)and baseball player Lou Gehrig both suffered from sporadic ALS.

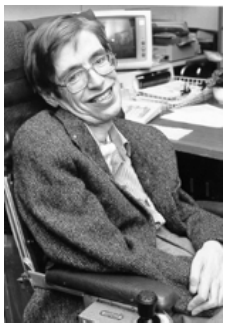


Figure 6-14
Stephen Hawking
(Wikipedia-NASA-PD)

D. Calculating Probabilities

Once the mode of inheritance of a disease or trait is identified, some inferences about the genotype of individuals in a pedigree can be made, based on their phenotypes and where they appear in the family tree. Given these genotypes, it is possible to calculate the probability of a particular genotype being inherited in subsequent generations. This can be useful in genetic counseling, for example when prospective parents wish to know the likelihood of their offspring inheriting a disease for which they have a family history.

Probabilities in pedigrees are calculated using knowledge of Mendelian inheritance and the same basic methods as are used in other fields. The first formula is the **product rule**: the joint probability of two independent events is the product of their individual probabilities; this is the probability of one event AND another event occurring. For example, the probability of a rolling a "five" with a single throw of a single six-sided die is 1/6, and the probability of rolling "five" in each of three successive rolls is 1/6 x 1/6 x 1/6 = 1/216.

The second useful formula is the **sum rule**, which states that the combined probability of two independent events is the sum of their individual probabilities. This is the probability of one event OR another event occurring. For example, the probability of rolling a five or six in a single throw of a dice is 1/6 + 1/6 = 1/3.

With these rules in mind, we can calculate the probability that two carriers (i.e. heterozygotes) of an AR disease will have a child affected with the disease as ½ x ½ = ¼, since for each parent, the probability of any gametes carrying the disease allele is ½. This is consistent with what we already know from calculating probabilities using a Punnett Square (e.g. in a monohybrid cross *A/a* x *A/a*, ¼ of the offspring are *a/a*).

We can likewise calculate probabilities in the more complex pedigree shown in We can likewise calculate probabilities in the more complex pedigree shown in We can likewise calculate probabilities in the more complex pedigree shown in We can likewise calculate probabilities in the more complex pedigree shown in We can likewise calculate probabilities in the more complex pedigree shown in **Figure 6-15**.

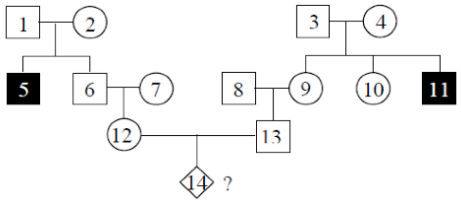


Figure 6-15
Individuals in this pedigree are labeled with numbers to make discussion easier.
(Original-Deyholos-CC BY-NC 3.0)

Assuming the disease has an AR pattern of inheritance, what is the probability that individual 14 will be affected? We can assume that individuals #1, #2, #3 and #4 are heterozygotes (*A/a*), because they each had at least one affected (*a/a*) child, but they are not affected themselves. This means that there is a 2/3 chance that individual #6 is also *A/a*. This is because according to Mendelian inheritance, when two heterozygotes mate, there is a 1:2:1 distribution of genotypes *A/A*:*A/a*:*a/a*. However, because #6 is unaffected, he can't be *a/a*, so he is either *A/a* or *A/A*, but the probability of him being *A/a* is twice as likely as *A/A*. By the same reasoning, there is likewise a 2/3 chance that #9 is a heterozygous carrier of the disease allele.

If individual 6 is a heterozygous for the disease allele, then there is a ½ chance that #12 will also be a heterozygote (i.e. if the mating of #6 and #7 is *A/a* × *A/A*, half of the progeny will be *A/a*; we are also assuming that #7, who is unrelated, does not carry any disease alleles). Therefore, the combined probability that #12 is also a heterozygote is 2/3 x 1/2 = 1/3. This reasoning also applies to individual #13, i.e. there is a 1/3 probability that he is a heterozygote for the disease. Thus, the overall probability that both individual #12 and #13 are heterozygous, and that a particular offspring of theirs will be homozygous for the disease alleles is 1/3 x 1/3 x 1/4 = 1/36.

Summary:

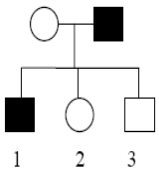
- Pedigree analysis can be used to determine the mode of inheritance of specific traits such as diseases.
- Loci can be X- or Y-linked or autosomal in location and alleles either dominant or recessive with respect to wild type.
- If the mode of inheritance is known, a pedigree can be used to calculate the probability of inheritance of a particular genotype by an individual.

Key terms:

Pedigree charts
mode of inheritance
genetic counseling
incomplete penetrance
variable expressivity
proband
affected
carrier
autosomal dominant
Achondoplasia
X-linked dominant
Fragile X-syndrome
Phenylketonuria (PKU)
autosomal recessive
X-linked recessive
Hemophilia A
Y-linked
hairy-ear-rim
chloroplast
mitochondrion
organellar
mitochondrial inheritance (mtDNA)
endopolyplody
sporadic
product rule
sum rule

Study Questions:

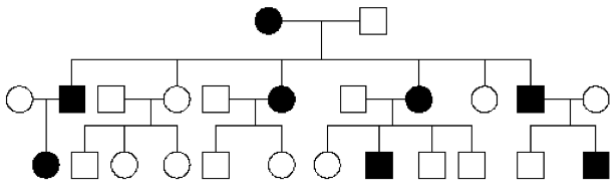
1) What are some of the modes of inheritance that are consistent with this pedigree?



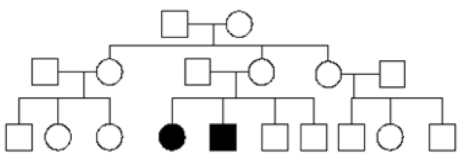
2) In this pedigree in question 1, the mode of inheritance cannot be determined unamibguously. What are some examples of data (e.g. from other generations) that, if added to the pedigree would help determine the mode of inheritance?

3) For each of the following pedigrees, name the most likely mode of inheritance (AR=autosomal recessive, AD=autosomal dominant, XR=X-linked recessive, XD=X-linked dominant). (These pedigrees were obtained from various external sources).

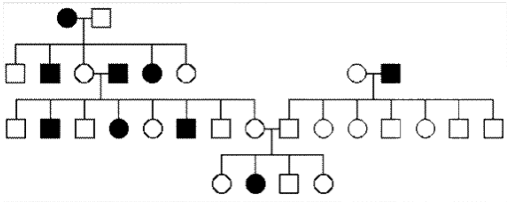
a)



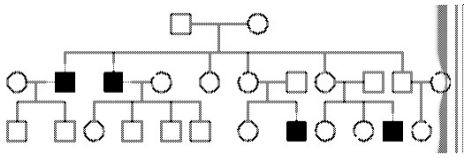
b)



c)

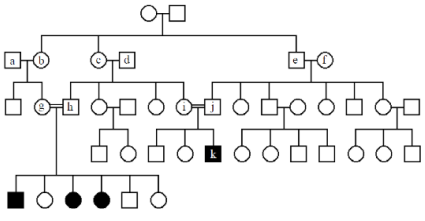


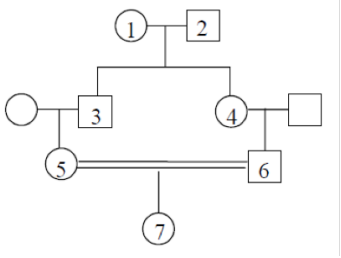
d)



4) The following pedigree represents a rare, autosomal recessive disease. What are the genotypes of the individuals who are indicated by letters?

5) If individual #1 in the following pedigree is a heterozygote for a rare, AR disease, what is the probability that individual #7 will be affected by the disease? Assume that #2 and the spouses of #3 and #4 are not carriers.





Chapter 7
Cell Cycles

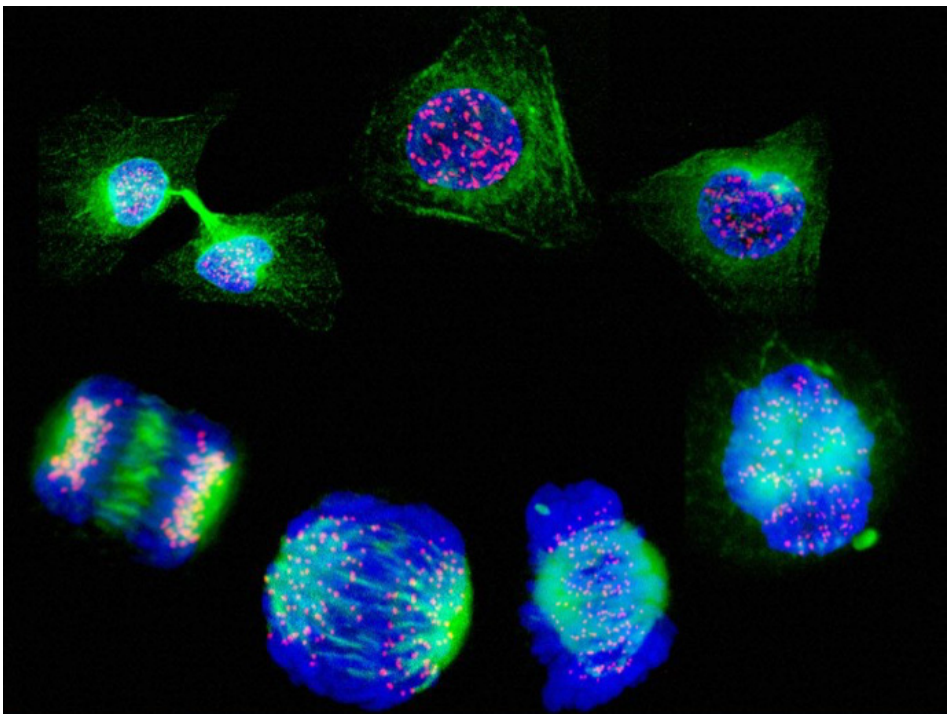


Figure 7-1
Confocal micrograph of human cells showing the stages of cell division. DNA is stained blue, microtubules stained green and kinetochores stained pink. Starting from the top and going clockwise you see an interphase cell with DNA in the nucleus. In the next cell, the nucleus dissolves and chromosomes condense in prophase. The next is prometaphase where microtubules are starting to attach, but the chromosomes haven't aligned. Next is metaphase where the chromosomes are all attached to microtubules and aligned on the metaphase plate. The next two are early and late anaphase, as the chromosomes start separating to their respective poles. Finally there is telophase where the cells are completing division to be two daughter cells. (Flickr-M. Daniels; Wellcome Images-CC BY-NC-ND 2.0)

INTRODUCTION

Cell growth and division is essential to asexual reproduction and the development of multicellular organisms. The transmission of genetic information is accomplished in a cellular process called mitosis. This process ensures that at cell division, each daughter cell inherits identical genetic material, i.e. exactly one copy of each chromosome present in the parental cell. An evolutionary adaptation of mitosis lead to a special type of cell division that reduces the number of chromosomes from diploid to haploid: this is meiosis, and is an essential step in sexual reproduction to avoid doubling the number of chromosomes each time progeny are generated through fertilization.

A BASIC STAGES OF A TYPICAL CELL CYCLE

The life cycle of eukaryotic cells can generally be divided into four stages (and a typical cell cycle is shown in Figure 7-2). When a cell is produced through fertilization or cell division it normally goes through four main stages: G1, S, G2, and M. The first stage of interphase is a lag period is called Gap 1 (G1), and is the first part of interphase. This is where the cell does its normal cellular functions and it grows in size, particularly after mitosis when the daughters are half the size of the mother cell. This stage ends with the onset of the DNA synthesis (S) phase, during which each chromosome is replicated (For more information on DNA replication, see the chapter on DNA and chromosome replication.) Though the chromosomes aren't condensed yet, because S phase is still part of interphase, they are replicated as two sister chromatids attached at the centromere. Still in interphase and following replication, there is another lag phase, called Gap 2 (G2). In G2, the cell is continues to grow and building up the

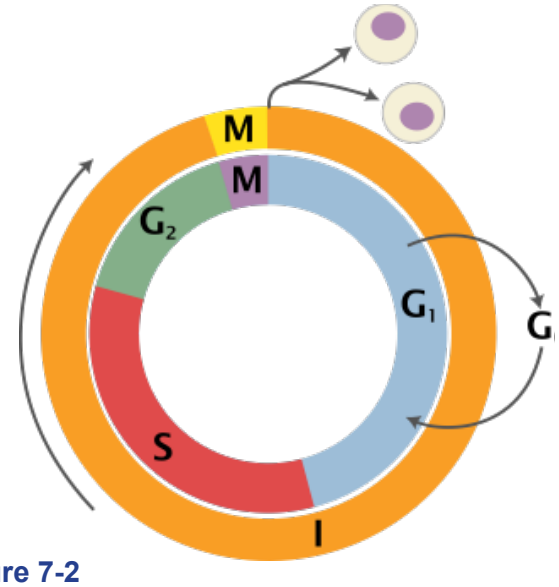


Figure 7-2

Stages of the cell cycle. The outer ring identifies when a cell is in interphase (I) and when it is in mitosis (M). The inner ring identifies the four major stages. Cells can enter G0 if they are not actively undergoing cell division, and may re-enter the cell cycle at a later time.

required proteins necessary for cell division to occur, and continuing to grow. This is a checkpoint stage, where if there are any problems with replication or acquiring the needed proteins for cell division doesn't occur, the cell cycle will arrest until it can fix itself or choose to die. The final stage is mitosis (M), where the cell undergoes cell division as is described in the last section.

Many variants of this generalized cell cycle also exist. Cells undergoing meiosis do not usually have a G2 phase. Cells like hematopoietic stem cells, that are found in the bone marrow and produce all the other blood cells will consistently go through these phases as they are constantly replicating. Others, like some cells in the nervous system will no longer divide in their lifetime. These cells never leave G1 phase, and are said to enter a permanent, non-dividing stage called Go. On the other hand some cells, like the larval tissues in *Drosophila*, undergo many rounds of DNA synthesis (S) without any mitosis or cell division, leading to endoreduplication (doubling of chromosomes within the nucleus). Understanding the control of the cell cycle is an active area of research, particularly because of the relationship between cell division and cancer.

B MITOSIS

During the S-phase of interphase the chromosomes replicate so that each chromosome has two sister chromatids attached at the centromere. After S-phase and G2, the cell enters Mitosis. The first step in mitosis is prophase where the nucleus dissolves and the replicated chromosomes condense into the visible structures we associated with chromosomes. Next is metaphase, where the microtubules attach to the kinetochore and the chromosomes align along the middle of the dividing cell, known as the metaphase plate. The kinetochore is the region in the chromosome where the microtubules attach, which contains the centromere and proteins that help the microtubules bind. Then in anaphase, each of the sister chromatids from each chromosome gets pulled towards opposite poles of the dividing cell. Finally in telophase, identical sets of unreplicated chromosomes (single chromatids) are completely separated from each other into the two daughter cells, and the nucleus re-forms around each of the two sets of chromosomes. Following this is the partitioning of the cytoplasm (cytokinesis) to complete the process and to make two identical daughter cells. Figure 7-1 on page 63 and Figure 7-3 show real pictures and a cartoon schematic of the process, respectively.

You should note that this is a dynamic and ongoing process, and cells don't just jump from one stage to the

You can practice chromosome movements of mitosis online.

http://tinyurl.com/oog-mitosis

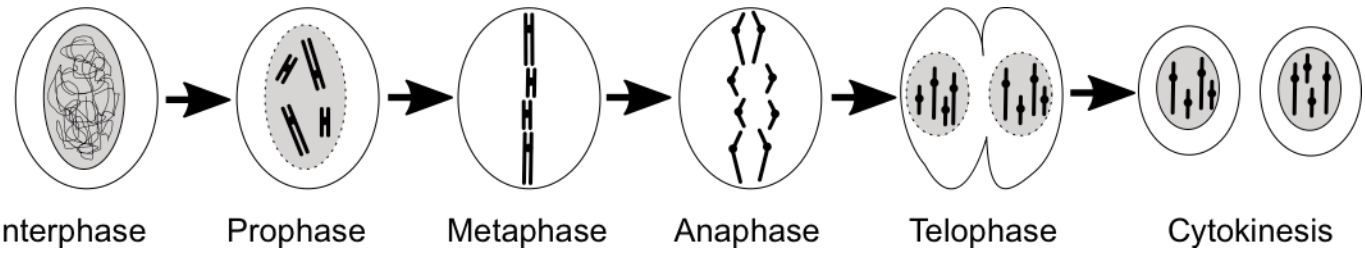


Figure 7-3

A cartoon diagram showing the main stages of Mitosis.

(Original-M. Deyholos/L. Canham-CC:AN)

next. When looking at snapshots of real cells, you will more often see cells between two stages, like is seen in some of the images in Figure 7-1.

Meiosis is a process which is just as dynamic as mitosis and shares some similar steps. That said, meiosis had developed, through evolutionary processes, extra steps which reduce the number of chromosomes in the daughter cells without losing any required chromosome types altogether. That is, the genetic instructions go from diploid to haploid in a carefully-controlled process.

C MEIOSIS

Most eukaryotes reproduce sexually - a cell from one individual joins with a cell from another to create offspring. In order for this to be successful, the cells that fuse must contain half the number of chromosomes as in the adult organism. Otherwise, the number of chromosomes would double with each generation, an unsustainable mechanism. The chromosome number is reduced through the process of meiosis. Meiosis is similar in many ways to (Figure 7-4), as the chromosomes are lined up along the metaphase plate and divided to the poles using microtubules. It also differs in many significant ways. Keep this in mind and try to note the differences as you read ahead. "N" refers to whether an organism is haploid or diploid; "C" is the quantity (mass) of DNA in each cell.

To begin, let's review what you should already know about meiosis. Much of this information should be review. It's fundamental to genetics, so keep in mind that we repeat it because it's important you understand it!

Meiosis has two main stages, designated by the roman numerals I and II. In Meiosis I homologous chromosomes segregate, while in Meiosis II sister chromatids segregate (Figure 7-4). Most multicellular organisms use meiosis to produce gametes, the cells that fuse to make offspring. Some single celled eukaryotes such as yeast also use meiosis to enter the haploid part of their life cycle. Cells that will undergo meiosis are called meiocytes and are diploid (2N)(Figure 7-4). You will hear of cells that have not yet undergone meiosis to become egg or sperm cells called oocytes or spermatocytes respectively.

Meiosis begins similarly to mitosis in that a cell has grown large enough to divide and has replicated its chromosomes. It differs in that Meiosis requires two rounds of division. In the first, known as meiosis I, the replicated, homologous chromo-somes segregate. During meiosis II the sister chromatids segregate. Note how meiosis I and II are both divided into prophase, metaphase, anaphase, and telophase, with those stages having similar features to mitosis (Figure 7-4). After two rounds of cytokinesis, four cells will be produced, each with a single copy of each chromosome in the set.

C.1 MEIOSIS I

Meiosis I is called a reductional division, because it reduces the number of chromosomes inherited in each of the daughter cells – the parent cell is 2N while the two daughter cells are each 1N. Meiosis I is further divided into Prophase I, Metaphase I, Anaphase I, and Telophase I, which are roughly similar to the corresponding stages of mitosis, except that in Prophase I and Metaphase I, homologous chromosomes pair up with each other, or synapse, and are called bivalents (Figure 7-6), in contrast with mitosis where the chromosomes line up individually during metaphase. This is an important difference between mitosis and meiosis, because it affects the segregation of alleles, and also allows for recombination to occur through crossing-over, which will be described later. During Anaphase I, one member of each pair of homologous chromosomes migrates to each daughter cell (1N) (Figure 7-8).

In meiosis I replicated, homologous chromosomes pair up, or synapse, during prophase I, line up in the middle of the cell during metaphase I, and separate during anaphase I. For this to happen the homologous chromosomes need to be brought together while they condense during prophase I. During synapsis, proteins bind to both homologous chromosomes along their entire length and form the synaptonemal complex (synapse means junction). These proteins hold the chromosomes in a transient structure called a bivalent (Figure 7-6). The proteins are released when the cell enters anaphase I.

C.2 Stages of prophase I

In meiosis, Prophase I is divided up into five visual stages, that are steps along a continuum of events (Figure 7-4 on page 66). Leptotene, zygotene, pachytene, diplotene and diakinesis. From interphase, a cell enters leptotene as the nuclear material begins to condense into long thin visible threads (chromosomes). During zygotene homologous chromosomes begin to pair up (synapse) and form an elaborate structure called the synaptonemal complex along their length. During zygotene the chromosomes are still quite long, but it

is more apparent that they are distinct now. At pachytene homologous chromosomes are thicker and fully synapsed (two chromosomes and four chromatids) to form bivalents. Crossing over (see section below) takes place in pachytene. After this, the pairing begins to loosen in diplotene. Remember that these are replicated chromosomes, but before this point this isn't apparent. This is also when the consequences of each crossing over event can be seen as a cross structure known as a chiasma (plural: *chiasmata*). Diakinesis follows as the chromosomes continue to fully condense and individualize. It is at this point that the nuclear membrane dissolves and the microtubules begin to form. This is followed by metaphase I were the paired chromosomes orient on the metaphase plate in preparation for segregation (reductional).

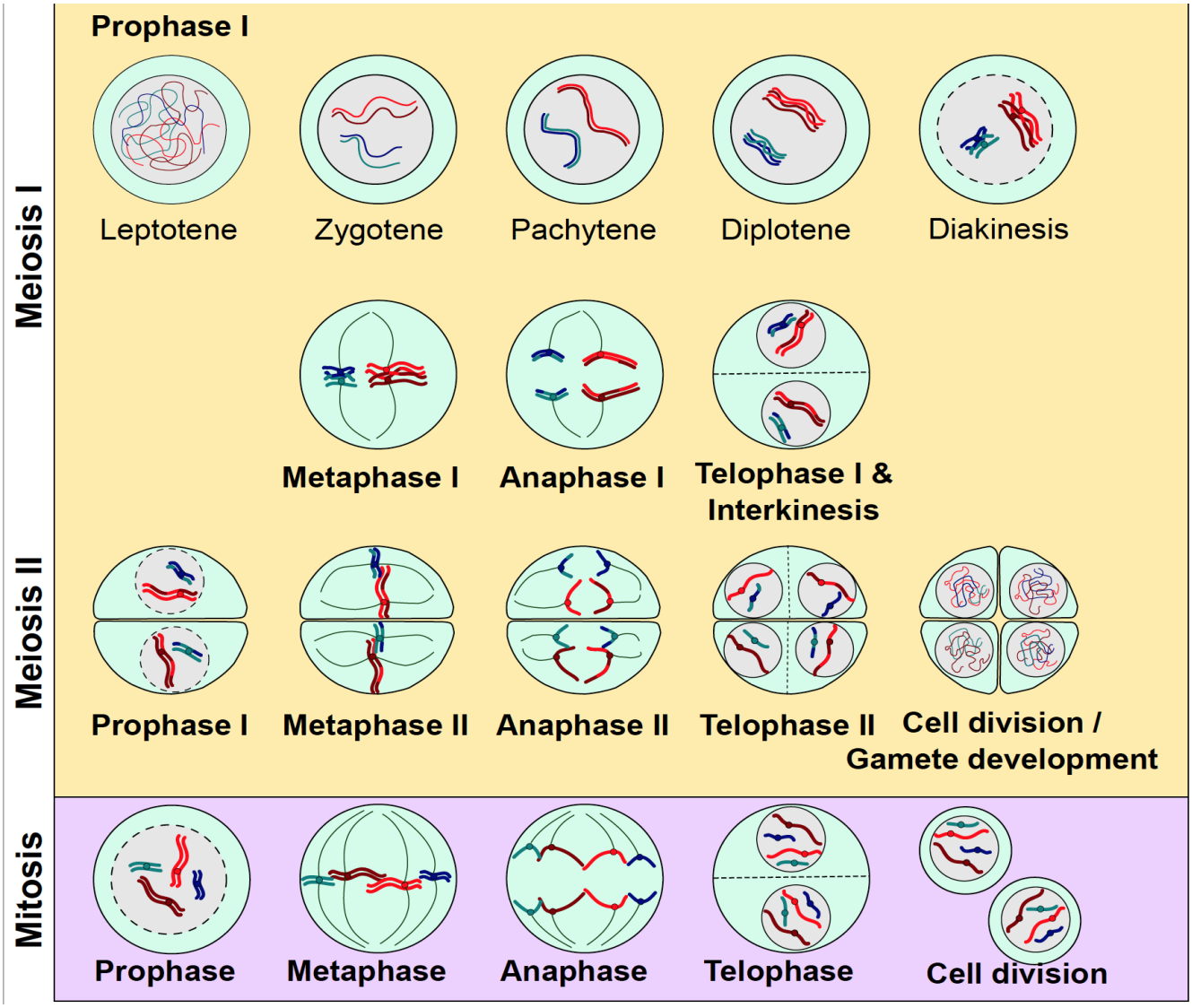


Figure 7-4

Stages of Prophase I and Meiosis with comparison to Mitosis. This example uses a diploid animal with 2 chromosome sets, so 4 chromosomes in total: Red, Maroon, Blue and Teal. Cross over events are shown between the two closest non-sister chromatids, but in reality can happen between all four chromatids.

Prophase I is divided into stages. Leptotene is defined by the beginning of chromosome condensation, though chromosomes are still long. Zygotene chromosomes are still long, but you can readily identify chromosomes as they are starting to pair. Pachytene chromosomes are thickening and fully synapsed. Diplotene you can begin to see the individual chromatids and chiasmata. Diakinesis, chromosomes are fully condensed and nuclear membrane dissolves. Metaphase I, the synapsed chromosomes align along the metaphase plate and then the synapse breaks in Anaphase I. Meiosis I is completed with Telophase I and potentially interkinesis, completing the reductional division. Meiosis II is an equational division where the chromosomes align in Metaphase II similarly to Mitosis, and complete Anaphase II and Telophase II, leaving with 4 haploid gametes formed. (Original–L. Canham–CC BY-NC 3.0)

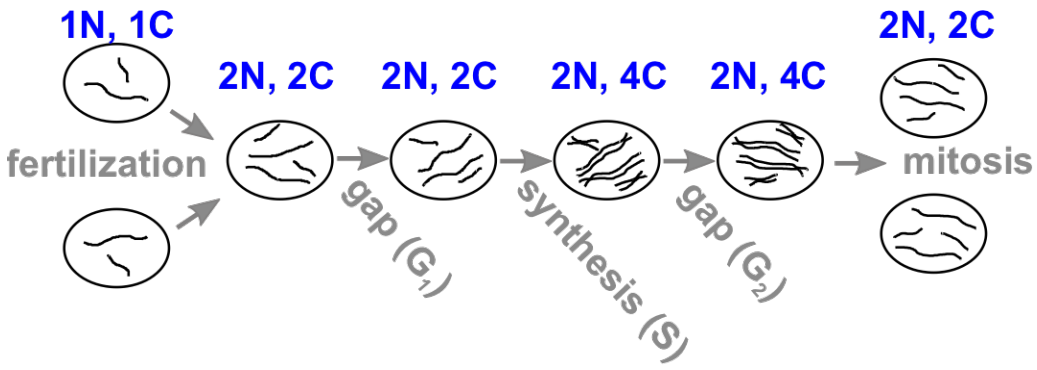


Figure 7-5

Changes in DNA and chromosome content during the cell cycle and mitosis. For simplicity, nuclear membranes are not shown, and all chromosomes are represented in a similar stage of condensation. (Original-M. Deyholos/L. Canham- CC BY-NC 3.0)

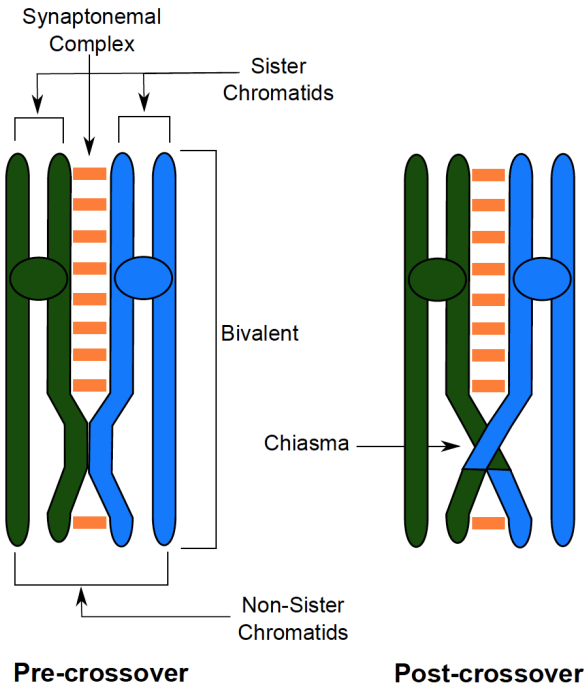
Figure 7-6

Diagram of a pair of homologous chromosomes during Prophase I.

Sister chromatids are chromatids found in one chromosome, so both blue chromatids are sister chromatids, and both green chromatids are sister chromatids. Non-sister chromatids are between chromosomes, so the green's non-sister chromatid is the blue.

When a pair of homologous chromosomes synapse during Prophase I they form a bivalent. Proteins known as the synaptonemal complex form between both chromosomes and join them together. Crossovers form between non-sister chromatids forming a cross-structure called a chiasma.

(Original-L. Canham- CC BY-NC 3.0)



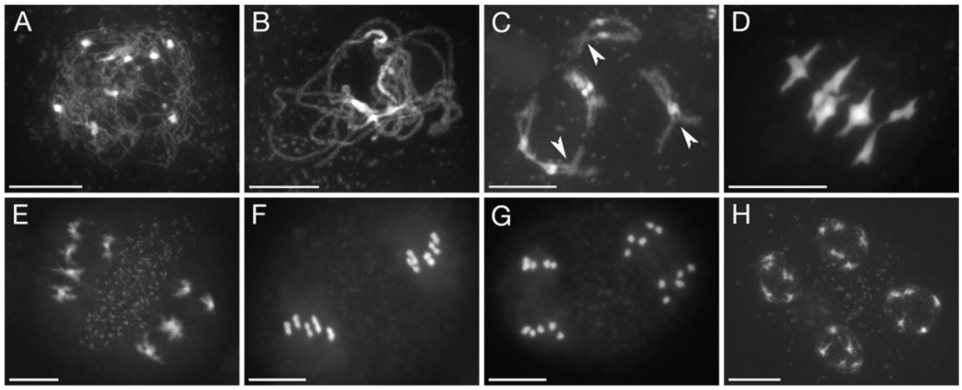


Figure 7-7

Meiosis in Arabidopsis (n=5). Panels A-C show different stages of prophase I, each with an increasing degree of chromosome condensation. Subsequent phases are shown: metaphase I (D), telophase I (E), metaphase II (F), anaphase II (G), and telophase II (H). (PLoS Genetics-Chelysheva, L. *et al* (2008) PLoS Genetics- CC BY 4.0)

C.3 METAPHASE I, ANAPHASE I AND TELOPHASE I

Metaphase I is where the major difference between mitosis and meiosis occurs. The homologous pairs, or bivalents, orient themselves along the metaphase plate and the microtubules attach themselves to each chromosome's centromere, one pole attaching to each respective homologous pair. Contrast to mitosis, where the chromosomes align individually and the microtubules from both poles attach them to an individual chromosome in preparation of separating the chromatids.

Anaphase I and Telophase I complete, separating the homologous pairs to their respective poles, but keeping the sister chromatids of each chromosome together. Telophase I completes with cell division to create two cells. Different organisms and cells behave differently after telophase I. In some, the nuclear membrane reforms around the chromosomes in each pole, the chromosomes become elongated again. These cells may stay in this state, known as interkinesis for some time. Other organisms the chromosomes will stay condensed, no nuclear membrane will form, and it will go directly into meiosis II.

C.4 MEIOSIS II

At the completion of meiosis I there are two cells, each with one, replicated copy of each chromosome (1N).

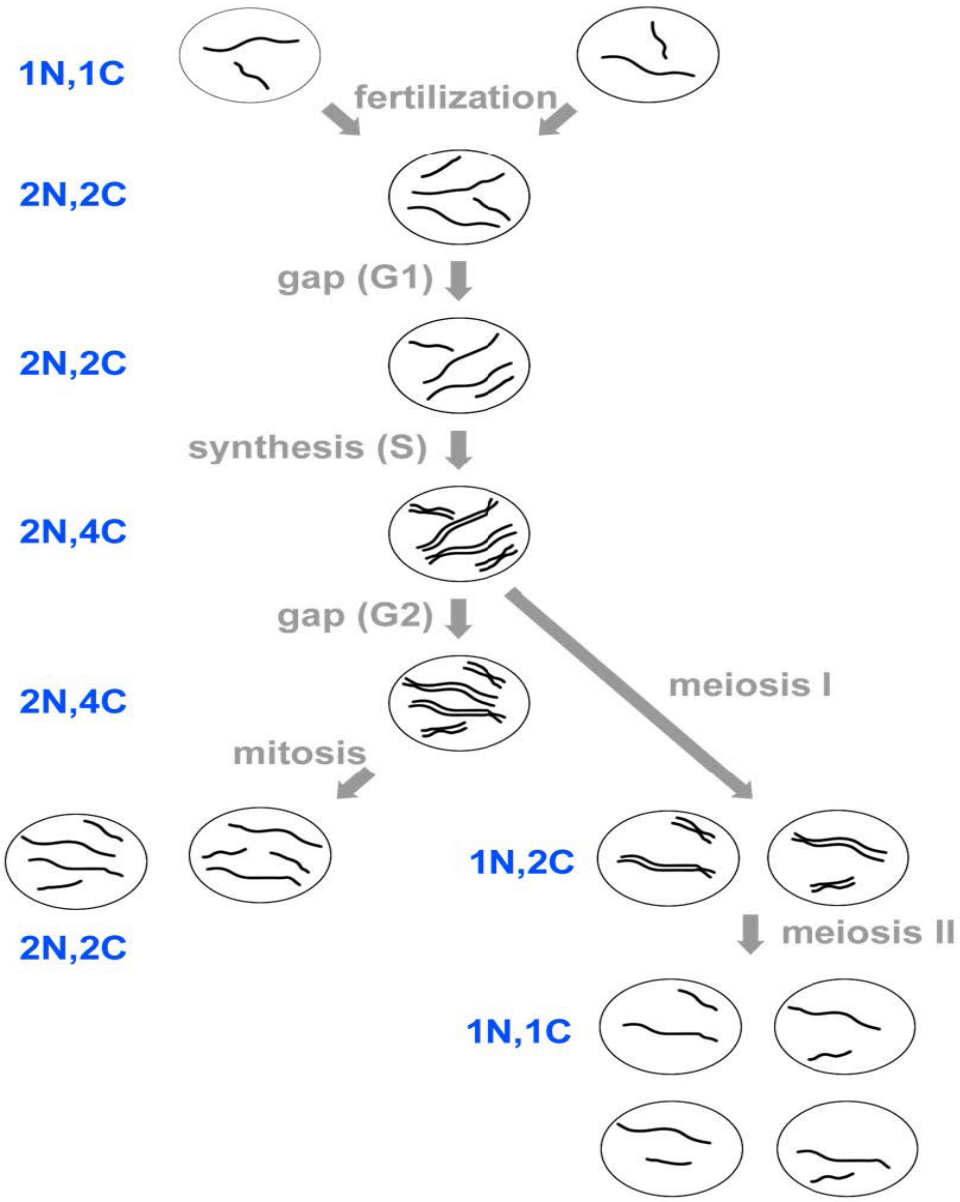


Figure 7-8

Changes in DNA and chromosome content during the cell cycle. For simplicity, nuclear membranes are not shown, and all chromosomes are represented in a similar stage of condensation. (Original-Deyholos- CC BY-NC 3.0)

Because the number of chromosomes per cell has decreased (2->1), meiosis I is called a reductional cell division. Meiosis II resembles mitosis, with one sister chromatid from each chromosome separating to produce two daughter cells. Because Meiosis II, like mitosis, results in the segregation of sister chromatids, Meiosis II is called an equational division.

If, after telophase I, the cells enter interkinesis, then during prophase II the haploid chromosomes will condense and the nuclear membrane will dissolve again. If interkinesis did not happen, then the cell will continue with meiosis II (Figure 1). Prophase II ends like in mitosis with the microtubules beginning to form. As metaphase II start, the pairs of sister chromatids align themselves along the metaphase plate, each chromatid attached to a microtubule from each pole. Anaphase II splits the sister chromatids and the microtubules pull them to the opposite poles. Telophase II reforms the nuclear membrane around the chromosomes, ending finally with cytokinesis and producing four cells with only one unreplicated chromosome of each type. There will be allelic differences among gametes based upon segregation of heterozygous alleles (Note the differences in colours of chromosomes in each of the gametes in Figure 7-4).

C.5 CROSSING OVER (INTRA-CHROMOSOMAL RECOMBINATION)

During prophase I the homologous chromosomes pair together and form a synaptonemal complex. Within the synaptonemal complex a second event, crossing over, occurs. These are places where DNA repair enzymes break the DNA of two non-sister chromatids in similar locations and then covalently reattach non-sister chromatids together to create a crossover between non-sister chromatids. This reorganization of chromatids will persist for the remainder of meiosis and result in recombination of alleles in the gametes. Crossover events can be seen as Chiasmata on the synapsed chromosomes in late Meiosis I.

Crossovers function to hold homologous chromosomes together during meiosis I so they orient correctly and segregate successfully. They also cause the reshuffling of gene/allele combinations to create genetic diversity, which can then be acted on in evolution.

C.6 GAMETE MATURATION

In animals and plants the cells produced by meiosis need to mature before they become functional gam-

etes. In male animals the four products of meiosis are called spermatids. They grow structures, like tails and become functional sperm cells. In female animals the gametes are eggs. In order that each egg contains the maximum amount of nutrients only one of the four products of meiosis becomes an egg. The other three cells end up as tiny disposable cells called polar bodies. In plants the products of meiosis reproduce a few times using mitosis as they develop into functional male or female gametes.

D MEASURES OF DNA CONTENT AND CHROMOSOME CONTENT

Figure 7-9
Marbled Lungfish (*Protopterus aethiopicus*) has a genome of ~133 x 10^9 base pairs, which is 200X that of a human. It is an example of the C-value paradox.

(Wikipedia-OpenCage- CC BY 2.5)

The amount of DNA within a cell changes during the following events: fertilization, DNA synthesis and mitosis (Figure 7-5). We use "C" to represent the DNA content in a cell, and "N" to represent the number of complete sets of chromosomes. In a gamete (i.e. sperm or egg), the amount of DNA is 1C, and the number of chromosomes is 1n. Upon fertilization, both the DNA content and the number of chromosomes doubles to 2C and 2N, respectively. Following DNA replication, the DNA content doubles again to 4C, but each pair of sister chromatids are still attached by the centromere, and so is still counted as a single chromosome (a replicated chromosome), so the number of chromosomes remains unchanged at 2N. If the cell undergoes mitosis, each daughter cell will return to 2C and 2N, because it will receive half of the DNA, and one of each pair of sister chromatids.

D.1 THE C-VALUE OF THE NUCLEAR GENOME

The complete set of DNA within the nucleus of any organism is called its nuclear genome and is measured as the C-value in units of either the number of base pairs or picograms of DNA. There is a general correlation between the nuclear DNA content of a genome (*i.e.* the C-value) and the physical size or complexity of an organism. Compare the size of *E. coli* and humans for example in **Table 1**. There are, however, many exceptions to this generalization, such as the human genome contains only 3.2 x 10^9 DNA bases, while the wheat genome contains 17 x 10^9 DNA bases, almost 6 times as much. The Marbled Lungfish (*Protopterus aethiopicus*, **Figure 7-9**) contains ~133 x 10^9 DNA bases, (~45 times as much as a human) and a fresh water amoeboid, *Polychaos dubium*, which has as much as 670 x 10^9 bases (200x a human).

D.2 THE C-VALUE PARADOX

This apparent paradox (called the C-value paradox) can be explained by the fact that not all nuclear DNA encodes genes – much of the DNA in larger genomes is non-gene coding. In fact, in many organisms, genes are separated from each other by long stretches of DNA that do not code for genes or any other genetic information. Much of this "non-gene" DNA consists of transposable elements of various types, which are an interesting class of self-replicating DNA elements discussed later in this textbook. Other non-gene DNA includes short, highly repetitive sequences of various types. Together, this non-functional DNA is often referred to as "Junk DNA".

D.3 THE "ONION TEST".

This "test" deals with any proposed explanation for the function(s) of non-coding (junk) DNA. For any proposed function for the excess of DNA in eukaryote genomes (C-value paradox) can it "explain why an onion needs about five times more non-coding DNA for this function than a human?" The onion Allium cepa has a haploid genome size of ~17 pg, while humans have only ~3.5 pg. Why? Also, onion species range from 7 to 31.5 pg, so why is there this range of genome size in organisms of similar complexity?

See : (`http://www.genomicron.evolverzone.com/2007/04/onion-test/`) for details.

You can practice chromosome positions during phases of meiosis online.

http://tinyurl.com/oog-meiosis

Summary:

- The asexual transmission of genetic information is accomplished in a process called Mitosis.
- The process of mitosis can be divided into Prophase, Metaphase, Anaphase, and Telophase.
- Mitosis reduces the C-number, but not the N-number of the daughter cells.
- Not all the DNA in an organism codes for genes. In most higher eukaryotes most DNA is non-gene coding and appears to have no specific function and is called "junk' DNA.
- Homologous chromosomes contain the same series of genes along their length, but not necessarily the same alleles. Sister chromatids initially contain the same alleles.
- Homologous chromosomes pair (synapse) with each other during meiosis, but not mitosis.
- A diploid can have up to two different alleles at a single locus. The alleles segregate equally between gametes during meiosis.
- The c-value paradox refers to the observation that the amount of DNA is not necessarily related to the complexity of the organism.

Key terms:

mitosis
interphase
G1 Phase
S Phase
G2 Phase
M Phase
G0 Phase
chromatids
prophase
metaphase
microtubules
kinetochore
metaphase plate
anaphase
telophase
unreplicated chromosome
synaptonemal complex
cytokinesis
meiosis I
meiosis II
gametes
meiocytes
reductional
synapse
bivalent leptotene
zygotene
pachytene
crossing over
diplotene
chiasma / chiasmata
diakinesis
metaphase I
anaphase I
telophase I
interkinesis
prophase II
metaphase II
anaphase II
telophase II
N-value
C-value
replicated chromosome
nuclear genome
C-value paradox

Study Questions:

1. Species A has N=4 chromosomes and Species B has N=6 chromosomes. Can you tell from this information which species has more DNA? Can you tell which species has more genes?

1. The answer to question 2 implies that not all DNA within a chromosome encodes genes. Can you name any examples of chromosomal regions that contain relatively few genes?
2. Consider the following:
   a) How many centromeres does a typical chromosome have?
   b) What would happen if there was more than one centromere per chromosome?
   c) What if a chromosome had zero centromeres?
3. For a diploid with 2n=16 chromosomes, how many chromosomes and chromatids are per cell present in the gamete, the zygote, and immediately following
   a) G_1,
   b) S,
   c) G_2,
   d) Mitosis
4. Refer to **Table 7-1 on page 71**.
   a) What is the relationship between DNA content of a genome, number of genes, gene density, and chromosome number?What feature of genomes explains the c-value paradox?
   b) Do any of the numbers in this Table show a correlation with organismal complexity?
5. You are working with a cell that has a genome of 2N=4 and is undergoing meiosis. Draw three large circles to represent the cell. The first circle should demonstrate the cell as a diploid in the G1 phase. The second circle should have the same cell after it reaches G2. The final circle should show the cell at prophase II (just draw one cell). Shade the chromosomes so that homologous chromosomes are different from each other, and be sure to draw nonhomologous chromosomes so they are also distinct from each other (you can make the chromosome arms larger or smaller, and move about the position of the centromere, for example. Assume no crossovers.

6. Does equal segregation of alleles into daughter cells happen during mitosis, meiosis, or both?
7. What is the maximum number of alleles for a given locus in a normal gamete of a diploid species?
8. A simple mnemonic for leptotene, zygotene, pachytene, diplotene, & diakinesis is Lame Zebras Pee Down Drains. Make another one yourself.

Table 7-1

Measures of genome size in selected organisms. The DNA content (1C) is shown in millions of basepairs (Mb). For eukaryotes, the chromosome number is the chromosomes counted in a gamete (1N) from each organism. The average gene density is the mean number of non-coding bases (in bp) between genes in the genome.

	DNA content (Mb, 1C)	Estimated gene number	Average gene density	Chromosome number (1N)
Homo sapiens	3,200	25,000	100,000	23
Mus musculus	2,600	25,000	100,000	20
Drosophila melanogaster	140	13,000	9,000	4
Arabidopsis thaliana	130	25,000	4,000	5
Caenorhabditis elegans	100	19,000	5,000	6
Saccharomyces cerevisiae	12	6,000	2,000	16
Escherichia coli	5	3,200	1,400	1

Chapter 8
Human Chromosomes

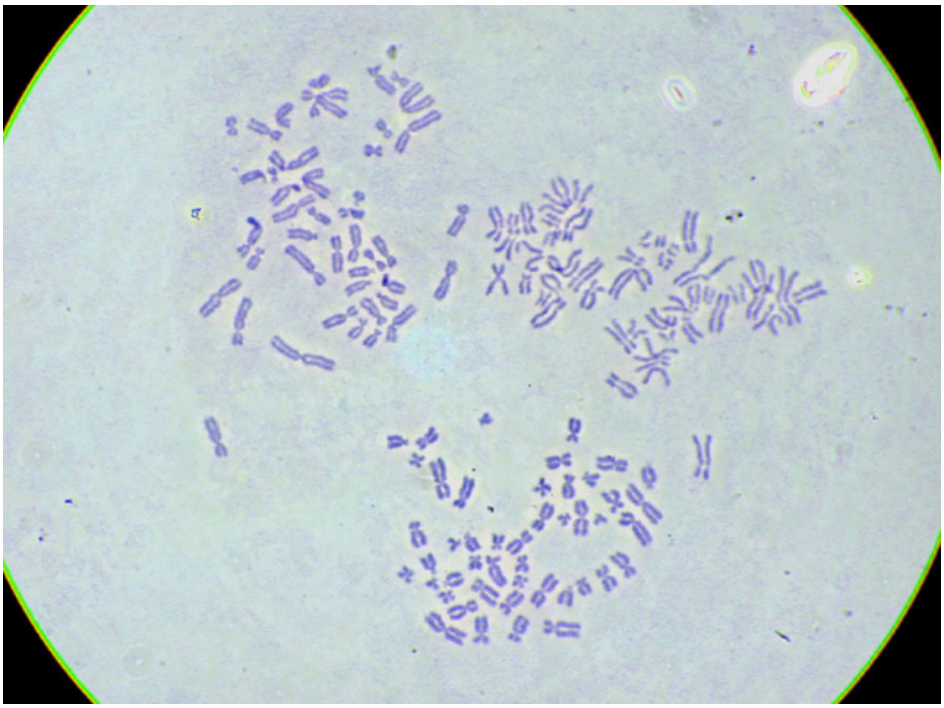


Figure 8-10
Human metaphase chromosome spread. To make these figures white blood cells in metaphase are dropped onto a slide. The cells burst open and the chromosomes can then be stained a purple colour. This image shows chromosomes from three cells that hit the slide close to one another.

(Original-Alexander Smith - CC BY-NC 3.0)

Humans, like all other species, store their genetic information in cells as large DNA molecules called chromosomes. Within each nucleus are 23 pairs of chromosomes, half from mother and half from father. In addition, our mitochondria have their own smaller chromosome that encodes some of the proteins found in this organelle.

A Metaphase Chromosome Spreads

A.1 The shape of the chromosomes

Figure 8-10 shows chromosomes from three cells. Each of the cells was in metaphase stage of mitosis, which is why the chromosomes appear **replicated** and **condensed**. We refer to chromosomes as being replicated when they consist of two sister chromatids held together at the centromeres. DNA replication occurs during S phase. These chromosomes are also condensed. Chromosomes are compacted at the start of mitosis in prophase. **Cytogeneticists** can observe chromosomes at any stage of the cell cycle but those from metaphase cells provide the most detail and clarity.

Figure 8-11 shows a more magnified view of a pair of chromosomes. On average a condensed human metaphase chromosome is 5 µm long and each chromatid is 700 nm wide. In contrast, a decondensed interphase chromosome is 2 mm long and only 30 nm wide, yet still fits into a single nucleus.

Figure 8-11
A pair of unreplicated metacentric human chromosome #1.

(Wikipedia- National Human Genome Research Institute-PD)

A.2 The Amount of DNA in a Cell (C-value)

To calculate how much DNA is seen in the nuclei in Figure 8-10, consider that a human gamete has about 3000 million base pairs. We can shorten this statement to 1c = 3000 Mb where c is the **C-value**, the DNA content in a gamete. When an egg and sperm join the resulting zygote is 2C = 6000 Mb. Before the zygote can divide and become two cells it must undergo DNA replication. This doubles the DNA content to 4C = 12 000 Mb. When the zygote does divide, each daughter cell inherits half the DNA and is therefore back to 2C = 6000 Mb. Then each cell will become 4C again (replication) before dividing themselves to become 2C each. From this point forwards, every cell in the embryo will be 2C = 6000 Mb before its S phase and 4C = 12 000 Mb afterwards. The same is true for the cells of fetuses, children, and adults. Because the cells used to prepare this chromosome spread were adult cells in metaphase each is 4C = 12 000 Mb. Note, there are some rare exceptions, such as some stages of meiocytes that make germ cells and other rare situations like the polyploidy of terminally differentiated liver cells. This is summarized in **Table 8-1**.

Table 8-1
DNA content in human cells at various phases.

Human cell	DNA content
gamete (egg or sperm)	1C = 3000 Mb
regular cell before S phase	2C = 6000 Mb
regular cell after S phase	4C = 12 000 Mb

A.3 The Number of Chromosomes (N-value)

Human gametes contain 23 chromosomes. We can summarize this statement as N = 23 where N is the **N-value**, the number of chromosomes in a gamete. When an N = 23 sperm fertilizes a N = 23 egg, the zygote will be 2N = 46. But, unlike DNA content (C) the number of chromosomes (N) does *not* change with DNA replication. One replicated chromosome is still just one chromosome. Thus the zygote stays 2N = 46 after S phase. When the zygote divides into two cells both contain 46 chromosomes and are still 2N = 46. Every cell in the embryo, fetus, child, and adult is also 2N = 46 (with the exceptions noted above).

Table 8-2 at right summarizes this.

Note that in the regular cell they are 2N before and 2N after chromosome replication. The N-value doesn't change although the C-value does.

B Human Karyotypes

B.1 Karyotypes

Human cytogenetists use metaphase chromosome spreads as a standard representation of the chromosomes in a cell, organism, or species. Comparisons permit them to identify chromosome abnormalities. Because it can be hard to distinguish individual chromosomes, cytogeneticists sort the photo to put the chromosomes into a standard pattern. The result is a karyotype ("nucleus picture"; **Figure 8-12**). In the past it was necessary to print a photograph of the metaphase spread, cut out each chromosome with scissors, and then glue each to a piece of cardboard to show the pattern. Now, computer software does much of this for us, but a qualified cytogeneticist usually reviews the karyotype assembly. But either way, the random collection of chromosomes seen in Figure 8-10 is converted to the organized pattern in **Figure 8-12.**

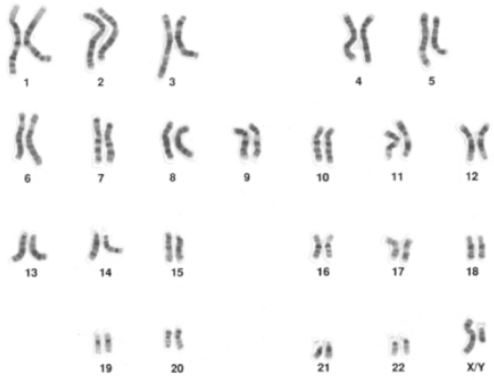


Figure 8-12
A normal human karyotype.
(Wikipedia- National Human Genome Research Institute-PD)

Table 8-2
Ploidy levels of a normal human cell. Note that if you use the rule "count the centromeres", replicated and unreplicated chromosomes each count as a single chromosome.

Human cell	Chromosome number
gamete (egg or sperm)	1N = 23
regular cell before S phase	2N = 46
regular cell after S phase	2N = 46

B.2 Human Chromosomes – Autosomes

The chromosomes are numbered to distinguish them. Chromosomes 1 through 22 are **autosomes**, which are present in two copies in both males and females. Because human chromosomes vary in size this was the easiest way to label them. Our largest chromosome is number 1, our next longest is 2, and so on. The karyotype above shows two copies of each of the autosomes. A karyotype from a normal female would also show these 22 pairs. There are also the sex-chromosomes, X and Y (see below). Normal females have two X-chromosomes, while normal males have an X and a Y each. They act as a homologous pair, similar to the autosomes. During meiosis only one of each autosome pair and one of the sex-chromosomes makes it into the gamete. This is how 2N = 46 adults can produce N = 23 eggs or sperm.

In addition to their length, Cytogenetists can distinguish chromosomes using their centromere position and banding pattern. Note that at the resolution in Figure 3 both chromosome 1s look identical even though at the base pair level there are small and often significant differences in the sequence that correspond to allelic differences between these homologous chromosomes.

Remember that in each karyotype there are **maternal chromosomes**, those inherited from their mother, and their **paternal chromosomes**, those from their father. For example, everyone has one maternal chromosome 1 and one paternal chromosome 1. In a typical karyotype it usually isn't possible to tell which is which. In some cases, however, there are visible differences between homologous chromosomes that do permit the distinction to be made.

Table 8-3
Table of terms describing different types and parts of chromosomes.

homologous chromosomes	the maternal and paternal copies of a chromosome	maternal chromosome 1 and paternal chromosome 1
non-homologous chromosomes	two different chromosomes within the same cell/organism	a chromosome 1 and a chromosome 8
sister chromatids	the identical chromatids within a single replicated chromosome	the two chromatids within maternal chromosome 1
non-sister chromatids	the similar but not identical chromatids from homologous chromosomes	a chromatid in maternal chromosome 1 and a chromatid in paternal chromosome 1

B.3 Relationships between Chromosomes and Chromatids

Figure 8-13 summarizes the terms we've covered so far, karyotypes depict replicated chromosomes and two copies of each chromosome because we usually model diploid cells. So how do we refer to all the pieces of DNA present?

B.4 Human Sex Chromosomes

Figure 8-12 shows that most of our chromosomes are present in two copies. Each copy has the same length, centromere location, and banding pattern. As mentioned before, these are called autosomes. However note that two of the chromosomes, the X and the Y, do not look alike. These are **sex chromosomes**. In mammals, males have one of each while females have two X chromosomes.

Autosomes are those chromosomes present in the same number in males and females while sex chromosomes are those that are not. When sex chromosomes were first discovered their function was unknown and the name X was used to indicate this mystery. The next one was named Y.

It is a popular misconception that the X and Y chromosomes were named based upon their shapes; physically each looks like any other chromosome. A Y-chromosome doesn't look like a Y any more than a chromosome 4 looks like a 4.

The combination of sex chromosomes for a species is associated with either male or female individuals. In mammals, fruit flies, and some flowering plants, XX individuals are females while XY individuals are males.

How do the sex chromosome behave during meiosis? Well, in those individuals with two of the same chromosome (i.e. XX females) the chromosomes pair and segregate during meiosis I the same as autosomes do. During meiosis in XY males the sex chromosomes pair with each other (**Figure 8-14** on the next page). In mammals the consequence of this is that all egg cells will carry an X chromosome while the sperm cells will

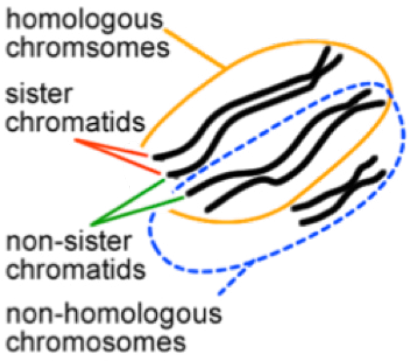


Figure 8-13
The relationships between chromosomes and chromatids.

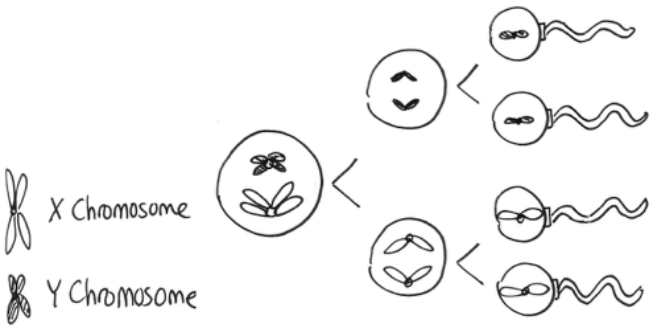


Figure 8-14

Meiosis in an XY mammal. The stages shown are anaphase I, anaphase II, and mature sperm. Note how half of the sperm contain Y chromosomes and half contain X chromosomes.

carry either an X or a Y chromosome. Half of the offspring will receive two X chromosomes and become female while half will receive an X and a Y and become male.

B.5 Human karyotypes

We can summarize the information shown in a karyotype such as Figure 8-12 with a written statement known as a **karyotype** ("nucleus features"). By convention we list (i) the total number of chromosomes, (ii) the sex chromosomes, and (iii) any abnormalities. The karyotype in Figure 8-12 would be 46,XY, which is typical for human males. Most human females are 46,XX.

If a cytogeneticist sees an abnormality, it may not be harmful or detrimental. For example many people in the world have a chromosome 9 with an inversion in the middle. They are therefore 46,XY,inv(9) or 46,XX-,inv(9). Other chromosomal abnormalities do have an effect on a person's health and well being. An example is 47,X/Y,+21 or 47,X/X,+21. These people have an extra copy of chromosome 21, a condition also known as trisomy-21 and Down Syndrome. These and other examples are described in the chapters on chromosome structure changes and chromosome number changes.

C Parts of a Typical Nuclear Chromosome

A functional chromosome requires four features. These are shown in **Figure 8-15.**

C.1 Thousands of genes

In the previous sections we mentioned human chromosome 1, but what exactly is it? Well, each chromosome is long molecule of double stranded DNA with one purpose. They contain **genes** in the cell. Chromosome 1, being our largest chromosome has the most genes, about 4778 in total. Many of these genes are transcribed into mRNAs, which encode proteins. Other genes are transcribed into tRNAs, rRNA, and other non-coding RNA molecules.

C.2 One centromere

A **centromere** ("middle part") is a place where proteins attach to the chromosome as required during the cell cycle. Cohesin proteins hold the sister chromatids together beginning in S phase. Kinetochore proteins form attachment points for microtubules during mitosis. The metaphase chromosomes shown in the karyotype in Figure 8-12 have both Cohesin and Kinetochore proteins at their centromeres. There are no genes within the centromere region DNA; all that is present is a simple repeated DNA sequence.

While all human chromosomes have a centromere it isn't necessarily in the middle of the chromosome. If it *is* in the centre the chromosome it is called a **metacentric** chromosome. If it is offset a bit it is **submetacentric**, and if it is towards one end the chromosome is **acrocentric**. In humans an example of each is chromosome 1, 5, and 21, respectively. Humans do not have any **telocentric** chromosomes, those with the centromere at one end, but mice and some other mammals do.

C.3 Two telomeres

The ends of a chromosome are called **telomeres** ("end parts"). Part of the DNA replication is unusual here, it is done with a dedicated DNA polymerase known as a Telomerase. Chapter 2 on DNA replication goes into more detail. As with the centromere region there are no genes in the telomeres, just simple, repeated DNA sequences.

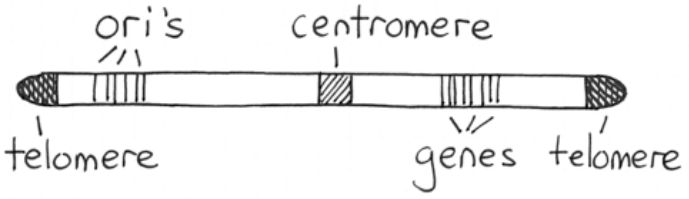


Figure 8-15

Parts of a typical human nuclear chromosome (not to scale). The ORIs and genes are distributed everywhere along the chromosome, except for the telomeres and centromere.

C.4 Thousands of Origins of Replication

At the beginning of S phase DNA polymerases begin the process of chromosome replication. The sites where this begins are called **origins of replication** (ori's). They are found distributed along the chromosome, about 40 kb apart. S phase begins at each ori as two replication forks leave travelling in opposite directions. Replication continues and replication forks travelling from one ori will collide with forks travelling towards it from the neighboring ori. When all the forks meet, DNA replication will be complete.

D Appearance of a Typical Nuclear Chromosome During the Cell Cycle

If we follow a typical chromosome in a typical human cell it alternates between unreplicated and replicated and between relatively uncondensed and condensed. The replication is easy to explain, if a cell has made the commitment to divide it first needs to replicate its DNA. This occurs during S phase. Before S phase chromosomes consist of a single piece of double stranded DNA and after they consist of two identical double stranded DNAs.

The condensation is a more complex story because eukaryote DNA is always wrapped around some proteins.

Figure 8-16 shows the different levels possible. During interphase a chromosome is mostly **30 nm fibre**. This allows it to fit inside the nucleus and still have the DNA be accessible for enzymes performing RNA synthesis, DNA replication, and DNA repair. At the start of mitosis these processes halt and the chromosome becomes even more condensed. This is necessary so that the chromosomes are compact enough to move to the opposite ends within the cell. When mitosis is complete the chromosome returns to its 30 nm fibre structure. Recall that each of our cells has a maternal and a paternal chromosome 1.

Figure 8-17 shows what these chromosomes look like during the cell cycle.

E DNA is Packaged into Chromatin

E.1 DNA can be Highly Compacted

If stretched to its full length, the DNA molecule of

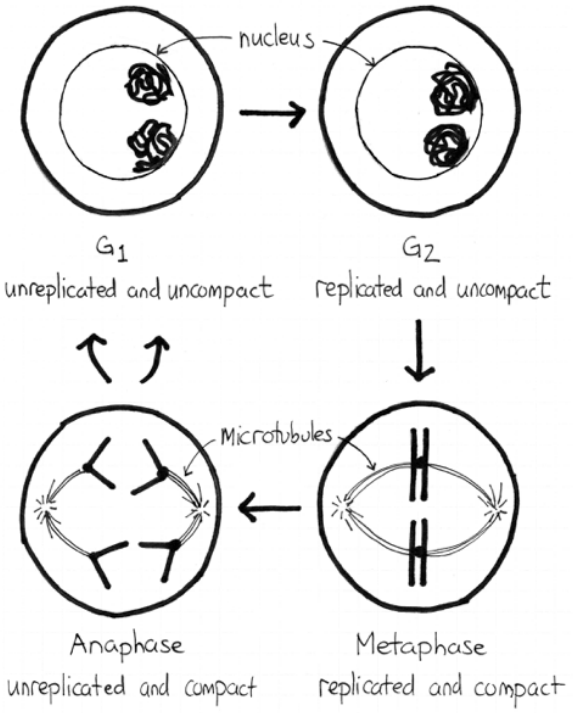


Figure 8-17
What maternal and paternal chromosomes 1 look like during the cell cycle. The other 44 chromosomes are not shown. Note that they are independent during both interphase (top) and mitosis (bottom). After anaphase there will be two cells in G_1.

(Original-Harrington- CC BY-NC 3.0)

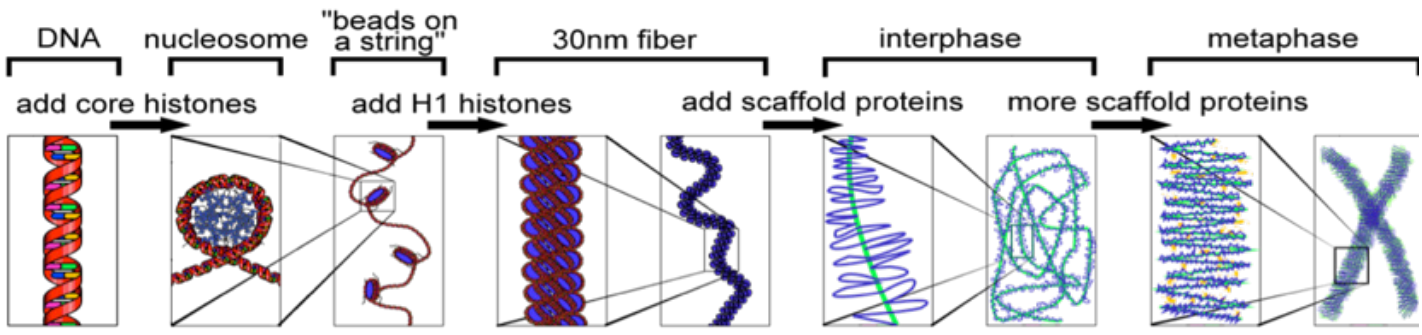


Figure 8-16
Successive stages of chromosome condensation depend on the introduction of additional proteins.

(Wikipedia-R. Wheeler- CC BY-SA 3.0)

the largest human chromosome would be 85mm. Yet during mitosis and meiosis, this DNA molecule is compacted into a chromosome approximately 5μm long. Although this compaction makes it easier to transport DNA within a dividing cell, it also makes DNA less accessible for other cellular functions such as DNA synthesis and transcription. Thus, chromosomes vary in how tightly DNA is packaged, depending on the stage of the cell cycle and also depending on the level of gene activity required in any particular region of the chromosome.

E.2 Levels of compaction

There are several different levels of structural organization in eukaryotic chromosomes, with each successive level contributing to the further compaction of DNA (Figure 8-16). For more loosely compacted DNA, only the first few levels of organization may apply. Each level involves a specific set of proteins that associate with the DNA to compact it. First, proteins called the **core histones** act as spool around which DNA is coiled twice to form a structure called the **nucleosome**. Nucleosomes are formed at regular intervals along the DNA strand, giving the molecule the appearance of "beads on a string". At the next level of organization, **histone H_1** helps to compact the DNA strand and its nucleosomes into a **30nm fibre**. Subsequent levels of organization involve the addition of **scaffold proteins** that wind the 30nm fibre into coils, which are in turn wound around other scaffold proteins.

E.3 Chromatin packaging varies inside the nucleus: euchromatin and heterochromatin

Chromosomes stain with some types of dyes, which is how they got their name (chromosome means "colored body"). Certain dyes stain some regions along a chromosome more intensely than others, giving some chromosomes a banded appearance. The material that makes up chromosomes, which we now know to be proteins and DNA, is called **chromatin**. Classically, there are two major types of chromatin, but these are more the ends of a continous and varied spectrum. **Euchromatin** is more loosely packed, and tends to contain genes that are being transcribed, when compared to the more densely compacted **heterochromatin,** which is rich in repetitive sequences and tends not to be transcribed. Heterochromatin sequences include short, highly-repetitive sequences called **satellite DNA**, which acquired their name because their bouyant density is distictly different from the main band of DNA following ultracentrifugation.

F Parts and Appearance of a Mitochondrial Chromosome

While most of our chromosomes are within the nucleus there is also DNA within the mitochondria. The human **mtDNA** is small, only 16.6 kb, and circular, although it is double stranded like most DNA molecules. It has only 37 genes, 13 of these make mitochondrial proteins and the rest encode tRNAs and rRNAs.

Each mtDNA has a single origin of replication. During DNA replication two replication forks leave the ori and halt when they bump into each other on the opposite side of the circle. DNA replication inside the mitochondria happens throughout interphase, not once during S phase as with the nuclear chromosomes. The consequence is that each mitochondrion has between 2 to 10 identical copies of each mitochondrial chromosome (**Figure 8-18**).

There are other differences compared with nuclear chromosomes. Organelles such as mitochrondria or

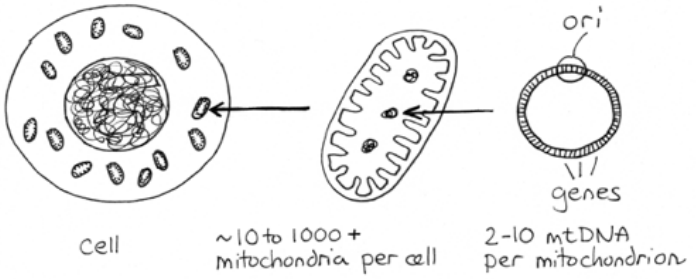


Figure 8-18

The relationship between cells, mitochondria, and mitochondrial DNA.

chloroplasts are likely the remnants of prokaryotic endosymbionts that entered the cytoplasm of ancient progenitors of today's eukaryotes (**endosymbiont theory**). These endosymbionts had their own, circular chromosomes (**Figure 8-19**), like most bacteria that exist today. Mitochondria typically have circular chromosomes that behave more like bacterial chromosomes than eukaryotic chromosomes, (i.e. mitochondrial genomes do not undergo mitosis or meiosis). Also, it lacks Histones or other proteins that compact it. It also lacks a centromere because mitochondrial replication is simpler than nuclear chromosome replication. The mitochondria just grow larger and splits in two, like the cells of its pro-

karyote origin. Because there are multiple mtDNAs and they are randomly distributed in the matrix both new mitochondria will end up inheriting some mtDNAs. And lastly because the mtDNA is circular there are no ends and thus no telomere regions.

Table 8-4 compares nuclear and mitochondrial chromosomes.

Table 8-4
Comparison of nuclear and mitochondrial chromosomes.

Feature	Nuclear chromosomes	Mitochondrial chromosome
DNA	linear double stranded DNA	circular double stranded DNA
genes	thousands	37
centromeres	1	0
telomeres	2	0
origins of replication	thousands	1
Mitosis/ Meiosis	Yes	No

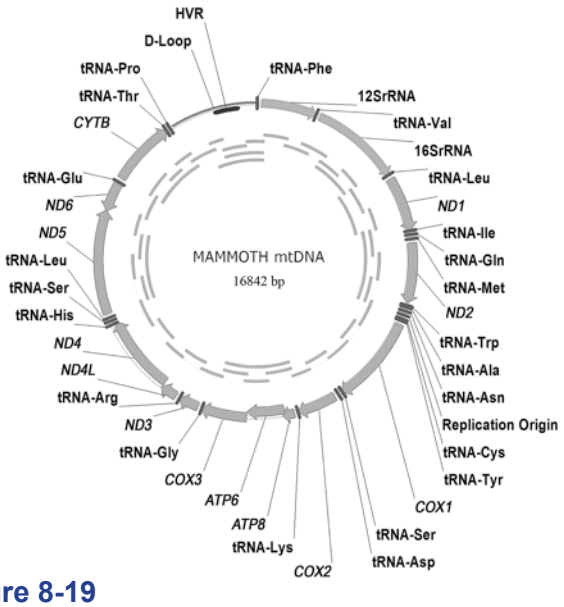


Figure 8-19
A map of the complete mitochondrial chromosome of the woolly mammoth (*Mammuthus primigenius*). The **mtDNA** that was used to produce this map was obtained from tissue of a mammoth that lived approximately 32,000 years ago. Circular organellar chromosomes such one as this are typical of almost all eukaryotes. (From Rogaev et al, 2006). Recent (Rohland et al, 2010) mtDNA work indicates that mammoths are more closely related to Indian elephants than to either of the African species.

G Example Genes

G.1 *LCT* - an autosomal gene

The *LCT* gene encodes the enzyme Lactase (see Chapter 9). This enzyme allows people to digest the milk sugar lactose. The *LCT* gene is on chromosome 2. Because this is an autosome everyone has a maternal and a paternal *LCT* gene. Genes come in different versions called alleles. The allele of the *LCT* gene you inherited from your mother will probably be slightly different from the allele you received from your father. So most people have two different alleles of this gene. If we consider a cell in G_1 there will be two pieces of DNA inside the nucleus with this gene. When this cell completes DNA replication there will be four of these genes. But because the chromatids on your maternal chromosome 2 are identical as are the chromatids on your paternal chromosome 2 this cell will still have just two different alleles. Because of this we simplify things by saying that humans have two copies of *LCT*. Because most genes are on autosomes you have two copies of most of your genes.

G.2 *HEMA* - an X chromosome gene

The *HEMA* gene makes a blood-clotting protein called Coagulation Factor VIII (F8) (see Chapter 22). Without F8 a person is unable to stop bleeding if injured. The *HEMA* gene is on the X chromosome. Females, with two X chromosomes, have two copies of the *HEMA* gene. Males only have one X chromosome and thus a single *HEMA* gene. This has an impact on male health, a topic discussed in Chapter 23 on pedigree analysis.

G.3 *SRY* - a Y chromosome gene

The *SRY* gene is only on the Y chromosome (see Chapter 22). This means that males have this gene and females do not. In fact it is the presence or absence of this gene that leads to humans being male or female. It all begins during embryogenesis. A pair of organs called the gonads can develop into either ovaries or testes. In XY embryos the *SRY* gene makes a protein that causes the gonads to mature into testes. Conversely, XX embryos do not have this gene and their gonads mature into ovaries instead. Once formed the testes produce sex hormones that direct the rest of the developing embryo to become male, while the ovaries make different sex hormones that promote female development. The testes and ovaries are also the organs where gametes (sperm or eggs) are produced. Whether a person is male or female is

decided at the moment of conception, if the sperm carries a Y chromosome the result is a male and if the sperm carries an X the result is a female.

G.4 MT-CO1 - A MITOCHONDRIAL GENE

The *MT-CO1* gene makes a protein in Complex IV of the mitochondrial electron transport chain. For reasons that aren't clear this protein must be made within the mitochondria. It can't be synthesized in the cytosol of the cell and then imported into the mitochondria afterwards like with most mitochondrial proteins. The *MT-CO1* gene is found on the mtDNA chromosome. Because humans get all of their mitochondria from their mother everyone has only one *MT-CO1* gene. It is the same one their mother (and her mother) possessed. (Technically speaking we have only one MT-CO1 *allele*, it will be identical on all of the mtDNA molecules in all of the mitochondria in all of the cells.)

Gene copy number is summarized in **Table 8-5**.

Table 8-5
Gene copy number for various human chromosomes.

Location of a gene	Number of this gene in males	Number of this gene in females
autosome chromosome	2	2
X chromosome	1	2
Y chromosome	1	0
mitochondria chromosome	1	1

SUMMARY:

- The C-value is the amount of DNA in a gamete. Humans are 1c = 3000 Mb.
- The N-value is the number of chromosomes in a gamete. Humans are N = 23.
- A typical cell in your body is 2C = 6000 Mb and 2N = 46 before DNA replication and 4C = 12 000 Mb and 2N = 46 after.
- A picture of metaphase chromosomes can be organized into a karyotype figure and described with a karyotype statement.
- Humans have two copies of each autosome chromosome. Females have two X chromosomes while males have one X and one Y chromosome.
- A typical nuclear chromosome has thousands of genes, one centromere, two telomeres, and thousands of origins of replication.
- A typical nuclear chromosome is replicated during S phase and consists of two chromatids up until the start of anaphase. It is condensed during prophase and remains condensed until the start of telophase. During metaphase a chromosome is both replicated and condensed for these reasons.
- The human mitochondrial chromosome has 37 genes, a single origin of replication, and no centromeres or telomeres.
- Humans have ~29 436 genes, most of which are on autosomal chromosomes.
- A typical human cell has two copies of each autosomal gene and one of each mitochondrial gene. Genes on sex chromosomes are different: females have two of each X chromosome gene while males have one; males have Y chromosome genes while females do not.

Key terms:

30 nm fibre
acrocentric
autosome
C-value
centromere
chromatin
condensed chromosome
cytogeneticist
endosymbiont theory
euchromatin
fibre
gene
heterochromatin
histone H1
histones
homologous chromosome
karyotype
karyotype
maternal chromosome
metacentric
mtDNA
n-value
non-homologous chomosome
non-sister chromatids
nucleosome
origin of replication
paternal chromosome
replicated chromosome
satellite DNA
scaffold proteins
sex chromosome
sister chromatids
submetacentric
telocentric
telomere

Questions:

1. Cytogeneticists use white blood cells in metaphase to make chromosome spread figures
   a) Why don't they use red blood cells?
   b) Why don't they use white blood cells in anaphase?
2. The human Y chromosome is smaller than the X chromosome. Does this mean that males have less DNA than females?
3. Are these statements true or false? For the false statements explain why.
   a) Everyone has a paternal chromosome 1.
   b) Everyone has a maternal chromosome 1.
   c) Everyone has a paternal X chromosome.
   d) Everyone has a maternal X chromosome.
   e) Everyone has a paternal Y chromosome.
   f) Everyone has a maternal Y chromosome.
   g) Everyone has a paternal mitochondrial chromosome.
   h) Everyone has a maternal mitochondrial chromosome.
4. Explain why centromeres do not have to be in the centre of a chromosome to function.
5. Why do nuclear chromosomes have to have multiple origins of replication?
6. Define chromatin. What is the difference between DNA, chromatin and chromosomes?
7. Have a look at Figure 8-17 on page 77. Which of these chromosomes would be associated with:
   a) Histone proteins (see Figure 8-16 on page 77)
   b) Condensin proteins (important scaffold proteins)
   c) Cohesin proteins (proteins which hold sister chromatids together)
   d) Kinetochore proteins (proteins which connect centromere DNA to Microtubules)
8. Where would you find these enzymes in a typical human cell?
   a) DNA polymerases
   b) RNA polymerases
   c) Ribosomes
9. Could the following genes continue to perform their function if they were moved next to the *LCT* gene on Chromosome 2?
   a) HEMA
   b) SRY
   c) MT-CO1

Chapter 9
Gene Interactions

Figure 9-20
The coat colour on this juvenile horse is called Bay Roan Tobiano. Bay is the brown base coat colour; Roan is the mixture of white hairs with the base coat, making a 'foggy' colour; and Tobiano is the white patches. The genes causing the Roan and Tobiano coat colours, respectfully, are found on the same chromosome and are linked. Knowing this, we can predict which coat colour genes are from which parents, and how those genes will be inherited in this horse's offspring.

(Wikimedia Commons-Kumana @ Wild Equines- CC BY 2.0)

As we learned in Chapter 3, Mendel reported that the pairs of loci he observed segregated independently of each other; for example, the segregation of seed color alleles was independent from the segregation of alleles for seed shape. This observation was the basis for his Second Law (Independent Assortment), and contributed greatly to our understanding of heredity as single units. However, further research showed that Mendel's Second Law did not apply to every pair of genes that could be studied. In fact, we now know that alleles of loci that are located close together on the same chromosome tend to be inherited together. This phenomenon is called **linkage**, and is a major exception to Mendel's **Second Law of Independent Assortment.** Researchers use linkage to determine the location of genes along chromosomes in a process called genetic mapping. The concept of gene linkage is important to the natural processes of heredity and evolution, as well as to our genetic manipulation of crops and livestock.

A Genetic Nomenclature & Symbols

Nomenclature and symbols have been covered in previous chapters. This will be a brief review to recover the topics.

A **gene** is a hereditary unit that occupies a specific position (**locus**) within the genome or chromosome and has one or more specific effects upon the phenotype of the organism and can mutate into various forms (**alleles**). A **genotype** is the specific allelic composition of a cell or organism. Normally only the genes under consideration are listed in a genotype and the alleles at the remaining gene loci are considered to be wild type. A **phenotype** is the detectable outward manifestation of a specific genotype. In describing a phenotype usually only the characteristics under consideration are listed while the remaining characters are assumed to be wild type (normal).

A.1 Gene names and symbols

Usually, gene names are unique and their corresponding symbols are unique letters or combinations of letters. So, for example, the "*vermillion*" gene in Drosophila is represented by the letter "*v*", while "*vg*" is the symbol for the "*vestigial*" gene and "*vvl*" is the symbol for the "*ventral veins lacking*" gene locus. Note however that the same letter symbols may represent a different gene in another organism. Gene symbols and gene names are typically shown in *italics* text. In lectures we may not always use italics for gene names and symbols.

The normal, or wild type, form of a gene is usually symbolized by superscript plus sign, "+". *e.g.* " a^+ ", " b^+ ", *etc*. or it is sometimes abbreviated to just "+". A forward slash is occasionally used to indicate that the two symbols are alleles of the same gene, but on homologous chromosomes (see **Figure 9-21**).

A typical mutant form of the gene, of which there can be many, can be symbolized by a superscript minus sign, "-". *e.g.* " a^- ", " b^- ", *etc*, or sometimes abbreviated to just "*a*", "*b*", *etc* (no superscript). Therefore if the genotype of a diploid organism is given as a^+/a^-, it means there is a wild type allele and mutant allele of the "*a*" gene at the "*a*" locus. This may also be abbreviated to *+/a*.

In some species of diploids, the dominant allele is typically designated with the upper case letter(s), while the recessive allele is given the lower case letter(s). For example, in Mendel's peas the dominant Rough allele is "*R*", while the recessive smooth allele is "*r*".

B Recombination

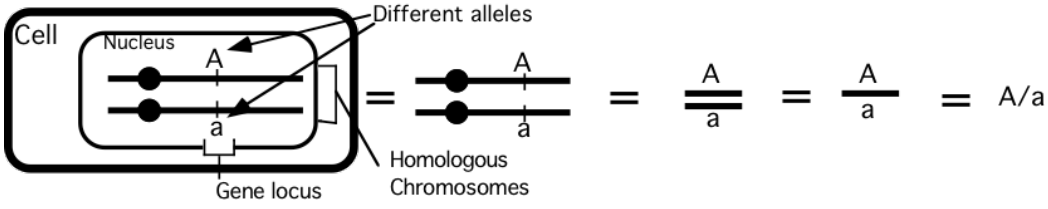


Figure 9-21

A diagram of how chromosomes, loci and alleles look in the cell, and how we depict them written.

(Original-J.Locke- CC BY-NC 3.0)

The process of meiosis leading to a separation of chromosomes, and crossing over is necessary for the understanding of this chapter. Refer to Chapter 7 for a review of these concepts.

The term "recombination" is used in several different contexts in genetics. In reference to heredity, **recombination** is defined as a process that results in gametes with combinations of alleles that were not present in the gametes from the parental generation (**Figure 9-22**). Recombination is important because it contributes to the genetic variation that may be observed between individuals within a population and that may be acted upon by selection for evolution.

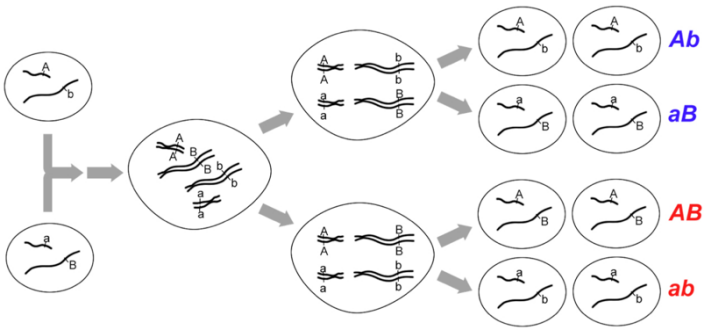


Figure 9-22

When two loci are on non-homologous chromosomes, their alleles will segregate in combinations identical to those present in the parental gametes (*A;b*, *a;B*), and in recombinant genotypes (*A;B*, *a;b*) that are different from the parental gametes.

(Original-Deyholos- CC BY-NC 3.0)

B.1 Inter- and intrachromosomal recombination

Interchromosomal recombination occurs either through **independent assortment** of alleles whose loci are on different chromosomes (Chapter 3). **Intrachromosomal recombination** occurs through **crossovers** between loci on the same chromosomes. It is important to remember that in both of these cases, recombination is a process that occurs during meiosis (mitotic recombination may also occur in some species, but it is relatively rare).

As an example of interchromosomal recombination, consider loci on two different chromosomes as shown in Figure 9-22.

We know that if these loci are on different chromosomes there is no physical connection between them, so they are **unlinked** and will segregate independently as did Mendel's traits. The segregation depends on the rel-

ative orientation of each pair of chromosomes at metaphase. Since the orientation is random and independent of other chromosomes, each of the arrangements (and their meiotic products) is equally possible for two unlinked loci as shown in Figure 9-22 on page 84.

Intrachromosomal recombination occurs through crossovers. Crossovers occur during prophase I of meiosis, when pairs of homologous chromosomes have aligned with each other in a process called **synapsis**. Crossing over begins with the breakage of DNA of a pair of non-sister chromatids. The breaks occur at corresponding positions on two non-sister chromatids, and then the ends of non-sister chromatids are connected to each other resulting in a reciprocal exchange of double-stranded DNA. Generally every pair of chromosomes has at least one crossover during meiosis, but often multiple crossovers occur in each chromatid during prophase I.

Because interchromosomal recombination happens through independent assortment, genes in this situation are always unlinked. Intrachromosomal recombination has instances of linked genes, and so they will be the focus of this chapter.

B.2 Inheriting Parental and Recombinant Gametes

If we consider only two loci and the products of meiosis results in recombination, then the meiotic products (gametes) are said to have a **recombinant genotype**. On the other hand, if no recombination occurs between the two loci during meiosis, then the products retain their original combinations and are said to have a non-recombinant, or **parental genotype.** The ability to properly identify parental and recombinant gametes is essential to apply recombination to experimental examples.

To properly identify recombinant and parental gametes, you need to know the genotype of the parents (the P generation) of the individual in question. This is most easily demonstrated in a dihybrid. If, for two genes, one parent has the genotype *AABB,* they can only produce one type of gamete: *AB* (**Figure 9-23** left). *Note: the absence of slashes and semicolons indicate that we don't know whether the genes are linked or not.* The same is true for the other parent. If it is *aabb*, then it can also only produce one type of gamete: *ab*

This is an encore presentation of the online exercise for writing out linked and unlinked genes.

http://tinyurl.com/oog-linked

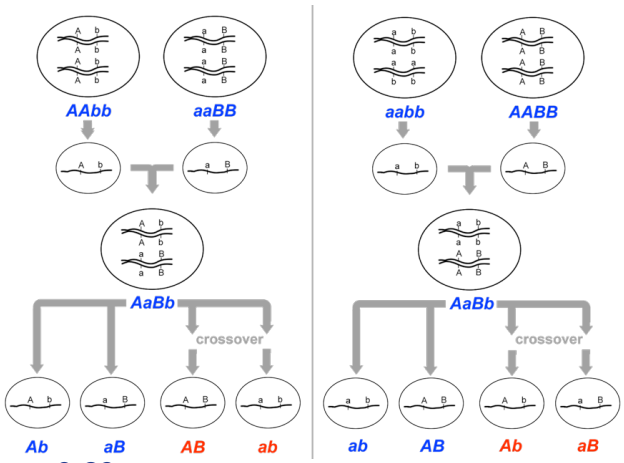


Figure 9-23

The genotype of gametes can be inferred unambiguously if the gametes are produced by homozygotes. However, recombination frequencies can only be measured among the progeny of heterozygotes (i.e. dihybrids). Note that the dihybrid on the left contains a different configuration of alleles (*Ab/aB*) than the dihybrid on the right (*AB/ab*) due to differences in the genotypes of their respective parents. Therefore, different gametes are defined as recombinant and parental among the progeny of the two dihybrids. In the cross at left, the recombinant gametes will be genotype *AB* and *ab*, and in the cross on the right, the recombinant gametes will be *Ab* and *aB*.

(Figure 9-23 right). When those two gametes combine, they create an individual (F_1) that has a genotype written as *AaBb*.

Only the F_1 dihybrid individual produces four different gametes: *AB*, *ab*, *Ab* and *aB*. The **parental gametes** are those that the F_1 obtained from its parents, in this case *AB* and *ab*. *Ab* and *aB* are **recombinant gametes** and are evidence of a recombination event happening, resulting in a different combination of alleles (Figure 9-23).

This idea also applies for independently assorting genes (which would be *A/a*; *B/b*). The frequency of each type of gamete from the dihybrid would be equal. Each would account for about ¼ each of the whole set of gametes. If the genes are linked (*ie.* on the same chromosome), the recombinants will be much less frequent than the parental arrangement.

For the above example, the P generation has one parent homozygous for both dominant alleles, and the other homozygous for both recessive alleles. It is very important to note that this will not always be the case. In some instances one parent will be homozygous with one gene dominant and the other gene recessive (*AAbb*) and the other parent will be the opposite (*aaBB*). This

situation will change which is the parental and recombinant gametes (compare left and right in Figure 9-23).

B.3 Coupling and Repulsion Configuration

When looking at an organism that is heterozygous at two loci, just by looking at them you cannot tell how the mutant and wild type alleles are arranged. Both mutant alleles could be on one homologous chromosome, and both wild type alleles could be on the other (e.g. *ab/a⁺b⁺).* This is known as a **coupling** (or cis) **configuration** (See **Figure 9-24**) When one wild type allele and one mutant allele are on one homologous chromosome, and the opposite is on the other, this is known as a **repulsion** (or trans) **configuration** (e.g. *a⁺b/ab⁺*). The way to determine the orientation is to look at the parents (or P generation) of that cross if you know the genotypes of them. If the parents are homozygous in both genes, and one shows both dominant phenotypes and the other shows both recessive phenotypes, then you know that the individual you are looking at is in coupling configuration. If one parent has one dominant and one recessive phenotype, and the other has the opposite, then you know the individual is in repulsion configuration.

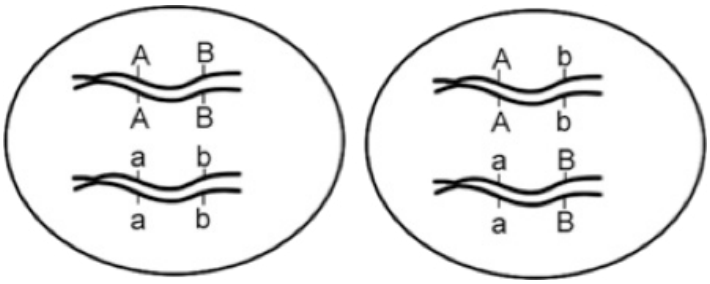


Figure 9-24
Alleles in coupling configuration (*AB/ab*,left) or repulsion configuration (*Ab/aB* right).
(Original-Deyholos-CC BY-NC 3.0)

B.4 Recombination Frequency

Recombination frequency represents distance between genes (**d**). It is a calculation to define the number of parental and recombinant gametes. The equation is as follows:

$$d = \frac{\#\ \text{Recombinant gametes}}{(\text{Total}\ \#\ \text{gametes})}\ X\ 100\%$$

The total number of gametes is, of course, the sum of both parental and recombinant gametes.

Through identifying and defining parental and recombinant gametes, you can calculate ***d*** and from there decide the degree of linkage.

Based upon the equation and independent assortment, you can see that the recombination frequency cannot be higher than 50%. The unit for this distance is more formally known as a centiMorgan (cM - note the capital M!). If alleles are assorting independently, there will a random distribution of the alleles in the progeny, and so 50% will be recombinant gametes and 50% will be parental gametes, making ***d*** approximately 50 cM. If a gene is linked you will see a higher percentage of parental gametes, making ***d*** < 50 cM. You will never see more recombinant gametes than parental, and so in no situation will recombination frequency be higher than 50 cM, except slightly with regards to standard experimental error. If you calculate a recombination frequency higher than 50 cM, you need to make sure you accurately defined parental and recombinant gametes.

C Unlinked Genes and Complete and Partial Linkage

When comparing any two genes, they can be varying distances apart. Their ***d*** allows us to categorize them into the degree of linkage. The amount of linkage can be placed on a sliding scale.

Table 9-2 shows generally how we categorize the degree linkage using recombination frequency. Because ***d*** is based upon experimental results that will have some experimental error, these should be treated as guidelines and not hard rules in determining the distance between genes.

C.1 Unlinked Genes

Unlinked genes are genes that appear to segregate and show independent assortment. There will be a ran-

Table 9-1
The linkage description is listed corresponding to its recombination frequency. Note: values between 0.30 and 0.50 may be partially linked, or many not be linked at all. It is often difficult to distinguish between these two possibilities because of experimental error.

Linkage Description	Distance
Unlinked	~50 cM (~50%)
Partial linkage	< 30 cM or 30%
Complete linkage	0.00 cM or 0%
You messed up!	>> 50 cM

dom and even distribution of gamete types, and a ***d*** of 50 cM is the expectation. This situation is describes two instances, either the genes are on completely different chromosomes, or they are far enough apart on a single chromosome that the crossovers are so numerous so as to randomly assort the alleles (Figure 9-22 on page 84). Either way, because the alleles are assorting independently you should see an equal number of recombinant and parental gametes, so ***d*** will be ~50 cM. Note, because of real life variability this value can be anywhere from ~40 cM to ~60 cM.

C.2 Complete Linkage

Having considered unlinked loci, let us turn to the opposite situation, in which two loci are so close together on a chromosome that the parental combinations of alleles always segregate together (**Figure 9-25**). This is because during meiosis they are so close that there are no crossover events between the two loci and the alleles at the two loci are physically attached on the same chromatid and so they always segregate together into the same gamete. In this case, no recombinants will be present following meiosis, and the recombination frequency will be 0.00. This is **complete** (or **absolute**) **linkage** and is rare, as the loci must be so close together that crossovers are never detected between them.

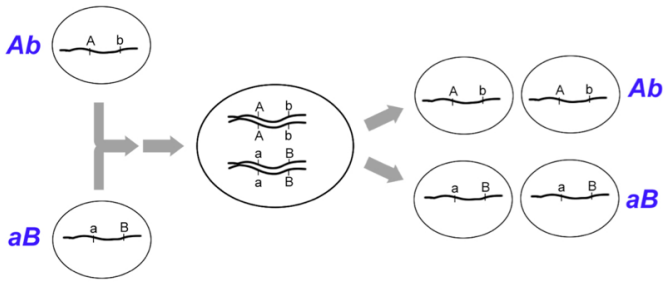


Figure 9-25

If two loci are completely linked, their alleles will segregate in combinations identical to those present in the parental gametes (*Ab*, *aB*). No recombinants will be observed.

(Original-Deyholos-CC BY-NC 3.0)

C.3 Partial Linkage

It is also possible to obtain recombination frequencies between 0% and 50%, which is a situation we call **incomplete** (or **partial**) **linkage**. Incomplete linkage occurs when two loci are located on the same chromosome but the loci are far enough apart so that crossovers occur between them during some, but not all, meioses (**Figure 9-26**). Genes that are on the same chromosome are said to be **syntenic** regardless of whether they are completely or incompletely linked or unlinked. All

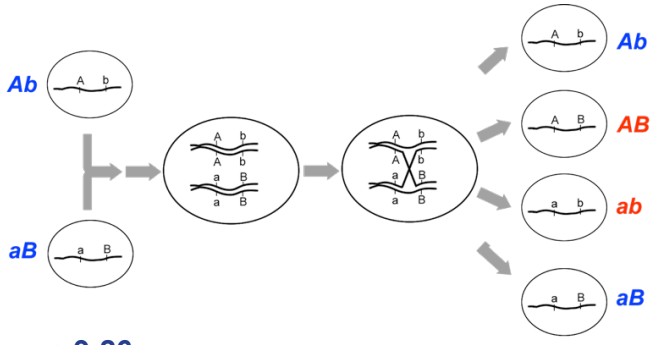


Figure 9-26

A crossover between two linked loci can generate recombinant genotypes (*AB*, *ab*), from the chromatids involved in the crossover. Remember that multiple, independent meioses occur in each organism, so this particular pattern of recombination will not be observed among all the meioses from this individual.

(Original-Deyholos-CC BY-NC 3.0)

linked genes are syntenic, but not all syntenic genes are linked (they may be so far apart that the proportion of gametes resemble independent assortment).

Because the location of crossovers is essentially random along the chromosome, the greater the distance between two loci, the more likely a crossover will occur between them. Furthermore, loci that are on the same chromosome, but are sufficiently far apart from each other, will on average have multiple crossovers between them and they will behave as though they are completely unlinked. A recombination frequency of 50% (***d*** = 50 cM) is therefore the maximum recombination frequency that can be observed, and is indicative of loci that are either on separate chromosomes, or are located very far apart on the same chromosome (*ie* syntenic, but not linked!)

D Experimentally Determining Recombination Frequency

Let us now consider a complete experiment in which our objective is to measure recombination frequency (**Figure 9-27**). We need at least two alleles for each of two genes, and we must know which combinations of alleles were present in the parental gametes. The simplest way

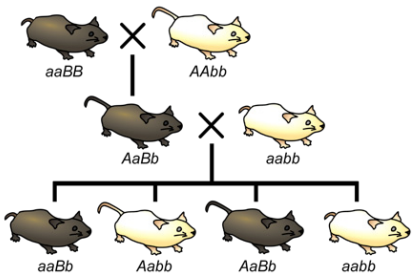


Figure 9-27

An experiment to measure recombination frequency between two loci. The loci affect coat color (B/b) and tail length (A/a).
(Wikipedia Mod by Deyholos-CC BY-NC 3.0)

to do this is to start with pure-breeding lines that have contrasting alleles at two loci. For example, we could cross short-tailed mice, brown mice (*aaBB*) with long-tailed, white mice (*AAbb*) (Figure 8, P cross). Based on the genotypes of the parents, we know that the parental gametes will be *aB* or *Ab* (but not *ab* or *AB*), and all of the progeny will be dihybrids, *AaBb*. We do not know at this point whether the two loci are on different chromosomes, or whether they are on the same chromosome, and if so, how close together they are. That is why the genotype in this paragraph does not contain slashes or semicolons.

The recombination events that may be detected will occur during meiosis in the dihybrid individual. If the loci are completely or partially linked, then prior to meiosis, alleles *aB* will be located on one chromosome, and alleles *Ab* will be on the other chromosome. These are the parental gametes based on our knowledge of the genotypes of the gametes that produced the dihybrid. Thus, recombinant gametes produced by the dihybrid will have the genotypes *ab* or *AB*.

Now that we have identified the parental and recombinant gametes, how do we determine the genotype of the gametes produced by the dihybrid individual? The most practical method is to use a testcross (Figure 8 shows F_1 crossed with a tester), in other words to mate *AaBb* to an individual that has only recessive alleles at both loci (*aabb*). This will give a different phenotype in the second generation for each of the four possible combinations of alleles in the gametes of the dihybrid (**Figure 9-28**).

We can then infer unambiguously the genotype of the gametes produced by the dihybrid individual, and therefore calculate the recombination frequency between these two loci. For example, if only two phenotypic classes were observed in the F_2 (i.e. short tails and brown fur (*aaBb*), and white fur with long tails (*Aabb*))

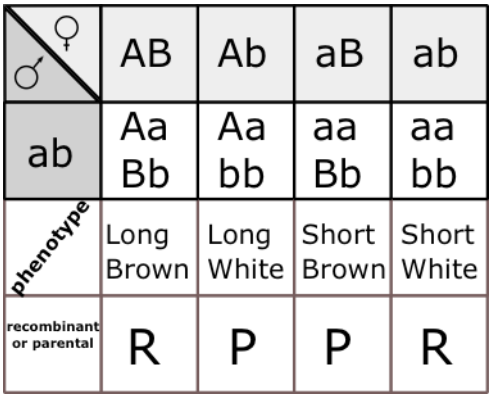


Figure 9-28

An experiment to measure recombination frequency between two loci. The loci affect coat color (*B/b*) and tail length (*A/a*).
(Wikipedia-Modified Deyholos-CC BY-NC 3.0)

we would know that the only gametes produced following meiosis of the dihybrid individual were of the parental type: *aB* and *Ab*, and the recombination frequency would therefore be 0%. Alternatively, we may observe multiple classes of phenotypes in the F_2 in ratios such as shown in **Table 9-2**. Given the data in this table, the calculation of recombination frequency is straightforward:

$$d = \frac{\text{\# recombinant offspring}}{\text{Total offspring}} \times 100$$

$$d = \frac{13+17}{48+42+13+17} \times 100\ cM$$

$$= 25\ cM$$

Because the recombination frequency is below 30%, we can say that the tail length gene and the fur colour gene are partially linked.

Note: The use of linkage and recombination frequency, will be extended to Genetic Mapping in the next chapter.

Table 9-2

An example of quantitative data that may be observed in a genetic mapping experiment involving two loci. The data correspond to the F_2 generation in the cross shown in Figure 9-27 on page 87.

tail phenotype	fur phenotype	number of progeny	gamete from dihybrid	genotype of F_2 from test cross	(P)arental or (R)ecombinant
short	brown	48	*aB*	*aaBb*	P
long	white	42	*Ab*	*Aabb*	P
short	white	13	*ab*	*aabb*	R
long	brown	17	*AB*	*AaBb*	R

Summary:

- Recombination is defined as any process that results in gametes with combinations of alleles that were not present in the gametes of a previous generation.
- The recombination frequency between any two loci depends on their relative chromosomal locations.
- Unlinked loci show a maximum 50% recombination frequency (a distance of 50 cM).
- Loci that are close together on a chromosome are linked and tend to segregate with the same combinations of alleles that were present in their parents.
- Crossovers are a normal part of most meioses, and allow for recombination between linked loci.
- Measuring recombination frequency is easiest when starting with pure-breeding lines with two alleles for each locus, and with suitable lines for test crossing.

Key terms:

allele
complete (absolute) linkage
coupling (cis) configuration
crossover
distance
    (d = #recomb/#total*100)
gene
genotype
incomplete (partial) linkage
independent assortment
interchromosomal recombination
intrachromosomal recombination
linkage
locus
parental genotype (and gametes)
phenotype
recombinant genotype (and gametes)
recombination
recombination frequency
    (see *distance*)
repulsion (trans) configuration
Second Law of
    Independent Assortment
synapsis
syntenic
unlinked

Study Questions:

1. Compare recombination and crossover. How are these similar? How are they different?
2. Explain why it usually necessary to start with pure-breeding lines when measuring genetic linkage by the methods presented in this chapter.
3. If you knew that a locus that affected earlobe shape was tightly linked to a locus that affected susceptibility to cardiovascular disease human, under what circumstances would this information be clinically useful?
4. In a previous chapter, we said a 9:3:3:1 phenotypic ratio was expected among the progeny of a dihybrid cross, in absence of gene interaction.
    a) What does this ratio assume about the linkage between the two loci in the dihybrid cross?
    b) What ratio would be expected if the loci were completely linked? Be sure to consider every possible configuration of alleles in the dihybrids.
5. Given a dihybrid with the genotype *CcEe*:
    a) If the alleles are in coupling (cis) configuration, what will be the genotypes of the parental and recombinant progeny from a test cross?
    b) If the alleles are in repulsion (trans) configuration, what will be the genotypes of the parental and recombinant progeny from a test cross?
6. In this question, white flowers (w) are recessive to purple flowers (W), and yellow seeds (y) are recessive to green seeds (Y). If a green-seeded, purple-flowered dihybrid is testcrossed, and half of the progeny have yellow seeds:
    a) What can you conclude about linkage between these loci?
    b) What do you need to know about the progeny in this case?

7. If the progeny of the cross *aaBB* x *AAbb* is testcrossed, and the following genotypes are observed among the progeny of the testcross, what is the frequency of recombination between these loci?

   a) *AaBb* 135
   b) *Aabb* 430
   c) *aaBb* 390
   d) *aabb* 120

8. Draw the cell in metaphase I of the cell cycle in a case where crossover did not occur in prophase I, and in a case where it did.

9. What is meant by the sentence "All linked genes are syntenic, but not all syntenic genes are linked."?

Chapter 10
Eukaryotic Gene Mapping

Figure 10-1
Two chromosomes with the same gene Thomas Hunt Morgan and his undergraduate Alfred Henry Sturtevant used fruit fly mutations like the ones in this figure to create the first recombination map.

Eye colors (clockwise): brown, cinnabar, sepia, vermilion, white, wild. Also, the white-eyed fly has a yellow body, the sepia-eyed fly has a black body, and the brown-eyed fly has an ebony body.

(Wikimedia-Ktbn-Public Domain)

Introduction

In previous chapters the relative location of two loci has been examined. We have used the frequency of **recombinants** vs **parentals** to determine the recombinant frequency (RF). Two loci could show **independent assortment** (**unlinked**, RF~ 0.5 or 50%) or were **linked** (RF<< 0.35 or 35%). If linked, the two must be located on the same chromosome (**syntenic**), but if unlinked they could be far apart on the same chromosome or on different chromosomes (**non-syntenic**). In this chapter we will learn how to construct genetic maps using 3-point crosses.

A. Genetic Mapping

A **genetic map** (or recombination map) is a representation of the linear order of genes (or loci), and their relative distances determined by crossover frequency, along a chromosome. The fact that such linear maps can be constructed supports the concept of genes being arranged in a fixed, linear order along a single duplex of DNA for each chromosome. We can use recombination frequencies to produce genetic maps of all the loci along each chromosome and ultimately in the whole genome.

A.1. Calculating map distance

The units of genetic distance are called **centiMorgans** (**cM**) in honor of **Thomas Hunt Morgan** by his undergraduate student, **Alfred Sturtevant**, who developed the concept of genetic maps. Thus, the recombination frequency in percent is the same as the map distance in cM. For example, if two loci have a recombination frequency of 25% they are said to be ~25cM apart on a chromosome (**Figure 10-2 on page 92**).

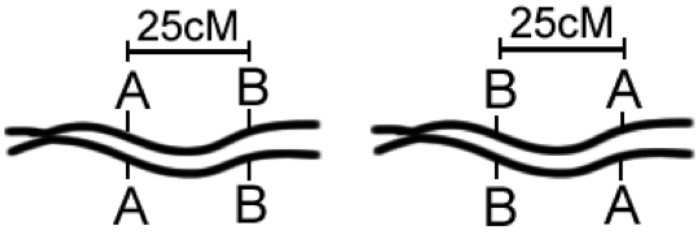


Figure 10-2

Two genetic maps consistent with a recombination frequency of 25% between A and B. Note the location of the centromere.

(Original-Deyholos-CC BY-NC 3.0)

Note, however, this approximation works well only for small distances (***d*** < 30 cM) but progressively fails at longer distances. This is because as the two loci get farther apart the RF reaches a maximum at 50%, like it would for two loci assorting independently (not linked). In fact, most chromosomes are >100 cM long but such loci at the tips only have an RF of 50%.

You might think that with a large enough distance, you would *guarantee* that a crossover would occur between them, thus making the number of recombinants larger than the parentals. What you'd be missing is that you can have second, third, and even more crossovers that will restore (even # crossovers) or remove (odd # chromosomes) parental arrangements. Each subsequent crossover is less probable, thus causing a plateau of measurable recombinants at 50%. See section A.3 for more details.

Calculating the map distance of the whole chromosome (end-to-end) of over 50cM comes from mapping of multiple loci dispersed along the chromosome, each with a value of less than 50%, with their total adding up to the value over 50cM (e.g. >100 cM as above). The method for mapping of these long chromosomes is described next.

Note that the map distance of two loci alone does not tell us anything about the orientation of these loci relative to other features, such as centromeres or telomeres, on the chromosome.

A.2. Map distance over long chromosomes

Map distances are always calculated for one pair of loci at a time. However, by combining the results of multiple pairwise calculations, a **genetic map** of many loci on a chromosome can be produced (**Figure 10-3**). A genetic map shows the map distance, in cM, that separates any two loci, and the position of these loci relative to all other mapped loci. The genetic map distance is roughly proportional to the physical distance, i.e. the amount of DNA between two loci. For example, in *Arabidopsis*, 1.0 cM corresponds to approximately 150,000bp and contains approximately 50 genes. The exact number of DNA base pairs in a cM depends on the organism, and on the particular position in the chromosome; some parts of chromosomes ("crossover hot spots") have higher rates of recombination than others, while other regions have reduced crossing over and often correspond to large regions of heterochromatin.

When a novel gene or locus is identified by mutation or polymorphism, crossing it with previously mapped

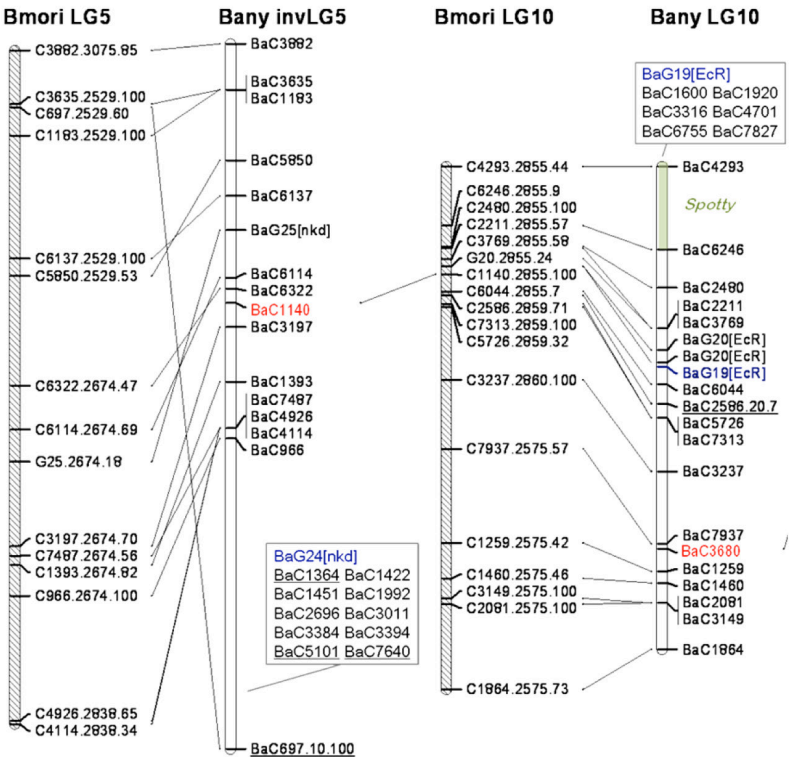


Figure 10-3

Genetic maps for regions of two chromosomes from two species of the moth, *Bombyx*. The scale at left shows distance in cM, and the position of various loci is indicated on each chromosome. Diagonal lines connecting loci on different chromosomes show the position of corresponding loci in different species. This is referred to as regions of **conserved synteny**.

(NCBI-NIH-PD)

genes and then calculating the recombination frequency can determine its approximate position on a chromosome.

If the novel gene and the previously mapped genes show complete or partial linkage with an existing locus, the recombination frequency will indicate the approximate position of the novel gene within the genetic map. This information is useful in isolating (i.e. cloning) the specific fragment of DNA that encodes the novel gene. This process called **map-based cloning**.

Genetic maps are also useful to (1) track genes/alleles when breeding crops and animals, (2) in studying evolutionary relationships between species, and (3) in determining the causes and individual susceptibility of some human diseases.

A.3. Genetic maps are an approximation

Genetic maps are useful for showing the order of loci along a chromosome, but the distances are only an approximation. The correlation between recombination frequency and actual chromosomal distance is more accurate for shorter distances (low RF values) than long distances. As the distance between two genes you are mapping increases, the actual number of crossovers is underestimated. As mentioned near the end of section A.1, this is because as the distance between loci increases, so does the possibility of having a second (third, or more) crossovers occur between the loci. This is a problem for geneticists, because with respect to the loci being studied, these **double-crossovers** produce gametes with the same genotypes as if no recombination events had occurred (**Figure 10-4**), so they have parental genotypes. Thus a double crossover will appear to be a parental type and not be counted as a recombinant, despite having two (or more) crossovers. Geneticists will sometimes use specific mathematical formulae to adjust large recombination frequencies to account for the possibility of multiple crossovers and thus get a better estimate of the actual distance between two loci.

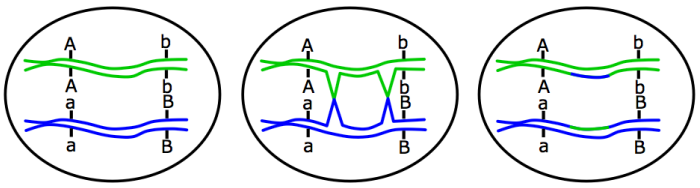


Figure 10-4

A double crossover between two loci will appear as a non-crossover in the progeny because the chromosomes contain parental arrangements of the genes being analyzed.

B. Mapping With Three-Point Crosses

B.1. Solving Gene Order Without Math

It is critical to get both the parental genotypes and the gene order correct; many students muff the mapping if they assume the genes are in alphabetical order or the order printed on the page.

The largest two classes of offspring represent the parentals; absence of crossover is the most common event. If you look ahead to **Table 1** (*Section B.2. An Example*), note that rows 2 and 5 have the most common offspring (38 and 42, respectively). Because this is a testcross, the phenotype will resemble whatever the F_1 trihybrid created through recombination (or not!). Parentals are most common because recombination is rare. The two largest classes indicate the genotype of the P generation, which is often true-breeding. In **Figure 5**, you are shown what the parents look like. (This is not always going to be done for you.) If you have a table with all the progeny and their frequencies, you have the tools to elucidate the parental gene combinations yourself. Try it out on **Table 1**. You should notice that the largest numbers reflected the original parents shown in **Figure 5**. In rows 2 and 5, alleles *A* and *C* (and also *a* and *c*) are in coupling. Likewise, *A* and *b* (also *a* and *B*) as well as *B* and *c* (and *b* and *C*) are in repulsion. Thus must it be in the parentals: *AbC/AbC* x *aBc/aBc*!

Next, you need to verify the gene order. While the strategy of mapping all three genes and figuring out the order works well (see section 2.2), for some reason this seems to not work out well for many students under exam conditions.

Let's start with an easy example where all the genes are in coupling (but note that is not always the case, as in section 2.2 and the example above). The largest classes of offspring reveal the parental genotypes (as you saw), but the *smallest* classes of offspring will tell us the order of the genes. Because double crossovers are the most rare, they have to be the smallest classes. In **Table 1**, these are in rows 3 and 8 (having only one member each).

Let's take a step back and use a simple example here. We'll use the superscript notation and start with two true-breeding parents with three linked genes: a^+, b^+ and c^+, all in coupling, and in that order (this would be written as $a^+b^+c^+/a^+b^+c^+$ x *abc/abc*). Let's put them on

homologous chromosomes that are aligned as shown:

$$\frac{a^+b^+c^+}{a^+b^+c^+} \times \frac{a\ b\ c}{a\ b\ c}$$

The resulting F_1 trihybrid would therefore look like:

$$\frac{a^+b^+c^+}{a\ b\ c}$$

A double crossover is identified when one crossover occurs between *a* and *b* as well as *b* and *c*. There are two kinds of gametes produced this way:

$\frac{a^+b^+c^+}{a\ b\ c}$ $\quad$ $\frac{a^+b^+c^+}{a\ b\ c}$

The double crossover gametes created will be ab^+c or a^+bc^+.

If we do a testcross, one chromosome of each offspring *must* contain all recessive alleles. It would look like this (the homozygous tester's chromosome is light blue):

$$\frac{a^+\ b\ c^+}{a\ b\ c} \text{ or } \frac{a\ b^+\ c}{a\ b\ c}$$

Phenotypically, the offspring would be a^+bc^+ *and* ab^+c. If that matches the phenotype of your least-populated rows, you have the order!

Try doing this with a different gene order that is wrong:

$$\frac{a^+c^+b^+}{a^+c^+b^+} \times \frac{a\ c\ b}{a\ c\ b} \rightarrow \frac{a^+c^+b^+}{a\ c\ b}$$

$\frac{a^+c^+b^+}{a\ c\ b}$ $\quad$ $\frac{a^+c^+b^+}{a\ c\ b}$

Note the outcome for this gene order would be different. Double crossovers would be ac^+b *and* a^+cb^+. These are different than what we found above. Only three orders are possible – try each gene in turn as the "middle". Only one can fit your data.

If the class with the lowest number of progeny resembles the phenotype you created using the parental genotypes match, you have the correct order. If it doesn't work, try a different order. There are only three possibilities.

Check http://universitygenetics-mapping.blogspot.ca/2012/02/determining-gene-order.html for a tutorial on this.

B.2. An Example

A genetic map consists of multiple loci distributed along a chromosome. An efficient method of mapping three genes at once is the **three-point testcross**, which allows the order and distance between three potentially linked genes to be determined in a single cross experiment ().

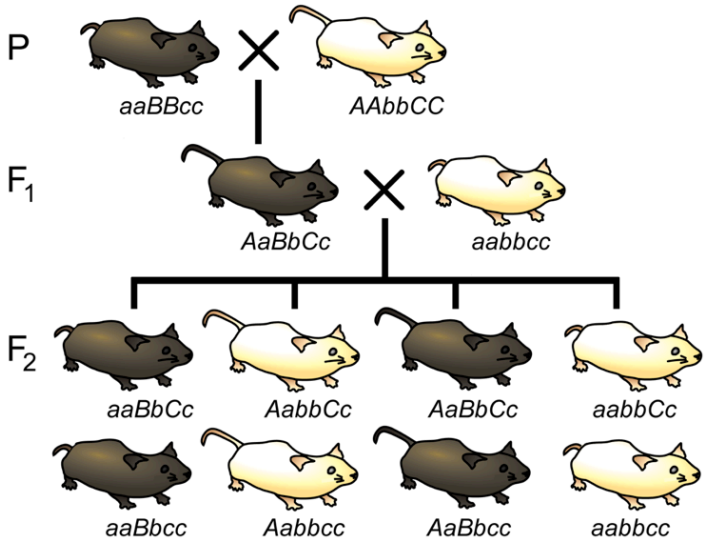


Figure 10-5
A three point cross for loci affecting tail length, fur color, and whisker length.
(Original-Modified Deyholos-CC BY-NC 3.0)

This is particularly useful when mapping a new mutation whose location is unknown relative to two previously mapped loci with known locations. The basic strategy is the same as for the dihybrid mapping experiment described previously, except pure breeding lines with contrasting genotypes are crossed to produce an individual heterozygous at three loci (a trihybrid), which is then testcrossed to a tester, which is homozygous recessive for all three genes, to determine the recombination frequency between each pair of genes, among the three loci. A Punnett square can be

♀ / ♂	aBC	AbC	abC	ABC	aBc	Abc	abc	ABc
abc	aa Bb Cc	Aa bb Cc	aa bb Cc	Aa Bb Cc	aa Bb cc	Aa bb cc	aa bb cc	Aa Bb cc
phenotype	Short tail Brown Long whis	Long tail White Long whis	Short tail White Long whis	Long tail Brown Long whis	Short tail Brown Short whis	Long tail White Short whis	Short tail White Short whis	Long tail Brown Short whis

Figure 10-6
Punnett square of the test cross for Figure 5 showing the predicted gametes possible from this cross, and their phenotypes.
(Original-L. Canham-CC BY-NC 3.0)

used to predict all the possible outcomes of the test cross (**Figure 10-6**). The progeny produced from the testcross is shown in **Table 10-1 on page 95**.

Table 10-1

An example of data that might be obtained from the F_2 generation of the three-point cross is shown in Figure 5. The rarest phenotypic classes correspond to double recombinant gametes *ABc* and *abC*. Each phenotypic class and the gamete from the trihybrid that produced it can also be classified as parental (P) or recombinant (R) with respect to each pair of loci (*A*,*B*), (*A*,*C*), (*B*,*C*) analyzed in the experiment.

	tail phenotype	fur phenotype	whisker phenotype	# of progeny	gamete from trihybrid	genotype of F_2 from test cross	loci A, B	loci A, C	loci B, C
1	Short tail	Brown fur	Long whiskers	5	*aBC*	*aBC/abc*	P	R	R
2	Long tail	White fur	Long whiskers	38	*AbC*	*AbC/abc*	P	P	P
3	Short tail	White fur	Long whiskers	1	*abC*	*abC/abc*	R	R	P
4	Long tail	Brown fur	Long whiskers	16	*ABC*	*ABC/abc*	R	P	R
5	Short tail	Brown fur	Short whiskers	42	*aBc*	*aBc/abc*	P	P	P
6	Long tail	White fur	Short whiskers	5	*Abc*	*Abc/abc*	P	R	R
7	Short tail	White fur	Short whiskers	12	*abc*	*abc/abc*	R	P	R
8	Long tail	Brown fur	Short whiskers	1	*ABc*	*ABc/abc*	R	R	P

The trihybrid, when crossed to a tester, should be able to make eight different gametes, to make eight possible different phenotype combinations in the offspring (2^n combinations, n = #traits). The next step would be to identify if the alleles are recombinant or parental gametes, comparing only two loci at one time to the parental combinations. In this example, the parents of the trihybrid are *aBc/aBc* and *AbC/AbC*, so the parental gametes would be *aBc* and *AbC* respectively. Now, by comparing two loci at once, you can determine if they are recombinant or parental. For example, the offspring in the first row in Table 10-1 came from gamete *aBC*. Comparing loci *A* and *B* we see that it matches one of the parental gametes and so it is parental. Comparing *A* and *C* we see that it matches neither parental, so it is recombinant. The same can be said for comparing *B* and *C*.

Once each class of progeny has been identified as being parental or recombinant for each gene (right three columns of Table 10-1), recombination frequencies may be calculated for each pair of loci individually, as we did before for one pair of loci in our dihybrid We can then use these numbers to build the map, placing the loci with the largest RF on the ends.

$$RF_{A,B} = \frac{5+16+5+1}{120} = 5\%$$

$$RF_{A,C} = \frac{5+1+5+1}{120} = 15\%$$

$$RF_{B,C} = \frac{5+16+5+12}{120} = 32\%$$

(not corrected for double crossovers)

B.3. Double Crossover Correction

Note that in the three-point cross, the sum of the distances between A-B and A-C (35%) is less than the distance calculated for B-C (32%). This is because of **double crossovers** between B and C, which were undetected when we considered only pairwise data for B and C. We can easily account for some of these double crossovers, and include them in calculating the map distance between B and C, as follows (**Figure 10-7**).

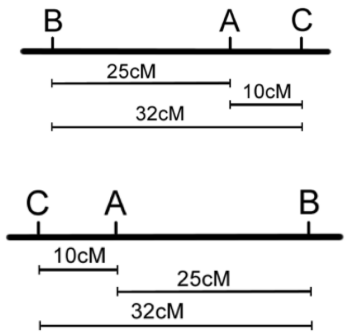


Figure 10-7

Two maps based on the data in Table 1 (without correction for double cross-overs). Both are equivalent.

We already deduced that the map order must be *BAC* (or *CAB*). However, these double recomb-inants, *ABc* and *abC,* were not included in our calculations of recombination frequency between loci *B* and *C*. If we included these double recombinant classes (multiplied by 2, since they each represent two recombination events), the calculation of recombination frequency between *B* and *C* is as follows, and the result is now more consistent with the sum of map distances between *A*-*B* and *A*-*C*.

$$RF_{B,C} = \frac{5+16+5+12+2(1)+2(1)}{120} = 35\%$$

(corrected for double crossovers)

Thus, the three-point cross was useful for:
(1) determining the order of three loci relative to each other,
(2) calculating map distances between the loci, and
(3) detecting some of the double crossover events that would otherwise lead to an underestimation of map distance.

However, it is possible that other double crossovers events remain undetected. Double crossovers between loci *A&B* or between loci *A&C* cannot be detected. Geneticists have developed a variety of mathematical procedures to try to correct for such double crossovers during large-scale mapping experiments. This, by the way, is also why the recombination frequency caps out at 50% as explained earlier in Chapter 9, sections B&C.

As more and more genes are mapped a better genetic map can be constructed. Then, when a new gene is discovered, it can be mapped relative to other genes of known location to determine its location. All that is needed to map a gene is two alleles: a wild type allele, and a mutant allele.

C. Analysis of Recombination Frequencies in a Three Point Test Cross

Now that we know what the map looks like, the frequency of each offspring type can be explained. Parental gametes (*AbC* and *aBc*) are the result of no crossovers, or double crossovers between two alleles. Because we know all three loci are linked, it is expected for this frequency to be relatively high, much like what we see in the example above.

There are recombinant gametes that are the result of one crossover between two alleles (*aBC*, *Abc*, *ABC* and *abc*) single crossover events are more common, but are more likely to happen between loci *B* and *A*, because they are 25 cM and so are farther apart than *A* and *C*, which are only 10 cM. So we expect to see more recombinant gametes with the former.

And lastly there are recombinant gametes that are a result of double crossover events (*ABc* and *abC*). Double crossovers between three linked genes like this is rare, so we don't expect to see many offspring from these recombinant gametes.

In the example given above, all the genes present are linked, with one pair more strongly linked than the other (A and C have stronger linkage than A and B). When choosing three genes to map, this will not always be the case. Sometimes you will have all genes linked, sometimes you may have two genes linked and one gene unlinked, and sometimes they all may be unlinked

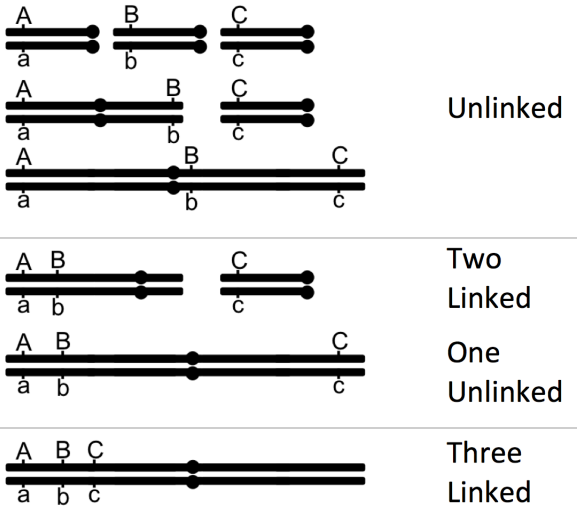
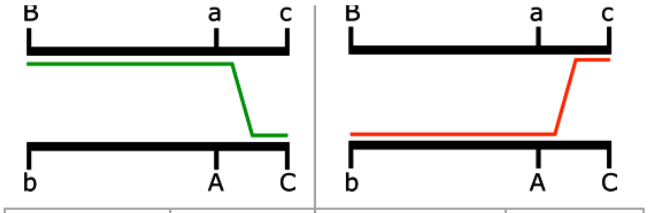


Figure 10-8
Three ways genes can be associated with each other based. All have the same genotype but linkage changes arrangements. Genes far apart are considered "unlinked".
(Original-J. Locke/L. Canham-CC BY-NC 3.0)

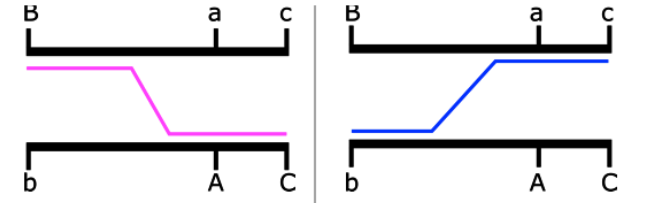
(**Figure 10-8 on page 96**). Much like what we did above, by comparing the ratios of offspring you should be able to predict if the genes in the trihybrid are linked or not.

The frequencies we see from this cross agree with our expectations. **Figure 10-9** (next page) shows a diagram of the crossover events that took place in regards to recombinant gametes and the number of offspring seen with that gamete type.

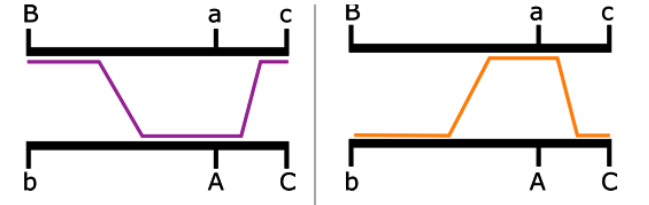
If all three genes are unlinked, then we expect independent assortment and an equal number of all progeny types. Like in the example, if all are linked,



gamete	# of progeny	gamete	# of progeny
aBC	5	Abc	5



gamete	# of progeny	gamete	# of progeny
ABC	16	abc	12



gamete	# of progeny	gamete	# of progeny
ABc	1	abC	1

Figure 10-9
Diagram of the crossover events to create the different recombinant gametes from the cross in **Figure 10-5 on page 94**. The parental alleles are seen on the black chromosomes. The coloured lines indicate show where the crossover event took place and underlines the alleles for that recombinant gamete. Below each diagram is the recombinant gamete and the number of progeny seen in that cross.

you expect there to be many parental genotypes, some recombinant genotypes if they are a result of a single recombination events. Recombinant genotypes that are a result of two recombination events will be rare. The actual numbers of each will differ depending if all the linked genes are equal distances from each other, or if one pair is more linked than the other. We used the largest and smallest classes to figure out what's happening earlier in this chapter.

In the case of having two genes linked and one gene unlinked the data will look different. Let's use the example as it was before, with the same parental gametes (*AC;b* and *ac;B*), but let's assume the genes *A* and *C* are linked and *B* is unlinked (as shown using the semicolon). In this case, because linkage causes a higher prevalence of parental gametes, we expect there to be more parental organizations of *A* and *C* (*ie. AC* or *ac*), and less recombinant organizations of *A* and *C* (*ie. Ac* and *aC*). The presence and or absence of parental *B* is not important here, because it is unlinked and will assort independently. Your data will show an equal number of *B* or *b* for each *AC* combination. If *A* and *C* are still 10 cM apart, your data may look like that shown in **Table 10-2**.

Table 10-2
Progeny ratios seen after a trihybrid test cross depending on whether they are all linked, only two are linked or if all are unlinked. The "All Linked" column is based upon the cross shown in **Figure 10-5 on page 94** as an example.

Gametes	Parental or Recombinant	A and C linked	All linked
AbC	P	more	many
aBc	P	more	many
ABC	R	more	some (big)
abc	R	more	some (big)
aBC	R	less	some (few)
Abc	R	less	some (few)
abC	R	less	rare
ABc	R	less	rare

Note how there are FOUR large classes and FOUR small classes instead of two (with two intermediates). This is your clue that there might be an unlinked gene.

D. Where Do Crossovers Occur on a Chromosome?

A crossover involves the reciprocal exchange between non-sister chromatids when synapsed at prophase I of meiosis. While this exchange can theoretically occur anywhere along the synapsed homologs, observations show us that some regions along a chromosome have higher rates of crossing over, while others are lower. In addition, the frequency of crossing over varies from species to species, and even from male to female within a species. For example, in *Drosophila melanogaster* there is no crossing over in males.

From *Drosophila* recombination data we know that the likelihood of a crossover is greatest in the middle of a chromosome arm and lower at the telomere and

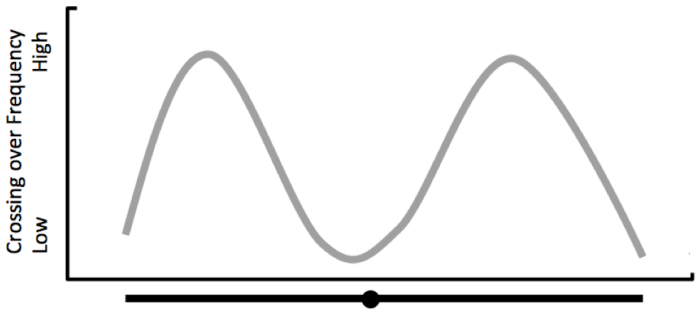


Figure 10-10

Diagram of the frequency of crossing over along a chromosome (bottom). The Y-axis shows the relative rate of crossing over. The two peaks are present in the middle of each chromosome arm, while the telomeres and centromeres have lower frequencies of exchanges.

(Original-J. Locke-CC BY-NC 3.0)

centromere regions (**Figure 10-10 on page 98**). This distribution would be expected if one of the functions of a crossover event were to hold the two synapsed chromosomes together so that they segregate correctly in metaphase I of Meiosis I.

E. Interference

Now we come to one last thing that we can determine from a three-point testcross: whether the incidence of one crossover influences the occurrence of a nearby one.

The product rule tells us that the probability of a double crossover is easily determined by multiplying the distances together. In Figure 10-7 on page 95, *B* and *A* are 25 cM apart and *A* and *C* are 10 cM apart. Each distance reflects the percentage of recombinants between the regions: 25% and 10%, respectively. The chance of getting a double crossover between both loci is 0.25 x 0.10 = 0.025, or 2.5%. For every 100 gametes, you expect 2-3 to be a result because of double crossover.

But recall that crossover is a physical event between chromosomes. Chapter 7 (Figure 7-4 on page 66) shows how homologous chromosomes approach each other (synapse) during the zygotene part of Prophase I. Physical connections between homologues occur during the diplotene phase, and chiasmata ("bridges") form between them. Figure 7-6 on page 67 in Chapter 7 shows that the chromosomes are bound together with the synaptonemal complex. What's not shown is that there are protein complexes at the site of chiasma that cause extremely precise cutting and splicing of DNA between nonsister chromatids. These cuts and splices are so precise that the genetic sequence exactly matches *atom to atom* at the crossover region.

This relates to crossover for this reason: connecting nonsister chromatids of homologous chromosomes involves enzymes and cellular tools that occupy space. This means that the machinery directing one crossover might influence an adjacent event. This is called **interference**. Positive interference means that the number of double crossover events is less than you expect: you have fewer than the number of predicted double crossover progeny. Perhaps the enzymes involved in one crossover event prevent another nearby event from happening.

Less common is negative interference. The presence of one crossover might actually increase another nearby event.

The exact molecular reasons for interference are still being worked out. You can, however, calculate the amount of interference with a simple formula:

$$I = 100\% - \frac{\#observed}{\#expected}$$

The number of observed double crossovers (*dco*) is simply all the double crossover progeny you found in your data. The expected number is calculated by multiplying the total number of progeny by the probability of a double crossover occurring. In the earlier example, we found that 2.5% of progeny should be double crossovers. In Table 1 there are 120 progeny in total. Therefore, you expect 0.025 x 120 = 3 double crossovers. You observed 2.

Ignoring that the number of progeny is very, very small and so the observed number is probably not statistically accurate, we could calculate interference as $I = 100\% - (2/3) = 33.3\%$. This means that 1/3 of expected double crossovers were interfered with: they did not occur.

Summary:

- A **genetic map** (or recombination map) is a representation of the linear order of genes (or loci), and their relative distances determined by crossover frequency, along a chromosome.
- Recombination frequency is usually proportional to the distance between loci, so recombination frequencies can be used to create genetic maps.
- Recombination frequencies tend to underestimate map distances, especially over long distances, since double crossovers may be indistinguishable from non-recombinants.
- Loci that are syntenic may appear to be unlinked. However, no genes can be mapped as farther than 50 cM apart.
- Three-point crosses can determine the order and map distance among three loci.
- In three-point crosses, a correction for the distance of the outside markers can be made to account for double crossovers between the two outer loci.
- Crossovers are not equally frequent all along a chromosome. In some regions, crossovers are more frequent while others are less.
- The resolution of genetic maps depends on the number of markers and the number of progeny.
- Interference between crossover events may occur. This can be calculated. Positive interference is when less crossovers occur than expected. Negative interference is when more crossovers appear than anticipated.

Key terms:

centimorgans (cM)
conserved synteny
double-crossover
genetic map
independent assortment
interference
linked
map units (mu)
map-based cloning
Morgan, Thomas Hunt
non-syntenic
parentals
recombinants
Sturtevant, Alfred
syntenic
synapse
three-point cross
unlinked

Study Questions:

1) In corn (i.e. maize, a diploid species), imagine that alleles for resistance to a particular pathogen are recessive and are linked to a locus that affects tassel length (short tassels are recessive to long tassels). Design a series of crosses to determine the map distance between these two loci. You can start with any genotypes you want, but be sure to specify the phenotypes of individuals at each stage of the process and specify which progeny will be considered recombinant. You do not need to calculate recombination frequency.

2) In a mutant screen in *Drosophila*, you identified a gene related to memory, as evidenced by the inability of recessive homozygotes to learn to associate a particular scent with the availability of food. Given another line of flies with an autosomal mutation that produces orange eyes, design a series of crosses to determine the map distance between these two loci and specify which progeny will be considered recombinant. You do not need to calculate recombination frequency.

3) Imagine that methionine heterotrophy, chlorosis (loss of chlorophyll), and absence of leaf hairs (trichomes) are each caused by recessive mutations at three different loci in Arabidopsis. Given a triple mutant, and assuming the loci are on the same chromosome, explain how you would determine the order of the loci relative to each other.

4) Three loci are linked in the order B-C-A. If the A-B map distance is 1cM, and the B-C map distance is 0.6cM, given the lines *AaBbCc* and *aabbcc*, what will be the frequency of *Aabb* genotypes among their progeny if one of the parents of the dihybrid had the genotypes *AABBCC*?

5) Genes for body color (B black dominant to b yellow) and wing shape (C straight dominant to c curved) are located on the same chromosome in flies. If single mutants for each of these traits are crossed (i.e. a yellow fly crossed to a curved-wing fly), and their progeny is testcrossed, the following phenotypic ratios are observed among their progeny.

black, straight	17
yellow, curved	12
black, curved	337
yellow, straight	364

a) Calculate the map distance between B and C.

b) Why are the frequencies of the two smallest classes not exactly the same?

6) Given the map distance you calculated between B-C in question 12, if you crossed a double mutant (i.e. yellow body and curved wing) with a wild-type fly, and testcrossed the progeny, what phenotypes in what proportions would you expect to observe among the F_2 generation?

7) Wild-type mice have brown fur and short tails. Loss of function of a particular gene produces white fur, while loss of function of another gene produces long tails, and loss of function at a third locus produces agitated behaviour. Each of these loss of function alleles is recessive. If a wild-type mouse is crossed with a triple mutant, and their F_1 progeny is test-crossed, the following recombination frequencies are observed among their progeny. Produce a genetic map for these loci.

Fur	Tail	Behaviour	Freq.
white	short	normal	16
brown	short	agitated	0
brown	short	normal	955
white	short	agitated	36
white	long	normal	0
brown	long	agitated	14
brown	long	normal	46
white	long	agitated	933

8) You discover an alien species that displays the same type of inheritance that is commonplace with diploid animals on Earth. You cross two true-breeding individuals and the F_1 displays these traits: *pear-shaped head, gray skin*, and *normal fingers*. The F_1 is testcrossed to give the following data:

91	pear-shaped head, glowing fingertip, and green skin
6	pear-shaped head, glowing fingertip, and gray skin
1	pear-shaped head, normal fingertip, and green skin
1	round head, glowing fingertip, and gray skin
7	round head, normal fingertip, and green skin
506	pear-shaped head, normal fingertip, and gray skin
85	round head, normal fingertip, and gray skin
491	round head, glowing fingertip, and green skin

Define gene symbols and create a genetic map that shows how these traits are arranged on the chromosomes. Be sure to mathematically correct for double crossover events. Calculate interference.

(Solution is at http://universitygenetics-mapping.blogspot.ca/2011/11/alien-gene-mapping.html)

9) Two true-breeding lizards were crossed. Three mutant traits were found in the parents: bent tail, missing thumb, and curled claws. The F1 lizards were all wild-type in appearance. The F1 females were testcrossed, and the offspring were sorted to obtain these data:

wild type	182
bent tail, missing thumb, curled claws	176
bent tail, missing thumb	5
missing thumb, curled claws	52
bent tail	55
bent tail, curled claws	2261
missing thumb	2279

a) Diagram the arrangement of alleles on the two homologous chromosomes for both parents (P generation) and the F_1.

b) Draw a genetic map based on these data. Be sure to mathematically correct for double-crossovers,

c) Calculate interference. Explain what this value means.

(Solution is at http://universitygenetics-mapping.blogspot.ca/2011/11/question-1b.html)

10) Two true-breeding minks were crossed. Between them there were three loci that were being analyzed: *drawn jowls, club foot*, and a behaviour of being *easily startled*. The F1 minks were all wild-type in appearance. The F1 females were test-crossed and the progeny sorted. The following data were obtained:

drawn jowls, normal foot, easily startled	157
normal jowls, club foot, calm	165
normal jowls, club foot, easily startled	15
drawn jowls, club foot, calm	139
drawn jowls, club foot, easily startled	14
normal jowls, normal foot, calm	21
drawn jowls, normal foot, calm	18
normal jowls, normal foot, easily startled	163

a) Define appropriate gene and allele symbols according to standard conventions.

b) Draw a genetic map based on these data.

c) What can you infer about these loci?

(Solution is at http://universitygenetics-mapping.blogspot.ca/2011/11/question-4.html)

CHAPTER 11 – DNA IS THE GENETIC MATERIAL

Figure 1.
Parent and offspring Wolf's Monkey.
(Flickr- Eric Heupel - CC BY-NC-ND 2.0)

INTRODUCTION

Genetics is the scientific study of heredity and the variation of inherited characteristics. It includes the study of genes, themselves, how they function, interact, and produce the visible and measurable characteristics we see in individuals and populations of species as they change from one generation to the next, over time, and in different environments.

Heredity is the concept that the characteristics of an individual plant or animal in a population could be passed down through the generations. Offspring look more like their parents (**Figure 1**). People learned that some heritable characteristics (such as the size or colour of fruit) varied between individuals, and that they could select or breed crops and animals for the most favorable traits. Knowledge of these hereditary properties has been of significant value in the history of human development. In the past, humans could only manipulate and select from naturally existing combinations of genes. More recently, with the discovery of the substance and nature of genetic material, DNA, we can now identify, clone, and create novel, better combinations of genes that will serve our goals. Understanding the mechanisms of genetics is fundamental to using it wisely and for the betterment of all.

Prior to **Mendel** (1865) heredity was considered to be of a "**blended inheritance**" but his work demonstrated that inheritance was particulate in nature (**particulate inheritance**). We now call these "particles" **genes** and their different forms, **alleles**. By the early 1900's, biochemists had isolated hundreds of different chemicals from living cells, but which of these was the genetic material? Proteins seemed like promising candidates, since they were abundant, diverse, and complex molecules. However, a few key experiments demonstrated that DNA, rather than protein, is the genetic material.

1. GRIFFITH'S TRANSFORMATION EXPERIMENT (1928)

Microbiologists identified two strains of the bacterium *Streptococcus pneumoniae*. The R-strain produced rough colonies on a bacterial plate, while the other S-strain was smooth (**Figure 2**). More importantly, the S-strain bacteria caused fatal infections when injected into mice, while the R-strain did not (**Figure 3**). Neither did "heat-treated" S-strain cells. **Griffith** in 1929 noticed that upon mixing "heat-treated" S-strain cells together with

some R-type bacteria (neither should kill the mice), the mice died and there were S-strain, pathogenic cells recoverable. Thus, some non-living component from the S-type strains contained genetic information that could be transferred to and **transform** the living R-type strain cells into S-type cells.

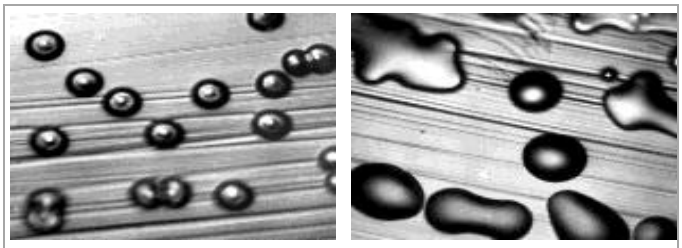


Figure 2.
Colonies of Rough (top) and Smooth (bottom) strains of *S. pneumoniae*.
(J. Exp.Med.98:21, 1953-R. Austrian-Pending)

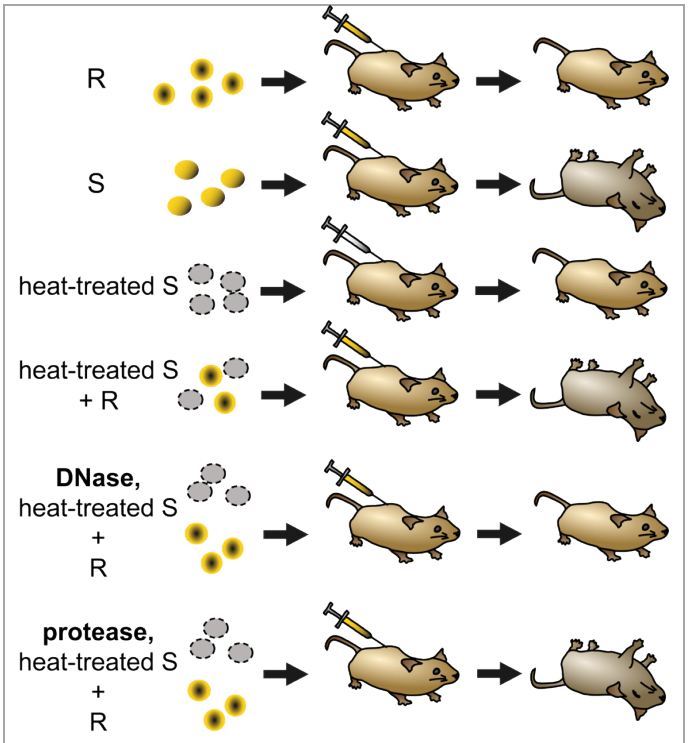


Figure 3.
Experiments of Griffith and of Avery, MacLeod and McCarty. R strains of *S. pneumoniae* do not cause lethality. However, DNA-containing extracts from pathogenic S strains are sufficient to make R strains pathogenic.
(Original-Deyholos- CC BY-NC 3.0)

2. AVERY, MACLEOD AND MCCARTY'S EXPERIMENT (1944)

What kind of molecule from within the S-type cells was responsible for the transformation? To answer this, researchers named **Avery, MacLeod and McCarty** separated the S-type cells into various components, such as proteins, polysaccharides, lipids, and nucleic acids. Only the nucleic acids from S-type cells were able to make the R-strains smooth and fatal. Furthermore, when cellular extracts of S-type cells were treated with **DNase** (an enzyme that digests DNA), the transformation ability was lost. The researchers therefore concluded that DNA was the genetic material, which in this case controlled the appearance (smooth or rough) and pathogenicity of the bacteria.

3. HERSHEY AND CHASE'S EXPERIMENT (1952)

Further evidence that DNA is the genetic material came from experiments conducted by **Hershey and Chase**. These researchers studied the transmission of genetic information in a virus called the T2 **bacteriophage**, which used *Escherichia coli* as its host bacterium (**Figure 4**).

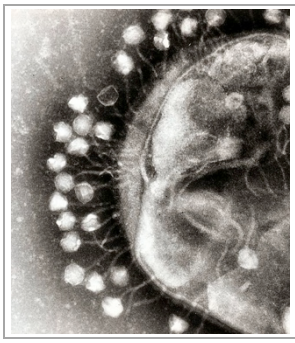


Figure 4.
Electronmicrograph of T2 bacteriophage on surface of E. coli.
(Wikipedia- Dr Graham Beards- CC BY-SA 3.0)

Like all viruses, T2 hijacks the cellular machinery of its host to manufacture more viruses. The T2 phage itself only contains both protein and DNA, but no other class of potential genetic material. To determine which of these two types of molecules contained the genetic blueprint for the virus, Hershey and Chase grew viral cultures in the presence of radioactive isotopes of either phosphorus (^{32}**P**) or sulphur (^{35}**S**). The phage incorporated these isotopes into their DNA and proteins, respectively (**Figure 5**). The researchers

then infected *E. coli* with the radiolabeled viruses, and looked to see whether ^{32}P or ^{35}S entered the bacteria. After ensuring that all viruses had been removed from the surface of the cells, the researchers observed that infection with ^{32}P labeled viruses (but not the ^{35}S labeled viruses) resulted in radioactive bacteria. This demonstrated that DNA was the material that contained genetic instructions.

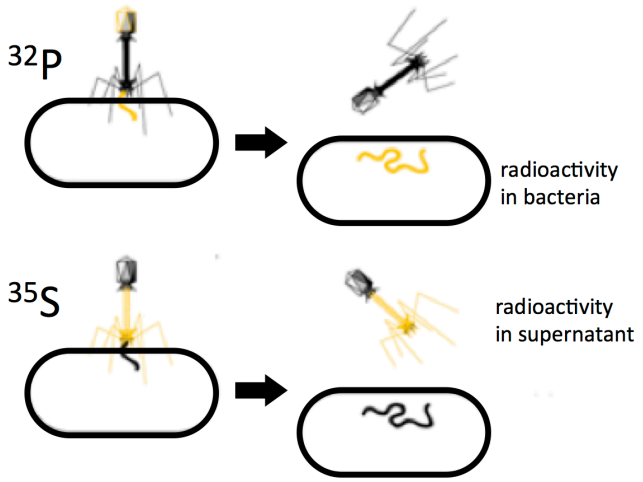


Figure 5.
When ^{32}P-labeled phage infects *E. coli*, radioactivity is found only in the bacteria, after the phage are removed by agitation and centrifugation. In contrast, after infection with ^{35}S-labeled phage, radioactivity is found only in the supernatant that remains after the bacteria are removed. (Original-Deyholos- CC BY-NC 3.0)

4. RNA AND PROTEIN

While DNA is the genetic material for the vast majority of organisms, there are some viruses that use **RNA** as their genetic material. These viruses can be either single or double stranded. Examples include SARS, influenza, hepatitis C and polio, as well as the retroviruses like HIV-AIDS. Typically there is DNA used at some stage in their life cycle to replicate their RNA genome.

Also, the **prion** protein is an infectious agent that transmits characteristics via only a protein (no nucleic acid present). Prions infect by transmitting a mis-folded protein state from one aberrant protein molecule to a normally folded molecule. These agents are responsible for Bovine Spongiform Encephalopathy (BSE, also known as "mad cow disease") in cattle, Chronic Wasting Disease in deer, Scrapie is sheep and Creutzfeldt–Jakob disease (CJD) in humans. All known prion diseases act by altering the structure of the brain or other neural tissue and all are currently untreatable and ultimately fatal.

SUMMARY:

- Genetics is the scientific study of heredity and the variation of inherted characteristics.
- Heredity is the concept that a trait of an individual can br passed down through generations
- A gene can be defined abstractly as a unit of inheritance.
- The experiments done by Griffitish and Hershey and Chase showed the ability of DNA from bacteria and viruses to transfer genetic information into bacteria demonstrates that DNA is the genetic material and that its universal.
- Some viruses use RNA as their genetic material and can be either single or double stranded.
- Prion is a misfolding protein that transmits its misfolding property to a normal one.

KEY WORDS:

genetics
heredity
Mendel
blending inheritance
particulate inheritance
gene
allele
Griffith
transform
Avery, MacLeod, & McCarty
DNase
Hershey and Chase
bacteriophage
^{35}S
^{32}P
prion

STUDY QUESTIONS:

1) Imagine that astronauts provide you with living samples of multicellular organisms discovered on another planet. These organisms reproduce with a short generation time, but nothing else is known about their genetics.
   a) How could you define laws of heredity for these organisms?
   b) How could you determine what molecules within these organisms contained genetic information?
   c) Would the mechanisms of genetic inheritance likely be similar for all organisms from this planet?
   d) Would the mechanisms of genetic inheritance likely be similar to organisms from earth?

2) It is relatively easy to extract DNA and protein from cells; biochemists had been doing this since at least the 1800's. Why then did Hershey and Chase need to use radioactivity to label DNA and proteins in their experiments?

3) Starting with mice and R and S strains of *S. pneumoniae*, what experiments in addition to those shown in **Figure 3** to demonstrate that DNA is the genetic material?

CHAPTER 12 – DNA STRUCTURE AND REPLICATION

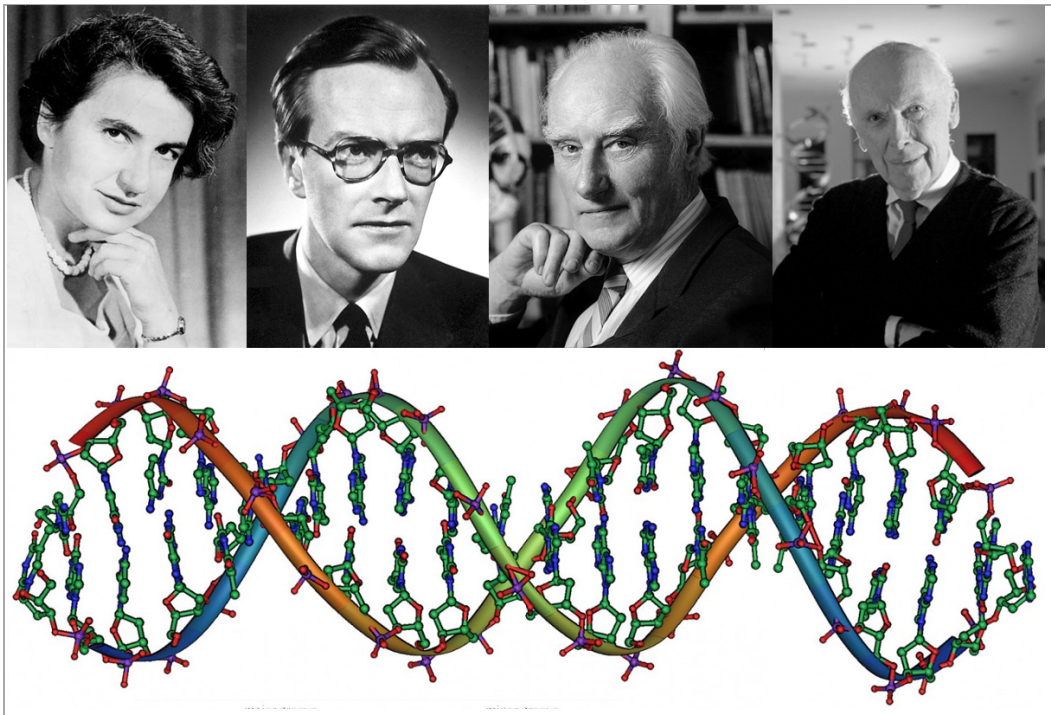


Figure 1.
The scientists responsible for determining the structure of DNA are Rosalind Franklin, Maurice Wilkins, Francis Crick and James Watson (left to right). Although the work was published in 1953, Wilkins, Crick, and Watson received the Nobel Prize in Physiology or Medicine in 1962, after Franklin had died in 1958 of ovarian cancer.

From left to right
(Wikipedia-Unknown-PD)
(Wikipedia-Unknown-PD)
(Wikiepdia-Marc Lieberman-CC BY 2.5)
(Wikipeida-Cold Spring Harbor Laborator-PD)
Bottom:
(Wikipedia-Michael Ströck-CC BY-SA 3.0)

INTRODUCTION

One of the fundamental things to know when studying genetics is the basic structure of DNA and how it is replicated. DNA is the "blueprint" that contains all the instructions for making the proteins that each cell needs, whether it is a single celled bacterium or a multicellular organism like humans. J. Watson, F. Crick, and M. Wilkins received the Nobel Prize (1962) for discovering the structure of DNA. (R. Franklin might have also received the prize for this discovery, but she died in 1958.)

The basic structure of DNA provides insight into its function. The main features of its structure are that it can reliably: (1) reproduce exact copies of this information to pass on to descendant cells, and (2) use the information to create proteins that produce and regulate the biochemistry of the cell. Remember however, DNA within the cell is more than just a loose strand within the nucleus. DNA interacts with proteins and is packaged into higher order structures (chromosomes) that will be discussed later in the textbook in Chapter 7. These proteins also regulate the expression of genes (information in the DNA).

This chapter will be going through the components of the DNA molecule, how the double helix structure was discovered, and how the method of replication was discovered.

1. DNA STRUCTURE - DOUBLE HELIX

1.1. NUCLEIC ACIDS AND PHOSPHATE SUGAR BACKBONE

In 1869 Johannes Friedrich Miesher, a Swiss physician and biologist, first isolated a substance he called 'nuclein' from the nuclei of a human white blood cell. He identified this substance to be weakly acidic with a high amount of phosphorus. This substance, after being further purified and studied was later called **deoxyribonucleic acid**, or DNA. Its name describes the three characteristics of the molecule, namely it has a ribose sugar with only one hydroxyl group called deoxyribose (**Figure 2**), it is found in the nucleus of a cell, and it is acidic.

After purifying the 'nuclein' to DNA they found it contained four different subunits that are linked in a chain. Those subunits were identified as **nucleotides**. A nucleotide contains three components, a phosphate group (PO_4^{3-}), a deoxyribose sugar, and one of four nitrogenous bases. Those bases fit into two groups based upon their structure. **Purines** have a double ring structure and include **adenine** and **guanine**. **Pyrimidines** have a single ring structure and include **cytosine** and **thymine** (**Figure 2**). The nature of the phosphate group and the single oxygen containing group of the deoxyribose sugar allows each nucleotide to chain together, forming the long DNA strand.

If you notice in **Figure 2**, one of the nucleotides has each carbon of the ribose labeled with a number followed by a prime, 1'-5'. The 1' position is where the base is attached. The 2' position is where the ribose is missing a hydroxyl group. The 5' position is attached to the phosphate group. When linked in a chain, that phosphate group can then link to the 3' oxygen of the next nucleotide using a **phosphodiester bond**. When a chain is formed, there will always be two termini, the free 5' phosphate and the free 3' oxygen on the ribose. These are known as the 5' and 3' ends, respectively, of the DNA strand.

Ribonucleic acid (RNA) is like DNA, in that it forms chains similarly, and has the bases attached to the same carbon. The extra hydroxyl group at the 2' position causes it to form a different conformation than DNA, becoming a more flexible molecule (DNA's conformation will be described later in this chapter). There are also **dideoxynucleotides** that are missing the hydroxyl group at both the 2' and 3' position. Because of this, a chain cannot form at the 3' carbon, terminating the chain. This feature of dideoxynucleotides is used in Sanger sequencing, which will be described in Chapter 33.

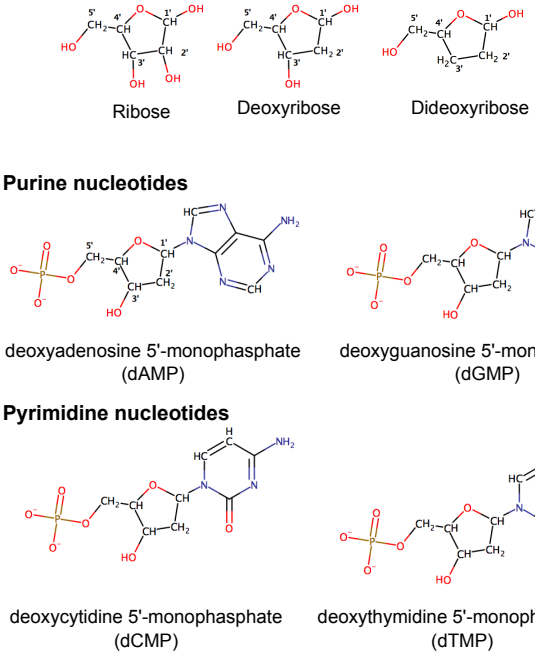


Figure 2.
Molecular models of the components of DNA. The top shows the three different types of ribose sugar found in nucleic acids. The bottom shows the purine and pyrimidine nucleotide monophosphates. Carbon numbers (1'-5') are labeled on the sugars and dAMP. (Original- L.Canham - CC BY-NC 3.0)

1.2. CHARGAFF'S RULES

When **Watson and Crick** set out in the 1940's to determine the structure of DNA, they already knew that DNA is made up of a series nucleotides with four different bases: adenine (A), cytosine (C), thymine (T), guanine (G). For DNA, the nucleotides are abbreviated as dNTPs (deoxyribonucleotide triphosphates), which include dATP, dCTP, dGTP, and dTTP. For RNA they are abbreviated as NTPs, which include ATP, CTP, GTP, and UTP. Watson and Crick also knew of **Chargaff's Rules**, which were a set of observations about the relative amount of each nucleotide that was present in almost any extract of DNA. Chargaff had observed that for any given species, the abundance of A was the same as T, and G was the same as C. This was essential to Watson & Crick's model.

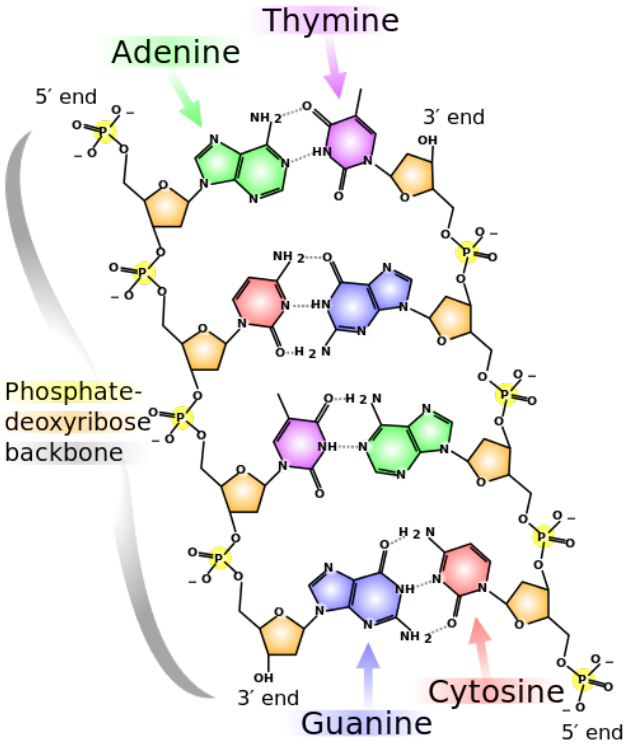


Figure 3.
Chemical structure of two pairs of nucleotides in a fragment of double-stranded DNA. Sugar, phosphate, and bases A,C,G,T are labeled. Hydrogen bonds between bases on opposite strands are shown by dashed lines. Note that the G-C pair has more hydrogen bonds than A-T. The polarity of each strand is indicated by the labels 5' and 3'. (Wikipedia- Madeleine Price Ball - CC0 1.0)

1.3. THE DOUBLE HELIX

Using proportional metal models of the individual nucleotides, Watson and Crick deduced a structure for DNA that was consistent with Chargaff's Rules and with x-ray crystallography data that was obtained (with some controversy) from another researcher named Rosalind Franklin. In Watson and Crick's famous **double helix**, each of the two strands contains DNA bases connected through covalent bonds to a sugar-phosphate backbone (**Figure 1** and **Figure 3**) Because one side of each sugar molecule is always connected to the opposite side of the next sugar molecule, each strand of DNA has polarity: these are called the 5' (5-prime) end and the 3' (3-prime) end. The two strands of the double helix run in **anti-parallel** (i.e. opposite) directions, with the 5' end of one strand adjacent to the 3' end of the other strand. The double helix has a **right-handed** twist, (rather than the left-handed twist that is often represented incorrectly in popular media). The DNA bases extend from the backbone towards the center of the helix, with a pair of bases from each strand forming hydrogen bonds that help to hold the two strands together. Because of the structure of the bases, A can only form hydrogen bonds with T, and G can only form hydrogen bonds with C (remember Chargaff's Rules). Each strand is therefore said to be complementary to the other, and so each strand also contains enough information to act as a template for the synthesis of the other. This complementary redundancy is important in DNA replication and repair.

Under most conditions, the two strands in the double helix are slightly offset, which creates a **major groove** on one face, and a **minor groove** on the other. In **Figure 1**, notice how if you look along the bottom edge of the figure you it makes a wave pattern, with a large dip followed by a small dip, followed by a large dip. That is where you can see the major and minor grooves. These grooves provide access for transcription regulating proteins (transcription factors), which bind to specific sequences of bases along the DNA.

2. SEMI-CONSERVATIVE REPLICATION (VS. CONSERVATIVE, DISPERSIVE)

From the complementary strands model of DNA, proposed by Watson and Crick in 1953, there were three straightforward possible mechanisms for DNA replication: (1) semi-conservative, (2) conservative, and (3) dispersive (**Figure 4**).

The **semi-conservative** model proposes the two strands of a DNA molecule separate during replication and then strand acts as a template for synthesis of a new, complementary strand.

The **conservative** model proposes that the entire DNA duplex acts as a single template for the synthesis of an entirely new duplex.

The **dispersive** model has the two strands of the double helix breaking into units that which are then replicated and reassembled, with the new duplexes containing alternating segments from one strand to the other.

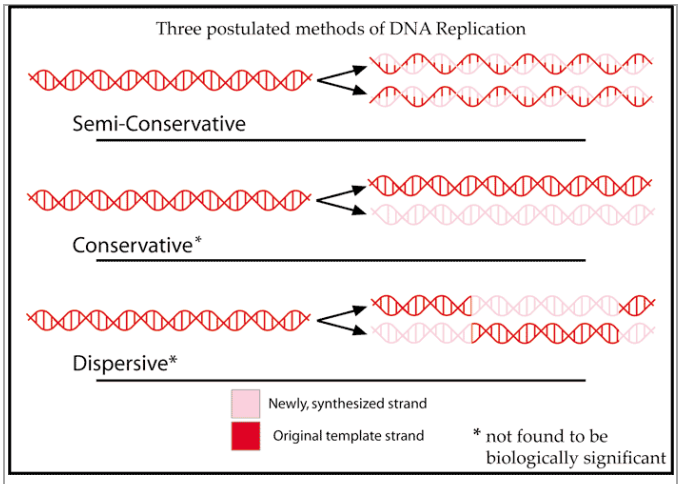


Figure 4.
The three models of DNA replication possible from the double helix model of DNA structure.
(Wikipedia-Adenosine- CC BY-SA 2.5)

Each of these three models makes a different prediction about the how DNA strands should be distributed following two rounds of replication. These predictions can be tested in the following experiment by following the nitrogen component in DNA in ***E. coli*** as it goes through several rounds of replication. Two scientists, **Meselson and Stahl** in 1958, used different isotopes of Nitrogen, which is a major component in DNA. **Nitrogen-14** (**^{14}N**) is the most abundant natural isotope, while **Nitrogen-15** (**^{15}N**) is rare, but also denser. Neither is radioactive; each can be followed by a difference in density – "**light**" 14 vs "**heavy**" 15 atomic weight in a CsCl density gradient ultra-centrifugation of DNA.

The experiment starts with *E. coli* grown for several generations on medium containing only ^{15}N. It will have denser DNA. When extracted and separated in a CsCl density gradient tube, this "heavy" DNA will move to a position nearer the bottom of the tube in the more dense solution of CsCl (left side in **Figure 5**). DNA extracted from *E. coli* grown on normal, ^{14}N containing medium will migrate more towards the less dense top of the tube.

If these *E. coli* cells are transferred to a medium containing only ^{14}N, the "light" isotope, and grown for one generation, then their DNA will be composed of one-half ^{15}N and one-half ^{14}N. If the this DNA is extracted and applied to a **CsCl gradient**, the observed result is that one band appears at the point midway between the locations predicted for wholly ^{15}N DNA and wholly ^{14}N DNA (**Figure 5**). This "single-band" observation is inconsistent with the predicted outcome from the conservative model of DNA replication (disproves this model), but is consistent with both that expected for the semi-conservative and dispersive models.

If the *E. coli* is permitted to go through another round of replication in the ^{14}N medium, and the DNA extracted and separated on a CsCl gradient tube, then two bands were seen by Meselson and Shahl: one at the ^{14}N-^{15}N intermediate position and one at the wholly ^{14}N position (**Figure 5**). This result is inconsistent with the dispersive model (a single band between the ^{14}N-^{15}N position and the wholly ^{14}N position) and thus disproves this model. The two band observation is consistent with the semi-conservative model which predicts one wholly ^{14}N duplex and one ^{14}N-^{15}N duplex. Additional rounds of replication also support the semi-conservative model/hypothesis of DNA replication. Thus, the semi-conservative model is the currently accepted mechanism for DNA replication. Note however, that we now also know from more recent experiments that whole chromosomes, which can be millions of bases in length, are also semi-conservatively replicated.

These experiments, published in 1958, are a wonderful example of how science works. Researchers start with three clearly defined models (hypotheses). These models were tested, and two (conservative and dispersive) were found to be inconsistent with the observations and thus disproven. The third hypothesis, semi-conservative, was consistent with the observations and thereby supported and accepted as mechanism of DNA replication. Note, however, this is not "proof" of the model, just strong evidence for it; hypotheses are not "proven", only disproven or supported

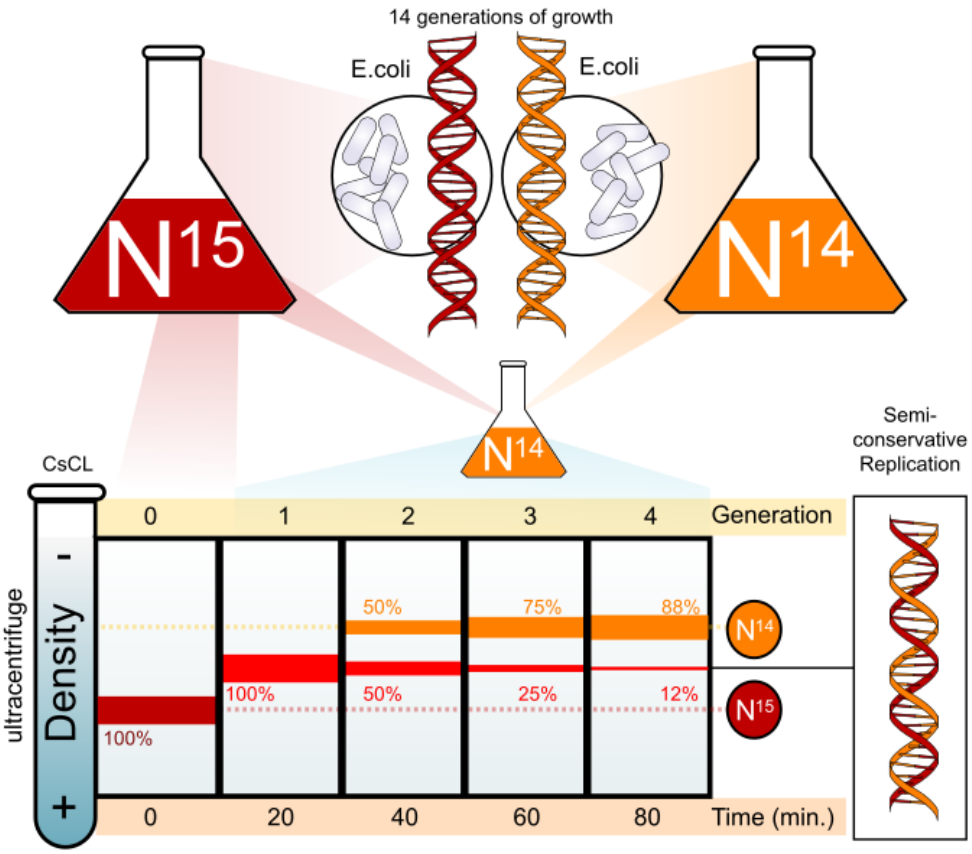


Figure 5.
The positions of the ^{14}N and ^{15}N containing DNA in the density gradient tube on the left. (Wikipedia-LadyofHats-PD)

3. CHROMOSOME REPLICATION (*E. COLI*) - CAIRNS EXPERIMENT

If the results of Meselson and Stahl were true and there was semi-conservative replication, then the two strands of DNA have to separate to provide the template for copying. This should be seen as a 'fork' in a linear model if you manage to see the DNA just as it's replicating. **John Cairns** in 1963 chose to test this.

To do this he took *E. coli* cells growing in a normal environment, and then allowed them to grow and replicate in the presence of radioactive **^{3}H-thymidine**. The hypothesis is that if the *E. coli*'s DNA or chromosome is semi-conservatively replicated then after the first round of replication there should be one newly made strand that is radioactive, or "hot", and the other strand that is the parental template strand with no radioactivity, so is "cold". The original parental DNA will have two strands, each not radioactive. After replication the daughter DNA will have two strands, one that is radioactive and one that is not. After a third round of replication there will be a two types of daughter DNA, one that has a non-radioactive strand and a radioactive strand, and one that has two radioactive strands.

After growth in the ^{3}H-thymidine, Cairns lysed the bacteria and collected the contents onto a microscope slide. He then covered the slide with a **photographic emulsion** and allowed exposure to film for 2 months. As the ^{3}H-thymidine decays it emits an electron with a lot of energy and speed, known as a beta particle. The emulsion reacts with the beta particle creating a black **silver grain** on the film. The density of grains should be indicative of whether one or two strands are radioactive.

After the first replication cycle, the film had a thin circular ring of grains (**Figure 6**). This was interpreted to be a daughter chromosome with one strand that is hot and one strand cold. This also provided physical evidence that the *E. coli* chromosome is circular, something that has only previously been shown genetically.

In the second replication cycle the replication fork was seen. Here Cairns saw the typical thin ring of grains much like the first replication cycle, but with a branch in the middle that had a thicker strand (**Figure 6**). This means that the branch seen was an actively replicating chromosome, using the radioactive strand of DNA as a template, and adding more radioactive thymidine as the DNA is being synthesized. Because of the shape these created on the film this replicating structure was called a **theta (Θ) structure**. Cairns observed many different molecules corresponding to the progression from starting replication to the completion of replication.

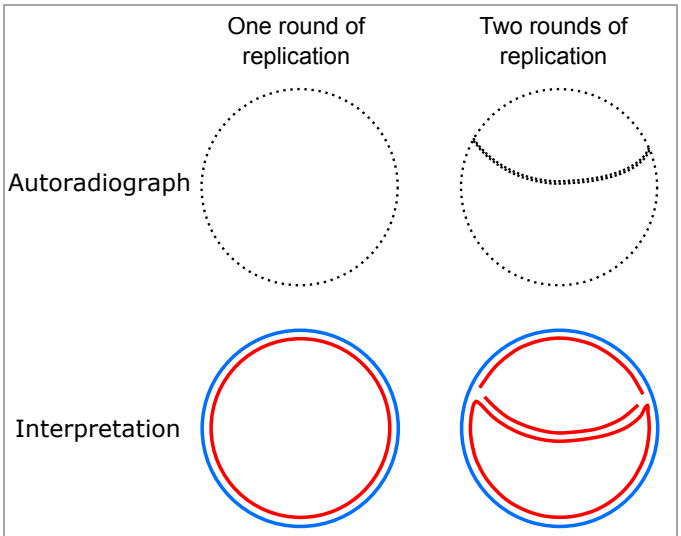


Figure 6.
In his experiment, Cairns looked at DNA with radioactive thymidine on an autoradiograph film, with the radioactive thymidine leaving dots on the film. This figure shows what the autoradiograph film would look like, and below what the interpretation of what the autoradiograph shows. The blue line represents the 'cold' DNA that has no radioactivity, while the red shows the 'hot' radioactive DNA. The density of the dots on the autoradiograph imply whether there is one strand or both strands of hot DNA. During the second round of replication, a theta structure can be seen, as the circular *E. coli* DNA is in the process of being replicated. (Original-L.Canham- CC BY-NC 3.0)

Here Cairns' results were able to further support the semi-conservative replication theory, showing the existence of replication forks, as well as the hypothesis that *E. coli* has a circular chromosome. What Cairns did not realize is that replication goes in both directions at the replication fork, where he thought one fork was static while the other strand went around the chromosome replicating. Scientists later went on to show that replication is in-fact **bidirectional**.

4. ORIGINS OF REPLICATION (PROKARYOTE - SINGLE ORGIN), REPLICATION FORK

When the cell enters S-phase in the cell cycle (See Chapter 14) the entire chromosomal DNA is replicated. This is done by enzymes called **DNA polymerases**. All DNA polymerases synthesize new strands by adding nucleotides to the 3'OH group present on the previous nucleotide. For this reason they are said to work in a 5' to 3' direction. DNA polymerases use a single strand of DNA as a template upon which it will synthesize the complementary sequence. This works fine for the middle of chromosomes. DNA-directed DNA polymerases travel along the original DNA strands making complementary strands (**Figure 7**a).

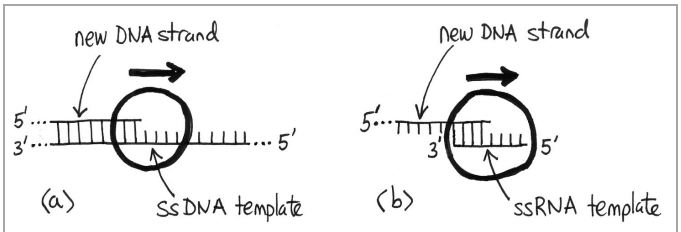


Figure 7.
DNA polymerases make new strands in a 5' to 3' direction. (a) Regular DNA polymerases are proteins or protein complexes that use a single strand of DNA as a template. For example, the main human DNA polymerase, Pol α, is large protein complex made of four polypeptides. (b) Telomerases use their own RNA as a template. The human telomerase is a complex made of one polypeptide and one RNA molecule.(Original-Harrington- CC BY-NC 3.0)

DNA replication in both prokaryotes and eukaryotes begins at an **Origin of Replication** (Ori). Origins are specific sequences on specific positions on the chromosome. In *E. coli*, the *OriC* origin is ~245 bp in size. Chromosome replication begins with the binding of the DnaA initiator protein to an AT-rich 9-mer in *OriC* and melts the two strands. Then DnaC loader protein helps DnaB helicase protein extend the single stranded regions such that the DnaG primase can initiate the synthesis of an RNA primer, from which the DNA polymerases

can begin DNA synthesis at the two replication forks. The forks continue in opposite directions until they meet another fork or the end of the chromosome (**Figure 8**).

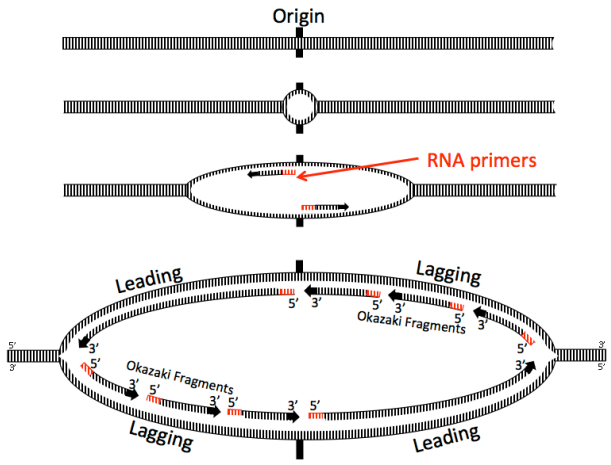


Figure 8.
An origin of replication. The sequence specific DNA duplex is melted then the primase synthesizes RNA primers from which bidirectional DNA replication begins as the two replication forks head off in opposite directions. The leading and lagging strands are shown along with Okazaki fragments. Note the 5' and 3' orientation of all strands.
(Original-Locke- CC BY-NC 3.0)

5. EUKARYOTE CHROMOSOME REPLICATION - MULTIPLE ORIGINS

In prokaryotes, with a small, simple, circular chromosome, only one origin of replication is needed to replicate the whole genome. For example, *E. coli* has a ~4.5 Mb genome (chromosome) that can be duplicated in ~40 minutes assuming a single origin, bi-directional replication, and a speed of ~1000 bases/second/fork for the polymerase.

However, in larger, more complicated eukaryotes, with multiple linear chromosomes, more than one origin of replication is required per chromosome to duplicate the whole chromosome set in the 8-hours of the replicative phase (S-phase) of the cell cycle. For example, the human diploid genome has 46 chromosomes (6 x 10^9 basepairs). The shortest chromosomes are ~50 Mbp long and so could not possibly be replicated from one origin. Additionally, the rate of replication fork movement is slower, only ~100 base/second. Thus, eukaryotes contain multiple origins of replication distributed over the length of each chromosome to enable the duplication of each chromosome within the observed time of S-phase (**Figure 9**).

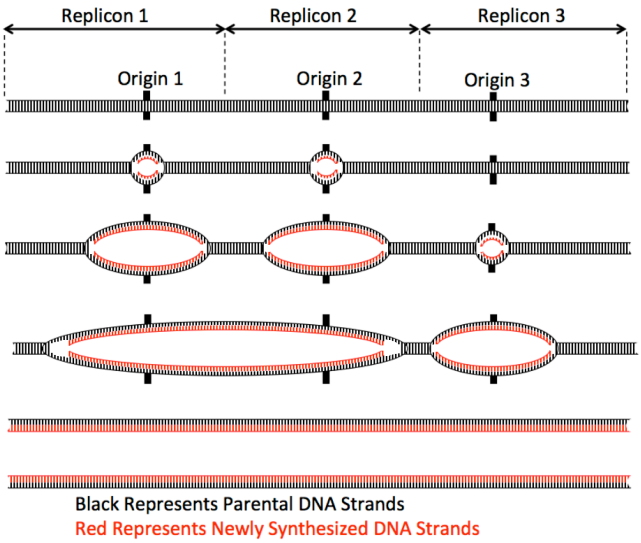


Figure 9.
Part of a eukaryote chromosome showing multiple Origins (1, 2, 3) of Replication, each defining a **replicon** (1, 2, 3). Replication may start at different times in S-phase. Here #1 and #2 begin first then #3. As the replication forks proceed bi-directionally, they create what are referred to as "**replication bubbles**" that meet and form larger bubbles. The end result is two semi-conservatively replicated duplex DNA strands.
(Original-Locke- CC BY-NC 3.0)

6. TELOMERES

The ends of linear chromosomes present a problem – at each end one strand cannot be completely replicated because there is no primer to extend. Although the loss of such a small sequence would not be a problem, the continued rounds of replication would result in the continued loss of sequence from the chromosome end to a point were it would begin to loose essential gene sequences. Thus, this DNA must be replicated. Most eukaryotes solve the problem of synthesizing this unreplicated DNA with a specialized DNA polymerase called **telomerase,** in combination with a regular polymerase. Telomerases are RNA-directed DNA polymerases. They are a **riboprotein,** as they are composed of both protein and RNA. As **Figure 7**b shows, these enzymes contain a small

piece of RNA that serves as a portable and reusable template from which the complementary DNA is synthesized. The RNA in human telomerases uses the sequence 3-AAUCCC-5' as the template, and thus our telomere DNA has the complementary sequence 5'-TTAGGG-3' repeated over and over 1000's of times. After the telomerase has made the first strand a primase synthesizes an RNA primer and a regular DNA polymerase can then make a complementary strand so that the telomere DNA will ultimately be double stranded to the original length (**Figure 10**). Note: the number of repeats, and thus the size of the telomere, is not set. It fluctuates after each round of the cell cycle. Because there are many repeats at the end, this fluctuation maintains a length buffer – sometimes it's longer, sometimes it's shorter – but the average length will be maintained over the generations of cell replication.

In the absence of telomerase, as is the case in human somatic cells, repeated cell division leads to the "**Hayflick limit**", where the telomeres shorten to a critical limit and then the cells enter a senescence phase of non-growth. The activation of telomerase expression permits a cell and its descendants to become immortal and bypass the Hayflick limit. This happens in cancer cells, which can form tumours as well as in cells in culture, such as **HeLa cells**, which can be propagated essentially indefinitely. HeLa cells have been kept in culture since 1951 (See Chapter 41).

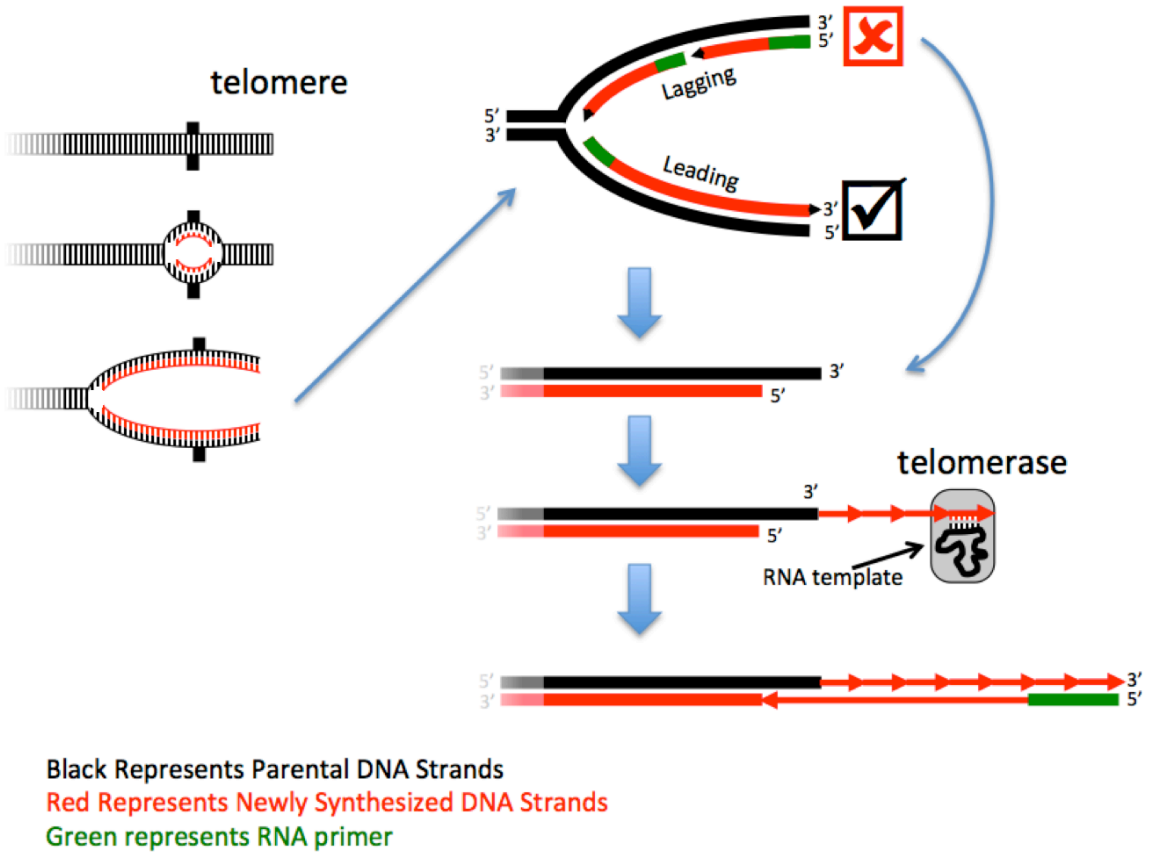


Figure 10.
Telomere replication showing the completion of the leading strand and incomplete replication of the lagging strand. The gap is replicated by the extension of the 3' end by telomerase and then filled in by extension of an RNA primer.

SUMMARY:

- DNA is a double helix made of two anti-parallel strands of bases on a sugar-phosphate backbone.
- Specific bases on opposite strands pair through hydrogen bonding (A=T and G=C), ensuring complementarity of the strands.
- The hereditary information is present as the sequence of bases along the DNA strand.
- Chromosome replication begins at an origin and proceeds by DNA polymerases at a replication fork.
- Replication proceeds bi-directionally.
- Typically eukaryotes have multiple origins along each chromosome, while prokaryotes have only one.
- Eukaryotes have telomerase to complete the replication of the ends of chromosomes.

KEY TERMS:

deoxyribonucleic acid
nucleotides
purine
adenine
guanine
pyrimidine
cytosine
thymine
phosphodiester bond
ribonucleic acid
dideoxynucleotide
Watson and Crick
Chargaff's Rules
double helix
anti-parallel
right-handed
major groove
minor groove
semi-conservative
conservative
dispersive
E. coli
Meselson and Stahl
Nitrogen-14
Nitrogen-15
light
heavy
CsCl gradient
John Cairns
^{3}H-thymidine
photographic emulsion
silver grain
theta structure
bidirectional
DNA polymerases
Origin of replication
replicon
replication bubble
telomerase
riboprotein
Hayflick limit
HeLa cells

STUDY QUESTIONS:

1) Compare Watson and Crick's discovery with Avery, MacLeod and McCarty's discovery.

a) What did each discover, and what was the impact of these discoveries on biology?

b) How did Watson and Crick's approach generally differ from Avery, MacLeod and McCarty's?

c) Briefly research Rosalind Franklin on the internet. Why is her contribution to the structure of DNA controversial?

2) List the information that Watson and Crick used to deduce the structure of DNA.

3) Refer to Watson and Crick'

a) List the defining characteristics of the structure of a DNA molecule.

b) Which of these characteristics are most important to replication?

c) Which characteristics are most important to the Central Dogma?

4) Refer to **Figure 3**.

a) Identify the part of the DNA molecule that would be radioactively labeled in the manner used by Hershey & Chase

b) DNA helices that are rich in G-C base pairs are harder to separate (e.g. by heating) than A-T rich helices. Why?

CHAPTER 12B – POLYMERASE CHAIN REACTION (PCR)

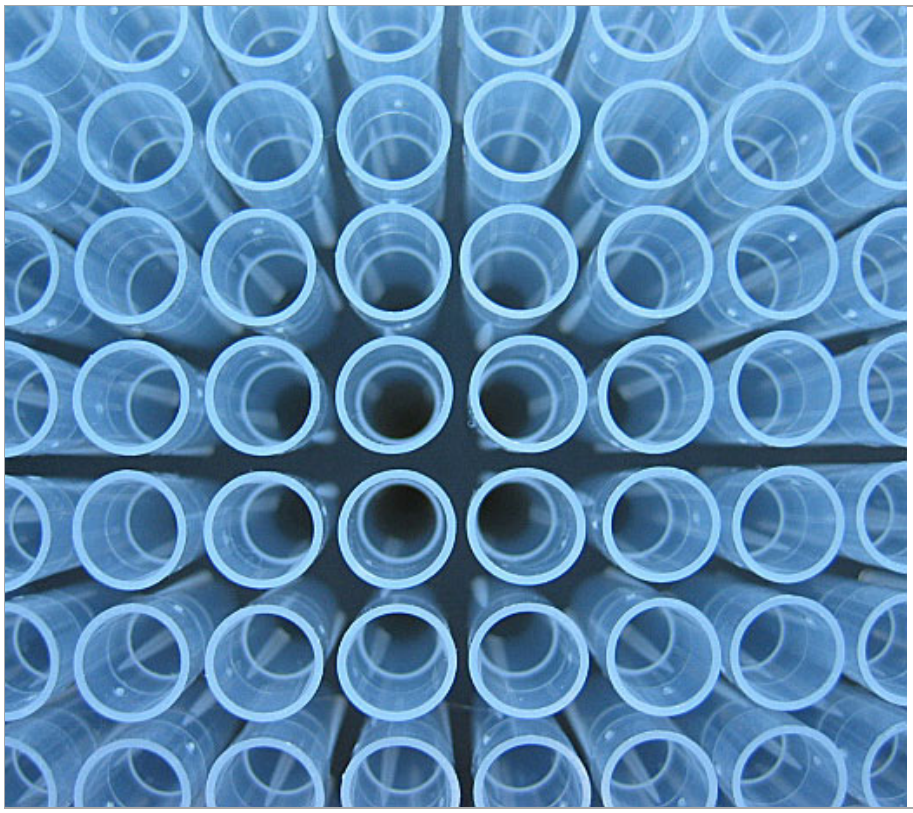


Figure 1.
Plastic disposable tips for a micro-pipettor are used to accurately distribute microliter volumes of liquid in molecular biology.
(Flickr-estherase- CC BY-NC-SA 2.0)

INTRODUCTION

While, genetics is the study of the inheritance and variation of biological traits, today, **classical genetics** is often complemented by **molecular biology**, to give **molecular genetics**, which involves the study of DNA and other **macromolecules** that have been isolated from an organism. Usually, molecular genetics experiments involve some combination of techniques to isolate, analyze, and characterize the DNA, RNA, and/or protein transcribed and translated from a particular gene. In some cases, the DNA may be subsequently manipulated by mutation or by recombination with other DNA fragments. Techniques of molecular genetics have wide application in many fields of biology, as well as forensics, biotechnology, and medicine. Polymerase Chain Reaction (PCR) is a widely used technique to amplify and isolate specific DNA sequences. It requires a "template" DNA, which is often genomic DNA. From this template, specific sequences can be amplified and many copies can be produced for analysis or manipulation.

1. ISOLATING GENOMIC DNA

DNA purification strategies rely on the chemical properties of DNA that distinguish it from other molecules in the cell, namely that it is a very long, negatively charged molecule. To extract purified DNA from a tissue sample, cells are broken open by grinding or **lysing** in a solution that contains chemicals that protect the DNA while disrupting other components of the cell (**Figure 2**). These chemicals may include **detergents**, which dissolve lipid membranes and denature proteins. A cation such as Na^+ helps to stabilize the negatively charged DNA and separate it from proteins, such as histones. A **chelating agent**, such as **EDTA**, is added to protect DNA by sequestering Mg^{2+} ions, which can otherwise serve as a necessary co-factor for **nucleases** (enzymes that digest DNA). As a result, free, double-stranded DNA molecules are released from the cell and from chromatin into the extraction

buffer, which also contains proteins and all other cellular components. (The basics of this procedure are simple enough that it can be done with household chemicals as presented on YouTube.)

The free DNA molecules are subsequently isolated by one of several methods. Commonly, proteins are removed by adjusting the salt concentration so they precipitate. The **supernatant**, which contains DNA and other, smaller metabolites, is then mixed with ethanol, which causes the DNA to precipitate. A small **pellet** of DNA can be collected by centrifugation, and after removal of the ethanol, the DNA pellet can be dissolved in water (usually with a small amount of EDTA and a pH buffer) for the use in other reactions. Note that this process has purified all of the DNA from a tissue sample (genomic and mitochondrial DNA); if we want to isolate a specific gene or DNA fragment, we must use additional techniques, such as PCR.

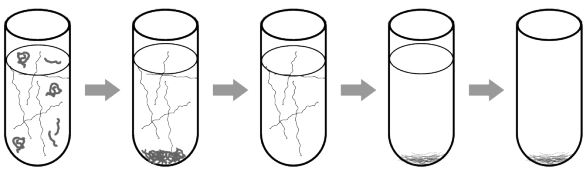


Figure 2.
Extraction of DNA from a mixture of solubilized cellular components by successive precipitations. Proteins are precipitated, then DNA (in the supernatant) is precipitated with ethanol, leaving a pellet of DNA.
(Original-Deyholos-CC BY-NC 3.0)

2. ISOLATING OR DETECTING A SPECIFIC SEQUENCE BY PCR

2.1. COMPONENTS OF THE PCR REACTION

The **Polymerase Chain Reaction (PCR)** is a method of DNA amplification that is performed in a test tube (i.e. *in vitro*). Here "polymerase" refers to a DNA polymerase enzyme extracted and purified from bacteria. The "chain reaction" refers to the ability of this technique produce billions of copies of a specific DNA molecule, by using each newly replicated double helix as a template to synthesize two new DNA double helices. PCR is therefore a very efficient method of amplifying a specific sequence of DNA from a small sample of a large, complex genome.

Besides its ability to make large amounts of DNA, there is a second characteristic of PCR that makes it extremely useful. Recall that most DNA polymerases can only add nucleotides to the end of an existing strand of DNA, and therefore require a **primer** to initiate the process of replication. For PCR, chemically synthesized primers of about 20 nucleotides are used. In an ideal PCR, primers only hybridize to their exact complementary sequence on the template strand (**Figure 3**).

Figure 3.
The primer-template duplex at the top part of the figure is perfectly matched, and will be stable at a higher temperature than the duplex in the bottom part of the figure, which contains many mismatches and therefore fewer hydrogen bonds. If the annealing temperature is sufficiently high, only the perfectly matched primer will be able to initiate extension (grey arrow) from this site on the template.
(Original-Deyholos-CC BY-NC 3.0)

The experimenter can therefore control exactly what region of a DNA template is amplified by specifying the sequence of the primers used in the reaction.
To conduct a PCR amplification, an experimenter combines in a small, thin-walled tube (**Figure 4**), all of the necessary components for DNA replication, including:

(1) DNA polymerase and solutions containing
(2) nucleotides (dATP, dCTP, dGTP, dTTP),
(3) a DNA template,
(4) DNA primers,
(5) a pH buffer, and
(6) ions (e.g. Mg^{2+}) required by the polymerase.

Successful PCR reactions have been conducted using only a single DNA molecule as a template, but in

practice, most successful PCR reactions contain many thousands of template molecules. The template DNA (e.g. total genomic DNA) has usually already been purified from cells or tissues using the techniques described above. However, in some situations it is possible to put whole cells directly in a PCR reaction for use as a template.

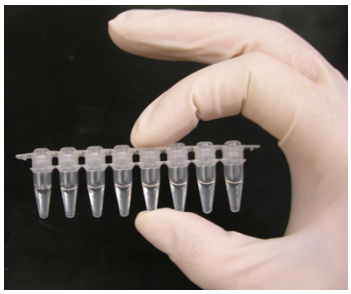


Figure 4.
A strip of PCR tubes
(Wikipedia-madprime- CC BY-SA 3.0)

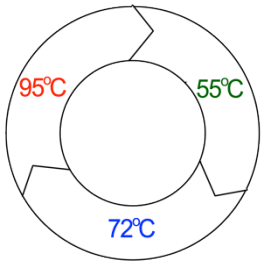


Figure 5.
Example of a thermal-cycle, in which the annealing temperature is 55°C.
(Original-Deyholos-CC BY-NC 3.0)

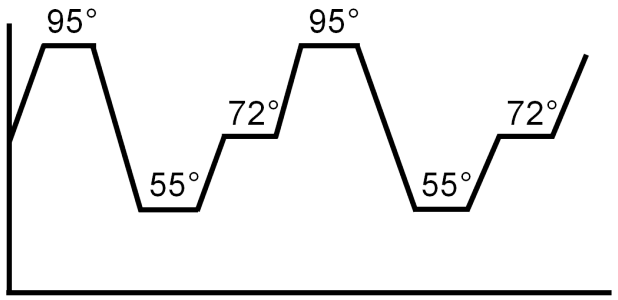


Figure 6.
A temperature vs. time graph showing two cycles of PCR.
(Original-Harrington-CC BY-NC 3.0)

An essential aspect of PCR is **thermal-cycling**, meaning the exposure of the reaction to a series of precisely defined temperatures (**Figure 5**). The reaction mixture is first heated to 95°C. This causes the hydrogen bonds between the strands of the template DNA molecules to melt, or **denature**. This produces two single-stranded DNA molecules from each double helix (**Figure 7**). In the next step (**annealing**), the mixture is cooled to 45-65°C. The exact temperature depends on the primer sequence used and the objectives of the experiment. This allows the formation of double stranded helices between complementary DNA molecules, including the annealing of primers to the template. In the final step (**extension**) the mixture is heated to 72°C. This is the temperature at which the particular DNA polymerase used in PCR is most active. During extension, the new DNA strand is synthesized, starting from the 3' end of the primer, along the length of the template strand. The entire PCR process is very quick, with each temperature phase usually lasting ~30 seconds or less. Each cycle of three temperatures (denaturation, annealing, extension) is usually repeated about 30 times, amplifying the target region approximately 2^{30}-fold. The amount of DNA product reaches a plateau at 20-40 cycles, usually because the nucleotide precursors have been exhausted. Notice from the figure that most of the newly synthesized strands in PCR begin and end with sequences either identical to or complementary to the primer sequences; although a few strands are longer than this, they are in such a small minority that they can almost always be ignored.

After completion of the thermalcycling (amplification), an aliquot from the PCR reaction is usually loaded onto an **electrophoretic agarose gel**

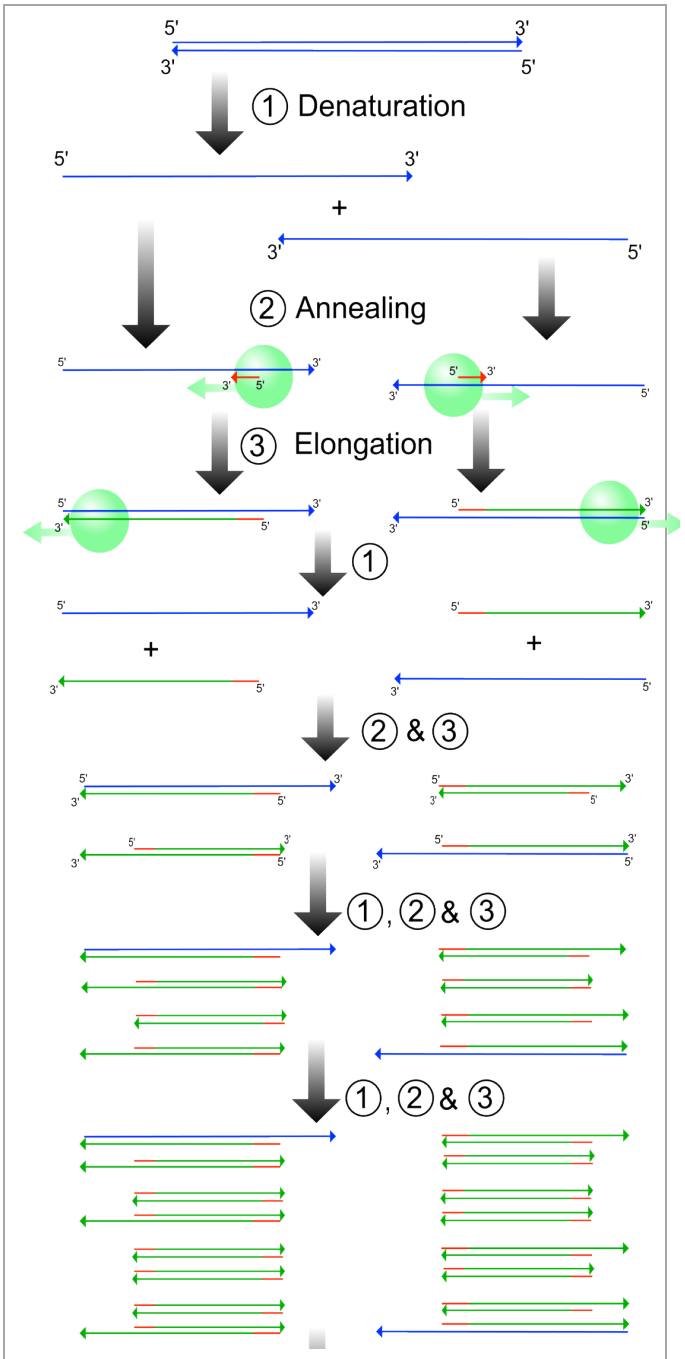


Figure 7.
PCR with the three phases of the thermal cycle numbered. The template strand (blue) is replicated using primers (red), to prime the newly synthesized strands (green). The green strands, which are flanked by the two primer sites, will increase in abundance exponentially through successive PCR cycles.
(Wikipedia-madprime- CC0 1.0)

(described in chapter 28) to determine whether a DNA fragment of the expected length was successfully amplified or not. Usually, the original template DNA will be so dilute that it will not be visible on the gel, only the amplified PCR product. The presence of a sharp band of the expected length indicates that PCR was able to amplify its target. If the purpose of the PCR was to test for the presence of a particular template sequence, this is the end of the experiment. Otherwise, the remaining PCR product can be used as starting material for a variety of other techniques such as sequencing or cloning.

2.2. REAL TIME PCR /QUANTITATIVE PCR (QPCR)

In a standard PCR reaction, the DNA molecule of interest is amplified and then the products are typically visualized at the end of the reaction on an electrophoresis gel. On the other hand, a procedure known as real-time PCR or quantitative PCR (qPCR) detects the replicated DNA molecules during the amplification process. qPCR uses florescent molecules and relies on the fluorescence of the amplified product measured over a number of cycles. However, the procedure of amplifying the DNA molecule is identical to the standard PCR procedures. There are two ways of processing qPCR.

(1) Using florescent chemical molecules known as **fluorochrome** that binds to all double stranded DNA molecule (Nonspecific).

In the first method, the florescent dye molecule binds to any double stranded DNA molecule. After each cycle of amplification, the amount of ds-DNA molecules synthesized can be quantified by measuring the florescence. The intensity of the florescence would indicate the amount of DNA molecule present.

(2) Using florescent reporter probe (Specific).

The second method is using a florescent reporter probe that hybridizes with the DNA sequence of interest. When the taq polymerase replicates the DNA molecule, it degrades the probe and the florescent molecule is released to the solution. This increases the intensity of florescence. The florescence is measured by the real-time PCR

machine and quantifies the DNA molecules being synthesized.

2.3. REVERSE TRANSCRIPTASE PCR (RT-PCR)

Reverse Transcriptase PCR (RT-PCR) can detect both the quality and quantity of mRNA molecules (gene transcription). As a result, we are able to find out the spatial (where the gene is expressed) and temporal (when the gene is expressed) level of gene expression.

Here is how it works (**Figure 8**):

(1) mRNA is extracted from the cell, tissue, or organism.

(2) An enzyme called **reverse transcriptase** (obtained from a retrovirus – see Chapter 30) is added, along with oligo-DT, which anneals to the poly-A tail and acts as a primer, to synthesize complementary DNA (cDNA) to the mRNA.

(3) mRNA template is degraded, and cDNA is added to a PCR reaction to amplify a specific gene sequence. If amplification occurs, the mRNA is present; if not, then it is absent. This permits the quantitation of a specific mRNA (gene) sequence.

The amplified products visualized on a gel verify the existence and the quantity of the gene of interest. By extracting mRNA at different stages, we can figure out the **temporal level** of gene expression. If we extract mRNAs from different cell types, we can figure out the **spatial level** of gene expression.

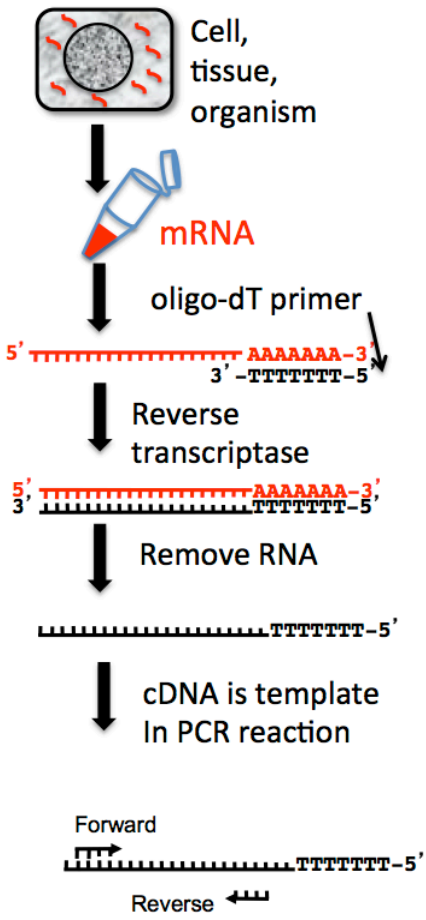


Figure 8.
The process of RT-PCR. Isolated mRNA can be reverse transcribed, and then used as a template in a PCR reaction to amplify one, or more, gene specific primers.
(Original-J. Locke-CC BY-NC 3.0)

2.4. AN APPLICATION OF PCR: THE STARLINK AFFAIR

PCR is very sensitive (meaning it can detect very small starting amounts of DNA), and specific (meaning it can amplify only the target sequence from a mixture of many DNA sequences). Due to these characteristic, PCR has many practical applications. For example, PCR can detect trace DNA contaminants in food, air, water or cells. The presence or absence and the type or species of the contaminant can be identified.

As an example, PCR was used as a tool to test whether genetically modified corn was present in consumer products on supermarket shelves. Although currently (2013) 85% of corn in the United States is genetically modified, and contains genes that government regulators have approved for human consumption, back in 2000, environmental groups showed that a strain of genetically modified corn, which had only been approved for use as animal feed, had been mixed in with corn used in producing human food, like taco shells. To do this, the groups purchased taco shells from stores in the Washington DC area, extracted DNA from the taco shells and used it as a template in a PCR reaction with primers specific for the unauthorized gene (*Cry9C*). Their suspicions were confirmed when they ran this PCR product on an agarose gel and saw a band of size expected for

Cry9C. The PCR test was sensitive enough to detect one transgenic kernel in a whole bushel of corn (1 per 100,000). The company (Aventis) that sold the transgenic seed to farmers had to pay for the destruction of large amounts of corn, and was the target of a class action law-suit by angry consumers who claimed they had been made sick by the taco shells. While no legitimate cases of harm were ever proven, and the plaintiffs were awarded $9 million, of which $3 million went to the legal fees, and the remainder of the judgment went to the consumers in the form of coupons for taco shells. The affair damaged the company, and exposed a weakness in the way the genetically modified crops were handled in the United States at the time.

PCR can be also used in medical diagnostic tests for detection of pathogens in blood, tissues and body fluids. More recently PCR has been used in the genotyping of patients to match their care with specific treatments for better outcomes.

PCR is also used for DNA genotyping of biological samples in forensic or criminal investigations. . People can be genotyped for identification purposes, so as to match with samples present at a crime scene or to establish family relationships in paternity/maternity cases. Genotypes also establish identity of people for future comparisons, much like taking fingerprints.

SUMMARY:

- Molecular biology involves the isolation and analysis of DNA and other macromolecules
- Isolation of total genomic DNA involves separating DNA from protein and other cellular components, for example by ethanol precipitation of DNA.
- PCR can be used as part of a sensitive method to detect the presence of a particular DNA sequence
- PCR can also be used as part of a method to isolate and prepare large quantities of a particular DNA sequence
- qPCR methodology allows the quantity of DNA product to be measured.
- RT-PCR methodology detects the quantity and quality of the mRNA, which indicates the spatial and temporal level gene expression.

KEY TERMS:

classical genetics
molecular biology
molecular genetics
macromolecules
lysis
detergent
chelating agent
EDTA
nuclease
supernatant
pellet
PCR
primer
thermalcycle
denature
anneal
extention
thermostable
Taq DNApol
electropheretic agarose gel
fluorochrome
Reverse Transcriptase PCR (RT-PCR)
Temporal level
Spacial level
StarLink affair
Cry9C gene

STUDY QUESTIONS:

1) What information, and what reagents would you need to use PCR to detect HIV in a blood sample?

2) If you started with 10 molecules of double stranded DNA template, what is the maximum number of molecules you would you have after 10 PCR cycles?

3) What is present in a PCR tube at the end of a successful amplification reaction? With this in mind, why do you usually only see a single, sharp band on a gel when it is analyzed by electrophoresis?

CHAPTER 13 – DNA SEQUENCING

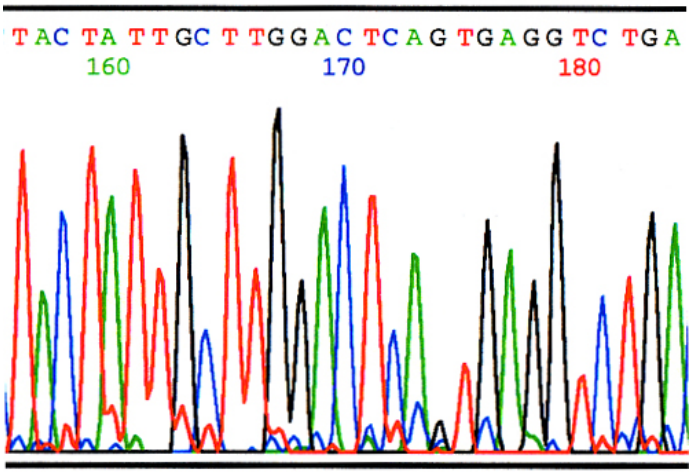


Figure 1.
Output from an automated Sanger DNA sequencer.
(Original-Harrington- CC BY-NC 3.0)

INTRODUCTION

DNA sequencing determines the order of nucleotide bases for a DNA molecule. These DNA molecules could be as small as a single restriction fragment, an entire gene, or as large as an organism's entire genome. Most DNA sequencing at the University of Alberta is done by the Molecular Biology Service Unit (MBSU). They use three machines: (1) an Applied Biosystems ABI 3730, (2) an Illumina MiSeq, and (3) an Illumina NextSeq 500, each has its own advantages and purposes. The 3730 uses an older technology called automated Sanger sequencing, while the Illumina machines perform next-generation DNA sequencing. This chapter will cover how these machines work and what they are used for.

1. AUTOMATED SANGER DNA SEQUENCING

1.1. HISTORICAL CONTEXT

DNA sequencing has had a long history. Beginning in the 1970s there have been many methods and improvements. Some dates that stand out are:

- 1977 - Frederick Sanger invents a popular method, later called manual Sanger sequencing.
- 1986 - Leroy Hood improves upon this method to invent **automated Sanger sequencing**.
- 1987 - **Applied Biosystems** begins selling a machine to perform automated Sanger sequencing, their ABI 370.
- 1995 to 2003 - Using ABI 370s, ABI 377s, and similar machines scientists in the US, UK, and other countries sequenced the human genome.
- 2002 - Applied Biosystems begins selling the **ABI 3730** (**Figure 2**) which became the most popular way to do automated Sanger sequencing and remains so to this day.

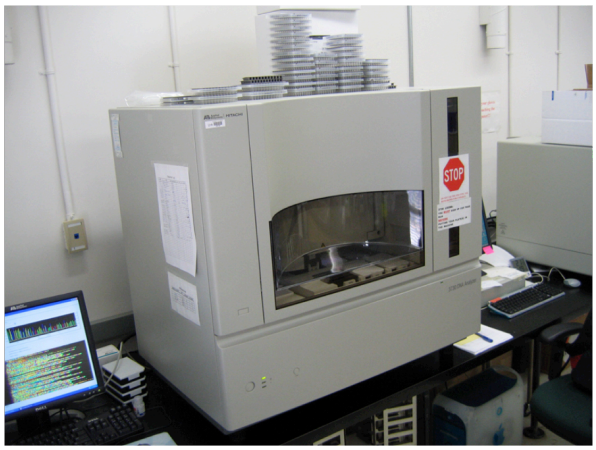


Figure 2.
The Applied Biosytems ABI 3730 in the MBSU (Molecular Biology Service Unit, Biological Sciences Department, U. of Alberta).
(Original-Harrington- CC BY-NC 3.0)

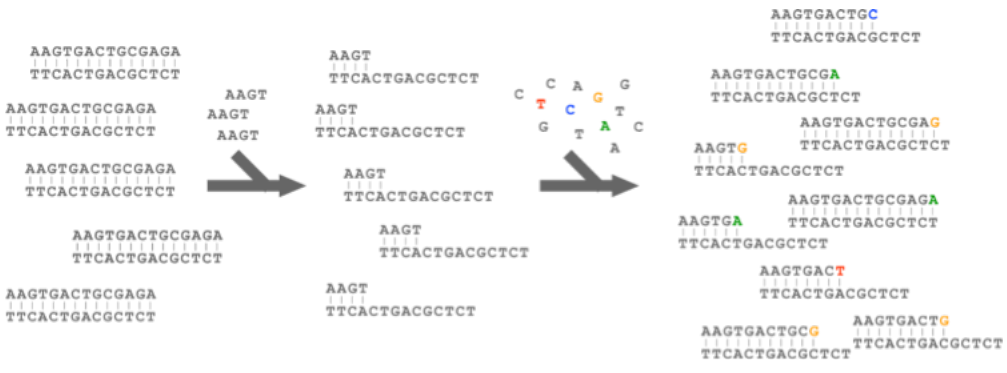


Figure 3.
An automated Sanger sequencing reaction. The regular dNTPs are shown here as black while the ddNTPs are in colour.
(Original-Deyholos- CC BY-NC 3.0)

1.2. HOW AUTOMATED SANGER DNA SEQUENCING WORKS

Recall that **DNA Polymerases** incorporate nucleotides (**dNTPs**) into a growing strand of DNA, based on the sequence of a template strand. DNA Polymerases add a new base only to the 3'-OH group of an existing strand of

DNA; this is why primers are required in natural DNA synthesis and in techniques such as PCR. Automated Sanger sequencing relies on the random incorporation of modified nucleotides called **dideoxy nucleotides (ddNTPs**, **Figure 3**).

These lack a 3'-OH group and therefore cannot serve as an attachment site for the addition of the next nucleotide. After a ddNTP is incorporated into a strand of DNA, no further elongation can occur. The ddNTPs are labelled with one of four fluorescent dyes, each specific for one the four nucleotide bases (**Figure 4**).

To sequence a DNA fragment, you need many copies of that fragment (**Figure 5**). Unlike PCR, Automated Sanger sequencing does not amplify the target sequence and only one **primer** is used. This primer is hybridized to the denatured template DNA, and determines where on the template strand the sequencing reaction will begin. A mixture of **regular dNTPs**, **fluorescently-labelled ddNTPs**, and DNA Polymerase is added to a tube containing the primer-template hybrid. The DNA Polymerase will then synthesize a new strand of DNA until a fluorescently-labelled ddNTP nucleotide is incorporated, at which point extension is terminated. Because the reaction contains millions of template molecules,

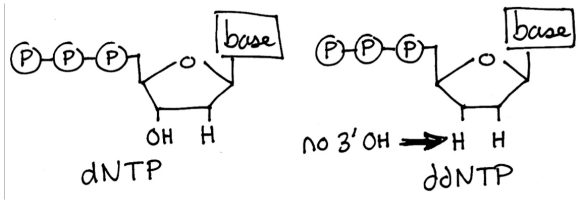


Figure 4.
ddNTPs are synthetic DNA nucleotides that lack a 3' hydroxyl group. If a DNA Polymerase uses one it can't continue.
(Original-Harrington- CC BY-NC 3.0)

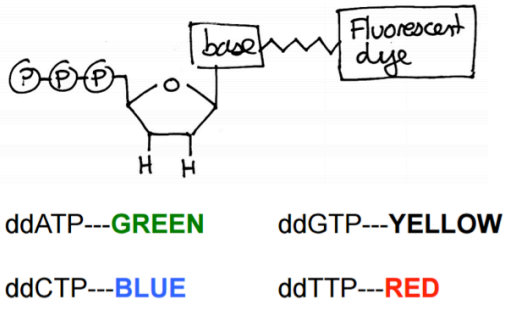


Figure 5.
Each ddNTP used in a sequencing reaction is attached to a different fluorescent molecule using a long chain of carbons. Note that black ink is used to represent yellow fluorescence.
(Original-Harrington- CC BY-NC 3.0)

a sufficient number of shorter molecules is synthesized, each ending in a fluorescent label that corresponds to the last base incorporated. The newly synthesized strands can be denatured from the template, and then separated electrophoretically based on their length (number of bases). The ABI machine is used for this electrophoresis step. While the original, old ABI 370 used a slab gel similar to the ones used in

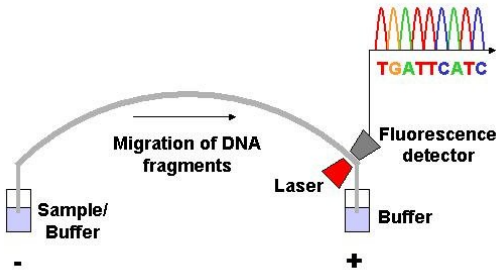


Figure 6.
Fluorescently labeled products can be separated by capillary electrophoresis, generating a chromatogram from which the sequence can be read.
(Wikimedia-Abizar-PD)

undergraduate labs, the newer ABI 3730 uses **capillary tube electrophoresis** (**Figure 6**). In this machine each sample travels through its own tube. Near the end of the tube is a laser, which excites any fluorescent dyes moving past and a detector that collects any emitted light. As each DNA molecule moves past the laser/detector it emits a specific colour. Because there will be several molecules with the same length and same colour the result appears as a peak of colour. A computer monitoring the results can add the sequence information to the colours since red = T and so on. In this way the DNA sequence can be read simply from the order of the colors in successive peaks.

The results from a sequencing reaction are presented as a **chromatograph**. While **Figure 6** only shows 9 peaks, a successful sequencing reaction will generate about 700 nucleotides worth of data. The figure shows the results from a single tube but in fact there can be 48 or 96 tubes in total. Thus in a single 'run' an ABI 3730 machine can sequence up to 67 000 bp of DNA.

1.3. USING AUTOMATED SANGER DNA SEQUENCING TO SEQUENCE A PLASMID

Making a new recombinant plasmid takes time and money. You will want to confirm that it has the DNA sequence it should before you use it for important experiments. A simple way to find out is to sequence it. Let's say you have put a 3.0 kb insert into a pBluescript II plasmid and right now the recombinant plasmid is in *E. coli* cells (**Figure 7**). The first step is to isolate plasmid DNA from some of the cells with a mini-prep protocol. This will be the template DNA. The primer will be oligonucleotides complementary to the pBluescript II vector adjacent to the insert. The sequencing reaction will tell you the sequence of the insert DNA within the plasmid.

1.4. USING AUTOMATED SANGER DNA SEQUENCING TO SEQUENCE A GENE

If you suspect that an organism has a mutation in a specific gene you can use automated Sanger sequencing to find out (**Figure 8**). Let's say you have a mouse strain and you think it has a mutation in a gene you are studying. As before, the first step is to isolate DNA. However, we can't sequence this DNA directly. Amongst all of the genomic DNA there just aren't enough copies of the gene to serve as the template DNA. To overcome this limitation a PCR reaction is used to amplify the gene sequence in question. Then we sequence the PCR product. Depending upon how large the gene is it may take several PCR products and several sequencing reactions to get the whole sequence.

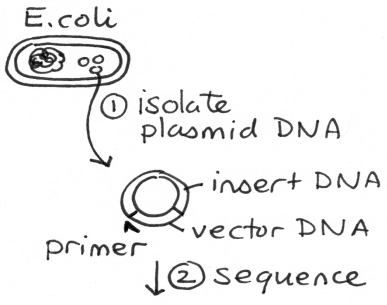


Figure 7
Using automated Sanger sequencing to sequence a plasmid.
(Original – Harrington- CC BY-NC 3.0)

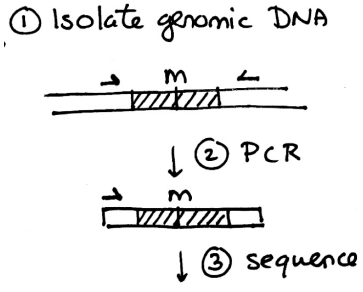


Figure 8.
Using automated Sanger sequencing to sequence a gene. The site of a point mutation is shown here as 'm'.
(Original –Harrington- CC BY-NC 3.0)

2. NEXT-GENERATION DNA SEQUENCING

2.1. HISTORICAL CONTEXT

Sequencing a single gene or plasmid with an ABI 3730 is quick and inexpensive. But sequencing a whole genome this way would be very slow and very expensive. There are two reasons for this.

The first is that automated Sanger sequencing requires many copies of the template DNA. A sample of purified plasmid DNA or purified PCR product has millions of copies of the target region. But a sample of genomic DNA has only a few copies of a specific target region. For many years the only way to sequence an organism was to isolate its genomic DNA, break the DNA into large pieces, and then clone these pieces into BAC (bacteria artificial chromosome) vectors. The BAC clones would then have to be sequenced one by one. Most of the 13 years and millions of dollars it took to sequence the human genome was spent making and organizing these BAC clones.

The second limitation of automated Sanger sequencing is that each reaction can only generate 700 nucleotides worth of data. It took literally millions of independent sequencing reactions to sequence the human genome.

Beginning in the late 1990s scientists realized that there was a need for a machine that could sequence genomic DNA directly and with a single reaction. In several instances a technology was invented in a university lab, developed in a small biotechnology company, and then purchased by a larger biotech company. An example of this is:

- 1996 - Swedish scientists invent a completely new way to sequence DNA called pyrosequencing. It is clever but very labour intensive.
- 2000 - An American inventor and entrepreneur, Jonathan Rothberg, refines their technique into automated pyrosequencing.
- 2004 - His company, 454 Life Sciences, markets the first so called **next-generation sequencing** machine.
- 2007 - The largest biotech company in the world, Roche, buys 454 Life Sciences.
- 2015 - 454, now a subsidiary of Roche, continues to develop and sell next-generation machines.

In 2015 there are several choices for next-generation sequencing. For a few hundred thousand dollars you can purchase a GS FLX (made by 454/Roche), an ABI 5500 (Applied Biosystems), or an Ion Proton (Life Technologies). Each is a fancy looking machine that uses a unique and proprietary technology.

2.2. HOW NEXT GENERATION DNA SEQUENCING WORKS

As mentioned in the introduction to this chapter, the MBSU recently purchased two next generation machines: an **Illumina MiSeq** and an **Illumina NextSeq 500** (**Figure 9**).

Both use a similar workflow (**Figure 10**). The scientist has to isolate genomic DNA from an organism (step 1) and then use a kit to break it into small fragments (step 2). The scientist then loads the fragments into the machine and turns it on. Once inside, the DNA fragments are isolated from each other (step 3), amplified in place (step 4), and finally sequenced (step 5). The technology is called **sequencing by synthesis**. Illumina has made animated movies of what happens within their machines:

www.youtube.com/watch?v=HMyCqWhwB8E

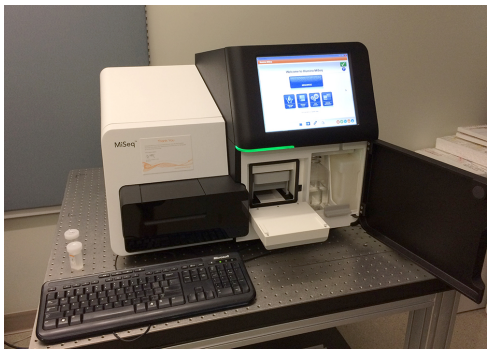


Figure 9.
The Illumina MiSeq in the MBSU (Biological Sciences Department, U. of Alberta). The door has been opened to show where the DNA sample and reaction mixtures are loaded. (Original-Harrington- CC BY-NC 3.0)

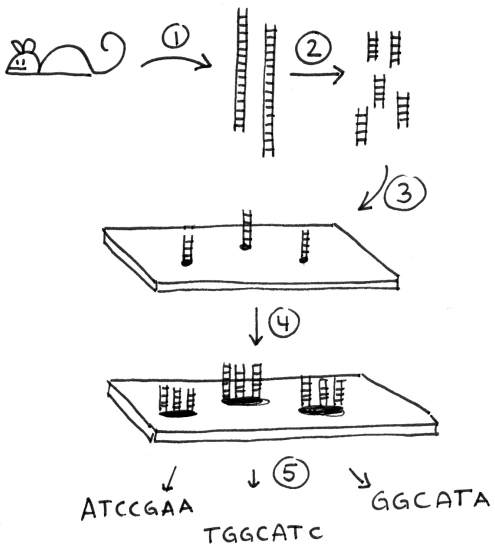


Figure 10.
Next generation DNA sequencing using Illumina's sequencing by synthesis technology. Steps 1 and 2 are done by the researcher and steps 3, 4, and 5 occur within the machine. (Original –Harrington- CC BY-NC 3.0)

Table 1. Comparison between different sequencing machines.

Machine	ABI 3730	Illumina MiSeq	Illumina NextSeq 500
DNA	plasmid or PCR product	genomic DNA	genomic DNA
Technology	automated Sanger	sequencing by synthesis	sequencing by synthesis
Data generated	700 bp	540 Mb to 15 Gb	16 Gb to 120 Gb
Price	$4.75	$1,250 - $1,850	$2,050 - $5,150

The output is just raw sequence data, there are no chromatograms. Powerful software is needed for **sequence assembly**, the process of joining these small pieces of sequence data into a continuous sequence (**Figure 11**). Ultimately there will be one sequence for each of the organism's chromosomes.

2.3. COMPARISON BETWEEN DNA SEQUENCING METHODS

Scientists all over the world now have a choice between automated Sanger sequencing and next-generation sequencing. For example, at the University of Alberta your choices at the MBSU are shown in **Table 1**.

Recall that DNA is measured in base pairs where:

1 kilobase (kb) = 1 000 base pairs (bp)
1 Megabase (Mb) = 1 000 000 bp
1 Gigabase (Gb) = 1 000 000 000 bp

Let's say you wanted to sequence a 2,000 bp long PCR product. You could do this with three sequencing reactions in the ABI 3730 or a single run in the Illumina MiSeq. The first method would cost $15 and the second would cost $1,250. Even though one machine is a decade older it is still the way to go! If you did use the MiSeq you'd end up sequencing the same PCR product over and over. It wouldn't produce any more data.

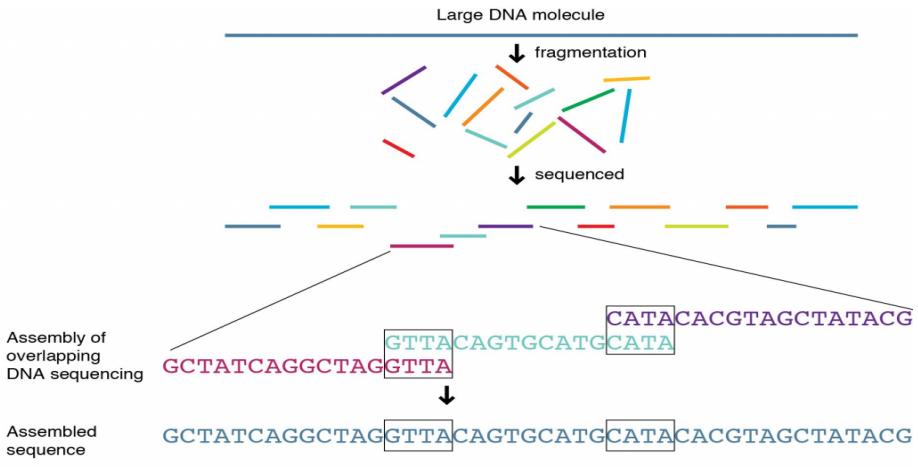


Figure 11.
Assembling the DNA sequence of a chromosome from many smaller sequences.
(NHGRI-Darryl Leja-PD)

On the other hand let's say you wanted to sequence your own DNA. Even if you don't consider the time and cost of making the BAC clones it would still cost millions of dollars to do all of the sequencing reactions in the ABI 3730. Conversely the MBSU could use their NextSeq 500 and have everything done in two days for about $4 000. Each of your 46 chromosomes would be sequenced about 30 times each. A more expensive machine, the Illumina HiSeq, can sequence human DNA for about $1 000 a person.

2.4. USING NEXT-GENERATION DNA SEQUENCING TO SEQUENCE HUMANS

Even though we know the *average* human DNA sequence, each of us is unique. There are two reasons why human DNA continues to be sequenced. (**Table 2**)

2.5. USING NEXT-GENERATION DNA SEQUENCING TO SEQUENCE OTHER ORGANISMS

Next-generation sequencing has made it feasible to sequence anything. Here are just a few examples. (**Table 3**)

Table 2. Using next-generation sequencing to sequence humans.

Use of next-generation sequencing	Description
Personalized genomics	If we sequence a person's DNA it can reveal information about their susceptibility to disease and their response to various medical treatments.
Tumour cell sequencing	If a person has cancer it is now possible to sequence individual cancer cells. This has revolutionized how physicians help their patients. Instead of treatments based upon the location of tumours, treatments can now be designed around the genetic defects that lead to the cells becoming cancerous in the first place.

Table 3. Using next-generation sequencing to sequence other organisms.

Use of next-generation sequencing	Description
***De novo* sequencing**	This is when an organism is sequenced for the first time. For example in 2014 researchers in Sierra Leone sequenced 99 Ebola virus genomes from 78 patients. They identified changes in the virus that caused the recent outbreak.
Metagenomics	This is when the entire collection of DNA in an environment is sequenced to determine which species are present. This technique has been used to show that a person's gut microbes vary with their diet.
RNA Seq	This is when RNAs from a tissue or organ are isolated, copied into DNA molecules, and then sequenced. This reveals which genes were active in the cell, tissue, or organ.

SUMMARY:

- Automated Sanger sequencing became commonplace in 1987 and is still used today. It is used to sequence plasmids and PCR products. The most popular machines are Applied Biosystem's ABI 3730s.
- Next-generation sequencing began in 2004 as a better way to sequence whole genomes. There are several competing technologies, for example Illumina's MiSeq machine and its sequencing by synthesis technology.
- Sequencing centres offer both types of sequencing today.
- Sequencing any DNA molecules, large or small, is now fast and inexpensive.

KEY TERMS:

automated Sanger sequencing
Applied Biosystems ABI 3730
DNA Polymerases
primer
regular dNTPs
fluorescently-labelled ddNTPs
capillary tube electrophoresis
chromatogram
next-generation sequencing
Ilumina MiSeq
Illumina NextSeq 500
sequencing by synthesis
sequence assembly
personalized genomics
tumour cell sequencing
de novo sequencing
metagenomics
RNA Seq

STUDY QUESTIONS:

1) What would the chromatogram look like if you set up an automated Sanger sequencing reaction with only template, primers, polymerase, and fluorescent ddNTPs?

2) How could you use DNA sequencing to identify new species of marine microorganisms?

3) An alternative name for automated Sanger sequencing is dye-terminator sequencing. Why is this term appropriate?

4) Ten years ago it would have cost $100,000,000 to sequence your DNA. Today it would cost as little as $1,000. Why did the cost go down so much?

5) Why haven't next-generation machines completely replaced the first generation of automated DNA sequencers?

6) True or false: Automated pyrosequencing and sequencing by synthesis are both considered next-generation DNA sequencing technologies.

Notes:

CHAPTER 14 – PROKARYOTE GENES: *E. COLI* LAC OPERON

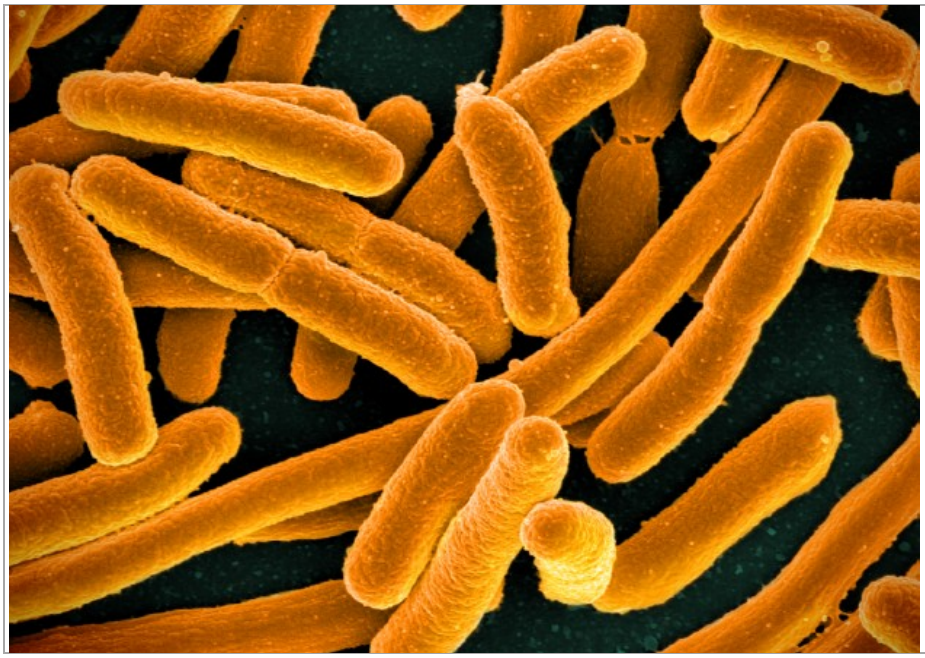


Figure 1.
Electron micrograph of growing *E. coli.* Some show the constriction at the location where daughter cells separate. The colouring is false.
(Flickr-NIAID-CC BY 2.0)

INTRODUCTION

With most organisms, every cell contains essentially the same genomic sequence. How then do cells develop and function differently from each other? The answer lies in the regulation of **gene expression**. Only a subset of all the genes is expressed (i.e. are functionally active) in any given cell participating in a particular biological process. Gene expression is regulated at many different steps along the process that converts DNA information into active proteins. In the first stage, transcript abundance can be controlled by regulating the rate of transcription initiation and processing, as well as the degradation of transcripts. In many cases, higher abundance of a gene's transcripts is correlated with its increased expression. We will focus on **transcriptional regulation** in *E. coli* (**Figure 1**). Be aware, however, that cells also regulate the overall activity of genes in other ways. For example, by controlling the rate of mRNA translation, processing, and degradation, as well as the post-translational modification of proteins and protein complexes.

1. THE *LAC* OPERON – A MODEL PROKARYOTE GENE

Early insights into mechanisms of transcriptional regulation came from studies of *E. coli* by researchers Francois Jacob & Jacques Monod (see Section 4 on page 4). In *E. coli,* and many other bacteria, genes encoding several different polypeptides may be located in a single transcription unit called an **operon**. The genes in an operon share the same transcriptional regulation, but are translated individually into separate polypeptides. Most prokaryote genes are not organized as operons, but are transcribed as single polypeptide units.

Eukaryotes do not group genes together as operons (an exception is *C. elegans* and a few other species).

1.1. BASIC LAC OPERON STRUCTURE

E. coli encounters many different sugars in its environment. These sugars, such as **lactose** and **glucose**, require different enzymes for their metabolism. Three of the enzymes for lactose metabolism are grouped in the ***lac* operon**: ***lacZ***,

lacY, and ***lacA*** (**Figure 2**). *LacZ* encodes an enzyme called **β-galactosidase**, which digests lactose into its two constituent sugars: glucose and galactose. *lacY* is a **permease** that helps to transfer lactose into the

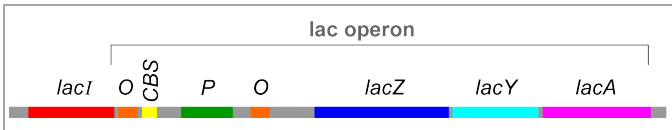


Figure 2.
Diagram of a segment of an E. coli chromosome containing the lac operon, as well as the lacI coding region. The various genes and cis-elements are not drawn to scale. (Original-Deyholos-CC BY-NC 3.0)

cell. Finally, *lacA* is a **trans-acetylase**; the relevance of which in lactose metabolism is not entirely clear. Transcription of the *lac* operon normally occurs only when lactose is available for it to digest. Presumably, this avoids wasting energy in the synthesis of enzymes for which no substrate is present. In the *lac* operon, there is a single mRNA transcript that includes coding sequences for all three enzymes and is called a polycistronic mRNA. A cistron in this context is equivalent to a gene.

1.2. *CIS*- AND *TRANS*- REGULATORS

In addition to these three protein-coding genes, the *lac* operon contains several short DNA sequences that do not encode proteins, but instead act as binding sites for proteins involved in transcriptional regulation of the operon. In the *lac* operon, these sequences are called **P (promoter)**, **O (operator)**, and **CBS (CAP-binding site).** Collectively, sequence elements such as these are called ***cis*-elements** because they must be located adjacently to the same piece of DNA in order to perform correctly. On the other hand, elements outside from the target DNA (such as the proteins that bind to these *cis*-elements) are called ***trans*-regulators** because (as diffusible molecules) they do not necessarily need to be encoded on the same piece of DNA as the genes they regulate.

2. NEGATIVE REGULATION – INDUCERS AND REPRESSORS

2.1. *lacI* ENCODES AN ALLOSTERICALLY REGULATED REPRESSOR

One of the major *trans*-regulators of the *lac* operon is encoded by ***lacI***, a gene located just upstream from the lac operon (**Figure 2**). Four identical molecules of *lacI* proteins assemble together to form a **homotetramer** called a **repressor** (**Figure 3**). This repressor is **trans-acting** and binds to two cis-acting operator sequences adjacent to the promoter of the *lac* operon. Binding of the repressor prevents RNA polymerase from binding to the promoter (**Figure 2**, **Figure 5.**). Therefore, the operon is not transcribed when the operator sequence is occupied by a repressor.

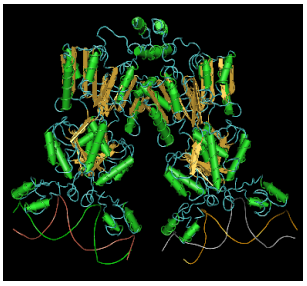


Figure 3.
Structure of *lacI* homotetramer bound to DNA.
(Original-Deyholos- CC BY-NC 3.0)

2.2. THE REPRESSOR ALSO BINDS LACTOSE (ALLOLACTOSE)

Besides its ability to bind to specific DNA sequences at the at the operator, another important property of the *lacI* protein *lacI* protein is its ability to bind to allolactose. If lactose is lactose is present, **β-galactosidase** (β-gal) enzymes convert a convert a few of the lactose molecules into allolactose (

Figure 4.).

This allolactose can then bind to the *lacI* protein. This alters the shape of the protein in a way that prevents it from binding to the operator. Proteins which change their shape and functional properties after binding to a ligand are said to be regulated through an **allosteric** mechanism.

Therefore, in the presence of lactose (which β-gal converts to allolatose), the repressor doesn't bind the operator sequence and thus RNA polymerase is

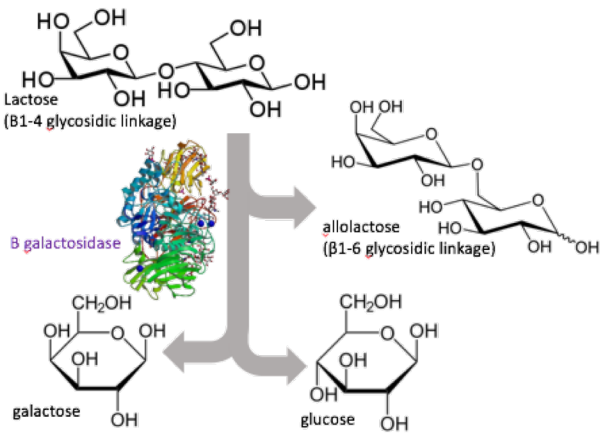


Figure 4.
β-galactosidase catalyzes two reactions: one converts lactose to allolactose; the other splits lactose into galactose and glucose. (Original-Nickle-CC BY-NC 4.0)

able to bind to the promoter and transcribe the *lac* operon. This leads to a moderate level of expression of the mRNA encoding the *lacZ*, *lacY*, and *lacA* genes. This kind of secondary molecule that binds to either activator or repressor and induces the production of specific enzyme is called an **inducer**. The role of *lacI* in regulating the *lac* operon is summarized in **Figure 5**.

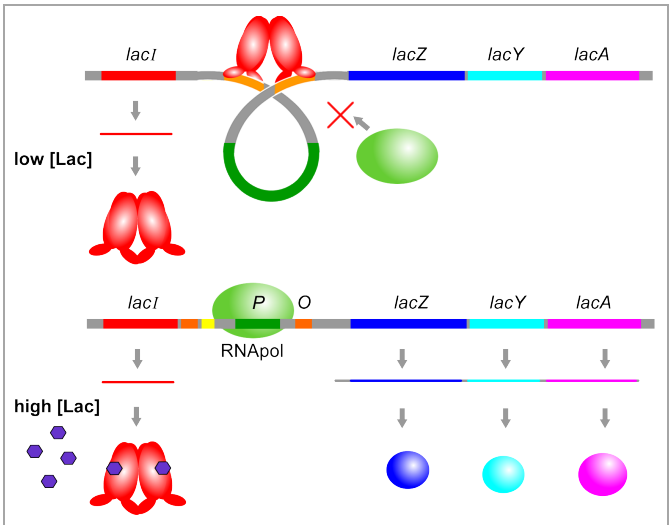


Figure 5.
When the concentration of lactose [Lac] is low, *lacI* tetramers bind to operator sequences (O), thereby blocking binding of RNApol (green) to the promoter (P). Alternatively, when [Lac] is high, lactose binds to *lacI*, preventing the repressor from binding to O, and allowing transcription by RNApol.
(Original-Deyholos-CC BY-NC 3.0)

3. POSITIVE REGULATION – *CAP*, cAMP & POLYMERASE

A second aspect of *lac* operon regulation is conferred by a *trans*-acting factor called **cAMP binding protein** (**CAP**, **Figure 6**). CAP is another example of an allosterically regulated *trans*-factor. Only when the CAP protein is bound to cAMP can another part of the protein bind to a specific *cis*-element within the *lac* promoter called the **CAP binding sequence (CBS)**. CBS is immediately in front of the promoter (P), and thus is a *cis-acting* element. When CAP is bound to at the CBS, RNA polymerase is better able to bind to the promoter and initiate transcription. Thus, the presence of cAMP ultimately leads to a further increase in *lac* operon transcription.

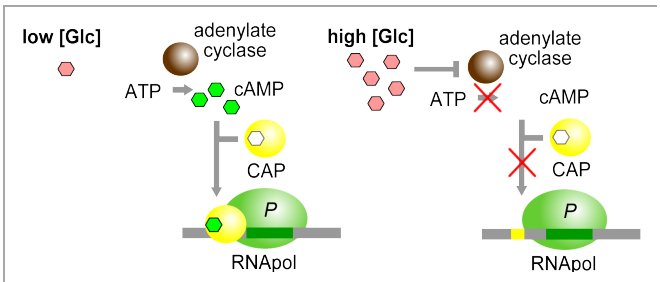


Figure 6.
CAP, when bound to cAMP, increases RNApol binding to the lac operon promoter. cAMP is produced only when glucose [Glc] is low. (Original-Deyholos-CC BY-NC 3.0)

The physiological significance of regulation by cAMP becomes more obvious in the context of the following information. The concentration of cAMP is inversely proportional to the abundance of glucose. When glucose concentrations are low, an enzyme called **adenylate cyclase** is able to produce cAMP from ATP. Evidently, *E. coli* prefers glucose over lactose, and so expresses the *lac* operon at high levels only when glucose is absent and lactose is present. This provides another layer of adaptive control of *lac* operon expression: only in the presence of lactose **and** in the absence of glucose is the operon expressed at its highest levels.

4. THE USE OF MUTANTS TO STUDY THE *LAC* OPERON

4.1. SINGLE MUTANTS OF THE *LAC* OPERON

The *lac* operon and its regulators were first characterized by studying mutants of *E. coli* that exhibited various abnormalities in lactose metabolism. Mutations can occur in any of the *lacZ*, *lacY*, and *lacA* genes. Such mutations result in altered protein sequences, and cause non-functional products. These are mutations in the protein coding sequences (non-regulatory).

Other mutants can cause the lac operon to be expressed constitutively, meaning the operon was transcribed whether or not lactose was present in the medium. Remember that normally the operon is only transcribed if lactose is present. Such mutants are called **constitutive** mutants. Constitutive mutants are always on and are unregulated by inducers. These include *lacO* and *lacI* genes.

4.2. OPERATOR MUTATIONS

The operator locus (*lacO*) - One example is ***O***$^{-}$, in which a mutation in an operator sequence reduces or precludes the repressor (the *lacI* gene product) from recognizing and binding to the operator sequence. Thus, in O^- mutants, *lacZ*, *lacY*, and *lacA* are expressed whether or not lactose is present. Note that this mutation is **cis dominant** (only affects the genes on the same chromosome) but not in trans (other DNA molecule). Another common name for a defective operator is O^c to refer to its constitutive expression. O^- and O^c are synonyms.

Note that constitutively expressed O$^-$ mutants may not be maximally expressed, and the extent of the mutation can also affect the level of expression. (**Table 1**)

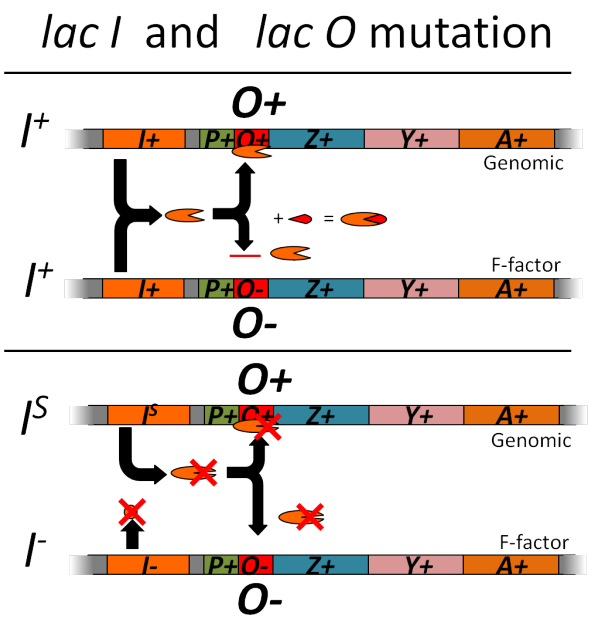


Figure 7.
Both *E. coli* strands with genotypes $I^+O^+Z^-Y^+A^+/F\ I^+\mathbf{O}^-Z^+Y^-A^+$ and $I^SO^+Z^+Y^+A^+/F\ I^-\mathbf{O}^-Z^+Y^+A^+$ will induce all the *lac* genes because repressor cannot bind to the O$^-$ sequence on the F-factor and cannot prevent transcription.
(Original-Locke-CC BY-NC 3.0)

Table 1. Constitutively expressed O^c mutants may not be maximally expressed and have various levels of expression.

Level	Genotype	Explanation
100%	lac I$^-$ O^c	no repressor
10-20%	lac I$^+$ O^c	repressor fails to bind tightly
~1%	P$^+$ O^c, high glucose	basal transcription, constitutive
0%	P$^-$ or Z$^-$	no transcription

4.3. INDUCER MUTATIONS (*lacI* LOCUS)

The *lacI* locus has two types of mutations: ***I***$^{-}$ and ***I***^{S}**.**

One class of mutant allele for *lacI* (called ***I***$^{-}$) either (1) prevents the production of a repressor polypeptide, or (2) produces a polypeptide that cannot bind to the operator sequence. Note that these two alleles would have different genetic sequences, but the phenotype is the same. Theoretically, we should have better locus names to properly identify specific alleles, but for now we'll group them under one name. With this form of mutation, the repressor *cannot* bind and transcription *can* occur without the presence of inducer (allolactose). This can also be referred to as a "constitutive expresser" of the *lac* operon: the absence of repressor binding permits transcription.

Note that I^+ is dominant over I^-. For example, in *E. coli* strain with $I^+Z^-Y^+A^+/F\ I^-Z^+Y^-A^+$, the *lac* genes will not be inducible because the I^+ allele will still

produce functional repressors that bind to operator sequence, preventing transcription. (**Figure 8**)

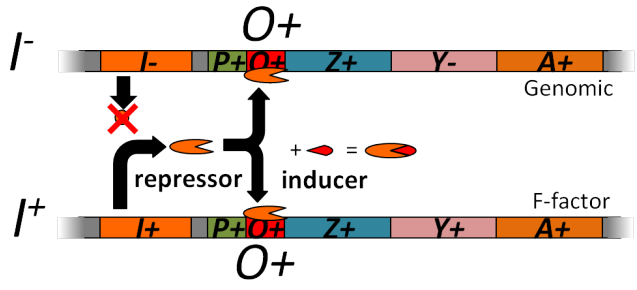


Figure 8.
E. coli strain with genotype $I^+Z^-Y^+A^+/F\ I^-Z^+Y^-A^+$ will not produce *lac Z, lac Y* and *lac A* products.
(Original-Locke-CC BY-NC 3.0)

The other class of mutant alleles for *lacI* are called I^s. The altered amino sequence of their proteins remove the "allosteric site". This means the repressor polypeptide cannot bind allolactose – and therefore it cannot change its shape. Even in the presence of lactose, it remains attached to the operator, so the *lac Z, lacY* and *lac A* genes cannot be expressed (no RNA polymerase, no protein). This mutant constitutively represses the *lac* operon whether lactose is present or not. The lac operon is not expressed at all and this mutant is called a "super-repressor". I^s is therefore dominant to both

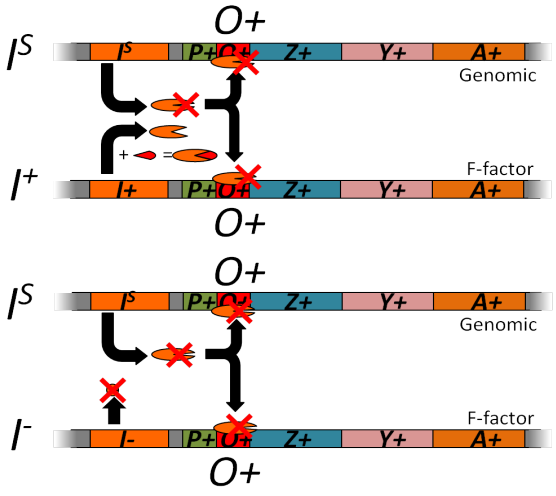


Figure 9.
E. coli strains with genotypes 1) $I^SZ^+Y^+A^+/F\ I^+Z^+Y^-A^+$ and 2) $I^SZ^+Y^+A^+/F\ I^-Z^+Y^+A^+$ will not produce *lac Z, lac Y* and *lac A* products. (Original-Locke-CC BY-NC 3.0)

I^+ and I^- in trans. Therefore, *E. coli* strains with the genotypes 1) $I^SZ^+Y^+A^+/F\ I^+Z^+Y^-A^+$ and 2) $I^SZ^+Y^+A^+/F\ I^-Z^+Y^+A^+$, the *lac Z*, *lac Y* and *lac A* genes will not be inducible (**Figure 9**).

The repressor protein encoded by *lacI* gene has at least two independent functional domains. This is the reason why it different mutations can give different classes of mutant phenotypes. **(Figure 10)**

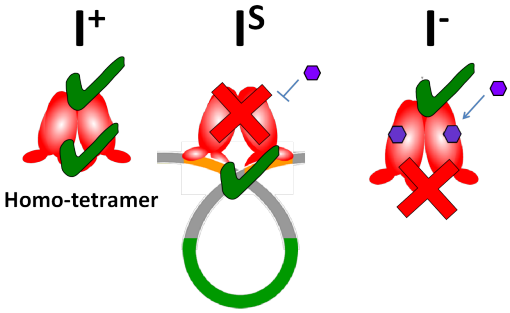


Figure 10.
Because the repressor encoded by lac I gene has independent domains, mutations can also occur independently.
(Original-Deyholos/Locke-CC BY-NC 3.0)

4.4. THE F-FACTOR AND TWO *lac* OPERONS IN A SINGLE CELL – PARTIAL DIPLOID IN *E. COLI*

More can be learned about the regulation of the lac operon when two different copies (each containing mutations) are present in one cell. This can be accomplished by using an **F-factor** (also known as a plasmid) to carry one copy, while the other is on the genomic *E. coli* chromosome. This results in a partial diploid in *E. coli* that contains two independent copies (alleles) of the lac operon and *lacI*.

An F-factor (named so because it creates "fertility" in the cell which contains it) is an **episome**. This is a self-replicating extrachromosomal piece of DNA: it is outside of the large, circular bacterial chromosome but has its own origin of replication. Jacob and Monod explored the regulation of the *lac* operon by introducing different genotypes into bacterial cells and noting how they responded in the presence of absence of sugars, namely glucose and lactose using the mutant classes discussed above.

For example, the genotype of a host bacterium that has a *lacI*$^-$ gene that is supplied with F factor containing *lacI*$^+$ can be written as *lacI*$^-$/*F*$^+$*lacI*$^+$. In fact, it doesn't matter which piece of DNA (genomic or episome) has which operon, so we can leave off the "F" altogether – equivalent notation is *lacI*$^-$/*lacI*$^+$.

Thus, cells with two copies of a gene sequence are partial diploids, called **merozygotes** (**Figure 11**). These allow researchers to test the regulation with different combinations of mutations in one cell. For example, the F-factor copy may have a I^S mutation while the genomic copy might have an O^C mutation. How would this cell respond to the presence/absence of lactose (or glucose)? This partial diploid can be used to determine that I^S is dominant to I^+, which in turn is dominant to I^-. It can also be used to show the O^C mutation only acts in *cis*- while the *lacI* mutation can act in *trans*- .

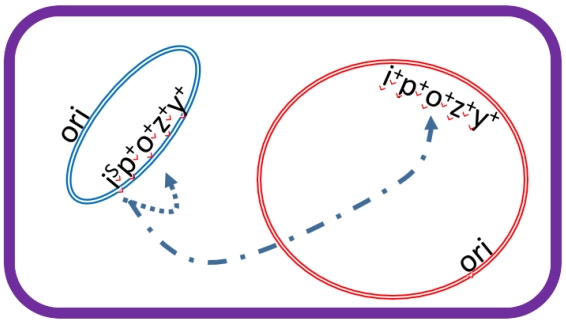


Figure 11.
A diagram showing a merozygote. This is a form of bacterial cell where a plasmid (blue) has a normal operon but a *lacI*^{s} allele for the repressor. Not all bacterial cells have plasmids. The chromosome (red) has a normal repressor and the operon has a *lacY*$^-$ allele. The super-repressor will inactivate both operons because it will bind to both operators (a trans-acting effect). Note that both the plasmid and chromosome have their own origins of replication. (Original-Nickle-CC BY-NC 4.0)

5. SUMMARY

In positive regulation, **low** levels of glucose (inducer) allow adenylate cyclase to produce cAMP from ATP, which binds to CAP protein. CAP protein can bind to DNA and increase the level of transcription. **High** glucose level halts adenylate cyclase from producing cAMP from ATP. Hence, cAMP will not bind to CAP protein and in turn, CAP will not bind to DNA and the level of transcription would be low.

In negative regulation, the repressor protein acts to prevent transcription. Inducer binds to repressor to alter conformation so it no longer binds to the operator sequence and transcription can take place. **High** levels of lactose (inducer) would allosterically inhibit repressor and therefore would not prevent transcription. **Low** levels of lactose would not cause inhibition to the repressor, so transcription would be prevented. Various forms of regulation in the lac operon are found in **Figure 12**.

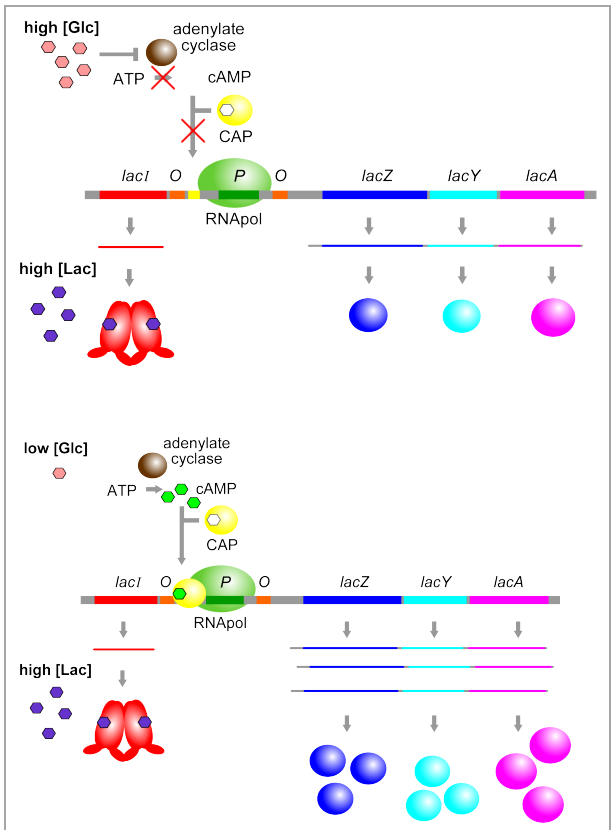


Figure 12.
When glucose [Glc] and lactose [Lac] are both high, the lac operon is transcribed at a moderate level, because CAP (in the absence of cAMP) is unable to bind to its corresponding cis-element (yellow) and therefore cannot help to stabilize binding of RNApol at the promoter. Alternatively, when [Glc] is low, and [Lac] is high, CAP and cAMP can bind near the promoter and increase further the transcription of the lac operon. (Original-Deyholos-CC BY-NC 3.0)

SUMMARY:

- Regulation of gene expression is essential to the normal development and efficient functioning of cells
- Gene expression may be regulated by many mechanisms, including those affecting transcript abundance, protein abundance, and post-translational modifications
- Regulation of transcript abundance may involve controlling the rate of initiation and elongation of transcription, as well as transcript splicing, stability, and turnover
- The rate of initiation of transcription is related to the presence of RNA polymerase and associated proteins at the promoter.
- RNApol may be blocked from the promoter by repressors, or may be recruited or stabilized at the promoter by other proteins including transcription factors
- The *lac* operon is a classic, fundamental paradigm demonstrating both positive and negative regulation through allosteric effects on *trans*-factors.

KEY TERMS:

gene expression
transcriptional regulation
operon
lactose
glucose
lac operon
lacZ
lacY
lacA
β-galactosidase
permease
trans-acetylase
P / promoter
O / operator
CBS
CAP-binding site
cis-elements
trans-regulators
lacI
homotetramer
repressor
inducer
allosteric
cAMP binding protein
CAP
CAP binding sequence
CBS
adenylate cyclase
constitutive
O^c / I^- / I^s
cis dominant
cis-acting factors
trans dominant
trans-acting factors
F-factor / episome
merozygotes

STUDY QUESTIONS:

1) List all points in the "central dogma" of gene action the mechanisms that can be used to regulate gene expression.

2) With respect to the expression of β-galactosidase, what would be the phenotype of each of the following strains of *E. coli*?

Use the symbols:

+++ Lots of β-galactosidase activity (100%)

\+ Moderate β-galactosidase activity (10-20%)

− No β-galactosidase activity (0-1%)

a) I^+, O^+, Z^+, Y^+ (no glucose, no lactose)
b) I^+, O^+, Z^+, Y^+ (no glucose, high lactose)
c) I^+, O^+, Z^+, Y^+ (high glucose, no lactose)
d) I^+, O^+, Z^+, Y^+ (high glucose, high lactose)
e) I^+, O^+, Z^-, Y^+ (no glucose, no lactose)
f) I^+, O^+, Z^-, Y^+ (high glucose, high lactose)
g) I^+, O^+, Z^+, Y^- (high glucose, high lactose)
h) I^+, O^c, Z^+, Y^+ (no glucose, no lactose)
i) I^+, O^c, Z^+, Y^+ (no glucose, high lactose)
j) I^+, O^c, Z^+, Y^+ (high glucose, no lactose)
k) I^+, O^c, Z^+, Y^+ (high glucose, high lactose)
l) I^-, O^+, Z^+, Y^+ (no glucose, no lactose)
m) I^-, O^+, Z^+, Y^+ (no glucose, high lactose)
n) I^-, O^+, Z^+, Y^+ (high glucose, no lactose)
o) I^-, O^+, Z^+, Y^+ (high glucose, high lactose)
p) I^s, O^+, Z^+, Y^+ (no glucose, no lactose)
q) I^s, O^+, Z^+, Y^+ (no glucose, high lactose)
r) I^s, O^+, Z^+, Y^+ (high glucose, no lactose)
s) I^s, O^+, Z^+, Y^+ (high glucose, high lactose)

3) In the *E. coli* strains listed below, some genes are present on both the chromosome, and the extrachromosomal F^-factor episome. The genotypes of the chromosome and episome are separated by a slash. What will be the β^- galactosidase phenotype of these strains? All of the strains are grown in media that lacks glucose.

Use the symbols:

+++ Lots of β-galactosidase activity (100%)

\+ Moderate β-galactosidase activity (10-20%)

− No β-galactosidase activity (0-1%)

a) $I^+, O^+, Z^+, Y^+ / I^+, O^-, Z^-, Y^-$ *(high lactose)*
b) $I^+, O^+, Z^+, Y^+ / I^+, O^-, Z^-, Y^-$ (no lactose)
c) $I^+, O^+, Z^-, Y^+ / I^+, O^-, Z^+, Y^+$ (high lactose)
d) $I^+, O^+, Z^-, Y^+ / I^+, O^-, Z^+, Y^+$ (no lactose)
e) $I^+, O^+, Z^-, Y^+ / I^-, O^+, Z^+, Y^+$ (high lactose)
f) $I^+, O^+, Z^-, Y^+ / I^-, O^+, Z^+, Y^+$ (no lactose)
g) $I^-, O^+, Z^+, Y^+ / I^+, O^+, Z^-, Y^+$ (high lactose)
h) $I^-, O^+, Z^+, Y^+ / I^+, O^+, Z^-, Y^+$ (no lactose)
i) $I^+, O^c, Z^+, Y^+ / I^+, O^+, Z^-, Y^+$ (high lactose)
j) $I^+, O^c, Z^+, Y^+ / I^+, O^+, Z^-, Y^+$ (no lactose)
k) $I^+, O^+, Z^-, Y^+ / I^+, O^c, Z^+, Y^+$ (high lactose)
l) $I^+, O^+, Z^-, Y^+ / I^+, O^c, Z^+, Y^+$ (no lactose)
m) $I^+, O^+, Z^-, Y^+ / I^s, O^+, Z^+, Y^+$ (high lactose)
n) $I^+, O^+, Z^-, Y^+ / I^s, O^+, Z^+, Y^+$ (no lactose)
o) $I^s, O^+, Z^+, Y^+ / I^+, O^+, Z^-, Y^+$ (high lactose)
p) $I^s, O^+, Z^+, Y^+ / I^+, O^+, Z^-, Y^+$ (no lactose)

4) What genotypes of *E. coli* would be most useful in demonstrating that the *lacO* operator is a *cis*-acting regulatory factor?

5) What genotypes of *E. coli* would be useful in demonstrating that the *lacI* repressor is a *trans*-acting regulatory factor?

6) What would be the effect of the following loss-of-function mutations on the expression of the *lac* operon?

a) loss-of-function of adenylate cyclase
b) loss of DNA binding ability of CAP
c) loss of cAMP binding ability of CAP
d) mutation of CAP binding site (CBS) *cis*-element so that CAP could not bind

CHAPTER 15 – RECOMBINANT DNA

Figure 1.
The manipulation of DNA often involves small quantities of liquids that are accurately dispensed using micro-pipettors into small plastic microfuge tubes. Volumes as small as 0.5 μl are routinely dispensed for some reactions. The use of clean, sterile plastic tubes and tips ensures the reactions work correctly and are reproducible.
(Flickr- University of Michigan School of Natural Resources and Environment-CC BY 2.0)

INTRODUCTION

Recombinant DNA is a general term to describe DNA that has been manipulated (recombined) somehow *in vivo*. It typically involves the breakage of DNA into fragments, using restriction enzymes, and the rejoining (ligation) of these fragments into various arrangements and into vectors, such as plasmids, to propagate the new arrangement for further analysis, like sequencing, or for insertion into other hosts, such as model organism as transgenes.

1. BASIC TERMINOLOGY

Before proceeding any further, there are some basic terminologies that students should know regarding recombinant DNA technology.

***in vivo* (in life)** experiments done within a living cell/organism

***in situ* (in place)** experiments done on cells and structures removed intact from an organism. (ex. Inserting RNA into a frog egg cell on a petri dish)

***in vitro* (in glass)** experiments done on individual molecules removed from an organism (ex. DNA in a test tube). These days most experiments are done ***in plastico* (in plastic).** See **Figure 1**.

***in silico* (in silicon)** Experiments done within a computer simulation.

Recombinant DNA: a composite DNA molecule created *in vitro* by joining a foreign DNA with a vector DNA molecule. (Note; technically recombinant DNA can be also made *in vivo* during meiosis in an organism, but this is usually not the typical meaning of these words.)

2. RECOMBINANT DNA TECHNIQUES:

There are many techniques for joining DNA molecules *in vitro* and introducing them into cells (usually bacteria) where the molecules are then replicated along with the host genomic DNA.

2.1. PLASMIDS ARE NATURALLY PRESENT IN SOME BACTERIA

Many bacteria contain extra-chromosomal DNA elements called **plasmids**. These are usually small (a few 1000 bp), circular, double stranded molecules that replicate independently of the chromosome and can be present in multiple copies within a cell. In the wild, plasmids can be transferred between

individuals during bacterial mating and are sometimes even transferred between different species. Plasmids are particularly important in medicine because they often carry genes for pathogenicity (making the bacteria more detrimental) and drug-resistance (able to survive various antibiotics). In the lab, plasmids are inserted into bacterial hosts in a process called **transformation**. These plasmids can be modified by the addition of foreign DNA so that both the plasmid vector and the target foreign DNA is replicated.

There are 3 main features of a plasmid (**Figure 2**):
(1) **Origin of replication (Ori)** which is similar in function to *oriC* in *E. coli* chromosome.
(2) **Selectable marker** gene that helps to screen the desired and undesired strains, which is usually an antibiotic resistance gene like ampR, tetR, or kanR. Cells with R6K are resistant to penicillin, ampicillin, etc.
(3) **Multiple cloning site (MCS)** that has many restriction enzyme sites in a short sequence.

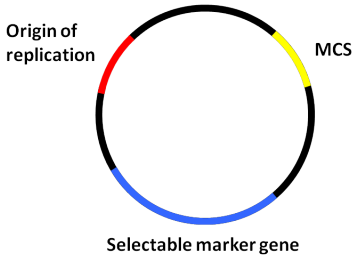


Figure 2.
Basic structure of a double stranded bacterial plasmid, represented by a circle.
(Original-Locke-CC BY-NC 3.0)

3. USING CLONING VECTORS

3.1. PLASMIDS VECTORS

There are multiple steps to using plasmids as cloning vectors. To insert a DNA fragment into a plasmid, both the fragment and the circular plasmid are cut using a restriction enzyme that produces compatible ends. Given the large number of restriction enzymes that are currently available, it is usually not too difficult to find an enzyme for which corresponding recognition sequences are present in both the plasmid and the DNA fragment, particularly because most plasmid vectors used in molecular biology have been engineered to contain recognition sites for a large number of restriction endonucleases in a segment called the Multiple Cloning Site (MCS).

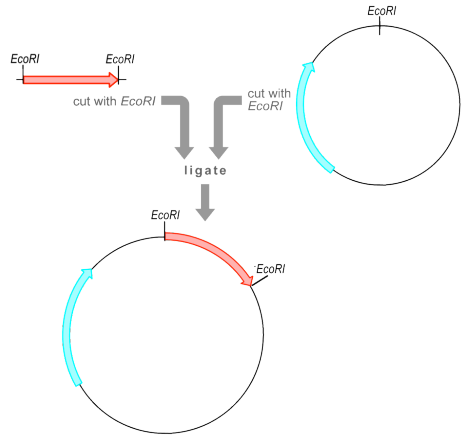


Figure 3.
Cloning of a DNA fragment (red) into a plasmid vector. The vector already contains a selectable marker gene (blue) such as an antibiotic resistance gene.
(Original-Deyholos-CC BY-NC 3.0)

After restriction digestion, the desired fragments may be further purified or selected before they are mixed together with ligase to join them together. Following a short incubation, the newly ligated plasmids, containing the gene of interest are **transformed** into *E. coli*.

Transformation is accomplished by mixing the ligated DNA with *E. coli* cells that have been specially prepared (i.e. made **competent**) to uptake DNA. Bacterial cells can be made competent by exposure to compounds such as $CaCl_2$ or to electrical fields (**electroporation**). Because only a small fraction of cells that are mixed with DNA will actually be transformed, a **selectable marker**, such as a gene for antibiotic resistance, is usually also present on the plasmid and used to select those few cells that have taken up the DNA. The rate of DNA uptake varies each time and is called **transformation efficiency**. This can range from ~10^5-10^{10} colonies per μg of DNA.

After transformation (combining DNA with competent cells), bacteria are spread on a bacterial agar plate containing an appropriate antibiotic so that only those cells that have actually incorporated the plasmid will be able to grow and form colonies.

Colonies (clone) can then be picked and used for further study.

Molecular biologists use plasmids as **vectors** to contain, amplify, transfer, and sometimes express genes of interest that are present in the cloned DNA. Often, the first step in a molecular biology experiment is to "**clone** a gene" (i.e. make a copy) into a plasmid, then transform this recombinant plasmid into bacteria so that essentially unlimited copies of the gene (and the plasmid that carries it) can be made as the bacteria reproduce. This is a practical necessity for further manipulations of the DNA, since most techniques of molecular biology require many copies of DNA to work. Even though small amounts are needed they are not sensitive enough to work with just a single molecule at a time.

Many molecular cloning and recombination experiments are therefore iterative (repetitive) processes. For example:

1. a DNA fragment (usually isolated by PCR and/or restriction enzyme digestion) is cloned into a plasmid cut with a compatible restriction enzyme
2. the recombinant plasmid is transformed into bacteria
3. the bacteria are allowed to multiply, usually in liquid culture
4. a large quantity of the recombinant plasmid DNA is isolated from the bacterial culture
5. further manipulations (such as site directed mutagenesis or the introduction of another piece of DNA) are conducted on the recombinant plasmid
6. the modified plasmid is again transformed into bacteria, prior to further manipulations, or for expression

3.2. OTHER VECTORS

Lambda phage is a bacteriophage that infects *E. coli* and can be used as a vector. Lambda phage is a linear DNA vector molecule that can typically hold a 15-20 kb fragment in each clone.

Cosmids are a hybrid vector system composed of part plasmid and part phage DNA. It can clone 30-45Kb fragments in each clone. The **lambda phage packaging system (**stuffs the recombinant DNA into the lambda bacteriophage heads) is used for higher transformation efficiency, but it also has the plasmid origin of replication so clones can be replicated in the host like plasmids.

BACs (Bacterial Artificial Chromosomes) is a circular DNA vector that uses a plasmid origin of replication to propagate. The insert DNA can be 100's of Kb so it has a very large cloned regions, but such large recombinant DNA molecules are difficult to transform so they are difficult to make.

4. DNA LIGATION

The process of **DNA ligation** occurs when DNA strands are covalently joined, end-to-end through the action of an enzyme called **DNA ligase**. Molecules with complementary overhanging sequences are said to have "**sticky**" or **compatible ends**, which facilitate their joining to form recombinant DNA. Likewise, two blunt-ended sequences are also considered compatible to join together, although they do not ligate together as efficiently as sticky-ends. Note: sticky-ended molecules with non-complementary sequences will not ligate together with DNA ligase.

The process of ligation is central to the production of recombinant DNA, including the insertion of a double stranded DNA fragment into a plasmid vector.

5. AN APPLICATION OF MOLECULAR CLONING: RECOMBINANT INSULIN

Purified insulin protein is critical to the treatment of diabetes. Prior to ~1980, insulin for clinical use was isolated from human cadavers or from slaughtered animals such as pigs. Human-derived insulin generally had better pharmacological properties, but was in limited supply and carried risks of disease transmission. By cloning the human insulin gene and expressing it in *E. coli*, large quantities of the insulin protein and identical to the human hormone sequence could be produced in fermenters, safely and efficiently. Production of recombinant insulin

also allows specialized variants of the protein to be produced: for example, by changing a few amino acids, longer-acting forms of the hormone can be made. The active insulin hormone contains two peptide fragments of 21 and 30 amino acids, respectively. Today, essentially all insulin is produced from recombinant sources (**Figure 4**), i.e. human genes and their derivatives expressed in bacteria or yeast.

Figure 4.
A vial of insulin. Note that the label lists the origin as "rDNA", which stands for recombinant DNA.
(Flickr-DeathByBokeh- CC BY-NC 2.0)

SUMMARY:

- DNA fragments can be cloned into vectors.
- Transformation of recombinant DNA is the transfer of DNA (usually recombinant plasmids) into bacteria.
- Cloning of genes into E. coli is a common technique that allows large quantities of a DNA for gene to made
- This allows further analysis or manipulation of the cloned sequences.
- Genomic DNA libraries contain fragments of genomic DNA.
- cDNA libraries contain shorter segments of DNA that correspond to the mRNA for each gene.
- Gene of interest can be identified using DNA probes to screen genomic or cDNA libraries.
- Cloning can also be used to produce useful proteins, such as insulin, in microbes.

KEY TERMS:

in vivo
in situ
in vitro
in plastico
in silico
Recombinant DNA
plasmid
transformation
Ori
selectable marker
Multiple cloning site (MCS)
competent
electroporation
vector
clone
Lambda phage
cosmid
lambda phage packaging system
BACs
DNA ligation
DNA ligase
sticky / compatible end
genomic library
cDNA library

STUDY QUESTIONS:

1) A coat protein from a particular virus can be used to immunize children against further infection. However, inoculation of children with proteins extracted from natural viruses sometimes causes a fatal disease, due to contamination with live viruses. How could you use molecular biology to produce an optimal vaccine?

2) How would cloning be different if there were no selectable markers?

Chapter 16 of the MRU textbook is used under Canadian Fair Dealing law but is removed from this online version to avoid copyright implications.

CHAPTER 17 – CHROMOSOME REARRANGEMENTS

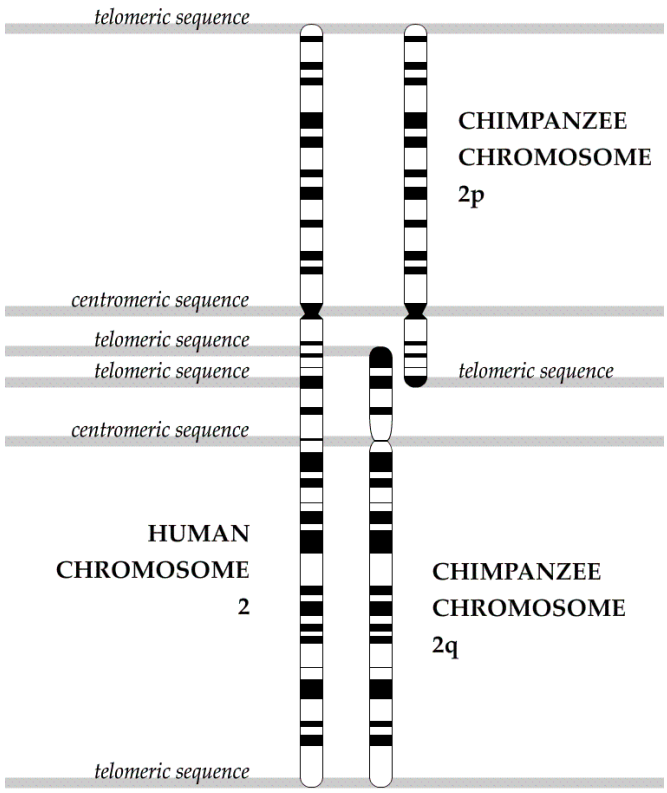


Figure 1.
Comparing an ideogram of the human chromosome 2 to the equivalent chromosomes in chimpanzees, we notice that the human chromosome 2 likely came from a fusion event that occurred since their common ancestor. This is supported by evidence finding telomeric and centromeric sequences in the middle of human chromosome 2 similar to that of the ends and middle of the chimpanzee chromosomes.
(Flickr- T. Michael Keesey- CC BY 2.0)

INTRODUCTION

Previous chapters described chromosomes as simple linear DNA molecules on which genes are located. For example, your largest chromosome, chromosome 1, has about 3536 genes. To ensure that each of your cells possesses these genes, the typical linear eukaryotic chromosome has three critical features that allow it to be passed on during cell division. (1) **Origins of replication** found along its length provide places for DNA replication to start, (2) **telomeres** protect each end of the chromosome, and (3) a single **centromere** near the middle provides a place for microtubules to attach and move the chromosome during mitosis and meiosis.

However, at various locations both strands of the double stranded DNA in a chromosome can break and the subsequent daughter cell(s) may not retain all the DNA and thus all the genes. For example, if a segment of the chromosome has been lost (a deletion), the cell may be missing many genes. The causes of chromosome structural abnormalities and the consequences they have for the cell and the organism are described below. They involve double stranded breaks in the DNA, meiotic crossover events, and rejoining of the broken ends. Human examples will be used to show the phenotypic consequences and methods for detection.

1. DNA DOUBLE STRAND BREAKS AND INCORRECT MEIOTIC CROSSOVERS CAUSE CHROMOSOMAL REARRANGEMENTS

1.1. DOUBLE STRAND BREAKS AND THEIR REPAIR

A chromosome is a very long but very thin molecule. In the phosphodiester backbone there are only two covalent bonds holding each base pair to the next. If one of these covalent bonds is

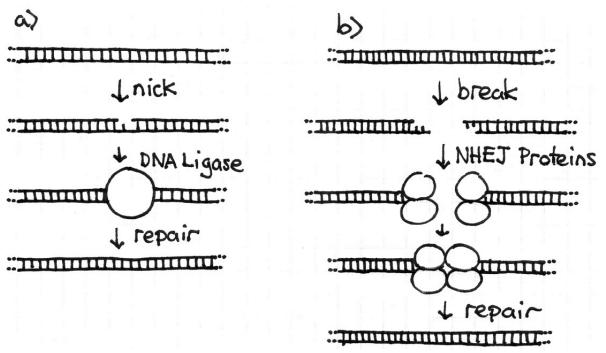


Figure 2.
Repair of single strand nicks and double strand breaks in DNA. (Original-Harrington-CC BY-NC 3.0)

broken the chromosome will still remain intact, although a DNA Ligase will be needed to repair the nick (**Figure 2**a). Problems arise when both strands are broken at or near the same location. This **double strand break** will cleave the chromosome into two independent pieces (**Figure 2**b). Because these events do occur in cells there is a repair system called the **non-homologous end joining** (NHEJ) **system** to fix them. Proteins bind to each broken end of the DNA and reattach them with new covalent bonds. This system is not perfect and sometimes leads to **chromosome rearrangements** (see next section).

The NHEJ system proteins only function if required. If the chromosomes within an interphase nucleus are all intact the system is not active. The telomeres at the natural ends of chromosomes prevent the NHEJ system from attempting to join the normal ends of chromosomes together. If there is one double strand break the two broken ends can be recognized and joined. But if there are two double strand breaks at the same time there will be four broken ends in total. The NHEJ system proteins may join the ends together correctly, but if they fail, the result is a **chromosome rearrangement** (**Figure 3**).

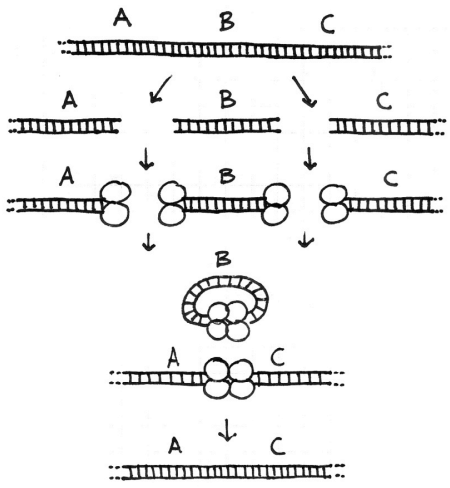


Figure 3.
Errors during DNA repair can cause a chromosome deletion. In this diagram A, B, and C are genes on the same chromosome. As in **Figure 2** there has been breaks in the DNA, recruitment of NHEJ proteins, and repair. After the repairs are completed the small piece of DNA with gene B is lost and the chromosome now only has genes A and C. (Original-Harrington-CC BY-NC 3.0)

1.2. INCORRECT MEIOTIC CROSSOVERS

Meiotic crossovers occur at the beginning of meiosis for two reasons. They help hold the homologous chromosomes together until separation occurs during anaphase I (see Chapter 7). They also allow recombination to occur between linked genes (see Chapter 9). The event itself takes place during prophase I when a double strand break on one piece of DNA is joined with a double strand break on another piece of DNA and the ends are put together (**Figure 4**a). Most of the time the breaks are on non-sister chromatids and most of the time the breaks are at the same relative locations.

Problems occur when the wrong pieces of DNA are matched up along the chromosomes during crossover events. This can happen if the same or similar DNA sequence is found at multiple sites on the chromosomes (**Figure 4**b). For example, if there are two ***Alu*** **transposable elements** on a chromosome. When the homologous chromosomes pair during prophase I the wrong *Alu* sequences might line up. A crossover may occur in this region. If so, when the chromosomes separate during anaphase I one of the chromatids will have a

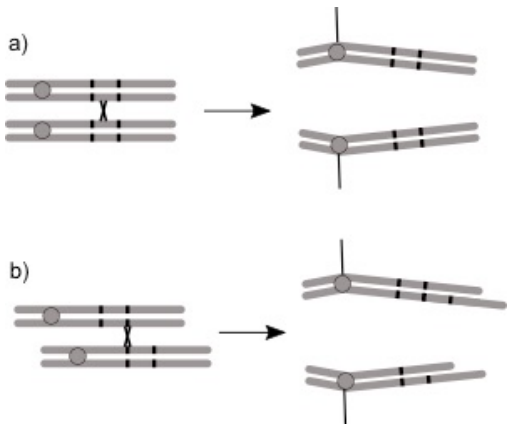


Figure 4 .
Errors during meiotic crossovers can cause duplications and deletions. This diagram shows homologous chromosomes pairing in prophase I and then separating in anaphase I. The shaded boxes are Alu transposable elements. a) The homologous chromosomes pair properly, a crossover occurs, and all four chromatids in anaphase I are normal. b) The pairing is incorrect, a crossover occurs in the mispaired region, and in anaphase I one chromatid has a duplication and another has a deletion.
(Original-Harrington/L. Canham-CC BY-NC 3.0)

duplication and one will have a deletion. Ultimately, of the four cells produced by this meiosis, two will be normal, one will have a chromosome with extra genes, and one will have a chromosome missing some genes. Errors of this type can also cause inversions and translocations.

Errors during the repair of multiple double strand breaks or incorrect meiotic crossovers can cause four types of chromosome rearrangements: deletion, inversion, duplication or translocation. The type of chromosome rearrangement is either dependent upon where the two breaks were originally and how they are rejoined, or on the location of the homology during meiosis. **Figure 3** shows some possibilities but more are shown in the following sections. The first part of each section shows a double strand DNA break between the B and C genes (shown here as a red X). A second DNA break occurs and the NHEJ proteins mend the damage incorrectly by joining the ends (shown with the blue arrows). The chromosomes are drawn as unreplicated as they are in G_1 phase but these events can happen anytime during interphase. The second part shows how meiosis can cause the rearrangements.

2. DELETIONS

There are two forms of deletions: **Terminal** and **Interstitial**. Terminal deletions are deletions off of the end of a chromosome. Interstitial deletions are deletions of a region in the middle of the chromosome, while the arms on each side remain normal. For example, with a chromosome that has the genes ABCDEF, an example of a terminal deletion will be CDEF. An example of an interstitial deletion will be ABCF.

2.1. DELETIONS FROM DOUBLE STRAND BREAK REPAIR

Deletions arise from double strand breaks when both breaks are on one chromosome. If the ends are joined in this way the piece of DNA with the B gene on it does not have a centromere and will be lost during the next cell division.

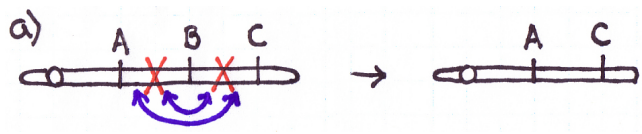


Figure 5.
Deletion can result from double strand break repair. (Original-Harrington- CC BY-NC 3.0)

2.2. DELETIONS FROM INCORRECT MEIOSIS

If meiotic rearrangement is the cause, **Deletion** chromosomes will pair up with a normal homolog along the shared regions and at the missing segment, the normal homolog will loop out (nothing to pair with) to form a **deletion loop**. This can be used to locate the deletion cytologically. The deleted region is also **pseudo-dominant**, in that it permits the mutant expression of recessive alleles on the normal homolog. Deletion mutations don't revert - nothing to replace the missing DNA.

3. INVERSIONS

3.1. INVERSIONS FROM DOUBLE STRAND BREAKS

Inversions also occur when both double strand breaks are on one chromosome. If the ends are joined in this way, part of the chromosome is inverted. This example shows a **paracentric inversion**, named because the inverted section does

not include the centromere (para = beside). If the breaks occur on different chromosome arms the inverted section includes the centromere and the result is a **pericentric inversion** (peri = around).

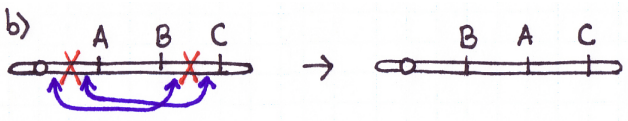


Figure 6.
Inversion can result from double strand break repair. (Original-Harrington- CC BY-NC 3.0)

3.2. INVERSIONS FROM INCORRECT MEIOSIS

In meiosis, when an **inversion** chromosome is paired up there is an **inversion loop** formed. If there is a crossover within the loop then abnormal products will result and abnormal, unbalanced gametes will be produced. For example, a crossover event within the loop of a **paracentric inversion** will lead to a di-centric product that will break into deletion products and produce unbalanced gametes (**Figure 7**). Similarly, with a **pericentric inversion**, a crossover event leads to duplicate/deletion products that are unbalanced (**Figure 8**).

If joined with a normal gamete, they will result in an unbalanced zygote, which are usually lethal. The consequence for this is that crossover products (recombinants) are lost and thus inversions appear to suppress crossovers within the inverted region.

Note: with both types of inversions, crossovers outside the loop are possible and fully viable, as they don't alter the gene balance.

4. DUPLICATIONS

There are two major forms of duplications: **tandem** and **inverse** duplications. Tandem duplications are when the duplicated genes are in the same order, and inverse duplications are where the duplicated genes are in the reverse order. For example if you have a chromosome that has the genes ABCDEFGH, and a duplication occurs in the BCD genes, then a tandem duplication would look like: ABCDBCDEFGH. An inverse duplication would look like: ABCDDCBEFGH.

Insertional duplications are also seen, where the duplicated region is inserted to a more distant location. e.g. ABCDEFBCDGH

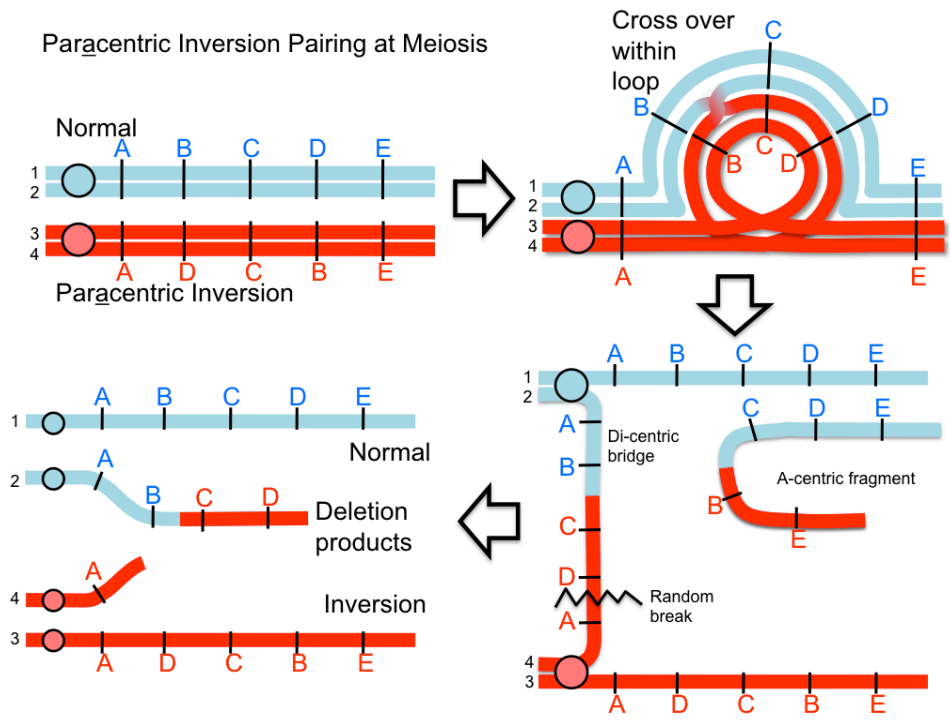


Figure 7.
A paracentric inversion pairing at meiosis. A crossover within the loop causes the production of an acentric and a dicentric chromatids, which leads to deletion product..
(Original-Locke-CC BY-NC 3.0)

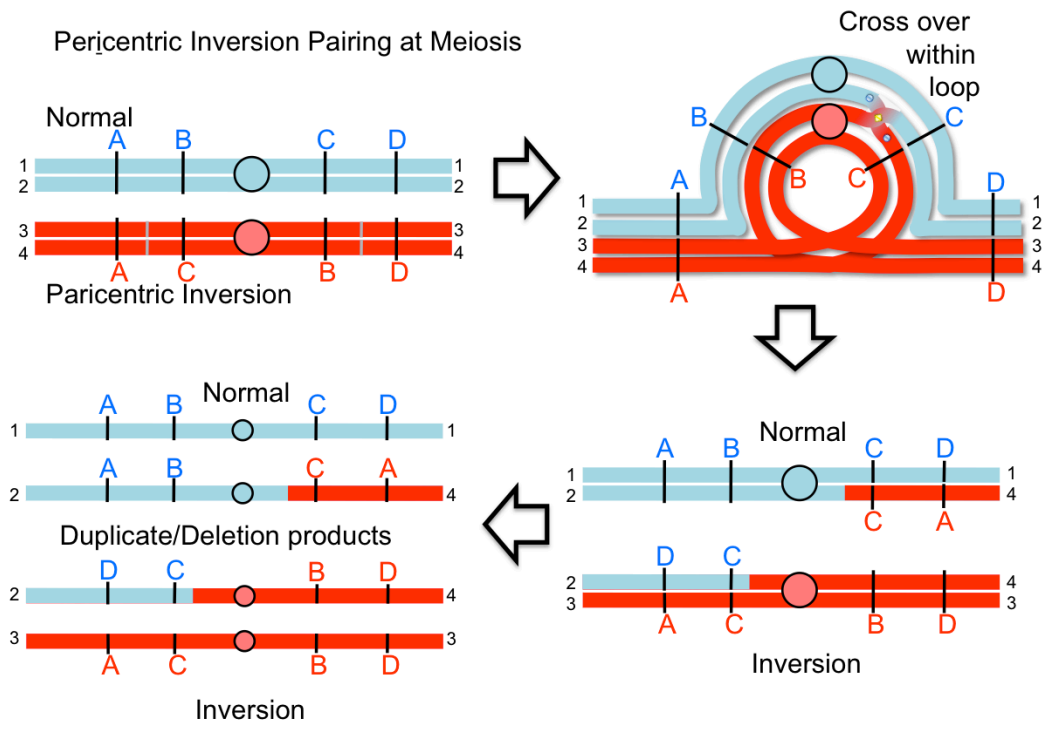


Figure 8.
A pericentric inversion pairing at meiosis. A crossover within the loop causes the production of duplicate and deletion products. (Original-Locke-CC BY-NC 3.0)

4.1. DUPLICATIONS FROM DOUBLE STRAND BREAKS

Duplications can occur from two DNA breaks at different places in sister chromatids (in a replicated chromosome). The ends are joined together incorrectly to create a chromosome with a duplication (two "B" regions as shown). Note: the reciprocal product has a deletion.

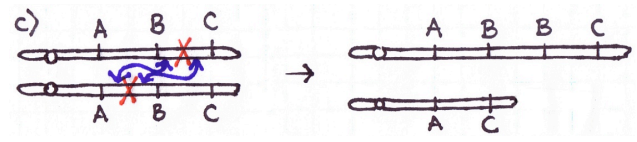


Figure 9.
Duplication can result from double strand break repair. (Original-Harrington- CC BY-NC 3.0)

4.2. DUPLICATIONS FROM INCORRECT MEIOSIS

Duplications also produce a cytologically visible loop at meiotic pairing. Duplications can revert at a relatively high frequency by unequal crossing over. Duplicated genes offer new possibilities for mutational divergence followed by natural selection in the course of evolution.

5. TRANSLOCATIONS

5.1. TRANSLOCATIONS FROM DOUBLE STRAND BREAKS

Translocations result from two breaks on different chromosomes (not homologs) and incorrect rejoining. This example shows a **reciprocal translocation** - two chromosomes have 'swapped' arms, the E gene is now part of the white chromosome and the C gene is now part of the shaded chromosome. **Robertsonian translocations** are those rare situations in which all of the genes end up together on one chromosome and the other chromosome is so small that it is typically lost.

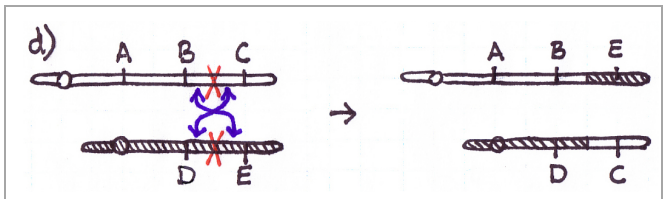


Figure 10.
Translocation can result from double strand break repair. (Original-Harrington- CC BY-NC 3.0)

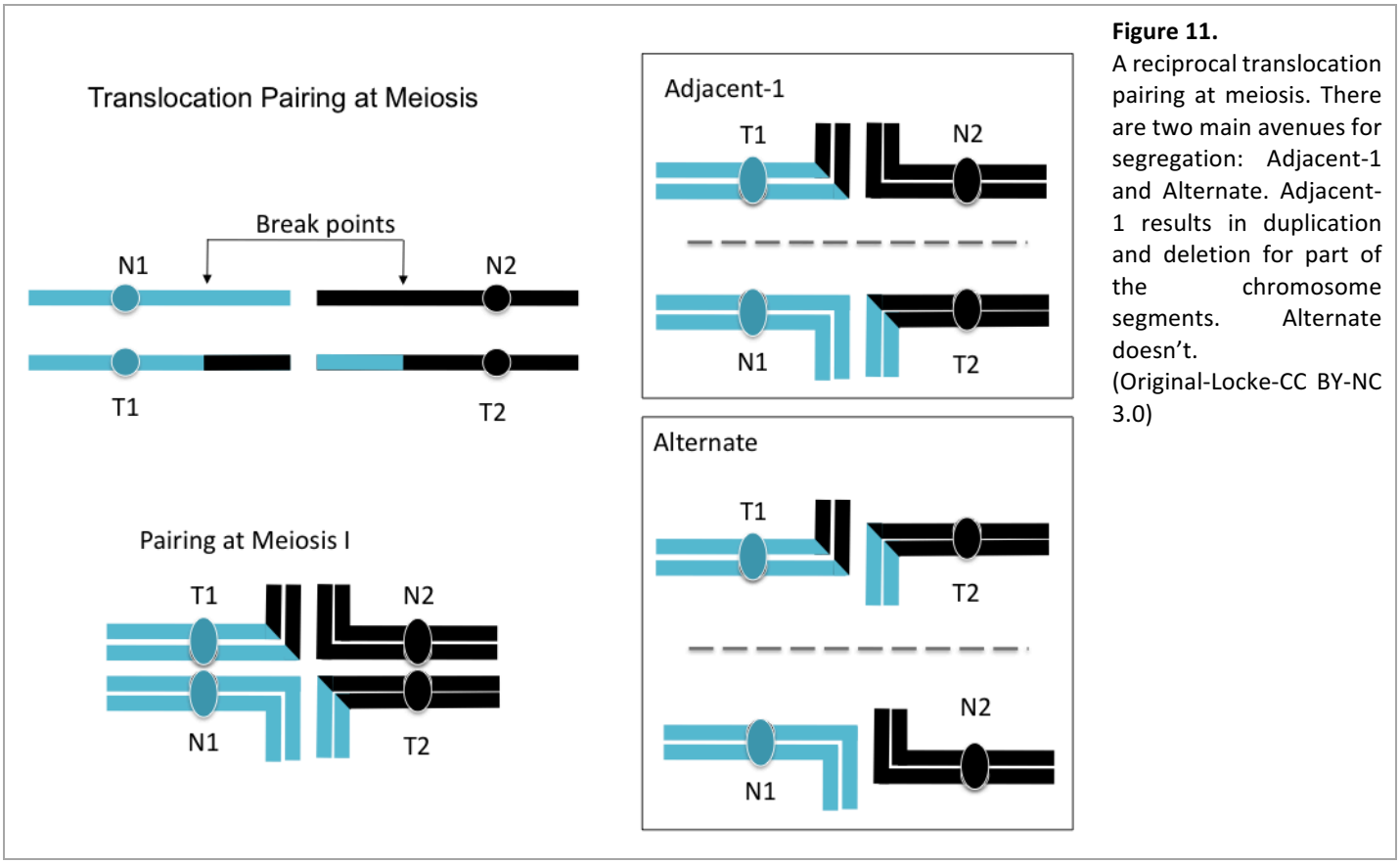


Figure 11.
A reciprocal translocation pairing at meiosis. There are two main avenues for segregation: Adjacent-1 and Alternate. Adjacent-1 results in duplication and deletion for part of the chromosome segments. Alternate doesn't.
(Original-Locke-CC BY-NC 3.0)

5.2. TRANSLOCATIONS FROM INCORRECT MEIOSIS

For **translocations** during meiosis, a consequence for the two chromosomes involved is that when they pair both replicated chromosome pairs will be together, which can be seen cytologically as a **tetrad**. This tetrad can segregate in three ways.

This set of paired, replicated chromosomes can segregate as **Alternate** (balanced) where both normal (N1 and N2) and both translocated chromosomes (T1 and T2) go to the same polls, respectively. The chromosomes can segregate as **Adjacent-1** (unbalanced) where the normal and translocation chromosomes segregate, with N2 and T1 segregate from N1 and T2. Alternate and Adjacent 1 both occur in approximate equal frequency and thus only about half the time do the gametes end up unbalanced (**Figure 11.**). Note how each daughter cell in Alternate has equal amounts of blue and black chromosomes, while in Adjacent-1 one daughter has extra black chromosomes, and the other has extra blue.

The third segregation possibility is known as Adjacent-2, where N1 and T1 go to one pole, while N2 and T2 go to the other. This way of segregating is extremely rare, and so will not be described in any further detail.

6. CONSEQUENCES OF CHROMOSOMAL REARRANGEMENTS

6.1. DECREASED VIABILITY

All of the chromosome rearrangements shown above produce functional chromosomes. Each has one centromere, two telomeres, and thousands of origins of replication. Because inversions and translocations do not change the number of genes in a cell or organism they are said to be **balanced rearrangements**. Unless one of the breakpoints occurred in the middle of a gene the cells will not be affected. On the other hand, deletions and duplications are **unbalanced rearrangements**. The larger they are (more genes involved) the more disruption they cause to the proper functioning of the cell or organism. Having too much or too little gene action for a large number of genes can disrupt

the cellular metabolism to generate a phenotype or reduce viability.

6.2. DECREASED FERTILITY

Recall that during meiosis I homologous chromosomes pair up. If a cell has a chromosome with a rearrangement this chromosome will have to pair with its normal homolog.

Cells heterozygous for balanced rearrangements actually have more difficulties in prophase I. Consider the chromosomes shown in **Figure 12**. There are different ways they might pair during prophase I - one is shown in **Figure 13**. But if a crossover occurs in the inverted region the result will be unbalanced gametes. Embryos made with unbalanced gametes rarely survive. The consequence is that the heterozygous organism will have **reduced fertility**.

Note that an organism homozygous for this inversion chromosome will not be affected in this way because no loops are formed. The chromosomes can pair along their entire length and crossovers will not produce any unbalanced gametes. This is a general property of inversions and translocations.

In heterozygotes there are problems during meiosis resulting in a lot of the gametes being unbalanced and an overall reduction in fertility. In homozygotes the rearranged chromosomes pair with one another just fine and there is no effect on fertility.

6.3. CANCER

Some chromosome rearrangements have breakpoints within genes leading to the creation of hybrid genes – the first part of one gene with the last part of another. If the hybrid gene inappropriately promotes cell replication, the cell can become cancerous.

6.4. EVOLUTION

Those chromosome changes that duplicate genes are important for evolution. If an organism has an extra copy of important genes, one gene can be retained for their original function while others can mutate and potentially acquire new functions (**Figure 14.**). An example of this is the multiple copies of the globin genes found in mammals.

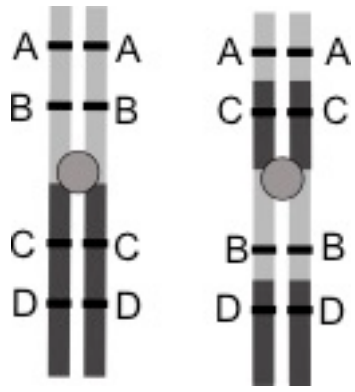


Figure 12.
A normally arranged chromosome (left) and a homolog with a pericentric inversion (right).
(Original-Harrington/Canham-CC BY-NC 3.0

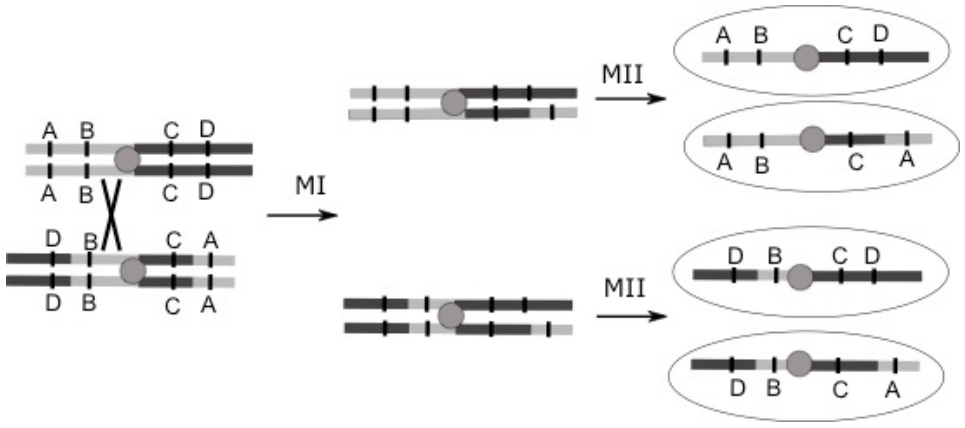


Figure 13.
Meiosis in a cell heterozygous for the chromosomes shown in **Figure 12**. Note that of the four gametes one has a deletion of the A gene and a duplication of the D gene while another gamete has a duplication of A and a deletion of D.
(Original-Harrington/Canham-CC BY-NC 3.0)

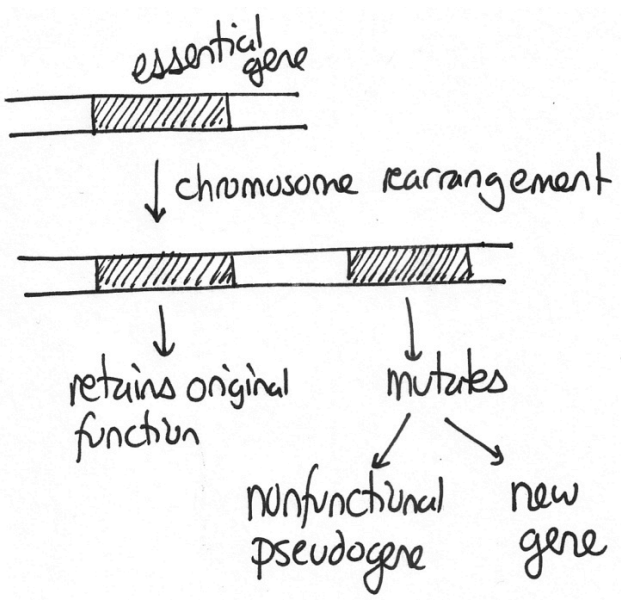


Figure 14.
Duplicated genes can mutate without compromising the viability of the organism. Occasionally the result is a new gene. (Original-Harrington-CC BY-NC 3.0)

Chromosome rearrangements that decrease fertility are also important for the origin of new species. If a rearrangement, such as the inversion shown in **Figure 12**, becomes common in a small isolated population, that population has 100% fertility if they mate within their group, but a reduced fertility if they mate with members of the larger population. As rearrangements accumulate the small population will become more and more reproductively isolated. When members are incapable of forming viable, fertile offspring with the original population the group will have become a new species.

Another example is shown in **Figure 1**, where the human chromosome 2 is a fusion of two chromosomes present in the common ancestor of humans and other great apes (chimpanzee, gorilla, orangutan). We do not know exactly when in human history (evolution) this fusion event occurred, except that, because it is absent in all other apes and present in all current humans, it must have occurred after the split between chimpanzee and humans.

7. CHROMOSOMAL REARRANGEMENTS IN HUMANS

The problems described above can affect all eukaryotes, unicellular and multicellular. To better understand the consequences let us consider those that affect people. The convention when describing a person's **karyotype** (chromosome composition) is to list the total number of chromosomes, then the sex chromosomes, and then anything out of the ordinary. Most of us are 46,XX or 46,XY. What follows are some examples of chromosome number and chromosome structure abnormalities.

7.1. CRI-DU-CHAT SYNDROME

Cri-du-chat syndrome occurs when a child inherits a defective chromosome 5 from one parent (**Figure 15**). This condition is rare - it is present in only 1 in 20 000 to 1 in 50 000 births but it does account for 1% of cases of profound intellectual disability. The specific defect is a deletion that removes 2 Mb or more from the tip of the short arm of the chromosome. In most cases the deletion is the result of a chromosomal rearrangement in one of the parent's germ line cells. People with cri-du-chat have a karyotype of 46,sex,deletion(5).

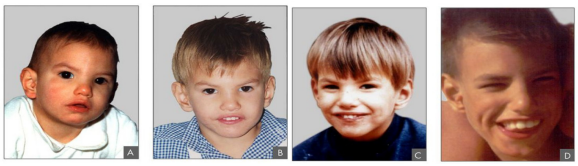


Figure 15.
A boy with cri-du-chat syndrome. The pictures were taken at 8 months (A), 2 years (B), 4 years (C), and 9 years (D). (Wikipedia-Paola Cerruti Mainardi/ changes: horizontally aligned the photos- CC BY 2.0)

As with Down syndrome this condition is associated with intellectual disability and other health problems. These problems include an improperly formed larynx which leads to infants making high pitched cat-like crying sounds (hence the name "cry of the cat"). It is suspected that at least some of the intellectual disability phenotype is due to having only a single copy of the CTNND2 gene. This gene is active during embryogenesis and makes a protein essential for neuron migration. Down syndrome and cri-du-chat syndrome are two examples of the need for genomes to contain the proper number of genes. Having too many copies of key genes (Down syndrome) or too few (cri-du-chat syndrome) can lead to substantial developmental problems.

7.2. INVERSION

The most common chromosome rearrangements in humans are inversions of chromosome 9. About 2% of the world's population is heterozygous or homozygous for inversion. This rearrangement does not affect a person's health because the genes on the chromosome are all present - all that has changed is their relative locations. Inversion is different from deletion in two main respects. As mentioned above because it is a balanced rearrangement it does not cause harm. And because of this nearly everyone with an inversion chromosome has inherited it from a parent who had inherited it from one of his or her parents and so on. In contrast, most cases of deletion are due to new mutations occurring in a parent.

7.3. DIAGNOSING HUMAN CHROMOSOME ABNORMALITIES

How can we confirm that a person has a specific chromosomal abnormality? The first method was simply to obtain a sample of their cells, stain the chromosomes with **Giemsa** dye, and examine the results with a light microscope (**Figure 16**). Each chromosome can be recognized by its length, the location of its centromere, and the characteristic pattern of purple bands produced by the Giemsa. Bright field microscopy has its limitations though - it only works with mitotic chromosomes and many chromosome rearrangements are either too subtle or too complex for even a skilled cytogeneticist to discern.

The solution to these problems was **fluorescence in situ hybridization** (FISH). A single stranded **fluorescent DNA probe** is allowed to hybridize to denatured target DNA. Because there are several fluorescent colours available it is common to use more than one probe at the same time.

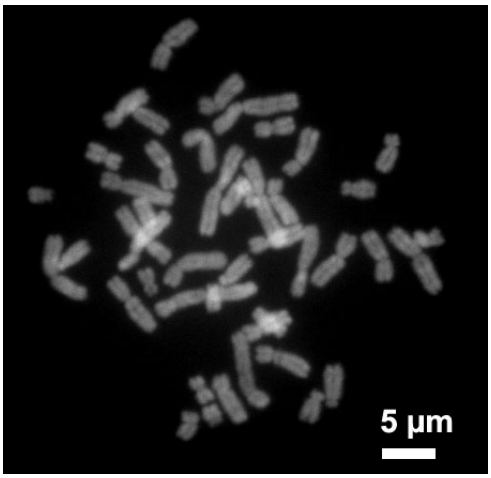


Figure 16.
Human chromosomes. One way to obtain chromosomes is to take a blood sample, culture the cells for three days in the presence of a T-cell growth factor, arrest the cells in metaphase with a microtubule inhibitor, and then drop the cells onto a slide. The cells burst and the chromosomes stick to the slide. The chromosomes can then be stained or probed. Because the cells are in metaphase it is possible to see 46 replicated chromosomes here. There will be dozens of collections of chromosomes like this over the entire slide.
(Wikipedia-Steffen Dietzel- CC BY-SA 3.0)

A physician may suspect that a patient has a specific genetic condition based upon the patient's physical appearance, mental abilities, health problems, and other factors. FISH can be used to confirm the diagnosis. For example, **Figure 17** shows a positive result for cri-du-chat syndrome. The probes are binding to two long arms of chromosome 5 but only one short arm. One of the chromosome 5s must therefore be missing part of its short arm.

FISH is an elegant technique that produces dramatic images of our chromosomes. Unfortunately, FISH is also expensive, time consuming, and requires a high degree of skill. For these reasons, FISH is slowly being replaced with PCR and DNA chip based methods. Versions of these techniques have been developed that can accurately quantify a person's DNA. For example DNA from a person with cri-du-chat syndrome will contain 50% less DNA from the end of chromosome 5. These techniques are very useful if the suspected abnormality is a deletion, a duplication, or a change in chromosome number. They are less useful for diagnosing chromosome

inversions and translocations because these rearrangements often involve no net loss or gain of genes.

In the future all of these techniques will likely be replaced with DNA sequencing. Each new generation of genome sequencing machines can sequence more DNA in less time. Eventually it will be cheaper just to sequence a patient's entire genome than to use FISH or PCR to test for specific chromosome defects.

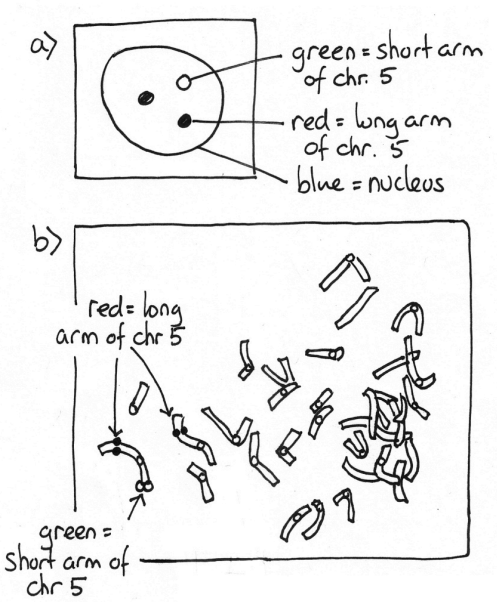


Figure 17.
A positive result for cri-du-chat syndrome. This diagram is based upon actual results. Cells from a patient's blood were prepared to show an interphase nucleus (a) and mitotic chromosomes (b). The DNA has been coloured blue with DAPI. The green fluorescent probe is binding to the tip of the short arm of chromosome 5 (shown here as open circles). This is the region absent in cri-du-chat. The red fluorescent probe is binding to the middle of the long arm of the same chromosome (filled circles). This probe is used as a control.

SUMMARY:

- A deletion in chromosome 5 causes a serious condition (cri-du-chat syndrome).
- Deletions are unbalanced chromosome rearrangements.
- Inversions tend to cause fewer health consequences than deletions because inversions are balanced chromosome rearrangements.
- Bright field microscopy can be used to detect chromosome number abnormalities and some chromosome rearrangements.
- Fluorescence *in situ* hybridization (*FISH*) can be used to detect all types of chromosome abnormalities.
- PCR and DNA chip based techniques can be used to detect chromosome number abnormalities, deletions, and duplications.

KEY TERMS:

origin of replication
telomere
centromere
double strand break
non-homologous end joining
chromosome rearrangement
meiotic crossover
Alu transposable elements
terminal deletion
interstitial deletion
deletion
deletion loop
pseudo-dominant
inversion
paracentric inversion
pericentric inversion
inversion loop
tandem duplication
inverse duplication
insertional duplication
duplication
translocation
reciprocal translocation
Robertsonian translocation
tetrad
reduced fertility
karyotype
bright field microscopy
Giemsa stain
fluorescence in situ hybridization (*FISH*)
fluorescent DNA probe

STUDY QUESTIONS:

1) Make diagrams showing how an improper crossover event during meiosis can lead to:
   a) an inversion
   b) a translocation.

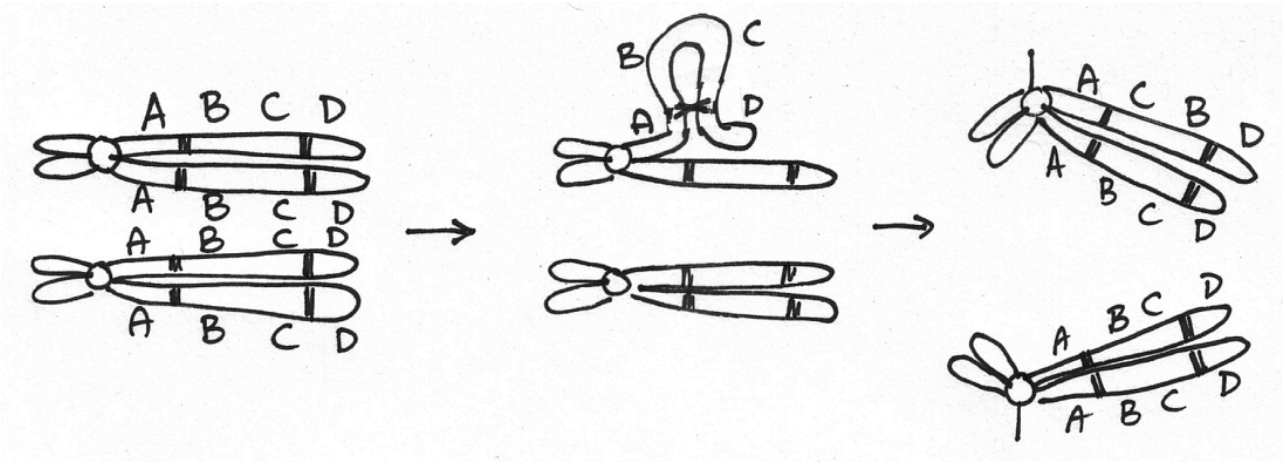
2) If Drosophila geneticists want to generate mutant strains with deletions they expose flies to gamma rays. What does this imply about gamma rays?

3) Design a FISH based experiment to find out if your lab partner is a 47,XXX female or a 47,XYY male.

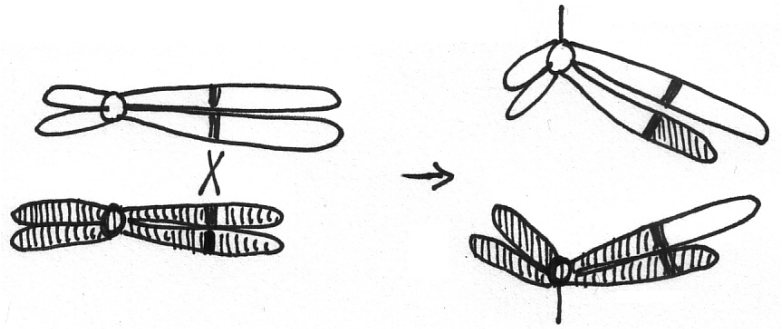
CHAPTER 17 – ANSWERS

1)

a) As in **Figure 4** homologous chromosomes pair during prophase I. The shaded boxes are regions of sequence similarity, for example Alu transposable elements. A crossover occurs between two of the Alu elements on the same chromatid leading to a chromosomal inversion.



b) A crossover occurs between Alu elements on different chromosomes leading to a chromosomal translocation. Note that the homologous chromosomes are not shown in this figure for simplicity.



2) Gamma rays are efficient at causing double strand DNA breaks, which are then more likely to rejoin and produce a deletion.

3) Obtain permission from the person (and ethical approval from the university), isolate some white blood cells, place the cells on a slide, denature the DNA, hybridize with fluorescent nucleic acid probes specific for the centromeres of the X chromosome and the Y chromosome, observe the results with a fluorescence microscope. If they are XXX there should be three X signals, if XXY, there should be two X signals and one Y in each cell nucleus.