Open Genetics - Fall 2013
1 September, 2013 Version

The first edition of this book 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 non-commercial use under the Creative Commons License (See below). Users are encouraged to make modifications and improvements to the book.

All text in the original edition was written by Michael Deyholos, Ph.D. Chapter 9 and portions of Chapters 2 and 3 in this edition were written by Mike Harrington, Ph.D. Additional content and editing by John Locke, Ph.D. and Mark Wolansky, M.Sc.

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 at the end of the book.

Under the terms of the license, you are free:

- **to Share** — to copy, distribute and transmit the work
- **to Remix** — to adapt the work

Under the following conditions:

- **Attribution**. You must include the name of the original author in books or excerpts derived from this work.
- **Noncommercial**. You may not use this work for commercial purposes.
- **Share Alike**. If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar licence to this one.
- For any reuse or distribution, you must make clear to others the licence terms of this work.
- Any of the above conditions can be waived if you get permission from the copyright holder.
- The author's moral rights are retained in this licence.

IMAGE, PHOTO AND FIGURE LICENSE DETAILS:

The source of each figure is given at the end of the legend and includes the following:

CC: AD Creative Commons Attribution-No Derivative Works 2.0 Generic

CC: AN Creative Commons Attribution-Noncommercial 2.0 Generic

CC: AND Creative Commons Attribution-Noncommercial-No Derivative Works 2.0 Generic

CC: AS Creative Commons Attribution-Share Alike 2.0 Generic

EDU Educational use explicitly allowed by author, who retains copyright

GFDL Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License

PD Public Domain because it was created by a US government agency or because the author has explicity released it into the public domain.

BIBLIOGRAPHY

Alberts, B. et al. 2004. "Molecular Biology of the Cell, fourth edition". Garland Science, New York

Griffiths, A. J. F. et al. 2005. "Introduction to Genetic Analysis, eighth edition." W. H. Freeman and Company, New York

King, R. C. and W. D. Stansfield. 1997. "A Dictionary of Genetics, fifth edition". Oxford University Press, Toronto

MendelWeb. R. B. Blumberg, {July 1, 2013}. World Wide Web URL: www.mendelweb.org/

Online Mendelian Inheritance in Man, OMIM®. McKusick---Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), { July 1, 2013}. World Wide Web URL: http://omim.org/

Watson, J. D. et al. 2008. "Molecular Biology of the Gene, sixth edition" Pearson Education, Inc., San Franciso

Your Genes, Your Health Dolan DNA Learning Center / Cold Spring Harbor Laboratory, { July 1, 2013}. World Wide Web URL: www.ygyh.org/

TABLE OF CONTENTS

CHAPTER 1 OVERVIEW, DNA, AND GENES

Fig.1-1.
Parent and offspring Wolf's Monkey.
(Flickr-eclectic echos-CC:AND)

1.1 OVERVIEW

Genetics – The 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.

Humans have always been aware that the characteristics of an individual plant or animal could be passed down through the generations. Offspring look more like their parents. Humans also knew 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.

1.2 DNA IS THE GENETIC MATERIAL

By the early 1900's, biochemists had isolated hundreds of different chemicals from living cells. 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.

1.2.1 Griffith's Transformation Experiment

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 (Fig. 1.2). More importantly, the S-strain bacteria caused fatal infections when injected into mice, while the R-strain did not (top, Fig. 1.3). 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.

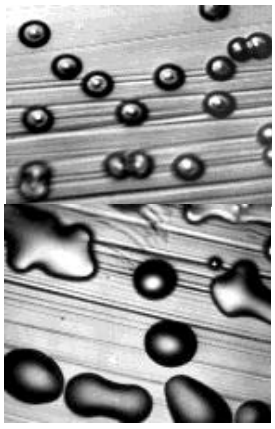


Figure 1.2
Colonies of Rough (top) and Smooth (bottom) strains of *S. pneumoniae*. (J. Exp.Med.98:21, 1953-R. Austrian-Pending)

1.2.2 Avery, MacLeod and McCarty's Experiment

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.

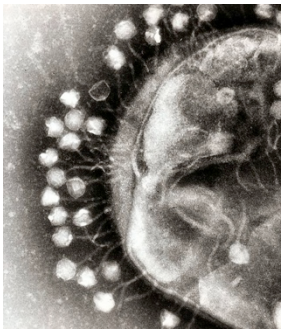


Figure 1.4
Electronmicrograph of T2 bacteriophage on surface of *E. coli*. (Wikipedia-G. Colm-PD)

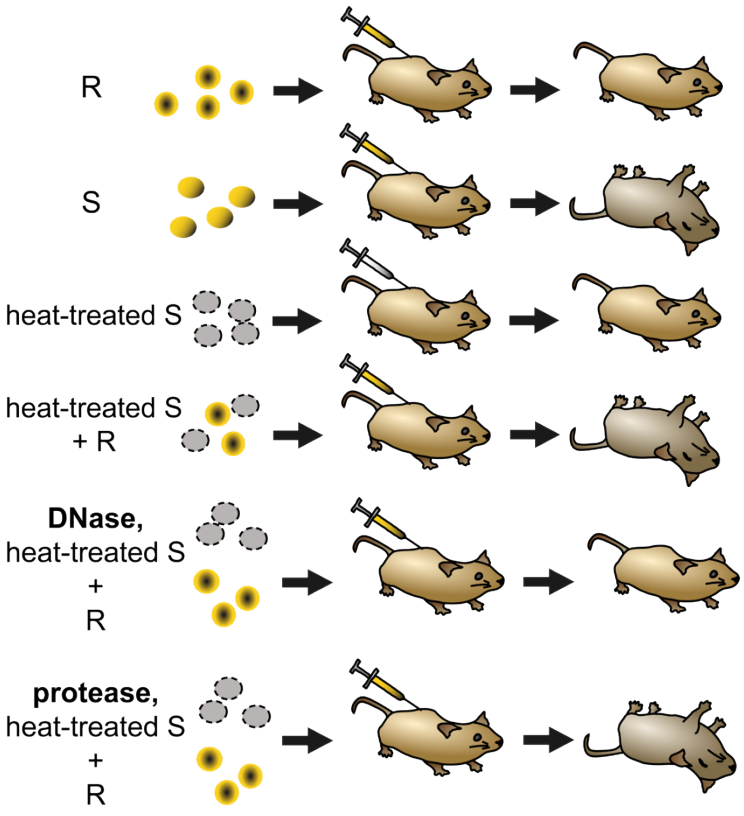


Figure 1.3 Experiments of Griffith and of Avery , MacLeod and McCarty. R strains of *S. pneumoniae* do not cause lethality. However, DNA-containing extracts from pathogenic S strains are sufficient to make R strains pathogenic. (Original-Deyholos-CC:AN)

1.2.3 Hershey and Chase's Experiment

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 (Fig. 1.4). 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 (Fig 1.5). 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.

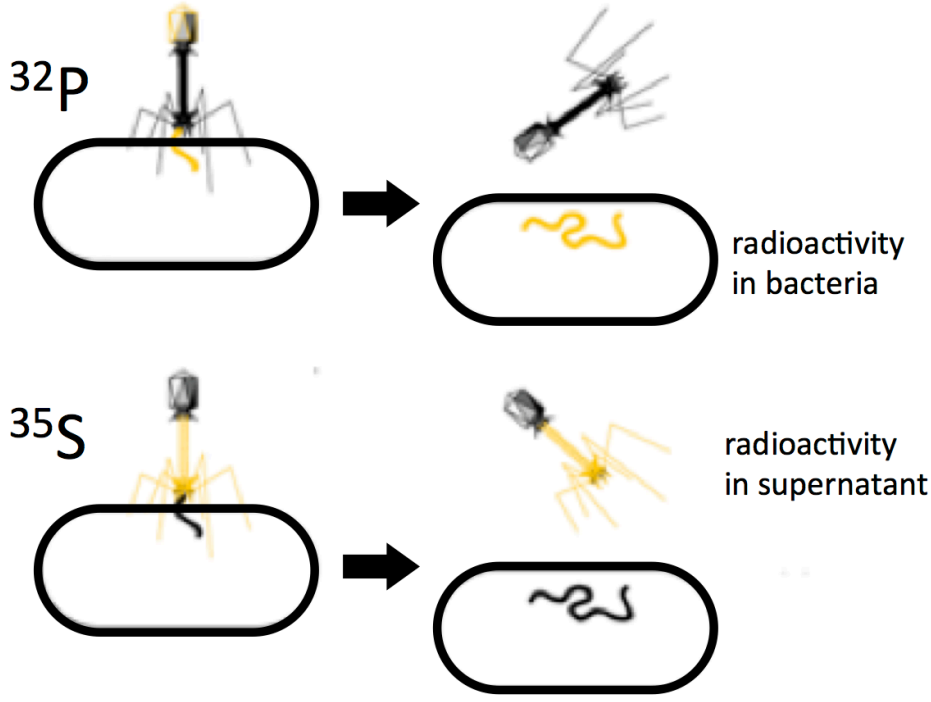


Figure 1.5
When ^{32}P-labeled phage infects *E. coli*, radioactivity is found only in the bacteria, after the phage are removed by agitation and centrifugation. In contrast, after infection with ^{35}S-labeled phage, radioactivity is found only in the supernatant that remains after the bacteria are removed. (Original-Deyholos-CC:AN)

1.2.4 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 and 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 case of **Prion** infections agents transmit characteristics via only a protein (no nucleic acid present). Prions infect by transmitting a misfolded 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 and deer 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.

1.3 The Structure of DNA

The experiments outlined in the previous sections proved that DNA was the genetic material, but very little was known about its structure at the time.

1.3.1 Chargaff's Rules

When **Watson and Crick** set out in the 1940's to determine the structure of DNA, it was already known that DNA is made up of a series four different types of molecules, called bases or nucleotides: adenine (A), cytosine (C), thymine (T), guanine (G). 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.

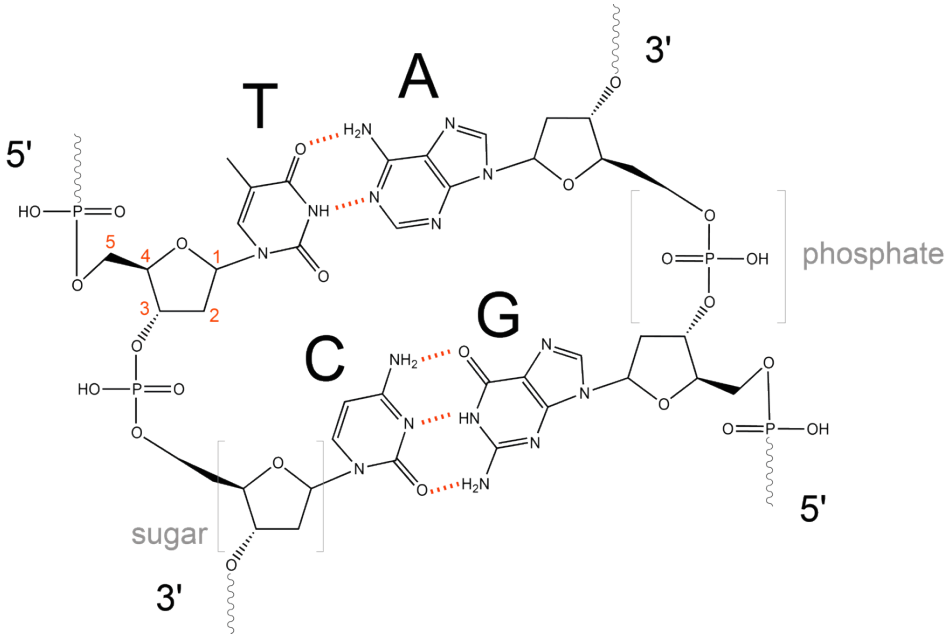


Figure 1.6 Chemical structure of two pairs of nucleotides in a fragment of double-stranded DNA. Sugar, phosphate, and bases A,C,G,T are labeled. Hydrogen bonds between bases on opposite strands are shown by dashed lines. Note that the G-C pair has more hydrogen bonds than A-T. The numbering of carbons within sugars is indicated by red numbers. Based on this numbering the polarity of each strand is indicated by the labels 5' and 3'. (Wikipedia-M. Strock-GFDL)

1.3.2 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 (Fig 1.6, 1.7). 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, in accordance with the nomenclature of the carbons in the sugars. 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. Under most conditions, the two strands are slightly offset, which creates a major groove on one face of the double helix, and a minor groove on the other. 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.

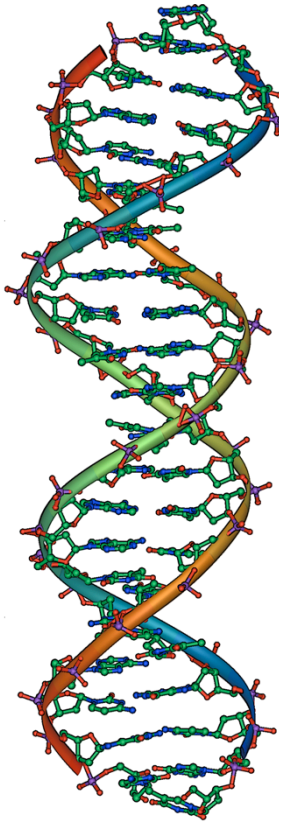


Figure 1.7
DNA double helix structure. (Original-unknown-?)

How can this molecule, DNA, contain the genetic material?

1.4 Genes are the Basic Units of Inheritance

1.4.1 Blending vs Particulate inheritance

The once prevalent (but now discredited) concept of **blending 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. However, **Gregor Mendel** (Fig 1.8) 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 (Fig 1.9). This was evidence that the 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.

Figure 1.8
Gregor Mendel. (Original-unknown-PD)

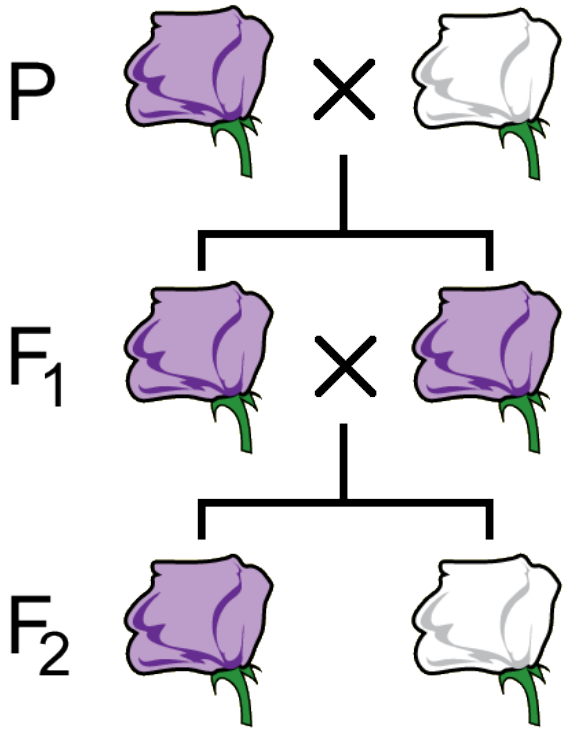


Figure 1.9
Inheritance of flower color in peas. Mendel observed that a cross between pure breeding, white and purple peas (generation P) produced only progeny (generation F_1) with purple flowers. However, white flowered plant reappeared among the F_2 generation progeny of a mating between two F_1 plants. The symbols P, F_1 and F_2 are abbreviations for parental, first filial, and second filial generations, respectively. (Original-Deyholos-CC:AN)

1.4.2 Genes and alleles

Mendel's discrete "factors of heredity" later became known as **genes**. Each hereditary factor could exist in one or more different versions or forms, which we now call **alleles**. In its narrowest definition, a gene is an abstract concept: a unit of inheritance. The connection between genes and substances like DNA and chromosomes was established largely through the experiments described in the remainder of this chapter. However, it is worth noting that Mendel and many researchers who followed him were able to provide great insights into biology, simply by observing the inheritance of specific traits – **genetics**.

1.5 The Function of Genes

1.5.1 Beadle and Tatum: one gene, one enzyme hypothesis

Life depends on (bio)chemistry to supply energy and to produce the molecules to construct and regulate cells. In 1908, A. Garrod described "in born errors of metabolism" in humans using the congenital disorder, alkaptonuria (black urine disease), as an example of how "genetic defects" led to the lack of an enzyme in a biochemical pathway and caused a disease (phenotype). Over 40 years later, in 1941, **Beadle and Tatum** built on this connection between genes and metabolic pathways. Their research led to the "**one gene, one enzyme (or protein)**" hypothesis, which states that each of the enzymes that act in a biochemical pathway is encoded by a different gene. Although we now know of many exceptions to the "one gene, one enzyme (or protein)" principle, it is generally true that each different gene produces a protein that has a distinct catalytic, regulatory, or structural function.

Beadle and Tatum used the fungus ***Neurospora crassa*** (a mold) for their studies because it had practical advantages as a laboratory organism. They knew that Neurospora was **prototrophic**, meaning that it could synthesize its own amino acids when grown on **minimal medium**, which lacked most nutrients except for a few minerals, simple sugars, and one vitamin (biotin). They also knew that by exposing Neurospora spores to X-rays, they could randomly damage its DNA to create mutations in genes. Each different spore exposed to X-rays potentially contained a mutation in a different gene. After genetically screening many, many spores for growth, most appeared to still be prototrophic and still able to grow on minimal medium. However, some spores had mutations that changed them into **auxotrophic** strains that could no longer grow on minimal medium, but did grow on **complete medium** supplemented with nutrients (Fig. 1.10). In fact, some auxotrophic mutations could grow on minimal medium with only one, single nutrient supplied, such as arginine.

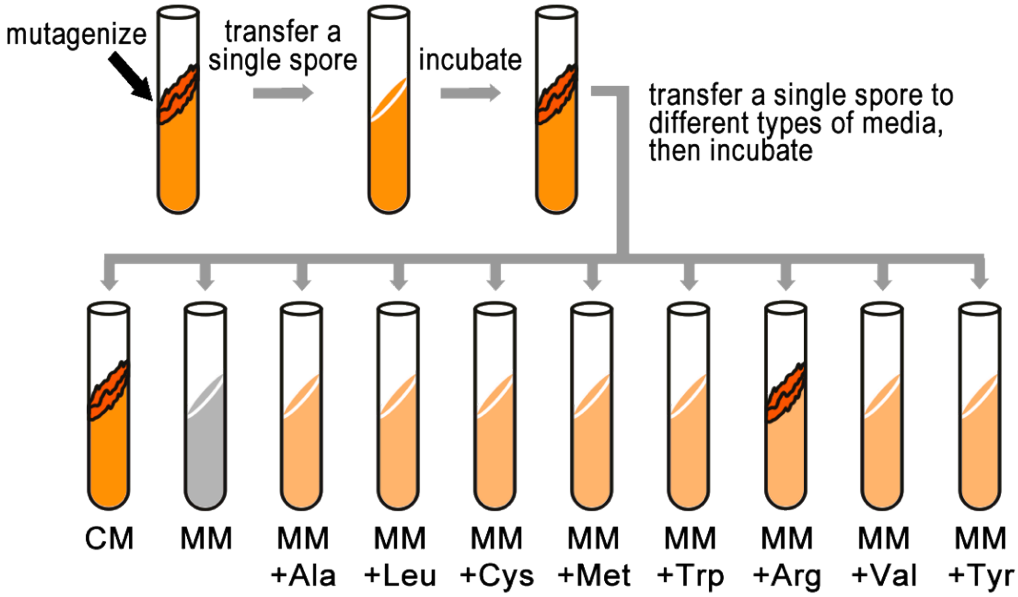


Figure 1.10
A single mutagenized spore is used to establish a colony of genetically identical fungi, from which spores are tested for their ability to grow on different types of media. Because spores of this particular colony are able to grown only on complete medium (CM), or on miminal medium supplemented with arginine (MM+Arg), they are considered Arg auxotrophs and we infer that they have a mutation in a gene in the Arg biosynthetic pathway. This type of screen is repeated many times to identify other mutants in the Arg pathway and in other pathways. (Original-Deyholos-CC:AN)

1.5.2 B&T's 1 gene: 1 enzyme hypothesis led to Biochemical Pathway dissection using genetic screens and mutations

Beadle and Tatum's experiments are important not only for its conceptual advances in understanding genes, but also because they demonstrate the utility of **screening for genetic mutants** to investigate a biological process – **genetic analysis**.

Beadle and Tatum's results were useful to investigate biological processes, specifically the metabolic pathways that produce amino acids. For example, Srb and Horowitz in 1944 tested the ability of the amino acids to **rescue** auxotrophic strains. They added one of each of the amino acids to minimal medium and recorded which of these restored growth to independent mutants.

For example, if the progeny of a mutagenized spore could grow on minimal medium only when it was supplemented with arginine (Arg), then the auxotroph must bear a

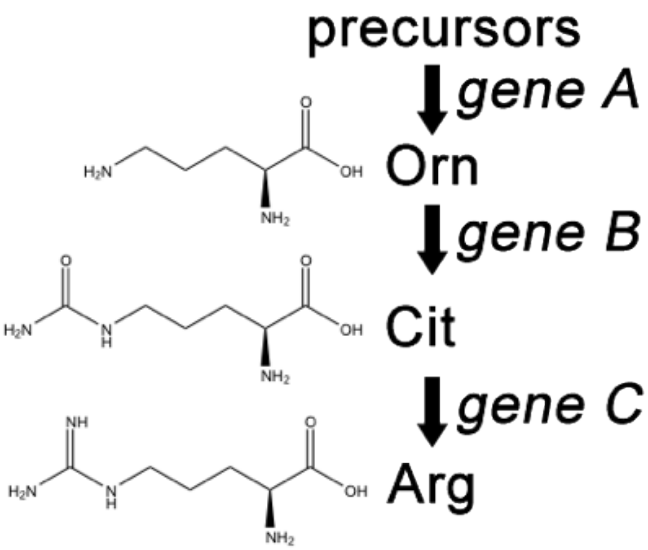


Figure 1.11
A simplified version of the Arg biosynthetic pathway, showing citrulline (Cit) and ornithine (Orn) as intermediates in Arg metabolism. These chemical reactions depend on enzymes represented here as the products of three different genes. (Original-Deyholos-CC:AN)

mutation in the Arg biosynthetic pathway and was called an "arginineless" strain (arg-).

Synthesis of even a relatively simple molecule such as arginine requires many steps, each with a different enzyme. Each enzyme works sequentially on a different intermediate in the pathway (Fig. 1.11). For arginine (Arg), two of the intermediates are ornithine (Orn) and citrulline (Cit). Thus, mutation of any one of the enzymes in this pathway could turn Neurospora into an Arg auxotroph (arg-). Srb and Horowitz extended their analysis of Arg auxotrophs by testing the intermediates of amino acid biosynthesis for the ability to restore growth of the mutants (Figure 1.12).

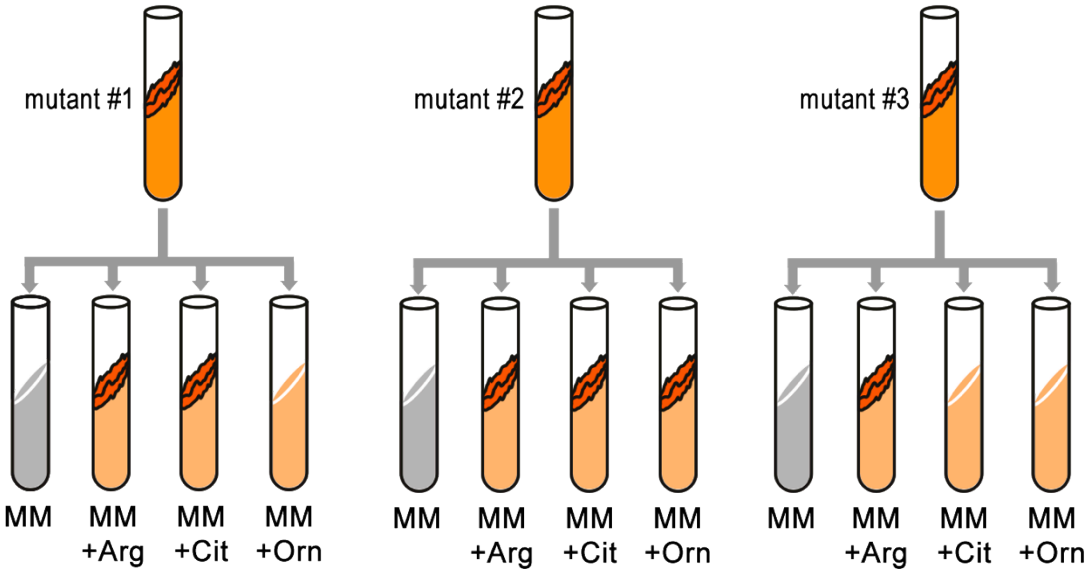


Figure 1.12
Testing different Arg auxotrophs for their ability to grow on media supplemented with intermediates in the Arg biosynthetic pathway. (Original-Deyholos-CC:AN)

They found that some of the Arg auxotrophs could be rescued only by Arg, while others could be rescued by either Arg or Cit, and still other mutants could be

rescued by Arg, Cit, or Orn (Table 1.1). Based on these results, they deduced the location of each mutation in the Arg biochemical pathway, (i.e. which gene was responsible for the metabolism of which intermediate).

1.5.3 Genetic screens for mutations help characterize biological pathways

Using many other mutations and the "one gene: one enzyme model" permitted the genetic dissection of many other biochemical and developmental pathways.

The general strategy for a **genetic screen for mutations** is to expose a population to a mutagen, then look for individuals among the progeny that have defects in the biological process of interest. There are many details that must be considered when designing a genetic screen (e.g. how can recessive alleles be made homozygous). Nevertheless, mutational analysis has been an extremely powerful and efficient tool in identifying and characterizing the genes involved in a wide variety of biological processes, including many genetic diseases in humans.

	MM + Orn	MM + Cit	MM + Arg
gene A mutants	**Yes**	**Yes**	**Yes**
gene B mutants	**No**	**Yes**	**Yes**
gene C mutants	**No**	**No**	**Yes**

Table 1.1 Ability of auxotrophic mutants of each of the three enzymes of the Arg biosynthetic pathways to grow on minimal medium (MM) supplemented with Arg or either of its precursors, Orn and Cit. Gene names refer to the labels used in Figure 1.11

1.5.4 The Central Dogma

How does the structure of DNA and genes relate to inheritance of biological traits such as the flower color of Mendel's peas? The answer lies in what has become known as molecular biology's **Central Dogma** (Fig 1.13), which has come to be described as the genetic information of each gene is encoded in DNA, and then, as needed, this information is **transcribed** into an RNA sequence, and then **translated** into a polypeptide (protein) sequence. The core of the Central Dogma is that genetic information is NEVER transferred from protein back to nucleic acids. In certain circumstances, the information in RNA may also be converted back to DNA through a process called **reverse transcription**. As well, DNA, and its information, can also be replicated (DNA➔DNA). The sequence of bases in DNA directly dictates the sequence of bases in the RNA, which in turn dictates the sequence of amino acids that make up a polypeptide. Proteins do most of the work in a cell. They (1) catalyze the formation and breakdown of most molecules within an organism as well as (2) form their structural components and (3) regulate the expression of genes. By dictating the structure of each protein, DNA affects the function of that protein, which can thereby affect the entire organism. Thus the genetic information, or **genotype**, defines the potential form, or **phenotype** of the organism.

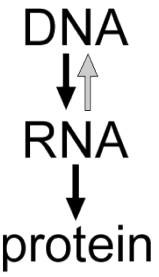


Figure 1.13 Central Dogma of molecular biology. (Original-Deyholos-CC:AN)

In the case of Mendel's peas, purple-flowered plants have a gene that encodes an enzyme that produces a purple pigment molecule. In the white-flowered plants (a purple-less mutant), the DNA for this gene has been changed, or mutated, so that it no longer encodes a functional protein. This is an example of a spontaneous, natural mutation in a biochemical pathway.

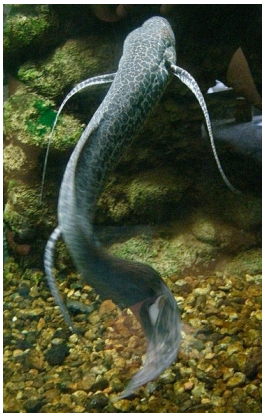


Fig 1.14
Marbled Lungfish. (Wikipedia-OpenCage-CC:AS)

1.6 The Nuclear Genome

1.6.1 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 the table below. 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* – Fig. 1.14) 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).

	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

Table 1.2
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."

1.6.2 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 in more detail in a subsequent chapter. Other non-gene DNA includes short, highly repetitive sequences of various types.

1.6.3 Other genomes

Organelles such as mitochondria and chloroplasts also have their own genomes. These are, compared to the nuclear genome, relatively small and are also circular, like the prokaryotes from which they originated (Endosymbiont hypothesis).

1.7 Model Organisms Facilitate Genetic Advances

1.7.1 Model organisms

Many of the great advances in genetics were made using species that are not especially important from a medical, economic, or even ecological perspective. Geneticists, from Mendel onwards, have sought the best organisms for their experiments. Today, a small number of species are widely used as **model organisms** in genetics (Fig 1.15). All of these species have specific characteristics

that make large number of them easy to grow and analyze in laboratories: (1) they are small, (2) fast growing with a short generation time, (3) produce lots of progeny from matings that can be easily controlled, (4) have small genomes (small c-value), and (5) are diploid (i.e. chromosomes are present in pairs).

The most commonly used model organism are:

- The prokaryote bacterium, ***Escherichia coli,*** is the simplest genetic model organism and is often used to clone DNA sequences from other model species.
- Yeast (***Saccharomyces cerevisiae***) is a good general model for the basic functions of eukaryotic cells.
- The roundworm, ***Caenorhabditis elegans*** is a useful model for the development of multicellular organisms, in part because it is transparent throughout its life cycle, and its cells undergo a well-characterized series of divisions to produce the adult body.
- The fruit fly (***Drosophila melanogaster***) has been studied longer, and probably in more detail, than any of the other genetic model organisms still in use, and is a useful model for studying development as well as physiology and even behaviour.
- The mouse (***Mus musculus***) is the model organism most closely related to humans, however some of the practical difficulties of working with mice led researchers more recently to develop zebrafish (***Danio rerio***) as a genetic model for vertebrates. Unlike mice, zebrafish embryos develop externally to their mothers and are transparent, making it easier to study their development.
- Finally, a small weed, ***Arabidopsis thaliana***, is the most widely studied plant genetic model organism. This provides knowledge that can be applied to other plant species, such as wheat, rice, and corn.

1.7.2 Society benefits from model organism research

The study of genetic model organisms has greatly increased our knowledge of genetics, and biology in general. Knowledge from model organisms has also provided **important implications in medical research, agriculture, and biotechnology**. By using these species genetic researchers can discover more knowledge, faster and cheaper than using humans, farm animals or crop plants directly. For example, at least 75% of the approximately 1,000 genes that have been associated with specific human diseases have similar genes in ***D. melanogaster***. Information about how these genes function in model organisms can usually be applied to other species, including humans. From research conducted thus far, we have learned that the main features of many biochemical, cellular, and developmental pathways tend to be common among all species. What is genetically and biochemically true in yeast, worms, flies and mice tends to be true in humans, too.

However, it is sometimes necessary to study important biological processes in non-model organisms. In humans, for example, there are some diseases or other traits for which no clear analog exists in model organisms. In these cases the tools of genetic analysis developed in model organisms can be applied to these other, non-

model species. Examples include the development of new types of gene discovery techniques, genetic mapping of desired traits, and whole genome sequencing.

Figure 1.15
Some of the most important genetic model organisms in use today. Clockwise from top left: yeast, fruit fly, arabidopsis, mouse, roundworm, zebrafish. (Original/Flickr/Wikipedia – Deyholos/M.Westby/D.Joly/Z.F.Altun/Masur/Azul – CC:AS/ANS/AN &GFDL)

SUMMARY

- Mendel demonstrated that heredity involved discrete, heritable factors that affected specific traits.
- A gene can be defined abstractly as a unit of inheritance.
- The ability of DNA from bacteria and viruses to transfer genetic information into bacteria demonstrated that DNA is the genetic material.
- 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, ensuring complementarity of the strands.
- The Central Dogma explains how DNA dictates heritable traits.
- Not all DNA in an organism contains genes.
- Model organisms accelerate the use of genetics in basic and applied research in biology, agriculture and medicine.

KEY TERMS

blending inheritance
particulate inheritance
Mendel
gene
allele
trait
P, F1, F2
Griffith
Avery, MacLeod, & McCarty
Hershey and Chase
DNase
proteinase
^{35}S
^{32}P
bacteriophage
Beadle & Tatum
auxotroph
prototroph
metabolic pathway
Neurospora crassa
Chargaff's Rules
Watson and Crick
DNA bases
sugar-phosphate backbone
anti-parallel
complementary
hydrogen bond
minor groove
major groove
adenine
cytosine
thymine
guanine
Central Dogma
transcription
reverse transcription
translation
RNA
prion
one-gene:one-enzyme
minimal medium
complete medium
arginine
genetic screen
nuclear genome
c-value paradox
model organism
Saccharomyces cerevisiae
Caenorhabditis elegans
Drosophila melanogaster
Mus musculus
Danio rerio
Arabidopsis thaliana
Escherichia coli

STUDY QUESTIONS

1.1 How would the results of the cross in Figure 1.8 have been different if heredity worked through blending inheritance rather than particulate inheritance?

1.2 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?

1.3 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?

1.4 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?

1.5 Starting with mice and R and S strains of *S. pneumoniae*, what experiments in additional to those shown in Figure 1.2 to demonstrate that DNA is the genetic material?

1.6 List the information that Watson and Crick used to deduce the structure of DNA.

1.7 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?

1.8 Compare Figure 1.11 and Table 1.1. Which of the mutants (#1, #2, #3) shown in Figure 1.11 matches each of the phenotypes expected for mutations in genes A, B,C?

1.9 Refer to Table 1.2

a) What is the relationship between DNA content of a genome, number of genes, gene density, and chromosome number?

b) What feature of genomes explains the c-value paradox?

c) Do any of the numbers in Table 1.2 show a correlation with organismal complexity?

1.10 a) List the characteristics of an ideal model organism.

b) Which model organism can be used most <u>efficiently</u> to identify genes related to:

i) eye development
ii) skeletal development
iii) photosynthesis
iii) cell division
iv) cell differentiation
v) cancer

1.11 Refer to Figure 1.5

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) that A-T rich helices. Why?

Notes:

Chapter 2 CHROMOSOMES, MITOSIS, AND MEIOSIS

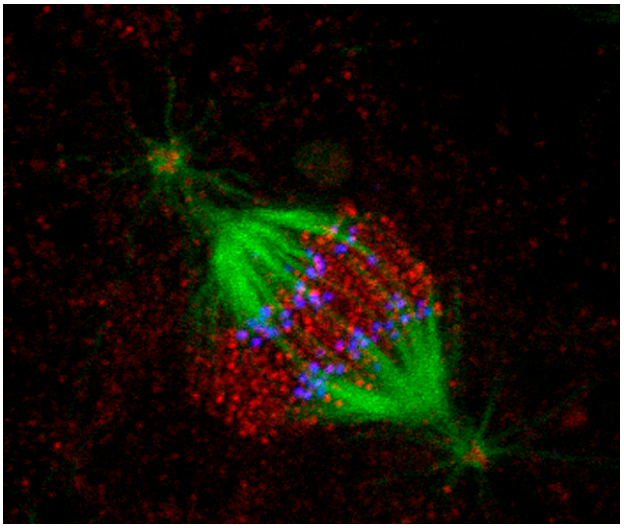


Figure 2.1
Moving chromosomes (blue) towards the poles at anaphase requires many proteins (red), all of which interact with microtubules (green). (Flickr: TheJCB-Zhang et al. (2007) J. Cell Biol. 177:231-242.- CC: ANS)

Chromosomes contain genetic information. We often take this fact for granted, but just over a century ago, even the best biologists in the world were uncertain of the function of these rod-shaped structures. We now know that most chromosomes contain a single molecule of double-stranded DNA that form a complex with proteins. This arrangement allows very long DNA molecules to be compacted into a small volume that can more easily be moved during mitosis and meiosis (Fig 2.1) and expressed during interphase. The compact structure also makes it easier for pairs of chromosomes to align with each other during meiosis. Finally, we shall see that chromosomal structure can affect whether genes are active or silent.

2.1 DNA IS PACKAGED INTO CHROMATIN

2.1.1 DNA CAN BE HIGHLY COMPACTED

If stretched to its full length, the DNA molecule of 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.

2.1.2 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 (Fig. 2.2). 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 H1** 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.

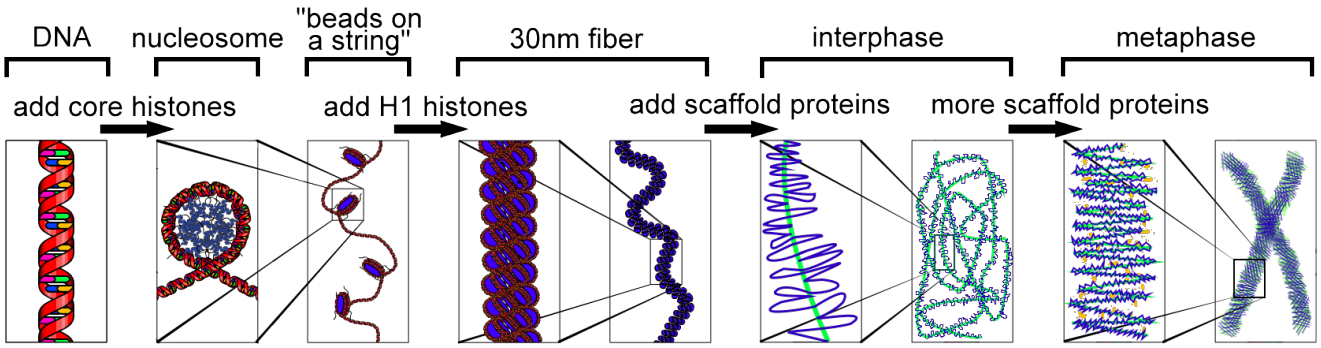


Figure 2.2
Successive stages of chromosome compaction depend on the introduction of additional proteins. (Wikipedia-R. Wheeler-GFDL)

2.1.3 Chromatin: 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.

2.1.4 Morphological features of chromosomes

Figure2.3
A pair of metacentric chromosomes. The arrow shows a centromeric region (primary constriction). (Unkonwn-Unknown-pending)

Chromosomes also contain other distinctive features such as centromeres and telomeres. Both of these are usually heterochromatin. In most cases, each chromosome contains one **centromere**. These sequences are bound by centromeric proteins that link the centromere to microtubules that transport chromosomes during cell division. Under the microscope, centromeres of metaphase chromosomes can sometimes appear as constrictions in the body of the chromosome (Fig. 2.3) and are called primary (1°) constrictions. If a centromere is located near the middle of a chromosome, it is said to be **metacentric**, while an **acrocentric** centromere is closer to one end of a chromosome, and a **telocentric** chromsome is at, or near, the very end. In contrast, some species have a **holocentric** centromere, where no single centromere can be defined and the entire chromsome acts as the centromere. **Telomeres** are repetitive sequences near the ends of linear chromosomes, and are important in maintaining the length of the chromosomes during replication, and protecting the ends of the chromosomes from alterations.

It is essential to describe the similarity between chromosomes using appropriate terminology (Fig 2.4). **Homologous chromosomes** are typically pairs of similar, but non-identical, chromosomes in which one member of the pair comes from the male parent, and the other comes from the female parent. Homologs contain the same gene loci but not necessarily the same alleles. **Non-homologous chromosomes** contain different gene loci, and are usually distinguishable based on cytological features such as length, centromere position, and banding patterns.

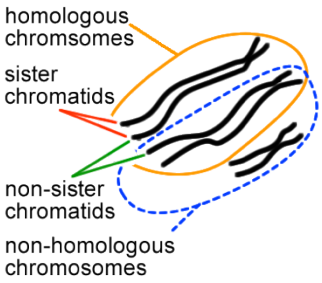


Figure 2.4
Relationships between chromosomes and chromatids.
(Original-Deyholos-CC:AN)

An **unreplicated chromosomes** can undergo replication, to produce a **replicated chromosome** that has two **sister chromatids,** which are physically connected to each other at the centromere and remain joined until cell division. Because a pair of sister chromatids is produced by the replication of a single DNA molecule, their sequences are essentially identical (same alleles), differing only because of DNA replication errors. On the other hand, **non-sister chromatids** come from two separate, but homologous chromosomes, and therefore usually contain the same gene loci in the same order, but do not necessarily have identical DNA sequences (allelic differences).

The decondensed chromosomes are not randomly arranged in within the interphase nucleus. They often have specific locations within the nucleus and relative to one another (Fig. 2.5)

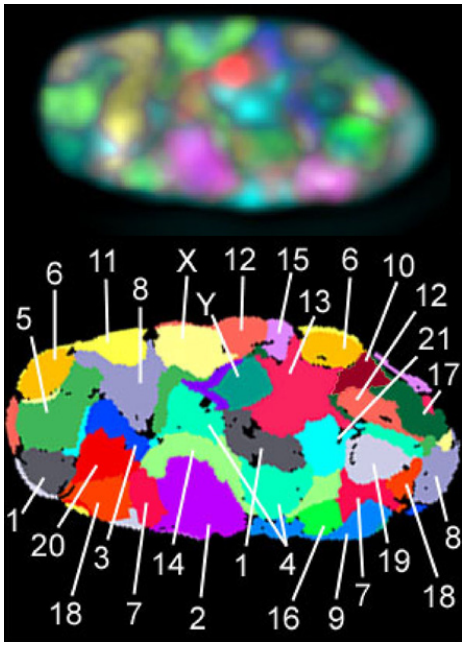


Figure 2.5
Top: FISH (Fluorescence in situ hybridization) labeling of all 24 different human chromosomes (1 - 22, X, and Y) in a fibroblast nucleus, each with a different combination of in total seven fluorochromes.
Bottom: False color representation of all chromosome territories visible in this mid-section after computer classification.
(BMC-Unknown-pending)

2.1.5 Chromosome and DNA Replication

When the cell enters S-phase in the cell cycle (G1-S-G2-M) all the 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 2.6a).

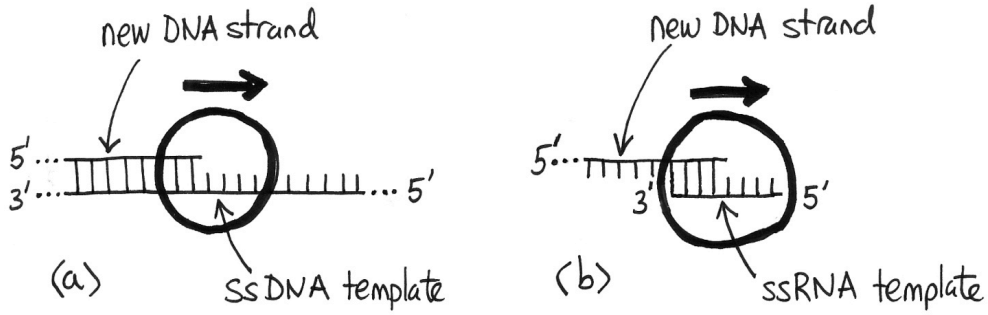


Figure 2.6 DNA polymerases make new strands in a 5' to 3' direction. (a) Regular DNA polymerases are proteins or protein complexes that use a single strand of DNA as a template. For example, the main human DNA polymerase, Pol α, is large protein complex made of four polypeptides. (b) Telomerases use their own RNA as a template. The human telomerase is a complex made of one polypeptide and one RNA molecule.(Original-Harrington-CC:AN)

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 (Fig. 2.7).

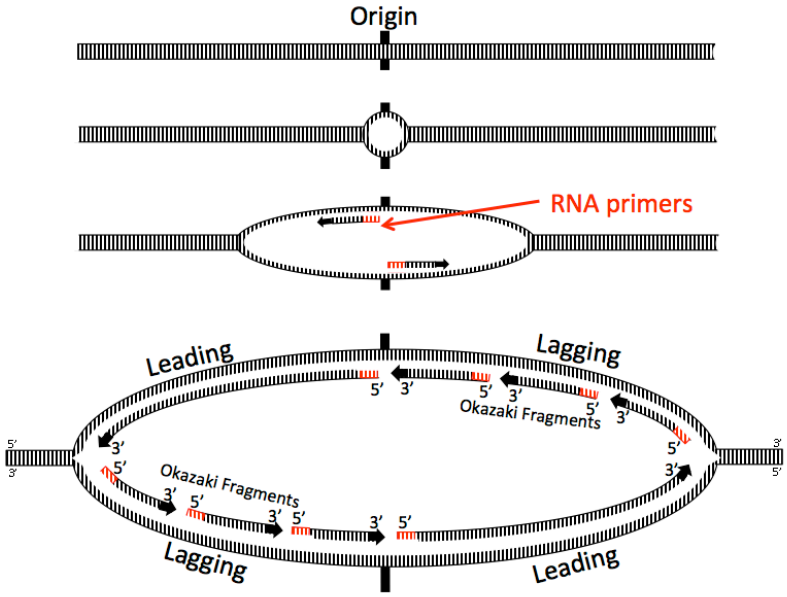


Figure 2.7
An origin of replication. The sequence specific DNA duplex is melted then the primase synthesizes RNA primers from which bidirectional DNA replication begins as the two replication forks head off in opposite directions. The leading and lagging strands are shown along with Okazaki fragments. Note the 5' and 3' orientation of all strands. (Original-Locke-CC:AN)

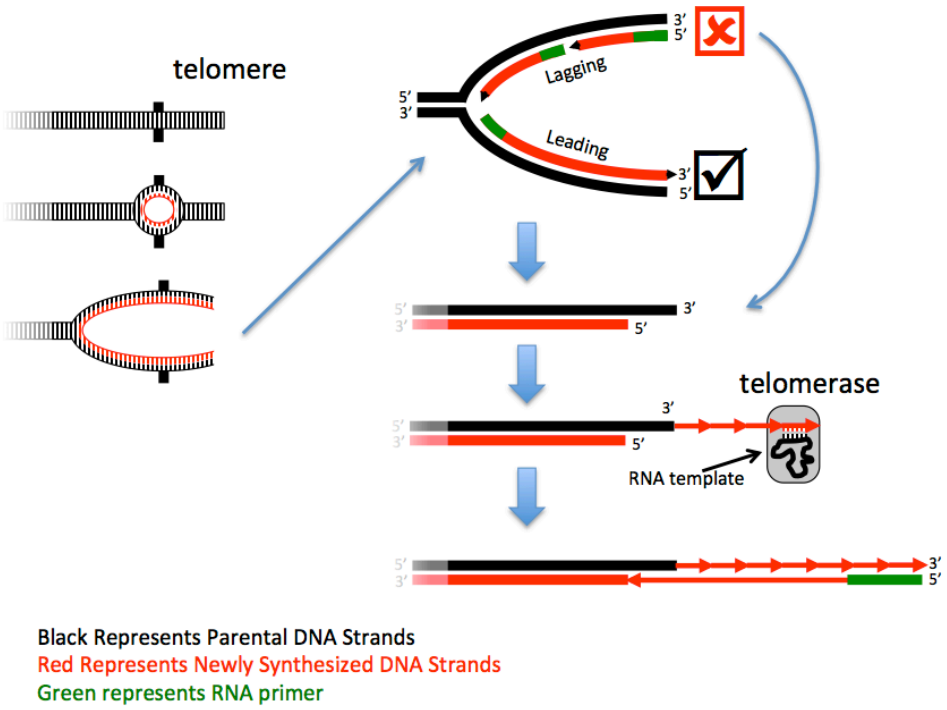


Figure 2.8
Telomere replication showing the completion of the leading strand and incomplete replication of the lagging strand. The gap is replicated by the extension of the 3' end by telomerase and then filled in by extension of an RNA primer.
(Original-Locke-CC:AN)

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 2.6b 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 (Fig. 2.8). 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.

2.1.6 Eukaryote Chromosomes have Multiple Origins of Replication

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 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 (Fig 2.9).

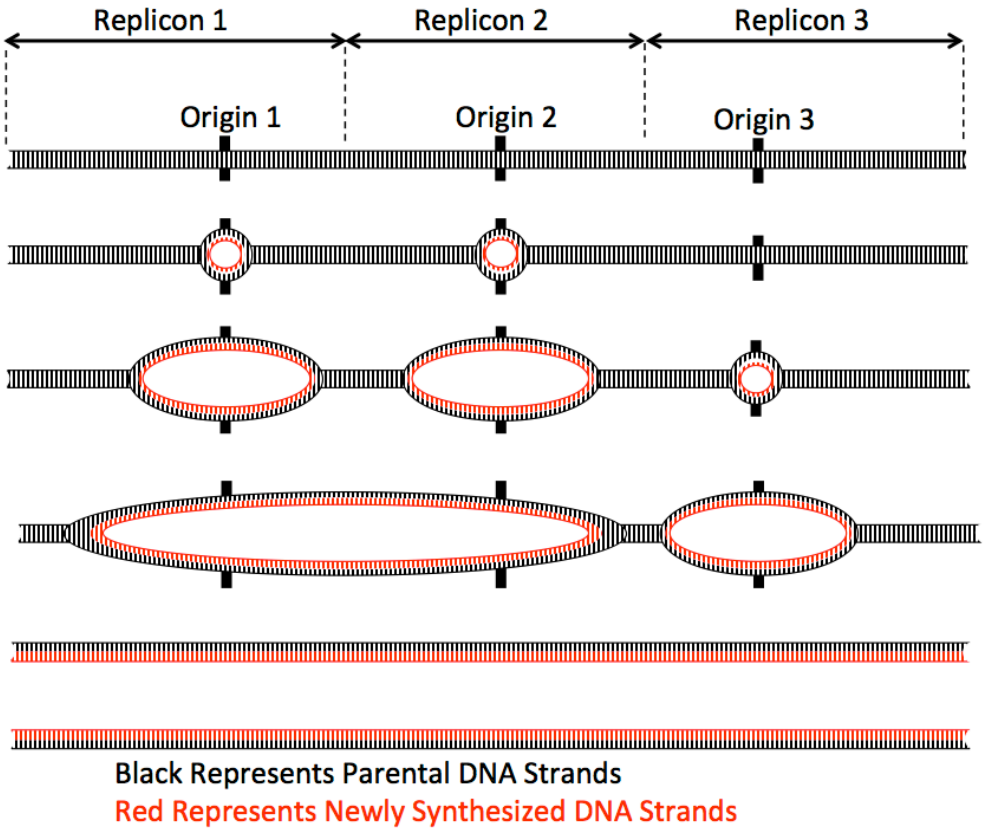


Figure 2.9. Part of a eukaryote chromosome showing multiple Origins (1, 2, 3) of Replication, each defining a **replicon** (1, 2, 3). Replication may start at different times in S-phase. Here #1 and #2 begin first then #3. As the replication forks proceed bi-directionally, they create what are referred to as "**replication bubbles**" that meet and form larger bubbles. The end result is two semi-conservatively replicated duplex DNA strands. (Original-Locke-CC:AN)

2.2 Mitosis

Cell growth and division is essential to asexual reproduction and the development of multicellular organisms. Accordingly, the primary function of mitosis is to ensure that at division each daughter cell inherits identical genetic material, i.e. exactly one

copy of each chromosome. To make this happen, replicated chromosomes condense (**prophase**), and are positioned near the middle of the dividing cell (**metaphase**), and then each of the sister chromatids from each chromosome migrates towards opposite poles of the dividing cell (**anaphase**), until the identical sets of unreplicated chromosomes are completely separated from each other within the two newly formed daughter cells (**telophase**) (see Figures 2.10 and 2.11 for diagrams of the process). This is followed by the division of the cytoplasm (**cytokinesis**) to complete the process. The movement of chromosomes occurs through microtubules that attach to the chromosomes at the centromeres.

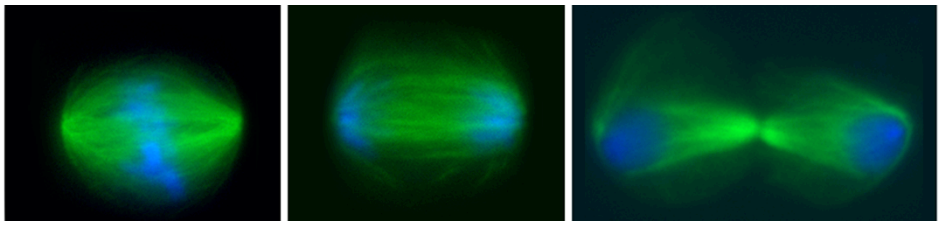


Figure 2.10. Mitosis in arabidopsis showing fluorescently labeled chromosomes (blue) and microtubules (green) at metaphase, anaphase and telophase (from left to right). (PLoS Genetics-Somma MP et al. (2008) PLoS Genets 4(7): e1000126 – PD)

2.3 MEIOSIS

Most eukaryotes replicate sexually - a cell from one individual joins with a cell from another to create the next generation. 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! The reduction in chromosome number is achieved by the process of **meiosis**. In meiosis, there are usually two steps, Meiosis I and II. In Meiosis I homologous chromosomes segregate, while in Meiosis II sister chromatids segregate . 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.

2.3.1 MEIOSIS I AND II

Meiosis begins similarly to mitosis (a cell has replicated its chromosomes and grown large enough to divide), but requires two rounds of division (Figure 2.8). In the first, known as meiosis I, the homologous chromosomes separate and segregate. During meiosis II the sister chromatids separate and segregate. Note how meosis I and II are both divided into prophase, metaphase, anaphase, and telophase. After two rounds of cytokinesis, four cells will be produced, each with a single copy of each chromosome.

Cells that will undergo the process of meiosis are called **meiocytes** and are diploid (2N). Meiosis is divided into two stages designated by the roman numerals I and II. Meiosis I is called a **reductional** division, because it reduces the number of chromosomes inherited by each of the daughter cells. 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 with each other, or **synapse**, and are called **bivalents** (Figs. 2.12). 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, as described later. During Anaphase I, one member of each pair of homologous chromosomes migrates to each daughter cell (1N). 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.

In meiosis I replicated, homologous chromosomes **pair up**, or **synapse**, during prophase I, lining up in the middle of the cell during metaphase I, and separating during anaphase I. For this to happen the homologous chromosomes need to be brought together while they condense during prophase I. These attachments are formed in two ways. 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**. The proteins are released when the cell enters anaphase I.

Within the synaptonemal complex a second event, **crossingover**, occurs. These are places where DNA repair enzymes break the DNA 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.

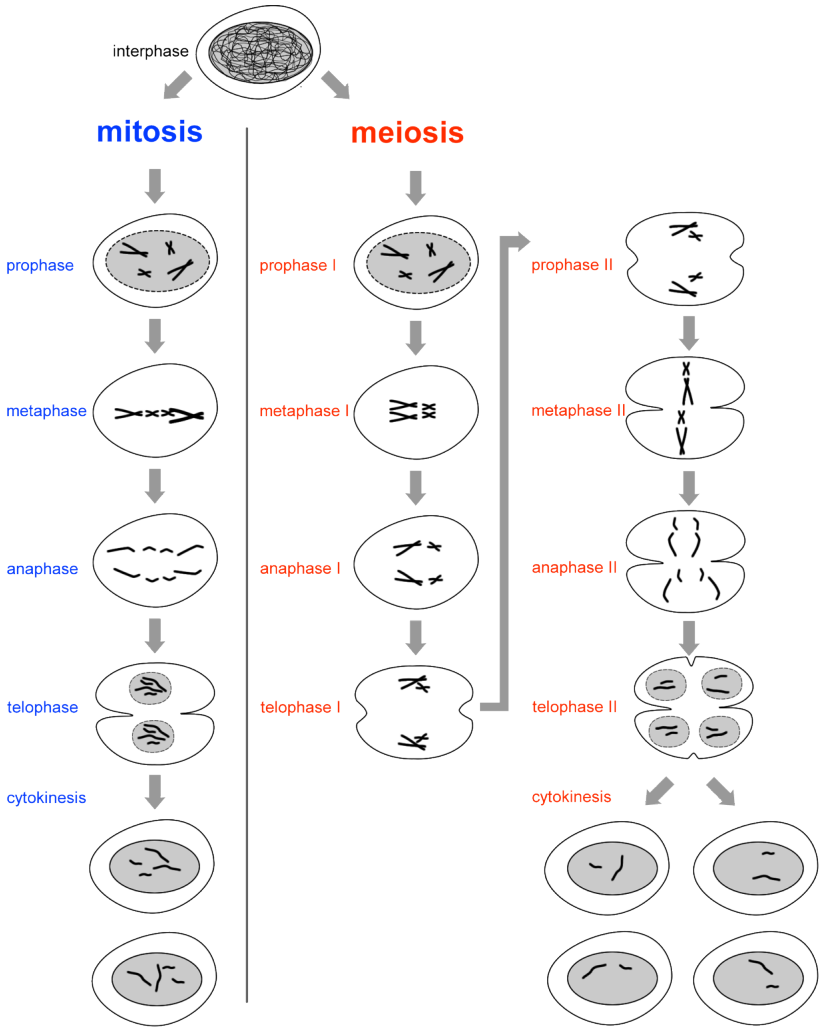


Figure 2.11
Mitosis and meiosis. Note the similarities and differences between metaphase in mitosis and metaphase I and II of meiosis. (Original-Deyholos-CC:AN)

Crossovers function to hold homologous chromosomes together during meiosis I so they segregate successfully and they also cause the reshuffling of gene/allele combinations to create genetic diversity, which can have an important effect on evolution (see Chapter 7).

2.3.2 Stages of Prophase I

In meiosis, Prophase I is divided up into five visual stages, that are steps along a continuum of events. Leptotene, zygotene, pachytene, diplotene and diakinesis. From interphase, a cell enters **leptotene** as the nuclear material begins to condense into long visible threads (chromosomes). During **Zygotene** homologous chromosomes begin to pair up (synapse) and form an elaborate structure called the synaptonemal complex along their length. At **pachytene** homologous chromosomes are fully synapsed (two chromosomes and four chromatids) to form **bivalents**. **Crossing over** takes place in pachytene. After this, the pairing begins to loosen and individual chromatids become apparent in **diplotene**. This is when the consequences of each crossing over event can be seen as a **chiasma** (plural: **chiasmata**). **Diakinesis** follows as the chromosomes continue to condense and individualize. This is followed by metaphase I were the paired chromosomes orient on the metaphase plate in preparation for segregation (reductional).

2.3.3 Meiosis II and Gamete Maturation

At the completion of meiosis I there are two cells, each with one, replicated copy of each chromosome (1N). Because the number of chromosomes per cell has decreased (2->1), meiosis I is called a **reductional cell division**. In the second part of meiosis the chromosomes will once again be brought to the middle of the cell, but this time it is the sister chromatids that will segregate during anaphase.

After cytokinesis there will be four cells, each containing only one unreplicated chromosome of each type. Meiosis II resembles mitosis in that the number of chromosomes per cell is unchanged - both are **equational cell divisions –** but in meiosis II all cells won't have the same genetic composition. There will be allelic differences among the gametes.

In animals and plants the cells produced by meiosis need to mature before they become functional gametes. In male animals the four products of meiosis are called spermatids. They grow 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

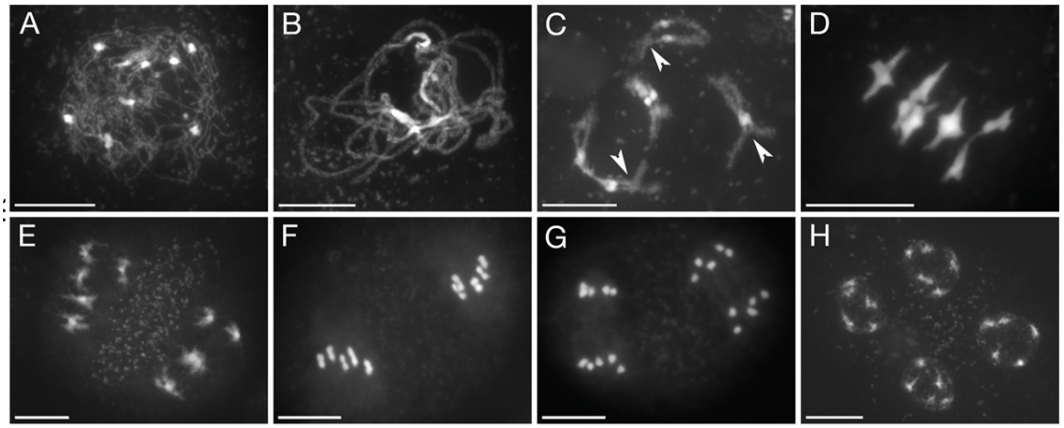


Figure 2.12 Meiosis in Arabidopsis (n=5). Panels A-C show different stages of prophase I, each with an increasing degree of chromosome condensation. Subsequent phases are shown: metaphase I (D), telophase I (E), metaphase II (F), anaphase II (G), and telophase II (H). (PLoS Genetics-Chelysheva, L. et al (2008) PLoS Genetics-PD)

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.

2.4 The Cell Cycle and Changes in DNA Content

2.4.1 Four stages of a typical cell cycle

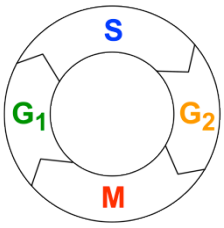


Figure 2.13 A typical eukaryotic cell cycle. (Original-Deyholos-CC:AN)

The life cycle of eukaryotic cells can generally be divided into four stages and a typical cell cycle is shown in Fig. 2.13. When a cell is produced through fertilization or cell division, there is usually a lag before it undergoes DNA synthesis (replication). This lag period is called Gap 1 (**G_1**), and ends with the onset of the DNA synthesis (**S**) phase, during which each chromosome is replicated. Following replication, there may be another lag, called Gap 2 (**G_2**), before mitosis (**M**). Cells undergoing meiosis do not usually have a G_2 phase. **Interphase** is as term used to include those phases of the cell cycle excluding mitosis and meiosis. Many variants of this generalized cell cycle also exist. Some cells never leave G_1 phase, and are said to enter a permanent, non-dividing stage called **G_0**. On the other hand, some cells undergo many rounds of DNA synthesis (S) without any mitosis or cell division, leading to endoreduplication. Understanding the control of the cell cycle is an active area of research, particularly because of the relationship between cell division and cancer.

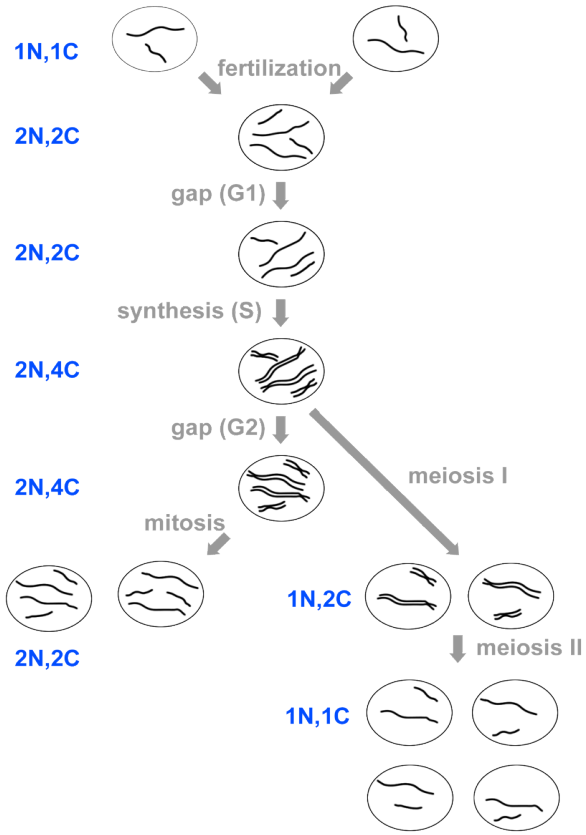


Figure 2.14 Changes in DNA and chromosome content during the cell cycle. For simplicity, nuclear membranes are not shown, and all chromosomes are represented in a similar stage of condensation. (Original-Deyholos-CC:AN)

2.4.2 Measures of DNA content and chromosome content

The amount of DNA within a cell changes following each of the following events: fertilization, DNA synthesis, mitosis, and meiosis (Fig 2.14). 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 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. In contrast, the 4 cells that come from meiosis of a 2n, 4c cell are each 1c and 1n, since each pair of sister chromatids, and each pair of homologous chromosomes, divides during meiosis.

2.5 Karyotypes Describe Chromosome Number and Structure

2.5.1 Karyograms are images of real chromosomes

Each eukaryotic species has its nuclear genome divided among a number of chromosomes that is characteristic of that species. For example, a haploid human nucleus (i.e. sperm or egg) normally has 23 chromosomes (n=23), and a diploid human nucleus has 23 pairs of chromosomes (2n=46). A **karyotype** is the complete set of chromosomes of an individual. The cell was in metaphase so each of the 46 structures is a replicated chromosome even though it is hard to see the two sister chromatids for each chromosome at this resolution. As expected there are 46 chromosomes. Note that the chromosomes have different lengths. In fact, human chromosomes were named based upon this feature. Our largest chromosome is called 1, our next longest is 2, and so on. By convention the chromosomes are arranged into the pattern shown in Figure 2.15 and the resulting image is called a **karyogram**. A karyogram allows a geneticist to determine a person's karyotype - a written description of their chromosomes including anything out of the ordinary.

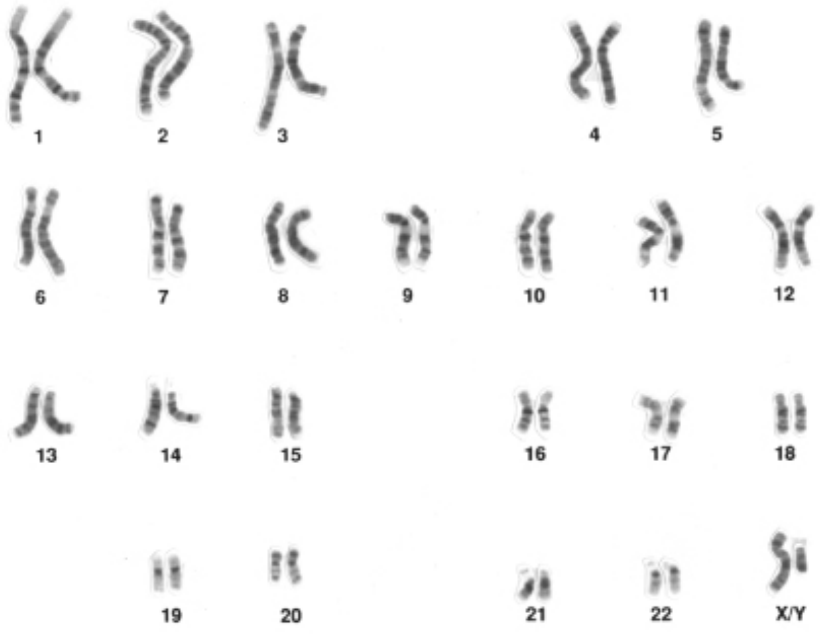


Figure 2.15
Karyogram of a normal human male karytype.
(Wikipedia-NHGRI-PD)

Various stains and fluorescent dyes are used to produce characteristic banding patterns to distinguish all 23 chromosomes. The number of chromosomes varies between species, but there appears to be very little correlation between chromosome number and either the complexity of an organism or its total amount genomic DNA.

2.5.2 Autosomes and Sex Chromosomes

In the figure above note that most of the chromosomes are paired (same length, centromere location, and banding pattern). These chromosomes are called **autosomes**. However note that two of the chromosomes, the X and the Y do not look alike. These are **sex chromosomes**. 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 flowering plants embryos, those with two X chromosomes develop into females while those with an X and a Y become males. In birds, moths, and butterflies males are ZZ and females are ZW. 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 .

In mammals, the sex chromosomes evolved just after the divergence of the monotreme lineage from the lineage that led to placental and marsupial mammals. Thus nearly every mammal species uses the same sex determination system. During embryogenesis the gonads will develop into either ovaries or testes. A gene present only on the Y chromosome called *TDF* encodes a protein that makes the gonads mature into testes. XX embryos do not have this gene and their gonads mature into ovaries instead (default). 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.

How do the sex chromosome behave during meiosis? Well, in those individuals with two of the same chromosome (i.e. **homogametic** sexes: XX females and ZZ males) the chromosomes pair and segregate during meiosis I the same as autosomes do. During meiosis in XY males or ZW females (**heterogametic** sexes) the sex chromosomes pair with each other (Figure 2.16). In mammals 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.

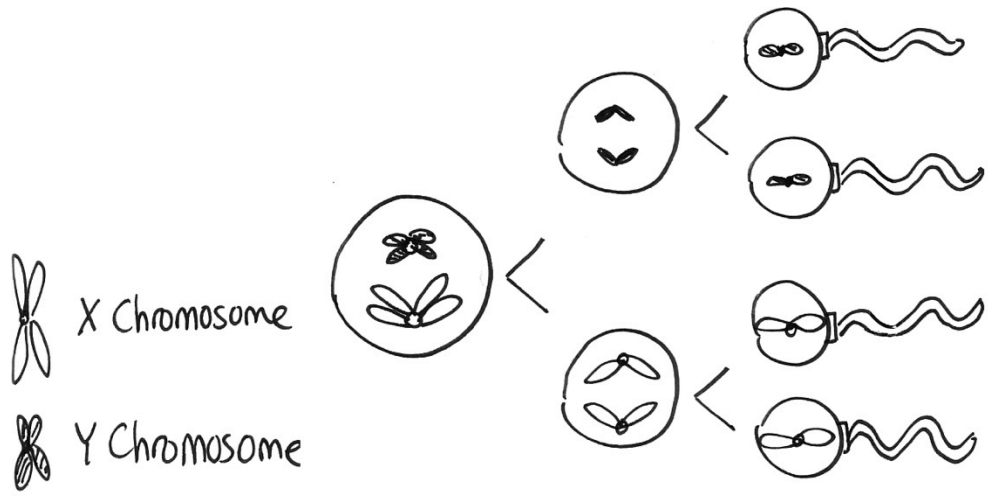


Figure 2.16
Meiosis in an XY mammal. The stages shown are anaphase I, anaphase II, and mature sperm. Note how half of the sperm contain Y chromosomes and half contain X chromosomes. (Original-Harrington-CC:AN)

2.5.2 Aneuploidy - Changes in Chromosome Number

Analysis of karyotypes can identify chromosomal abnormalities, including **aneuploidy**, which is the addition or subtraction of a chromosome from a pair of homologs. More specifically, the absence of one member of a pair of homologous chromosomes is called **monosomy** (only one remains). On the other hand, in a **trisomy**, there are three, rather than two (disomy), homologs of a particular chromosome. Different types of aneuploidy are sometimes represented symbolically; if **2n** symbolizes the normal number of chromosomes in a cell, then **2n-1** indicates monosomy and **2n+1** represents trisomy. The addition or loss of a whole chromosome is a mutation, a change in the genotype of a cell or organism.

The most familiar human aneuploidy is trisomy-21 (i.e. three copies of chromosome 21), which is one cause of **Down syndrome**. Most (but not all) other human aneuploidies are lethal at an early stage of embryonic development. Note that aneuploidy usually affects only one set of homologs within a karyotype, and is therefore distinct from **polyploidy**, in which the entire chromosome set is duplicated (see below). Aneuploidy is almost always deleterious, whereas polyploidy appears to be beneficial in some organisms, particularly many species of food plants.

Aneuploidy can arise due to a **non-disjunction** event, which is the failure of at least one pair of chromosomes or chromatids to segregate during mitosis or meiosis. Non-disjunction will generate gametes with extra and missing chromosomes.

2.5.3 Chromosomal abnormalities

Structural defects in chromosomes are another type of abnormality that can be detected in karyotypes (Fig 2.17). These defects include **deletions, duplications**, and **inversions**, which all involve changes in a segment of a single chromosome. **Insertions** and **translocations** involve two non-homologous chromosomes. In an insertion, DNA from one chromosome is moved to a non-homologous chromosome in a unidirectional manner. In a translocation, the transfer of chromosomal segments is bidirectional and reciprocal – a reciprocal translocation.

Structural defects affect only part of a chromosome (a subset of genes), and so tend to be less harmful than aneuploidy. In fact, there are many examples of ancient chromosomal rearrangements in the genomes of species including our own. Duplications of some small chromosomal segments, in particular, may have some evolutionary advantage by providing extra copies of some genes, which can then evolve in new, potentially beneficial, ways.

Chromosomal abnormalities arise in many different ways, some of which can be traced to rare errors in natural cellular processes such as DNA replication. **Chromosome breakage** also occurs infrequently as the result of physical damage (such as ionizing radiation), movement of some types of transposons, and other factors. During the repair of a broken chromosome, deletions, insertions, translocations and even inversions can be introduced.

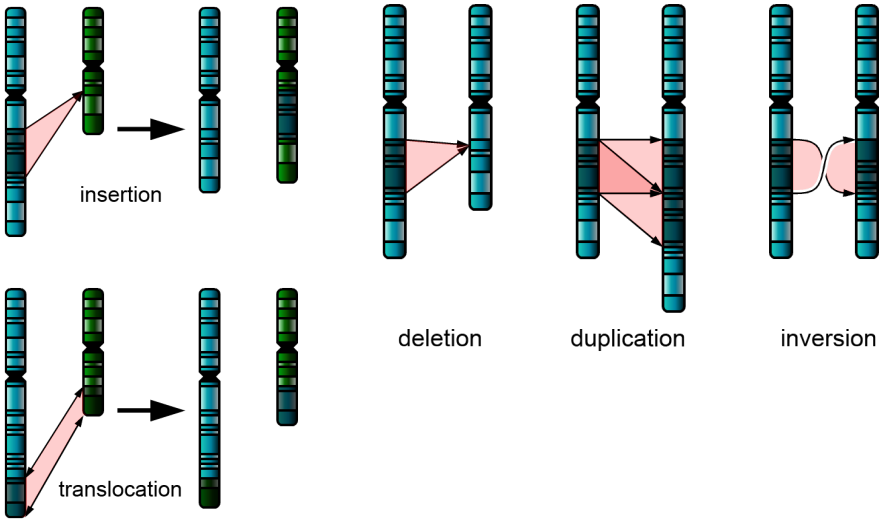


Figure 2.17
Structural abberations in chromosomes. (Wikipedia-Zephyris-GFDL)

2.5.4 Structural defects and pairing of chromosomes in meiosis.

The loss or gain of chromosome segments interferes with the pairing of homologous chromosomes in prophase I of meiosis. With deletions and duplications there is a region of unpaired chromosome that forms a loop out from the rest of the paired chromosome. These loops can be seen cytologically and used to determine the location of the aberration along the chromosome. In inversions, the pairing also forms a loop and crossovers within the loop region led to the formation of unbalanced gamete products that, when combined with a normal gamete at fertilization, produce lethal zygotes (see Gene Balance below). Translocations also can produce unbalance gametes because the proper segregation of chromosomes is compromised. There are three possible ways that two translocated chromosomes can segregate (alternate, adjacent-I, adjacent-2) and only one leads to balanced gametes.

2.6 Polyploidy arises from changes in whole sets of chromosomes

If you refer back to Fig. 2.15 you can see that humans, like most animals and most eukaryotic genetic model organisms, have two copies of each autosome. This situation is called **diploidy**. This means that most of their cells have two homologous copies of each chromosome. In contrast, many plant species and even a few animal species are **polyploids**. This means they have more than two

chromosome sets, and so have more than two homologs of each chromosome in each cell.

When the nuclear content changes by a whole chromosome set we call it a change in ploidy. Gametes are haploid (1n) and thus most animals are diploid (2n), formed by the fusion of two haploid gametes. However, some species can exist as monoploid (1x), triploid (3x), tetraploid (4x), pentaploid (5x), hexaploid (6x), or higher.

2.6.1 Notation of ploidy

When describing polyploids, we use the letter "**x**" (not "n") to define the level of ploidy. A diploid is 2x, because there are two basic sets of chromosomes, and a tetraploid is 4x, because it contains four chromosome sets. For clarity when discussing polyploids, geneticists will often combine the "x" notation with the "n" notation already defined previously in this chapter. Thus for both diploids and polyploids, "n" is the number of chromosomes in a gamete, and "2n" is the number of chromosomes following fertilization. For a diploid, therefore, n=x, and 2n=2x. But for a tetraploid, n=2x, and 2n=4x and for a hexaploid, n=3x, and 2n=6x.

2.6.2 Male Bees are Monoploid

Monoploids, with only one set, are usually inviable in most species, however, in many species of hymenoptera (bees, wasps, ants) the males are monoploid and develop from unfertilized eggs. These males don't undergo meiosis for gametes; mitosis produces sperm. Females are diploid (from fertilized eggs) and produce eggs via meiosis. This is the basis for the haploid-diploid sex determination system (not the X/Y chromosome system). Female bees are diploid (2n=32) and are formed when an egg (n=16) is fertilized by a sperm (n=16). If an egg isn't fertilized it can still develop and the result is a n=16 male drone. Males are described as haploid (because they have the same number of chromosomes as a gamete) or monoploid (because they have only one chromosome set). Females produce eggs by meiosis while males produce sperm by mitosis. This form of sex determination produces more females – workers, which do the work (Fig. 2-18) than males, who are only needed for reproduction.

Figure 2.18
A worker caste European honey bee, which is female and diploid. Male drones are haploid. (Wikipedia-J. Severns-PD)

2.6.3 Polyploids can be stable or sterile

Like diploids (2n=2x), stable polyploids generally have an even number of copies of each chromosome: tetraploid (2n=4x), hexaploid (2n=6x), and so on. The reason for this is clear from a consideration of meiosis. Remembering that the purpose of meiosis is to reduce the sum of the genetic material by half, meiosis can equally divide an even number of chromosome sets, but not an odd number. Thus, polyploids with an odd number of chromosomes (e.g. triploids, 2n=3x) tend to be **sterile**, even if they are otherwise healthy.

The mechanism of meiosis in stable polyploids is essentially the same as in diploids: during metaphase I, homologous chromosomes pair with each other. Depending on the species, all of the homologs may be aligned together at metaphase, or in multiple separate pairs. For example, in a tetraploid, some species may form **tetravalents** in which the four homologs from each chromosome align together, or alternatively, two pairs of homologs may form two bivalents. Note that because that mitosis does not involve any pairing of homologous chromosomes, mitosis is equally effective in diploids, even-number polyploids, and odd-number polyploids.

2.6.4 Many Crop Plants are Hexaploid or Octoploid

Polyploid plants tend to be larger and healthier than their diploid counterparts. The strawberries sold in grocery stores come from **octoploid** (8x) strains and are much larger than the strawberries formed by wild diploid strains. An example is bread wheat which is a **hexaploid** (6x) strain. This species is derived from the combination of three other wheat species, *T. monococcum* (chromosome sets = AA), *T. searsii* (BB), and *T. tauschii* (DD). Each of these chromosome sets has 7 chromosomes so the diploid species are 2n=2x=14 and bread wheat is 2n=6x=42 and has the chromosome sets AABBDD. Bread wheat is viable because each chromosome behaves independently during mitosis. The species is also fertile because during meiosis I the A chromosomes pair with the other A chromosomes, and so on. Thus, even in a polyploid, homologous chromosomes can segregate equally and gene balance can be maintained.

2.6.5 Bananas, Watermelons, and Other Seedless Plants are Triploid

Figure 2.19
Seedless watermelon is triploid, with the white, aborted seeds within the flesh.
(Flickr-Darwin Bell-CC:AN)

The bananas found in grocery stores are a seedless variety called Cavendish. They are a **triploid** variety (chromosome sets = AAA) of a normally diploid species called *Musa acuminata* (AA). Cavendish plants are viable because mitosis can occur. However they are sterile because the chromosomes cannot pair properly during meiosis I. During prophase I there are three copies of each chromosome trying to "pair" with each other. Because proper chromosome segregation in meiosis fails, seeds cannot be made and the result is a fruit that is easier to eat because there are no seeds to spit out. Seedless watermelons (Figure 2.19) have a similar explanation.

If triploids cannot make seeds, how do we obtain enough triploid individuals for cultivation? The answer depends on the plant species involved. In some cases, such as banana, it is possible to propagate the plant asexually; new progeny can simply be grown from cuttings from a triploid plant. On the other hand, seeds for seedless watermelon are produced sexually: a tetraploid watermelon plant is crossed with a diploid watermelon plant. Both the tetraploid and the diploid are fully fertile, and produce gametes with two (1n=2x) or one (1n=1x) sets of chromosomes, respectively. These gametes fuse to produce a zygote (2n=3x) that is able to develop normally into an adult plant through multiple rounds of mitosis, but is unable to compete normal meiosis or produce seeds.

Polyploids are often larger in size than their diploid relatives (Figure 2.20). This feature is used extensively in food plants. For example, most strawberries you eat are not diploid, but octoploid (8x).

Polyploidy in animals is rare, essentially limited to lower forms, which often reproduce by parthenogenesis.

Credit: www.jamesandthegiantcorn.com/wp-content/uploads/2009/11/strawberries2.jpg

Figure 2.20 Polyploidy in Strawberries. The sweet, flavorful, wild diploid is on the left, while the huge, cultivated octoploid is on the right.

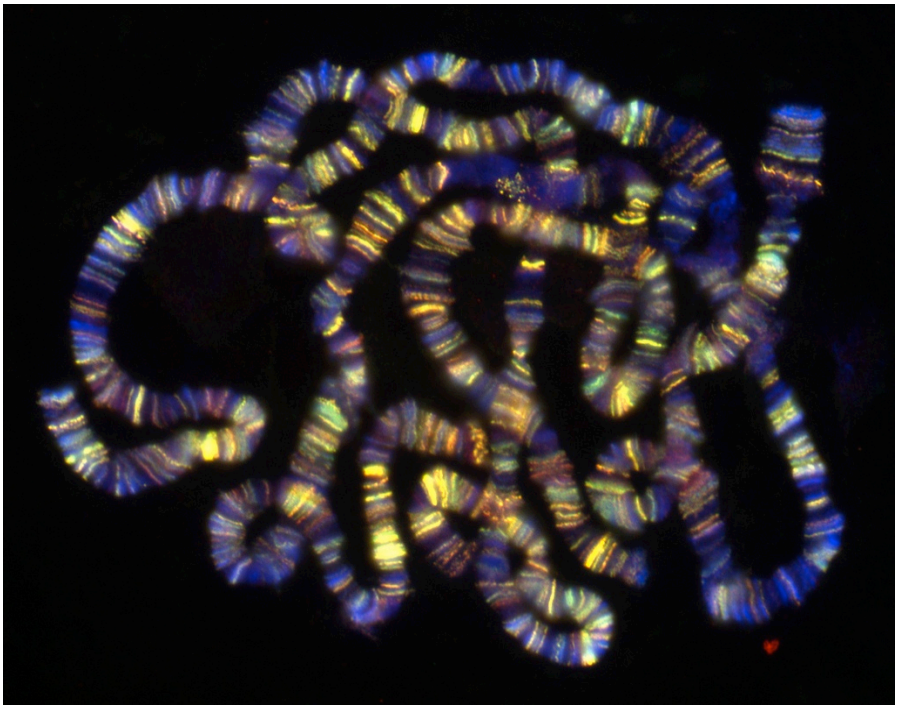


Figure 2.21. Endoreduplicated chromosomes from a Drosophila salivary gland cell. The banding pattern is produced with fluorescent labels. (Flickr-Elissa Lei, Ph.D. @ NIH-CCA)

2.7 ENDOREDUPLICATION

Endoreduplication, is a special type of tissue-specific genome amplification that occurs in many types of plant cells and in specialized cells of some animals including humans. Endoreduplication does not affect the germline or gametes, so species with endoreduplication are not considered polyploids. Endoreduplication occurs when a cell undergoes extra rounds of DNA synthesis (S-phase) without any mitosis or cytokinesis to produce an endopolyploid cell. This produces multiple chromatids of each chromosome. Endopolyploidy seems to be associated with cells that are metabolically very active, and produce a lot of enzymes and other proteins in a short period of time. An example is the highly endoreduplicated **salivary gland polytene** chromosomes of *D. melanogaster* (Figure 2.21) which can have over 1,000 chromatids that align together and form giant chromosomes that show a banding pattern that reflects the underlying DNA sequence and genes in that chromosome region. These chromosomes have been wonderful research models in genetics, since their relatively large, amplified size makes it easy to identify and study a wide variety of chromosome aberrations under the microscope.

2.8 GENE BALANCE

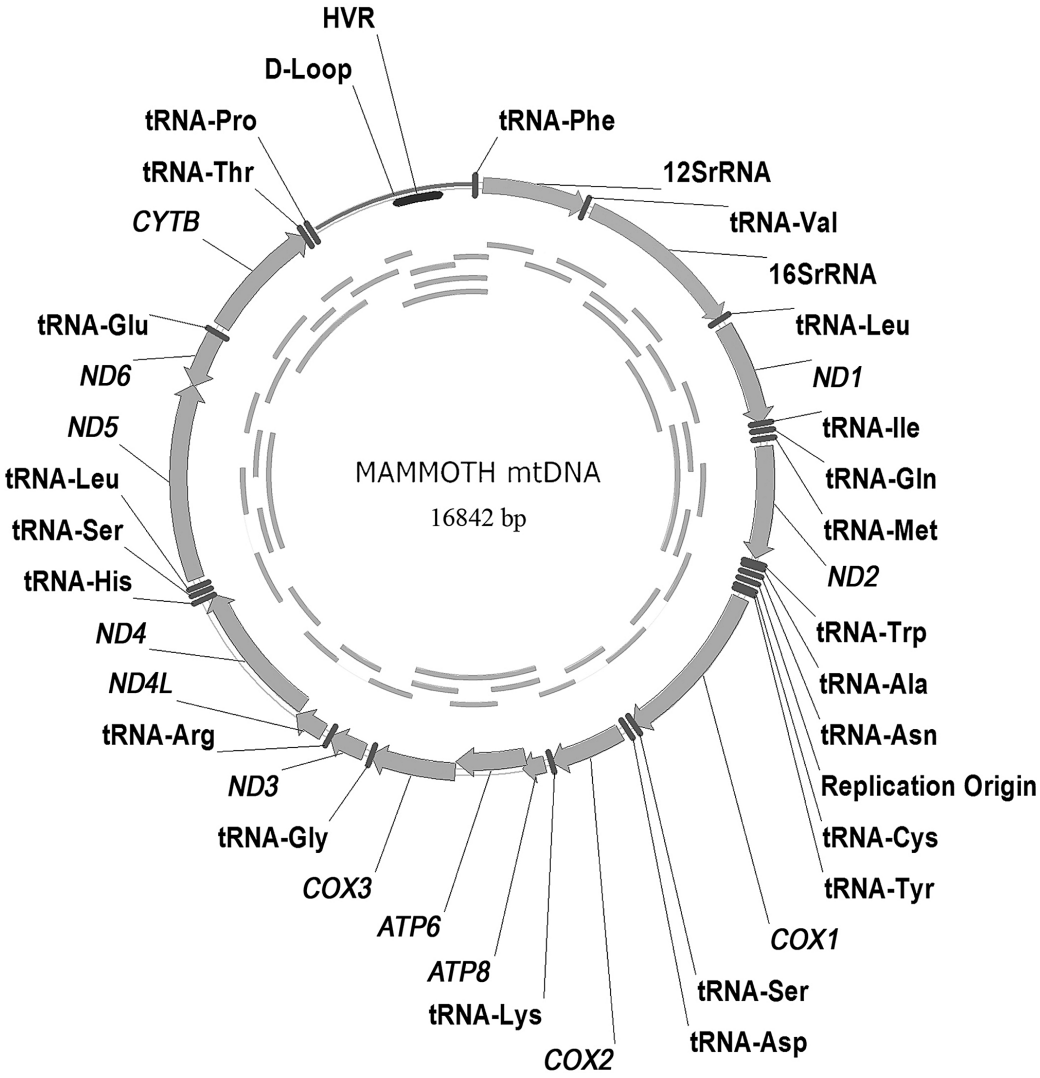
Why do trisomies, duplications, and other chromosomal abnormalities that alter gene copy number often have a negative effect on the normal development or physiology of an organism? This is particularly intriguing because in many species, aneuploidy is detrimental or lethal, while polyploidy is tolerated or even beneficial. The answer probably differs in each case, but is probably related to the concept of **gene balance**, which can be summarized as follows: genes, and the proteins they produce, have evolved to function in complex metabolic and regulatory networks. Some of these networks function best when certain enzymes and regulators are present in specific ratios to each other. Increasing or decreasing the gene copy number for just one part of the network may throw the whole network out of balance, leading to increases or decreases of certain metabolites, which may be toxic in high concentrations or which may be limiting in other important processes in the cell. The activity of genes and metabolic networks is regulated in many different ways besides changes in gene copy number, so duplication of just a few genes will usually not be harmful. However, trisomy and large segmental duplications of chromosomes affect the dosage of so many genes that **cellular networks** are unable to compensate for the changes and an abnormal or lethal phenotype results.

2.9 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 (Fig. 2.22). 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).

These organelles 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, like most bacteria that exist today. Chloroplasts and mitochondria typically have circular chromosomes that behave more like bacterial chromosomes than eukaryotic chromosomes, i.e. these organellar genomes do not undergo mitosis or meiosis.

Figure 2.22 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. The map shows the position of enzymes encoded on the chromosome including components of the NADH dehydrogenase (ND) complex and cytochrome oxidases (COX), all of which function during energy metabolism in the mitochondrion. The mitochondrial chromosome also encodes various tRNAs and rRNAs used in translation of the genes encoded on this chromosome. Other proteins required by the mitochondrion are encoded in the nuclear genome, and are translated in the cytoplasm and imported into the organelle. 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.



SUMMARY

- Chromosomes are complex and dynamic structures consisting of DNA and proteins (chromatin).
- The degree of chromatin compaction involves proteins and varies between heterochromatic and euchromatic regions and among stages of the cell cycle.
- Chromosomes can be distinguished cytologicaly based on their length, centromere position, and banding patterns when stained dyes or labeled with sequence-specific probes.
- Homologous chromosomes contain the same series of genes along their length, but not necessarily the same alleles. Sister chromatids initially contain the same alleles.
- Chromosomes are replicated by DNA polymerases and begin at an origin. Replication is bi-directional. Eukaryotes have multiple origins along each chromosome and have telomerase to replicate the ends.
- Mitosis reduces the c-number, but not the n-number. Meiosis reduces both c and n.
- Homologous chromosomes pair (sysnapse) with each other during meiosis, but not mitosis.
- Several types of structural defects in chromosomes occur naturally, and can affect cellular function and even evolution.
- Aneuploidy results from the addition or subtraction of one or more chromosomes from a group of homologs, and is usually deleterious to the cell.
- Polyploidy is the presence of more than two complete sets of chromosomes in a genome. Even-numbered multiple sets of chromosomes can be stably inherited in some species, especially plants.
- Endopolyploidy is tissue-specific type of polyploidy observed in some species, including diploids.
- Both aneuploidy and structural defects such as duplications can affect gene balance.
- Organelles also contain chromosomes, but these are much more like prokaryotic chromosomes than the nuclear chromosomes of eukaryotes.

Key Terms

chromosome
core histones
nucleosome
30nm fiber
histone H1
scaffold proteins
heterochromatin
euchromatin
satellite DNA
chromatid
centromere
metacentric
acrocentric
telocentric
holocentric
telomere
homologous
non-homologous
chromatid
sister chromatid
non-sister chromatid
interphase
mitosis
prophase
metaphase
anaphase
telophase
DNA polymerase
origin of replication
telomerase
riboprotein
Hayflick limit
HeLa cells
cytokinesis
meiosis
gametes
prophase (I, II)
metaphase (I, II)
anaphase (I, II)
telophase (I, II)
cytokinesis
meiocyte
bivalent
syanapse, pair up
synaptonemal complex
reductional division
equational division
leptotene
zygotene
pachytene
diplotene
diakinesis
crossing over
chiasma (chiasmata)
polar bodies
G_1
G_2
S
M
G_0
interphase
n
c
replicated chromosome
karyotype/karyogram
autosome
sex-chromosome
homogametic
heterogametic
aneuploidy
monsomic
trisomic
Down syndrome
deletion
duplication
insertion
inversion
translocation
non-disjunction
chromosome breakage
polyploidy
x
monoploid
sterile
tetravalent
octoploid
hexaploid
triploid
endoreduplication
endopolyploidy
salivary gland chromosome
polytene
gene balance
cellular network
chloroplast
mitochondria
endosymbiont
endosymbiont theory
organellar chromosome
mtDNA

STUDY QUESTIONS

2.1 Define chromatin. What is the difference between DNA, chromatin and chromosomes?

2.2 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?

2.3 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?

2.4 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?

2.5 For a diploid with 2n=16 chromosomes, how many chromosomes and chromatids are per cell present in the gamete, and zygote and immediately following G_1, S, G_2, mitosis, and meiosis?

2.6 Bread wheat (*Triticum aestivum*) is a hexaploid. Using the nomenclature presented in class, an ovum cell of wheat has n=21 chromosomes. How many chromosomes in a zygote of bread wheat?

2.7 For a given gene:

a) What is the maximum number of alleles that can exist in a 2n cell of a given diploid individual?

b) What is the maximum number of alleles that can exist in a 1n cell of a tetraploid individual?

c) What is the maximum number of alleles that can exist in a 2n cell of a tetraploid individual?

d) What is the maximum number of alleles that can exist in a population?

2.8 a) Why is aneuploidy more often lethal than polyploidy?

b) Which is more likely to disrupt gene balance: polyploidy or duplication?

2.9 For a diploid organism with 2n=4 chromosomes, draw a diagram of all of the possible configurations of chromosomes during normal anaphase I, with the maternally and paternally derived chromosomes labelled.

2.10 For a triploid organism with 2n=3x=6 chromosomes, draw a diagram of all of the possible configurations of chromosomes at anaphase I (it is not necessary label maternal and paternal chromosomes).

2.11 For a tetraploid organism with 2n=4x=8 chromosomes, draw all of the possible configurations of chromosomes during a normal metaphase.

2.12 A simple mnemonic for leptotene, zygotene, pachytene, diplotene, & diakinesis is **L**ame **Z**ebras **P**ee **D**own **D**rains. Make another one yourself.

Chapter 3 GENETIC ANALYSIS OF SINGLE GENES

Figure 3.1
Pea plants were used by Gregor Mendel to discover some fundamental laws of genetics.
(Flicker-Christian Guthier-CC:A)

Before Mendel, the basic rules of heredity were not understood. For example, it was known that green-seeded pea plants occasionally produced offspring that had yellow seeds; but were the hereditary factors that controlled seed color somehow changing from one generation to the next, or were certain factors disappearing and reappearing? And did the same factors that controlled seed color also control things like plant height?

3.1 MENDEL'S FIRST LAW

3.1.1 CHARACTER TRAITS EXIST IN PAIRS THAT SEGREGATE AT MEIOSIS

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.

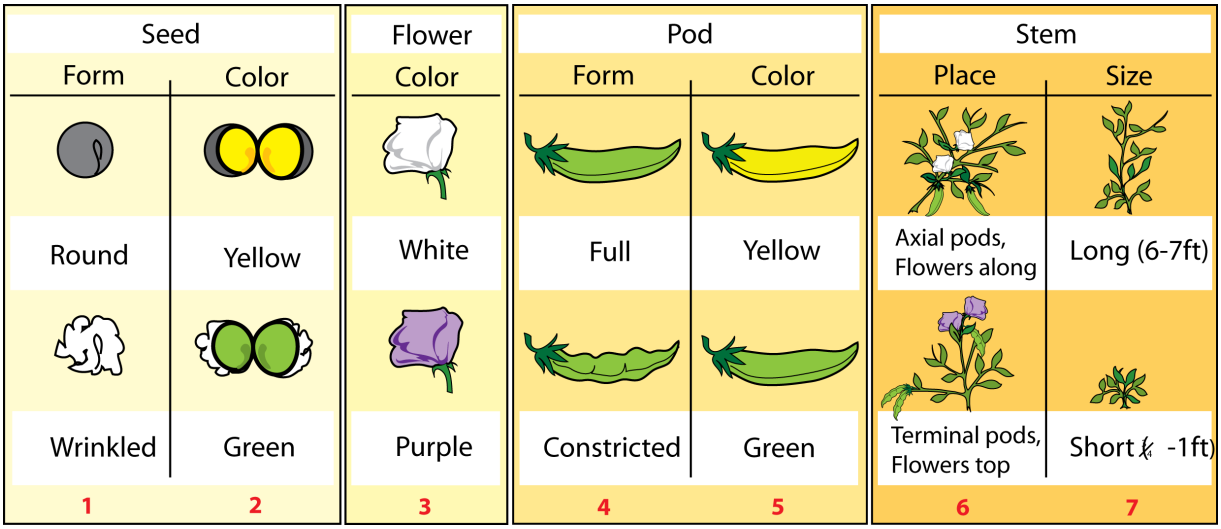


Figure 3.2
Seven traits Mendel studied in peas. (Wikipedia-Mariana Ruiz-PD)

3.1.2 Hetero-, Homo-, Hemizygosity

Mendel's First Law 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 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 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 identical, the individual is said to be **homozygous** for that gene. 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 on the homologous chromosome, we use the term **hemizygous**.

Although a typical diploid individual can have at most two different alleles of a particular gene, many more than two different 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. Such collections of mutations are known as an **allelic series**.

3.2 Relationships Between Genes, Genotypes and Phenotypes

3.2.1 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. The visible or detectable effect of these 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, then various interactions between them may influence their expression in the phenotype.

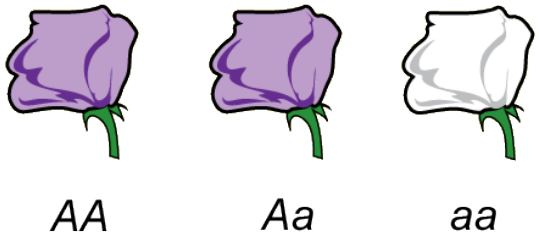


Figure 3.3
Relationship between genotype and phenotype for an allele that is completely dominant to another allele.
Original-Deholos (Fireworks)-CC:AN)

3.2.2 Complete Dominance

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 (see Figures 1.8 and 3.3). 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 experimental 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. To represent this relationship, often, a dominant allele will be represented by a capital letter (e.g. *A*) while a recessive allele will be represented in lower case (e.g. *a*). However, many different systems of genetic symbols are in use. The most common are shown in Table 3.1. Also note that genes and 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.

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, but on homologous chromosomes.

Table 3.1
Examples of symbols used to represent genes and alleles.

3.2.3 Incomplete Dominance

Besides dominance and recessivity, other relationships can exist between alleles. In **incomplete dominance** (also called **semi-dominance,** Figure 3.4), both alleles affect the trait additively, and the phenotype of the heterozygote is intermediate between either of the homozygotes. 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.

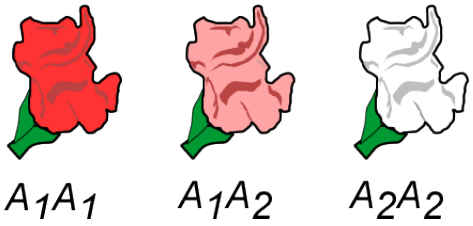


Figure 3.4
Relationship between genotype and phenotype for incompletely dominant alleles affecting petal colour in carnations.
(Original-Deholos-CC:AN)

3.2.4 Co-Dominance

Co-dominance is another type of allelic relationship, in which a heterozygous individual expresses the phenotype of both alleles simultaneously. An example of co-dominance is found within the **ABO blood group** of humans. The *ABO* gene has three common alleles which 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 3.5). Heterozygous I^AI^B individuals have both A and B antigens on their cells, and so have type AB blood. Notice that the heterozygote expresses both alleles simultaneously, and is not some kind of novel intermediate between A and B. Co-dominance is therefore distinct from incomplete dominance, although they are sometimes confused.

It is also important to note that the third allele, i, does not make either antigen and is recessive to the other alleles. I^A/i or I^B/i individuals display 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.

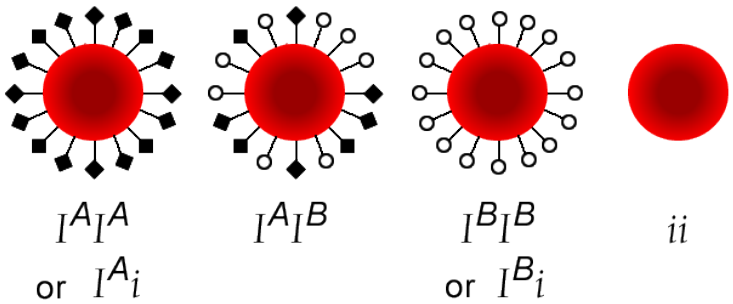


Figure 3.5
Relationship between genotype and phenotype for three alleles of the human ABO gene. The I^A and I^B alleles show co-dominance. The I^A allele is completely dominant to the i allele. The I^B allele is completely dominant to the i allele.
(Original-Deholos - CC:AN)

3.3 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 semi-dominant) 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.

3.4 Crossing Techniques Used in Classical Genetics

3.4.1 Classical Genetics

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 take off until Thomas Morgan began working with fruit flies in 1908. Later, starting with Watson and Crick's structure of DNA in 1953, classical genetics was joined by **molecular genetics**, 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.

3.4.2 True Breeding Lines

Geneticists make use of **true breeding lines** just as Mendel did (Figure 3.6a). 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 of their 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.

3.4.3 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 3.6b). Recall from chapter 1 that the generations in a cross are named P (parental), F_1 (first filial), F_2 (second filial), and so on.

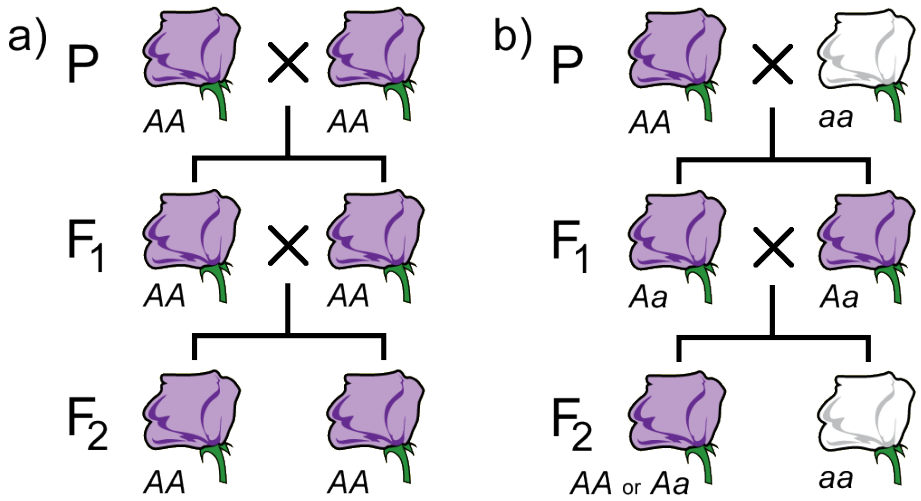


Figure 3.6
(a) A true-breeding line **(b)** A monohybrid cross produced by mating two different pure-breeding lines. (Original-Deholos-CC:AN)

	A	***a***
A	*AA*	*Aa*
a	*Aa*	*aa*

Figure 3.7 A Punnett square showing a monohybrid cross. (Original-Deholos (Fireworks)-CC:AN)

3.4.4 Punnett Squares

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. A convenient method for calculating the expected genotypic and phenotypic ratios from a cross was invented by Reginald Punnett. 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. The F_1 cross from Figure 3.6b would be drawn as in Figure 3.7. 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.

3.4.5 Test Crosses

Knowing the genotypes of an individual is usually 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 homozgyote based on phenotype alone (e.g. see the purple-flowered F_2 plants in Figure 3.6b). 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 (*AA*) or a heterozygote (*Aa*). You could cross this purple-flowered plant to a white-flowered plant as a **tester**, since you know the genotype of the tester is *aa*. Depending on the genotype of the purple-flowered parent (Figure 3.8), you will observe different phenotypic ratios in the F_1 generation. If the purple-flowered parent was a homozgyote, all of the F_1 progeny will be purple. If the purple-flowered parent was a heterozygote, the F_1 progeny should segregate purple-flowered and white-flowered plants in a 1:1 ratio.

	A	***A***
a	*Aa*	*Aa*
a	*Aa*	*Aa*

	A	***a***
a	*Aa*	*aa*
a	*Aa*	*aa*

Figure 3.8
Punnett Squares showing the two possible outcomes of a test cross. (Original-Deholos (Fireworks)-CC:AN)

3.5 Sex-Linkage: An Exception to Mendel's First Law

In the previous chapter we introduced sex chromosomes and autosomes. For loci on autosomes, the alleles follow the normal Mendelian pattern of inheritance. However, for loci on the sex chromosomes this is mostly not true, because most 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.

3.5.1 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*$^-$ (w^-) have white eyes because the red pigments are absent. Because this mutation is recessive to the wild type w^+ 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.

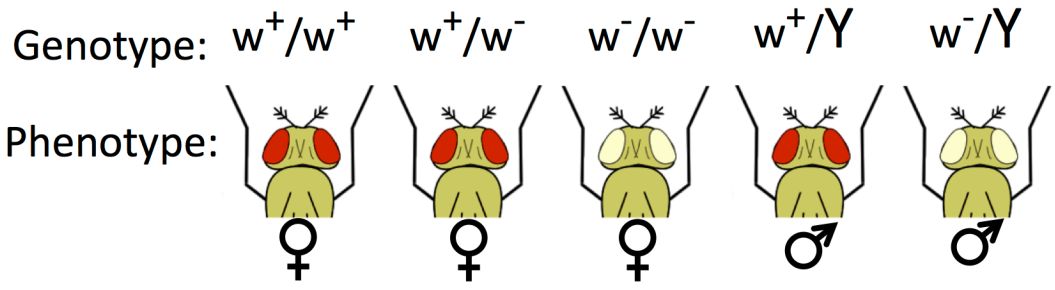


Figure 3.9
Relationship between genotype and phenotype for a the *white* gene on the X-linked gene in *Drosophila melanogaster*. The Y chromosome is indicated with a capital Y because it does not have a copy of the *white* gene. (Original-Harrington/Locke-CC:AN)

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 3.10). 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 3.10. 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 XX-XY sex chromosome system, including mammals and humans. The ZZ-ZW system is similar, but reversed (see below).

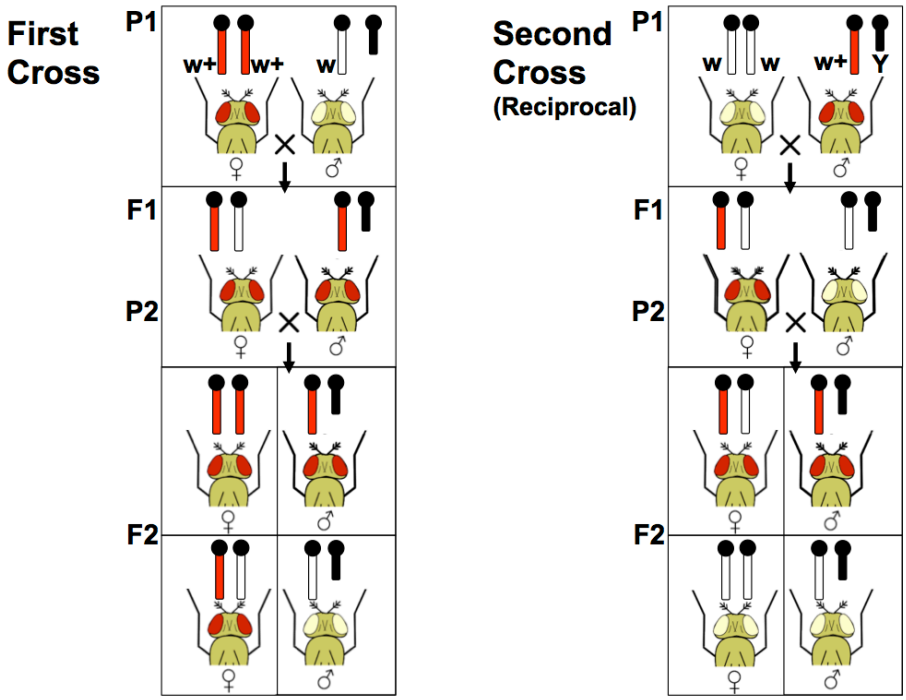


Figure 3.10 Reciprocal crosses involving an X-linked gene in *Drosophila melanogaster*. In the first cross (left) all of the offspring have red eyes. In second (reciprocal) cross (right) all of the female offspring have red eyes and the male offspring all have white eyes. If the F1 progeny are crossed (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 won the Nobel Prize for using these crosses to demonstrate that genes (such as *white*) were on chromosomes (in this case the X-chromosome). (Wikipedia-PAR-PD)

3.5.2 Sex Determination in Animals.

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

For example in humans and other mammals XY embryos develop as males while XX embryos become females. This difference in development is due to the presence of only a **single gene**, the TDF-Y gene, on the Y-chromosome. Its presence and expression dictates that the sex of the individual will be male. Its absence results in a female phenotype.

Although *Drosophila melanogaster* also has an XX-XY sex chromosomes, its sex determination system uses a different 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) will develop as females, while those with only one

X-chromosome (2A:1X) will develop as males. The presence/absence of the Y-chromosome and its genes are not significant.

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.

In other species, the environment can determine an individuals sex. In alligators (and some other reptiles) the temperature of development dictates the sex, while in many reef fish, the population sex ratio can cause some individuals to change sex.

3.5.3 Dosage Compensation for Loci on Sex Chromosomes.

Mammals and *Drosophila* both have XX - XY sex determination systems. However, because these systems evolved independently they work differently with regard to compensating for the difference in gene dosage (and sex determination – see above). 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 loosing 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. XX 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.

In Drosophila and many other insects, to make up for the males only having a single X chromosome the genes on it are expressed at twice the normal rate. This mechanism of dosage compensation restores a balance between proteins encoded by **X-linked genes** and those made by **autosomal genes**.

In mammals a different mechanism is used, called **X-chromosome inactivation**.

3.5.4 X-chromosome Inactivation in Mammals

In mammals the dosage compensation system operates in females, not males. In XX embryos one X in each cell is randomly chosen and marked for inactivation. From this point forward this chromosome will be inactive, hence its name $X_{inactive}$ (X_i). The other X chromosome, the X_{active} (X_a), is unaffected. The X_i is replicated during S phase and transmitted during mitosis the same as any other chromosome but most of its genes are never allowed to turn on. The chromosome appears as a condensed mass within interphase nuclei called the Barr body. With the inactivation of genes on one X-chromosome, females have the same number of functioning X-linked genes as males.

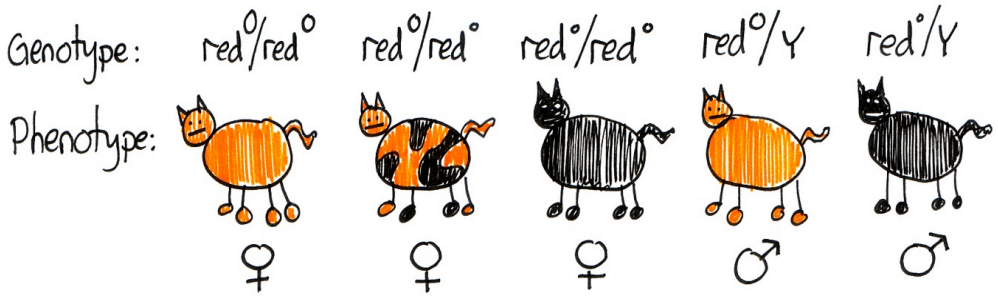


Figure 3.11
Relationship between genotype and phenotype for an X-linked gene in cats. The 'red capital-O' allele = orange while the 'red lower-case-O' allele = black. (Original-Harrington-CC:AN)

This random inactivation of one X-chromosome leads to a commonly observe phenomenon in cats. A familiar X-linked gene is the *red* gene in cats. The red^{O} allele encodes an enzyme that results in orange pigment for the hair. The red^{o} allele causes the hairs to be black. The phenotypes of various genotypes of cats are shown in Figure 3.11. 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 chromosome carrying the red^{o} allele is inactivated. The red^{O} allele on the X_a is functional and orange pigments are made. In black hairs the reverse is true, the X chromosome with the red^{O} allele is inactive and the X chromosome with the red^{o} 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.

The *red* gene in cats is a good demonstration of how the mammalian dosage compensation systems affects gene expression. However, most X-linked genes do not produce such dramatic mosaic phenotypes in heterozygous females. A more typical example 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 the result is hemophilia type A. Females homozygous for mutant alleles will also have hemophilia. 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 3.12). Because some of their liver cells are exporting Factor VIII proteins into the blood stream they have the ability to form blood clots throughout their bodies. The genetic mosaicism in the cells of their bodies does not produce a visible mosaic phenotype.

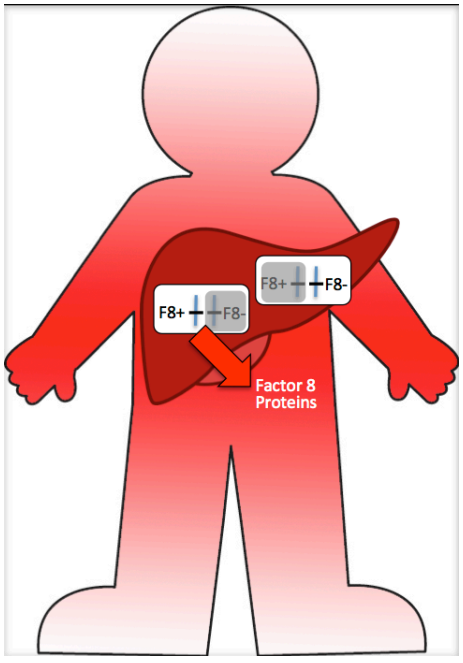


Figure 3.12
This figure shows the two types of liver cells in females heterozygous for an *F8* mutation. Because people with the $F8^{+}/F8^{-}$ genotype have the same phenotype, normal blood clotting, as $F8^{+}/F8^{+}$ people the $F8^{-}$ mutation is classified as recessive. . (Original-Harrington/Locke-CC:AN)

3.5.5 Other Sex-Linked Genes – Z-linked genes

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 3.13). 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.

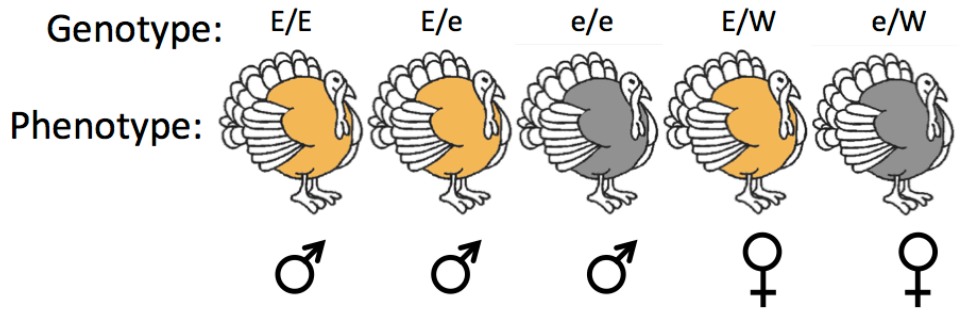


Figure 3.13 Relationship between genotype and phenotype for a Z-linked gene in turkeys. The W chromosome does not have an *E/e*-gene so it is just indicated with a capital W. (Original-Harrington/Locke-CC:AN)

3.6 Phenotypes May Not Be As Expected from the Genotype

3.6.1 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 phenotypes are not determined entirely by genotype alone. They are instead determined by an interaction between genotype and non-genetic, environmental factors and can be conceptualized in the following relationship:

Genotype + Environment $\Rightarrow$ Phenotype (G + E $\Rightarrow$ P)

Or:

Genotype + Environment + Interaction$_{GE}$ $\Rightarrow$ Phenotype (G + E + I$_{GE}$ $\Rightarrow$ P)

(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 pre-dispose an individual to cancer, but cancer may only develop if the individual is exposed to certain DNA-damaging chemicals. Therefore, not all individuals with the particular genotype will develop the cancer phenotype.

3.6.2 Penetrance and Expressivity

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

Penetrance is the proportion of individuals (usually expressed as a percentage) with a particular genotype that display a corresponding phenotype (Figure 3.14). Because all pea plants that are homozygous for the allele for white flowers (e.g. *aa* in Figure 3.3) actually have white flowers, this genotype is completely 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.

Expressivity describes the variability in mutant phenotypes observed in individuals with a particular phenotype (Figure 3.14). 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).

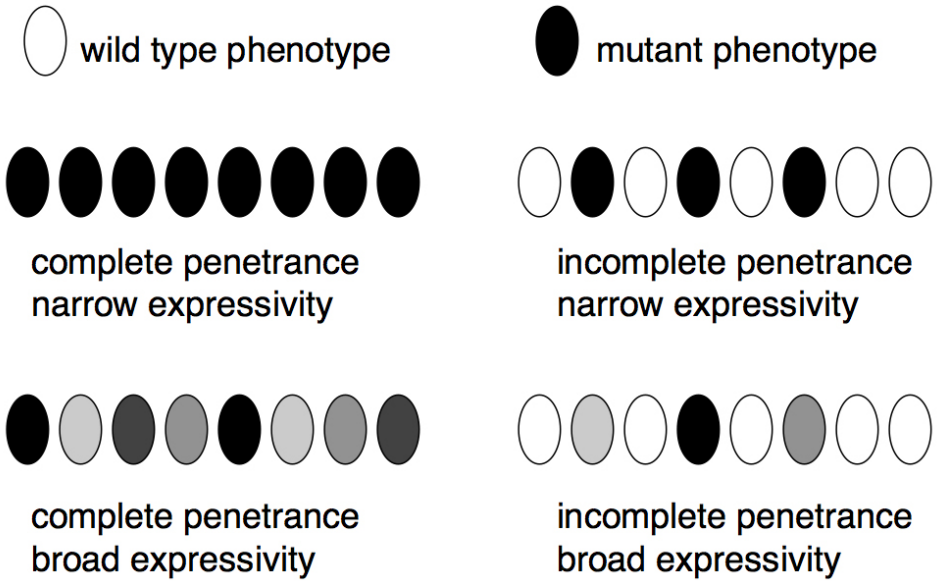


Figure 3.14
Relationship between penetrance and expressivity in eight individuals that all have a mutant genotype. Penetrance can be complete (all eight have the mutant phenotype) or incomplete (only some have the mutant phenotype). Amongst those individuals with the mutant phenotype the expressivity can be narrow (very little variation) to broad (lots of variation). (Original-Locke-CC:AN)

3.7 Phenotypic Ratios May Not Be As Expected

3.7.1 Explanation

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. On the other hand, it may be because certain genotypes have a less than 100% survival rate. For example, *Drosophila* crosses sometimes give unexpected results because the more mutant alleles a zygote has the less likely it is to survive to become an adult. Genotypes that cause death for embryos or larvae are underrepresented when adult flies are counted.

3.7.2 The x^2 Test For Goodness-of-fit

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 covered in the labs.

SUMMARY

- A diploid can have up to two different alleles at a single locus. The alleles segregate equally between gametes during meiosis.
- 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.
- Sex-linked genes are an exception to standard Mendelian inheritance. Their phenotypes are influenced by the type of sex chromosome system and the type of dosage compensation system found in the species.

KEY TERMS

allele
Mendel's First Law
Law of Equal Segregation
homozygous
heterozygous
hemizygous
wild-type
variant
locus
genotype
phenotype
dominant
recessive
complete dominance
incomplete (semi) dominance
co-dominance
ABO blood group
haplosufficiency
haploinsufficiency
classical genetics
molecular genetics
true breeding lines
monohybrid cross
Punnett Square
test cross
tester
sex-linked
dosage compensation
X-linked genes
autosomal genes
reciprocal cross
Z-linked genes
G + E = P
penetrance
expressivity
sampling effects
chi-square χ^2 test

QUESTIONS

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

3.2 Wirey hair (*W*) is dominant to smooth hair (*w*) in dogs.
a) If you cross a homozygous, wirey-haired dog with a smooth-haired dog, what will be the genotype and phenotype of the F_1 generation?

b) If two dogs from the F_1 generation mated, what would be the most likely ratio of hair phenotypes among their progeny?

c) When two wirey-haired *Ww* dogs actually mated, they had alitter of three puppies, which all had smooth hair. How do you explain this observation?

d) Someone left a wirey-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?

3.3 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?

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

3.5 If your blood type is B, what are the possible genotypes of your parents at the locus that controls the ABO blood types?

3.6 In the table below, 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.

3.7 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?

3.8 Make Punnett Squares to accompany the crosses shown in Figure 3.10.

3.9 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)

3.10 Draw reciprocal crosses that would demonstrate that the turkey E-gene is on the Z chromosome.
3.11 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?

3.12 What is the relationship between the red^{0} and red^{o} alleles of the *red* gene in cats?

3.13 Make a diagram similar to those in Figures 3.9, 3.11, and 3.13 that shows the relationship between genotype and phenotype for the *F8* gene in humans.

Table for Question 3.6

	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

Notes

Chapter 4 MUTATION AND VARIATION

Figure 4.1
The difference in appearance between pigmented and white peacocks is due to mutation. (Flickr-ecstaticist-CCANS)

The techniques of genetic analysis discussed in the previous chapters depend on the availability of two or more alleles for a gene of interest. Where do these alleles come from? The short answer is **mutation**.

Humans have an interesting relationship with mutation. From our perspective, mutations can be extraordinarily useful, since mutations are need for evolution to occur. Mutation is also essential for the domestication and improvement of almost all of our food. On the other hand, mutations are the cause of many cancers and other diseases, and can be devastating to individuals. Yet, the vast majority of mutations probably go undetected. In this section, we will examine some of the causes and effects of mutations.

4.1 MUTATION AND POLYMORPHISM

We have previously noted that an important property of DNA is its fidelity: most of the time it accurately passes the same information from one generation to the next. However, DNA sequences can also change. Changes in DNA sequences are called **mutations**. If a mutation changes the phenotype of an individual, the individual is said to be a **mutant**. Naturally occurring, but rare, sequence variants that are clearly different from a normal, wild-type sequence are also called mutations. On the other hand, many naturally occurring variants exist for traits for which no clearly normal type can be defined; thus, we use the term **polymorphism** to refer to variants of DNA sequences (and other phenotypes) that co-exist in a population at

relatively high frequencies (>1%). Polymorphisms and mutations arise through similar biochemical processes, but the use of the word "polymorphism" avoids implying that any particular allele is more normal or abnormal. For example, a change in a person's DNA sequence that leads to a disease such as cancer is appropriately called a mutation, but a difference in DNA sequence that explains whether a person has red hair rather than brown or black hair is an example of polymorphism. Molecular markers, which we will discuss in Chapter 9, are a particularly useful type of polymorphism for some areas of genetics research.

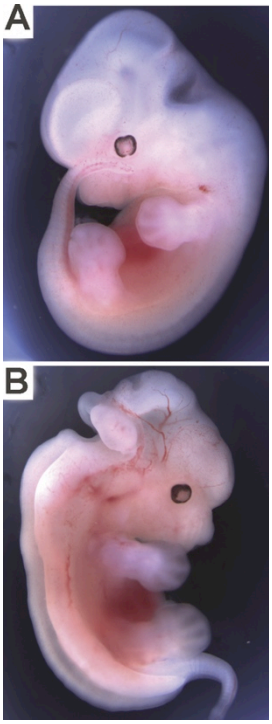


Figure 4.2 Examples of wild-type (A) and mutant (B) mouse embryos observed while screening for genes affecting cranium development.(PLoS Biology- Zarbalis, K. et al (2004) PLoS Biology-PD)

4.2 ORIGINS OF MUTATIONS

4.2.1 TYPES OF MUTATIONS

Mutations may involve the loss (**deletion**), gain (**insertion**) of one or more base pairs, or else the **substitution** of one or more base pairs with another DNA sequence of equal length. These changes in DNA sequence can arise in many ways, some of which are **spontaneous** and due to natural processes, while others are **induced** by humans intentionally (or unintentionally) using mutagens. There are many ways to classify **mutagens**, which are the agents or processes that cause mutation or increase the frequency of mutations. We will classify mutagens here as being (1) biological, (2) chemical, or (3) physical.

4.2.2 MUTATIONS OF BIOLOGICAL ORIGIN

A major source of spontaneous mutation is errors that arise during DNA replication. DNA polymerases are usually very accurate in adding a base to the growing strand that is the exact complement of the base on the template strand. However, occasionally, an incorrect base is inserted. Usually, the machinery of DNA replication will recognize and repair mispaired bases, but nevertheless, some errors become permanently incorporated in a daughter strand, and so become mutations that will be inherited by the cell's descendents (Figure 4.3).

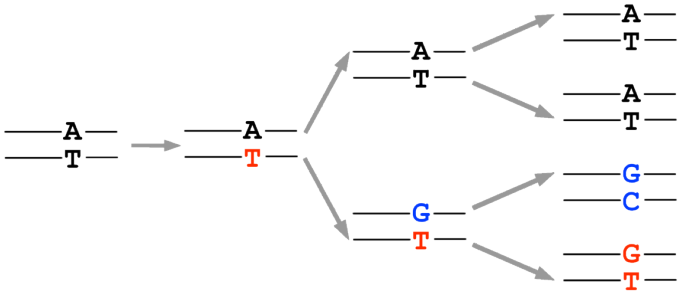


Figure 4.3 Mispairing of bases (e.g. G with T) can occur due to tautomerism, alkylating agents, or other effects. As a result, in this example the AT base pair in the original DNA strand will become permanently substituted by a GC based pair in some progeny. The mispaired GT basepair will likely be repaired or eliminated before further rounds of replication. (Original-Deyholos-CC:AN)

Another type of error introduced during replication is caused by a rare, temporary misalignment of a few bases between the template strand and daughter strand (Figure 4.4). This **strand-slippage** causes one or more bases on either strand to be temporarily displaced in a **loop** that is not paired with the opposite strand. If this loop forms on the template strand, the bases in the loop may not be replicated, and a **deletion** will be introduced in the growing daughter strand. Conversely, if a region of the daughter strand that has just been replicated becomes displaced in a loop, this

region may be replicated again, leading to an **insertion** of additional sequence in the daughter strand, as compared to the template strand.

Consequences: Regions of DNA that have several repeats of the same few nucleotides in a row are especially prone to this type of error during replication. Thus regions with **short-sequence repeats** (**SSRs**) are tend to be highly polymorphic, and are therefore particularly useful in genetics. They are called **microsatellites**.

Mutations can also be caused by the insertion of viruses, transposable elements (transposons), see below, and other types of DNA that are naturally added at more or less random positions in chromosomes. The insertion may disrupt the coding or regulatory sequence of a gene, including the fusion of part of one gene with another. These insertions can occur spontaneously, or they may also be intentionally stimulated in the laboratory as a method of mutagenesis called **transposon-tagging**. For example, a type of transposable element called a ***P* element** is widely used in Drosophila as a biological mutagen. **T-DNA**, which is an insertional element modified from a bacterial pathogen, is used as a mutagen in some plant species.

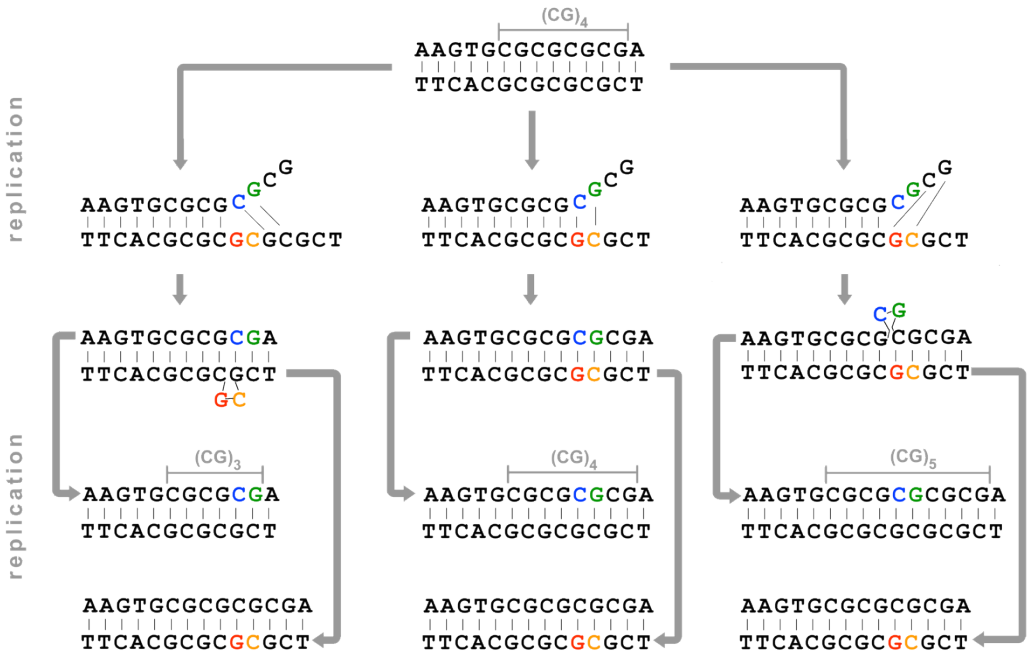


Figure 4.4 Strand-slippage can occur occasionally during replication, especially in regions with short, repeated sequences. This can lead to either deletion (left) or insertion (right) of sequences compared to the products of normal replication (center), depending on whether the template strand or daughter strand "loops-out" during replication. (Original-Deyholos-CC:AN)

4.2.3 Transposable Elements

Transposable elements (TEs) are also known as **mobile genetic elements**, or more informally as **jumping genes**. They are present throughout the chromosomes of almost all organisms. These DNA sequences have a unique ability to be cut or

copied from their original location and inserted into new locations in the genome. This is called transposition. These insert locations are not entirely random, but TEs can, in principle, be inserted into almost any region of the genome. TEs can therefore insert into genes, disrupting its function and causing a mutation. Researchers have developed methods of artificially increasing the rate of transposition, which makes some TEs a useful type of mutagen. However, the biological importance of TEs extends far beyond their use in mutant screening. TEs are also important causes of disease and phenotypic instability, and they are a major mutational force in evolution.

There are two major classes of TEs in eukaryotes (Figure 4-5). **Class I** elements include **retrotransposons**; these transpose by means of an RNA intermediate. The TE transcript is reverse transcribed into DNA before being inserted elsewhere in the genome through the action of enzymes such as **integrase**. **Class II** elements are known also as **transposons**. They do not use reverse transcriptase or an RNA intermediate for transposition. Instead, they use an enzyme called **transposase** to cut DNA from the original location and then this excised dsDNA fragment is inserted into a new location. Note that the name transposon is sometimes used incorrectly to refer to any type of TEs, but in this book we use transposon to refer specifically to Class II elements.

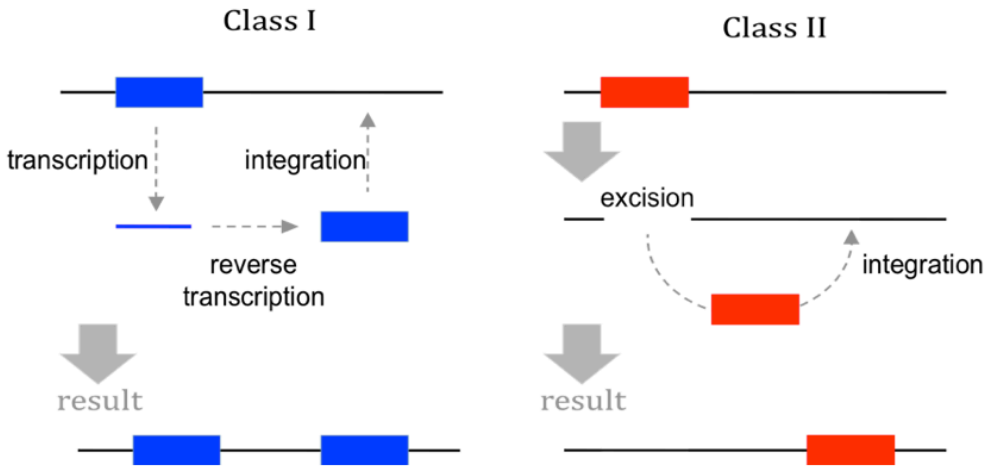


Figure 4-5 . Diagrams of the two main types of transposable elements. (TEs) Class I elements transpose via an ssRNA intermediate, which is reverse transcribed to dsDNA prior to insertion of this copy in a new site in the genome. Class II elements do not involve an RNA intermediate; most Class II elements are cut from their original location as dsDNA, prior to being inserted into a new site in the genome. Although the diagram shows TEs being inserted on the same chromosome as they originated from, TEs can also move to other chromosomes within the same cell.

TEs are relatively short DNA sequences (100-10,000 bp), and encode no more than a few proteins (if any). Normally, the protein-coding genes within a TE are all related to the TE's own transposition functions. These proteins may include **reverse transcriptase**, transposase, and integrase. However, some TEs (of either Class I or II) do not encode any proteins at all. These **non-autonomous** TEs can only transpose if they are supplied with enzymes produced by other, **autonomous** TEs located elsewhere in the genome. In all cases, enzymes for transposition

recognize conserved nucleotide sequences within the TE, which dictate where the enzymes begin cutting or copying.

The human genome consists of nearly 45% TEs, the vast majority of which are families of Class I elements called **LINEs** and **SINEs**. The short, ***Alu*** type of SINE occurs in more than one million copies in the human genome (compare this to the approximately 21,000, non-TE, protein-coding genes in humans). Indeed, TEs make up a significant portion of the genomes of almost all eukaryotes. Class I elements, which usually transpose via an RNA **copy-and-paste** mechanism, tend to be more abundant than Class II elements, which mostly use a **cut-and-paste** mechanism. But even the cut-paste mechanism can lead to an increase in TE copy number. For example, if the site vacated by an excised transposon is repaired with a DNA template from a homologous chromosome that itself contains a copy of a transposon, then the total number of transposons in the genome will increase.

Besides greatly expanding the overall DNA content of genomes, TEs contribute to genome evolution in many other ways. As already mentioned, they may disrupt gene function by insertion into a gene's coding region or regulatory region. More interestingly adjacent regions of chromosomal DNA are sometimes mistakenly transposed along with the TE; this can lead to gene duplication. The duplicated genes are then free to evolve independently, leading in some cases to the development of new functions. The breakage of strands by TE excision and integration can disrupt genes, and can lead to chromosome rearrangement or deletion if errors are made during strand rejoining. Furthermore, having so many similar TE sequences distributed throughout a chromosome sometimes allows mispairing of regions of homologous chromosomes at meiosis, which can cause unequal crossing-over, resulting in deletion or duplication of large segments of chromosomes. Thus, TEs are a potentially important evolutionary force, and may not be included as merely "junk DNA", as they once were.

4.2.4 CHEMICAL MUTATGENS

Many chemical compounds, whether natural or synthetic, can react with DNA to cause mutations. In some of these reactions the chemical structure of particular bases may change, so that they are misread during replication. In others the chemical mutagens distort the double helix causing it to be replicated inaccurately, while still others may cause breaks in chromosomes that lead to deletions and other types of aberrations. The following are examples of two classes of chemical mutagens: that are important in genetics and medicine: **alkylating agents**, and **intercalating agents**.

Ethane methyl sulfonate (**EMS**) is an example of an alkylating agent that is commonly used by geneticists to induce mutations in a wide range of both prokaryotes and eukaryotes. The organism is fed or otherwise exposed to a solution of EMS. It reacts with some of the G bases it encounters in a process called alkylation, where the addition of an alkyl group to G changes the base pairing properties so that the next time the alkylated DNA strand is replicated, a T instead of a C will be inserted opposite to the alkylated-G in the daughter strand (Figure 4.6). The new strand therefore bears a mutation (base substitution), which will be inherited in all the strands that are subsequently replicated from it.

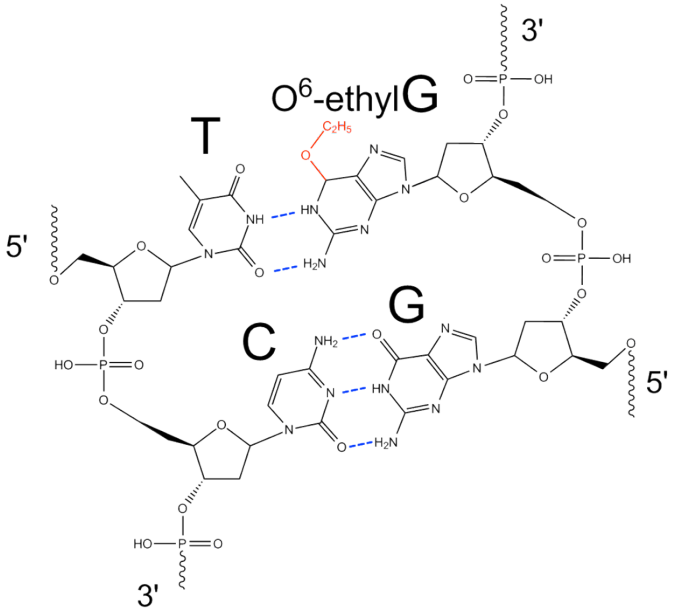


Figure 4.6
Alkylation of G (shown in red) allows G to bond with T, rather than with C. (Original-Deyholos-CC:AN)

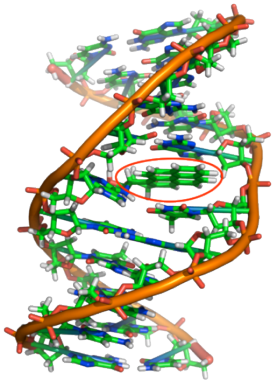


Figure 4.7
Benzo[*a*]pyrene (circled in red) is an example of an intercalating agent.(Wikipedia-Zephyris-GFDL)

Intercalating agents are another type of chemical mutagen. They tend to be flat, planar molecules like **benzo[*a*]pyrene**, a component of wood and tobacco smoke, and induce mutations by inserting between the stacked bases at the center of the DNA helix (Figure 4.7). This intercalation distorts the shape of the DNA helix, which can cause the wrong bases to be added to a growing DNA strand during DNA synthesis.

There are a large number of chemicals that act as intercalating agents, can mutate DNA, and are **carcinogenic** (can cause cancer). Many of these are also used to treat cancer, as they preferentially kill actively dividing cells. Another important intercalating agent is **ethidium bromide**, the dye that fluorescently stains DNA in laboratory assays. For this reason, molecular biologists are trained to handle this chemical carefully.

4.2.4 Mutations of Physical Origin

Anything that damages DNA by transferring energy to it can be considered a physical mutagen. Usually this involves radioactive particles, x-rays, or UV light. The smaller, fast moving particles may cause base substitutions or delete a single bases, while larger, slightly slower particles may induce larger deletions by breaking the double stranded helix of the chromosome. Physical mutagens can also create unusual structures in DNA, such as the thymine dimers formed by UV light (Figure 4.8). Thymine dimers disrupt normal base-pairing in the double helix, and may block replication altogether if not repaired by the cell's DNA repair enzymes.

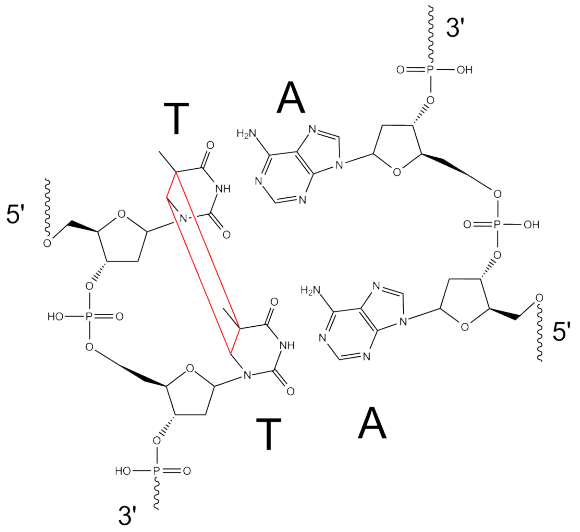


Figure 4.8 Thymine dimers are formed when adjacent thymine bases on the same DNA strand become covalently linked (red bonds) follow exposure to mutagens such as UV light. The dimers distort base pairing and can interrupt processes such as replication. (Source?)

4.3 Genetic screening for mutations: Forward genetics

One way to identify genes that affect a particular biological process is to induce random mutations in a large population, and then look for mutants with phenotypes that might be caused by a disruption of a particular biochemical pathway. This strategy of **mutant screening** has been used very effectively to identify and understand the molecular components of hundreds of different biological processes. For example, to find the basic biological processes of memory and learning, researchers have screened mutagenized populations of *Drosophila* to recover flies (or larvae) that lack the normal ability to learn. They lack the ability to associate a particular odor with an electric shock. Because of the similarity of biology among all organisms, some of the genes identified by this mutant screen of a model organism may be relevant to learning and memory in humans, including conditions such as Alzheimer's disease.

4.3.1 – Genetic Screens

In a typical mutant screen, researchers treat a parental population with a mutagen. This may involve soaking seeds in EMS, or mixing a mutagen with the food fed to flies. Usually, no phenotypes are visible among the individuals that are directly exposed to the mutagen because in all the cells every strand of DNA will be affected independently. Thus the induced mutations will be heterozygous and limited to single cells. However, what is most important to geneticists are the mutations in the germline of the mutagenized individuals. The **germline** is defined as the gametes and any of their developmental precursors, and is therefore distinct from the **somatic cells** (i.e. non-reproductive cells) of the body. Because most induced mutations are recessive, the progeny of mutagenized individuals must be mated in a way that allows the new mutations to become homozygous (or hemizygous). Strategies for doing this vary between organisms. In any case, the generation in which induced mutations are expected to occur can be examined for the presence of novel phenotypes. Once a relevant mutant has been identified, geneticists can begin to make inferences about what the normal function of the mutated gene is, based on its mutant phenotype. This can then be investigated further with molecular genetic techniques.

Exposure of an organism to a mutagen causes mutations in essentially random positions along the chromosomes. Most of the mutant phenotypes recovered from a genetic screen are caused by **loss-of-function** mutations. These alleles are due to changes in the DNA sequence that cause it to no longer produce the same level of active protein as the wild-type allele. Loss-of-function alleles tend to be recessive because the wild-type allele is haplosufficient (see Chapter 3). 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 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.

4.4 TYPES OF MUTATIONS

4.4.1 – MULLER'S MORPHS

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. Any mutant allele can be classified into one of five types: (1) amorph, (2) hypomorph, (3) hypermorph, (4) neomorph, and (5) antimorph.

Amorph alleles are complete loss-of-function. They make no active product – zero function. The absence of function can be due to a lack of transcription (gene regulation mutation) or due to the production of a malfunctioning (protein coding mutation) product. These are also sometimes referred to as a **Null** allele.

Hypomorph alleles are only a partial loss-of-function. They make an incompletely functioning product. This could occur via reduced transcription or via the production of a product that lacks complete activity. These alleles are sometimes referred to as **Leaky** mutations, because they provide some function, but not complete function.

Both amorphs and hypomorphs tend to be recessive to wild type because the wild type allele is usually able to supply sufficient product to produce a wild type phenotype (called haplo-sufficient - see Chapter 6). If the mutant allele is not haplo-sufficient, then it will be dominant to the wild type.

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.

Hypermorph alleles produce more of the same, active product. This can occur via increased transcription or by changing the product to make it more efficient/effective at its function.

Neomorph alleles produce an active product with a new, different function, something that the wild type allele doesn't do. It can be either new expression (new tissue or time) or a mutation in the product to create a new function (additional substrate or new binding site), not present in the wild type product.

Antimorph alleles are relatively rare, and have an activity that is dominant and opposite to the wild-type function. These alleles usually have no normal function of their own and they interfere with the function from the wild type allele. Thus, when an antimorph allele is heterozygous with wild type, the wild type allele function is reduced. While at the molecular level there are many ways this can happen, the simplest model to explain antimorph effect is that the product acts as a dimer (or any multimer) and one mutant subunit poisons the whole complex. Antimorphs are also known as dominant negative mutations.

Identifying Muller's Morphs - All mutations can be sorted into one of the five morphs base on how they behave when heterozygous with other alleles – deletion alleles (zero function), wild type alleles (normal function), and duplication alleles (double normal function).

4.5 SOME MUTATIONS MAY NOT HAVE DETECTABLE PHENOTYPES

4.5.1 SILENT CHANGES

After mutagen treatment, the vast majority of base pair changes (especially substitutions) have no effect on the phenotype. Often, this is because the change occurs in the DNA sequence of a non-coding region of the DNA, such as in **inter-genic regions** (between genes) or within an intron region. Also, even if the change occurs in a base within a codon, it may not change the amino acid that it encodes (recall that the genetic code is degenerate; for example, GCT, GCC, GCA, and GCG all encode alanine) and is referred to as a **silent** mutation. Additionally, the base substitution may change an amino acid, but this doesn't alter the function of the product, so no phenotypic change would occur.

4.5.2 ENVIRONMENT AND GENETIC REDUNDANCY

There are also situations where a mutation can cause a complete loss-of-function of a gene, yet not produce a change in the phenotype, even when the mutant allele is homozygous. The lack of a phenotypic change can be due to environmental effects: the loss of that gene product may not be apparent in that environment, but might in another. Alternatively, the lack of a phenotype might be attributed to genetic **redundancy**, i.e. the encoding of similarly functioning genes at more than one locus in the genome. Thus the loss of one gene is compensated by another. This important limitation of mutational analysis should be remembered: genes with redundant functions cannot be easily identified by mutant screening.

4.5.3 ESSENTIAL GENES AND LETHAL ALLELES

Some phenotypes require individuals to reach a particular developmental stage before they can be scored. For example, flower color can only be scored in plants that are mature enough to make flowers, and eye color can only be scored in flies that have developed eyes. However, some alleles may not develop sufficiently to be included among the progeny that are scored for a particular phenotype. Mutations in **essential genes** create **recessive lethal alleles** that arrest the development of an individual at an embryonic stage. This type of mutation may therefore go unnoticed in a typical mutant screen because they are absent from the progeny being screened.

Furthermore, the progeny of a monohybrid cross involving an embryonic lethal recessive allele may therefore all be of a single phenotypic class, giving a phenotypic ratio of 1:0 (which is the same as 3:0). In this case the mutation may not be detected.

4.5.4 Naming Genes

Many genes have been first identified in mutant screens, and so they tend to be named after their mutant phenotypes, not the normal function or phenotype. This can cause some confusion for students of genetics. For example, we have already encountered an X-linked gene named *white* in fruit flies. Null mutants of the *white* gene have white eyes, but the normal *white*$^{+}$ allele has red eyes. This tells us that the wild type (normal) function of this gene is actually to help make red eyes. Its product is a protein that imports a pigment precursor into developing cells of the eye. Why don't we call it the "red" gene, since that is what its product does? Because there are more than one-dozen genes that when mutant alter the eye colour; e.g. *violet, cinnabar, brown, scarlet*, etc. For all these genes their function is also needed to make the eye wild type red and not the mutant colour. If we used the name "red" for all these genes it would be confusing, so we use the distinctive mutant phenotype as the gene name. However, this can be problematic, as with the "lethal" mutations described above. This problem is usually handled by giving numbers or locations to the gene name, or making up names that describe how they die (e.g. *even-skipped, hunchback, hairy, runt*, etc.) .

4.6 Complementation tests and Allelism

4.6.1 – One, or more than one gene?

As explained earlier in this chapter, mutant screening is one of the beginning steps geneticists use to investigate biological processes. When geneticists obtain two independently derived mutants (either from natural populations or during a mutant screen) with similar phenotypes, an immediate question is whether or not the mutant phenotype is due to a loss of function in the same gene, or are they mutant in different genes that both affect the same phenotype (e.g., in the same pathway). That is, are they **allelic mutations**, or **non-allelic mutations**, respectively? This question can be resolved using **complementation tests**, which bring together, or combine, the two mutations under consideration into the same organism to assess the combined phenotype.

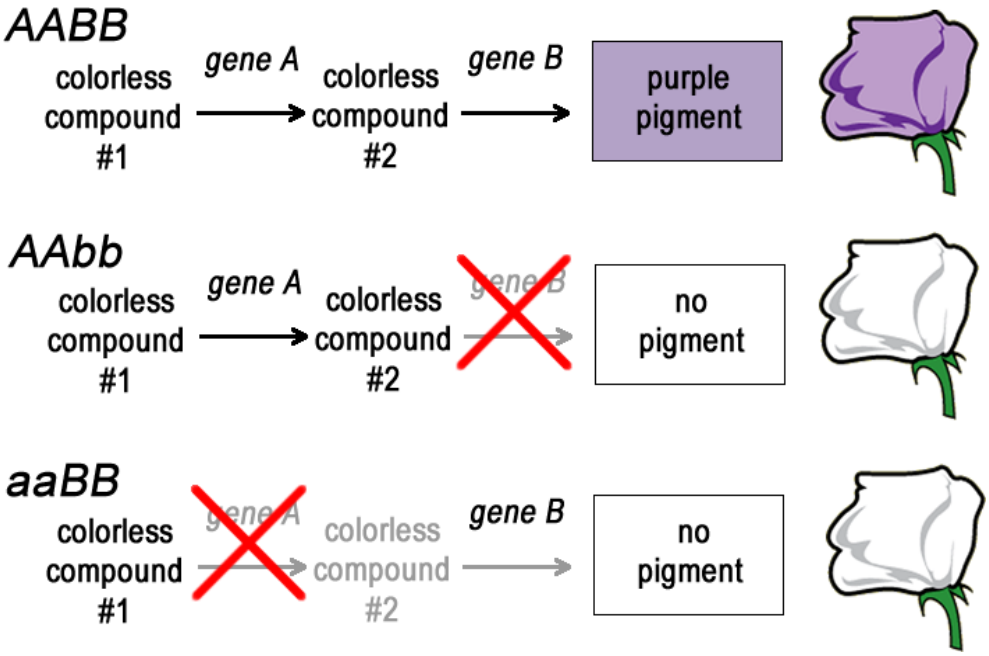


Figure 4.9 In this simplified biochemical pathway, two enzymes encoded by two different genes modify chemical compounds in two sequential reactions to produce a purple pigment. Loss of either of the enzymes disrupts the pathway and no pigment is produced. (Original-Deyholos-CC:AN)

4.6.2 - A HYPOTHETICAL EXAMPLE OF PURPLE FLOWERS

The easiest way to understand a complementation test is by example (Fig.4.9). The pigment in a purple flower could depend on a biochemical pathway much like the **biochemical pathways** leading to the production of arginine in Neurospora (review in Chapter 1). A plant that lacks the function of gene A (genotype *aa*) would produce mutant, white flowers that looked just like the flowers of a plant that lacked the function of gene B (genotype *bb*). (The genetics of two loci are discussed more in the following chapters.) Both A and B are enzymes in the same pathway that leads from a colorless compound#1, thorough colorless compound#2, to the purple pigment. Blocks at either step will result in a mutant white, not wild type purple, flower.

Strains with mutations in gene A can be represented as the genotype *aa*, while strains with mutations in gene B can be represented as *bb*. Given that there are two genes here, A and B, then each of these mutant strains can be more completely represented as *aaBB* and *AAbb* . (LEARNING NOTE: Student often forget that genotypes usually only show mutant loci, however, one must remember all the other genes are assumed to be wild type.)

If these two strains are crossed together the resulting progeny will all be *AaBb*. They will have both a wild type, functional A gene and B gene and will thus have a pigmented, purple flower, a wild type phenotype. This is an example of **complementation**. Together, each strain provides what the other is lacking (*AaBb*). The mutations are in different genes and are thus called non-allelic mutations.

Now, if we are presented with a third pure-breeding, independently derived white-flower mutant strain, we won't initially know if it is mutant in gene A or gene B (or possibly some other gene altogether). We can use complementation testing to determine which gene is mutated. To perform a complementation test, two homozygous individuals with similar mutant phenotypes are crossed (Figure 4.10).

If the F1 progeny all have the same mutant phenotype (Case 1 - Figure 4.10A), then we infer that the same gene is mutated in each parent. These mutations would then be called allelic mutations - in the same gene locus. These mutations FAIL to COMPLEMENT one another (still mutant). These could be either the exact same mutant alleles, or different mutations in the same gene (allelic).

Conversely, if the F1 progeny all appear to be wild-type (Case 2 - Figure 4.10B), then each of the parents most likely carries a mutation in a different gene. These mutations would then be called non-allelic mutations - in a different gene locus. These mutations do COMPLEMENT one another.

Note: For mutations to be used in complementation tests they are (1) usually true-breeding (homozygous at the mutant locus), and (2) must be recessive mutations. Dominant mutation CANNOT be used in complementation tests. Also, remember, some mutant strains may have more than one gene locus mutated and thus would fail to complement mutants from more than one other locus (or group).

A.

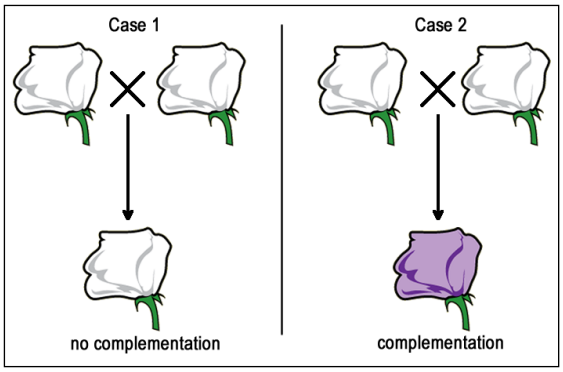


Figure 4.10A – Observation:
In a typical complementation test, the genotypes of two parents are unknown (although they must be pure breeding, homozygous mutants). If the F1 progeny all have a mutant phenotype (Case 1), there is no complementation. If the F1 progeny are all wild-type, the mutations have successfully complemented each other.

B.

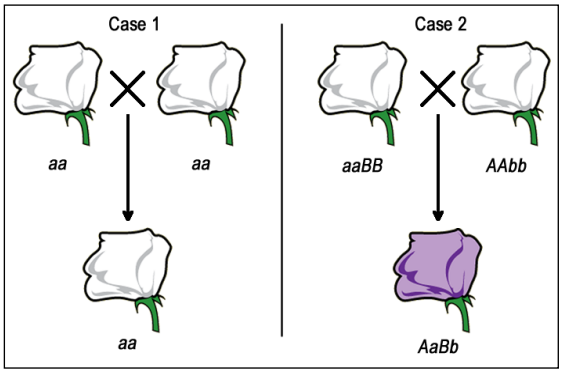


Figure 4.10B – Interpretation: The pure breeding, homozygous mutant parents had unknown genotypes before the complementation test, but it could be assumed that they had either mutations in the same genes (Case 1) or in different genes (Case 2). In Case 1, all of the progeny would have the mutant phenotype, because they would all have the same, homozygous genotype as the parents. In Case 2, each parent has a mutation in a different gene, therefore none of the F_1 progeny would be homozygous mutant at either locus. Note that the genotype in Case 1 could be written as either *aa* or *aaBB*. (Original-Deyholos-CC:AN)

4.6.3 - COMPLEMENTATION GROUPS = GROUPS OF ALLELIC MUTATIONS.

So, with the third mutant strain above, we could assign it to be allelic with either gene A or gene B, or some other locus, should it complement both gene A and gene B mutations. If we have a fourth, fifth, sixth, etc. (e.g. they came from different natural populations or from independently mutagenized individuals) white flower strains, then we could begin to group the allelic mutations into what are called **complementation groups**. These are groups of mutations that FAIL TO COMPLEMENT one another (a group of NON-complementing mutations).

A "group" can consist of as few as one mutation and as many as all the mutants under study. Each group represents a set of mutations in the same gene (allelic). The number of complementation groups represents the number of genes that are represented in the total collection of mutations. It all depends on how many mutations you have in that gene. For example, the *white* gene in Drosophila has >300 different mutations described in the literature. If you were to obtain and cross all these mutations to themselves you would find they all belonged to the same complementation group. Each complementation group represents a gene.

If, however, you obtained a different mutation, *vestigial* for example, and crossed it to a *white* mutation, the double heterozygote would result in red eyes and normal wings (wild type for both characters) so the two would complement and represent two different complementation groups: (1) *white*, (2) *vestigial*. The same would be true for the other eye-colour mutations mentioned in Section 4.5.4 above. For example, if you crossed a *scarlet* mutant to a *white* mutant, the double heterozygote would have wild type red eyes.

SUMMARY

- When a variation in DNA sequence originated recently, and is rare in a population, we call that change a mutation.
- When variations in DNA sequence co-exist in a population, and neither one can be meaningfully defined as wild-type, we call the variations polymorphisms.
- Mutations may either occur spontaneously, or may be induced by exposure to mutagens.
- Mutations may result in either substitutions, deletions, or insertions.
- Mutation usually causes either a partial or complete loss of function, but sometimes results in a gain of function, including new functions.
- Spontaneous mutations arise from many sources including natural errors in DNA replication, usually associated with base mispairing, or else insertion deletion especially within repetitive sequences.
- Induced mutations result from mispairing, DNA damage, or sequence interruptions caused by chemical, biological, or physical mutagens.
- By randomly inducing mutations, then screening for a specific phenotype, it is possible to identify genes associated with specific biological pathways.
- Transposable elements are dynamic, abundant components of eukaryotic genomes and important forces in evolution.
- Transposable elements are dynamic, abundant components of eukaryotic genomes and important forces in evolution.
- The efficiency of mutant screening is limited by silent mutations, redundancy, and embyronic lethality.
- Mutation of different genes can produce a similar phenotype.
- Complementation testing determines whether two mutants are the result of mutation of the same gene (allelic mutations), or if each mutant is caused by mutation of a different gene (non-allelic mutations).

KEY TERMS

mutation
mutant
polymorphism
insertion
deletion
substitution
mutagen
DNA replication error
Strand slippage
biological mutagen
chemical mutagen
physical mutagen
mispairing
loop
SSR
insertional mutagen
Class I, Class II
transposon
retrotransposon
reverse transcriptase
transposase
non-autonomous
autonomous
SINE, LINE, Alu
P-element
T-DNA
copy-and-paste
cut-and-paste
alkylation agent
EMS
intercalating agent
benzopyrene
carcinogenic
ethidium bromide
thymine dimer
mutant screen

loss-of-function
gain-of-function
amorph
null
hypomorph
hypermorph
neopmorph
dominant negative
somatic cells
germline cells
silent mutation
inter-genic region
redundancy
essential gene
recessive lethal allele
complementation testing
allelic / non-allelic
cM_0, M_1, M_2
redundancy
lethality
allelic
non-allelic
complementation group

STUDY QUESTIONS

4.1 How are polymorphisms and mutations alike? How are they different?

4.2 What are some of the ways a substitution can occur in a DNA sequence?

4.3 What are some of the ways a deletion can occur in a DNA sequence?

4.4 What are all of the ways an insertion can occur in a DNA sequence?

4.5 In the context of this chapter, explain the health hazards of smoking tobacco.

4.6 You have a female fruit fly, whose father was exposed to a mutagen (she, herself, wasn't). Mating this female fly with another non-mutagenized, wild type male produces offspring that all appear to be completely normal, except there are twice as many daughters as sons in the F_1 progeny of this cross.

a) Propose a hypothesis to explain these observations.

b) How could you test your hypothesis?

4.7 You decide to use genetics to investigate how your favourite plant makes its flowers smell good.

a) What steps will you take to identify some genes that are required for production of the sweet floral scent? Assume that this plant is a self-pollinating diploid.

b) One of the recessive mutants you identified has fishy-smelling flowers, so you name the mutant (and the mutated gene) *fishy*. What do you hypothesize about the normal function of the wild-type *fishy* gene?

c) Another recessive mutant lacks floral scent altogether, so you call it *nosmell*. What could you hypothesize about the normal function of this gene?

4.8 Suppose you are only interested in finding dominant mutations that affect floral scent.

a) What do you expect to be the relative frequency of dominant mutations, as compared to recessive mutations, and why?

b) How will you design your screen differently than in the previous question, in order to detect dominant mutations specifically?

c) Which kind of mutagen is most likely to produce dominant mutations, a mutagen that produces point mutations, or a mutagen that produces large deletions?

4.9 Which types of transposable elements are transcribed?

4.10 You are interested in finding genes involved in synthesis of proline (Pro), an amino acid that is normally synthesizes by a particular model organism.

a) How would you design a mutant screen to identify genes required for Pro synthesis?

b) Imagine that your screen identified ten mutants (#1 through #10) that grew poorly unless supplemented with Pro. How could you determine the number of different genes represented by these mutants?

c) If each of the four mutants represents a different gene, what will be the phenotype of the F1 progeny if any pair of the four mutants are crossed?

d) If each of the four mutants represents the same gene, what will be the phenotype of the F1 progeny if any pair of the four mutants are crossed?

Chapter 5 PEDIGREES AND POPULATIONS

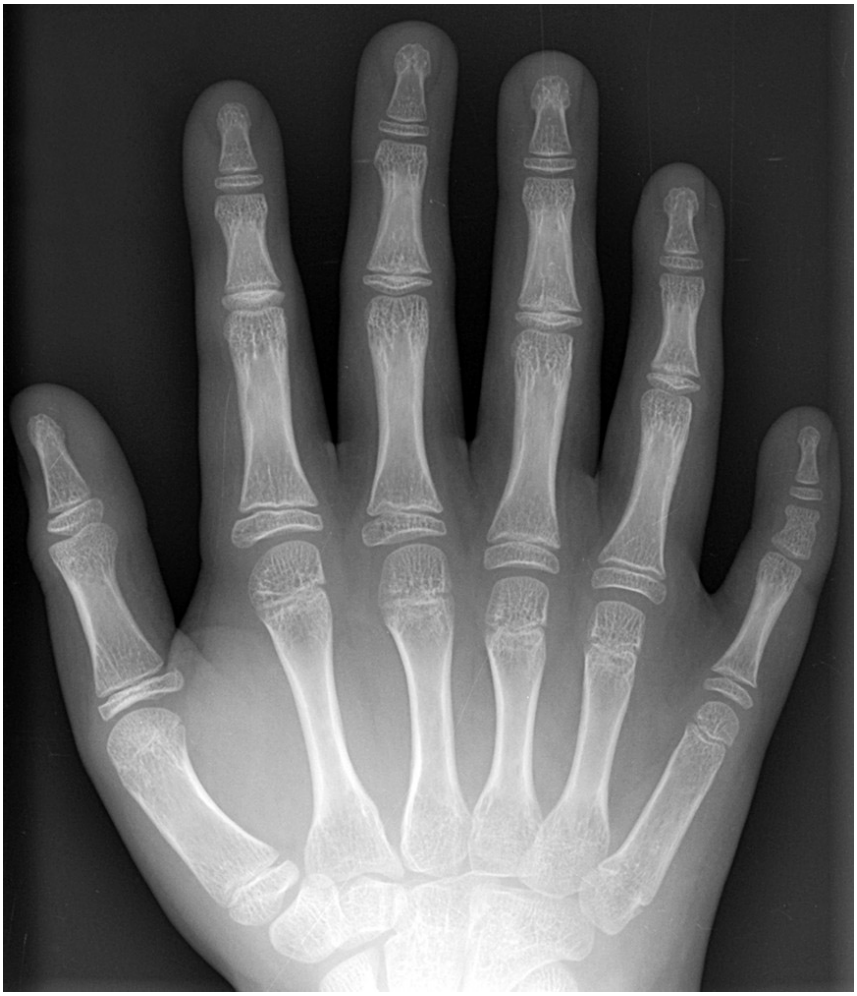


Figure 5.1
Polydactyly (six fingers in this case) is an example of a human trait that can be studied by pedigree analysis. (Wikipedia-en:User:DRGNU23-GFDL)

The basic concepts of genetics described in the preceding chapters can be applied to almost any eukaryotic organism. However, some techniques, such as test crosses, 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.

5.1 PEDIGREE ANALYSIS

Pedigree charts are diagrams that show the phenotypes and/or genotypes for a particular organism and its ancestors. 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 both basic research and **genetic counseling**.

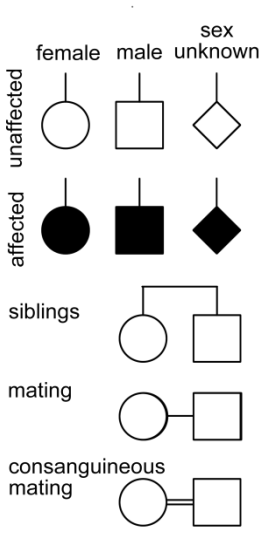


Figure 5.2
Symbols used in drawing a pedigree. (Original-Deyholos-CC:AN)

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. 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 5.2, 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 is two lines.

The affected individual that brings the family to the attention of a geneticist is called the **proband** (or propositus). 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. Note that when a pedigree is constructed, it is often unknown whether a particular individual is a carrier or not, so not all carriers are always explicitly indicated in a pedigree. For simplicity, in this chapter we will assume that the pedigrees presented are accurate, and represent fully penetrant traits.

5.2 Inferring the Mode 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).

5.2.1 Autosomal Dominant (AD)

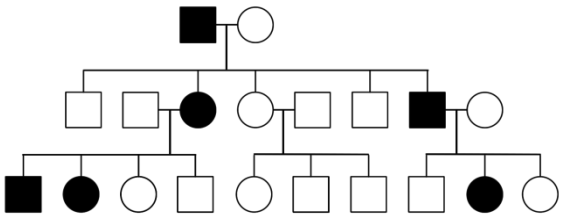


Figure 5.3
A pedigree consistent with AD inheritance. (Original-Deyholos-CC:AN)

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. Thus, every affected individual must have an affected parent. A pedigree with affected

individuals in every generation is typical of AD 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 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.

5.2.2 X-LINKED DOMINANT (XD)

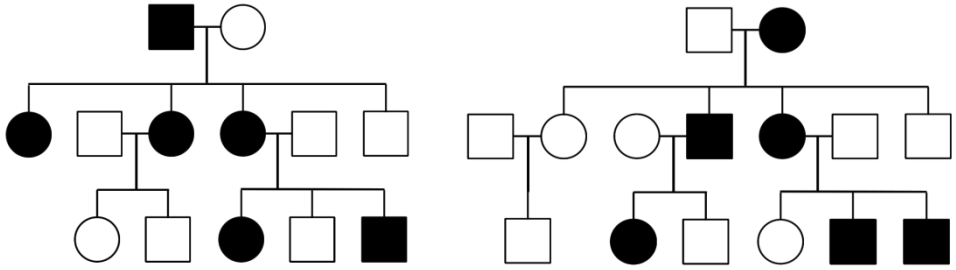


Figure 5.4
Two pedigrees consistent with XD inheritance. (Original-Deyholos_CC:AN)

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 <u>not</u> possible with XD, since males inherit their X chromosome from their mothers.

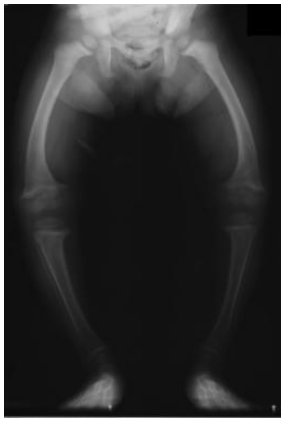


Figure 5.5
Some types of rickets may follow an XD mode of inheritance. (Wikipedia-Mrish-CC:AS)

5.2.3 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.

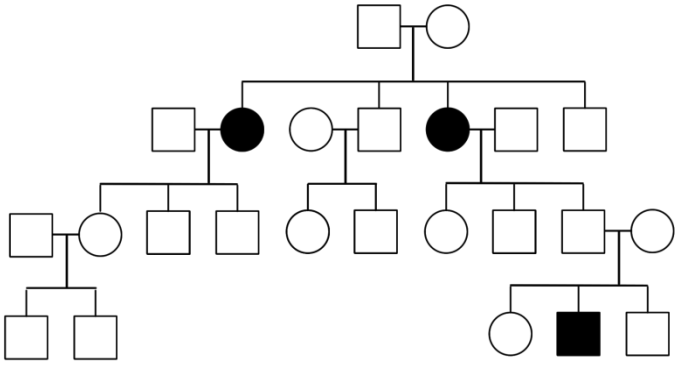


Figure 5.6
A pedigree consistent with AR inheritance. (Original-Deyholos_CC:AN)

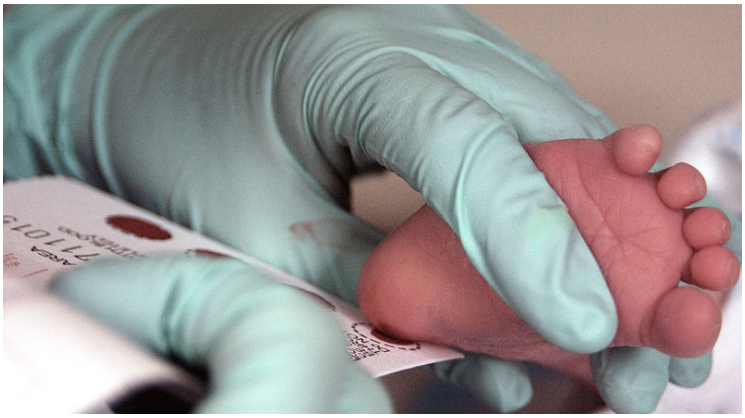


Figure 5.7
Many inborn errors of metabolism, such as phenylketonuria (PKU) are inherited as AR. Newborns are often tested for a few of the most common metabolic diseases. (Wikipedia-U.S. Air Force photo/Staff Sgt. Eric T. Sheler-PD)

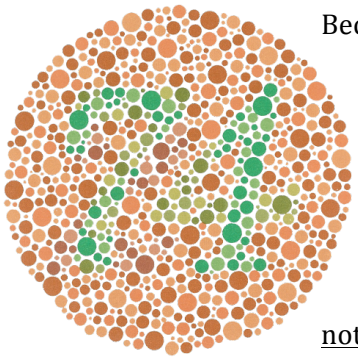


Figure 5.8
Some forms of colour blindness are inherited as XR-traits. Colour blindness is diagnosed using tests such as this. Ishihara Test. (Wikipedia-unknown-PD)

5.2.4 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. 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. On the other hand, one feature of a pedigree that can be used to definitively establish 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.

5.2.5 Y-LINKED AND MITOCHONDRIAL INHERITANCE.

Two additional modes are **Y-linked** and **Mitochondrial inheritance**.

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 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).

Mutations in Mitochondrial DNA are inherited through the maternal line (from your mother). 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 the mitochondria are inherited via the egg, not the sperm. Mitochondrial DNA polymorphisms are also used to investigate evolutionary lineages, both ancient and recent. Because of the relative similarity of sequence mtDNA is also used in species identification in ecology studies.

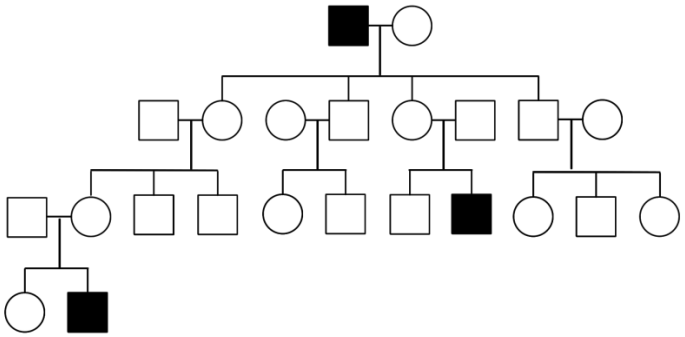


Figure 5.9
A pedigree consistent with XR inheritance. (Original-Deyholos-CC:AN)

5.3 SPORADIC AND NON-HERITABLE DISEASES

Figure 5.10
Stephen Hawking (Wikipedia-NASA-PD)

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 (Fig. 5.10) and baseball player Lou Gehrig both suffered from sporadic ALS.

5.4 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 *Aa* x *Aa*, ¼ of the offspring are *aa*).

We can likewise calculate probabilities in the more complex pedigree shown in Figure 5.11.

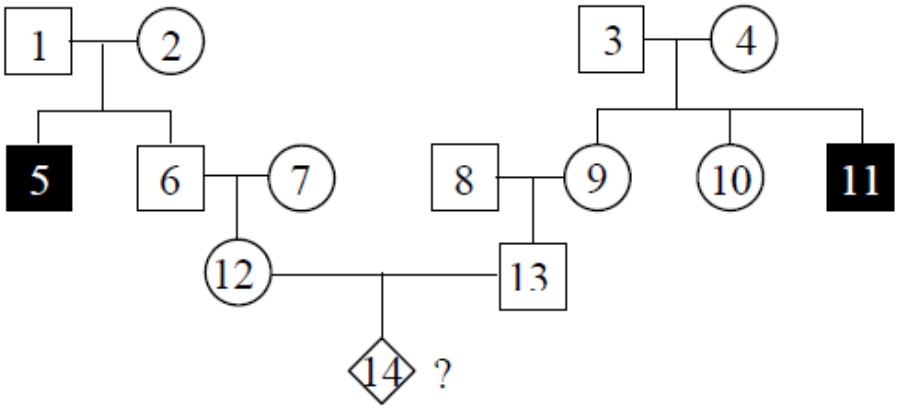


Figure 5.11 Individuals in this pedigree are labeled with numbers to make discussion easier. (Original-Deyholos-CC:AN)

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 (*Aa*), because they each had at least one affected (*aa*) child, but they are not affected themselves. This means that there is a 2/3 chance that individual #6 is also *Aa*. This is because according to Mendelian inheritance, when two heterozygotes mate, there is a 1:2:1 distribution of genotypes *AA*:*Aa*:*aa*. However,

because #6 is unaffected, he can't be *aa*, so he is either *Aa* or *AA*, but the probability of him being *Aa* is twice as likely as *AA*. 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 *Aa* × *AA*, half of the progeny will be *Aa*; 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.

Figure 5.12. (Flickr-Zach Stern-CC:AND)

5.5 Population Genetics

A **population** is a large group of individuals of the same species, who are capable of mating with each other. It is useful to know the frequency of particular alleles within a population, since this information can be used to calculate disease risks. Population genetics is also important in ecology and evolution, since changes in allele frequencies may be associated with migration or natural selection.

5.5.1 Allele frequencies may also be studied at the population level

The frequency of different alleles in a population can be determined from the frequency of the various phenotypes in the population. In the simplest system, with two alleles of the same locus (e.g. *A*,*a*), we use the symbol **p** to represent the frequency of the dominant allele within the population, and **q** for the frequency of the recessive allele. Because there are only two possible alleles, we can say that the frequency of p and q together represent 100% of the alleles in the population (**p+q=1**).

We can calculate the values of p and q, in a representative sample of individuals from a population, by simply counting the alleles and dividing by the total number of alleles examined. For a given allele, homozygotes will count for twice as much as heterozygotes.

For example:

genotype	number of individuals
AA	320
Aa	160
aa	20

p = (2(AA) + Aa) / (total alleles counted)
= (2(320) + 160) / (2(320) + 2(160) + 2(20)) = 0.8

q = (2(aa) + Aa) / (total alleles counted)
= (2(20) + 160) / (2(320) + 2(160) + 2(20)) = 0.2

	A (p)	*a* (q)
A (p)	p^2	pq
a (q)	pq	q^2

5.5.2 HARDY-WEINBERG FORMULA

With the allele frequencies of a population we can use an extension of the Punnett Square, and the product rule, to calculate the expected frequency of each genotype following random matings within the entire population. This is the basis of the **Hardy-Weinberg formula**:

$$\mathbf{p^2 + 2pq + q^2 = 1}.$$

Here p^2 is the frequency of homozygotes *AA*, 2pq is the frequency of the heterozygotes, and q^2 is the frequency of homozygotes *aa*.

Notice that if we substitute the allele frequencies we calculated above (p=0.8, q=0.2) into the formula $\mathbf{p^2 + 2pq + q^2 = 1}$, we obtain expected probabilities for each of the genotypes that exactly match our original observations:

$p^2=0.8^2=0.64$	0.64 x 500 = 320
2pq= 2(0.8)(0.2)=0.32	0.32 x 500 = 160
$q^2=0.2^2=0.04$	0.04 x 500 = 20

This is a demonstration of the **Hardy-Weinberg Equilibrium**, where both the genotype frequencies and allele frequencies in a population remain unchanged following successive matings within a population, *if* certain conditions are met. These conditions are listed in Table 5.1. Few natural populations actually satisfy all of these conditions. Nevertheless, large populations of many species, including humans, appear to approach Hardy-Weinberg equilibrium for many loci. In these situations, deviations of a particular gene from Hardy-Weinberg equilibirum can be an indication that one of the alleles affects the reproductive success of organism, for example through natural selection or **assortative mating**.

Table 5.1. Conditions for the Hardy-Weinberg equilibrium

- Random mating: Individuals of all genotypes mate together with equal frequency. Assortative mating, in which certain genotypes preferentially mate together, is a type of non-random mating.
- No natural selection: All genotypes have equal fitness.
- No migration: Individuals do not leave or enter the population.
- No mutation: The allele frequencies do not change due to mutation.

The Hardy-Weinberg formula can also be used to estimate allele frequencies, when only the frequency of one of the genotypic classes is known. For example, if 0.04% of the population is affected by a particular genetic condition, and all of the affected individuals have the genotype aa, then we assume that $q^2 = 0.0004$ and we can calculate p, q, and 2pq as follows:

$q^2 = 0.04\% = 0.0004$

$q = \sqrt{0.0004} = 0.02$

$p = 1-q = 0.98$

$2pq = 2(0.98)(0.02) = 0.04$

Thus, approximately 4% of the population is expected to be heterozygous (i.e. a carrier) of this genetic condition. Note that while we recognize that the population is probably not exactly in Hardy-Weinberg equilibrium for this locus, application of the Hardy-Weinberg formula nevertheless can give a reasonable estimate of allele frequencies, in the absence of any other information.

SUMMARY

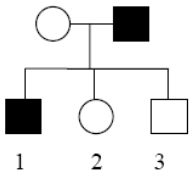
- Pedigree analysis can be used to determine the mode of inheritance of specific traits such as diseases. Loci can be X- or Y-linked or autosomal in location and alleles either dominant or recessive with respect to wild type.
- If the mode of inheritance is known, a pedigree can be used to calculate the probability of inheritance of a particular genotype by an individual.
- The frequencies of all alleles and genotypes remain unchanged through successive generations of a population in Hardy-Weinberg equilibrium.
- Populations in true Hardy-Weinberg equilibrium have random mating, and no genetic drift, no migration, no mutation, and no selection with respect to the gene of interest.
- The Hardy-Weinberg formula can be used to estimate allele and genotype frequencies given only limited information about a population.

KEY TERMS

Pedigree charts	X-linked dominant	$p+q=1$
mode of inheritance	X-linked recessive	Hardy-Weinberg
genetic counseling	Y-linked inheritance	formula
incomplete penetrance	Y-linked	$p^2 + 2pq + q^2=1$
variable expressivity	mitochondrial	Hardy-Weinberg
proband	inheritance (mtDNA)	equilibrium
affected	sporadic	assortative mating
carrier	product rule	random mating
autosomal dominant	sum rule	migration
autosomal recessive	population	genetic drift

STUDY QUESTIONS

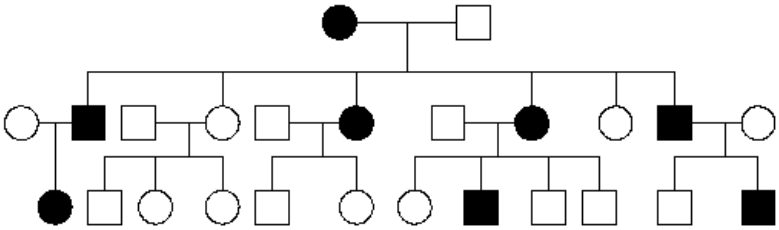
1. What are some of the modes of inheritance that are consistent with this pedigree?



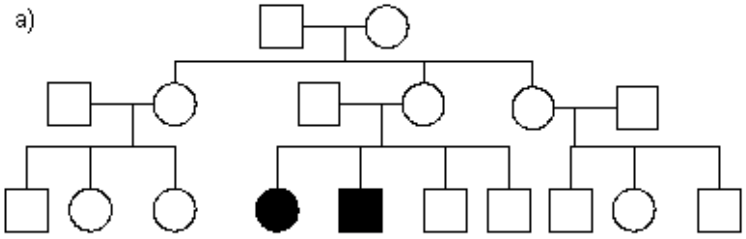
2. In this pedigree in question 1, the mode of inheritance cannot be determined unamibguously. What are some examples of data (e.g. from other generations) that, if added to the pedigree would help determine the mode of inheritance?

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)

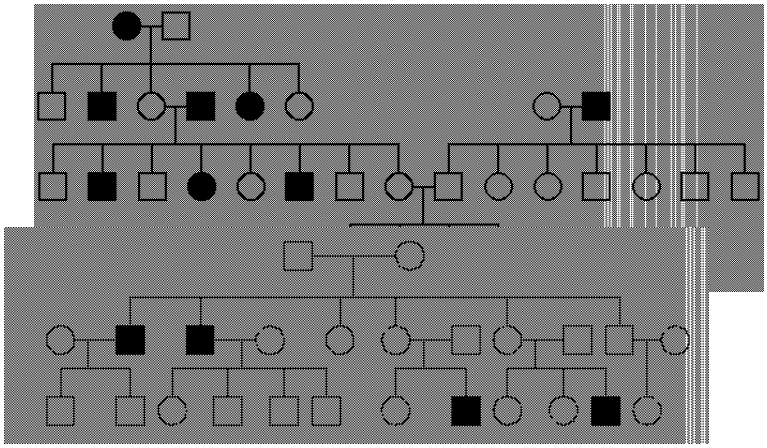


b)

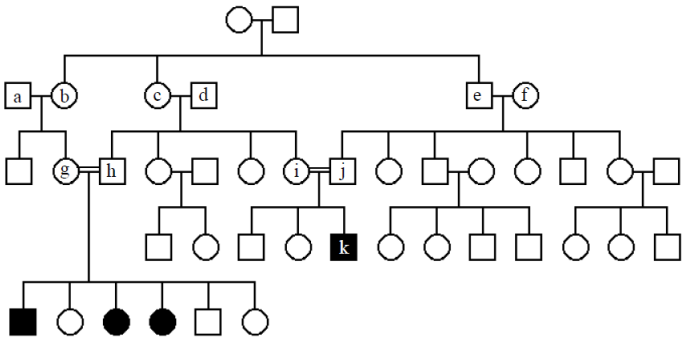


c)

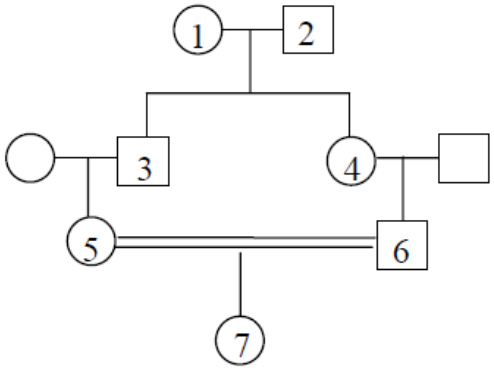
d)



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.



6. You are studying a population in which the frequency of individuals with a recessive homozygous genotype is 1%. Assuming the population is in Hardy-Weinberg equilibrium, calculate:
a) The frequency of the recessive allele.
b) The frequency of dominant allele.
c) The frequency of the heterozygous phenotype.
d) The frequency of the homozygous dominant phenotype.

7. Determine whether the following population is in Hardy-Weinberg equilibrium.

genotype	number of individuals
AA	432
Aa	676
aa	92

8. Out of 1200 individuals examined, 432 are homozygous dominant (*AA*)for a particular gene. What numbers of individuals of the other two genotypic classes (*Aa, aa*) would be expected if the population is in Hardy-Weinberg equilibrium?

9. Propose an explanation for the deviation between the genotypic frequencies calculated in question 8 and those observed in the table in question 7.

Chapter 6 GENETIC ANALYSIS OF MULTIPLE GENES

Figure 6.1
Coat color in animals is an example of a trait that is controlled by more than one locus. (Flickr-Gossamer1013-CC:AND)

The principles of genetic analysis that we have described for a single locus can be extended to the study of alleles at two loci simultaneously. Analysis of two loci in parallel is required for genetic mapping and can also reveal gene interactions. These techniques are very useful for both basic and applied research. Before discussing these techniques, we will first revisit Mendel's classical experiments.

6.1 DIHYBRID CROSSES

6.1.1 MENDEL'S SECOND LAW

Before Mendel, it had not yet been established that heritable traits were controlled by discrete factors. Therefor an important question was therefore whether distinct traits were controlled by discrete factors that 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. 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.) 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 6.2), 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 were in the same individual?

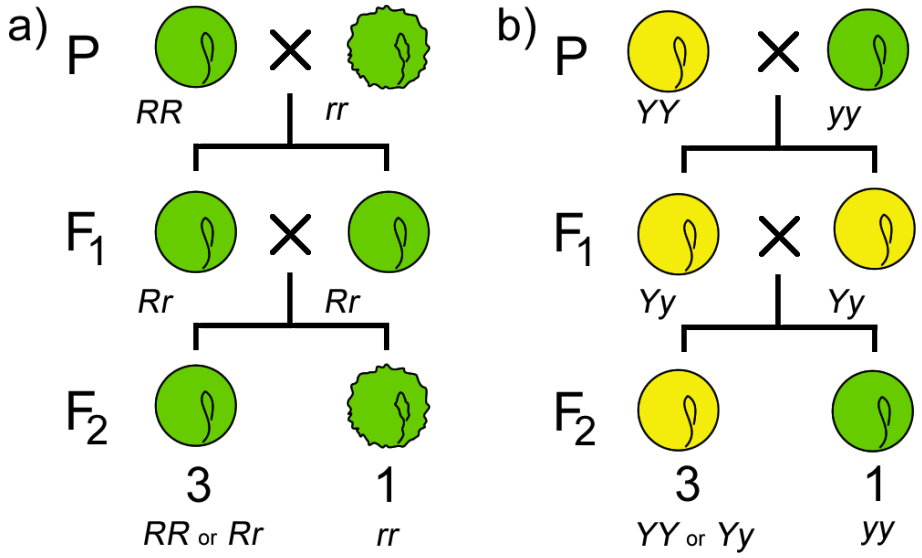


Figure 6.2
Monohybrid crosses involving two distinct traits in peas. a) is R/r and b) is Y/y. (Original-Deyholos-CC:AN)

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, and which were also **dihybrids**; they carried two alleles at each of two loci (Figure 6.3),.

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 6.3). 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 (Table 6.1). Likewise, ¾ × ¼ = 3/16 of the progeny would be both round and yellow, and so on. By applying the product 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.

Table 6.1
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

The 9:3:3:1 phenotypic ratio that we calculated using the product rule can also be obtained using Punnett Square (Figure 6.4). First, we list the 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 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

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 *R* allele and one *Y* allele will be round and yellow; these genotypes are shown in the nine, green-shaded cells in Figure 6.4. We can represent all of four of the different genotypes shown in these cells with the notation (*R_Y_*), where the blank line (_), means "any allele". The three offspring that have at least one R allele and are homozygous recessive for *y* (i.e. *R_yy*) will have a round, green phenotype. Conversely the three progeny that are homozygous recessive *r*, but have at least one *Y* allele (*rrY_*) will have wrinkled, yellow seeds. Finally, the rarest phenotypic class of wrinkled, yellow seeds is produced by the doubly homozygous recessive genotype, *rryy*, which is expected to occur in only one of the sixteen possible offspring represented in the square.

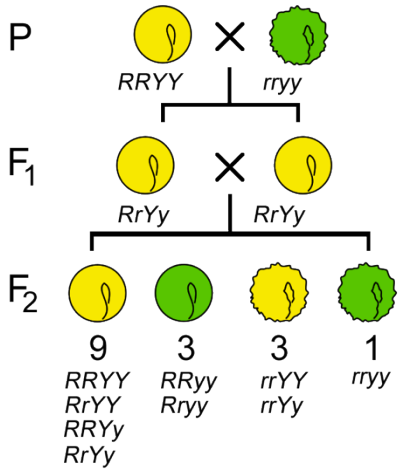


Figure 6.3
Pure-breeding lines are crossed to produce dihybrids in the F_1 generation. The cross of these particular dihybrids produces four phenotypic classes.
(Original-Deyholos-CC:AN)

	RY	***Ry***	***rY***	***ry***
RY	*RRYY*	*RRYy*	*RrYY*	*RrYy*
Ry	*RRYy*	*RRyy*	*RrYy*	*Rryy*
rY	*RrYY*	*RrYy*	*rrYY*	*rrYy*
ry	*RrYy*	*Rryy*	*rrYy*	*rryy*

Figure 6.4
A Punnett Square showing the results of the dihybrid cross from Figure 6.3. Each of the four phenotypic classes is represented by a different color of shading.
(Original-Deyholos-CC:AN)

6.1.2 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 *RrYy* × *RrYy*. In making these calculations, we assumed that: **(1)** both loci assort independently; **(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. Modified ratios in the progeny of a dihybrid cross can therefore reveal useful information about the genes involved.

Linkage is one of the most important reasons for distortion of the ratios expected from independent assortment. Linked genes 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 7. Deviations from 9:3:3:1 ratios can also be due to interactions between genes. These interactions will be discussed in the remainder of this chapter. For simplicity, we will focus on examples that involve easily scored phenotypes, such as pigmentation. Nevertheless, keep in mind that the analysis of segregation ratios of any markers can provide insight into a wide range of biological processes they represent.

6.2 Epistasis and Other Gene Interactions

Some dihybrid crosses produce a phenotypic ratio that differs from 9:3:3:1, such as 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 interactions between genes.

Figure 6.5
Labrador Retrievers with different coat colors: (from left to right) black, chocolate, yellow: an example of recessive epistasis.
(Flickr-John Curley/PhilRomans/Miss Chien-CC:AND)

6.2.1 Recessive epistasis

Epistasis (which means "standing upon") occurs when the phenotype of one locus masks, or prevents, the phenotype 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_bb, aaB_,* and *aabb*. If either of the singly homozygous recessive genotypes (i.e. *A_bb* or *aaB_*) has the same phenotype as the double homozygous recessive (*aabb*), then a **9:3:4** phenotypic ratio will be obtained. For example, in the Labrador Retriever breed of dogs (Figure 6.5), 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 *bb* hair appears brown. A second locus, which we will call *E*, controls the deposition of melanin in the hairs. At least one functional *E* allele is required to deposit any pigment, whether it is black or brown. Thus, all retrievers that are *ee* fail to deposit any melanin (and so appear pale yellow), regardless of the genotype at the *B* locus (Figure 6.6).

The *ee* genotype is therefore said to be **epistatic** to both the B and b alleles, since the homozygous *ee* phenotype masks the phenotype of the B locus. The B/b locus is said to be **hypostatic** to the *ee* genotype. Because the masking allele is in this case is recessive, this is called **recessive epistasis**.

	EB	*Eb*	*eB*	*eb*
EB	*EEBB*	*EEBb*	*EeBB*	*EeBb*
Eb	*EEBb*	*EEbb*	*EeBb*	*Eebb*
eB	*EeBB*	*EeBb*	*eeBB*	*eeBb*
eb	*EeBb*	*Eebb*	*eeBb*	*eebb*

Figure 6.6

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, *E_B_*); chocolate coat (brown, *E_bb*); yellow coat (yellow, *eeB_* or *eebb*). (Original-Locke-CC:AN)

6.2.2 Dominant epistasis

In some cases, a dominant allele at one locus may mask the phenotype of a second locus. This is called **dominant epistasis**, which produces a segregation ratio such as **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 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 6.7). Alleles of a locus that we will call *B* produce either yellow (*B_*) or green (*bb*) 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 *bb* combinations (Figure 6.8). 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 neither yellow or green pigments are produced.

Figure 6.7

Green, yellow, and white fruits of squash. (Flickr-Unknown-CC:AN)

	AB	***Ab***	***aB***	***ab***
AB	*AABB*	*AABb*	*AaBB*	*AaBb*
Ab	*AABb*	*AAbb*	*AaBb*	*Aabb*
aB	*AaBB*	*AaBb*	*aaBB*	*aaBb*
ab	*AaBb*	*Aabb*	*aaBb*	*aabb*

Figure 6.8

Genotypes and phenotypes among the progeny of a dihybrid cross of squash plants heterozygous for two loci affecting fruit color. (Original-Deyholos-CC:AN)

6.2.3 Duplicate gene action

When a dihybrid cross produces progeny in two phenotypic classes in a 15:1 ratio, this can be because the two loci's gene products have the same (redundant) functions within the same biological pathway. Yet another pigmentation pathway, in this case in wheat, provides an example of this **duplicate gene action**. The biosynthesis of red pigment near the surface of wheat seeds (Figure 6.9) 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 *aaB_* or *A_bb*). Only the doubly recessive mutant (*aabb*), which lacks function of both genes, shows a phenotype that differs from that produced by any of the other genotypes (Figure 6.10). 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 6.9

Red (left) and white (right) wheat seeds. (cropwatch.unl.edu-pending)

	AB	***Ab***	***aB***	***ab***
AB	*AABB*	*AABb*	*AaBB*	*AaBb*
Ab	*AABb*	*AAbb*	*AaBb*	*Aabb*
aB	*AaBB*	*AaBb*	*aaBB*	*aaBb*
ab	*AaBb*	*Aabb*	*aaBb*	*aabb*

Figure 6.10

Genotypes and phenotypes among the progeny of a dihybrid cross of a wheat plants heterozygous for two loci affecting seed color. (Original-Deyholos-CC:AN)

6.2.4 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, due to **complementary gene action** (meaning that the functions of both genes work together to produce a final product). 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 6.11). Loss of function of either A or B, or both, will have the same result: no pigment production. Thus *A_bb*, *aaB_*, and *aabb* will all be colorless, while only *A_B_* genotypes will produce pigmented product (Figure 6.12). 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.

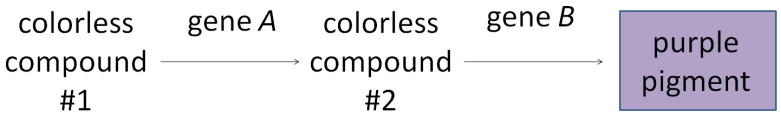


Figure 6.11

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. (Original-Deyholos-

	AB	***Ab***	***aB***	***ab***
AB	*AABB*	*AABb*	*AaBB*	*AaBb*
Ab	*AABb*	*AAbb*	*AaBb*	*Aabb*
aB	*AaBB*	*AaBb*	*aaBB*	*aaBb*
ab	*AaBb*	*Aabb*	*aaBb*	*aabb*

Figure 6.12

Genotypes and phenotypes among the progeny of a dihybrid cross of a hypothetical plant heterozygous for two loci affecting flower color. (Original-Deyholos-CC:AN)

SUMMARY

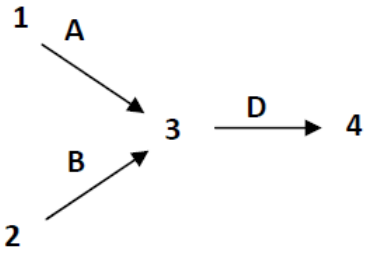
- The alleles of different loci are inherited independently of each other, unless they are genetically linked.
- The expected phenotypic ratio of a dihybrid cross is 9:3:3:1, except in cases of linkage or gene interactions that modify this ratio.
- Modified ratios from 9:3:3:1 are seen in the case of recessive and dominant epistasis, duplicate genes, and complementary gene action. This usually indicates that the two genes interact within the same biological pathway.

KEY TERMS

Mendel's Second Law	linkage
Dihybrid	recessive epistasis
9:3:3:1	dominant epistasis
9:3:4	complementary action
12:3:1	redundancy
independent assortment	duplicate gene action

STUDY QUESTIONS

Answer questions 6.1 -6.3 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.



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

a) *aa*
b) *bb*
c) *dd*
d) *aabb*
e) *aadd*
f) *bbdd*
g) *aabbdd*
h) What will be the phenotypic ratios among the offspring of a cross *AaBb* × *AaBb*?
i) What will be the phenotypic ratios among the offspring of a cross *BbDd* × *BbDd*?
j) What will be the phenotypic ratios among the offspring of a cross *AaDd* × *AaDd*?

6.2 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) *aa*
b) *bb*
c) *dd*
d) *aabb*
e) *aadd*
f) *bbdd*
g) *aabbdd*
h) What will be the phenotypic ratios among the offspring of a cross *AaBb* × *AaBb*?
i) What will be the phenotypic ratios among the offspring of a cross *BbDd* × *BbDd*?
j) What will be the phenotypic ratios among the offspring of a cross *AaDd* × *AaDd*?

6.3 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) *aa*
b) *bb*
c) *dd*
d) *aabb*
e) *aadd*
f) *bbdd*
g) *aabbdd*
h) What will be the phenotypic ratios among the offspring of a cross *AaBb* × *AaBb*?

i) What will be the phenotypic ratios among the offspring of a cross *BbDd* × *BbDd*?

j) What will be the phenotypic ratios among the offspring of a cross *AaDd* × *AaDd*?

6.4 Which of the situations in questions 6.1 – 6.3 demonstrate epistasis?

6.5 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 6.6
b) Figure 6.8
c) Figure 6.10
d) Figure 6.12

6.6 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.

6.7 With reference to question 6.6, if the recessive allele, *g* is mutated again to make allele *g**, what are the possible phenotypes of a homozygous *g* g** individual?

6.8 Again, in reference to question 6.7, what are the possible phenotypes of a homozygous *aagg* 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 *AaGg* x *AaGg*

6.9 Calculate the phenotypic ratios from a dihybrid cross involving the two loci shown in Figure 6.13. There may be more than one possible set of ratios, depending on the assumptions you make about the phenotype of allele *b*.

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

Chapter 7 LINKAGE & MAPPING

Figure 7.1
Linkage affects the frequency at which some combinations of traits are observed.
(Wikipedia-Abiyoyo-CC:AN)

7.1. LINKAGE

As we learned in Chapter 6, Mendel reported that the pairs of loci he observed behaved 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. 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.

7.2 RECOMBINATION

The term "recombination" is used in several different contexts in genetics. In reference to heredity, **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 (see Figure 7.2). **Interchromosomal recombination** occurs either through **independent assortment** of alleles whose loci are on different chromosomes (Chapter 6). **Intrachromosomal recombination** occurs through **crossovers** between loci on the same chromosomes (as described below). 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). If meiosis results in recombination, the products are said to have a **recombinant genotype**. On the other hand, if no recombination occurs during meiosis, the products have their original combinations and are said to have a non-recombinant, or **parental genotype**. Recombination is important because it contributes to the genetic variation that may be observed between individuals within a population and acted upon by selection to produce evolution.

As an example of interchromosomal recombination, consider loci on two different chromosomes as shown in Figure 7.2. We know that if these loci are on different chromosomes, there are no physical connections between them, so they are **unlinked** and will segregate independently as did Mendel's traits. The segregation depends on the relative 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 7.2. More precisely, there is a 50% probability for recombinant genotypes, and a 50% probability for parental genotypes within the gametes produced by a meiocyte with unlinked loci. Indeed, if we examined all of the gametes that could be produced by this individual (which are the products of multiple independent meioses), we would note that approximately 50% of the gametes would be recombinant, and 50% would be parental. **Recombination frequency** (RF) is simply the number of recombinant gametes, divided by the total number of gametes. A frequency of approximately 50% recombination is therefore a defining characteristic of unlinked loci. Thus the greatest recombinant frequency expected is ~50%.

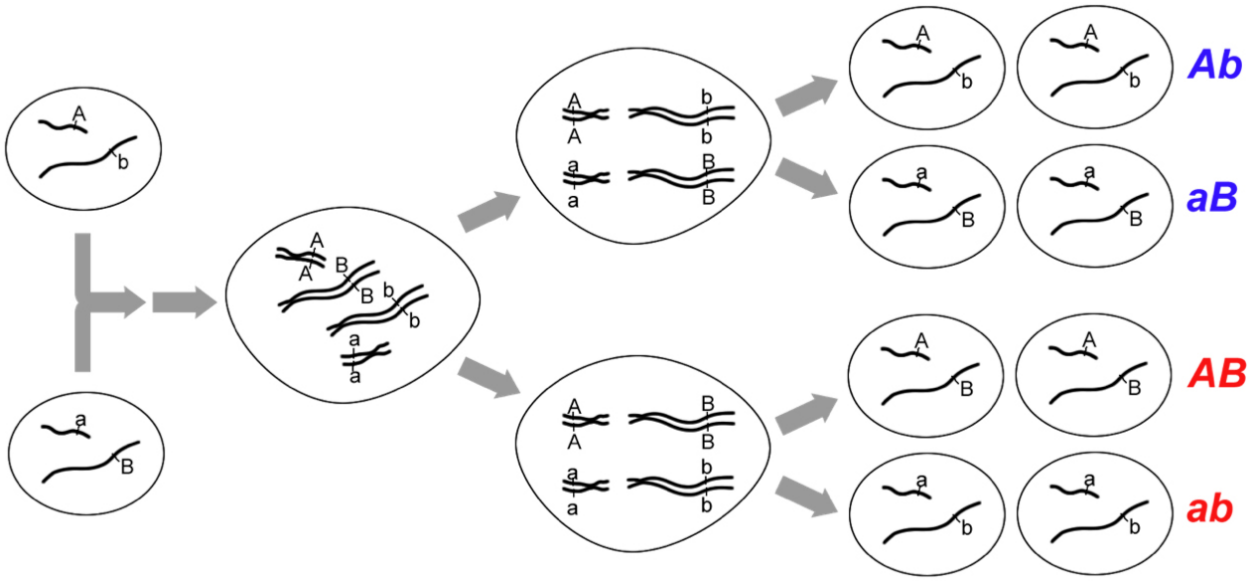


Figure 7.2 When two loci are on non-homologous chromosomes, their alleles will segregate in combinations identical to those present in the parental gametes (*Ab*, *aB*), and in recombinant genotypes (*AB*, *ab*) that are different from the parental gametes. (Original-Deyholos-CC:AN)

7.3 Linkage Reduces Recombination Frequency

Having considered unlinked loci above, 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 7.3). 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%. This is **complete** (or **absolute**) **linkage** and is rare, as the loci must be so close together that crossovers are never detected between them.

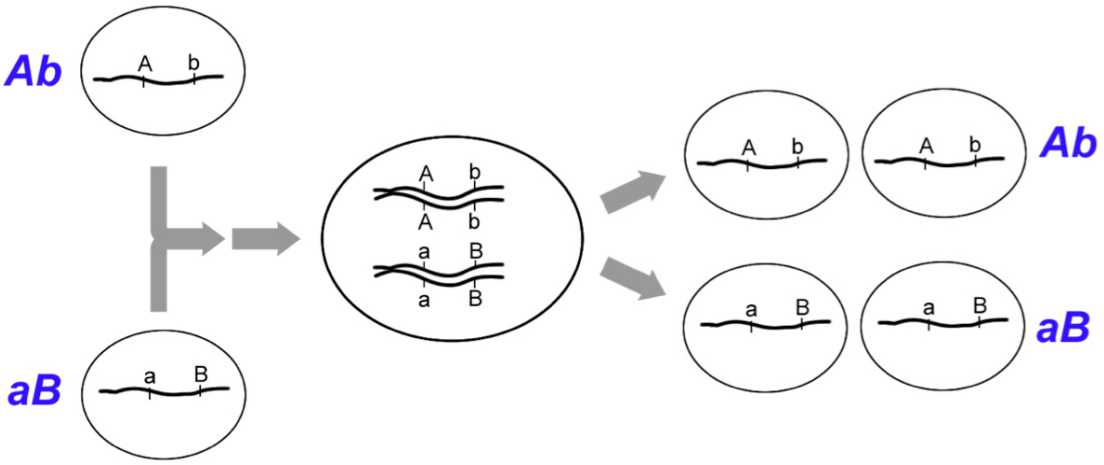


Figure 7.3 If two loci are completely linked, their alleles will segregate in combinations identical to those present in the parental gametes (*Ab*, *aB*). No recombinants will be observed. (Original-Deyholos-CC:AN)

7.4 Crossovers Allow Recombination of Linked Loci

Thus far, we have only considered situations with either no linkage (50% recombination) or complete linkage (0% recombination). 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. Genes that are on the same chromosome are said to be **syntenic** regardless of whether they are completely or incompletely linked. All linked genes are syntenic, but not all syntenic genes are linked, as we will learn later.

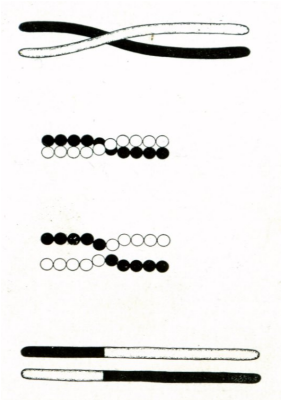


Figure 7.4
A depiction of a crossover from Morgan's 1916 manuscript. Only one pair of non-sister chromatids is shown, but remember there are 4 chromatids for each homologous pair in meiosis. (Wikipedia-Morgan-PD)

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 (Figure 7.4). Generally every pair of chromosomes has at least one (and often more) crossovers during meioses (Figure 7.5).

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% 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.

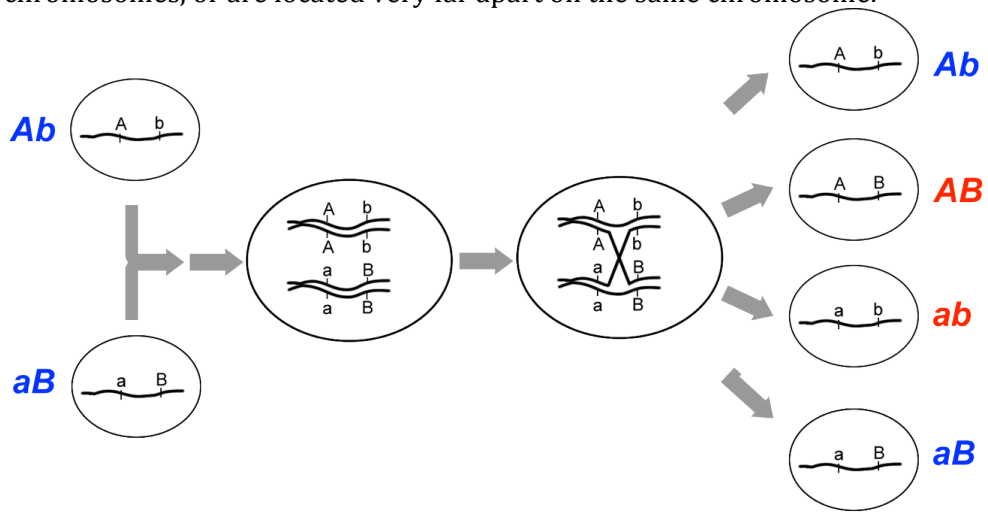


Figure 7.5 A crossover between two linked loci can generate recombinant genotypes (*AB*, *ab*), from the chromatids involved in the crossover. Remember that multiple, independent meioses occur in each organism, so this particular pattern of recombination will not be observed among all the meioses from this individual. (Original-Deyholos-CC:AN)

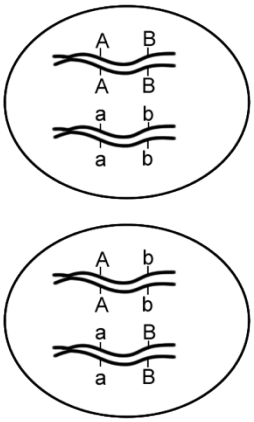


Figure 7.6 Alleles in coupling configuration (top) or repulsion configuration (bottom). (Original-Deyholos-CC:AN)

7.5 Inferring Recombination From Genetic Data

In the preceding examples, we had the advantage of knowing the approximate chromosomal positions of each allele involved, before we calculated the recombination frequencies. Knowing this information beforehand made it relatively easy to define the parental and recombinant genotypes, and to calculate recombination frequencies. However, in most experiments, we cannot directly examine the chromosomes, or even the gametes, so we must infer the arrangement of alleles from the phenotypes over two or more generations. Importantly, it is generally not sufficient to know the genotype of individuals in just one generation; for example, given an individual with the genotype *AaBb*, we do not know from the genotype alone whether the loci are located on the same chromosome, and if so, whether the arrangement of alleles on each chromosome is *AB* and *ab* or *Ab* and *aB* (Figure 7.6). The top cell has the two dominant alleles together and the two recessive alleles together and is said to have the genes in the **coupling** (or **cis**) **configuration**. The alternative shown in the cell below is that the genes are in the **repulsion** (or **trans**) **configuration**.

Fortunately for geneticists, the arrangement of alleles can sometimes be inferred if the genotypes of a previous generation are known. For example, if the parents of *AaBb* had genotypes *AAB*B and *aabb* respectively, then the parental gametes that fused to produce *AaBb* would have been genotype *AB* and genotype *ab*. Therefore, prior to meiosis in the dihybrid, the arrangement of alleles would likewise be *AB* and *ab* (Figure 7.7). Conversely, if the parents of *AaBb* had genotypes *aaBB* and *AAbb*, then the arrangement of alleles on the chromosomes of the dihybrid would be

aB and *Ab*. Thus, the genotype of the previous generation can determine which of an individual's gametes are considered recombinant, and which are considered parental.

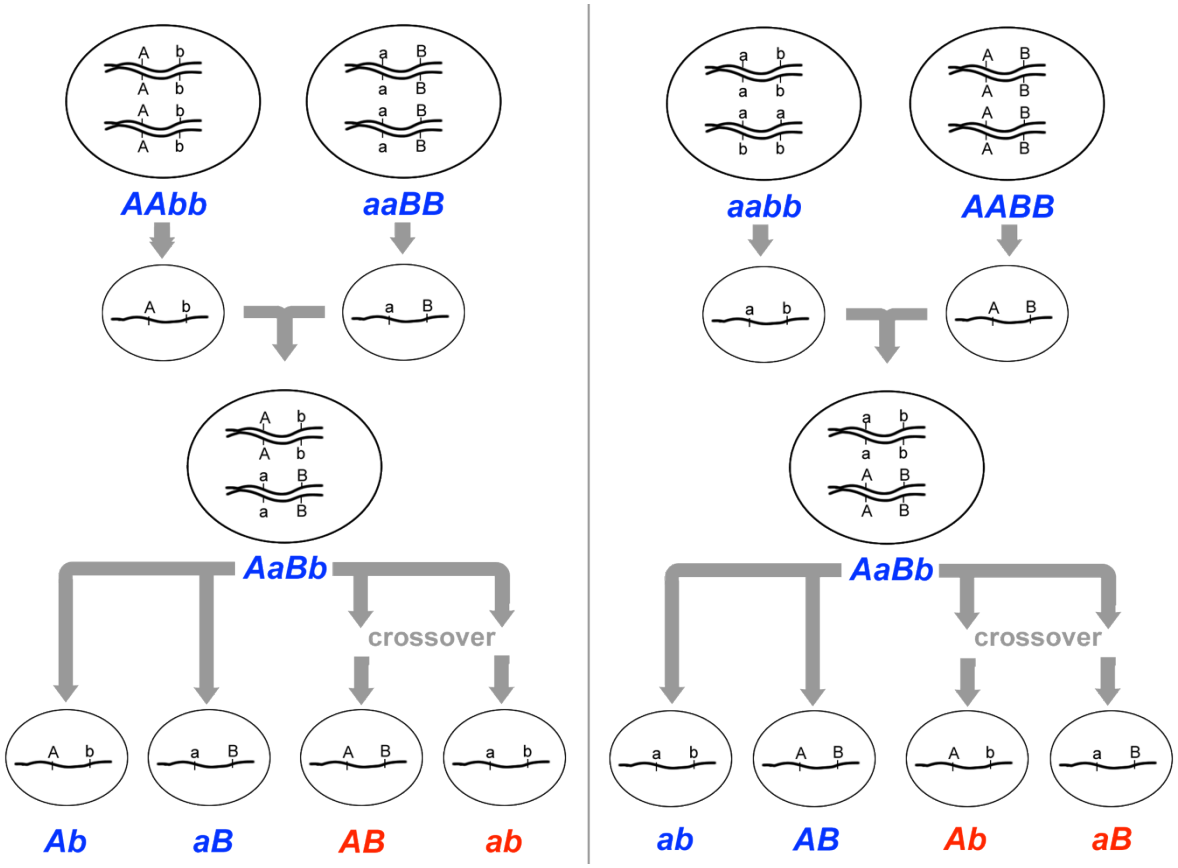


Figure 7.7 The genotype of gametes can be inferred unambiguously if the gametes are produced by homozygotes. However, recombination frequencies can only be measured among the progeny of heterozygotes (i.e. dihybrids). Note that the dihybrid on the left contains a different configuration of alleles than the dihybrid on the rightdue 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*. (Original-Deyholos-CC:AN)

Let us now consider a complete experiment in which our objective is to measure recombination frequency (Figure 7.8). 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 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*). 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 pairs of homologous chromosomes, or whether they are on the same chromosome, and if so, how close together they are.

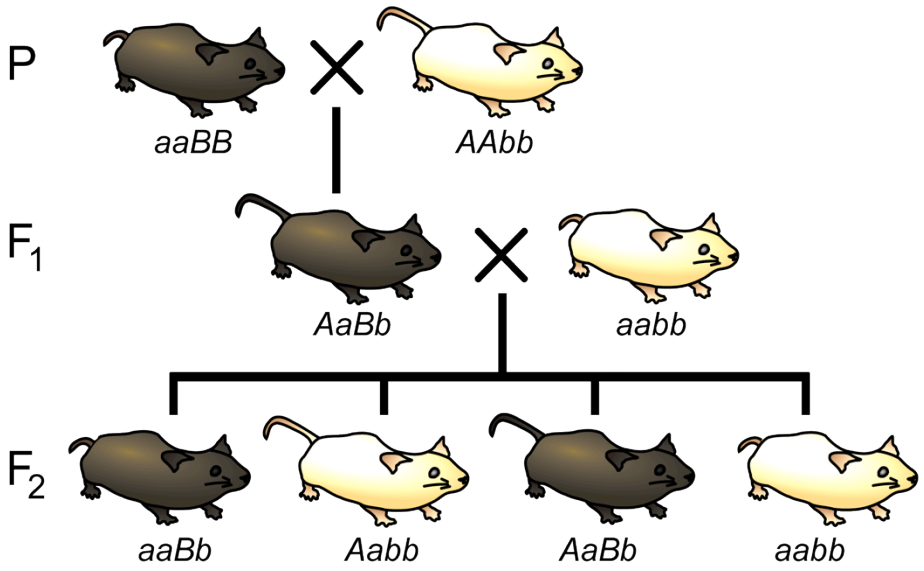


Figure 7.8
An experiment to measure recombination frequency between two loci. The loci affect coat color (B/b) and tail length (A/a). (Wikipedia-Modified

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 (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*, and non-recombinant (i.e. parental) gametes will have the genotypes *aB* or *Ab*.

How do we determine the genotype of the gametes produced by the dihybrid individual? The most practical method is to use a testcross (Figure 7.8), 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 F_2 generation for each of the four possible combinations of alleles in the gametes of the dihybrid. 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*) 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 7.1:

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

Table 7.1
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 7.8.

Given the data in Table 7.1, the calculation of recombination frequency is straightforward:

$$\text{recombination frequency} = \frac{\text{number of recombinant gametes}}{\text{total number of gametes scored}}$$

$$\text{R.F.} = \frac{13+17}{48+42+13+17}$$

$$= 25\ \%$$

7.6 GENETIC MAPPING

Because the frequency of recombination between two loci (up to 50%) is roughly proportional to the chromosomal distance between them, we can use recombination frequencies to produce genetic maps of all the loci along a chromosome and ultimately in the whole genome. The units of genetic distance are called **map units** (mu) or **centiMorgans** (cM), in honor of **Thomas Hunt Morgan** by his student, **Alfred Sturtevant**, who developed the concept. Geneticists routinely convert recombination frequencies into cM: the recombination frequency in percent is approximately 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 7.9). Note: this approximation works well for small distances (RF<30%) but progressively fails at longer distances because the RF reaches a maximum at 50%. Some chromosomes are >100 cM long but loci at the tips only have an RF of 50%. The method for mapping of these long chromosomes is shown below.

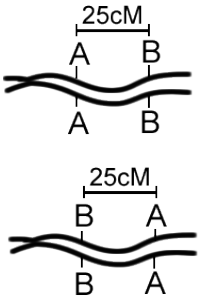


Figure 7.9
Two genetic maps consistent with a recombination frequency of 25% between A and B. Note the location of the centromere. (Original-Deyholos-CC:AN)

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.

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 7.10). 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 bases 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, its approximate position on a chromosome can be determined by crossing it with previously mapped genes, and then calculating the recombination frequency. If the novel gene and the previously mapped genes show complete or partial linkage, 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, through a process called **map-based cloning**.

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

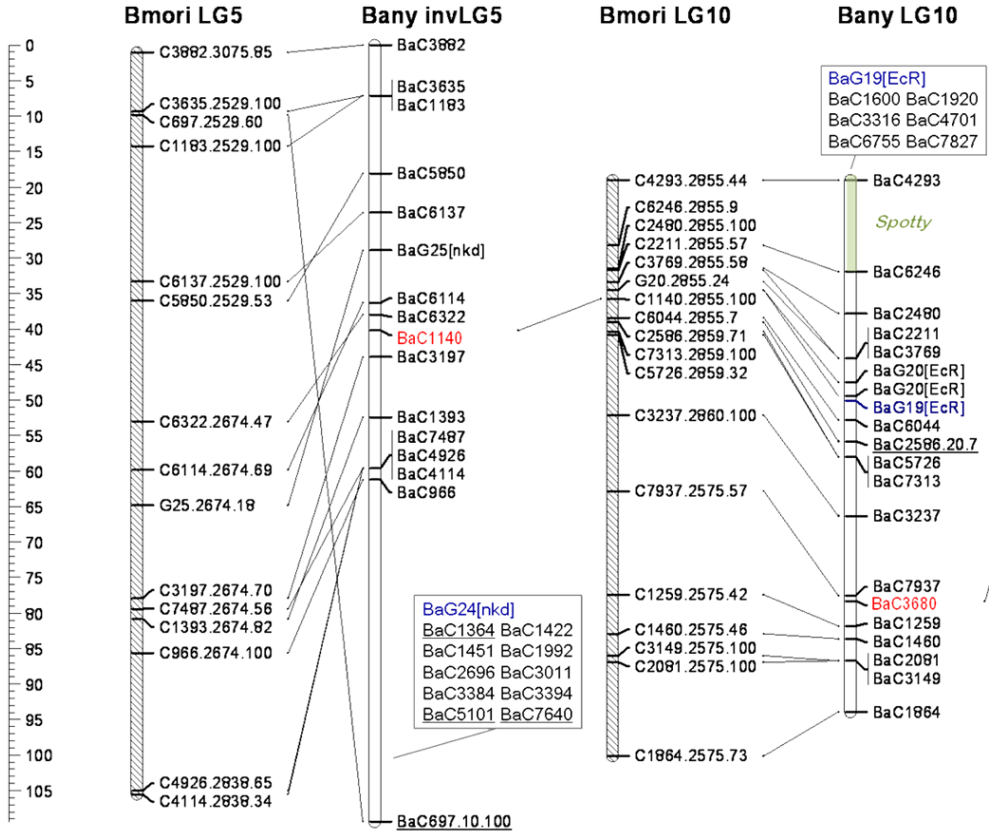


Figure 7.10 Genetic maps for regions of two chromosomes from two species of the moth, *Bombyx*. The scale at left shows distance in cM, and the position of various loci is indicated on each chromosome. Diagonal lines connecting loci on different chromosomes show the position of corresponding loci in different species. This is referred to as regions of **conserved synteny**. (NCBI-NIH-PD)

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 short distances (low RF values) than long distances. Observed recombination frequencies between two relatively distant markers tend to underestimate the actual number of crossovers that occurred. This is because as the distance between loci increases, so does the possibility of having a second (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 7.11) – 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.

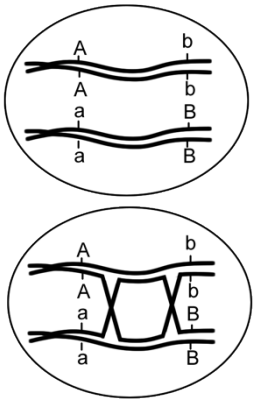


Figure 7.11
A double crossover between two loci will produce gametes with parental genotypes. (Original-Deyholos-CC:AN)

7.7 Mapping With Three-Point Crosses

A particularly efficient method of mapping three genes at once is the **three-point cross**, which allows the order and distance between three potentially linked genes to be determined in a single cross experiment (Figure 7.12). This is particularly useful when mapping a new mutation with an unknown location to two previously mapped loci. The basic strategy is the same as for the dihybrid mapping experiment described above; pure breeding lines with contrasting genotypes are crossed to produce an individual heterozygous at three loci (a trihybrid), which is then testcrossed to determine the recombination frequency between each pair of genes.

One useful feature of the three-point cross is that the order of the loci relative to each other can usually be determined by a simple visual inspection of the F_2 segregation data. If the genes are linked, there will often be two phenotypic classes that are much more infrequent than any of the others. In these cases, the rare phenotypic classes are usually those that arose from two crossover events, in which the locus in the middle is flanked by a crossover on either side of it. Thus, among the two rarest recombinant phenotypic classes, the one allele that differs from the other two alleles relative to the parental genotypes likely represents the locus that is in the middle of the other two loci. For example, based on the phenotypes of the pure-breeding parents in Figure 7.12, the parental genotypes are *aBC* and *AbC* (remember the order of the loci is unknown, and it is not necessarily the alphabetical order in which we wrote the genotypes). Because we can deduce from the outcome of the testcross (Table 7.2) that the rarest genotypes were *abC* and *ABc*, we can conclude that locus *A* that is most likely located between the other two loci, since it would require a recombination event between both *A* and *B* and between *A* and *C* in order to generate these gametes. Thus, the order of loci is *BAC* (which is equivalent to *CAB*).

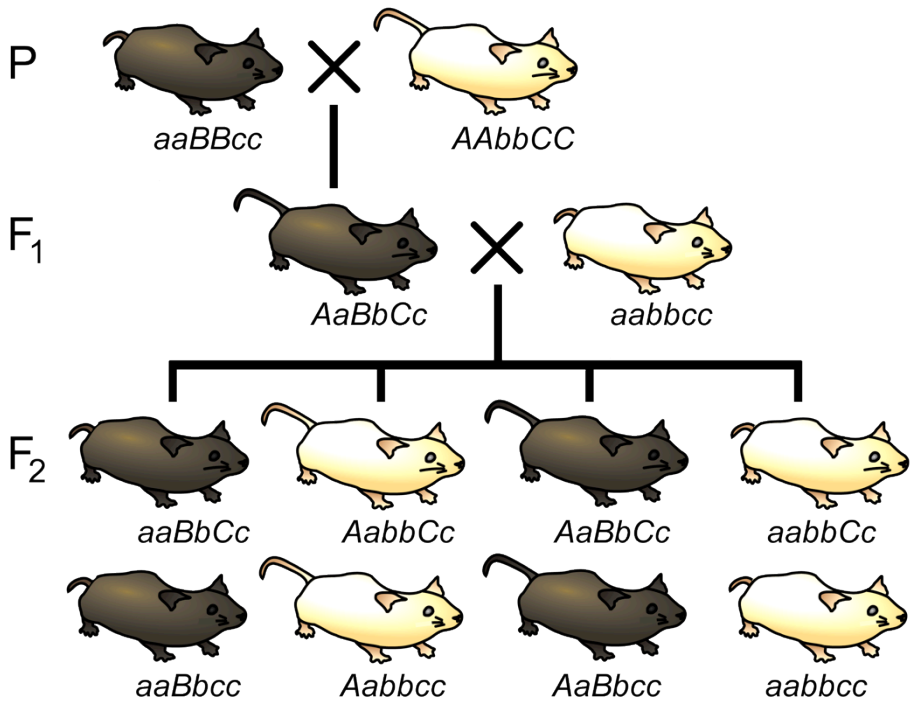


Figure 7.12 A three point cross for loci affecting tail length, fur color, and whisker length. (Original-Modified Deyholos-CC:AN)

tail phenotype	fur phenotype	whisker phenotype	number of progeny	gamete from trihybrid	genotype of F_2 from test cross	loci A, B	loci A, C	loci B, C
short	brown	long	5	*aBC*	*aaBbCc*	P	R	R
long	white	long	38	*AbC*	*AabbCc*	P	P	P
short	white	long	1	*abC*	*aabbCc*	R	R	P
long	brown	long	16	*ABC*	*AaBbCc*	R	P	R
short	brown	short	42	*aBc*	*aaBbcc*	P	P	P
long	white	short	5	*Abc*	*Aabbcc*	P	R	R
short	white	short	12	*abc*	*aabbcc*	R	P	R
long	brown	short	1	*ABc*	*AaBbcc*	R	R	P

Table 7.2 An example of data that might be obtained from the F_2 generation of the three-point cross shown in Figure 7.12. 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.

Recombination frequencies may be calculated for each pair of loci in the three-point cross as we did before for one pair of loci in our dihybrid (Figure 7. 8).

loci A,B R.F. = $\frac{1+16+12+1}{120}$ = 25%

loci A,C R.F. = $\frac{1+5+1+5}{120}$ = 10%

loci B,C R.F. = $\frac{5+16+12+5}{120}$ = 32%
(not corrected for double crossovers)

However, 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%)(Figure 7.13). 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. We already deduced that the map order must be *BAC* (or *CAB*), based on the genotypes of the two rarest phenotypic classes in Table 7.2. However, these double recombinants, *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.

loci B,C R.F. = $\frac{5+16+12+5+2(1)+2(1)}{120}$ = 35%
(corrected for double recombinants)

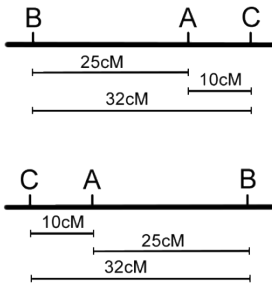


Figure 7.13
Equivalent maps based on the data in Table 7.2 (without correction for double crossovers). (Original-Deyholos-CC:AN)

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, for example double crossovers between loci A,B or between loci A,C. Geneticists have developed a variety of mathematical procedures to try to correct for things like double crossovers during large-scale mapping experiments.

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 (e.g. A) and a mutant allele (e.g. 'a').

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.
- 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.
- Because recombination frequency is usually proportional to the distance between loci, 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 genetically indistinguishable from non-recombinants.
- Three-point crosses can be used to determine the order and map distance between of three loci, and can correct for some of the double crossovers between the two outer loci.

KEY TERMS

linkage
recombination
independent assortment
crossover
recombinant genotype
parental genotype
unlinked
recombination frequency (RF)
complete (absolute) linkage
incomplete (partial) linkage
syntenic
synapsis
coupling (cis) configuration
repulsion(trans) configuration
repulsion
map units (mu)
centiMorgans (cM)
genetic map
conserved synteny
double-crossover
three-point cross

STUDY QUESTIONS

7.1 Compare recombination and crossover. How are these similar? How are they different?

7.2 Explain why it usually necessary to start with pure-breeding lines when measuring genetic linkage by the methods presented in this chapter.

7.3 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?

7.4 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.

7.5 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?

7.6 Imagine the white flowers are recessive to purple flowers, and yellow seeds are recessive to green seeds. If a green-seeded, purple-flowered dihybrid is testcrossed, and half of the progeny have yellow seeds, what can you conclude about linkage between these loci? What do you need to know about the parents of the dihybrid in this case?

7.7 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. Outline the crosses similar to what is shown in Figure 7.8, and specify which progeny will be considered recombinant. You do not need to calculate recombination frequency.

7.8 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. Outline the crosses similar to what is shown in Figure 7.8, and specify which progeny will be considered recombinant. You do not need to calculate recombination frequency.

7.9 Image 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.

7.10 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?

AaBb	135
Aabb	430
aaBb	390
aabb	120

7.11 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*?

7.12 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?

7.13 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?

7.14 In a three-point cross, individuals *AAbbcc* and *aaBBCC* are crossed, and their F_1 progeny is testcrossed. Answer the following questions based on these F_2 frequency data.

aaBbCc	480
AaBbcc	15
AaBbCc	10
aaBbcc	1
aabbCc	13
Aabbcc	472
AabbCc	1
aabbcc	8

a) Without calculating recombination frequencies, determine the relative order of these genes.
b) Calculate pair-wise recombination frequencies (without considering double cross overs) and produce a genetic map.
c) Recalculate recombination frequencies accounting for double recombinants.

7.15 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.

Chapter 8 TECHNIQUES OF MOLECULAR GENETICS

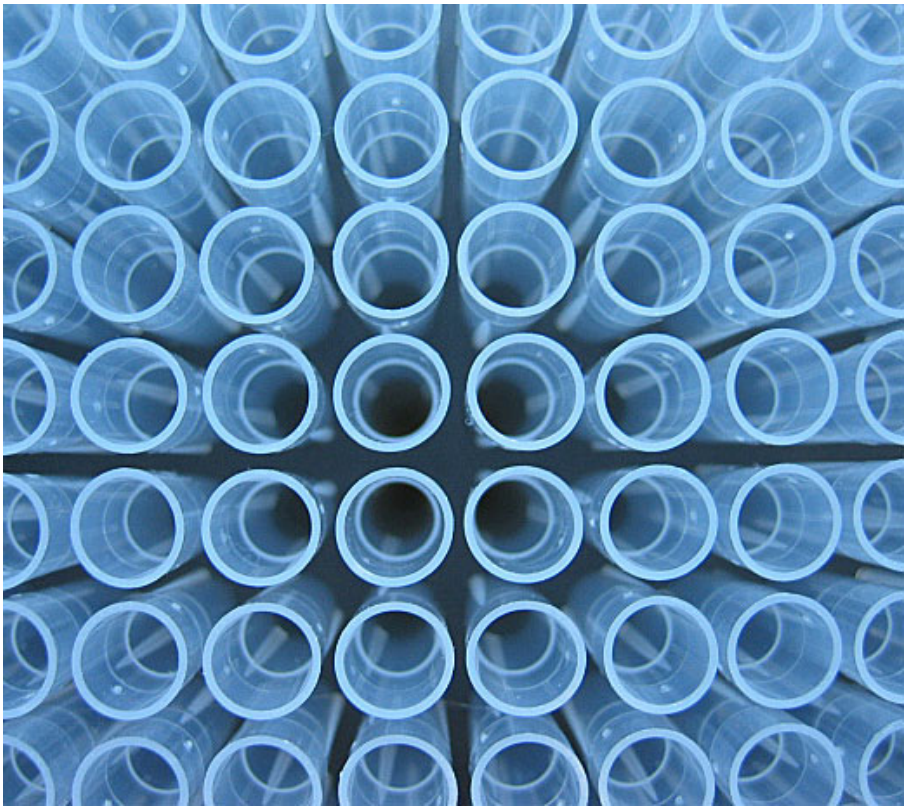


Figure 8.1
Disposable tips for a pipette are used to distribute microliter volumes of liquid in molecular biology. (Flickr-estherase-CC:ANS)

Genetics is the study of the inheritance and variation of biological traits. We have previously noted that it is possible to conduct genetic research without directly studying DNA. Indeed some of the greatest geneticists had no special knowledge of DNA at all, but relied instead on analysis of phenotypes, inheritance patterns, and their ratios in carefully designed crosses. 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 and analyze the DNA or RNA transcribed 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.

8.1 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 8.2). 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 chromatin into the extraction buffer, which also contains proteins and all other cellular components. (The basics of this procedure can be done with household chemicals and are 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; if we want to further isolate a specific gene or DNA fragment, we must use additional techniques, such as PCR.

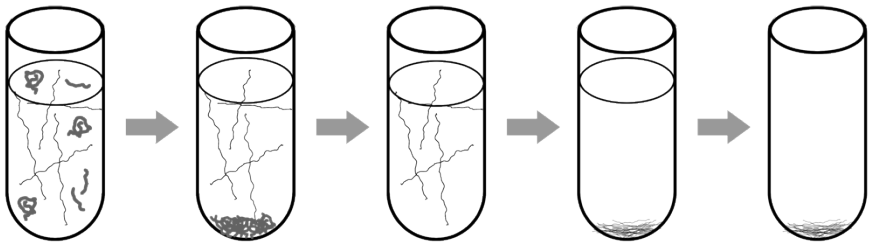


Figure 8.2
Extraction of DNA from a mixture of solubilized cellular components by successive precipitations. Proteins are precipitated, then DNA (in the supernatant) is precipitated in ethanol, leaving a pellet of DNA.
(Original-Deyholos-CC:AN)

8.2 Isolating or Detecting a Specific Sequence by PCR

8.2.1 Components of the PCR Reaction

The **Polymerase Chain Reaction (PCR)** is a method of DNA replication that is performed in a test tube (i.e. *in vitro*). Here "polymerase" refers to a DNA polymerase enzyme extracted and purified from bacteria, and "chain reaction" refers to the ability of this technique produce millions of copies of a 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 DNA.

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 8.3).

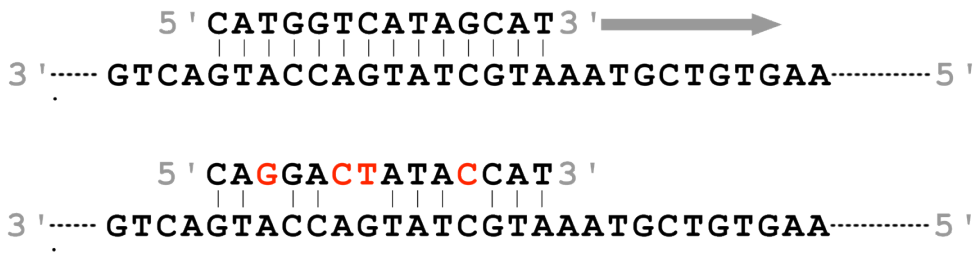


Figure 8.3 The primer-template duplex at the top part of the figure is perfectly matched, and will be stable at a higher temperature than the duplex in the bottom part of the figure, which contains many mismatches and therefore fewer hydrogen bonds. If the annealing temperature is sufficiently high, only the perfectly matched primer will be able to initiate extension (grey arrow) from this site on the template. (Original-Deyholos-CC:AN)

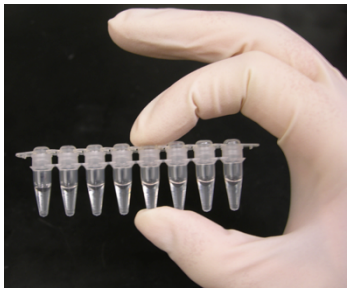


Figure 8.4
A strip of PCR tubes (Wikipedia-madprime-GFDL)

The experimenter can therefore control exactly what region of a DNA template is amplified by controlling the sequence of the primers used in the reaction.
To conduct a PCR amplification, an experimenter combines in a small, thin-walled tube (Figure 8.4), all of the necessary components for DNA replication, including DNA polymerase and solutions containing nucleotides (dATP, dCTP, dGTP, dTTP), a DNA template, DNA primers, a pH buffer, and 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 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.

An essential aspect of PCR is **thermalcycling**, meaning the exposure of the reaction to a series of precisely defined temperatures (Figure 8.5). 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 8.6). 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

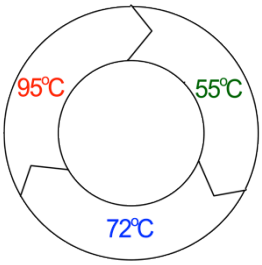


Figure 8.5
Example of a thermalcycle, in which the annealing temperature is 55°C.
(Original-Deyholos-CC:AN)

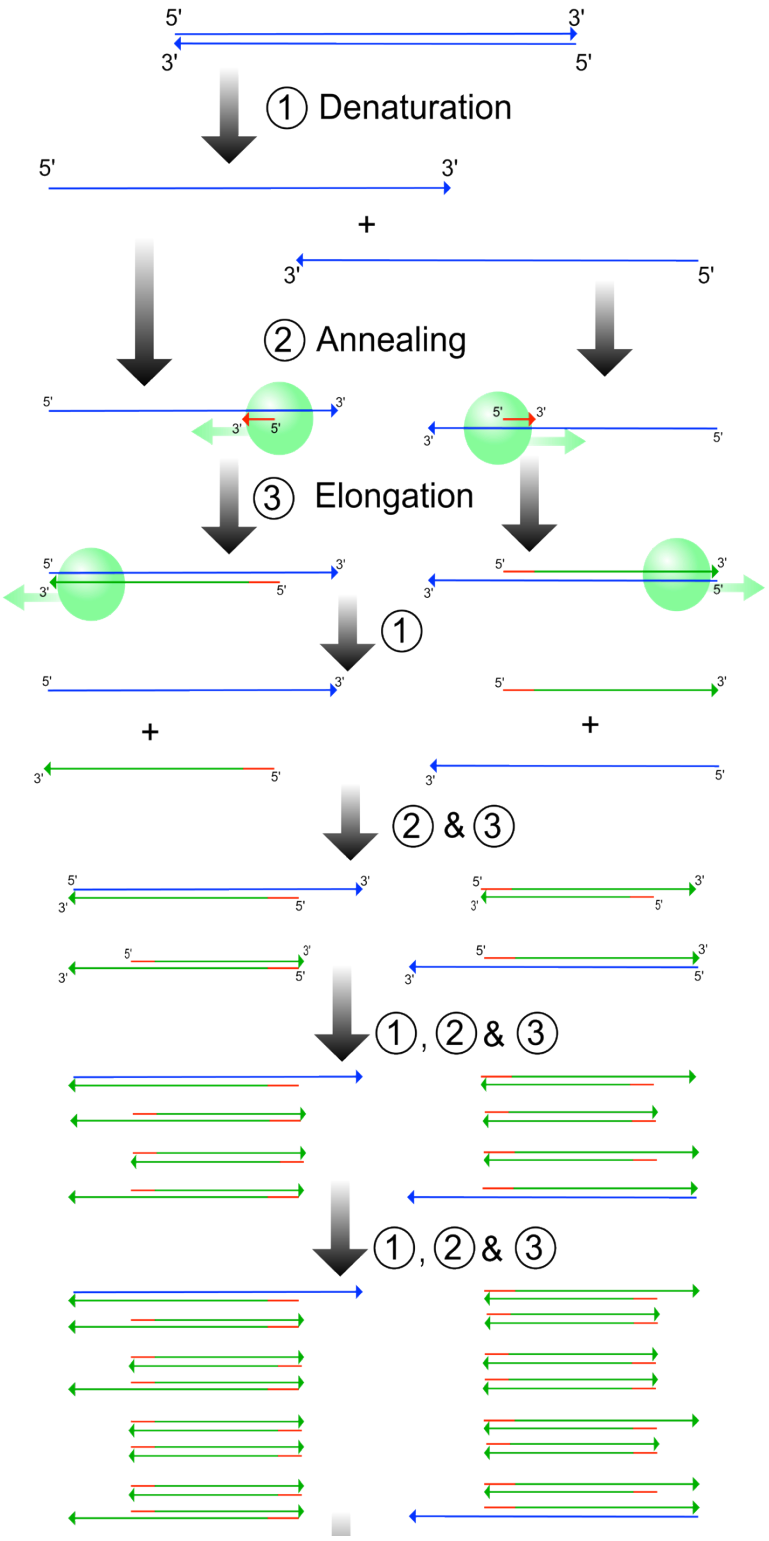


Figure 8.6
PCR with the three phases of the thermalcycle numbered. The template strand (blue) is replicated from primers (red), with newly synthesized strands in green. The green strands flanked by two primer binding sites will increase in abundance exponentially through successive PCR cycles.
(Wikipedia-madprime-GFDL)

2^{30}-fold. 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.

The earliest PCR reactions used a polymerase from *E. coli*. Because the high temperature of the denaturation step destroyed the enzyme, new polymerase had to be added after each cycle. To overcome this, researchers identified **thermostable** DNA polymerases such as **Taq DNA pol**, from *Thermus acquaticus*, a thermophilic bacterium that lives in hot springs. Taq, and similar thermostable polymerases from other hot environments, are able to remain functional in the repeated cycles of amplification. Taq polymerase cannot usually amplify fragments longer than about 3kbp, but under some specialized conditions, PCR can amplify fragments up to approximately 10kbp. Other polymerases, either by themselves or in combination with Taq, are used to increase the length of amplified fragments or to increase the fidelity of the replication.

After completion of the thermalcycling (amplification), an aliquot from the PCR reaction is usually loaded onto an **electrophoretic gel** (described below) 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.

8.2.3 An Application of PCR: the StarLink Affair

PCR is very sensitive (meaning it can amplify very small starting amounts of DNA), and specific (meaning it can amplify only the target sequence from a mixture of many DNA sequences). This made PCR the perfect 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 to the unauthorized gene (*Cry9C*). Their suspicions were confirmed when they ran this PCR product on an agarose gel and saw a band of expected size. The PCR test was able 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.

8.3 Cutting and Pasting DNA: Restriction Digests and DNA Ligation

8.3.1 Restriction Enzymes

Figure 8.7
An *EcoRI* dimer (blue, purple) sits like a saddle on a double helix of DNA (one strand is green, one is brown). This image is looking down the center of the helix.
(NCBI-?-PD)

Many bacteria have enzymes that recognize specific DNA sequences (usually 4 or 6 nucleotides) and then cut the double stranded DNA helix at this sequence (Figure 8.7). These enzymes are called site-specific **restriction endonucleases**, or more simply "**restriction enzymes**", and they naturally function as part of bacterial defenses against viruses and other sources of foreign DNA. To cut DNA at known locations, researchers use restriction enzymes that have been purified from various bacterial species, and which can be purchased from various commercial sources. These enzymes are usually named after the bacterium from which they were first isolated. For example, ***EcoRI*** and *EcoRV* are both enzymes from *E. coli*. *EcoRI* cuts double stranded DNA at the sequence GAATTC, but note that this enzyme, like many others, does not cut in exactly the middle of the restriction sequence (Figure 8.8). The ends of a molecule cut by *EcoRI* have an overhanging region of single stranded DNA, and so are sometimes called **sticky-ends**. On the other hand, *EcoRV* is an example of an enzyme that cuts both strands in exactly the middle of its recognition sequence, producing what are called **blunt-ends**, which lack overhangs.

Figure 8.8
The recognition sequence for *EcoRI* (blue) is cleaved by the enzyme (grey). This particular enzyme cuts DNA at a position offset from the center of the restriction site. This creates an overhanging, sticky-end.
(Original-Deyholos-CC:AN)

8.3.2 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**. Sticky-ended molecules with complementary overhanging sequences are said to have **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. Ligation is therefore central to the production of recombinant DNA, including the insertion of a double stranded DNA fragment into a plasmid vector.

8.4 Cloning DNA: Plasmid Vectors

8.4.1 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 high copy numbers 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 and drug-resistance. In the lab, plasmids can be inserted into bacteria in a process called **transformation**.

8.4.2 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 (Figure 8.9). Given

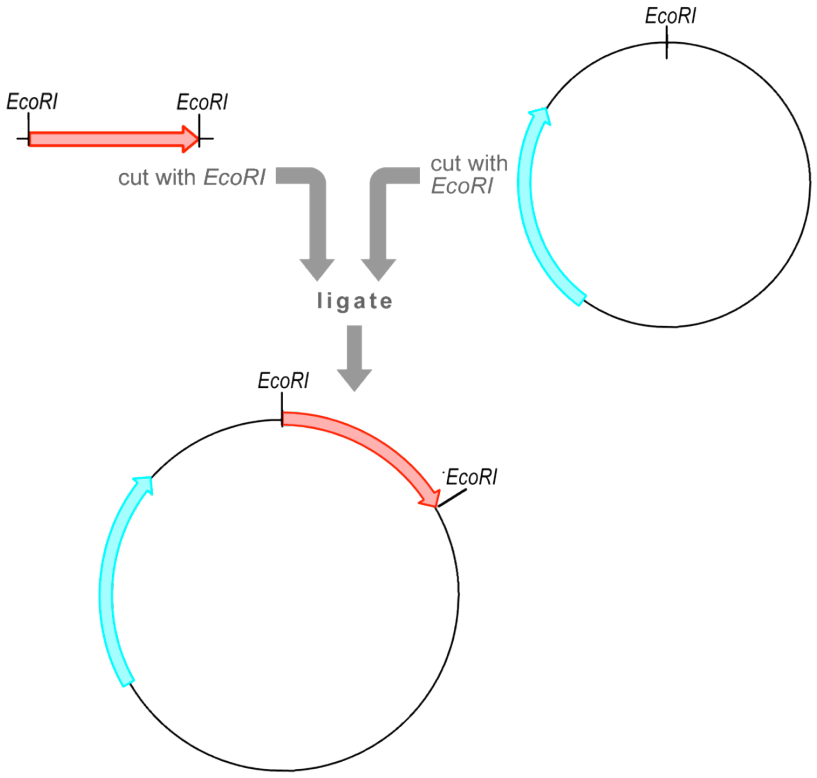


Figure 8.9
Cloning of a DNA fragment (red) into a plasmid vector. The vector already contains a selectable marker gene (blue) such as an antibiotic resistance gene.
(Original-Deyholos-CC:AN)

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.

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. Competent cells can be made 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. 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. This 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** (i.e. copy) a gene into a plasmid, then transform this recombinant plasmid back 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 are not sensitive enough to work with just a single molecule at a time. Many molecular cloning and recombination experiments are therefore iterative processes in which:

1. a DNA fragment (usually isolated by PCR and/or restriction 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

Figure 8.10
A vial of insulin. Note that the label lists the origin as "rDNA", which stands for recombinant DNA. (Flickr-DeathByBokeh-CC:AN)

8.4.3 An Application of Molecular Cloning: Production of 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 insulin identical to the human hormone 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 8.10), i.e. human genes and their derivatives expressed in bacteria or yeast.

8.5 DNA Analysis: Gel Electrophoresis

A solution of DNA is colorless, and except for being viscous at high concentrations, is visually indistinguishable from water. Therefore, techniques such as **gel electrophoresis** have been developed to detect and analyze DNA (Figure 8.11). This analysis starts when a solution of DNA is deposited at one end of a gel slab. This gel is made from polymers such as **agarose**, which is a polysaccharide isolated from seaweed. The DNA is then forced through the gel by an electrical current, with DNA molecules moving toward the positive electrode (Figure 8.12).

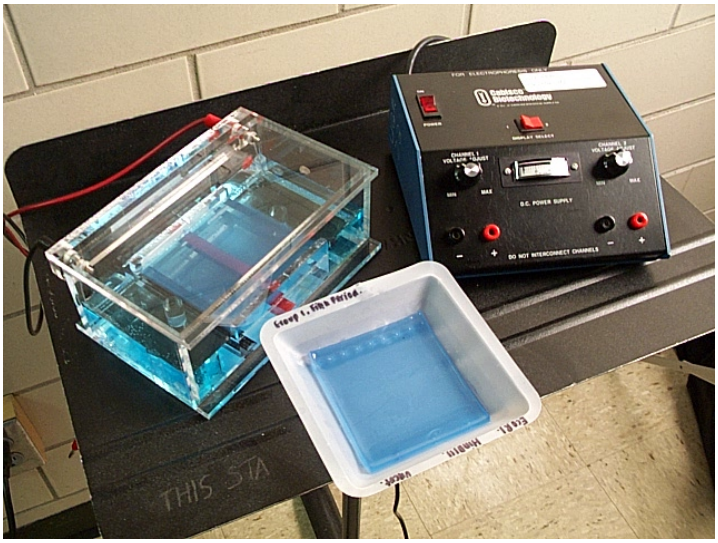


Figure 8.11
Apparatus for agarose gel electrophoresis. A waterproof tank is used to pass current through a slab gel, which is submerged in a buffer in the tank. The current is supplied by an adjustable power supply. A gel (stained blue by a dye sometimes used when loading DNA on the gel) sits in a tray, awaiting further analysis, such as photography under a UV light source. (Flickr-457088634_585df11af5_o-pending)

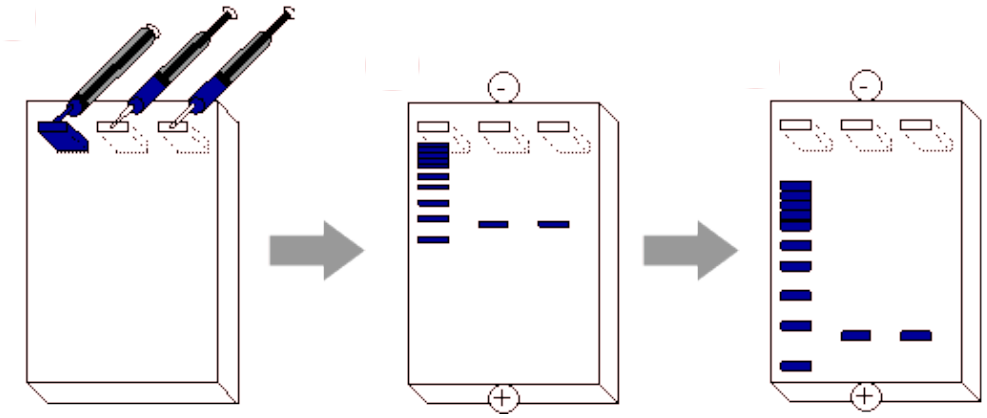


Figure 8.12
Agarose gel electrophoresis. DNA is loaded into wells at the top of a gel. A current is passed through the gel, pulling DNA towards the positively charged electrode. The DNA fragments are separated by size, with smaller fragments moving fastest towards the electrode. (Wikipedia-Magnus Manske_PD)

As it migrates, each piece of DNA threads its way through the pores, which form between the polymers in the gel. Because shorter pieces can move through these pores faster than longer pieces, gel electrophoresis separates molecules based on their size (length), with smaller DNA pieces moving faster than long ones. DNA

molecules of a similar size migrate to a similar location in each gel, called a **band**. This feature makes it easy to see DNA after staining the DNA with a fluorescent dye such as **ethidium bromide** (Figure 8.13). By separating a mixture of DNA molecules of known size (**size markers**) in adjacent lanes on the same gel, the length of an uncharacterized DNA fragment can be estimated. Gel segments containing the DNA bands can also be cut out of the gel, and the size-selected DNA extracted and used in other types of reactions, such as sequencing and cloning.

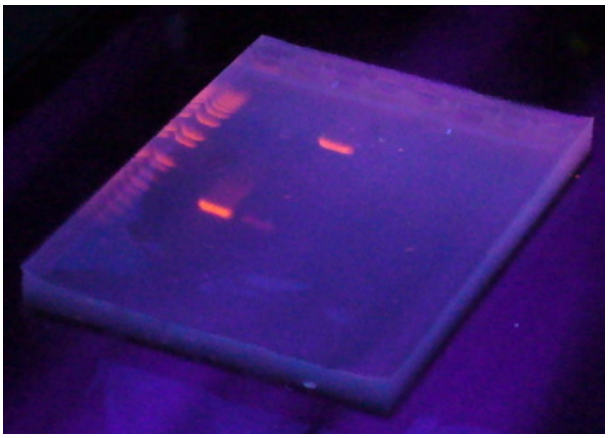


Figure 8.13
An agarose gel stained with ethidium bromide and illuminated by UV light. The stain associated with DNA is fluorescent. (Wikipedia-Transcontrol-GFDL)

8.6 DNA Analysis: Blotting and Hybridization

Bands of DNA in an electrophoretic gel form only if most of the DNA molecules are of the same size, such as following a PCR reaction, or restriction digestion of a plasmid. In other situations, such as after restriction digestion of chromosomal (genomic) DNA, there will be a large number of variable size fragments in the digest and it will appear as a continuous smear of DNA, rather than distinct bands. In this case, it is necessary to use additional techniques to detect the presence of a specific DNA sequence within the smear of DNA separated on an electrophoretic gel. This can be done using a "Southern Blot".

8.6.1 Southern Blots

A **Southern blot** (also called a **Southern Transfer**) is named after Ed Southern, its inventor. In the first step, DNA is digested with restriction enzymes and separated by gel electrophoresis (as discussed above). Then a sheet or **membrane** of nylon or similar material is laid under the gel and the DNA, in its separated position (bands or smear), is transferred to the membrane by drawing the liquid out of the gel, in a process called **blotting** (Figure 8.14). The blotted DNA is usually covalently attached to the nylon membrane by briefly exposing the blot to UV light. Transferring the DNA to the sturdy membrane is necessary because the fragile gel would fall apart during the next two steps in the process. Next, the membrane is bathed in a solution to **denature** (double stranded made single stranded) the attached DNA. Then a **hybridization** solution containing a small amount of single-stranded **probe** DNA that is complementary in sequence to a target molecule on the membrane. This probe DNA is labeled using fluorescent or radioactive molecules, and if the hybridization is performed properly, the probe DNA will form a stable duplex only with those DNA molecules on the membrane that are exactly complementary to it.

Then, the unhybridized probe is washed off and remaining radioactive or fluorescent signal will appear in a distinct band when appropriately detected. The band represents the presence of a particular DNA sequence within the mixture of DNA fragments.

The probe is sequence specific (requires complementarity). However, variation in hybridization temperature and washing solutions can alter the **stringency** of the probe. At maximum stringency (higher temperature) hybridization conditions, probes will only hybridize with the exact target sequences that are perfectly complementary (maximum number of hydrogen bonds). At lower temperatures, probes will be able to hybridize to targets to which they do not match exactly, but only are roughly complementary for part of the sequence.

Southern blotting is useful not only for detecting the presence of a DNA sequence within a mixture of DNA molecules, but also for determining the size of a restriction fragment in a DNA sample. Southern blots are useful for detecting fragments larger than those normally amplified by PCR, and when trying to detect fragments that may be only distantly related to a known sequence. Applications of Southern blotting will be discussed further in the context of molecular markers in a subsequent chapter. Southern blotting was invented before PCR, but PCR has replaced blotting in many applications because of its simplicity, speed, and convenience.

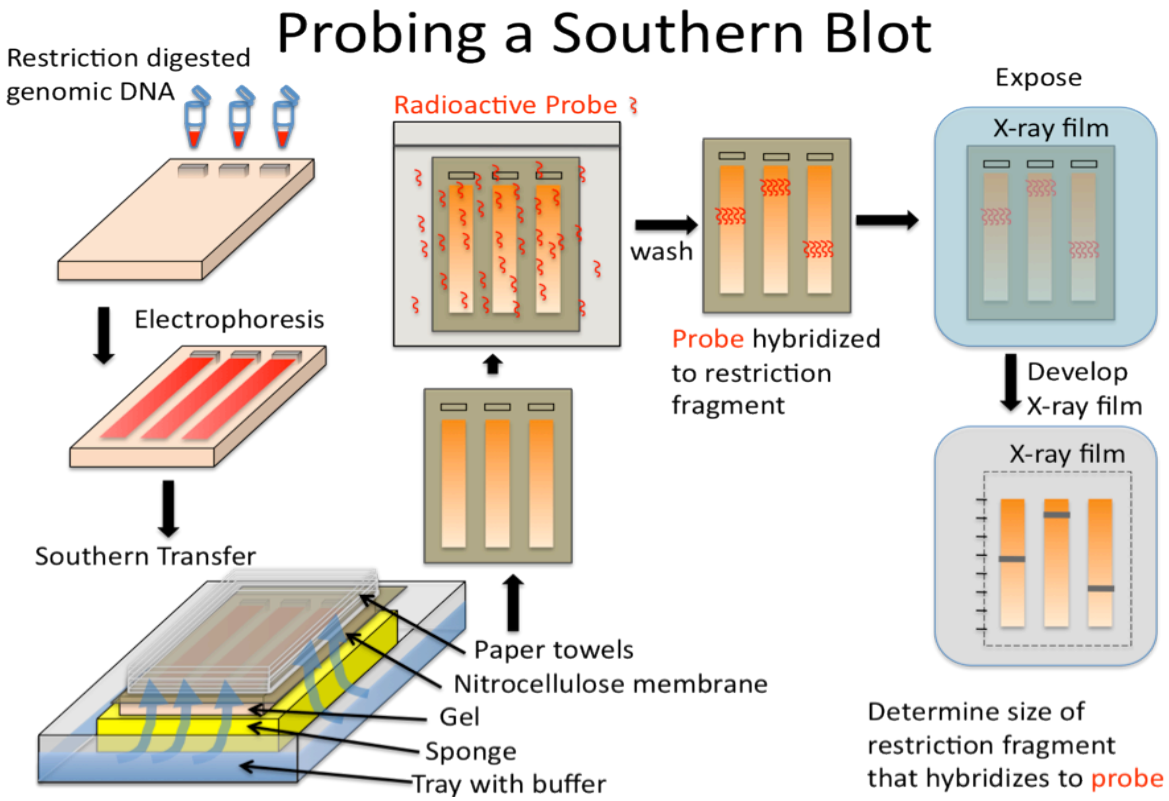
8.6.2 Other Blotting Based Techniques

Following the development of the Southern blot, other types of blotting techniques were invented.

The **Northern blot** involves the size separation of RNA in gels like that of DNA. Because we wish to determine the native size of the RNA transcript (and because RNA is single stranded) no restriction enzymes are ever used. Because most RNA is single stranded and can fold into various conformations thorough intra-molecular base pairing, the electrophoresis separation is more haphazard and the bands are often less sharp, compared to that of double stranded DNA.

In a **Western blot,** protein is size separated on a gel (usually an acrylamide gel) before transferring to a membrane, which is then probed with an antibody that specifically binds to an antigenic site on the target protein. This antibody is then detected by other antibodies with some fluorescent or colour production marker system. It will also give bands proportional to the amount and size of the target protein.

Figure 8.14 A diagram of Southern blotting. Genomic DNA that has been digested with a restriction enzyme is separated on an agarose gel, then the DNA is transferred from the gel to a nylon membrane (grey sheet) by blotting. The DNA is immobilized on the membrane, then probed with a radioactively labeled DNA fragment that is complementary to a target sequence. After stringent washing, the blot is exposed to X-ray film to detect what size frament the probe is bound. In this case, the probe bound to different-sized fragments in lanes 1, 2, and 3. In the last image the orange represent the position of the digested DNA, but it is not actually present on the X-ray film. (Original-J. Locke-CC:AN)



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
- Restriction enzymes are natural endonucleases used in molecular biology to cut DNA sequences at specific sites.
- DNA fragments with compatible ends can be joined together through ligation. If the ligation produces a sequence not found in nature, the molecule is said to be recombinant.
- Transformation is the introduction of DNA (usually recombinant plasmids) into bacteria.
- Cloning of genes in *E. coli* is a common technique in molecular biology, since it allows large quantities of a DNA for gene to made, which allows further analysis or manipulation
- Cloning can also be used to produce useful proteins, such as insulin, in microbes.
- Southern blotting can be used to detect the presence of any sequence that matches a probe, within a mixture of DNA (such as total genomic DNA).
- The stringency of hybridization in blotting and in PCR is dependent on physical parameters such as temperature and washing solution content.

KEY TERMS:

macromolecules
lysis
detergent
chelating agent
EDTA
nuclease
pellet
PCR
primer
thermalcycle
denature
anneal
extend
Taq DNApol
electrophoretic gel
restriction endonuclease
EcoRI
sticky end
blunt end
compatible end
ligation
ligase
plasmid
transformation
competent
electroporation
selectable marker
agarose
vector
band
ethidium bromide
Southern blot
membrane
denaturation
hybridization
washing
probe
stringency
northern blot
western blot

STUDY QUESTIONS:

8.1 What information, and what reagents would you need to use PCR to detect HIV in a blood sample?

8.2 A 6.0 kbp PCR fragment flanked by recognition sites for the *HindIII* restriction enzyme is cut with *HindIII* then ligated with a 3kb plasmid vector that has also been cut with *HindIII*. This recombinant plasmid is transformed into *E. coli.* From one colony a plasmid is prepared and digested with HindIII.

a) When the product of the *HindIII* digestion is analyzed by gel electrophoresis, what will be the size of the bands observed?

b) What bands would be observed if the recombinant plasmid was cut with *EcoRI*, which has only one site, directly in the middle of the PCR fragment?

c) What bands would be observed if the recombinant plasmid was cut with both *EcoRI* and *HindIII* at the same time?

8.3 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?

8.4 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?

8.5 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 fatal disease, due to contamination with live viruses. How could you use molecular biology to produce an optimal vaccine?

8.6 How would cloning be different if there were no selectable markers?

8.7 Research shows that a particular form of cancer is caused by a 200bp deletion in a particular human gene that is normally 2kb long. Only one mutant copy is needed to cause the disease.

a) Explain how you would use Southern blotting to diagnose the disease.

b) How would any of the blots appear if you hybridized and washed at very low temperature?

8.8 Refer to question 8.7.

a) Explain how you would detect the presence of the same deletion using PCR, rather than a Southern blot.

b) How would PCR products appear if you annealed at very low temperature?

8.9 You have a PCR fragment for a human olfactory receptor gene (perception of smells). You want to know what genes a dog might have that are related to this human gene. How can you use your PCR fragment and genomic DNA from a dog to find this out? Do you think dogs have more or less of these genes?

8.10 You add ligase to a reaction containing a sticky-ended plasmid and sticky-ended insert fragment, which both have compatible ends. Unbeknownst to you, someone in the lab left the stock of ligase enzyme out of the freezer overnight and it degraded (no longer works). Explain in detail what will happen in your ligation experiment in this situation should you try and transform with it.

Chapter 9
CHANGES IN CHROMOSOME NUMBER & STRUCTURE

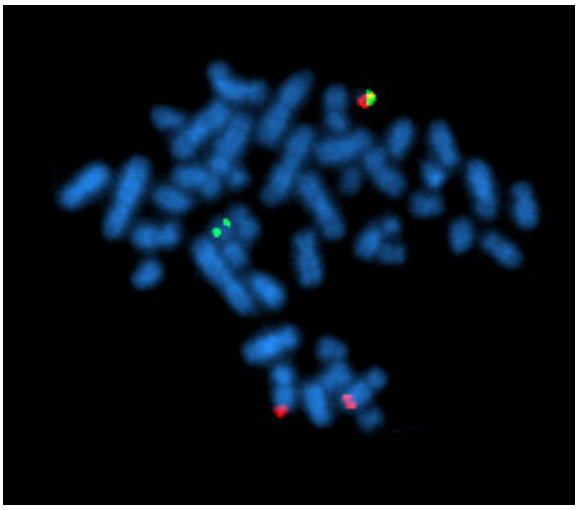


Figure 9.1
Fluorescence in situ hybridization of mitotic chromosomes from a human cancer cell. (Wikipedia-Pmx-CC:AS)

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 chromosome has features that allow it to be passed on during cell division. **Origins of replication** found along its length provide places for DNA replication to start, **telomeres** protect each end of the chromosome, and a single **centromere** near the middle provides a place for microtubules to attach and move the chromosome during mitosis and meiosis.

This chapter examines (1) changes in the number of whole chromosomes and how they affect the phenotype of an organism and (2) changes in the structure of individual chromosomes and how they affect meiotic pairing. Human examples will be used to show the phenotypic consequences and methods for detection.

9.1 CHANGES IN CHROMOSOME NUMBER

If something goes wrong during cell division, an entire chromosome may be lost and the cell will lack all of these genes. The causes behind these chromosome abnormalites and the consequences they have for the cell and the organism is the subject of this section.

9.1.1 CAUSE: NONDISJUNCTION DURING MITOSIS OR MEIOSIS

Segregation occurs in anaphase. In mitosis and meiosis II, sister chromatids (of replicated chromosomes) are normally pulled to opposite ends of the cell (see Figure 2.12). In Meiosis I, it is homologous chromosomes, which are synapsed at that time, that segregate and move apart.

In rare cases, the sister chromatids (or paired chromosomes in Meiosis I) fail to separate, or dis-join. This failure to segregate properly is called **nondisjunction** and it can happen during mitosis, meiosis I, or meiosis II. This nondisjunction results in

both chromatids (or chromosomes) moving to one pole and none at the other. One cell will have an extra copy and the other will lack a copy. Thus failure to segregate properly leads to unbalanced products.

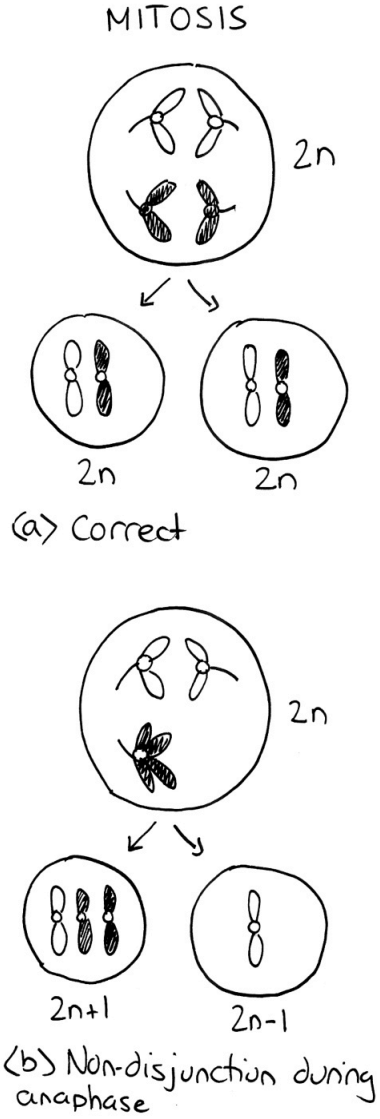


Figure 9.2
Mitosis done successfully (a) and unsuccessfully (b). The cell is diploid and the homologs of one chromosome are shown in white and black. (Original-Harrington-CC:AN)

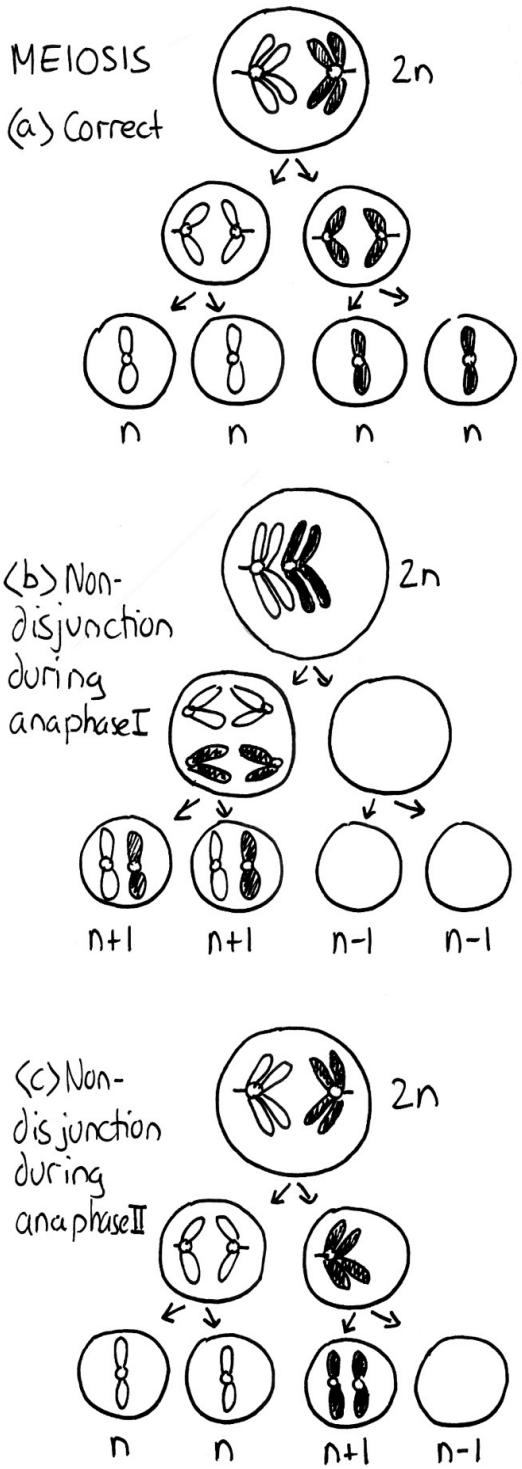


Figure 9.3
Meiosis done successfully (a) and unsuccessfully (b and c). (Original-Harrington-CC:AN)

9.1.2 Consequence: Decreased Viability

The result of a non-disjunction event is daughter cells that have an abnormal number of chromosomes. Cells, such as the parent cell in Figure 9.2a, which have the proper number of chromosomes, are said to be **euploid**. The daughter cells have one too many or one too few chromosomes and are thus **aneuploid**. Even though both product cells have at least one copy of all genes, both cells will probably die. The reason is due to the loss or gain of a large number of genes. Genes produce an standard amount of product - either functional RNAs or proteins. The parent cell shown has a **balanced** genotype because it has two copies of all of its genes. Each of its genes produces half of the products needed by the cell. But if one of these cells suddenly had only one copy (or three copies) of an important gene, the amount of product would be either 50% (or 150%) of what was required. Such a change for a single gene could probably be tolerated by the cell and it would probably survive. But the sudden change to one copy (or three copies) of the hundreds or thousands of genes on an entire chromosome the results would be more than tolerable and be catastrophic for the daughter cells. They have what's called an **unbalanced** genotype, which usually decreases their viability.

If a **first division** or **second division nondisjunction** event occurs during meiosis the result is an unbalanced gamete (Figure 9.3b and c). The gamete will often be functional, but after fertilization the embryo will be genetically unbalanced. This usually leads to the death of the embryo. There are some exceptions to this in humans and these will be presented later in this chapter.

9.2 Changes in Chromosome Structure

If the chromosome is altered, but still retains the three critical features of a chromosome (centromeres, telomeres, and origin of replication), it will continue to be inherited during subsequent cell divisions, however the daughter cell may not retain all the genes. For example, if a segment of the chromosome has been lost, the cell may be missing some genes. The causes of chromosome structural abnormalites and the consequences they have for the cell and the organism is shown below. They all involve breaks in the DNA that makes up the chromosome.

9.2.1 Cause #1: Incorrect Repair of Double Strand DNA Breaks During Interphase

A chromosome is a very long but very thin molecule. In the phopho-diester backbone there are only two covalent bonds holding each base pair to the next. If one of these covalent bonds is broken the chromosome will still remain intact, although a DNA Ligase will be needed to repair the nick (Figure 9.4a). 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 9.4b). 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 do not the result is a **chromosome rearrangement** (Figure 9.5).

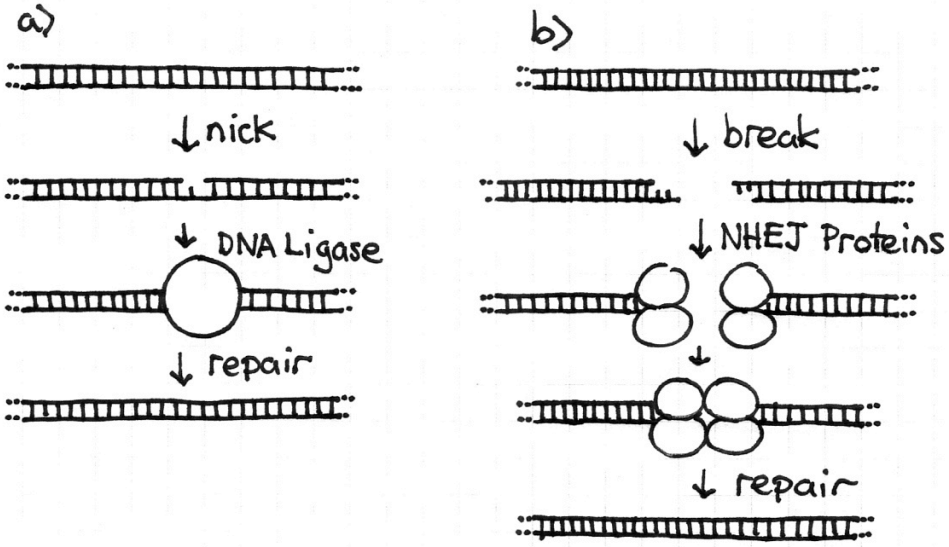


Figure 9.4 Repair of single strand nicks and double strand breaks in DNA. (Original-Harrington-CC:AN)

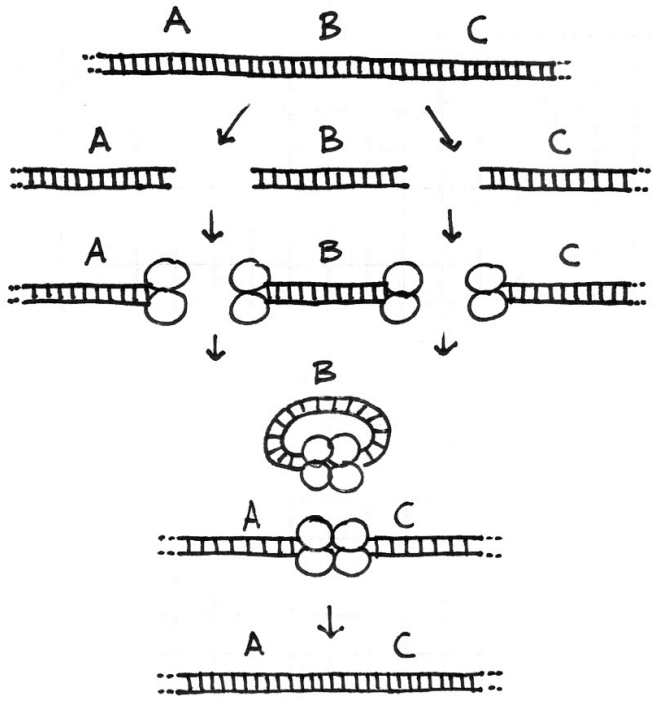
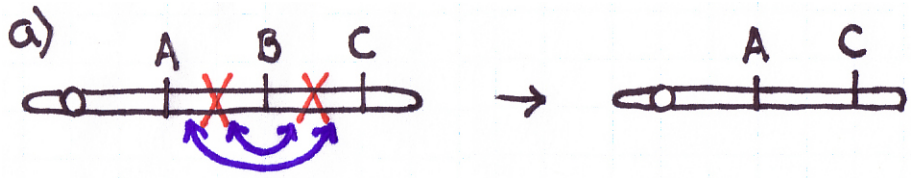


Figure 9.5 Errors during DNA repair can cause a chromosome deletion. In this diagram A, B, and C are genes on the same chromosome. As in Figure 9.4 there has been breaks in the DNA, recruitment of NHEJ proteins, and repair. After the repairs are completed the small piece of DNA with gene B is lost and the chromosome now only has genes A and C. (Original-Harrington-CC:AN)

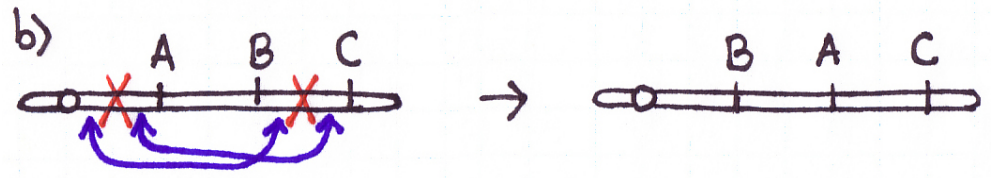
9.2.2 The Four Types of Chromosome rearrangements

Errors during the repair of multiple double strand breaks can cause four types of chromosome rearrangements. The type of chromosome rearrangement is dependent upon where the two breaks were originally and how they are rejoined. Figure 9.5 shows some possibilities but more are shown below. In these there is 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 unreplicated as they are in G_1 phase but these events can happen anytime during interphase.

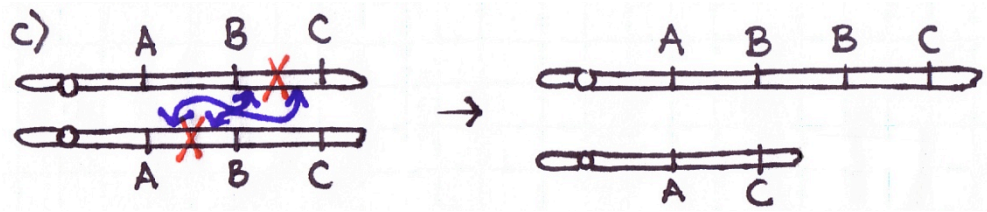
There are four major types of rearrangements:



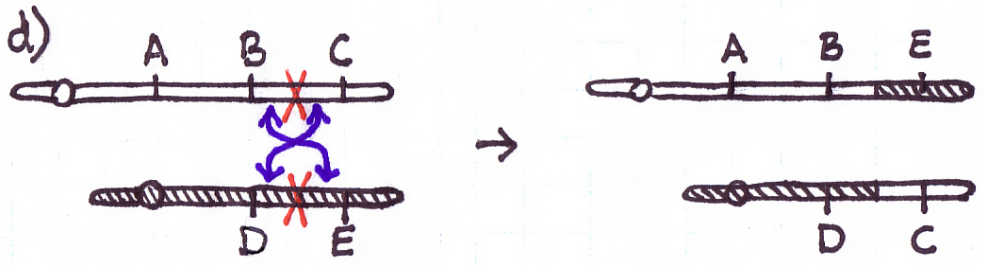
a) Deletions arise when both breaks are on one chromosome. If the ends are joined in this way the piece of DNA with the B gene on it does not have a centromere and will be lost during the next cell division.



b) Inversions also occur when both 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).



c) Duplications can occur from two DNA breaks at different places in sister chromatids (in a replicated chromosome). The ends are joined together incorrectly to give a chromosome with a duplication (two "B" regions as shown above). Note: the reciprocal product has a deletion.



d) 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.

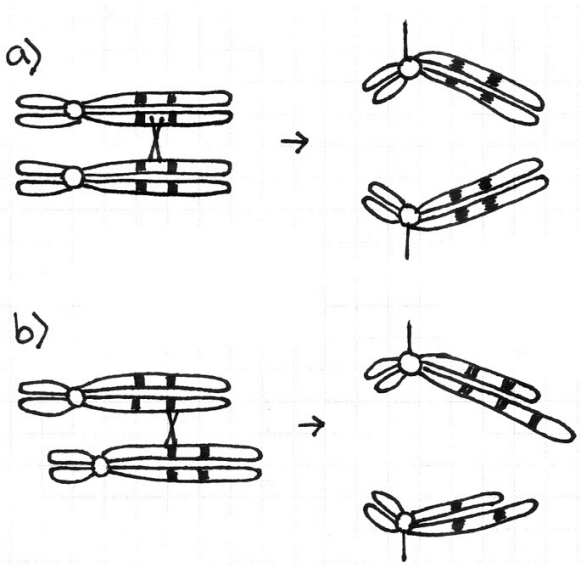


Figure 9.7 Errors during meiotic crossovers can cause duplications and deletions. This diagram shows homologous chromosomes pairing in prophase I and then separating in anaphase I. The shaded boxes are alu transposable elements. a) The homologous chromosomes pair properly, a crossover occurs, and all four chromatids in anaphase I are normal. b) The pairing is incorrect, a crossover occurs in the misspaired region, and in anaphase I one chromatid has a duplication and another has a deletion. (Original-Harrington-CC:AN)

9.2.3 Cause #2: Incorrect Crossovers During Meiosis

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 2). They also allow recombination to occur between linked genes (see Chapter 7). 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 9.7a). 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 9.7b). 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 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.

9.2.4 Consequence #1 - Rearrangements Show Abnormal Pairing at Meiosis

Homologous regions of chromosomes pair at meiosis I (prophase I). With rearranged chromosomes this can lead to visible abnormalities and segregation abnormalities.

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.

When an **inversion** chromosome is paired up in meiosis 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 9.8n). Similarly, with a **pericentric inversion**, a crossover event leads to duplicate/deletion products that are unbalanced (Figure 9.9n).

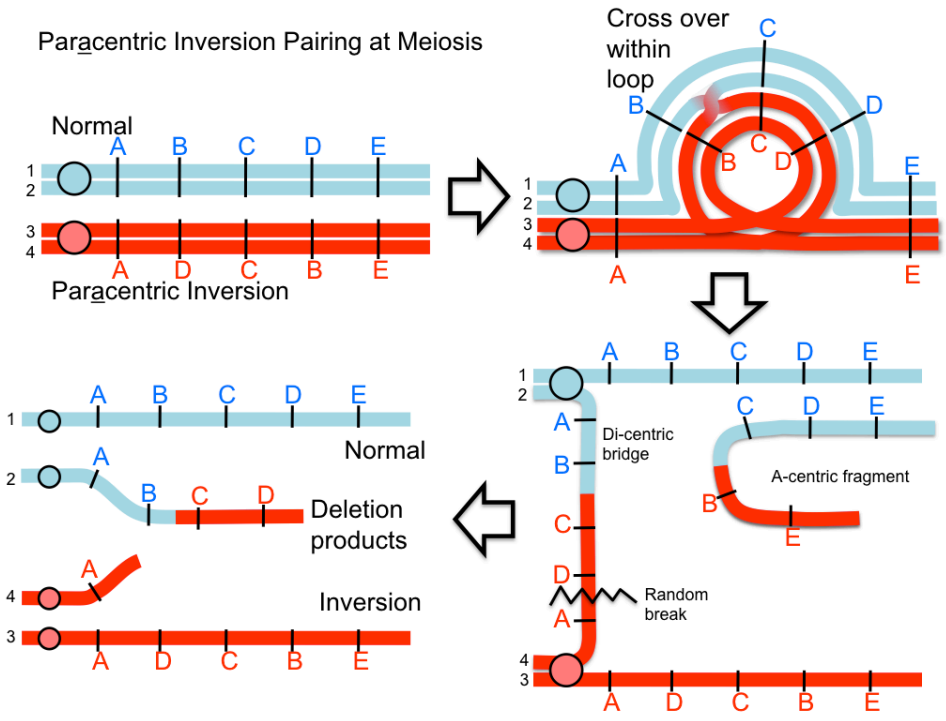


Figure 9.8n
A paracentric inversion pairing at meiosis. A crossover within the loop causes the production of an acentric and a dicentric chromatids which leads to deletion product..
(Original-Locke-CC:AN)

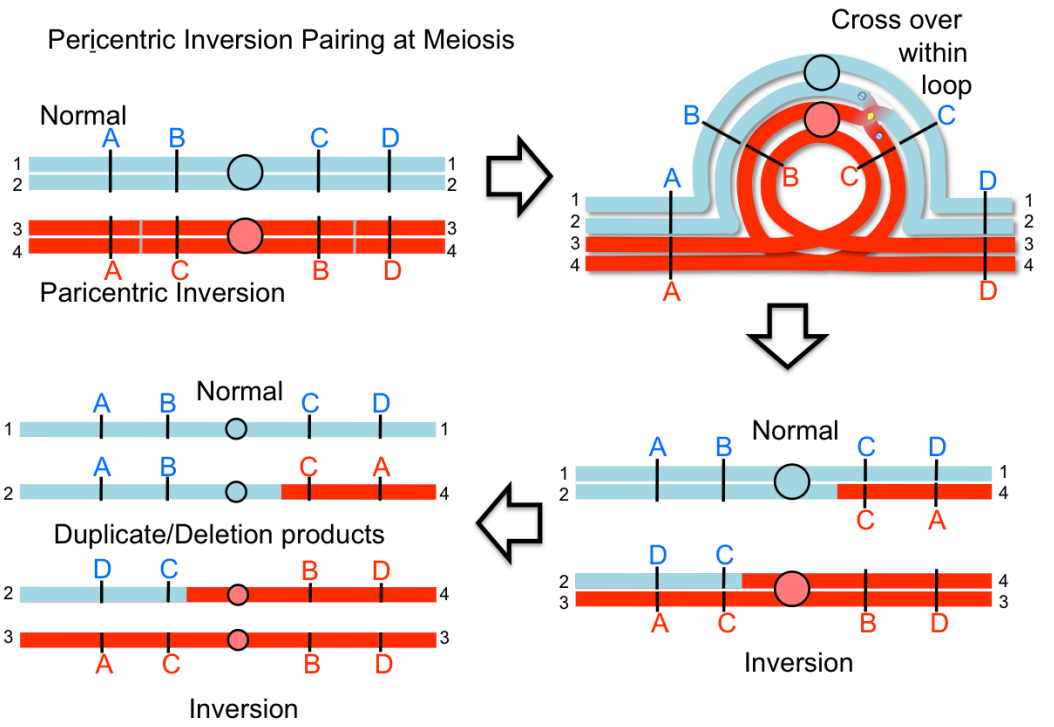


Figure 9.9n
A pericentric inversion pairing at meiosis. A crossover within the loop causes the production of duplicate and deletion products.
(Original-Locke-CC:AN)

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.
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.

For **translocations**, a consequence for the two chromosomes involved is that when they pair at meiosis both replicated chromosome pairs will be together, which can be seen cytologically as a **tetrad**. This tetrad can segregate in three ways. Two of which are shown below. This set of paired, replicated chromosomes can segregate as **Alternate** (balanced) where both normal and both translocated chromosomes go to the same polls. Or the chromosomes can segregate as **Adjacent-1** (unbalanced) where the normal and translocation chromosomes segregate as shown below. Each of these possibilities is approximately equally frequent and thus only about half the time do the gametes end up unbalanced (Figure 10n).

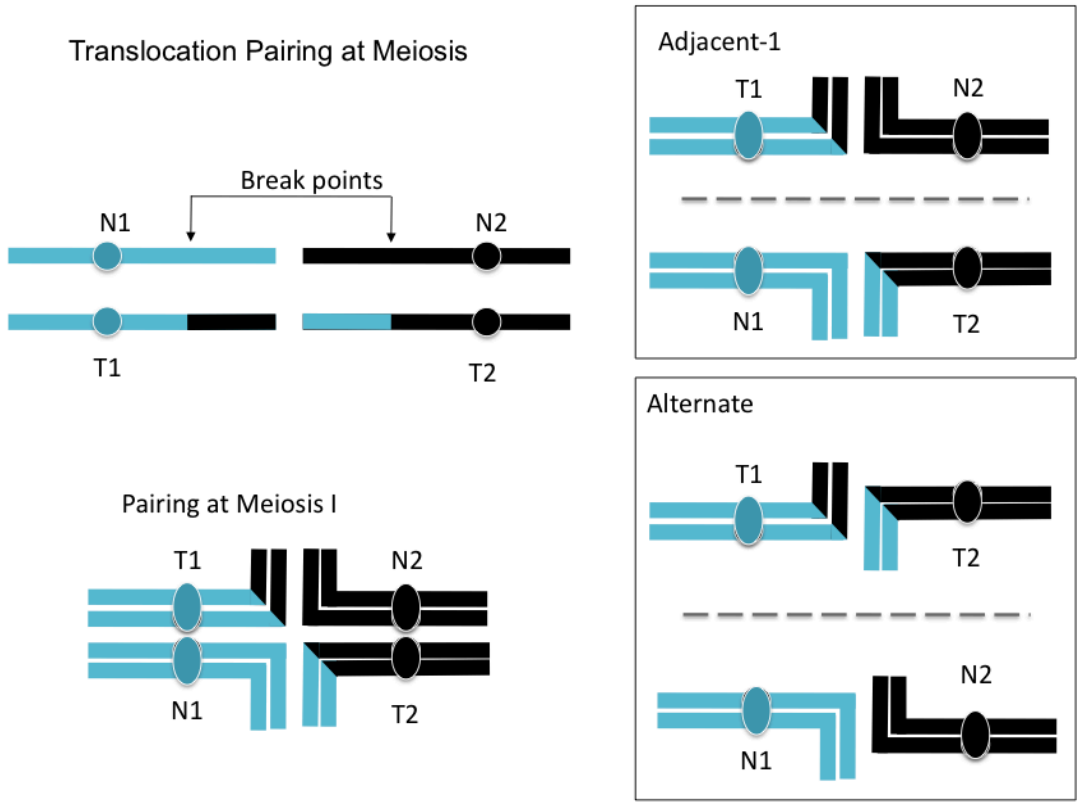


Figure 9.10n
A reciprocal translocation pairing at meiosis. There are two main avenues for segregation: Adjacent-1 and Alternate. Adjacent-1 results in duplication and deletion for part of the chromosome segments. Alternate doesn't. (Original-Locke-CC:AN)

9.2.5 Consequence #2: 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. As explained in Section 9.1.2 above 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.

9.2.6 Consequence #3: 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.

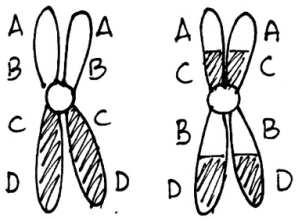


Figure 9.11n
An unrearranged chromosome (left) and a homolog with a pericentric inversion (right). (Original-Harrington-CC:AN)

Cells heterozygous for balanced rearrangements actually have more difficulties in prophase I. Consider the chromosomes shown in Figure 9.11n. There are different ways they might pair during prophase I - one is shown in Figure 9.12n. 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**.

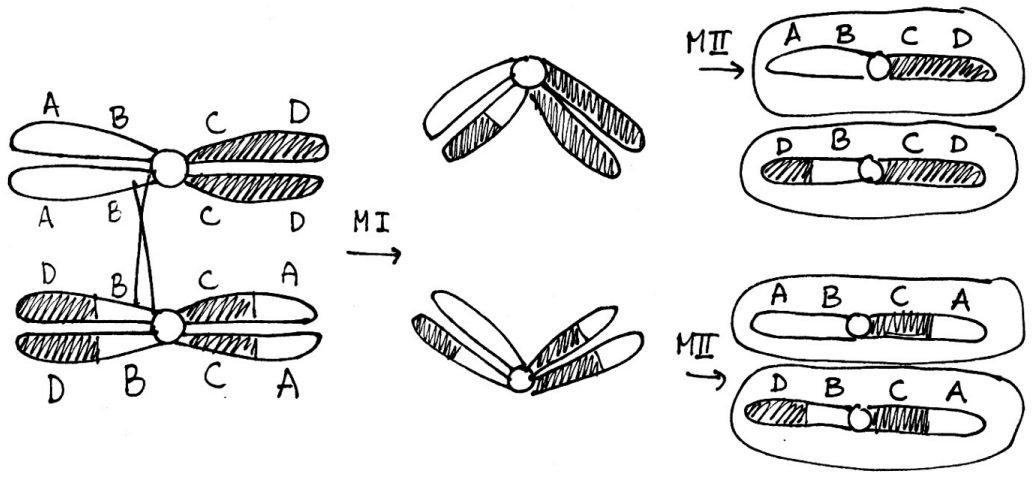


Figure 9.12n
Meiosis in a cell heterozygous for the chromosomes shown in Figure 9.11. Note that of the four gametes one has a deletion of the A gene and a duplication of the D gene while another gamete has a duplication of A and a deletion of D. (Original-Harrington-CC:AN)

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.

9.2.7 Consequence #4: 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. An example of this is shown in Figure 9.1 where the chromosomes are from a patient with leukemia caused by a translocation between chromosomes 9 and 22 (the red and green spots side-by-side).

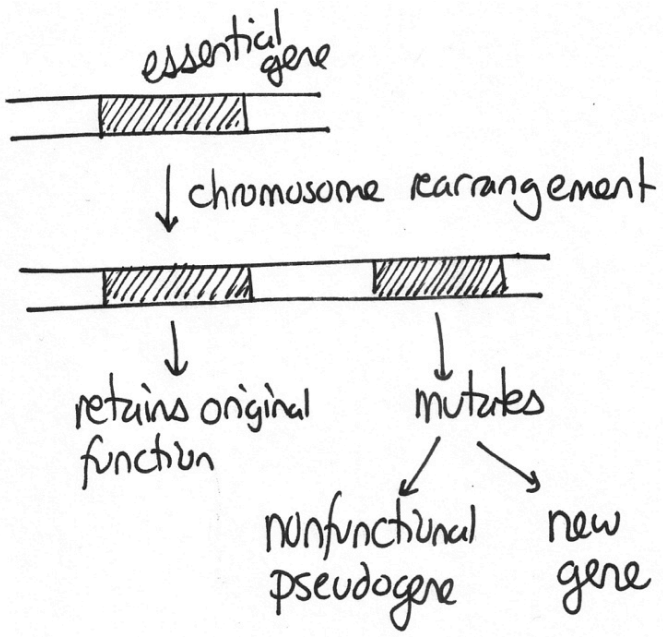


Figure 9.13n
Duplicated genes can mutate without compromising the viability of the organism. Occasionally the result is a new gene. (Original-Harrington-CC:AN)

9.2.8 Consequence #5: 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 9.13n). An example of this is the multiple copies of the globin genes found in mammals (see Figure 12.13).

Chromosome rearrangements that decrease fertility are also important for the origin of new species. If a rearrangement, such as the inversion shown in Figure 9.11n, 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.

9.3 Chromosome Abnormalities 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. As you will recall from Figure 2.12, humans are 2n=46. 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.

9.3.1 Down Syndrome

The most common chromosome number abnormality is trisomy-21 or, as it is more commonly known, **Down syndrome**. Having an extra copy of the smallest human chromosome, chromosome 21, causes substantial health problems.It is present in about 1 in 800 births. Infants with this condition have three copies of chromosome 21 rather than the normal two. Don't confuse **trisomy** - having three copies of one chromosome (i.e. 2n+1) with triploidy - having three entire chromosome sets (3x; see Section 2.6.) Females with trisomy-21 are 47,XX,+21 while males are 47,XY,+21. In general, people with Down syndrome are 47,sex,+21 where the word 'sex' signifies that the sex chromosomes may be XX or XY.

Trisomy-21 may arise from a nondisjunction event during meioisis in either parent or during mitosis very early during embryogenesis. However, most cases are due to a first division non-disjunction event occurring in the female parent (Figure 9.14n).

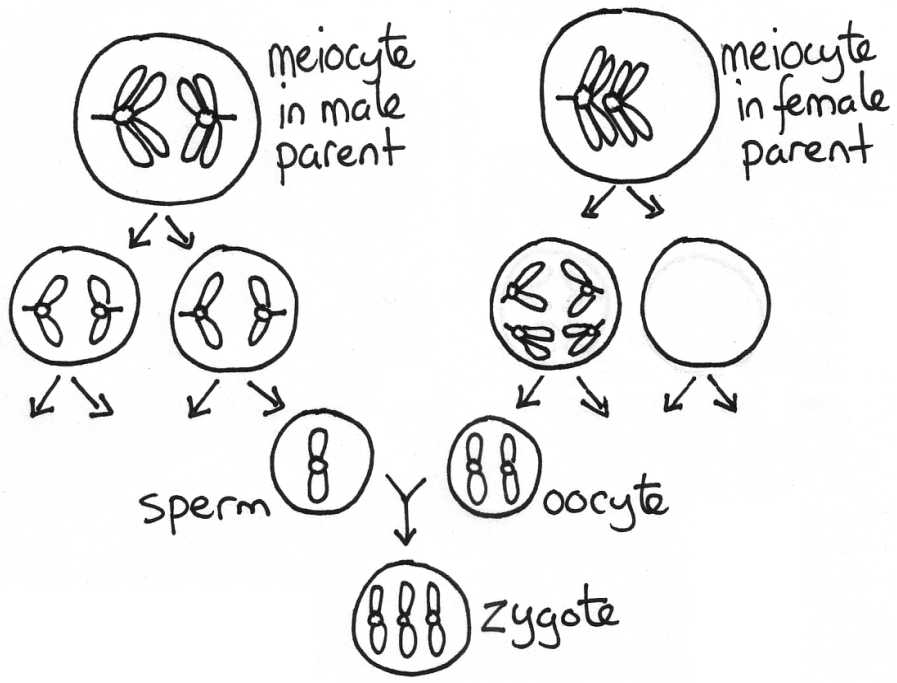


Figure 9.14n This diagram shows the errors during chromosome segregation that cause Down syndrome during meiosis in both parents and fusion of the gametes. Note that the cells that begin meiosis are called meiocytes and that this diagram only shows one of the four cells produced by meiosis. Meiosis occurred properly in the male parent but there was a nondisjunction event in the female parent in anaphase I. (Original-Harrington-CC:AN)

Having an extra copy of the smallest human chromosome, chromosome 21, causes substantial health problems. People with Down syndrome have various degrees of intellectual disability and often have other health problems such as heart defects. The disease was first described by John Down in 1866 but it was not until 1959 when its chromosomal basis was discovered. Current research suggests that at least some of the mental problems are due to having three copies of the *DYRK* gene on chromosome 21. This gene is active in the brain and there is evidence from humans and from mice that neurons are damaged if there is too much DYRK protein synthesized.

9.3.2 XYY AND XXX

While fetuses trisomic for one of the other larger autosomes seldom survive to term, the situation is quite different for the sex chromosomes. Approximately 1 in 1000 males have an extra Y chromosome and yet most are unaware of it! There is little harm in having two Y chromosomes because they have relatively few genes. Similarly, 1 in 1000 females have an extra X chromosome. This situation is also relatively harmless although for a different reason. Normally in female mammals one of the two X chromosomes is inactivated in each cell so that there can be genetic balance (Figure 9.15n and see Section 3.5.2). In 46,XX females one of the X chromosomes is inactivated while in 47,XXX females two are.

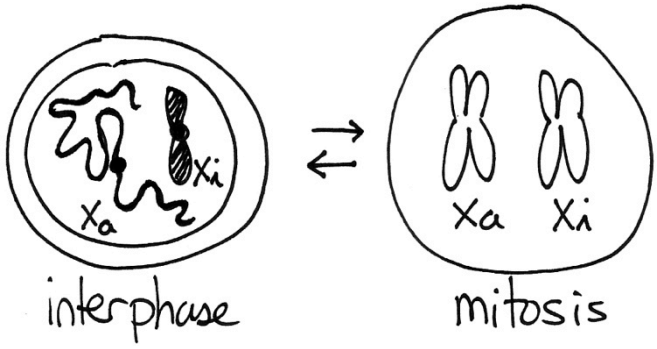


Figure 9.15n
A simplified view of dosage compensation in mammals. Because females have more X chromosomes than males in females one X chromosome is marked as inactive (X_i). It is replicated and transmitted during cell division but most of its genes are silent. It appears as a condensed mass called the Barr body within the interphase nucleus.
(Original-Harrington-CC:AN)

9.3.3 Turner Syndrome

Monosomy (2n-1) for autosomal chromosomes does occur at conception but these embryos almost never survive to term. Similarly, embryos that are 45,Y are also non-viable because they lack many essential genes found on the X chromosome. The only viable monosomy in humans is 45,X, also known as **Turner syndrome**. These people are phenotypically female because they do not have a Y chromosome (see Section 2.5.2). They are viable because in females only one X is active in most cells anyways. People with this condition do have health problems though: they are shorter than average, they have an elevated risk of heart defects, and they are infertile.

The reason for the health problems is that there are a few genes that are found on both the X and the Y chromosome. Because these genes are found in two copies in both XY males and XX females they are in what is called the **pseudoautosomal region**. This region escapes X chromosome inactivation. One of the genes in this region is called *SHOX*. It makes a protein that promotes bone growth. 46,XX and 46,XY people have two functioning copies and have average height. People with 47,XYY and 47,XXX genomes have three copies and are taller than average. And people with 45,X have one copy and are short. It is the single copy of *SHOX* and a few of the other genes in the pseudoautosomal region that causes health problems for women with Turner syndrome.

The reason for the infertility is that the X chromosome inactivation system only acts in somatic cells - it is not needed in the germline cells. Ovaries naturally have two functional X chromosomes. Women with Turner syndrome can not perform oogenesis because this process only works if there are two active X chromosomes. Recently, it has become possible for these women to become pregnant with donated eggs and in vitro fertilization.

9.3.4 Klinefelter Syndrome

There are four common sex-chromosome aneuploidies: 47,XYY, 47,XXX, 45,X, and 47,XXY. This last situation is known as **Klinefelter syndrome**. These people are male (because they have a Y chromosome) and tall (because they have three *SHOX* genes). They do not have health problems because the X chromosome inactivation system is independent of sex. In the embryonic nuclei the X chromosome are counted and all but one are shut down. It doesn't matter whether the embryo is male or female. Men with Klinefelter syndrome have a Barr bodies in their nuclei, the same as 46,XX females. They do have fertility problems because there are two

active X chromosome in their testes and this interferes with spermatogenesis. They make enough sperm to conceive children using intracytoplasmic sperm injection though.

9.3.5 Cri-du-Chat Syndrome

Cri-du-chat syndrome occurs when a child inherits a defective chromosome 5 from one parent (Figure 9.16n). 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).

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.

Figure 9.16n
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-CC:AN)

9.3.6 Inversion(9)

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(9). 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(9) is different from deletion(5) 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(9) chromosome has inherited it from a parent who had inherited it from one of their parents and so on. In contrast, most cases of deletion(5) are due to new mutations occurring in a parent.

9.4 Diagnosing Human Chromosome Abnormalities

9.4.1 Bright Field Microscopy

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 9.17n). Each chromosome can be recognized by its length, the location of its centromere, and the characteristic pattern of purple bands produced by the Giemsa. For example, if mitotic cells from a person consistently contained forty seven chromosomes in total with three chromosome 21s this would be indicative of Down syndrome. Bright field microscopy does 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.

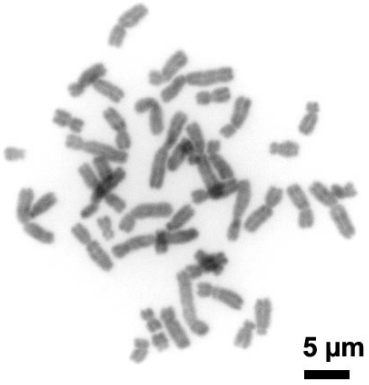


Figure 9.17n
Human chromosomes. One way to obtain chromosomes is to take a blood sample, culture the cells for three days in the presence of a T-cell growth factor, arrest the cells in metaphase with a microtubule inhibitor, and then drop the cells onto a slide. The cells burst and the chromosomes stick 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:AS)

9.4.2 Fluorescence In Situ Hybridization

The solution to these problems was **fluorescence in situ hybridization** (FISH). The technique is similar to a Southern blot in that a single stranded **DNA probe** is allowed to hybridize to denatured target DNA (see Section 8.6). However, instead of the probe being radioactive it is fluorescent and instead of the target DNA being restriction fragments on a nylon membrane it is denatured chromosomes on a glass slide. Because there are several fluorescent colours available it is common to use more than one probe at the same time. Typically the chromosomes are also labeled with a fluorescent stain called **DAPI** which gives them a uniform blue colour. If the chromosomes have come from a mitotic cell it is possible to see all forty six of them spread out in a small area. Alternatively, if the chromosomes are within the nucleus of an interphase cell they appear together within a large blue circle.

9.4.3 Using FISH to Diagnose Down Syndrome

Most pregnancies result in healthy children. However in some cases there is an elevated chance that the fetus has trisomy-21. Older women are at a higher risk because the non-disjunction events that lead to trisomy become more frequent with age. The second consideration is what the fetus looks like during an ultrasound examination. Fetuses with trisomy-21 and some other chromosome abnormalities have a swelling in the back of the neck called a nuchal translucency. If either or both factors is present the woman may choose amniocentesis. In this test some amniotic fluid is withdrawn so that the fetal cells within it can be examined. Figure 9.18n

shows a positive result for trisomy-21. Based upon this image the fetus has two X chromosomes and three chromosome 21s and therefore has a karyotype of 47,XX,+21.

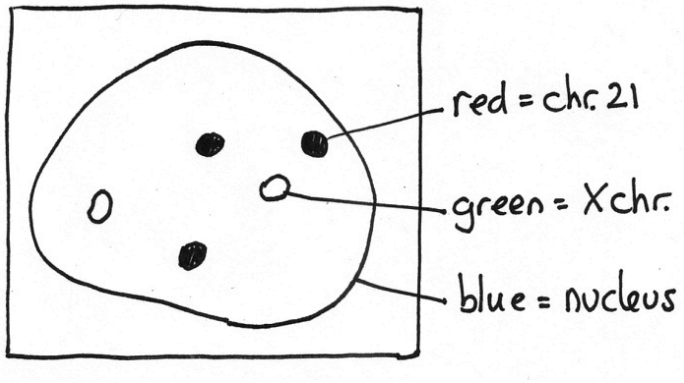


Figure 9.18n
Confirmation of Down syndrome in a fetal cell. This diagram is based upon actual results. The DNA has been stained blue with DAPI. A red fluorescent probe is binding to the centromeres of chromosome 21 (shown here as filled circles). A green fluorescent probe is binding to the centromeres of the X chromosome (open circles). Source: Figure 4 in Antonarakis, S. E. et al. 2004. Chromosome 21 and Down syndrome: From genomics to pathophysiology. *Nature Reviews Genetics* 5:725-738 PubMed ID: 15510164.

9.4.4 Using FISH to Diagnose Cri-du-Chat Syndrome

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 9.19n 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.

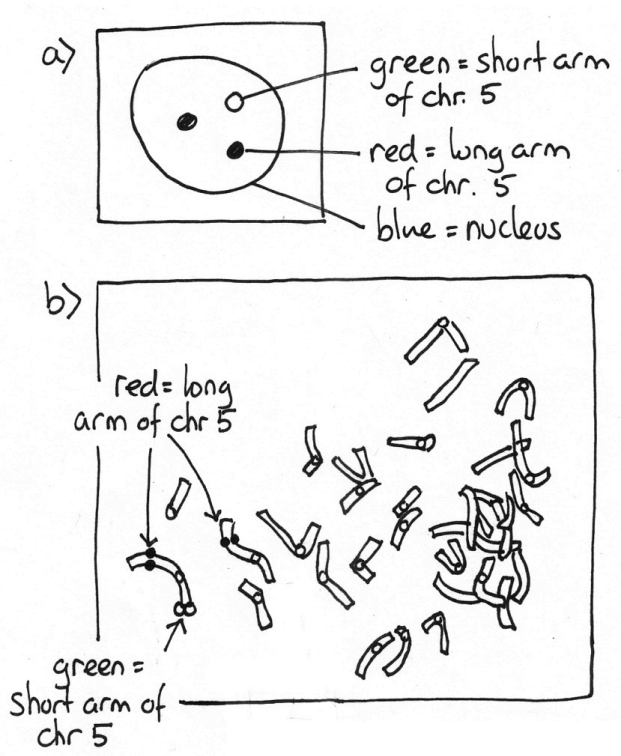


Figure 9.19n
A positive result for cri-du-chat syndrome. This diagram is based upon actual results. Cells from a patient's blood were prepared to show an interphase nucleus (a) and mitotic chromosomes (b). The DNA has been coloured blue with DAPI. The green fluorescent probe is binding to the tip of 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. Source: Figure 1 in Fang J.-S. et al. 2008 Cytogenetic and molecular characterization of three-generation family with chromosome 5p terminal deletion. Clinical Genetics 73:585-590 PubMed ID: 18400035.

9.4.5 Newer Techniques

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 a sample of DNA from a person with Down syndrome will contain 150% more DNA from chromosome 21 than the other chromosomes. Likewise 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.

SUMMARY:

- Errors during anaphase in mitosis or meiosis can lead to trisomy and other forms of aneuploidy.
- Errors during the repair of DNA breaks or during meiotic crossing over can lead to chromosome rearrangements.
- Five common forms of aneuploidy in humans are 47,sex,+21 (Down syndrome), 47,XYY, 47,XXX, 45,X (Turner syndrome) and 47,XXY (Klinefelter syndrome).
- Deletion(5) causes a serious condition (cri-du-chat syndrome) because deletions are unbalanced chromosome rearrangements.
- Inversion(9) causes few health consequences 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 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
non-disjunction
euploid
aneuploid
balanced
unbalanced
first division nondisjunction
second division nondisjunction
double strand break
nonhomologous end joining DNA repair system
chromosome rearrangement
deletion
inversion
paracentric inversion
pericentric inversion
tandem duplication
translocation
reciprocal translocation
Robertsonian translocation
meiotic crossover
deletion loop
karyotype
46,XX
46,XY
47,sex,+21 (Down syndrome)
trisomy
47,XYY
47,XXX
monosomy
45,X (Turner syndrome)
pseudoautosomal region
47,XXY (Klinefelter syndrome)
46,sex,deletion(5) (cri-du-chat syndrome)
46,sex,inversion(9)
bright field microscopy
Giemsa stain
fluorescence in situ hybridization
fluorescent probe
DAPI stain
amniocentesis

STUDY QUESTIONS:

9.1 Make diagrams showing how an improper crossover event during meiosis can lead to: (a) an inversion or (b) a translocation.

9.2 Make a diagram showing how a nondisjunction event can lead to a child with a 47,XYY karyotype.

9.3 How many Barr bodies would you expect to see in cells from people who are: (a) 46, XY, (b) 46,XX, (c) 47, XYY, (d) 47,XXX, (e) 45,X, and (f) 47,XXY ?

9.4 Why can people survive with trisomy-21 (47,sex,+21) but not monosomy-21 (45,sex,-21)?

9.5 If *Drosophila* geneticists want to generate mutant strains with deletions they expose flies to gamma rays. What does this imply about gamma rays?

9.6 What would happen if there was a nondisjunction event involving chromosome 21 in a 46,XY zygote?

9.7 Design a FISH based experiment to find out if your lab partner is a 47,XXX female or a 47,XYY male.

9.8 What would Figure 9.18 look like if it also showed metaphase chromosomes from another cell?

Notes:

Chapter 10
MOLECULAR MARKERS & QUANTITATIVE TRAITS

Figure 10.1
Many traits, such as body mass, show continuous variation, rather than discrete variation. Although environment obviously also affects this trait, some of the variation observed between individuals is heritable, and is dependent on the interactions of multiple alleles at multiple loci. The study of quantitative traits is one of many applications of molecular markers.
(Flickr-Jamie Golombek-CC:AND)

10.1 SOME VARIATIONS IN THE GENOME AFFECT COMPLEX TRAITS

Imagine that you could compare the complete genomic DNA sequence of any two people you meet today. Although their sequences would be very similar on the whole, they would certainly not be identical at each of the 3 billion base pair positions you examined (unless, perhaps, your subjects were identical twins – but even they may have some somatic differences). In fact, the genomic sequences of almost any two unrelated people differ at millions of nucleotide positions. Some of these differences would be found in the regions of genes that code for proteins. Others might affect the amount of transcript that is made for a particular gene. A person's health, appearance, behavior, and other characteristics depend in part on these polymorphisms.

Most difference, however, have no effect at all. They have no effect on gene sequences or expression, because they occur within regions of DNA that neither encode proteins, nor regulate the expression of genes. These polymorphisms are nevertheless very useful because they can be used as **molecular markers** in medicine, forensics, ecology, agriculture, and many other fields. In most situations, molecular markers obey the same rules of inheritance that we have already described for other types of loci, and so can be used to create genetic maps and to identify linked genes.

10.2 ORIGINS OF MOLECULAR POLYMORPHISMS

Mutations of DNA sequences can arise in many ways (Chapter 4). Some of these changes occur during DNA replication processes, resulting in an insertion, deletion, or substitution of one or a few nucleotides. Larger mutations can be caused by mobile genetic elements such as transposons, which are inserted more or less

randomly into chromosomal DNA, sometimes occurring in clusters. In these and other types of **repetitive DNA** sequences, the number of repeated units is highly prone to change through unequal crossovers and other replication events.

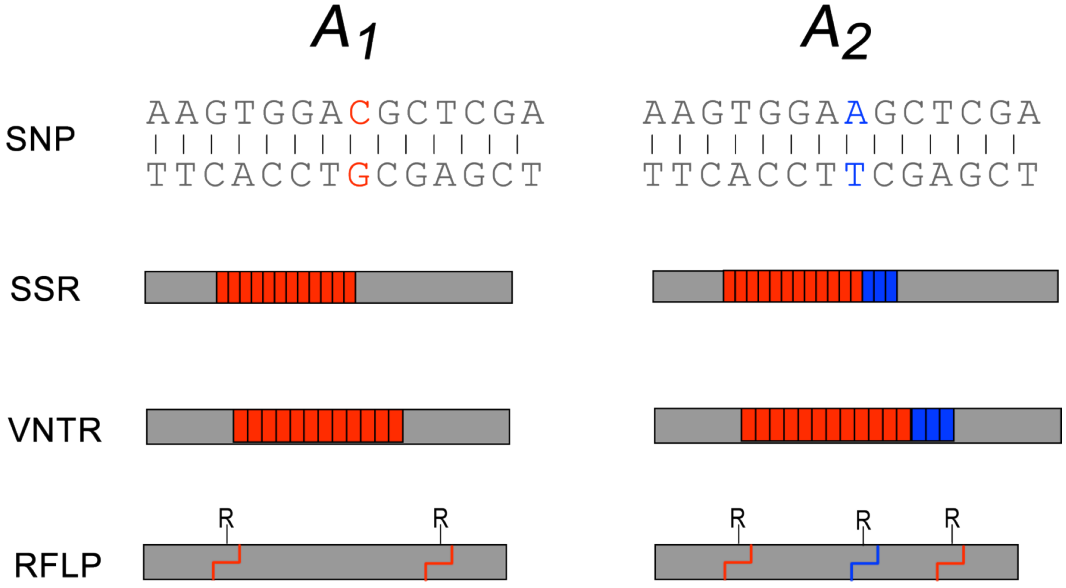


Figure 10.2 Some examples of DNA polymorphisms. The variant region is marked in blue, and each variant sequence is arbitrarily assigned one of two allele labels. Abbreviations: SNP (Single Nucleotide Polymorphism); SSR (Simple Sequence Repeat) = SSLP (Simple Sequence Length Polymorphism); **VNTR** (Variable Number of Tandem Repeats); **RFLP** (Restriction Fragment Length Polymorphisms. VNTRs and SSRs differ in the size of the repeat unit; VNTRs are larger than SSRs. (Original-Deyholos-CC:AN)

10.3 Classification and Detection of Molecular Markers

Regardless of their origins, molecular markers can be classified as polymorphisms that either vary in the length of a DNA sequence, or vary only in the identity of nucleotides at a particular position on a chromosome (Figure 10.2). In both cases, because two or more alternative versions of the DNA sequence exist, we can treat each variant as a different allele of a single locus. Each allele gives a different **molecular phenotype**. For example, polymorphisms of **SSRs** (short sequence repeats) can be distinguished based on the length of PCR products: one allele of a particular SSR locus might produce a 100bp band, while the same primers used with a different allele as a template might produce a 120bp band (Figure 10.3). A different type of marker, called a **SNP** (single nucleotide polymorphism), is an example of polymorphism that varies in nucleotide identity, but not length. SNPs are the most common of any molecular markers, and the genotypes of thousands of SNP loci can be determined in parallel, using new, hybridization based instruments. Note that the alleles of most molecular markers are co-dominant, since it is possible to distinguish the molecular phenotype of a heterozygote from either homozygote.

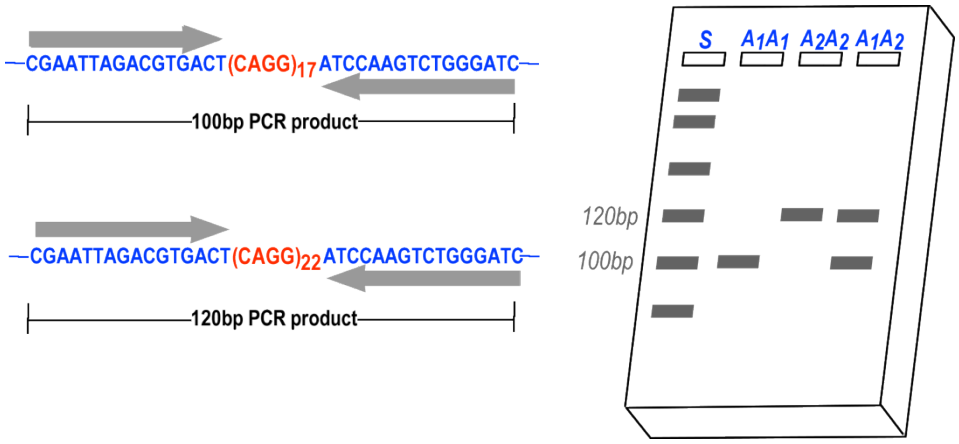


Figure 10.3 Determining the genotype of an individual at a single SSR locus using a specific pair of PCR primers and agarose gel electrophoresis. S= size standard
(Original-Deyholos-CC:AN)

Mutations that do not affect the function of protein sequences or gene expression are likely to persist in a population as polymorphisms, since there will be no selection either in favor or against them (i.e. they are **neutral**). Note that the although the rate of spontaneous mutation in natural populations is sufficiently high so as to generate millions of polymorphisms that accumulate over thousands of generations, the rate of mutation is on the other hand sufficiently low that existing polymorphisms are stable throughout the few generations we study in a typical genetic experiment.

10.4 Applications of Molecular Markers

Several characteristics of molecular markers make them useful to geneticists. First, because of the way DNA polymorphisms arise and are retained, they are frequent throughout the genome. Second, because they are phenotypically neutral, it is relatively easy to find markers that differ between two individuals. Third, their neutrality also makes it possible to study hundreds of loci without worrying about gene interactions or other influences that make it difficult to infer genotype from phenotype. Lastly, unlike visible traits such as eye color or petal color, the phenotype of a molecular marker can be detected in any tissue or developmental stage, and the same type of assay can be used to score molecular phenotypes at millions of different loci. Thus, the neutrality, high density, high degree of polymorphism, co-dominance, and ease of detection of molecular markers has lead to their wide adoption in many areas of research.

It is worth emphasizing again that DNA polymorphisms are a natural part of most genomes. Geneticists discover these polymorphisms in various ways, including comparison of random DNA sequence fragments from several individuals in a population. Once molecular markers have been identified, they can be used in many ways, including:

10.4.1 DNA fingerprinting

By comparing the allelic genotypes at multiple molecular marker loci, it is possible to determine the likelihood of similarity between two DNA samples. If markers differ, then clearly the DNA is from different sources. If they don't differ, then one

can estimate the unlikelihood of them coming from different sources – eg they are from the same source. For example, a forensic scientist can demonstrate that a blood sample found on a weapon came from a particular suspect. Similarly, that leaves in the back of a suspect's pick-up truck came from a particular tree at a crime scene. DNA fingerprinting is also useful in paternity testing (Figure 10.4) and in commercial applications such as verification of species of origin of certain foods and herbal products.

10.4.2 Construction of genetic linkage maps

By calculating the recombination frequency between pairs of molecular markers, a map of each chromosome can be generated for almost any organism (Figure 10.5). These maps are calculated using the same mapping techniques described for genes in Chapter 7, however, the high density and ease with which molecular markers can be genotyped makes them more useful than other phenotypes for constructing genetic maps. These maps are useful in further studies, including map-based cloning of protein coding genes that were identified by mutation.

10.4.3 Population studies

As described in Chapter 5, the observed frequency of alleles, including alleles of molecular markers, can be compared to frequencies expected for populations in Hardy-Weinberg equilibrium to determine whether the population is in equilibrium. By monitoring molecular markers, ecologists and wildlife biologists can make inferences about migration, selection, diversity, and other population-level parameters.

Molecular markers can also be used by anthropologists to study migration events in human ancestry. There is a large commercial business available that will genotype people and determine their deep genetic heritage for ~$100. This can be examined through the maternal line via sequencing their mitochondrial genome and through the paternal line via genotyping their Y-chromosome.

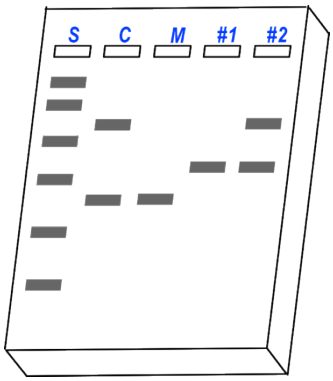


Figure 10.4 Paternity testing. Given the molecular phenotype of the child (C) and mother (M), only one of the possible fathers (#2) has alleles that are consistent with the child's phenotype. (Original-Deyholos-CC:AN)

For example, about 8% of the men in parts of Asia (about 0.5% of the men in the world) have a Y-chromosomal lineage belonging to Genghis Khan (the haplogroup C3) and his decendents.

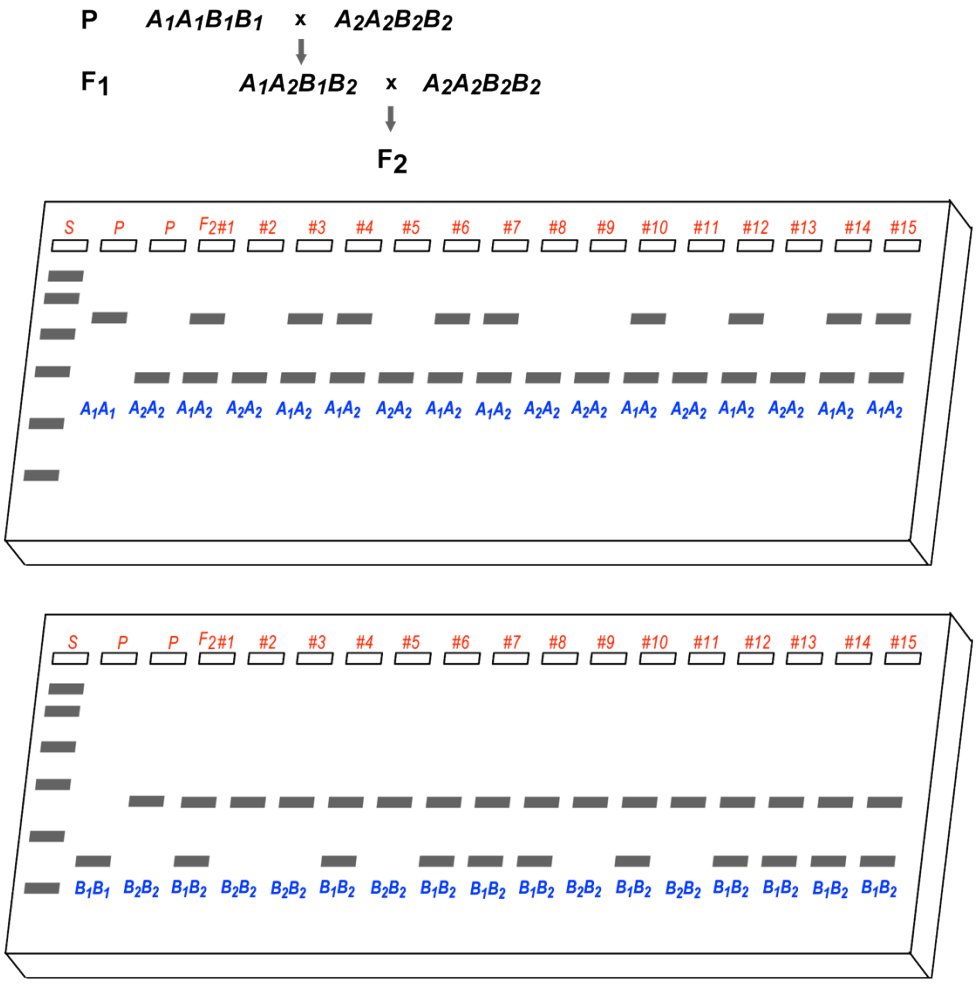


Figure 10.5 Measuring recombination frequency between two molecular marker loci, *A* and *B*. A different pair of primers is used to amplify DNA from either parent (P) and 15 of the F_2 offspring from the cross shown. Recombinant progeny will have the genotype $A_1A_2B_2B_2$ or $A_2A_2B_1B_2$. Individuals #3, #8, #13 are recombinant, so the recombination frequency is 3/15=20%. (Original-Deyholos-CC:AN)

10.4.4 Identification of linked traits

It is often possible to correlate or link an allele of a molecular marker with a particular disease or other trait of interest. One way to make this correlation is to obtain genomic DNA samples from hundreds of individuals with a particular disease, as well as samples from a control population of healthy individuals. The genotype of each individual is scored at hundreds or thousands of molecular marker loci (e.g. SNPs), to find alleles that are usually present in persons with the disease, but not in healthy subjects. The molecular marker is presumed to be tightly linked

to the gene that causes the disease, although this protein-coding gene may itself be as yet unknown. The presence of a particular molecular polymorphism may therefore be used to diagnose a disease, or to advise an individual of susceptibility to a disease.

Molecular markers may also be used in a similar way in agriculture to track desired traits. For example, markers can be identified by screening both the traits and molecular marker genotypes of hundreds of individuals. Markers that are linked to desirable traits can then be used during breeding to select varieties with economically useful combinations of traits, even when the genes underlying the traits are not known.

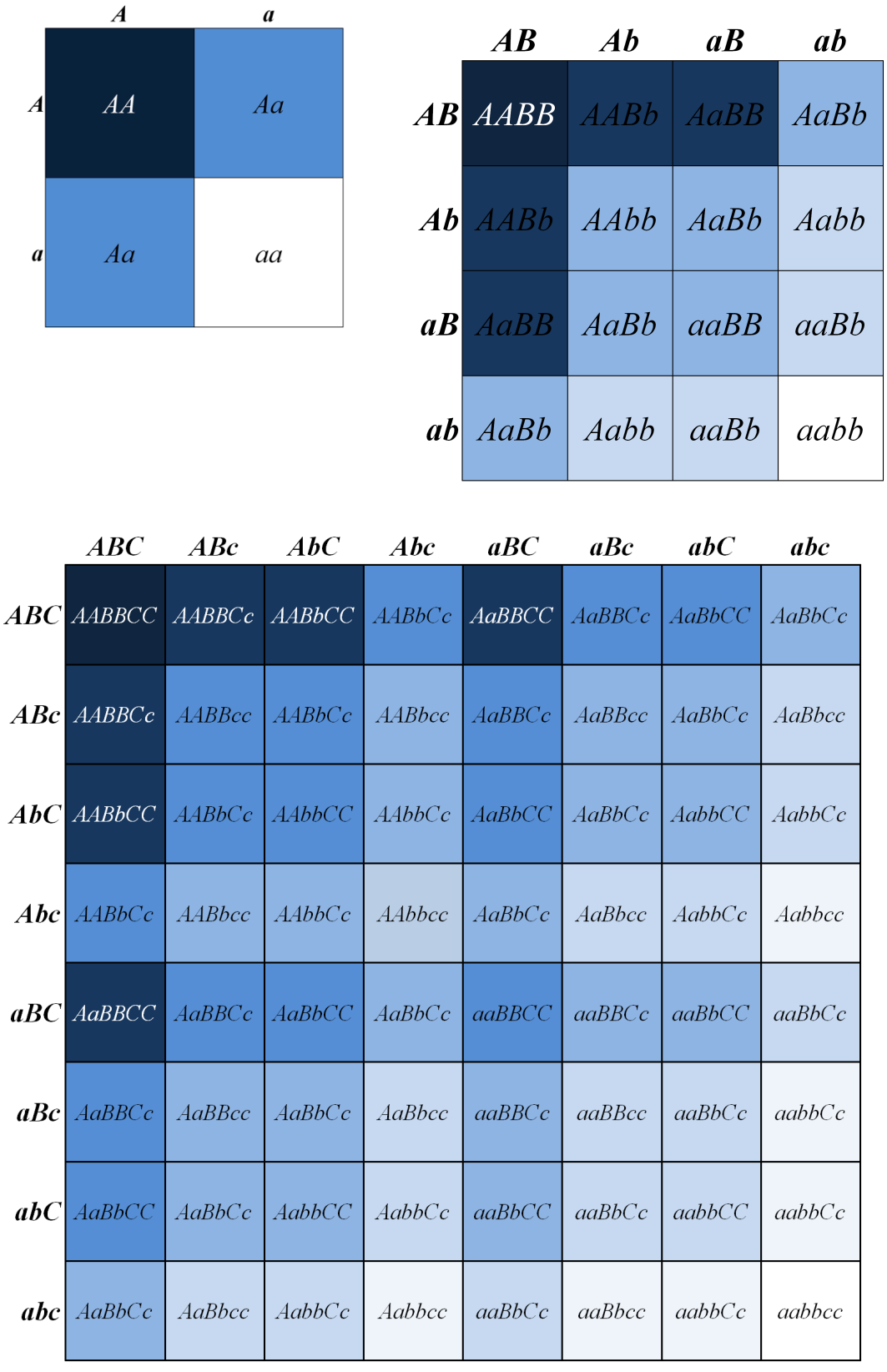
10.4.5 Quantitative trait locus (QTL) mapping

Molecular markers can be used to identify multiple different regions of chromosomes that contain genes that act together to produce complex traits. This process involves finding combinations of alleles of molecular markers that are correlated with a quantitative phenotype such as body mass, height, or intelligence. QTL mapping is described in more detail in the following section.

10.5 Quantitative Trait Locus (QTL) Analysis

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. Many complex traits are also influenced heavily by the environment. 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 10.6).



	A	a
A	AA	Aa
a	Aa	aa

	AB	Ab	aB	ab
AB	AABB	AABb	AaBB	AaBb
Ab	AABb	AAbb	AaBb	Aabb
aB	AaBB	AaBb	aaBB	aaBb
ab	AaBb	Aabb	aaBb	aabb

	ABC	ABc	AbC	Abc	aBC	aBc	abC	abc
ABC	AABBCC	AABBCc	AABbCC	AABbCc	AaBBCC	AaBBCc	AaBbCC	AaBbCc
ABc	AABBCc	AABBcc	AABbCc	AABbcc	AaBBCc	AaBBcc	AaBbCc	AaBbcc
AbC	AABbCC	AABbCc	AAbbCC	AAbbCc	AaBbCC	AaBbCc	AabbCC	AabbCc
Abc	AABbCc	AABbcc	AAbbCc	AAbbcc	AaBbCc	AaBbcc	AabbCc	Aabbcc
aBC	AaBBCC	AaBBCc	AaBbCC	AaBbCc	aaBBCC	aaBBCc	aaBbCC	aaBbCc
aBc	AaBBCc	AaBBcc	AaBbCc	AaBbcc	aaBBCc	aaBBcc	aaBbCc	aaBbcc
abC	AaBbCC	AaBbCc	AabbCC	AabbCc	aaBbCC	aaBbCc	aabbCC	aabbCc
abc	AaBbCc	AaBbcc	AabbCc	Aabbcc	aaBbCc	aaBbcc	aabbCc	aabbcc

Figure 10.6 Punnett Squares for one, two, or 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:AN)

If the number of phenotypic classes is sufficiently large (as with three or more loci), individual classes may become indistinguishable from each other (particularly when environmental effects are included), and the phenotype appears as a continuous variation (Figure 10.7). 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.

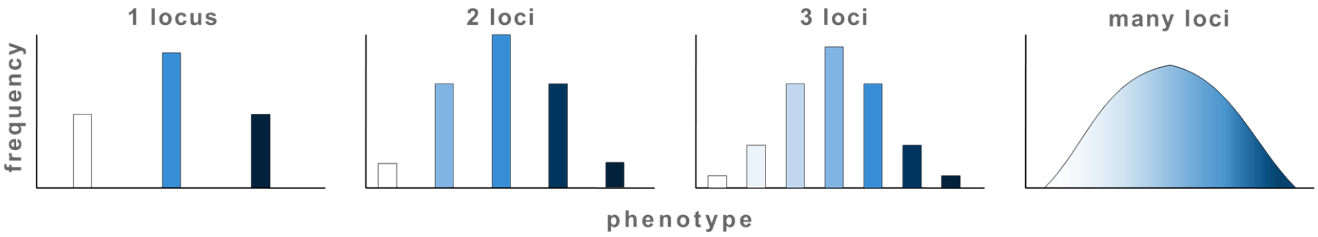


Figure 10.7 The more loci that affect a trait, the larger the number of phenotypic classes that can be expected. For some traits, the number of contributing loci is so large that the phenotypic classes blend together in apparently continuous variation. (Original-Deyholos-CC:AN)

We can use molecular markers to identify at least some of the genes (those with a major influence) that affect a given quantitative trait. This is essentially an extension of the mapping techniques we have already considered for discrete traits. A QTL mapping experiment will ideally start with two pure-breeding lines that differ greatly from each other in respect to one or more quantitative traits (Figure 10.8). The parents and all of their progeny should be raised under as close to the same environmental conditions as possible, to ensure that observed variation is due to genetic rather than external environmental factors. These parental lines must also be polymorphic for a large number of molecular loci, meaning that they must have different alleles from each other at hundreds of loci. The parental lines are crossed, and then this F_1 individual, in which recombination between parental chromosomes has occurred is self-fertilized (or back-crossed). Because of recombination (both crossing over and independent assortment), each of the F_2 individuals will contain a different combination of molecular markers, and also a different combination of alleles for the genes that control the quantitative trait of interest (Table 10.1). By comparing the molecular marker genotypes of several hundred F_2 individuals with their quantitative phenotypes, a researcher can identify molecular markers for which the presence of particular alleles is always associated with extreme values of the trait. In this way, regions of chromosomes that contain genes that contribute to quantitative traits can be identified. (Figure 10.9) It then takes much more work (further mapping and other experimentation) to identify the individual genes in each of the regions that control the quantitative trait.

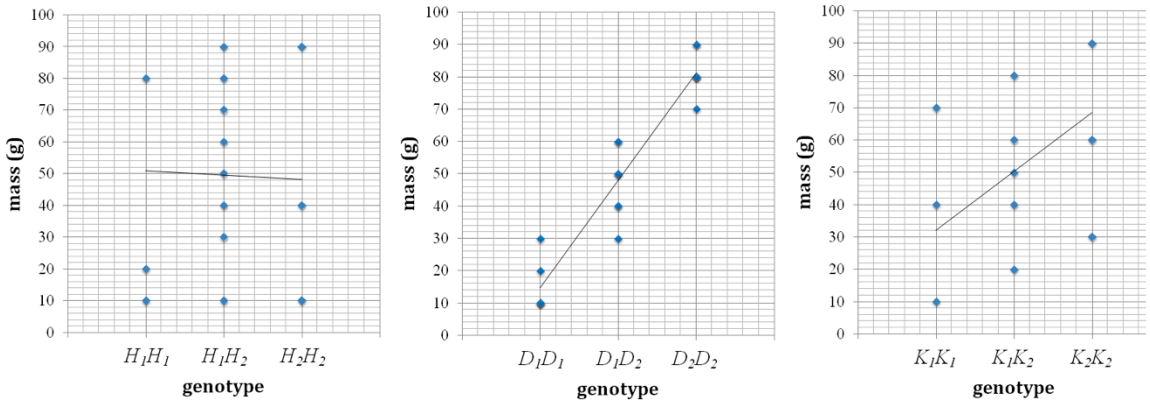


Figure 10.9 Plots of fruit mass and genotype for selected loci from Table 10.1. For most loci (e.g. *H*), the genotype shows no significant correlation with fruit weight. However, for some molecular markers, the genotype will be highly correlated with fruit weight. Both *D* and *K* influence fruit weight, but the effect of genotype at locus *D* is larger than at locus *K*. (Original-Deyholos-CC:AN)

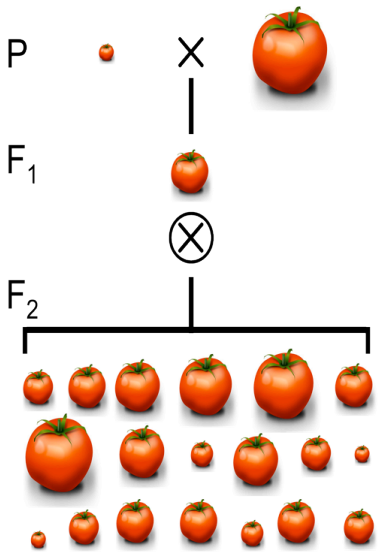


Figure 10.8 Strategy for a typical QTL mapping experiment. Two parents that differ in a quantitative trait (e.g. fruit mass) are crossed, and the F_1 is self-fertilized (as shown by the cross-in-circle symbol). The F_2 progeny will show a range of quantitative values for the trait. The task is then to identify alleles of markers from one parent that are strongly correlated with the quantitative trait. For example, markers from the large-fruit parent that are always present in large-fruit F_2 individuals (but never in small-fruit individuals) are likely linked to loci that control fruit mass.

	genotype	fruit mass
P	$A_1A_1B_1B_1C_1C_1D_1D_1E_1E_1F_1F_1G_1G_1H_1H_1J_1J_1K_1K_1$	10g
P	$A_2A_2B_2B_2C_2C_2D_2D_2E_2E_2F_2F_2G_2G_2H_2H_2J_2J_2K_2K_2$	90g
F_1	$A_1A_2B_1B_2C_1C_2D_1D_2E_1E_2F_1F_2G_1G_2H_1H_2J_1J_2K_1K_2$	50g
F_2 #001	$A_1A_1B_1B_2C_1C_1D_2D_2E_1E_2F_1F_2G_1G_2H_1H_1J_1J_2K_1K_2$	80g
F_2 #002	$A_1A_2B_1B_2C_1C_2D_1D_1E_1E_2F_1F_2G_2G_2H_1H_2J_2J_2K_1K_1$	10g
F_2 #003	$A_2A_2B_1B_2C_2C_2D_1D_2E_1E_2F_1F_2G_1G_1H_1H_2J_1J_2K_1K_2$	50g
F_2 #004	$A_1A_2B_1B_2C_1C_2D_1D_2E_1E_2F_1F_2G_1G_2H_1H_2J_1J_2K_2K_2$	60g
F_2 #005	$A_1A_2B_1B_1C_1C_2D_2D_2E_1E_2F_1F_2G_1G_2H_1H_2J_1J_1K_2K_2$	90g
F_2 #006	$A_1A_2B_2B_2C_1C_2D_1D_2E_1E_2F_1F_2G_2G_2H_1H_2J_1J_2K_1K_2$	60g
F_2 #007	$A_2A_2B_1B_1C_1C_2D_2D_2E_1E_2F_1F_2G_1G_1H_1H_2J_1J_1K_1K_2$	80g
F_2 #008	$A_1A_1B_1B_2C_1C_2D_1D_2E_1E_2F_1F_2G_1G_2H_1H_2J_1J_2K_1K_2$	50g
F_2 #009	$A_1A_2B_1B_2C_2C_2D_1D_2E_1E_2F_1F_2G_1G_2H_1H_2J_1J_2K_1K_2$	50g
F_2 #010	$A_1A_2B_1B_2C_1C_2D_1D_2E_1E_2F_1F_2G_1G_2H_1H_2J_1J_2K_2K_2$	30g
F_2 #011	$A_1A_2B_1B_2C_1C_2D_2D_2E_1E_1F_1F_2G_1G_2H_1H_2J_1J_2K_1K_2$	80g
F_2 #012	$A_1A_1B_1B_2C_1C_2D_1D_1E_1E_2F_2F_2G_1G_2H_1H_2J_1J_2K_2K_2$	30g
F_2 #013	$A_2A_2B_1B_1C_1C_2D_1D_1E_1E_2F_1F_1G_1G_2H_2H_2J_1J_1K_1K_1$	10g
F_2 #014	$A_2A_2B_1B_1C_1C_1D_2D_2E_1E_2F_1F_2G_1G_2H_1H_2J_1J_1K_1K_1$	70g
F_2 #015	$A_2A_2B_2B_2C_1C_2D_1D_2E_1E_2F_2F_2G_1G_2H_1H_1J_2J_2K_1K_2$	40g
F_2 #016	$A_1A_2B_2B_2C_1C_2D_1D_1E_1E_2F_1F_1G_2G_2H_1H_1J_1J_2K_1K_1$	10g
F_2 #017	$A_1A_2B_2B_2C_1C_2D_2D_2E_2E_2F_1F_1G_2G_2H_1H_2J_1J_2K_2K_2$	90g
F_2 #018	$A_1A_2B_2B_2C_1C_2D_1D_2E_1E_2F_1F_1G_2G_2H_1H_2J_1J_2K_1K_1$	40g
F_2 #019	$A_1A_1B_1B_2C_1C_2D_1D_1E_1E_2F_2F_2G_1G_1H_1H_1J_1J_2K_1K_2$	20g
F_2 #100	$A_1A_1B_1B_2C_1C_2D_2D_2E_1E_2F_1F_2G_2G_2H_1H_2J_2J_2K_1K_2$	80g

Table 10.1 Genotypes and quantitative data for some individuals from the crosses shown in Figure 10.8

SUMMARY:

- Natural variations in the length or identity of DNA sequences occur at millions of locations throughout most genomes.
- DNA polymorphisms are often neutral, but because of linkage may be used as molecular markers to identify regions of genomes that contain genes of interest.
- Molecular markers are useful because of their neutrality, co-dominance, density, allele frequencies, ease of detection, and expression in all tissues.
- Molecular markers can be used for any application in which the identity of two DNA samples is to be compared, or when a particular region of a chromosome is to be correlated with inheritance of a trait.
- Many important traits show continuous, rather than discrete variation. These are also called quantitative traits.
- Many quantitative traits are influenced by a combination of environment and genetics.
- The heritable component of quantitative traits can best be studied under controlled conditions, with pure-breeding parents that are polymorphic for both a quantitative trait and a large number of molecular markers.
- Molecular markers can be identified for which specific alleles are tightly correlated with the quantitative value of a particular phenotype. The genes that are linked to these markers can be identified through subsequent research.

KEY TERMS:

molecular marker
repetitive DNA
SSR
SSLP
VNTR
SNP
RFLP
neutral mutation
QTL
discrete variation
continuous variation
polygenic

STUDY QUESTIONS:

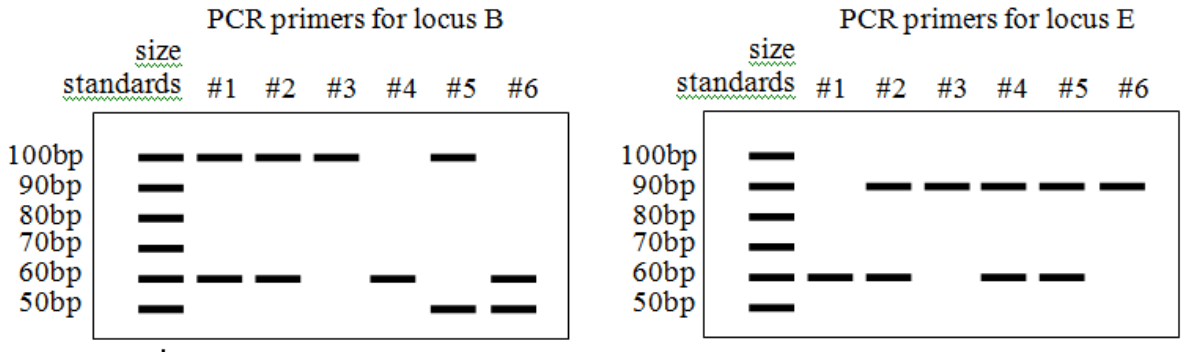
10.1 Three different polymorphisms have been identified at a particular molecular marker locus. A single pair of PCR primers will amplify either a 50bp fragment (*B2*), a 60bp fragment (*B3*), or a 100bp fragment (*B4*).

Draw the PCR bands that would be expected if these primers were used to amplify DNA from individuals with each of the following genotypes:

a) B_2B_2
b) B_4B_4
c) B_2B_3
d) B_2B_4

10.2 In addition to the primers used to genotype locus *B* (described above), a separate pair of primers can amplify another polymorphic SSR locus *E,* with either a 60bp product (E_1) or a 90bp (E_2) product. DNA was extracted from six individuals (#1- #6), and DNA from each individual was used as a template in separate PCR reactions with primers for either locus *B* or primers for locus *E*, and the PCR products were visualized on electrophoretic gels as shown below.

Based on the following PCR banding patterns, what is the full genotype of each of the six individuals?



10.3 Based on the genotypes you recorded in Question 10.2, can you determine which of the individuals could be a parent of individual #1?

10.4 Here is part of the DNA sequence of a chromosome:

```
TAAAGGAATCAATTACTTCTGTGTGTGTGTGTGTGTGTGTGTGTTCTTAGTTGTTTAAGTTTTAA
GTTGTGA
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
|||||||
ATTTCCTTAGTTAATGAAGACACACACACACACACACACACACAAGAATCAACAAATTCAAAATT
CAACACT
```

Identify the following features on the sequence:

a) the region of the fragment that is most likely to be polymorphic
b) any simple sequence repeats
c) the best target sites for PCR primers that could be used to detect polymorphisms in the length of the simple sequence repeat region in different individuals

10.5 In a particular diploid plant, seed color is a polygenic trait. If true-breeding plants that produce red seeds are crossed with true breeding plants that produce white seeds, the F_1 produces seeds that are intermediate in color (i.e. pink). When an F_1 plant self-fertilizes, white seeds are observed in the next generation. How many genes are involved in seed color for each of the following frequencies of white seeds in the F_2 generation?

a) 1/4 white seeds
b) 1/16 white seeds
c) 1/64 white seeds
d) 1/256 white seeds

10.6 If height in humans is a polygenic trait, explain why it occasionally happens that two tall parents have a child who grows up to be much shorter than either of them.

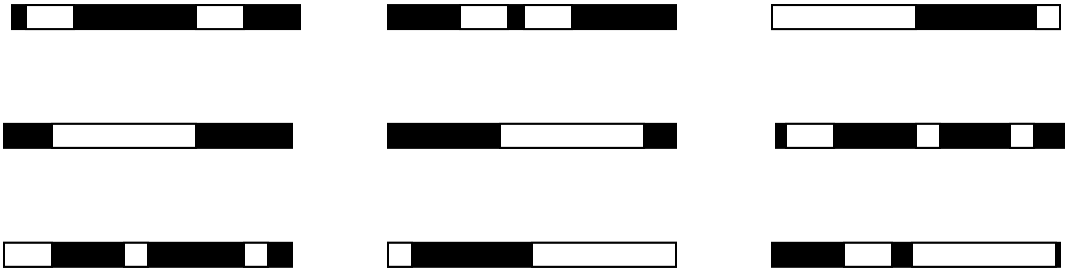
10.7 In quantitative trait (QTL) mapping, researchers cross two parents that differ in expression of some quantitative trait, then allow chromosomes from these parents to recombine randomly, and after several generations of inbreeding, produce a large number of offspring ("recombinant inbred lines"). Because the position of crossovers is random, each of the offspring contain a different combination of chromosomal regions from each of the two parents. The researchers then use molecular markers to determine which chromosomal regions have the greatest influence on the quantitative trait, e.g. in tall offspring, which chromosomal regions always come from the tall parent?

Imagine that two mice strains have been identified that differ in the time required to complete a maze, which may be an indication of intelligence. The time for maze completion is heritable and these parental strains "breed true" for the same completion time in each generation. Imagine also that their chromosomes are different colors and we can track the inheritance of chromosomal regions from each parent based on this color.

Based on the following diagrams of one chromosome from each individual in a pedigree, identify a chromosomal region that may contain a gene that affects time to complete a maze. The time for each individual is shown below each chromosome. Assume that all individuals are homozygous for all loci.

Parents:

Selected individuals from among F8 progeny of the above parents:



10.8 In a more realistic situation (as compared to question 7), where you could not distinguish the parental origin of different chromosomal regions just by appearance of chromosomes, explain how you could identify which parent was the source of a particular region of a chromosome in recombinant offspring.

Chapter 11 GENOMICS AND SYSTEMS BIOLOGY

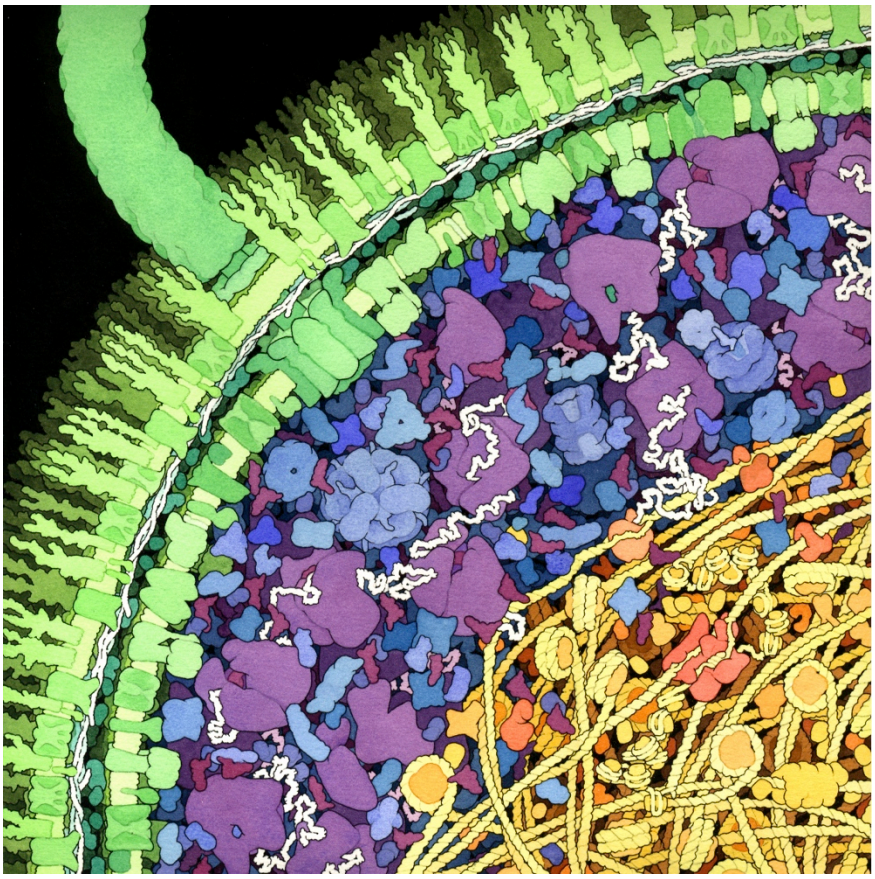


Figure 11.1
An artist's depiction of part of an *E.coli* cell, showing many different types of molecules in their typical abundance. mRNA appears as white lines associated with purple ribosomes, while DNA and proteins such as histones are yellow. (Goodsell, Scripps-EDU)

Imagine that you could identify and quantify every molecule within a cell (Figure 11.1) in a single assay. You could use this ability to better understand almost any aspect of biology. For example, by comparing the molecular profiles of plants that differed in their resistance to drought, you might discover which combination of genes or proteins makes a crop drought tolerant. Although it is not currently possible to study literally every molecule in a cell in a single experiment, recent advances in molecular biology have made it possible to study many genes (or their products) in parallel.

11.1 'OMICS TECHNOLOGIES

The complete set of DNA within an organism is called its **genome**. **Genomics** is therefore the large-scale description, using techniques of molecular biology of many genes or even whole genomes at once. This type of research is facilitated by technologies that increase throughput (i.e. rate of analysis), and decrease cost. The **–omics** suffix has been used to indicate high-throughput analysis of many types of molecules, including transcripts (**transcriptomics**), proteins (**proteomics**), and the products of enzymatic reactions, or metabolites (**metabolomics**; Fig. 11.2). Interpretation of the large data sets generated by –omics research depends on a combination of computational, biological, and statistical knowledge provided by

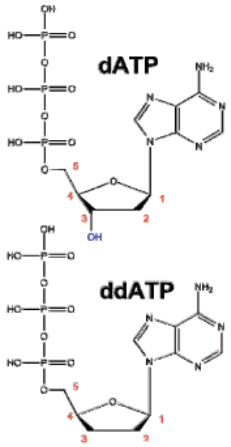
experts in **bioinformatics**. Attempts to combine information from different types of –'omics studies is sometimes called **systems biology**.

Figure 11.2
A word cloud listing some of the different –omics technologies. Terms that are more widely used are written in the largest characters. There is no significance to the color of each word. (Original-Deyholos-CC:AN)

11.2 DNA SEQUENCING

DNA sequencing determines the order of nucleotide bases within a given fragment of DNA. This information can be used to infer the RNA or protein sequence encoded by the gene, from which further inferences may be made about the gene's function and its relationship to other genes and gene products. DNA sequence information is also useful in studying the regulation of gene expression. If DNA sequencing is applied to the study of many genes, or even a whole genome, it is considered an example of genomics.

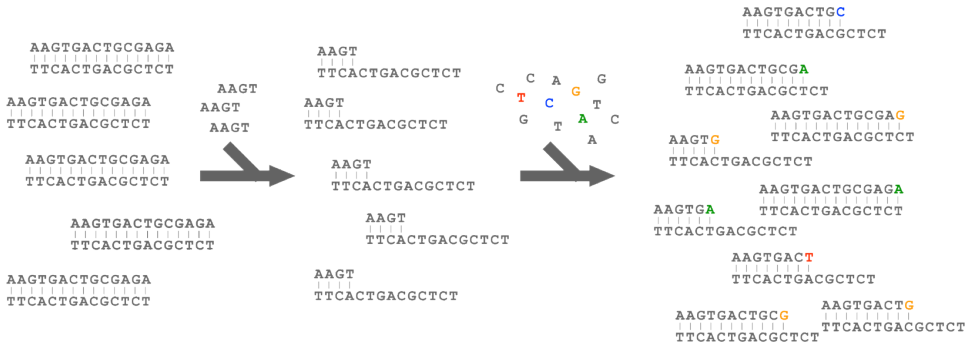
Figure 11.3
ddNTPs (Original-Deyholos-CC:AN)



11.2.1 DIDEOXY SEQUENCING

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. Most of the currently used DNA sequencing techniques rely on the random incorporation of modified nucleotides called terminators. Examples of terminators are the **dideoxy nucleotides** (**ddNTPs**), which lack a 3'-OH group and therefore cannot serve as an attachment site for the addition of new bases to a growing strand of DNA (Figure 11.3). After a ddNTP is incorporated into a strand of DNA, no further elongation can occur. Terminators are labeled with one of four fluorescent dyes, each specific for one the four nucleotide bases.

Figure 11.4
A sequencing reactions begins with many identical copies of a template DNA fragment. The template is denatured, then primers are annealed to the template. Following the addition of polymerase, regular dNTPS, and fluorescently labeled terminators, extension begins at the primer site. Elongation proceeds until a fluorescently labeled terminator (shown here in color) is incorporated. (Original-Deyholos-CC:AN)



To sequence a DNA fragment, you need many copies of that fragment (Fig. 11.4). Unlike PCR, DNA 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 dNTPs, fluorescently labeled terminators, 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 labeled nucleotide is incorporated, at which point extension is terminated. Because the reaction contains millions of template molecules, a corresponding 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 (Fig. 11.5). Since each band differs in length by one nucleotide, and the identity of that nucleotide is known from its fluorescence, the DNA sequence can be read simply from the order of the colors in successive bands. In practice, the maximum length of sequence that can be read from a single sequencing reaction is about 700bp.

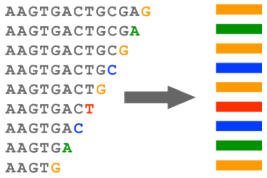


Figure 11.5
Fluorescently labeled products can be separated electrophoretically based on their length. (Original-Deyholos-CC:AN)

A particularly sensitive electrophoresis method used in the analysis of DNA sequencing reactions is called capillary electrophoresis (Fig. 11.6). In this method, a current pulls the sequencing products through a gel-like matrix that is encased in a fine tube of clear plastic. As in conventional electrophoresis, the smallest fragments move through the capillary the fastest. As they pass through a point near the end of the capillary, the fluorescent intensity of each dye is read. This produces a graph called a chromatogram. The sequence is determined by identifying the highest peak (i.e. the dye with the most intense fluorescent signal) at each position.

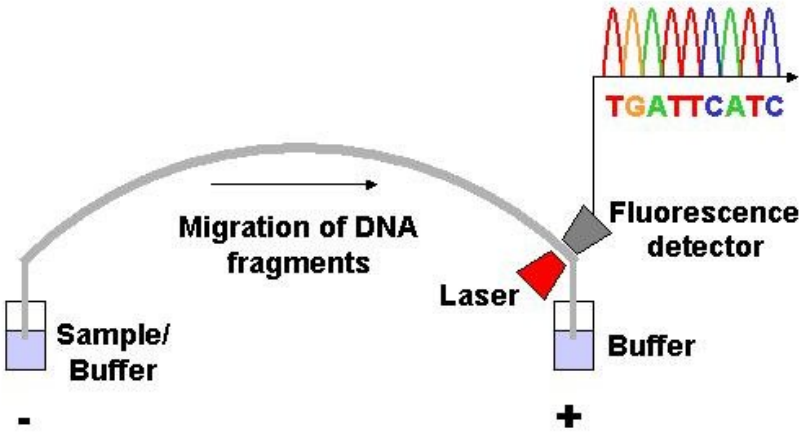


Figure 11.6
Fluorescently labeled products can be separated by capillary electrophoresis, generating a chromatogram from which the sequence can be read.(Wikipedia-Abizar Lakdawalla-PD)

11.2.2 Next-generation sequencing

Advances in technology over the past two decades have increased the speed and quality of sequencing, while decreasing the cost. This has become especially true with the most recently developed methods called next-generation sequencing. Not all of these new methods rely on terminators, but one that does is a method used in instruments sold by a company called **Illumina**. Illumina sequencers use a special variant of PCR called bridge PCR to make many thousands of copies of a short (45bp) template fragment. Each of these short template fragments is attached in a cluster in a small spot on a reaction surface. Millions of other clusters, each made by different template fragment, are located at other positions on the reaction surface. DNA synthesis at each template strand then proceeds using dye-labeled terminators

that are used are reversible. Synthesis is therefore terminated (temporarily) after the incorporation of each nucleotide. Thus, after the first nucleotide is incorporated in each strand, a camera records the color of fluorescence emitted from each cluster. The terminators are then modified, and a second nucleotide is incorporated in each strand, and again the reaction surface is photographed. This cycle is repeated a total of 45 times. Because millions of 45bp templates are sequenced in parallel in a single process, Illumina sequencing is very efficient compared to other sequencing techniques. However, the short length of the templates currently limits the application of this technology.

11.3 Whole Genome Sequencing

11.3.1 The need for assembly

Given that the length of a single, individual sequencing read is somewhere between 45bp and 700bp, we are faced with a problem determining the sequence of longer fragments, such as the chromosomes in an entire genome of humans (3 x10^9 bp). Obviously, we need to break the genome into smaller fragments. There are two different strategies for doing this: (1) clone-by-clone sequencing, which relies on the creation of a physical map first then sequencing, and (2) whole genome shotgun sequencing, which sequences first and does not require a physical map.

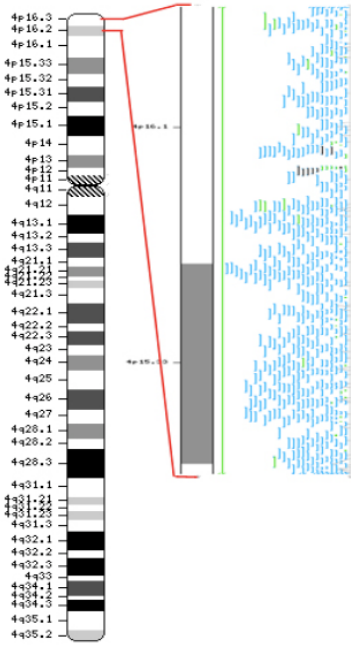


Figure 11.7
A portion of the physical map for human chromosome 4. The entire chromosome is shown at left. The physical map is made up of small blue lines, each of which represents a cloned piece of DNA approximately 100kb in length. (NCBI-unknown-PD)

11.3.2 Physical mapping

A physical map is a representation of a genome, comprised of cloned fragments of DNA. The map is therefore made from physical entities (pieces of DNA) rather than abstract concepts such as the linkage frequencies and genes that make up a genetic map (Fig. 11.7). It is usually possible to correlate genetic and physical maps, for example by identifying the clone that contains a particular molecular marker. The connection between physical and genetic maps allows the genes underlying particular mutations to be identified through a process call map-based cloning.

To create a physical map, large fragments of the genome are cloned into plasmid vectors, or into larger vectors called bacterial artificial chromosomes (BACs). BACs can contain approximately 100kb fragments. The set of BACs produced in a cloning reaction will be redundant, meaning that different clones will contain DNA from the same part of the genome. Because of this redundancy, it is useful to select the minimum set of clones that represent the entire genome, and to order these clones respective to the sequence of the original chromosome. Note that this is all to be done without knowing the complete sequence of each BAC. Making a physical map may therefore rely on techniques related to Southern blotting: DNA from the ends of one BAC is used as a probe to find clones that contain the same sequence. These clones are then assumed to overlap each other. A set of overlapping clones is called a **contig**.

11.3.3 Clone-by-clone sequencing

Physical mapping of cloned sequences was once considered a pre-requisite for genome sequencing. The process would begin by breaking the genome into BAC-sized pieces, arranging these BACs into a map, then breaking each BAC up into a

series of smaller clones, which were usually then also mapped. Eventually, a minimum set of smaller clones would be identified, each of which was small enough to be sequenced (Fig. 11.8). Because the order of clones relative to the complete chromosome was known prior to sequencing, the resulting sequence information could be easily assembled into one complete chromosome at the end of the project. Clone-by-clone sequencing therefore minimizes the number of sequencing reactions that must be performed, and makes sequence assembly straightforward and reliable. However, a drawback of this strategy is the tedious process of building physical map prior to any sequencing.

11.3.4 Whole genome shotgun sequencing

This strategy breaks the genome into fragments that are small enough to be sequenced, then reassembles them simply by looking for overlaps in the sequence of each fragment. It avoids the laborious process of making a physical map (Fig. 11.8). However, it requires many more sequencing reactions than the clone-by-clone method, because, in the shotgun approach, there is no way to avoid sequencing redundant fragments. There is also a question of the feasibility of assembling complete chromosomes based simply on the sequence overlaps of many small fragments. This is particularly a problem when the size of the fragments is smaller than the length of a repetitive region of DNA. Nevertheless, this method has now been successfully demonstrated in the nearly complete sequencing of many large genomes (rice, human, and many others). It is the current standard methodology.

However, shotgun assemblies are rarely able to complete entire genomes. The human genome, for example, relied on a combination of shotgun sequence and physical mapping to produce contiguous sequence for the length of each arm of each chromosome. Note that because of the highly repetitive nature of centromeric and telomeric DNA, sequencing projects rarely include these heterochromatic, gene poor regions.

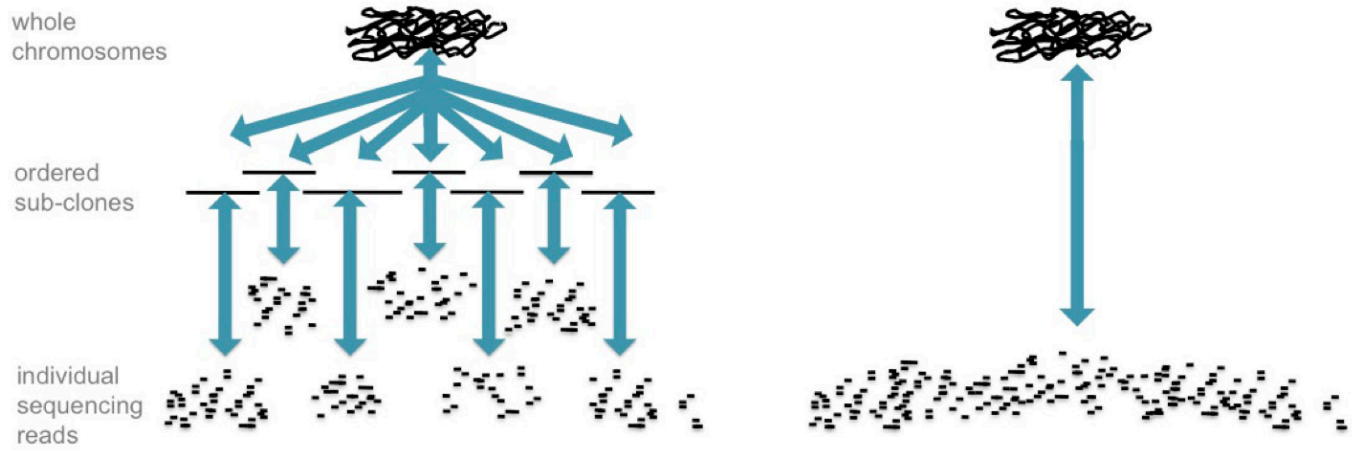


Figure 11.8
Genome sequencing statagies. A clone-by-clone strategy (left) in which the genome is divided into progressively smaller units (clones) before seqeuencing. Whole genome shotgun strategy (right) assembles the sequence from all the smaller reads. (Origianl-Deyholos-CC:AN)

11.3.5 Genome analysis

An assembled genome is a string of millions of A's,C's,G's,T's. Which of these represent nucleotides that encode proteins, and which of these represent other features of genes and their regulatory elements? The process of **genome annotation** relies on computers to define features such a start and stop codons,

introns, exons, and splice sites. However, few of the predictions made by these programs is entirely accurate, and most must be verified experimentally for any gene of particular importance or interest.

11.4 FUNCTIONAL GENOMICS – DETERMINING FUNCTION(S)

Having identified putative genes within a genome sequence, how do we determine their function? Techniques of functional genomics are an experimental approach to address this question. One widely used technique in functional genomics is called **microarray analysis** (Fig. 11.9). Microarrays can measure the abundance of mRNA for hundreds or thousands of genes at once. The abundance of mRNA of a particular gene is usually correlated with the activity of that gene. For example, genes that are involved in neuronal development likely produce more mRNA in brain tissue than in heart tissue. We can therefore learn about the relationship between particular genes and particular processes by comparing transcript abundance under different conditions. This can identify tissue specific expression (e.g. the nerve/heart example above), as well as differences in temporal expression (development), or exposure to external agents (eg, disease, hormones, drugs, etc.).

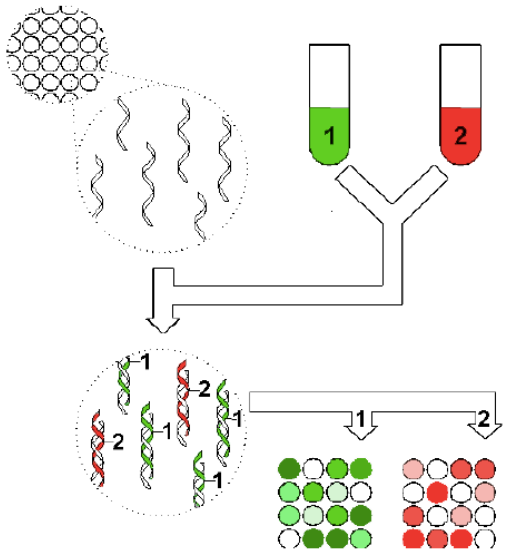


Figure 11.9 An example of a type of DNA microarray. Fluorescently labeled molecules derived from the transcripts of two tissue samples are hybridized to immobilized DNA molecules on the surface of an array. The labeled molecules bind in proportion to their abundance in the original tissue samples. The amount of green or red fluorescence at each spot can be measured using filters for the appropriate wavelengths. Thus, spots on the micoarray that are more green than red represent genes that are more abundant in the tissue sample from which green-labeled molecules were derived. (Original-Deyholos-CC:AN)

SUMMARY

- Genomics and related technologies differ from other techniques of molecular biology largely because of their scale; they allow many different genes (or gene products) to be studied in parallel.
- DNA sequencing can be applied to either a single gene, or in the case of genomics, to a large number of genes.
- Most DNA sequencing relies on the incorporation of dye-labeled terminator molecules, which create products that differ in length and end in a known nucleotide. The products can then be separated based on length, and the identity of the last based in each fragment can be determined based on fluorescence.
- Next-generation sequencing technologies have further reduced costs of sequencing, through miniaturization and parallelization.
- Physical maps are ordered sets of clones containing overlapping pieces of DNA, which together represent large pieces of chromosomes.
- Whole genomes may be sequenced using either a clone-by-clone approach, which required a physical map, or by a shotgun approach, in which small fragments are randomly sequenced.
- Genome analysis does not end after sequence acquisition; various features of the genome including genes (and their introns, exons, etc.) must be identified through a process called annotation.
- Functional genomics techniques including microarray analysis correlate transcript abundance with particular tissue samples. Genes whose transcripts are highly abundant under certain biological conditions may cause or respond to that condition.

KEY TERMS

genome
genomics
proteomics
transcriptomics
ddNTP
terminator nucleotide
capillary electrophoresis
chromatogram
next-generation sequencing
Ilumina
physical map
BAC
clone-by-clone sequencing
whole genome shotgun
genome annotation
functional genomics
microarray

STUDY QUESTIONS

11.1 What are the advantages of high-throughput –omics techniques compared to studying a single gene or protein at a time? What are the disadvantages

11.2 What would the chromatogram from a capillary sequencer look like if you accidentally added only template, primers, polymerase, and fluorescent terminators to the sequencing reaction?

11.3 What are the advantages and disadvantages of clone-by-clone vs. whole genome shotgun sequencing?

11.4 How could you use DNA sequencing to identify new species of marine microorganisms?

11.5 Explain how you could use a microarray to identify wheat genes that have altered expression during drought?

11.6 A microarray identified 100 genes whose transcripts are abundant in tumors, but absent in normal tissues. Do any or all of these transcripts cause cancer? Explain your answer.

11.7 How could you ensure that each spot printed on a microarray contains DNA for only one gene?

11.8 What would the spots look like on a microarray after hybridization, if each spot contained a random mixture of genes?

11.9 What would the spots look like if the hybridization of green and red labeled DNA was done at low stringency?

Chapter 12 REGULATION OF GENE EXPRESSION

Figure 12.1

The stickleback is an example of an organism in which mutations cause changes in the regulation of gene expression. These mutations confer a selective advantage in some environments. Natural selection acts on mutations altering gene expression as well as those changing the coding regions of genes. (Flickr-frequency-CC:AND)

Within most multicellular 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. In this chapter, we will focus on **transcriptional regulation**. 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.

12.1 THE *lac* OPERON

Early insights into mechanisms of transcriptional regulation came from studies of *E. coli* by researchers Francois Jacob & Jacques Monod. In *E. coli,* and many other bacteria, genes encoding several different proteins may be located on a single transcription unit called an **operon**. The genes in an operon share the same transcriptional regulation, but are translated individually. Eukaryotes generally do not group genes together as operons (exception is *C. elegans* and a few other species).

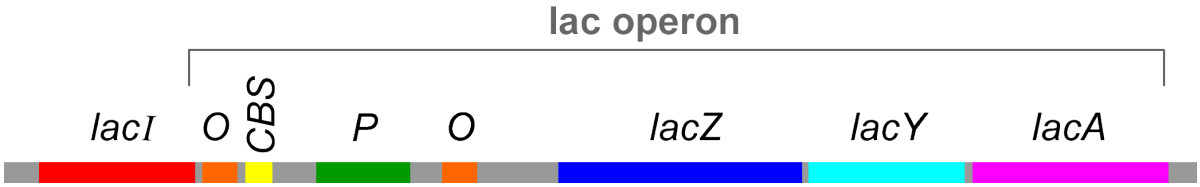


Figure 12.2

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. (Origianl-Deyholos-CC:AN)

12.1.1 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 12.2). *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 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. A single mRNA transcript includes all three enzyme-coding sequences and is called polycistronic. A cistron is equivalent to a gene.

12.1.2 *cis*- and *trans*- Regulators

In addition to the three protein-coding genes, the *lac* operon contains short DNA sequences that do not encode proteins, but are instead 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 on the same piece of DNA as the genes they regulate. On the other hand, 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.

12.1.3 *lacI* is an allosterically regulated repressor

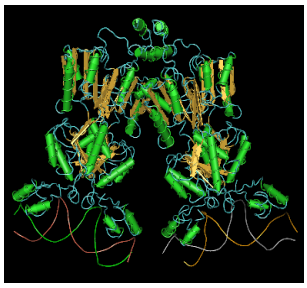


Figure 12.3

Structure of *lacI* homotetramer bound to DNA (Origianl-Deyholos-CC:AN)

One of the major *trans*-regulators of the *lac* operon is encoded by ***lacI***. Four identical molecules of *lacI* proteins assemble together to form a **homotetramer** called a **repressor** (Figure 12.3). This repressor binds to two operator sequences adjacent to the promoter of the *lac* operon. Binding of the repressor prevents RNA polymerase from binding to the promoter (Figure 12.4). Therefore, the operon will not be transcribed when the operator is occupied by a repressor.

Besides its ability to bind to specific DNA sequences at the operator, another important property of the *lacI* protein is its ability to bind to lactose. When lactose is bound to *lacI*, the shape of the protein changes in a way that prevents it from binding to the operator. Therefore, in the presence of lactose, RNA polymerase is able to bind to the promoter and transcribe the *lac* operon, leading to a moderate level of expression of the *lacZ*, *lacY*, and *lacA* genes. Proteins such as *lacI* that change their shape and functional properties after binding to a ligand are said to be regulated through an **allosteric** mechanism. The role of *lacI* in regulating the *lac* operon is summarized in Figure 12.4.

12.1.4 *CAP* IS AN ALLOSTERIC ACTIVATOR OF THE *LAC* OPERON

A second aspect of *lac* operon regulation is conferred by a *trans*-factor called **cAMP binding protein** (**CAP**, Figure 12.5). 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 located very close to the promoter (P). When CAP is bound to at 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.

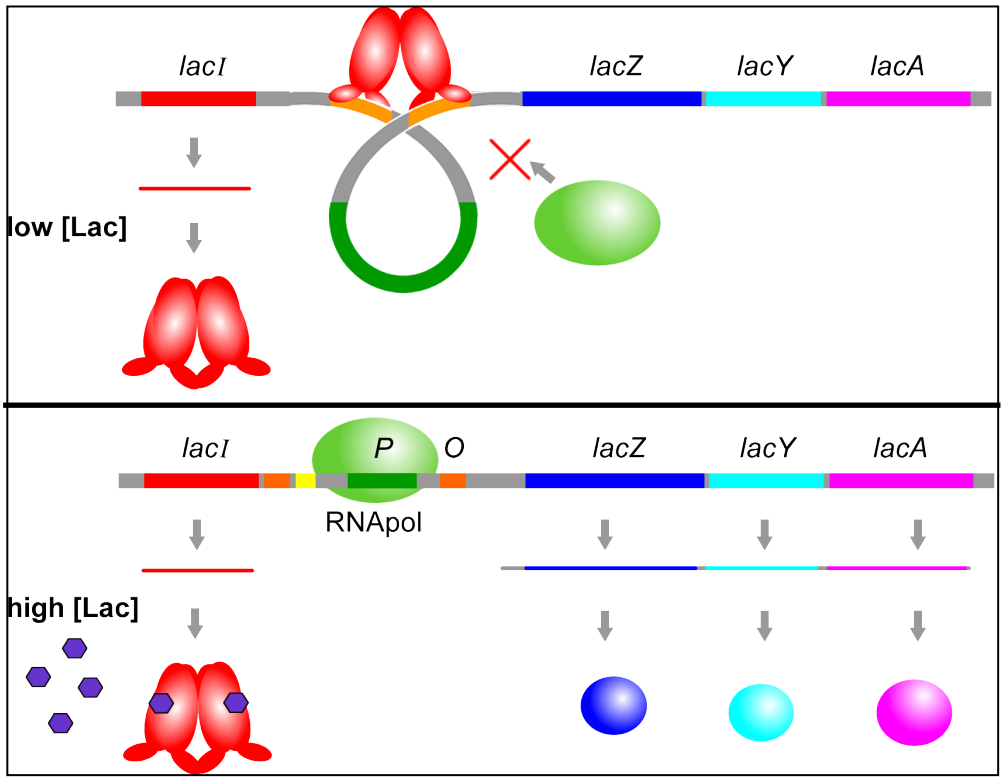


Figure 12.4 When the concentration of lactose [Lac] is low, *lacI* tetramers bind to operator sequences (O), thereby blocking binding of RNApol (green) to the promoter (P). Alternatively, when [Lac] is high, lactose binds to *lacI*, preventing the repressor from binding to O, and allowing transcription by RNApol. (Origianl-Deyholos-CC:AN)

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 logical 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.

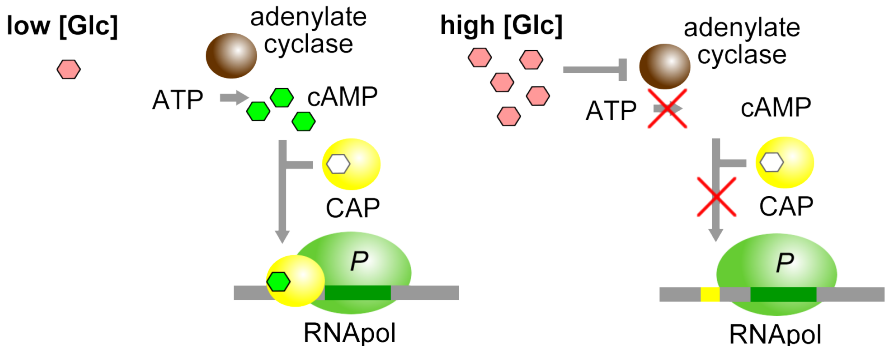


Figure 12.5 CAP, when bound to cAMP, helps RNApol to bind to the *lac* operon. cAMP is produced only when glucose [Glc] is low. (Origianl-Deyholos-CC:AN)

12.2 THE USE OF MUTANTS TO STUDY THE *lac* OPERON

12.2.1 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. Some mutants expressed the *lac* operon genes constitutively, meaning the operon was expressed whether or not lactose was present in the medium. Such mutant are called **constitutive** mutants.

The operator locus (*lacO*) - One example is O^c, in which a mutation in an operator sequence and reduces or precludes the repressor (the *lacI* gene product) from recognizing and binding to the operator sequence. Thus, in O^c mutants, *lacZ*, *lacY*, and *lacA* are expressed whether or not lactose is present.

The *lacI* locus – One type of mutant allele of *lacI* (callled I^-) prevents either the production of a repressor polypeptide or produces a polypeptide that cannot bind to the operator sequence. This is also a constitutive expresser of the *lac* operon because absence of repressor binding permits transcription.

Another type of mutant of *lacI* called I^s prevents the repressor polypeptide from binding lactose, and thus will bind to the operator and be non-inducible.. This mutant constitutively represses the *lac* operon whether lactose is present or not. The lac operon is not expressed and this mutant is called a "super-suppressor".

12.2.2 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 are present in one cell. This can be accomplished by using the **F-factor** to carry one copy, while the other is on the genomic *E. coli* chromosome. This results in a partial diploid in *E. coli.*

The F-factor is an **episome** that is capable of being either a free plasmid or integrated into the host bacterial chromosome. This switching is accomplished by IS elements where unequal crossing over can recombine the F-factor and adjacent DNA sequences (genes) in and out of the host chromosome. Researchers have used

this genetic tool to create partial diploids (merozygotes) that allow them to test the regulation with different combinations of different 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*- .

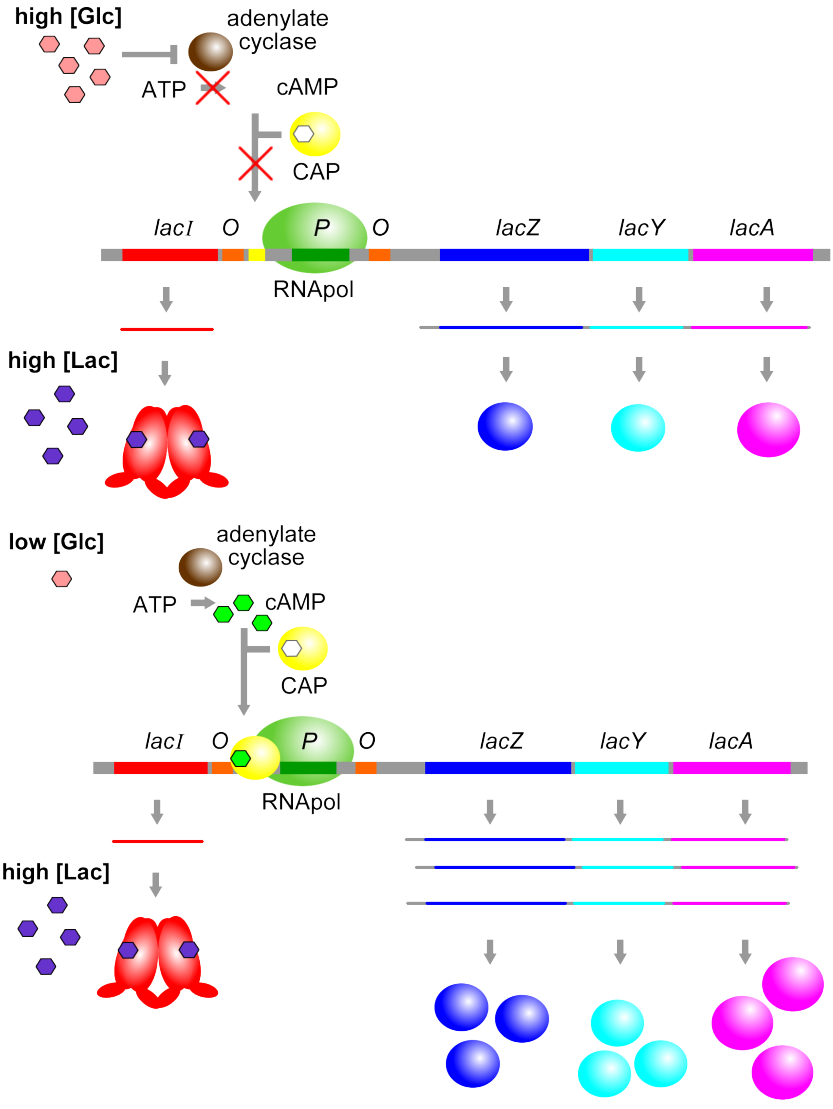


Figure 12.6 When glucose [Glc] and lactose [Lac] are both high, the *lac* operon is transcribed at a moderate level, because CAP (in the absence of cAMP) is unable to bind to its corresponding *cis*-element (yellow) and therefore cannot help to stabilize binding of RNApol at the promoter. Alternatively, 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. (Origianl-Deyholos-CC:AN)

12.3 EUKARYOTIC GENE REGULATION

Like prokaryotes, transcriptional regulation in eukaryotes involves both *cis*-elements and *trans*-factors, only there are more of them and they interact in a more complex way. A diagram of a typical eukaryotic gene, including several types of *cis*-elements, is shown in Figure 12.7.

12.3.1 PROXIMAL REGULATORY SEQUENCES.

As in prokaryotes the RNA polymerase binds to the gene at its promoter to begin transcription. In eukaryotes, however, RNApol is part of a large protein complex that includes additional proteins that bind to one or more specific *cis*-elements in the promoter region, including **GC-rich boxes**, **CAAT boxes**, **and TATA boxes**. High levels of transcription require both the presence of this protein complex at the promoter, as well as their interaction with other *trans*-factors described below. The approximate position of these elements relative to the **transcription start site** (+1) is shown in Figure 12.7, but it should be emphasized that the distance between any of these elements and the transcription start site can vary, but are typically within ~200 base pairs of the start of transcription. This contrasts the next set of elements.

12.3.2 DISTAL REGULATORY ELEMENTS

Even more variation is observed in the position and orientation of the second major type of *cis*-regulatory element in eukaryotes, which are called **enhancer elements**. Regulatory *trans*-factor proteins called **transcription factors** bind to enhancer sequences, then, while still bound to DNA, these proteins interact with RNApol and other proteins at the promoter to enhance the rate of transcription. There are a wide variety of different transcription factors and each recognizes a specific DNA sequence (enhancer element) to promote gene expression in the adjacent gene under specific circumstances. Because DNA is a flexible molecule, enhancers can be located near (~100s of bp) or far (~10K of bp), and either upstream or downstream, from the promoter (Figure 12.7 and 12.8).

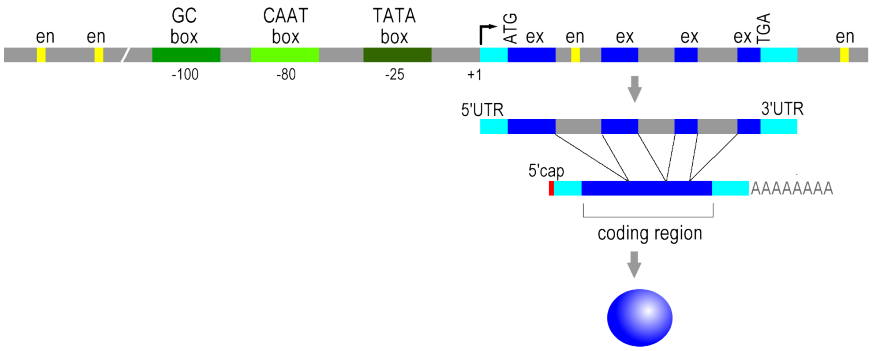


Figure 12.7
Structure of a typical eukaryotic gene. RNA polymerase binding may involve one or more *cis*-elements within the proximal region of a promoter (green boxes). Enhancers (yellow boxes) may be located any distance upstream or downstream of the promoter and are also involved in regulating gene expression. The processing of a primary transcript to a mature mRNA is also shown. (Origianl-Deyholos-GC:AN)

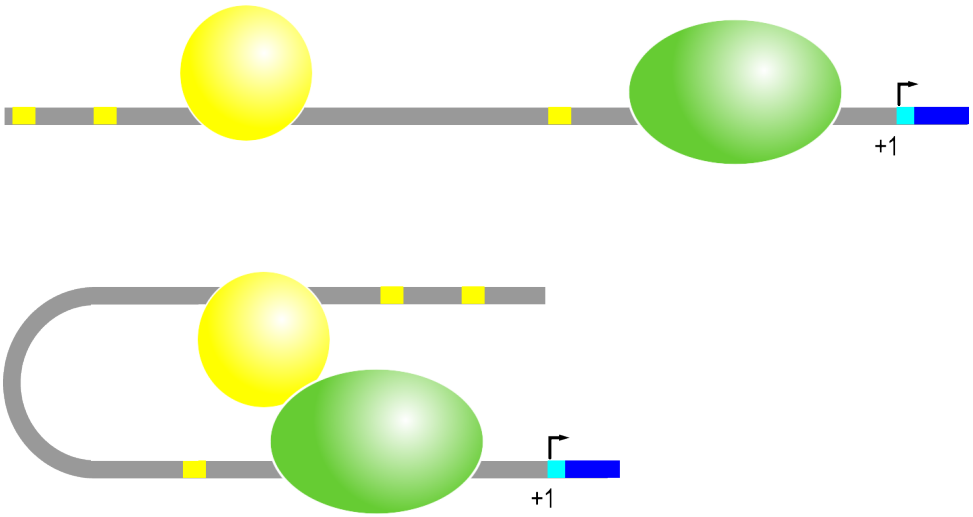


Figure 12.8 A transcription factor (yellow) bound to an enhancer that is located far from a promoter. Because of the flexibility of the DNA molecule, the transcription factor and RNApol are able to interact physically, even though the *cis*-elements to which they are bound are located far apart. In eukaryotic cells, RNApol is actually part of a large complex of proteins (not shown here) that assembles at the promoter. (Origianl-Deyholos-CC:AN)

12.3.3 Example 1: Drosophila yellow gene

The *yellow* gene of Drosophila provides an example of the modular nature of enhancers. This gene encodes an enzyme in the pathway that produces a dark pigment in the insect's exoskeleton. Mutants have a yellow cuticle rather than the wild type darker pigmented cuticle. (Why call the gene "yellow": recall that genes are often named after their mutant phenotype.) Figure 12.9 shows three enhancer elements (left side - wing, body, mouth), each binds a different tissue specific transcription factor to enhancer transcription of *yellow*$^+$ in that tissue and makes the pigment. So, the wing cells will have a transcription factor that binds to the wing enhancer to drive expression; likewise in the body and mouth cells. Thus, specific combinations of *cis*-elements and *trans*-factors control the differential, tissue-specific expression of genes. This type of combinatorial action of enhancers is typical of the transcriptional activation of most eukaryotic genes: specific transcription factors activate the transcription of target genes under specific conditions.

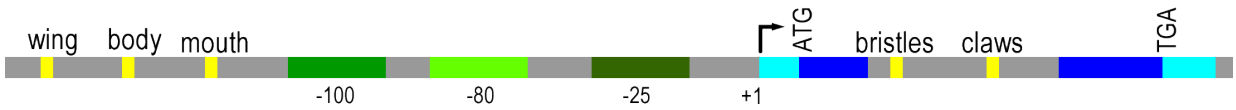


Figure 12.9
Tissue-specific *cis*-regulatory elements within a simplified representation of the *yellow* gene of Drosophila. (Origianl-Deyholos-CC:AN)

While enhancer sequences promote expression, there is an oppositely acting type of element, called **silencers**. These elements function in much the same manner, with transcription factors that bind to DNA sequences, but they act to silence or reduce transcription from the adjacent gene.

Again, a gene's expression profile (transcription level, tissue specific, temporal specific) is a combination of various enhancer and silencer elements.

12.3.4 Example 2: Gal4-UAS system from yeast – a genetic tool

Genetic researchers have taken advantage of a yeast distal enhancer sequence to make the **GAL4-UAS system**, a powerful technique for studying the expression of genes in other eukaryotes. It relies on two parts: a "**driver**" and a "**responder**" (Figure 12.10). The driver part is a gene encoding a yeast transcriptional activator protein called Gal4. It is separate from the responder part, which contains the enhancer sequence, or upstream activation sequence (UAS, as it is called in yeast) to which the Gal4 protein specifically binds. This UAS is placed upstream (using genetic engineering) from a promoter transcribing a reporter gene, or other gene of interest, such as GFP (green fluorescent protein).

Both parts must be present in the same cell for the system to express the responder gene. If the driver is absent, the responder product will not be expressed. However, both are in the same cell (or organism) the pattern of expression of the driver part will induce the responder part's expression in the same pattern.

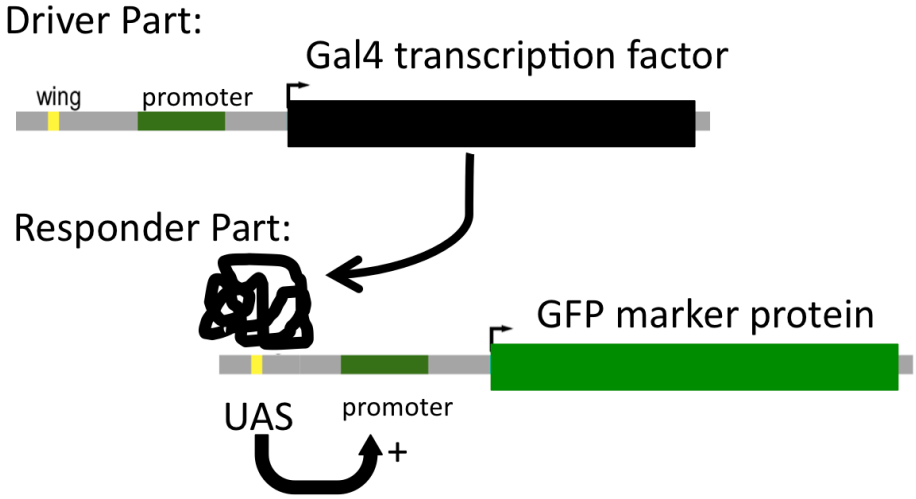


Figure 12.10

The GAL4-UAS system. The driver, with a wing enhancers, expresses the Gal4 protein that then binds to the UAS element upstream of a marker gene, GFP. This would express the GFP in the wing tissues. The modular aspect of this system would let the wing enhancer be replaced by any other enhancer and the GFP marker replaced with any other gene. (Origianl-Locke-CC:AN)

This system works is a variety of eukaryotes, including humans. It has been especially well exploited in Drosophila where 100's (1,000's ?) of differently expressing driver lines are available. These lines permit the tissue specific expression of any responder gene to examine its effect on development.

12.4 Regulatory elements in evolution

Mutations can occur in both *cis*-elements and *trans*-factors; both can result in altered patterns of gene expression. If an altered pattern of gene expression results in a selective advantage (or at least do not produce a major disadvantage), they may be selected and maintained in future populations. They may even contribute to the evolution of new species. An example of a sequence change in an enhancer is found in the *Pitx* gene.

12.4.1 EXAMPLE: *PITX* EXPRESSION IN STICKLEBACK

The three-spined **stickleback** (Figure 12.11) provides an example of natural selection of a mutation in a *cis*-regulatory element. This fish occurs in two forms: (1) populations that inhabit deep, open water and have a spiny pelvic fin that deters larger predator fish from feeding on them; (2) populations from shallow water environments and lack this spiny pelvic fin. In shallow water, it appears that a long, spiny pelvic fin would be a disadvantage because it frequently contacts the sediment at the bottom of the pond and allows parasitic insects in the sediment to invade the stickleback. Researchers compared gene sequences of individuals from both deep and shallow water environments as shown in Figure 12.11. They observed that in embryos from the deep-water population, a gene called *Pitx* was expressed in several groups of cells, including those that developed into the pelvic fin. Embryos from the shallow-water population expressed *Pitx* in the same groups of cells as the other population, with an important exception: *Pitx* was not expressed in the pelvic fin **primordium** in the shallow-water population. Further genetic analysis showed that the absence of *Pitx* gene expression from the developing pelvic fin of shallow-water stickleback was due to the absence (mutation) of a particular enhancer element upstream of *Pitx*.

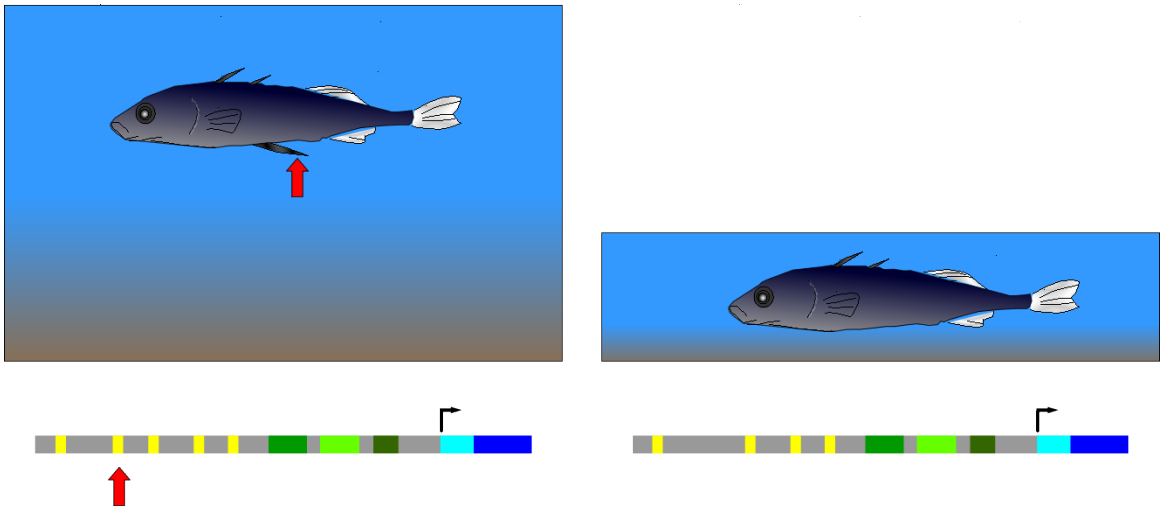


Figure 12.11 Development of a large, spiny pelvic fin in deep-water stickleback (left) depends on the presence of a particular enhancer element upstream of a gene called *Pitx*. Mutants lacking this element, and therefore the large pelvic fin (right), have been selected for in shallow-water environments. (Wikipedia-Richard Wheeler-GFDL)

12.4.2 Example: Hemoglobin expression in placental mammals.

Hemoglobin is the oxygen-carrying component of red blood cells (erythrocytes). Hemoglobin usually exists as tetramers of four non-covalently bound hemoglobin molecules (Fig 12.12). Each hemoglobin molecule consists of a **globin** polypeptide with a covalently attached heme molecule. Heme is made through a specialized metabolic pathway and is then bound to globin polypeptide through **post-translational modification**.

The composition of the tetramers changes during development (Fig. 12.13). From early childhood onward, most tetramers are of the type $\alpha_2\beta_2$, which means they contain of two copies of each of two slightly different globin proteins named α and β. A small amount of adult hemoglobin is $\alpha_2\delta_2$, which has δ globin instead of the more common β globin. Other tetrameric combinations predominate before birth: $\zeta_2\varepsilon_2$ is most abundant in embryos, and $\alpha_2\gamma_2$ is most abundant in fetuses. Although the six globin proteins (α = alpha, β = beta , γ = gamma, δ =delta, ε =epsilon , ζ = zeta) are very similar to each other, they do have slightly different functional properties. For example, fetal hemoglobin has a higher oxygen affinity than adult hemoglobin, allowing the fetus to more effectively extract oxygen from maternal blood. The specialized γ globin genes that are characteristic of fetal hemoglobin are found only in placental mammals.

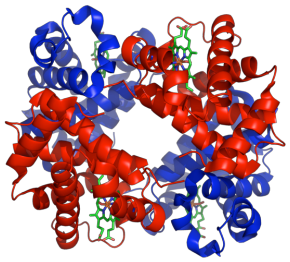


Figure 12.12

A tetramer of human hemoglobin, type α2β2. The α chains are labeled red, and the β chains are labeled blue. Heme groups are green. (Wikipedia-from Wood, W.G. 1976 Br. Med. Bull. 32, 282-CC:AS)

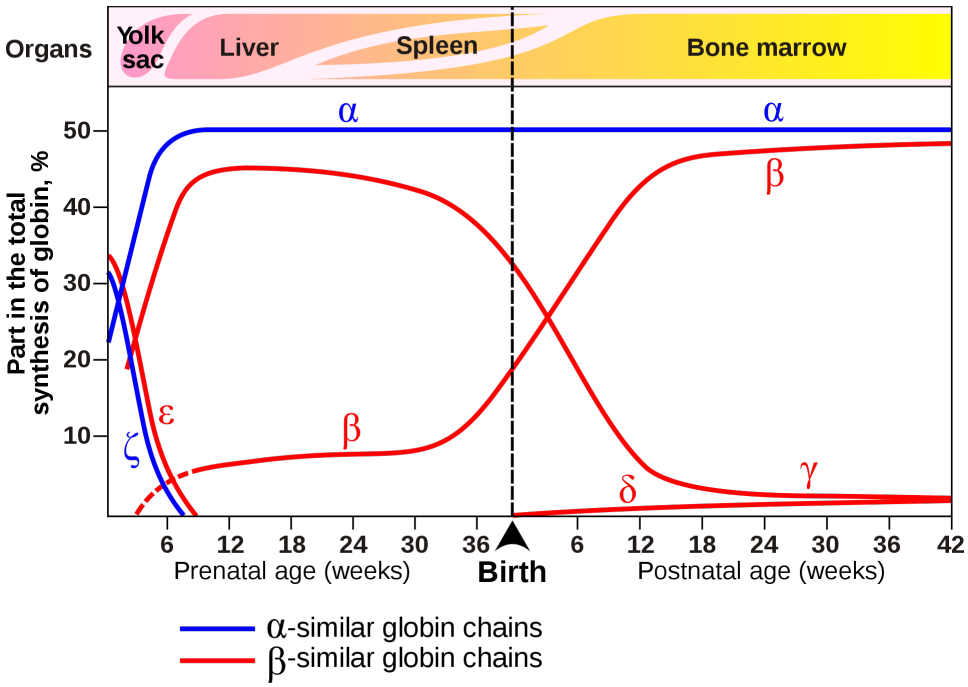


Figure 12.13 Expression of globin genes during prenatal and postnatal development in humans. The organs in which globin genes are primarily expressed at each developmental stage are also indicated. (Origianl-Deyholos-CC:AN)

Each of these globin polypeptides is encoded by a different gene. In humans, globin genes are located in clusters on two chromosomes (Figure 12.14). We can infer that these clusters arose through a series of duplications of an ancestral globin gene. **Gene duplication** events can occur through rare errors in processes such as DNA replication, meiosis, or transposition. The duplicated genes can accumulate mutations independently of each other. Mutations can occur in either the regulatory regions (e.g. promoter regions), or in the coding regions, or both. In this way, the promoters of globin genes have evolved to be expressed at different phases of development, and to produce proteins optimized for the prenatal environment.

Of course, not all mutations are beneficial: some mutations can lead to inactivation of one or more of the products of a gene duplication. This can produce what is called a **pseudogene**. Examples of pseudogenes (ψ) are also found in the globin clusters. Pseudogenes have mutations that prevent them from being expressed at all. The globin genes provide an example of how gene duplication and mutation, followed by selection, allows genes to evolve specialized expression patterns and functions. Many genes have evolved as **gene families** in this way, although they are not always clustered together as are the globins.

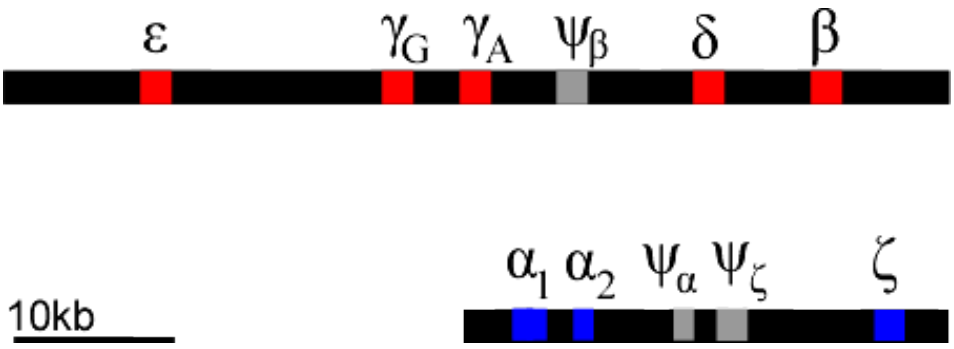


Figure 12.14 Fragments of human chromosome 11 and human chromosome 16 on which are located clusters of β-like and α-like goblin genes, respectively. Additional globin genes (θ, μ) have also been described by some researchers, but are not shown here. (Origianl-Deyholos-CC:AN)

12.5 ADDITIONAL LEVES OF REGULATING TRANSCRIPTION

Eukaryotes regulate transcription via promoter sequences close to the transcription unit (as in prokaryotes) and also use more distant enhancer sequences to provide more variation in the timing, level, and location of transcription, however, there are still additional levels of genetic control. This consists of two major mechanism: (1) large-scale changes in chromatin structure, and (2) modification of bases in the DNA sequence. These

12.5.1 CHROMATIN DYNAMICS

Despite the simplified way in which we often represent DNA in figures such as those in this chapter, DNA is almost always associated with various chromatin proteins. For example, histones remain associated with the DNA even during transcription. Thus the rate of transcription is also controlled by the accessibility of DNA to RNApol and regulatory proteins. So, in regions were the chromatin is highly compacted, it is unlikely that any gene will be transcribed, even if all the necessary *cis*- and *trans*- factors are present in the nucleus. The extent of chromatin compaction in various regions is regulated through the action of **chromatin remodeling** proteins. These protein complexes include enzymes that add or

remove chemical tags, such as methyl or acetyl groups, to various DNA bound proteins. These modifications alter the local chromatin density and thus the availability for transcription. **Acetylated** histones, for example, tend to be associated with actively transcribed genes, whereas **deacetylated** histone are associated with genes that are silenced.

Likewise, **methylation** of DNA itself is also associated with transcription regulation. Cytosine bases, particularly when followed by a guanine (**CpG sites**) are important targets for DNA methylation. Methylated cytosine within clusters of CpG sites is often associated with transcriptionally inactive DNA.

The modification of DNA and its associated proteins is enzymatically reversible (acetylation/deacetylation; methylation/demethylation) and thus a cyclical activity. Regulation of this provides another layer by which eukaryotic cells control the transcription of specific genes.

12.5.2 EPIGENETICS

Interestingly, some changes in gene expression appear heritable, but are not due to specific DNA sequence changes. For example, the grandchildren of famine victims are known to have lower birth weight than children without a family history of famine. This heritability of altered state of gene expression is surprising, since it appears not usually involve typical changes in the sequence of DNA. The term **epigenetics** describes any heritable change in phenotype that is associated with something other than chromosomal DNA sequence.

The basis of at least some types of epigenetic inheritance appears to be the replication of patterns of histone modifications and DNA methylation in parallel with the replication of the primary DNA sequence (see Section 12.5.3).

Epigenetics, in a simple way, can be used to describe the events occurring in cells of a multicellular organisms during development. Here, dividing cells commit to different differentiation pathways (e.g. blood cells vs nerve cells) and various genes are irreversibly silenced, through epigenetic mechanisms, in some cells, but not others. Additionally, some epigenetic information is inherited transgenerationally, but the permanence of this "change" is not the same as changes in the DNA sequence itself. What is clear is that epigenetics is an important part of regulating gene expression, and can serve as a type of cellular memory, certainly within an individual, or sometimes across a few generations.

Figure 12.15

A winter wheat crop (green) in early spring in the English countryside. (Flickr-Beardy Git-CC:AND)

12.5.3 VERNALIZATION AS AN EXAMPLE OF EPIGENETICS

Many plant species in temperate regions are **winter annuals**, meaning that their seeds germinate in the late summer, and grow vegetatively through early fall before entering a dormant phase during the winter, often under a cover of snow. In the spring, the plant resumes growth and is able to produce seeds before other species that germinated in the spring. In order for this life strategy to work, the winter annual must not resume growth or start flower production until winter has ended. **Vernalization** is the name given to the requirement to experience a long period of cold temperatures prior to flowering.

How does a plant sense that winter has passed? The signal for resuming growth cannot simply be warm air temperature, since occasional warm days, followed by long periods of freezing, are common in temperate climates. Researchers have discovered that winter annuals use epigenetic mechanisms to sense and "remember" that winter has occurred

Fortunately for the researchers who were interested in vernalization, some varieties of Arabidopsis are winter annuals. Through mutational analysis of Arabidopsis, researchers found that a gene called *FLC* (*FLOWERING LOCUS C*) encodes a transcription repressor acting on several of the genes involved in early stages of flowering (Figure 12.16). In the fall and under other warm conditions, the histones associated with *FLC* are acetylated and so *FLC* is transcribed at high levels; expression of flowering genes is therefore entirely repressed. However, in response to cold temperatures, enzymes gradually deacetylate the histones associated with *FLC*. The longer the cold temperatures persist, the more acetyl groups are removed from the *FLC*-associated histones, until finally the *FLC* locus is no longer transcribed and the flowering genes are free to respond to other environmental and hormonal signals that induce flowering later in the spring. Because the deacetylated state of *FLC* is inherited as cells divide and the plant grows in the early spring, this is an example of a type of cellular memory mediated by an epigenetic mechanism.

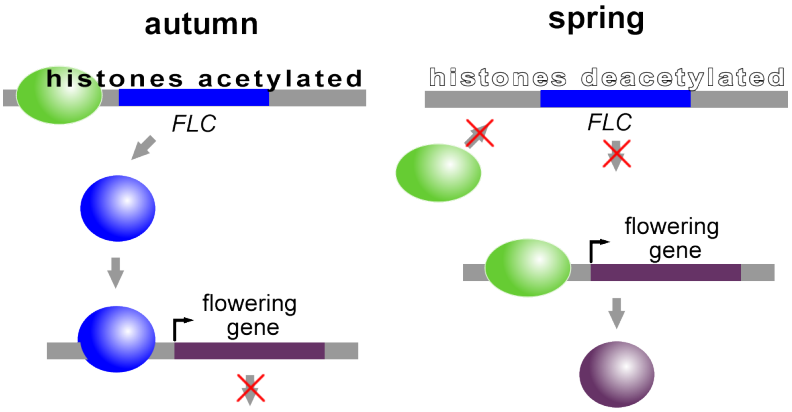


Figure 12.16

In the autumn, histones associated with FLC are acetylated, allowing this repressor of flowering genes to be expressed. During winter, enzymes progressive deacetylate FLC, preventing it from being expressed, and therefore allowing flowering genes to respond to other signals that induce flowering. (Origianl-Deyholos-CC:AN)

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.
- In eukaryotes, *cis*-elements that are usually called enhancers bind to specific *trans*-factors to regulate transcriptional initiation.
- Enhancers may be modular, with each enhancer and its transcription factor regulating a distinct component of a gene's expression pattern, as in the *yellow* gene.
- Sticklebacks provide examples of recent evolutionary events in which mutation of an enhancer produced a change in morphology and a selective advantage.
- Chromatin structure, including reversible modifications such as acetylation of histones, and methylation DNA CpG sites also regulates the initiation of transcription.
- Chromatin modifications or DNA methylation of some genes are heritable over many mitotic, and sometimes even meiotic divisions.
- Heritable changes in phenotype that do not result from a change in DNA sequence are called epigenetic. Many epigenetic phenomena involve regulation of gene expression by chromatin modification and/or DNA methylation.

KEY TERMS:

gene expression
transcriptional regulation
operon
lactose
glucose
lac operon
lacZ
lacY
lacA
galactosidase
permease
trans-acetylase
P / promoter
O / operator
CBS
CAP-binding site
cis-elements
trans-regulators
lacI
homotetramer
repressor
allosteric
cAMP binding protein
CAP
CAP binding sequence
CBS
adenylate cyclase
constitutive
O^c / I^- / I^s
F-factor / episome
GC boxes
CAAT boxes
TATA boxes
GAL4-UAS
Driver/responder
transcription start site
enhancers/silencers
transcription factors
hemoglobin/heme/globin
pseudogene
gene families
stickleback
primordium
chromatin remodeling
acetylation/deacetylation
methylation/demethylation
CpG sites
epigenetics
winter annual
vernalization
FLC

STUDY QUESTIONS:

12.1 List all the mechanisms that can be used to regulate gene expression in eukaryotes.

12.2 With respect to the expression of β-galactosidase, what would be the phenotype of each of the following strains of *E. coli*?

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^+, Oc, Z^+, Y^+ (no glucose, no lactose)
i) I^+, Oc, Z^+, Y^+ (no glucose, high lactose)
j) I^+, Oc, Z^+, Y^+ (high glucose, no lactose)
k) I^+, Oc, 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)

12.3 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.

a) $I^+, O^+, Z^+, Y^+ / O^-, Z^-, Y^-$ (high lactose)
b) $I^+, O^+, Z^+, Y^+ / O^-, Z^-, Y^-$ (no lactose)
c) $I^+, O^+, Z^-, Y^+ / O^-, Z^+, Y^+$ (high lactose)
d) $I^+, O^+, Z^-, Y^+ / O^-, Z^+, Y^+$ (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) $I^+, Oc, Z^+, Y^+ / I^+, O^+, Z^-, Y^+$ (high lactose)
j) $I^+, Oc, Z^+, Y^+ / I^+, O^+, Z^-, Y^+$ (no lactose)
k) $I^+, O^+, Z^-, Y^+ / I^+, Oc, Z^+, Y^+$ (high lactose)
l) $I^+, O^+, Z^-, Y^+ / I^+, Oc, Z^+, Y^+$ (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)

12.4 What genotypes of *E. coli* would be most useful in demonstrating that the *lacO* operator is a *cis*-acting regulatory factor?

12.5 What genotypes of E. coli would be useful in demonstrating that the *lacI* repressor is a *trans*-acting regulatory factor?

12.6 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

12.7 How are eukaryotic and prokaryotic gene regulation systems similar? How are they different?

12.8 Deep-water sticklebacks that are heterozygous for a loss-of-function mutation in the coding region of *Pitx* look just like homozygous wild-type fish from the same population. What phenotype or phenotypes would be observed if a wild-type fish from a deep-water population mated with a wild-type fish from a shallow-water population?

12.9 Some varieties of Arabidopsis, including those adopted for lab use, do not require vernalization before flowering. How might these varieties have evolved?

12.10 Histone deacetylase (HDAC) is an enzyme involved in gene regulation. What might be the phenotype of a winter annual plant that lacked HDAC function?

Chapter 13 CANCER GENETICS

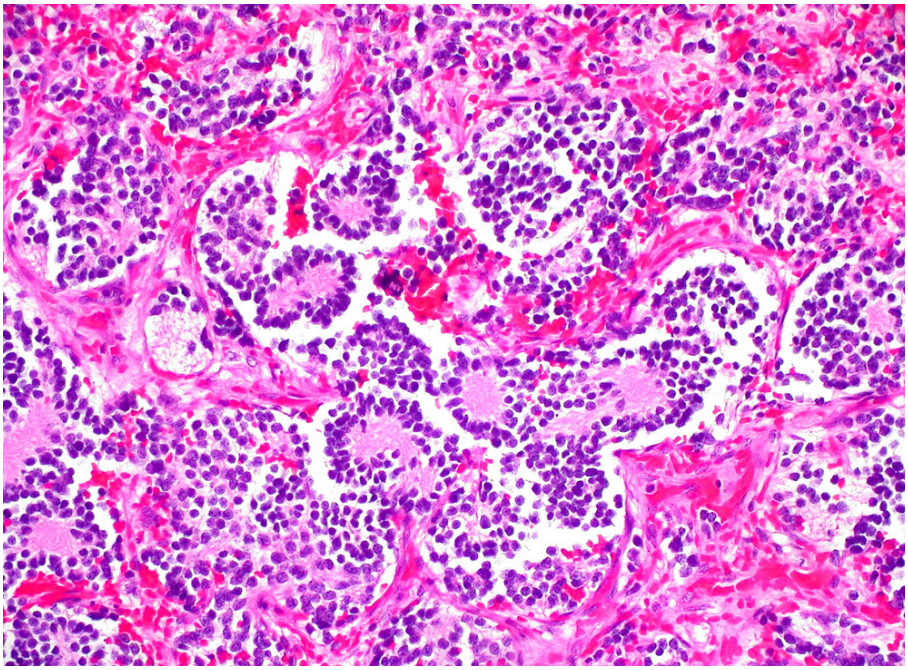


Figure 13.1
Stained histological section of a neuroblastoma in an adrenal gland. (Flickr-Ed Uthman-CC:AS)

Cancer is a group of diseases that exhibit uncontrolled growth, invasion of adjacent tissues, and sometimes **metastasis** (the movement of cancer cells through the blood or lymph). In cancer cells, the regulatory mechanisms that control cell division and limit abnormal growth have been disrupted, usually by the accumulation of several mutations. Cancer is therefore essentially a genetic disease. Although some cancer---related mutations may be heritable, most cancers are sporadic, meaning they arise from new mutations that occur in the individual who has the disease. In this chapter we will examine the connection between cancer and genes.

13.1 CLASSIFICATION OF CANCERS

Cancers can be classified based on the tissues in which they originate. **Sarcomas** are cancers that originate in mesoderm tissues, such as bone or muscle, and cancers arising in glandular tissues (e.g. breast, prostate) are classified as **adenocarcinomas**. **Carcinomas** originate in epithelial cells (both inside the body and on its surface) and are the most common types of cancer (~85%). Each of these classifications may be further sub---divided. For example, **squamous cell carcinoma (SCC)**, **basal cell carcinoma (BCC)**, and **melanoma** are all types of skin cancers originating respectively in the squamous cells, basal cells, or melanocytes of the skin.

13.2 CANCER CELL BIOLOGY

Cancer is a progressive disease that usually begins with increased frequency of cell division (Figure 13.2). Under the microscope, this may be detectable as increased cellular and nuclear size, and an increased proportion of cells undergoing mitosis.

As the disease progresses, cells typically lose their normal shape and tissue organization. Tissues with increased cell division and abnormal tissue organization exhibit **dysplasia**. Eventually a tumor develops, which can grow rapidly and expand into adjacent tissues. As cellular damage accumulates and additional control mechanisms are lost, some cells may break free of the primary tumor, pass into the blood or lymph system, and be transported to another organ, where they develop into new tumors (Figure 13.3). The early detection of tumors is important so that they can be treated or removed before the onset of metastasis, but note that not all tumors will lead to cancer. Tumors that do not metastasize are classified as **benign,** and are not usually considered life threatening. In contrast, **malignant** tumors become invasive, and ultimately result in cancer.

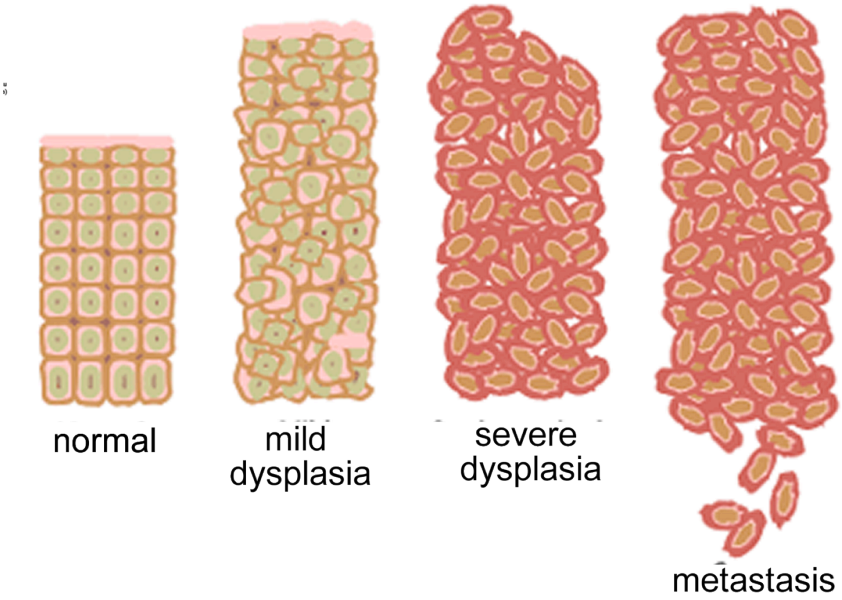


Figure 13.2
Progressive increases in cell division and abnormal cell morphology associated with cancer.
(Wikipedia-NIH-PD)

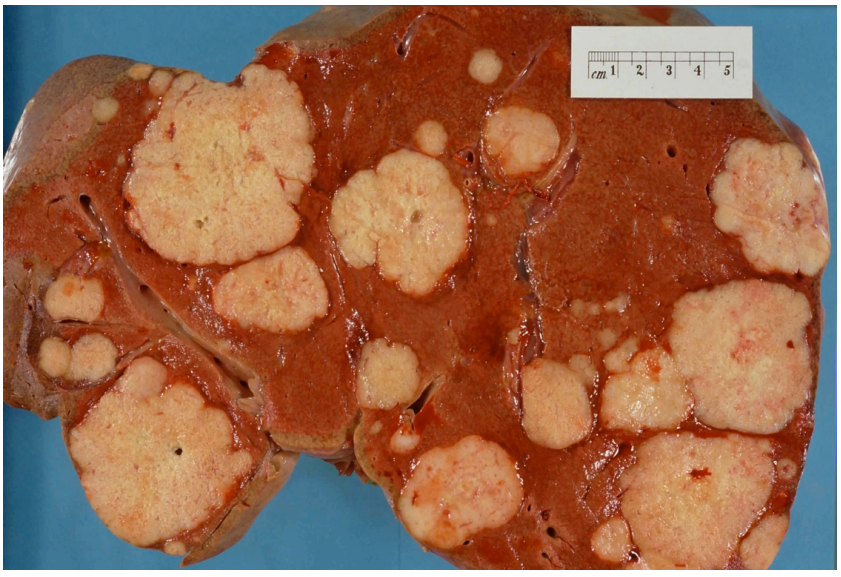


Figure 13.3
Secondary tumors (white) develop in the liver from cells of a metastatic pancreatic cancer.
(Wikipedia-J. Hayman-PD)

.

13.3 HALLMARKS OF CANCER

Researchers have identified six molecular and cellular traits that characterize most cancers. These six hallmarks of cancer are summarized in Table 13.1. In this chapter, we will focus on the first two hallmarks, namely growth signal autonomy and insensitivity to anti---growth signals.

Table 13.1 Ten Hallmarks of Cancer (Hanahan and Weinberg, 2000; Hanahan 2011)

1. Growth signal autonomy
   Cancer cells can divide without the external signals normally required to stimulate division.
2. Insensitivity to growth inhibitory signals
   Cancer cells are unaffected by external signals that inhibit division of normal cells.
3. Evasion of apoptosis
   When excessive DNA damage and other abnormalities are detected, apoptosis (a type of programmed cell death) is induced in normal cells, but not in cancer cells.
4. Reproductive potential not limited by telomeres
   Each division of a normal cell reduces the length of its telomeres. Normal cells arrest further division once telomeres reach a certain length. Cancer cells avoid this arrest and/or maintain the length of their telomeres.
5. Sustained angiogenesis
   Most cancers require the growth of new blood vessels into the tumor. Normal angiogenesis is regulated by both inhibitory and stimulatory signals not required in cancer cells.
6. Tissue invasion and metastasis
   Normal cells generally do not migrate (except in embryo development). Cancer cells invade other tissues including vital organs.
7. Deregulated metabolic pathways
   Cancer cells use an abnormal metabolism to satisfy a high demand for energy and nutrients.
8. Evasion of the immune system
   Cancer cells are able to evade the immune system.
9. Chromosomal instability
   Severe chromosomal abnormalities are found in most cancers.
10. Inflammation
    Local chronic inflammation is associated with many types of cancer.

13.4 MUTAGENS AND CARCINOGENS

A **carcinogen** is any agent that directly increases the incidence of cancer. Most, but not all carcinogens are mutagens. Carcinogens that do not directly damage DNA include substances that accelerate cell division, thereby leaving less opportunity for cell to repair induced mutations, or errors in replication. Carcinogens that act as mutagens may be biological, physical, or chemical in nature, although the term is most often used in relation to chemical substances.

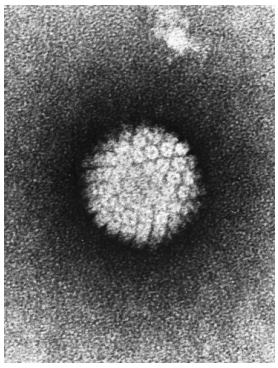


Figure 13.4
Electron micrograph of HPV.
(Wikipedia-Unknown-PD)

Human Papilloma Virus (**HPV**, Figure 13.4) is an example of a biological carcinogen. Almost all cervical cancers begin with infection by HPV, which contains genes that disrupt the normal pattern of cell division within the host cell. Any gene that leads to an uncontrolled increase in cell division is called an **oncogene**. The HPV E6 and E7 genes are considered oncogenes because they inhibit the host cell's natural tumor suppressing proteins (include p53, described below). The product of the E5 gene mimics the host's own signals for cell division, and these and other viral gene products may contribute to dysplasia, which is detected during a Pap smear (Figure 13.5). Detection of abnormal cell morphology in a Pap smear is not necessarily evidence of cancer. It must be emphasized again that cells have many regulatory mechanisms to limit division and growth, and for cancer to occur, each of these mechanisms must be disrupted. This is one reason why only a minority of individuals with HPV infections ultimately develop cancer. Although most HPV-related cancers are cervical, HPV infection can also lead to cancer in other tissues, in both women and men.

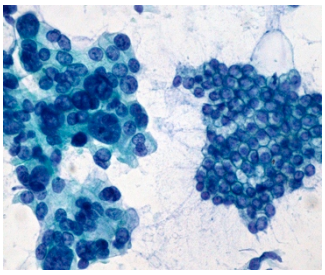


Figure 13.5
Dysplastic (left) and normal (right) cells from a Pap smear.
(Flickr-Ed Uthman-CC:AS)

.

Radiation is a well-known physical carcinogen, because of its potential to induce DNA damage within the body. The most damaging type of radiation is **ionizing**, meaning waves or particles with sufficient energy to strip electrons from the molecules they encounter, including DNA or molecules that can subsequently react with DNA. Ionizing radiation, which includes x-rays, gamma rays, and some wavelengths of ultraviolet rays, is distinct from the non-ionizing radiation of microwave ovens, cell phones, and radios. As with other carcinogens, mutation of multiple, independent genes that normally regulate cell division is required before cancer develops.

Chemical carcinogens (Table 13.2) can be either natural or synthetic compounds that, based on animal feeding trials or **epidemiological** (i.e. human population) studies, increase the incidence of cancer. The definition of a chemical as a carcinogen is problematic for several reasons. Some chemicals become carcinogenic only after they are metabolized into another compound in the body; not all species or individuals may metabolize chemicals in the same way. Also, the carcinogenic properties of a compound are usually dependent on its dose. It can be difficult to define a relevant dose for both lab animals and humans. Nevertheless, when a correlation between cancer incidence and chemical exposure is observed, it is usually possible to find ways to reduce exposure to that chemical.

Table 13.2 Some classes of chemical carcinogens (Pecorino 2008)
1. PAHs (polycyclic aromatic hydrocarbons) e.g. benzo[a]pyrene and several other components of the smoke of cigarettes, wood, and fossil fuels 2. Aromatic amines e.g. formed in food when meat (including fish, poultry) are cooked at high temperature 3. Nitrosamines and nitrosamides e.g. found in tobacco and in some smoked meat and fish 4. Azo dyes e.g. various dyes and pigments used in textiles, leather, paints. 5. Carbamates e.g. ethyl carbamate (urethane) found in some distilled beverages and fermented foods 6. Halogenated compounds e.g. pentachlorophenol used in some wood preservatives and pesticides. 7. Inorganic compounds e.g. asbestos; may induce chronic inflammation and reactive oxygen species 8. Miscellaneous compounds e.g. alkylating agents, phenolics

13.5 ONCOGENES

The control of cell division involves many different genes. Some of these genes act as signaling molecules to activate normal progression through the cell cycle. One of the pre-requisites for cancer occurs when one or more of these activators of cell division become mutated.

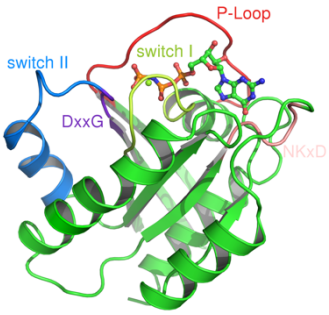


Figure 13.6
Structure of the *ras* protein.
(Wikipedia-Mark"AbsturZ"-PD)

The mutation may involve a change in the coding sequence of the protein, so that it is more active than normal, or a change in the regulation of its expression, so that it is produced at higher levels than normal, or persists in the cell longer than normal. Genes that are a part of the normal regulation of cell division, but which after mutation contribute to cancer, are called **proto-oncogenes**. Once a proto-oncogene has been abnormally activated by mutation, it is called an oncogene. More than 100 genes have been defined as proto-oncogenes. These include genes at almost every step of the signaling pathways that normally induce cell to divide, including growth factors, **receptors**, **signal transducers**, and transcription factors.

ras is an example of a proto-oncogene. *ras* acts as a switch within signal transduction pathways, including the regulation of cell division. When a receptor protein receives a signal for cell division, the receptor activates *ras*, which in turn activates other signaling components, ultimately leading to activation of genes involved in cell division. Certain mutations of the *ras* sequence causes it to be in a permanently active form, which can lead to constitutive activation of the cell cycle. This mutation is dominant as are most oncogenes. An example of the role of ras in relaying a signal for cell division in the EGF pathway is shown in Figure 13.7.

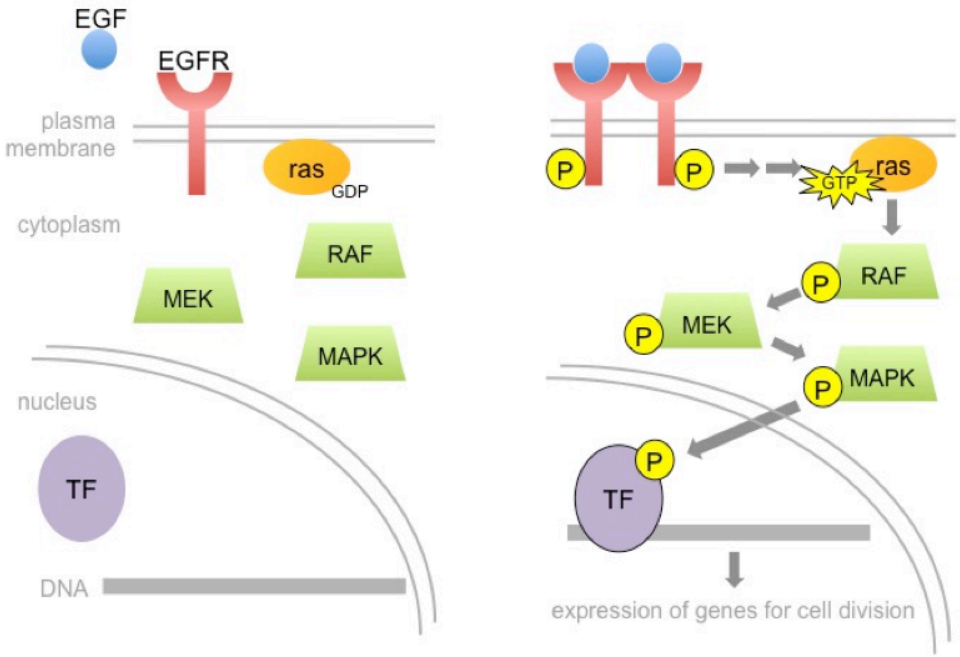


Figure 13.7 Simplified representation of the epidermal growth factor (EGF) signaling pathway. In the panel on the left, the components are shown in their inactive forms, prior to stimulation of the pathway. The components include the soluble ligand, EGF, its receptor (EGFR, a tyrosine kinase), ras (a G protein), several kinases (RAF, MEK, MAPK), and a transcription factor (TF). In the right panel, the activate pathway is shown. Binding of the ligand to its receptor leads to autophosphorylation of the receptor. Through a series of proteins not shown here, the phosphorpylated simulates conversion of ras to its active, GTP-bound form. The activated ras then stimulates phosphorylation of a series of kinases, which ultimately activate transcription factors and the expression of genes required for cell proliferation. (Original-Deyholos-CC:AN)

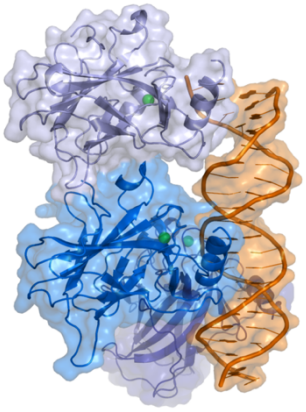


Figure 13.8 p53 bound to its target site on a DNA molecule. (Wikipedia-Thomas Spettstoesser from Cho etal, Science 265PP346, 1994-CC:AS)

13.6 TUMOR SUPPRESSOR GENES

More than 30 genes are classified as **tumor suppressors**. The normal functions of these genes include repair of DNA, induction of programmed cell death (**apoptosis**) and prevention of abnormal cell division. In contrast to proto-oncogenes, in tumor suppressors it is loss-of-function mutations that contribute to the progression of cancer. This means that tumor suppressor mutations tend to be recessive, and thus both alleles must be mutated in order to allow abnormal growth to proceed. It is perhaps not surprising that mutations in tumor suppressor genes, are more likely than oncogenes to be inherited. An example is the tumor suppressor gene, ***BRCA1***, which is involved in DNA-repair. Inherited mutations in *BRCA1* increase a woman's lifetime risk of breast cancer by up to seven times, although these heritable mutations account for only about 10% of breast cancer. Thus, sporadic rather than inherited mutations are the most common sources of both oncogenes and disabled tumor suppressor genes.

An important tumor suppressor gene is a transcription factor named **p53**. Other proteins in the cell sense DNA damage, or abnormalities in the cell cycle and activate

p53 through several mechanisms including **phosphorylation** (attachment of phosphate to specific site on the protein) and transport into the nucleus. In its active form, p53 induces the transcription of genes with several different types of tumor suppressing functions, including DNA repair, cell cycle arrest, and apoptosis. Over 50% of human tumors contain mutations in p53. People who inherit only one function copy of p53 have a greatly increased incidence of early onset cancer. However, as with the other cancer related genes we have discussed, most mutations in p53 are sporadic, rather than inherited. Mutation of p53, through formation of pyrimidine dimers in the genes following exposure to UV light, has been causally linked to squamous cell and basal cell carcinomas (but not melanomas, highlighting the variety and complexities of mechanisms that can cause cancer).

SUMMARY:

- Cancer is the name given to a class of different diseases that share common properties.
- Most cancers require accumulation of mutations in several different genes.
- Most cancer causing mutations are sporadic, rather than inherited, and most are caused by environmental carcinogens, including virus, radiation, and certain chemicals.
- Oncogenes are hyperactivated regulators of cell division, and are often derived from gain-of-function mutations in proto-oncogenes.
- Tumor suppressor genes normal help to repair DNA damage, arrest cell division, or to kill over proliferating cells. Loss-of-function of these genes contributes to the progression of cancer.

KEY TERMS:

metastasis
dysplasia
carcinogen
HPV
oncogene
ionizing
epidemiology
proto-oncogene
receptor
signal transduction
ras
apoptosis
BRC1A
p53
tumor suppressor
phosphorylation

STUDY QUESTIONS:

13.1 Why do oncogenes tend to be dominant, but mutations in tumor suppressors tend to be recessive?

13.2 What tumor suppressing functions are controlled by p53? How can a single gene affect so many different biological pathways?

13.3 Are all carcinogens mutagens? Are all mutagens carcinogens? Explain why or why not.

13.4 Imagine that a laboratory reports that feeding a chocolate to laboratory rats increases the incidence of cancer. What other details would you want to know before you stopped eating chocolate?

13.5 Do all women with HPV get cancer? Why or why not? Do all women with mutations in *BRCA1* get cancer? Why or why not?

Answers for Open Genetics
Fall 2013 version

CHAPTER 1 - ANSWERS

1.1 If genetic factors blended together like paint then they could not be separated again. The white flowered phenotype would therefore not reappear in the F_2 generation, and all the flowers would be purple or maybe light purple.

1.2 **a)** Identify pure breeding lines of the individuals that differed in some detectable trait, then cross the lines with the different traits and see how the traits were inherited over several generations.

b) Purify different biochemical components, then see if any of the components were sufficient to transfer traits from one individual to another.

c) It depends in part whether the organisms all evolved from the same ancestor. If so, then it seems likely.

d) The extraterrestrials would not necessarily (and perhaps would be unlikely) to have the same types of reductional divisions of chromosome-like material prior to sexual reproduction. In other words, there are many conceivable ways to accomplish what sex, meiosis, and chromosomes accomplish on earth.

1.3 Hershey and Chase wanted to be able to track DNA and protein molecules from a specific source, within a mixture of other protein and DNA molecules. Radioactivity is a good way to label molecules, since detection is quite sensitive and the labeling does not interfere with biological function**.**

1.4 **a)** Avery and colleagues demonstrated that DNA was likely the genetic material, while Watson and Crick demonstrated the structure of the molecule. By knowing the structure, it was possible to understand how DNA replicated, and how it encoded proteins, etc.

b) Avery and colleagues performed experiments, while Watson and Crick mostly analyzed the data of others and used that to build models.

c) Watson and Crick relied on Franklin's data in building their model. It is controversial whether Watson and Crick should have been given access to these data.

1.5 The experiments shown in Figure 1.4 show that DNA is necessary for transformation, (since removing the DNA by nuclease treatment removes the competency for transformation). However, this does not demonstrate that DNA is sufficient to transfer genetic information; you could therefore try to purify S strain DNA and see if injecting that DNA alone could transform R strains into S strains.

1.6 Chargaff's Rules, X-ray crystallography data, and Avery, MacLeod & McCarty and Hershey & Chase's data, as well as other information (e.g. specific details about the structure of the bases).

1.7 **a)** Right-handed, anti-parallel double helix with a major and minor groove. Each strand is composed of sugar-nucleotide bases linked together by covalent phosphodiester bonds. Specific bases on opposite strands of the helix pair together through hydrogen bonding, so that each strand contains the same information in a complementary structure.

b) The complimentarity of the bases and the redundant nature of the strands.

c) The order of the bases.

1.8 Mutant strain #1 has a mutation in gene B (but genes A and C are functional).
Mutant strain #2 has a mutation in gene A (but genes B and C are functional).
Mutant strain #3 has a mutation in gene C (but genes A and B are functional).

1.9 a) There is little correlation between any of these.
b) Genomes have different amounts of non-coding DNA between genes.
c) No.

1.10 a) Fast and simple to grow in high density, diploid,
b) i) zebrafish (for vertebrate eyes); flies for eyes in general
ii) zebrafish
iii) Arabidopsis
iii) yeast
iv) *C. elegans*
v) arguably, any of the organisms, but the vertebrates would be most relevant

1.11 a) Hershey & Chase labeled the phosphate groups that join the bases
b) G-C pairs have more hydrogen bonds, so more energy is required to break the larger number of bonds in a G-C rich region as compared to an A-T rich region.

Chapter 2 - Answers

2.1 Chromatin is the material from which chromosomes are made (mostly DNA + protein). DNA is a component of both chromatin and chromosomes.

2.2 No. Since chromosomes vary greatly in size, the number of chromosomes does not correlate with the total DNA content. For reasons discussed in Chapter 1, the number of genes does not correlated closely to DNA content either.

2.3 Heterochromatic regions with repetitive DNA, centromeres, and telomeres are examples of gene-poor regions of chromosomes.

2.4 a) One, except for holocentric chromosomes.
b) The chromosome might get pulled apart.
c) There would have to be another way to control its movement at mitosis and meiosis.

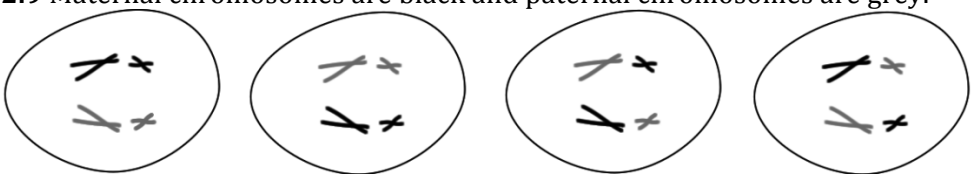
2.5 Immediately following G_1, 16 chromosomes (8 pairs) with 1 chromatid each.
Immediately following S, 16 chromosomes with 2 chromatids each.
Immediately following G_2, 16 chromosomes with 2 chromatids each.
Immediately following mitosis, 16 chromosomes with 1 chromatid each.
Immediately following meiosis 8 chromosomes with 1 chromatids each.

2.6 2n=6x=42

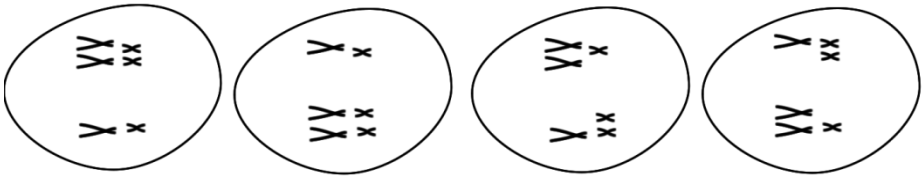
2.7 a) Two is the maximum number of alleles that can exist for a given gene in a 2n cell of a given diploid individual.
b) Two is the maximum number of alleles that can exist in a 1n cell of a tetraploid individual.
c) Four is the maximum number of alleles that can exist in a 2n cell of a tetraploid individual. .
d) The maximum number of alleles that can exist in a population is theoretically limited only by the population size.

2.8 a) Aneuploidy can disrupt gene balance and disrupt meiosis, whereas even-numbered polyploids (e.g. tetraploid, hexaploid) can be stable through meiosis, and can retain normal gene balance.
b) Duplication is more likely than polyploidy to disrupt gene balance since only some genes will increase their copy number following duplication of a chromosomal segment.

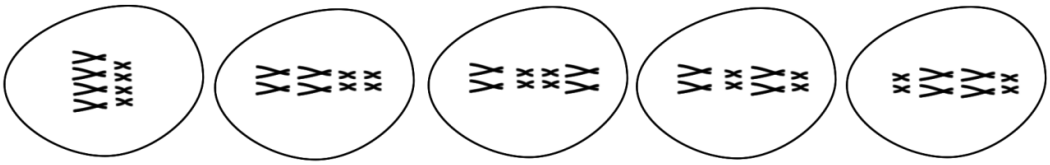
2.9 Maternal chromosomes are black and paternal chromosomes are grey.



2.10



2.11



CHAPTER 3 - ANSWERS

3.1 There is a maximum of two alleles for a normal autosomal locus in a diploid species.

3.2 a) In the F_1 generation, the genotype of all individuals will be *Ww* and all of the dogs will have wirey hair.
b) In the F_2 generation, there would be an expected 3:1 ratio of wirey-haired to smooth-haired dogs.
c) Although it is expected that only one out of every four dogs in the F2 generation would have smooth hair, large deviations from this ratio are possible, especially with small sample sizes. These deviations are due to the random nature in which gametes combine to produce offspring. Another example of this would be the fairly common observation that in some human families, all of the offspring are either girls, or boys, even though the expected ratio of the sexes is essentially 1:1.
d) You could do a test cross, i.e. cross the wirey-haired dog to a homozygous recessive dog (*ww*). Based on the phenotypes among the offspring, you might be able to infer the genotype of the wirey-haired parent.
e) From the information provided, we cannot be certain which, if either, allele is wild-type. Generally, dominant alleles are wild-type, and abnormal or mutant alleles are recessive.

3.3 Even before the idea of a homozygous genotype had really been formulated, Mendel was still able to assume that he was working with parental lines that contained the genetic material for only one variant of a trait (e.g. EITHER green seeds of yellow seeds), because these lines were pure-breeding. Pure-breeding means that the phenotype doesn't change over several generations of self-pollination. If the parental lines had not been pure-breeding, it would have been very hard to make certain key inferences, such as that the F_1 generation could contain the genetic information for two variants of a trait, although only one variant was expressed. This inference led eventually to Mendel's First Law.

3.4 Equal segregation of alleles occurs only in meiosis. Although mitosis does produce daughter cells that are genetically equal, there is no segregation (i.e. separation) of alleles during mitosis; each daughter cell contains both of the alleles that were originally present.

3.5 If your blood type is B, then your genotype is either I^Bi or I^BI^B. If your genotype is I^Bi, then your parents could be any combination of genotypes, as long as one parent had at least one *i* allele, and the other parent had at least one I^B allele. If your genotype was $I^B\ I^B$, then both parents would have to have at least one I^B allele.

3.6 case 1 co-dominance
case 2 incomplete-dominance
case 3 incomplete penetrance
case 4 pleiotropy
case 5 haplosufficiency

case 6 haploinsufficiency
case 7 broad (variable) expressivity

3.7 If the gene is autosomal, the probability is 50%. If it is sex-linked, 100%. In both situations the probability would decrease if the penetrance was less than 100%.

3.8

Cross (a)

♂ \ ♀	w^+	w^+
w^1	w^+/w^1 red eyed female	w^+/w^1 red eyed female
Y	w^+/Y red eyed male	w^+/Y red eyed male

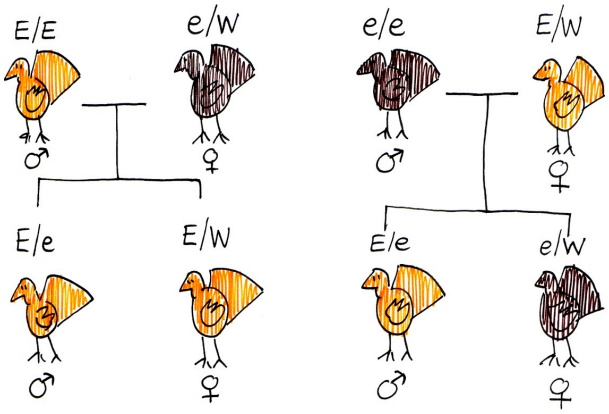
Cross (b)

♂ \ ♀	w^1	w^1
w^+	w^+/w^1 red eyed female	w^+/w^1 red eyed female
Y	w^1/Y white eyed male	w^1/Y white eyed male

3.9 Note that a semicolon is used to separate genes on different chromosomes.

Phenotype	**Genotype(s)**
a) entirely black	*red*^{o} / *red*^{o} ; *s* / *s* *red*^{o} / Y ; *s* / *s*
b) entirely orange	*red*^{0} / *red*^{0} ; *s* / *s* *red*^{0} / Y ; *s* / *s*
c) black and white	*red*^{o} / *red*^{o} ; *S* / _ *red*^{o} / Y ; *S* / _
d) orange and white	*red*^{0} / *red*^{0} ; *S* / _ *red*^{0} / Y ; *S* / _
e) orange and black (tortoiseshell)	*red*^{0} / *red*^{o} ; *s* / *s*
f) orange, black, and white (calico)	*red*^{0} / *red*^{o} ; *S* / _

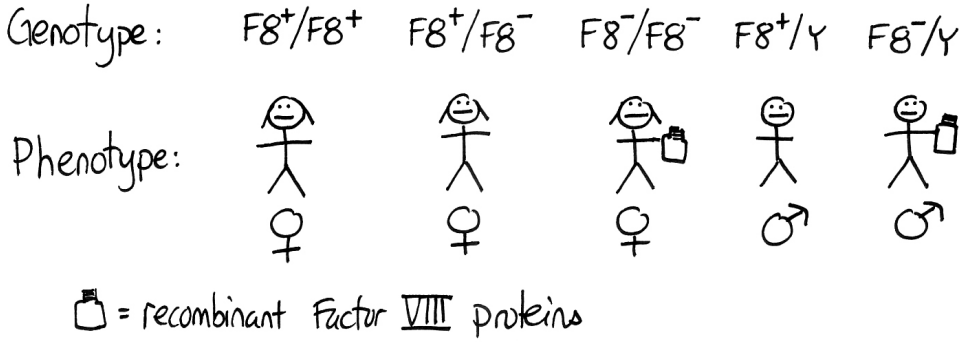
3.10



3.11 Because each egg or sperm cell receives exactly one sex chromosome (even though this can be either an X or Y, in the case of sperm), it could be argued that the sex chromosomes themselves do obey the law of equal segregation, even though the alleles they carry may not always segregate equally. However, this answer depends on how broadly you are willing to stretch Mendel's First Law.

3.12 Co-dominance

3.13 People with hemophilia A use injections of recombinant Factor VIII proteins on demand (to control bleeding) or regularly (to limit damage to joints).



notes

Chapter 4 - Answers

4.1 Polymorphisms and mutations are both variations in DNA sequence and can arise through the same mechanisms. We use the term polymorphism to refer to DNA variants that are relatively common in populations. Mutations affect the phenotype.

4.2 Misreading of bases during replication can lead to substitution and can be caused by things like tautomerism, DNA alkylating agents, and irradiation.

4.3 Looping out of DNA on the template strand during replication; strand breakage, due to radiation and other mutagens; and (discussed in earlier chapters) chromosomal aberrations such as deletions and translocations.

4.4 Looping out of DNA on the growing strand during replication; transposition; and (discussed in earlier chapters) chromosomal aberrations such as duplications, insertions, and translocation.

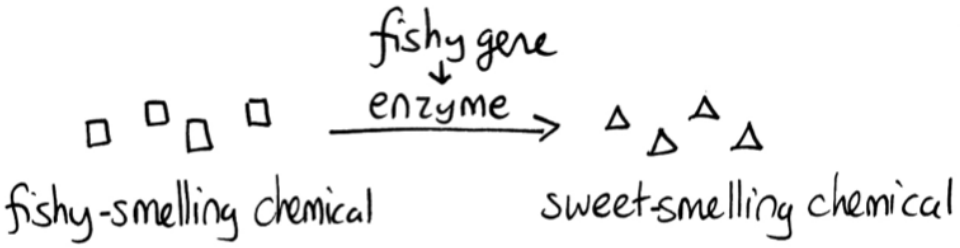
4.5 Benzopyrene is one of many hazardous compounds present in smoke. Benzopyrene is an intercalating agent, which slides between the bases of the DNA molecule, distorting the shape of the double helix, which disrupts transcription and replication and can lead to mutation.

4.6 a) One possible explanation is that original mutagenesis resulted in a loss-of-function mutation in a gene that is essential for early embryonic development, and that this mutation is X-linked recessive in the female. Because half of the sons will inherit the X chromosome that bears this mutation, half of the sons will fail to develop beyond very early development and will not be detected among the F_1 progeny. The proportion of male flies that were affected depends on what fraction of the female parent's gametes carried the mutation. In this case, it appears that half of the female's gametes carried the mutation.

b) To test whether a gene is X-linked, you can usually do a reciprocal cross. However, in this case it would be impossible to obtain adult male flies that carry the mutation; they are dead. If the hypothesis proposed in a) above is correct, then half of the females, and none of the living males in the F_1 should carry the mutant allele. You could therefore cross F_1 females to wild type males, and see whether the expected ratios were observed among the offspring (e.g. half of the F_1 females should have a fewer male offspring than expected, while the other half of the F_1 females and all of the males should have a roughly equal numbers of male and female offspring).

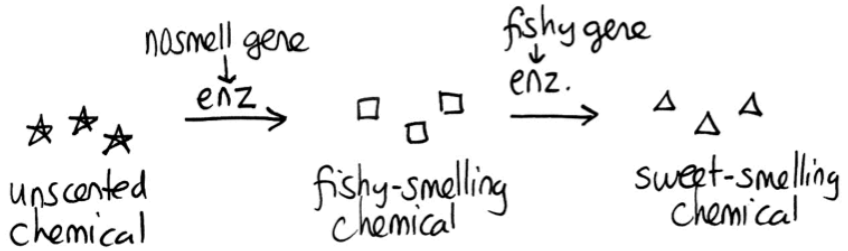
4.7 a) Treat a population of seeds with a mutagen such as EMS. Allow these seeds to self-pollinate, and then allow the F_1 generation to also self-pollinate. In the F_2 generation, smell each flower to find individuals with abnormal scent.

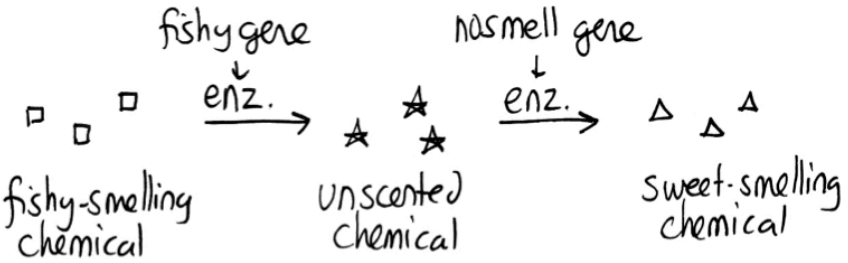
b) The fishy gene appears to be required to make the normal floral scent. Because the flowers smell fishy in the absence of this gene, <u>one</u> possibility explanation of this is that fishy makes an enzyme that converts a fishy-smelling intermediate into a chemical that gives flowers their normal, sweet smell.



Note that although we show this biochemical pathway as leading from the fishy-smelling chemical to the sweet-smelling chemical in one step, it is likely that there are many other enzymes that act after the fishy enzyme to make the final, sweet-smelling product. In either case, blocking the pathway at the step catalyzed by the *fishy* enzyme would explain the fishy smell.

c) In *nosmell* plants, the normal sweet smell disappears. Unlike *fishy*, the sweet smell is not replaced by any intermediate chemical that we can easily detect. Thus, we cannot conclude where in the biochemical pathway the *nosmell* mutant is blocked; *nosmell* may normally therefore act either before or after *fishy* normally acts in the pathway:





Alternatively, *nosmell* may not be part of the biosynthetic pathway for the sweet smelling chemical at all. It is possible that the normal function of this gene is to transport the sweet-smelling chemical into the cells from which it is released into the air, or maybe it is required for the development of those cells in the first place. It could even be something as general as keeping the plants healthy enough that they have enough energy to do things like produce floral scent.

4.8 a) Dominant mutations are generally much rarer than recessive mutations. This is because mutation of a gene tends to cause a loss of the normal function of this gene. In most cases, having just one normal (wt) allele is sufficient for normal biological function, so the mutant allele is recessive to the wt allele. Very rarely, rather than destroying normal gene function, the random act of mutation will cause a gene to gain a new function (e.g. to catalyze a new

enzymatic reaction), which can be dominant (since it performs this new function whether the wt allele is present or not). This type of gain-of-function dominant mutation is very rare because there are many more ways to randomly destroy something than by random action to give it a new function (think of the example given in class of stomping on an iPod).

b) Dominant mutations should be detectable in the F_1 generation, so the F_1 generation, rather than the F_2 generation can be screened for the phenotype of interest.

c) Large deletions, such as those caused by some types of radiation, are generally less likely than point mutations to introduce a new function into a protein: it is hard for a protein to gain a new function if the entire gene has been removed from the genome by deletion.

4.9 Class I. see Figure 4.5 on Transposable Elements.

4.10 a) Mutagenize a wild type (auxotrophic) strain and screen for mutations that fail to grow on minimal media, but grow well on minimal media supplemented with proline.

b) Take mutants #1-#10) and characterize them, based on (1) genetic mapping of the mutants (different locations indicate different genes); (2) different response to proline precursors (a different response suggests different genes); (3) complementation tests among the mutations (if they complement then they are mutations in different genes).

c) If the mutations are in different genes then the F1 progeny would be wild type (able to grow on minimal medium without proline).

d) If the mutations are in the same gene then the F1 progeny would NOT be wild type (unable to grow on minimal medium without proline).

Notes:

CHAPTER 5 - ANSWERS

5.1 The pedigree could show an AD, AR or XR mode of inheritance. It is most likely AD. It could be AR if the mother was a carrier, and the father was a homozygote. It could be XR if the mother was a carrier, and the father was a hemizygote. It cannot be XD, since the daughter (#2) would have necessarily inherited the disease allele on the X chromosome she received from her father.

5.2 There are many possible answers. Here are some possibilities: if neither of the parents of the father were affected (i.e. the paternal grandparents of children 1, 2, 3), then the disease could not be dominant. If only the paternal grandfather was affected, then the disease could only be X-linked recessive if the paternal grandmother was a heterozygote (which would be unlikely given that this is a rare disease allele).

5.3 a) The mode of inheritance is most likely AD, since every affected individual has an affected parent, and the disease is inherited even in four different matings to unrelated, unaffected individuals. It is very unlikely that it is XD or XR, in part because affected father had an affected son.

b) The mode of inheritance cannot be AD or XD, because affected individuals must have an affected parent when a disease allele is dominant. Neither can it be XR, because there is an affected daughter of a normal father. Therefore, it must be AR, and this is consistent with the pedigree.

c) The mode of inheritance cannot be AD or XD, because, again, there are affected individuals with unaffected parents. It is not XR, because there are unaffected sons of an affected mother. It is therefore likely AR, but note that the recessive alleles for this condition appear to be relatively common in the population (note that two of the marriages were to unrelated, affected individuals).

d) The mode of inheritance cannot be AD or XD, because, again, there are affected individuals with unaffected parents. It could be either XR or AR, but because all of the affected individuals are male, and no affected males pass the disease to their sons, it is likely XR.

5.4 If *a* represents the disease allele, individuals a, d, f (who all married into this unusual family) are *AA*, while b, c, e, g, h, i, j are all *Aa*, and k is *aa*.

5.5 There is a ½ chance that an offspring of any mating Aa x AA will be a carrier (Aa). So, there is a ½ chance that #3 will be Aa, and likewise for #4. If #3 is a carrier, there is again a ½ chance that #5 will be a carrier, and likewise for #6. If #5 and #6 are both Aa, then there is a ¼ chance that this monohybrid cross will result in #7 having the genotype aa, and therefore being affected by the disease. Thus, the joint probability is 1/2 x 1/2 x 1/2 x 1/2 x 1/4 =**1/64**.

5.6 a) q = *,-0.01.* = 0.1
b) 1-q = p; 1-0.1 = 0.9

c) $2pq = 2(0.1)(0.9) = 0.18$
d) $p^2 = 0.81$

5.7 First, calculate allele frequencies:

p = 2(AA) + (Aa) / total number of alleles scored = 2(432) + 676 / 2(432+676+92) = 0.6417

q = 2(aa) + (Aa) / total number of alleles scored = 2(92) + 676 / 2(432+676+92) = 0.3583

Next, given these observed allele frequencies, calculate the genotypic frequencies that would be expected if the population was in Hardy-Weinberg equilibrium.

$p^2 = 0.6417^2 = 0.4118$
$2pq = 2(0.6417)(0.3583) = 0.4598$
$q^2 = 0.3583^2 = 0.1284$

Finally, given these expected frequencies of each class, calculate the expected numbers of each in your sample of 1200 individuals, and compare these to your actual observations.

	expected	observed (reported in the original question)
AA	0.4118 × 1200 = 494	432
Aa	0.4598 × 1200 = 552	676
aa	0.1284 × 1200 = 154	92

The population does not appear to be at Hardy-Weinberg equilibrium, since the observed genotypic frequencies do not match the expectations. Of course, you could do a chi-square test to determine how significant the discrepancy is between observed and expected.

5.8 If in this theoretical question, the frequency of genotype of *AA* is set at 432/1200 and we are asked what frequencies of the other classes would fit a Hardy-Weinberg equilibrium. So, given that $p^2 = 432/1200$, then p=0.6, and q=0.4. Given these allele frequencies and a sample size of 1200 individuals, then there should be 576 *Aa* individuals ($2pq \times 1200 = 2(0.6)(0.4) \times 1200 = 576$) and 192 *aa* individuals ($q^2 \times 1200 = 0.4^2 \times 1200 = 192$), if the population was at Hardy-Weinberg equilibrium with 432 *AA* individuals.

5.9 The actual population appear to have more heterozygotes and few recessive homozygotes than would be expected for Hardy-Weinberg equilibrium. There are many possible reasons that a population may not be in equilibrium (see Table 5.1). In this case, there is possibly some selection against homozygous recessive genotypes, in favour of heterozygotes in particular. Perhaps the heterozygotes have some selective advantage that increases their fitness.

It is also worth noting the discrepancies between the allele frequencies calculated in 5.8 and 5.7. In question 5.7, we calculated the frequencies directly from the genotypes (this is the most accurate method, and does not require the population to be in equilibrium). In 5.8, we essentially estimated the frequency base on one of the phenotypic classes. The discrepancy between these calculations shows the limitations of using phenotypes to estimate allele frequencies, when a population is not in equilibrium.

CHAPTER 6 - ANSWERS

6.1 If 1 and 2 and 3 are all colorless, and 4 is red, what will be the phenotypes associated with the following genotypes? All of these mutations are recessive. As always, if the genotype for a particular gene is not listed, you can assume that alleles for that gene are wild-type.

a) red (because A and B are redundant, so products 3 and then 4 can be made)

b) red (because A and B are redundant, so products 3 and then 4 can be made)

c) white (because product 3 will accumulate and it is colorless)

d) white (because only product 1 and 2 will be present and both are colorless)

e) white (because only product 1 and 3 will be present and both are colorless)

f) white (because only product 2 and 3 will be present and both are colorless)

g) white (because only product 1 and 2 will be present and both are colorless)

h) 15 red : 1 white

	AB	Ab	aB	ab
AB	AABB	AABb	AaBB	AaBb
Ab	AABb	AAbb	AaBb	Aabb
aB	AaBB	AaBb	aaBB	aaBb
ab	AaBb	Aabb	aaBb	aabb

i) 12 red : 4 white

	BD	Bd	bD	bd
BD	BBDD	BBDd	BbDD	BbDd
Bd	BBDd	BBdd	BbDd	Bbdd
bD	BbDD	BbDd	bbDD	bbDd
bd	BbDd	Bbdd	bbDd	bbdd

j) 12 red :4 white

	AD	Ad	aD	ad
AD	AADD	AADd	AaDD	AaDd
Ad	AADd	AAdd	AaDd	Aadd
aD	AaDD	AaDd	aaDD	aaDd
ad	AaDd	Aadd	aaDd	aadd

6.2

a) red (because A and B are redundant, so products 3 and then 4 can be made)

b) red (because A and B are redundant, so products 3 and then 4 can be made)

c) blue (because product 3 will accumulate, and it is blue)

d) white (because only product 1 and 2 will be present and both are colorless)

e) blue (because only product 1 and 3 will be present and 1 is colorless and 3 is blue)

f) blue(because only product 2 and 3 will be present and 2 is colorless and 3 is blue)

g) white (because only product 1 and 2 will be present and both are colorless)

h) 15 red : 1 white

	AB	Ab	aB	ab
AB	AABB	AABb	AaBB	AaBb
Ab	AABb	AAbb	AaBb	Aabb
aB	AaBB	AaBb	aaBB	aaBb
ab	AaBb	Aabb	aaBb	aabb

i) 12 red : 4 blue

	BD	Bd	bD	bd
BD	BBDD	BBDd	BbDD	BbDd
Bd	BBDd	BBdd	BbDd	Bbdd
bD	BbDD	BbDd	bbDD	bbDd
bd	BbDd	Bbdd	bbDd	bbdd

j) 12 red : 4 blue

	AD	Ad	aD	ad
AD	AADD	AADd	AaDD	AaDd
Ad	AADd	AAdd	AaDd	Aadd
aD	AaDD	AaDd	aaDD	aaDd
ad	AaDd	Aadd	aaDd	aadd

6.3

a) red (because A and B are redundant, so products 3 and then 4 can be made)

b) red (because A and B are redundant, so products 3 and then 4 can be made)

c) blue (because product 3 will accumulate, and it is blue)

d) yellow (because only product 1 and 2 will be present and 1 is colorless and 2 is yellow)

e) blue (because only product 1 and 3 will be present and 1 is colorless and 3 is blue)

f) green? (because only product 2 and 3 will be present and 2 is yellow and 3 is blue, so probably the fruit will be some combination of those two colors)

g) yellow (because only product 1 and 2 will be present and 1 is colorless and 2 is yellow)

h) 15 red : 1 yellow

	AB	Ab	aB	ab
AB	AABB	AABb	AaBB	AaBb
Ab	AABb	AAbb	AaBb	Aabb
aB	AaBB	AaBb	aaBB	aaBb
ab	AaBb	Aabb	aaBb	aabb

i) 12 red : 3 blue:1 green

	BD	Bd	bD	bd
BD	BBDD	BBDd	BbDD	BbDd
Bd	BBDd	BBdd	BbDd	Bbdd
bD	BbDD	BbDd	bbDD	bbDd
bd	BbDd	Bbdd	bbDd	bbdd

j) 12 red : 4 blue

	AD	Ad	aD	ad
AD	AADD	AADd	AaDD	AaDd
Ad	AADd	AAdd	AaDd	Aadd
aD	AaDD	AaDd	aaDD	aaDd
ad	AaDd	Aadd	aaDd	aadd

6.4 Epistasis is demonstrated when the phenotype for a homozygous mutant in one gene is the same as the phenotype for a homozygous mutant in two genes. So, the following situations from questions 6.1-6.3 demonstrated epistasis:

6.1 No epistasis is evident from the phenotypes, even though we know from the pathway provided that D is downstream of A and B.

6.2 The phenotypes show that D is epistatic to A, because *aadd* looks like *AAdd* or *Aadd*. Also D is epistatic to B, because *bbdd* looks like *BBdd* or *Bbdd*.

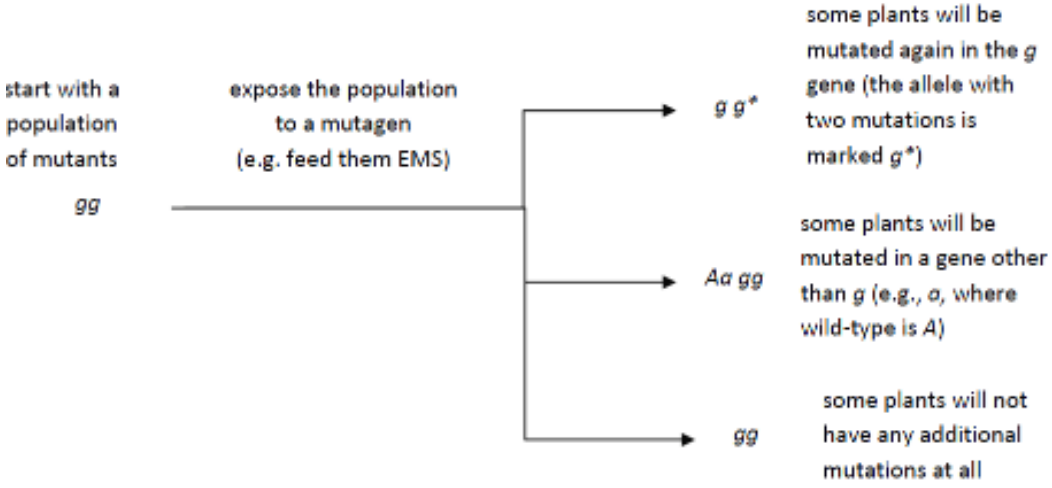
6.3 The phenotypes show that D is epistatic to A, because aadd looks like AAdd or Aadd. The phenotypes do not provide evidence for epistasis between B and D.

6.5 The answer is the same for a) – d)

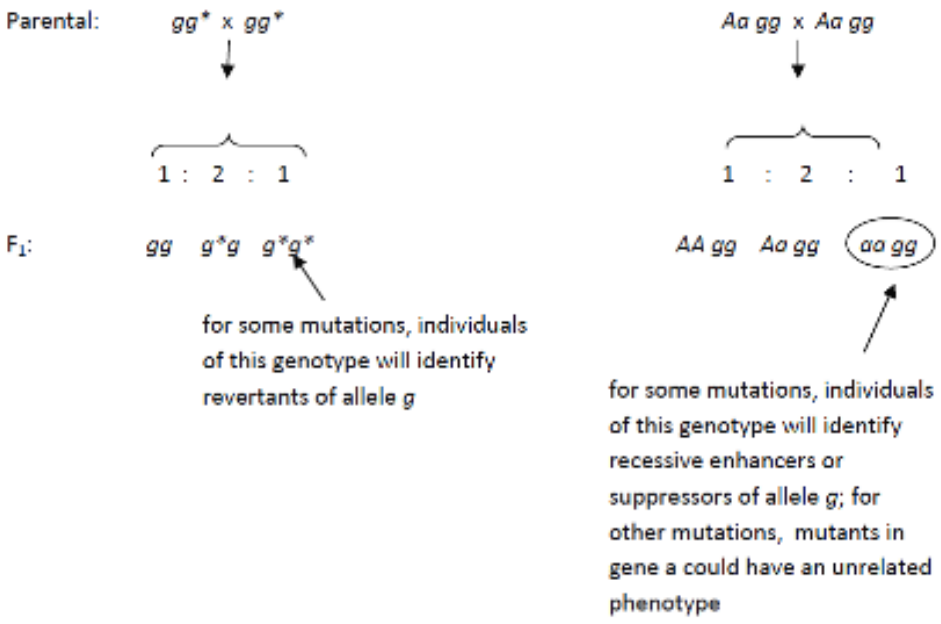
P could have been either: *AABB* x *aabb* or *aaBB* x *AAbb*;

F_1 was : *AaBb* x *AaBb*

6.6 Conduct an enhancer/suppressor screen (which can also result in the identification of revertants, as well)



allow the plants to self-pollinate in order to make any new, recessive mutations homozygous



6.7 Depending which amino acids were altered, and how they were altered, a second mutation in *g*g** could either have no effect (in which case the phenotype would be the same as *gg*), or it could possibly cause a reversion of the phenotype to wild-type, so that *g*g** and *GG* have the same phenotype.

6.8 Depending on the normal function of gene *A,* and which amino acids were altered in allele *a*, there are many potential phenotypes for *aagg*:

Case 1: If the normal function of gene *A* is in an unrelated process (e.g. *A* is required for
root development, but not the development of leaves), then the phenotype of *aagg* will
be: short roots and narrow leaves. The phenotypic ratios among the progeny of a
dihybrid cross will be:

9	3	3	1
*A_G_*	*A_gg*	*aaG_*	*aagg*
wild-type	tubular leaves normal roots	short roots normal leaves	tubular leaves short roots

Case 2: If the normal function of gene *A* is in the same process as *G,* such that *a* is a recessive allele that **increases** the severity of the *gg* mutant (**i.e. *a* is an enhancer of *g***) then the phenotype of *aagg* could be : no leaves. The phenotypic ratios among the progeny of a dihybrid cross depend on whether *aa* mutants have a phenotype independent of *gg,* in other words, do *aaG_* plants have a phenotype that is different from wild-type or from *A_gg*. There is no way to know this without doing the experiment, since it depends on the biology of the particular gene, mutation and pathway involved, so there are three possible outcomes:

Case 2a) If *aa* is an enhancer of *gg*, and *aaG_* plants have a mutant phenotype that differs from wild-type or (*A_gg*) then the phenotypic ratios among the progeny of a dihybrid cross will be:

9	3	3	1
*A_G_*	*A_gg*	*aaG_*	*aagg*
wild-type	tubular leaves (some phenotype that differs from gg; maybe small twisted leaves)	abnormal leaves	no leaves

Case 2b) If *aa* is an enhancer of *gg*, and *aaG_* plants have a mutant phenotype that is the same as *A_gg* , the phenotypic ratios among the progeny of a dihybrid cross will be:

9	6	1
*A_G_*	*A_gg aaG_*	*aagg*
wild-type	tubular leaves	no leaves

Case 2c) If *aa* is an enhancer of *gg*, and *aaG_* do not have a phenotype that differs from wild-type then the phenotypic ratios among the progeny of a dihybrid cross will be:

12	3	1
*A_G_ aaG_*	*A_gg*	*aagg*
wild-type	tubular leaves	no leaves

Case 3: If the normal function of gene *A* is in the same process as *G,* such that *a* is a recessive allele that **decreases** the severity of the *gg* mutant (**i.e. *a* is an suppressor of *g***) then the phenotype of *aagg* could be : wild-type. The phenotypic ratios among the progeny of a dihybrid cross depend on whether *aa* mutants have a phenotype independent of *gg,* in other words, do *aaG_* plants have a phenotype that is different from wild-type or from *A_gg*. There is no way to know this without doing the experiment, since it depends on the biology of the particular gene, mutation and pathway involved, so there are three possible outcomes:

Case 3a) If *aa* is a suppressor of *gg*, and *aaG_* plants have a mutant phenotype that differs from wild-type or (*A_gg*) then the phenotypic ratios among the progeny of a dihybrid cross will be:

10	3	3
A_G_ aagg	*A_gg*	*aaG_*
wild-type	tubular leaves (some phenotype that differs from *gg*)	no leaves

Case 3b) If *aa* is an suppressor of *gg*, and *aaG_* plants have a mutant phenotype that is the same as *A_gg* the phenotypic ratios among the progeny of a dihybrid cross will be:

10	6
A_G_ aagg	*A_gg aaG_*
wild-type	tubular leaves

Case 3c) If *aa* is an suppressor of *gg*, and *aaG_* plants do not have a phenotype that differs from wild-type then the phenotypic ratios among the progeny of a dihybrid cross will be:

13	3
A_G_ aaG_ aagg	*A_gg*
wild-type	tubular leaves

Case 4: If the normal function of gene *A* is in the same process as *G,* such that *a* is a recessive allele that with a phenotype that **is epistatic** to the *gg* mutant then the phenotype of both *aaG_* and *aagg* could be : no leaves. The phenotypic ratios among the progeny of a dihybrid cross will be:

9	4	3
*A_G_*	*aaG_ aagg*	*A_gg*
wild-type	no leaves	tubular leaves

Case … ?: There are many more phenotypes and ratios that could be imagined (e.g. different types of dominance relationships, different types of epistasis, lethality…etc).
Isn't genetics wonderful? It is sometimes shocking that more people don't want to become geneticists.

The point of this exercise is to show that many different ratios can be generated, depending on the biology of the genes involved. On an exam, you could be asked to calculate the ratio, given particular biological parameters. So, this exercise is also meant to demonstrate that it is better to learn how to calculate ratios than just trying to memorize which ratios match which parameters. In a real genetic screen, you would observe the ratios, and then try to deduce something about the biology from those ratios.

6.9 Calculate the phenotypic ratios from a dihybrid cross involving the two loci shown in Figure 6.13. There may be more than one possible set of ratios, depending on the assumptions you make about the phenotype of allele *b*.

Assuming that bb has no phenotype on its own (i.e. *A_bb* looks like *A_B_),* then *aaB_* will have the mutant phenotype, and *A_bb*, *A_B_*, and aabb will appear phenotypically wild-type. The phenotypic ratio will be 13 wild-type : 3 mutant.

6.10 For a dihybrid cross, there are 4 classes, 9:3:3:1. In a trihybrid cross without gene interactions, each of these 4 classes will be further split into a 3:1 ratio based on the phenotype at the third locus. For example, 9 x 3 =27 and 9 x 1 = 9. This explains the first two terms of the complete ratio: 27:9:9:9:3:3:3:1.

End

CHAPTER 7 - ANSWERS

7.1

Crossovers are defined cytologically; they are observed directly under the microscope.

Recombination is defined genetically; it is calculated from observed phenotypic proportions.

Some crossovers lead to recombination, but not all crossovers result in recombination.

Some recombinations involve crossovers, but not all recombinations result from crossovers.

Crossovers happen between sister and non-sister chromatids. If the chromatids involved the crossover have identical alleles, there will not be any recombination.

Crossovers can also happen without causing recombination when there are two crossovers between the loci being scored for recombination.

Recombination can occur without crossover when loci are on different chromosomes.

7.2

The use of pure breeding lines allows the researcher to be sure that he/she is working with homozygous genotypes. If a parent is known to be homozygous, then all of its gametes will have the same genotype. This simplifies the definition of parental genotypes and therefore the calculation of recombination frequencies.

7.3

This would suggest that individuals with a particular earlobe phenotype may also carry one or more alleles that increased their risk of cardiovascular disease. These individuals could therefore be informed of their increased risk and have an opportunity to seek increased monitoring and reduce other risk factors.

7.4 a)

It assumes that the loci are completely unlinked.

b)

If the parental gametes were *AB* and *ab*, then the gametes produced by the dihybrids would also be *AB* and *ab*, and the offspring of a cross between the two dihybrids would all be genotype *AABB*:*AaBb*:*aabb*,in a 1:2:1 ratio.

If the parental gametes were *Ab* and *aB*, then the gametes produced by the dihybrids would also be *Ab* and *aB*, and the offspring of a cross between the two dihybrids would all be genotype *AAbb*:*AaBb*:*aaBB*, in a 1:2:1 ratio.

fur	tail	behaviour	
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

7.5 a) Parental: *CcEe* and *ccee;* Recombinant: *Ccee* and *ccEe*.
b) Parental: *Ccee* and *ccEe;* Recombinant: *CcEe* and *ccee*.

7.6 Let *WwYy* be the genotype of a purple-flowered (*W*), green seeded (*Y*) dihybrid . Half of the progeny of the cross *WwYy* × *wwyy* will have yellow seeds whether the loci are linked or not. The proportion of the seeds that are also either white or purple flowered would help you to know about the linkage between the two loci only if the genotypes of the parents of the dihybrid were also known.

7.7
Let *tt* be the genotype of a short tassels, and *rr* is the genotype of pathogen resistant plants. We need to start with homozygous lines with contrasting combinations of alleles, for example:

P: *RRtt* (pathogen sensitive, short tassels) × *rrTT* (pathogen resistant, long tassels)

F_1: *RrTt* (sensitive, long) × *rrtt* (resistant, short)

F_2: parental *Rrtt* (sensitive, short) , *rrTt* (resistant, long)
Recombinant *rrtt* (resistant, short) , *RrTt* (sensitive, long)

7.8 Let *mm* be the genotype of a mutants that fail to learn, and *ee* is the genotype of orange eyes. We need to start with homozygous lines with contrasting combinations of alleles, for example (wt means wild-type):

P: *MMEE* (wt eyes, wt learning) × *mmee* (orange eyes, failure to learn)

F_1: *MmEe* (wt eyes, wt learning) × *mmee* (orange eyes, failure to learn)
F_2: parental *MmEe* (wt eyes, wt learning) , *mmee* (orange eyes, failure to learn)
Recombinant *Mmee* (wt eyes, failure to learn) , *mmEe* (orange eyes, wt learning)

7.9 Given a triple mutant *aabbcc* , cross this to a homozygote with contrasting genotypes, i.e. *AABBCC*, then testcross the trihybrid progeny, i.e.

P: *AABBCC* × *aabbcc*
F_1: *AaBbCc* × *aabbcc*

Then, in the F_2 progeny, find the two rarest phenotypic classes; these should have reciprocal genotypes, e.g. *aaBbCc* and *AAbbcc*. Find out which of the three possible orders of loci (i.e. *A-B-C*, *B-A-C*, or *B-C-A*) would, following a double crossover that flanked the middle marker, produce gametes that correspond to the two rarest phenotypic classes. For example, if the rarest phenotypic classes were produced by genotypes *aaBbCc* and *AAbbcc,* then the dihybrid's contribution to these genotypes was *aBC* and *Abc*. Since the parental gametes were *ABC* and *abc* the only gene order that is consistent with *aBC* and *Abc* being produced by a double crossover flanking a middle marker is *B-A-C* (which is equivalent to *C-A-B*).

7.10 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?

AaBb 135
Aabb 430
aaBb 390
aabb 120
(135 + 120)/(135+120+390+430)= 24%

7.11
Based on the information given, the recombinant genotypes with respect to these loci will be *Aabb* and *aaBb*. The frequency of recombination between A-B is 1cM=1%, based on the information given in the question, so each of the two recombinant genotypes should be present at a frequency of about 0.5%. Thus, the answer is 0.5%.

7.12
a) 4cM
b) Random sampling effects; the same reason that many human families do not have an equal number of boys and girls.

7.13
There would be approximately 2% of each of the recombinants: (yellow, straight) and (black, curved), and approximately 48% of each of the parentals: (yellow, curved) and (black, straight).

7.14
a) Without calculating recombination frequencies, determine the relative order of these genes.
A-C-B
b)
A-B 4.6%
A-C 2%
B-C 3%

B C A
|----------|---------|
3cM 2cM

	A-B	A-C	B-C
aBC	0	0	0
ABc	15	0	15
ABC	10	10	0
aBc	0	1	1
abC	13	0	13
Abc	0	0	0
AbC	0	1	1
abc	8	8	0
TOTAL	46	20	30
%	4.6	2	3

c) Recalculate recombination frequencies accounting for double recombinants

	A-B	A-C	B-C

aBC	0	0	0
ABc	15	0	15
ABC	10	10	0
aBc	1 x 2	1	1
abC	13	0	13
Abc	0	0	0
AbC	1 x 2	1	1
abc	8	8	0
TOTAL	50	20	30
%	5	2	3

7.15

A is fur color locus
B is tail length locus
C is behaviour locus

fur (A)	tail (B)	behaviour (C)			AB	AC	BC
white	short	normal	16	*aBC*	R	R	P
brown	short	agitated	0	*ABc*	P	R	R
brown	short	normal	955	*ABC*	P	P	P
white	short	agitated	36	*aBc*	R	P	R
white	long	normal	0	*abC*	P	R	R
brown	long	agitated	14	*Abc*	R	R	P
brown	long	normal	46	*AbC*	R	P	R
white	long	agitated	933	*abc*	P	P	P

B C A
|----------|---------|
4.1cM 1.5cM

Pairwise recombination frequencies are as follows (calculations are shown below):
A-B 5.6%
A-C 1.5%
B-C 4.1%

AB	AC	BC
16	16	0
0	0	0
0	0	0
36	0	36
0	0	0
14	14	0
46	0	46
0	0	0
112	30	82
5.6%	1.5%	4.1%

CHAPTER 8 - ANSWERS

8.1 You would need to know that HIV is an RNA-virus, but you should be able to detect the DNA pro-virus in infected white blood cells. You would have to be able to extract DNA from white blood cells, then use HIV-specific primers to see if HIV pro-virus DNA could be amplified. Thus, you would need to know some of the sequence of the HIV genome. You would probably want to compare your primers to the human genome sequence too, to make sure the primers are complementary only to HIV-DNA, but not human DNA. You would probably want to try to amplify some known HIV-free human DNA with the primers as a negative control, just to be sure that the primers were HIV-specific. And amplify the sequences from a known positive sample to know you can detect the sequences (positive control).

For the PCR reaction, you would need primers (as mentioned), dNTPs, Taq polymerase, and other buffers or salts as required by the polymerase. You would need an agarose gel, ethidium bromide, and electrophoresis buffers to analyze the PCR products to detect a band in a control positive sample, have it absent in a negative sample, and then test your experimental to obtain a valid result.

8.2 a) There will be a 6kb band (the insert) and a 3kb band (the plasmid vector)
b) There would be a 9kb band.
c) There would only be a 3kb band, which represents the plasmid, and both insert fragments.

8.3 The amplification factor is 2^n, where n is the number of cycles. So after 10 cycles, starting with 10 molecules, you would have $10 \times 2^{10} = 10{,}240$ molecules.

8.4 Present are polymerase, some dNTPs, and primers, as well as the original template, and PCR products. By far the most abundant products will be the one flanked at both ends by the primer sequences; these will be the only thing observed on the gel, since the template and other PCR products are present in much lower abundance and are usually not visualized.

8.5 Identify the gene encoding the antigenic fragment of the virus. Clone this gene into *E. coli* and produce lots of recombinant protein, purify, and use as a vaccine.

8.6 Without a selectable marker, you would have to individually test millions of bacterial colonies to find one that contained your cloned fragment. Furthermore, you could not keep the plasmid in the *E. coli* because the retention of the plasmid is dependent upon the antibiotic resistance selectable marker.

8.7 a) Radioactively label a piece of DNA that hybridized to the gene, outside of the part of the gene contained in the deletion. Extract DNA from the suspect cancer cells of individuals, digest with a restriction enzyme (the best choice would be ones that cleave just outside the gene), and separate the DNA by electrophoresis. Southern blot the gel and probe with DNA complementary to the gene. Be sure the probe spans the 200bp deleted region. Wash at high stringency and expose to a sheet of X-ray film (or equivalent). Individuals with the deletion will have two

bands, one at a lower position (200 bp lower) on the gel than those samples that do not have a deletion (or the cancer).

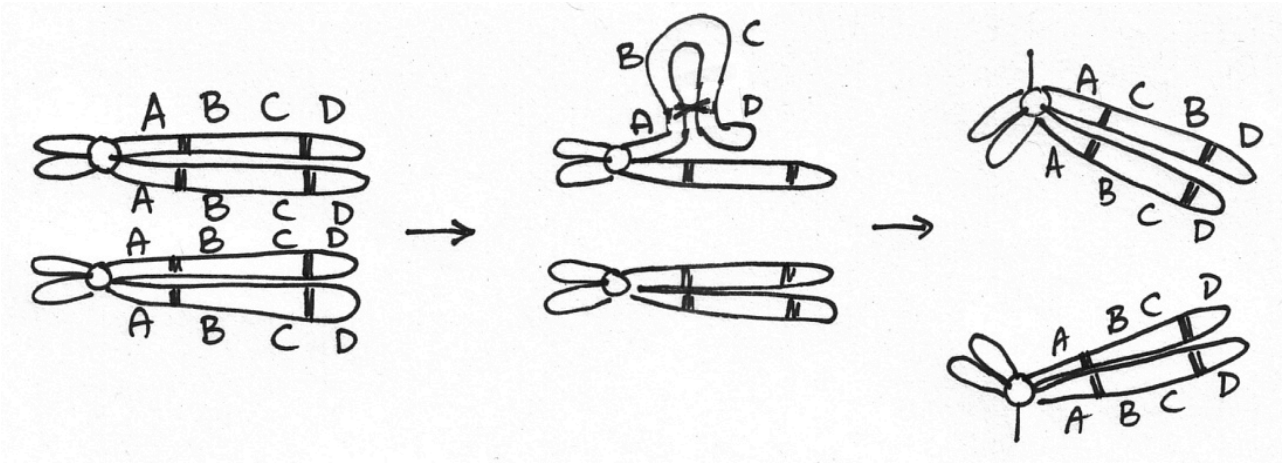
b) You would probably get hybridization to extra bands, or even just a big smear, since the probe would hybridize non-specifically.

8.8 a) Use PCR primers that flank the deletion. Extract DNA from cancer cell samples for use a template (one sample per reaction), and analyze the PCR products by gel electrophoresis.
Cancer cells will have two bands, one full length, one 200bp shorter. Unaffected tissue will only have full length products.
b) If the temperature was too low, the PCR products would probably appear as smears nearly the entire length of each lane, since the primers would bind to the genomic templates at many different positions and amplify fragments of many different lengths.

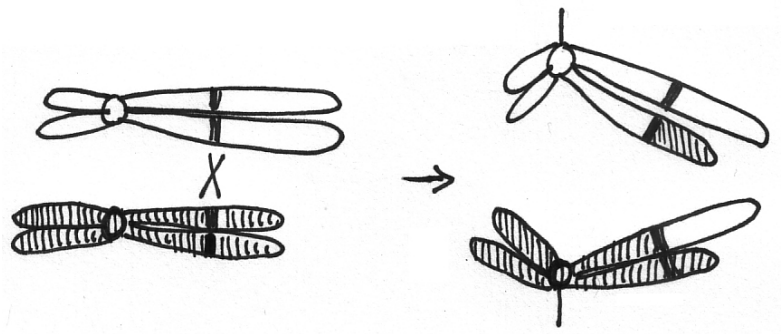
8.9 Label the PCR fragment for use as a probe. Hybridize the probe to a Southern blot of dog DNA. Cut out and clone any bands that hybridize to the probe.
Or, more recently, ignore the fragment and dog DNA sample, and take the sequence of the human olfactory receptor gene and BLAST it against the dog genome sequence. Compare the sequence output results to identify the dog olfactory receptor genes.
More? Do the test.
8.10 The complementary, sticky ends of the insert and plasmid vector will anneal together, but the non-functional ligase will not be able to covalently link the insert and vector together. Thus the almost-recombinant DNA will not be stable enough to be transformed, and thus unable to replicate -> no transformants.

ANSWERS - CHAPTER 9

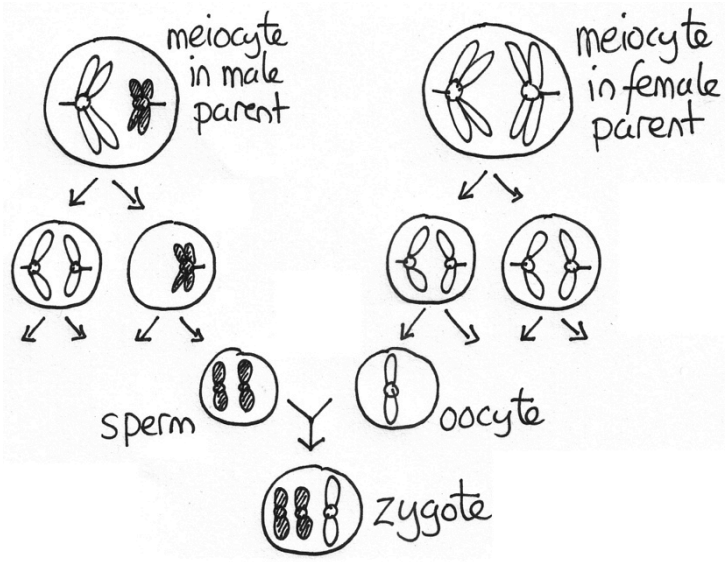
9.1 a) As in Figure 9.7 homologous chromosomes pair during prophase I. The shaded boxes are regions of sequence similarity, for example Alu transposable elements. A crossover occurs between two of the Alu elements on the same chromatid leading to a chromosomal inversion.



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.



9.2 As in Figure 9.12 there is a nondisjunction event during gamete formation. The larger X chromosomes are shown using open symbols and the smaller Y chromosomes are shown with shaded symbols. A second division nondisjunction event in the male parent leads to a zygote with an XYY karyotype.



9.3 a) 46, XY - zero Barr bodies, **b)** 46,XX - one, **c)** 47, XYY - zero, **d)** 47,XXX - two,

e) 45,X - zero, **f)** 47,XXY - one.

9.4 Having a shortage of key proteins is usually more detrimental than having an excess.

9.5 Gamma rays are efficient at causing double strand DNA breaks, which are then more likely to rejoin and produce a deletion.

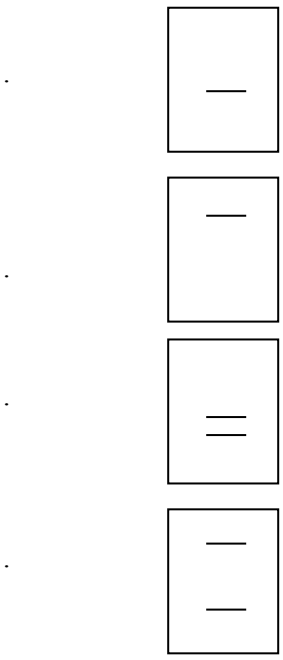
9.6 At the two cell stage, one of the embryo's cells will be 45,XY,-21 while the other will be 47,XY,+21. As embryogenesis continues most of the monosomy-21 cells will die and the embryo will ultimately be made of mostly trisomy-21 cells. The child will be born with Down syndrome.

9.7 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.

9.8 The results would be similar to Figure 9.19b. There would be 47 chromosomes glowing blue. The centromeres of two of the chromosomes would be glowing green and the centromeres of three other chromosomes would be glowing red.

CHAPTER 10 - ANSWERS:

10.1



a) B_2B_2

b) B_4B_4

c) B_2B_3

d) B_2B_4

10.2

#1 $B_3B_4E_1E_1$
#2 $B_3B_4E_1E_2$
#3 $B_4B_4E_2E_2$
#4 $B_3B_3E_1E_2$
#5 $B_2B_4E_1E_2$
#6 $B_2B_3E_2E_2$

10.3 #3 and #6 cannot be a parent, since neither #3 and #6 have any alleles in common with #1 at locus E.

10.4 a) the region of the fragment that is most likely to be polymorphic

```
TAAAGGAATCAATTACTTCTGTGTGTGTGTGTGTGTGTGTGTTCTTAGTTGTTTAAGTTTTAA
GTTGTGA
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
|||||||
ATTTCCTTAGTTAATGAAGACACACACACACACACACACACAAGAATCAACAAATTCAAAATT
CAACACT
```

b) any simple sequence repeats

```
TAAAGGAATCAATTACTTCTGTGTGTGTGTGTGTGTGTGTGTTCTTAGTTGTTTAAGTTTTAA
GTTGTGA
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
|||||||
```

```
ATTTCCTTAGTTAATGAAGACACACACACACACACACACACAAGAATCAACAAATTCAAAATT
CAACACT
```

c) the best target sites for PCR primers that could be used to detect polymorphisms in the length of the simple sequence repeat region in different individuals

```
TAAAGGAATCAATTACTTCTGTGTGTGTGTGTGTGTGTGTGTTCTTAGTTGTTTAAGTTTTAA
GTTGTGA
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
|||||||
ATTTCCTTAGTTAATGAAGACACACACACACACACACACACAAGAATCAACAAATTCAAAATT
CAACACT
```

10.5

a) 1/4 white seeds 1 gene
b) 1/16 white seeds 2 genes
c) 1/64 white seeds 3 genes
d) 1/256 white seeds 4 genes

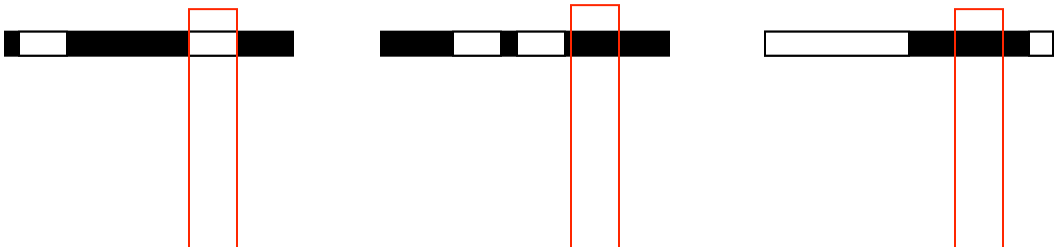
10.6 It is possible that the parents are recessive at several loci that together contribute to height. If the offspring inherits recessive alleles from each parent at several of these loci, and the recessive alleles tend to make an individual not as tall as the dominant alleles, then the child may be shorter than the parents.

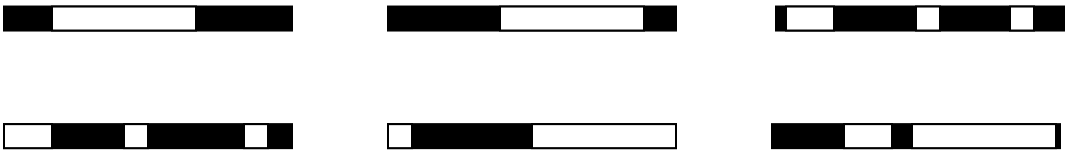
10.7 Imagine that two mice strains have been identified that differ in the time required to complete a maze, which may be an indication of intelligence. The time for maze completion is heritable and these parental strains "breed true" for the same completion time in each generation. Imagine also that their chromosomes are different colors and we can track the inheritance of chromosomal regions from each parent based on this color.
Based on the following diagrams of one chromosome from each individual in a pedigree, identify a chromosomal region that may contain a gene that affects time to complete a maze. The time for each individual is shown below each chromosome. Assume that all individuals are homozygous for all loci.

Answer marked in red on diagram.

Parents:

Selected individuals from among F8 progeny of the above parents:





10.8 You could use molecular markers that were polymorphic between each of the parents, and that were present at high density along the length of the chromosome.

Notes:

CHAPTER 11 - ANSWERS:

11.1 -omics techniques allow you to examine thousands of genes or protein at the same time. On a per-gene basis, this is more cost effective than examining a single gene at a time. However, the total cost of –omics technologies can be quite high (especially the instrumentation). Besides being more efficient, by examining many genes in parallel, you may identify patterns or combinations of genes that work together.

11.2 If only terminators (but no regular dNTPs) were added to the reaction, the reaction would always terminate at the first base added after the primer, and chromatogram would be essentially flat lines (no peaks) for all but the first position, which would show a Brobdingnagian peak.

11.3 Clone-by-clone sequencing requires less sequencing than WGS, since the objective is to minimize sequencing of redundant fragments. However, clone-by-clone is ultimately less efficient, since it takes so long to determine the minimum number of clones to sequence, through physical mapping. WGS is faster and more cost-effective, but is more prone to result in gaps in the sequence than is clone-by-clone sequencing.

11.4 You could extract raw, naked DNA from seawater in various different places in the world and then sequence all of this DNA. Next, use computer comparisons to identify DNA that did not belong to any known species. This is an example of meta-genomics, and is already being done by some scientists. Remember, having the sequence is not the same as having the organism or understanding the sequence.

11.5 One way is to extract RNA from (1) a drought-stressed and (2) a well-watered wheat plant. Reverse transcribe (make a DNA copy) the RNA samples independently and label each with a different fluorescent dye, then hybridize the two sample simultaneously to a microarray of wheat genes. If the drought-stressed sample was labeled one colour, say red, and the watered plant another colour, say green, then those gene transcripts that increase in abundance specifically under drought will be reder, while those that decreased would be greener, relatively speaking.

11.6 All that microarrays and similar types of analyses demonstrate (whether high-throughput or not) is a correlation between transcript presence/abundance and a phenotype. This does not mean that the genes cause cancer, although some of them could. The alternate probable explanation is that the gene's change in expression is a response to being cancerous.

11.7 PCR amplify DNA from only a single bacterial clone containing single insert of part of a gene.

11.8 Spots with a random mix would be approximately equally green and red, which appears yellow or brown in the microarray analysis equipment.

11.9 At low stringency there would be excessive non-specific hybridization so many (all?) the spots would be approximately equally green and red, which appears yellow or brown in the microarray analysis equipment.

CHAPTER 12 - ANSWERS

12.1 Transcriptional: initiation, processing & splicing, degradation
Translational: initiation, processing, degradation
Post-translational: modifications (e.g. phosphorylation), localization
Others: histone modification, other chromatin remodeling, DNA methylation

12.2 Legend:

+++ Lots of β-galactosidase activity
+ Moderate β-galactosidase activity
-- No β-galactosidase activity

-- 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^+, Oc, Z^+, Y^+ (no glucose, no lactose)
+++ i) I^+, Oc, Z^+, Y^+ (no glucose, high lactose)
+ j) I^+, Oc, Z^+, Y^+ (high glucose, no lactose)
+ k) I^+, Oc, 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)

12.3 Legend:

+++ Lots of β-galactosidase activity
+ Moderate β-galactosidase activity
-- No β-galactosidase activity

+++ a) $I^+, O^+, Z^+, Y^+ / O^-, Z^-, Y^-$ (high lactose)
-- b) $I^+, O^+, Z^+, Y^+ / O^-, Z^-, Y^-$ (no lactose)
+++ c) $I^+, O^+, Z^-, Y^+ / O^-, Z^+, Y^+$ (high lactose)
+ d) $I^+, O^+, Z^-, Y^+ / O^-, Z^+, Y^+$ (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) $I^+, Oc, Z^+, Y^+ / I^+, O^+, Z^-, Y^+$ (high lactose)
+++ j) $I^+, Oc, Z^+, Y^+ / I^+, O^+, Z^-, Y^+$ (no lactose)
+++ k) $I^+, O^+, Z^-, Y^+ / I^+, Oc, Z^+, Y^+$ (high lactose)
+++ l) $I^+, O^+, Z^-, Y^+ / I^+, Oc, Z^+, Y^+$ (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)

12.4 You could demonstrate this with just $I^+O^cZ^-/I^+O^+Z^+$. The fact that this does not have constitutive lactose expression shows that the operator only acts on the same piece of DNA on which it is located. There are also other possible answers.

12.5 You could also demonstrate this with just $I^+O^+Z^-/I^-O^+Z^+$. The fact that this has the same lactose-inducible phenotype as wild-type hows that a functional lacI gene can act on operators on both the same piece of DNA from which it is transcribed, or on a different piece of DNA. There are also other possible answers.

12.6 For all of these, the answer is the same: The l*ac* operon would be inducible by lactose, but only moderate expression of the lac operon would be possible, even in the absence of glucose
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

12.7 Both involve *trans*-factors binding to corresponding *cis*-elements to regulate the initiation of transcription by recruiting or stabilizing the binding of RNApol and related transcriptional proteins at the promoter. In prokaryotes, genes may be regulated as a single operon. In eukaryotes, enhancers may be located much further from the promoter than in prokaryotes.

12.8 These fish would all have spiny tales like the deep-water population.

12.9 These could have arisen from loss-of-function mutation in *FLC*, or in the cis-element to which *FLC* normally binds.

12.10 If there was no deacetylation of *FLC* by HDAC, transcription of FLC might continue constantly, leading to constant suppression of flowering, even after winter.

CHAPTER 13 – ANSWERS

13.1 Oncogenes usually arise from gain-of-function mutations, which tend to be haplosufficient. Mutations in tumor supressors are usually loss-of-function mutations, which tend to be haploinsufficient.

13.2 p53 activates DNA repair, apoptosis, and inhibitors of cell division. Different genes involved in each of these pathways have enhancer elements to which p53 binds; therefore they call all be activated by p53.

13.3 Some substances can promote cancer without causing a mutation, for example by inducing the cell cycle or accelerating it so that there is less time to repair DNA damage. All mutagens are potentially carcinogens, although some potential mutagens may not cause significant damage to cells in the body due to detoxification or other reasons that limit their efficacy.

13.4 Was the dose fed to the rats relevant? Were similar effects seen in other organisms? Do epidemiological studies support these conclusions?

13.5 Cancer results from an accumulation of mutations that activate cell division and disable tumor suppression. HPV infection or *BRCA1* mutation, alone do not satisfy all of these requirements. Also, not all strains of HPV are equally carcinogenic, and the body's defense may be able to suppress the activity of the virus.

Notes: