Open Genetics Lectures

Fall 2017

Department of Biological Sciences – University of Alberta, Canada

OGL
Open Genetics Lectures

Open Genetics Lectures (OGL)
September, 2017 Version
The **Open Genetic Lecture** textbook is derived from the **Open Genetics** textbook.

OPEN GENETICS (OG) - History

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

Open Genetic Lectures (OGL) – Origin 2015, Updated Summer 2016, 2017

OGL is an alternative approach to an open source textbook. Much of its content is derived from the OG textbook. The 13 chapters in OG were cut up and distributed into 41 shorter chapters that parallel the current lecture topics in BIOL 207 (Molecular Genetics and Heredity) at the University of Alberta. More text content, figures, and chapter-end questions were added in this revision. The most recent version of OG had ~76,000 words, while the Fall 2015 version of OGL had ~128,000 words, a 68% increase.

This reorganization of OG content into OGL was accomplished during the summer of 2015 by John Locke, Mike Harrington, Lindsay Canham, and Min Ku Kang. This project was funded in part by the Alberta Open Educational Resources (ABOER) Initiative, which is made possible through an investment from the Alberta government. Lindsay Canham was supported by a grant from the Alberta Open Education Initiative (OEI) through the University of Alberta. Min Ku Kang was supported by a Summer Student Scholarship from the Centre for Teaching and Learning (CTL), University of Alberta. Without these sources of financial aid this project would not have been possible. John Locke and Michael Harrington appreciate their help, as well as that of Michael Deyholos, who initiated this endeavor. Typographical errors, rewording, and additional questions were added in the summers of 2016 & 2017 (J. Locke, M. Harrington, and K. King-Jones).

Access to OGL text files through DataVerse

The final version of this work is available via a DataVerse link:
    https://dataverse.library.ualberta.ca/dvn/dv/OpenGeneticsLectures
This includes all the .docx files for each chapter and other relevant files. This is made available for anyone to use, adapt, or improve for educational purposes. If you have edits, improvements or additions that you wish to share under the same license terms, please contact John Locke, University of Alberta.

OGL Cover Images:

Girl: Flickr-Kate Andrews-CC BY-NC-SA 2.0   Peas: Wikipedia-Bill Ebbesen-CC BY-SA 3.0
Fly: Wikimedia Commons-Aka-CC BY-SA 2.5   DNA: Wikimedia Commons-Jerome Walker-PD

Table of Contents

Introduction

Also available:
Online Open Genetics!

To be successful in Introductory Genetics, you are encouraged to use the supplementary electronic resources provided by the website for Online Open Genetics **(http://opengenetics.net)**. These resources will help you learn and practice problem-solving skills and self-assess your knowledge as you progress through the course.

The website provides:
> (1) access to short instructional videos and
> (2) supplementary readings, as well as
> (3) interactive exercises.

All these will help deepen your understanding of basic concepts in genetics, as well as to practice and refine the skills needed to solve common problems in genetic analysis.

These supplementary materials can be accessed using the internet via web browsers on Windows or Mac computers, or on tablets and iPads. More information is available on the web page.

ONLINE OPEN GENETICS

Definitions:

Gene - a <u>hereditary unit</u> that <u>occupies a specific position</u> (locus) within the genome or chromosome and has one or more <u>specific effects upon the phenotype</u> of the organism and can <u>mutate into various forms</u> (alleles) and can <u>recombine</u> with similar such units.

Gene locus -(plural = loci) The specific place on a chromosome where a gene is located

Allele - refers to one of the different forms of a gene that can exist at a single gene locus

Genotype - the specific allelic composition of a cell or organism. Normally only the genes under consideration are listed in a genotype and the alleles at all the remaining gene loci are considered to be wild type.

Phenotype - the detectable outward manifestation of a specific genotype. In describing a phenotype usually only the characteristics under consideration are listed while the remaining characters are assumed to be wild type (normal).

Allelic mutations- two mutations at the same gene locus

Non-allelic mutations- two mutations that affect different gene loci

Genetic Nomenclature & Symbols - What you need to know

Geneticists use a variety of different nomenclature systems to represent genes and their mutations in different organisms. You will need to become familiar with these different systems in order to understand genetics and answer questions on an exam.
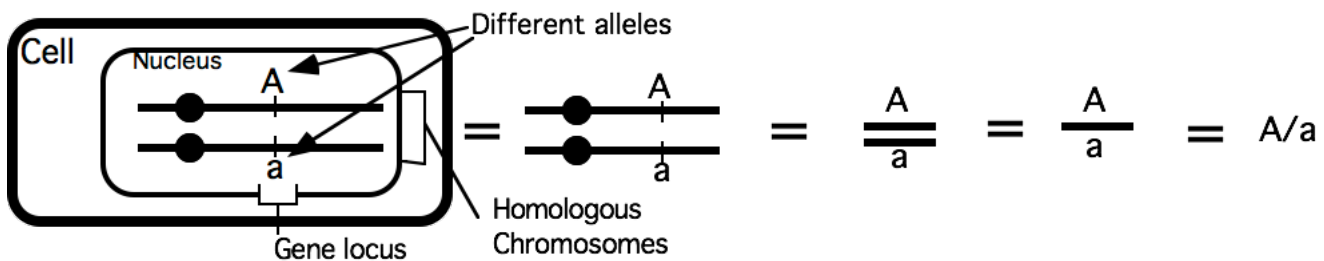
(Definitions are taken with modification from Griffiths et al, 2000 and <u>A Dictionary of Genetics</u> 4th Ed., King & Stansfield, 1990.)

Gene names and symbols

Different organisms have different nomenclature systems to symbolize their genes

Because different genetic model organisms have historically developed their own nomenclature systems for denoting genes and alleles, there is a variety of different nomenclature systems in use today. This can be very confusing for students trying to learn the basics of genetics (genes, alleles, and mutations). However, all systems have two main parts: (1) a gene name, and (2) an allele names.

Gene Names

Usually, genes have a full name (e.g. the *white* locus in Drosophila) as well as a short symbol form (e.g. *w* in the case of the *white* locus) that is a unique letter or combinations of letters.

- So, the letters "a", "b", and "c" would represent different named genes.
- Each named gene would have a unique letter, or combination of letters for an organism.
  - For example, the "*vermillion*" gene in Drosophila is represented by the letter "*v* ".
  - While "*vg* " is the symbol for the "*vestigial*" gene.
  - And "*vvl* " is the symbol for the "*ventral veins lacking*" gene locus.

- Note: the same letter symbols may represent different gene loci in different organisms, but often the same or similar symbol is used in different organisms.
- Sometimes letters **and numbers** are used, especially for different loci that have similar phenotypes.
  - For example, the *arg-1*, *arg-2*, and *arg-3* loci described by Beadle and Tatum in Neurospora have a similar arginine auxotrophic phenotype but are three separate gene loci.
- Also, formally, gene symbols and gene names are **always shown in *italics* text**, but in the lecture portion of BIOL 207, we may not require or use italics in gene names and symbols all the time.

Allele names.

1) Superscripts - usually denote different allelic forms of a gene locus

The normal copy of a gene is known as wild type and is usually symbolized by superscript plus sign, "+".

E.g. " a^+ ", " b^+ ", or *white* $^+$, etc. or it is sometimes abbreviated to just "+"

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

2) UPPER vs. lower case letters are often used to denote dominant and recessive alleles

In diploid heterozygotes, the dominant allele is typically (but not always) designated with the upper case letter(A) while the recessive allele is given the lower case letter(a). An example of this is Mendel's round (*R*) vs. wrinkled (*r*) alleles at the pea shape locus. Note that not all wild type alleles are dominant (capital). Some mutant alleles can be dominant to wild type, even though most are recessive.

Note: Nomenclature is covered more extensively in Chapter 13.

Symbols of genes vs. proteins.

We also need to use symbols to describe the protein derived from a specific gene. Typically, the name of the protein uses the same name as the gene, only with all CAPITAL letters. For example, the Drosophila *white* gene codes for a polypeptide involved in pigment precursor transport across the cell membrane (see Chapter 10). Thus, the *white* gene codes for the WHITE polypeptide/protein.

Attempts to unify the gene name system:

There have been recent attempts to unify the gene and allele naming systems of the various genetic model organisms (e.g. Alberts et al. Molecular Biology of the Cell, 6/e), but such attempts can lead to greater student confusion, rather than clarity.

In BIOL 207, we will try and use consistent naming systems to facilitate student learning and leave the intricacies of multi-organism nomenclature to senior courses where it is necessary and more appropriate.

Creative Commons Copyright License

Figures in this textbook

All figures in this textbook are open source and/or under the Creative Commons License. Many are obtained from various, non-copyrighted sources, including Flickr, Wikipedia, Public Library of Science, and Wikimedia Commons. They are available for re-use, as per their licenses. Each caption lists the origin, creator, and license for that figure. If any are incorrect, please contact John Locke for correction.

Guide for Image, Photo and Figure License Details in this work

In the caption for each image, photo, and figure there is a citation. This includes the source or origin, creator, and license for each work. We have made a best effort to correctly attribute these works; if there are any errors, please contact John Locke.

The license is given at the end of the caption in short form as the following:

CC BY 2.0	Creative Commons Attribution 2.0 Generic
CC BY 2.5	Creative Commons Attribution 2.5 Generic
CC BY 3.0	Creative Commons Attribution 3.0 Unported
CC BY 4.0	Creative Commons Attribution 4.0 International
CC BY-NC 2.0	Creative Commons Attribution-NonCommercial 2.0 Generic
CC BY-NC 3.0	Creative Commons Attribution-NonCommercial 3.0 Unported
CC BY-NC-ND 2.0	Creative Commons Attribution-NonCommercial-NoDerivs 2.0 Generic
CC BY-NC-SA 2.0	Creative Commons Attribution-NonCommercial-ShareAlike 2.0 Generic
CC BY-ND 2.0	Creative Commons Attribution-NoDerivs 2.0
CC BY-SA 1.0	Creative Commons Attribution-ShareAlike 1.0 Generic
CC BY-SA 2.0	Creative Commons Attribution-ShareAlike 2.0
CC BY-SA 2.5	Creative Commons Attribution-ShareAlike 2.5 Generic
CC BY-SA 3.0	Creative Commons Attribution-ShareAlike 3.0 Unported
CC BY-SA 4.0	Creative Commons Attribution-ShareAlike 4.0 International
CC0 1.0	CC0 1.0 Universal Public Domain Dedication
PD	Public Domain because it was created by a US government agency or because the author has explicitly released it into the public domain.

Creative Commons Copyright Licenses

	Attribution	This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation.
	Attribution-ShareAlike	This license lets others remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms.
	Attribution-NoDerivs	This license allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to you.
	Attribution-NonCommercial	This license lets others remix, tweak, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial, they don't have to license their derivative works on the same terms.
	Attribution-NonCommercial-ShareAlike	This license lets others remix, tweak, and build upon your work non-commercially, as long as they credit you and license their new creations under the identical terms.
	Attribution-NonCommercial-NoDerivs	This license is the most restrictive of our six main licenses, only allowing others to download your works and share them with others as long as they credit you, but they can't change them in any way or use them commercially.
	Public Domain	No longer restricted by copyright.

Taken from https://creativecommons.org/licenses/ on May 25, 2015

Bibliography

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

Felix Ratjen, Gerd Döring, Cystic fibrosis, The Lancet, Volume 361, Issue 9358, 22 February 2003, Pages 681-689

Francis, Richard C. "Epigenetics". 2011, Norton, 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

Lap-Chee Tsui and Ruslan Dorfman, The Cystic Fibrosis Gene: A Molecular Genetic Perspective Cold Spring Harb Perspect, February 2013;3

MendelWeb. R. B. Blumberg, {August 1, 2012}. World Wide Web URL: www.mendelweb.org/

Online Mendelian Inheritance in Man, OMIM®.

McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), {August 1, 2012}. World Wide Web URL: http://omim.org/

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

Your Genes, Your Health

Dolan DNA Learning Center / Cold Spring Harbor Laboratory, {August 1, 2012}. World Wide Web URL: www.ygyh.org/

Wikipedia, URL: www.wikipedia.com

Introduction

Notes:

Chapter 01 – DNA is the Genetic Material

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

Introduction

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

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

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

1. Griffith's Transformation Experiment (1928)

Microbiologists identified two strains of the bacterium *Streptococcus pneumoniae*. The R-strain produced rough colonies on a bacterial plate, while the other S-strain was smooth (**Figure 2**). More importantly, the S-strain bacteria caused fatal infections when injected into mice, while the R-strain did not (**Figure 3**). Neither did "heat-treated" S-strain cells. **Griffith** in 1928 noticed that upon mixing "heat-treated" S-strain cells together with some R-type bacteria (neither individually should kill the mice), the mice died. Furthermore, there

were S-strain, pathogenic cells recoverable. Thus, some non-living component from the S-type strains contained genetic information that could be transferred to and **transform** the living R-type strain cells into S-type cells.
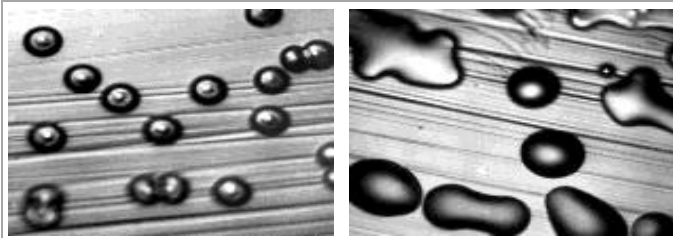
Figure 2.
Colonies of Rough (left) and Smooth (right) strains of *S. pneumoniae*.
(J. Exp.Med.98:21, 1953-R. Austrian-Pending)
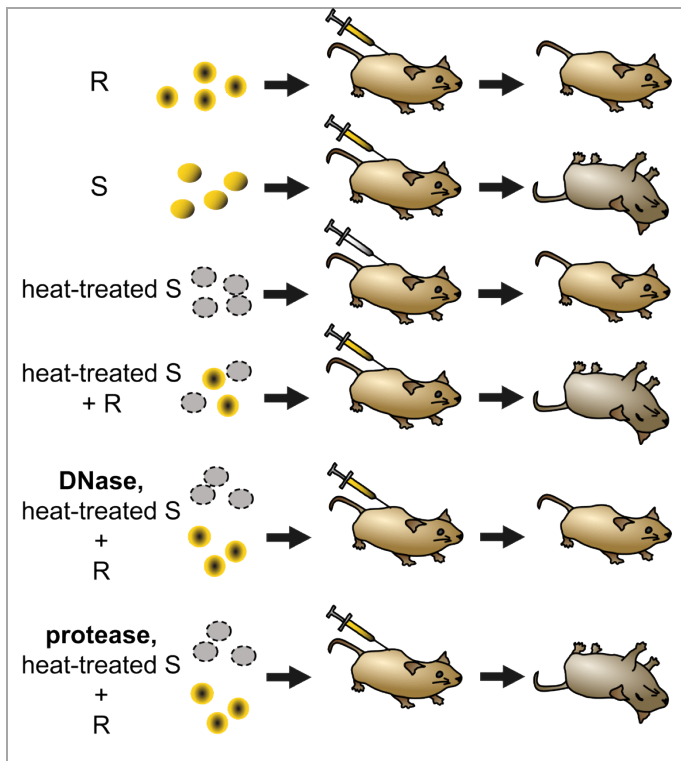
Figure 3.
Experiments of Griffith and of Avery, MacLeod and McCarty. R strains of *S. pneumoniae* do not cause lethality. However, DNA-containing extracts from pathogenic S strains are sufficient to make R strains pathogenic.
(Wikipedia - Modified by Deyholos- CC BY-NC 3.0)

2. AVERY, MACLEOD AND MCCARTY'S EXPERIMENT (1944)

What kind of molecule from within the S-type cells was responsible for the transformation? To answer this, researchers named **Avery, MacLeod and McCarty** separated the S-type cells into various components, such as proteins, polysaccharides, lipids, and nucleic acids. Only the nucleic acids from S-type cells were able to make the R-strains smooth and fatal. Furthermore, when cellular extracts of S-type cells were treated with **DNase** (an enzyme that digests DNA), the transformation ability was lost. The researchers therefore concluded that DNA was the genetic material, which in this case controlled the appearance (smooth or rough) and pathogenicity of the bacteria.

3. HERSHEY AND CHASE'S EXPERIMENT (1952)

Further evidence that DNA is the genetic material came from experiments conducted by **Hershey and Chase**. These researchers studied the transmission of genetic information in a virus called the T2 **bacteriophage**, which uses *Escherichia coli* as its host bacterium (**Figure 4**).

Figure 4.
Electronmicrograph of T2 bacteriophage on surface of E. coli.
(Wikipedia- Dr Graham Beards- CC BY-SA 3.0)

Like all viruses, T2 hijacks the cellular machinery of its host to manufacture more viruses. The T2 phage itself only contains both protein and DNA, but no other class of potential genetic material. To determine which of these two types of molecules contained the genetic blueprint for the virus, Hershey and Chase grew viral cultures in the presence of radioactive isotopes of either phosphorus (^{32}P) or sulphur (^{35}S). The phage incorporated these isotopes into their DNA and proteins, respectively (**Figure 5**). The researchers

then infected *E. coli* with the radiolabeled viruses, and looked to see whether ^{32}P or ^{35}S entered the bacteria. After ensuring that all viruses had been removed from the surface of the cells, the researchers observed that infection with ^{32}P labeled viruses (but not the ^{35}S labeled viruses) resulted in radioactive bacteria. This demonstrated that DNA was the material that contained genetic instructions.
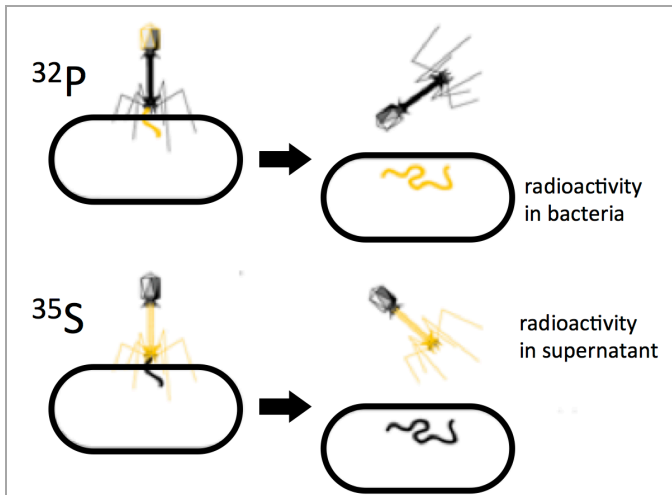
Figure 5.
When ^{32}P-labeled phage infects *E. coli*, radioactivity is found only in the bacteria, after the phage are removed by agitation and centrifugation. In contrast, after infection with ^{35}S-labeled phage, radioactivity is found only in the supernatant that remains after the bacteria are removed. (Wikipedia –Modified by Deyholos- CC BY-NC 3.0)

4. RNA and Protein

While DNA is the genetic material for the vast majority of organisms, there are some viruses that use **RNA** as their genetic material. These viruses can be either single- or double-stranded. Examples include SARS virus, influenza virus, hepatitis C virus and polio virus, as well as the retroviruses like HIV-AIDS. Typically, there is DNA used at some stage in their life cycle to replicate their RNA genome.

Also, the **prion** protein is an infectious agent that transmits characteristics via only a protein (no nucleic acid present). Prions infect by transmitting a mis-folded protein state from one aberrant protein molecule to a normally folded molecule. These agents are responsible for Bovine Spongiform Encephalopathy (BSE, also known as "mad cow disease") in cattle, Chronic Wasting Disease in deer, Scrapie is sheep and Creutzfeldt–Jakob disease (CJD) in humans. All known prion diseases act by altering the structure of the brain or other neural tissue and all are currently untreatable and ultimately fatal.

_____

SUMMARY:

- Genetics is the scientific study of heredity and the variation of inherited characteristics.

- Heredity is the concept that a trait of an individual can be passed down through generations

- A gene can be defined abstractly as a unit of inheritance.

- The experiments done by Griffith and Hershey and Chase showed the ability of DNA from bacteria and viruses to transfer genetic information into bacteria demonstrates that DNA is the genetic material and that its universal.

- Some viruses use RNA as their genetic material and can be either single or double stranded.

- Prion is a mis-folding protein that transmits its mis-folding property to a normal one.

KEY WORDS:

genetics	transform
heredity	Avery, MacLeod, & McCarty
Mendel	DNase
blending inheritance	Hershey and Chase
particulate inheritance	bacteriophage
gene	^{35}S
allele	^{32}P
Griffith	prion

STUDY QUESTIONS:

1) Imagine that retuning astronauts provide you with living samples of multicellular organisms discovered on another planet. These organisms reproduce with a short generation time like our standard yeast species. Initial observations about their reproduction indicate that they also require two "sexual types" to mate, but nothing else is known about how their genetics works.
   a) How could you define laws of heredity for these organisms?
   b) How could you determine what molecules within these organisms contained genetic information?
   c) Would the mechanisms of genetic inheritance likely be similar for all organisms from this planet?
   d) Would the mechanisms of genetic inheritance likely be similar to organisms from earth?

2) It is relatively easy to extract DNA and protein from cells; biochemists have been doing this since at least the 1800's. Why then did Hershey and Chase need to use radioactivity to label DNA and proteins in their experiments?

3) Starting with mice and R and S strains of *S. pneumoniae*, what experiments, in addition to those shown in **Figure 3** can be used to demonstrate that DNA is THE genetic material and the only genetic material?

4) Mendel put forth a "particulate inheritance" model – alleles, dominant, recessive, etc. At the time there was a "blended inheritance" model, which is like mixing paint colours (analogy). Suggest an analogy for Mendel's particulate model, taking into account the dominant and recessive characters of alleles.

Research Experience

Courses:
BIOL 298, Understanding Biological Research *3 (fi 6) (either term, 3-0-3)
BIOL 398, Research Project. *3 (fi 6) (either term, 0-0-6)
BIOL 399, Research Project. *6 (fi 12) (two term, 0-0-6)
BIOL 498, Research Project. *3 (fi 6) (either term, 0-0-6).
BIOL 499, Research Project. *6 (fi 12) (two term, 0-0-6).

A <u>Research Certificate</u> in Science

will provide an opportunity for undergraduate students to engage in authentic research in their discipline and acquire skills beyond what a normal research experience in an Honors or Specialization program may allow.
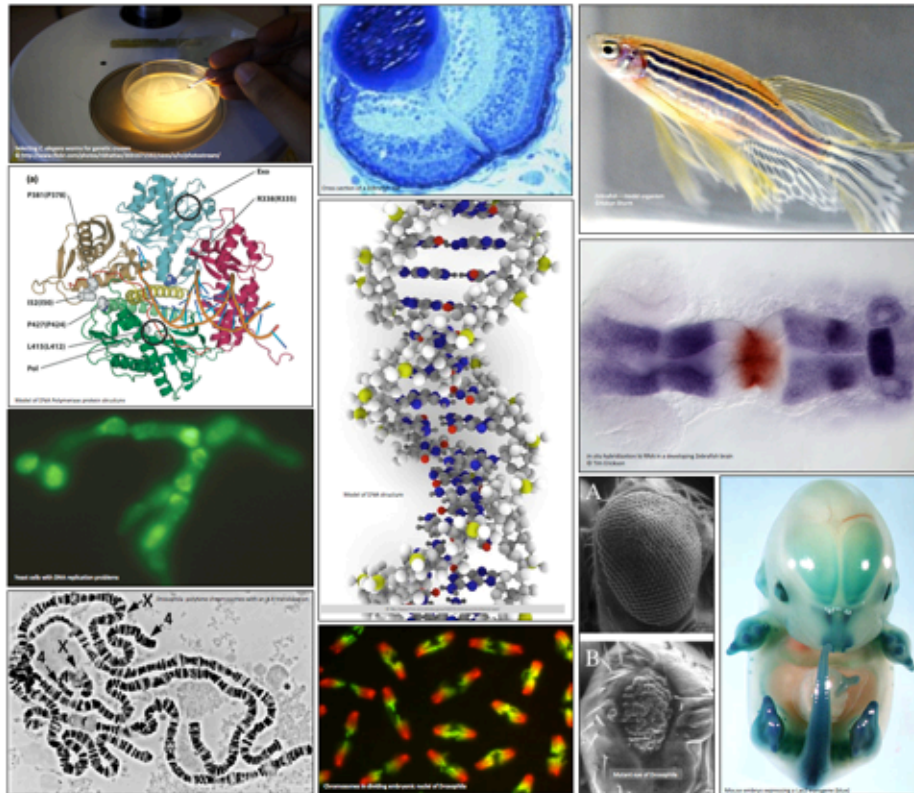
Enrollment: Ask your instructor

Link:
https://www.ualberta.ca/biological-sciences/undergraduate-studies/research-certificate-in-science

CHAPTER 02 – DNA STRUCTURE AND REPLICATION
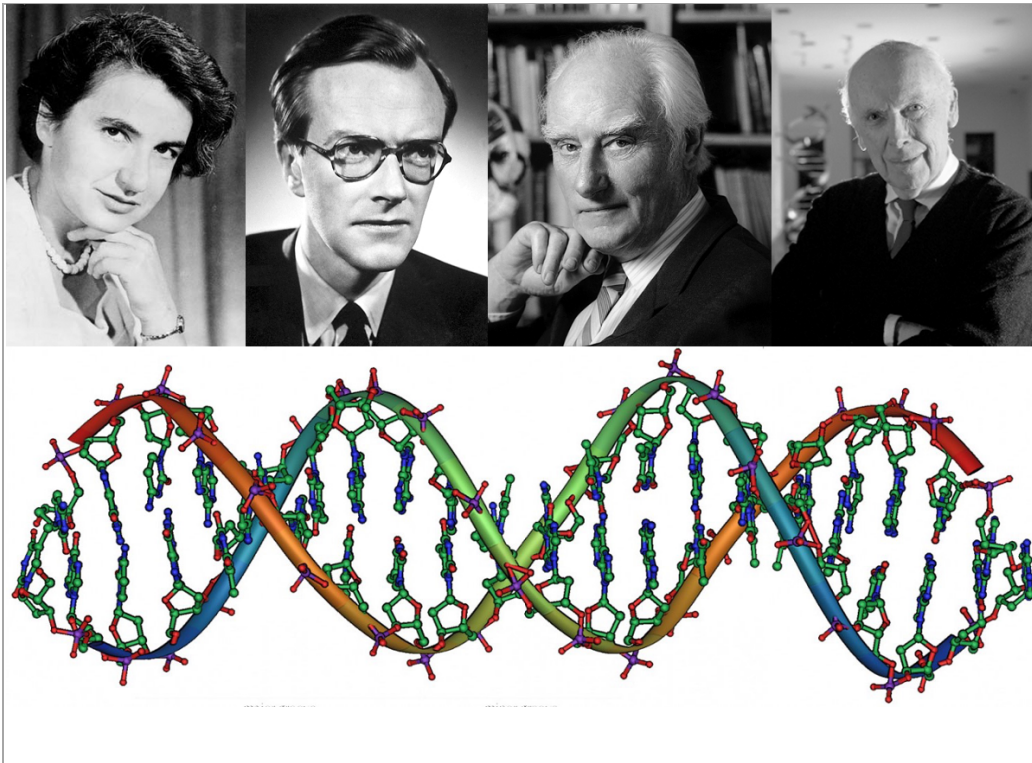
Figure 1.
The scientists responsible for determining the structure of DNA are Rosalind Franklin, Maurice Wilkins, Francis Crick and James Watson (left to right). Although the work was published in 1953, Wilkins, Crick, and Watson received the Nobel Prize in Physiology or Medicine in 1962, after Franklin had died in 1958 of ovarian cancer.

From left to right
(Wikipedia-Unknown-PD)
(Wikipedia-Unknown-PD)
(Wikiepdia-Marc Lieberman-CC BY 2.5)
(Wikipeida-Cold Spring Harbor Laborator-PD)
Bottom:
(Wikipedia-Michael Ströck-CC BY-SA 3.0)

INTRODUCTION

One of the fundamental things to know when studying genetics is the basic structure of DNA and how it is replicated. DNA is the "blueprint" that contains all the instructions for making the proteins that each cell needs, whether it is a single celled bacterium or a multicellular organism like humans. J. Watson, F. Crick, and M. Wilkins received the Nobel Prize (1962) for discovering the structure of DNA. (R. Franklin might have also received the prize for this discovery, but she died in 1958.)

The basic structure of DNA provides insight into its function. The main features of its structure are that it can reliably: (1) reproduce exact copies of itself to pass on to descendant cells, and (2) use the information to create proteins that produce and regulate the biochemistry of the cell. Remember however, DNA within the cell is more than just a loose strand within the nucleus. DNA interacts with proteins and is packaged into higher order structures (chromosomes) that will be discussed later in the textbook in Chapter 7. These proteins also regulate the expression of genes (information in the DNA).

This chapter will cover the components of the DNA molecule, how the double helix structure was discovered, and how the mechanisms of replication were discovered and characterized.

1. DNA STRUCTURE - DOUBLE HELIX

1.1. NUCLEIC ACIDS AND PHOSPHATE SUGAR BACKBONE

In 1869 Johannes Friedrich Miesher, a Swiss physician and biologist, first isolated a substance he called 'nuclein' from the nuclei of human white blood cells. He identified this substance to be weakly acidic with a high amount of phosphorus. This substance, after being further purified and studied was later called **deoxyribonucleic acid**, or DNA. Its name describes the three characteristics of the molecule: it has a ribose sugar with only one hydroxyl group called deoxyribose (**Figure 2**), it is found in the nucleus of a cell, and it is acidic.

After purifying the 'nuclein' to DNA they found it contained four different subunits that are linked in

a chain. Those subunits were identified as **nucleotides**. A nucleotide contains three components, a phosphate group (PO_4^{3-}), a deoxyribose sugar, and one of four nitrogenous bases. Those bases fit into two groups based upon their structure. **Purines** have a double ring structure and include **adenine** and **guanine**. **Pyrimidines** have a single ring structure and include **cytosine** and **thymine** (**Figure 2**). The nature of the phosphate group and the deoxyribose sugar allows each nucleotide to chain together, forming the long DNA strand.

Notice in **Figure 2**, dAMP has each carbon of the ribose labeled with a number followed by a prime, e.g. 1'-5'. The 1' position is where the base is attached. The 2' position of the ribose is missing a hydroxyl group. The 5' position is attached to the phosphate group. When linked in a chain, the phosphate group is linked to the 3' oxygen of the next nucleotide using a **phosphodiester bond**. When a polynucleotide chain is formed, there will always be a free 5' phosphate at one end, and one the free 3' oxygen on the ribose at the other. These are known as the 5' and 3' ends, respectively, of the DNA strand.

Ribonucleic acid (RNA) is like DNA, in that it forms chains similarly, and has the bases attached to the same carbon. The extra hydroxyl group at the 2' position causes it to form a different conformation than DNA, becoming a more flexible molecule (DNA's conformation will be described later in this chapter). There are also **dideoxynucleotides** that are missing the hydroxyl group at both the 2' and 3' position. Because of this, a chain cannot form at the 3' carbon, terminating the chain. This feature of dideoxynucleotides is used in Sanger sequencing, which will be described in Chapter 33.

Figure 2.
Molecular models of the components of DNA. The top shows the three different types of ribose sugar found in nucleic acids. The bottom shows the purine and pyrimidine nucleotide monophosphates. Carbon numbers (1'-5') are labeled on the sugars and dAMP. (Original- L.Canham - CC BY-NC 3.0)
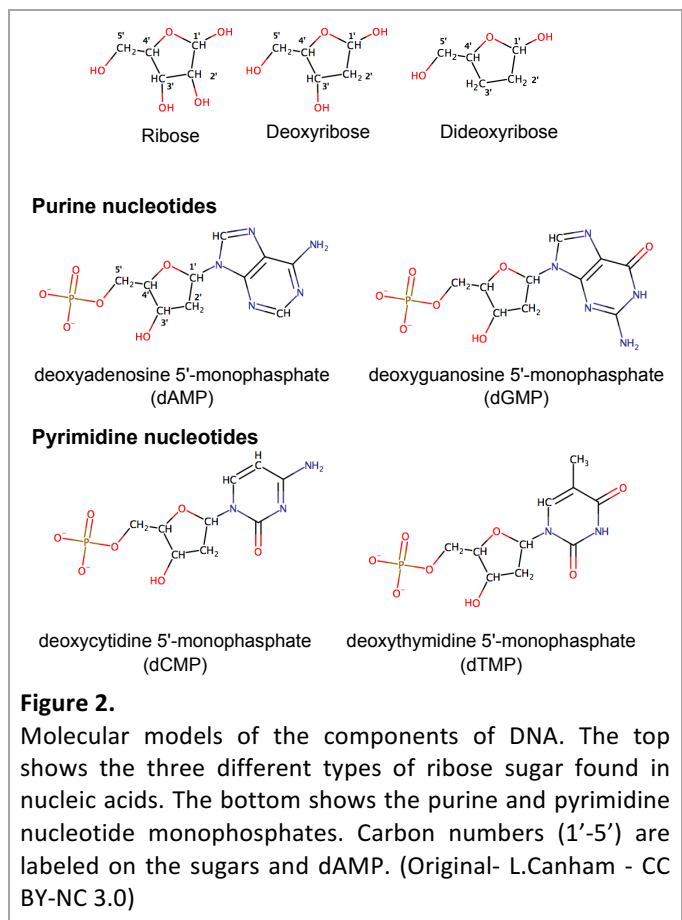
1.2. CHARGAFF'S RULES

When **Watson and Crick** set out in the 1940's to determine the structure of DNA, they already knew that DNA is made up of a series nucleotides with four different bases: adenine (A), cytosine (C), thymine (T), guanine (G). For DNA, the nucleotides are abbreviated as dNTPs (deoxyribonucleotide triphosphates), which include dATP, dCTP, dGTP, and dTTP. For RNA they are abbreviated as NTPs, which include ATP, CTP, GTP, and UTP. Watson and Crick also knew of **Chargaff's Rules**, which were a set of observations about the relative amount of each nucleotide that was present in almost any extract of DNA. Chargaff had observed that for any given species, the abundance of A was the same as T, and G was the same as C. This was essential to Watson & Crick's model.

Figure 3.
Chemical structure of two pairs of nucleotides in a fragment of double-stranded DNA. Sugar, phosphate, and bases A,C,G,T are labeled. Hydrogen bonds between bases on opposite strands are shown by dashed lines. Note that the G-C pair has more hydrogen bonds than A-T. The polarity of each strand is indicated by the labels 5' and 3'. (Wikipedia- Madeleine Price Ball - CC0 1.0)

1.3. THE DOUBLE HELIX

Using proportional metal models of the individual nucleotides, Watson and Crick deduced a structure for DNA that was consistent with Chargaff's Rules and with x-ray crystallography data that was obtained (with some controversy) from another researcher named Rosalind Franklin. In Watson and Crick's famous **double helix**, each of the two strands contains DNA bases connected through covalent bonds to a sugar-phosphate backbone (**Figure 1** and **Figure 3**). Because one side of each sugar molecule is always connected to the opposite side of the next sugar molecule, each strand of DNA has polarity: these are called the 5' (5-prime) end and the 3' (3-prime) end. The two strands of the double helix run in **anti-parallel** (i.e. opposite) directions, with the 5' end of one strand adjacent to the 3' end of the other strand. The double helix has a **right-handed** twist, (rather than the left-handed twist that is often represented incorrectly in popular media). The DNA bases extend from the backbone towards the center of the helix, with a pair of bases from each strand forming hydrogen bonds that help to hold the two strands together. Because of the structure of the bases, A can only form hydrogen bonds with T, and G can only form hydrogen bonds with C (remember Chargaff's Rules). Each strand is therefore said to complement to the other, and so each strand also contains enough information to act as a template for the synthesis of the other. This complementary redundancy is important in DNA replication and repair.

Under most conditions, the two strands in the double helix are slightly offset, which creates a **major groove**, and a **minor groove**. In **Figure 1**, notice how if you look at the bottom edge of the helix you can see it makes a wave pattern, with a large dip followed by a small dip, followed by a large dip, etc. The "peaks" are not equidistant and you can see the major and minor grooves. These grooves provide access for transcription regulating proteins (transcription factors), which bind to specific sequences of bases along the DNA.

2. SEMI-CONSERVATIVE REPLICATION (VS. CONSERVATIVE, DISPERSIVE)

From the complementary strands model of DNA, proposed by Watson and Crick in 1953, there were three straightforward possible mechanisms for DNA replication: (1) semi-conservative, (2) conservative, and (3) dispersive (**Figure 4**).

The **semi-conservative** model proposes the two strands of a DNA molecule separate during replication and then strand acts as a template for synthesis of a new, complementary strand.

The **conservative** model proposes that the entire DNA duplex acts as a single template for the synthesis of an entirely new duplex.

The **dispersive** model has the double helix breaking into segments that which are then replicated and reassembled, with the new duplexes containing alternating segments from one strand to the other.
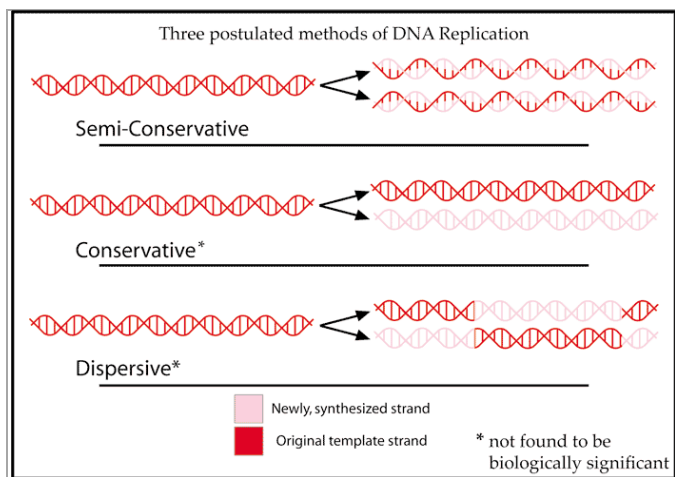
Figure 4.
The three models of DNA replication possible from the double helix model of DNA structure.
(Wikipedia-Adenosine- CC BY-SA 2.5)

Each of these three models makes a different prediction about the how DNA strands should be distributed following two rounds of replication. These predictions can be tested in the following experiment by following the nitrogen component in DNA in *E. coli* as it goes through several rounds of replication. Two scientists, **Meselson and Stahl** in 1958, used different isotopes of Nitrogen, which is a major component in DNA. **Nitrogen-14 (^{14}N)** is the most abundant natural isotope, while **Nitrogen-15 (^{15}N)** is rare, but also heavier. Neither is radioactive; each can be followed by a difference in density – "**light**" 14 vs "**heavy**" 15 atomic weight in a CsCl density gradient ultra-centrifugation of DNA.

The experiment starts with *E. coli* grown for several generations on medium containing only ^{15}N. It will have denser DNA. When extracted and separated in a CsCl density gradient tube, this "heavy" DNA will move to a position nearer the bottom of the tube in the more dense solution of CsCl (left side in **Figure 5**). DNA extracted from *E. coli* grown on normal (^{14}N containing) medium will migrate towards the less dense top of the tube.

If these *E. coli* cells are transferred to a medium containing only ^{14}N, the "light" isotope, and grown for one generation, then their DNA will be composed of one-half ^{15}N and one-half ^{14}N. If the this DNA is extracted and applied to a **CsCl gradient**, the observed result is that one band appears at the point midway between the locations predicted for wholly ^{15}N DNA and wholly ^{14}N DNA (**Figure 5**). This "single-band" observation is inconsistent with the predicted outcome from the conservative model of DNA replication (disproves this model), but is consistent with both that expected for the semi-conservative and dispersive models.

If the *E. coli* is permitted to go through another round of replication in the ^{14}N medium, and the DNA extracted and separated on a CsCl gradient tube, then two bands were seen by Meselson and Shahl: one at the ^{14}N-^{15}N intermediate position and one at the wholly ^{14}N position (**Figure 5**). This result is inconsistent with the dispersive model (a single band between the ^{14}N-^{15}N position and the wholly ^{14}N position) and thus disproves this model. The two band observation is consistent with the semi-conservative model which predicts one wholly ^{14}N duplex and one ^{14}N-^{15}N duplex. Additional rounds of replication also support the semi-conservative model/hypothesis of DNA replication. Thus, the semi-conservative model is the currently accepted mechanism for DNA replication. Note however, that we now also know from more recent experiments that whole chromosomes, which can be millions of bases in length, are also semi-conservatively replicated.

These experiments, published in 1958, are a wonderful example of how science works. Researchers start with three clearly defined models (hypotheses). These models were tested, and two (conservative and dispersive) were found to be inconsistent with the observations and thus disproven. The third hypothesis, semi-conservative, was consistent with the observations and thereby supported and accepted as mechanism of DNA replication. Note, however, this is not "proof" of the model, just strong evidence for it; hypotheses are not "proven", only disproven or supported
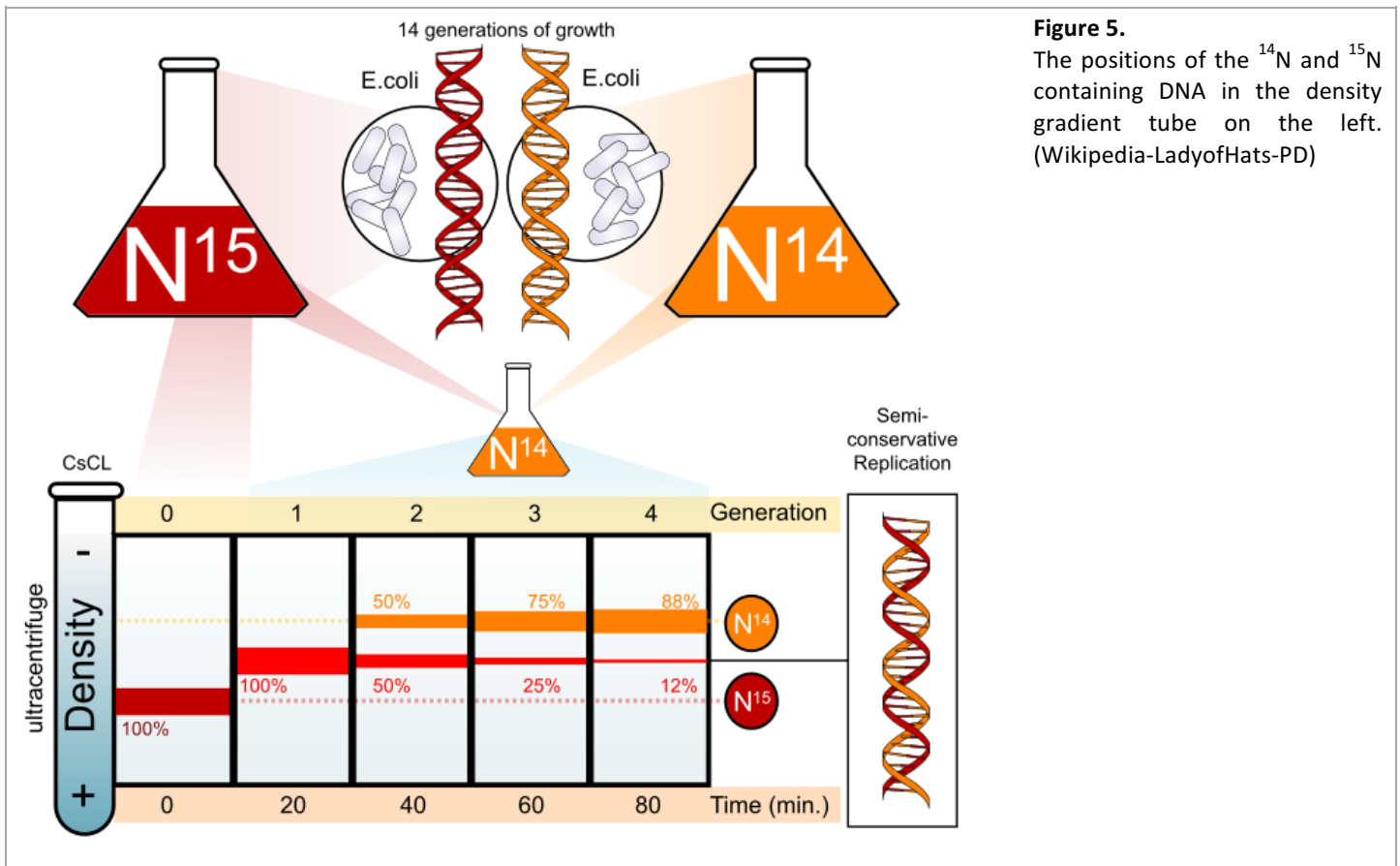
3. CHROMOSOME REPLICATION (*E. COLI*) - CAIRNS EXPERIMENT

If the results of Meselson and Stahl were true and there was semi-conservative replication, then the two strands of DNA have to separate to provide the template for copying. This should be seen as a 'fork' in a linear model if you manage to see the DNA just as it's replicating. **John Cairns** in 1963 chose to test this.

To do this he took *E. coli* cells growing in a normal environment, and then allowed them to grow and replicate in the presence of radioactive **^{3}H-thymidine**. The hypothesis is that if the *E. coli*'s DNA or chromosome is semi-conservatively replicated then after the first round of replication there should be one newly made strand that is radioactive, or "hot", and the other strand that is the parental template strand with no radioactivity, so is "cold". The original parental DNA will have two strands, each not radioactive. After replication the daughter DNA will have two strands, one that is radioactive and one that is not. After a third round of replication there will be a two types of daughter DNA, one that has a non-radioactive strand and a radioactive strand, and one that has two radioactive strands.

After growth in the ^{3}H-thymidine, Cairns lysed the bacteria and collected the contents onto a microscope slide. He then covered the slide with a **photographic emulsion** and allowed exposure to film for 2 months. As the ^{3}H-thymidine decays it emits an electron with a lot of energy and speed, known as a beta particle. The emulsion reacts with the beta particle creating a black **silver grain** on the film. The density of grains should be indicative of whether one or two strands are radioactive.

After the first replication cycle, the film had a thin circular ring of grains (**Figure 6**). This was interpreted to be a daughter chromosome with one strand that is hot and one strand cold. This also provided physical evidence that the *E. coli* chromosome is circular, something that has only previously been shown genetically.

In the second replication cycle the replication fork was seen. Here Cairns saw the typical thin ring of grains much like the first replication cycle, but with a branch in the middle that had a thicker strand (**Figure 6**). This means that the branch seen was an actively replicating chromosome, using the radioactive strand of DNA as a template, and adding more radioactive thymidine as the DNA is being synthesized. Because of the shape these created on the film this replicating structure was called a **theta (Θ) structure**. Cairns observed many different molecules corresponding to the progression from starting replication to the completion of replication.
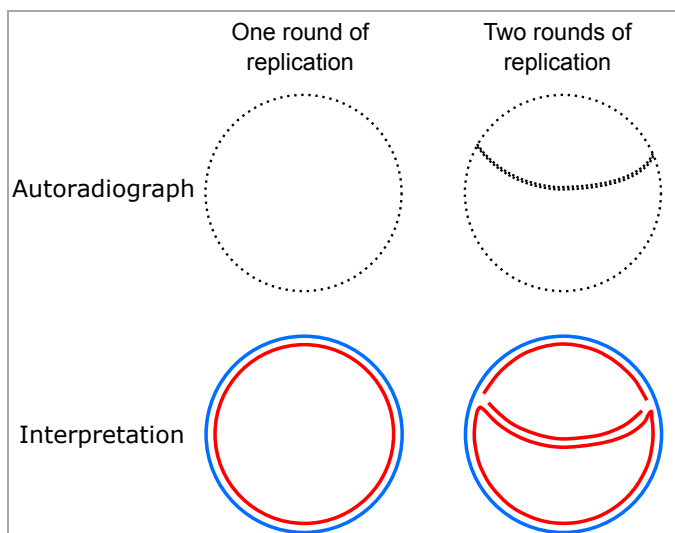
Figure 6.
In his experiment, Cairns looked at DNA with radioactive thymidine on an autoradiograph film, with the radioactive thymidine leaving dots on the film. This figure shows what the autoradiograph film would look like, and below what the interpretation of what the autoradiograph shows. The blue line represents the 'cold' DNA that has no radioactivity, while the red shows the 'hot' radioactive DNA. The density of the dots on the autoradiograph imply whether there is one strand or both strands of hot DNA. During the second round of replication, a theta structure can be seen, as the circular *E. coli* DNA is in the process of being replicated. (Original-L.Canham- CC BY-NC 3.0)

Here Cairns' results were able to further support the semi-conservative replication theory, showing the existence of replication forks, as well as the hypothesis that *E. coli* has a circular chromosome. What Cairns did not realize is that replication goes in both directions at the replication fork, where he

thought one fork was static while the other strand went around the chromosome replicating. Scientists later went on to show that replication is in-fact **bidirectional**.

4. ORIGINS OF REPLICATION (PROKARYOTE - SINGLE ORIGIN), REPLICATION FORK

When the cell enters S-phase in the cell cycle (See Chapter 14) the entire chromosomal DNA is replicated. This is done by enzymes called **DNA polymerases**. All DNA polymerases synthesize new strands by adding nucleotides to the 3'OH group present on the previous nucleotide. For this reason they are said to work in a 5' to 3' direction. DNA polymerases use a single strand of DNA as a template upon which it will synthesize the complementary sequence. This works fine for the middle of chromosomes. DNA-directed DNA polymerases travel along the original DNA strands making complementary strands (**Figure 7**a).
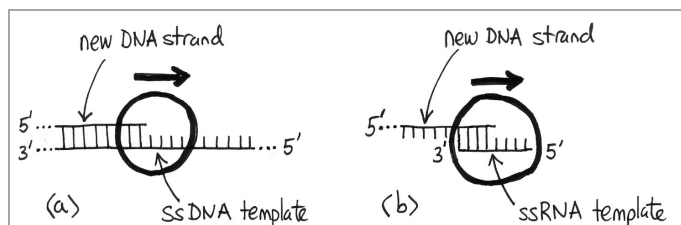
Figure 7.
DNA polymerases make new strands in a 5' to 3' direction. (a) Regular DNA polymerases are proteins or protein complexes that use a single strand of DNA as a template. For example, the main human DNA polymerase, Pol α, is large protein complex made of four polypeptides. (b) Telomerases use their own RNA as a template. The human telomerase is a complex made of one polypeptide and one RNA molecule.(Original-Harrington- CC BY-NC 3.0)

DNA replication in both prokaryotes and eukaryotes begins at an **Origin of Replication** (Ori). Origins are specific sequences on specific positions on the chromosome. In *E. coli*, the *OriC* origin is ~245 bp in size. Chromosome replication begins with the binding of the DnaA initiator protein to an AT-rich 9-mer in *OriC* and melts the two strands. Then DnaC loader protein helps DnaB helicase protein extend the single stranded regions such that the DnaG primase can initiate the synthesis of an RNA primer, from which the DNA polymerases

can begin DNA synthesis at the two replication forks. The forks continue in opposite directions until they meet another fork or the end of the chromosome (**Figure 8**).

Figure 8.
An origin of replication. The sequence-specific DNA duplex is melted, then the primase synthesizes RNA primers from which bidirectional DNA replication begins as the two replication forks head off in opposite directions. The leading and lagging strands are shown along with Okazaki fragments. Note the 5' and 3' orientation of all strands. (Original-Locke- CC BY-NC 3.0)

5. Eukaryote Chromosome Replication - Multiple Origins

In prokaryotes, with a small, simple, circular chromosome, only one origin of replication is needed to replicate the whole genome. For example, *E. coli* has a ~4.5 Mb genome (chromosome) that can be duplicated in ~40 minutes assuming a single origin, bi-directional replication, and a speed of ~1000 bases/second/fork for the polymerase.

However, in larger, more complicated eukaryotes, with multiple linear chromosomes, more than one origin of replication is required per chromosome to duplicate the whole chromosome set in the 8-hours of the replicative phase (S-phase) of the cell cycle. For example, the human diploid genome has 46 chromosomes (6×10^9 basepairs). The shortest chromosomes are ~50 Mbp long and so could not possibly be replicated from one origin. Additionally, the rate of replication fork movement is slower,

only ~100 base/second. Thus, eukaryotes contain multiple origins of replication distributed over the length of each chromosome to enable the duplication of each chromosome within the observed time of S-phase (**Figure 9**).

Figure 9.
Part of a eukaryotic chromosome showing multiple Origins (1, 2, 3) of Replication, each defining a **replicon** (1, 2, 3). Replication may start at different times in S-phase. Here #1 and #2 begin first then #3. As the replication forks proceed bi-directionally, they create what are referred to as "**replication bubbles**" that meet and form larger bubbles. The end result is two semi-conservatively replicated duplex DNA strands. (Original-Locke- CC BY-NC 3.0)

6. Telomeres

The ends of linear chromosomes present a problem – at each end one strand cannot be completely replicated because there is no primer to extend and replace the end RNA primer. While the loss of such a small sequence might not be a problem, the continued rounds of replication would result in the continued loss of sequence from the chromosome end. Ultimately, the losses would reach a point where essential gene sequences would be lost and the organism would die. Thus, this end DNA must be replicated. Most eukaryotes solve the problem of synthesizing this unreplicated, end 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 10.** 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 telomeric DNA has the complementary sequence 5'-TTAGGG-3' repeated over and over 1000's of times. After the telomerase has made the first strand, a primase synthesizes an RNA primer and a regular DNA polymerase can then make a complementary strand so that the telomere DNA will ultimately be double stranded to the original length (**Figure 10**). Note: the number of repeats, and thus the size of the telomere, is not set. It fluctuates after each round of the cell cycle. Because there are many repeats at the end, this fluctuation maintains a length buffer – sometimes it's longer, sometimes it's shorter – but the average length will be maintained over the generations of cell replication.

In the absence of telomerase, as is the case in human somatic cells, repeated cell division leads to the "**Hayflick limit**", where the telomeres shorten to a critical limit and then the cells enter a senescence phase of non-proliferation. The inappropriate 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. **HeLa cells**, which can be propagated essentially indefinitely, have been kept in culture since 1951 (See Chapter 41).

Black Represents Parental DNA Strands
Red Represents Newly Synthesized DNA Strands
Green represents RNA primer

Figure 10.
Telomere replication showing the completion of the leading strand and incomplete replication of the lagging strand. The gap is replicated by the extension of the 3' end by telomerase and then filled in by extension of an RNA primer.
(Original-Locke- CC BY-NC 3.0)

_____

SUMMARY:

- DNA is a double helix made of two anti-parallel strands of bases on a sugar-phosphate backbone.

- Specific bases on opposite strands pair through hydrogen bonding (A=T and G=C), ensuring complementarity of the strands.

- The hereditary information is present as the sequence of bases along the DNA strand.

- Chromosome replication begins at an origin and proceeds by DNA polymerases at a replication fork.

- Replication proceeds bi-directionally.

- Typically eukaryotes have multiple origins along each chromosome, while prokaryotes have only one.

- Eukaryotes have telomerase to complete the replication of the ends of chromosomes.

KEY TERMS:

deoxyribonucleic acid
nucleotides
purine
adenine
guanine
pyrimidine
cytosine
thymine
phosphodiester bond
ribonucleic acid
dideoxynucleotide
Watson and Crick
Chargaff's Rules
double helix
anti-parallel
right-handed
major groove
minor groove
semi-conservative
conservative
dispersive

E. coli
Meselson and Stahl
Nitrogen-14
Nitrogen-15
light
heavy
CsCl gradient
John Cairns
^{3}H-thymidine
photographic emulsion
silver grain
theta structure
bidirectional
DNA polymerases
Origin of replication
replicon
replication bubble
telomerase
riboprotein
Hayflick limit
HeLa cells

STUDY QUESTIONS:

1) Compare Watson and Crick's discovery with Avery, MacLeod and McCarty's discovery.
   a) What did each discover, and what was the impact of these discoveries on biology?
   b) How did Watson and Crick's approach generally differ from Avery, MacLeod and McCarty's?
   c) Briefly research Rosalind Franklin on the internet. Why is her contribution to the structure of DNA controversial?

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

3) Refer to Watson and Crick's
   a) List the defining characteristics of the structure of a DNA molecule.
   b) Which of these characteristics are most important to replication?
   c) Which characteristics are most important to the Central Dogma?

4) Refer to **Figure 3**.
   a) Identify the part of the DNA molecule that would be radioactively labeled in the manner used by Hershey & Chase
   b) DNA helices that are rich in G-C base pairs are harder to separate (e.g. by heating) than A-T rich helices. Why?

5) Are the ends of eukaryote, linear chromosomes static and fixed in length? Explain.

CHAPTER 03 – GENES ENCODE PROTEINS

Figure 1.
Most, but not all, genes code for proteins. They are transcribed into mRNA, which is then translated into polypeptides.
(pixabay-PublicDomainPictures-   CC0 1.0)

INTRODUCTION

How is the genetic information in DNA (genes) expressed as 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**. While not all genes code for proteins, most do. This chapter describes the Central Dogma and some experiments that were used to support this concept.

1. CENTRAL DOGMA

The Central Dogma of Biology describes the concept that genetic information is encoded in DNA in the form of genes (**Figure 2**). This information is then transferred as needed, in a process called **transcription** into a messenger RNA (mRNA) sequence. The information is then transferred again, in a process called **translation** into a polypeptide (protein) sequence. 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.

The original 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 be converted back to DNA through a process called **reverse transcription**. As well, DNA, and its information, can also be replicated (DNA➔DNA).

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 sequence and thus structure of each protein, DNA directs 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. Note, however, that the environment can also influence phenotype.
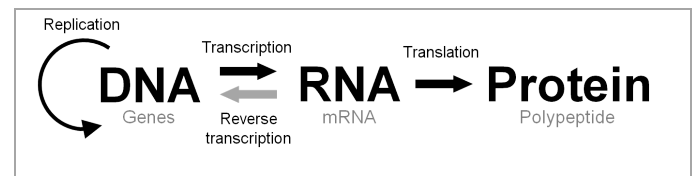
Figure 2.
Central Dogma of molecular biology.
(Original-Locke/Kang- CC BY-NC 3.0)

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 gene coding for an enzyme in a biochemical pathway.

2. GENES CODE FOR ENZYMES – A. GARROD

Life depends on (bio)chemistry to supply energy and to produce the molecules that construct and regulate cells. In 1908, Archibald Garrod described "in-born errors of metabolism" in humans, using the congenital disorder, alkaptonuria (black urine disease), as an example of how "genetic defects" (genotype) led to the lack of an enzyme in a biochemical pathway and caused a disease (phenotype). The reason why people with alkaptonuria have black urine is because a chemical, called "alkapton", makes urine black when exposed to air. In normal people, enzymes catalyze the reaction to break down alkapton, but people who are born with the disease, due to genetic defect, cannot make such enzymes and therefore cannot break down alkapton. Garrod's work gave huge impact to modern genetics as it attempted to explain the biochemical mechanism behind the genes proposed in Mendelian genetics.

3. BEADLE AND TATUM: PROTOTROPHIC AND AUXOTROPHIC MUTANTS

In 1941, over 30 years after Garrod's discovery, **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 enzyme that acts in a biochemical pathway is encoded by a different gene. Although we now know of many exceptions to the "one gene, one enzyme" 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 bread mold) for their studies because it had practical advantages as a laboratory model organism. They knew that Neurospora was **prototrophic**, meaning that it could grow on **minimal medium (MM)**. Minimal medium lacked most nutrients, except for a few minerals, simple sugars, and one vitamin (biotin). Prototrophs can synthesize the amino acids, vitamin, etc. necessary for normal growth**.**

They also knew that by exposing *Neurospora* spores to X-rays, they could randomly induce mutations in genes (now known as damage to the DNA leading to DNA sequence change). Each spore exposed to X-rays potentially contained a mutation in a different gene. While most mutagenized spores were still able to grow (prototrophic), some spores had mutations that changed their phenotype from a prototroph into an **auxotrophic** strain, which could no longer grow on minimal medium. Instead these auxotrophs could grow on **complete medium** (**CM**), which was MM supplemented with nutrients, such as amino acids and vitamins, etc. (**Figure 3**). In fact, some auxotrophic mutations could grow on minimal medium with only one, single nutrient supplied, such as the amino acid arginine. This implied that each auxotrophic mutant was blocked at a specific step in a biochemical pathway and that by adding an essential compound, such as arginine, that block could be circumvented.
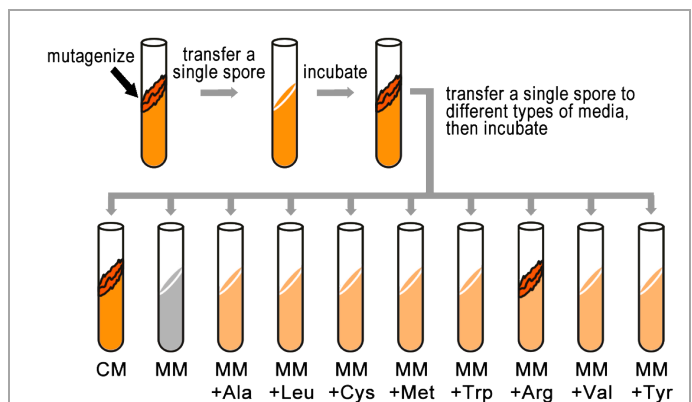
Figure 3.
A single mutagenized spore is used to establish a colony of genetically identical fungi, from which spores are tested for their ability to grow on different types of media. Because spores of this particular colony are able to grown only on complete medium (CM), or on minimal medium supplemented 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 BY-NC 3.0)

4. ONE GENE: ONE ENZYME HYPOTHESIS LED TO BIOCHEMICAL PATHWAY DISSECTION USING GENETIC SCREENS AND MUTATIONS

Beadle and Tatum's experiments are important not only for their 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.

Figure 4.
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 BY-NC 3.0)
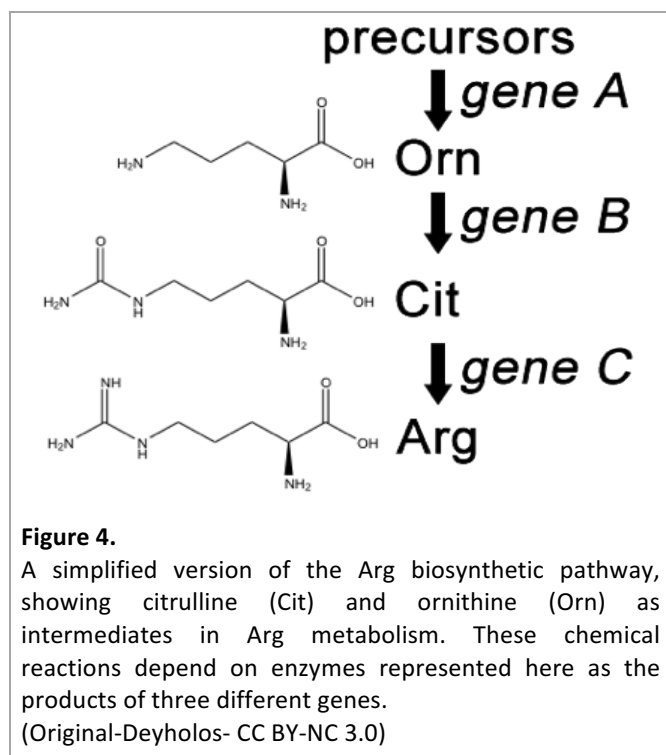
A convenient example is arginine. If the progeny of a mutagenized spore could grow on minimal medium only when it was supplemented with **arginine** (**Arg**), then the auxotroph must bear a mutation in the Arg biosynthetic pathway and was called an "arginineless" strain (arg-).

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 (**Figure 4**). For arginine (Arg), two of the biochemical 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 5**).

Figure 5.
Testing different Arg auxotrophs for their ability to grow on media supplemented with intermediates in the Arg biosynthetic pathway. (Original-Deyholos- CC BY-NC 3.0)
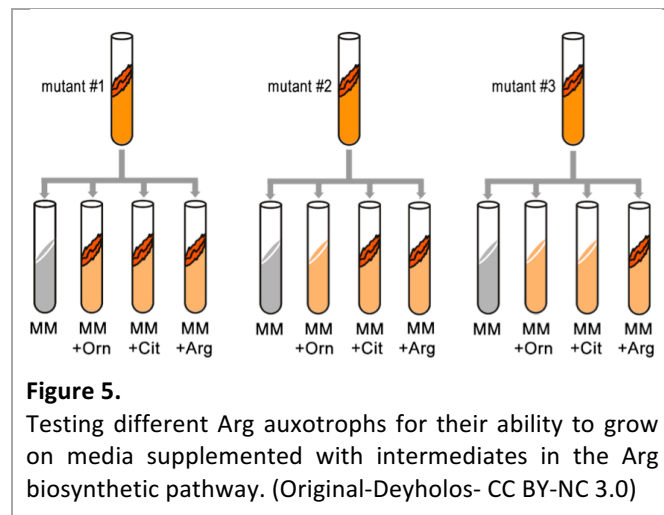
They found that only Arg could rescue all of the Arg auxotrophs, while either Arg or Cit could rescue some (**Table 1**). Based on these results, they deduced the location of each mutation in the Arg biochemical pathway, (i.e. which gene was responsible for the metabolism of which intermediate).

Mutants In:	MM + Orn	MM + Cit	MM + Arg
gene A	**Yes**	**Yes**	**Yes**
gene B	**No**	**Yes**	**Yes**
gene C	**No**	**No**	**Yes**

Table 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 4.

5. 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. Genetic screens are covered in more detail in Chapter 12.

_____

SUMMARY:

- The Central Dogma describes the information flow from nucleic acids to proteins.

- Garrod's observations showed that there is a connection between genes and enzymes.

- Beadle and Tatum proposed that one gene encoded one enzyme,

- It was an example of how to screen for genetic mutants, and therefore characterize biochemical pathways or biological processes.

KEY TERMS:

Central Dogma

transcription

translation

reverse transcription

genotype

phenotype

Beadle & Tatum

metabolic pathway

one-gene:one-enzyme

Neurospora crassa

prototroph

minimal medium

auxotroph

complete medium

genetic screen

genetic analysis

rescue

arginine

genetic screen for mutations

STUDY QUESTIONS:

1) Compare **Figure 4** and **Table 1**. Suppose you created three new arg⁻ mutation called mutants #1, #2, & #3. #1 grew on MM+cit and MM+arg, #2 grew on only MM+arg, while #3 grew on MM+ orn, cit or arg. Which genes are #1, 2, & 3 mutant in (A, B, or C)?

2) Why was the Vitamin biotin (see Section #3) always added the MM?

3) Last century, A. Garrod, and later Beadle and Tatum, showed that genes encode enzymes. From what we know now, do all genes encode enzymes? Explain.

4) Most mutant proteins differ from wild type (normal) by a single substitution at a specific amino acid site. Explain how some amino acid changes result in:
   a) no loss of protein function,
   b) only partial loss-of-function,
   c) complete loss-of-function,
   d) and how do changes at different amino acid sites result in the same complete loss-of-function.

5) Some mutants result in the loss of a specific enzyme activity. Does this mean that no protein product is produced from that mutant gene?

6) The molecular weight of the A and B chains of *E. coli* tryptophan synthase are 29,500 and 49,500, respectively. The size of the entire enzyme is 159,000.
   a) If the average molecular weight of each amino acid is 110, then how many amino acids are present in each chain?
   b) How many chains does the whole enzyme contain? Explain.

7) Recall that *Neurospora* is orange coloured bread mould. This biochemical pathway below is how wild type cells become orange. None of the compounds are essential. Cells containing W are white, cells with Y are yellow, and cells with O are orange. Assume that the reactions will go to completion if possible.
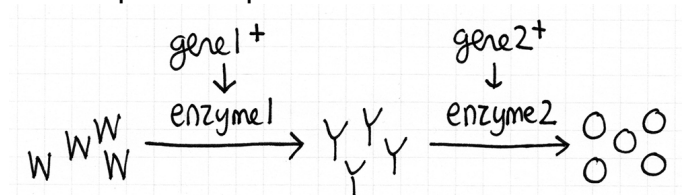
Fill in this table with the colours of the cell cultures.

Strain	MM+W	MM+Y	MM+O
gene1⁺ gene2⁺			
gene1⁻ gene2⁺			
gene1⁺ gene2⁻			
gene1⁻ gene2⁻			

CHAPTER 04 – COMPLEMENTATION
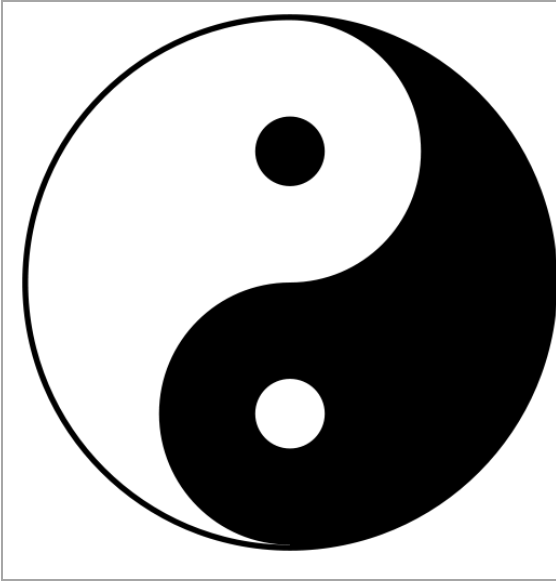
Figure 1.
In Chinese philosophy the *yin yang* symbol suggests opposite forces, such as two mutations, can actually be complementary and how together they can give rise to a whole, as with complementation in genetics.
(Wikipedia-Gregory Maxwell-PD)

INTRODUCTION

How do genetic researchers determine whether two mutants that have similar phenotypes are mutant in the same gene or in different genes? One way is by determining if the genes are located at a similar or different location. If they are different, they must be in different genes and thus are not allelic. If they are located in the same region then a complementation test is used. These consist of classical Mendelian genetic crosses to see if one mutant can complement another, or give a wild type phenotype. More recently, transformation of DNA with a gene has been used to see if putting a single gene into a cell/organism can rescue a mutant phenotype.

1. COMPLEMENTATION TESTS AND ALLELISM

As explained earlier in the previous chapter, mutant screening is one of the starting points geneticists use to investigate biological processes. Geneticists can observe two independently derived mutants with similar phenotypes, through a mutant screen or in natural populations. An immediate question from this observation 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 cause the same phenotype (e.g., in the same pathway). In other words, 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.
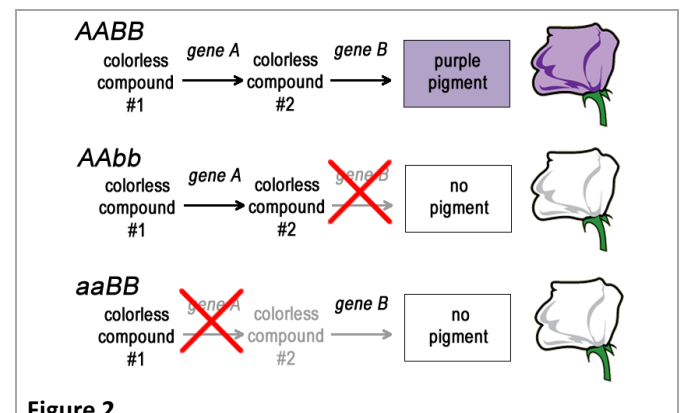
Figure 2.
In this simplified biochemical pathway, two enzymes encoded by two different genes modify chemical compounds in two sequential reactions to produce a purple pigment. Loss of either of the enzymes disrupts the pathway and no pigment is produced. (Original-Deyholos-CC BY-NC 3.0)

The easiest way to understand a complementation test is by example (**Figure 2**). 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* (Chapter 3). A diploid plant that lacks the function of gene A (genotype *aa*) would produce mutant white flowers that phenotypically looked just like the white flowers of a plant that lacked the function of gene B (genotype *bb*). Both A and B are enzymes in the same pathway that leads from a colorless compound #1, through colorless compound #2, to the purple pigment. Blocks at either step will result in a mutant white flower instead of the 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: Students often forget that genotypes usually only show mutant loci, however, one must remember all the other genes in the diploid genome 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, gene B or 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 3**).

If the F1 progeny all have the same mutant phenotype (Case 1 - **Figure 3**A), then we infer that the same gene is mutated in each parent. These mutations would then be called **allelic mutations** – mutant in the same gene locus. These two mutations FAIL to COMPLEMENT one another (still

mutant). These could be either the exact same mutant alleles (same base pair changes), or different mutations (different base pair changes, but in the same gene - allelic).

Conversely, if the F1 progeny all appear to be wild type (Case 2 - **Figure 3**B), then each of the parents most likely carries a mutation in a different gene. These mutations would then be called **non-allelic mutations** - mutant in a different gene locus. These mutations DO COMPLEMENT one another.
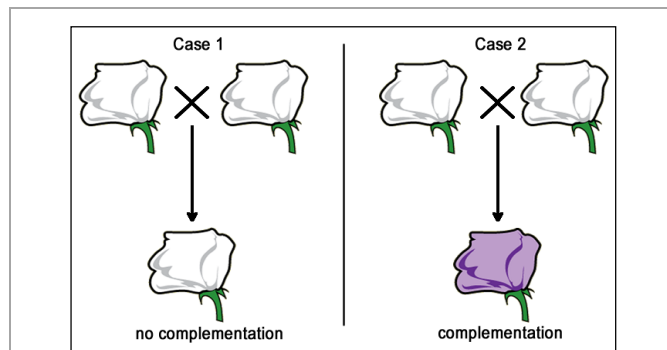
Figure 3A – Observation:
In a typical complementation test, the genotypes of two parents are unknown (although they must be pure breeding, homozygous mutants). If the F1 progeny all have a mutant phenotype (Case 1), there is no complementation. If the F1 progeny are all wild-type, the mutations have successfully complemented each other. (Original-Deyholos-CC BY-NC 3.0)
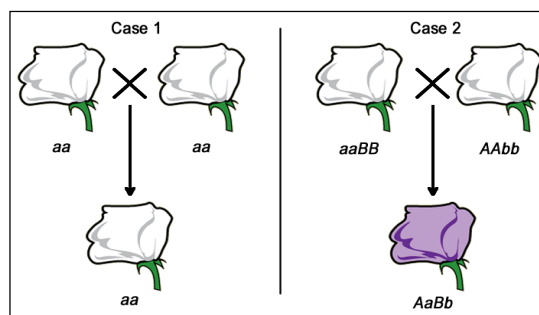
Figure 3B – Interpretation:
The pure breeding, homozygous mutant parents had unknown genotypes before the complementation test, but it could be assumed that they had either mutations in the same genes (Case 1) or in different genes (Case 2). In Case 1, all of the progeny would have the mutant phenotype, 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 F1 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 BY-NC 3.0)

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 and semi-dominant mutations CANNOT be used in complementation tests, since these mutations won't show complementation effects of two non-allelic genes. (3) Note that haploid organisms like Neurospora cannot be used in complementation test since they have only one set of chromosome (4). 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).

2. 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 they came from different natural populations or from independently mutagenized individuals, we could have a fourth, fifth, sixth, etc. white flower strain, then we could begin to organize the allelic mutations into groups, which are called **complementation groups**. These are groups of mutations that FAIL TO COMPLEMENT one another (a group of NON-complementing mutations) and are assumed to have mutations in the SAME gene; hence they are grouped as complementation group. 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 a gene. For example, the *white* gene in *Drosophila* has >300 different mutations within the *white* gene 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 or same *white* gene. Each complementation group represents a gene.

If, however, you obtained a different mutation, *vestigial* for example, which affects wing growth, and crossed it to a *white* eye colour 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 elsewhere in this text. For example, if you crossed a *scarlet* eye-colour mutant to a *white* eye-colour mutant, the double heterozygote would have wild type red eyes. Each mutant has the wild type allele of the other. Again, remember that all the other genes in the diploid genome are assumed to be wild type.

To drive home the concept of complementation groups, we will look at a two hypothetical examples.

2.1. EXAMPLE ONE: MULTIPLE MUTANT COMPLEMENTATION TEST

The first example, shows the results of a series of crosses as a complementation test table (**Figure 4**) with six mutants labeled a to f. The mutants fall into three complementation groups in total: (1) a (2) b, c, f and (3) d, e. Notice that a complementation group can consist of only one mutant, or more than one.

#	a	b	c	d	e	f
a	w					
b	p	w				
c	p	w	w			
d	p	p	p	w		
e	p	p	p	w	w	
f	p	w	w	p	p	w

Figure 4.
Complementation test table showing which flower mutant strains complement each other and vice versa. "w" stands for the white flowers, which is mutant (no complementation) and "p" stands for purple which represents wild type (complementation). Blanks are for crosses not done. (Original-Di Cara-CC BY-NC 3.0)

2.2. EXAMPLE TWO: DOUBLE HIT STRAIN

The second example is similar, but has a twist (**Figure 5**). It has five mutants labeled 1-5, with 1-4 being mutations in only a single gene each, while mutant #5 has mutations in two different genes, and thus is unable to complement the mutations in two, different genes. A **double-hit strain** like strain #5 is normally a very rare event, but is included here to make a point. A double-hit strain may appear to belong in two different groups. In this case, mutants #3 and #4 complement (different genes) but #5 fails to complement both #3 and #4, indicating it has mutations in both the mutant genes in #3 (gene B) and #4 (gene C) (**Figure 6**).

Example 2:

Figure 5.
Complementation test table with pink as mutant and green as wild type (black is for crosses not done). Note mutant #5 has two mutations. (Original-Locke-CC BY-NC 3.0)

Figure 6.
Chromosomes of the organisms that are used in complementation tests to decide if the genes are allelic or non-allelic. (Original-Locke-CC BY-NC 3.0)
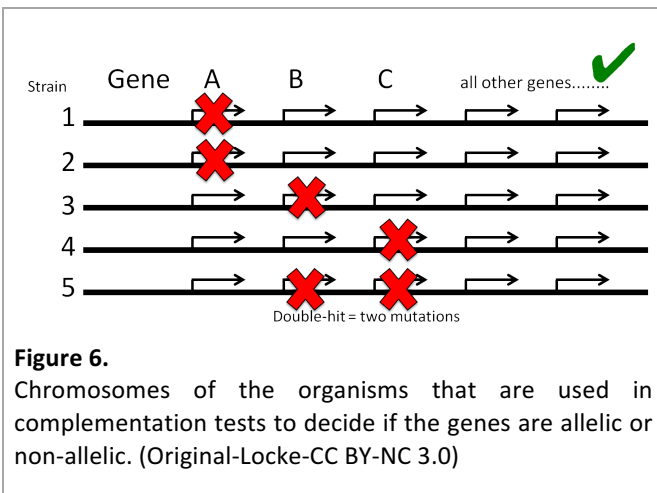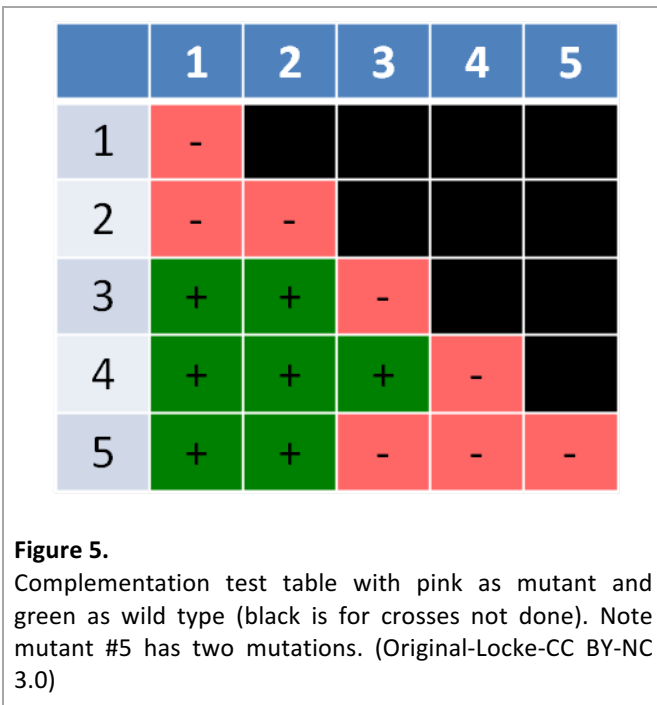
3. TRANSFORMATION RESCUE

In a normal rescue experiment (See chapter 3), arginine auxotrophic strands of single-celled *Neurospora crassa* were "rescued" when supplemented with the amino acids that they could not synthesize and that were essential for the organism's metabolism. In transformation rescue, rather than giving supplementary metabolic pathway products, it supplies the needed genes that can complement the mutant allele. The process of taking in foreign DNA (**transformation**) that contains the normal version of the gene and thereby **rescuing** the auxotrophic strain is called **transformation rescue**.

Let's say that there is an *E. coli* auxotrophic mutant in a gene called "a" (**Table 1**).

E. coli Strain	MM (Minimal medium)	MM + supplement
a^-	Auxotrophic (no growth)	Growth
a^+	Growth	Growth

Table 1.
The auxotrophic strand (a^-) cannot grow on MM (minimal medium) but the prototrophic strand (a^+) can grow.

In order to transform this auxotrophic strain and rescue, we need to:

(1) Make the *E. coli* auxotrophic cells competent so that it can incorporate foreign DNA molecules. We can form a competent cell via **heat shock** or **electroporation** that can slightly damage the membrane and therefore provide passageways for DNA molecules to enter the cell.

(2) Extract DNA molecules from a wild type strain of *E. coli* and break them down into short fragments using enzymes.

(3) Insert these short fragments of *E. coli* DNA into a **DNA vector**, which is a DNA molecule that can contain, amplify, and transfer the inserted DNA fragments into the host cell. This combined DNA molecule is called **recombinant DNA**. **Plasmids** are small circular DNA molecule that are mostly found in bacteria and are suitable as DNA vector. The

[vector + DNA insert] molecule can be replicated and the result would be multiple **clones** of the original DNA insert.

(4) After the *E. coli* DNA fragments that were once a single long DNA molecule are inserted into DNA vectors, we have a collection of recombinant DNA molecules, which when transformed, can be called a **DNA library.** Among all the recombinant DNA molecules in the library, there are three possibilities (**Figure 7**): (1) DNA clones that contain gene *a*, (2) DNA clones that don't contain gene *a*, which will be collectively presented by the letter *b* and (3) DNA clones that don't contain any foreign genes.

(5) Combine the recombinant DNA molecules and host *E. coli* strain together so that the auxotrophic strain can incorporate those DNA molecules through transformation. Growing the strains on minimal and complete media will let us decide if the transformation rescue worked or not.

The host strain's genotype is a⁻b⁺. It needs a wild type a⁺ in order to grow on minimal medium. Therefore, plasmids that have the a⁺ allele would grow (prototrophic), and other strains that have

either plasmids with no transgene or have plasmids with gene b′ would be still auxotrophic.

Notice that the plasmids contain an antibiotic resistance gene called **Anti^R** and that the strains were actually grown on minimal medium that contained antibiotics. Why was this so? This is because we want to select for the ones that actually incorporated the plasmid that contained the wild-type "*a*" gene.

Only a small fraction of cells is actually transformed by foreign DNA. Therefore, if we grow those strains on agar plate without antibiotics, we cannot guarantee that the growth was due to the complementation between the host DNA and the recombinant DNA or by some reversion back to wild type. There is a small possibility that the cells that weren't transformed could somehow synthesize the essential substrate due to a spontaneous mutation. Adding the antibiotic selection will remove cells that weren't transformed and therefore don't contain a plasmid with the antibiotic resistance gene, and select for the cells that were successfully transformed and complemented by the recombinant DNA.

Figure 7.
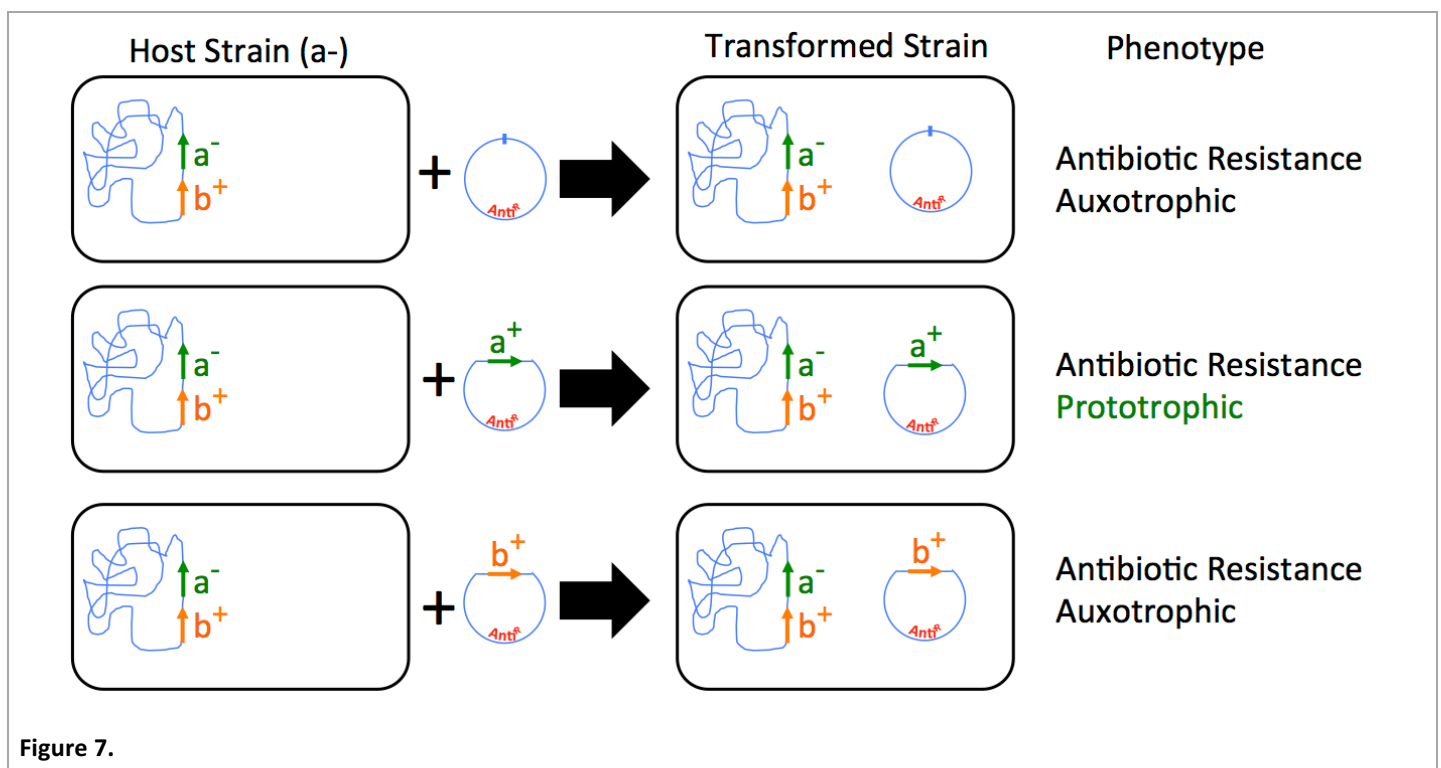
Transformation rescue diagram (Original-Locke-CC BY-NC 3.0)

_____

SUMMARY:

- 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).

- Complementation group contains mutants that cannot complement each other (allelic mutations) and therefore are assumed to have mutations at the same gene loci.

- Transformation rescue refers to the incorporation of recombinant DNA molecule that contains a gene that is able to complement the mutated gene in another organism.

KEY TERMS:

allelic mutations

non-allelic mutations

complementation test

biochemical pathways

complementation

complementation group

double-hit strain

transformation

rescue

transformation rescue

heat shock

electroporation

DNA vector

recombinant DNA

plasmids

clones

DNA library

Anti^R

STUDY QUESTIONS:

1) You are working with a prototrophic model organism (e.g. a fungus). You are interested in finding genes involved in synthesis of proline (Pro), an amino acid that is normally synthesizes by this organism.

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

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

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

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

2) Draw the expected results of a series of complementation tests (crosses), in the form of a table, for five yeast mutant strains where there are at least three different mutant loci, and one of the mutations involves a double hit (two loci are mutant in the same strain).

3) Students create a mutant *E. coli* strain that is auxotrophic for methionine. Three students build plasmid DNA libraries from wild type DNA from the parental strain. Student A uses *EcoRI* to clone the restriction fragments. Student B uses *HindIII* and student C uses *XhoI*. Each transforms the auxotrophic mutant strain with their library. Student A gets lots of prototrophic colonies on minimal medium, while students B and C don't get any. Explain what might have

happened. The student's control experiments indicate that the transformation protocol worked.

4) Figure 7 shows how we can rescue an a^- strain with a plasmid carrying an a^+ gene. Could we also rescue this strain by growing the cells on media containing Enzyme A (the product of the a^+ gene)? How about the product of Enzyme A?

Recommended course:

Foundations of Molecular Genetics

GENETICS 270, Foundations of Molecular Genetics

★ 3 (fi 6) (either term, 3-1.5s-0)

Basic concepts on the organization of genetic material and its expression will be developed from experiments on bacteria and viruses.

Why study Bacterial Genetics?

- Model systems for higher organisms
- Bacteria are highly complex cells
- Bacteria are interesting
  - Essential to ecology of the Earth
  - Symbiotic with higher organisms
  - Play a role in disease

Prerequisites: BIOL 207

Class schedule: T/Th 11:00AM - 12:20PM

Offered: Either Term

Contact: Fall - Dr. J. Dennis, jon.dennis@ualberta.ca

Winter – Dr. T. Raivio, traivio@ualberta.ca

Calendar Link:
http://calendar.ualberta.ca/preview_course_nopop.php?catoid=6&coid=44621

Chapter 05 - Genes to Genomes
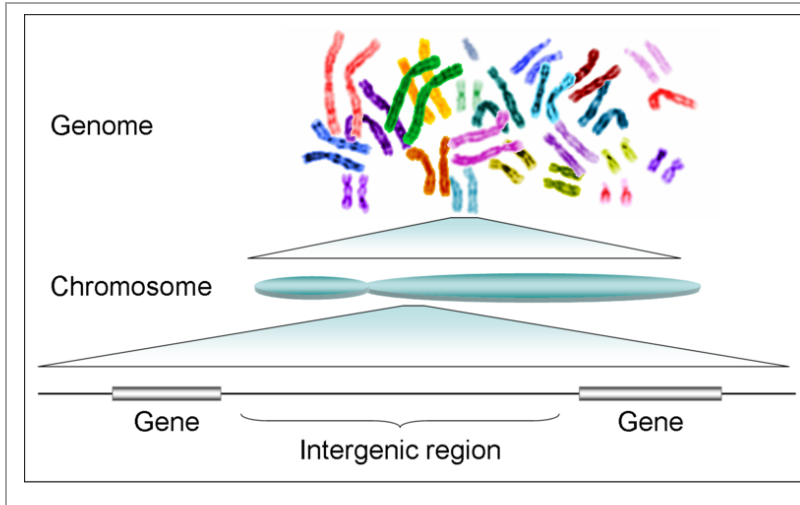
Figure 1.
The DNA in most genomes begins at the level of nucleotides, then to genes, chromosomes and finally to a whole genome.
(Wikipedia- Plociam- CC BY 2.0)

Introduction

With Gregor Mendel's work (1865), we transitioned from a "blended" concept of inheritance to a "particulate" concept. The particles were given the name "gene" in 1909 by Wilhelm Johannsen and William Bateson coined the word *genetics* at about that time. A first understanding in gene regulation came from Jacob and Monods' work on the *lac* operon. Marshall Nirenberg and Heinrich J. Matthaei in 1961 cracked the "genetic code" (codons). With the advent of DNA cloning and recombinant DNA in the 1970s we first glimpsed the interrupted nature of eukaryote genes in work by Phillip Allen Sharp and Richard J. Roberts. More recently, the Human Genome Project determined the entire nucleotide sequence of humans. Analysis predicts around 20,000-25,000 genes, but the actual number rides on deciding what is a gene. Since then many more genomes have been sequenced, both higher, multi-cellular eukaryotes, as well as hundreds (thousands?) of prokaryotes. Each contributes thousands of new genes to the databases.

What is a "gene" and how are they organized in a genome?

1. Central Dogma - Review

Molecular biology's **Central Dogma** (see Chapter 03) states the genetic information of each gene is encoded in the nucleotide sequence, 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. The protein coding genes: (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. Thus, nucleic acids (DNA and RNA) dictate the structure of each protein and the structure affects the function of that protein, which can thereby affect the entire organism. Thus the sum of all the genes present in the **genome** of an organism, or **genotype**, defines the potential form, or **phenotype** of the organism. In prokaryotes, nucleotides are normally a single, circular chromosome, while in eukaryotes it is present as multiple linear chromosomes.

2. What is a Gene?

Mendel's work showed that inheritance was particulate (not blended). These particles became known as **genes** and could exist in one or more different versions or forms, which we now call **alleles**. In its broadest and most general definition, a gene is an abstract concept with five components.

A gene is:

(1) a unit of inheritance
(2) that occupies a specific position (locus) within the genome or chromosome and
(3) has one or more specific effects upon the phenotype of the organism and
(4) can mutate into various forms (alleles) and
(5) can recombine with similar such units
(with modification from: King & Stansfield (1968) A dictionary of Genetics, Third Ed. Oxford University Press, New York).

Note that this definition of a gene isn't limited to just protein coding genes, but includes other types that will be discussed below. Today we know more about the different roles that DNA sequences play in the expression of genetic information. Often it is difficult to say for some sequences whether they are "genes" or not.

3.   BASIC TYPES OF DNA SEQUENCE

First, not all DNA sequences in a genome are parts of a gene. In prokaryotes, a minority of DNA sequences are not genes, i.e. most are. Those sequences not in genes have no apparent function other than linking one gene to the next (**intergenic regions**) (**Figure 2**). In eukaryotes, however, especially the multi-cellular species such as us, there appears to be a large percentage of non-functional DNA (~75-90%; sometimes called "Junk" DNA), depending upon what is called non-functional. Here we call it non-functional if it fails to influence the phenotype, even if it might have some biochemical activity. Remember all DNA is replicated (a biochemical activity) and all DNA is bound up in histone proteins (another biochemical activity), so not all biochemical activity should be considered as functional for the phenotype.

While most references to genes usually involves protein coding genes, the functional (gene containing) sequences can actually be divided into three main types:
(1) DNA acting directly
(2) DNA transcribed into RNA, which functions directly, and
(3) DNA transcribed into mRNA, which is translated into a polypeptide, which has a function.

4.   GENES: DNA ACTING DIRECTLY

DNA does not have to be transcribed and/or translated in order to have a function. The DNA itself has a function. Examples under this type

Figure 2.
DNA sequence can be divided into two main categories in a general sense: Functional and Non-functional. The Functional sequences, or genes, can be further divided. (Original-J. Locke- CC BY-NC 3.0)
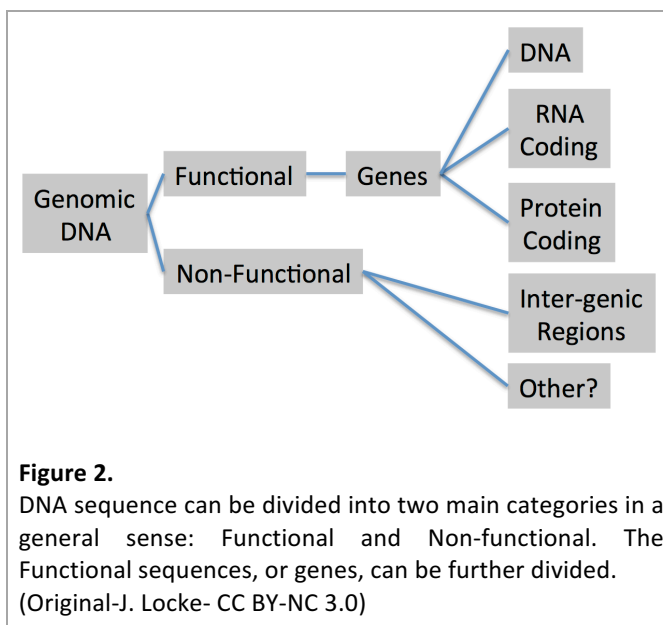
include origins of replication (Ori), centromeres, and telomeres. These sequences are essential, can be mutated, and occupy a specific location on a chromosome. They are unusual and not typically thought of as genes, but they are fit the definition above. They are typically the least frequent in the genome.

5.   GENES: DNA TRANSCRIBED INTO RNA (RNA-CODING GENES)

Some genes are transcribed but not translated into a protein. It is the RNA that functions - **functional RNA** molecules. There are many types of RNA molecules but for now, we will only list a few of them.

5.1.   PROTEIN SYNTHESIS RELATED
Ribosomes contain **rRNA**; the large and small subunits each have RNA molecules as components of their structure. These rRNAs are encoded by the rRNA genes, which are usually located as multi-gene repeats in clusters. **tRNA** is also a functional RNA that is involved in the transfer of amino acids to the elongating polypeptide chain. tRNA genes

are usually dispersed around the genome (not clustered).

5.2. SMALL RNAS (AN INCOMPLETE LIST):

snRNA (small nuclear RNA) reside in the nucleus and form an RNA-protein complex called "spliceosomes" that process the primary mRNA transcripts into the mature mRNA. In this complex, it is the RNA molecule, not the protein component that has the catalytic activity.

snoRNA (small nucleolar RNA) act as guides for other RNA molecules such as snRNA or rRNA molecules in modification process.

miRNA (microRNA) is a single-stranded RNA molecule that is about 22 bp long, and regulate gene expression in both transcription and post-transcriptional level.

siRNA (small interfering RNA) are short double stranded RNA molecules (about 21-24 bp long) and are also involved in **RNA interference (RNAi)** regulation of genes and post-transcriptional gene silencing (PTGS). These small interfering RNAs are similar in size and function with miRNAs but come from a different RNA precursor. Double-stranded DNA molecules are chopped into shorter fragments that are still double-stranded, which are called siRNAs. For example, RNA is transcribed from centromeric sequences and modified into siRNA fragments. This double stranded siRNA is broken down into two strands. One of the strands forms a complex with other proteins and this complex finds its precursor (the centromeric sequence) and modifies it into a highly condensed chromatin, producing **heterochromatin**.

piRNA (Piwi-interacting RNA) are single stranded RNA molecules of 24-32 bp in length that interact with a protein called *piwi* and this RNA-protein complex affects epigenetic and post-transcriptional gene silencing. For example, it can block the transcription from DNA transposons by turning the normal chromatin into heterochromatin.

Each of these types of small RNAs is transcribed from a gene.
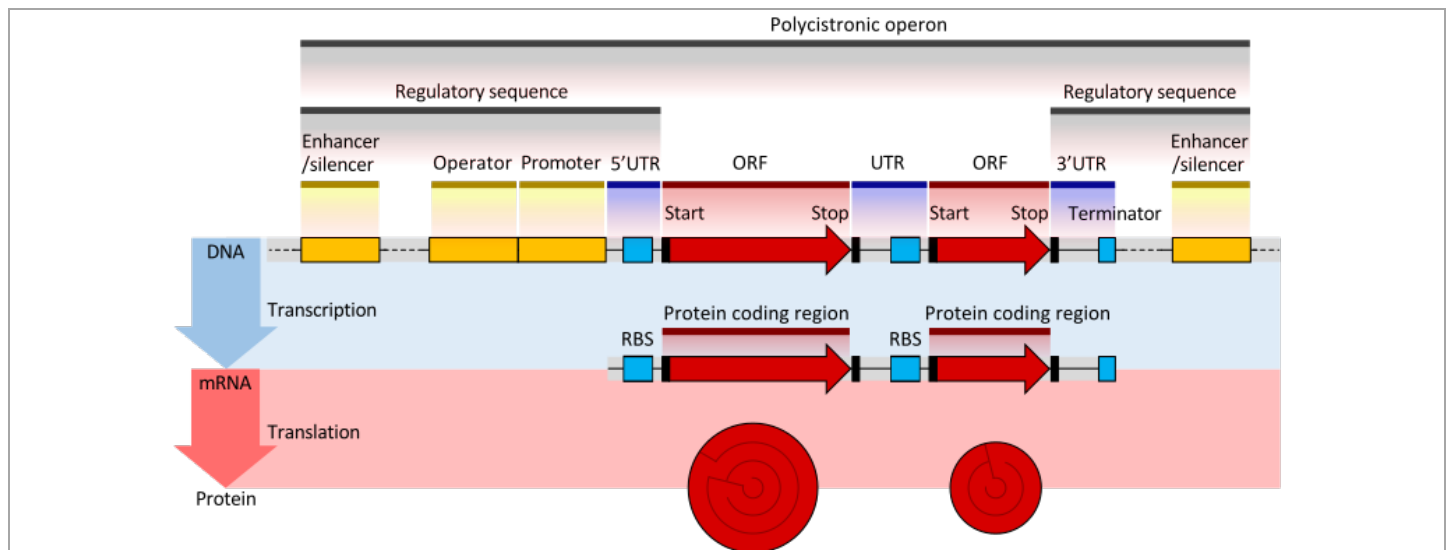
Figure 3.
Structure of a gene contains many components. This is a mix of prokaryote and eukaryote gene structure. For example, a polycistronic operon is found in prokaryotes, while the distant enhancer/silencer elements are found in eukaryotes. ORF is open reading frame; UTR is untranslated region; RBS is ribosome binding site.
(Wikipedia-Thomas Shafee-CC BY-SA 4.0)

6. GENES: DNA TRANSCRIBED INTO MRNA, TRANSLATED INTO A POLYPEPTIDE (PROTEIN CODING GENES)

Protein coding genes consist of both a regulatory and a transcribed sequence. The DNA is transcribed into an mRNA that is then translated into polypeptides as directed by the cis-regulatory elements in combination with various trans-acting factors. One thing to note is that a gene is depicted as simply a "block" or a "line" in various diagrams but it actually contains more than that; there are many components inside a gene (**Figure 3**). For example, in prokaryotes, a gene consists of:
- regulatory sequences: enhancer/silencer + operator + promoter

- transcribed sequences (transcription unit): 5'UTR (untranslated region) + open reading frame + 3'UTR (includes the terminator).

The **open reading frame (ORF)** refers to the sequence beginning at the start codon, through to the stop codon. Also, many diagrams depict these components as a "single unit" or as a cluster sequence on the same chromosome, but in eukaryotes, regulatory regions can be very far away (kilobases upstream/downstream) from the transcription unit.

7. HOW ARE GENES AND OTHER SEQUENCES DISTRIBUTED IN THE GENOME?

In our genome, genes are interspaced by inter-genic regions that contain interspersed repeats such as **SINE** (short interspersed elements) or **LINE** (long interspersed elements). Organisms that have smaller genes, such as bacteria, tend to have less inter-genic DNA compared to organisms that have larger genes, such as yeast, Drosophila and mammals.

Figure 4.
Comparison of the gene distribution between prokaryote and different eukaryotic organisms.
(Original-Locke- CC BY-NC 3.0)
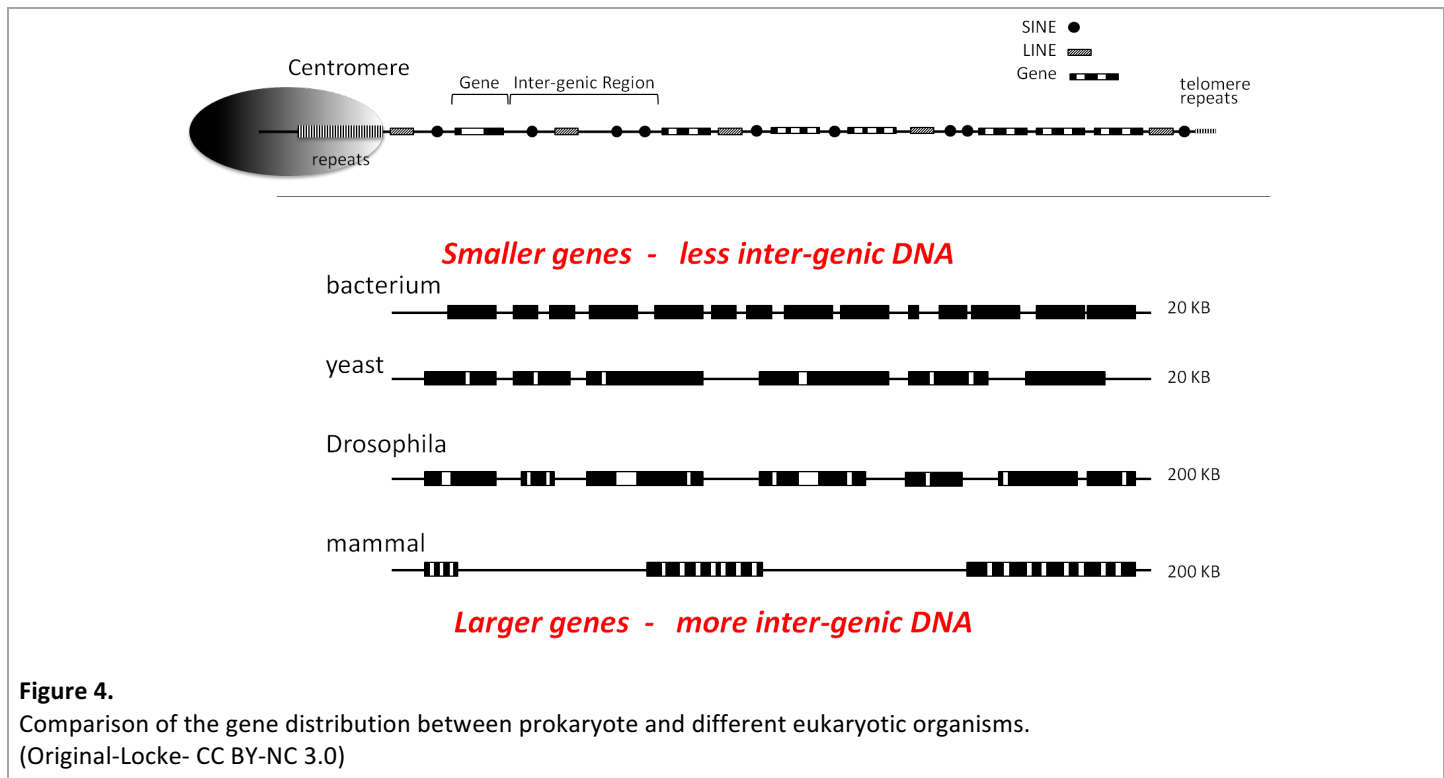
Major components of the Human Genome

Type of DNA Sequence	Essential/ Functional	Unknown	Junk/ Nonfunctional
Transposable Elements:			44%
DNA transposons	<0.1%		3%
Retrotransposons	<0.1%	<0.1%	41%
Viruses			9%
DNA viruses	<0.1%		~1%
RNA viruses	<0.1%	<0.1%	8%
Pseudogenes			1.2%
From Protein coding genes			1.2%
Co-opted pseudogenes to form a new function	<0.1%		
Ribosomal RNA genes	0.22%		0.19%
Other RNA encoding genes	~2.0%		
tRNA genes	<0.1%		
small RNA genes	<0.1%		
putative regulatory RNA genes	~2.0%		
Protein-encoding genes	1.8%		9.6%
Exon regions	1.8%		
Intron regions (including transposons, etc)			30%
Intron regions (not including other sequences)			9.6%
Regulatory sequences	0.6%		
Origins of replication	**<0.1%**		
Scaffold Attachment Regions (SARS)	**<0.1%**		
Highly Repetitive DNA			
α-satellite DNA (Centromeres)	**2.0%**		**1.0%**
Telomeres	<0.1%		
Intergenic DNA (not included in the above)			
Conserved	**2.0%**		
Non-conserved		**26.3%**	
Totals	**8.7%**	**26.3%**	**65%**

From: http://sandwalk.blogspot.ca/2011/05/whats-in-your-genome.html

_____

SUMMARY:

- The Central Dogma states that information in nucleic acids (DNA and RNA) is translated into protein, but it can never go back in the opposite direction. More recently it has been described as information flows from DNA->RNA->protein.

- The definition of a gene is changing as new discoveries are made but in general, a gene is a unit of inheritance that has a locus on a chromosome, can affect an organism's phenotype, can exist in various forms, and can recombine with other such units.

- DNA sequences can be divided into two main categories: functional and non-functional. The functional DNA can be divided into (1) DNA acting directly, (2) DNA transcribed into RNA, which functions directly, and (3) DNA transcribed into mRNA, which is translated into a polypeptide, which has a function.

- There are various kinds of RNA molecules that are involved in protein synthesis, DNA replication, post-transcriptional modification, and gene regulation.

- Protein coding genes contain regulatory sequences and transcribed sequences. The transcript (mRNA) contains a 5' untranslated region (5'UTR), the open reading frame (ORF), and the 3' untranslated region (3'UTR).

KEY TERMS:

Central Dogma
transcription
translation
reverse transcription
genotype
phenotype
genes
alleles
genetics
ORF
protein coding

non-protein coding
RNA encoding
Structural DNA
rRNA
tRNA
snRNA
snoRNA
miRNA
siRNA
piRNA

STUDY QUESTIONS:

1) Provide a definition of a gene that includes all types of genes (e.g. more than just a protein coding gene).

2) Is all DNA in a genome part of a gene? Does all DNA have a function? Explain.

3) Do all transcribed RNA molecules end up as mRNA transcripts?

4) What is the UTR on an mRNA?

5) Does a segment of DNA have to be transcribed in order to be a gene?

Recommended course:

Foundations of Molecular Genetics

GENETICS 270, Foundations of Molecular Genetics
★ 3 (fi 6) (either term, 3-1.5s-0)
Basic concepts on the organization of genetic material and its expression will be developed from experiments on bacteria and viruses.

Why study Bacterial Genetics?

- Model systems for higher organisms
- Bacteria are highly complex cells
- Bacteria are interesting
  - Essential to ecology of the Earth
  - Symbiotic with higher organisms
  - Play a role in disease

Prerequisites: BIOL 207
Class schedule: T/Th 11:00AM - 12:20PM
Offered: Either Term
Contact: Fall - Dr. J. Dennis, jon.dennis@ualberta.ca
                 Winter – Dr. T. Raivio, traivio@ualberta.ca
Calendar Link:
http://calendar.ualberta.ca/preview_course_nopop.php?catoid=6&coid=44621

CHAPTER 06 – PROKARYOTIC GENES: *E. COLI* LAC OPERON
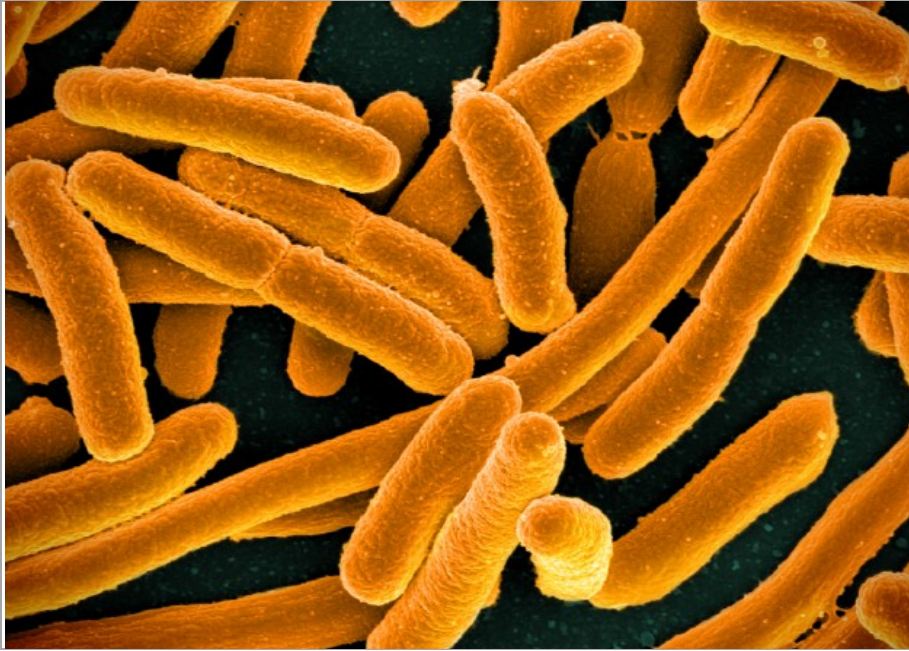
Figure 1.
Electron micrograph of growing *E. coli.* Some show the constriction at the location where daughter cells separate. The colouring is false.
(Flickr-NIAID-CC BY 2.0)

INTRODUCTION

With most organisms, every cell contains essentially the same genomic sequence. How then do cells develop and function differently from each other? The answer lies in the regulation of **gene expression**.   Only a subset of all the genes is expressed (i.e. 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 proteins. In the first stage, transcript abundance can be controlled by regulating the rate of transcription initiation and processing, as well as the degradation of transcripts. In many cases, higher abundance of a gene's transcripts is correlated with its increased expression. We will focus on **transcriptional regulation** in *E. coli* (**Figure 1**). Be aware, however, that cells also regulate the overall activity of genes in other ways.   For example, by controlling the rate of mRNA translation, processing, and degradation, as well as the post-translational modification of proteins and protein complexes.

1. THE *LAC* OPERON – A MODEL PROKARYOTE GENE

Early insights into mechanisms of transcriptional regulation came from studies of *E. coli* by Francois Jacob & Jacques Monod. In *E. coli,* and many other bacteria, genes encoding several different polypeptides may be located in a single transcription unit called an **operon**. The genes in an operon share the same transcriptional regulation, but are translated individually into separate polypeptides. Most prokaryote genes are not organized as operons, but are transcribed individually yielding single peptide units.

Eukaryotes do not group genes together as operons (an exception is *C. elegans* and a few other species).
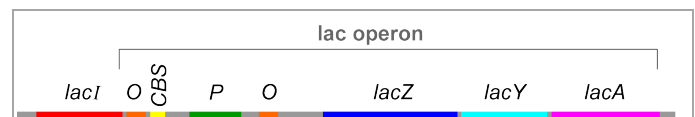
Figure 2.
Diagram of a segment of an E. coli chromosome containing the *lac* operon, as well as the *lacI* coding region. The various genes and cis-elements are not drawn to scale.
(Original-Deyholos-CC BY-NC 3.0)

1.1. BASIC LAC OPERON STRUCTURE

E. coli encounters many different sugars in its environment. These sugars, such as **lactose** and **glucose**, require different enzymes for their metabolism. Three of the enzymes for lactose metabolism are grouped in the *lac* **operon**: *lacZ*, *lacY*, and *lacA* (**Figure 2**). *LacZ* encodes an enzyme called **β-galactosidase**, which digests lactose into its two constituent sugars: glucose and galactose. *lacY* is a **permease** that helps to transfer lactose into the cell. Finally, *lacA* is a **trans-acetylase**; the relevance of which in lactose metabolism is not entirely clear. Transcription of the *lac* operon normally occurs only when lactose is available for it to digest. Presumably, this avoids wasting energy in the synthesis of enzymes for which no substrate is present. In the lac operon there is a single mRNA transcript that includes coding sequences for all three enzymes and is called a polycistronic mRNA. A cistron in this context is equivalent to a gene.

1.2. *CIS-* AND *TRANS-* REGULATORS

In addition to these three protein-coding genes, the *lac* operon contains several short DNA sequences that do not encode proteins, but instead act as binding sites for proteins involved in transcriptional regulation of the operon. In the *lac* operon, these sequences are called **P (promoter)**, **O (operator)**, and **CBS (CAP-binding site).** Collectively, sequence elements such as these are called *cis*-elements because they must be located on the same piece of DNA as the genes they regulate. On the other hand, intermolecular elements outside from the target DNA such as the proteins that bind to these *cis*-elements are called **trans-regulators** because (as diffusible molecules) they do not necessarily need to be encoded on the same piece of DNA as the genes they regulate.

2. NEGATIVE REGULATION – INDUCERS AND REPRESSORS

LacI encodes an allosterically regulated repressor
One of the major *trans*-regulators of the *lac* operon is encoded by *lacI*, a gene located just upstream from the lac operon (**Figure 2**). Four identical molecules of *lacI* proteins assemble together to form a **homotetramer** called a **repressor (Figure 3)**.

This repressor is **trans-acting** and binds to two cis-acting operator sequences adjacent to the promoter of the *lac* operon. Binding of the repressor prevents RNA polymerase from binding to the promoter (**Figure 2**, **Figure 4.**). Therefore, the operon is not transcribed when the operator sequence is occupied by a repressor.

Figure 3.
Structure of *lacI* homotetramer bound to DNA.
(Original-Deyholos- CC BY-NC 3.0)

2.1. THE REPRESSOR ALSO BINDS LACTOSE (ALLOLACTOSE)

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 allolactose. If lactose is present, **β-galactosidase** enzymes convert a few of the lactose molecules into allolactose. This allolactose can then be allosterically bound to the lacI protein. This alters the shape of the protein in a way that prevents it from binding to the operator. Therefore, in the presence of lactose (allolatose) the repressor doesn't bind the operator sequence and thus RNA polymerase is able to bind to the promoter and transcribe the *lac* operon. This leads to a moderate level of expression of the mRNA encoding the *lacZ*, *lacY*, and *lacA* genes. This kind of secondary molecule that binds to either activator or repressor and induces the production of specific enzyme is called an **inducer**. Also, 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 4**.
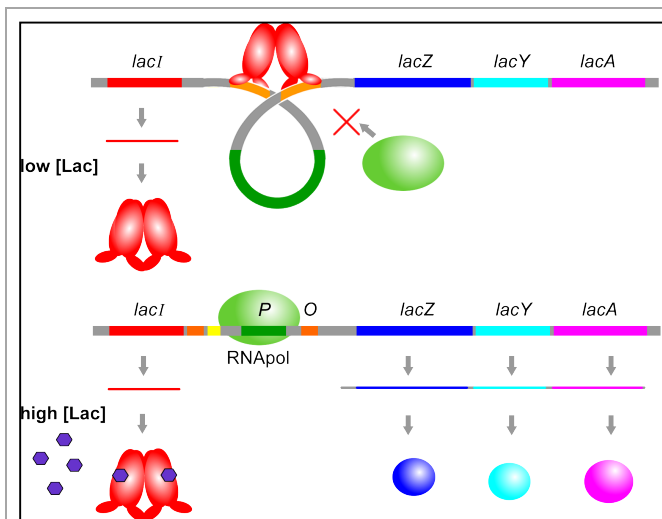
Figure 4.
When the concentration of lactose [Lac] is low, lacI tetramers bind to operator sequences (O), thereby blocking binding of RNApol (green) to the promoter (P). Alternatively, when [Lac] is high, lactose binds to lacI, preventing the repressor from binding to O, and allowing transcription by RNApol.
(Original-Deyholos-CC BY-NC 3.0)

Figure 5.
CAP, when bound to cAMP, helps RNApol to bind to the lac operon. cAMP is produced only when glucose [Glc] is low. (Original-Deyholos-CC BY-NC 3.0)

3. POSITIVE REGULATION – *CAP,* cAMP & POLYMERASE

A second aspect of *lac* operon regulation is conferred by a *trans*-acting factor called **cAMP binding protein** (**CAP**, **Figure 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, but upstream, to the promoter (P). When CAP is bound to at the CBS, RNA polymerase is better able to bind to the promoter and initiate transcription. Thus, the presence of cAMP ultimately leads to a further increase in *lac* operon transcription.

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 (inducer in this case): when glucose concentrations are low, an enzyme called **adenylate cyclase** is able to produce cAMP from ATP. Evidently, *E. coli* prefers glucose over lactose, and so expresses the *lac* operon at high levels only

when glucose is absent and lactose is present. This provides another layer of adaptive control of *lac* operon expression: only in the presence of lactose, and in the absence of glucose is the operon expressed at its highest levels.

4. THE USE OF MUTANTS TO STUDY THE *LAC* OPERON

4.1. SINGLE MUTANTS OF THE *LAC* OPERON
The *lac* operon and its regulators were first characterized by studying mutants of *E. coli* that exhibited various abnormalities in lactose metabolism. Mutations can occur in any of the lacZ, lacY, and lacA genes. Such mutations result in altered protein sequences, and cause non-functional products. These are mutations in the protein coding sequences (non-regulatory).

Other mutants can cause the lac operon to be expressed constitutively, meaning the operon was transcribed whether or not lactose was present in the medium. Remember that normally the operon is only transcribed if lactose is present. Such mutants are called **constitutive** mutants. Constitutive mutants are always on and are unregulated by inducers. These include *lacO* and *lacI* genes.

4.2. INDUCER MUTATIONS (*LACI* LOCUS)
The <u>*lacI* locus</u> has two types of mutations: I^- and I^S.

(1) One type of mutant allele of *lacI* (called I^-) either (a) prevents the production of a repressor polypeptide or (b) produces a polypeptide that cannot bind to the operator sequence. Therefore,

there is no repressor binding and transcription can occur without the presence of inducer (allolactose). This is also a constitutive expresser of the *lac* operon because absence of repressor binding

Figure 6.
Both *E. coli* strands with genotypes $I^+O^+Z^+Y^+A^+/F\ I^+O^-Z^+YA^+$ and $I^SO^+Z^+Y^+A^+/F\ I^-O^-Z^+Y^+A^+$ will induce all the *lac* genes because repressor cannot bind to the O^- sequence on the F-factor and cannot prevent transcription.
(Original-Locke-CC BY-NC 3.0)

Figure 7.
E. coli strain with genotype $I^+Z^+Y^+A^+/F\ I^-Z^+Y^+A^+$ will not produce *lacZ, lacY* and *lacA* products.
(Original-Locke-CC BY-NC 3.0)
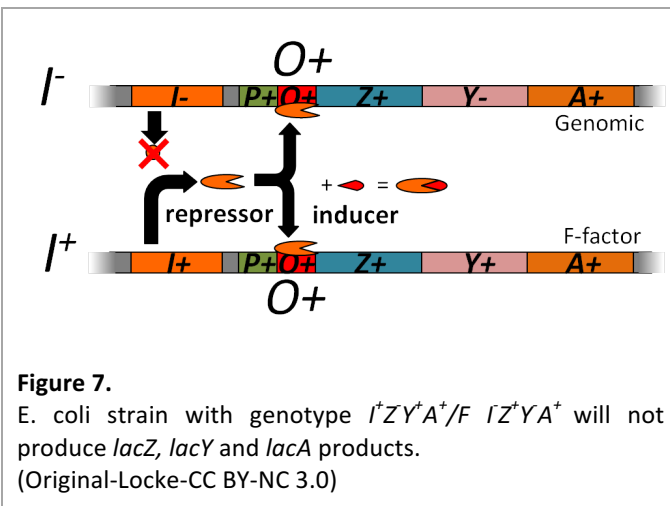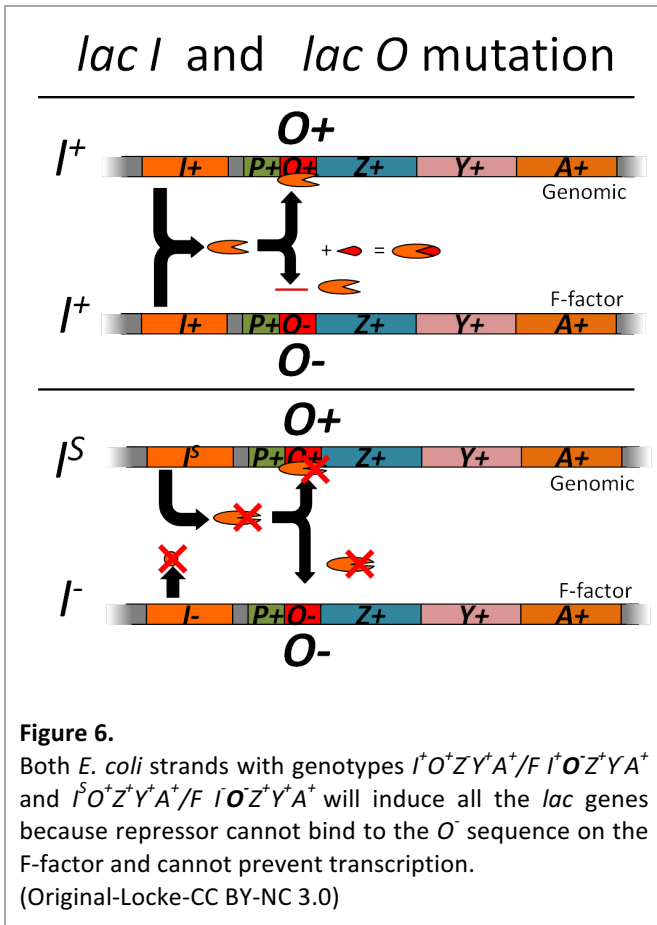
permits transcription.

Note that I^+ is dominant over I^-. For example, in *E. coli* strain with $I^+Z^-Y^+A^+/F\ I^-Z^+Y^-A^+,$ the *lac* genes will not be transcribed because the I^+ allele will still produce functional repressors that bind to all operator sequences, preventing transcription. (**Figure 7**)

(2) The other type of mutant of *lacI* called I^S prevents the repressor polypeptide from binding allolactose, and thus will only bind to the operator and the *lacZ, lacY* and *lacA* genes would be non-inducible. This mutant constitutively represses the *lac* operon whether lactose is present or not. The *lac* operon is not expressed at all and this mutant is called a "super-suppressor". I^S is therefore dominant to both I^+ and I^- in trans. Therefore, *E. coli* strains with the genotypes 1) $I^SZ^+Y^+A^+/F\ I^+Z^+YA^+$ and 2) $I^SZ^+Y^+A^+/F\ I^-Z^+Y^+A^+,$ the *lac Z, lac Y* and *lac A* genes will not be inducible (**Figure 8**).

Figure 8.
E. coli strains with genotypes 1) $I^SZ^+Y^+A^+/F\ I^+Z^+YA^+$ and 2) $I^SZ^+Y^+A^+/F\ I^-Z^+Y^+A^+$ will not produce *lacZ, lacY* and *lacA* products. (Original-Locke-CC BY-NC 3.0)
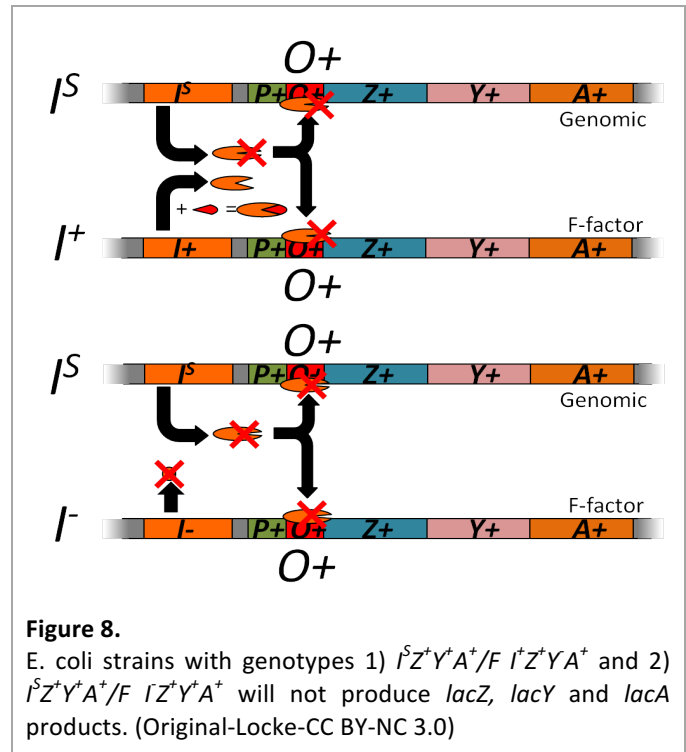
The repressor protein encoded by lac I gene has at least two independent functional domains. This is the reason why it can mutate independently to give two different types of mutants. **(Figure 9)**

I⁺   **Iˢ**   **I⁻**

Homo-tetramer

Figure 9.
Because the repressor encoded by *lacI* gene has independent domains, mutations can also occur independently.
(Original-Deyholos/Locke-CC BY-NC 3.0)
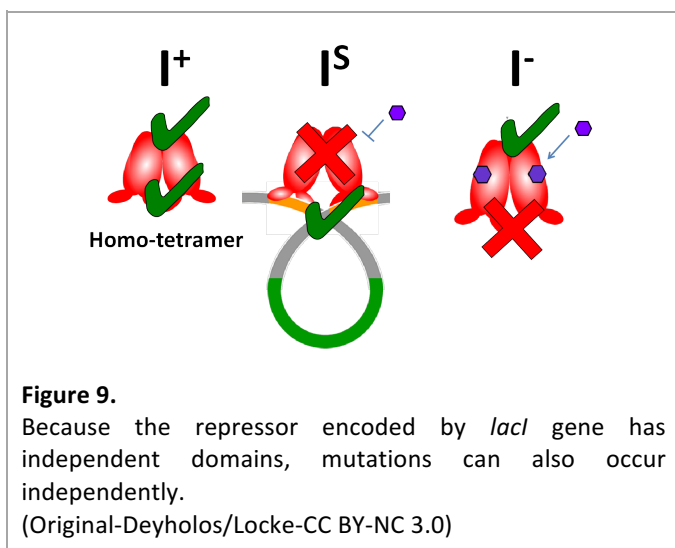
4.3. OPERATOR MUTATIONS

The operator locus (*lacO*) is an example of a mutation in an operator sequence. The base pair change 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. Note that this mutation is **cis dominant** (only affects the genes on the same chromosome) but not in trans (other DNA molecule).

Note that while O^c mutants will be constitutively expressed (not regulated by lactose), some may not be maximally expressed. Some alleles may partially bind the repressor and thus have some measure of inhibition. For example, a deletion of the operator sequence will result in maximal expression, while a single base pair change might only reduce binding slightly and affect the level of expression slightly (**Table 1**).

Table 1. Constitutively expressed O^c mutants may not be maximally expressed and have various levels of expression relative to wild type levels.

Level	Genotype	Explanation
100%	*lacI⁻ Oᶜ*	no repressor
10-20%	*lacI⁺ Oᶜ*	repressor fails to bind tightly
~≤1%	$P^+ \, O^c$, high glucose	basal transcription, constitutive
0%	P^- or Z^-	no transcription or no protein

4.4. THE *F*-FACTOR AND TWO LAC OPERONS IN A SINGLE CELL – PARTIAL DIPLOID IN *E. COLI*

More can be learned about the regulation of the lac operon when two different copies (each containing mutations) are present in one cell. This can be accomplished by using the **F-factor** to carry one copy, while the other is on the genomic *E. coli* chromosome. This results in a partial diploid *E. coli* cell, that is, one that contains two, independent copies (alleles) of the lac operon and *lacI*.

The *F*-factor or Fertility factor is an **episome**, which is capable of being either a free plasmid or integrated into the host bacterial chromosome. This switching is accomplished by IS (Insertion sequence) elements where unequal crossing over can recombine the *F*-factor and adjacent DNA sequences (genes) in and out of the host chromosome. If the *F* factor is present, then the strain is an F^+ strain. For example, the genotype of a host bacterium that has lac⁻ gene that is supplied with F factor containing lac⁺ can be written as *lac⁻ /F⁺lac⁺*. Researchers have used this genetic tool to create partial diploids (**merozygotes**) that allow them to test the regulation with 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*.

5. SUMMARY

In positive regulation, a **low** glucose level allows adenylate cyclase to produce cAMP from ATP, which binds to CAP protein. CAP protein can then bind to DNA and increase the level of *lac* operon transcription. A **high** glucose level halts adenylate cyclase from producing cAMP from ATP. Hence, cAMP will not bind to CAP protein and in turn, CAP will not bind to DNA and the level of transcription would be low.

In negative regulation, repressor protein acts to prevent transcription. Inducer binds to repressor to

alter conformation so it no longer binds to the operator sequence and transcription can take place. **High** levels of lactose (inducer) would allosterically inhibit repressor and therefore would not prevent transcription. **Low** levels of lactose would not cause inhibition to the repressor, so transcription would be prevented. Various forms of regulation in the lac operon are found in **Figure 10**.
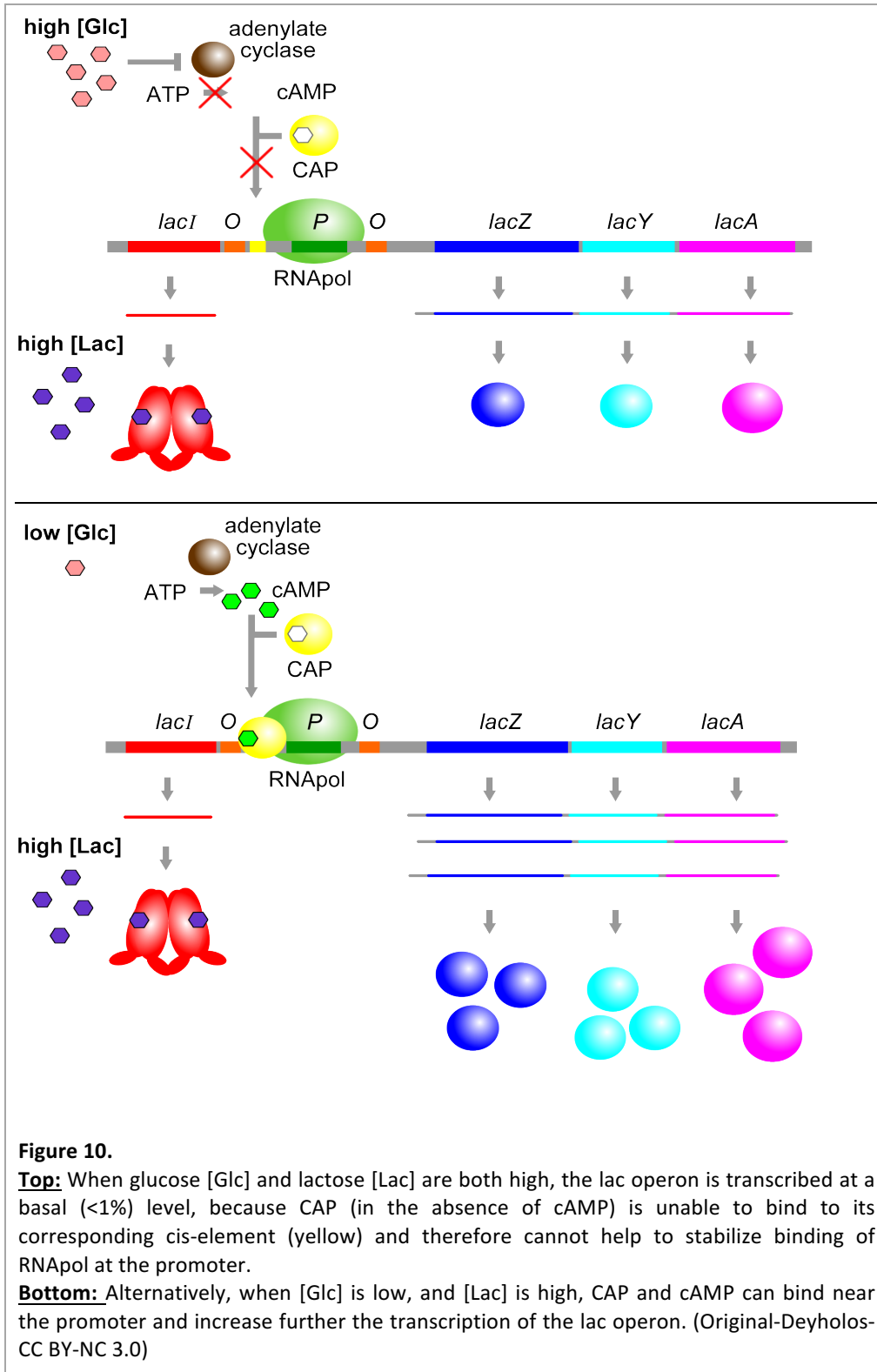
Figure 10.
Top: When glucose [Glc] and lactose [Lac] are both high, the lac operon is transcribed at a basal (<1%) level, because CAP (in the absence of cAMP) is unable to bind to its corresponding cis-element (yellow) and therefore cannot help to stabilize binding of RNApol at the promoter.
Bottom: Alternatively, when [Glc] is low, and [Lac] is high, CAP and cAMP can bind near the promoter and increase further the transcription of the lac operon. (Original-Deyholos-CC BY-NC 3.0)

_____

SUMMARY:

- Regulation of gene expression is essential to the normal development and efficient functioning of cells

- Gene expression may be regulated by many mechanisms, including those affecting transcript abundance, protein abundance, and post-translational modifications

- Regulation of transcript abundance may involve controlling the rate of initiation and elongation of transcription, as well as transcript splicing, stability, and turnover

- The rate of initiation of transcription is related to the presence of RNA polymerase and associated proteins at the promoter.

- RNApol may be blocked from the promoter by repressors, or may be recruited or stabilized at the promoter by other proteins including transcription factors

- The *lac* operon is a classic, fundamental paradigm demonstrating both positive and negative regulation through allosteric effects on *trans*-factors.

KEY TERMS:

gene expression

transcriptional regulation

operon

lactose

glucose

lac operon

lacZ

lacY

lacA

β-galactosidase

permease

trans-acetylase

P / promoter

O / operator

CBS

CAP-binding site

cis-elements

trans-regulators

lacI

homotetramer

repressor

inducer

allosteric

cAMP binding protein

CAP

CAP binding sequence

CBS

adenylate cyclase

constitutive

O^c / I^- / I^s

cis dominant

F-factor / episome

merozygotes

STUDY QUESTIONS:

1) 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^+, O^C, Z^+, Y^+ (no glucose, no lactose)

i) I^+, O^C, Z^+, Y^+ (no glucose, high lactose)

j) I^+, O^C, Z^+, Y^+ (high glucose, no lactose)

k) I^+, O^C, Z^+, Y^+ (high glucose, high lactose)

l) I^-, O^+, Z^+, Y^+ (no glucose, no lactose)

m) I^-, O^+, Z^+, Y^+ (no glucose, high lactose)

n) I^-, O^+, Z^+, Y^+ (high glucose, no lactose)

o) I^-, O^+, Z^+, Y^+ (high glucose, high lactose)

p) I^S, O^+, Z^+, Y^+ (no glucose, no lactose)

q) I^S, O^+, Z^+, Y^+ (no glucose, high lactose)

r) $I^S, O+, Z^+, Y^+$ (high glucose, no lactose)

s) $I^S, O+, Z^+, Y^+$ (high glucose, high lactose)

Use Answer Legend:
+++ Lots of β-galactosidase activity (100%)
++   Moderate β-galactosidase activity (10-20%)
+   Basal β-galactosidase activity (~≤1%)
-    No β-galactosidase activity (0%)

2) 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) $I^+, O^+, Z^+, Y^+ / I^+, O^C, Z^-, Y^-$ (high lactose)

b) $I^+, O^+, Z^+, Y^+ / I^+, O^C, Z^-, Y^-$ (no lactose)

c) $I^+, O^+, Z^-, Y^+ / I^+, O^C, Z^+, Y^+$ (high lactose)

d) $I^+, O^+, Z^-, Y^+ / I^+, O^C, Z^+, Y^+$ (no lactose)

e) $I^+, O+, Z^-, Y^+ / I^-, O^+, Z^+, Y^+$ (high lactose)

f) $I^+, O^+, Z^-, Y^+ / I^-, O^+, Z^+, Y^+$ (no lactose)

g) $I^-, O^+, Z^+, Y^+ / I^+, O^+, Z^-, Y^+$ (high lactose)

h) $I^-, O^+, Z^+, Y^+ / I^+, O^+, Z^-, Y^+$ (no lactose)

i) $I^+, O^C, Z^+, Y^+ / I^+, O^+, Z^-, Y^+$ (high lactose)

j) $I^+, O^C, Z^+, Y^+ / I^+, O^+, Z^-, Y^+$ (no lactose)

k) $I^+, O^+, Z^-, Y^+ / I^+, O^C, Z^+, Y^+$ (high lactose)

l) $I^+, O^+, Z^-, Y^+ / I^+, O^C, Z^+, Y^+$ (no lactose)

m) $I^+, O^+, Z^-, Y^+ / I^S, O^+, Z^+, Y^+$ (high lactose)

n) $I^+, O^+, Z-, Y^+ / I^S, O^+, Z^+, Y^+$ (no lactose)

o) $I^S, O^+, Z^+, Y^+ / I^+, O^+, Z^-, Y^+$ (high lactose)

p) $I^S, O^+, Z^+, Y^+ / I^+, O^+, Z^-, Y^+$ (no lactose)

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

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

5) What would be the effect of the following loss-of-function mutations on the expression of the lac operon?

a) loss-of-function of adenylate cyclase

b) loss of DNA binding ability of CAP

c) loss of cAMP binding ability of CAP

d) mutation of CAP binding site (CBS) *cis*-element so that CAP could not bind

CHAPTER 07 – EUKARYOTIC GENES: STRUCTURE
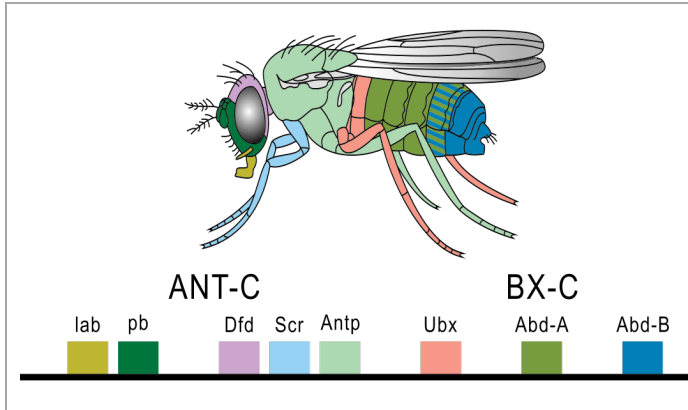
Figure 1.
Some genes are expressed in a segmental pattern and dictate the development of cells in that segment. This tissue specific patterning happens through the temporal and spatial regulation of these genes. (Wikipedia-PhiLiP-PD)

INTRODUCTION

While prokaryote protein-coding genes are relatively simple with a promoter driving a transcribed mRNA sequence (or a multiple protein coding mRNA in the case of an operon), the expression of a eukaryote protein-coding gene is much more complex. There are intron sequences, which are spliced out during processing, or they may be alternately spliced, and there are three levels of transcriptional regulation. All these make the typical eukaryote gene much larger than the typical prokaryote one and more complex.

1. THE EUKARYOTIC GENOME CONTAINS VARIOUS TYPES OF SEQUENCES

There are three main types of sequences in eukaryote genome, which are: (1) single copy genes, (2) multiple copy genes and (3) repeated sequences. **Single copy genes** have a single copy in the genome and include most protein-coding genes. **Multiple copy genes** have multiple copies in the genome and include rRNA- and tRNA-coding genes, and some protein coding genes. **Repeated sequences** can be either tandem repeats or interspersed repeats (**Figure 2**). Tandem repeats are followed directly after one another, whereas interspersed repeats are scattered randomly. Tandem repeats include (a) short centromeric-tandem arrays, which are sequence repeats at the centromere region and (b) VNTR (Variable Number Tandem Repeats). VNTR include microsatellites

(short tandem repeats) and mini-satellites (longer tandem repeats). Interspersed repeats include SINEs (Short Interspersed Elements), LINEs (Long interspersed elements).
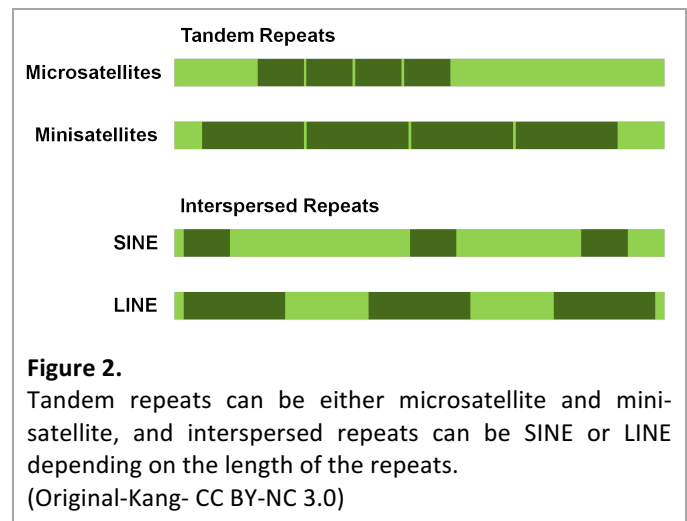
Figure 2.
Tandem repeats can be either microsatellite and mini-satellite, and interspersed repeats can be SINE or LINE depending on the length of the repeats.
(Original-Kang- CC BY-NC 3.0)

2. TRANSCRIPTS OF PROTEIN CODING GENES – PROCESSING

2.1. 5' CAP, POLY(A) TAIL
An mRNA is transcribed by RNA polymerase II using its complementary DNA strand as a template. As it is synthesized, it undergoes processing before transport to the cytoplasm. Here are the major steps of during transcription of eukaryote mRNA:

a) mRNA transcript is synthesized by RNA polymerase II.

b) While the mRNA is being synthesized, a 7-methyl guanosine cap is added to the 5' end by an enzyme called guanylyltransferase.

c) Transcription proceeds past the poly(A) addition site (AATAAA).

d) Endonuclease cleaves the mRNA strand 11-30 nucleotides downstream the AAUAAA signal sequence to create the 3' end.

e) At this 3' end, a poly(A) tail of 150-200 nucleotides is added by poly(A) polymerase.

f) This results in a mRNA **primary transcript**, which is not yet mature.

2.2. INTRON AND EXON

Primary transcripts undergo **RNA splicing and** are shortened by the removal of intervening sequences called **introns** before being transported to the cytoplasm. Only the sequences that are retained, called the **exons,** are joined together to make the **mature transcript** mRNA. This is done by a large multi-protein structure called the **spliceosome,** which also contains small nuclear ribonucleoproteins (snRNPs = small nuclear RNA).

Figure 3.
The steps of synthesizing a primary mRNA transcript. (Original-Locke-CC BY-NC 3.0)

Figure 4.
For each intron spliced out, there are three sites that are essential. They are the 5' donor site, branch point, and 3' acceptor site. The number below each nucleotide represents the percentage of that nucleotide at that site. (Original-Locke-CC BY-NC 3.0)
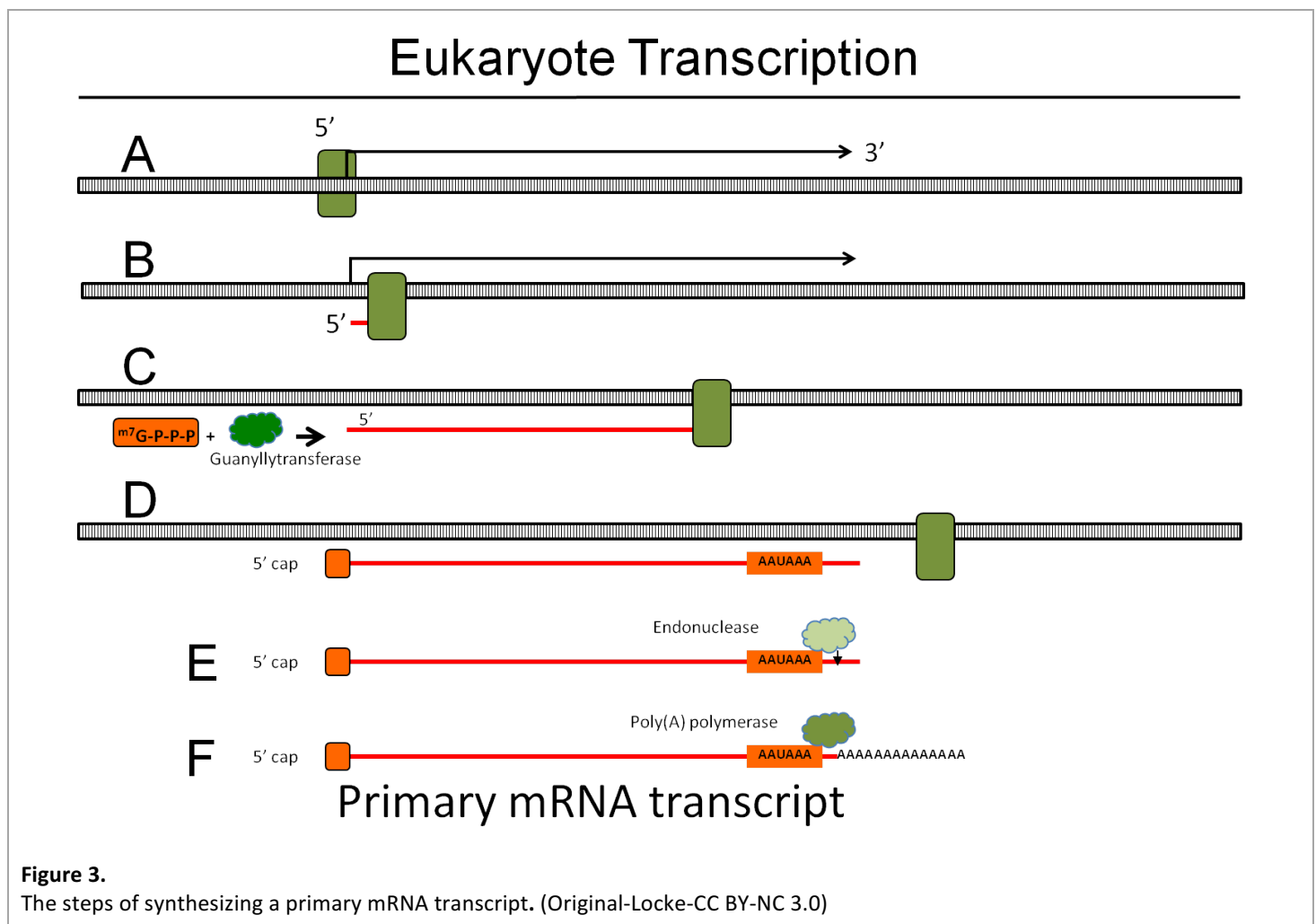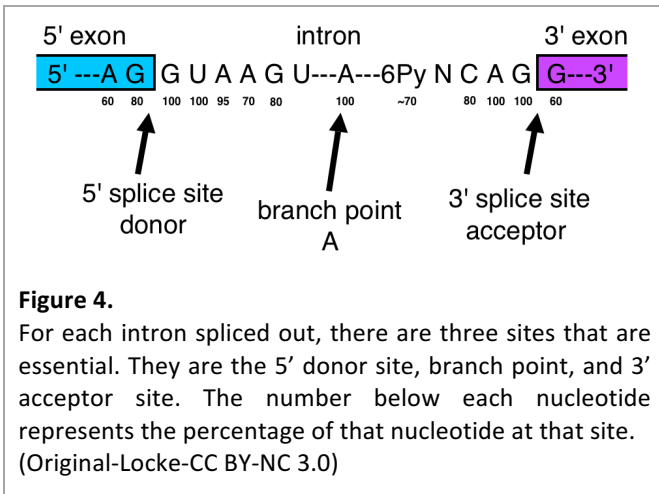
For each intron on the primary transcript RNA, there exists (1) 5' splice donor site, (2) branch point A, and (3) 3' splice acceptor site. (Note that the directionality in these names (ex. 5', 3') are referenced to the mRNA sequence). The snRNA of the spliceosome base pairs with the RNA sequences at the 5' splice donor site and cuts it. This cut 5' end of the intron "donates" or attaches to the branch point A via 2'-5' phosphodiester bond and forms a **lariat**. Next, a second cut is made at the 3' splice site acceptor (3' end of the intron) (**Figure 5.**) RNA ligase attaches the two exon ends together, and the intron sequence is are degraded leaving the mature mRNA.

Note: Almost all eukaryote genes have introns, but for some rare genes, like the Heat Shock Protein 70 (HSP70) gene, the primary transcript is the mature mRNA. Prokaryote genes do NOT have introns.

Figure 5.
(1) Spliceosome makes a cut at 5' splice donor site, and (2) the 5' end of the intron attaches to branch A point, and (3) the second cut is made at the 3' splice acceptor site, ultimately forming a mature RNA.
(Original-Kang-CC BY-NC 3.0)
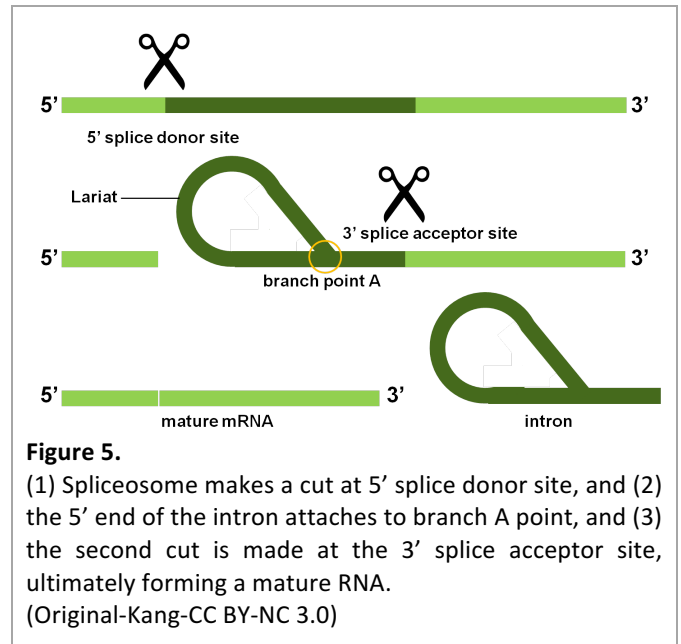
2.3. ALTERNATIVE SPLICING

Many genes have primary transcripts that are processed differently to produce more than one type of mature mRNA. This is called **alternative splicing**, which often results in the production of more than one type of protein product from the same gene (**Figure 6.**). Alternative splicing is another means of gene regulation, but it happens at a post-transcriptional level.

To further complicate this process, in some organisms it is even possible for exons from different gene transcripts to be ligated together through a process called **trans-splicing**. Although rare, an example comes from the worm, *C. elegans*, where an identical short leader sequence, the spliced leader (SL), is trans-spliced onto the 5'ends of multiple mRNAs.

With alternative splicing, it is possible for organisms with 25,000 genes (e.g. humans) to produce a much larger variation of polypeptides.

Figure 6.
Alternative splicing produces different combinations of exons, which result in different mRNA products, thus different proteins.
(Wikipedia- National Human Genome Research Institute-PD)
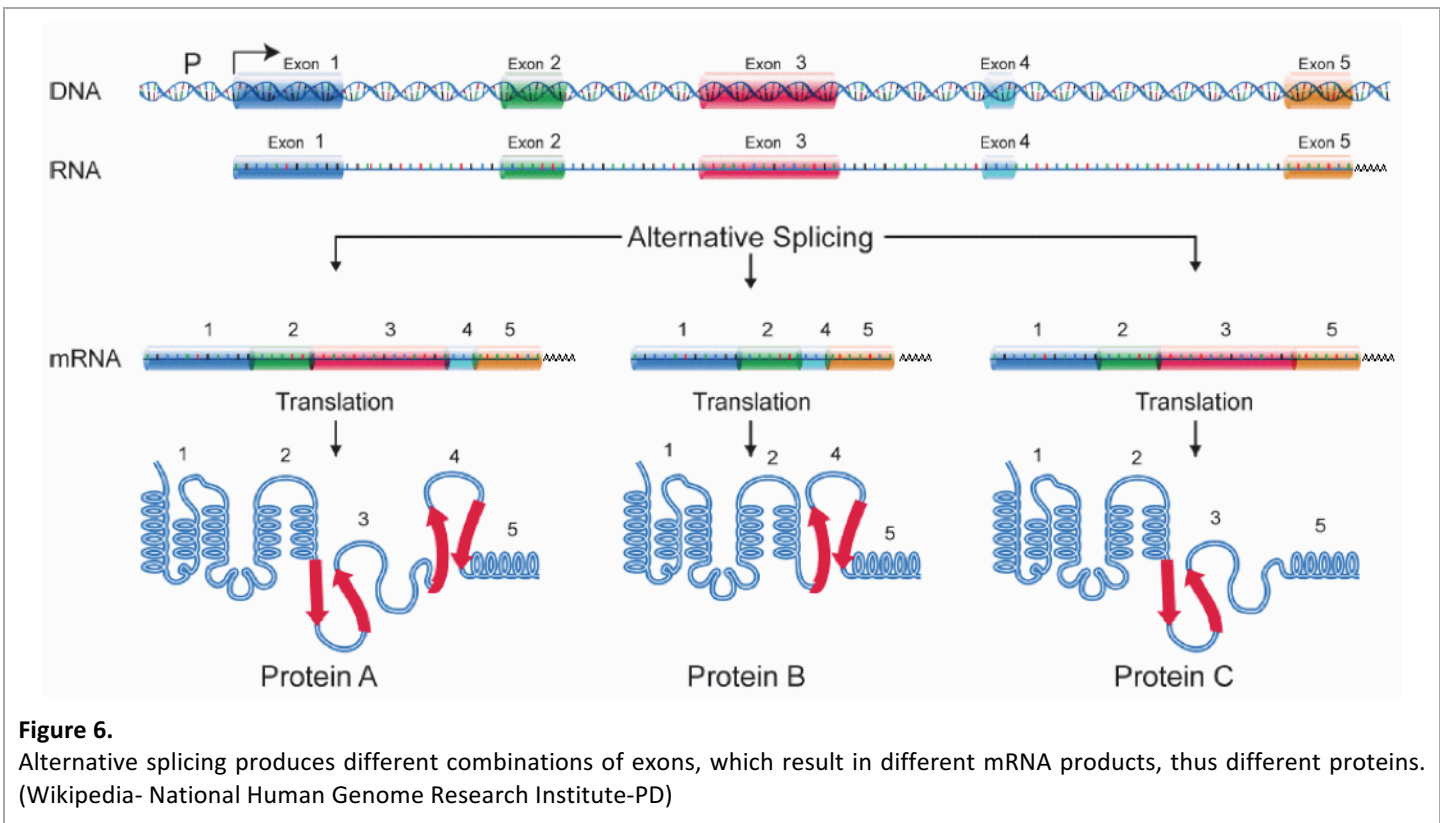
3. TRANSCRIPTION REGULATION – PROMOTERS, ENHANCERS/SILENCERS

3.1. PROXIMAL REGULATORY SEQUENCES.

As in prokaryotes the RNA polymerase binds to the DNA at the gene's **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, which includes **GC-rich boxes**, **CAAT boxes, and TATA boxes**. **Cis-elements** are intramolecular elements that exist and act within the same DNA molecule. However, ***trans*-elements** are intermolecular elements that are distinct molecules from their target DNA; it could be RNA or proteins. 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 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.

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 is a wide variety of different transcription factors and each recognizes a specific DNA sequence (enhancer elements) 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 7** and **Figure 8**).
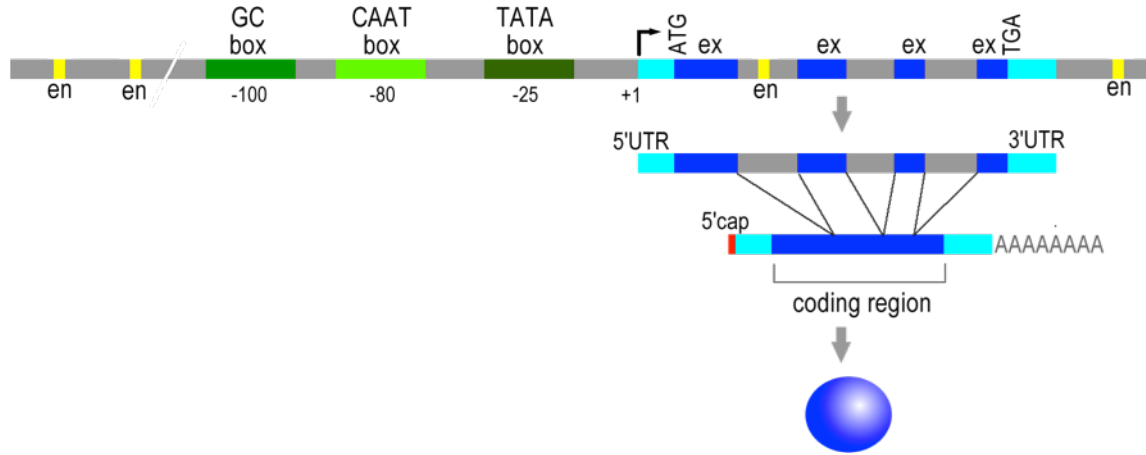
Figure 7.
Structure of a typical eukaryotic gene. RNA polymerase binding may involve one or more *cis*-elements within the proximal region of a promoter (green boxes). Enhancers (yellow boxes) may be located any distance upstream or downstream of the promoter and are also involved in regulating gene expression. The processing of a primary transcript to a mature mRNA is also shown. Note: not to scale. (Original-Deyholos- CC BY-NC 3.0)

3.3. EXAMPLE: GAL4-UAS SYSTEM FROM YEAST – A GENETIC TOOL

Yeasts use the Gal regulon to convert galactose to glucose-1-phosphate for glycolysis. Geneticists have taken advantage of a yeast distal enhancer sequence to make the **GAL4-UAS system**, a powerful technique for studying genes in other eukaryotes. It relies on two parts: a "**driver**" and a "**responder**" (**Figure 9.**). 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 and activate the *Gal* genes. 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. This system works is a variety of

eukaryotes, including humans. It has been especially well exploited in Drosophila where >10,000 differently expressing driver lines are available. These lines permit the tissue specific expression of any responder gene to examine its effect on development or cellular functions.
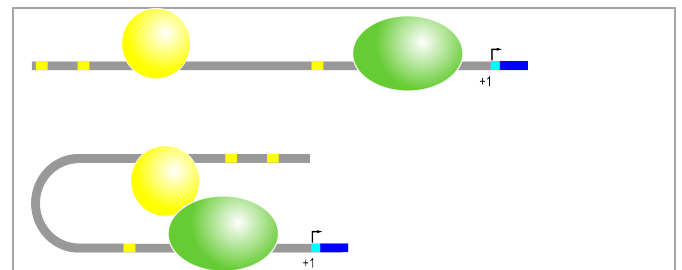
Figure 8.
A transcription factor (yellow) bound to an enhancer that is located far from a promoter. Because of the flexibility of the DNA molecule, the transcription factor and RNApol (green) are able to interact physically, even though the *cis*-elements to which they are bound are located far apart. In eukaryotic cells, RNApol is actually part of a large complex of proteins (not shown here) that assembles at the promoter. (Original-Deyholos- CC BY-NC 3.0)

Figure 9.
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. (Original-Locke- CC BY-NC 3.0
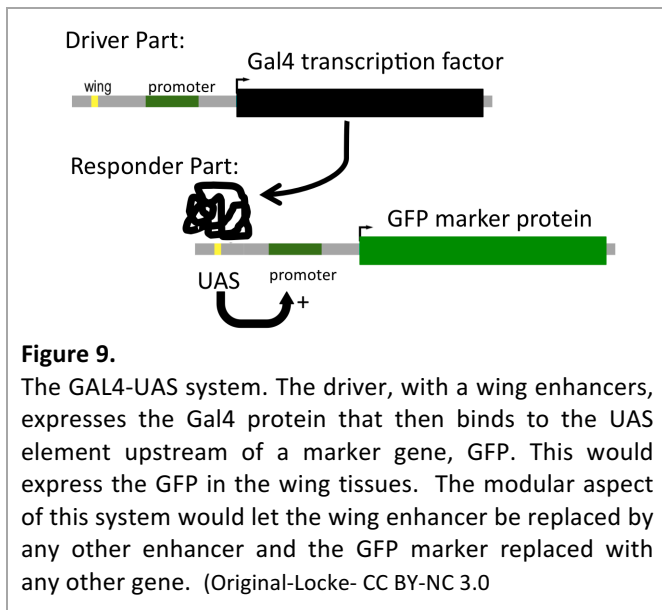
4. HIGHER ORDER CHROMATIN - ADDITIONAL LEVELS OF REGULATING TRANSCRIPTION

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 mechanisms: (1) large-scale changes in chromatin structure, and (2) modification of bases in the DNA sequence. These two are often inter-connected.

4.1. 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 where 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 (**Figure 10**).

Figure 10.
Acetylation of histone proteins is associated with more open chromatin configuration. Acetylation is a reversible process. (Original-Deyholos- CC BY-NC 3.0)
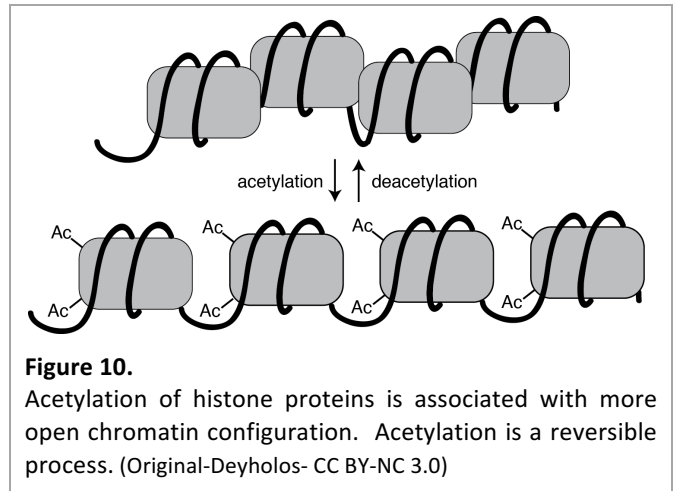
4.2. MODIFICATION OF DNA BASES

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 (**Figure 11**). Methylated cytosine within clusters of CpG sites is often associated with transcriptionally inactive DNA.

Figure 11.
methylation reaction shown here produces 5-methylcytosine (5mC). Methyl groups may also be removed by various processes.
(Original-Deyholos- CC BY-NC 3.0)
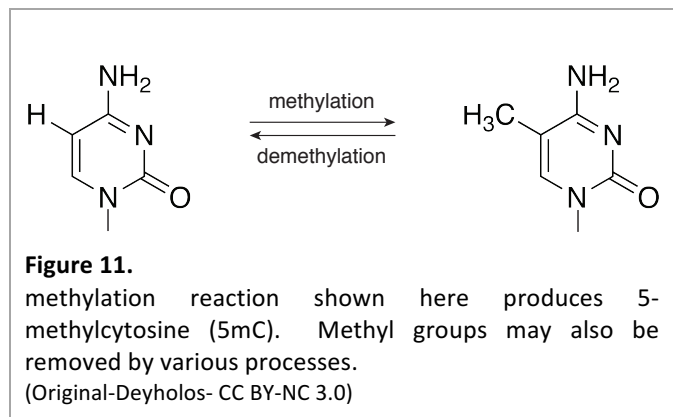
The modification of DNA and its associated proteins is enzymatically reversible (acetylation/deacetylation; methylation/demethylation) and thus a cyclical activity. Regulation of this provides another layer

through which eukaryotic cells control the transcription of specific genes.

5. EPIGENETICS

5.1. THE BACKGROUND OF EPIGENETICS

The word "**epigenetics**" has become popular in the last decade and its meaning has become confused. The term **epigenetics** describes any heritable change in phenotype that is not associated with a change in the chromosomal DNA sequence.

Originally it meant the processes through which the genes were expressed to give the phenotype; that is, the changes in gene expression that occur during normal development of multicellular organisms. This includes the change in transcriptional state of a DNA sequence (gene) via DNA or chromatin protein reversible modifications. Thus, DNA methylation and chromatin protein methylation, phosphorylation, and acetylation have been targeted as mechanisms for "heritable" changes in cells as they grow from a single cell (zygote) and differentiate to a multicellular organism. Here, dividing cells commit to differentiate into different tissues such as muscle, neuron, and fibroblast due to the genes that they express or silence. Some genes are irreversibly silenced, through epigenetic mechanisms, in some cell types, but not in others. All of this doesn't involve any change in DNA sequence.

Remember, these epigenetic effects are not permanent changes and thus are not selectable in an evolutionary context. However, mutations in the genes that regulate the epigenetic effect can be selected.

5.2. SOME HERITABLE INFORMATION CAN BE PASSED ON INDEPENDENT OF THE DNA SEQUENCE

More recently however, researchers have found many cases of environmentally induced changes in gene expression that can be passed on to the next generation – a potential multi-generational effect. These cases have also been called "**epigenetics**", and probably involve similar reversible changes to the DNA and chromatin proteins. These altered expression patterns represent the diversity of expression for a genome. This "extended"

phenotype, the ability to influence traits in the next generation, is a topic of current research and only some examples will be discussed here.

One example comes from the grandchildren of famine victims. They 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 to involve typical changes in the sequence of DNA. The term **epigenetics** is applied here since the apparently heritable change in phenotype is associated with something other than DNA sequence.

This change is inherited from one generation to the next and is thus **transgenerational**, for at least one generation. In developmental epigenetics, the expression state (developmentally differentiated state) is conserved only from one mitosis to the next, but is erased or reset at meiosis (the boundary of one generation to the next). The basis of at least some types of epigenetic inheritance appears to be replication of patterns of histone and DNA methylation that occurs in parallel with the replication of the primary DNA sequence. The permanence of this "epigenetic 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 across a few generations in some cases. It is becoming clear that epigenetics is an important part of biology.

5.3. IMPRINTING AND PARENT-OF-ORIGIN EFFECTS

For some genes, the allele inherited from the female parent is expressed differently than the allele that is inherited from the male parent. This is distinct from sex-linkage and is true even if both alleles are wild-type and autosomal. During gamete development (**gametogenesis**), each parent **imprints** epigenetic information on some genes that will affect the activity of the gene in the offspring. Imprinting does not change the DNA sequence, but does involve methylation of DNA and histones, and generally silences the expression of one of the parent's alleles. In humans, some

genes are expressed only from the paternal allele, and other genes are expressed only from the maternal allele. The imprinting marks are reprogrammed before the next generation of gametes are formed. Thus, although a male inherits epigenetic information from both his mother and father, this information is erased before sperm development, and he passes only one pattern of imprinting to both his sons and daughters. Most examples of imprinting come from placental mammals, and many imprinted genes control growth rate, such as *IGF2* (insulin-like growth factor 2).

Imprinting appears to explain many different **parent-of-origin** effects. For example, **Prader-Willi Syndrome** (**PWS**) and **Angelman Syndrome** (**AS**) are two phenotypically different conditions in humans that result from deletion of a specific region of chromosome 15, which contains several genes. Whether the deletion results in PWS or in AS depends on the parent-of-origin. If the deletion is inherited from the father, PWS results. Conversely, if the deletion is inherited from the mother, AS is the result. The gene(s) associated with PWS is maternally silenced by imprinting, therefore the deletion of its paternally-inherited allele results in a complete deficiency of a required protein. On the other hand, the paternal allele of the gene involved in AS is silenced by imprinting, so deletion of the maternal allele results in deficiency of the protein encoded by that gene.

5.4. TRANSGENERATIONAL INHERITANCE OF NUTRITIONAL INFLUENCES

Nutrition is one aspect of the environment that has been particularly well-studied from an epigenetic perspective in both mice and humans. People alive today who experienced the Dutch famine of 1944-1945 as fetuses have *IGF2* genes that are less methylated than their siblings. Methylation of *IGF2* (and birth rate) is also lower in children of mothers who do not take folic acid supplements as compared those who do. Furthermore, an individual's phenotype can be influenced by the nutrition of parents or even grandparents. This transgenerational inheritance of nutritional effects appears to involve epigenetic mechanisms.

The mouse *agouti* gene produces a signaling molecule that regulates pigment-producing cells and brain cells that affect feeding and body weight. Normally, *agouti* is silenced by methylation, and these mice are brown and have a normal weight. When *agouti* is demethylated by feeding certain chemicals or by mutating a gene that controls methylation, some mice become yellow and overweight, although their DNA sequence remains unchanged. Methylation of *agouti* and normal weight and pigmentation of offspring can be restored if their mothers are fed folic acid and other vitamins during pregnancy.

A study of an isolated Swedish village called Överkalix provides an example of transgenerational inheritance of nutritional factors. Detailed historical records allowed researchers to infer the nutritional status of villagers going back to 1890. The researchers then studied the health of two generations of these villagers' offspring, using medical records. A significant correlation was found between the mortality risk of grandsons and the food availability of their paternal grandfathers. This effect was not seen in the granddaughters. Furthermore, the nutrition of paternal grandmothers, or either of the maternal grandparents did not affect the health of the grandsons. It was therefore proposed that epigenetic information affecting health (specifically diabetes and heart disease) was passed from the grandfathers, to the grandsons, through the male line.

5.5. 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.

Figure 12.
A winter wheat crop (green) in early spring in the English countryside. (Flickr-Beardy Git- CC BY-NC-ND 2.0)

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 13**). 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.
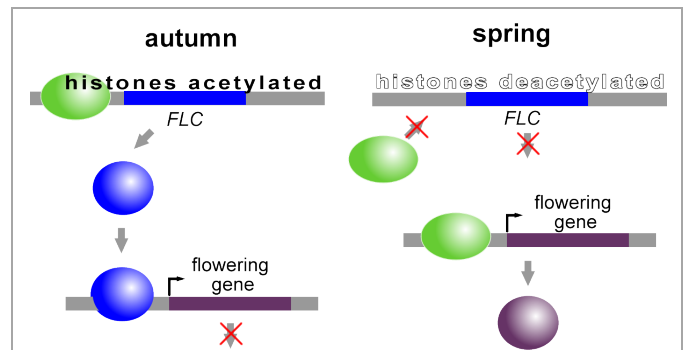
Figure 13.
In the autumn, histones associated with FLC are acetylated, allowing this repressor of flowering genes to be expressed. During winter, enzymes progressive deacetylate FLC, preventing it from being expressed, and therefore allowing flowering genes to respond to other signals that induce flowering.
(Original-Deyholos- CC BY-NC 3.0)

6. EXAMPLES OF CHROMATIN STRUCTURE: X-CHROMOSOME INACTIVATION

6.1. MAMMALIAN X-CHROMOSOME INACTIVATION – CALICO CATS, HUMAN EXAMPLE

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 (Xi)**. The other X chromosome, the **X-active (Xa)**, is unaffected. The Xi is replicated during S phase and transmitted during mitosis the same as any other chromosome but most of its genes are never expressed. 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.

This random inactivation of one X-chromosome leads to a commonly observed phenomenon in cats. A familiar X-linked gene is the Orange gene (*O*) in cats. The O^O allele encodes an enzyme that results in orange pigment for the hair. The O^B allele causes the hairs to be black. The phenotypes of various genotypes of cats are shown in **Figure 14**. 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 Xi chromosome carrying the O^B allele is inactivated. The O^O allele on the Xa is functional and orange pigments are made. In black hairs the reverse is true, the Xi chromosome with the O^O allele is inactive and the Xa chromosome with the O^B allele is active. Because the inactivation decision happens early during embryogenesis, the cells continue to divide to make large patches on the adult cat skin where one or the other X is inactivated.

Figure 14
Relationship between genotype and phenotype for an X-linked gene in cats. The O^O allele = orange while the O^B allele = black.
(Original-Harringtion- CC BY-NC 3.0)
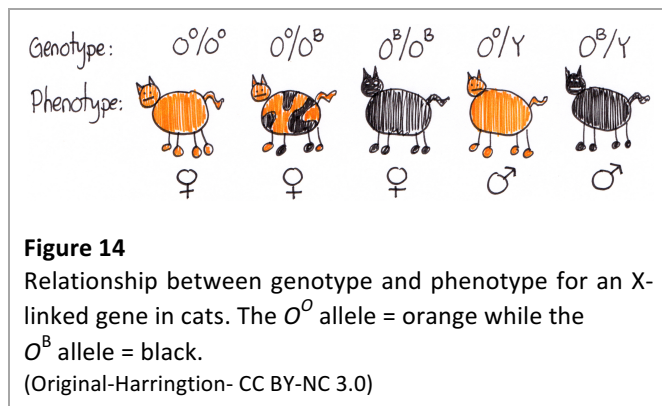
6.2. FACTOR VIII BLOOD CLOTTING PROTEINS

Another mammalian X-inactivation system 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, *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 15**). 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. Even though their liver cells are a genetic mosaic, this does not produce a visible mosaic phenotype.
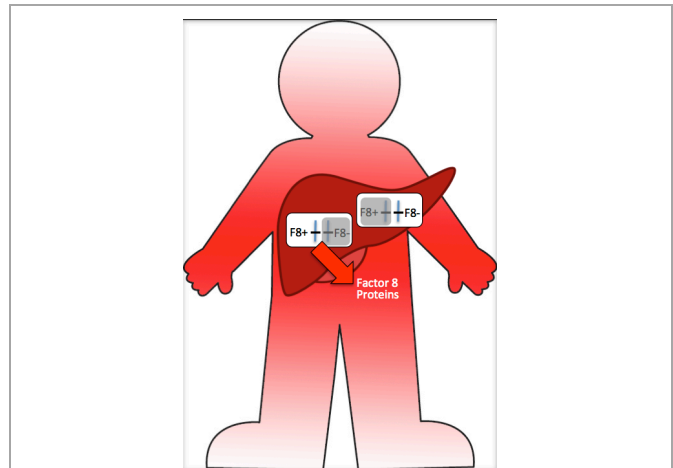
Figure 15.
This figure shows the two types of liver cells in females heterozygous for an F8 mutation. Because people with the F8+/F8- genotype have the same phenotype, normal blood clotting, as F8+/F8+ people the F8- mutation is classified as recessive. .
(Original-Harringtion/Locke- CC BY-NC 3.0)

_____

SUMMARY:

- In eukaryotic genome, there are single-copy and multi-copy genes and repeated sequences that are subdivided in various categories. Repeated sequences have tandem and interspersed repeats with varying lengths.

- The primary mRNA transcript undergoes some modification and processing before being exported to the cytoplasm: 5'cap, poly (A) tail, and splicing.

- Alternative splicing allows maximum number of products (proteins) with limited amount of resources (genes)

- In eukaryotes, enhancers bind to specific trans-factors, RNA polymerase and additional proteins to regulate transcriptional initiation in the promoter region.

- GAL4-UAS system in yeast uses driver (transcriptional activator/Gal4) and responder system (enhancer sequence / upstream activation sequence UAS) that can be integrated into other genes and be used as a biomarker.

- 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 and allow higher level of transcription

- During gamete development, each parent imprints epigenetic information on some genes that will affect the activity of the gene in the offspring.

- Heritable changes in phenotype that do not result from a change in DNA sequence are called epigenetics. Many epigenetic phenomena involve regulation of gene expression by chromatin modification and/or DNA methylation.

- When there are two X chromosomes in a female, X-inactivation compensates for overdosage; examples are calico cats and factor VIII Blood clotting protein in humans.

KEY TERMS:

Single copy genes	GC boxes / CAAT boxes / TATA	epigenetics
Multiple copy genes	boxes	transgenerational
Repeated sequences	Cis-elements	gametogenesis
primary transcript	Trans-elements	imprint
RNA splicing	transcription start site	parent-of-origin
introns	enhancer elements	Prader-Willi Syndrome (PWS)
exons	transcription factorsGAL4-UAS	Angleman Syndrome (AS)
mature transcript	Driver/responder	agouti
spliceosome	chromatin remodeling	winter annuals
lariat	acetylation/deacetylation	vernalization
alternative splicing,	methylation/demethylation	X-inactive (Xi) / X-active (Xa)
trans-splicing	CpG sites	Barr body

STUDY QUESTIONS:

1) List all the mechanisms that can be used to regulate gene expression in eukaryotes.
2) How are eukaryotic and prokaryotic gene regulation systems similar?
   How are they different?
3) 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 08 – EUKARYOTE GENES: HUMAN BETA-GLOBIN GENES
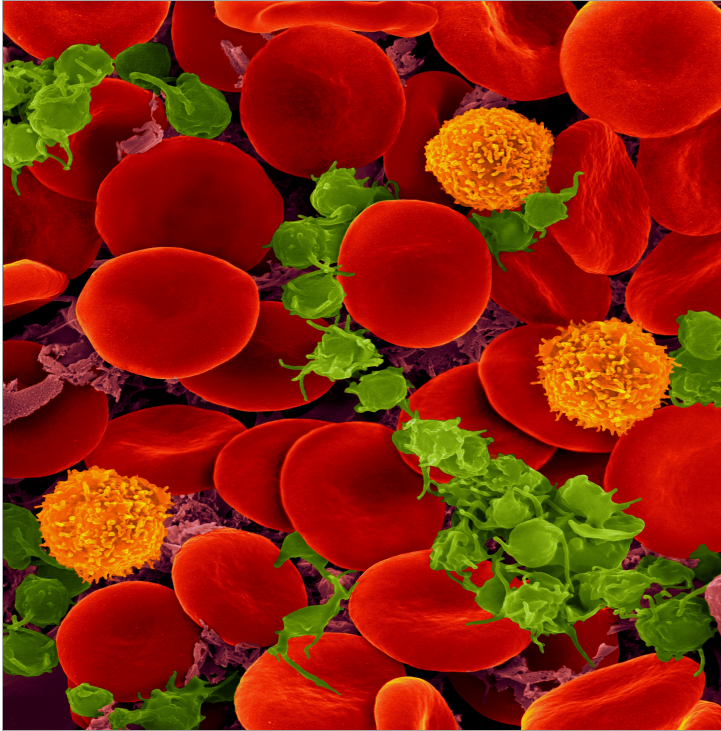
Figure 1.
Image of red blood cells (red), platelets (green) and T cells (orange) with a scanning electron microscope. The cells in this image are artificial coloured.
(Flickr-ZEISS Microscopy-CC BY-NC-ND 2.0)

INTRODUCTION

Genes encoding the globin polypeptides (component of red blood cell - **Figure 2**) are found in most species of higher eukaryotes. The human β-globin genes can be used as an example of a classic eukaryotic gene because they have most of the features needed for understanding basic eukaryote gene structure, expression, and regulation. There are several β-like genes in each cluster and each gene is expressed as part of a developmental program. Each gene's polypeptide product functions as part of a multimer protein.

1.   BETA-GLOBIN – PROTEIN AND GENE STRUCTURE, CLUSTERS, PSEUDO-GENES

1.1.   HEMOGLOBIN IS A HETEROTETRAMER WITH TWO 2 α-GLOBIN AND 2 β-GLOBIN POLYPEPTIDES

Hemoglobin is an oxygen-transporting or storing protein. This protein, or something similar, is found in most animals and many plants. In higher vertebrates, hemoglobin is a component of red blood cells (erythrocytes) that transports O_2 from the environment to the body cells.

The hemoglobin protein usually exists as a heterotetramer of four non-covalently bound hemoglobin polypeptides (**Figure 2**). In adults, each hemoglobin protein consists of a dimer of **α-globin** and another dimer of **β-globin** polypeptides, each with a bound heme molecule. Together these four form the hemoglobin tetramer = 2 **α-globin-like** + 2 **β-globin-like polypeptides.** The heme molecule is made through an independent metabolic pathway and then bound to the globin polypeptide through the iron (Fe) ion, which is covalent attachment of Fe as a **post-translational modification** to the polypeptide.

Figure 2.
A heterotetramer of human hemoglobin, type α2β2. The α chains are labeled red, and the β chains are labeled blue. The heme groups are green.
(Wikipedia-Zephyris- CC BY-SA 3.0)
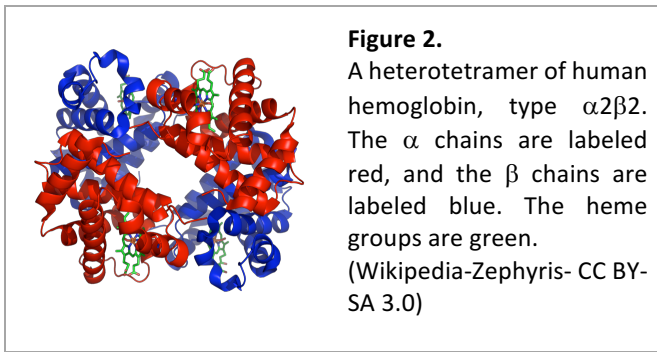
1.2. HUMAN GLOBIN GENES ARE FOUND IN TWO CLUSTERS

α- and β-like genes form a family of genes in most vertebrates. The positions of the introns/exons of those two genes are very similar. These genes have 3 exons and 2 introns in total. Comparison of other α and β globin genes from other species shows the intron positions are conserved. Both polypeptides are similar, too. Human α-globin is 141 amino acids long and β is 146 amino acids long.

Note that there are other proteins that are similar to α-globin such as **zeta (ζ)-globin** protein. There are also β-globin-like proteins such as **epsilon (ε), gamma (γ) and delta (δ) globin** proteins. In a tetramer, the zeta can take the place of the alpha, while the epsilon, gamma, or delta can take the place of the beta – the ratio is always 2 α-like to 2 β-like polypeptides. Each of these types of proteins (α- vs β-) are encoded by different genes. Each cluster of genes is referred to as a "**locus**" – the α-globin locus and the β-globin locus. Each set of genes is located as a single gene cluster (**Figure 3**)

In humans, the beta-globin cluster is located at chromosome 11 and includes 5 genes; epsilon, G-gamma, A-gamma, delta and beta genes. The alpha-globin cluster is located at chromosome 16 and includes 3 genes; zeta, alpha-1 and alpha-2 genes. Other vertebrates have similar clusters of α- and β-like genes. These clusters have arisen through a series of duplications of an ancestral globin gene. In general, **gene duplication** events can occur through rare errors in normal processes such as DNA replication, meiosis (crossing over), or transposition. Through time, 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 the current globin genes have evolved and are expressed at different phases of development to produce proteins optimized for the prenatal/postnatal environment.

1.3. PSEUDO-GENES

Of course, not all mutations are beneficial: some mutations can lead to inactivation of one or more of the products of a gene duplication event. This can result in what is called a **pseudogene**. Examples of pseudogenes (ψ) are also found within the globin clusters. Pseudogenes have mutations that prevent them from being expressed. They frequently lack the *cis*-acting regulatory elements (promoter and enhancer sequences) that are required for expression, but still retain similarity in the protein coding sequences, which permits their identification

Figure 3.
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 theta, μ mu) have also been described by some researchers, but are not shown here.
(Wikipedia –Modified by Kang- CC BY-NC 3.0)
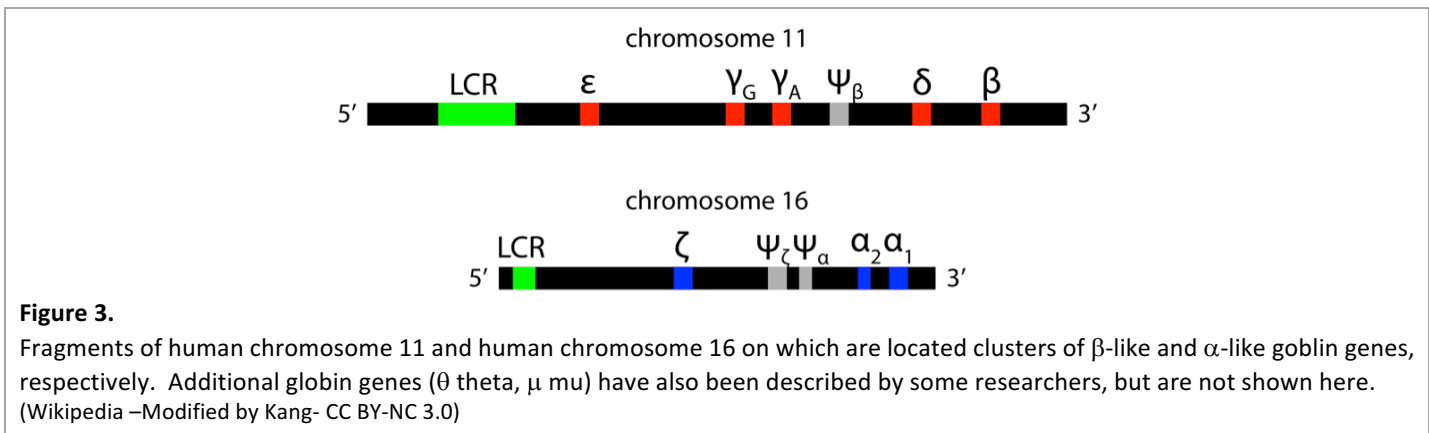
as a globin gene. The ψ (psi) symbol represents the designation as a pseudogene. The globin genes provide an example of how gene duplication and mutation, followed by selection, allows genes to evolve specialized expression patterns and functions. In general, many genes have evolved as **gene families** in this way, although they are not always clustered together as are the globins.

2. HEMOGLOBIN EXPRESSION CHANGES DURING DEVELOPMENT IN HUMANS.

In humans the composition of the globin tetramer changes during development (**Figure 6**). There are 3 distinct time periods that differ in globin gene expression. In embryos, $\zeta_2\varepsilon_2$ (zeta, epsilon) is the most abundant type, which means the globin tetramer contain two copies of each of zeta and epsilon proteins, which are similar but slightly different from each other. Next, in fetuses, $\alpha_2\gamma_2$ (alpha, gamma) is most abundant form. From early

childhood onward, most tetramers are of the type $\alpha_2\beta_2$ (alpha, beta). A small amount of adult hemoglobin is $\alpha_2\delta_2$ (alpha, delta), which has δ globin instead of the more common β globin. 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 $\alpha_2\gamma_2$ has a higher oxygen affinity than adult hemoglobin, allowing the fetus to more effectively extract oxygen from maternal blood, which is $\alpha_2\beta_2$. The specialized γ globin genes that are characteristic of fetal hemoglobin are found only in placental mammals.

Note that in humans the developmental changes in gene expression from zeta to alpha and from epsilon to gamma to beta parallel the location along the chromosome. This correlation is found in other species with clusters of globin genes, although not as rigidly.

Figure 4.
Expression of globin genes during prenatal and postnatal development in humans. The organs where the globin genes are primarily expressed at each developmental stage are also indicated on top.
Data: Wood, W.G. 1976 Br. Med. Bull. 32, 282
Original: (Wkipedia-Furfur- CC BY-SA 3.0)
Derivative work/Translation: (Wikipedia-Leonid2- CC BY-SA 3.0)
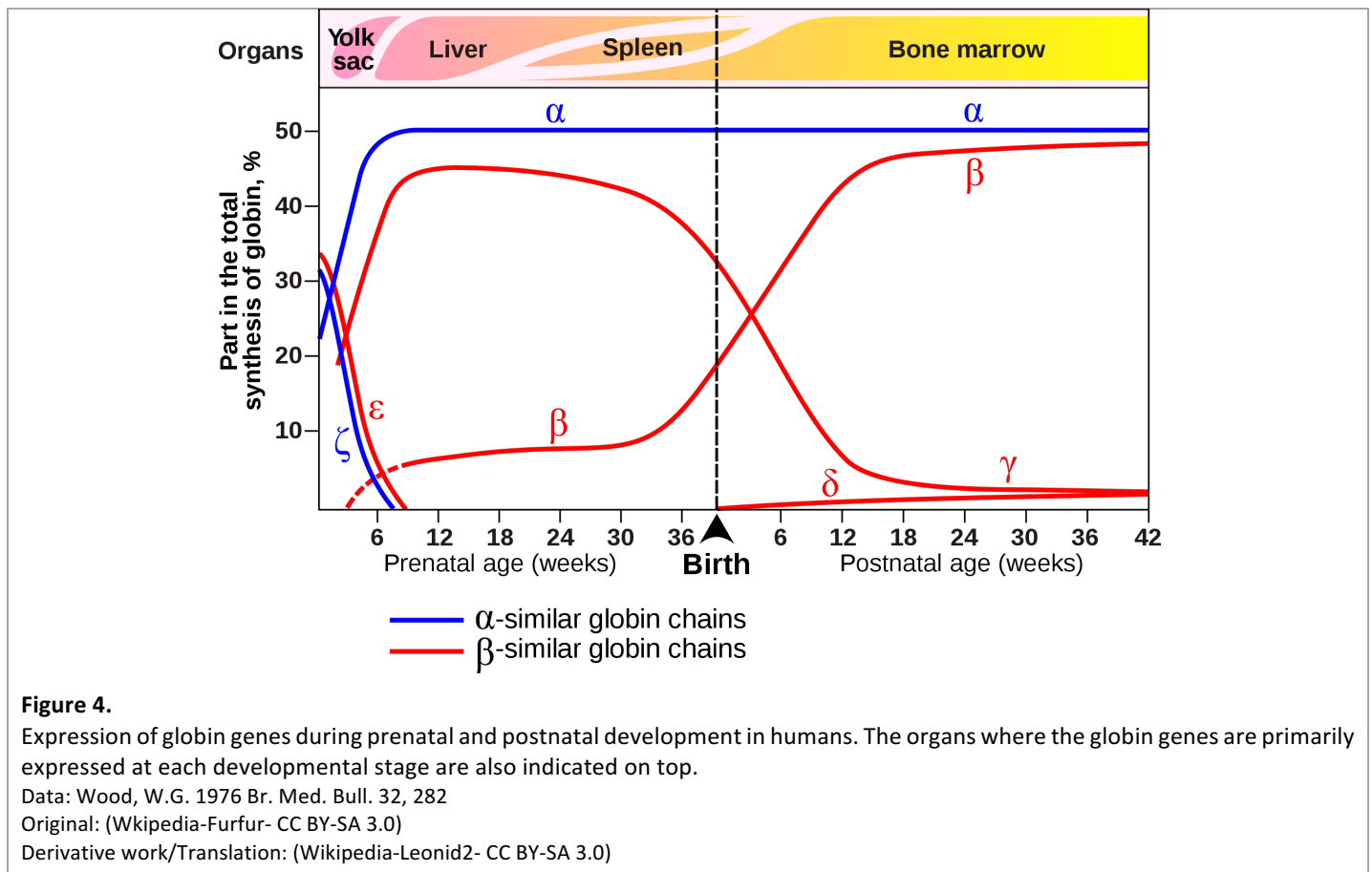
3. LOCUS CONTROL REGION (LCR) – ANOTHER LEVEL OF REGULATION

The transcription of the globin genes is controlled at multiple levels. First, the promoter dictates the mRNA start position and provides a basal level of transcription. Second, there are enhancer/silencer elements that act on the promoter to determine tissue- and temporal-specific transcription. Third, the β-globin locus is also regulated by higher order chromatin structure changes.

At the chromatin level, there is a region upstream from the cluster of β-globin genes that regulates and controls the expression of all the genes in the cluster. It is called a **locus control region (LCR).** LCR region can transcriptionally activate distal globin genes and its exact mechanism hasn't been fully identified. There are many models proposed to explain how LCR works, and one of them is by forming a loop and interacting with transcription factors to form a complex called an **enhancesome**. During development, this complex associates with other transcription factors in sequential manner. The LCR region contains sequences that regulate the conformation of chromatin for all the adjacent globin genes (**Figure 5**).

The change in conformation is recognized through differences in this region's sensitivity to added nucleases. The LCR contains 4 nuclease hypersensitive sites (HS4, HS3, HS2, and HS1) that influences can be detected when isolated nuclei are treated with added nucleases. If the DNA is in a "closed" conformation the nucleases cannot cleave the DNA. If the DNA is in an "open" conformation

the nucleases can cleave the DNA (**Figure 6.**). This nuclease sensitivity assay is done in vitro and reflects the open/closed chromatin found *in vivo*. Also, these hypersensitive sites aid in the recruitment of molecular factors that are needed for transcription; each hypersensitive region independently affects, in an additive fashion, the activation of gene expression. Note that LCR regions do not necessarily exist at a single site like the beta globin gene; in other cases, LCR can be found in multiple sites.

This kind of regional change in chromatin conformation permits the various globin genes to be regulated and expressed by their own individual promoters, enhancers, as well as having a dynamic chromatin regulation, too.

The LCR-dependent chromatin changes develop in erythroid precursor cells long before any globin gene is expressed. It begins with the opening (become nuclease sensitive) of the sites 5' to the globin genes first, then sites that are 3' open later on. Thus the change in chromatin structure at this locus is a developmental planned series of events.

A deletion mutation that removes the LCR region, prevents the 5' site from forming and also prevents the subsequent 3' site formation, thus the 5' site is needed for the 3' site to gain nuclease sensitivity. Note that LCR sites must be present in order to activate the expression globin genes; non red blood cell precursors do not open this region of chromatin so the β-globin genes are not expressed.
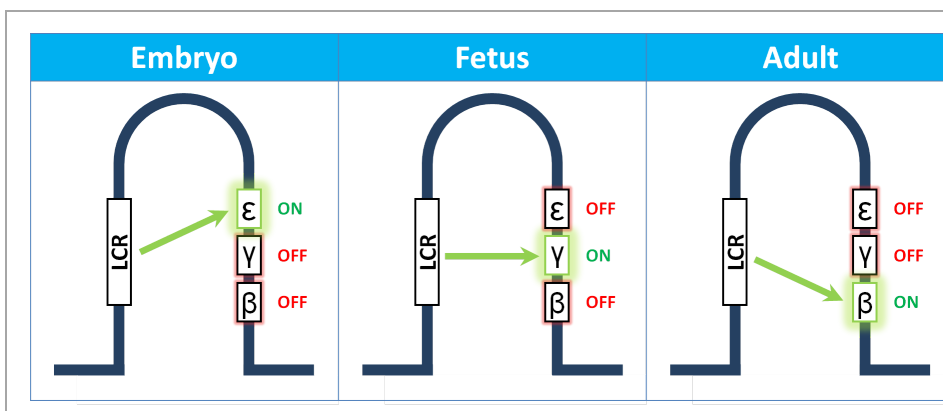
Figure 5.
Diagram showing the role of the LCR in development. The LCR regulates which of the globin genes in the cluster is expressed at different times during development of the red blood cell.
(Original-Kang-CC BY-NC 3.0)

Nuclease insensitive

Nuclease

Nuclease hypersensitive site

Figure 6.
The upper part of this diagram represents the nuclease insensitive chromatin form where the nucleases (scissors) cannot access or cleave the DNA. The lower part represents the nuclease hypersensitive chromatin form where the nucleases can access and cleave the DNA. The sensitivity/insensitivity of the DNA to nucleases is just a method to reveal the difference in chromatin structure along the DNA molecule. Some areas are "open" (sensitive) while others are "closed" (insensitive).
(Original-Locke- CC BY-NC 3.0)

.

4. ADDITIONAL INFORMATION-MYOGLOBIN

Globin gene expression is tissue-specific; α- and β-like globin genes are expressed in red blood cell precursors. A different globin-like gene, **myoglobin** is expressed and found only in muscle cells. Just like hemoglobin, myoglobin is an oxygen-binding protein and acts as temporary storage in muscle cells. The main difference between myoglobin and hemoglobin is that hemoglobin is mainly found in the blood stream, but myoglobin is only found in skeletal and heart muscles to provide oxygen for metabolically active cells. Therefore, myoglobin can act as a biomarker for detecting muscle injuries as high concentrations of myoglobin in the bloodstream indicate internal bleeding from the muscles. Also, myoglobin has higher affinity for oxygen than hemoglobin; this feature allows myoglobin to take up oxygen from hemaglobin.

_____

SUMMARY:

- Hemoglobin is a tetramer that transports and stores oxygen; it is composed of 2 α-globin-like + 2 β-globin-like polypeptides. α-globin can be replaced by zeta (ζ) globin, and beta-globin can be replaced by epsilon (ε), gamma (γ) and delta (δ) globin proteins. Each protein has different affinity for oxygen.

- Pseudogenes are version of a normal gene that frequently lacks the cis-acting regulatory elements but still possess the protein coding sequences.

- Expression can be tissue specific; globin is expressed in red blood cells and myoglobin is expressed only in muscle cells.

- Expression is developmental specific; each globin gene is expressed at a limited time during development.

- Expression is coordinately controlled; alpha and beta genes are expressed to the same level so that there is a 1:1 ratio of globin polypeptides.

- Promoter, enhancer/silencer elements, and locus control region regulates gene expression.

- Hemoglobin can be found in the blood stream, and myoglobin is only expressed in the muscles. Myoglobin acts as temporary oxygen storage for metabolically active cells and has higher affinity for oxygen than hemoglobin.

KEY TERMS:

hemoglobin/heme/ α , β globin

post-translational modification

zeta (ζ)-globin

epsilon (ε)/gamma (γ)/delta (δ) globin

locus

Pseudogene

Gene duplication

gene families

ζ2ε2 / α2γ2 /α2β2/ α2δ2

locus control region (LCR)

myoglobin

STUDY QUESTIONS:

1) The various α- and β-globin genes are expressed at various times during development (embryo, fetus, adult). This might reflect their various physiological roles during development. What might those roles be, and how might this be tested?

2) Why might the α-globin and β-globin genes have the same intron/exon structure?

3) Draw a simple cartoon showing the organization of the globin polypeptides in the functional hemoglobin molecule.

4) **Figure 3.** shows the organization of the human globin genes on chromosomes 11 and 16. The figure lacks a scale bar. Search the internet for a similar figure and show the length of 10 kbp on your figure.

5) Some adults have a condition called hereditary persistence of fetal hemoglobin (HPFE). What causes it? Does it affect their health?

Recommended course:

Organization of Simple Genomes

GENET 301, Organization of Simple Genomes

★ 3 (fi 6) (first term, 3-0-0)

Two models of simple genomes will be examined. One model will focus on the function and transmission of mitochondrial DNA, the evolution of mitochondria, and the role of mitochondria in human disease and aging. The other model will focus on the application of genomics, molecular biology, and cell biology to understand chromosome structure, DNA replication, cell division, and cell-cell communication in yeast.
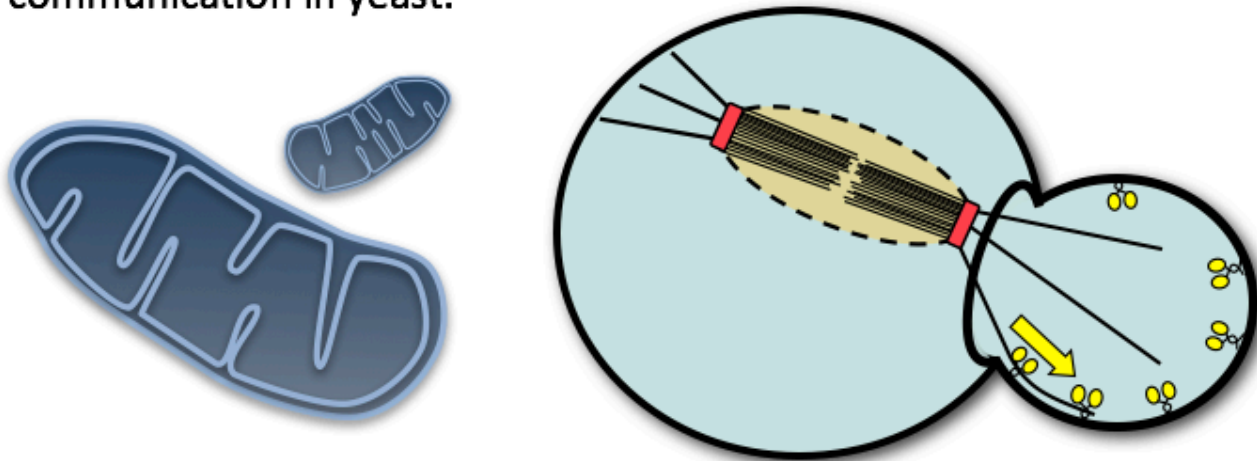
Prerequisites: GENET 270

Class schedule: MWF 9:00-9:50 AM

Offered: Fall Term

Contact: Dr. M. Srayko, srayko@ualberta.ca

Calendar Link:
http://calendar.ualberta.ca/preview_course_nopop.php?catoid=6&coid=44622

CHAPTER 09 – EUKARYOTIC GENES:

THE HUMAN LACTASE (LCT) GENE

Figure 1.
Unlike most dairy products, these lack the sugar lactose. (Flickr-USDAgov-CC BY 2.0)

INTRODUCTION

Young mammals get nourishment from their mother's milk. Human milk for example contains 4% fat, 1% protein, and 7% carbohydrates. There are 30+ different types of carbohydrates but the most abundant is the disaccharide **lactose** ("milk sugar"). How does the infant digest this lactose? When the lactose reaches the start of the small intestine it encounters an enzyme called **Lactase**. As shown in **Figure 2** Lactase is a membrane protein on the surface of intestinal epithelial cells. Lactase performs a hydrolysis reaction, which separates the disaccharide into two monosaccharides, galactose and glucose. These are then imported into the cells by a second membrane protein named the Sodium-Glucose Transporter 1 (**SGLT1**). It transports some monosaccharides (glucose and galactose) but not others (fructose for example). As its name suggests, it is powered by a sodium gradient, there are more sodium ions outside the cell than inside. Each time the protein allows two sodium ions to enter one monosaccharide can be imported. Once inside the cell some of the sugars are consumed by the cell itself. Most however are transported a second time at the other side of the cell where they enter the blood. Once in the circulatory system the sugars will travel to the other cells of the body.
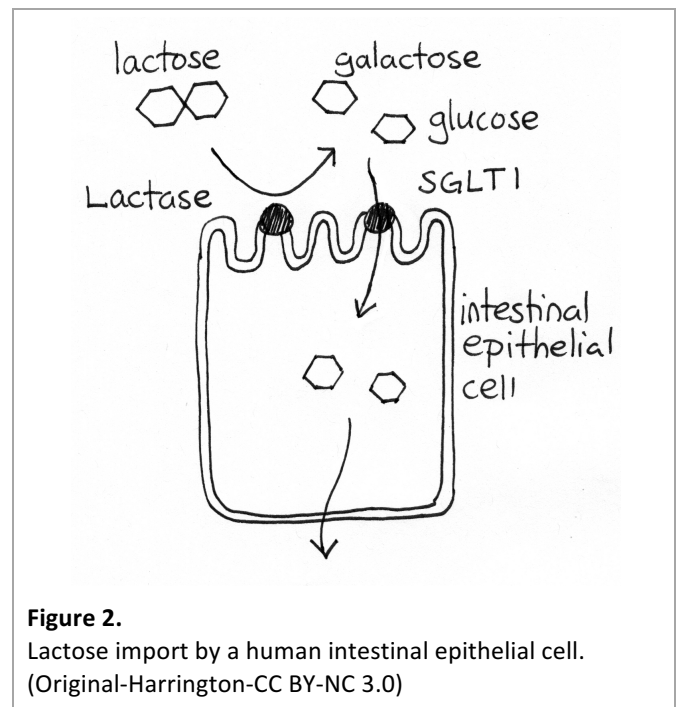
Figure 2.
Lactose import by a human intestinal epithelial cell. (Original-Harrington-CC BY-NC 3.0)

1. THE LACTASE PROTEIN

The purposes of this chapter are (i) to review how Eukaryotic genes make proteins and (ii) to demonstrate the importance of mutations in human evolution. Let's start with the protein itself. As shown in **Figure 2**, Lactase is a plasma membrane protein. Like many others it is synthesized by ribosomes at the rough ER, modified by enzymes in the Golgi apparatus, and finally moved in a transport vesicle to the surface of the cell. The mature protein is 1060 amino acids long. Most of the protein, including its active site, is outside the cell (**Figure 3**). It has a single **trans-membrane domain,** which anchors it in the membrane. There is also a small cytosolic portion at the carboxyl-end of the protein.

2. THE LCT GENE AND mRNA

In humans, the gene that encodes Lactase is called *LCT* and is found on chromosome 2. Its cytogenetic location is 2q21 (chromosome 2, long arm, region 2, band 1; **Figure 4**). Because chromosome 2 is an autosome everyone has two copies of the *LCT* gene, one on their maternal chromosome 2 and one on their paternal chromosome 2. Information about this and other human genes can be found on the Online Mendelian Inheritance in Man website. This website is easy to find if you search for "OMIM".

The *LCT* gene is about 55 kilobases (kb) long (**Figure 5**). When it is transcribed, RNA polymerases travel along the DNA from the promoter through the AATAAA polyadenylation site. The RNA made is known as a pre-mRNA (or primary transcript) as it still requires processing. For this particular pre-mRNA, 16 introns are removed yielding a much shorter mature mRNA. The *LCT* mature mRNA contains a 5' cap, 6274 nucleotides, and a poly(A) tail. (**Figure 6**).

The structure of the *LCT* mRNA is typical (**Figure 7**). It has untranslated regions (**UTRs**) at each end and a 5784 nucleotide long **coding sequence** in the middle. Since three nucleotides are equal to one codon this mRNA encodes a protein that is 1927
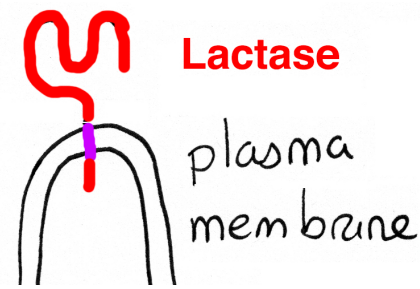
Figure 3.
The Lactase protein. Its trans-membrane domain is shown here in purple.
(Original-Harrington-CC BY-NC 3.0)
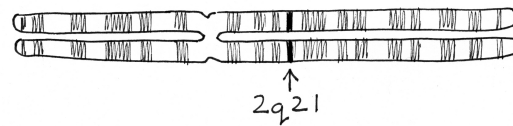
Figure 4.
The location of the *LCT* gene on *Homo sapiens* chromosome 2 (HSA2).
(Original-Harrington-CC BY-NC 3.0)
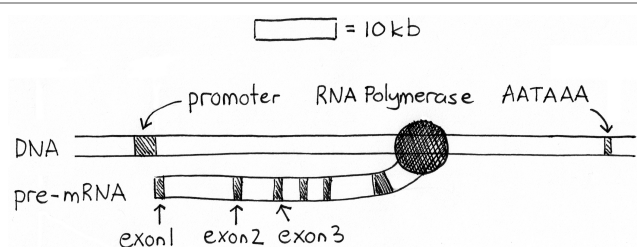
Figure 5.
The *LCT* gene and pre-mRNA. (Original-Harrington-CC BY-NC 3.0)
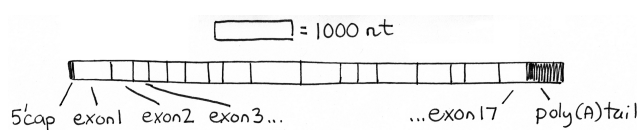
Figure 6.
The LCT mature mRNA. (Original-Harrington-CC BY-NC 3.0)

amino acids long. However, recall from earlier in this chapter that the mature protein is only 1060 amino acids long

The structure of a typical human protein coding mRNA including the untranslated regions (UTRs)

Figure 7.
A typical human mRNA shown approximately to scale. In the *LCT* mRNA the lengths of the 5'UTR, coding sequence, and 3'UTR are 11, 5784, and 479 nucleotides, respectively. Codon 1 is AUG = methionine/start and codon 1928 is UGA = stop. (Wikipedia-Daylite-PD)
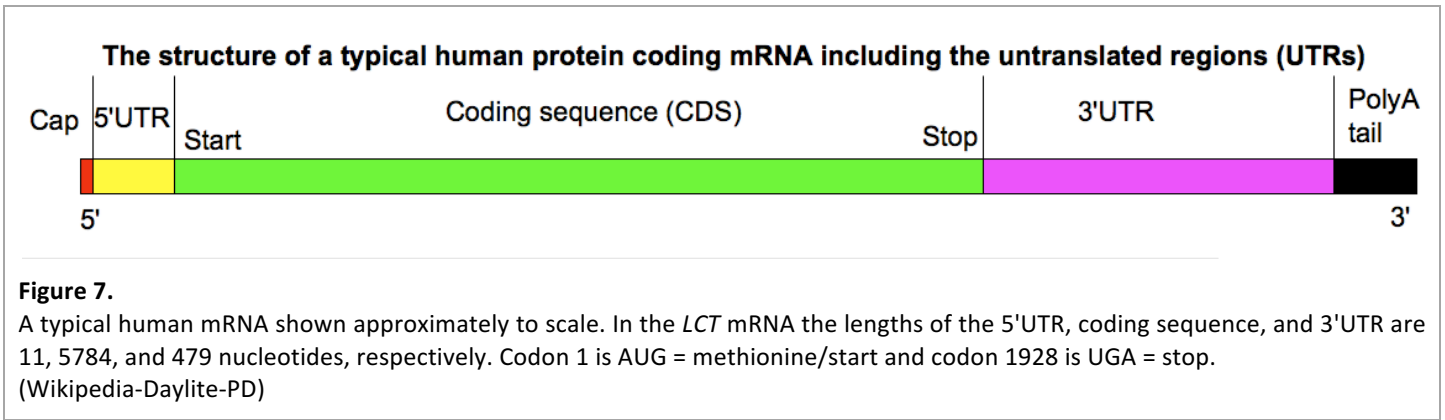
The resolution to this mystery comes from how the Lactase protein is made (**Figure 8**). When the ribosome first binds to the mRNA, both are floating free in the cytosol. The first 20 amino acids of the Lactase protein are an **ER signal sequence**. This signal attracts an RNA-protein complex called the Signal Recognition Complex. It brings the ribosome to the surface of the rough ER. The ribosome continues protein synthesis but now the protein is fed into the ER lumen (**Figure 8a**). Since its job is done, the ER signal sequence is cut off and its amino acids are recycled. These are typical events in the synthesis of membrane proteins such as Lactase.

Towards the end of the coding sequence comes a **stop transfer sequence**. This portion is left in the ER membrane and becomes the trans-membrane domain (**Figure 8b**). The ribosome soon reaches the stop codon in the mRNA and departs.

At this point we have a membrane protein but it is still longer than expected. This so-called pro-Lactase protein travels to the Golgi apparatus. There, an enzyme cuts the protein a second time releasing an 847 amino acid long **pro region** (**Figure 8c**). The remaining protein is the 1060 amino acid long mature Lactase. It is modified a bit more in the Golgi apparatus before being sent onwards to the plasma membrane (**Figure 8d**).

Figure 8.
Synthesis of Lactase proteins in a human intestinal epithelial cell.
(Original-Harrington-CC BY-NC 3.0)
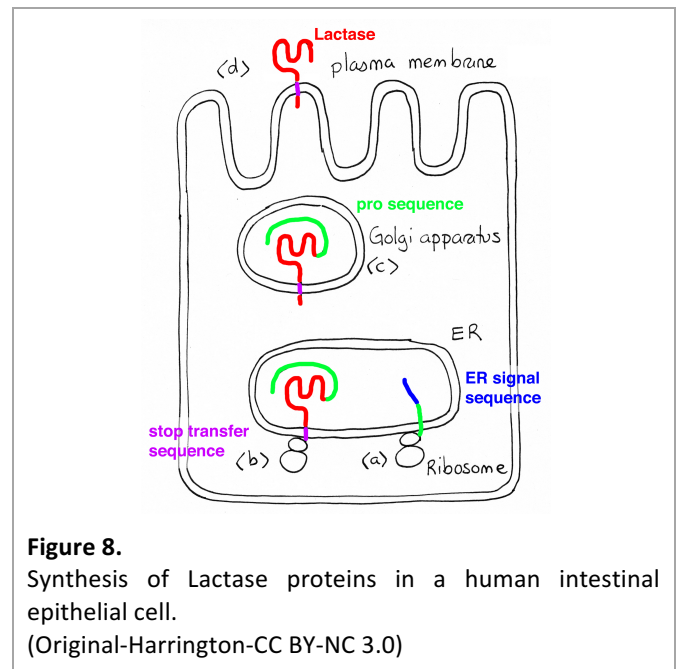
The purpose of this pro region was a mystery until 1994 when scientists made synthetic Lactase proteins that lacked them. These Lactase proteins were unable to fold into their correct shapes. Thus the pro regions are necessary so that normal Lactase proteins can fold properly in the ER. Afterwards in the Golgi apparatus, the pro regions are removed so that the enzymatic active sites are exposed. Some other proteins, for example Insulin, also contain pro regions when they are first made. Once the pro-proteins fold into their proper shapes the pro regions are removed and their amino acids recycled.

3. LCT GENE EXPRESSION DURING DEVELOPMENT

Because young mammals are completely dependent upon mother's milk their *LCT* genes are very active in the intestinal epithelial cells (**Figure 9**). This supplies these cells with enough Lactase enzymes to digest the lactose sugars in the gut. The resulting glucose and galactose can then be imported into the cells. Other cells do not turn on the *LCT* genes because they have no use for Lactase proteins. This is an example of **spatial gene expression** - when a gene is active in only some cells in a multicellular organism. In liver, muscle, and other cells the *LCT* genes remain off to not waste nucleotides on unneeded mRNAs and amino acids on unneeded proteins. This also saves energy that would be required by the transcription and translation machinery too.
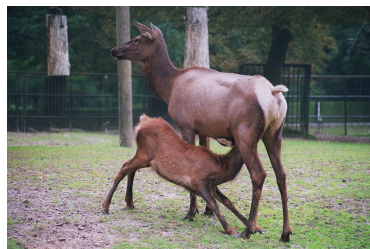
Figure 9.
A young elk drinking milk (top) and an older elk eating grass (bottom).
(Wikipedia-Left: Norbert Kaiser, Right: Jonathunder-CC BY-SA 3.0)

When the mammal is older it will switch from mother's milk to eating regular food, a process called weaning. For most mammals this means that there will be no more lactose in the diet, and thus no more need for Lactases. The *LCT* genes are turned off in the intestinal cells because their job is done. This is an example of **temporal gene expression** - when a gene is only active during specific stages during an organism's development.

In about 65% of people the *LCT* genes follow this pattern of temporal gene expression. By the time they turn eight the *LCT* genes have turned off. Without Lactases they can no longer break down lactose. If they drink cow milk or eat too much cheese, ice cream, or other dairy products the result is diarrhea. However in 35% of people this doesn't happen. The *LCT* genes remain active, Lactases continue to be produced, and milk and dairy products do not cause gastric problems. Each of us is thus either **lactose intolerant** or **lactose tolerant**. Either we have stopped making Lactase (a phenotype also known as **Lactase non-persistence**) or we continue to make it (**Lactase persistence**). The explanation for this difference reveals much about human genetics and human evolution.

Before we move on, an option for lactose intolerant people are lactose-reduced dairy products such as those shown in **Figure 1**. How are they made? The answer is simple, Lactases are purified from yeast such as *Kluyveromyces fragilis* and added to food during its processing. Any lactose will be broken down into monosaccharides by the time the person eats the food. All people, lactose tolerant and intolerant, can import glucoses and galactoses into their intestinal cells using their SGLT1 transporters.

4. EVOLUTION OF THE LCT GENE

If we look at the distribution of lactase persistence (LP) in the human population we see three hotspots - Northern Europe, Eastern Africa, and Arabia/India (**Figure 10**). After much searching scientists found that in each case there was a single mutation in *LCT* responsible. In the European population a CG to TA base pair substitution created the LP allele. The reason it took so long to find was the mutation was far upstream of the transcribed region, 13 910 bp in fact!

The LP alleles in the other populations were different bp substitutions very close by (**Figure 11**). In African populations it was a TA --> GC bp substitution at –13 915 while in Arabia/India was a CG to GC bp substitution at –13 907.
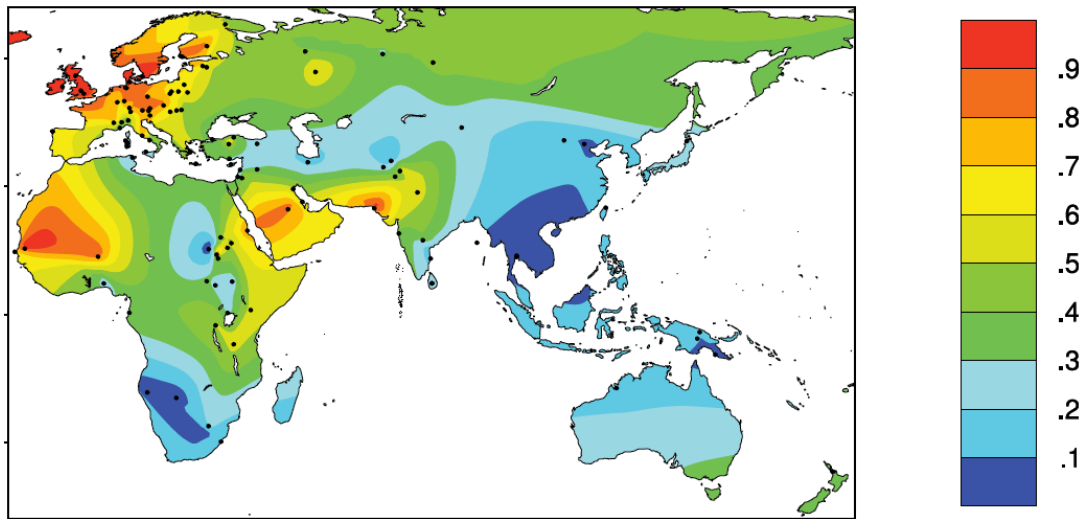
Figure 10.
Distribution of the lactose tolerant phenotype. Dots represent collection locations. Colours show the frequency of the lactose tolerant phenotype from 0-10% to 90-100% of the local population.
(BMC Evolutionary Biology, 2010, 10:36-Itan et al-CC BY 2.0)

How can these mutations have an effect? What each does is to turn this stretch of DNA into a binding site for a **positive transcription factor**. Positive transcription factors bind to genes and activate them, while **negative transcription factors** bind to genes and have the opposite effect. In this case the positive transcription factor is a protein known as Oct1. It is Oct1 that is keeping the *LCT* genes active long after they would otherwise be turned off. All it took to change the *LCT* gene's temporal expression pattern were mutations in the gene's regulatory region. The result had a dramatic effect on a person's phenotype.

Like other mutations these three occurred randomly. But why did these mutations become so common? The answer comes from the food consumed in these places over the past thousands of years. All three groups of people raised animals; goats, sheep, cows, or camels; that could be milked. Milk and milk-products offered a new year round food source. People in these communities with the lactose tolerance phenotype would have had more food available and been able to have more children. Their children would have inherited the LP alleles and also had this advantage. Over many generations the population shifted to where most if not all of the people had the LP alleles and

were Lactase persistent. In other places in the world, places where lactose tolerance had no benefit, any LP alleles that arose would have not been selected for and would remain rare.
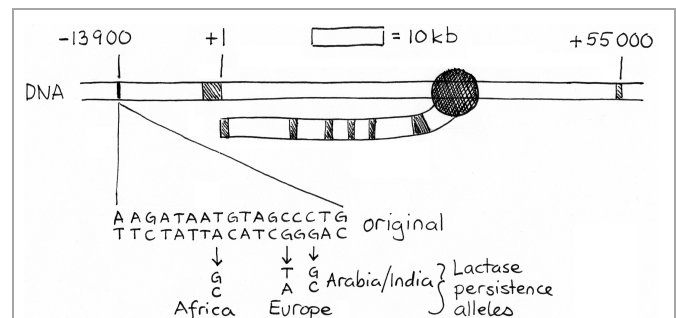
Figure 11.
Mutations in the LCT gene that produce a Lactose persistence allele. Note that locations on genes are numbered relative to the first base pair read by the RNA Polymerase being +1.
(Original-Harrington-CC BY-NC 3.0)

In summary, while there are different alleles of the *LCT* gene in the human population neither type is "better". The original allele turns off after weaning and thus conserves nucleotides, amino acids, and energy. The LP alleles remain on and allow a person to eat a greater variety of foods. Ultimately, the reason you are either lactose tolerant or intolerant has to do with what your ancestors ate and drank thousands of years ago!

SUMMARY:

- In the human gut the dissaccharide lactose is hydrolyzed by an intestinal epithelial cell membrane protein named Lactase. The resulting monosaccharides, galactose and glucose, are then imported into the cell by the SGLT1 transport protein.

- The *LCT* gene makes a long pre-mRNA which is processed (5' cap added, introns removed, poly(A) tail added) to produce a much shorter mature mRNA.

- The Lactase protein contains regions that control where it is synthesized (ER signal sequence), become a membrane domain (stop transfer sequence), and assist with its folding (pro sequence). Some of these regions are removed as the protein is formed and delivered to its final location.

- The *LCT* gene shows both spatial gene expression (it is only active in some cells) and, in many people, temporal gene expression (it is only active during some developmental stages).

- During human history three independent mutations have generated Lactose persistence (LP) alleles of the *LCT* gene. These mutations have altered how the gene is regulated. People with one of these alleles can digest lactose during their whole lives and not just as infants.

KEY TERMS:

lactose
Lactase
SGLT1
trans-membrane domain
UTR
coding sequence
ER signal sequence
stop transfer sequence

pro region
spatial gene expression
temporal gene expression
lactose intolerant /Lactase non-persistence
lactose tolerant / Lactase persistence
positive transcription factor
negative transcription factor

STUDY QUESTIONS:

1) If a person was heterozygous for *LCT*, i.e. they had one Lactase persistence allele and one Lactase non-persistence allele, what would their phenotype be? In other words, is the LP allele dominant or recessive to the original allele?

2) Go to OMIM and find the entry for the *LCT* gene. What are the alternative symbols for this gene? Why is it necessary for the HUGO Gene Nomenclature Committee to approve only one symbol for each gene?

3) One way to find out if a person is lactose tolerant or intolerant is to feed them some lactose and then monitor their blood glucose levels. How does this work?

4) Insulin proteins are synthesized and exported from human pancreatic cells. Consult **Figure 8** and describe how these proteins are made.

5) In *E. coli* a protein called Lac Permease imports lactose into cell so that a protein called Beta-Galactosidase can turn it into galactose and glucose. What are the similarities and differences between how *E. coli* imports lactose and how you do?

6) How do you suppose other dissaccharides are digested in humans? Note that in our diet the most common dissaccharides are lactose (galactose + glucose), sucrose (glucose + fructose), and maltose (glucose + glucose).

7) Why do people with a lactose intolerant phenotype have gastric troubles if they drink milk or eat dairy products? More specifically, what problem is all that undigested lactose causing in their large intestines?

Recommended course:

Organization of Complex Genomes

GENET 302, Organization of Complex Genomes
★ 3 (fi 6) (second term, 3-0-0)
Analysis of how eukaryotic chromosomes are organized, inherited, studied, and manipulated. Topics include: classical and current techniques, mouse genetics, epigenetics, sex chromosomes, dosage compensation, genomic imprinting, transposable elements, centromeres, telomeres, and stem cells.
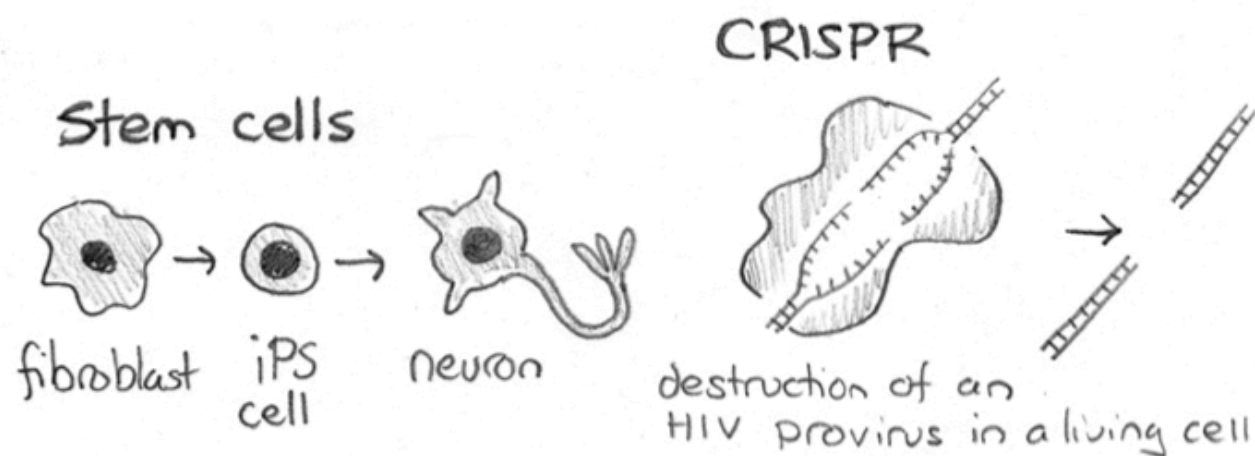
Prerequisites: GENET 270

Class schedule: MWF 9:00-9:50 AM

Offered: Winter Term

Contact: Dr. Harrington, mjh@ualberta.ca

Calendar Link:
http://calendar.ualberta.ca/preview_course_nopop.php?catoid=6&coid=44623

CHAPTER 10 – EUKARYOTIC GENES:
THE DROSOPHILA *WHITE* (*W*) GENE

Figure 1.
A *Drosophila melanogaster* adult male
(Wikipedia- André Karwath - CC BY-SA 2.5)

INTRODUCTION

One of the most striking features of *Drosophila melanogaster* is the adult's large red eyes (**Figure 1**). As with other insects, these are **compound eyes**. Each *Drosophila* eye is made of about 800 tubes called **ommatidia** arranged in a hemisphere (**Figure 2**). Light enters the outwards facing side of the ommatidium and activates a light sensitive photoreceptor cell at the base. This cell sends a nerve impulse to the brain. In order for compound eyes to function the sides of each ommatidium have to be opaque - otherwise light coming from other directions will activate the photoreceptor cell. Thus each ommatidium has three parts, a **lens** at the top, **pigment cells** along the sides, and a **photoreceptor cell** at the base.
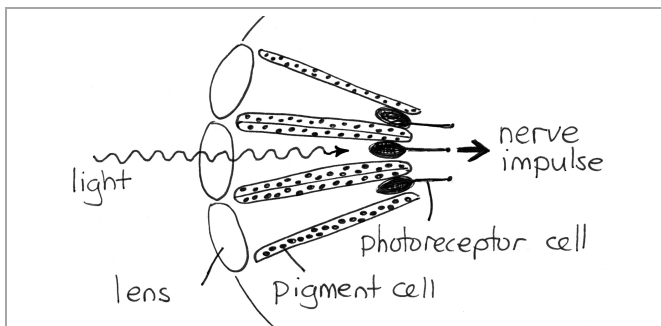
Figure 2.
Three ommatidia within a Drosophila eye. Arthropod compound eyes have multiple lenses in contrast to human eyes, which have a single lens each.
(Original-Harrington- CC BY-NC 3.0)

In *Drosophila melanogaster,* two types of **pigments** are used: orange-coloured drosopterins and brown-coloured ommochromes. Eyes that contain both pigments have the wild type, bright red colour. Synthesizing these pigments requires a set of **transporters** and enzymes. If any of the genes encoding these proteins is mutated, the result will be a fly with an altered eye colour. In the wild this would be detrimental, however, a fly confined within a laboratory vial does not require vision to find food and mates. Eye colour mutations therefore do not compromise the viability and fertility of lab strains.

Because eye colour mutants are easy to isolate and propagate, scientists have used them to make many scientific discoveries. The best example of this is a gene called *white* (*w*), with mutants giving a white coloured eye. This chapter describes how this gene functions, some of its mutant alleles, and why it is important in the history of genetics. The study of fly genes provides insight into gene expression, function, control for other genes, including human genes and diseases.

1. THE *WHITE* GENE PROTEIN

Each fly eye begins as a clump of cells called an imaginal disc inside the larva. During pupation these imaginal discs grow and mature into eyes. One of the developmental steps in the future pigment cells is to import tryptophan and guanine

molecules (**Figure 3**). These will be converted into brown (ommochrome) and orange (drosopterin) pigments, respectively. Tryptophan and guanine are imported by transporter proteins in the cell's plasma membrane. Both transporter proteins are **heterodimers**, proteins made of two different polypeptides. The transporter made with the W and S polypeptides imports tryptophan while the guanine transporter is made with the same W polypeptide joined with a B polypeptide.
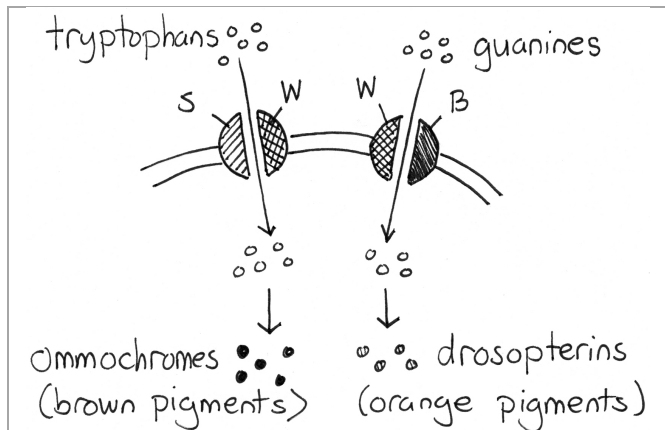
Figure 3.
Import of pigment precursor molecules into a future eye pigment cell.
(Original-Harrington- CC BY-NC 3.0)
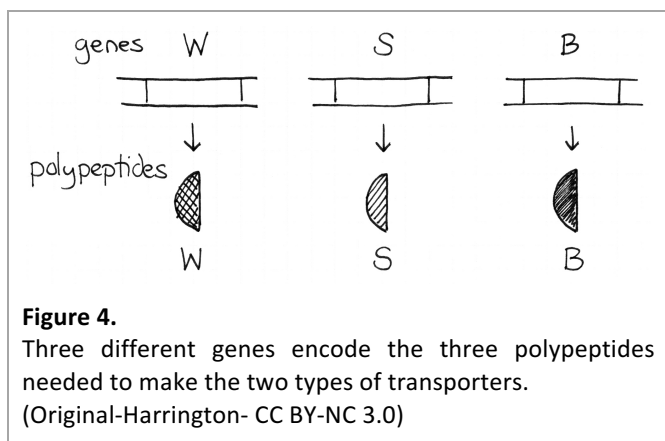
Figure 4.
Three different genes encode the three polypeptides needed to make the two types of transporters.
(Original-Harrington- CC BY-NC 3.0)

Each of these three polypeptides is encoded by a different gene, there being three genes in total (**Figure 4**). From these figures we can predict what would happen if any one of these genes were non-functional.

- If the S gene was mutated, pigment cells would not make any S polypeptides, but the synthesis of W and B polypeptides would be unaffected. These cells would therefore be lacking tryptophan transporters but would have functional guanine transporters. During pupation the cells would be able to synthesize orange (drosopterin) but not brown (ommochrome) pigments. The cells, and the fly eyes as a whole, would be orange.

- If the B gene was mutated the opposite situation would happen, the future pigment cells would be able to import tryptophan but not guanine. They would contain brown but not orange pigments. These flies would have brown eyes.

- If the W gene was mutated neither transporter could be produced. With no precursors imported there would be no pigments and the flies would have white eyes.

These figures are missing one piece of information, the true names of the three genes. Each was discovered decades before their protein's cellular function was revealed. So what did geneticists name a gene that, when mutated, makes the eyes white? Well, it was named the *white* gene. To reduce confusion *Drosophila* geneticists use a system of italics and capital letters when referring to DNA, RNA, and proteins. In this system the *white* gene is transcribed into the *white* mRNA, which translated into the WHITE protein. The wild type (functional) allele of the *white* gene can be depicted *white*[+] or just *w*[+].

The two other genes were named the same way. The actual name of the S gene is *scarlet (st)*, while the B gene is officially the *brown (bw)* gene. **Figure 5** shows the actual names of the three genes and their polypeptide products. There are a few other genes that also mutate to produce an unusual eye colour. Many of these encode enzymes which turn the tryptophans into ommochromes or guanines into drosopterins. One example is the *rosy* gene which makes the Xanthine Dehydrogenase enzyme. Flies without this enzyme can not synthesize the

orange pigments and have brown-coloured eyes as a consequence.

2. THE *WHITE* GENE

2.1. THE FUNCTIONAL (WILD TYPE) ALLELE

The *white* gene is a typical eukaryotic gene. It makes a pre-mRNA that will have five introns removed during processing to yield a shorter mature mRNA. In **Figure 6** the transcribed region on the DNA has an arrow indicating where the RNA polymerase starts and the direction it travels. The mature mRNA is shown below where boxes
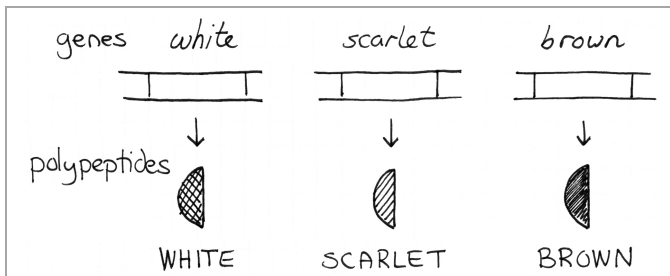
Figure 5.
The actual names of the genes that make the two transporters.
(Original-Harrington- CC BY-NC 3.0)
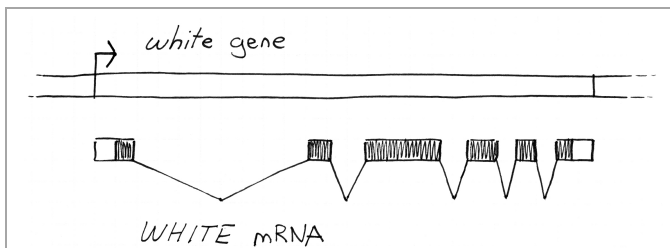
Figure 6.
The *Drosophila white* gene and its mRNA.
(Original-Harrington- CC BY-NC 3.0)
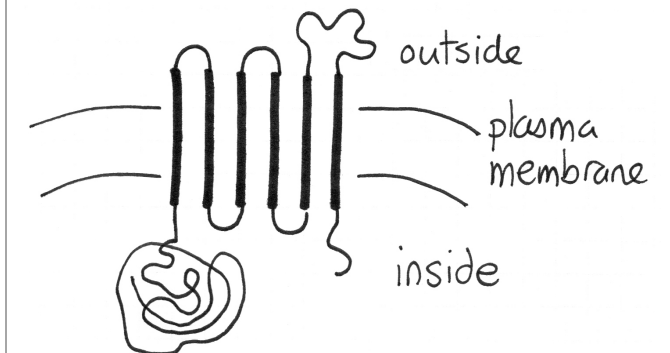
Figure 7.
The *Drosophila* WHITE polypeptide.
(Original-Harrington- CC BY-NC 3.0)

represent exons and the filled in sections are the protein coding region. In this mRNA the start codon is in the first exon and the stop codon is in the sixth (last) exon. The V's below and connecting the boxes represent the introns which were removed. This representation allows the mature mRNA and gene sequences to line up vertically. The figure omits the 5' cap and poly(A) tail that are present in the mature mRNA.

Figure 7 shows the WHITE polypeptide. It has been flattened into two-dimensions so that the six trans-membrane domains and large cytosolic domain can be seen. The actual polypeptide would join with a similarly structured BROWN or SCARLET polypeptide to form a cylindrical membrane protein.

2.2. THE FIRST MUTANT ALLELE TO BE DISCOVERED

In 1910 T. H. Morgan, the father of *Drosophila* genetics, described his discovery of an unusual fly:

> "In a pedigree culture of *Drosophila* which had been running for nearly a year through a considerable number of generations, a male appeared with white eyes. The normal flies have brilliant red eyes."

This fly carried a mutation later named *white*[1] (w^1). The #1 indicates that this was the first of many mutant alleles (>300 now). While you might think that this mutation was due to a simple base pair substitution, many years later it was determined that this mutation was actually due to the insertion of a **transposable element** into the *white* gene. Transposable elements, in this case one called *Doc*, are pieces of DNA that jump from one location and insert into new locations on chromosomes. If they happen to insert into a gene the gene is mutated (non-functional), which is what happened to the *white* gene in one of Morgan's flies.

We now know that transposable elements are very active in *Drosophila* and are responsible for 50% of the spontaneous mutations discovered by early *Drosophila* geneticists. Humans also have transposable elements but they are much less

active. They only cause 0.2% of our spontaneous mutations.

2.3. THE *WHITE-APRICOT* ALLELE

Geneticists have isolated over a thousand mutant alleles of the *white* gene. Not all of these have a white-eyed phenotype though. The *white*apricot (w^a) allele has apricot-coloured eyes for example (**Figure 8**). It was also caused by a spontaneous transposable element insertion, in this case one called *copia*. Because of the location and orientation of the copia element insert, the w^a allele can still make mRNAs but not at wild type levels. This allows the future pigment cells to synthesize some WHITE polypeptides. Consequently, there are fewer transporter proteins on the surface of the cells and thus not enough precursors are imported to make pigments. With fewer of both types of pigment molecules, the cells and eyes have a pale orange colour.
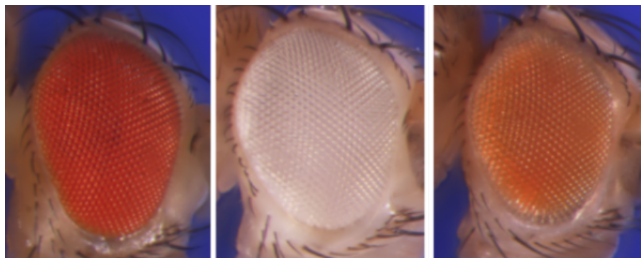
Figure 8.
Drosophila with *white*+ (left), *white*- mutation (middle) and *white*apricot mutation. (right) (FlyBase-PD)

2.4. OTHER MUTANT ALLELES

Mutant alleles of the *white* gene produce fly eyes that range from white to yellow, to orange, and to red with every shade in between. *Drosophila* geneticists can tell these mutant alleles apart using some tricks:

- Observe young adults when pigment deposition is incomplete. Many of the mutant colours darken with age and become harder to distinguish in older flies.

- Look for **pseudopupils**. Pseudopupils are shadows that appear to float in the middle of insect compound eyes (**Figure 9 left**). In most *Drosophila* eye colour mutants the pseudopupil is absent and the eye has a flat, uniform colour.

- Look at the **ocelli**. These are three simple eyes on the top of insect heads (**Figure 9 right**). *Drosophila* adults use these to keep themselves upright as they fly. In *Drosophila* the ocelli are normally red but they are unpigmented in certain mutant strains.

Figure 9.
A mantis with prominent pseudopupils (top) and a wasp with its ocelli circled (bottom).
(Wikipedia-right: Luc Viatour/left: Assafn- CC BY-SA 3.0)

3. THE WHITE GENE IS X-LINKED

T. H. Morgan noticed that when single white-eyed male flies were mated to female flies with regular red eyes, the offspring all had red eyes. When he mated the offspring and observed the next generation he found that 3/4 of the flies had red eyes and 1/4 had white eyes. But this was different from the typical 3:1 Mendelian ratio because only the males had white eyes. Initially he suspected that the white-eyed phenotype was only possible in males, but soon he had more data. He had also taken that same white-eyed male and mated it to some of its female offspring. From this cross he found a 1:1:1:1 ratio of white-eyed males, red-eyed males, white-eyed females, and red-eyed females.

The only explanation that fit was that males had one copy of this eye colour gene while females must have two. This parallels the situation with the X chromosome, males have one and females have two. His explanation for the results was that the gene for eye colour was *on* the X chromosome. We now call genes such as this **X-linked**. In a famous statement Morgan concluded:

> *"The fact is that this R [the allele for red eyes] and X [the X chromosome] are combined, and have never existed apart."*

His three crosses, using modern nomenclature, are shown in **Figure 10**.

Morgan used these results as confirmation of the **chromosomal theory of inheritance**. Other biologists had proposed that genes were on chromosomes but here was the evidence – the gene for *Drosophila* eye colour was inherited as if on a specific chromosome, the X chromosome. For this and other research T. H. Morgan won the Nobel prize in Physiology or Medicine in 1933.

In many cases, the white-eyed flies students work with in genetics labs are the descendants of that one male fly he discovered over one hundred years ago (or 2500+ generations). Also, the experiment students do today to demonstrate X-linked inheritance would have given them a Nobel prize had they done it before Morgan in 1910!

4. THE IMPORTANCE OF THE *WHITE* GENE

4.1. OTHER DISCOVERIES

Since Morgan's time geneticists have made further discoveries using the *white* gene and its mutant alleles. Some highlights include:

- Much of what we know of heterochromatin comes from the analysis of the *white*$^{mottled\#4}$ (w^{m4}) allele. A chromosome rearrangement has placed the *white* gene too close to the centromere. In some ommatidia the gene is able to function but in most the gene is silenced. The result is a fly with a mosaic of red and white ommatidia (**Figure 11**).

- The first *Drosophila* transposable element discovered was in the *white*^{ivory} (w^i) allele.

- The first *Drosophila* gene cloned and sequenced was *white*. The *white*^{apricot} allele played an important part.

- A functional *white*$^+$ gene is used as a marker for small pieces of DNA when scientists introduce DNA into *Drosophila* or move DNA from chromosome to chromosome. These techniques are called transfection and transposition, respectively.
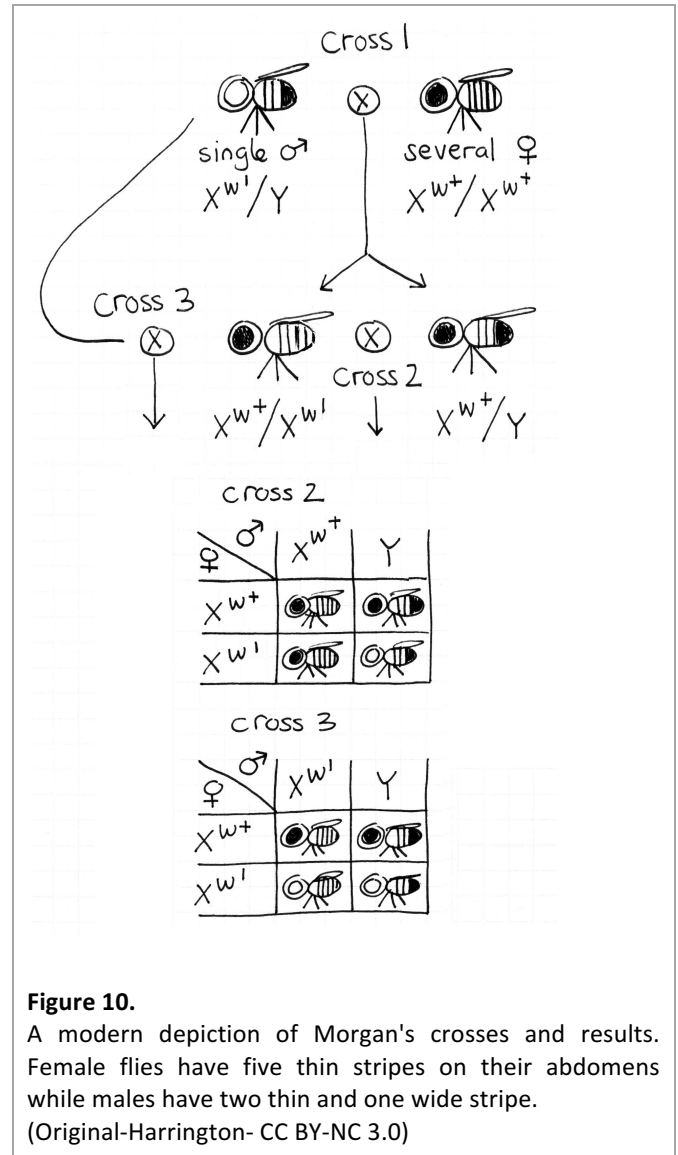
Figure 10.
A modern depiction of Morgan's crosses and results. Female flies have five thin stripes on their abdomens while males have two thin and one wide stripe.
(Original-Harrington- CC BY-NC 3.0)

Figure 11.
A fly with the *white*^mottled#4 mutation.
(PLoS Genet-Piacentini *et. al* (2009)-PD)

4.2. RELEVANCE TO HUMAN GENETICS

Are there any direct connections to human genetics? The answer is yes and no. Eye colour in humans, or more specifically, the colour of the iris, is due to a different pigment, melanin. Irises with lots of melanin are brown or black, while those with small amounts are hazel, green, or blue depending upon how light entering the eye is scattered. The colour of the iris does not affect what we see, although people with lighter eyes are more sensitive to bright light than people with darker eyes. In short, eye colour in humans is not related to eye colour in *Drosophila*.

On the other hand we do have many of the same genes as *Drosophila*. The *Drosophila* WHITE, SCARLET, and BROWN proteins are in the ABC transporter family. Humans have proteins in this family, for example the CFTR protein which is non-functional in people with cystic fibrosis. Of the human ABC transporters the one most similar to WHITE is named ABCG1. It is encoded by a gene on our chromosome 21. ABCG1 is a transport protein, although its job is to export cholesterol and phospholipids from macrophage cells.

The same is true of the enzymes involved in *Drosophila* eye colour. The *Drosophila rosy* gene mentioned in Section 1 is called *XDH* in humans. Both genes make enzymes necessary for purine metabolism. There are many examples of medically important human genes that have a *Drosophila* counterpart. *Drosophila* geneticists can reveal much about human health by studying and manipulating these genes.

_____

SUMMARY:

- Insects have compound eyes. Pigment-containing cells line the side of each ommatidia and serve a crucial function in vision.

- Heterodimer transport proteins import pigment precursor molecules into these cells during development.

- *Drosophila* adults that are unable to make one or both of the transporters do not have the normal red eye colour.

- Mutations in the *white* gene reduce or eliminate eye pigmentation.

- The *Drosophila white* gene was the first X-linked gene discovered and was used by Morgan to confirm the chromosomal theory of inheritance.

- Geneticists continue to use the *white* gene as a tool in their research.

KEY TERMS:

compound eye	transposable element
ommatidia	pseudopupil
pigment cell	ocelli
pigment	X-linked gene
transporter protein	chromosomal theory of inheritance
heterodimer protein	

QUESTIONS:

1) What is the difference between white, *white*, and WHITE?

2) How do *Drosophila* adults with the *white*[1] mutation perceive the world? Is it darker or lighter?

3) What colour eyes would a fly have if it had homozygote mutations in both the *sepia* and the *brown* genes?

4) Other *Drosophila* cells import tryptophan and guanine but use them to make the neurotransmitters serotonin and dopamine. Does this mean that flies with the *white*[1] mutation have altered behaviour?

Chapter 11 – Mutations Originate as Damage to DNA

Figure 1.
The difference in appearance between pigmented and white peacocks is due to mutation.
(Flickr-ecstaticist- CC BY-NC-SA 2.0)

Introduction

The techniques of genetic analysis discussed in the chapters on Mendelian inheritance 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**, or a change to the DNA sequence.

Humans have an interesting relationship with mutations. From our perspective, mutations can be extraordinarily useful because they are essential for the domestication and improvement of almost all the organisms we use as food. On the other hand, mutations are the cause of many cancers and other diseases that can be devastating to individuals. Yet, the vast majority of mutations probably go unnoticed and undetected. In this section, we will examine some of the causes of mutations.

1. Mutation and Polymorphism

We have previously noted that an important property of DNA is its sequence fidelity: most of the time its sequence accurately passes the same information from one generation to the next.

However, DNA sequences can 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** (as opposed to **wild type**).

In a typical population of individuals (e.g. a classroom of students), not all members will have the same DNA sequence – there is genetic variation. The extent of this variation can be divided into two categories. First, naturally occurring but rare (<1%), sequence variants that are clearly different from a normal, wild-type sequence are called mutations. Second, in a population there may be many naturally occurring variants for a trait for which no wild type can be defined. In this case 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 albinism is appropriately called a mutation, but a difference in DNA sequence that explains whether a person has red hair rather than brown, black, or blond hair is an example of polymorphism.

Molecular markers, which we will discuss in the chapters on DNA variation, are a particularly useful type of polymorphism for some areas of genetic research.

2. TYPES OF MUTATIONS

Mutations, or **lesions**, 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 in the next section.

Mutations can occur in many locations, with respect to a gene. They can occur within genes, and so can possibly change the polypeptide sequence from that gene. The severity of that mutation, and how it affects the genes function, can be described using Muller's morphs, which is explained in Chapter 13. They can also occur in regions that are transcribed but not translated. These are non-protein coding genes, which can include tRNA, rRNA or siRNA. Mutations that can still affect gene function can occur in regions that are not transcribed or translated, such as in the promoter or regulatory regions (enhancer/silencer) of genes.

Mutations here will affect the ability of RNA polymerase to bind and transcribe that gene, and so will ultimately affect the overall levels of the protein. Lastly, mutations can occur in regions between genes, or within introns. These mutations will not affect the functions of any genes, and so the organism will appear wild type.

2.1. DELETION AND INSERTION MUTATIONS - FRAME SHIFT

A deletion or insertion mutation may cause dramatic changes in the sequence of the protein. A deletion is removing base pair(s) from the DNA, and an insertion is inserting new base pair(s) into the DNA. Remember that three nucleotides, or a **codon**, code for a single amino acid. If the insertion or deletion is only three nucleotides, it will maintain the sequence reading frame so the protein will have one extra or one missing amino acid (**Figure 2**). The same will occur for multiples of three (6,9,12, etc.). The location of the insertion/deletion will affect the severity of the mutant allele, but this type of mutation is generally less harmful than non-multiples of three.

If a deletion or insertion mutation is not a multiple of three, it will cause a **frame shift**. The typical codon next to the insertion or deletion will be shifted over, and the ribosome will start placing incorrect amino acids after the mutation. This will lead to a severely disrupted protein that will likely not be able to function properly. A frame shift is also very likely to cause a **premature stop codon**. This will lead to a **truncated**, or shortened polypeptide (**Figure 2**). If this happens near the end of the polypeptide sequence, it is likely that a frame shift will not have major effects on the polypeptide function. If it happens near the start then the protein will likely be non-functional.

```
            ATG CCG AAA ATA AGT TTC AGG GGT ...
            Met Pro Lys Ile Ser Phe Arg Gly ...


Three base insertion    ATG CCG AAA CTC ATA AGT TTC AGG GGT ...
  with no frameshift     Met Pro Lys Leu Ile Ser Phe Arg Gly ...


 Two base insertion      ATG CCG AAG CAA TAA GTT TCA GGG GT ...
  with frameshift        Met Pro Lys Glu XXX Val Ser Gly    ...
   and stop codon
```

Figure 2.
The top is an example sequence, and the amino acids produced from those codons. When a three base insertion, there isn't a frameshift, but just an insertion of a new amino acid. When a two base insertion occurs, it causes a frame shift of every base after, in this situation leading to a premature stop codon.
Purple shows inserted bases, green shows affected bases and amino acids, red XXX is the stop codon.
(Original-L. Canham- CC BY-NC 3.0)

Figure 3.
Diagram that shows what changes lead to transition mutations and what changes lead to transversion mutations.
(Wikipedia- Petulda-PD)

2.2. SUBSTITUTION MUTATIONS

Mutations don't always add or remove pieces of DNA. Sometimes they can also just change one nucleotide to another. This is called a substitution. There are two ways that a substitution can change a nucleotide:

(1) A purine can be changed to another purine (A to G or G to A), or a pyrimidine to another pyrimidine (C to T or T to C). This is called a **transition** mutation.

(2) A purine can be changed to a pyrimidine, or *vice versa*. This is called a **transversion** mutation.

A substitution changes the genetic sequence, but this simple change to the DNA sequence can cause three different changes to the polypeptide sequence (**Figure 4**).

(1) The first is known as a **silent mutation**. This is where a substitution in the DNA does not affect the amino acid made. This is because the degeneracy of the genetic code: different codons can code for the same amino acid. For example, glutamic acid can be translated from GAA or GAG, so a transition from A to G in the third position will still produce glutamic acid.

(2) A **missense mutation** is a mutation that changes the amino acid translated. For example if the codon AGC undergoes a transversion mutation to AGG, then it changes the amino acid from a serine to an arginine. Amino acids are put into groups based on their features, hydrophobic, polar uncharged, charged, and other. If the mutation changes the amino acid to another in a similar group (**conservative** missense mutation), the protein may still have partial function. But if a charged amino acid changes to a hydrophobic one (**non-conservative** missense mutation), this is more likely to cause major changes to the function and/or folding of the protein.

(3) A **nonsense mutation** occurs when the substitution leads to a stop codon. For example, a UCA for serine has a transversion mutation to a UAA, which is one of the three stop codons. A stop codon will lead to a truncated polypeptide and, much like in the last section, its position will affect the severity of this mutation.
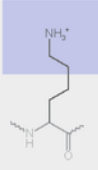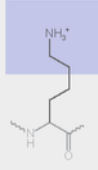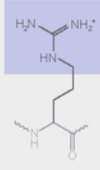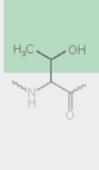
Figure 4.
Examples of substitution mutations that lead to silent, nonsense or missense mutations.
(Wikipedia-Jonsta247- CC BY-SA 4.0)

2.3. CHROMOSOMAL REARRANGEMENT MUTATIONS

There can also be mutations that create major rearrangements of the chromosome. These include **translocations**, and **inversions.** Translocations are a situation where large segments of DNA are swapped between non-homologous chromosomes. Inversions have large segments within a chromosome 'flipped' so the DNA stays within the same chromosome, but its orientation is in the opposite direction. In both of these situations, if a break occurs in the middle of a gene, that gene will be disrupted and unable to make a normal polypeptide. More details on translocations and inversions can be found in the Chromosomal Rearrangements Chapter (Chapter 24).

3. SPONTANEOUS MUTATIONS OF BIOLOGICAL ORIGIN

3.1. ERRORS DURING DNA REPLICATION – SINGLE BASE SUBSTITUTIONS

A major source of spontaneous mutations is **DNA replication errors**. 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, generating a mismatched base pair. Usually, the DNA replication machinery will recognize and repair mismatched
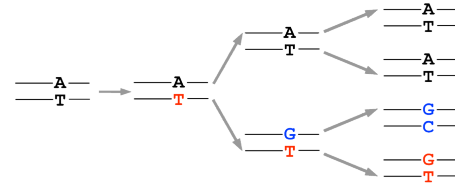
Figure 5.
Mispairing of bases (e.g. G with T) can occur due to tautomerism, alkylating agents, or other effects. As a result, in this example the AT base pair in the original DNA strand will become permanently substituted by a GC based pair in some progeny. The mispaired GT basepair will likely be repaired or eliminated before further rounds of replication. (Original-Deyholos- CC BY-NC 3.0)

or **mispaired bases**, through exonuclease activity of the polymerase, or other repair systems, such as base excision repair or mismatch repair. Nevertheless, some errors become permanently incorporated in a daughter strand, and so become mutations that will be inherited by the cell's descendants (**Figure 5**).

3.2. ERRORS DURING DNA REPLICATION – STRAND SLIPPAGE

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 6**).

Figure 6.
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 BY-NC 3.0)

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. Frame shift mutations account for **mutation hot spots** in some genes. Hot spots are sites in a gene that are significantly more mutable than other sites.

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

3.3. TRINUCLEOTIDE REPEATS
Some regions of the genome have repeated sequences. Dinucleotide repeats (eg. AGAGAGAG) are common throughout the genome, as well as larger repeats such as VNTRs (Variable Number of Tandem Repeats). Trinucleotide repeats (eg. ..CGGCGGCGGCGG..) though have a tendency to expand the region of their repeat, which leads to genetic diseases. These are known as **trinucleotide**

repeat diseases. If there are a low number of repeats, the gene can be stable and the polymerase is able to faithfully replicate the repeats. **Strand slippage**, as described in the last section, can cause an increase in the number of repeats in that region. If it only increases slightly, this usually doesn't cause instability. Once a threshold is reached, the ability of polymerases to faithfully replicate the repeated region becomes more difficult, and the repeats can grow, often very quickly. This mostly occurs in the germline of individuals, and often leads to their offspring making individuals with more and more repeats. This threshold is different for different genes.

An example of this can be seen in with the trinucleotide repeat disease **Huntington's Disease**. Huntington's has the repeat CAG within the reading frame of **Huntingtin gene (*HTT*)**. A normal *HTT* gene will have fewer than 28 repeats. If more repeats are gained, through strand slippage, to 28-35 repeats, then this is called a pre-mutation. At this point, the DNA polymerase cannot faithfully reproduce the repeats and strand slippage occurs more frequently. Such a person is unaffected, but their children may inherit the *HTT* gene with an increased numbers of repeats, and their children after that even more. The more repeats, the more severe the disease. Once the repeats get above 40, the *HTT* gene will cause the disease, which is a neurodegenerative disorder that leads to a

decreased life expectancy. With the CAG repeat within the *HTT* gene reading frame, one can understand why increasing the repeats will cause a decrease in protein function, as it will be gaining more and more of the amino acid glutamine.

Not all trinucleotide repeat disorders are caused by repeats within the reading frame though, for example **Fragile X syndrome**. Fragile X syndrome is a genetic congenital disease characterized by both mental retardation and physical abnormalities like long face and large ears in men. The gene associated with this disorder is ***FMR-1***. At the 5' end of the *FMR-1* mRNA, before the translational start site is a CGG repeat. Most individuals have between 5-54 repeats of CGG. Parents with a premutation, where they are prone to pass on the disease but are not affected themselves, have between 60-200 repeats. Children of individuals with this premutation will have a greatly expanded region of around 200-4000 of repeats and will exhibit the syndrome. Since this is not in the reading frame, why does this expanded repeat cause the disease? It is predicted that the expanded repeat causes aberrant methylation of base pairs the 5' upstream region, which leads to changes in chromatin structure and gene silencing. The karyogram of affected individual exhibits a constricted region on the X chromosome, which is fragile and is prone to chromosomal breakage. This constriction is caused by the **hypermethylation** on the large number of *FMR-1* repeats.

3.4. SPONTANEOUS LESIONS FROM ENDOGENOUS METABOLITES

Occasionally mutagenic lesions are caused by naturally occurring damage to the DNA. These lesions do <u>not</u> occur during DNA replication.

Depurination is a chemical reaction where the bond attaching the purine nucleotide (adenine or guanine) to the deoxyribose sugar is hydrolytically cleaved, leaving a deoxyribose sugar without its nucleic base. This is called an **apurinic site** (**Figure 7**). Unless the apurinic site is repaired correctly, a mutation will occur. Loss of pyrimidines can also occur but the chemical structure of purines makes them a good leaving group, so depurination is more

common. This reaction occurs because of endogenous metabolites within the cell.

Deamination is the removal of an amine group through a hydrolysis reaction. When looking at the individual nucleotides (See Chapter 2), adenine, guanine and cytosine all have amine groups on the base. Like depurination, deamination can occur spontaneously due to metabolites within the cell. In cytosine deamination the cytosine will change to a uracil (**Figure 8**) and free ammonia. This change can be easily corrected, as uracil is only found in RNA, not DNA. So when a uracil is found in DNA, it is removed and replaced again with a cytosine. If this does not occur though, the uracil will pair with adenine, leading to a GC to AT transition mutation.

5-methylcytosine is a cytosine with a methyl-tag. Deamination of 5-methylcytosine is the most common deamination mutation. It leads to thymine and ammonia. In this situation a T would
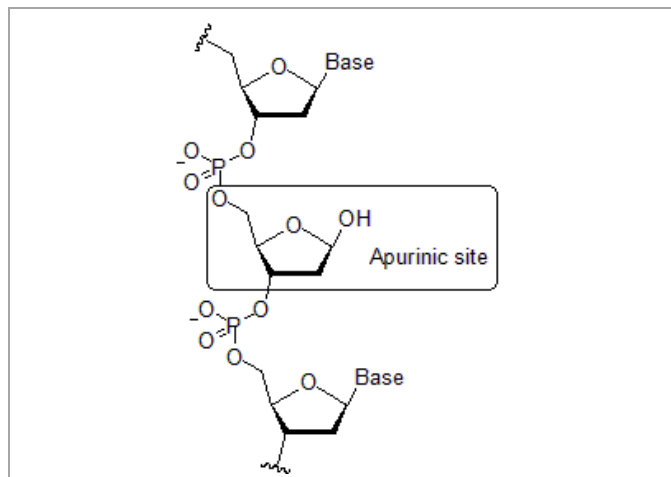
Figure 7.
Structure of an apurinic site in a strand of DNA
Wikipedia-Chemist234- CC BY-SA 3.0)
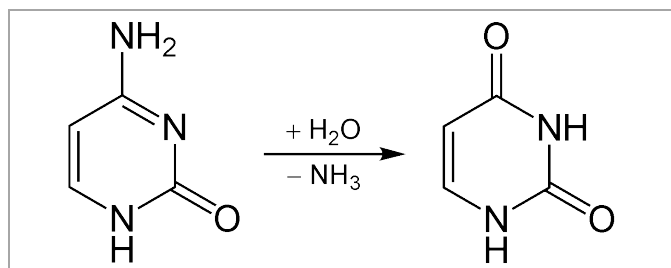
Figure 8.
Deamination of cytosine to uracil.
(Wikipedia-Yikrazuul-PD)

then be opposite a T. This is fixed prior to the passage of the replication fork, but because the repair systems do not know where the original C was located, this can lead to a point mutation at that location.

Deamination of guanine creates the base xanthine. Xanthine is more prone to pair with thymine instead of cytosine. If unfixed, this will cause a GC to AT mutation. Deamination of adenine leads to hypoxanthine is more prone to pair with cytosine instead of thymine. This causes an AT to GC mutation.

Oxidative damage is caused by oxygen free radicals such as superoxide (O_2^-), hydrogen peroxide (H_2O_2) and hydroxyl radicals (OH^-). These are normal byproducts in cells, but are also increased in areas of inflammation as the body uses oxidizing species to attack pathogens. A common product of oxidative damage is 8-oxo-7-hydrodeoxyguanosine (8-oxo dG), which is formed from oxidative damaged guanine. 8-oxo dG mispairs with A, causing GC to AT mutations.

3.5. MUTATIONS FROM TRANSPOSABLE ELEMENTS

Mutations can also be caused by the insertion of viruses, transposable elements (see below), and other types of DNA that are naturally inserted 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 (see Chapter 30). **T-DNA**, which is a transposable element modified from a bacterial pathogen, is used as a mutagen in some plant species.

Transposable elements (TEs) are also known as **mobile genetic elements**, or more informally as jumping genes. They are present throughout the genome 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. The 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 their function and causing mutations.

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.
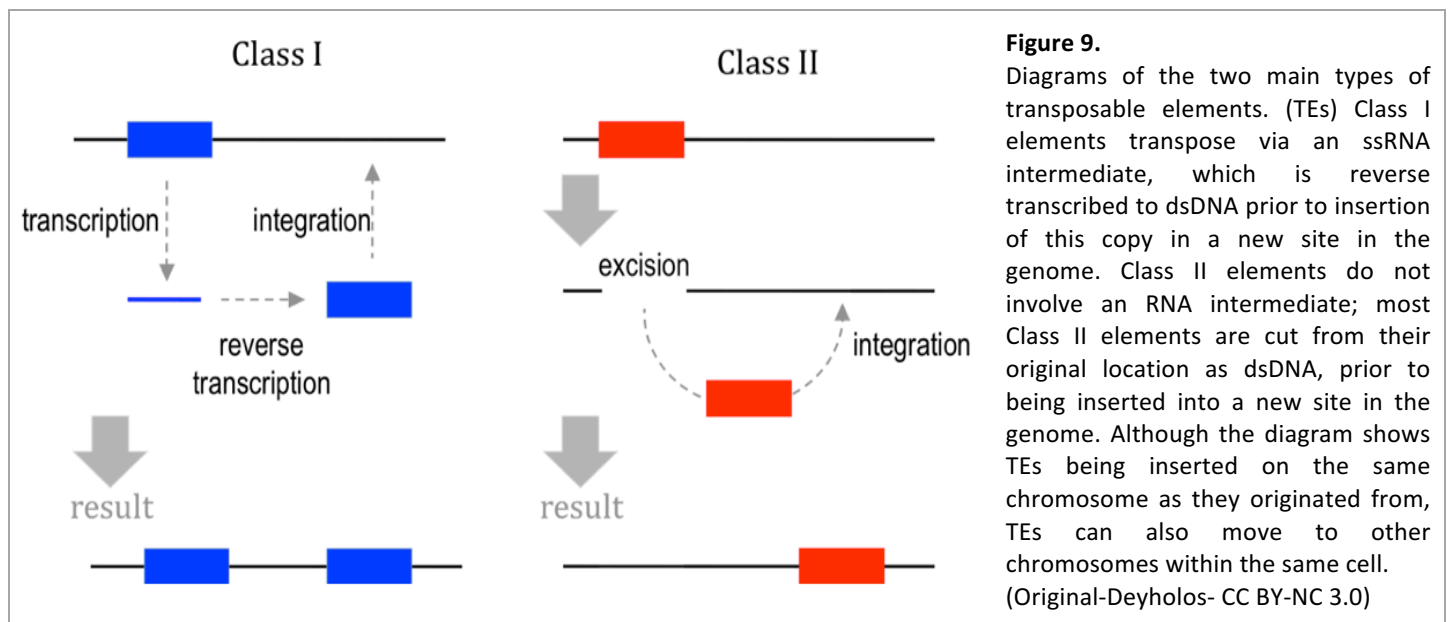
Figure 9.
Diagrams of the two main types of transposable elements. (TEs) Class I elements transpose via an ssRNA intermediate, which is reverse transcribed to dsDNA prior to insertion of this copy in a new site in the genome. Class II elements do not involve an RNA intermediate; most Class II elements 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. (Original-Deyholos- CC BY-NC 3.0)

There are two major classes of TEs in eukaryotes (**Figure 9**).

Class I elements include **retrotransposons**; these are mobile 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 **integrases**.

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.

TEs are relatively short DNA stretches of 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 L̲ong I̲nterspersed E̲lements (**LINEs**) and S̲hort I̲nterspersed E̲lements (**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 the 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. INDUCED MUTATIONS OF CHEMICAL ORIGIN

Many chemical compounds, whether natural or synthetic, can react with DNA and cause mutations. In some of these reactions the chemical structure of particular bases may change, so that they are misread during replication. In other cases the chemical mutagens distort the double helix causing it to be replicated inaccurately, while still other compounds 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**.

4.1. ALKYLATING 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. The compound 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 10**). The new strand therefore bears a C to T transition mutation, which will be inherited in all the strands that are subsequently replicated from it.
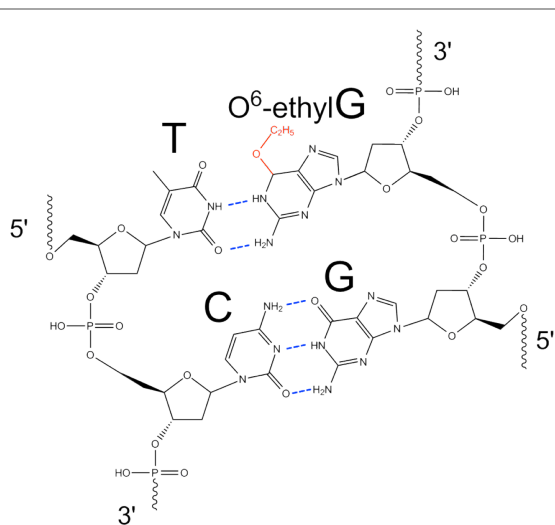
Figure 10.
Alkylation of G (shown in red) allows G to bond with T, rather than with C.
(Original-Deyholos- CC BY-NC 3.0)
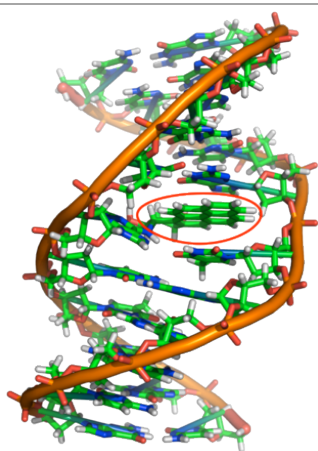
Figure 11.
Benzo[*a*]pyrene (circled in red) is an example of an intercalating agent.
(Wikipedia-Zephyris- CC BY-SA 3.0)

4.2. Intercalating agents

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 11**). 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.

4.3. Other chemical agents

Aflatoxins are a group of fungal metabolites that contaminate corn and peanuts. The metabolites they produce can be toxic and carcinogenic. Aflatoxin B1 (**AFB1**) is the most mutagenic. As it is metabolized for excretion, some of the byproducts are toxic, including endo epoxide and exo epoxide. Exo epoxide intercalates with DNA and then catalyzes a reaction, attaching itself to the N^7 position of guanine. This leads to the removal of the guanine product creating an apurinic site. Additionally, metabolism of AFB1 produces reactive oxygen species, cause oxidative damage as described previously. Combining all the mutagenic properties, aflatoxin mostly creates G to T transversions, but can also create G to A or G to C mutations in low frequencies.

Another chemical mutagen is a **base analog**. Base analogs are chemicals that look similar to a normal nucleotide and so they can be falsely incorporated into the DNA during replication. One example of a common base analog is **5-bromouracil** (5BU). When 5BU is incorporated into DNA it will pair with adenine, but can spontaneously shift into another
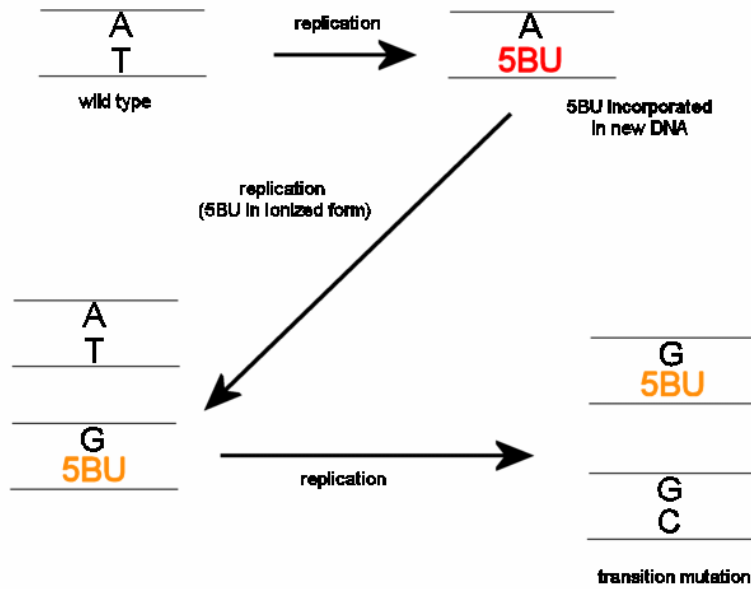
Figure 12.
Mutagenesis with 5-Bromouracil
(Wikipedia-Allen Gathman-CC BY-SA 2.5)
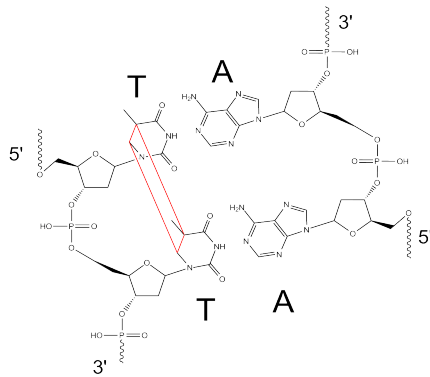
Figure 13.
Thymine dimers are formed when adjacent thymine bases on the same DNA strand become covalently linked (red bonds) follow exposure to mutagens such as UV light. The dimers distort base pairing and can interrupt processes such as replication. (Original-Deyholos- CC BY-NC 3.0)
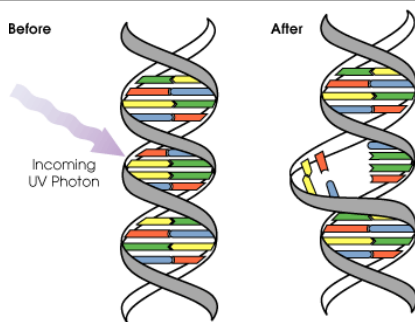
Figure 14.
UV photons can cause adjacent pyrimidines to bond with each other. This distorts the DNA, creating a bulge that prevents polymerases from passing by the lesion.
(Wikipedia- NASA/David Herring-PD)

isomer that pairs with guanine. This ultimately causes an A to G transition (**Figure 12**).

5. INDUCED 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 ultraviolet (UV) light. The smaller, fast moving particles may cause base substitutions or delete a single base, while larger, slightly slower particles may induce larger deletions by breaking the double stranded helix of the chromosome. For example, X-rays can cause DNA double stranded breaks.

Physical mutagens can also create unusual structures in DNA, such as the **pyrimidine dimers** formed by **UV light** (**Figure 13**). Pyrimidine dimers are covalent linkages between two adjacent pyrimidines, with thymine dimers being the most common. When a cell is trying to replicate its DNA, it cannot go through the dimer and so is forced to stop. Replication can only proceed if DNA repair enzymes fix the damage. Pyrimidine dimers cause conformations changes in the DNA, so they are easily recognized by DNA repair enzymes, but are often repaired incorrectly.

The most common types of mutations from UV light are GC to AT transitions but GC to TA, AT to

TA, AT to CG and CG to GC can all be caused by UV mutagenesis.

UV light is a very broad mutagen that can cause many mutation types. Compare this with EMS, which mostly creates GC to AT mutations; or AFB1, which mostly creates GC to TA mutations, but can cause some other mutations in low frequencies as well.

6. FAILURE OF REPAIR SYSTEMS

For each type of damage cells have a way to fix it. These repair systems include but aren't limited to base excision repair, nucleotide excision repair, mismatch repair, non-homologous end joining and homologous recombination. The mechanisms of each repair are not important at this time. All these systems require multiple enzymes to recognize the specific type of mutagenic lesion and repair it as accurately as possible. In certain situations, **DNA repair systems** are unable to cope with DNA damage, either because the damage is too numerous for the enzymes to be able to recognize and repair all of them, or because there is damage to the DNA repair systems itself.

If the final efforts to rescue a cell from DNA damage are successful, the cell will be able to survive replication, but will be full of mutations that are potentially deleterious to its long-term health. If it is not successful, cells will enter one of three possible states: (1) they will enter a state of dormancy, or **senescence**, where the cell is still living but no longer functional; (2) programmed cell death, or **apoptosis**, will be activated and the cell will die; or (3) unregulated cell division, where the cell will divide rapidly despite numerous mutations and chromosomal abnormalities (**Figure 15**). This can lead to **cancerous** tumours (Chapter 41).

6.1. EXCESSIVE DNA DAMAGE

A cell may be exposed to DNA damage past a **threshold** that it is normally capable of dealing with. Such DNA damage usually prevents the DNA polymerase from normally replicating the DNA and it becomes 'stuck'. If the cell is unable to find a way to continue replicating it will lead to cellular death. Alternatively, it may enter an **error-prone DNA repair** system.

In prokaryotes this is called SOS repair, and when induced will recruit error-prone **DNA translesion polymerases** to continue to replicate the DNA past the mutagenic lesion usually causing errors in the place of the DNA damage. A similar example is seen in eukaryotes called translesion synthesis.
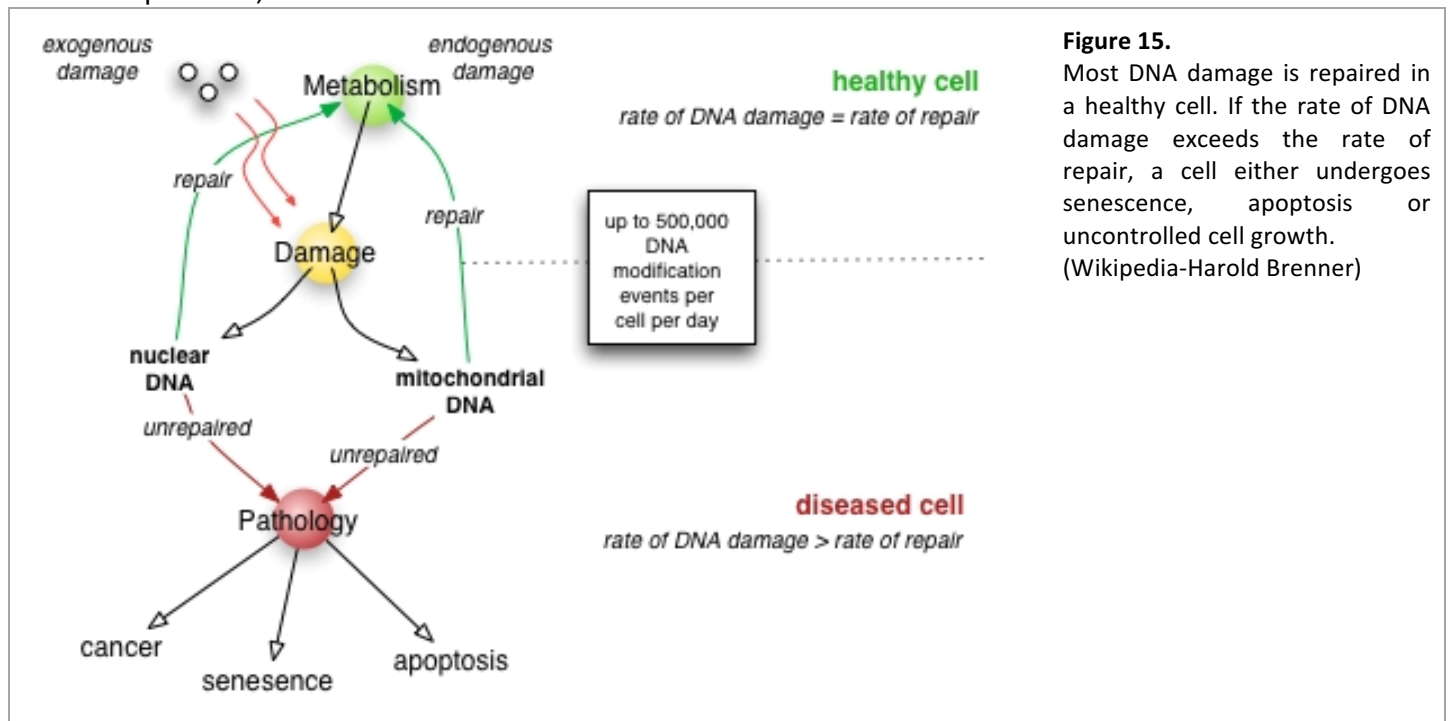
Figure 15.
Most DNA damage is repaired in a healthy cell. If the rate of DNA damage exceeds the rate of repair, a cell either undergoes senescence, apoptosis or uncontrolled cell growth.
(Wikipedia-Harold Brenner)

This is often used when there are excessive DNA lesions from thymine dimers or apurinic sites. If these lesions are not able to be repaired through their normal mechanism, specific translesion polymerases. Like in the SOS response, the translesion polymerase is error prone and will often insert incorrect bases in the areas of the lesion.

6.2. DAMAGED DNA REPAIR SYSTEMS

Another source of mutations is when the DNA repair systems fail. The genes that make the proteins for the various DNA repair machinery are just like any other genes, they can be mutated and cease to function. Many people are homozygous wild type for these DNA repair genes. Also, most of these gemes are haplosufficient, meaning that only one copy of the normal allele is required to produce a normal DNA repair system. Those who are heterozygous are at a greater risk though. They maybe one mutation away from completely losing function of that DNA repair systems gene. When that gene is lost, DNA repair mechanisms might not be able to work as well, or at all, and mutagenic lesions can occur within the genome, leading to permanent mutations.

If an individual is heterozygous for DNA repair genes, then individual cells are at a higher risk when exposed to mutagens. Since UV light is one of the most common mutagens we encounter, the loss of DNA repair proteins in skin cells exposed to UV light can make those cells more susceptible to cancer. Similarly, individuals who smoke increase mutagens in their lungs. Thus, if heterozygous they are more susceptible to DNA damage and cancer in their lungs compared to a homozygous wild type individual.

If damage to DNA repair system genes happens in the gametes of both parents, then they can pass that on to their child, making the child homozygous mutant in that specific DNA repair gene. Instead of individual cells being at risk, every cell in the homozygous mutant child will be missing that DNA repair gene, causing problems with DNA repair. An example of this is the disease **Xeroderma pigmentosum**, which is caused by a mutant version of a nucleotide excision repair enzyme. Individuals with this inherited disease are extremely susceptible to UV light, and develop skin cancer very easily and usually die at a young age.

Xeroderma pigmentosum is just one example of many when people who are born with defects in one of the DNA repair system genes. Most of these diseases lead to various types of cancers, particularly many hereditary colorectal cancers, like hereditary nonpolyposis colorectal cancer (HNPCC). The breast cancer genes, BRCA1 and BRCA2 are associated with DNA repair as well, and when mutant lead to early onset breast and ovarian cancers.

Summary:

- Variations in DNA sequence that originated recently, and are rare in a population, are called mutations.

- Variations in DNA sequence that co-exist in a population, and neither one can be meaningfully defined as wild type, are called polymorphisms.

- Mutations may either occur spontaneously, or may be induced by exposure to mutagens.

- Mutations may result in substitutions, deletions, insertions or chromosomal rearrangements.

- 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. Occasionally metabolites within a cell can catalyze spontaneous mutations.

- Transposable elements are dynamic, abundant components of eukaryotic genomes and important forces in evolution.

- Induced mutations result from mispairing, DNA damage, or sequence interruptions caused by chemical, or physical mutagens.

- DNA repair systems can fail through excessive mutations so the cell cannot cope, or by loss of function of the DNA repair genes themselves.

- When DNA repair systems fail, a cell will senesce, undergo apoptosis or become a cancer.

Key terms:

mutation	mispairing	LINEs
mutant	strand slippage	SINEs
wild type	loop	Alu
polymorphism	mutation hot spot	copy-and-paste
lesion	short-sequence repeats (SSRS)	cut-and-paste
deletion	microsatellites	alkylating agent
insertion	trinucleotide repeat diseases	intercalating agent
substitution	depurination	EMS
spontaneous	apurinic site	benzopyrene
induced	deamination	carcinogenic
mutagen	oxidative damage	ethidium bromide
codon	transposon-tagging	aflatoxin
frame shift	P element	AFB1
premature stop codon	T-DNA	base analog
truncated	transposable elements (TEs)	5-bromocuracil
transition	mobile genetic elements	UV light
transversion	Class I TE	pyrimidine dimer
silent mutation	retrotransposon	DNA repair systems
missense mutation	integrase	senescence
conservative	Class II TE	apoptosis
non-conservative	transposon	cancer
nonsense mutation	transposase	DNA damage threshold
translocation	reverse transcriptase	error-prone DNA repair
inversion	non-autonomous	translesion polymerases
DNA replication error	autonomous	Xeroderma pigmentosum

STUDY QUESTIONS:

1) How are polymorphisms and mutations alike? How are they different?
2) What are some of the ways a substitution can occur in a DNA sequence?
3) What are some of the ways a deletion can occur in a DNA sequence?
4) What are all of the ways an insertion can occur in a DNA sequence?
5) In the context of this chapter, explain the health hazards of smoking tobacco.
6) How was the first mutation in the *white* gene of Drosophila, w^1, caused? (See Chapter 10).
7) Which types of transposable elements are transcribed?

CHAPTER 12 – MUTATIONS: CONSEQUENCES

Figure 1.
A breed of cat, the Canadian Sphynx, lack hair due to a genetic mutation. The Sphynx breed originated in Minnesota, but the Canadian Sphynx line was started in Toronto in 1966 through a selective breeding program from a spontaneous mutation that gave naked kittens. This mutation is inherited in an autosomal recessive manner for the *hairlessness* gene. (Flickr- Weimar Meneses -CC BY 2.0)

INTRODUCTION

When we think of the word "mutation", we automatically think of it as something negative or detrimental. However, a mutation, which is a change in the DNA sequence, may have one or more effects on an organism, depending on what it is and in which gene it occurs. While detrimental effects are most common, sometimes mutations can create new features. These mutations give us a tool with which to investigate the gene and the biological processes in which it is involved. In this chapter we will first take a look at how scientists perform genetic screening for mutations, and the various consequences of those mutation.

1. GENETIC SCREENING FOR MUTATIONS: FORWARD GENETICS, REVERSE GENETICS

Forward genetic screening refers to the process of finding the gene or genes responsible for a certain phenotype or biochemical process. 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 is the strategy of **mutant screening,** and 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 (yes Drosophila can learn). Mutants lack the ability to associate a particular odor with an electric shock. Because of the similarity of biology among all organisms (common descent), 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.

On the other hand, **reverse genetic screening** refers to the process of creating a mutation in a gene, then identifying the phenotypic consequences of that specific mutant gene on the organism. This method is becoming more useful with the advent of whole genome sequencing. Here, we have identified the gene sequences, but are unsure of what each gene does.

1.1. 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 newly induced 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 show a phenotype can be examined for the presence of novel traits. 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 to connect the gene function with the external appearance.

2. SOME MUTATIONS MAY NOT HAVE DETECTABLE MUTANT PHENOTYPES

Not all DNA sequence changes result in mutant phenotypes. Various reasons are described below.

2.1. SILENT CHANGES

After mutagen treatment, the vast majority of base pair changes (especially substitutions) have no obvious 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 **intergenic regions** (between genes) or within an intron where the sequence does not code for protein and is not essential for proper mRNA splicing. Also, even if the change affects the coding region, it may not alter the amino acid sequence

(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 does not quantitatively or qualitatively alter the function of the product, so no phenotypic change would occur.

2.2. ENVIRONMENT AND GENETIC REDUNDANCY

There are 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 visible phenotypic change can be due to **environmental effects**: the loss of that gene product may not be apparent in that specific environment, but might in another. An example is an auxotrophic mutant on complete medium. Conversely, researchers can alter the environment to reveal such mutants (e.g. auxotrophs on minimal media).

Alternatively, the lack of a phenotype might be attributed to genetic **redundancy**. That is. the mutant gene's lost function is compensated by another gene, at another locus, encoding a similarly functioning product. Thus, the loss of one gene is compensated by the presence of another. The concept of genetic redundancy is an important consideration in genetic screens. A gene whose function can be compensated for my another gene, cannot be easily identified in a genetic screen for loss of function mutations.

2.3. ESSENTIAL GENES AND LETHAL ALLELES

Some mutant maybe required to reach a particular developmental stage before the phenotype can be seen or 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 to the adult stage. However, some mutant organisms may not develop sufficiently to reach a stage that can be scored for a particular phenotype. Mutations in **essential genes** create recessive lethal alleles that arrest or derail the development of an individual at an immature (embryonic, larval, or pupal) 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. Nevertheless, the study of recessive lethal mutations (those in essential genes) has elucidated many important biochemical pathways.

An example is the identification of whole classes of genes involved in early embryonic development. Three Drosophila geneticists, Eric Wieschaus, Edward Lewis, and Christiane Nüsslein-Volhard, who were awarded a Nobel Prize (1995), identified pair-rule, gap, and segment polarity genes that have corresponding homologs in all segmented organisms, including humans.

2.4. Naming genes

Many genes are 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 required to make red eyes. We now know its product is a protein that imports a colourless 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.).

3. Example of Human Mutations

3.1. Cystic Fibrosis (CF) – autosomal recessive
Cystic fibrosis (CF) is one of many diseases that geneticists have shown to be primarily caused by mutation in a single, well-characterized gene. Cystic fibrosis is the most common (1/2,500) life-limiting autosomal recessive disease among people of European heritage, with ~ 1 in 25 people being carriers. The frequency varies in different populations. Most of the deaths caused by CF are the result of lung disease, but many CF patients also suffer from other disorders including infertility and gastrointestinal disease. The disease is due to a mutation in the **CFTR** (Cystic Fibrosis Transmembrane Conductance Regulator) gene, which was first identified by **Lap-chee Tsui**'s group at the University of Toronto.
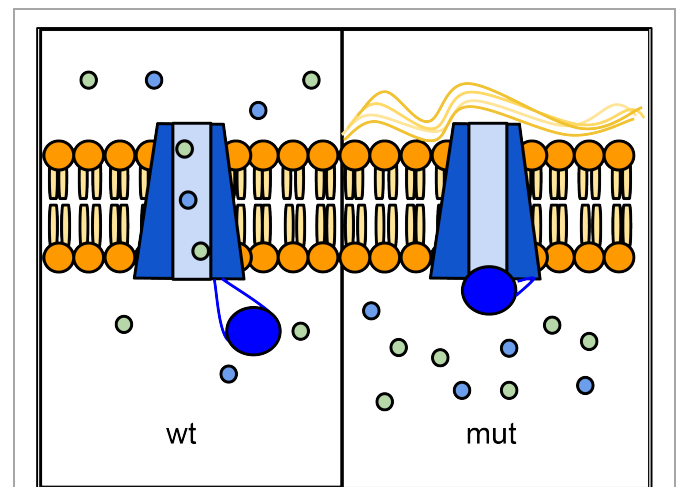
Figure 2.
Wild-type and mutant forms of CFTR in the cell membrane. In wild-type, the CFTR ion channel is gated; when activated by ATP, the channel opens and allows ions to move across the membrane. In some *CFTR* mutants, the channel does not open. This prevents the movement of ions and water and allows mucus to build up on the lung epithelium. (Wikipedia- Lbudd14- CC BY-SA 3.0--modified)

Epithelial tissues in some organs rely on the CFTR protein to transport ions (especially Cl⁻) across their cell membranes. The passage of ions through a six-sided channel is gated by another part of the CFTR protein, which binds to ATP. If there is insufficient activity of CFTR, an imbalance in ion concentration

results, which disrupts the properties of the liquid layer that normally forms on the epithelial surface. In the lungs, this causes mucus to accumulate and can lead to infection. Defects in CFTR also affect pancreas, liver, intestines, and sweat glands, all of which need this ion transport. CFTR is also expressed at high levels in the salivary gland and bladder, but defects in CFTR function do not cause problems in these organs, probably because other ion transporters are able to compensate.

Over one thousand different mutant alleles of *CFTR* have been described. Any mutation that prevents CFTR from sufficiently transporting ions can lead to cystic fibrosis (CF). Worldwide, the most common *CFTR* allele among CF patients is called **ΔF508** (delta-F508; or PHE508DEL), which is a deletion of three nucleotides that eliminates a phenylalanine from position 508 of the 1480 aa wild-type protein. Mutation *ΔF508* causes CFTR to be folded improperly in the endoplasmic reticulum (ER), which then prevents CFTR from reaching the cell membrane. *ΔF508* accounts for approximately 70% of CF cases in North America, with ~1/25 people of European descent being carriers. The high frequency of the *ΔF508* allele has led to speculation that it may confer some selective advantage to heterozygotes, perhaps by reducing dehydration during cholera epidemics, or by reducing susceptibility to certain pathogens that bind to epithelial membranes.

CFTR is also notable because it is one of the well-characterized genetic diseases for which a drug has been developed that compensates for the effects of a specific mutation. The drug, **Kalydeco (Ivacaftor)**, was approved by the FDA and Health Canada in 2012, decades after the *CFTR* gene was first mapped to DNA markers (in 1985) and cloned (in 1989). Kalydeco is effective on only some *CFTR* mutations, most notably *G551D* (i.e. where glycine is substituted by aspartic acid at position 551 of the protein; *GLY551ASP*). This mutation is found in less than 5% of CF patients. The *G551D* mutation affects the ability of ATP to bind to CFTR and open the channel it for transport. Kalydeco compensates for this mutation by binding to CFTR and holding it in an open conformation. Kalydeco is expected to cost approximately $250,000 per patient per year.

_____

SUMMARY:

- Forward genetic screening aims to find the molecular basis for a certain phenotype whereas reverse genetic screening aims to find the phenotypic effects that a gene might have on the organism.

- Somatic mutations occur in non-reproductive cells which affect the current individual, while germline mutations occur in the gametes which affect future generations and not the individual.

- Mutation can alter a gene into different levels and types of expression.

- Not all base pair changes (mutations) cause detectable changes in an organism. The efficiency of mutant screening is limited by silent mutations, redundancy, and embryonic lethality.

- Cystic Fibrosis is a genetic disease caused by the mutation in the CTFR gene.

KEY TERMS:

mutant screen
loss-of-function
gain-of-function
null
dominant negative
somatic cells
germline cells
silent mutation
inter-genic region
redundancy
essential gene

recessive lethal allele
double strand break
non-homologous end joining
DNA repair system
chromosome rearrangement
CFTR
Cystic Fibrosis (CF)
ΔF508(PHE508DEL)
Kalydeco

STUDY QUESTIONS:

1) 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?
2) 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?
3) 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) 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 13 – ALLELES AT A SINGLE LOCUS

Figure 1.
A flower called Camellia showing co-dominance of the red and white alleles of flower colour.
(Flickr- darwin cruz-CC BY 2.0)

INTRODUCTION

The previous chapter described the consequences of mutations. We will now use the mutant forms of a gene to investigate the interactions of alleles at a single locus. This will begin with the difference between somatic and germ line mutations. Then it will deal with simple dominance/recessive relationships, which many students have encountered before. It will end with more sophisticated interactions that can be described by "Muller's Morphs", which deal with the interrelationships of mutant and wild type alleles at a more detailed level.

1. TERMINOLOGY

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

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

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

Figure 2.
Relationship between genotype and phenotype for an allele that is completely dominant to another allele. (Original-M. Deyholos -CC BY-NC 3.0)
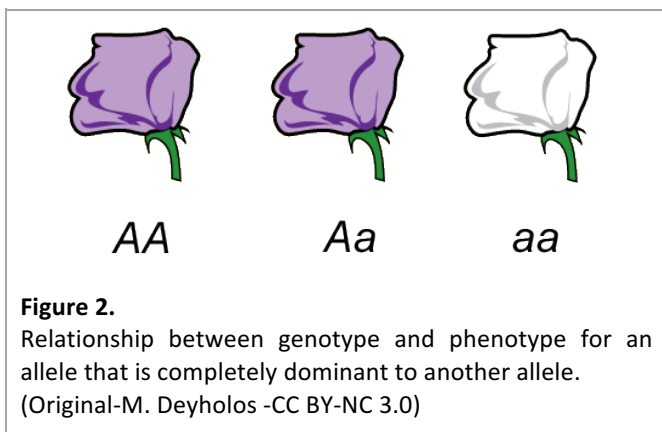
2. SOMATIC VS. GERMLINE MUTATIONS

A mutation occurs in the DNA of a single cell. In single-cell organisms, that mutation is passed on directly to its descendants, typically through the process of mitosis. In multicellular animals, there is a partitioning early in development into **somatic cells**, which form the body cells, and **germline cells**, which form the gametes for the next generation. Mutations may be passed on to somatic cells via mitosis and to gametes via meiosis. In plants, this somatic/germline separation occurs later, in the cells that form the flower.

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

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

Figure 3.
Patch of brown eye colour in a green eye. (Wikipedia-Sheila.lorquiana-CC BY-SA 3.0)
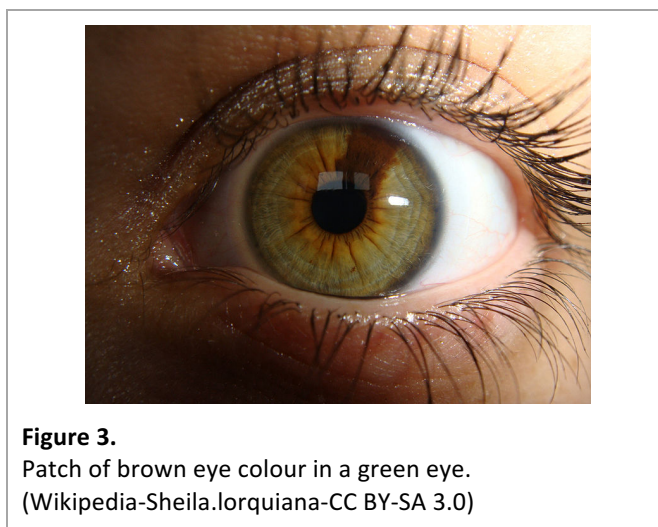
In animals, somatic cells are segregated from germ line cells. In plants, somatic cells become germline cells; so somatic mutations can become germline mutations.

2.3. HAPLOID VS. DIPLOID ORGANISMS
Haploid organisms, have only one copy of a gene, thus a mutation will directly affect the organism's phenotype. Therefore, the phenotype can be used to directly infer the genotype of the organism.

However, in diploid organisms, there are two copies of each gene. The phenotype depends upon an interaction between the two alleles. Thus, any mutation may not have a direct impact on the organism's phenotype. The interaction of the two alleles can show complete dominance, incomplete dominance, co-dominance, or recessiveness. Therefore, inferring the genotype based upon its phenotype is not as simple as in diploids.

3. ALLELES: HETERO-, HOMO-, HEMIZYGOSITY

Mendel's First Law (segregation of alleles) is especially remarkable because he made his observations and conclusions (1865) without knowing about the relationships between genes, chromosomes, and DNA. We now know the reason why more than one allele of a gene can be present in an individual: most eukaryotic organisms are diploid and have at least two sets of homologous chromosomes. For organisms that are predominantly diploid, such as humans or Mendel's peas, chromosomes exist as pairs, with one copy

inherited from each parent. Diploid cells therefore can contain two different alleles of each gene, with one allele part of each member of a pair of homologous chromosomes. If both alleles of a particular gene are the same (indistinguishable), the individual is said to be **homozygous** at that gene or locus. On the other hand, if the alleles are different (can be distinguished) from each other, the genotype is **heterozygous**. In cases where there is only one copy of a gene present, for example if there is a deletion of the locus on the homologous chromosome, we use the term **hemizygous**. In another example is single X-chromosome in X/Y males were almost all the loci on that chromosome are hemizygous. (The exception is the pseudo-autosomal region – see the chapter on sex chromosomes.)

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

4. PLEIOTROPY AND POLYGENIC INHERITANCE

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

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

5. COMPLETE DOMINANCE AND RECESSIVE

An example of a simple phenotype is flower color in Mendel's peas. We have already said that one allele as a homozygote produces purple flowers, while the other allele as a homozygote produces white flowers (**Figure 2**). But what about a heterozygous individual that has one purple allele and one white allele? What is the phenotype of a heterozygote?

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. The dominant/recessive character is a relationship between two alleles and must be determined by observation of the heterozygote phenotype.

Sometimes, to represent this relationship, a dominant allele will be written as a capital letter (e.g. *A*) while a recessive allele will be written in lower case (e.g. *a*). However, this is not the only system. Many different systems of genetic symbols are in use. The most common are shown in Table 3.1. Also note that genotypes (alleles) are usually written in *italics* and chromosomes and proteins are not. For example, the *white* gene in *Drosophila melanogaster* on the X chromosome encodes a protein called WHITE, which is a pigment precursor transmembrane transporter enzyme.

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

Examples	Interpretation
A and *a*	Uppercase letters represent dominant alleles and lowercase letters indicate recessive alleles. Mendel invented this system but it is not commonly used because not all alleles show complete dominance and many genes have more than two alleles.
a^+ and a^1	Superscripts or subscripts are used to indicate alleles. For wild type alleles the symbol is a superscript +.
AA or *A/A*	Sometimes a forward slash is used to indicate that the two symbols are alleles of the same gene locus, but on homologous chromosomes.
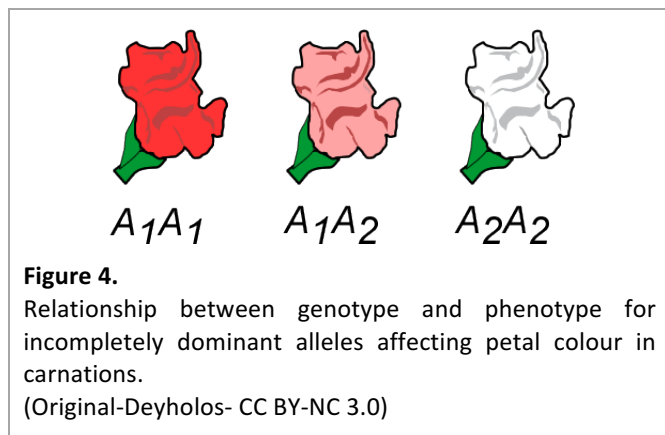
Figure 4.
Relationship between genotype and phenotype for incompletely dominant alleles affecting petal colour in carnations.
(Original-Deyholos- CC BY-NC 3.0)
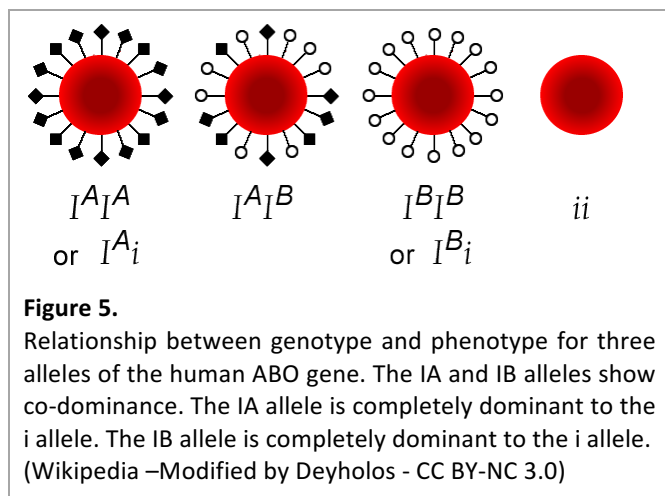
Figure 5.
Relationship between genotype and phenotype for three alleles of the human ABO gene. The IA and IB alleles show co-dominance. The IA allele is completely dominant to the i allele. The IB allele is completely dominant to the i allele.
(Wikipedia –Modified by Deyholos - CC BY-NC 3.0)

6. INCOMPLETE DOMINANCE

Besides the complete dominant and recessive relationship, other relationships can exist between alleles. In **incomplete dominance** (also called **semi-dominance**), both alleles affect the trait additively, and the phenotype of the heterozygote shows a typically intermediate between the homozygotes, which is often referred to as blended phenotype. For example, alleles for color in carnation flowers (and many other species) exhibit incomplete dominance. Plants with an allele for red petals (A_1) and an allele for white petals (A_2) have pink petals. We say that the A_1 and the A_2 alleles show incomplete dominance because neither allele is completely dominant over the other (**Figure 4**).

7. CO-DOMINANCE

Co-dominance is another type of allelic relationship in which a heterozygous individual expresses the phenotype of <u>both</u> alleles simultaneously. An example of co-dominance is found within the **ABO blood group** of humans. The *ABO* gene has three common alleles that were named (for historical reasons) I^A, I^B, and *i*. People homozygous for I^A or I^B display only A or B type antigens, respectively, on the surface of their blood cells, and therefore have either type A or type B blood (**Figure 5**). Heterozygous $I^A I^B$ individuals have both A and B antigens on their cells, and so have type AB blood. Note that the heterozygote expresses both alleles simultaneously, and is not some kind of novel intermediate between A and B. 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 thus is recessive to the other alleles. I^A/*i* or I^B/*i* individuals display only A or B antigens, respectively. People homozygous for the *i* allele have type O blood.

This is a useful reminder that different types of dominance relationships can exist, even for alleles of the same gene. Many types of molecular markers, which we will discuss in a later chapter, display a co-dominant relationship among alleles.

Another example of co-dominance is shown in the first figure of this chapter – flower colour in *Camellia sp*.

8. 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 **haplo-sufficient**. 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 **haplo-insufficient** 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.

9. MUTANT CLASSIFICATION

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

9.2. LETHAL
A lethal mutation causes the premature death of an organism. For example, in Drosophila lethal mutations can result in the death during the embryonic, larval, or pupal stage. Lethal mutations

are usually recessive, so both copies of a gene have to be lost for the premature death to occur (homozygous lethal alleles will not be viable). Heterozygotes which have one lethal allele and one wild type allele are typically viable.

9.3. BIOCHEMICAL
Auxotrophic mutants can be derived from prototrophic parents. This type of mutation blocks a step in a biochemical pathway as discussed for the *arg-* mutants of Beadle and Tatum in the chapter on biochemical pathways. Such biochemical mutations are a specific type of the conditional mutation class (next).

9.4. CONDITIONAL
Conditional mutations rely on the concept of: phenotype = genotype + environment + interaction. Organisms with this kind of mutation express a mutant phenotype, but only under specific environmental conditions. Under **restrictive conditions**, they express the mutant phenotype while under **permissive conditions**, they show a wild type phenotype. One example of a conditional mutation is the temperature-sensitive pigmentation of Siamese cats. Siamese cats have temperature sensitive fur colour; their fur appears unpigmented (light coloured) when grown in a, warm temperature environment. The hair appears pigmented (dark) when grown at a cooler temperature. This is seen at the peripheral regions of the feet, snout, and ears (**Figure 6**). This is because in warm temperature, the enzyme that is needed for melanin pigment synthesis becomes nonfunctional. However, in cooler temperature, the enzyme needed for melanin synthesis is functional and the deposition of melanin makes the fur look dark.

Figure 6.
Siamese cats have temperature sensitive pigmentation due to genetic mutation. (Wikimedia-Telekokopelli-CC BY-SA 3.0)

10. MULLER'S MORPHS

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

Thus, mutations (changes in a gene sequence) can result in mutant alleles that no longer produce the same <u>level</u> or <u>type</u> 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.

10.1. AMORPH

Amorphic alleles have a complete loss-of-function. They make <u>no active product</u> – zero function. They are known as a "**Null**" mutation or a "loss-of-function" mutation.

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

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

(2) Gene is present, but is not transcribed because of a sequence change in the promoter or enhancer/regulatory elements.

(3) Gene is present but the transcript is aberrantly processed. There is normal transcription but

base pair changes cause the mature mRNA to incorrectly splice introns, therefore the translated amino acid sequence would be altered and nonfunctional.

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

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

Genetic/phenotypic explanation - Amorphic mutations of most genes usually act as recessive to wild type (case #1). However, with some genes the amorphic mutations are dominant to wild type. (case #2).

case #1: *white* gene in Drosophila	
w^+/w^+	wildtype and red eyed
w^+/w^-	wildtype and red eyed
w^-/w^-	mutant and white eyed
case #2: *Minute* locus in Drosophila	
M^+/M^+	wildtype and long bristeld
M^+/M^-	mutant and short bristled
M^-/M^-	dead, recessive lethal

For the *Minute* gene, we concluded that the organism needs both copies to have a wild type phenotype. Loss of one copy (an amorphic mutation) produces a dominant visible mutant phenotype. Deletion of the gene is an example of a classic amorphic mutation.

10.2. HYPOMORPH

Hypomorphic alleles show only a <u>partial</u> loss-of-function. These alleles are sometimes referred to as "**leaky**" mutations, because they provide some function, but not complete, normal function.

Molecular explanation - Changes in the DNA base pair sequence of the hypomorphic allele may cause

one or more of the following, with gene still being present:

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

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

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

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

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

$white^{apricot}$ allele in Drosophila	
w^+/w^+	wildtype and red eyed
w^+/w^a	wildtype and red eyed
w^a/w^a	mutant and apricot eye colour

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

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

10.3. HYPERMORPH
Hypermorphic alleles produce quantitatively more of the same, active product.

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

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

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

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

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

10.4. NEOMORPH
Neomorphic alleles produce a product with a new, different function, something that the wild type allele does not do.

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

(1) new transcription – changed DNA sequence in the promoter or enhancer/regulatory elements that makes new transcription either temporally or in a tissue-specific manner.

(2) new function protein product – normal transcription, processing, translation but base pair changes alter certain amino acids in the polypeptide sequence so it acquires a new function (activity) that is different from the normal function (e.g. additional substrate or new binding site).

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

10.5. ANTIMORPH

Antimorphic alleles are relatively rare, and have a new activity that is dominant and opposite to the wildtype function. These alleles usually interfere with the function from the wild type allele. (They often lose their normal function as well.) The new function works against the normal expression of the wild type allele. This can happen at the transcriptional, translational, or later level of expression. Thus, when an antimorphic allele is heterozygous with wild type, the wild type allele function is reduced or prevented. At the molecular level, there are many ways this can happen. The simplest model to explain an antimorphic effect is that the protein acts as a dimer (or any multimer) and the inclusion of a mutant subunit poisons the whole complex, thereby preventing or reducing its level of function. Antimorphs are also known as **dominant-negative** mutations because they are usually dominant and act negatively against the wild type function.

10.6. IDENTIFYING MULLER'S MORPHS

All mutations can be sorted into one of the five morphs base on how they behave when heterozygous with three other standard alleles (**Figure 7**): (1) deletion alleles (zero function), (2) wild type alleles (normal function), and (3) duplication alleles (double normal function).
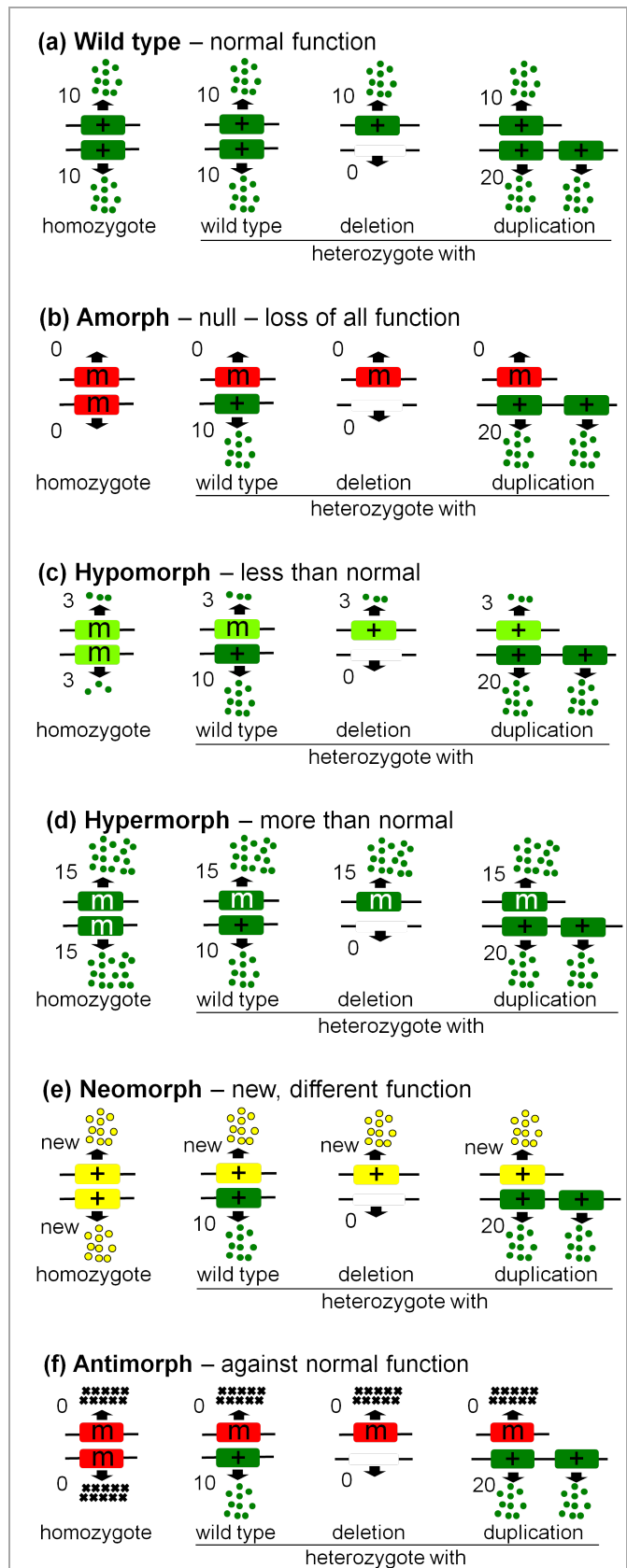
Figure 7.
Five classes of mutants designated as morphs (forms) by a Nobel prize winner, H.J. Muller, which are known as Muller's Morphs. (Original-Locke- CC BY-NC 3.0)

_____

SUMMARY:

- Symbols are used to denote the alleles, or genotype, of a locus.

- Phenotype depends on the alleles that are present, their dominance relationships, and sometimes also interactions with the environment and other factors.

- A somatic mutation affects the individual but not the progeny, whereas a germline mutation affects the progeny in the next generation but not the individual in which they occur.

- In a diploid organism, alleles can be homozygous, heterozygous or hemizygous.

- Allelic interactions at a locus can be described as dominant vs. recessive, incomplete dominance, or co-dominance.

- Muller's morphs classify all types of mutations including: amorph, hypomorph, hypermorph, neomorph, and antimorph.

KEY TERMS:

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

STUDY QUESTIONS:

1) Distinguish amongst the following terms: (1) gene, (2) locus, (3) allele, (4) transcription unit.
2) A flower geneticist crosses a red flowered diploid plant with a white flower diploid plant and all the progeny are red. Use two different forms of symbols to show this cross and its progeny. What if all the progeny were pink?
3) If your blood type is B, what are the possible genotypes of your parents at the locus that controls the ABO blood types?
4) 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:
   haplo-sufficiency,
   haplo-insufficiency,
   pleiotropy,
   incomplete dominance,
   co-dominance,
   incomplete penetrance,
   broad (variable) expressivity.
5) In this hypothetical example of Drosophila bristle mutations, when various, true-breeding mutant strains (all at a single locus) are crossed to a wild type strain the following phenotypes are observed in the progeny:
   Mutant#1 = bristles 20% shorter
   Mutant#2 = bristles 30% longer
   Mutant#3 = bristles 50% shorter
   Mutant#4 = bristles kinked and misshapen
   Mutant#5 = bristles are missing
   What is the best characterization, using Muller's Morphs, for each?

Table for Question 2

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

CHAPTER 14 – MITOSIS AND THE CELL CYCLE
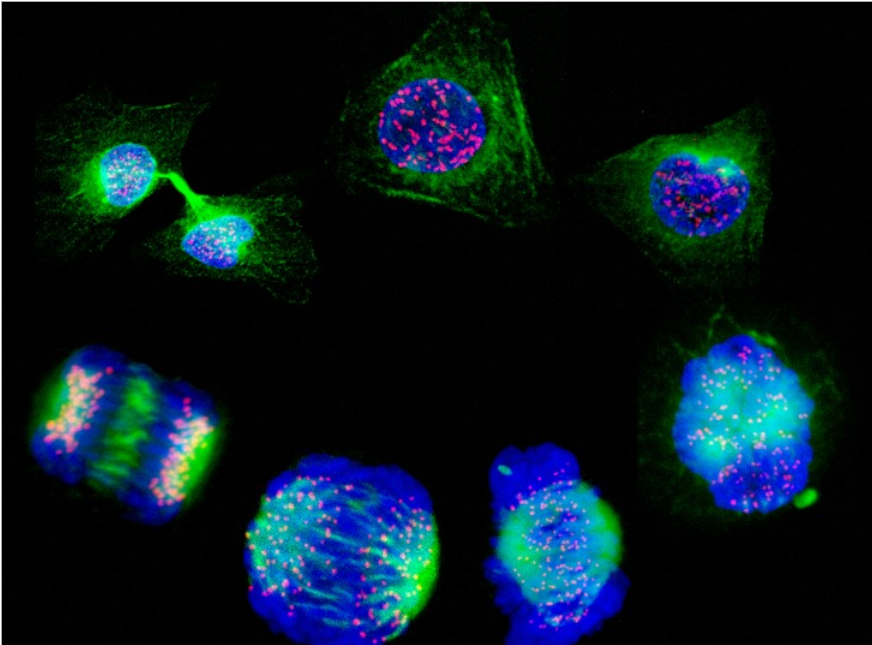
Figure 1.
Confocal micrograph of human cells showing the stages of cell division. DNA is stained blue, microtubules stained green and kinetochores stained pink. Starting from the top and going clockwise you see an interphase cell with DNA in the nucleus. In the next cell, the nucleus dissolves and chromosomes condense in prophase. The next is prometaphase where microtubules are starting to attach, but the chromosomes haven't aligned. Next is metaphase where the chromosomes are all attached to microtubules and aligned on the metaphase plate. The next two are early and late anaphase, as the chromosomes start separating to their respective poles. Finally there is telophase where the cells are completing division to be two daughter cells.
(Flickr-M. Daniels; Wellcome Images- CC BY-NC-ND 2.0)

INTRODUCTION

Cell growth and division is essential to asexual reproduction and the development of multicellular organisms. The transmission of genetic information is accomplished in a cellular process called **Mitosis**. This process ensures that a cell division with each daughter cell inheriting identical genetic material, i.e. exactly one copy of each chromosome present in the parental cell.

1. FOUR STAGES OF A TYPICAL CELL CYCLE

The life cycle of eukaryotic cells can generally be divided into four stages (and a typical cell cycle is shown in **Figure 1**). When a cell is produced through fertilization or cell division it normally goes through four main stages: G_1, S, G_2, and M. The first stage of **interphase** is a lag period is called Gap 1 (G_1), and is the first part of **interphase**. This is where the cell does its normal cellular functions and it grows in size, particularly after mitosis when the daughters are half the size of the mother cell. This stage ends with the onset of the DNA synthesis (**S**) phase, during which each chromosome is replicated (For more information on DNA replication, see the chapter on DNA and chromosome replication). Though the

chromosomes are not condensed yet, because S phase is still part of interphase, they are replicated as two sister chromatids attached at the centromere. Still in interphase and following replication, there is another lag phase, called Gap 2 (G_2). In G_2, the cell continues to grow and acquire the proteins necessary for cell division. There is a checkpoint stage, where, if there are any problems with replication or acquiring the needed proteins, the cell cycle will arrest until it can fix itself or die. The final stage is mitosis (**M**), where the cell undergoes cell division as is described in the last section.

Many variants of this generalized cell cycle also exist. Cells undergoing meiosis do not usually have a G_2 phase. Cells, like hematopoietic stem cells, which are found in the bone marrow and produce all the other blood cells, will consistently go through these phases as they are constantly replicating. Other cells, as in the nervous system, will no longer divide. These 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, like the larval tissues in *Drosophila,* undergo many rounds of DNA synthesis (S) without any mitosis or cell division, leading to

endoreduplication (See Chapter 2). Understanding the control of the cell cycle is an active area of research, particularly because of the relationship between cell division and cancer.

2. MITOSIS

During the S-phase of interphase the chromosomes replicate so that each chromosome has two sister **chromatids** attached at the centromere. After S-phase and G2, the cell enters Mitosis. The first step in mitosis is **prophase** where the nucleus dissolves and the replicated chromosomes condense into the visible structures we associate with chromosomes. Next is **metaphase**, where the **microtubules**
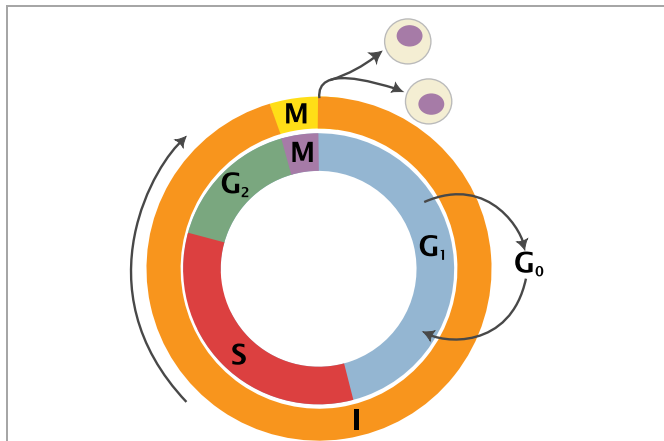
Figure 2.
Stages of the cell cycle. The outer ring identifies when a cell is in interphase (I) and when it is in mitosis (M). The inner ring identifies the four major stages. Cells can enter G_0 if they are not actively undergoing cell division, and may re-enter the cell cycle at a later time.
(Wikimedia Commons - R. Wheeler - CC BY-SA 3.0

attach to the **kinetochore** and the chromosomes align along the middle of the dividing cell, known as the **metaphase plate**. The kinetochore is the region on the chromosome where the microtubules attach. It contains the centromere and proteins that help the microtubules bind. Then in **anaphase**, each of the sister chromatids from each chromosome gets pulled towards opposite poles of the dividing cell. Finally in **telophase**, identical sets of **unreplicated chromosomes** (single chromatids) are completely separated from each other into the two daughter cells, and the nucleus re-forms around each of the two sets of chromosomes. Following this is the partitioning of the cytoplasm (**cytokinesis**) to complete the process and to make two identical daughter cells. **Figure 1** and **Figure 3** show real pictures and a cartoon schematic of the process, respectively.

You should note that this is a dynamic and ongoing process, and cells don't just jump from one stage to the next. When looking at snapshots of real cells, you will more often see cells between two stages, like is seen in some of the images in **Figure 1**.

An acronym to remember the main stages of mitosis is iPMAT, where the little i stands for interphase, which will be described next.

In contrast, Meiosis, which may appear similar, is a very different process. Read through the Chapter 16 and try to identify the similarities and differences between the two processes.

Interphase   Prophase   Metaphase   Anaphase   Telophase   Cytokinesis
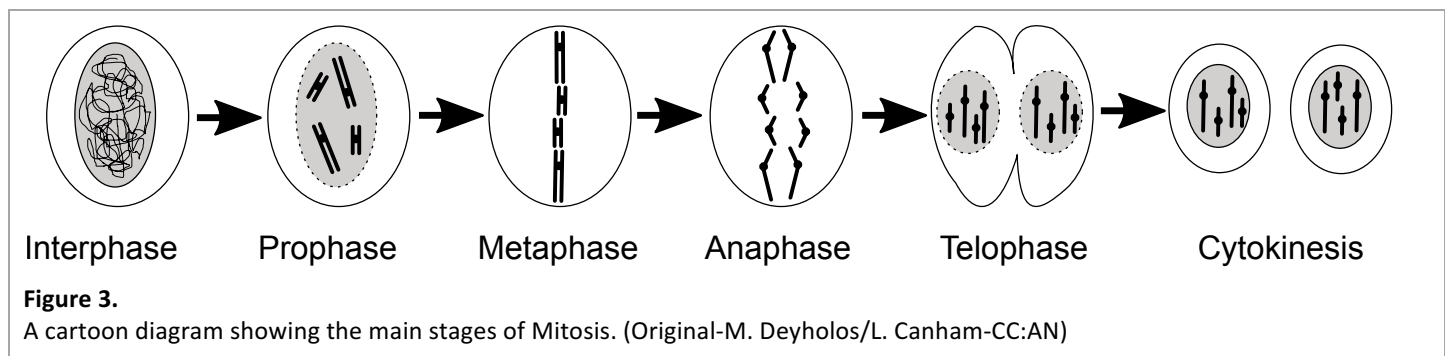
Figure 3.
A cartoon diagram showing the main stages of Mitosis. (Original-M. Deyholos/L. Canham-CC:AN)

1N, 1C **2N, 2C** **2N, 2C** **2N, 4C** **2N, 4C** **2N, 2C**

fertilization gap (G₁) synthesis (S) gap (G₂) mitosis

Figure 4.
Changes in DNA and chromosome content during the cell cycle and mitosis. For simplicity, nuclear membranes are not shown, and all chromosomes are represented in a similar stage of condensation. (Original-M. Deyholos/L. Canham- CC BY-NC 3.0)
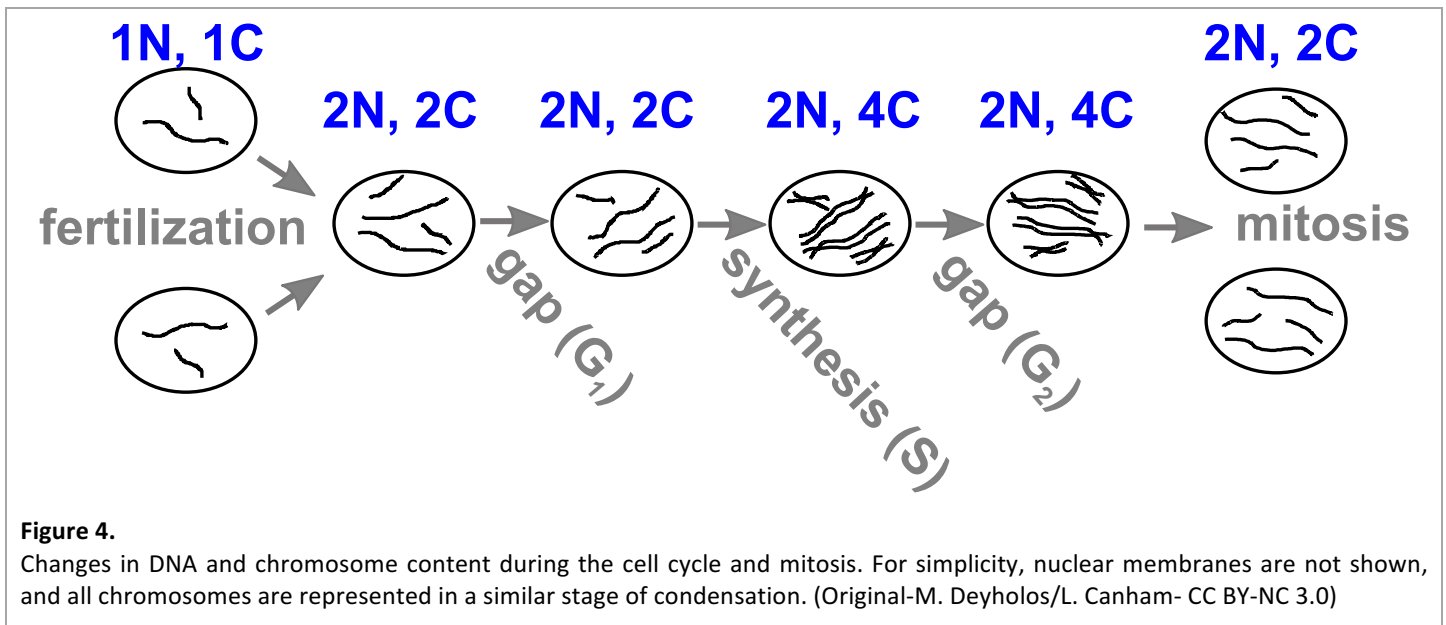
3. MEASURES OF DNA CONTENT AND CHROMOSOME CONTENT

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

3.1. 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
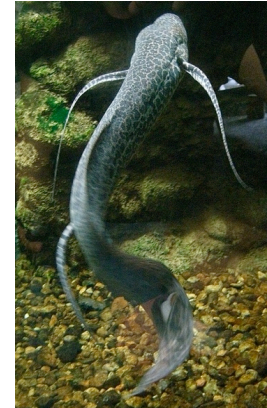
Figure 5.
Marbled Lungfish (*Protopterus aethiopicus)* has a genome of ~133 x 10⁹ base pairs, which is ~45X that of a human. It is an example of the C-value paradox.
(Wikipedia-OpenCage- CC BY 2.5)

physical size or complexity of an organism. Compare the size of *E. coli* and humans for example in the **Table 1**. There are, however, many exceptions to this generalization, such as the human genome contains only 3.2×10^9 DNA bases, while the wheat genome contains 17×10^9 DNA bases, almost 6 times as much. The Marbled Lungfish (*Protopterus aethiopicus* - **Figure 5**) 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 670×10^9 bases (200x a human).

3.2. 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 Chapter 30. Other non-gene DNA includes short, highly repetitive sequences of various types. Together, this non-functional DNA is often referred to as "Junk DNA".

3.3. THE "ONION TEST".

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

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

	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.
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."

_____

SUMMARY:

- The asexual transmission of genetic information is accomplished in a process called Mitosis.

- The process of mitosis can be divided into Prophase, Metaphase, Anaphase, and Telophase.

- Mitosis reduces the c-number, but not the n-number of the daughter cells.

- Not all the DNA in an organism codes for genes. In most higher eukaryotes most DNA is non-gene coding and appears to have no specific function and is called "junk' DNA.

- The c-value paradox refers to the observation that the amount of DNA is not necessarily related to the complexity of the organism.

KEY TERMS:

mitosis	metaphase plate
interphase	anaphase
G_1 Phase	telophase
S Phase	unreplicated chromosome
G_2 Phase	cytokinesis
M Phase	n-value
G_0 Phase	c-value
chromatids	replicated chromosome
prophase	nuclear genome
metaphase	c-value paradox
microtubules	
kinetochore	

STUDY QUESTIONS:

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

2) The answer to question 1 implies that not all DNA within a chromosome encodes genes. Can you name any examples of chromosomal regions that contain relatively few genes

3)
    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 no centromeres?

4) For a diploid organism with 2n=16 chromosomes, how many chromosomes and chromatids are present per cell at the end of:
    a) G_1,
    b) S,
    c) G_2,
    d) mitosis,

5) Refer to **Table 1**.
    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 this Table show a correlation with organismal complexity?

Chapter 15 – Human Chromosomes
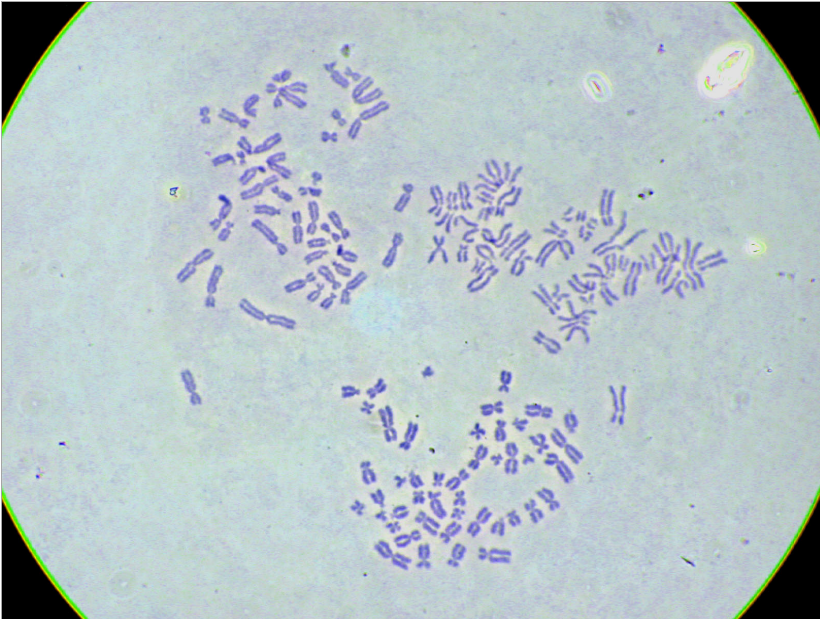
Figure 1.
Human metaphase chromosome spreads. To make these figures, white blood cells in metaphase were dropped onto a slide. The cells burst open and the chromosomes can then be stained with giemsa (a purple colour). This image shows chromosomes from three cells that hit the slide close to one another. They can be distinguished by the size difference among the chromosome sets, which is due to the differences in condensation during the stages of mitosis (prophase).
(Original-Alexander Smith - CC BY-NC 3.0)

INTRODUCTION

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

This chapter will provide information on human chromosomes that will be referred to in various other chapters, lectures, and in the lab.

1. METAPHASE CHROMOSOME SPREADS

1.1. THE SHAPE OF CHROMOSOMES
Figure 1 shows chromosomes from three cells. Each of the cells was in the metaphase stage of mitosis, which is why the chromosomes appear **replicated** and **condensed**. We refer to chromosomes as being replicated when they consist of two sister chromatids held together at the centromeres. DNA replication occurs during S phase. These chromosomes are also condensed. Chromosomes are compacted at the start of mitosis in prophase. **Cytogeneticists** can observe chromosomes at any stage of the cell cycle but those from metaphase cells provide the most detail

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

Figure 2.
A pair of metacentric human chromosome #1.
(Wikipedia- National Human Genome Research Institute-PD)

1.2. THE AMOUNT OF DNA IN A CELL (C-VALUE)
To calculate how much DNA is seen in the nuclei in **Figure 1**, consider that a human gamete has about 3000 million base pairs. We can shorten this statement to $1c = 3000$ Mb where c is the **c-value**, the DNA content in a gamete. When an egg and sperm join the resulting zygote is $2c = 6000$ Mb. Before the zygote can divide and become two cells it must undergo DNA replication. This doubles the DNA content to $4c = 12\,000$ Mb. When the zygote divides, each daughter cell inherits half the DNA and is therefore back to $2c = 6000$ Mb. Then each

cell will become 4c again (replication) before dividing themselves to become 2c each. From this point forward, every cell in the embryo will be 2c = 6000 Mb before its S phase and 4c = 12 000 Mb afterwards. The same is true for the cells of fetuses, children, and adults. Because the cells used to prepare this chromosome spread were adult cells in metaphase each is 4c = 12 000 Mb. Note, there are some rare exceptions, such as some stages of meiocytes that make germ cells and other rare situations like the polyploidy of terminally differentiated liver cells. In summary:

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

1.3. THE NUMBER OF CHROMOSOMES (N-VALUE)

Human gametes contain 23 chromosomes. We can summarize this statement as 1n = 23 where n is the **n-value**, the number of chromosomes in a gamete. When a 1n = 23 sperm fertilizes a 1n = 23 egg, the zygote will be 2n = 46. But, unlike DNA content (c), the number of chromosomes (n) does *not* change with DNA replication. A replicated chromosome is still just one chromosome. Thus the zygote stays 2n = 46 after S phase. When the zygote divides into two cells both contain 46 chromosomes and are still 2n = 46. Every cell in the embryo, fetus, child, and adult is also 2n = 46 (with the exceptions noted above).

In summary:

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

Note, that in a normal cell, the chromosome number is 2n before and after chromosome

replication. The n-value does not change while the c-value does.
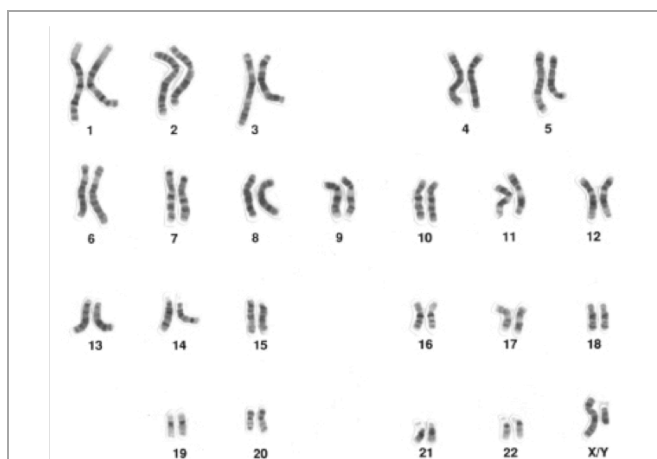
Figure 3.
Karyogram of a normal human male.
(Wikipedia-National Human Genome Research Institute – PD)

2. HUMAN KARYOGRAMS AND KARYOTYPES

2.1. KARYOGRAMS

Human cytogenetists use metaphase chromosome spreads as a standard representation of the chromosomes in a cell, organism, or species. Comparisons permit them to identify chromosome abnormalities. Because it can be hard to distinguish individual chromosomes, cytogeneticists sort the photo to put the chromosomes into a standard pattern. The result is a **karyogram** ("nucleus picture"; **Figure 3**). In the past it was necessary to print a photograph of the metaphase spread, cut out each chromosome with scissors, and then glue each to a piece of cardboard to show the pattern. Now, computer software does much of this for us, but the karyogram assembly is usually reviewed by a qualified cytogeneticist. But either way, the random collection of chromosomes seen in **Figure 1** is converted to the organized pattern in **Figure 3**.

2.2. HUMAN CHROMOSOMES – AUTOSOMES

The chromosomes are numbered to distinguish them. Chromosomes 1 through 22 are **autosomes**, which are present in two copies in both males and females. Because human chromosomes vary in size this was the easiest way to label them. Our largest chromosome is number 1, our next longest is 2,

and so on. The karyogram above shows two copies of each of the autosomes. A karyogram from a normal female would also show these 22 pairs. There are also the sex-chromosomes, X and Y (see below). Normal females have two X-chromosomes, while normal males have an X and a Y each. They act as a homologous pair, similar to the autosomes. During meiosis only one of each autosome pair and one of the sex-chromosomes makes it into the gamete. This is how 2n = 46 adults can produce 1n = 23 eggs or sperm.

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

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

2.3. RELATIONSHIPS BETWEEN CHROMOSOMES AND CHROMATIDS

To summarize what we have covered so far, karyograms depict replicated chromosomes (because the cells had past S phase in the cell cycle) and two copies of each chromosome (because the cells were diploid). So how do we refer to all the pieces of DNA present? **Figure 4** summarizes the terms used.

2.4. HUMAN SEX CHROMOSOMES

Figure 3 shows that most of our chromosomes are present in two copies. Each copy has the same length, centromere location, and banding pattern. As mentioned before, these are called autosomes. However, note that two of the chromosomes, the X and the Y, do not look alike. These are **sex chromosomes**. In mammals, males have one of each while females have two X chromosomes.

Figure 4.
The relationships between chromosomes and chromatids.
(Original Deyholos- CC BY-NC 3.0)
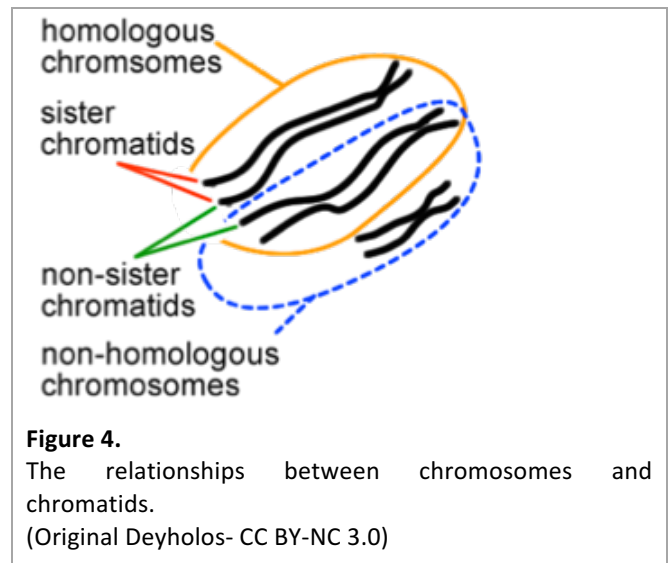
Term	Definition	Example
homologous chromosomes	the maternal and paternal copies of a chromosome	maternal chromosome 1 and paternal chromosome 1
non-homologous chromosomes	two different chromosomes within the same cell/organism	a chromosome 1 and a chromosome 8
sister chromatids	the identical chromatids within a single replicated chromosome	the two chromatids within maternal chromosome 1
non-sister chromatids	the similar but not identical chromatids from homologous chromosomes	a chromatid in maternal chromosome 1 and a chromatid in paternal chromosome 1

Autosomes are those chromosomes present in the same number in males and females while sex chromosomes are those that are not. When sex chromosomes were first discovered their function was unknown and the name X was used to indicate this mystery. The next one was named Y (then Z, and then W – see Chapter 21).

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

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

How do the sex chromosome behave during meiosis? Well, in those individuals with two of the same chromosome (i.e. XX females) the chromosomes pair and segregate during meiosis I the same as autosomes do. During meiosis in XY males the sex chromosomes pair with each other (**Figure 5**). 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 5.
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 BY-NC 3.0)
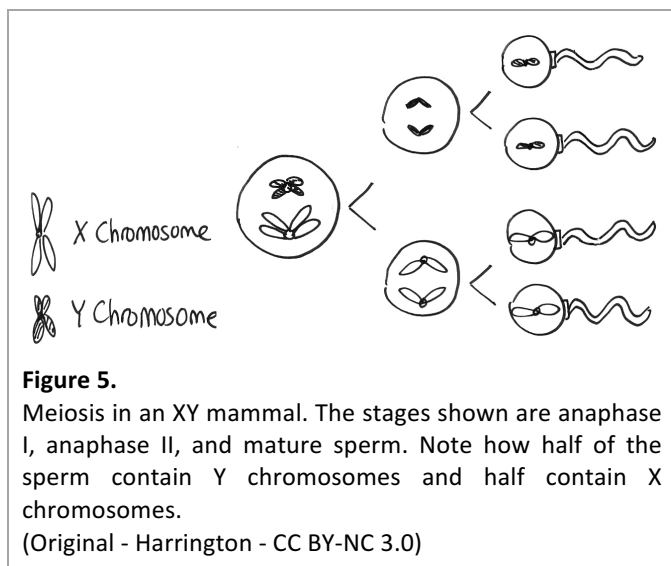
2.5. HUMAN KARYOTYPES

We can summarize the information shown in a karyogram such as **Figure 3** with a written statement known as a **karyotype** ("nucleus features"). By convention we list (i) the total number of chromosomes, (ii) the sex chromosomes, and (iii) any abnormalities. The karyotype in **Figure 3** would be 46,XY, which is typical for human males. Most human females are 46,XX.

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

3. PARTS OF A TYPICAL NUCLEAR CHROMOSOME

A functional chromosome requires four features. These are shown in **Figure 6.**

Figure 6.
Parts of a typical human nuclear chromosome (not to scale). The ori's and genes are distributed everywhere along the chromosome, except for the telomeres and centromere.
(Original-Harrington- CC BY-NC 3.0)
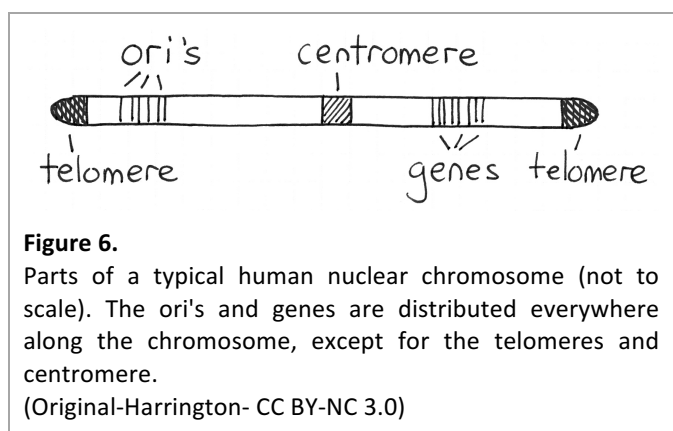
3.1. THOUSANDS OF GENES

In the previous sections we mentioned human chromosome 1, but what exactly is it? Well, each chromosome is long molecule of double stranded DNA. They carry **genetic information** (genes). Chromosome 1, being our largest chromosome has the most genes, about 4778 in total. Many of these

genes are transcribed into mRNAs, which encode proteins. Other genes are transcribed into tRNAs, rRNA, and other non-coding RNA molecules (see Chapter 07).

3.2. ONE CENTROMERE

A **centromere** ("middle part") is a place where proteins attach to the chromosome as required during the cell cycle. Cohesin proteins hold the sister chromatids together beginning in S phase. Kinetochore proteins form attachment points for microtubules during mitosis. The metaphase chromosomes shown in **Figure 3** have both Cohesin and Kinetochore proteins at their centromeres. There are no genes within the centromere region DNA; rather it is composed of a simple repeated DNA sequence.

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

3.3. TWO TELOMERES

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

3.4. THOUSANDS OF ORIGINS OF REPLICATION

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

4. APPEARANCE OF A TYPICAL NUCLEAR CHROMOSOME DURING THE CELL CYCLE

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

The condensation is a more complex story because eukaryotic DNA is always wrapped around some proteins. **Figure 7** shows the different levels commonly found in cells. During interphase, a chromosome exists mostly as a **30 nm fibre**. This allows it to fit inside the nucleus and still have the DNA be accessible for enzymes performing RNA synthesis, DNA replication, and DNA repair. At the start of mitosis these processes halt and the chromosome becomes even more condensed. This is necessary so that the chromosomes are compact enough to move to the opposite ends within the cell. When mitosis is complete the chromosome returns to its 30 nm fibre structure. Recall that each of our cells has a maternal and a paternal chromosome 1. **Figure 8** shows what these chromosomes look like during the cell cycle

Figure 7.
Successive stages of chromosome condensation depend on the introduction of additional proteins. (Wikipedia-R. Wheeler- CC BY-SA 3.0)

Figure 8.
Appearance of maternal and paternal chromosome 1 look like during the cell cycle. The other 44 chromosomes are not shown. Note that they are independent during both interphase (top) and mitosis (bottom). After anaphase there will be two cells in G_1.
(Original-Harrington- CC BY-NC 3.0)

5. DNA IS PACKAGED INTO CHROMATIN

5.1. DNA CAN BE HIGHLY COMPACTED

If stretched to its full length, the DNA molecule of the largest human chromosome would be 85mm long. 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.

5.2. LEVELS OF COMPACTION

There are several different levels of structural organization in eukaryotic chromosomes, with each successive level contributing to the further compaction of DNA (**Figure 7**). 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.

5.3. CHROMATIN PACKAGING VARIES INSIDE THE NUCLEUS: EUCHROMATIN AND HETEROCHROMATIN

Chromosomes can be stained with certain 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 continuous and varied spectrum. **Euchromatin** is more loosely packed, and tends to

contain genes that are actively being transcribed. **Heterochromatin,** is more densely compacted and tends not to be transcribed; the genes are inactive. Heterochromatin sequences also include short, highly-repetitive sequences called **satellite DNA**, which acquired their name because their buoyant density, as determined by ultracentrifugation, is distinctly different from the main band of DNA.

6. PARTS AND APPEARANCE OF A MITOCHONDRIAL CHROMOSOME

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

Each mtDNA has a single origin of replication. During DNA replication two replication forks leave the ori and halt when they bump into each other on the opposite side of the circle. DNA replication inside the mitochondria happens throughout interphase, not once during S phase as with the nuclear chromosomes. The consequence is that each mitochondrion has between 2 to 10 identical copies of its chromosome (**Figure 9**).
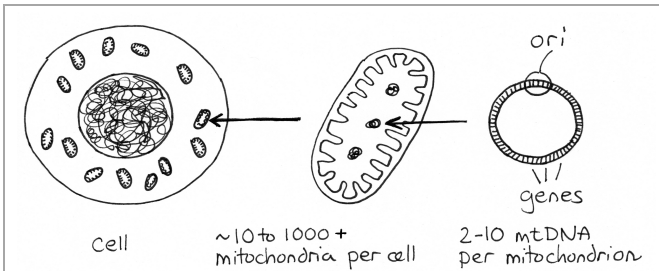
Figure 9.
The relationship between cells, mitochondria, and mitochondrial DNA. (Original-Harrington- CC BY-NC 3.0)

There are other differences when compared to nuclear chromosomes. Organelles such as mitochondria or chloroplasts are likely the remnants of prokaryotic endosymbionts that entered the cytoplasm of ancient progenitors of today's eukaryotes (**endosymbiont theory**). These endosymbionts had their own, circular

chromosomes (**Figure 10**), like most bacteria that exist today. Mitochondria typically have circular chromosomes that behave more like bacterial chromosomes than eukaryotic chromosomes, (i.e. mitochondrial genomes do not undergo mitosis or meiosis). Also, the mitochondrial chromosome is not associated with histones or other proteins that compact it. It also lacks a centromere because mitochondrial replication is simpler than nuclear chromosome replication. Mitochondria just grow larger and split in two, like the cells of its prokaryote origin. Because there are multiple mtDNA copies that are randomly distributed in the matrix, both new mitochondria will end up inheriting some mtDNAs. And lastly because the mtDNA is circular there are no ends and thus no telomeres.

In summary:

Feature	Nuclear chromosomes	Mitochondrial chromosome
DNA	linear double stranded DNA	circular double stranded DNA
genes	thousands	37
centromeres	1	0
telomeres	2	0
origins of replication	thousands	1
Mitosis/ Meiosis	Yes	No
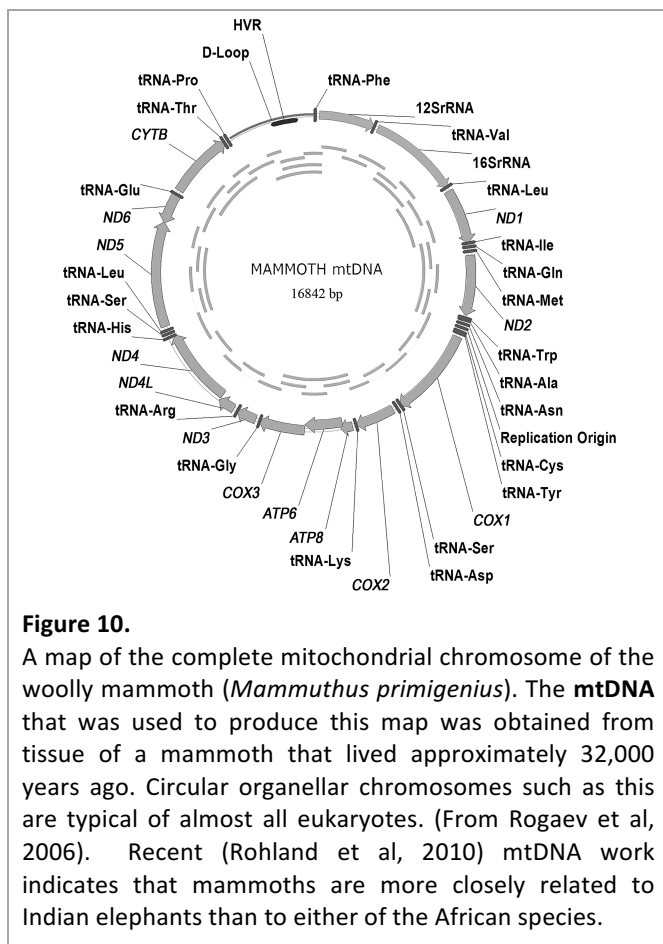
Figure 10.
A map of the complete mitochondrial chromosome of the woolly mammoth (*Mammuthus primigenius*). The **mtDNA** that was used to produce this map was obtained from tissue of a mammoth that lived approximately 32,000 years ago. Circular organellar chromosomes such as this are typical of almost 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.

7. EXAMPLE GENES

7.1. *LCT* - AN AUTOSOMAL GENE

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

Because most genes are on autosomes you have two copies of most of your genes.

7.2. *F8* - AN X CHROMOSOMAL GENE

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

7.3. *SRY* - A Y CHROMOSOMAL GENE

The SRY gene is only found in males, because it is located on the Y chromosome (see Chapter 22). Males have this gene and females do not. In embryogenesis, the presence this gene leads to being male. Its absence leads to being female. A pair of organs called the gonads can develop into either ovaries or testes. In XY embryos the *SRY* gene makes a protein that causes the gonads to develop into testes. Conversely, XX embryos do not have this gene and their gonads develop into ovaries instead. Once formed the testes produce sex hormones that direct the rest of the developing embryo to become male, while the ovaries make different sex hormones that promote female development. The testes and ovaries are also the organs where gametes (sperm or eggs) are produced. Whether a person is genetically male or female is decided at the moment of conception, if the sperm carries a Y chromosome the result is a male and if the sperm carries an X the result is a female.

7.4. *MT-CO1* - A MITOCHONDRIAL GENE

The *MT-CO1* gene is located on the mtDNA chromosome. It encodes a protein in Complex IV of the mitochondrial electron transport chain. For reasons that are not clear this protein must be made in the mitochondria. It cannot be synthesized in the cytosol of the cell and then imported into the mitochondria as is the case with most mitochondrial proteins. Because humans generally receive their mitochondria from their mother, everyone has only one *MT-CO1* gene. It is the same

one found in their mother (and her mother). Technically speaking we have only one MT-CO1 *allele*, it will be identical on all of the mtDNA molecules in all of the mitochondria in all of the cells.

In summary:

Location of a gene	Number of this gene in males	Number of this gene in females
autosomal chromosome	2	2
X chromosome	1	2
Y chromosome	1	0
mitochondrial chromosome	1	1

_____

SUMMARY:

- The c-value is the amount of DNA in a gamete. Humans are 1c = 3000 Mb.

- The n-value is the number of chromosomes in a gamete. Humans are 1n = 23.

- A typical cell in your body is 2c = 6000 Mb and 2n = 46 before DNA replication and 4c = 12 000 Mb and 2n = 46 after.

- A picture of metaphase chromosomes can be organized into a karyogram figure and described with a karyotype statement.

- Humans have two copies of each autosomal chromosome. Females have two X chromosomes while males have one X and one Y chromosome.

- A typical nuclear chromosome has thousands of genes, one centromere, two telomeres, and thousands of origins of replication.

- A typical nuclear chromosome is replicated during S phase and consists of two chromatids up until the start of anaphase. It is condensed during prophase and remains condensed until the start of telophase. During metaphase a chromosome is both replicated and condensed for these reasons.

- The human mitochondrial chromosome has 37 genes, a single origin of replication, and neither centromeres nor telomeres.

- Humans have ~29 000 genes, most of which are on autosomal chromosomes.

- A typical human cell has two copies of each autosomal gene and one of each mitochondrial gene. Genes on sex chromosomes are different: females have two of each X-chromosomal gene while males have one; males have Y-chromosomal genes while females do not.

KEY TERMS:

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

QUESTIONS:

1) Cytogeneticists use white blood cells to obtain metaphase chromosomes for karyotyping.
   a) Why don't they use red blood cells?
   b) Why don't they use white blood cells in anaphase?

2) The human Y chromosome is smaller than the X chromosome. Does this mean that males have less DNA than females?

3) Are these statements true or false? For the false statements explain why.
   a) Everyone has a paternal chromosome 1.
   b) Everyone has a maternal chromosome 1.
   c) Everyone has a paternal X chromosome.
   d) Everyone has a maternal X chromosome.
   e) Everyone has a paternal Y chromosome.
   f) Everyone has a maternal Y chromosome.
   g) Everyone has a paternal mitochondrial chromosome.
   h) Everyone has a maternal mitochondrial chromosome.

4) Explain why centromeres do not have to be in the centre of a chromosome to function.

5) Why do nuclear chromosomes have to have multiple origins of replication?

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

7) Have a look at **Figure 8** Which of these chromosomes would be associated with:
   a) Histone proteins (see **Figure 7**)
   b) Condensin proteins (important scaffold proteins)
   c) Cohesin proteins (proteins which hold sister chromatids together)
   d) Kinetochore proteins (proteins which connect centromere DNA to Microtubules)

8) Where would you find these enzymes in a typical human cell?
   a) DNA polymerases
   b) RNA polymerases
   c) Ribosomes

9) Could the following genes continue to perform their normal developmental function if they were moved next to the *LCT* gene on Chromosome 2?
   a) *F8*
   b) *SRY*
   c) *MT-CO1*

Recommended course:

Regulation of Gene Expression

GENET 304, Gene Expression and its Regulation.
★3 (fi 6) (fall term, 3-0-0)

One of the greatest challenges in biomedical research today lies in understanding the function for each of the ~22,000 genes in the human genome. Perhaps not surprisingly, thousands of these genes are devoted to regulate and coordinate the expression of other genes. In this course we will explore fundamental processes by which genes are regulated, both in simple and complex organisms. We will discuss examples for how organisms respond to nutrients and toxins, how cells acquire memory, and how human diseases are caused when gene regulation goes wrong. We will also cover a range of molecular techniques, including the usage of reporter genes, ChIP-Seq, chromosome conformation capture and CRISPR-Cas9, and how they can be utilized to study gene regulation.
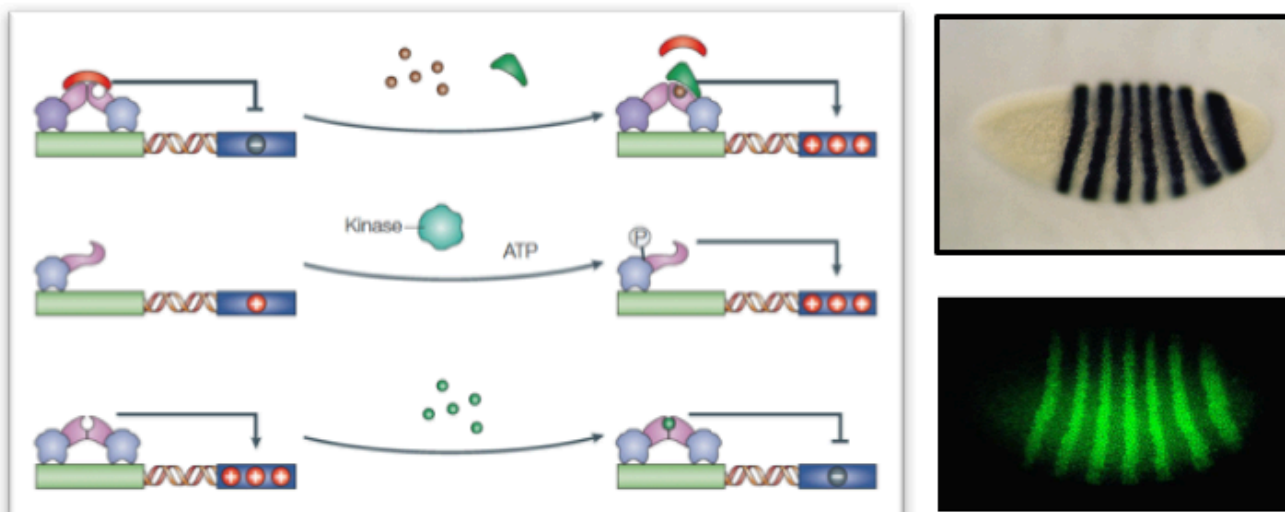
Prerequisite: GENET 270

Offered: Fall Term. Monday, Wednesday, Friday 2:00-3:00 PM

Contact: Dr. Kirst King-Jones, kirst.king-jones@ualberta.ca

Calendar link: http://calendar.ualberta.ca/preview_course_nopop.php?catoid=6&coid=44624

CHAPTER 16 – MENDEL'S FIRST LAW: SEGREGATION OF ALLELES

Figure 1.
Pea plants were used by Gregor Mendel to discover fundamental laws of genetics.
His first law, the segregation of alleles, is covered in this chapter. His second law, independent assortment, is covered in the next chapter. (Wikimedia commons-B. Ebbesen-CC BY-SA 3.0)

INTRODUCTION

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

However, **Gregor Mendel** (**Figure 2**) was one of the first to take a quantitative, scientific approach to the study of heredity. He started with well-characterized strains, repeated his experiments many times, and kept careful records of his observations. Working with peas, Mendel showed that white-flowered plants could be produced by crossing two purple-flowered plants, but only if the purple-flowered plants themselves had at least one white-flowered parent (**Figure 3**). This was evidence that a discrete genetic factor that produced white-flowers had *not* blended

irreversibly with the factor for purple-flowers. Mendel's observations disproved blending inheritance and favor an alternative concept, called **particulate inheritance**, in which heredity is the product of discrete factors that control independent traits.

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 2.
Gregor Johann Mendel (1822-1884), an Augustinian Friar, who lived in Moravia (now part of the Czech Republic), published his work in 1866 on what has become known as the laws of Mendelian Inheritance.
(Wikipedia-Hugo Iltis- CC BY 4.0)
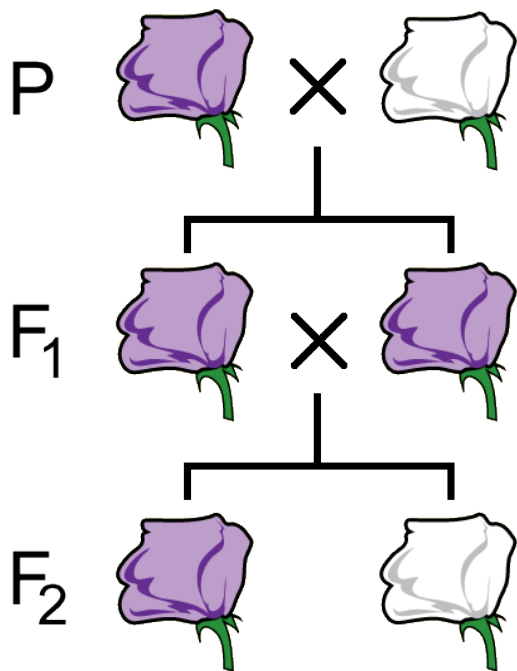
Figure 3.
Inheritance of flower color in peas. Mendel observed that a cross between pure breeding, white and purple peas (generation P) produced only progeny (generation F_1) with purple flowers. However, white flowered plant reappeared among the F_2 generation progeny of a mating between two F_1 plants. The symbols P, F_1 and F_2 are abbreviations for parental, first filial, and second filial generations, respectively.
(Original-Deyholos- CC BY-NC 3.0)

1. OVERVIEW

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

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

1.1. DOMINANT AND RECESSIVE ALLELES

The concepts of dominant and recessive alleles were introduced in Chapter 13. Remember, alleles are different versions of a gene. The relationship of different alleles of a gene can be described as complete dominance, incomplete dominance or co-dominance. The traits Mendel studied with his peas were all completely dominant, and therefore will only be briefly reviewed here.

In a diploid organism, if an allele is **dominant** only one copy of that allele is necessary to express the dominant phenotype. If an allele is **recessive**, then the gene needs to have two copies (or be homozygous) to express the recessive phenotype. If an organism is a heterozygote, or has one copy of each allele type, then it will show the dominant phenotype. When representing these in written form, a dominant allele is written as a capital letter (e.g. *A*), while a recessive allele will be written in lower case (e.g. *a)*. If these are alleles of the same gene, they should be written with the same letter. This is the most common way of writing genotypes (Table 1), but there are many different systems

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

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

that often deviate from these general rules. Note that genes and alleles are usually written in *italics* and chromosomes and proteins are not, proteins often written in all capitals. For example, the *white* gene in *Drosophila melanogaster* on the X chromosome encodes a protein called WHITE.

1.2. MEIOSIS OVERVIEW

Most eukaryotes reproduce sexually - a cell from one individual joins with a cell from another to create offspring. In order for this to be successful, the cells that fuse must contain half the number of chromosomes as in the adult organism. Otherwise, the number of chromosomes would double with each generation, which would be unsustainable. The chromosome number is reduced through the process of **meiosis**. Meiosis is similar in many ways to mitosis (**Figure 4**), as the chromosomes are lined up along the metaphase plate and divided to the poles using microtubules. It also differs in many significant ways from mitosis. Keep this in mind and try to note the differences as you read ahead. Note also that during this chapter that we will be discussing N and C values. Refer back to Chapter 14 to refresh yourself with this concept.

Meiosis has two main stages, designated by the roman numerals I and II. In **Meiosis I** homologous chromosomes segregate, while in **Meiosis II** sister chromatids segregate (**Figure 5**). Most multicellular organisms use meiosis to produce **gametes**, the cells that fuse to make offspring. Some single celled eukaryotes such as yeast also use meiosis to enter the haploid part of their life cycle. Cells that will undergo meiosis are called **meiocytes** and are diploid (2N)(**Figure 6**). You will hear of cells that have not yet undergone meiosis to become egg or sperm cells called oocytes or spermatocytes respectively.

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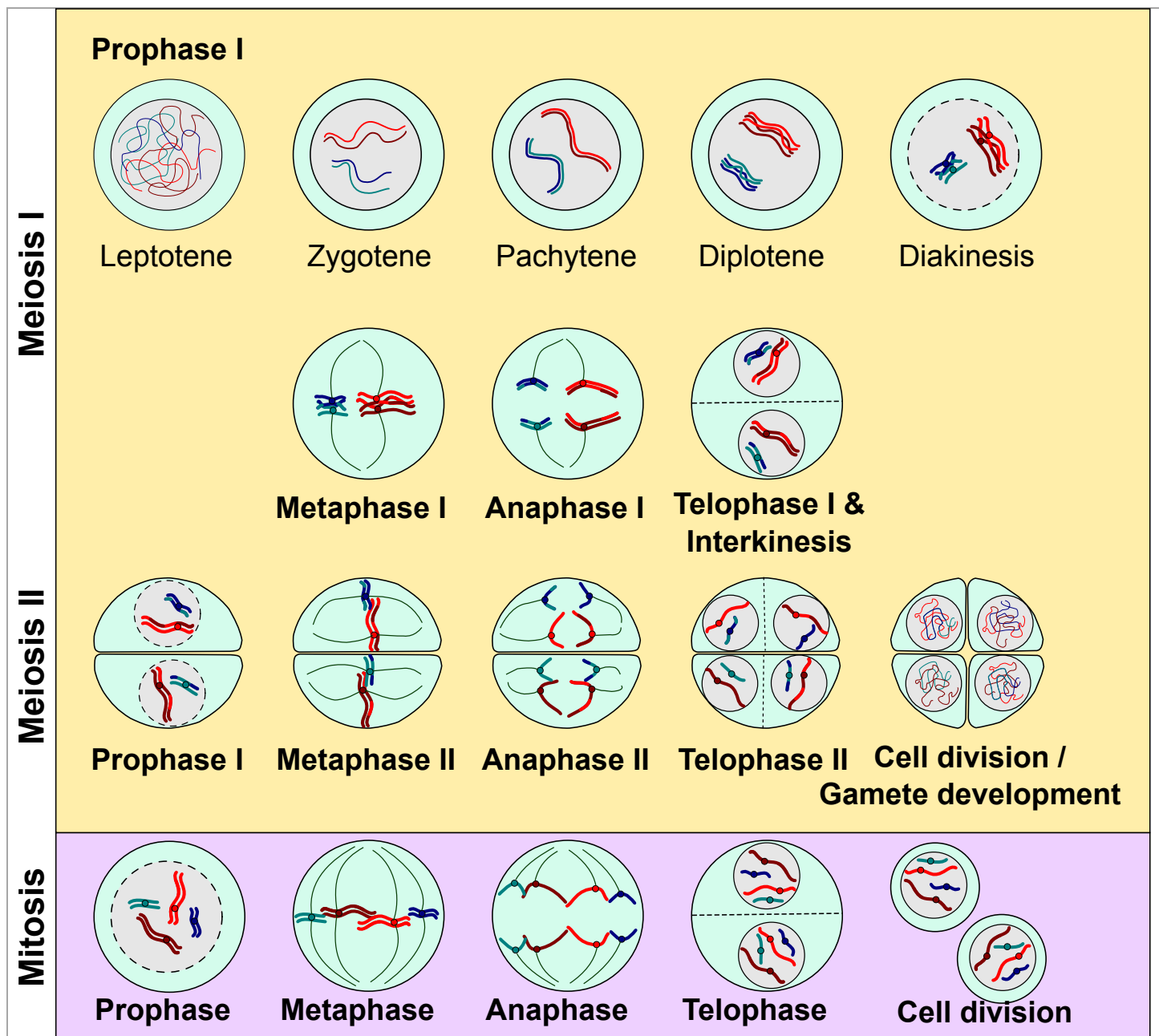
Figure 4.

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

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

Mitosis is listed for comparison. See the chapter on Mitosis for more details on the stages.

(Original–L. Canham–CC BY-NC 3.0)

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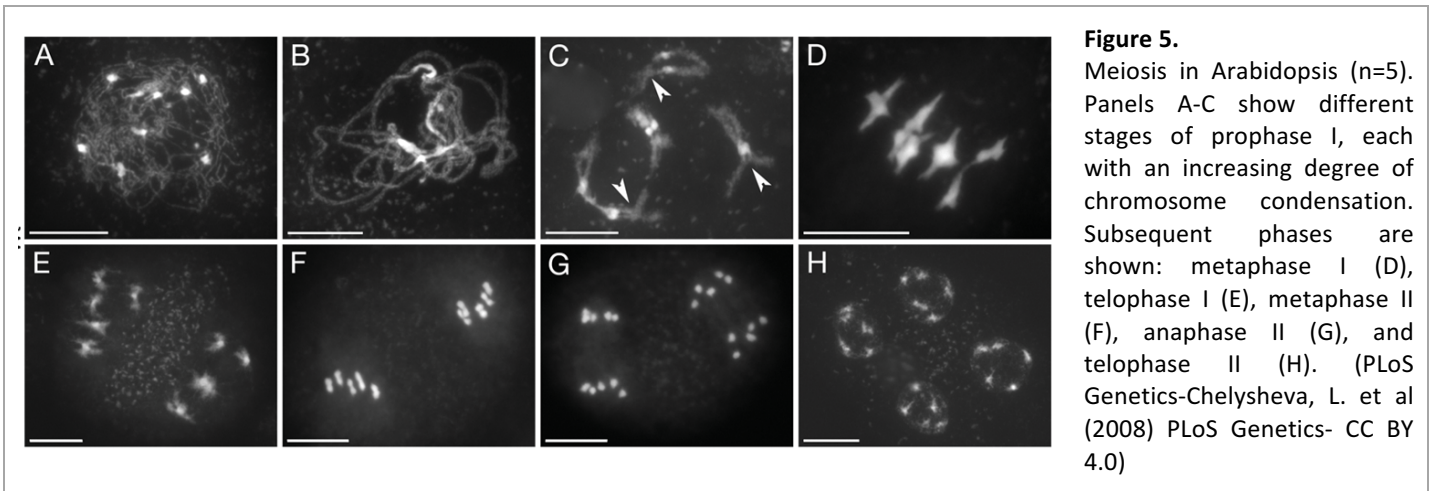
2. MEIOSIS I

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

In meiosis I replicated, homologous chromosomes pair up, or synapse, during prophase I, line up in the middle of the cell during metaphase I, and separate during anaphase I. For this to happen the homologous chromosomes need to be brought together while they condense during prophase I. During synapsis, proteins bind to both homologous chromosomes along their entire length and form the **synaptonemal complex** (synapse means junction). These proteins hold the chromosomes in

Figure 6.
Changes in DNA and chromosome content during the cell cycle. For simplicity, nuclear membranes are not shown, and all chromosomes are represented in a similar stage of condensation. (Original-Deyholos- CC BY-NC 3.0)
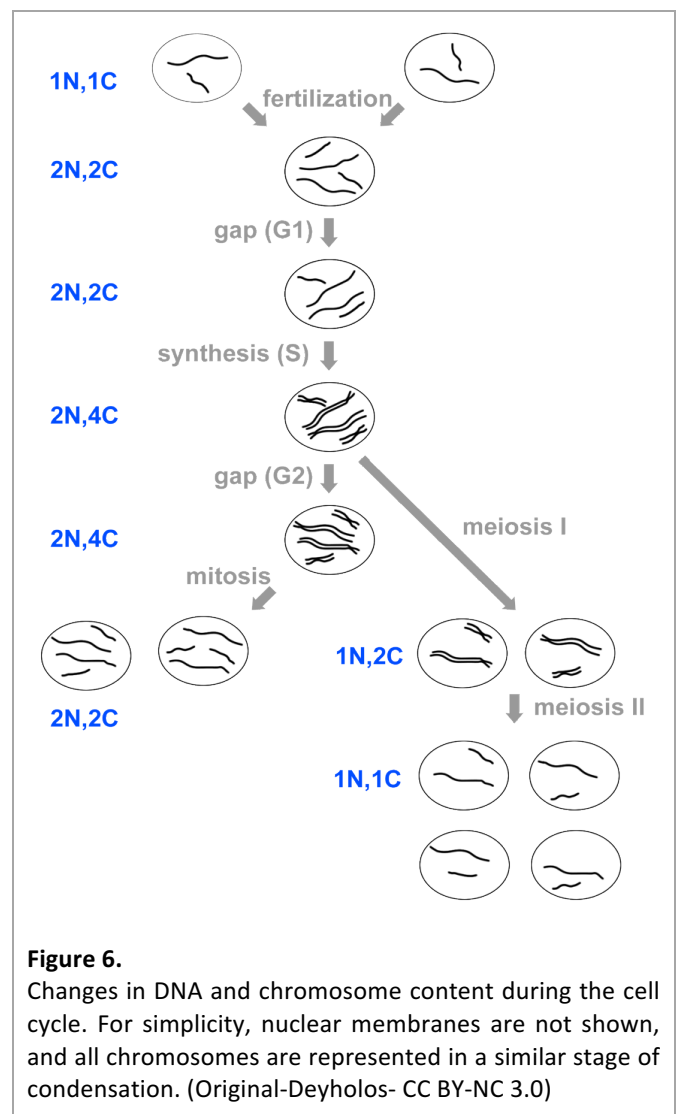
the transient structure of a bivalent (**Figure 7**). The proteins are released when the cell enters anaphase I.
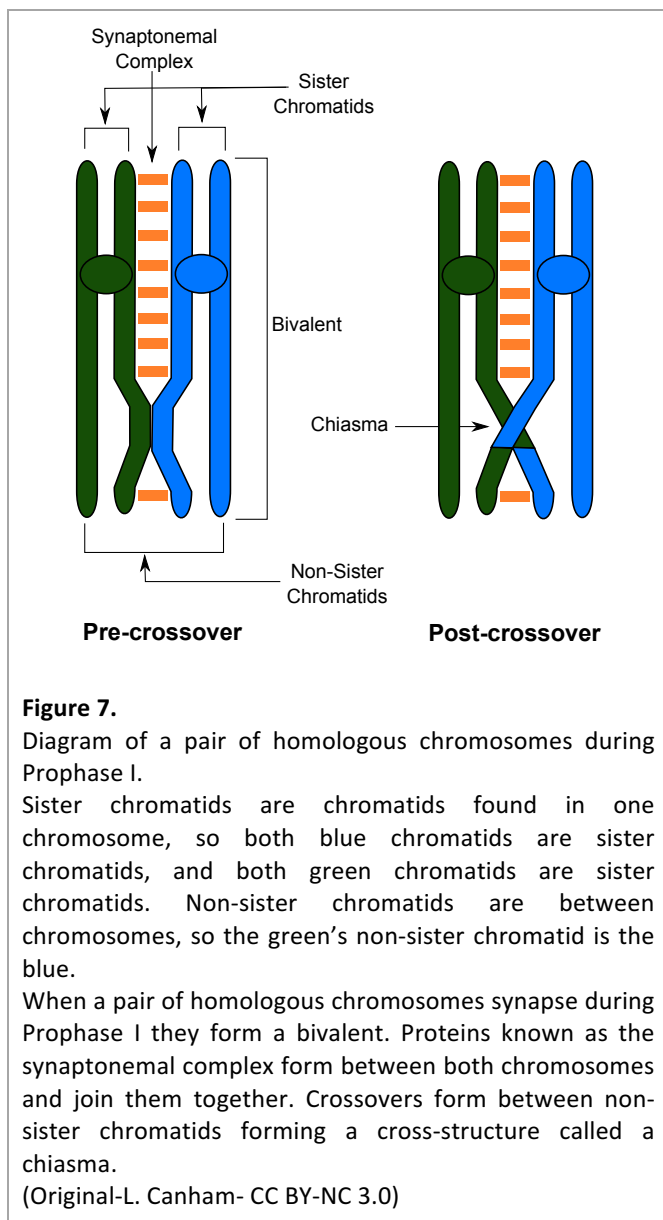
Figure 7.
Diagram of a pair of homologous chromosomes during Prophase I.
Sister chromatids are chromatids found in one chromosome, so both blue chromatids are sister chromatids, and both green chromatids are sister chromatids. Non-sister chromatids are between chromosomes, so the green's non-sister chromatid is the blue.
When a pair of homologous chromosomes synapse during Prophase I they form a bivalent. Proteins known as the synaptonemal complex form between both chromosomes and join them together. Crossovers form between non-sister chromatids forming a cross-structure called a chiasma.
(Original-L. Canham- CC BY-NC 3.0)

2.1. STAGES OF PROPHASE I

In meiosis, Prophase I is divided up into five visual stages, that are steps along a continuum of events (**Figure 4**). Leptotene, zygotene, pachytene, diplotene and diakinesis. From interphase, a cell enters **leptotene** as the nuclear material begins to condense into long thin visible threads (chromosomes). During **zygotene** homologous chromosomes begin to pair up (synapse) and form an elaborate structure called the synaptonemal complex along their length. During zygotene the chromosomes are still quite long, but it is more apparent that they are distinct now. At **pachytene** homologous chromosomes are thicker and fully synapsed (two chromosomes and four chromatids)

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

2.2. METAPHASE I, ANAPHASE I AND TELOPHASE I

Metaphase I is where the major difference between mitosis and meiosis becomes apparent. The homologous pairs, or bivalents, orient themselves along the metaphase plate and the microtubules attach themselves to each chromosome's centromere, one pole attaching to each respective homologous pair (**Figure 4**). This is different from mitosis, where the chromosomes align individually and the microtubules from both poles attach themselves to an individual chromosome in preparation for separating the chromatids.

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

3. MEIOSIS II

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**. 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 (**Figure 6**).

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

3.1. GAMETE MATURATION
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 structures, like tails and become functional sperm cells. In female animals the gametes are eggs. For each egg to contain the maximum amount of nutrients, typically only one of the four products of meiosis becomes an egg. The other three cells end up as tiny disposable cells called **polar bodies**. In plants the products of meiosis reproduce a few times using mitosis as they develop into functional male or female gametes.

4. CROSSING OVER (INTRA-CHROMOSOMAL RECOMBINATION)

During prophase I the homologous chromosomes pair together and form a synaptonemal complex.

Crossing over occurs within the synaptonemal complex. A crossover is a place where DNA repair enzymes break the DNA of two non-sister chromatids in similar locations and then covalently reattach non-sister chromatids together to create a crossover between non-sister chromatids. This reorganization of chromatids will persist for the remainder of meiosis and result in recombination of alleles in the gametes. Crossover events can be seen as **Chiasmata** on the synapsed chromosomes in late Meiosis I.

Crossovers function to hold homologous chromosomes together during meiosis I so they orient correctly and segregate successfully. Crossing over also reshuffles the allele combinations along a chromosome resulting in genetic diversity, that can be selected in a population over time (evolution).

5. ONE LOCUS ON A CHROMOSOME - SEGREGATION - MONOHYBRID

Not only did Mendel solve the mystery of inheritance as units (genes), he also invented several testing and analysis techniques still used today. **Classical genetics** is the science of examining biological questions using controlled matings of model organisms. It began with Mendel in 1865 but did not attain widespread usage until Mendel's work was rediscovered in 1903 by four researchers (E. von Tschermak, H. de Vries, C. Correns, and W. J. Spillman). Then Thomas Morgan began working with fruit flies in 1908 and used this work. Later, starting with Watson and Crick's structure of DNA in 1953, classical genetics was joined by **molecular genetics**, the science of solving biological problems using DNA, RNA, and proteins. The genetics of **DNA cloning** began in 1970 with the discovery of restriction enzymes and plasmids as cloning vectors.

Knowing what we now know about the process of meiosis, we can better understand the mechanisms underlying Mendel's First Law. The Law of Segregation states that every individual contains a pair of alleles for each gene, which segregate during the formation of gametes, and so for every gene pair each parent passes on a random allele to

its offspring. The series of experiments that led to the formulation of Mendel's first law where based on the process of **Monohybrid crosses**, which will be described below.

5.1. Terminology

A specific position, region, or segment 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 – <u>all the others are present and assumed to be wild type</u> but are normally not written in the genotype. The observable or detectable effect of these alleles on the structure or function of that individual is called its **phenotype**. 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.

5.2. True breeding lines

Geneticists make use of **true-breeding lines** just as Mendel did (**Figure 8**a). These are in-bred populations of plants or animals in which all parents and their offspring (over many

Figure 8.
(**a**) A true-breeding line (**b**) A monohybrid cross produced by mating two different pure-breeding lines.
(Original-Deyholos-CC BY-NC 3.0)

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.

5.3. 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 8**b). Recall from **Figure 3** that the generations in a cross are named P (parental), F_1 (first filial), F_2 (second filial), and so on.

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

6. Punnett Squares - 3:1 Ratio

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

Given the genotypes of any two parents, we can predict all of the possible genotypes of the offspring. Furthermore, if we also know the
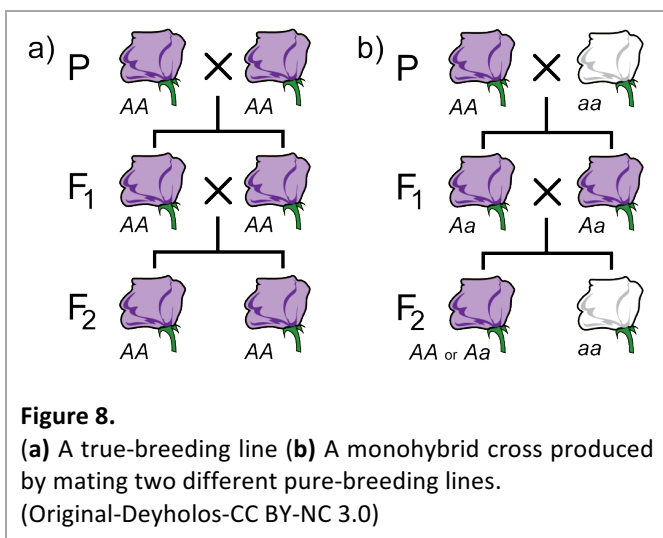
dominance relationships for all alleles, we can predict the phenotypes of the offspring. This provides a convenient method for calculating the expected genotypic and phenotypic ratios from a cross.

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

The F_1 cross from **Figure 8**b would be drawn as in **Figure 9**. As you can see, in a Monohybrid cross, the offspring ratios will be 3:1 of dominant phenotype (purple) : recessive phenotype (white). Punnett squares can also be used to calculate the frequency of offspring. The frequency of each offspring is the frequency of the male gametes multiplied by the frequency of the female gamete.

Figure 9.
A Punnett square showing a monohybrid cross. The purple boxes represent the purple colour of the dominant (A) allele, while the white box represents the recessive (aa) allele homozygote. (Original-L. Canham- CC BY-NC 3.0)

Figure 10.
Punnett squares showing the two possible outcomes of a single locus test cross. (Original-L. Canham- CC BY-NC 3.0)
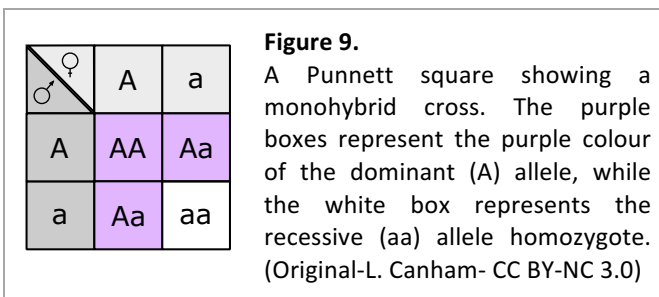
7. SINGLE LOCUS TEST CROSSES

Knowing the genotypes of an individual is an important part of a genetic experiment. However, genotypes cannot be observed directly; they must be inferred based on phenotypes. Because of dominance, it is often not possible to distinguish between a heterozygote and a homozygote based on phenotype alone (e.g. see the purple-flowered F_2 plants **Figure 8**b). To determine the genotype of a specific individual, a **test cross** can be performed, in which the individual with an unknown 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 10**), you will observe one of two phenotypic ratios in the F_1 generation. If the purple-flowered parent was a homozygote *AA*, all of the F_1 progeny will be purple. If the purple-flowered parent was a heterozygote *Aa*, the F_1 progeny should segregate purple-flowered and white-flowered plants in a 1:1 ratio.

_____
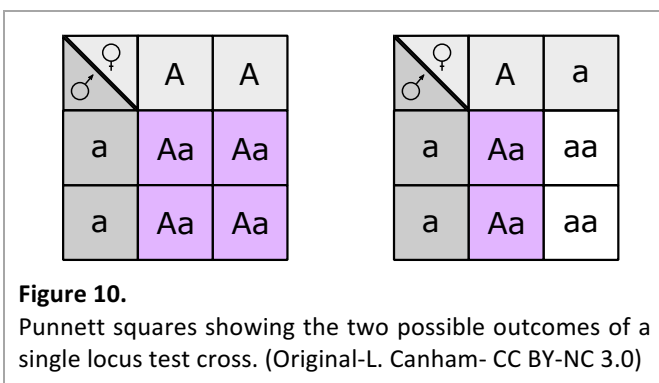
SUMMARY:

- Mendel demonstrated that heredity involved discrete, heritable factors that affected specific traits.

- A gene can be defined operationally as a unit of inheritance.

- Homologous chromosomes contain the same series of genes along their length, but not necessarily the same alleles. Sister chromatids initially contain the same alleles.

- Homologous chromosomes pair (sysnapse) with each other during meiosis, but not mitosis.

- A diploid organisms 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, and reciprocal crosses.

KEY TERMS:

blending inheritance	chiasma / chiasmata
Gregor Mendel	diakinesis
particulate inheritance	metaphase I
alleles	anaphase I
Mendel's First Law	telophase I
The Law of Equal Segregation	interkinesis
dominant	prophase II
recessive	metaphase II
meiosis I	anaphase II
meiosis II	telophase II
gametes	polar bodies
meiocytes	classical genetics
reductional	molecular genetics
synapse	DNA cloning
bivalent	monohybrid cross
equational	locus
pair up	genotype
synaptonemal complex	phenotype
leptotene	true-breeding lines
zygotene	punnett square
pachytene	test cross
crossing over	tester
diplotene	

STUDY QUESTIONS:

1) How would the results of the cross in **Figure 3** have been different if heredity worked through blending inheritance rather than particulate inheritance?

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

3) What is the maximum number of alleles at a given autosomal locus in a normal gamete from a diploid individual? In the whole population of a species?

4) 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 a litter 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?

5) 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?

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

Recommended course:

Genetic Analysis

GENET 305, Genetic Analysis

★ 3 (fi 6) (second term, 3-0-0).

Learn techniques used to analyze the functions of genes in animal model systems. Mutant analysis; gene dosage; chromosome mechanics; transgenics; forward genetic screens; dominant phenotypic modifiers; epistasis; genetic mosaics, meiotic recombination.
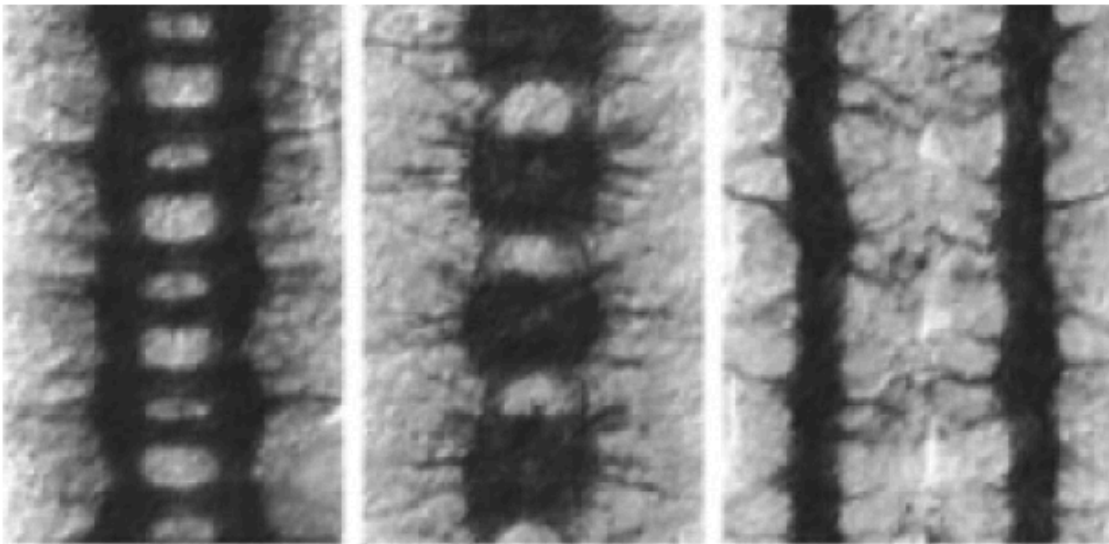
Example: Correct gene expression is required for embryonic nervous system development. Wild type expression in the ventral nerve cord (left) and two different mutants (center, right).

Prerequisites: GENET 270

Class schedule: T/Th 9:30-10:50 AM

Offered: Winter Term

Contact: Dr. DiCara, dicara@ualberta.ca

Chapter 17 – Mendel's Second Law:

Independent Assortment

Figure 1.
Hand pollination of a pumpkin flower. When Mendel was doing his crosses with pea plants, he pollinated each flower by hand in a similar way in order to be sure he knew the parents of each cross.
(Flickr-S. Hurmerinta- CC BY-NC 2.0)

Introduction

The principles of genetic analysis that we have described for a single locus in Chapter 16 will be extended to the study of alleles at two loci in this Chapter. The analysis of two loci in the same cross provides information for genetic mapping (Chapter 18) and testing gene interactions (Chapter 26). These techniques are very useful for both basic and applied research. Before discussing these techniques, we will first revisit Mendel's classical experiments.

Before Mendel's 1865 publication, **blended inheritance** was the accepted model to explain the transmission of traits. It was Mendel's work that established that **heritable traits** were controlled by discrete factors, which we now call alleles, in a **particulate inheritance** model. At the time it was an important question as to whether heritable traits, controlled by discrete factors, were inherited independently of each other? To answer this, Mendel took two apparently unrelated traits, such as seed shape and seed color, and studied their inheritance together in one individual. For example, he studied two variants of each trait: seed color was either green or yellow, and seed shape was either round or wrinkled. (He studied

seven traits in all, each on a different chromosome.) When either of these traits was studied individually, the phenotypes segregated in the classical 3:1 ratio among the progeny of a monohybrid cross (**Figure 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 hybrids were in the same individual?
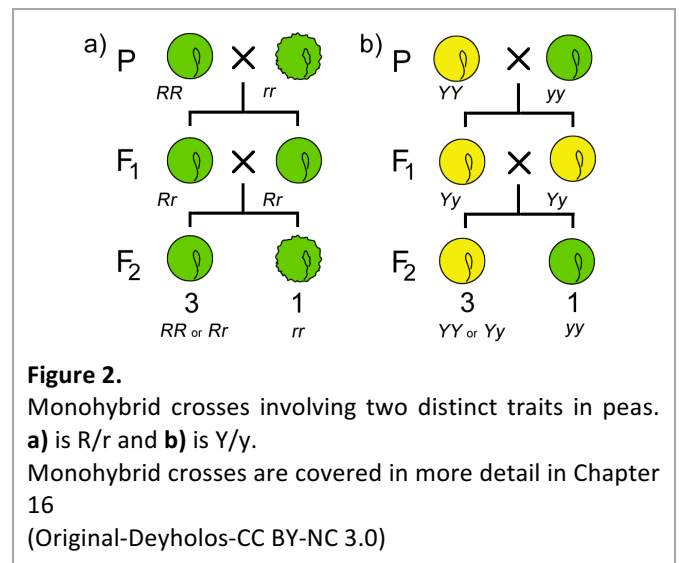
Figure 2.
Monohybrid crosses involving two distinct traits in peas.
a) is R/r and **b)** is Y/y.
Monohybrid crosses are covered in more detail in Chapter 16
(Original-Deyholos-CC BY-NC 3.0)

Like in the previous chapter, we will first walk through how a dihybrid cross works on at the DNA level, and then we will explain the results that

Mendel saw that led him to his law, the Law of Independent Assortment.

When dealing with alleles at two different loci, we have to use nomenclature that makes the arrangement clear. There are three possible arrangements: Both loci are on the same chromosome (AB/ab), different chromosomes (A/a; B/b), or unknown (AaBb).

1. TWO LOCI ON DIFFERENT CHROMOSOMES

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

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

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

2. TWO LOCI ON ONE CHROMOSOME

Based on the description in the last section, it would be expected that if the genes were on the same chromosome the alleles would travel together through meiosis (**Figure 4** top). However, when tested this is not always the case. The recombination of alleles can be explained through the phenomenon of **crossing over**, which occurs during prophase I as described in chapter 16.
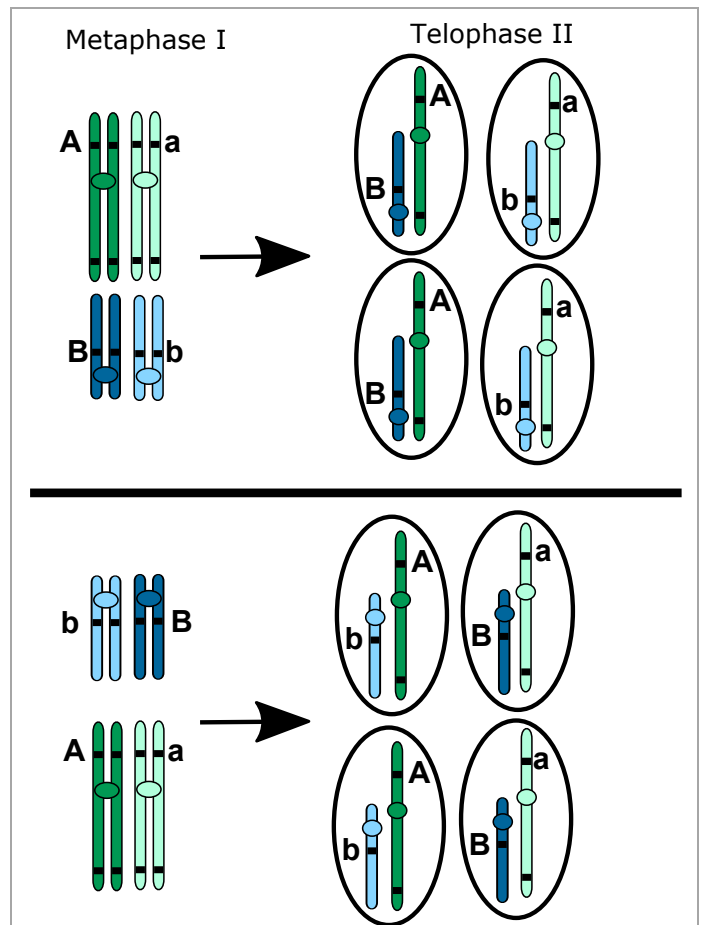
Figure 3.
Independent assortment as seen on two different chromosomes. Gene A is found on the short chromosome and Gene B is found on the long chromosome, and both genes are heterozygotes for the dominant (A and B) and recessive (a and b) alleles. The orientation that the chromosomes align themselves during metaphase I affect the alleles found in the 4 gametes produced after telophase II.
These are just two of many orientations the chromosomes can arrange themselves in at metaphase I. The full stages of meiosis were removed for simplicity; refer to chapter 16 to understand the divisions that lead to the 4 gametes seen in telophase II.
(Original-L. Canham-CC BY-NC 3.0)

Crossing over is an exchange between non-sister chromatids that can occur at any position along the entire chromosome. If the two loci that are being considered are sufficiently separated from each other on the chromosome, crossover events can occur between the two loci.

This coupled with the random orientation that the chromosomes align during metaphase I, will allow

the other combination of alleles in the gametes (**Figure 4** bottom).

While not shown in **Figure 4**, if the two loci are very far apart, multiple crossover events can also take place, further increasing the shuffling of alleles.

Figure 4.
Independent assortment as seen on the same chromosome. On the top is an example of what would happen if cossovers do not occur. The dominant alleles of gene *A* and gene *D* would travel together, not leading to independent assortment. Crossovers do occur in most situations though, like in the bottom half of the figure. If a crossover occurs between the two genes, then the alleles will transfer to the other non-sister chromatid, thus rearranging alleles. This allows for independent assortment, despite being on the same chromosome.
This is just one of the many arrangements or crossover events that could occur during meiosis, with every meiocyte arranging themselves differently with different crossovers.
(Original-L. Canham-CC BY-NC 3.0)

The farther apart on the chromosome the more crossover events will take place between the two

loci. Ultimately, this will result in similar allele combinations to those observed in independent assortment shown above, even if they are on the same chromosome.

If the loci are very close together on the same chromosome, fewer crossovers are likely occur between them. We will not discuss this situation in here, but will do later in chapter 18.

3. A DIHYBRID CROSS SHOWING MENDEL'S SECOND LAW (INDEPENDENT ASSORTMENT)

Mendel found that each locus had two alleles, that segregated from each other during the creation of gametes. He wondered whether dealing with multiple traits at a time would affect this segregation, so he created a **dihybrid cross**. The distribution of offspring from his experiments led him to formulate **Mendel's Second Law**, the **Law of Independent Assortment**, which states that the segregation of alleles at one locus will not influence the segregation of alleles at another locus during gamete formation – the alleles segregate independently. Next, we will discuss how he came to this understanding, given that independent assortment occurs.

3.1. MENDEL'S SECOND LAW

To analyze the simultaneous segregation of two 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. This produced F_1 progeny that had all green and round peas. They were called **dihybrids** because they carried two alleles at each of the two loci (**Figure 5**).

From **Figure 2** we know that yellow and round are dominant, and green and wrinkled are recessive. If the inheritance of seed color was truly independent of seed shape, then when the F_1 dihybrids were crossed to each other, a 3:1 ratio of one trait should be observed within each phenotypic class of the other trait (**Figure 5**). Using the product law, we would therefore predict
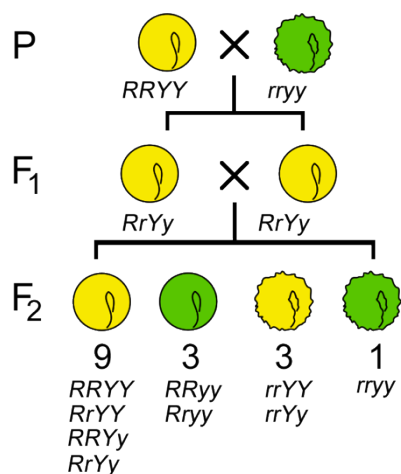
Figure 5.
Two pure-breeding lines are crossed to produce dihybrids in the F1 generation. These F1 are crossed to produce four phenotypic classes, which appear in a 9:3:3:1 ratio. (Original-Deyholos-CC BY-NC 3.0)

Figure 6.
A Punnett Square showing the results of the dihybrid cross from **Figure 5**. Each of the four phenotypic classes is represented by a different color of shading. (Original-Deyholos-CC BY-NC 3.0)

that if ¾ of the progeny were yellow, and ¾ of the progeny were round, then ¾ × ¾ = 9/16 of the progeny would be both round and yellow (**Table 1**).

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

The 9:3:3:1 phenotypic ratio that we calculated using the product rule could also be obtained using Punnett Square (**Figure 6**). 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.

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

Table 1.
Phenotypic classes expected in monohybrid and dihybrid crosses for two seed traits in pea.

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**. 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 genotypic classes 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 classes 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, green 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.

3.2. 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) alleles at each locus segregate independently of the alleles at the other;
(2) one allele at each locus is completely dominant (the other recessive); and
(3) each of four possible phenotypes can be distinguished unambiguously, with no interactions between the two genes that would interfere with determining the genotype correctly.

For simplicity, most student examples involve easily scored phenotypes, such as pigmentation or other changes in visible structures. However, keep in mind that the analysis of segregation ratios of any two marker loci can provide insight into their relative positions on chromosomes.

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

Linkage is one of the most important reasons for distortion of the ratios expected from independent assortment. Two loci show linkage if they are located close together on the same chromosome. This close proximity alters the frequency of allele combinations in the gametes. We will return to the concept of linkage in Chapter 18. Deviations from 9:3:3:1 ratios can also be due to interactions between genes, such as epistasis, duplicate gene action and complementary gene action. These interactions are discussed in Chapter 26.

4. THE DIHYBRID TEST CROSS
While the cross of an F$_1$ x F$_1$ gives a ratio of 9:3:3:1, there is a better, easier cross to test for independent assortment: the dihybrid test cross. In a dihybrid test cross, independent assortment is seen as a ratio of 1:1:1:1, which is easier to score than the 9:3:3:1 ratio. This test cross will also be easier to use when testing for linkage (Chapter 18).

Like in monohybrid crosses (Chapter 16), you can do test crosses with dihybrids to determine the genotype of an individual with dominant phenotypes, to see if they are heterozygous or homozygous dominant. This type of cross is set up in the same fashion: an individual with an unknown genotype in two loci is crossed to an individual that is homozygous recessive for both loci.

Punnett squares should be done ahead of the crosses, so you know what to expect for any of the possible outcomes. Using the example from the rest of this chapter, you cross a double homozygous recessive pea plant (r/r ; y/y. green and wrinkled) to an unknown individual that has two dominant phenotypes (R/_ ; Y/_. yellow and round). There are four possible genotypes the unknown individual could be: R/R ; Y/Y or R/R ; Y/y or R/r ; Y/Y or R/r; Y/y. The Punnett squares for the first two are listed below (**Figure 7**). Notice on the left you only get the dominant phenotype for both, so you know both genes in the unknown are

homozygous dominant. On the right you get only the dominant phenotype for round peas, but you get 50% yellow and 50% green peas, showing that the unknown is homozygous for round, but heterozygous for colour of the peas. **Figure 8** is blank for you to fill in the two other gamete and genotype possibilities.

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

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

Figure 7.
Punnett square for a test cross. The tester in both cases is the male with the genotype r/r ; y/y .
On the left, the unknown has a genotype of R/R; Y/Y.
On the right, the unknown has the genotype R/R ; Y/y.
(Original-L. Canham-CC BY-NC 3.0)

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

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

Figure 8.
Blank Punnett squares to fill in the other two possibilities of the test cross.

_____

SUMMARY:

- The alleles of loci in different chromosomes are inherited independently of each other.

- The expected phenotypic ratio of a dihybrid cross is 9:3:3:1.

- The 9:3:3:1 ratio can be modified if the loci are not simple Dominant/recessive to each other, or if there are gene interactions, or if the two loci are linked.

- A test cross gives a ratio of 1:1:1:1 for loci that assort independently.

KEY TERMS:

blended inheritance	dihybrid cross
heritable traits	Mendel's Second Law
particulate inheritance	Law of Independent Assortment
Independent Assortment (IA)	9:3:3:1
crossing over	Linkage
dihybrid	

STUDY QUESTIONS:

1) **Figure 7** shows Punnett squares for two of the four possible test crosses. Fill in the Punnett squares in **Figure 8** for the other two possible genotypes of the unknown that aren't shown.

2) Based on meiosis, when dealing with two loci, there will always be four distinct gamete types. But if the organism is homozygous, like the tester, all those gametes will look the same. In this situation, when writing a Punnett square, is it necessary to write out the four similar gametes? How would you re-draw the Punnett Square on the right in **Figure 7**?

3) If two loci assort independently, then the AABB x aabb cross will result in dihybrid progeny, which when crossed together will give ratios of 9:3:3:1 in the F2, assuming "A" and "B" are dominant to "a" and "b", respectively.
   Now, assume that locus "A" and "B" are somewhat linked and thus will **NOT** assort independently. That is the "AB" and "ab" combinations are more likely. How will this affect (change) the 9:3:3:1 ratio?

4) Do the same first cross as Question#3 but make the second cross a test cross (x aabb), with expectation of a 1:1:1:1 ratio. How would the ratio be changed if the two loci were not assorting independently but are somewhat linked?

CHAPTER 18 – GENES ON THE SAME CHROMOSOME: LINKAGE

Figure 1.
The coat colour on this juvenile horse is called Bay Roan Tobiano. Bay is the brown base coat colour; Roan is the mixture of white hairs with the base coat, making a 'foggy' colour; and Tobiano is the white patches. The genes causing the Roan and Tobiano coat colours, respectfully, are found on the same chromosome and are linked. Knowing this, we can predict which coat colour genes are from which parents, and how those genes will be inherited in this horse's offspring.
(Wikimedia Commons-Kumana @ Wild Equines- CC BY 2.0)

INTRODUCTION

As we learned in Chapter 17, Mendel reported that the pairs of loci he observed segregated independently of each other; for example, the segregation of seed color alleles was independent from the segregation of alleles for seed shape. This observation was the basis for his Second Law (Independent Assortment), and contributed greatly to our understanding of heredity as single units. However, further research showed that Mendel's Second Law did not apply to every pair of genes that could be studied. In fact, we now know that alleles of loci that are located close together on the same chromosome tend to be inherited together. This phenomenon is called **linkage**, and is a major exception to Mendel's **Second Law of Independent Assortment.** Researchers use linkage to determine the location of genes along chromosomes in a process called genetic mapping. The concept of gene linkage is important to the natural processes of heredity and evolution, as well as to our genetic manipulation of crops and livestock.

1. GENETIC NOMENCLATURE & SYMBOLS

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

A **gene** is a hereditary unit that occupies a specific position (**locus**) within the genome or chromosome and has one or more specific effects upon the phenotype of the organism and can mutate into various forms (**alleles**) (A Dictionary of Genetics 3rd Ed., King & Stansfield,1985) . A **genotype** is the specific allelic composition of a cell or organism. Normally only the genes under consideration are listed in a genotype and the alleles at the remaining gene loci are considered to be wild type. A **phenotype** is the detectable outward manifestation of a specific genotype. In describing a phenotype usually only the characteristics under consideration are listed while the remaining characters are assumed to be wild type (normal).

1.1. GENE NAMES AND SYMBOLS

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

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

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

and mutant allele of the "a" gene at the "a" locus. This may also be abbreviated to *+/a.*

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

2. RECOMBINATION

The process of meiosis leading to a separation of chromosomes, and crossing over is necessary for the understanding of this chapter. Refer to Chapter 16 and 17 for a review of these concepts.

The term "recombination" is used in several different contexts in genetics. In reference to heredity, **recombination** is defined as a process that results in gametes with combinations of alleles that were not present in the gametes from the parental generation (**Figure 3**). Recombination is important because it contributes to the genetic variation that may be observed between individuals within a population and that may be acted upon by selection for evolution.

Figure 2.
A diagram illustrating how chromosomes, loci and alleles in a cell, and how we depict them as text.
(Original-J.Locke- CC BY-NC 3.0)

Figure 3.
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 BY-NC 3.0)

2.1. INTER- AND INTRACHROMOSOMAL RECOMBINATION

Interchromosomal recombination occurs either through **independent assortment** of alleles whose loci are on different chromosomes (Chapter 17). **Intrachromosomal recombination** occurs through **crossovers** between loci on the same chromosomes. It is important to remember that in both of these cases, recombination is a process that occurs during meiosis (mitotic recombination may also occur in some species, but it is relatively rare).

As an example of interchromosomal recombination, consider loci on two different chromosomes as shown in **Figure 3**. We know that if these loci are on different chromosomes there is no physical connection between them, so they are **unlinked** and will segregate independently as did Mendel's traits. The segregation depends on the 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 3**.

Intrachromosomal recombination occurs through crossovers. Crossovers occur during prophase I of meiosis, when pairs of homologous chromosomes have aligned with each other in a process called **synapsis**. Crossing over begins with the breakage of DNA of a pair of non-sister chromatids. The breaks occur at corresponding positions on two non-sister chromatids, and then the ends of non-sister chromatids are connected to each other resulting in a reciprocal exchange of double-stranded DNA. Generally, every pair of chromosomes has at least one crossover during meiosis, but often multiple crossovers occur in each chromatid during prophase I. Further details and figures of crossovers are shown in Chapter 16 and 17.

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

2.2. INHERITING PARENTAL AND RECOMBINANT GAMETES

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

To properly identify recombinant and parental gametes from an individual, you need to know the genotype of its parents (the P generation). This is most easily demonstrated in a dihybrid. If, for two genes, one parent has the genotype *A/A B/B,* they can only produce one type of gamete: *AB*. Similarly, if they are *a/a b/b,* then they can also only produce one type of gamete: *ab* (**Figure 4** right). However, if those two gametes (*AB* and *ab*) combine, they create an individual (F$_1$) that has a genotype written as *A/a B/b*. It can be easier to keep track of the parental combinations of gametes by keeping them together when writing the genotype, for this example *AB/ab* (**Figure 4**).

So the above dihybrid individual can produce four different gametes: *AB*, *ab*, *Ab* and *aB*. The **parental gametes** are those that the individual obtained from their parents, in this case *AB* and *ab*. *Ab* and *aB* are **recombinant gametes** and are evidence of a recombination event happening, resulting in a different combination of alleles (**Figure 4** right).

For the above example, the P generation has one parent homozygous for both dominant alleles, and the other homozygous for both recessive alleles. It is very important to note that this will not always be the case. In some instances, one parent will be homozygous with one gene dominant and the other gene recessive (*A/A b/b*) and the other parent will be the opposite (*a/a B/B*). This situation will change which is the parental and recombinant gametes (compare left and right in **Figure 4**).
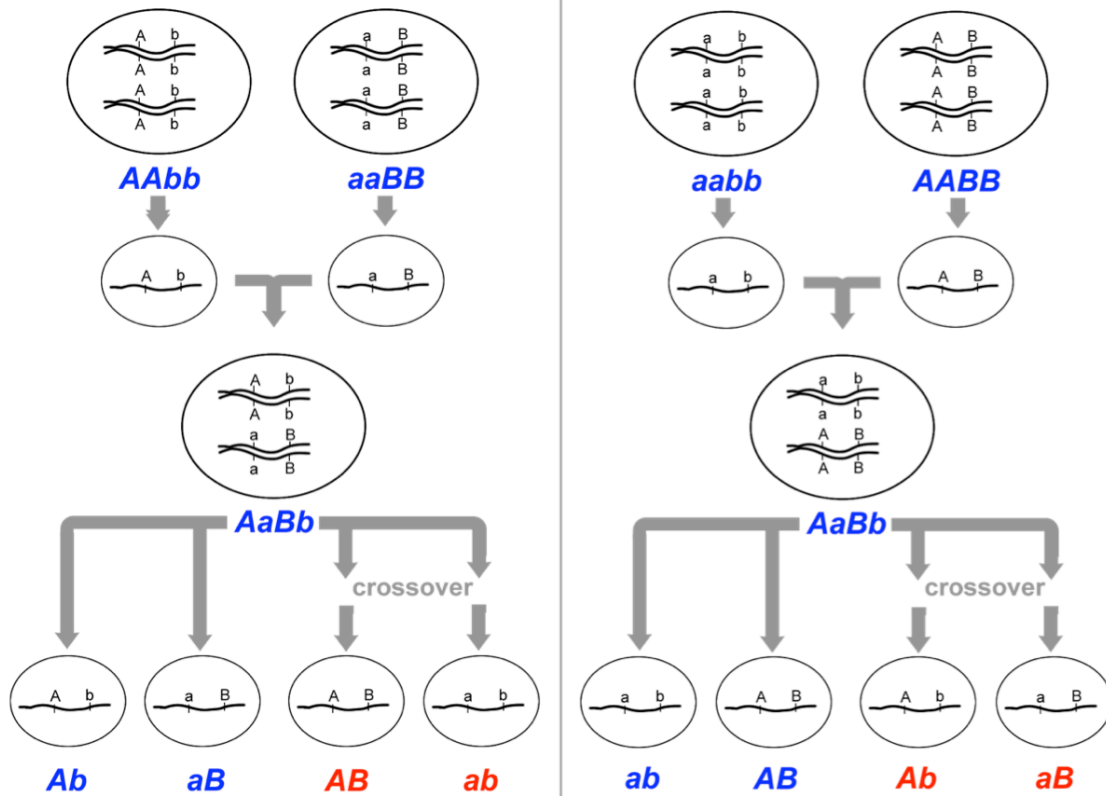
Figure 4.
The genotype of gametes can be inferred unambiguously if the gametes are produced by homozygotes. However, recombination frequencies can only be measured among the progeny of heterozygotes (i.e. dihybrids). Note that the dihybrid on the left contains a different configuration of alleles than the dihybrid on the right due to differences in the genotypes of their respective parents. Therefore, different gametes are defined as recombinant (red) and parental (blue) 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 BY-NC 3.0)

2.3. COUPLING AND REPULSION CONFIGURATION

When looking at an organism that is heterozygous at two loci, just by looking at them you cannot tell how the mutant and wild type alleles are arranged. Both mutant alleles could be on one homologous chromosome, and both wild type alleles could be on the other (e.g. a^-b^- / A^+B^+). This is known as a **coupling** (or cis) **configuration**. When one wild type allele and one mutant allele are on one homologous chromosome, and the opposite is on the other, this is known as a **repulsion** (or trans) **configuration** (e.g. A^+b^- / a^-B^+). The way to determine the orientation is to look at the parents (or P generation) of that cross if you know the genotypes of them. If the parents are homozygous for both genes, and one shows both dominant phenotypes and the other shows both recessive
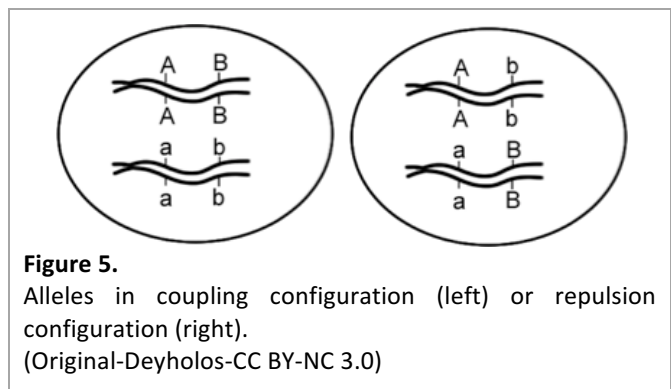
Figure 5.
Alleles in coupling configuration (left) or repulsion configuration (right).
(Original-Deyholos-CC BY-NC 3.0)

phenotypes, then you know that the individual you are looking at is in coupling configuration. If one parent has one dominant and one recessive phenotype, and the other has the opposite, then you know the individual is in repulsion configuration.

2.4. RECOMBINATION FREQUENCY

Recombination frequency (**RF**) is a calculation to define the number of parental and recombinant gametes. The equation is as follows:

$$RF = \frac{\#\ \textbf{Recombinant gametes}}{(\#\ \textbf{Recombinant} + \#\ \textbf{Parental})}$$

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

Based upon the equation and independent assortment, you can see that the recombination frequency cannot be higher than 0.50. If alleles are assorting independently, there will a random distribution of the alleles in the progeny, and so 50% will be recombinant gametes and 50% will be parental gametes, making the RF approximately 0.50. If a gene is linked you will see a higher percentage of parental gametes, making the RF < 0.50. You will never see recombinant gametes more than parental, and so in no situation will recombination frequency be higher than 0.50, except slightly with regards to standard experimental error. If you calculate a recombination frequency higher than 0.50, you need to make sure you accurately defined parental and recombinant gametes.

3. UNLINKED GENES AND COMPLETE AND PARTIAL LINKAGE

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

Table 1 shows generally how we categorize the degree linkage using recombination frequency. Because RF is based upon experimental results that will have some experimental error, these should be treated as guidelines and not hard rules in determining the distance between genes.

Linkage Description	Recombination Frequency
Unlinked	~0.50 or ~50%
Partial linkage	<0.30 or 30%
Complete linkage	0.00 or 0%

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

3.1. UNLINKED GENES

Unlinked genes appear to segregate and show independent assortment. There will be a random and even distribution of gamete types, and an RF of 0.50 is the expectation. This situation occurs in two instances: either when the genes are on completely different chromosomes, or when they are far enough apart on a single chromosome that the crossovers are so numerous that the alleles are distributed randomly (**Figure 3**). Either way, because the alleles are assorting independently you should observe an equal number of recombinant and parental gametes, with an RF near ~0.50. Note, because of real life variability this value can be anywhere from ~0.40 to ~0.60.

3.2. COMPLETE LINKAGE

Having considered unlinked loci, let us turn to the opposite situation, in which two loci are so close together on a chromosome that the parental combinations of alleles always segregate together (**Figure 6**). This is because the physical distance between the two loci is so short that crossover events become extremely rare. Therefore, the alleles at the two loci are physically attached to the same chromatid and will nearly always segregate together into the same gamete. In this case, no recombinants will be present following meiosis, and the recombination frequency will be 0.00. This is **complete** (or **absolute**) **linkage** and is rare, as the loci must be so close together that crossovers are virtually impossible to detect.

Figure 6.
If two loci are completely linked, their alleles will segregate in combinations identical to those present in the parental gametes (*Ab*, *aB*). No recombinants will be observed.
(Original-Deyholos-CC BY-NC 3.0)

Figure 7.
A crossover between two linked loci can generate recombinant genotypes (*AB*, *ab*), from the chromatids involved in the crossover. Remember that multiple, independent meioses occur in each organism, so this particular pattern of recombination will not be observed among all the meioses from this individual.
(Original-Deyholos-CC BY-NC 3.0)

3.3. PARTIAL LINKAGE

It is also possible to obtain recombination frequencies between 0% and 50%, which is a situation we call **incomplete** (or **partial**) **linkage**. Incomplete linkage occurs when two loci are located on the same chromosome but the loci are far enough apart so that crossovers occur between them during some, but not all, meioses (**Figure 7**). Genes that are on the same chromosome are said to be **syntenic** regardless of whether they are completely or incompletely linked or unlinked. Thus, all linked genes are syntenic, but not all syntenic genes are linked.

Because the location of crossovers is essentially random for any given base pair of 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 separated

from each other, will on average have multiple crossovers between them and they will behave indistinguishably from physically unlinked loci. 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 sufficiently separated on the same chromosome.

4. EXPERIMENTALLY DETERMINING RECOMBINATION FREQUENCY

Let us now consider a complete experiment in which our objective is to measure recombination frequency (**Figure 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

Figure 8.
An experiment to measure recombination frequency between two loci. The loci affect coat color (B/b) and tail length (A/a).
(Wikipedia-Modified Deyholos-CC BY-NC 3.0)
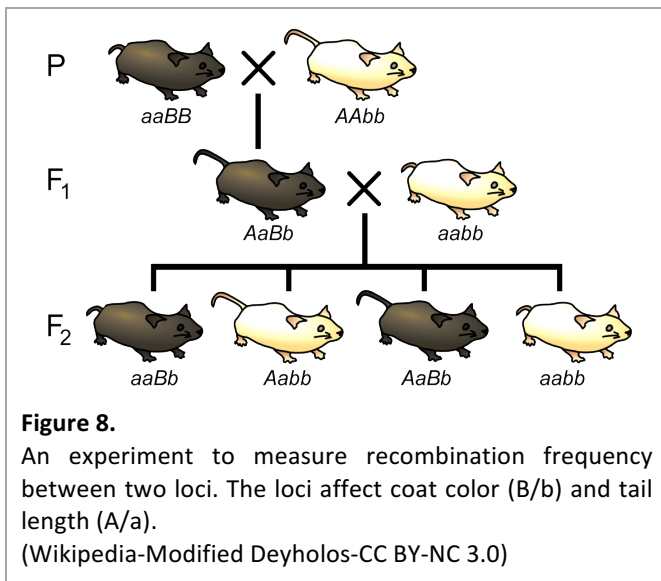
alleles at two loci. For example, we could cross short-tailed (aa), brown mice (BB) with long-tailed (AA), white mice (bb). Thus, (aaBB) are short-tailed and brown, while (AAbb) are long-tailed and white (**Figure 8** P cross). Based on the genotypes of the parents, we know that the parental gametes will be *aB* or *Ab* (but <u>not</u> *ab* or *AB*), and all of the progeny will be dihybrids, *AaBb*. We do not know at this point whether the two loci are on different chromosomes, or whether they are on the same chromosome, and if so, how close together they are.

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

Now that we have identified the parental and recombinant gametes, how do we determine the genotype of the gametes produced by the dihybrid individual? The most practical method is to use a testcross (**Figure 8** F$_1$ to tester), in other words to mate *AaBb* to an individual that has only recessive alleles at both loci (*aabb*). This will give a different phenotype in the second generation for each of the

four possible combinations of alleles in the gametes of the dihybrid (**Figure 9**).

Figure 9.
Punnett Square of example test cross. Homozygous recessive tester can only produce one gamete type so only one is listed. Phenotypes are listed below. Using the phenotypes and what we know of the parents, we can identify which phenotypes came from recombinant or parental gametes. (Original-L. Canham- CC BY-NC 3.0)

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 F2 in ratios such as shown in **Table 2**. Given the data in **Table 2**, the calculation of recombination frequency is straightforward:

RF = # recombinant offspring / Total offspring

RF = 13+17 / 48+42+13+17

= 0.25

Because the recombination frequency is below 0.30, we can say that the tail length gene and the fur colour gene are partially linked.

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

tail phenotype	fur phenotype	number of progeny	gamete from dihybrid	genotype of F₂ 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 2.
An example of quantitative data that may be observed in a genetic mapping experiment involving two loci. The data correspond to the F₂ generation in the cross shown in **Figure 8**.

_____

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.

KEY TERMS:

linkage	unlinked
Second Law of Independent Assortment	synapsis
gene	recombinant genotype (and gametes)
locus	parental genotype (and gametes)
allele	coupling (cis) configuration
genotype	repulsion (trans) configuration
phenotype	recombination frequency (RF)
recombination	complete (absolute) linkage
interchromosomal recombination	incomplete (partial) linkage
independent assortment	syntenic
intrachromosomal recombination	
crossover	

STUDY QUESTIONS:

1) Compare the terms "recombination" and "crossover". How are they similar? How are they different?

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

3) Suppose you knew that in a population, a trait (allele at a locus) that dominantly affected earlobe shape was tightly linked to a trait that dominantly affected susceptibility to cardiovascular disease in humans. Under what circumstances would this information be clinically useful?

4) In a previous chapter, we said a 9:3:3:1 phenotypic ratio was expected among the progeny of a dihybrid cross, in absence of gene interaction.
   a) What does this ratio assume about the linkage between the two loci in the dihybrid cross?
   b) What ratio would be expected if the loci were completely linked? Be sure to consider every possible configuration of alleles in the dihybrids.

5) Given a dihybrid with the genotype *CcEe*:
   a) If the alleles are in coupling (cis) configuration, what will be the genotypes of the parental and recombinant progeny from a test cross?
   b) If the alleles are in repulsion (trans) configuration, what will be the genotypes of the parental and recombinant progeny from a test cross?

6) In this question the white flowers (w) are recessive to purple flowers (W), and yellow seeds (y) are recessive to green seeds (Y). If a green-seeded, purple-flowered dihybrid is testcrossed, and half of the progeny have yellow seeds.
   a) What can you conclude about linkage between these loci?
   b) What do you need to know about the progeny in this case?

7) If the progeny of the cross *aaBB* x *AAbb* is testcrossed, and the following genotypes are observed among the progeny of the testcross, what is the frequency of recombination between these loci?
   AaBb   135
   Aabb   430
   aaBb   390
   aabb   120

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

CHAPTER 19 – RECOMBINATION MAPPING OF GENE LOCI

Figure 1.
Thomas Hunt Morgan and his undergraduate Alfred Henry Sturtevant used fruit fly mutations like the ones in this figure to create the first recombination map.
Eye colors (clockwise): brown, cinnabar, sepia, vermilion, white, wild type. Also, the white-eyed fly has a yellow body, the sepia-eyed fly has a black body, and the brown-eyed fly has an ebony body.
(Wikimedia-Ktbn-Public Domain)

INTRODUCTION

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

1. GENETIC MAPPING

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

1.1. CALCULATING MAP DISTANCE

The units of genetic distance are called **map units** (**mu**) or **centiMorgans** (**cM**), in honor of **Thomas Hunt Morgan** by his undergraduate student, **Alfred Sturtevant**, who developed the concept of genetic maps. Geneticists routinely directly convert the recombination frequencies of two loci into cM. Thus, 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 2**).
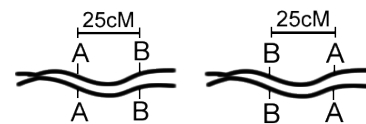
Figure 2.
Two genetic maps consistent with a recombination frequency of 25% between A and B. Note the location of the centromere. (Original-Deyholos-CC BY-NC 3.0)

Note, however, this approximation works well only for small distances (RF<30%) but progressively fails at longer distances. This is because as the two loci get farther apart the RF reaches a maximum at 50%, like it would for two loci assorting independently (not linked). In fact, some chromosomes are >100 cM long but such loci at the tips only have an RF of 50%. Calculating the map distance of the whole chromosome (end-to-end) of over 50cM comes from mapping of multiple loci dispersed along the chromosome, each with a value of less than 50%, with their total adding up to the value over 50cM (e.g. >100cM as above). The method for mapping of these long chromosomes is described next.

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

.2. MAP DISTANCE OVER LONG CHROMOSOMES

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

When a novel gene or locus is identified by mutation or polymorphism, crossing it with previously mapped genes, and then calculating the recombination frequency can determine its approximate position on a chromosome.

Figure 3.
Genetic maps for regions of two chromosomes from two species of the moth, *Bombyx*. The scale at left shows distance in cM, and the position of various loci is indicated on each chromosome. Diagonal lines connecting loci on different chromosomes show the position of corresponding loci in different species. This is referred to as regions of **conserved synteny**.
(NCBI-NIH-PD)
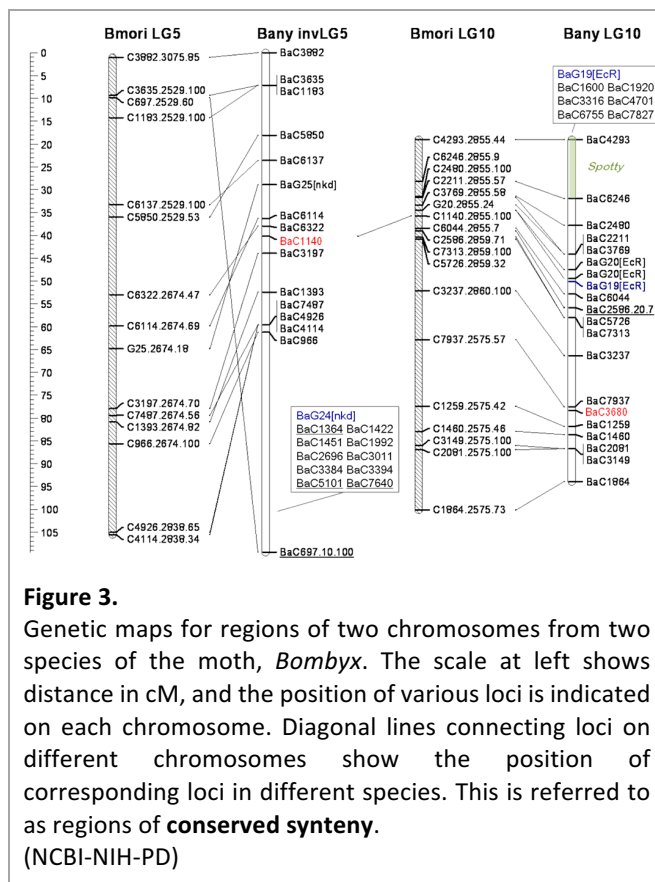
If the novel gene and the previously mapped genes show complete or partial linkage with an existing locus, the recombination frequency will indicate the approximate position of the novel gene within the genetic map. This information is useful in isolating (i.e. cloning) the specific fragment of DNA that encodes the novel gene. This process called **map-based cloning**.

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

1.3. GENETIC MAPS ARE AN APPROXIMATION

Genetic maps are useful for showing the order of loci along a chromosome, but the distances are only a relative 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 (third, or more) crossovers occur between the loci. This is a problem for geneticists, because with respect to the loci being studied, these **double-crossovers** produce gametes with the same genotypes as if no recombination events had occurred (**Figure 4**), so they have parental genotypes. Thus, a double crossover will appear to be a parental type and not be counted as a recombinant, despite having two (or more) crossovers. Geneticists will sometimes use specific mathematical formulae to adjust large recombination frequencies to account for the possibility of multiple crossovers and thus get a better estimate of the actual distance between two loci.

2. MAPPING WITH THREE-POINT CROSSES

A genetic map consists of multiple loci distributed along a chromosome. 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 5**).
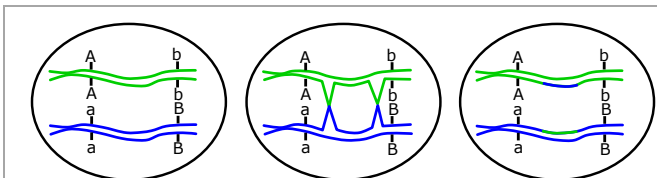
Figure 4.
A double crossover between two loci will produce gametes with parental genotypes, even though TWO crossovers have occurred between the loci.
(Original-Deyholos/Canham-CC BY-NC 3.0)

This is particularly useful when mapping a new mutation for which the location is unknown relative to two previously mapped loci with known locations. The basic strategy is the same as for the dihybrid mapping experiment described previously,

except pure breeding lines with contrasting genotypes are crossed to produce an individual heterozygous at three loci (a trihybrid), which is then testcrossed to a tester, which is homozygous recessive for all three genes, to determine the recombination frequency between each pair of genes, among the three loci. A Punnett square can be used to predict all the possible outcomes of the test cross (**Figure 6**). The progeny produced from the testcross is shown in **Table 1**.

When the trihybrid is crossed to a tester, it should be able to make eight different gametes, to make eight possible different phenotype combinations in the offspring. The next step would be to identify if the alleles are recombinant or parental gametes. This can be done by comparing only two loci at one time to the parental gametes. In this example, the parents of the trihybrid are a/a B/B c/c, and A/A b/b C/C, so the parental gametes would be aBc and AbC respectively. Now by comparing two loci at once you can determine if, between the two, they are recombinant or parental. For example, the offspring in the first row in **Table 1** came from gamete aBC.
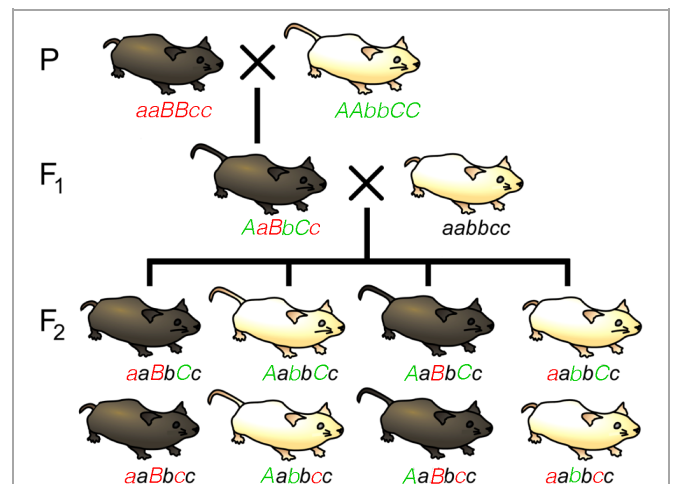
Figure 5.
A three point cross for loci affecting tail length, fur color, and whisker length.
(Original-Modified Deyholos/Locke-CC BY-NC 3.0)

♀ / ♂	P2 aBC	AbC	abC	ABC	P1 aBc	Abc	abc	ABc	Figure 6.
abc	aa Bb Cc	Aa bb Cc	aa bb Cc	Aa Bb Cc	aa Bb cc	Aa bb cc	aa bb cc	Aa Bb cc	Punnett square of the test cross for **Figure 5**, showing the predicted gametes possible from this cross, and the resulting phenotypes. (Original-L. Canham-CC BY-NC 3.0)
phenotype	Short tail Brown Long whis	Long tail White Long whis	Short tail White Long whis	Long tail Brown Long whis	Short tail Brown Short whis	Long tail White Short whis	Short tail White Short whis	Long tail Brown Short whis	

tail phenotype	fur phenotype	whisker phenotype	number of progeny n=120	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 (P2)	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 (P1)	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 1.
An example of data that might be obtained from the F_2 generation of the three-point cross is shown in **Figure 5**. The rarest phenotypic classes correspond to double recombinant gametes *ABc* and *abC*. Each phenotypic class and corresponding gamete can also be classified as parental (P) or recombinant (R) with respect to how each pair of loci (A,B), (A,C), (B,C) are arranged on the chromosome.

Comparing loci A and B, we see that it matches one of the parental gametes and therefore it is parental. Comparing A and C we see that it matches neither parental, so it is recombinant. The same can be said for comparing B and C.

loci A,B	$RF = \frac{1+16+12+1}{120} = 25\%$
loci A,C	$RF = \frac{1+5+1+5}{120} = 10\%$
loci B,C	$RF = \frac{5+16+12+5}{120} = 32\%$ (not corrected for double crossovers)

Once the classes of progeny have been identified as each pair of locus being parental or recombinant,

recombination frequencies may be calculated for each pair of loci individually, as we did before for one pair of loci in our dihybrid cross (Chapter 18). We can then use these numbers to build the map, placing the loci with the largest RF on the ends.

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%). This is because of **double crossovers** between B and C, which were undetected when we considered only pairwise data for B and C. We can easily account for some of these double crossovers, and include them in calculating the map distance between B and C, as follows (**Figure 7**).

Figure 7.
Two possible maps based on the data in Table 1 (without correction for double crossovers).
(Original-Deyholos-CC BY-NC 3.0)

gene is discovered, it can be mapped relative to other genes of known location to determine its location. All that is needed to map a gene is two alleles, a wild type allele and a mutant allele.

3. ANALYSIS OF RECOMBINATION FREQUENCIES IN A THREE POINT TEST CROSS

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

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

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

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

In the example given above, all the genes present are linked, with one pair more strongly linked than the other (A and C have stronger linkage than A and B). When choosing three genes to map, this will not always be the case. Sometimes you will have all genes linked, sometimes you may have two genes linked and one gene unlinked, and sometimes they all may be unlinked (**Figure 8**). Much like what we did above, by comparing the ratios of offspring you should be able to predict if the genes in the trihybrid are linked or not.

We already deduced that the map order must be *BAC* (or *CAB*). 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	$RF = \frac{5+16+12+2(1)+2(1)}{120} = 35\%$ (corrected for double crossovers)

Thus, the three-point cross was useful for:

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

However, it is possible that other, double crossovers events remain undetected, 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 such 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

Figure 8.
Examples of how three genes can be associated with each other, based on whether all three are unlinked, all three are linked or two linked and one unlinked.
(Original-J. Locke/L. Canham-CC BY-NC 3.0)

If all three genes are unlinked, then we expect independent assortment and an equal number of all progeny types. Like in the example, if all are linked, you expect there to be <u>many</u> parental genotypes, <u>some</u> recombinant genotypes if they are a result of a single recombination events. Recombinant genotypes that are a result of two recombination events will be <u>rare</u>. The actual numbers of each will differ depending if all the linked genes are equal distances from each other, or if one pair is more linked than the other.

In the case where two genes are linked and one gene is unlinked the following applies: as in the example before we will use the same parental gametes (AbC and aBc), but will assume the genes A and C are linked and B is unlinked. In this case, because linkage causes a higher prevalence of parental gametes, we expect there to be <u>more</u> parental organizations of A and C, and <u>fewer</u> recombinant organizations of A and C. The presence and or absence of parental B is not important here, because it is unlinked and will assort independently.

This information is summarized in **Table 2**. You can use this to look at the offspring of a trihybrid test cross and predict the linkage ahead of time.

gamete	# of progeny	gamete	# of progeny
aBC	5	Abc	5
ABC	16	abc	12
ABc	1	abC	1

Figure 9.
Diagram of the crossover events to create the different recombinant gametes from the cross in Figure 5. The parental alleles are seen on the black chromosomes. The coloured lines indicate show where the crossover event took place and underlines the alleles for that recombinant gamete. Below each diagram is the recombinant gamete and the number of progeny seen in that cross per Table 1.
(Original-L. Canham-CC BY-NC 3.0)

4. WHERE DO CROSSOVERS OCCUR ON A CHROMOSOME?

4.1. GENERAL INFORMATION

A crossover involves the reciprocal exchange between non-sister chromatids when synapsed at prophase I of meiosis. While this exchange can theoretically occur anywhere along the synapsed homologs, observations show us that some regions along a chromosome have higher rates of crossing

over, while others are lower. In addition, the frequency of crossing over varies from species to species, and even from male to female within a species. For example, in *Drosophila melanogaster* there is no crossing over in males.

From Drosophila recombination data, we know that the likelihood of a crossover is greatest in the middle of a chromosome arm and lower at the telomere and centromere regions (**Figure 7**). This distribution would be expected if one of the functions of a crossover event were to hold the two synapsed chromosomes together so that they segregate correctly in metaphase I of Meiosis I.

4.2. RESOLUTION OF GENETIC MAPS

The resolution of genetic maps depends on two factors: (1) the <u>number</u> of marker loci and, (2) the number of progeny.

For species with a high number of marker loci (those which have a phenotype that permits the alleles to be distinguished), more locations can be plotted ,

resulting in a higher resolution map compared to species with fewer markers.

For species with a greater number of progeny, a better map is possible. The ability to score recombinants among 100's, 1000's, etc. means that one can identify rare or very rare recombinants and thus map loci that are very close together. For example, with the mapping of bacteriophage, it is possible to map mutations down to the level of single base pairs using certain selectable marker systems.

Because of these two factors, the genetic maps of simple prokaryote genomes are more refined than those of the larger and more complex eukaryote genomes.

These days, most laboratory species have had their genomes sequenced. This knowledge provides another means to locate the specific gene(s) responsible for a desired trait(s).

Gametes	Parental or Recombinant	Unlinked	A and C linked	All linked
AbC	P	1	more	many
aBc	P	1	more	many
ABC	R	1	more	some
abc	R	1	more	some
aBC	R	1	less	some
Abc	R	1	less	some
abC	R	1	less	rare
ABc	R	1	less	rare

Table 2.
Progeny ratios seen after a trihybrid test cross depending on whether they are all linked, only two are linked or if all are unlinked.
This table is based upon the cross done in **Figure 5** as an example.
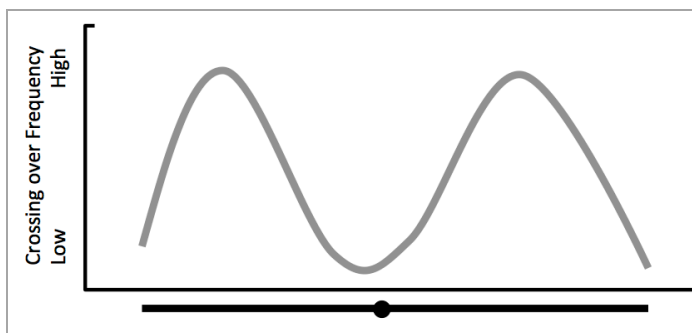
Figure 10.
Diagram of the frequency of crossing over along a chromosome (bottom). The Y-axis shows the relative rate of crossing over. The two peaks are present in the middle of each chromosome arm, while the telomeres and centromeres have lower frequencies of exchanges. (Original-J. Locke-CC BY-NC 3.0)

_____

SUMMARY:

- A **genetic map** (or recombination map) is a representation of the linear order of genes (or loci), and their relative distances determined by crossover frequency, along a chromosome.

- Recombination frequency is usually proportional to the distance between loci, so recombination frequencies can be used to create genetic maps.

- Recombination frequencies tend to underestimate map distances, especially over long distances, since double crossovers may be indistinguishable from non-recombinants.

- Three-point crosses can determine the order and map distance among three loci.

- In three-point crosses, a correction for the distance of the outside markers can be made to account for double crossovers between the two outer loci.

- Crossovers are not equally frequent all along a chromosome. In some regions, crossovers are more frequent while others are less.

- The resolution of genetic maps depends on the number of markers and the number of progeny.

KEY TERMS:

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

STUDY QUESTIONS:

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

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

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

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

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

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

a) Calculate the map distance between B and C.
b) Why are the frequencies of the two smallest classes not exactly the same?

6) Given the map distance you calculated between B-C in question 5, if you crossed a double mutant (i.e. yellow body and curved wing) with a wild-type fly, and testcrossed the progeny, what phenotypes in what proportions would you expect to observe among the F_2 generation?

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

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

Recommended course:

Introduction to Molecular Genetic Techniques

GENET 375, Introduction to Molecular Genetic Techniques
★3 (fi 6) (0-1s-6)
A laboratory course in which students will be introduced to modern techniques in molecular genetics. These may include cytogenetics, recombinant DNA techniques, PCR, DNA sequencing, methods of detecting gene expression, and genome analysis. Prerequisites: GENET 270, MICRB 265, and a 300-level GENET course. Enrolment is limited, and registration is by consent of instructor.
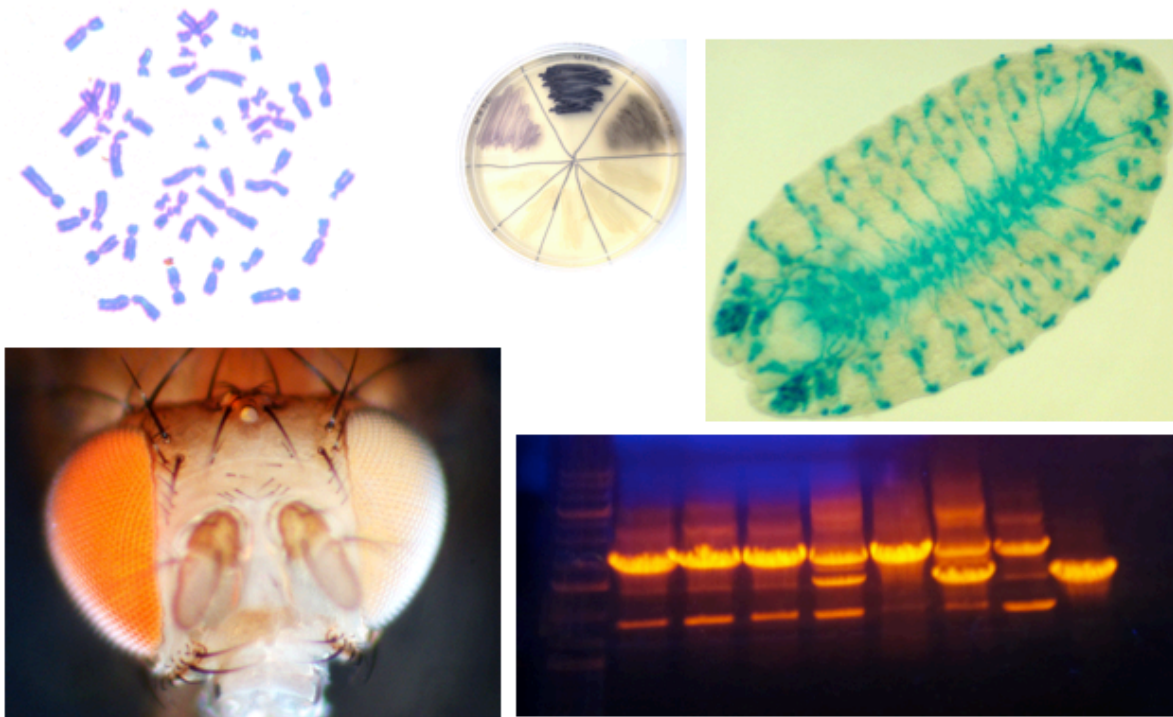
Prerequisite: GENET 270, MICRB 265, and a 300-level GENET
Offered: Winter Term. Seminar Monday 10:00-10:50AM

Lab MT 2:00-5:00 PM or ThF 2:00-5:00 PM
Contact: Dr. J. Locke, john.locke@ualberta.ca

Calendar link:http://calendar.ualberta.ca/preview_course_nopop.php?catoid=6&coid=44627

Chapter 20 – Sex Chromosomes: Sex Linkage

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

Introduction

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

1. Autosomes and Sex Chromosomes

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

The combination of sex chromosomes within a species is associated with either male or female individuals. In mammals, fruit flies, and some dioecious plants, those with two X chromosomes are females while those with an X and a Y are males. In birds, moths, and butterflies, males are 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 (see the next chapter).

How do the sex chromosomes behave during meiosis? Well, in those individuals with two of the same chromosome (i.e. **homogametic** sexes: 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.

In mammals (XX, XY) 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

Figure 2.
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 BY-NC 3.0)

Figure 3.
X and Y chromosome have pseudoautosomal regions, which are capable of pairing during meiosis and recombination. (Original-Locke/Kang-CC BY-NC 3.0)
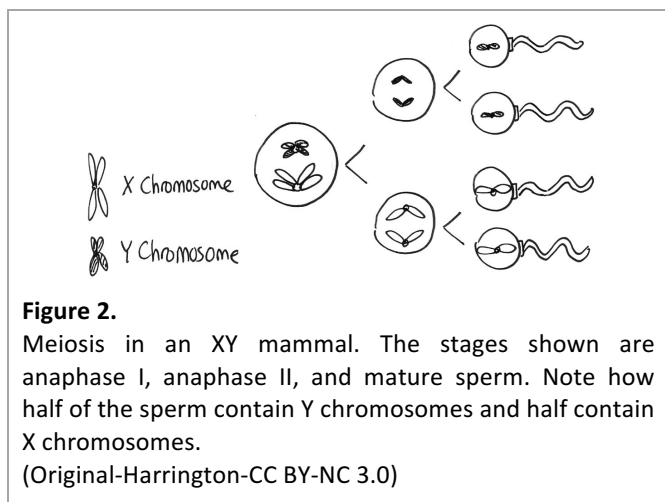
X chromosomes and become female while half will receive an X and a Y and become male (**Figure 2**). In species with ZZ males, all sperm carry a Z chromosome, while in females, ZW, half will have a Z and half a W.

2. PSEUDO-AUTOSOMAL REGIONS ON THE X AND Y CHROMOSOMES

In evolution, before the X and Y chromosomes differentiated, they used to be equivalent homologs, like an autosome. Over time, the Y chromosome lost most of its genes (hence the reduced size), but the X chromosome retained all its genes. Thus, even though the Y chromosome has lost most of its genes, it still shares some regions with the X chromosome. This is the reason why although X and Y chromosomes are **heteromorphic** (morphologically dissimilar), they are able to act as a homologous pair in meiosis and undergo crossover. These common regions, contain similar genes, permit the X and Y to pair up and are called the "**pseudoautosomal regions**". The name comes from the observation that genes in these regions behave like autosomes in their inheritance. Alleles of the genes in this region crossover just like those on the autosomes. Thus, genes in this region are not inherited in a sex-linked pattern, even though they are located on the X chromosome.

The genes found in pseudo-autosomal region are present in two copies in both XY males and XX females and thus if expressed from both active and
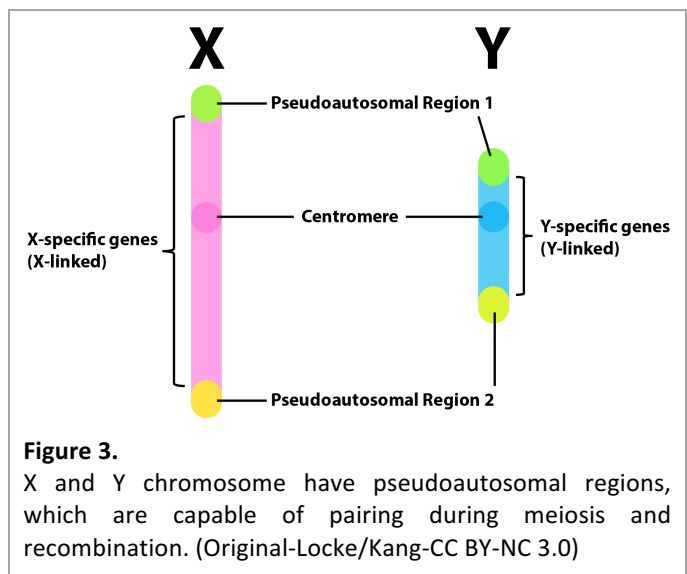
inactive X chromosomes. These genes may explain clinical features in sex chromosome aneuploidy (addition or subtraction of a sex chromosome; e.g. XXY) as gene products may be either under or over expressed in relation to normal females and males. One of the genes in this region is called *SHOX*. It makes 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 pseudo-autosomal region that causes health problems for women with Turner syndrome.

3. SEX LINKAGE: AN EXCEPTION TO MENDEL'S FIRST LAW

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

3.1. 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.

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 5**). 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 5**. 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).

4. Y-LINKED GENE

Genes located on the Y-chromosome exhibit Y-linkage. For example, the TDF gene that is responsible for sex determination and hairy ear rim phenotype show only father to son inheritance pattern.
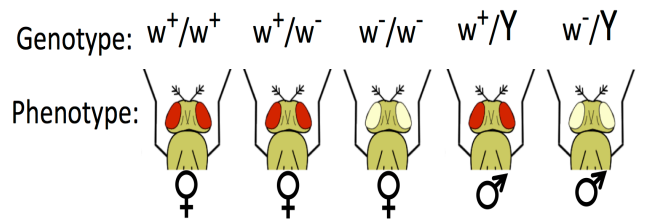
Figure 4.
Relationship between genotype and phenotype for the *white* gene on the X-linked gene in *Drosophila melanogaster*. The Y chromosome is indicated with a capital Y because it does not have a copy of the *white* gene.
(Original-Deyholos/Harrington/Locke-CC BY-NC 3.0)
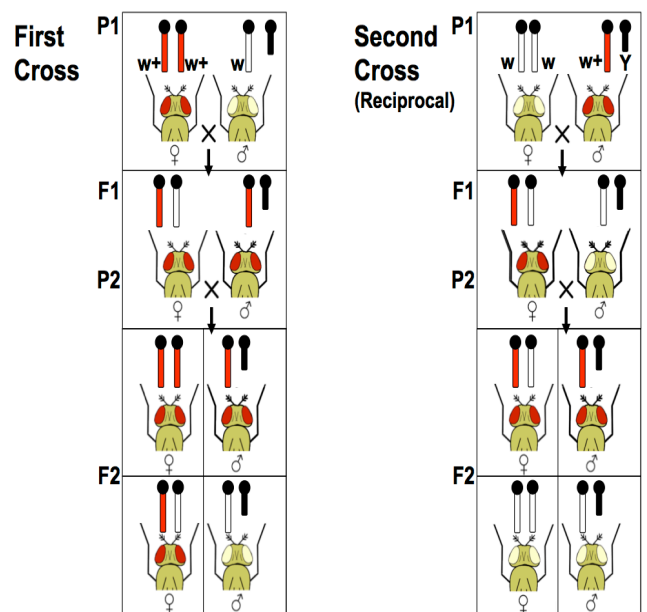
Figure 5.
Reciprocal crosses involving an X-linked gene in *Drosophila melanogaster*. In the first cross (left) all of the offspring have red eyes. In second (reciprocal) cross (right) all of the female offspring have red eyes and the male offspring all have white eyes. If the F1 progeny are crossed (to make the P2), the F2 progeny will be different in each cross. The first cross has all red-eyed females and half red-eyed males. The reciprocal cross has half red-eyed males and females.
Thomas Morgan was awarded the Nobel Prize, in part, for using these crosses to demonstrate that genes (such as *white*) were on chromosomes (in this case the X-chromosome).
(Wikipedia- Deyholos/Harrington/Locke -PD)

5. Z-linked Genes in Birds

One last example is a **Z-linked gene** that influences feather colour in turkeys. Turkeys are birds, which use the ZZ-ZW sex chromosome system. The *E* allele makes the feathers bronze and the *e* allele makes the feathers brown (**Figure 6**). 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 6.
Relationship between genotype and phenotype for a Z-linked gene in turkeys. The W chromosome does not have an *E/e*-gene so it is just indicated with a capital W. (Original-Harrington/Locke-CC BY-NC 3.0)

_____
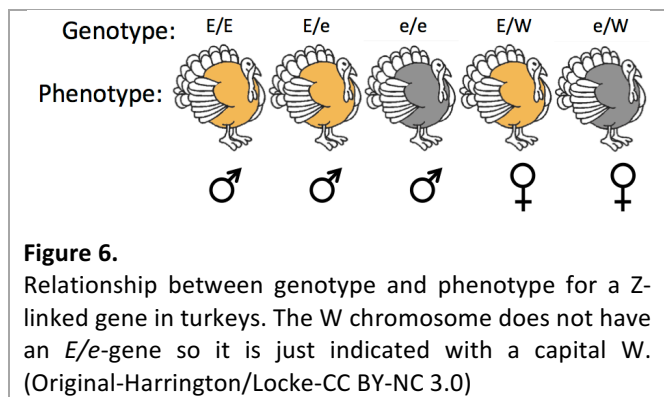
SUMMARY:

- Autosomes and sex chromosomes differ in that the former exist in pairs but the latter depends on the sex of the chromosome.

- Pseudo-autosomal regions are regions on X and Y chromosome that can pair up and recombine.

- Sex-linked genes are an exception to standard Mendelian inheritance. Their phenotypes are influenced by the type of sex chromosome system and the type of dosage compensation system found in the species.

- Some of the examples of sex-linked genes are: *white* gene on the Drosophila's X chromosome, *TDF* gene on Y chromosome, *E/e* gene on Z chromosome.

KEY TERMS:

autosome	heteromorphic
sex chromosome	sex-linked
homogametic	X-linked genes
heterogametic	reciprocal cross
pseudoautosomal regions	Z-linked genes

STUDY QUESTIONS:

1) 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?
2) Make Punnett Squares to accompany the crosses shown in **Figure 5**.
3) Draw reciprocal crosses that would show that the turkey E-gene is on the Z-chromosome.

CHAPTER 21 – SEX CHROMOSOMES: SEX DETERMINATION

Figure 1.
Not all species determine sex using the same mechanism. There are many factors that can determine a species' sex and one of them is growth temperature. For alligators, sex is determined by the temperature of the eggs in their nest. (Flickr-Florida Fish and Wildlife-CC BY-ND 2.0)

INTRODUCTION

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

1. SEX DETERMINATION MECHANISMS IN ANIMALS

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

1.1. SEX CHROMOSOME SYSTEMS:

a) XY system

Different combinations of the X and Y sex chromosomes can determine the sex of an organism. 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 **Sex-determining Region Y** (**SRY**) gene, also known as **Testis-Determining Factor** (**TDF**) gene, on the Y-chromosome. Its presence in the genome and expression in gonad tissues dictates that the sex of that individual will be male. Its absence or lack of correct expression results in a female phenotype for that individual.

In mammals, the sex chromosomes evolved just after the divergence of the monotreme lineage (mammals that lay eggs) from the lineage that led to marsupial mammals (young are carried in a pouch) and placental mammals. Thus nearly every mammal species uses the same sex determination system. In this system, during embryogenesis, the gonads will develop into either ovaries or testes. **(Figure 2)**
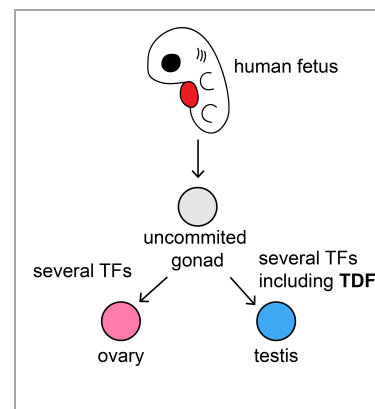
Figure 2.
Gonad differentiation is under the control of several genes including *Testis-determining factor (TDF, SRY)* at Yp11.3. (y chromosome, p arm, region 1, band 1, sub-band 3).
(Original-Harrington/Kang-CC BY-NC 3.0)

A gene, present only on the Y chromosome called *SRY,* encodes a protein that directs the gonads to mature into testes. XX embryos do not have this gene and their gonads mature into ovaries instead, a default (**Figure 2**).

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

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

b) ZW system

In birds, some fish, some insects (butterflies and moths) and reptiles, they use different chromosome for sex determination, Z- and W-chromosomes. Z-chromosome is larger and has more genes than the W-chromosome. ZZ embryos become male and ZW embryo become females. This sex linkage pattern is backwards of the X and Y sex linkage pattern. It is currently unknown if the presence of W chromosome induces female features or two copies of Z chromosome induces male features. In birds, researches have not yet found a ZZW or Z0 individual.

c) X/O system

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

1.2. Chromosome dosage

a) X-Autosome Ratio

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

b) Ploidy Level

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

2. Environmental Factors

2.1. Growth temperature

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

Tuatara (**Figure 3**) – These reptiles look like lizards but are a distinctly separate Order, which has survived for over 200 million years. There are currently only two extant species. Embryo's development temperature determines the animal's

sex; low temperatures (below a threshold) develop into females. High temperatures (above a threshold) develop into males. Global warming will affect the sex ratio in the population. By 2080 there will be conditions that produce 100% males.

2.2. SOCIAL ORGANIZATION

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

2.3. PARTHENOGENETIC SPECIES

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

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

Figure 4.
Moon Wrasse (*Thalassoma lunare*) can change sex.
(Flickr- Nick Hobgood- CC BY-NC 2.0)
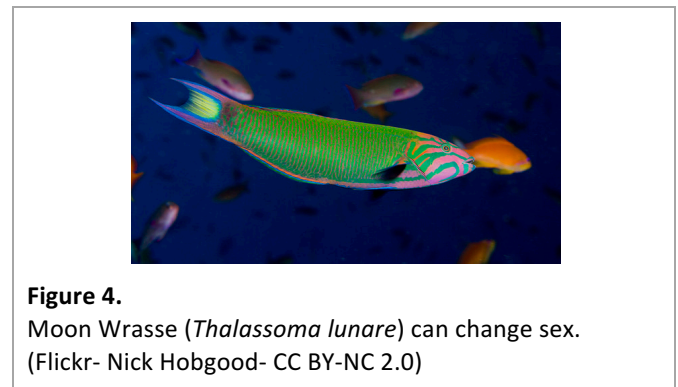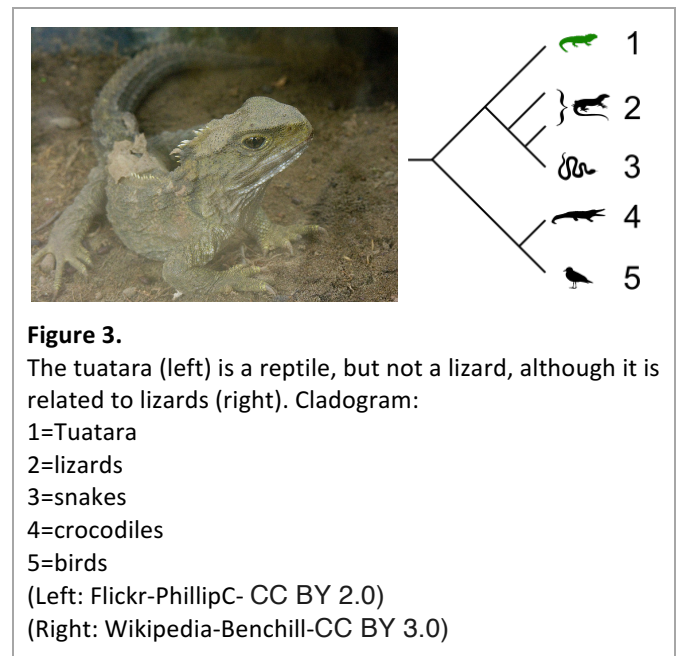
Determining Factors	Genetic Mechanism	Cell Response Mechanism
Chromosomal: • XX/XY • ZW/ZZ • XX/XO • Haploid/Diploid	• Single gene • X-Autosome Ratio (gene dosage)	• Hormonal: directs cells to sex phenotype • Cell-autonomous (each cell "knows" what sex it is)
Environment: • Rearing temp. • Social interactions • Parthenogenesis	Not genetic	Hormonal?

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

_____

SUMMARY:

- The sex of an individual can be determined by sex chromosomes

- This includes the X/Y, Z/W, and X/O system

- Also, differences in the ploidy level (haploid vs diploid) determine sex in some species

- Lastly, environmental factors such as rearing temperature or social organization (male vs female ratio) can determine sex.

KEY TERMS:

single gene
Testis-Determining Factor (*TDF*)
Sex-determining Region Y (*SRY*)
therians
XY system
ZW system

X/O system
X:Autosome (X:A) ratio
haploid-diploid system
Tuatara
Sex-ratio
parthenogenetic

STUDY QUESTIONS:

1) Draw reciprocal crosses that would demonstrate that the turkey E-gene is on the Z chromosome.

2) Mendel's First Law (as stated in class) does not apply to alleles of most genes located on sex chromosomes. Does the law apply to the chromosomes themselves?

Recommended course:

Research Techniques in Molecular Genetics

GENET 420, Research Techniques in Molecular Genetics
★6 (fi 12) (0-1s-12) Fall Term
A laboratory course teaching modern techniques in molecular biology with emphasis on the analysis of gene expression in animal systems.

The zebrafish is the central experimental system in this senior molecular biology laboratory. We have made this as authentic a research experience as an instructional lab can be. A specific technique is taught each week with each student working on a different gene or protein, allowing them to take ownership of their individual successes and removing the expectation of a single correct answer. At the end of the term, students present their data in the form of an original manuscript, complete with figures and discussion.

The students have access to the lab outside of scheduled class time, and repeat experiments that do not work (with guidance from the instructor, technologist and TA). The pedagogy of the course is critical; oral exams and frequent discussions are used to track the students understanding, their progress, and the significance of their results.

Note that this is a **6-credit** course, two days per week, designed for fourth or fifth year undergraduate and graduate students in molecular genetics or cellular biology programs. Prerequisites for undergrads are GENET 390 (may be co-requisite) and one other GENET 3XX lecture course, and GENET 375 is recommended. Enrolment is limited and registration is by consent of instructor. See link below for application form.

Offered: Fall Term. Labs are Thursday & Friday 2-5 pm and Seminars Tuesday 2-3 pm

Contact: Dr. Dave Pilgrim
dave.pilgrim@ualberta.ca

Link: http://calendar.ualberta.ca/preview_course_nopop.php?catoid=6&coid=44632
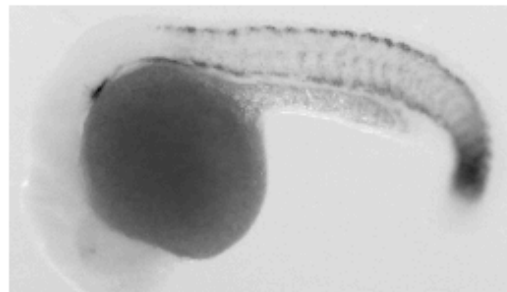
CHAPTER 22 – SEX CHROMOSOMES: DOSAGE COMPENSATION
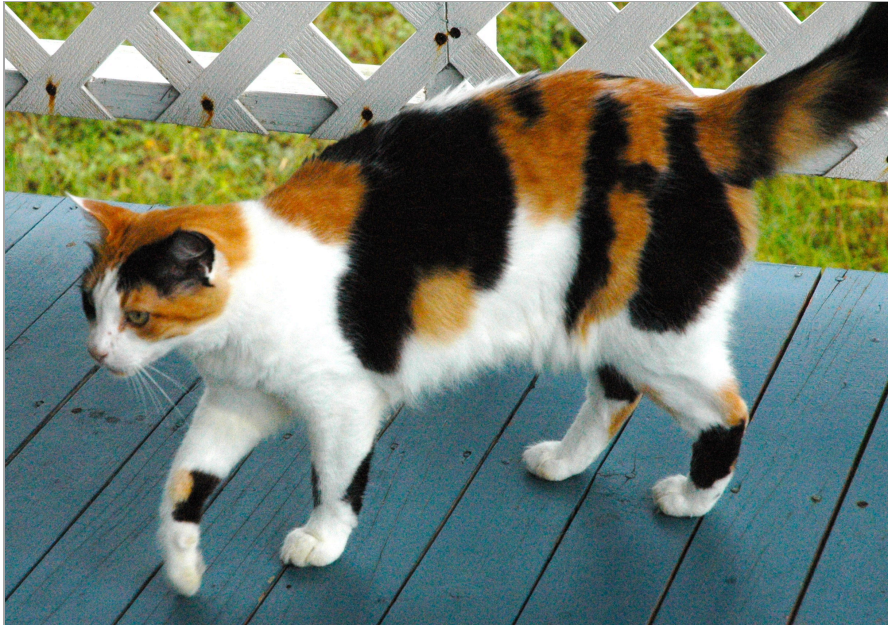
Figure 1.
A calico cat showing the random inactivation (X-inactivation) of one or the other X-chromosome giving either an orange or black fur colour. The inactivation is a mechanism of dosage compensation. (Note: the white colour pattern is due to another gene.)
(Original-J. Locke-CC:AS)

INTRODUCTION

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

1. GENE DOSAGE PROBLEM

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

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

2. DOSAGE COMPENSATION IN DROSOPHILA

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

3. X-CHROMOSOME INACTIVATION IN MAMMALS

3.1. BASICS

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

The X_i chromosome is replicated during S phase and transmitted during mitosis the same as any other chromosome, but most of its genes are not transcribed (**Figure 2**). The chromosome appears as a condensed mass within interphase nuclei and is called the **Barr body** (**Figure 3**) and does not decondense to be expressed. (The Barr body is named after Canadian researcher, Murray Barr, who along with his graduate student Ewart Bertram at Western University in London Ontario discovered it in 1948.) With the inactivation of genes on one X-chromosome, females have the same number of functioning X-linked genes as males. However, some genes and particularily those in the pseudoautosomal regions escape

Figure 2.
X chromosome inactivation during mitosis and after mitosis.
(Original-Harrington/Kang-CC BY-NC 3.0)

Figure 3.
X chromosomes detected by FISH method in a female cell's nucleus.
(Wikipedia-Steffen Dietzel-CC BY-SA 3.0)
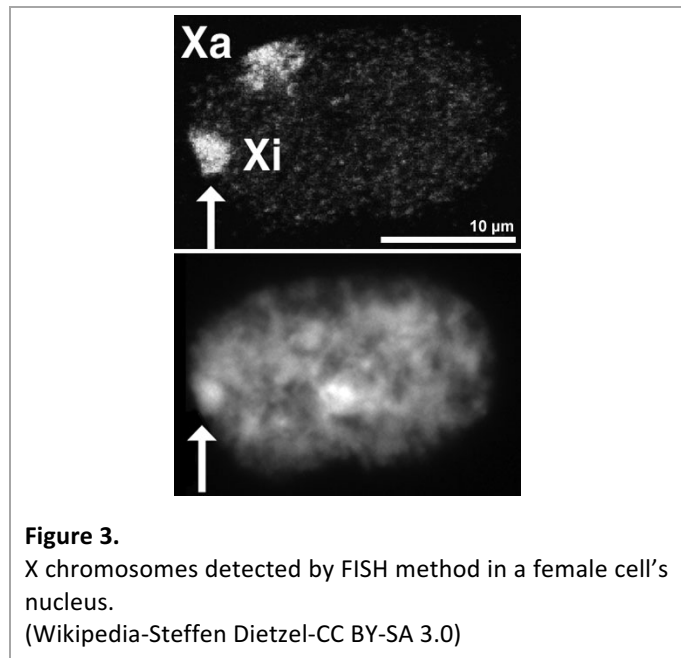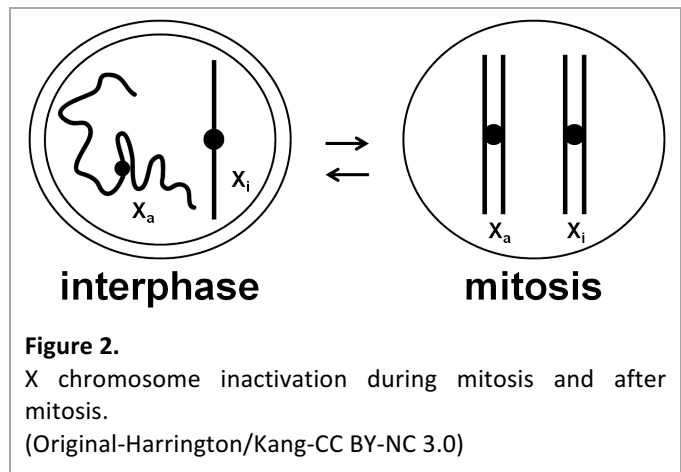
inactivation and the alleles are expressed from both active and inactive X chromosomes. These genes may explain clinical features in sex chromosome aneuploidy as gene products may be either under or over expressed in relation to normal females and males.

3.2. X-LINKED GENES – ORANGE GENE IN CATS

A classic X-linked gene that shows X-inactivation is the *Orange* gene (*O*) in cats. The O^O allele encodes an enzyme that results in orange pigment in the fur hairs. The O^B allele results in the hairs being black. The phenotypes of various genotypes of cats are shown in **Figure 4**. 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-

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

Figure 5.
This figure shows the two types of liver cells in females heterozygous for an *F8* mutation. Because people with the $F8^+/F8^-$ genotype have the same phenotype, normal blood clotting, as $F8^+/F8^+$ people the $F8^-$ mutation is classified as recessive. .
(Original-Harrington/Locke-CC BY-NC 3.0)

chromosomes inactivated. In each orange hair the X_i chromosome carrying the O^B allele is inactivated. The O^O allele on the X_a is functional and orange pigments are made. In black hairs the reverse is true, the X_i chromosome with the O^O allele is inactive and the X_a chromosome with the O^B allele is active. Because the inactivation decision happens early during embryogenesis, the cells continue to divide to make large patches on the adult cat skin where one or the other X is inactivated.

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

3.3.   A TYPICAL X-LINKED GENE – F8 GENE IN HUMANS

A more typical example of an X-linked gene is the *F8* gene in humans. It makes Factor VIII blood clotting proteins in liver cells. If a male is hemizygous for a mutant allele ($F8^-/Y$) the result is hemophilia type A. Females homozygous for mutant alleles ($F8^-/F8^-$) will also have hemophilia. However, heterozygous females, those people who are $F8^+/F8^-$, do not have hemophilia because even though half of their liver cells do not make Factor VIII (because the X with the $F8^+$ allele is inactive) the other 50% can (**Figure 5**). Because some of their liver cells are producing and exporting Factor VIII proteins into the blood stream they have the ability to form blood clots throughout their bodies. The genetic mosaicism in the liver cells of their bodies does not result in a visible mosaic phenotype2
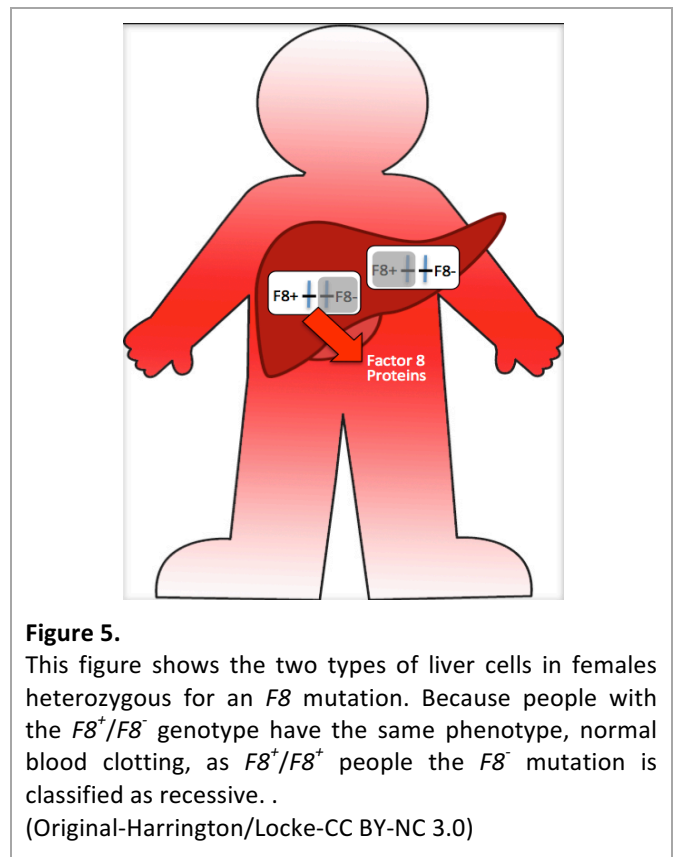
4.   MECHANISMS OF SEX DETERMINATION SYSTEMS

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

For each species, the genetic determination relies on one of several gene or chromosome based mechanisms. See **Figure 5** for a summary. There are, for other species, also a variety of environmental mechanisms, too (rearing temperature, social interactions, **parthenogenesis**). Whatever the sex choice mechanism, however, there are two different means by which the cells of an organism carry out this decision: **hormonal** or **cell-autonomous.**

Different types of chromosomal based sex determination

Figure 6.
Different types of chromosomal (or gene) based sex determination. From top to bottom, there is the archetypal XX/XY system found in humans (and most mammals) with the TDF-Y gene leading to a male phenotype; the ZW/ZZ system found in chickens (birds, moths, and butterflies); the same XX/XY system in Drosophila (sex is determined by the X-chromosome:autosome ratio); the XX/XO system as found in grasshoppers; and the diploid/haploid system as found in bees (and ants, and wasps). Also, the hormonal mechanism is used in humans, while all the other examples use the cell-autonomous mechanism for development of the male or female sex phenotype. (Wikipedia-original - CFCF with additions and corrections by J. Locke- CC BY-SA 3.0)

4.1. HORMONAL MECHANISM:

With this system, used by mammals for example, including humans, the zygote initially develops into a sexually undifferentiated embryo that can become either sex. Then, depending on the sex choice of the genital ridge cells, they grow and differentiate into male (testis) or female (ovary) gonads, which will then produce the appropriate hormones (e.g. testosterone or estrogen). This hormone will circulate throughout the body, enter cells, and activate transcription factors, which will cause all the other tissues to develop and differentiate accordingly, into a male or female phenotype for that individual. Simply put, the circulating hormone induces <u>all</u> the cells and tissues to be the appropriate sexual phenotype.

Sometime this hormone inducing system fails. In cattle (and some other mammals), a **freemartin** is a type of heifer (female) that becomes masculinized because of hormone transfer from a bull (male) twin. Externally a freemartin appears as a female but it is infertile over 90% of the time, has masculinized behavior, and non-functioning ovaries. The animal originates as a female (XX) but the female reproductive development is altered by anti-Müllerian hormone from the male twin, acquired via vascular connections between placentas. Thus, the freemartin has conflicting hormonal cues, which leads to the intermediate phenotype.

4.2. CELL-AUTONOMOUS MECHANISM:

With this system, used by many animals, including birds and insects, the zygote cell initially has a sex phenotype set at the cell level (not whole organism level). A cell intrinsically determines, individually, their sex and then develops accordingly, giving the appropriate sexual characteristics and phenotype to the whole organism. Each cell is autonomous with respect to its sex; there are no sex hormone cues to determine the sex expressed by the organism.

This cell autonomy mechanism can lead to **sexual gynandromorphs**, which are **genetic mosaics** (a single organism composed of genetically distinct cells derived from <u>the same</u> zygote) that display both male and female characteristics in a mosaic fashion. They often are phenotypically split down the midline of the organism. These rare individuals are thought to be the result of an improper sex chromosome segregation that occurs in a cell very

early in development so that one half of the individual has cells with a male chromosome set while the other half has cells with a female set. If a species is **sexually dimorphic** (external morphology easily distinguishes males from females) they are easily visible. See **Figure 8** for a local example. Gynandromorphs are so common that they are even sometimes seen in the wild. A search on the internet will bring up many more examples.

While gynandromorphs are seen in cell-autonomous species, such as insects and birds, they are not seen in hormonally determined species, such as mammals. This is because all the cells in the body display the same sex phenotype caused by the circulating sex hormones. Sexual gynandromorphs appear to be absent in reptiles, amphibians, and fish indicating that they don't use a cell-autonomous mechanism. Nevertheless, there are genetic mosaic individuals in these groups but they do not appear to involve sex determined traits, which is required for a true gynandromorph. They often involve mosaicism of alleles at a single gene locus (somatic mutation) that affect external morphology (e.g. colour).

4.3.   MIXED CELL INDIVIDUALS – MOSAIC VS CHIMERA

Both Mosaic and chimeras have genetically different cells. However, the difference between the two is the origin of those cells. (**Figure 10**)

a) Mosaic

A **mosaic** is an organism or a tissue that contains two or more types of genetically different cells derived from the same (single) zygote. Since the cell is derived from the same organism, most of the loci will be identical in all cell populations except for some loci. Now the genetic change within the zygote might occur due to mutations, changes in the number or structure of chromosomes, or X-chromosome inactivation.

In a clinical sense, if the mosaic individual has large amounts of genetically abnormal cells and few of normal cells, that individual will manifest disease. However, if the individual has small amounts of abnormal cells but sufficient amount of normal cells, the severity of the disease will be reduced.
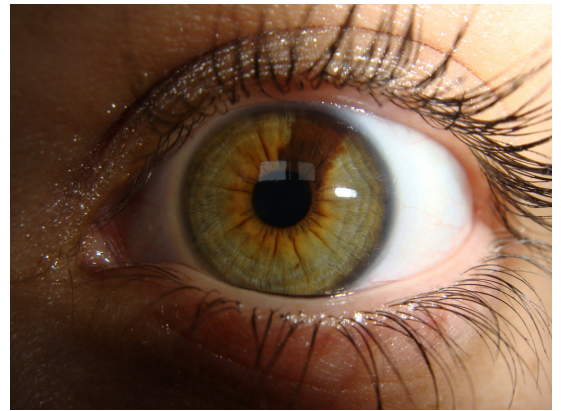
Figure 7.
Mosacism can cause conditions such as partial heterochromia in eyes, which is a condition that causes difference in pigmentation in the iris.
(Wikipedia-Sheila.lorquiana- CC BY-SA 3.0)

Figure 8.
Drosophila sexual gynadromorph. The upper side is female with the most distal abdominal segments being not heavily pigmented, while the lower side is male and the two most distal segments are heavily pigmeted (red circle). Note the curvature of the abdomen caused by the longer, female half above and shorter male half below.
This example was found by a U. of Alberta student in the GENET 375 lab course, Introductions to Molecular Genetic Techniques.
(Original – Locke – CC BY-NC 3.0).

For example, most of the embryos with mosaic turner syndrome experience death prior to birth. This is because they have only one X chromosome. (45, X). However those who survive are known to have more number of normal cells that have 46, XX genotype than the abnormal 45, X genotype cells.

Another example is **X chromosome mosaicism**. During embryogenesis, one of the female's X chromosome is randomly inactivated. Now some cells might have their paternal X chromosome inactivated while others might have their maternal

X chromosome inactivated. Since these two cell lines have different genetic composition, they are also considered as mosaicism. All females have roughly the same amount of two genetically different cell lines.

A **gynandromorph** is an organism that is made up of mosaic tissues of male and female genotypes and displays both male and female characteristics.

b) Chimera

A chimera is an organism composed of genetically distinct cells derived from <u>different (more than one) zygotes</u>. Because the cells are derived from different organisms, the cell populations will have more divergent genotypes when compared to those of a mosaic. The different sources can sometimes even be different species such as a goat and a sheep, which when mixed makes a "shoat" or a "geep".

A chimeric cattle is another example, in an outside the lab setting, to explain this concept. When cows conceive fraternal (non-identical) twins, the circulatory systems of the twins can be connected via a joining called an **anastomosis**. Because of this, blood, cells, and tissue can be exchanged between the fetuses. This is how an organism can contain genetically distinct cells from another organism, its fraternal twin. Now if the blood and cells are "shared", a female fetus will be exposed to male hormones. The result is a masculinized female cow, which is called **free martins** (visibly female, but with male behavior and also sterile).

Also, patients who have undergone cellular transplant such as bone marrow transplant are also considered to be chimeras.

Figure 9.
A chimeric mouse on the very right, made in NIMH's Transgenic Core Facility. The two coloured fur shows the two types of cells present.
(Wikipedia- Staff at NIMH's Transgenic Core Facility-PD)
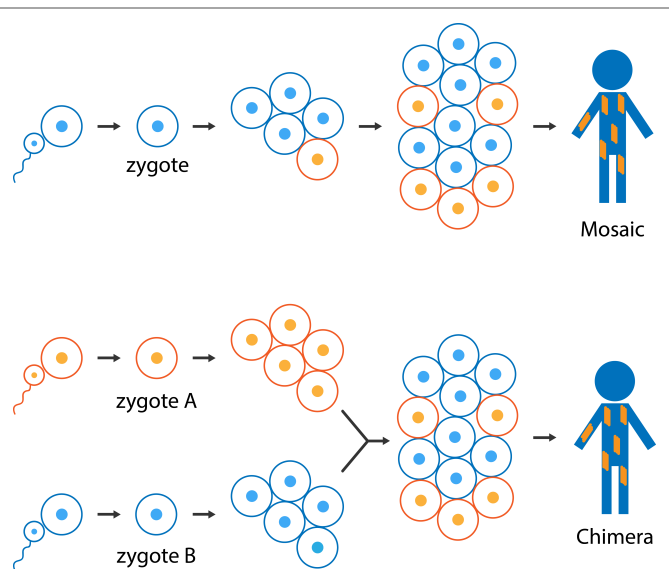
Figure 10.
The main difference between mosaic and chimeric organisms is that the former arises from the same zygote whereas the latter arises from different (more than one) zygotes. (Original-Kang-CC BY-NC 3.0)

_____

Summary:

- In order to compensate for under or over dosage of gene products, organisms use various methods such as expressing genes twice the normal rate or inactiving one X chromosome.

- X-chromosome inactivation occurs randomly (except for special circumstances), and during interphase the inactivated chromosome appears as a condensed mass in the nucleus called the Barr body.

- Orange gene in cats and F8 gene in humans are examples of X-linked genes.

- Sex determination can be either hormonal or cell-autonomous. Abnormality in the cell-autonomous mechanism may result in gynandromorphs.

- Both mosaic and chimeric organisms are composed of genetically distinct cells, but their origins of those cells are different.

Key terms:

dosage compensation

X-linked genes

autosomal genes

X-chromosome inactivation

Barr body

Orange gene

F8 gene

hermaphrodites

parthenogenesis

hormonal

cell-autonomous

freemartin

chimera

sexual gynandromorphs

genetic mosaics

sexually dimorphic

X chromosome mosaicism

gynandromorph

anastomoses

STUDY QUESTIONS:

1) What is the relationship between the O^O and O^B alleles of the *Orange* gene in cats?

2) Another cat hair colour gene is called *White Spotting*. This gene is autosomal. Cats that have the dominant "*S*" allele have white spots, while the "*s*" allele doesn't. Taking the *Orange* locus (O^B and O^O) into account, 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) Make a diagram similar to **Figure 4**, but with the F8 alleles/genotypes, that shows the relationship between genotype and phenotype in females and males and which would use the purified Factor VIII protein.

CHAPTER 23 – PEDIGREE ANALYSIS

Figure 1.
Polydactyly (six fingers in this case – count them) is an example of a human trait that can be studied by pedigree analysis.
(Wikipedia- Drgnu23- CC BY-SA 3.0)

INTRODUCTION

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

1.   PEDIGREE ANALYSIS

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

Pedigree analysis is therefore an important tool in basic research, agriculture, and **genetic counseling**.

Each pedigree chart represents all of the available information about the inheritance of a single trait (most often a disease) within a family. The pedigree chart is therefore drawn using factual information, but there is always some possibility of errors in this information, especially when relying on family members' recollections or even clinical diagnoses. In real pedigrees, further complications can arise due to **incomplete penetrance** (including age of onset) and **variable expressivity** of disease alleles, but for the examples presented in this book, we will presume complete accuracy of the pedigrees – that is, the phenotype accurately reflects the genotype. A pedigree may be drawn when trying to determine the nature of a newly discovered disease, or when an individual with a family history of a disease wants to know the probability of passing the disease on to their children. In either case, a tree is drawn, as shown in **Figure 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 the individual is unaffected, they are called the consultand. If an individual is known to have symptoms of the disease (**affected**), the symbol is filled in. Sometimes a half-filled in symbol is used to indicate a known **carrier** of a disease; this is someone who does not have any symptoms of the disease, but who passed the disease on to subsequent generations because they are a heterozygote. Female carriers of X-linked traits are indicated by a circle with a dot in the centre. 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.

1.2. PEDIGREE CHART CONVENTION SYMBOLS
In pedigree analysis, standardized human pedigree nomenclature is used.

If possible, male partner should be left of female partner on relationship line. Siblings should be listed from left to right in birth order, oldest to youngest.

2.   MODES OF INHERITANCE

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

2.1. AUTOSOMAL DOMINANT (AD)

When a disease is caused by a dominant allele of a gene, every person with that allele will show
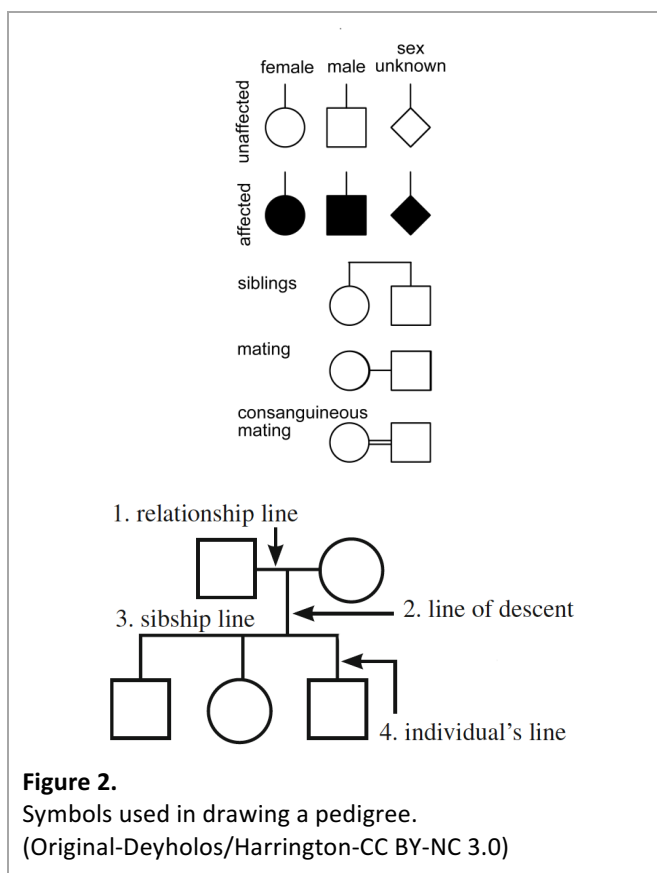
Figure 2.
Symbols used in drawing a pedigree.
(Original-Deyholos/Harrington-CC BY-NC 3.0)

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.

Figure 3.
A pedigree consistent with AD inheritance. (Unknown)
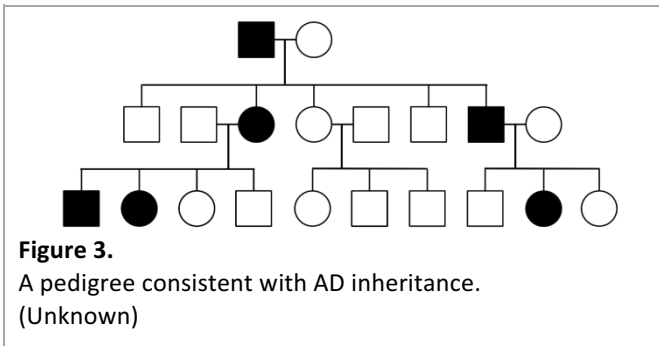
Nomenclature	
A	mutant allele
a	WT allele
A / a	affected
A / A	affected to the same or greater extent

Table 1.
Genotype nomenclature consistent with AD inheritance. (Original-Harrington-CC BY-NC 3.0)
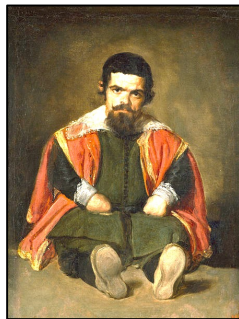
Figure 4.
Portrait of Sebastián de Morra by Diego Velázquez, a court dwarf and was painted ~1645. He likely had achondroplasia, a condition that has autosomal dominant inheritance. (Wikimedia Commons-Diego Velázquez-PD)
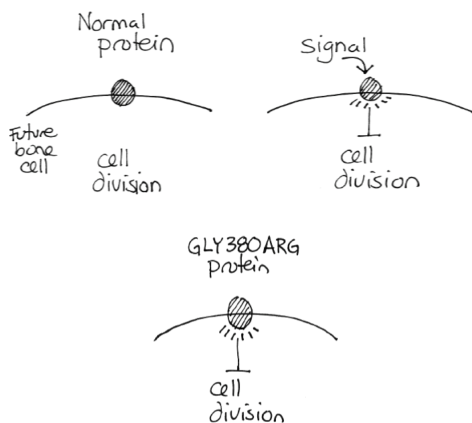
Figure 5.
Diagram showing the mechanism of achondroplasia. (Original-Harrington-CC BY-NC 3.0)

AD EXAMPLE: ACHONDROPLASIA

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

2.2. X-LINKED DOMINANT (XD)

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

Figure 6.
Two pedigrees consistent with XD inheritance. (Unknown)
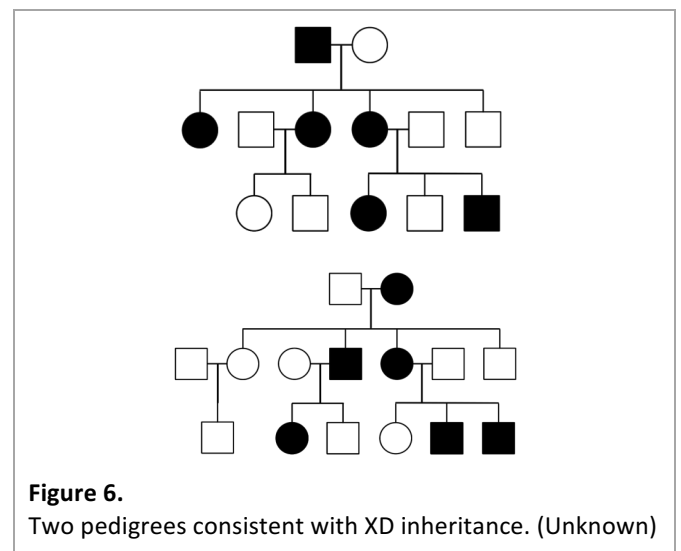
Nomenclature	
A	mutant allele
a	WT allele
X^A / Y	affected male
X^A / X^A	affected females
X^A / X^a	

Table 2.
Genotype nomenclature consistent with XD inheritance. (Original-Harrington-CC BY-NC 3.0)
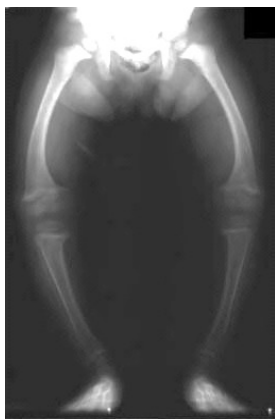
Figure 7.
Some types of rickets may follow an XD mode of inheritance.
(Wikipedia-Mrish-CC BY-SA 1.0)

XD EXAMPLE: FRAGILE X SYNDROME

The *FMR1* gene at Xq21 (X chromosome, q arm, region 2, band 1) encodes a protein needed for neuron development. There is a (CGG)$_n$ repeat array in the 5'UTR (untranslated region). If there is expansion of the repeat in the germline cell the child will inherit a non-functional allele. X^A / Y males have fragile X mental retardation (IQ < 50) because none of their neurons can make FMR1 proteins. Fragile X syndrome is considered X-linked dominant because only some neurons in X^A / X^a females can make FMR1 proteins. The severity (IQ 50 – 70) in these females depends upon the number and location of these cells within in the brain.

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
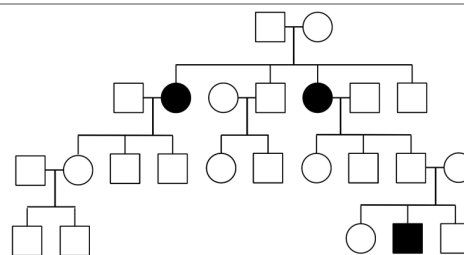
Figure 8.
A pedigree consistent with AR inheritance. (Unknown)

Nomenclature	
A	WT allele
a	mutant allele
a / a	affected
A / a	carrier
A / _	status unknown

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

pedigree can be carriers, probably without knowing it. Compared to pedigrees of dominant traits, AR pedigrees tend to show fewer affected individuals and are more likely than AD or XD to "skip a generation". Thus, the major feature that distinguishes AR from AD or XD is that unaffected individuals can have affected offspring. Attached earlobes is a human condition that may follow an AR mode of inheritance.

AR EXAMPLE: PHENYLKETONURIA (PKU)

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

Figure 9.
Many inborn errors of metabolism, such as phenylketonuria (PKU) are inherited as AR. Newborns are often tested for a few of the most common metabolic diseases.
(Wikipedia-U.S. Air Force photo/Staff Sgt. Eric T. Sheler-PD)
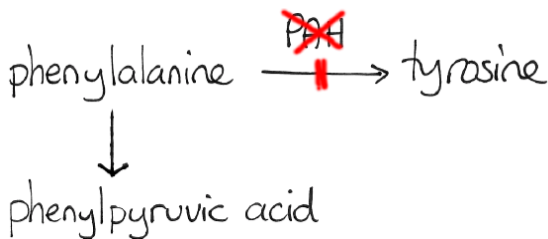
Figure 10.
Mutation in the *PAH* gene cannot catalyze the breakdown of phenylalanine into tyrosine. This causes a buildup of phenylpyruvic acid, which would damage the central nervous system.
(Original-Harrington-CC BY-NC 3.0)

Figure 11.
A pedigree consistent with XR inheritance. (Unknown)

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

Table 4.
Genotype nomenclature consistent with XR inheritance.
(Original-Harrington-CC BY-NC 3.0)
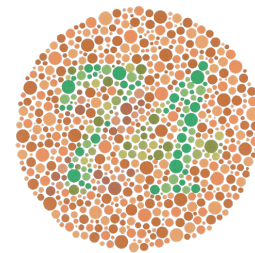
Figure 12.
Some forms of colour blindness are inherited as XR-traits. Colour blindness is diagnosed using tests such as this Ishihara Test, which is shown above. Note: if printed in B&W, this image will not show the number. See the online image in colour. (Wikipedia-unknown-PD)

heterozygotes. Compound heterozygotes have two different mutant alleles (different base pair changes) at a given locus, in this case the *PAH* gene.

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 <u>not</u> 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.

XR EXAMPLE: HEMOPHILIA A

F8 gene at Xq28 (X chromosome, q arm, region 2, band 8) encodes blood clotting factor VIIIc. Without Factor VIIIc, internal and external bleeding can't be stopped. Back in the 1900s, X^a / Y male's
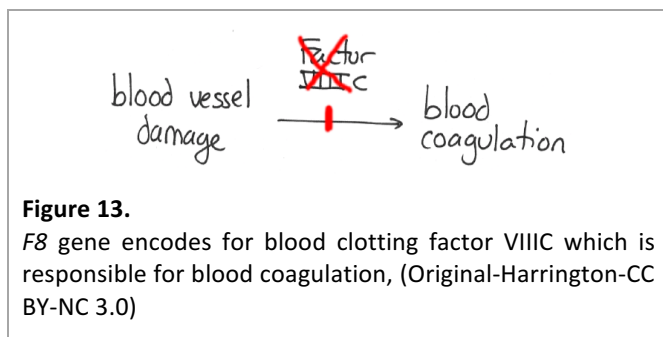
Figure 13.
F8 gene encodes for blood clotting factor VIIIC which is responsible for blood coagulation, (Original-Harrington-CC BY-NC 3.0)

average life expectancy was 1.4 years, but in the 2000s it has increased to 65 years with the advent of Recombinant Human Factor VIIIc. Hemophilia A is recessive because X^A / X^a females have normal blood coagulation, while X^a / X^a females have hemophilia.

2.5. Y-LINKED

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

A common, but incorrect, example of Y-linked inheritance is the **hairy-ear-rim** phenotype seen in some Indian families. A better example are the **Y-chromosome DNA polymorphisms** that have been 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).

2.6. ORGANELLE GENOMES

In eukaryotes, DNA and genes also exist outside of the chromosomes found in the nucleus. Both the **chloroplast** and **mitochondrion** have circular chromosomes. These organelle genomes are often present in multiple copies within each organelle. In most sexually reproducing species, **organelle chromosomes** are inherited from only one parent, usually the one that produces the largest gamete. Thus, in mammals, angiosperms, and many other organisms, mitochondria and chloroplasts are

inherited only through the mother (maternally). Therefore, mutations in mitochondrial DNA (mtDNA) are inherited through the maternal line. There are some human diseases associated with mutations in mitochondria genes. These mutations can affect both males and females, but males cannot pass them on as all mitochondria are inherited via the egg, not the sperm. Mitochondrial DNA polymorphisms are also used to investigate evolutionary and historical lineages, both ancient and recent. Because of the relative similarity of sequence mtDNA is also used in species identification in ecology studies. An example of Mitochondrial inheritance is the Leber hereditary optical neuropathy (LHON). Mitochondria are very important in retinal cells for ATP and/or a specialized function. Mutations in several mtDNA genes result in blindness during early childhood.

3. SPORADIC AND NON-HERITABLE DISEASES

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

On the other hand, some non-genetic diseases may appear to be heritable because they affect multiple members of the same family, but this is due to the family members being exposed to the same toxins or other environmental factors (e.g. in their homes).

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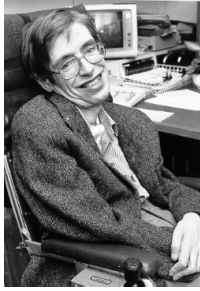
Figure 14.
Stephen Hawking
(Wikipedia-NASA-PD)

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 15**.
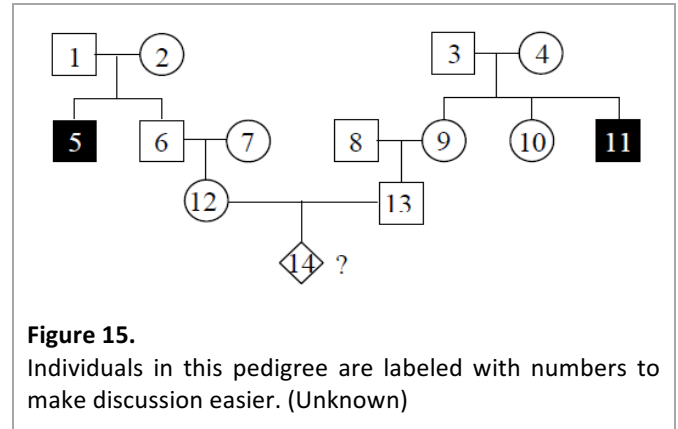
Figure 15.
Individuals in this pedigree are labeled with numbers to make discussion easier. (Unknown)

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.

_____

SUMMARY:

- Pedigree analysis can be used to determine the mode of inheritance of specific traits such as diseases.

- Loci can be X- or Y-linked or autosomal in location and alleles either dominant or recessive with respect to wild type.

- If the mode of inheritance is known, a pedigree can be used to calculate the probability of inheritance of a particular genotype by an individual.

KEY TERMS:

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

STUDY QUESTIONS:

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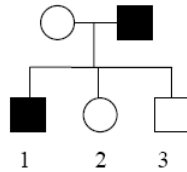
2) In this pedigree in question 1, the mode of inheritance cannot be determined unambiguously. What are some examples of data (e.g. from other generations) that, if added to the pedigree would help determine the mode of inheritance?

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

a)
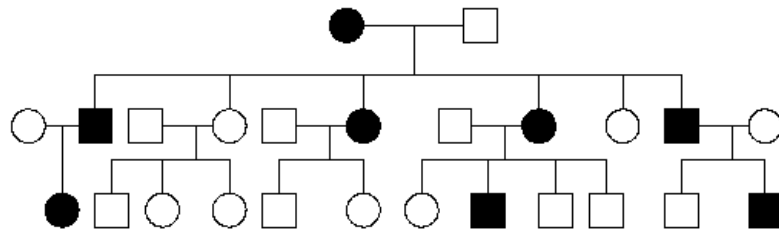
b)
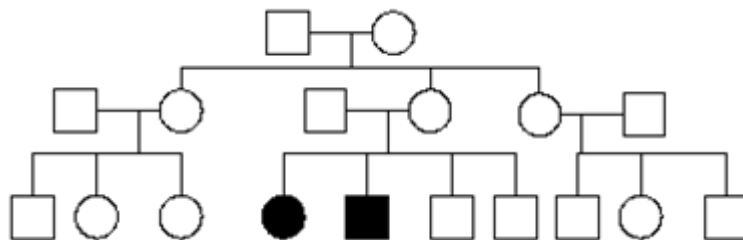
c)
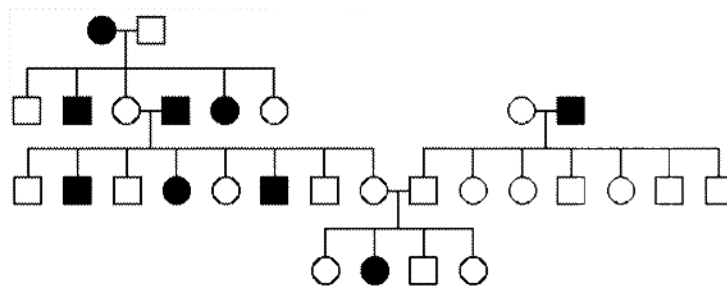
d)
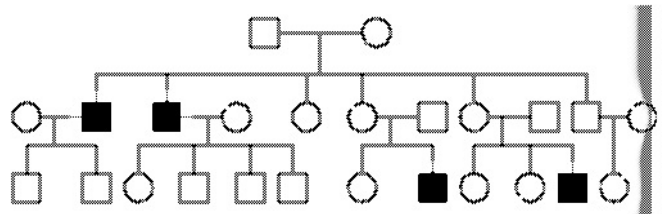
4) The following pedigree represents a rare, autosomal recessive disease. What are the genotypes of the individuals who are indicated by letters?
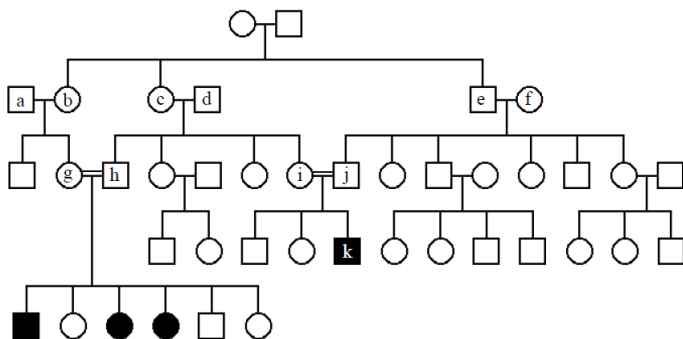
5) If individual #1 in the following pedigree is a heterozygote for a rare, AR disease, what is the probability that individual #7 will be affected by the disease? Assume that #2 and the spouses of #3 and #4 are not carriers.
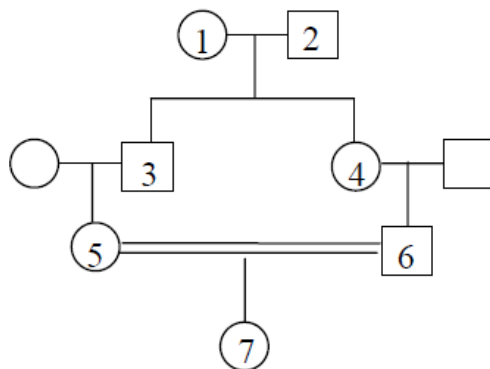
CHAPTER 24 – CHROMOSOME REARRANGEMENTS

Figure 1.
Comparing an ideogram of the human chromosome 2 to the equivalent chromosomes in chimpanzees, we notice that the human chromosome 2 likely came from a fusion event that occurred since their common ancestor. This is supported by evidence finding telomeric and centromeric sequences in the middle of human chromosome 2 similar to that of the ends and middle of the chimpanzee chromosomes.
Note: The actual formation of HSA2 was more complex (and interesting) than this. As drawn it can't have occurred this way by either
(i) loss of telomeres followed by chromosome fusion or by
(ii) a Robertsonian translocation because in either case the telomere DNA would be lost.
For details see:
http://genome.cshlp.org/content/22/6/1036.full
(Flickr- T. Michael Keesey- CC BY 2.0)

INTRODUCTION

Previous chapters described chromosomes as simple linear DNA molecules on which genes are located. For example, your largest chromosome, chromosome 1, has about 3536 genes. To ensure that each of your cells possesses these genes, the typical linear eukaryotic chromosome has three critical features that allow it to be passed on during cell division. (1) **Origins of replication** found along its length provide places for DNA replication to start, (2) **telomeres** protect each end of the chromosome, and (3) a single **centromere** near the middle provides a place for microtubules to attach and move the chromosome during mitosis and meiosis.

However, at various locations both strands of the double stranded DNA in a chromosome can break and the subsequent daughter cell(s) may not retain all the DNA and thus all the genes. For example, if a segment of the chromosome has been lost (a deletion), the cell may be missing many genes. The <u>causes</u> of chromosome structural abnormalities and the <u>consequences</u> they have for the cell and the organism are described below. They involve double stranded breaks in the DNA, meiotic crossover events, and rejoining of the broken ends. Human examples will be used to show the phenotypic consequences and methods for detection.

1. DNA DOUBLE STRAND BREAKS AND INCORRECT MEIOTIC CROSSOVERS CAUSE CHROMOSOMAL REARRANGEMENTS

1.1. DOUBLE STRAND BREAKS AND THEIR REPAIR
A chromosome is a very long but very thin molecule. In the phopho-diester backbone there are only two covalent bonds holding each base pair to the next. If one of these covalent bonds is

Figure 2.
Repair of single strand nicks and double strand breaks in DNA. (Original-Harrington-CC BY-NC 3.0)

Figure 3.
Errors during DNA repair can cause a chromosome deletion. In this diagram A, B, and C are genes on the same chromosome. As in **Figure 2** there has been breaks in the DNA, recruitment of NHEJ proteins, and repair. After the repairs are completed the small piece of DNA with gene B is lost and the chromosome now only has genes A and C. (Original-Harrington-CC BY-NC 3.0)

broken the chromosome will still remain intact, although a DNA Ligase will be needed to repair the nick (**Figure 2**a). Problems arise when both strands are broken at or near the same location. This **double strand break** will cleave the chromosome into two independent pieces (**Figure 2**b). Because these events do occur in cells there is a repair system called the **non-homologous end joining** (NHEJ) **system** to fix them. Proteins bind to each broken end of the DNA and reattach them with new covalent bonds. This system is not perfect and sometimes leads to **chromosome rearrangements** (see next section).

The NHEJ system proteins only function if required. If the chromosomes within an interphase nucleus are all intact the system is not active. The telomeres at the natural ends of chromosomes prevent the NHEJ system from attempting to join the normal ends of chromosomes together. If there is one double strand break the two broken ends can be recognized and joined. But if there are two double strand breaks at the same time there will be four broken ends in total. The NHEJ system proteins may join the ends together correctly, but if they fail, the result is a **chromosome rearrangement** (**Figure 3**).
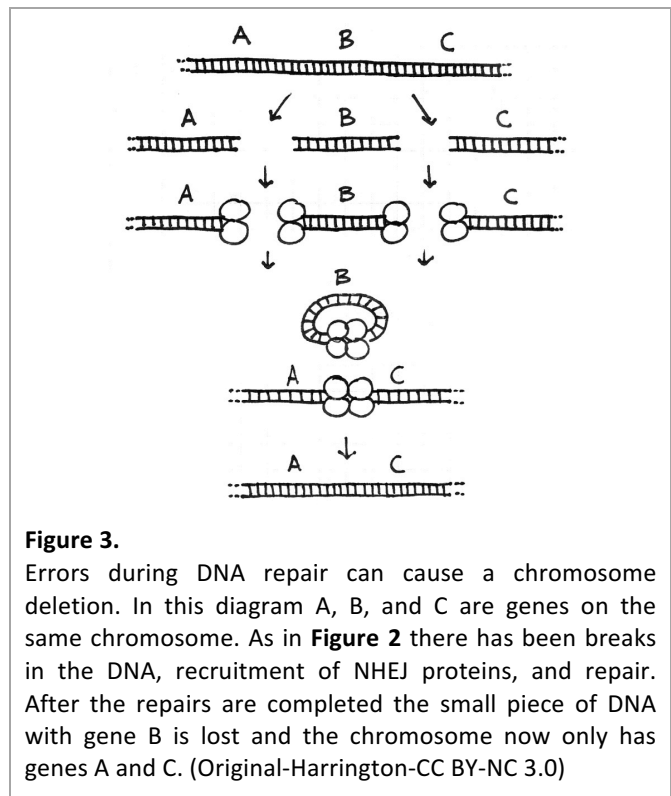
1.2. INCORRECT MEIOTIC CROSSOVERS
Meiotic crossovers occur at the beginning of meiosis for two reasons. They help hold the homologous chromosomes together until separation occurs during anaphase I (see Chapter 16). They also allow recombination to occur between linked genes (see Chapter 17). The event itself takes place during prophase I when a double strand break on one piece of DNA is joined with a double strand break on another piece of DNA and the ends are put together (**Figure 4**a). Most of the time the breaks are on non-sister chromatids and most of the time the breaks are at the same relative locations.

Problems occur when the wrong pieces of DNA are matched up along the chromosomes during crossover events. This can happen if the same or similar DNA sequence is found at multiple sites on the chromosomes (**Figure 4**b). For example, if there are two **_Alu_ transposable elements** on a chromosome. When the homologous chromosomes pair during prophase I, the wrong _Alu_ sequences might line up. A crossover may occur in this region. If so, when the chromosomes
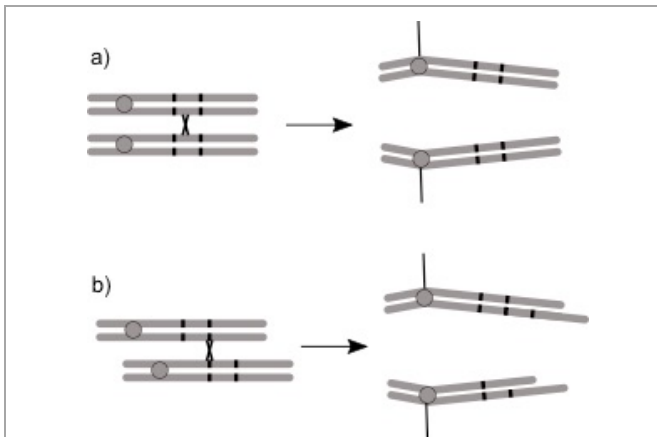
Figure 4 .
Errors during meiotic crossovers can cause duplications and deletions. This diagram shows homologous chromosomes pairing in prophase I and then separating in anaphase I. The shaded boxes are Alu transposable elements. a) The homologous chromosomes pair properly, a crossover occurs, and all four chromatids in anaphase I are normal. b) The pairing is incorrect, a crossover occurs in the mispaired region, and in anaphase I one chromatid has a duplication and another has a deletion.
(Original-Harrington/L. Canham-CC BY-NC 3.0)

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.

Errors during the repair of multiple double strand breaks or incorrect meiotic crossovers can cause four types of chromosome rearrangements: deletion, inversion, duplication or translocation. The type of chromosome rearrangement is either dependent upon where the two breaks were originally and how they are rejoined, or on the location of the homology during meiosis. **Figure 3** shows some possibilities but more are shown in the following sections. The first part of each section shows a double strand DNA break between the B and C genes (shown here as a red X). A second DNA break occurs and the NHEJ proteins mend the damage incorrectly by joining the ends (shown with the blue arrows). The chromosomes are drawn unreplicated as they are in G_1 phase but these events can happen anytime during interphase. The

second part shows how meiosis can cause the rearrangements.

2. DELETIONS

There are two forms of deletions: **Terminal** and **Interstitial**. Terminal deletions are deletions off of the end of a chromosome. Interstitial deletions are deletions of a region in the middle of the chromosome, while the arms on each side remain normal. For example, with a chromosome that has the genes ABCDEF, an example of a terminal deletion will be CDEF. An example of an interstitial deletion will be ABCF.

2.1. DELETIONS FROM DOUBLE STRAND BREAK REPAIR
Deletions arise from double strand breaks when both breaks are on one chromosome. If the ends are joined in this way the piece of DNA with the B gene on it does not have a centromere and will be lost during the next cell division.
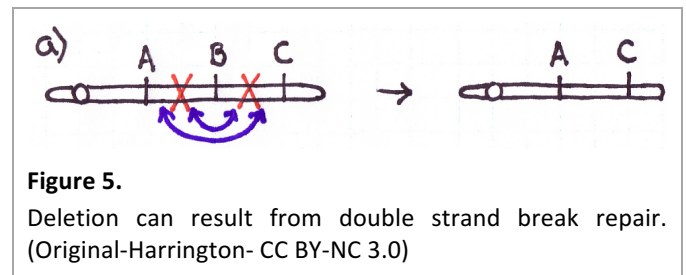
Figure 5.
Deletion can result from double strand break repair.
(Original-Harrington- CC BY-NC 3.0)

2.2. DELETIONS FROM INCORRECT MEIOSIS
If meiotic rearrangement is the cause, **Deletion** chromosomes will pair up with a normal homolog along the shared regions and at the missing segment, the normal homolog will loop out (nothing to pair with) to form a **deletion loop**. This can be used to locate the deletion cytologically. The deleted region is also **pseudo-dominant**, in that it permits the mutant expression of recessive alleles on the normal homolog. Deletion mutations don't revert - nothing to replace the missing DNA.

3. INVERSIONS

3.1. INVERSIONS FROM DOUBLE STRAND BREAKS
Inversions also occur when both double strand breaks are on one chromosome. If the ends are joined in this way, part of the chromosome is inverted. This example shows a **para<u>c</u>entric inversion**, named because the inverted section

does not include the centromere (para = beside). If the breaks occur on different chromosome arms the inverted section includes the centromere and the result is a **pericentric inversion** (peri = around).

Figure 6.
Inversion can result from double strand break repair. (Original-Harrington- CC BY-NC 3.0)
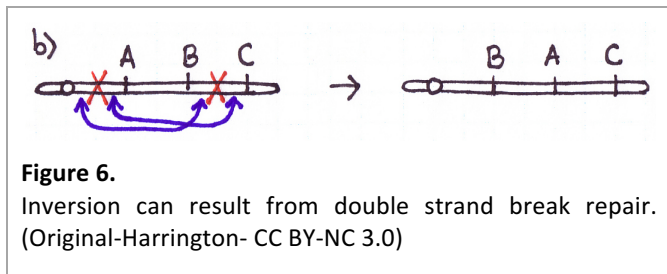
3.2. INVERSIONS FROM INCORRECT MEIOSIS

In meiosis, when an **inversion** chromosome is paired up there is an **inversion loop** formed. If there is a crossover within the loop then abnormal products will result and abnormal, unbalanced gametes will be produced. For example, a crossover event within the loop of a **paracentric inversion** will lead to a di-centric product that will break into deletion products and produce unbalanced gametes (**Figure 7**). Similarly, with a **pericentric inversion**, a crossover event leads to duplicate/deletion products that are unbalanced (**Figure 8**).

If joined with a normal gamete, they will result in an unbalanced zygote, which are usually lethal. The consequence for this is that crossover products (recombinants) are lost and thus inversions appear to suppress crossovers within the inverted region.

Note: with both types of inversions, crossovers outside the loop are possible and fully viable, as they don't alter the gene balance.

4. DUPLICATIONS

There are two major forms of duplications: **tandem** and **inverse** duplications. Tandem duplications are when the duplicated genes are in the same order, and inverse duplications are where the duplicated genes are in the reverse order. For example if you have a chromosome that has the genes ABCDEFGH, and a duplication occurs in the BCD genes, then a tandem duplication would look like: ABCDBCDEFGH. An inverse duplication would look like: ABCDDCBEFGH.

Insertional duplications are also seen, where the duplicated region is inserted to a more distant location. e.g. ABCDEFBCDGH

Figure 7.
A paracentric inversion pairing at meiosis. A crossover within the loop causes the production of an acentric and a dicentric chromatids, which leads to deletion product..
(Original-Locke-CC BY-NC 3.0)
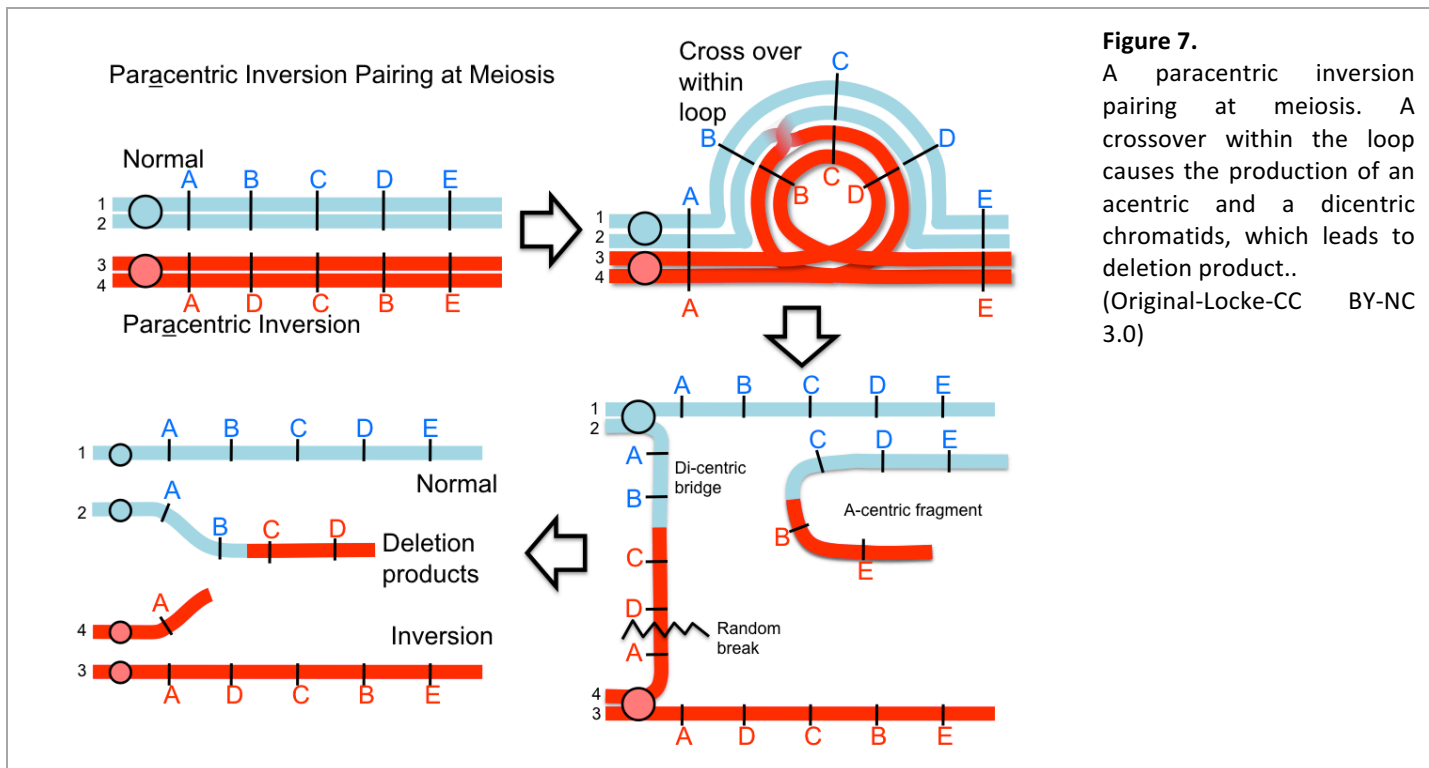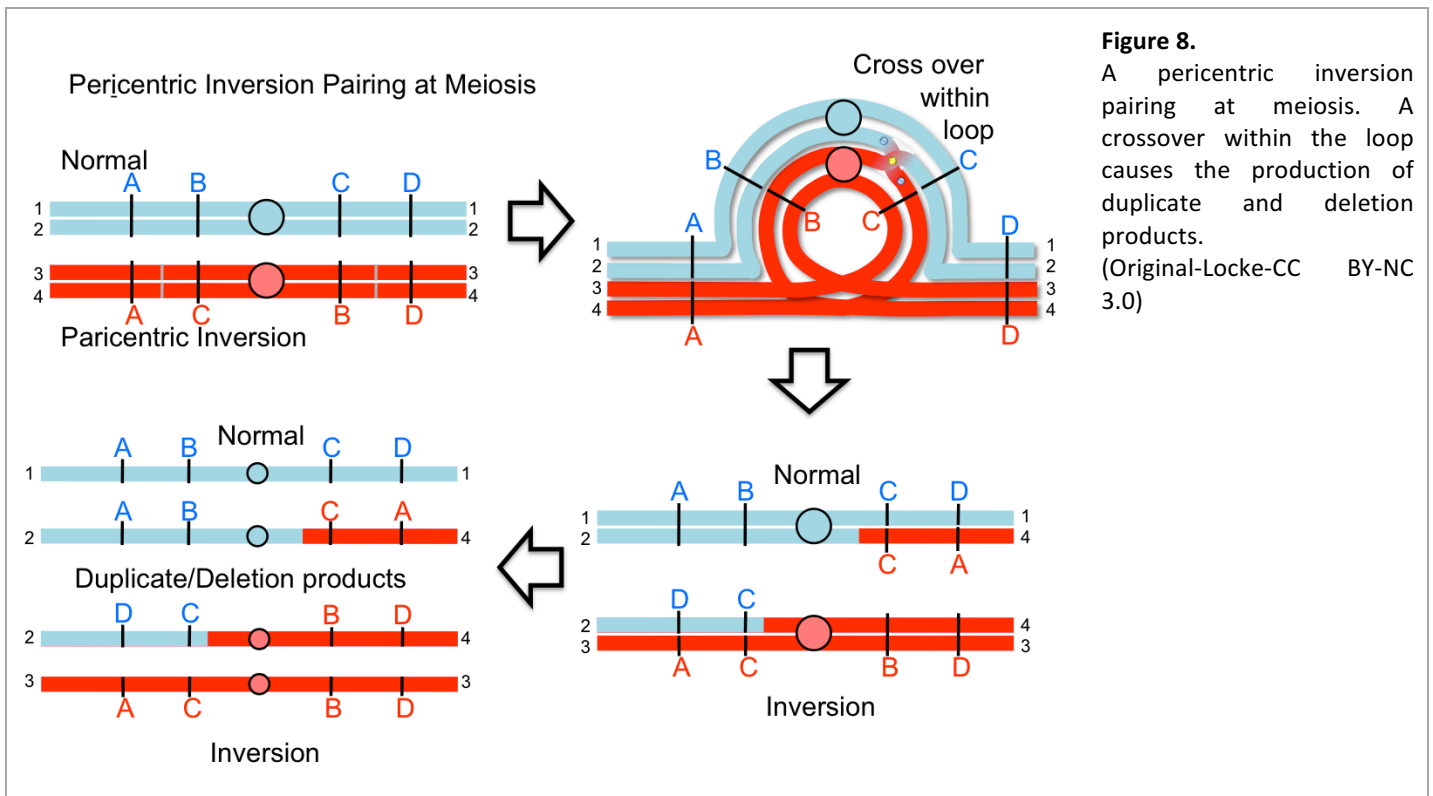
Figure 8.
A pericentric inversion pairing at meiosis. A crossover within the loop causes the production of duplicate and deletion products.
(Original-Locke-CC BY-NC 3.0)

4.1. DUPLICATIONS FROM DOUBLE STRAND BREAKS

Duplications can occur from two DNA breaks at different places in sister chromatids (in a replicated chromosome). The ends are joined together incorrectly to create a chromosome with a duplication (two "B" regions as shown). Note: the reciprocal product has a deletion.
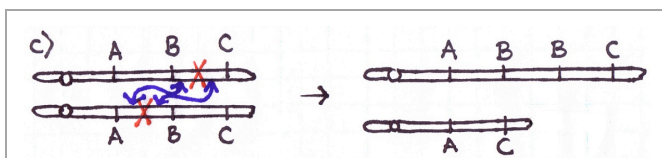
Figure 9.
Duplication can result from double strand break repair.
(Original-Harrington- CC BY-NC 3.0)

4.2. DUPLICATIONS FROM INCORRECT MEIOSIS

Duplications also produce a cytologically visible loop at meiotic pairing. Duplications can revert at a relatively high frequency by unequal crossing over. Duplicated genes offer new possibilities for mutational divergence followed by natural selection in the course of evolution.

5. TRANSLOCATIONS

5.1. TRANSLOCATIONS FROM DOUBLE STRAND BREAKS

Translocations result from two breaks on different chromosomes (not homologs) and incorrect rejoining. This example shows a **reciprocal translocation** - two chromosomes have 'swapped' arms, the E gene is now part of the white chromosome and the C gene is now part of the shaded chromosome. **Robertsonian translocations** are those rare situations in which all the genes end up together on one chromosome and the other chromosome is so small that it is typically lost.
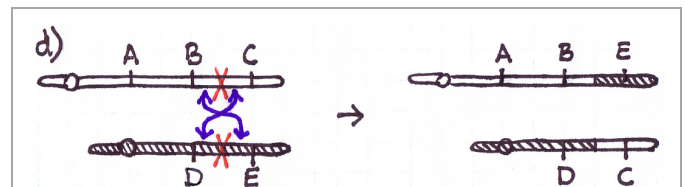
Figure 10.
Translocation can result from double strand break repair.
(Original-Harrington- CC BY-NC 3.0)

Figure 11.
A reciprocal translocation pairing at meiosis. There are two main avenues for segregation: Adjacent-1 and Alternate. Adjacent-1 results in duplication and deletion for part of the chromosome segments. Alternate doesn't.
(Original-Locke-CC BY-NC 3.0)
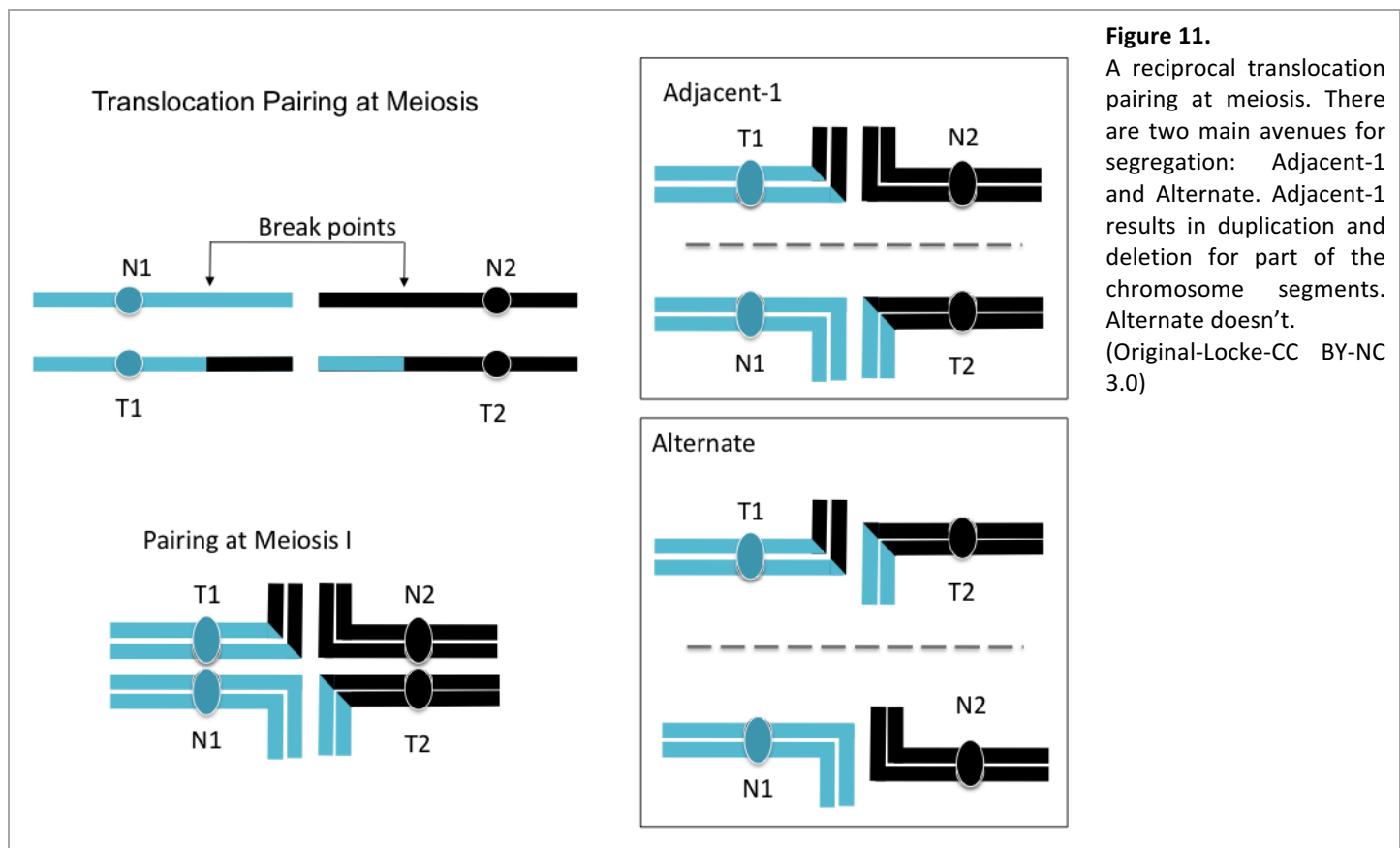
5.2. TRANSLOCATIONS FROM INCORRECT MEIOSIS

For **translocations** during meiosis, a consequence for the two chromosomes involved is that when they pair both replicated chromosome pairs will be together, which can be seen cytologically as a **tetrad**. This tetrad can segregate in three ways.

This set of paired, replicated chromosomes can segregate as **Alternate** (balanced) where both normal (N1 and N2) and both translocated chromosomes (T1 and T2) go to the same polls, respectively. The chromosomes can segregate as **Adjacent-1** (unbalanced) where the normal and translocation chromosomes segregate, with N2 and T1 segregate from N1 and T2. Alternate and Adjacent 1 both occur in approximate equal frequency and thus only about half the time do the gametes end up unbalanced (**Figure 11.**). Note how each daughter cell in Alternate has equal amounts of blue and black chromosomes, while in Adjacent-1 one daughter has extra black chromosomes, and the other has extra blue.

The third segregation possibility is known as Adjacent-2, where N1 and T1 go to one pole, while N2 and T2 go to the other. This way of segregating is extremely rare, and so will not be described in any further detail.

6. CONSEQUENCES OF CHROMOSOMAL REARRANGEMENTS

6.1. DECREASED VIABILITY

All the chromosome rearrangements shown above produce functional chromosomes. Each has one centromere, two telomeres, and thousands of origins of replication. Because inversions and translocations do not change the number of genes in a cell or organism they are said to be **balanced rearrangements**. Unless one of the breakpoints occurred in the middle of a gene the cells will not be affected. On the other hand, deletions and duplications are **unbalanced rearrangements**. The larger they are (more genes involved) the more disruption they cause to the proper functioning of the cell or organism. Having too much or too little gene action for a large number of genes can disrupt

the cellular metabolism to generate a phenotype or reduce viability.

6.2. DECREASED FERTILITY

Recall that during meiosis I homologous chromosomes pair up. If a cell has a chromosome with a rearrangement this chromosome will have to pair with its normal homolog.

Cells heterozygous for balanced rearrangements actually have more difficulties in prophase I. Consider the chromosomes shown in **Figure 12**. There are different ways they might pair during prophase I - one is shown in **Figure 13**. But if a crossover occurs in the inverted region the result will be unbalanced gametes. Embryos made with unbalanced gametes rarely survive. The consequence is that the heterozygous organism will have **reduced fertility**.

Note that an organism homozygous for this inversion chromosome will not be affected in this way because no loops are formed. The chromosomes can pair along their entire length and crossovers will not produce any unbalanced

gametes. This is a general property of inversions and translocations.

In heterozygotes there are problems during meiosis resulting in a lot of the gametes being unbalanced and an overall reduction in fertility. In homozygotes the rearranged chromosomes pair with one another just fine and there is no effect on fertility.

6.3. CANCER

Some chromosome rearrangements have breakpoints within genes leading to the creation of hybrid genes – the first part of one gene with the last part of another. If the hybrid gene inappropriately promotes cell replication, the cell can become cancerous.

6.4. EVOLUTION

Those chromosome changes that duplicate genes are important for evolution. If an organism has an extra copy of important genes, one gene can be retained for their original function while others can mutate and potentially acquire new functions (**Figure 14.**). An example of this is the multiple copies of the globin genes found in mammals.
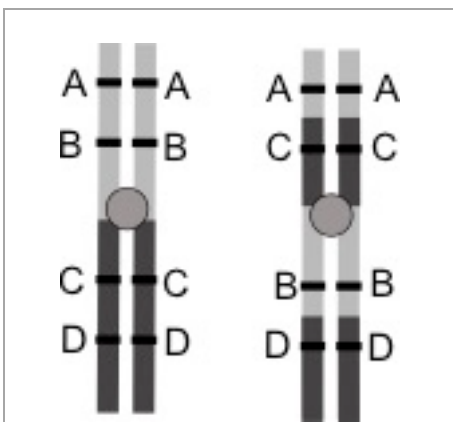
Figure 12.
A normally arranged chromosome (left) and a homolog with a pericentric inversion (right).
(Original-Harrington/Canham-CC BY-NC 3.0
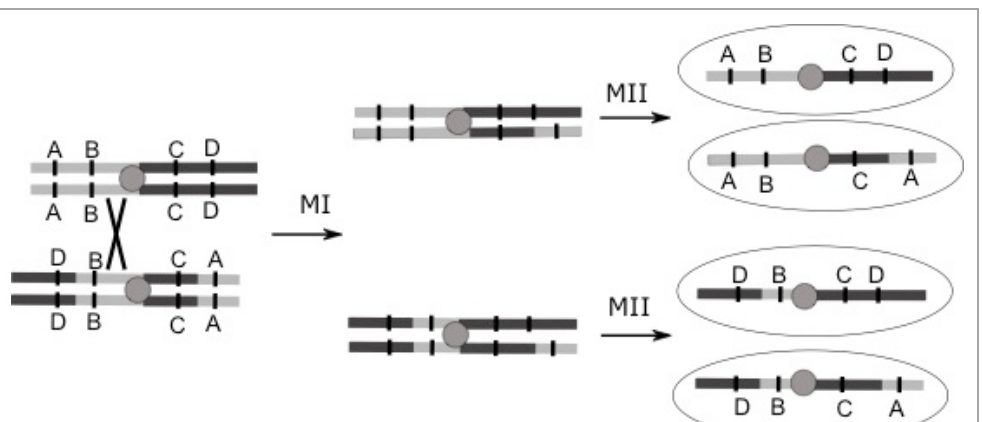
Figure 13.
Meiosis in a cell heterozygous for the chromosomes shown in **Figure 12**. Note that of the four gametes one has a deletion of the A gene and a duplication of the D gene while another gamete has a duplication of A and a deletion of D.
(Original-Harrington/Canham-CC BY-NC 3.0)

Figure 14.
Duplicated genes can mutate without compromising the viability of the organism. Occasionally the result is a new gene. (Original-Harrington-CC BY-NC 3.0)
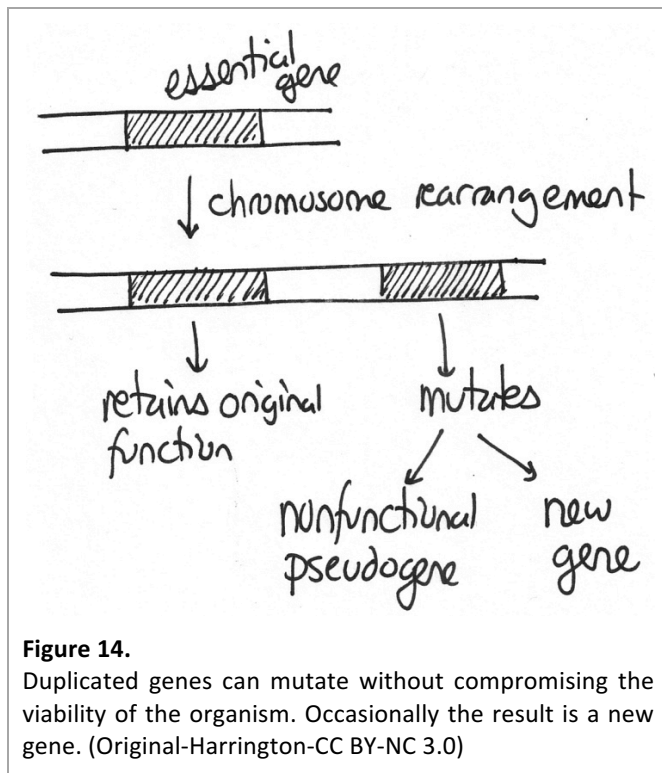
Chromosome rearrangements that decrease fertility are also important for the origin of new species. If a rearrangement, such as the inversion shown in **Figure 12**, becomes common in a small isolated population, that population has 100% fertility if they mate within their group, but a reduced fertility if they mate with members of the larger population. As rearrangements accumulate the small population will become more and more reproductively isolated. When members are incapable of forming viable, fertile offspring with the original population the group will have become a new species.

Another example is shown in **Figure 1**, where the human chromosome 2 is a fusion of two chromosomes present in the common ancestor of humans and other great apes (chimpanzee, gorilla, orangutan). We do not know exactly when in human history (evolution) this fusion event occurred, except that, because it is absent in all other apes and present in all current humans, it must have occurred after the split between chimpanzee and humans.

7. CHROMOSOMAL REARRANGEMENTS IN HUMANS

The problems described above can affect all eukaryotes, unicellular and multicellular. To better understand the consequences let us consider those that affect people. The convention when describing a person's **karyotype** (chromosome composition) is to list the total number of chromosomes, then the sex chromosomes, and then anything out of the ordinary. Most of us are 46,XX or 46,XY. What follows are some examples of chromosome number and chromosome structure abnormalities.

7.1. CRI-DU-CHAT SYNDROME
Cri-du-chat syndrome occurs when a child inherits a defective chromosome 5 from one parent (**Figure 15**). This condition is rare - it is present in only 1 in 20 000 to 1 in 50 000 births but it does account for 1% of cases of profound intellectual disability. The specific defect is a deletion that removes 2 Mb or more from the tip of the short arm of the chromosome. In most cases the deletion is the result of a chromosomal rearrangement in one of the parent's germ line cells. People with cri-du-chat have a karyotype of 46,sex,deletion(5).

Figure 15.
A boy with cri-du-chat syndrome. The pictures were taken at 8 months (A), 2 years (B), 4 years (C), and 9 years (D). (Wikipedia-Paola Cerruti Mainardi/ changes: horizontally aligned the photos- CC BY 2.0)
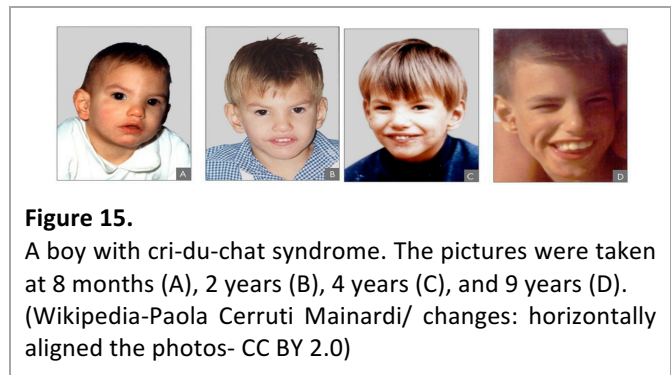
As with Down syndrome this condition is associated with intellectual disability and other health problems. These problems include an improperly formed larynx which leads to infants making high pitched cat-like crying sounds (hence the name "cry of the cat"). It is suspected that at least some of the intellectual disability phenotype is due to having only a single copy of the CTNND2 gene. This gene is active during embryogenesis and makes a protein essential for neuron migration. Down syndrome and cri-du-chat syndrome are two

examples of the need for genomes to contain the proper number of genes. Having too many copies of key genes (Down syndrome) or too few (cri-du-chat syndrome) can lead to substantial developmental problems.

7.2. INVERSION(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 his or her parents and so on. In contrast, most cases of deletion(5) are due to new mutations occurring in a parent.

7.3. DIAGNOSING HUMAN CHROMOSOME ABNORMALITIES

How can we confirm that a person has a specific chromosomal abnormality? The first method was simply to obtain a sample of their cells, stain the chromosomes with **Giemsa** dye, and examine the results with a light microscope (**Figure 16**). Each chromosome can be recognized by its length, the location of its centromere, and the characteristic pattern of purple bands produced by the Giemsa. Bright field microscopy has its limitations though - it only works with mitotic chromosomes and many chromosome rearrangements are either too subtle or too complex for even a skilled cytogeneticist to discern.

The solution to these problems was **fluorescence in situ hybridization** (FISH). A single stranded **fluorescent DNA probe** is allowed to hybridize to denatured target DNA. Because there are several fluorescent colours available it is common to use more than one probe at the same time. A more detailed explanation of the FISH technique can be found in the chapter 32.
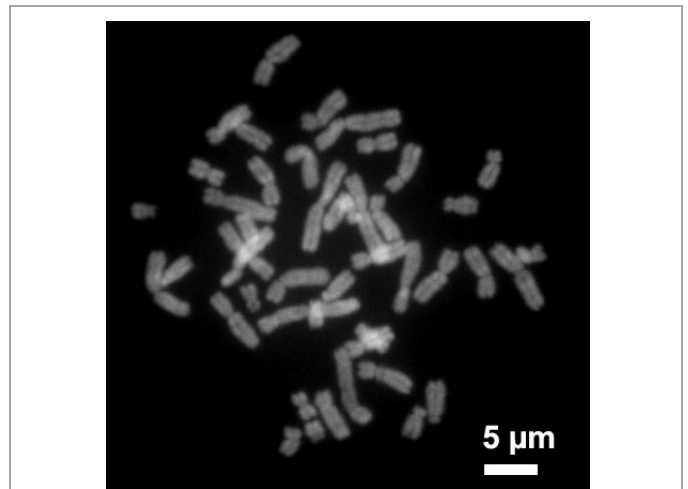
Figure 16.
Human chromosomes. One way to obtain chromosomes is to take a blood sample, culture the cells for three days in the presence of a T-cell growth factor, arrest the cells in metaphase with a microtubule inhibitor, and then drop the cells onto a slide. The cells burst and the chromosomes stick to the slide. The chromosomes can then be stained or probed. Because the cells are in metaphase it is possible to see 46 replicated chromosomes here. There will be dozens of collections of chromosomes like this over the entire slide.
(Wikipedia-Steffen Dietzel- CC BY-SA 3.0)

A physician may suspect that a patient has a specific genetic condition based upon the patient's physical appearance, mental abilities, health problems, and other factors. FISH can be used to confirm the diagnosis. For example, **Figure 17** shows a positive result for cri-du-chat syndrome. The probes are binding to two long arms of chromosome 5 but only one short arm. One of the chromosome 5s must therefore be missing part of its short arm.

FISH is an elegant technique that produces dramatic images of our chromosomes. Unfortunately, FISH is also expensive, time consuming, and requires a high degree of skill. For these reasons, FISH is slowly being replaced with PCR and DNA chip based methods. Versions of these techniques have been developed that can accurately quantify a person's DNA. For example DNA from a person with cri-du-chat syndrome will contain 50% less DNA from the end of chromosome 5. These techniques are very useful if the suspected abnormality is a deletion, a duplication, or a change

in chromosome number. They are less useful for diagnosing chromosome inversions and translocations because these rearrangements often involve no net loss or gain of genes.

In the future, all of these techniques will likely be replaced with DNA sequencing. Each new generation of genome sequencing machines can sequence more DNA in less time. Eventually it will be cheaper just to sequence a patient's entire genome than to use FISH or PCR to test for specific chromosome defects. More details on DNA sequencing can be found in the chapter 33.

Figure 17.
A positive result for cri-du-chat syndrome. This diagram is based upon actual results. Cells from a patient's blood were prepared to show an interphase nucleus (a) and mitotic chromosomes (b). The DNA has been coloured blue with DAPI. The green fluorescent probe is binding to the tip of the short arm of chromosome 5 (shown here as open circles). This is the region absent in cri-du-chat. The red fluorescent probe is binding to the middle of the long arm of the same chromosome (filled circles). This probe is used as a control.
(Original-Harrington- CC BY-NC 3.0)

_____

SUMMARY:

- 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
double strand break
non-homologous end joining
chromosome rearrangement
meiotic crossover
Alu transposable elements
terminal deletion
interstitial deletion
deletion
deletion loop
pseudo-dominant
inversion
paracentric inversion
pericentric inversion
inversion loop
tandem duplication

inverse duplication
insertional duplication
duplication
translocation
reciprocal translocation
Robertsonian translocation
Tetrad
Alternate (balanced)
Adjacent-1 (unbalanced)
reduced fertility
karyotype
46,sex,deletion(5)
(cri-du-chat syndrome)
46,sex,inversion(9)
bright field microscopy
Giemsa stain
fluorescence in situ hybridization
fluorescent DNA probe

STUDY QUESTIONS:

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

2) If Drosophila geneticists want to generate mutant strains with deletion mutations, they expose flies to gamma rays. What does this imply about gamma rays?

3) Design a FISH based experiment to find out if someone is a 47,XXX female or a 47,XYY male.

CHAPTER 25 – CHANGES IN CHROMOSOME NUMBER

Figure 1.
Xenopus laevis, and other species in the *Xenopus* genus, are one of the few animals that are polyploidy. *X. laevis* is tetraploid (4n), but other species can get up to dodecaploid (12n).
(Wikimedia Commons-P.Narbonne, D.Simpson, J..Gurdon- CC BY 2.5

INTRODUCTION

So far in this textbook, we have talked about cells and organisms that are haploid and diploid. Having the appropriate number of chromosomes is important for allowing mitosis and meiosis to occur. Having too many or too few individual chromosome, or whole sets of chromosomes can lead to cell replication or fertility problems. This chapter we will discuss the repercussions of having too many or too few individual chromosomes, known as **aneuploidy**, or having multiples of whole chromosome sets, known as **polyploidy**.

Most organisms of all kingdoms are haploid or diploid. Occasionally though, particularly in plants, you will see chromosomes sets higher than diploid. This is known as polyploidy. When coming from a typically diploid plant, and increasing the ploidy in even numbers, the resulting plant is typically healthy, and often with larger fruits produced. However, when increasing to an odd number, it makes it difficult for gamete production and often leads to infertility (seedless varieties).

As opposed to polyploidy, where the plant is often healthy, aneuploidy plants and animals (losses or multiples of individual chromosomes) often see

more deleterious problems. Having the correct expression levels of genes is important for the function of the organism. Since chromosomes have large numbers of genes on them, missing or gaining whole chromosomes can cause more serious gene dosage problems. Aneuploidy is caused through incorrect segregation in meiosis or mitosis, and if there are living organisms with aneuploidy, they often have difficulty with meiosis or mitosis as well.

1.   PLOIDY NOTATION

1.1.   NOTATION OF DNA CONTENT AND CHROMOSOME CONTENT IN DIPLOID ORGANISMS

The amount of DNA within a cell changes following each of the following events: fertilization, DNA synthesis, mitosis, and meiosis (**Figure 2**). 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**),
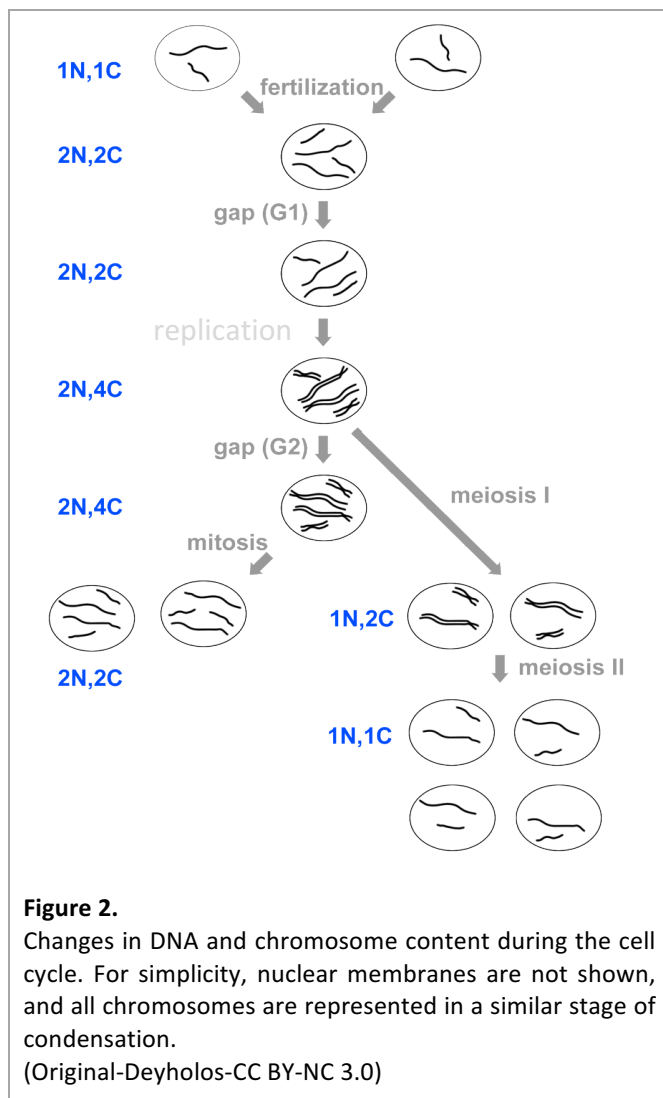
Figure 2.
Changes in DNA and chromosome content during the cell cycle. For simplicity, nuclear membranes are not shown, and all chromosomes are represented in a similar stage of condensation.
(Original-Deyholos-CC BY-NC 3.0)

Figure 3.
Here is an example of a diploid (2x), tetraploid (4x) and hexaploid (6x) cell. Each cell has 2 chromosomes, a long chromosome with acrocentric centromere and a short chromosome with metacentric centromere. Diploids have 2 copies of this chromosome, tetraploids have 4 and hexaploids have 6.
(Original-L. Canham-CC BY-NC 3.0)

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.

N and C values were introduced in Chapter 14.

1.2. NOTATION IN POLYPLOID ORGANISMS

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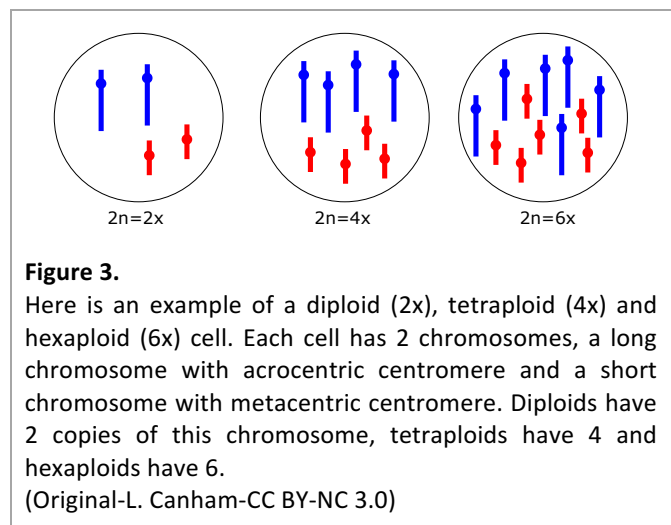
combine the "x" notation with the "n" notation already defined previously in this chapter. Thus, for both diploids and polyploids, "n" is the number of chromosomes in a gamete, and "2n" is the number of chromosomes following fertilization. For a diploid, therefore, n=x, and 2n=2x. But for a tetraploid,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

2. POLYPLOIDY

2.1. POLYPLOIDY FROM CHANGES IN WHOLE SETS OF CHROMOSOMES

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) (**Figure 2**), 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.

2.2. 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.

2.3. 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 (**Figure 4**). 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.
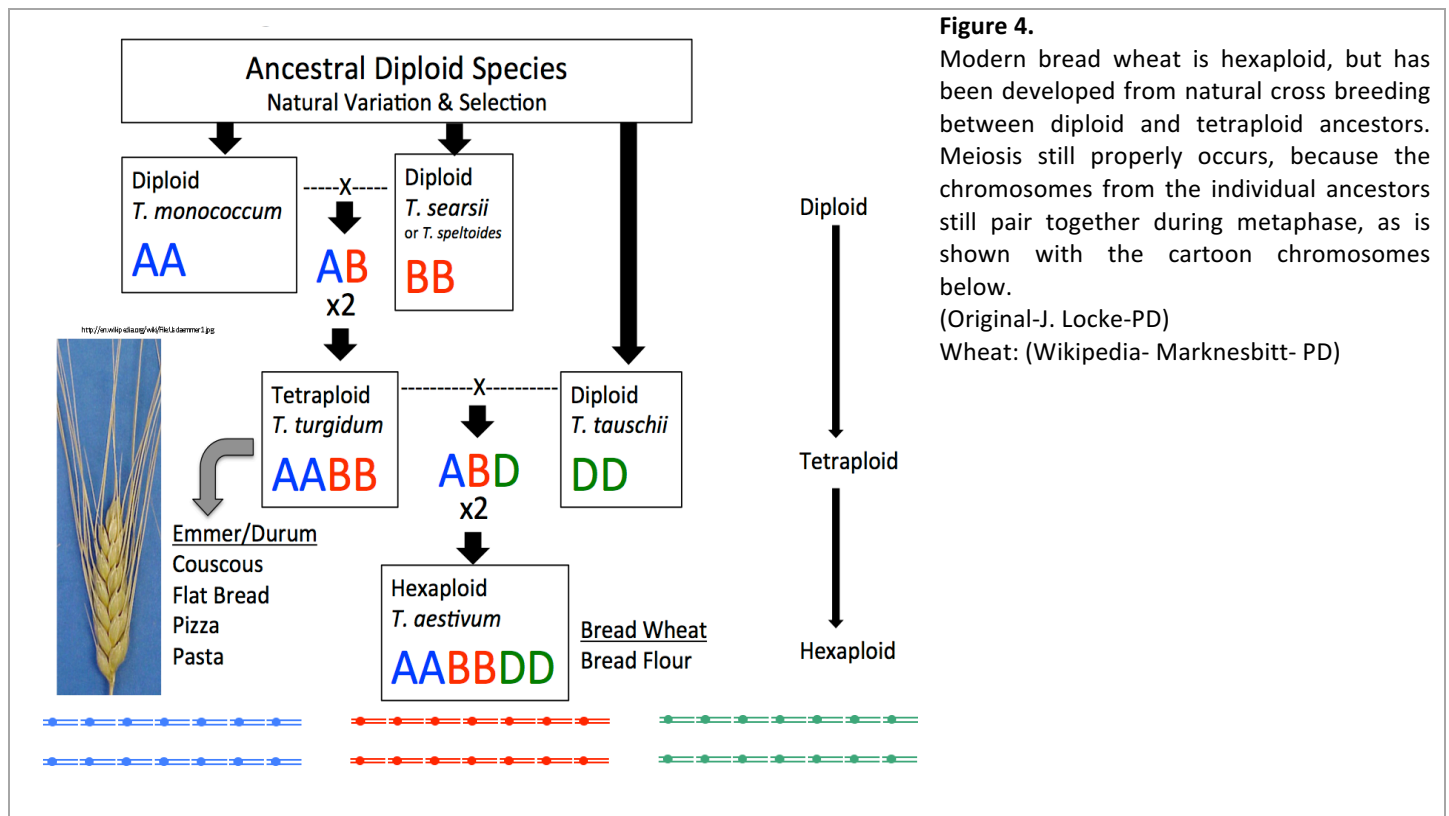
Figure 4.
Modern bread wheat is hexaploid, but has been developed from natural cross breeding between diploid and tetraploid ancestors. Meiosis still properly occurs, because the chromosomes from the individual ancestors still pair together during metaphase, as is shown with the cartoon chromosomes below.
(Original-J. Locke-PD)
Wheat: (Wikipedia- Marknesbitt- PD)

2.4. BANANAS, WATERMELONS, AND OTHER SEEDLESS PLANTS ARE TRIPLOID

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 5**) have a similar explanation.
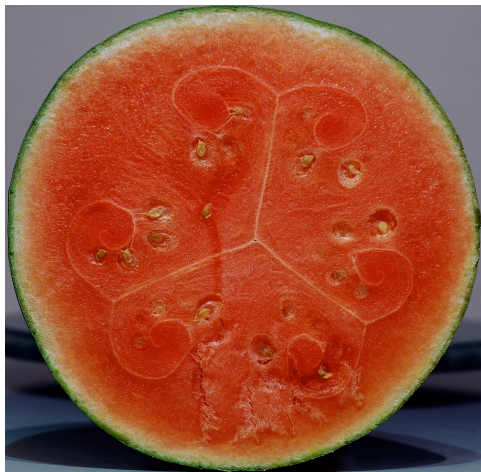
Figure 5.
Seedless watermelon is triploid, with white, aborted seeds within the flesh.
(Flickr- justmakeit- CC BY-NC 2.0)

Figure 6.
Polyploidy in strawberries. The sweet, flavorful, wild diploid is on the left, while the huge, cultivated octoploid is on the right. (Wikipedia- Left: Ivar Leidus/Right: David Monniaux - CC BY-SA 3.0)

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 6**). 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.

2.5. 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 an 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 (**Figure 7**) than males, who are only needed for reproduction
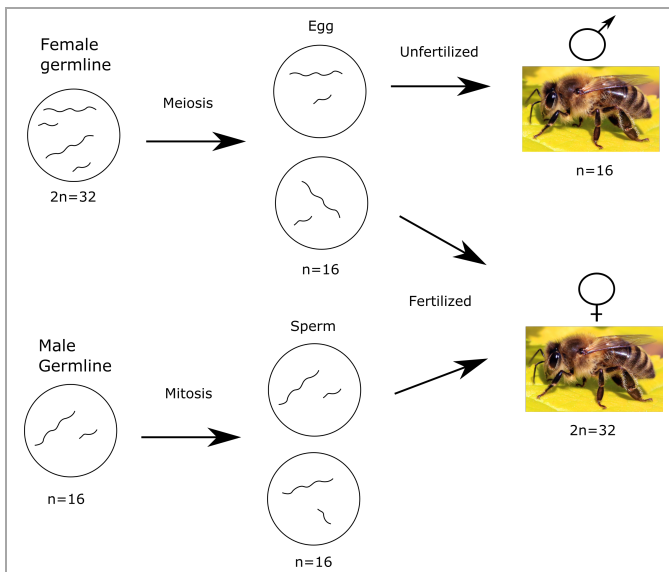
Figure 7.
Hymenoptera are haploid males and diploid females.
Bee: (Sharp Photography – Charlesjsharp- CC BY-SA 3.0)
(Original-L. Canham- CC BY-SA 3.0)
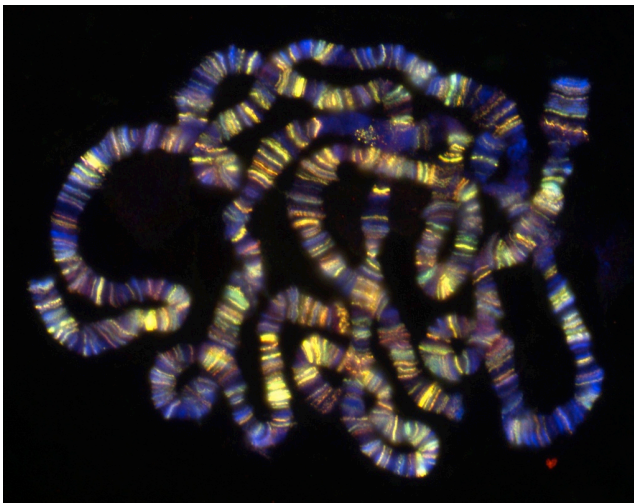
Figure 8.
Endoreduplicated chromosomes from a Drosophila salivary gland cell. The banding pattern is produced with fluorescent labels.
(Flickr-Elissa Lei, Ph.D. @ NIH- CC BY 2.0)

3. 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 8**) 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.

4. ANEUPLOIDY

4.1. NOMENCLATURE

If something goes wrong during cell division, an entire chromosome may be lost and the cell will lack all of these genes. Conversely, an entire chromosome may be improperly included into the new cell. These chromosomal abnormalities are known as **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 widely known 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 autosomal aneuploidies are lethal at an early stage of embryonic development.

Aneuploidy can arise through 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/or missing chromosomes.

Note that aneuploidy usually affects the number of only one type of chromosome and is therefore distinct from **polyploidy**, in which the entire chromosome set is duplicated (see previous section). Unlike aneuploidy, which is almost always deleterious, polyploidy can be beneficial in some organisms, particularly many species of food plants. Higher ploidy levels often result in larger plants and fruits (**Figure 6**).

This section will go into the details of the causes of aneuploidy and the consequences and diseases associated with them.

4.2. 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. In Meiosis I, it is homologous chromosomes, which are synapsed at that time, that segregate and move apart.

4.3. CONSEQUENCE: DECREASED VIABILITY

A non-disjunction event results in daughter cells having an abnormal number of chromosomes. Cells, such as the parent cell in **Figure 9**a, 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 called **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 on the chromosome. Genes normally produce a 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 (on its autosomes). But if one of these cells suddenly had only one copy (or three copies) of all the genes on a whole chromosome, the amount of product would be either 50% (or 150%) of what was normal. The cell could probably tolerate such a change for a single gene 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 would be more than tolerable for the daughter cells. They have what is called an **unbalanced** genotype, which usually kills the cell (decreases their viability).

If a **first division** or **second division nondisjunction** event occurs during meiosis the result is an unbalanced gamete (**Figure 10**b and c). The gamete in this case can often be functional, but after fertilization the embryo will be genetically unbalanced. This usually leads to the death of the cell or embryo at some point in development. There are some exceptions to this in humans and these will be presented later in this chapter.

5. 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, 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.
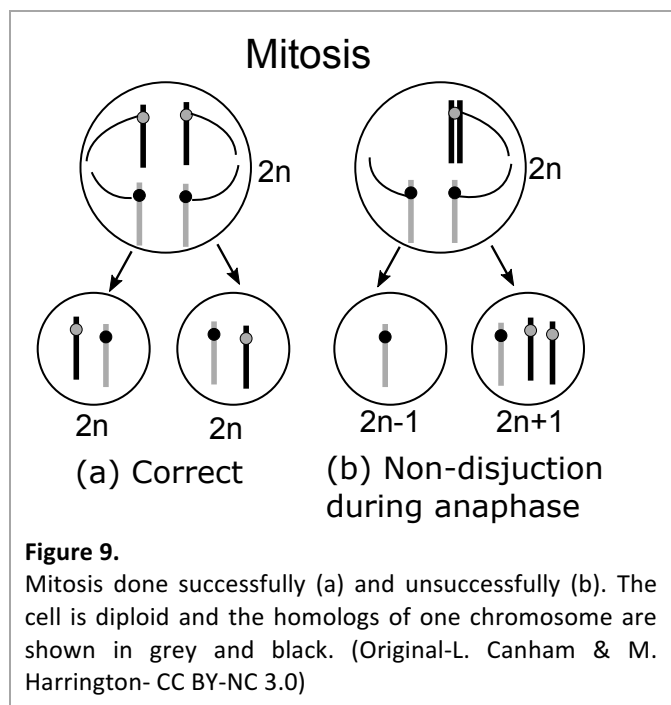
Figure 9.
Mitosis done successfully (a) and unsuccessfully (b). The cell is diploid and the homologs of one chromosome are shown in grey and black. (Original-L. Canham & M. Harrington- CC BY-NC 3.0)

Meiosis

(a) Correct

(b) Non-disjunction during anaphase I

(c) Non-disjunction during anaphase II

Figure 10.
Meiosis done successfully (a) and unsuccessfully (b and c) (Original-L. Canham & M. Harrington- CC BY-NC 3.0)
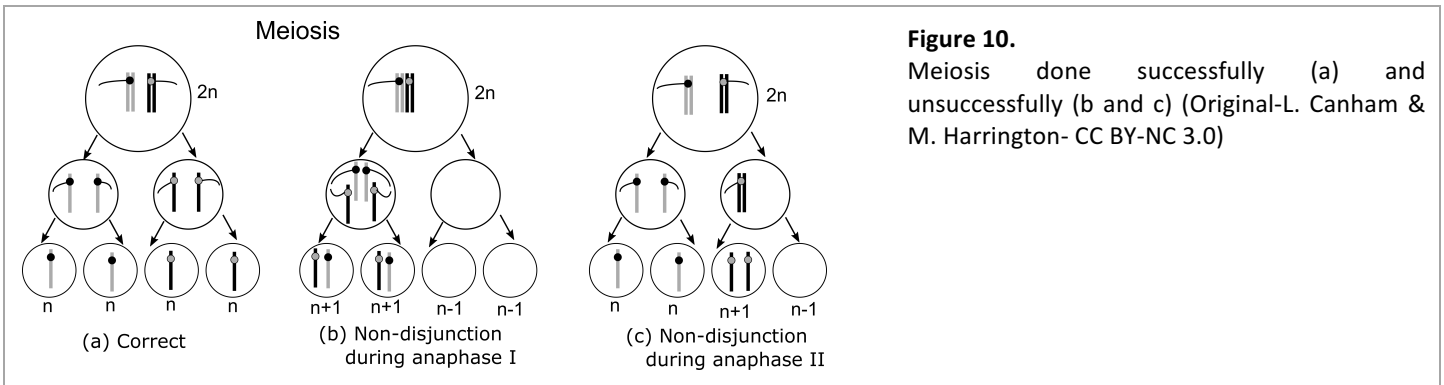
5.1. AUTOSOMAL CHROMOSOME ABNORMALITIES - DOWN SYNDROME

The most common chromosome number abnormality is trisomy-21 or, as it is more commonly known, **Down syndrome** (**Figure 11**). 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) 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 meiosis 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 12**).

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. John Down first described the disease in 1866, but it was not until 1959 when its chromosomal basis was discovered.

Figure 11.
A young girl with Down Syndrome, which can result from an extra chromosome 21. (Flickr-Andreas-Photography-CCBY-NC 3.0)

Meiocyte in male parent

Meiocyte in female parent

sperm   oocyte

zygote

Figure 12.
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-M. Harrington/L. Canham-CC BY-NC 3.0)

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.

5.2. SEX CHROMOSOME ABNORMALITIES - XYY AND XXX

While fetuses trisomic for any one of the other autosomes seldom survive to term, the situation is quite different for the sex chromosomes. Approximately 1 in 1000 males has an extra Y chromosome and yet most are unaware of it! There appears to be little harm in having two Y chromosomes because they have relatively few genes. Similarly, 1 in 1000 females has an extra X

Figure 13.
Simplified dosage compensation for XXX female mammals. In mitosis, both X chromosomes are condensed, as normal. In interphase, one chromosome is selectively silenced and remains condensed. The condensed chromosome is called a Barr Body. Mammals that are XXX compensate for the extra X chromosome the same way, but instead silences 2 of the Xs while the third stays active.
For more information of mammalian dosage compensation, see Chapter 22
(Original-L. Canham & M. Harrington-CC BY-NC 3.0)

chromosome. This situation also appears relatively harmless, although for a different reason. Normally in female mammals (humans 46,XX), one of the two X chromosomes is inactivated in each cell so that there can be genetic balance with males (see Chapter 22). In 47,XXX females, two of the X chromosomes are inactivated, leaving one active, just like in normal 46,XX females (**Figure 13**).

5.3. SEX CHROMOSOME ABNORMALITIES - 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 the 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 lack a Y chromosome (see Chapter 21). They

are viable because the one X is active in most cells. People with this condition do have health problems though: they are typically shorter than average, have an elevated risk of heart defects, and are infertile.

The reason for the health problems is that there are a few genes that have allelic copies on both the X and the Y chromosome. They are found in what is called the **pseudo-autosomal region** of the X and Y chromosome. This region escapes X chromosome inactivation. One of the genes in this region is called *SHOX*. It makes a protein that promotes bone growth. The normal 46,XX and 46,XY individuals have two functioning copies and have average height. People with 47,XYY and 47,XXX genomes have three copies and are typically taller than average, while people with 45,X have one copy and are typically shorter. It is the single copy of *SHOX* and a few of the other genes in the pseudo-autosomal region that causes health problems for women with Turner syndrome.

5.4. SEX CHROMOSOME ABNORMALITIES - 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 (happens in phenotypic males, as well as females). In the embryonic nuclei, all but one of the X chromosome are inactivated. It doesn't matter whether the embryo is male or female. Cells from men with Klinefelter syndrome have a Barr body 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 may make enough sperm to conceive children using intracytoplasmic sperm injection though.

6. 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 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 such changes and an abnormal or lethal phenotype results.

_____

SUMMARY:

- 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.

- Aneuploidy can affect gene balance.

- Errors during anaphase in mitosis or meiosis can lead to trisomy and other forms of aneuploidy.

- Five common forms of aneuploidy in humans are 47,XY,+21 or 47,XX,+21 (Down syndrome), 47,XYY, 47,XXX, 45,X (Turner syndrome) and 47,XXY (Klinefelter syndrome).

KEY TERMS

aneuploidy
polyploidy
n
c
replicated chromosome
x
monoploid
sterile
tetravalent
octoploid
hexaploid
triploid
gene balance
cellular network
non-disjunction
euploid

balanced
unbalanced
first division nondisjunction
second division nondisjunction
karyotype
46,XX
46,XY
47,sex,+21 (Down syndrome)
trisomy
47,XYY
47,XXX
monosomy
45,X (Turner syndrome)
pseudo-autosomal region
47,XXY (Klinefelter syndrome)

STUDY QUESTIONS

1) 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) 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?

3)
   a) Why is aneuploidy more often lethal than polyploidy?
   b) Which is more likely to disrupt gene balance: polyploidy or duplication?

4) 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 labeled.

5) 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).

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

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

8) 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,
   f) 47,XXY

9) Why can people survive with trisomy-21 (47,sex,+21) but not monosomy-21 (47,XY,-21 or 47,XX,-21)?

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

Recommended course:

Ethical Issues in Genetics

GENETICS 424: Ethical Issues in Genetics

★ 3 (fi 6) (second term, 0-3s-0). A seminar and discussion course where students will use their existing knowledge of genetics to investigate, evaluate, and discuss how the field of genetics affects society. Students participate in classroom presentations, written submissions and discussions that may include medical research ethics, genetically modified organisms (GMOs), gene patenting, and other current topics. Enrollment is limited and is by permission of the instructor(s). Prerequisite: Any two GENET 300-level lecture courses.

Offered: Winter Term. Wednesday 2:00-5:00 PM
Contact: Dr. J. Locke, john.locke@ualberta.ca or Dr. H. McDermid, hmcdermi@ualberta.ca

Calendar link:
http://calendar.ualberta.ca/preview_course_nopop.php?catoid=6&coid=44634

CHAPTER 26 – GENE INTERACTIONS

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

INTRODUCTION

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

1. MENDELIAN DIHYBRID CROSSES

1.1. MENDEL'S SECOND LAW (A QUICK REVIEW)

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

If the inheritance of seed color was truly independent of seed shape, then when the F_1 dihybrids were crossed to each other, a 3:1 ratio of one trait should be observed within each phenotypic class of the other trait (**Figure 2**). Using the product law, we would therefore predict that if ¾ of the progeny were green, and ¾ of the progeny were round, then ¾ × ¾ = 9/16 of the progeny would be both round and green. Likewise, ¾ × ¼ = 3/16 of the progeny would be both round and yellow, and so on. By applying the product 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, <u>if</u> 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.
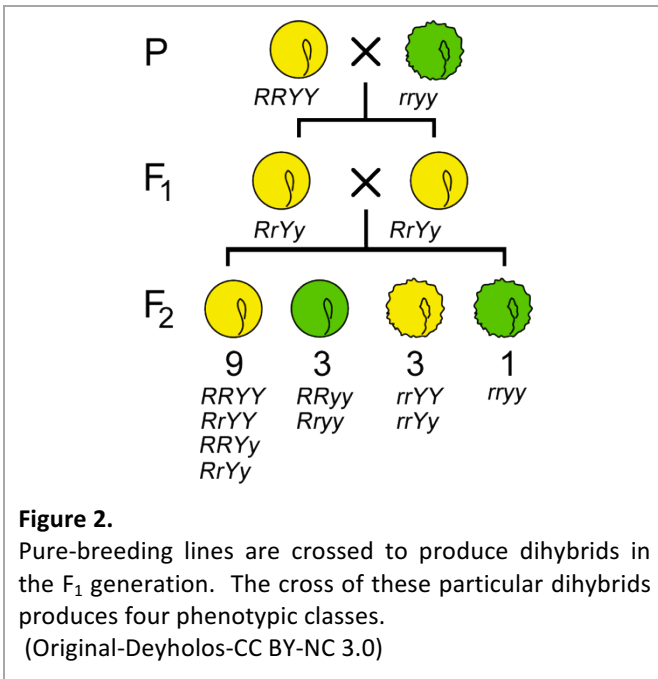
Figure 2.
Pure-breeding lines are crossed to produce dihybrids in the F$_1$ generation. The cross of these particular dihybrids produces four phenotypic classes.
(Original-Deyholos-CC BY-NC 3.0)

Figure 3.
A Punnett Square showing the results of the dihybrid cross from **Figure** 2. Each of the four phenotypic classes is represented by a different color of shading.
(Original-Deyholos-CC BY-NC 3.0)

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 expectations, 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.

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

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

2. Epistasis and Other Gene Interactions

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

Figure 4.
Retrievers with different coat colors: (from left to right) black, chocolate, yellow: an example of recessive epistasis phenotypes. (Flickr- Pirate Scott - CC BY-NC 2.0)
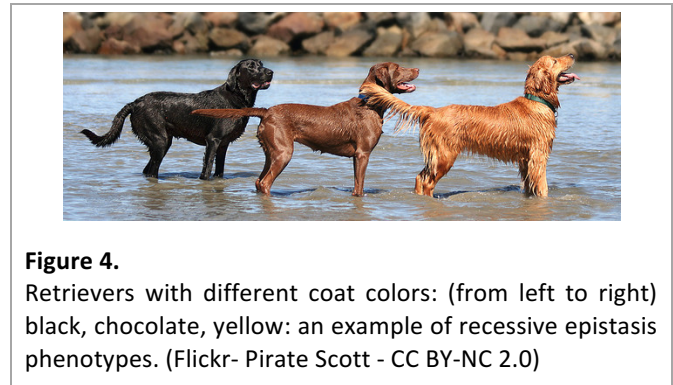
2.1. Recessive epistasis

Epistasis (which means "standing upon") occurs when the phenotype of one locus masks, or prevents, the phenotypic expression of another locus. Thus, following a dihybrid cross fewer than

the typical four phenotypic classes will be observed with epistasis. As we have already discussed, in the absence of epistasis, there are four phenotypic classes among the progeny of a dihybrid cross. The four phenotypic classes correspond to the genotypes: *A_B_, A_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 4**), 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-white), regardless of the genotype at the *B* locus (**Figure 4**, right side).

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**. A table showing all the possible progeny genotypes and their phenotypes is shown in **Figure 5**.

	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 5.
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 BY-NC 3.0)

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 of **12:3:1**, which can be viewed as a modification of the 9:3:3:1 ratio in which the *A_B_* class is combined with one of the other genotypic classes (9+3) that contains a dominant allele. One of the best known examples of a 12:3:1 segregation ratio is fruit color in some types of squash (**Figure 6**). 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 7**). One possible biological interpretation of this segregation pattern is that the function of the *A* allele somehow blocks an early stage of pigment synthesis, before either yellow or green pigments are produced.

Figure 6.
Green, yellow, and white fruits of squash. (Flickr-Unknown-CC BY-NC 3.0)

	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 7.
Genotypes and phenotypes among the progeny of a dihybrid cross of squash plants heterozygous for two loci affecting fruit color. (Original-Deyholos-CC BY-NC 3.0)

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. With yet another pigmentation pathway example, wheat shows this **duplicate gene action**. The biosynthesis of red pigment near the surface of wheat seeds (**Figure 8**) 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 <u>both</u> genes, shows a phenotype that differs from that produced by any of the other genotypes (**Figure 9**). 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 8.
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 9.
Genotypes and phenotypes among the progeny of a dihybrid cross of a wheat plants heterozygous for two loci affecting seed color. (Original-Deyholos-CC BY-NC 3.0)

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 <u>either</u> *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 10).**
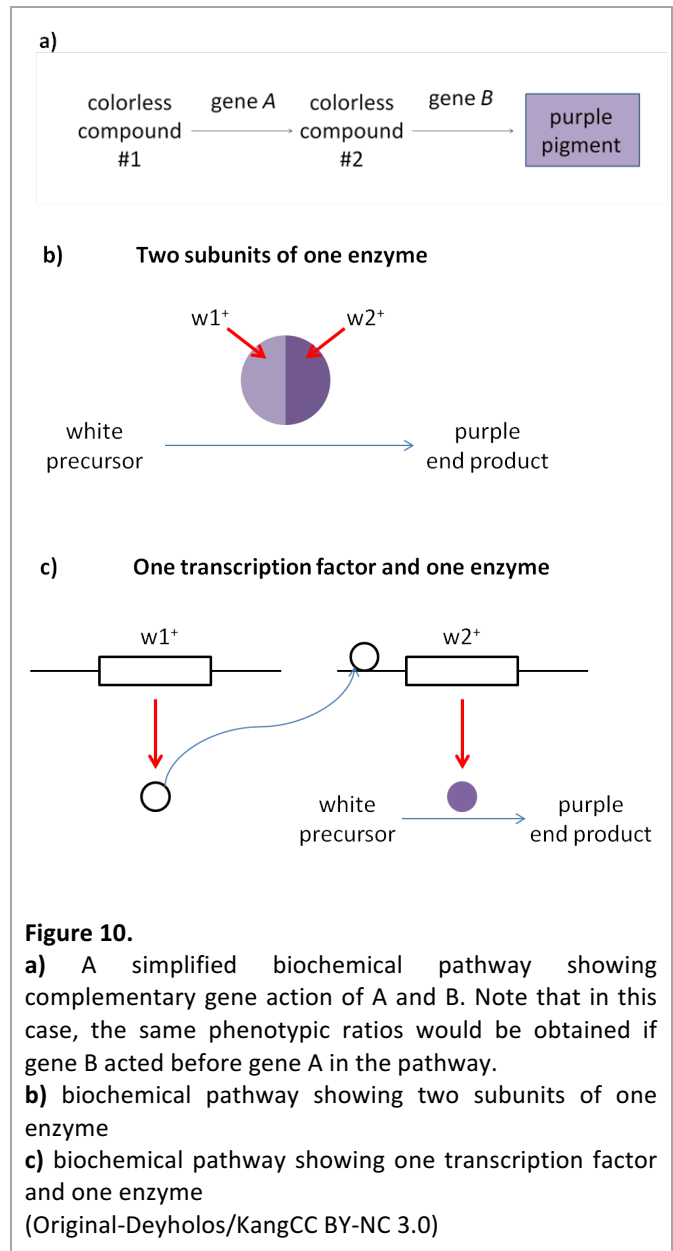
Figure 10.
a) A simplified biochemical pathway showing complementary gene action of A and B. Note that in this case, the same phenotypic ratios would be obtained if gene B acted before gene A in the pathway.
b) biochemical pathway showing two subunits of one enzyme
c) biochemical pathway showing one transcription factor and one enzyme
(Original-Deyholos/KangCC BY-NC 3.0)

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 11**). The modified 9:7 ratio may therefore be obtained when two genes act together in the same biochemical pathway, and when their loss of function phenotypes are indistinguishable from each other or from the loss of both genes. There are also other possible biochemical explanations for complementary gene action.

	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 11.
Genotypes and phenotypes among the progeny of a dihybrid cross of a hypothetical plant heterozygous for two loci affecting flower color. (Original-Deyholos-CC BY-NC 3.0)

2.5. Genetic Suppression and Enhancment

A **suppressor mutation** is a type of mutation that usually had no phenotype of its own, but act to suppress (makes more wildtype, less mutant) the phenotypic expression of another mutation that already exists in an organism. On the other hand, **enhancer mutations** have the opposite effect of suppressor mutations. They make the phenotype more mutant and less wild type (enhance the mutant phenotype).

For example, if a fly has a **white^mottled** (w^m) phenotype, it can be suppressed to look more like **white^+** phenotype by a dominant Suppressor mutation (S^-), or Enhanced to look more like **white^-** by a dominant enhancer mutation (E^-) (**Figure 12**). Note that the w^m allele is <u>recessive</u> to *white^+* (w^+) but <u>dominant</u> to *white^-* (w^-).

Figure 12.
Mutation in the *white* gene impacts the pigmentation in *Drosophila* eyes. Note that white^mottled is recessive to white^+ and dominant to white^-.
(Original-Locke-CC BY-NC 3.0)
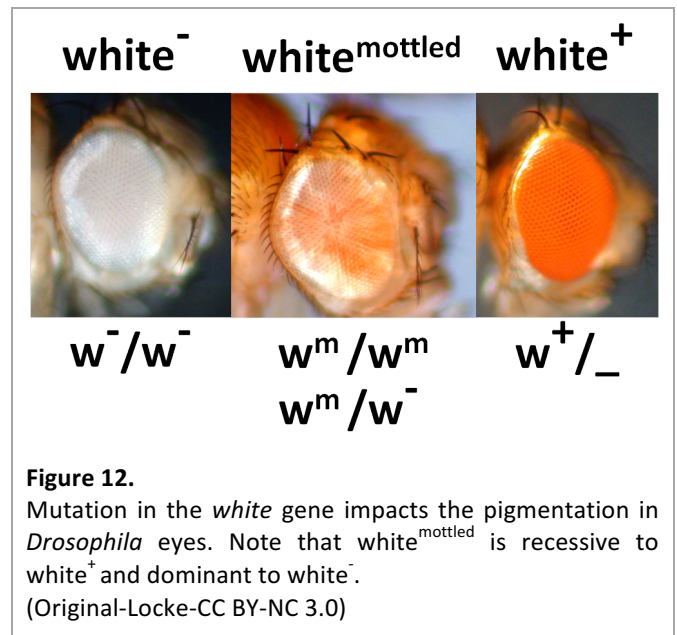
The suppressor mutation can be within the original gene itself (**intragenic**) or outside the gene, at some other gene elsewhere in the genome (**extragenic**). For example, a frameshift mutation caused by a deletion in a gene can be reverted, or suppressed, by an insertion in the same gene to restore the original reading frame (<u>intragenic</u> suppressor mutation). A case of an <u>extragenic</u> suppressor mutation, on the other hand, a can occur when a mutant phenotype caused by mutation in gene A is suppressed by a mutation in gene B. In extragenic suppressor mutation, there are two types of suppressor mutations: (1) **dominant suppression** and (2) **recessive suppression.**

2.6. Dominant suppression

In **dominant suppression**, the <u>mutant</u> suppressor allele is dominant to the <u>wild type</u> suppressor allele. Therefore, one mutant suppressor allele is sufficient to suppress the mutant phenotype. For example, in **Figure 13**, the *Su* gene represents the suppressor gene. Flies that have at least one *Su^-* allele, even though they have homozygous recessive w^m/w^m genotype, will show a wild-type (w^+) phenotype. A fly will have w^m phenotype only if it has homozygous recessive *Su^+/Su^+* genotype. If $w^+/w^{mottled}$; *Su^+/Su^-* flies are crossed together, the ratio of white^+ (wild type) to white^mottled (mutant) would be 15:1.

Figure 13 – Dominant Suppression.
Drosophila cross and its Punnett square showing the effects of dominant suppression of *Su* gene on the *white* gene. Note that A = white$^+$, a = white^{mottled,} B = Su$^+$ b = Su$^{-,}$ and ___ (blank) = any allele. (Original-Kang-CC BY-NC 3.0)

Figure 14 – Recessive Suppression.
Drosophila cross and its Punnett square showing the effects of recessive suppression of *Su* gene on the *white* gene. Note that A = white$^+$, a = white^{mottled,} B = Su$^+$ b = Su$^{-,}$ and ___ (blank) = any allele. (Original-Kang-CC BY-NC 3.0)

2.7. RECESSIVE SUPPRESSION

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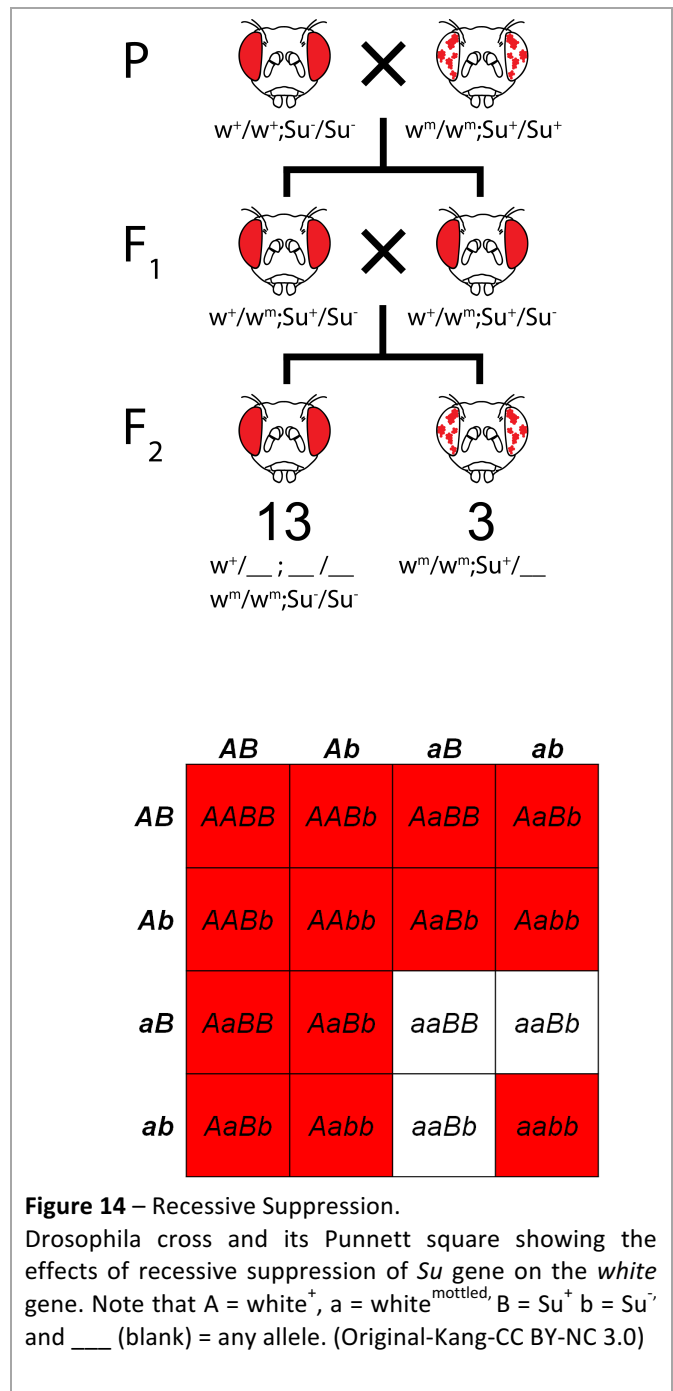
mutation. On the other hand, flies that have the *w^{m}w^m* alleles will have mottled phenotype unless they have homozygous *su$^-$* alleles. If *w$^+$/w^{mottled}; Su$^+$/Su$^-$* flies are crossed together, the ratio of white$^+$ (wild type) to white^{mottled} (mutant) would be 13:3.

2.8. SUMMARY

Table 1. Summary table showing gene interactions and their genotypic and phenotypic ratios.					
Ratio:	9	3	3	1	Ratio
Genotype"	A⁻B⁻	A⁻bb	aaB⁻	Aabb	
None	9	3	3	1	9:3:3:1
	AB	B	aB	ab	
Recessive epistasis	9	3	4		9:3:4
of aa acting on B and b alleles	AB	B	A		
Dominant epistasis	12		3	1	12:3:1
of A acting on B and b alleles	A		aB	ab	
Duplicate genes	15			1	15:1
	A			a	
Complementary genes	9	7			9:7
	A	A			
Recessive suppression	9	3	4		13:3
by aa acting on bb	B	b	B		
Dominant suppression	15			1	15:1
by A acting on bb	B			b	

Shading represents combined classes.

3. EXAMPLE OF MULTIPLE GENES AFFECTING ONE CHARACTER (POLYGENIC INHERITANCE)

3.1. CONTINUOUS VARIATION

Most of the phenotypic traits commonly used in introductory genetics are qualitative, meaning that the phenotype exists in only two (or possibly a few more) discrete, alternative forms, such as either purple or white flowers, or red or white eyes. These qualitative traits are therefore said to exhibit **discrete variation**. On the other hand, many interesting and important traits exhibit **continuous variation**; these exhibit a continuous range of phenotypes that are usually measured quantitatively, such as intelligence, body mass, blood pressure in animals (including humans), and yield, water use, or vitamin content in crops. Traits with continuous variation are often complex,

and do not show the simple Mendelian segregation ratios (e.g. 3:1) observed with some qualitative traits. 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 15**).
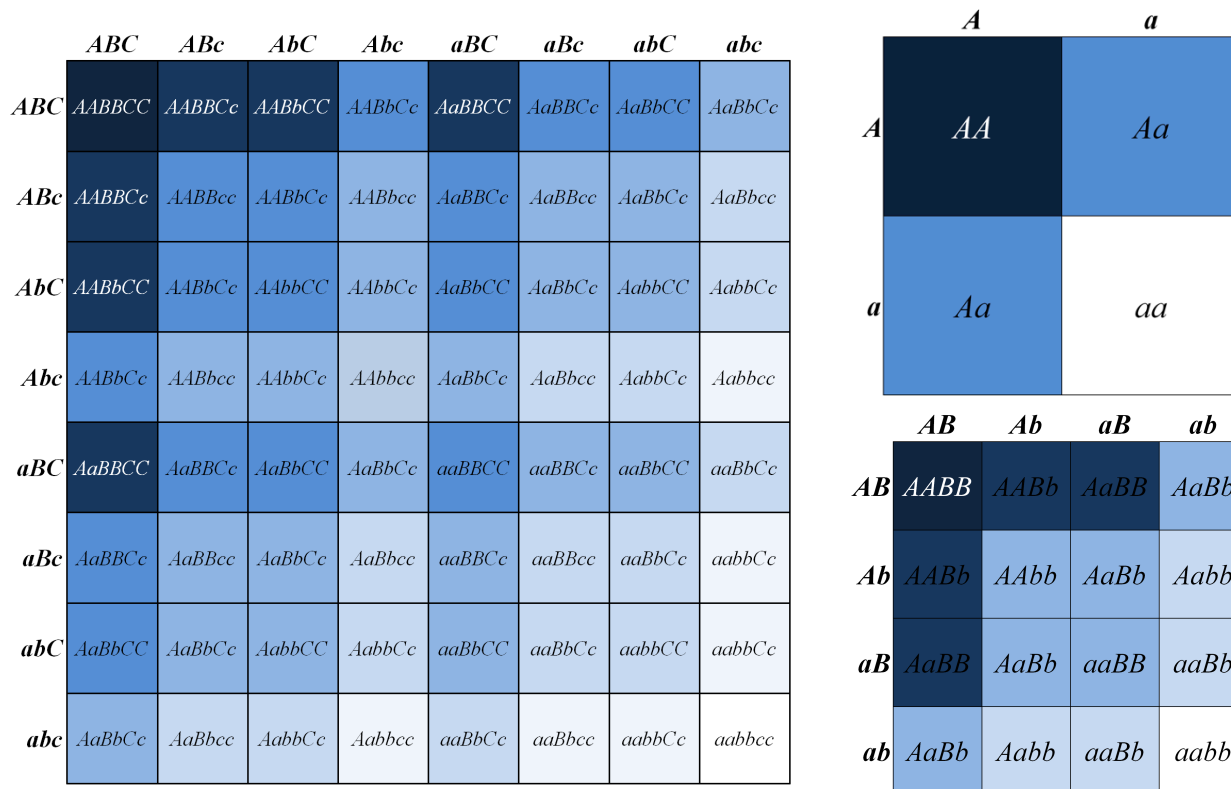
Figure 15.
Punnett Squares for one, two, or three loci. We are using a simplified example of up to three semi-dominant genes, and in each case the effect on the phenotype is additive, meaning the more "upper case" alleles present, the stronger the phenotype. Comparison of the Punnett Squares and the 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 BY-NC 3.0)
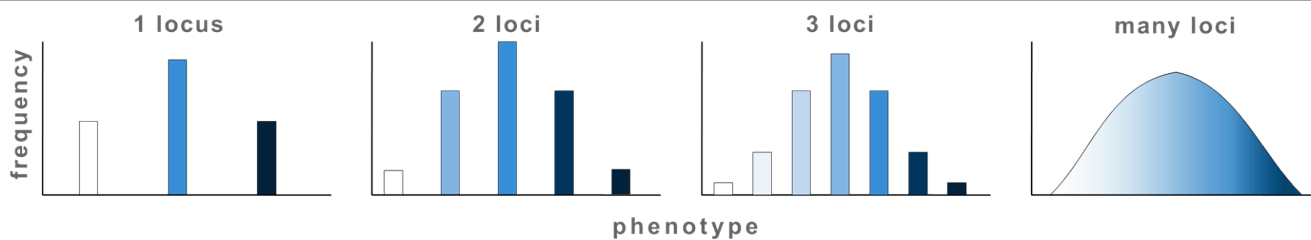
Figure 16.
The more loci that affect a trait, the larger the number of phenotypic classes that can be expected. For some traits, the number of contributing loci is so large that the phenotypic classes blend together in apparently continuous variation. (Original-Deyholos-CC BY-NC 3.0)

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 16**). 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.

3.2. CAT FUR GENETICS –

(ADAPTED FROM CHRISTENSEN (2000) GENETICS 155:999-1004)

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

For example, the X-linked **Orange** gene has two allelic forms. The O^O allele produces orange fur, while the O^B alleles produce non-orange (often black) fur. Note however, that because of X-chromosome inactivation the result is mosaicism in expression. In O^O / O^B female heterozygotes patches of black and orange are seen, which produces the tortoise shell pattern (**Figure 17** A,B). This is a rare example of **co-dominance** since the phenotype of both alleles can be seen. Note that the cat in part A has short fur compared to the cat in part B; recessive alleles at an independent locus (L/l) produce long (*ll*) rather than short (*L_*) fur.

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

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

Piebald spotting is the occurrence of patches of white fur. These patches vary in size due to many reasons, including genotype. Homozygous cats with genotype *ss* do not have any patches of white, while cats of genotype *Ss* and *SS* do have patches of white, and the homozygotes tend to have a larger proportion of white fur than heterozygotes (part F). The combination of piebald spotting and tortoise shell patterning produce a **calico cat**, which has separate patches of orange, black, and white fur.
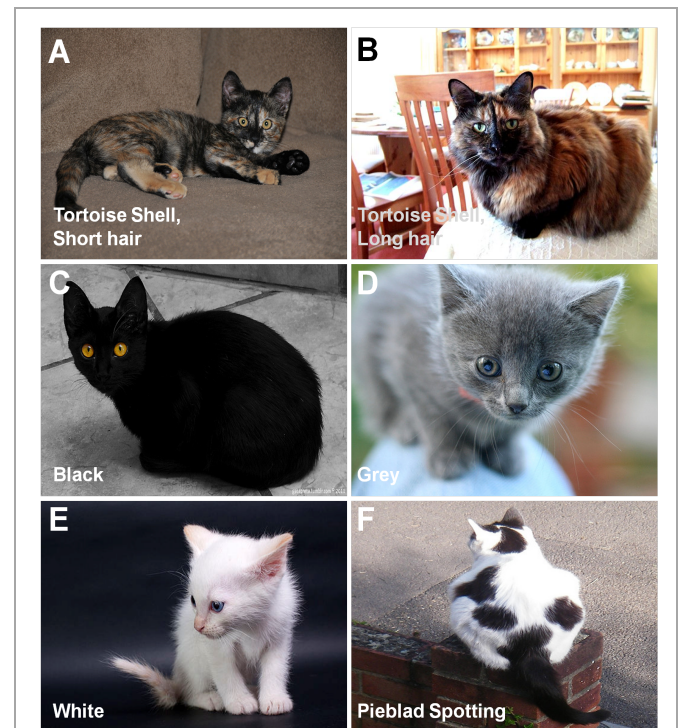
Figure 17.
Representatives of various fur phenotypes in cats. Tortoise shell (A,B) pigmentation in cats with short (A) and long (B) fur; black (C) and grey (D) cats that differ in genotype at the *dilute* locus. The pure white pattern (E) is distinct from piebald spotting (F).
A: (Flickr-Bill Kuffrey-CC BY 2.0), B: (Wikipedia-Dieter Simon-PD), C: (Flickr-atilavelo-CC BY 2.0), D: (Flickr-Waldo Jaquith-CC BY-SA 2.0), E: (Wikipedia-Valerius Geng-CC BY-SA 3.0), F: (Flickr-Denni Schnapp-CC BY-NC-SA 2.0) *Changes: Letters and descriptions were added on the pictures

Table 2. Summary of simplified cat fur phenotypes and genotypes.

Trait	Phenotype	Genotype	Comments
fur length	short	*LL* or *Ll*	*L* is completely dominant
	long	*ll*	
all white fur (non-albino)	100% white fur	*WW* or *Ww*	If the cat has red eyes it is albino, not *W_*. *W* is epistatic to all other fur color genes; if cat is *W_*, can't infer genotypes for any other fur color genes.
	<100% white fur	*ww*	
piebald spotting	> 50% white patches (but not 100%)	*SS*	*S* is incompletely dominant and shows variable expressivity
	< 50% white patches	*Ss*	
	no white patches	*ss*	
orange fur	all orange fur	X^OX^O or X^OY	*O* is X-linked
	tortoise shell variegation	X^OX^B	
	no orange fur (often black)	X^BX^B or X^BY	
dilute pigmentation	pigmentation is intense	*Dd* or *DD*	*D* is completely dominant
	pigmentation is dilute (e.g. gray rather than black; cream rather than orange; light brown rather than brown)	*dd*	
tabby	tabby pattern	*AA* or *Aa*	This is a simplification of the tabby phenotype, which involves multiple genes
	solid coloration	*aa*	
sex	female	XX	
	male	XY	

Note: Phenotypes May Not Be As Expected from the Genotype

4. ENVIRONMENTAL FACTORS

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

Genotype + Environment

$\Rightarrow$ **Phenotype (G + E $\Rightarrow$ P)**

Or:

Genotype + Environment + Interaction$_{GE}$

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

*GE = Genetics and Environment

This interaction is especially relevant in the study of economically important phenotypes, such as human diseases or agricultural productivity. For

example, a particular genotype may pre-dispose an individual to cancer, but cancer may only develop if the individual is exposed to certain DNA-damaging chemicals or carcinogens. Therefore, not all individuals with the particular genotype will develop the cancer phenotype, only those who experience a particular environment.

Penetrance and Expressivity

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

4.1. PENETRANCE

Penetrance is the proportion of individuals with a particular genotype that display a corresponding phenotype (**Figure 18**). It is usually expressed as a percentage of the population. Because all pea plants that are homozygous for the allele for white flowers (e.g. *aa* in **Figure 2** of Chapter 12) actually have white flowers, this genotype is completely (100%) penetrant. In contrast, many human genetic diseases are incompletely penetrant, since not all individuals with the disease genotype actually develop symptoms associated with the disease (less than 100%).

4.2. EXPRESSIVITY

Expressivity describes the variability in mutant phenotypes observed in individuals with a particular phenotype (**Figure 18** and **Figure 19.**). 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).

5. MENDELIAN PHENOTYPIC RATIOS MAY NOT BE AS EXPECTED

5.1. OTHER FACTORS

There are other factors that that affects organism's phenotype and thus appear to alter Mendelian inheritance.
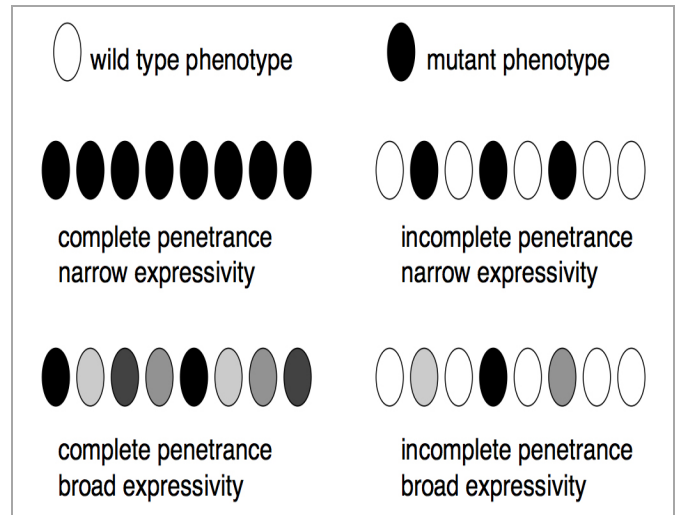
Figure 18.
Relationship between penetrance and expressivity in eight individuals that all have a mutant genotype. Penetrance can be complete (all eight have the mutant phenotype) or incomplete (only some have the mutant phenotype). Amongst those individuals with the mutant phenotype the expressivity can be narrow (very little variation) to broad (lots of variation).
(Original-Locke-CC BY-NC 3.0)
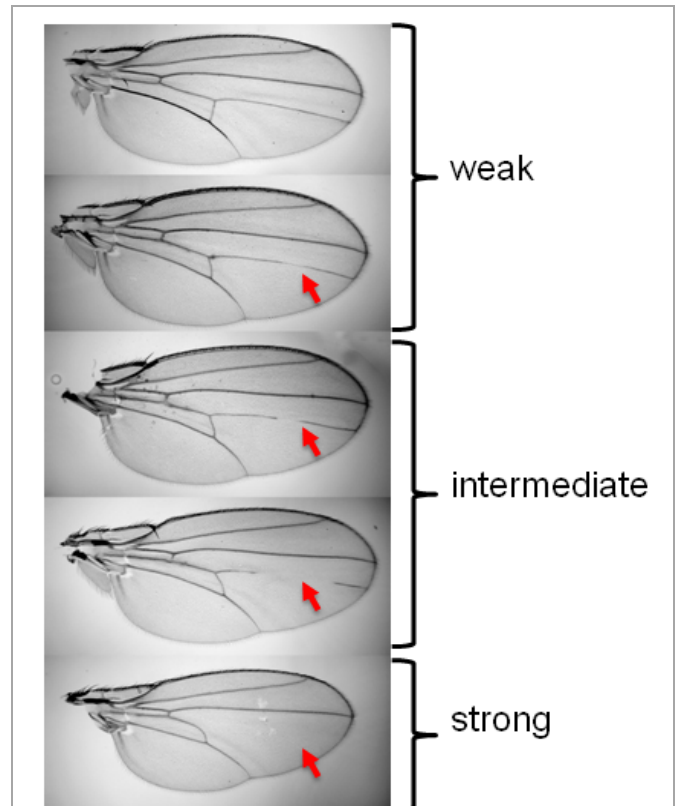
Figure 19.
Mutations in wings of *Drosophila melanogaster* showing weak to strong expressivity.
(Original-J. Locke-CC;AN)

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

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

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

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

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

5.2. THE x^2 TEST FOR GOODNESS-OF-FIT

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

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

_____

SUMMARY:

- Phenotype depends on the alleles that are present, their dominance relationships, and sometimes also interactions with the environment and other factors.

- The alleles of different loci are inherited independently of each other, unless they are genetically linked.

- Many important traits show continuous, rather than discrete variation. These are called quantitative traits.

- Many quantitative traits are influenced by a combination of environment and genetics.

- The expected phenotypic ratio of a dihybrid cross is 9:3:3:1, except in cases of linkage or gene interactions that modify this ratio.

- Modified ratios from 9:3:3:1 are seen in the case of recessive and dominant epistasis, duplicate genes, and complementary gene action. This usually indicates that the two genes interact within the same biological pathway.

- There are other factors that alter the expected Mendelian ratios.

KEY TERMS:

Mendel's Second Law
independent assortment
linkage
dihybrid
Modified Mendelian Ratios
9:3:3:1
9:3:4
12:3:1recessive epistasis
dominant epistasis
complementary action
redundancy
duplicate gene action
Orange

long
dilute
White masking
piebald spotting
calico Discrete variation
Continuous variation
Polygenic traits
G + E = P
penetrance
expressivity
recessive lethal mutations

STUDY QUESTIONS:

1) In the table on the opposite page, 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.

Answer questions 2-4 using the following biochemical pathway for fruit color. Assume all mutations (lower case allele symbols) are recessive, and that *either* precursor 1 or precursor 2 can be used to produce precursor 3.   If the alleles for a particular gene are not listed in a genotype, assume that they are wild-type.
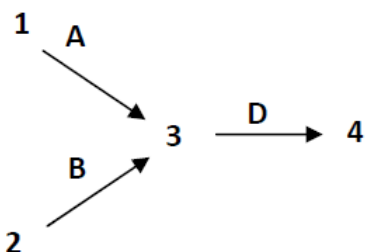
2) If 1 and 2 and 3 are all colorless, and 4 is red, what will be the phenotypes associated with the following genotypes?
a) aa
b) bb
c) dd
d) aabb
e) aadd
f) bbdd
g) aabbdd
h) What will be the phenotypic ratios among the offspring of a cross AaBb × AaBb?
i) What will be the phenotypic ratios among the offspring of a cross BbDd × BbDd?
j) What will be the phenotypic ratios among the offspring of a cross AaDd × AaDd?

3) 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?

4) 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?

5) Which of the situations in questions 2 – 4 demonstrate epistasis?

6) If the genotypes written within the Punnett Square are from the F_2 generation, what would be the phenotypes and genotypes of the F_1 and P generations for:
a) **Figure 5**
b) **Figure 7**
c) **Figure 9**
d) **Figure 11**

7) To better understand how genes control the development of three-dimensional structures, you conducted a mutant screen in Arabidopsis plant 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.

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

9) Again, in reference to question 8, 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*

10) Calculate the phenotypic ratios from a dihybrid cross involving the two loci shown in **Figure 17**. There may be more than one possible set of ratios, depending on the assumptions you make about the phenotype of allele *b*.

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

Table for Question 1

	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

Recommended course:

Human Genetics

GENETICS 418: Human Genetics

★ 3 (fi 6) (second term, 3-1s-0). A survey of human genetic variation and mutation in a molecular genetics context. Molecular basis of diseases and applications to genetic counseling and screening, chromosomal abnormalities, genomic imprinting, cancer genetics, gene mapping, population genetics, multifactorial inheritance, gene therapy, and ethical issues. Prerequisites: any two GENET 300-level lecture courses, GENET 302 is recommended. Credit cannot be obtained for both GENET 418 and 518.
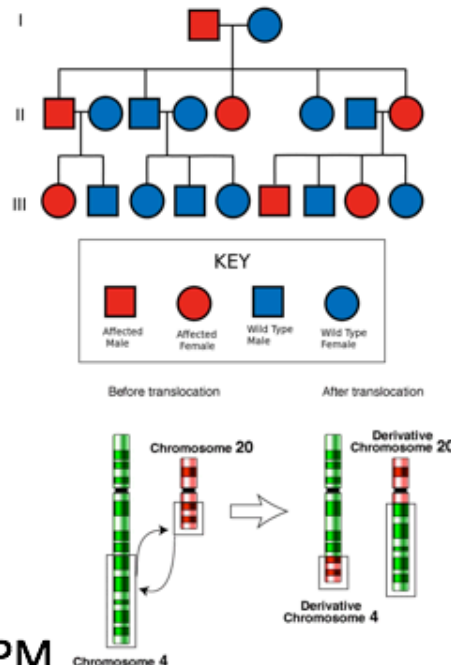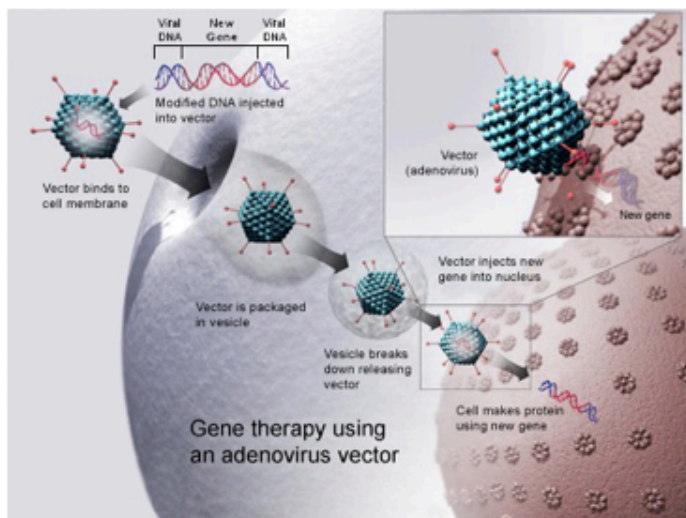
Offered: Winter Term. MWF 1:00-1:50 PM

Contact: Dr. Heather McDermid, hmcdermi@ualberta.ca

Calendar link:
http://calendar.ualberta.ca/preview_course_nopop.php?catoid=6&coid=44631

CHAPTER 27 – PHYSICAL MAPPING OF CHROMOSOMES AND GENOMES

Figure 1.
Genetic map of human chromosome 1 showing a region expanded to the point of showing the genes within that region.
(Wikipedia- SpelgroepPhoenix-CC BY-SA 3.0)
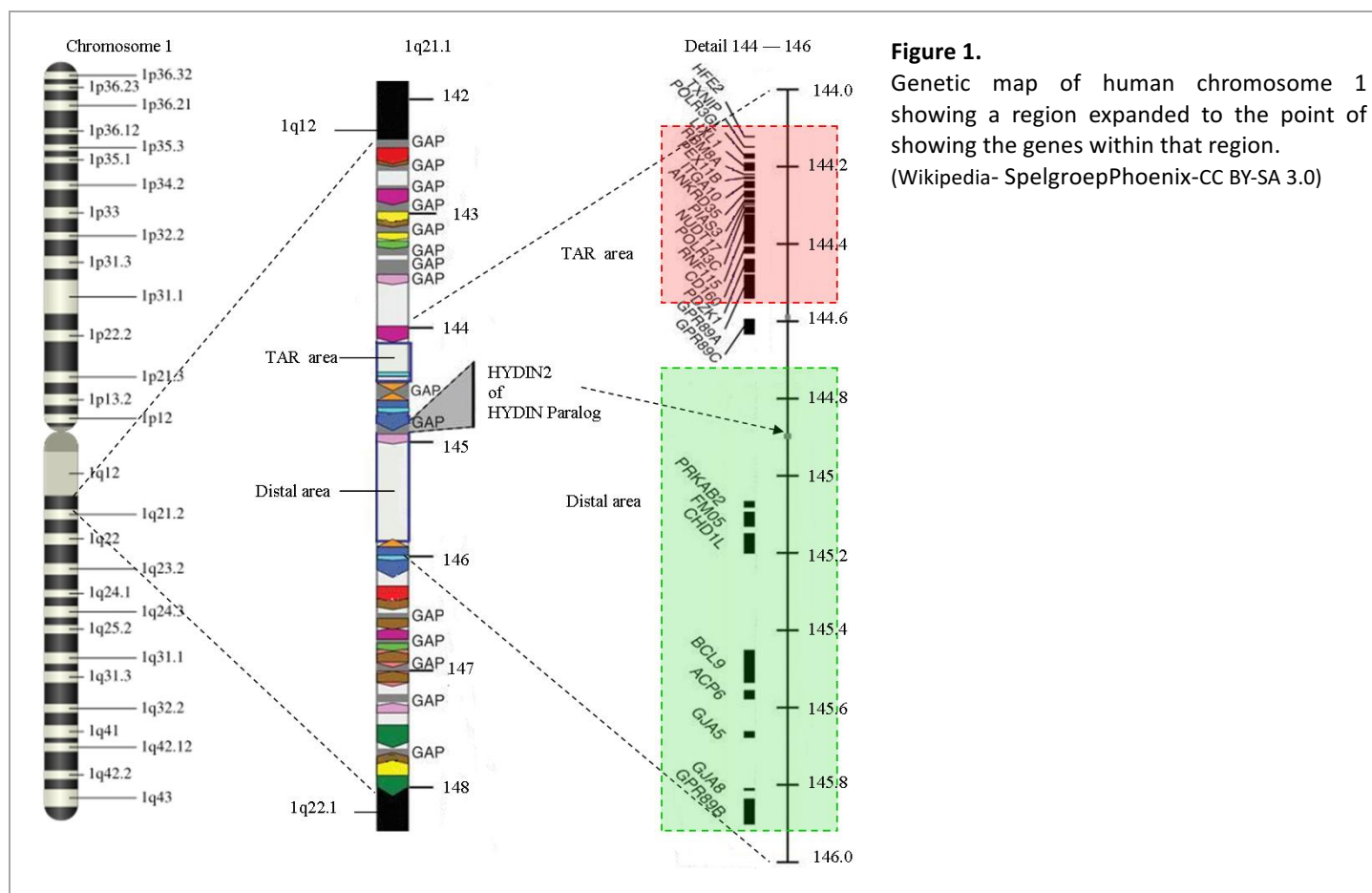
INTRODUCTION

Chromosomes are long duplex molecules of DNA that are either linear or circular and composed of a relatively constant sequence of nucleotides. There are three different ways of describing the linear contents of a chromosome: (1) genetic map, (2) cytogenetic map, and (3) physical map (ultimately the sequence).

1. GENETIC MAP (DISTANCE IN cM, RECOMBINATION FREQUENCY)

In Chapter 18, we described the units of genetic distance (**map units / centiMorgans, cM**) and how this relates to recombination frequency. We can use this information in order to produce a **genetic map**, which is a "map" that shows the locations of

genes along a linear chromosome. Note that map distances are always calculated for one pair of loci at a time. However, by combining the results of multiple pair-wise calculations, a genetic map of many loci on a chromosome can be produced (**Figure 2**). 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
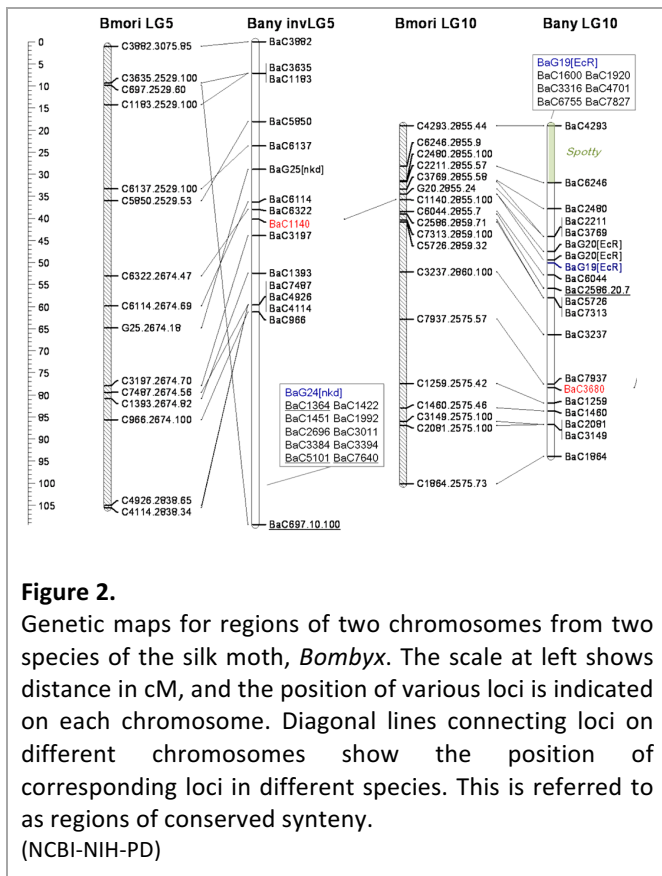
Figure 2.
Genetic maps for regions of two chromosomes from two species of the silk moth, *Bombyx*. The scale at left shows distance in cM, and the position of various loci is indicated on each chromosome. Diagonal lines connecting loci on different chromosomes show the position of corresponding loci in different species. This is referred to as regions of conserved synteny.
(NCBI-NIH-PD)

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.

2. CYTOGENETIC MAP

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. In **Figure 3**, 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.

2.1. CENTROMERE LOCATION

A chromosome has a telomere and centromere, which are usually in a heterochromatin state. Centromere is DNA sequences that are bound by centromeric proteins that link the centromere to microtubules. Centromere can be in the middle (**metacentric**), near to the middle (**submetacentric**), near the end (**acrocentric**), at the end (**telocentric**) or the entire chromosome can act as a chromosome (**holocentric**). Telomeres are repetitive sequences like TTAGGG at the end of the chromosomes that help maintain the length of the chromosome. Another feature is that in a chromosome there are **p arm** (**p**etite = small) and **q arm** (**q**ueue = tail or just the next letter in the alphabet).[*]

Centromere Location	Name	Shape
middle	metacentric	
around middle	submetacentric	
towards one end	acrocentric	
at one end	telocentric	

Table 1. Table showing four types of centromere location. (Original-Harrington/Kang-CC BY-NC 3.0)

---

[*] See https://thednaexchange.com/2011/05/02/p-q-solved-being-the-true-story-of-how-the-chromosome-got-its-name/

Figure 3.

Karyogram of a normal human male karyotype.

(Wikipedia-NHGRI-PD)
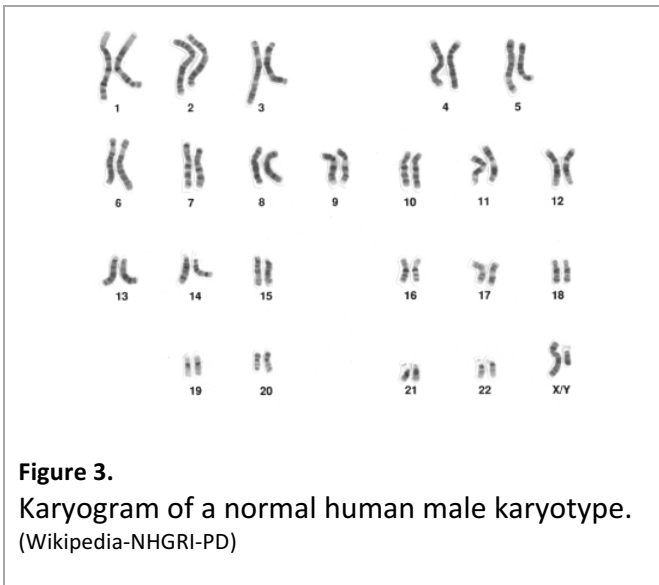
2.2. Karyogram

By convention the chromosomes are arranged into the pattern shown in **Figure 3** 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. Therefore, karyotype is a description of the complete set of chromosomes, and karyogram is an image that visually describes the karyotype.

2.3. Banding

Various stains and fluorescent dyes like Trypsin+Giesma and Quinacrine are used to produce characteristic banding patterns to distinguish all 23 chromosomes. These bands are first grouped in regions, sectioned into bands, and further divided into sub-bands. Notice that the band numbers are start from the centromere and extend towards the tip of each arm (**Figure 4**). 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.

3. Physical Map (DNA Sequence, Restriction Sites)

3.1. Basics

The ultimate physical map is an accurate representation of the DNA sequence of a genome.

CHROMOSOME 5

Figure 4.

Fictional diagram of a human chromosome and its bands. A chromosome has p and q arm, which are both divided by regions. These regions are divided by bands, and these bands are subdivided into sub-bands. The bands are numbered away from the centromere, and sub-bands are renumbered for each bands. Notice that this fictional diagram was made for educational purposes.

(Original-Kang-CC BY-NC 3.0)

These days that sequence is usually held in a computer database and is accessible via the Internet. This wasn't always the case. The first genome sequences were constructed from a series of large, cloned physical fragments of DNA. The map was therefore made from physical entities (pieces of DNA) rather than abstract concepts such as the linkage frequencies between genes that make up a genetic map. 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**.

3.2. Contig construction

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. Typically the set of sequences in a BAC clone library will contain redundant, over lapping sequences, meaning that different clones will contain DNA from the same part of the genome so there are going to be some overlaps. Because of these overlaps, it is possible 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. A set of overlapping clones is called a **contig**. Making a contig map can 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 in another, overlapping BAC clone. These clones are then assumed to overlap each other. This process of finding overlaps can progress to position all the clones into overlapping series that span the genome. Also, if we already know the sequence of one strain of a simple organism, it can be used as a reference for mutant strains and can identify the differences in the sequences.

Small sized genome like Lambda DNA is only 48kb long, but most chromosomes are Mb long. Currently, the only way to construct physical maps of large regions is through the joining of smaller regions to map a larger or whole portion of the chromosome. In order to do this, small, multiple

Figure 5.
A portion of the physical map for human chromosome 4. The entire chromosome is shown at left. The physical map is derived from the small blue lines, each of which represents a cloned piece of DNA approximately 100kb in length. (NCBI-unknown-PD)
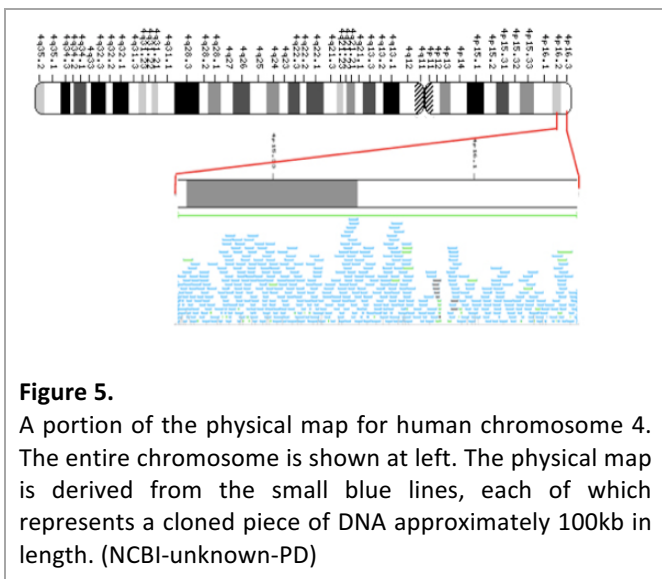
copies of the chromosome have to be broken down into little pieces with different length and frames using restriction enzymes, so that they can partially overlap with each other. The continual overlaps of the fragments will eventually form a whole map of the chromosome. This contiguous assembly of clones is called contig.

3.3. RESTRICTION MAPPING PROCEDURE
Restriction mapping is an inexpensive, quick, and easy method to describe a sample of cloned DNA. It is preferred over DNA sequencing for these reasons, but the sequence is still the ultimate description.

Restriction mapping is the technique for identifying the location of restriction sites, relative to other sites on a DNA molecule. Typically a sample of purified cloned DNA is aliquoted into several tubes and each is treated with several different restriction enzymes or combination of enzymes. These are then separated by agarose gel electrophoresis and the restriction fragment sizes determined by comparison to known size markers. By trial and error, the combination of fragments can be assembled like a linear jigsaw puzzle into a map of the restrictions sites – a **restriction map** (**Figure 7**). One can increase the resolution of the restriction site map by mapping more restriction sites.

3.4. USES OF A RESTRICTION MAP
Restriction mapping is a quick, easy and inexpensive way to characterize and distinguish DNA samples without actually sequencing the DNA; sequences can be represented by series of restriction sites and using this knowledge, one can tell if the DNA of interest is similar or different from others by comparing their degree of overlaps. Also, restriction sites offer positions for convenient manipulation of the DNA. Restriction fragments that contain the gene of interest can be cut out and once the gene is purified from the fragments, it can be sequenced or used as a probe. This is the reason why restriction mapping is still routinely used today, even though sequencing technologies allows us to sequence the whole genome.

Figure 6.
A series of overlapping cloned sequences can be combined to eventually span much larger regions, including whole chromosomes . (Original-Locke-CC BY-NC 3.0)

M = Size marker

C = Cvn   CC•TNAGG

K = Kpn   GGTAC•C

A = Apa   GGGCC•C

N = Nar   GG•CGCC

Figure 7.
By looking at the size of the fragments produced by one restriction enzyme or combination of the restriction enzymes, the location and the order of the restriction site on a chromosome can be identified, forming a restriction map.
(Original-Locke-CC BY-NC 3.0)

SUMMARY:

- There are different types of chromosome maps: genetic (recombination), cytogenetic (metaphase chromosome), and physical maps.

- Recombination frequency is usually proportional to the distance between loci, and so recombination frequencies can be used to create genetic maps.

- Chromosomes can be distinguished cytologically based on their length, centromere position, and banding patterns when stained with dyes.

- Single clones can be restriction mapped and then combined into a contig that represents a larger region of DNA, ultimately the whole chromosome.

- The ultimate physical map is the DNA sequence of the whole chromosome or genome.

KEY TERMS:

map units
centiMorgans
genetic map
recombination frequency
map-based cloning
karytotype

karyogram
contig
physical map
restriction map
contig construction

STUDY QUESTIONS:

1) 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*?

2)   Given the restriction digests and with the fragment sizes shown in the gel diagram, can you construct a map of this linear DNA molecule (Lambda DNA)?

	M	C	C/K	K	K/A	A	A/N	N	C/N
Origin —	□	□	□	□	□	□	□	□	□
								45.7	
						38.4	35.7		
			29.9	29.9					
		26.7							26.7
			17.0	17.0					
		14.2	14.2						11.4
				10.0	10.0	10.0			
			8.2						
		7.6	7.6						7.6
				7.0					
						2.8	2.8		2.8
			1.5	1.5	1.5				

M = Size marker
C = Cvn   CC•TNAGG
K = Kpn   GGTAC•C
A = Apa   GGGCC•C
N = Nar   GG•CGCC

Scale: 00.0 — 10.0 — 20.0 — 30.0 — 40.0 — 50.0

CvnI ————————————————————

CvnI/KpnI ————————————————————

KpnI ————————————————————

KpnI/ApaI ————————————————————

ApaI ————————————————————

ApaI/NarI ————————————————————

NarI ————————————————————

CvnI/NarI ————————————————————

Overall Map ————————————————————
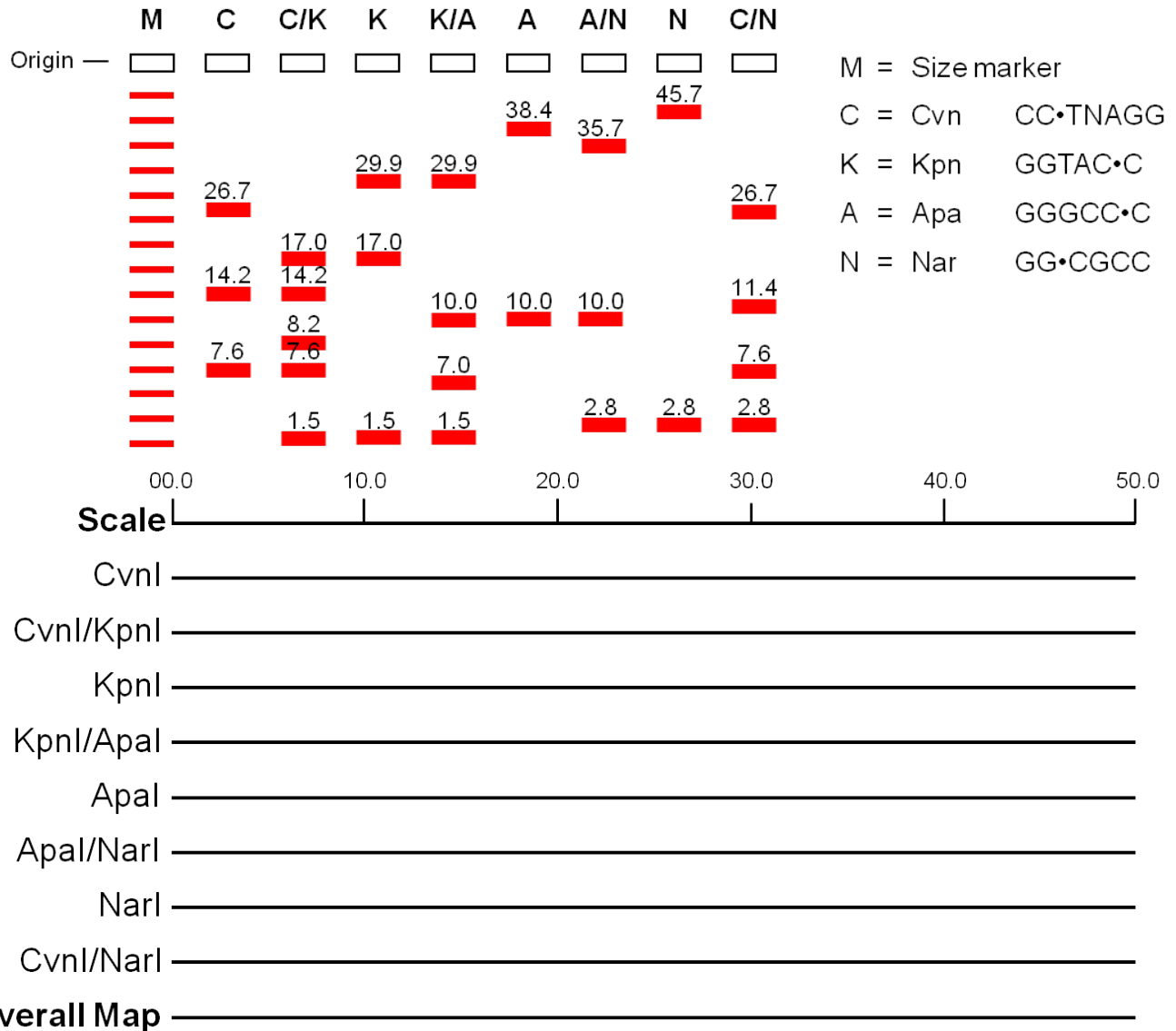
TIPS ON SOLVING RESTRICTION MAPPING QUESTIONS

1) Start with the Nar and Apa digests, each has only one site. This will help get a simple starting map.
2) Next, try and add in the Cvn sites using the double digests with Nar and Cvn.
3) Next, try and add in the Kpn sites using the double digests with Apa and Cvn.
4) There is no formal method to solve these maps. Instead, think of them like a jigsaw puzzle (only linear) and use trial and error to solve the puzzle.
5) Use the class notes for help.

Recommended course:

Independent Studies in Genetic Counselling

BIOL 490 – Individual Study
Independent Studies in Genetic Counselling
★ 3 (fi 6) (either term, 3-0-0) To provide undergraduate students with insight into the role of a genetic counsellor through observation of genetic counselling sessions and researching key topics in the area. Enrollment is limited and is by permission of the instructor(s). There is a limit of 2 students/term.
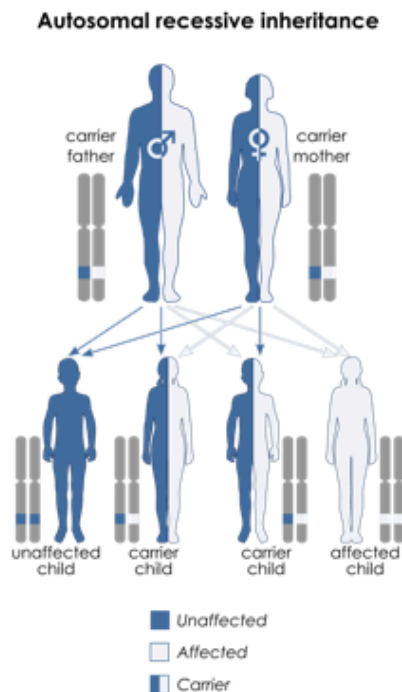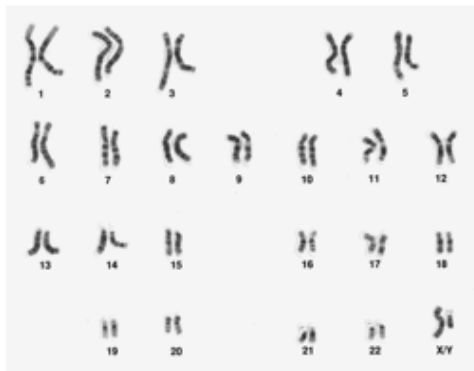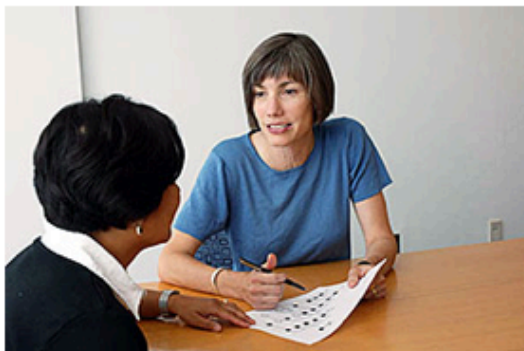
Recommended preparation: Intro stats, psych courses, molecular genetics techniques course, population genetics, one-on-one counselling experience

Offered: Fall and Winter term, Time arranged each term

Contact: Dr. Heather McDermid, hmcdermi@ualberta.ca or Mr. Sajid Merchant, sajidm@ualberta.ca

Website: http://geneticcounselling.biology.ualberta.ca

CHAPTER 28 – RESTRICTION MAPPING AND
GEL ELECTROPHORESIS

Figure 1.
Restriction enzymes that are available on a vending machine. (Flickr- Jun Seita CC BY-NC 2.0)

INTRODUCTION

Molecular Genetics techniques involve the isolation, purification, and manipulation of DNA. DNA can come in the form of genomic DNA, plasmids, or oligonucleotides.

1. ISOLATING 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 **(1)** grinding or **lysing** in a solution that contains chemicals that protect the DNA while disrupting other components of the cell (**Figure 2**). These chemicals may include **detergents**, which dissolve lipid membranes and denature proteins. A cation such as Na$^+$ helps to stabilize the negatively charged DNA and separate it from proteins such as histones. **(2)** 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. **(3)** Commonly, proteins are removed by adjusting the salt concentration so they precipitate. **(4)** The **supernatant**, which contains DNA and other, smaller metabolites, is then mixed with ethanol, which causes the DNA to precipitate. **(5)** A small **pellet** of DNA can be collected by centrifugation, and **(6)** 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.
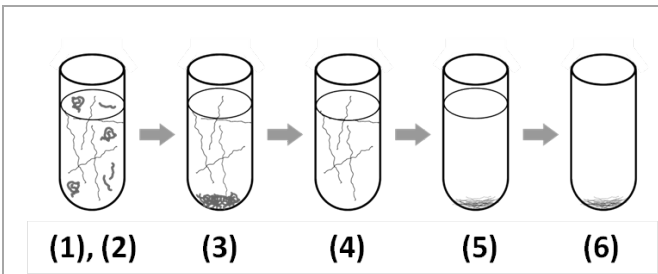
Figure 2.
Extraction of DNA from a mixture of solubilized cellular components by successive precipitations. Proteins are precipitated, then DNA (in the supernatant) is precipitated in ethanol, leaving a pellet of DNA.
(Original-Deyholos-CC BY-NC 3.0)

Figure 3.
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)

2. RESTRICTION ENZYMES AND DNA METHYLATION

2.1. RESTRICTION ENZYMES

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 via hydrolysis of phosphodiester backbone 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.

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.
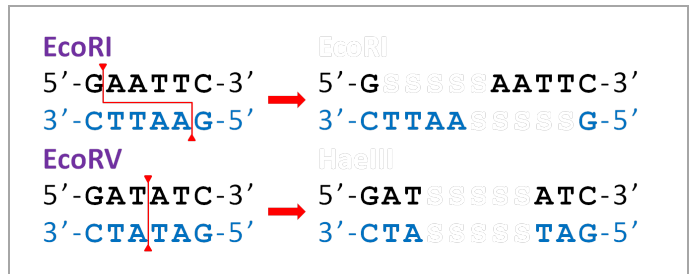
Figure 4.
Each recognition sequences are cleaved by restriction enzyme. It can either cut DNA at a position offset from the center of the restriction site creating an overhanging region or it can cut directly in the middle to create blunt end.
(Original-Deyholos/Kang-CC BY-NC 3.0)

Figure 5.
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 BY-NC 3.0)

For example, *EcoRI* (**Figure 3**) and *EcoRV* are both enzymes from *E. coli*. *EcoRI (pronounced eco-r-1)* cuts double stranded DNA at the recognition sequence 5'-GAATTC-3', but note that this enzyme, like many others, does not cut in exactly the middle of the restriction sequence (**Figure 4**). The ends of a molecule cut by *EcoRI* have an overhanging region of single stranded DNA, and so are sometimes called **sticky-ends**. These "sticky ends" are short stretch of complementary base pairs that anneal together and aid in the formation of recombinant molecules. On the other hand, *EcoRV* is an example of an enzyme that cuts both strands in exactly the middle of its recognition sequence 5'-

GATATC-3', producing what are called **blunt-ends**, which lack overhangs: 5'-GAT   ATC-3'

Many different kinds of restriction sites exist in genome, and it takes time for the enzymes to cut up all of the restriction sites. If the concentration of the enzyme and the exposure time is low, this will result in partial digest and produce longer fragments. If the DNA is exposed to restriction enzyme long enough, this will result in complete digest and produce shorter fragments.

2.2. DNA METHYLATION
Bacteria keep their DNA safe from their own restriction enzymes by methylating (adding a CH3 group) using methyl transferase (methylase) enzyme. For each different restriction enzyme, its matching methylase enzymes are produced to methylate the host DNA. After each replication of DNA, the enzyme has to methylate the newly synthesized DNA.

3. DNA LIGATION

The process of **DNA ligation** occurs when DNA strands are covalently joined, end-to-end forming a phosphodiester bond between the 5' phosphate end and 3' hydroxyl end through the action of an enzyme called **DNA ligase**. Typically, sticky-ended molecules with complementary overhanging sequences (**compatible ends**) 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. The sticky-ended molecules with non-complementary sequences will not ligate together with DNA ligase. This function of joining two fragments of DNA pieces together by DNA ligase is essential when connecting Okazaki fragments during DNA replication, or repairing breaks in either single or double stranded DNA molecules during recombination. Therefore, if a mutation occurs in the genes that encode for ligase enzymes, the result will impact the organism immensely in a negative way. In molecular genetics, ligation is particularly important as DNA ligase facilitates the insertion of double stranded DNA fragment into a

plasmid vector: Ligation is therefore central to the production of recombinant DNA.

4. AGAROSE GEL ELECTROPHORESIS

4.1. BASICS
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 6**). 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 molecules that compose the gel are linked by hydrogen bonds not covalent bonds, so the experimenter can mold the shape of the gel by heating and cooling. The DNA is then forced through the gel by an electrical current, with DNA molecules moving toward the positive electrode (**Figure 7**). This is because the phosphate backbone of DNA or RNA has negative charge on it. Therefore, rather than moving in a vertical manner, the DNA or RNA molecule will move by its horizontal side.
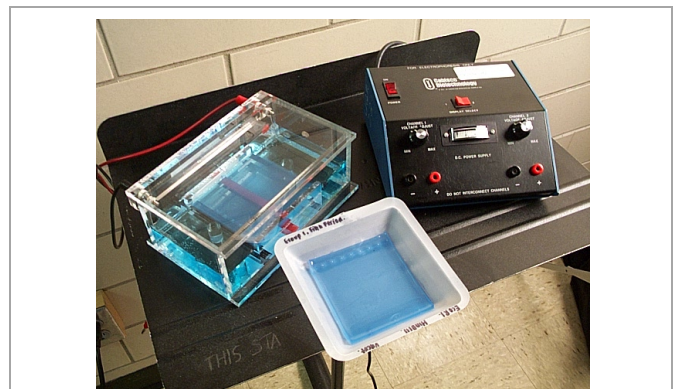
Figure 6.
Apparatus for agarose gel electrophoresis. A waterproof tank is used to pass current through a slab gel, which is submerged in a buffer in the tank. The current is supplied by an adjustable power supply. A gel (stained blue by a dye sometimes used when loading DNA on the gel) sits in a tray, awaiting further analysis, such as photography under a UV light source.
(Flickr- camerazn - CC BY 2.0)
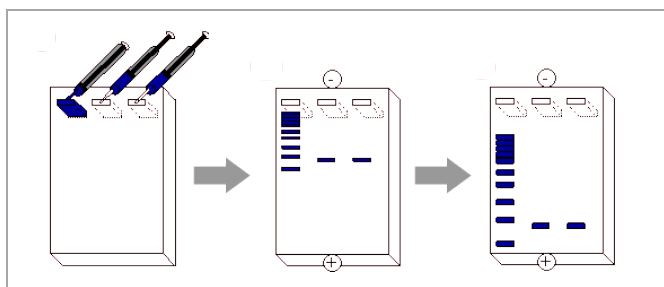
Figure 7.
 Agarose gel electrophoresis. DNA is loaded into wells at the top of a gel. A current is passed through the gel, pulling DNA towards the positively charged electrode. The DNA fragments are separated by size, with smaller fragments moving fastest towards the electrode. The left lane is a series of known size marker fragments, while the centre and right lanes contain fragments whose size can be approximated in comparison to the known marker sizes. (Wikipedia-Magnus Manske-PD)
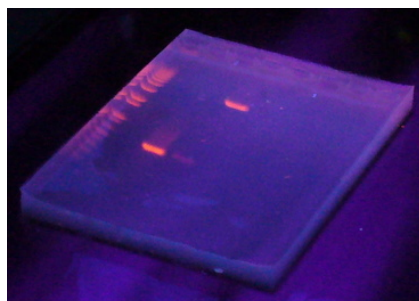
Figure 8.
A photograph of an agarose gel stained with ethidium bromide and illuminated by UV light. The stain associated with DNA is fluorescent orange. Each band represents a different sized cloned fragment. As in the previous figure, the size of the unknown fragments can be determined by comparison to the known marker fragments on the left. (Wikipedia-Transcontrol- CC BY-SA 3.0)
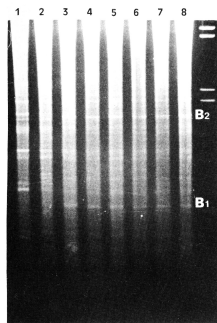
Figure 9.
Genomic DNA of Rice on agarose gel produces a smear of DNA that represents the fragments in a whole genome, where as the cloned DNA on the very right lane has sharp bands that are separated by distance.
(Flickr-IRRI Photos-CC BY-NC-SA 2.0)

As it migrates, each piece of DNA threads its way through the pores, which form between the polymers in the gel. Note that the mobility of the molecules is affected by the molecular weight, which with a linear molecule like DNA is primarily length. Because shorter pieces can move through these pores faster than longer pieces, gel electrophoresis separates molecules based on their size (length), with shorter DNA pieces moving faster than longer ones. Circular DNA molecules like plasmids move according to their conformation. Open circular will dsDNA will move slower than super-coiled dsDNA due to size difference and vice versa. DNA molecules of a similar size migrate at a similar rate and thus will arrive at a similar location in each gel, called a **band**. This feature makes it easy to see specific sized DNA after staining with a fluorescent dye, such as **ethidium bromide (EtBr)** that   acts as intercalating agent that is inserted between two bases (**Figure 8**). 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.

4.2.   GENOMIC VS CLONED DNA ON A GEL
Samples of cloned DNA will have few bands, while genomic DNA will appear as a smear of DNA that represents the many sized fragments present in a whole genome. Since the fragments that are run on the gel represent the whole genome, there would be minimal gaps in between in terms of the length of the various DNA fragments and therefore appear as a continuous band on the gel.

5.   OTHER APPLICATIONS OF GEL ELECTROPHORESIS

5.1.   SEPARATION OF RNA BY GEL ELECTROPHORESIS.
Like DNA, samples of RNA can be separated on a gel. The major difference between RNA and DNA is that RNA is single stranded and DNA is double stranded.   Therefore,   RNA   molecules   are susceptible to intramolecular base paring which

produces a secondary structure that forms loops. This can affect its mobility and can therefore provide wrong information regarding its size. Also, the bands are going to look less sharp. Therefore, denaturing agents have to be used when running on the gel in order to break the hydrogen bond that is holding the secondary structure of RNA. This way, most of the secondary structure of RNA can be prevented. Some of the molecules still might have its secondary structure and some even might re-coil back. This is why EtBr, which is an intercalating agent that is inserted between two planar bases, still works on RNA molecules but with a lower efficiency. Therefore, in order to produce similar quality with DNA on a gel, a lot more RNA molecules have to be used.

RNA gel electrophoresis can be used when scientists want to identify the existence of certain mRNAs and therefore certain genes that are expressed in the cell compared to other cell types or the same cell type but in a different stage in life. Also, RNA molecules can be quantified so the level of gene expression can be identified as well. Finally, it can be purified just like DNA molecules.

5.2. CONTAMINATION IN GEL ELECTROPHORESIS
DNA samples don't always separate correctly in agarose gel electrophoresis. Typically, the DNA sample is contaminated with other macro-molecules or chemicals. **Figure 10** shows the consequences of various forms of contamination.

6. RESTRICTION MAPPING

6.1. PROCEDURE
Restriction mapping is the technique of identifying the location of restriction sites, relative to other sites, on a DNA molecule. Typically, a sample of purified plasmid DNA is aliquoted into several tubes and each is treated with several different restriction enzymes or combination of enzymes. These are then separated by agarose gel electrophoresis and their restriction fragment sizes determined. By trial and error, the combination of fragments can be assembled like a linear jigsaw puzzle into a map of the restrictions sites – a **restriction map**.

Lane	Appearance	Salvageable?
1. a good DNA prep	---	---
2. RNA contamination	RNA molecules run faster than the DNA	yes, add RNase to the prep
3. salt contamination	DNA runs slower when coated in positive ions	yes, redo the last step in the prep
4. protein contamination	proteins remain in the wells and make EtBr or RedSafe glow	yes, redo the last couple of steps in the prep
5. DNA fragmentation	DNA fragments less than 20 kb long run faster	no, toss it
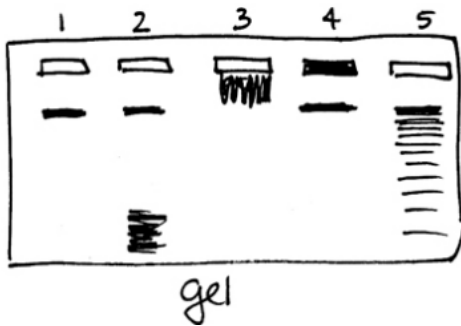
Figure 10.
The consequences of various forms of contamination on the separation of DNA in an agarose gel.
(Original-Harrington-CC BY-NC 3.0)
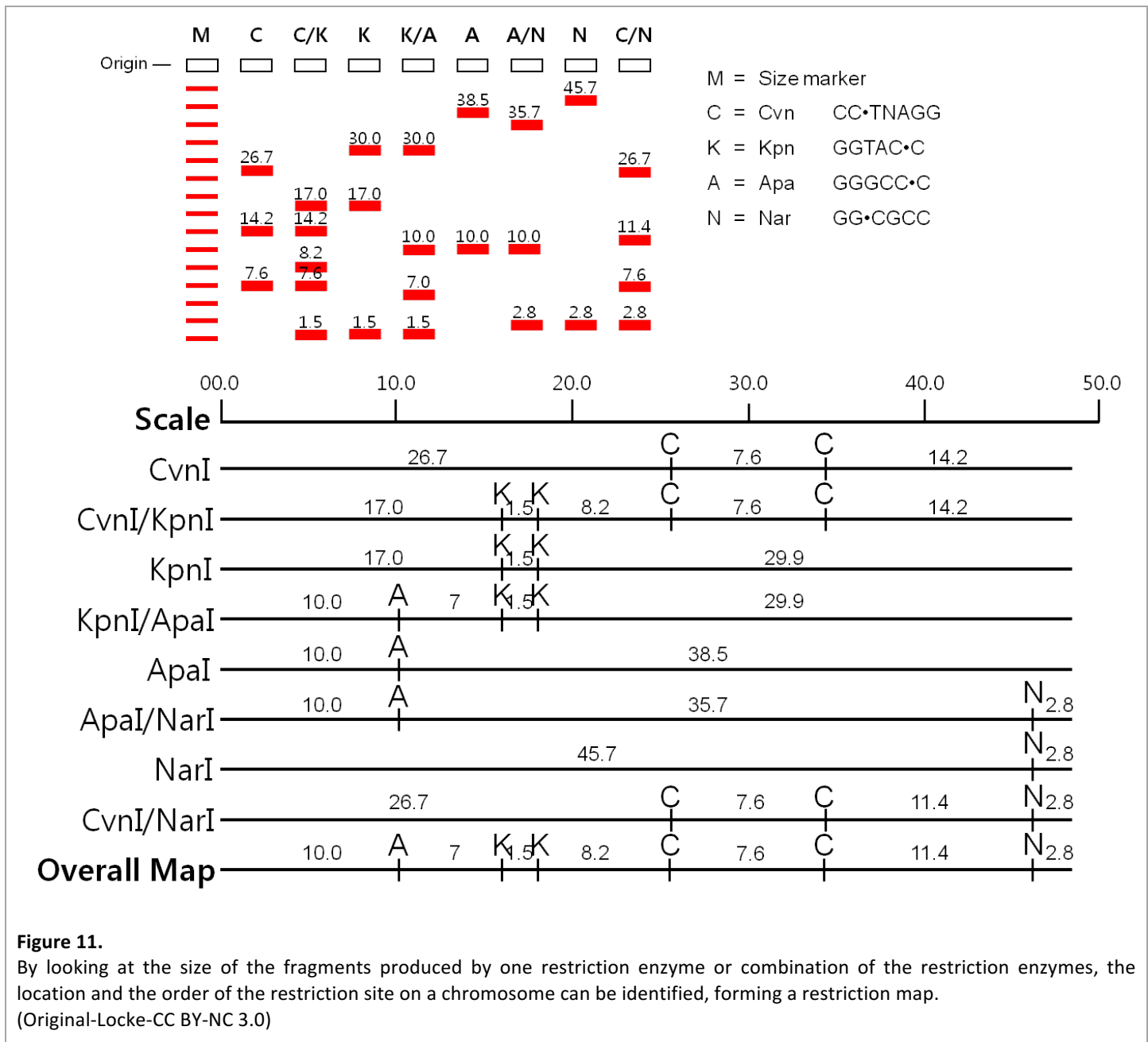
Figure 11.
By looking at the size of the fragments produced by one restriction enzyme or combination of the restriction enzymes, the location and the order of the restriction site on a chromosome can be identified, forming a restriction map.
(Original-Locke-CC BY-NC 3.0)

6.2. USES

Restriction mapping is a quick, easy and inexpensive way to characterize and distinguish DNA samples without actually sequencing the DNA; sequences can be represented by series of restriction sites and using this knowledge, one can tell if the DNA of interest is similar or different from others by comparing their degree of overlaps.

Also, restriction sites offer positions for convenient manipulation of the DNA. Restriction fragments that contain the gene of interest can be cut out and once the gene is purified from the fragments, it can be sequenced or used as a probe. This is the reason why restriction mapping is still routinely used today, even though sequencing technologies allows us to sequence the whole genome.

SUMMARY:

- 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.

- DNA or RNA molecules can be identified, quantified, and separated on electrophoresis gel.

- Contamination in DNA samples such as RNA, salt, or protein that can affect the bandings on a electrophoresis gel.

KEY TERMS:

lysing
detergents
chelating agent
EDTA
nucleases
supernatant
pellet
restriction endonucleases
restriction enzymes
EcoRI
EcoRV
sticky-ends

blunt-ends
DNA methylation
DNA ligation
DNA ligase
compatible ends
gel electrophoresis
agarose
band
ethidium bromide
size markers
restriction map

STUDY QUESTIONS:

1) A 6.0 kbp PCR fragment flanked by recognition sites for the *HindIII* restriction enzyme is cut with *HindIII* then ligated into 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 band(s) observed?

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

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

2) 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.

3) Which would move faster during agarose gel electrophoresis, a 1.0 kbp duplex DNA molecule or a 1,000 nt of RNA (single stranded) molecule?

CHAPTER 29 – RECOMBINANT DNA

Figure 1.
The manipulation of DNA often involves small quantities of liquids that are accurately dispensed using micro-pipettors into small plastic microfuge tubes. Volumes as small as 0.5 μl are routinely dispensed for some reactions. The use of clean, sterile plastic tubes and tips ensures the reactions work correctly and are reproducible.
(Flickr- University of Michigan School of Natural Resources and Environment-CC BY 2.0)

INTRODUCTION

Recombinant DNA is a general term to describe DNA that has been manipulated (recombined) somehow *in vivo*. It typically involves the breakage of DNA into fragments, using restriction enzymes, and the rejoining (ligation) of these fragments into various arrangements and into vectors, such as plasmids, to propagate the new arrangement for further analysis, like sequencing, or for insertion into other hosts, such as model organism as transgenes.

1. BASIC TERMINOLOGY

Before proceeding any further, there are some basic terminologies that students should know regarding recombinant DNA technology.

in vivo **(in life)** experiments done within a living cell/organism

in situ **(in place)** experiments done on cells and structures removed intact from an organism. (ex. Inserting RNA into a frog egg cell on a petri dish)

in vitro **(in glass)** experiments done on individual molecules removed from an organism (ex. DNA in a test tube). These days most experiments are done *in plastico* **(in plastic).** See **Figure 1**.

in silico **(in silicon)** Experiments done within a computer simulation.

Recombinant DNA: a composite DNA molecule created *in vitro* by joining a foreign DNA with a vector DNA molecule. (Note; technically recombinant DNA can be also made *in vivo* during meiosis in an organism, but this is usually not the typical meaning of these words.)

2. RECOMBINANT DNA TECHNIQUES:

There are many techniques for joining DNA molecules *in vitro* and introducing them into cells (usually bacteria) where the molecules are then replicated along with the host genomic DNA.

2.1. PLASMIDS ARE NATURALLY PRESENT IN SOME BACTERIA

Many bacteria contain extra-chromosomal DNA elements called **plasmids**. These are usually small (a few 1000 bp), circular, double stranded molecules that replicate independently of the chromosome and can be present in multiple copies within a cell. In the wild, plasmids can be

transferred between individuals during bacterial mating and are sometimes even transferred between different species. Plasmids are particularly important in medicine because they often carry genes for pathogenicity (making the bacteria more detrimental) and drug-resistance (able to survive various antibiotics). In the lab, plasmids are inserted into bacterial hosts in a process called **transformation**. These plasmids can be modified by the addition of foreign DNA so that both the plasmid vector and the target foreign DNA is replicated.

There are 3 main features of a plasmid (**Figure 2**):
(1) **Origin of replication (Ori)** which is similar in function to *oriC* in *E. coli* chromosome.
(2) **Selectable marker** gene that helps to screen the desired and undesired strains, which is usually an antibiotic resistance gene like amp^R, tet^R, or kan^R. Some cells have plasmids that contain resistance to multiple, different antibiotics.
(3) **Multiple cloning site (MCS)** that has many restriction enzyme sites in a short sequence.

Figure 2.
Basic structure of a double stranded bacterial plasmid, represented by a circle.
(Original-Locke-CC BY-NC 3.0)
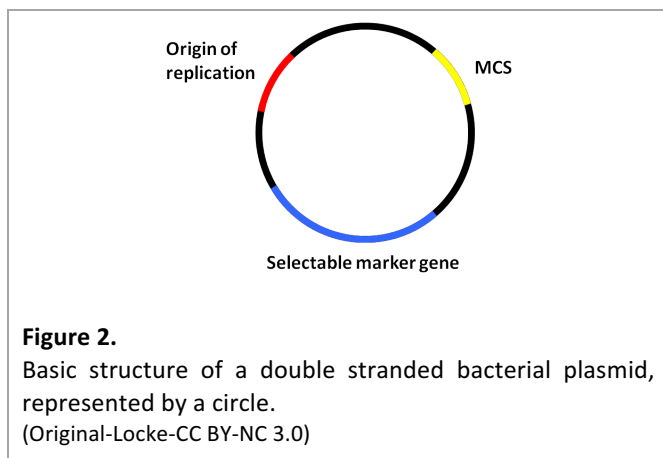
3. USING CLONING VECTORS

3.1. PLASMIDS VECTORS

There are multiple steps to using plasmids as cloning vectors. To insert a DNA fragment into a plasmid, both the fragment and the circular plasmid are cut using a restriction enzyme that produces compatible ends. Given the large number of restriction enzymes that are currently available, it is usually not too difficult to find an enzyme for which corresponding recognition sequences are present in both the plasmid and the DNA fragment,

particularly because most plasmid vectors used in molecular biology have been engineered to contain recognition sites for a large number of restriction endonucleases in a segment called the Multiple Cloning Site (MCS).

Figure 3.
Cloning of a DNA fragment (red) into a plasmid vector. The vector already contains a selectable marker gene (blue) such as an antibiotic resistance gene.
(Original-Deyholos-CC BY-NC 3.0)
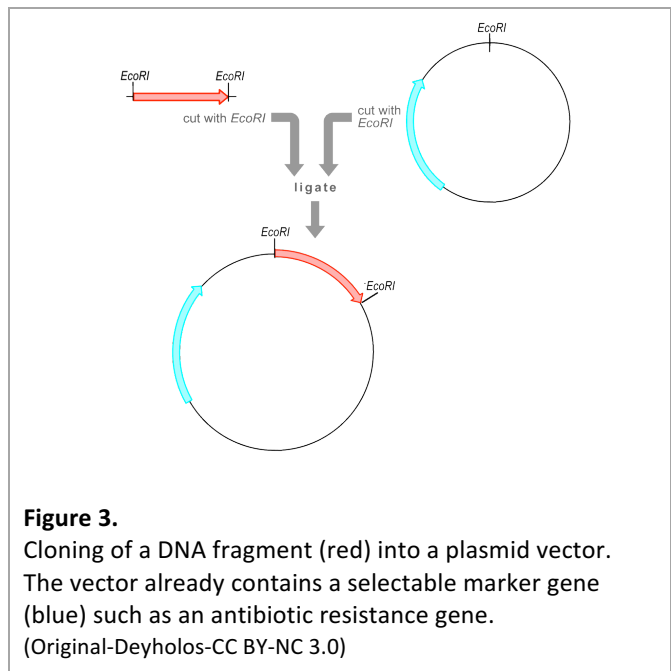
After restriction digestion, the desired fragments may be further purified or selected before they are mixed together with ligase to join them together. Following a short incubation, the newly ligated plasmids, containing the gene of interest are **transformed** into *E. coli*.

Transformation is accomplished by mixing the ligated DNA with *E. coli* cells that have been specially prepared (i.e. made **competent**) to uptake DNA. Bacterial cells can be made competent by exposure to compounds such as $CaCl_2$ or to electrical fields (**electroporation**). Because only a small fraction of cells that are mixed with DNA will actually be transformed, a **selectable marker**, such as a gene for antibiotic resistance, is usually also present on the plasmid and used to select those few cells that have taken up the DNA. The rate of DNA uptake varies each time and is called **transformation efficiency**. This can range from ~10^5-10^{10} colonies per µg of DNA.

After transformation (combining DNA with competent cells), bacteria are spread on a bacterial agar plate containing an appropriate antibiotic so

that only those cells that have actually incorporated the plasmid will be able to grow and form colonies. Colonies (clone) can then be picked and used for further study.

Molecular biologists use plasmids as **vectors** to contain, amplify, transfer, and sometimes express genes of interest that are present in the cloned DNA. Often, the first step in a molecular biology experiment is to "**clone** a gene" (i.e. make a copy) into a plasmid, then transform this recombinant plasmid into bacteria so that essentially unlimited copies of the gene (and the plasmid that carries it) can be made as the bacteria reproduce. This is a practical necessity for further manipulations of the DNA, since most techniques of molecular biology require many copies of DNA to work. Even though small amounts are needed they are not sensitive enough to work with just a single molecule at a time.

Many molecular cloning and recombination experiments are therefore iterative (repetitive) processes. For example:

1. a DNA fragment (usually isolated by PCR and/or restriction enzyme digestion) is cloned into a plasmid cut with a compatible restriction enzyme

2. the recombinant plasmid is transformed into bacteria

3. the bacteria are allowed to multiply, usually in liquid culture

4. a large quantity of the recombinant plasmid DNA is isolated from the bacterial culture

5. further manipulations (such as site directed mutagenesis or the introduction of another piece of DNA) are conducted on the recombinant plasmid

6. the modified plasmid is again transformed into bacteria, prior to further manipulations, or for expression

3.2. OTHER VECTORS
Lambda phage is a bacteriophage that infects *E. coli* and can be used as a vector. Lambda phage is a linear DNA vector molecule that can typically hold a 15-20 kb fragment in each clone.

Cosmids are a hybrid vector system composed of part plasmid and part phage DNA. It can clone 30-45Kb fragments in each clone. The **lambda phage packaging system (**stuffs the recombinant DNA into the lambda bacteriophage heads) is used for higher transformation efficiency, but it also has the plasmid origin of replication so clones can be replicated in the host like plasmids.

BACs (Bacterial Artificial Chromosomes) is a circular DNA vector that uses a plasmid origin of replication to propagate. The insert DNA can be very large, 100's of kbp, so it may contain many genes. But, such large recombinant DNA molecules are difficult to transform so BACs are difficult to make.

4. DNA LIGATION

The process of **DNA ligation** occurs when DNA strands are covalently joined, end-to-end through the action of an enzyme called **DNA ligase**. Molecules with complementary overhanging sequences are said to have "**sticky**" or **compatible ends**, which facilitate their joining to form recombinant DNA. Likewise, two blunt-ended sequences are also considered compatible to join together, although they do not ligate together as efficiently as sticky-ends. Note: sticky-ended molecules with non-complementary sequences will not ligate together with DNA ligase.

The process of ligation is central to the production of recombinant DNA, including the insertion of a double stranded DNA fragment into a plasmid vector.

5. AN APPLICATION OF MOLECULAR CLONING: RECOMBINANT INSULIN

Purified insulin protein is critical to the treatment of diabetes. Prior to ~1980, insulin for clinical use was isolated from human cadavers or from slaughtered animals such as pigs. Human-derived insulin generally had better pharmacological properties, but was in limited supply and carried risks of disease transmission. By cloning the human insulin gene and expressing it in *E. coli*, large

quantities of the insulin protein and identical to the human hormone sequence could be produced in fermenters, safely and efficiently. Production of recombinant insulin also allows specialized variants of the protein to be produced: for example, by changing a few amino acids, longer-acting forms of the hormone can be made. The active insulin hormone contains two peptide fragments of 21 and 30 amino acids, respectively. Today, essentially all insulin is produced from recombinant sources (**Figure 4**), i.e. human genes and their derivatives expressed in bacteria or yeast.

Figure 4.
A vial of insulin. Note that the label lists the origin as "rDNA", which stands for recombinant DNA.
(Flickr-DeathByBokeh- CC BY-NC 2.0)

6. GENOMIC DNA LIBRARIES AND CDNA LIBRARIES

6.1. GENOMIC DNA LIBRARY

The human genome is large and complex. It is much easier to break it down into little fragments to study. This is true for all organisms – deal with a gene at a time.

Here is the process of constructing a genomic DNA library (**Figure 5**):

1. Genomic DNA is broken down into short fragments by partial restriction enzyme digestion. The size is dictated by the vector used. Plasmids will need short fragments (~5 kbp) while cosmid vectors will need larger ones (30-45 kbp). .

2. Circular plasmid or cosmid vector DNA is opened with the same restriction enzymes that were used in (1) or another enzyme that yields compatible, sticky ends.

3. The DNA fragment and vector are mixed together in the same test tube and ligated together. (The ligation occurs by random chance so not all molecules will be appropriate

for the next step.) Each independent assembly of a DNA segment in a vector is a **clone**

4. The recombinant DNA molecules are transformed into a competent bacteria host cell. For plasmids, this is a direct process, while for cosmids, a lambda in vitro packaging system is used to increase the efficiency of the process. (Lambda *in vitro* packaging system refers to the packaging of recombinant DNA into the head of the bacteriophage, and then transferring this package to the host cell.)

5. The transformed cells that contain plasmid or cosmid with antibiotic resistance gene are selected and propagated.

6. Amplified recombinant plasmid DNA molecules can be purified, and collected. This collection of many different clones (genomic DNA fragments) makes up a **DNA library** or clone library that can contain the entire DNA sequence of an organism in the fragmented form of multiple clones.

These clones can then be stored and the fragment of interest can be retrieved at a later time, hence the name "genomic library." Gene libraries can be constructed using different vectors, but almost all work is done with plasmids these days.

How many clones are needed to include every sequence in a library? The number of clones needed to have 1 genome equivalent can be calculated by dividing the number of sequences of the genome by the number of sequences of the clone. For example, if the *E. coli* genome is about 4,500,000 bp and cosmid clone contains 45,000 bp, then 4,500,000/45,000 = 100 clones would be needed, end to end, to cover the whole genome. However, in real life, some sequences in the genome might not be cloned at all (others may be cloned more than once) and therefore the process is not 100% efficient. To get a 99% chance of finding a specific gene or sequence of interest, one needs about 5 genome equivalents. That is: 500 clones for the *E. coli* example above.

Figure 5.
Process of making a genomic clone library of an organism.
(Original-Locke-CC BY-NC 3.0)
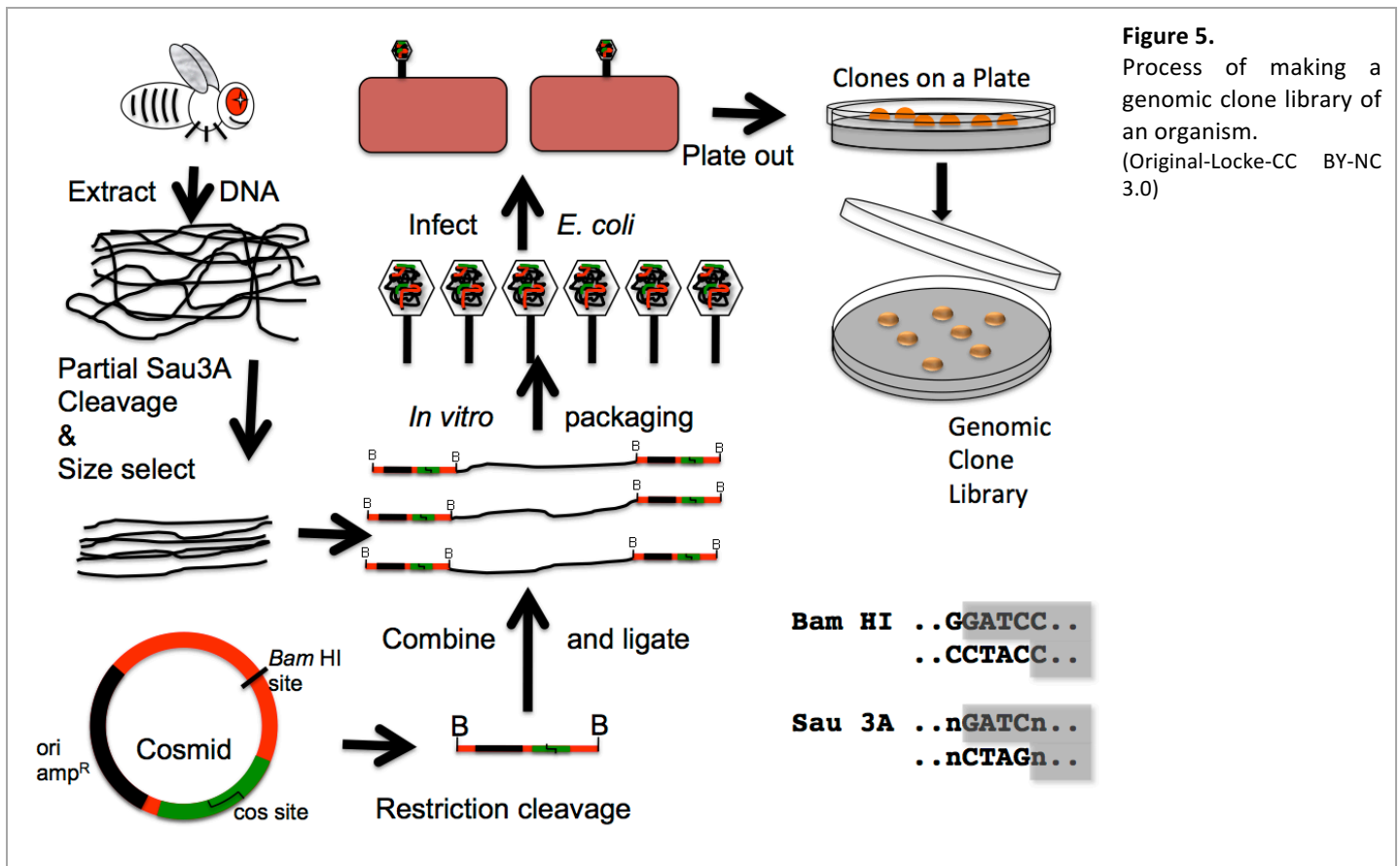
6.2. cDNA LIBRARY

Genomic library above hypothetically contains all the sequences of the target's DNA, but a **cDNA** library only contains the sequences that are expressed in a particular cell or tissue.

To create a cDNA library, RNA is collected from the cell or tissue of interest. Primers, nucleotides, and RNA transcriptase enzyme are added so that complimentary DNA, or cDNA that is complimentary to the RNA is synthesized. The result is a RNA-cDNA hybrid, and these two strands are separated by adding heat, and RNA can be denatured by adding RNase enzyme or NaOH. The remaining cDNA can act as a template and its complementary DNA strand is synthesized, each forming a double helix. From this point, the rest of the procedure to create a library is equivalent with that above. Here, however, the cloned DNA

corresponds to the mRNA present in the cell. The DNA between genes, and intron DNA, is absent from this library. These clones are often used to express the gene to make a protein.

The main difference between Genomic DNA library and cDNA libray is that the genomic library contains DNA with exons, introns, and intergenic sequences, so the number of different clones in the library is much bigger. On the other hand, cDNA library contains only the sequences present after transcription and processing (e.g. splicing exons), which are translated into polypeptides. Therefore, by looking at the cDNA library we can identify which genes are expressed in particular cell types, and to what level of expression, too.

cDNA Library

Cell, tissue, organism

oligo-dT primer

5' ⊥⊥⊥⊥⊥⊥⊥⊥⊥⊥⊥⊥ AAAAAAA–3'
3' –TTTTTTT–5'

Reverse transcriptase

5' ⊥⊥⊥⊥⊥⊥⊥⊥⊥⊥⊥⊥ AAAAAAA–3'
3' ⊥⊥⊥⊥⊥⊥⊥⊥⊥⊥⊥⊥ TTTTTTT–5'

Remove RNA

3'
⊥⊥⊥⊥⊥⊥⊥⊥⊥⊥⊥⊥ TTTTTTT–5'

DNA polymerase

⊥⊥⊥⊥⊥⊥⊥⊥⊥⊥⊥⊥ 3'
⊥⊥⊥⊥⊥⊥⊥⊥⊥⊥⊥⊥ TTTTTTT–5'

S1 nuclease

5' ⊥⊥⊥⊥⊥⊥⊥⊥⊥⊥⊥⊥ 3'
3' ⊥⊥⊥⊥⊥⊥⊥⊥⊥⊥⊥⊥ 5'

Clone DNA into vector

mRNA

Plasmid vector

Library of clones representing all the mRNAs in the sample

Transform *E.coli*

○ +
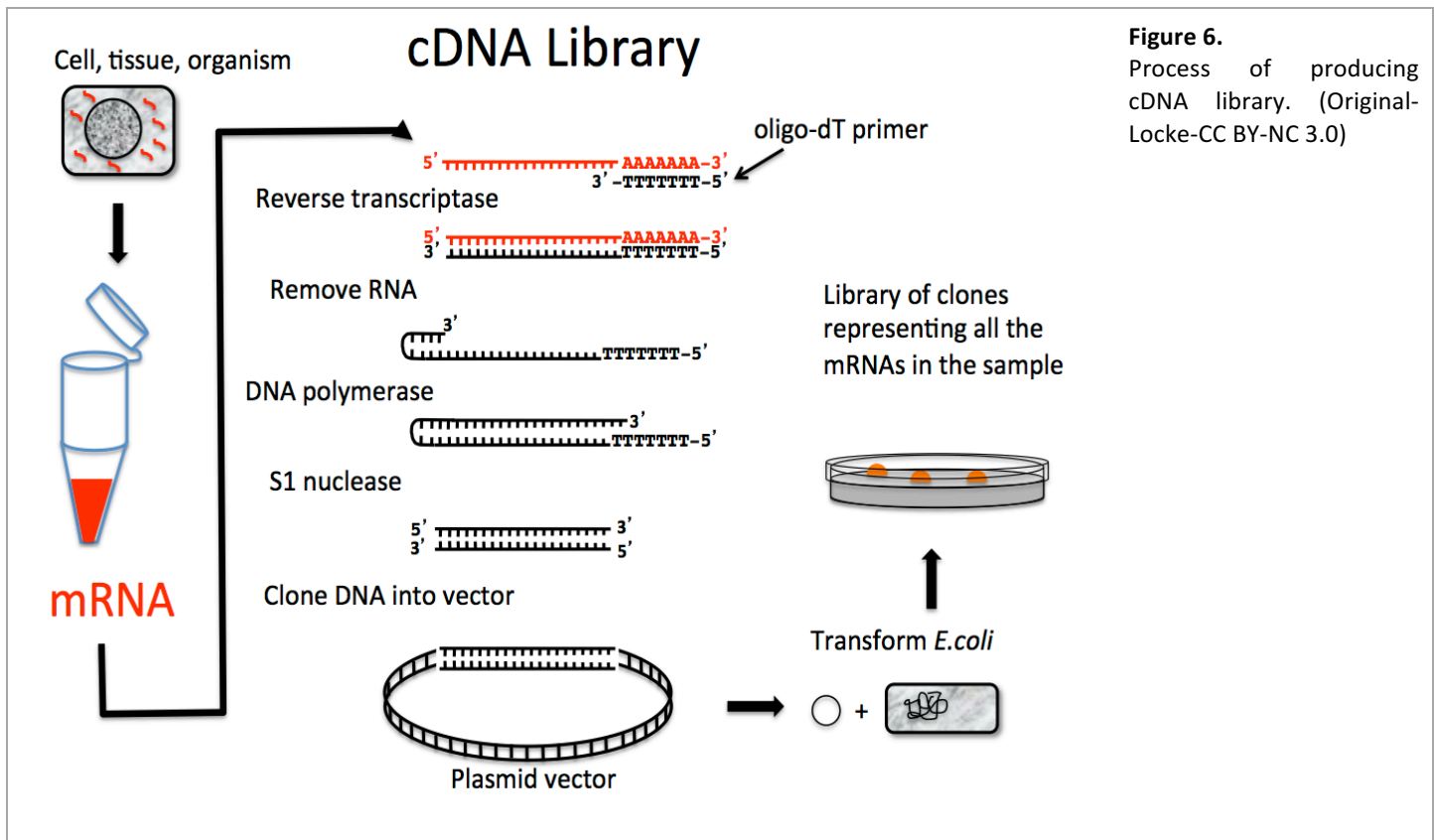
Figure 6.
Process of producing cDNA library. (Original-Locke-CC BY-NC 3.0)

7.   SCREENING A CLONE LIBRARY

After genomic or cDNA libraries have been constructed, clones containing a particular gene, or DNA sequence, can be identified and recovered using the process of hybridization and labeled DNA probes. DNA labeling involves putting a tag on the DNA molecule that is going to be complementary to the DNA sequence of interest, in some manner that permits one to detect its presence in minute quantities at some later point in an experiment. DNA can be labeled in several different ways; one widely used technique is to replace the normal Phosphorous of the DNA with a radioactive atom of Phosphorous, ^{32}P (normal isotope = ^{31}P). This radioactivity can be detected by photographic emulsion. A cloned DNA sequence will hybridize to only its complementary sequences and thus provides an almost unique probe. Once the appropriate probes are made, the following procedures are performed:

1. Plate out library - each colony on the bacterial plate is a clone.

2. Lift clones (DNA) onto Nitrocellulose filter. Lyse the cells and fix the clone DNA onto the filter. Denature the clone DNA, so as to make it able to form hybrids with probe.

3. Place filter in a hybridization bag with solution containing labeled, denatured probe DNA. Incubate to permit the probe strands to form hybrids with the clone strands.

4. Wash away unhybridized probe.

5. Expose probed filter to X-ray film (autoradiography) to detect the presence of clones with labeled probe.

6. From the X-ray film determine which clone hybridized to the probe and recover that clone for further analysis.

Screening a clone library

Clones on a Plate

Nitrocellulose membrane

Clones DNA on membrane

Probe hybridized to clone DNA

Radioactive Probe = ⸗

wash

Expose
X-ray film

Developed
X-ray film

Pick clone
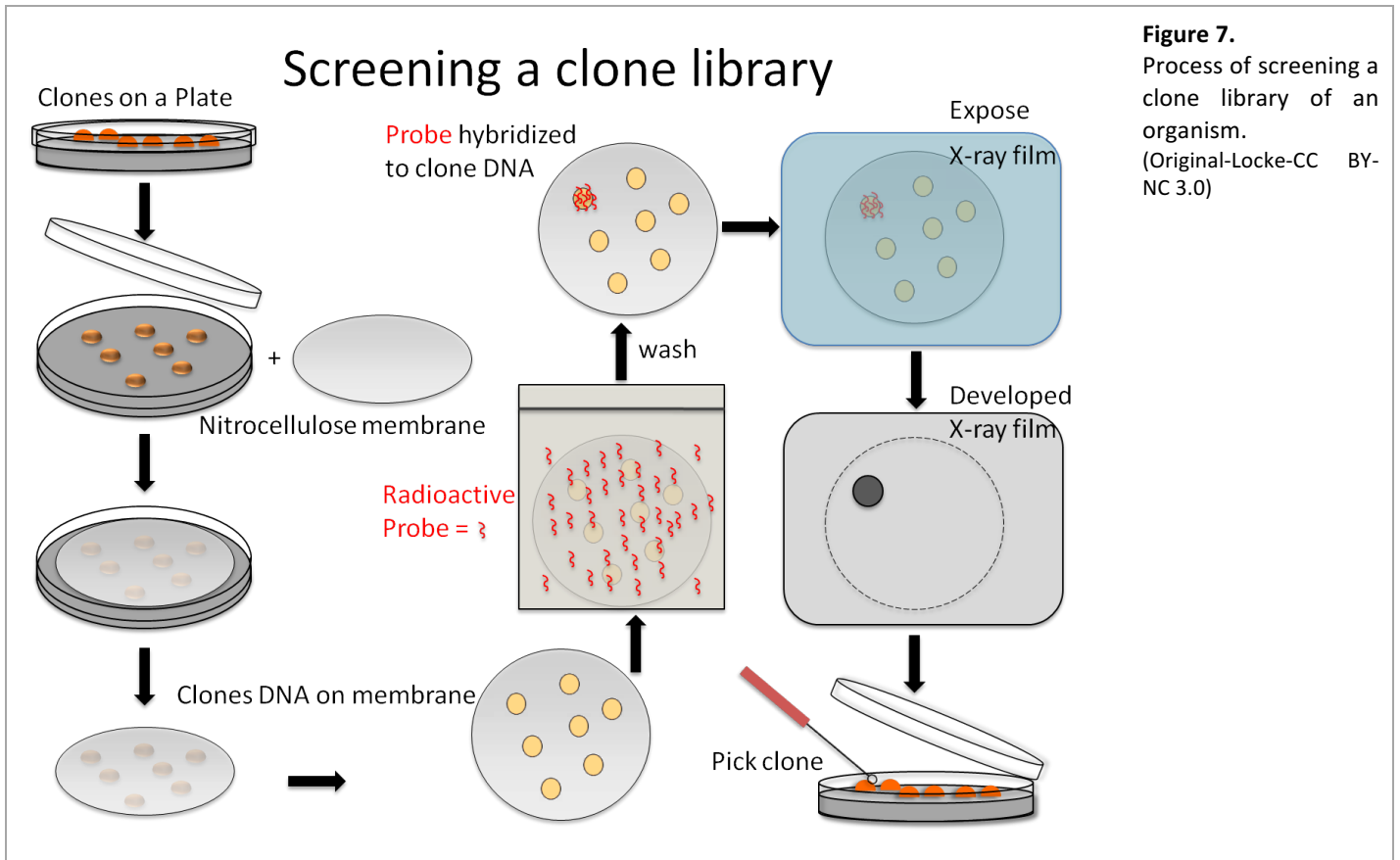
Figure 7.
Process of screening a clone library of an organism.
(Original-Locke-CC BY-NC 3.0)

_____

SUMMARY:

- DNA fragments can be cloned into vectors.

- Transformation of recombinant DNA is the transfer of DNA (usually recombinant plasmids) into bacteria.

- Cloning of genes into E. coli is a common technique that allows large quantities of a DNA for gene to made

- This allows further analysis or manipulation of the cloned sequences.

- Genomic DNA libraries contain fragments of genomic DNA.

- cDNA libraries contain shorter segments of DNA that correspond to the mRNA for each gene.

- Gene of interest can be identified using DNA probes to screen genomic or cDNA libraries.

- Cloning can also be used to produce useful proteins, such as insulin, in microbes.

KEY TERMS:

in vivo

in situ

in vitro

in plastico

in silico

Recombinant DNA

plasmid

transformation

Ori

selectable marker

Multiple cloning site (MCS)

competent

electroporation

vector

clone

Lambda phage

cosmid

lambda phage packaging system

BACs

DNA ligation

DNA ligase

sticky / compatible end

genomic library

cDNA library

STUDY QUESTIONS:

1) A coat protein from a particular virus can be used to immunize children against further infection. However, inoculation of children with proteins extracted from natural viruses sometimes causes a fatal disease, due to contamination with live viruses. How could you use molecular biology to produce an optimal vaccine?

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

Notes:

CHAPTER 30 – CLONING A GENE
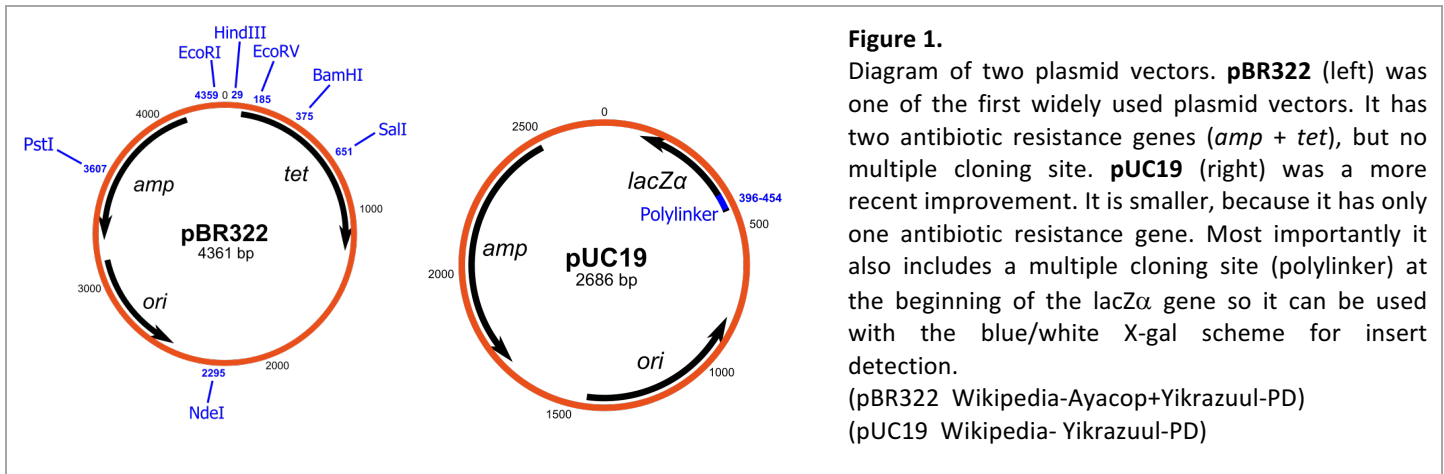
Figure 1.
Diagram of two plasmid vectors. **pBR322** (left) was one of the first widely used plasmid vectors. It has two antibiotic resistance genes (*amp* + *tet*), but no multiple cloning site. **pUC19** (right) was a more recent improvement. It is smaller, because it has only one antibiotic resistance gene. Most importantly it also includes a multiple cloning site (polylinker) at the beginning of the lacZα gene so it can be used with the blue/white X-gal scheme for insert detection.
(pBR322 Wikipedia-Ayacop+Yikrazuul-PD)
(pUC19 Wikipedia- Yikrazuul-PD)

INTRODUCTION

As a geneticist, suppose you have created a strain that has a mutation in a gene that involves a biological process you wish to learn more about. This could be a gene dealing with cancer, development, or heart disease, or something as simple as the colour of an eye or flower. All you have to start with is this mutant strain, and you know that it's genetically different from wild type (or its parents).

How do you find the gene (that is mutant) amongst the ~20,000 that are present in the typical eukaryote genome?

The process of cloning a gene involves several steps and has been done in multiple different ways. Only three basic methods will be presented here. All involve the construction of a **genomic DNA** clone library, which is the cloning of all the sequences in a genome into a **DNA vector** (a plasmid for example, **Figure 1**). Then this library is screened, using one of several methods, to find the specific DNA sequence (clone) that contains the gene of interest to you.

The three methods for screening a library will be (1) cloning by complementation; (2) hybridization of DNA Probes; and (3) transposon tagging.

1. CLONING BY COMPLEMENTATION – A HYPOTHETICAL AUXOTROPHIC MUTATION IN *E. COLI*

The concept of genetic **complementation** was covered previously in Chapter 4. This cloning method is primarily used with single-cell organism, such as bacteria or yeast (fungi). It involves the introduction of one wild type gene (allele) into a mutant strain, which is then able to supply sufficient product to result in a wild type phenotype. The easiest explanation is a hypothetical example dealing with an auxotrophic mutation in *E. coli*.

1.1. BUILDING A GENOMIC DNA LIBRARY CONTAINING THE WILD TYPE GENE

To complete this method we need both an autotrophic mutant (A-) and a wild type (A+) strain.

The first step is to **clone** the wild type allele of the gene of interest. Step (1) is to **restriction enzyme** digest the DNA of the wild type strain into clone-sized fragments. One fragment should contain the A+ gene. These fragments are cloned into a vector, such as a plasmid by means of **ligation** (**Figure 2**). The ligated *E. coli* genomic DNA and plasmid vector constructs are a **genomic DNA library**.
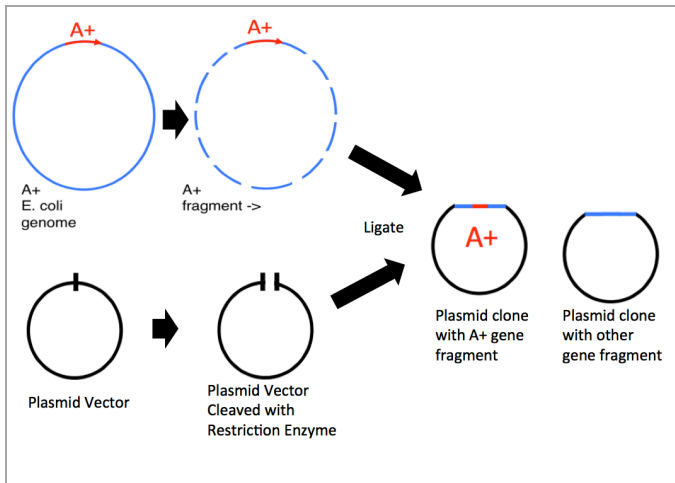
Figure 2.
E. coli genomic DNA (blue) is digested with a restriction enzyme to generate fragments for cloning into the plasmid vector (black). The ligation products will include some clones with the A+ gene (red) and many without (blue). (Original-J. Locke-CC BY-NC 3.0)

1.2. TRANSFORMATION AND CLONE SELECTION VIA COMPLEMENTATION

The ligated genomic library is then **transformed** into the auxotrophic mutant host (A-). This will result in a variety of cells with different genotypes as shown in **Figure 3**.

The untransformed cells (left) will not grow on any **Minimal Media** (MM) with or without **antibiotic**. The cells transformed with a plasmid containing a genomic fragment (blue) will grow on the antibiotic containing media because of the antibiotic resistance on the plasmid, but only if it contains the supplement for the **auxotrophic** strain. The transformants that have the A+ gene (red) will grow on antibiotic media that lacks the supplement. Thus the Minimal Media plate, with or without antibiotic, can be used to select (screen for) the clone with the A+ gene, a **prototroph**.

Typically, when doing an experiment like this, a researcher would recover several independent clones, each having the same A+ gene fragment cloned in it. This recurring result would provide evidence that this is the gene that of interest.

1.3. SELECTION OF THE APPROPRIATE RESTRICTION ENZYME

Note in this example the restriction enzyme cut the genomic DNA such that the entire A+ gene was contained on one fragment. In actual practice a researcher would not know ahead of time which enzyme to use. Several different enzymes would need to be tried to identify an enzyme that would not cleave within the gene. Separating the A+ gene into two fragments would not result in the recovery of the gene.

Once cloned, this fragment can be sequenced and the genes within it characterized and used for further experiments to determine its function and role in the biological process under investigation.

Host Cells: Plasmid:	E. coli A– no plasmid	E. coli A+ plasmid without A+ gene	E. coli A+ plasmid with A+ gene
Plate on MM: with antibiotic with supplement	no growth	growth	growth
Plate on MM: with antibiotic no supplement	no growth	no growth	growth
Plate on MM: no antibiotic no supplement	no growth	no growth	growth

MM = minimal medium
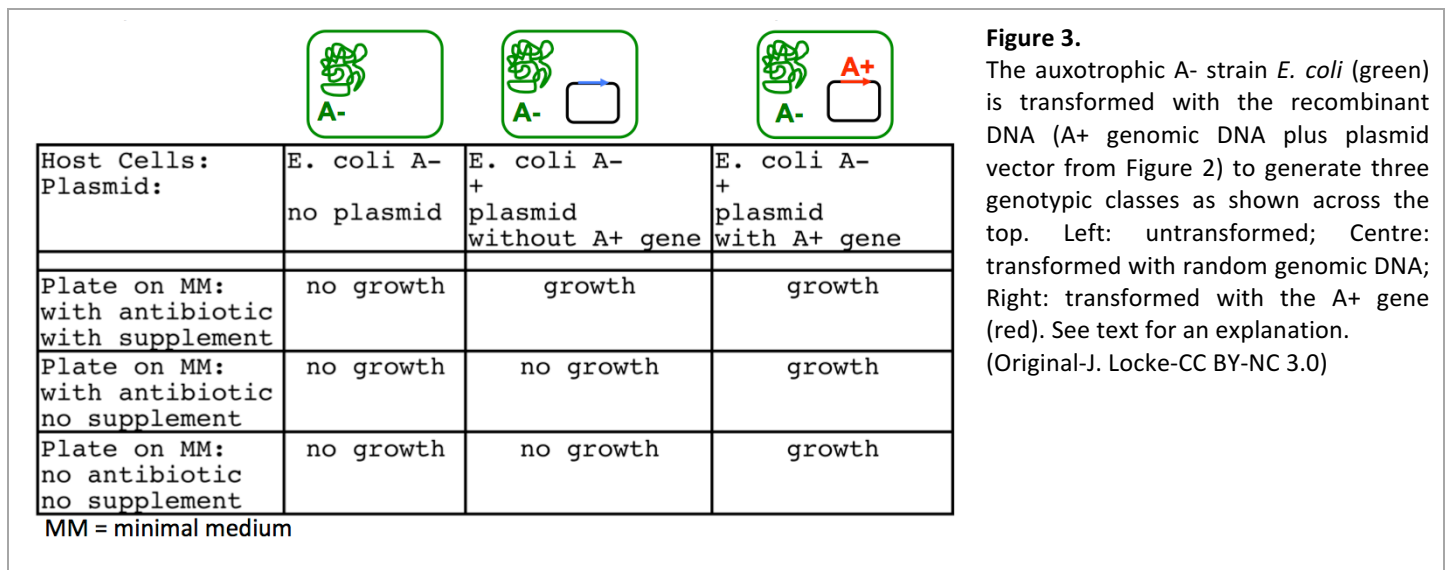
Figure 3.
The auxotrophic A- strain *E. coli* (green) is transformed with the recombinant DNA (A+ genomic DNA plus plasmid vector from Figure 2) to generate three genotypic classes as shown across the top. Left: untransformed; Centre: transformed with random genomic DNA; Right: transformed with the A+ gene (red). See text for an explanation. (Original-J. Locke-CC BY-NC 3.0)

1.4. HOW MANY CLONES ARE NEEDED TO FIND THE ONE OF INTEREST?

Given that the construction of the recombinant genomic DNA library is random, in that the fragment with the A+ gene is only one of many in the genome, how many clones do we need to make and screen to be sure of finding the one we seek? There is a simple formula that researchers use to estimate this probability:

(1) If each plasmid contains ~4.5 Kb of insert DNA and the *E. coli* genome contains ~4.5 Mb DNA then ~1000 plasmid clones, if arranged end-to-end could contain one *E. coli* genome's worth of DNA.

(2) Because of the random nature of which fragment are cloned, probability says we need to screen the equivalent of 5 genomes worth of clones (e.g. 5,000 clones in this case) to provide a 99% chance of finding the A+ gene (or any/every other gene in the genome), taking the conditions in the previous section into account. Because cloning is a statistical probability, there can never be a 100% chance, but the 99% chance is almost always sufficient to find the gene of interest.

Note that 5,000 bacterial clones can be produced easily and screened quickly on a single Petri dish plate. This method is relatively easy and straightforward.

2. CLONING BY HYBRIDIZATION OF DNA PROBES

2.1. DNA HYBRIDIZATION IS SEQUENCE SPECIFIC

The use of DNA probes to recover clones from genomic, or other libraries, relies on the principle of **DNA hybridization.** DNA is normally a duplex with the two strands are held together via hydrogen bonds (A=T   G=C) and base stacking interactions.

The two strands can be separated in a process called **denaturation**. This can be easily done by heating (e.g. boiling water 100°C) or alkali (e.g. 50 mM Na0H) to raise the pH.

```
         <duplex DNA>

5'....GAATTCGGATCC....3'
3'....CTTAAGCCTAGG....5'
            ↓

5'....GAATTCGGATCC....3'
            +
3'....CTTAAGCCTAGG....5'
```

These dissociated single strands can **reassociate** (or **anneal**) to reform the duplex DNA. This process is sequence specific, in that the duplex will only form if the two strands are complementary.

When the two strands reform a duplex, a hybrid is formed, hence the name hybridization.

```
5'....GAATTCGGATCC....3'
3'....CTTAAGCCTAGG....5'
    (duplex DNA again)
```

Note: hybrids can form between DNA/DNA (two complementary DNA strands), DNA/RNA (a DNA strand and its complementary RNA sequence), or between RNA/RNA (not useful here since clones are DNA).

Hybrid formation only requires that the complementary sequences be similar, not a perfect match. A hybrid can form some "**mismatch**" in the pairing.

```
5'....GAATTTAGATCC....3'
      |||||  |||||
3'....CTTAAGCCTAGG....5'
    (duplex DNA again)
```

The extent of mismatch possible in a hybrid depends upon the hybridization conditions in the solution (temperature, salt concentration etc). Under some conditions 30-40% mismatched strands are able to form a stable hybrid. This can be useful for finding a similar gene across species. The DNA sequence from one species can be used to find a similar gene in another species. For example, see **Figure 4**.

Figure 4.
A DNA sequence alignment showing a comparison between mouse and Drosophila DNA sequences for part of their actin gene. There are seven mismatches in this 120 base sequence (94% similar), which would probably easily form a hybrid. (Original-J. Locke-PD)

2.2. SOURCE OF A DNA PROBE

Typically a cloned DNA sequence is used as a **probe** into a DNA library. There must be a source for the DNA and it can be from many different sources, but we will present only four common ones.

First, these sequences can come from "natural" (pre-existing) sources, such as a similar gene cloned from another species. See the mouse/Drosophila actin gene in **Figure 4** above. The cloned Drosophila actin gene sequence can be used as a probe into a mouse genomic DNA library to find the mouse actin gene, or *vice versa*. Another source would be a

cDNA clone, which is a plasmid vector containing the DNA complementary to an mRNA sequence. The process of making a cDNA library is similar to that of a genomic DNA library. The result is a set of clones that contain DNA sequences corresponding to the mRNA molecules in a sample, cell, or tissue. The mRNA of a sample is extracted and reverse transcribed into a complementary DNA (cDNA) sequence that is then cloned. This procedure typically makes a mixture of clones that represents the diversity of mRNA in the sample. See **Figure 5** for a diagrammatic description of the method.

Another source of DNA for a probe can come from **synthetic oligo-nucleotides**. The sequence can be derived from the amino acid sequence of the polypeptide the gene encodes. The amino acid sequence can be determined by various biochemical techniques and then reverse translated to determine what specific DNA sequence to chemically synthesize. The degeneracy of the triplet code usually makes it impossible to obtain a unique sequence. Instead, a degenerate oligo-sequence is used.
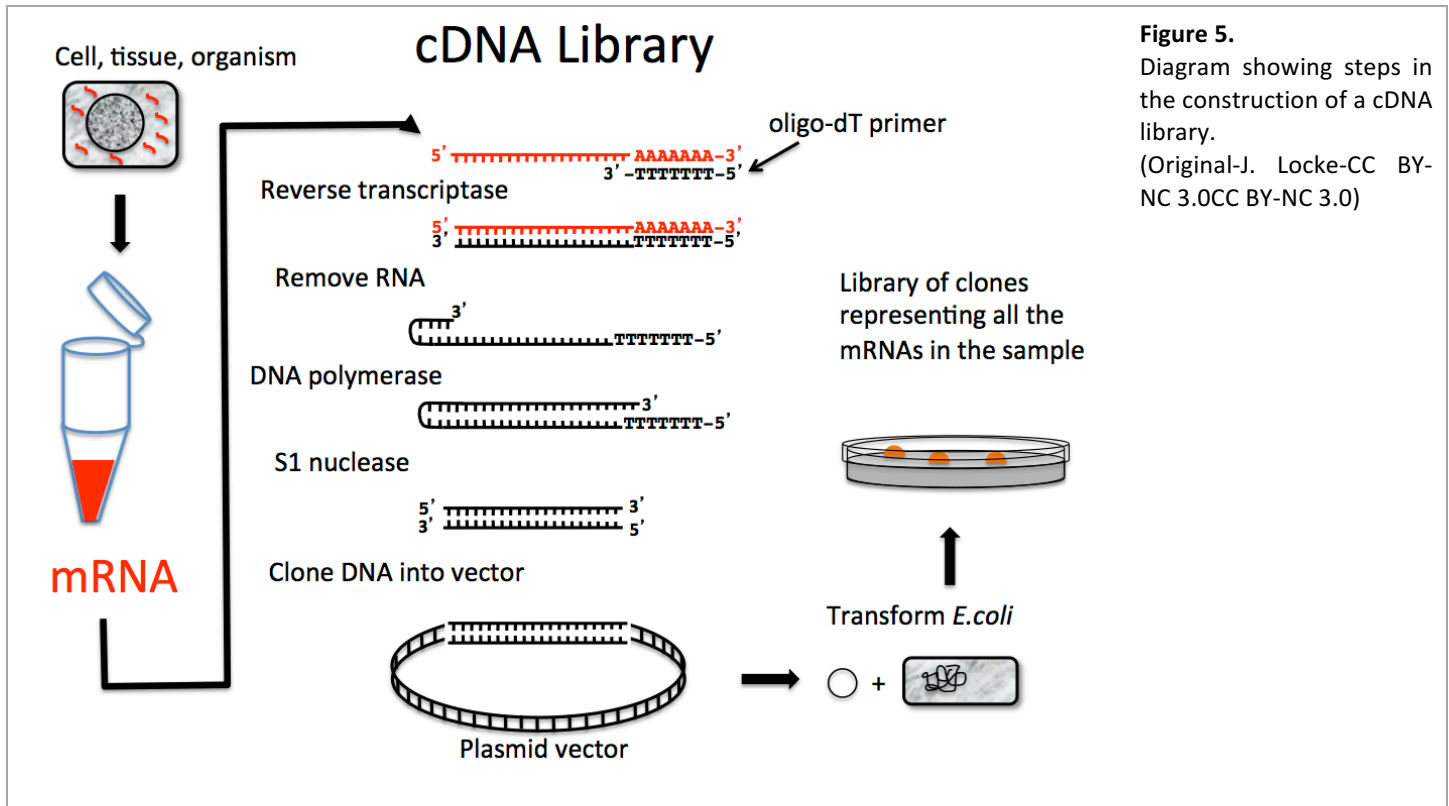
Figure 5.
Diagram showing steps in the construction of a cDNA library.
(Original-J. Locke-CC BY-NC 3.0CC BY-NC 3.0)

Example:

```
Protein sequence: Met – Lys – Asn — Glu        Lys=Lysine
           codon: AUG – AAA – AAU – GAA        Asn=Asparagine,
  alternate codon:     – AAG – AAC — GAG       Glu=Glutamic Acid
   Probe sequence: ATG    AAA    AAT    GAA
                          G      C      G
                          |      |      |
                  use both bases in the oligo–nucleotide
                     a 50:50 mix of each base
```

Another source of DNA for probes comes from the amplification of polymerase chain reaction (PCR products). The methodology of PCR is covered in Chapter 31 and will not be presented here. Basically PCR amplification is just another method to synthesize sufficient DNA for use as a probe. It uses two primer oligonucleotides to synthesize (amplify) a specific sequence of duplex DNA. This can then be used as a probe as described in the next section.

2.3. SCREENING A LIBRARY TO FIND A CLONE

Once a DNA fragment has been obtained (see above) it can be labeled and used as a probe. "**Labeling**" involves putting a tag on the DNA in some manner that permits one to detect its presence, in minute quantities, at some later point in an experiment. DNA can be "labeled" in several different ways. One widely used technique is to replace the normal Phosphorous of the DNA with a radioactive atom of Phosphorous, ^{32}P (normal isotope = ^{31}P). This radioactivity can be detected by photographic emulsion (e.g. **auto-radiography**) on sheets of X-ray film. There are several methods to do this (nick-translation, random priming, PCR). All produce a single strand of ^{32}P labeled DNA that will hybridize with its complementary sequence and thus localize any DNA with that sequence.

We now have the probe labeled and ready to screen a genomic library for a gene of interest. The method of screening a library of DNA clones relies on the probe's sequence specificity to hybridize with only the clones with a complementary sequence. The procedure is diagrammed in **Figure 6**.

The goal of screening a clone library is to identify and recover clone(s) that have a sequence complementary to the probe (e.g.: a specific gene sequence)
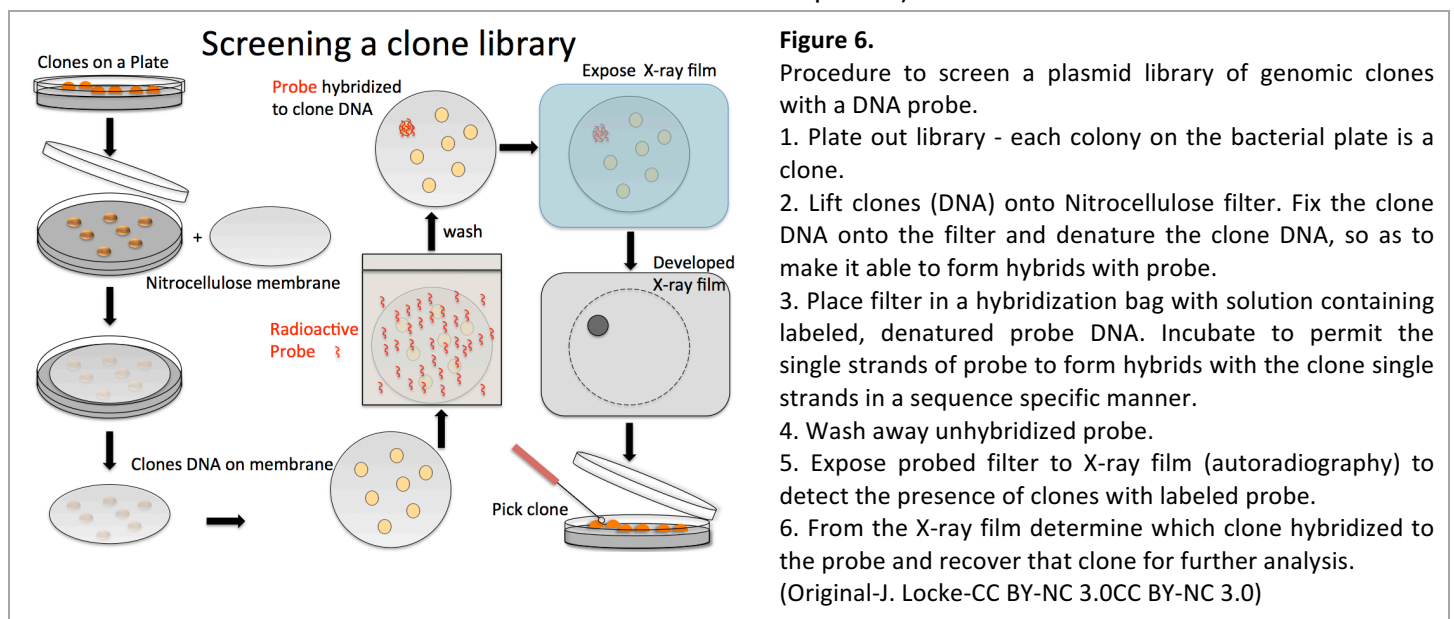
Figure 6.
Procedure to screen a plasmid library of genomic clones with a DNA probe.
1. Plate out library - each colony on the bacterial plate is a clone.
2. Lift clones (DNA) onto Nitrocellulose filter. Fix the clone DNA onto the filter and denature the clone DNA, so as to make it able to form hybrids with probe.
3. Place filter in a hybridization bag with solution containing labeled, denatured probe DNA. Incubate to permit the single strands of probe to form hybrids with the clone single strands in a sequence specific manner.
4. Wash away unhybridized probe.
5. Expose probed filter to X-ray film (autoradiography) to detect the presence of clones with labeled probe.
6. From the X-ray film determine which clone hybridized to the probe and recover that clone for further analysis.
(Original-J. Locke-CC BY-NC 3.0CC BY-NC 3.0)

3. CLONING A GENE USING THE TRANSPOSON TAGGING METHOD

The final method of cloning a gene uses the creation of a mutation by inserting a **mobile element** (ME), also called a **transposable element** (TE) or **transposon**, into the gene of interest. This provides a DNA sequence tag (a known sequence) inserted into the unknown gene, and thus a means through which the gene of interest can be identified and cloned. That is, you can **transposon tag** a gene to clone it.

3.1. BASICS OF TRANSPOSABLE ELEMENTS

All eukaryotes have multiple types of transposable elements as single or multiple copies in their genomes. These segments of DNA, usually hundreds or thousands of base pairs long, are usually distributed throughout the genome as randomly inserted sequences. Typically, they are mobile, in the sense that some of them are able to move and/or replicate within the genome. There are two main classes: (1) **Retro-transposons** and (2) **DNA transposons**.

Retro-transposons (or **retroposons**) move from one site to another in the genome via RNA intermediates (**Figure 7**). The sequence is transcribed from the DNA sequence at one locus into RNA, then reverse-transcribed back into DNA (DNA->RNA->DNA), which is then inserted back into the genomic DNA at another locus. Within this class there are two subclasses. The first is the **retrovirus-like class**, which has sequences similar in organization to a retrovirus, such as HIV. Their replication is also like a retrovirus. Then there is the **non-viral class** of retroposons. The human *Alu* element is an example of a s̲hort i̲nterspersed element (SINE). There are also l̲ong i̲nterspersed elements (LINES).

DNA transposons move via DNA intermediates (DNA-> DNA). No RNA (transcription) is involved. One of the best-studied elements is the **P element** in Drosophila. It is very well characterized and used as a **genetic tool** to create insertion mutations (mutagenesis) and as a transformation vector to move constructed genes into the germline of
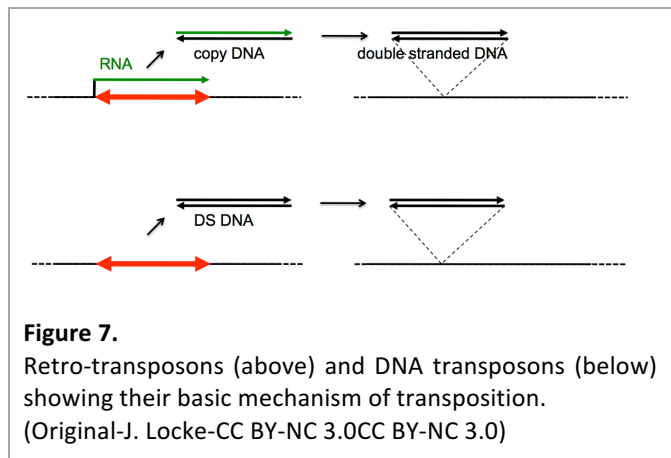
Figure 7.
Retro-transposons (above) and DNA transposons (below) showing their basic mechanism of transposition.
(Original-J. Locke-CC BY-NC 3.0CC BY-NC 3.0)

Drosophila (transgenes, etc.). We will use the P element as an insertion mutation in our model example here.

P elements are 2907 bp long, have 31 bp inverted repeats at either end, and code for either a **transposase** or **repressor protein**, depending on the mRNA splicing alternative. A transposase is an enzyme that binds to the transposon and catalyzes the movement from its current location to another part of the genome. It does this by cleaving the strands surrounding the region, and then cutting and inserting the transposon in a new location in the genome (**Figure 7**). The transposase transcript has 4 exons (Exon 0 - Exon 3). The repressor polypeptide is made if the last intron (2-3) is not spliced out. The resulting mRNA will be translated such that the **2-3 intron** is translated and a premature stop codon prevents translation of the sequence in exon 3 (**Figure 8**). A simplistic description proposes that this truncated product will bind to the P element ends, but not cleave the DNA as the transposase polypeptide does. This binding prevents other transposase molecules from binding and cleaving, thereby acting as a repressor of mobilization. The real situation is more complex. In most somatic tissues, the 2-3 intron is not removed from the primary transcript so only the repressor is made. In germline tissues, the intron can be removed and the transposase produced to cause mobilization. The mobilization of the transposon will cause insertions, which can cause mutations in genes. This is explained in the next section.

3.2. CREATING A TE INDUCED MUTANT IN A GENETIC SCREEN

This needs wild type and mutant stocks for the gene of interest. If a P element induced mutation is already available from a stock collection, then this step is completed. However, this is not the case with most situations and so you must make your own. To create a P element induced mutation a P element containing stock (P-stock) is crossed with one that lacks P elements. Both must have wild type alleles of the gene of interest. If done correctly, this cross will cause the P elements to mobilize, and create random insertions into genes all over the genome. Next is a genetic screen for this rare insert into the gene of interest. It can be identified in the screen by a failure to complement an existing mutant allele of the gene of interest (See Chapter 4).

This new allele is tagged with the P element and can now be recovered and isolated for further use.

3.3. CLONING A TE INSERTION SITE & THE WILD TYPE ALLELE

From the stock with the P element induced allele in the gene of interest, the genomic DNA can be extracted and a genomic library built. This library can be screened and a clone containing the P element sequence identified (see the previous section on using a DNA probe to screen a library).

These clones can be characterized and the genomic sequences adjacent to the P element can be subcloned and used as a probe into a library made from a wild type stock. This will identify clones containing the wild type DNA of the gene of interest.

3.4. CONFIRMING THE INSERT IS RIGHT GENE BY REVERSION ANALYSIS

Because the insertion of P element (and other TE) are not always the causal event in the type of mutagenesis described above, it is essential that the cause of the mutation in the gene of interest be established to be due to the insert that you have cloned. This is usually done through **reversion analysis.**

The allele with the TE in the gene of interest (above) is reverted in a manner similar to that of the initial insertion. The expectation is that the excision of the TE will be associated with a reversion of the mutation back to wildtype. The presence/absence of the TE insert can be monitored by Southern Blots in the reverted stocks and compared to the original, parental line and the insert mutant line.

4. CURRENT APPROACHES TO MATCHING GENES TO MUTATIONS

The methods described above have been used for several decades to successfully find many new genes. However, the current ease and quickness of **whole genome sequencing** is changing this methodology. If sufficient money and equipment is available, the mutant genome can be compared to the parental or wild type genome and single base pair changes can be identified by computer analysis. While theoretically straightforward, the technical details make it challenging and unclear. Often there are many changes in the sequence and it is not clear which are causative and which are just random changes.
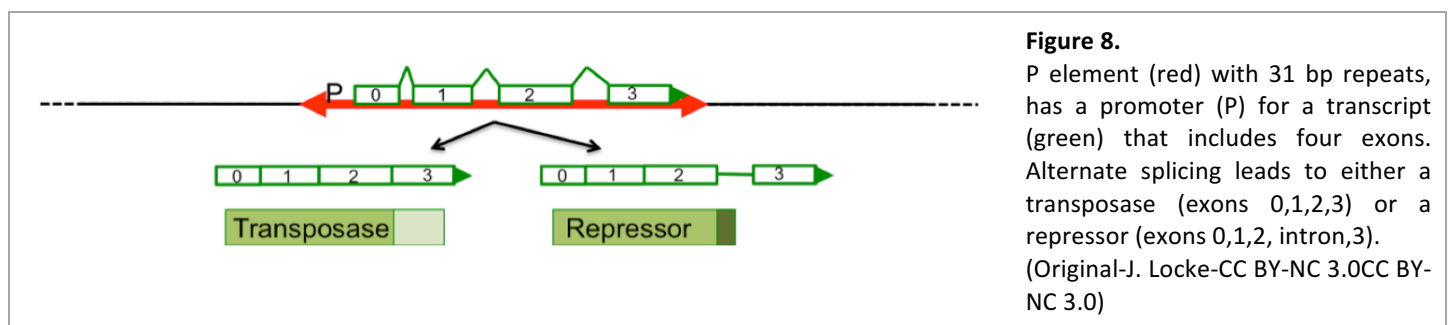
Figure 8.
P element (red) with 31 bp repeats, has a promoter (P) for a transcript (green) that includes four exons. Alternate splicing leads to either a transposase (exons 0,1,2,3) or a repressor (exons 0,1,2, intron,3). (Original-J. Locke-CC BY-NC 3.0CC BY-NC 3.0)

_____

SUMMARY:

- Genes in simple organisms (single cells) can be cloned by the complementation of host mutation by the transformation of a plasmid containing a wild type copy of the host's mutant locus.

- In screening a library for a clone, five genomes worth of cloned DNA needs to be screened in order to have a 99% probability of finding the clone of interest.

- Libraries can be screened with a labeled probe using the DNA-DNA hybridization to bind the probe to a target sequence in a clone.

- Genes can also be cloned via transposon tagging.

KEY TERMS:

pBR322
pUC19
genomic DNA
DNA vector
complementation
clone
restriction enzyme
ligation
genomic DNA library
transformed
minimal media
antibiotic
auxotroph
prototroph
DNA hybridization
denaturation
reassociation
annealing
mis-match
probe

cDNA clone
synthetic oligo-nucleotide
labeling
^{32}P
autoradiography
mobile element
transposable element
transposon
transposon tag
retro-transposon
DNA transposon
retrovirus-like class
non-viral class
P element
Genetic tool
transposase
repressor
2-3 intron
reversion analysis
whole genome sequencing

QUESTIONS:

1) In **Figure 4**, it is intentionally not stated which is the mouse and which is the Drosophila DNA sequence. With a computer and internet access, how could you determine which is which?

2) In the example showing how the amino acid sequence is used to make an oligonucleotide probe, there are four amino acids shown. Assume the next is proline.
   a) Find a codon usage chart on the internet and determine what the next three nucleotides should be in the oligonucleotide.
   b) With the addition of the proline, how many different sequence oligo-nucleotides are needed to cover all the possible gene sequences (assuming there is no exon-intron site in this region)?

Notes:

CHAPTER 31 – POLYMERASE CHAIN REACTION (PCR)
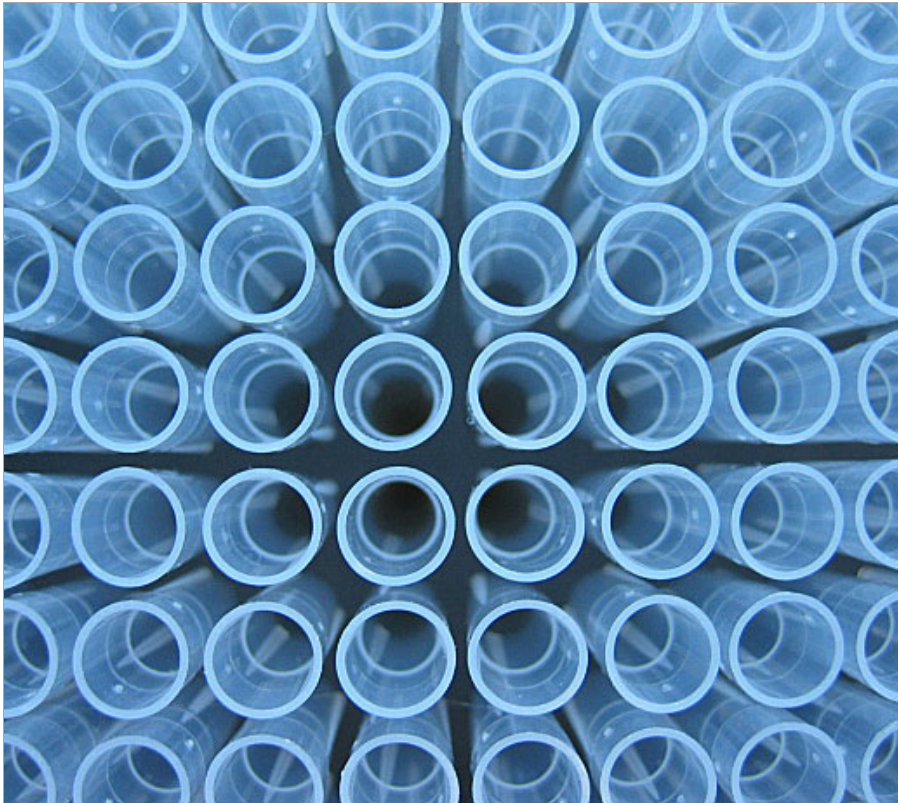
Figure 1.
Plastic disposable tips for a micro-pipettor are used to accurately distribute microliter volumes of liquid in molecular biology. (Flickr-estherase- CC BY-NC-SA 2.0)

INTRODUCTION

While, genetics is the study of the inheritance and variation of biological traits, today, **classical genetics** is often complemented by **molecular biology**, to give **molecular genetics**, which involves the study of DNA and other **macromolecules** that have been isolated from an organism. Usually, molecular genetics experiments involve some combination of techniques to isolate, analyze, and characterize the DNA, RNA, and/or protein transcribed and translated from a particular gene. In some cases, the DNA may be subsequently manipulated by mutation or by recombination with other DNA fragments. Techniques of molecular genetics have wide application in many fields of biology, as well as forensics, biotechnology, and medicine. Polymerase Chain Reaction (PCR) is a widely used technique to amplify and isolate specific DNA sequences. It requires a "template" DNA, which is often genomic DNA. From this template, specific sequences can be amplified and many copies can be produced for analysis or manipulation.

1. ISOLATING GENOMIC DNA

DNA purification strategies rely on the chemical properties of DNA that distinguish it from other molecules in the cell, namely that it is a very long, negatively charged molecule. To extract purified DNA from a tissue sample, cells are broken open by grinding or **lysing** in a solution that contains chemicals that protect the DNA while disrupting other components of the cell (**Figure 2**). These chemicals may include **detergents**, which dissolve lipid membranes and denature proteins. A cation such as Na^+ helps to stabilize the negatively charged DNA and separate it from proteins, such as histones. A **chelating agent**, such as **EDTA**, is added to protect DNA by sequestering Mg^{2+} ions, which can otherwise serve as a necessary co-factor for **nucleases** (enzymes that digest DNA). As a result, free, double-stranded DNA molecules are released from the cell and from chromatin into the

extraction buffer, which also contains proteins and all other cellular components. (The basics of this procedure are simple enough that it can be done with household chemicals as presented on YouTube.)

The free DNA molecules are subsequently isolated by one of several methods. Commonly, proteins are removed by adjusting the salt concentration so they precipitate. The **supernatant**, which contains DNA and other, smaller metabolites, is then mixed with ethanol, which causes the DNA to precipitate. A small **pellet** of DNA can be collected by centrifugation, and after removal of the ethanol, the DNA pellet can be dissolved in water (usually with a small amount of EDTA and a pH buffer) for the use in other reactions. Note that this process has purified all of the DNA from a tissue sample (genomic and mitochondrial DNA); if we want to isolate a specific gene or DNA fragment, we must use additional techniques, such as PCR.

Figure 2.
Extraction of DNA from a mixture of solubilized cellular components by successive precipitations. Proteins are precipitated, then DNA (in the supernatant) is precipitated with ethanol, leaving a pellet of DNA.
(Original-Deyholos-CC BY-NC 3.0)
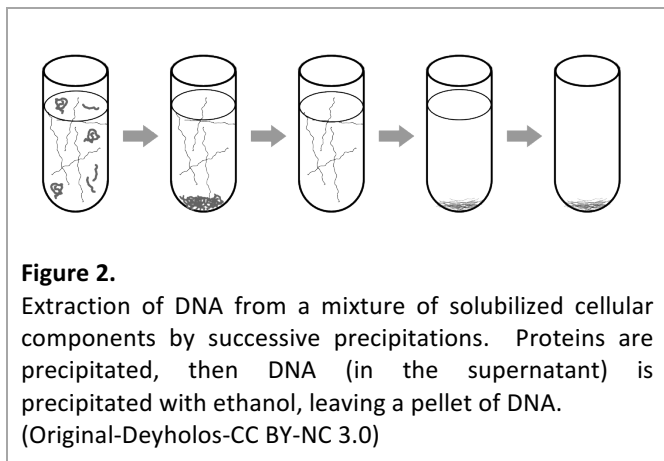
2. ISOLATING OR DETECTING A SPECIFIC SEQUENCE BY PCR

2.1. COMPONENTS OF THE PCR REACTION

The **Polymerase Chain Reaction (PCR)** is a method of DNA amplification that is performed in a test tube (i.e. *in vitro*). Here "polymerase" refers to a DNA polymerase enzyme extracted and purified from bacteria. The "chain reaction" refers to the ability of this technique produce billions of copies of a specific DNA molecule, by using each newly replicated double helix as a template to synthesize two new DNA double helices. PCR is therefore a very efficient method of amplifying a specific sequence of DNA from a small sample of a large, complex genome.

Besides its ability to make large amounts of DNA, there is a second characteristic of PCR that makes it extremely useful. Recall that most DNA polymerases can only add nucleotides to the end of an existing strand of DNA, and therefore require a **primer** to initiate the process of replication. For PCR, chemically synthesized primers of about 20 nucleotides are used. In an ideal PCR, primers only hybridize to their exact complementary sequence on the template strand (**Figure 3**).

Figure 3.
The primer-template duplex at the top part of the figure is perfectly matched, and will be stable at a higher temperature than the duplex in the bottom part of the figure, which contains many mismatches and therefore fewer hydrogen bonds. If the annealing temperature is sufficiently high, only the perfectly matched primer will be able to initiate extension (grey arrow) from this site on the template.
(Original-Deyholos-CC BY-NC 3.0)

The experimenter can therefore control exactly what region of a DNA template is amplified by specifying the sequence of the primers used in the reaction.

To conduct a PCR amplification, an experimenter combines in a small, thin-walled tube (**Figure 4**), all of the necessary components for DNA replication, including:

(1) DNA polymerase and solutions containing

(2) nucleotides (dATP, dCTP, dGTP, dTTP),

(3) a DNA template,

(4) DNA primers,

(5) a pH buffer, and

(6) ions (e.g. Mg^{2+}) required by the polymerase.

Successful PCR reactions have been conducted using only a single DNA molecule as a template, but in practice, most successful PCR reactions contain many thousands of template molecules. The template DNA (e.g. total genomic DNA) has usually already been purified from cells or tissues using the techniques described above. However, in some situations it is possible to put whole cells directly in a PCR reaction for use as a template.
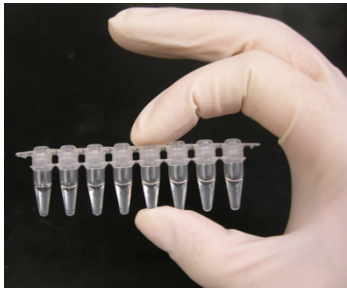
Figure 4.
A strip of PCR tubes
(Wikipedia-madprime- CC BY-SA 3.0)
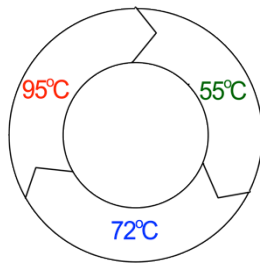
Figure 5.
Example of a thermal-cycle, in which the annealing temperature is 55°C.
(Original-Deyholos-CC BY-NC 3.0)
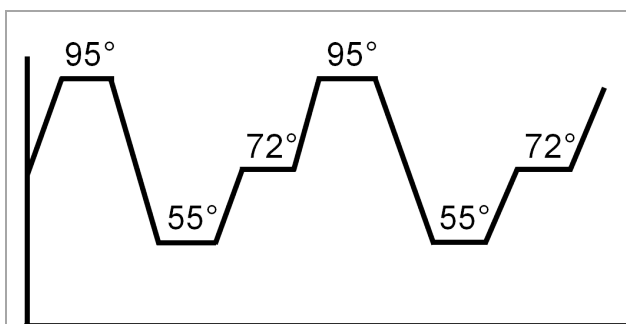
Figure 6.
A temperature vs. time graph showing two cycles of PCR.
(Original-Harrington-CC BY-NC 3.0)

An essential aspect of PCR is **thermal-cycling**, meaning the exposure of the reaction to a series of precisely defined temperatures (**Figure 5**). The reaction mixture is first heated to 95°C. This causes the hydrogen bonds between the strands of the template DNA molecules to melt, or **denature**. This produces two single-stranded DNA molecules from each double helix (**Figure 7**). In the next step (**annealing**), the mixture is cooled to 45-65°C. The exact temperature depends on the primer sequence used and the objectives of the experiment. This allows the formation of double stranded helices between complementary DNA molecules, including the annealing of primers to the template. In the final step (**extension**) the mixture is heated to 72°C. This is the temperature at which the particular DNA polymerase used in PCR is most active. During extension, the new DNA strand is synthesized, starting from the 3' end of the primer, along the length of the template strand. The entire PCR process is very quick, with each temperature phase usually lasting ~30 seconds or less. Each cycle of three temperatures (denaturation, annealing, extension) is usually repeated about 30 times, amplifying the target region approximately 2^{30}-fold. The amount of DNA product reaches a plateau at 20-40 cycles, usually because the nucleotide precursors have been exhausted. Notice from the figure that most of the newly synthesized strands in PCR begin and end with sequences either identical to or complementary to the primer sequences; although a few strands are longer than this, they are in such a small minority that they can almost always be ignored.

After completion of the thermal-cycling (amplification), an aliquot from the PCR reaction is usually loaded onto an **electrophoretic agarose gel**

Figure 7.
PCR with the three phases of the thermal cycle numbered. The template strand (blue) is replicated using primers (red), to prime the newly synthesized strands (green). The green strands, which are flanked by the two primer sites, will increase in abundance exponentially through successive PCR cycles.
(Wikipedia-madprime- CC0 1.0)
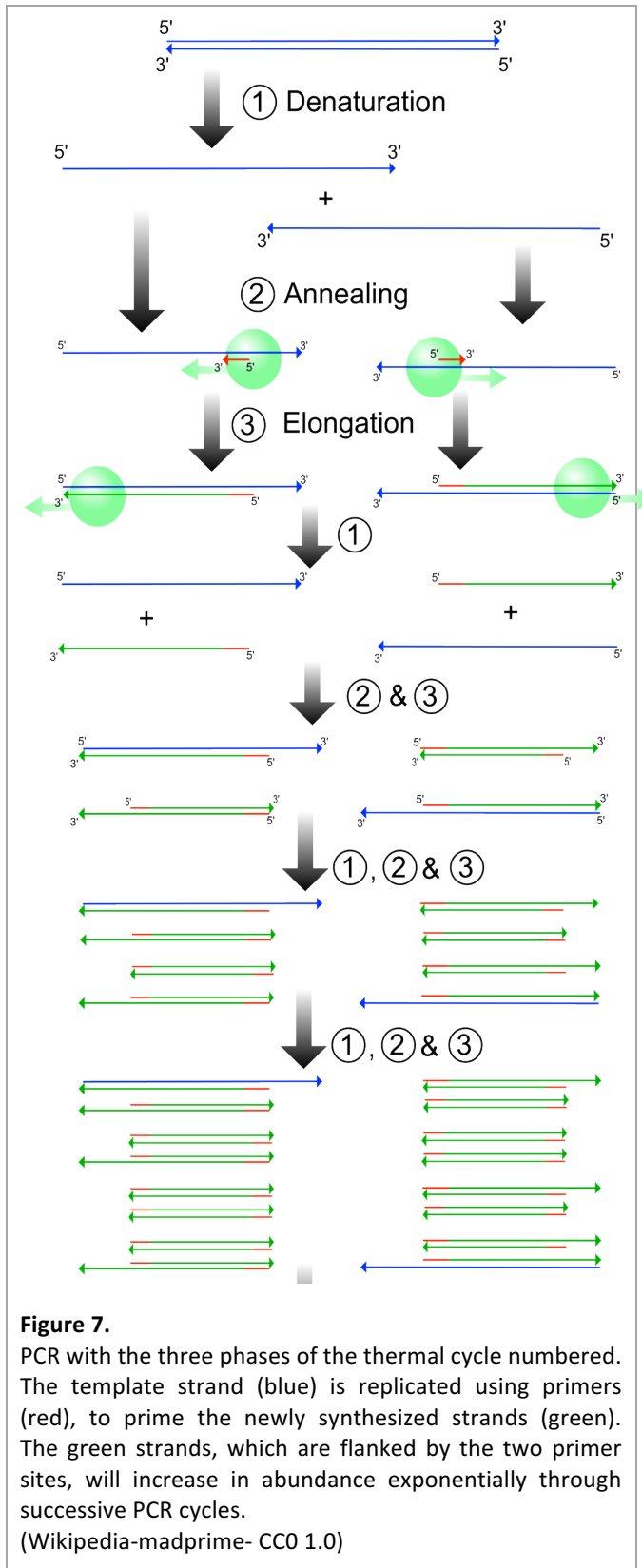
(described in chapter 28) to determine whether a DNA fragment of the expected length was successfully amplified or not. Usually, the original template DNA will be so dilute that it will not be visible on the gel, only the amplified PCR product. The presence of a sharp band of the expected length indicates that PCR was able to amplify its target. If the purpose of the PCR was to test for the presence of a particular template sequence, this is the end of the experiment. Otherwise, the remaining PCR product can be used as starting material for a variety of other techniques such as sequencing or cloning.

2.2. REAL TIME PCR / QUANTITATIVE PCR (qPCR)
In a standard PCR reaction, the DNA molecule of interest is amplified and then the products are typically visualized at the end of the reaction on an electrophoresis gel. On the other hand, a procedure known as real-time PCR or quantitative PCR (qPCR) detects the replicated DNA molecules <u>during</u> the amplification process. qPCR uses florescent molecules and relies on the fluorescence of the amplified product measured over a number of cycles. However, the procedure of amplifying the DNA molecule is identical to the standard PCR procedures. There are two ways of processing qPCR.

(1) Using florescent chemical molecules known as **fluorochrome** that binds to all double stranded DNA molecule (Nonspecific).

In the first method, the florescent dye molecule binds to any double stranded DNA molecule. After each cycle of amplification, the amount of ds-DNA molecules synthesized can be quantified by measuring the florescence. The intensity of the florescence would indicate the amount of DNA molecule present.

(2) Using florescent reporter probe (Specific).

The second method is using a florescent reporter probe that hybridizes with the DNA sequence of interest. When the *taq* polymerase replicates the DNA molecule, it degrades the probe and the florescent molecule is released to the solution. This increases the intensity of florescence. The florescence is measured by the real-time PCR

machine and quantifies the DNA molecules being synthesized.

2.3. REVERSE TRANSCRIPTASE PCR (RT-PCR)

Reverse Transcriptase PCR (RT-PCR) can detect both the quality and quantity of mRNA molecules (gene transcription). As a result, we are able to find out the spatial (where the gene is expressed) and temporal (when the gene is expressed) level of gene expression.

Here is how it works (**Figure 8**):

(1) mRNA is extracted from the cell, tissue, or organism.

(2) An enzyme called **reverse transcriptase** (obtained from a retrovirus – see Chapter 30) is added, along with oligo-DT, which anneals to the poly-A tail and acts as a primer, to synthesize complementary DNA (cDNA) to the mRNA.

(3) mRNA template is degraded, and cDNA is added to a PCR reaction to amplify a specific gene sequence. If amplification occurs, the mRNA is present; if not, then it is absent. This permits the quantitation of a specific mRNA (gene) sequence.

The amplified products visualized on a gel verify the existence and the quantity of the gene of interest. By extracting mRNA at different stages, we can figure out the **temporal level** of gene expression. If we extract mRNAs from different cell types, we can figure out the **spatial level** of gene expression.

2.4. AN APPLICATION OF PCR: THE STARLINK AFFAIR

PCR is very sensitive (meaning it can detect very small starting amounts of DNA), and specific (meaning it can amplify only the target sequence from a mixture of many DNA sequences). Due to these characteristic, PCR has many practical applications. For example, PCR can detect trace DNA contaminants in food, air, water or cells. The presence or absence and the type or species of the contaminant can be identified.

As an example, PCR was used as a tool to test whether genetically modified corn was present in consumer products on supermarket shelves. Although currently (2013) 85% of corn in the
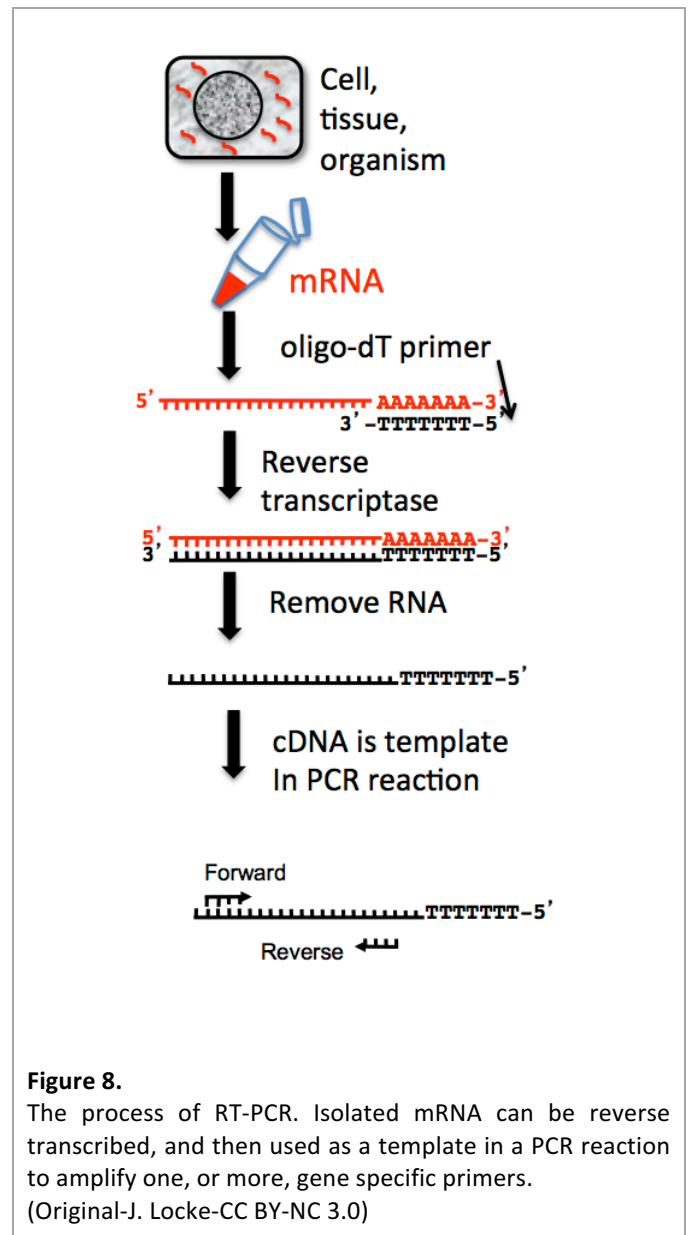
Figure 8.
The process of RT-PCR. Isolated mRNA can be reverse transcribed, and then used as a template in a PCR reaction to amplify one, or more, gene specific primers.
(Original-J. Locke-CC BY-NC 3.0)

United States is genetically modified, and contains genes that government regulators have approved for human consumption, back in 2000, environmental groups showed that a strain of genetically modified corn, which had only been approved for use as animal feed, had been mixed in with corn used in producing human food, like taco shells. To do this, the groups purchased taco shells from stores in the Washington DC area, extracted DNA from the taco shells and used it as a template in a PCR reaction with primers specific for the unauthorized gene (*Cry9C*). Their suspicions were confirmed when they ran this PCR product on an agarose gel and saw a band of size expected for

Cry9C. The PCR test was sensitive enough to detect one transgenic kernel in a whole bushel of corn (1 per 100,000). The company (Aventis) that sold the transgenic seed to farmers had to pay for the destruction of large amounts of corn, and was the target of a class action law-suit by angry consumers who claimed they had been made sick by the taco shells. While no legitimate cases of harm were ever proven, and the plaintiffs were awarded $9 million, of which $3 million went to the legal fees, and the remainder of the judgment went to the consumers in the form of coupons for taco shells. The affair damaged the company, and exposed a weakness in the way the genetically modified crops were handled in the United States at the time.

PCR can be also used in medical diagnostic tests for detection of pathogens in blood, tissues and body fluids. More recently PCR has been used in the genotyping of patients to match their care with specific treatments for better outcomes.

PCR is also used for DNA genotyping of biological samples in forensic or criminal investigations. People can be genotyped for identification purposes, so as to match with samples present at a crime scene or to establish family relationships in paternity/maternity cases. Genotypes also establish identity of people for future comparisons, much like taking fingerprints.

_____

Summary:

- Molecular biology involves the isolation and analysis of DNA and other macromolecules

- Isolation of total genomic DNA involves separating DNA from protein and other cellular components, for example by ethanol precipitation of DNA.

- PCR can be used as part of a sensitive method to detect the presence of a particular DNA sequence

- PCR can also be used as part of a method to isolate and prepare large quantities of a particular DNA sequence

- qPCR methodology allows the quantity of DNA product to be measured.

- RT-PCR methodology detects the quantity and quality of the mRNA, which indicates the spatial and temporal level gene expression.

Key terms:

classical genetics	thermalcycle
molecular biology	denature
molecular genetics	anneal
macromolecules	extention
lysis	thermostable
detergent	Taq DNApol
chelating agent	electropheretic agarose gel
EDTA	fluorochrome
nuclease	Reverse Transcriptase PCR (RT-PCR)
supernatant	Temporal level
pellet	Spacial level
PCR	StarLink affair
primer	*Cry9C* gene

STUDY QUESTIONS:

1) What information, and what reagents would you need in order to use PCR to detect HIV in a human blood sample?

2) If you started with 10 molecules of double stranded DNA template, what is the maximum number of molecules you would you have after 10 PCR cycles?

3) What is present in a PCR tube at the end of a successful amplification reaction? With this in mind, why do you usually only see a single, sharp band on a gel when it is analyzed by electrophoresis?

Chapter 32 – Observing Intact Chromosomes
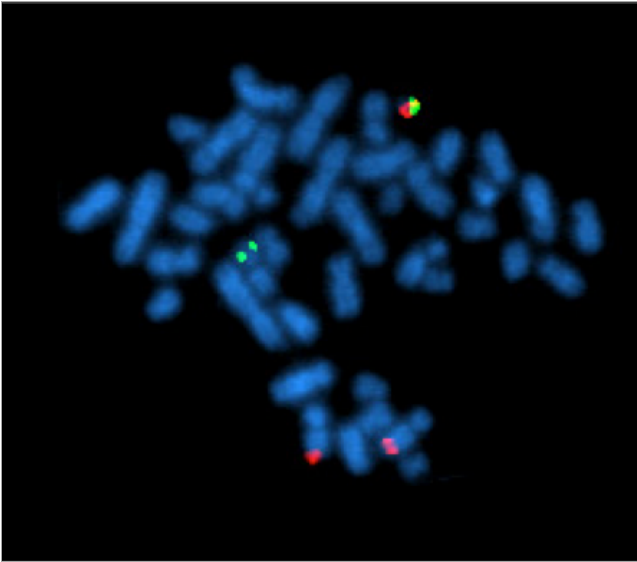
Figure 1.
Fluorescence *in situ* hybridization of mitotic chromosomes from a human cancer cell.
(Wikipedia - Pmx- CC BY-SA 3.0)

Introduction

A lot of information can be obtained from a visual observation of human chromosomes. This chapter will discuss two techniques: bright field microscopy and fluorescence *in situ* hybridization (**Figure 1**). While all the examples come from humans these methods can be applied to any organism.

1. Bright Field Microscopy

1.1. Making a metaphase chromosome spread

The most commonly observed chromosomes are those from white blood cells in the metaphase stage of mitosis. The protocol is as follows:

Step 1 Obtain a sample of whole blood from a person and add to culture medium.

Step 2 Add lymphocyte growth factor proteins to stimulate white blood cells. (Remember, red blood cells lack nuclei and chromosomes.) After three days, the cells have reproduced several times and become more numerous and many are in the process of mitosis.

Step 3 Add colcemid, a Microtubule inhibitor. The cells continue through the cell cycle and will arrest in metaphase. This is because a cell can't enter anaphase without functional Microtubules. More white blood cells will arrest in metaphase. Metaphase cells are best because their chromosomes are condensed and there are no nuclear envelope to get in the way.

Step 4 Transfer the cells to a hypotonic environment. Water enters the cells and they swell up and become delicate. Fix the cells with a mix of acetic acid and methanol.

Step 5 Drop the solution containing the cells onto a glass slide. The cells burst on contact and leaving behind clusters of chromosomes for each cell.

Step 6 Soak the slide in a chromosome staining solution. If **Giemsa** is used by itself the chromosome become a uniform dark purple colour. If Giemsa and Trypsin are used together the chromosomes take on dark purple and light purple bands. This pattern of Giemsa-dark and Giemsa-light bands is consistent and can be used to identify chromosomes and chromosome regions. These protocols are known as **Giemsa staining** and **G-banding**, respectively.

Step 7 Observe the slide with a bright field microscope.

Figure 2 shows an example of Giemsa stained chromosomes. These images are called **metaphase chromosome spreads** because each cell's chromosomes are randomly displayed on the surface of the slide.

1.2. USING BRIGHT FIELD MICROSCOPY TO DIAGNOSE DOWN SYNDROME

Recall that Down syndrome is usually due to a person having three copies of chromosome 21, a situation known as trisomy-21. If a newborn has the physical and mental properties suggestive of Down syndrome a physician will likely order a chromosome test. A cytogeneticist will take a blood sample and make a slide of metaphase spreads. Each spread would show 47 chromosomes in total for a person with trisomy-21. Chromosome 21 can be recognized by its characteristic length, centromere location, and Giemsa banding pattern. **Idiograms** are maps showing this information (**Figure 3**).

To confirm that it is in fact trisomy-21 at least one of these spreads will have its chromosomes arranged into a **karyogram** pattern (see the Chapter 15 on Human Chromosomes). **Figure 4** shows what the cytogeneticist is looking for.
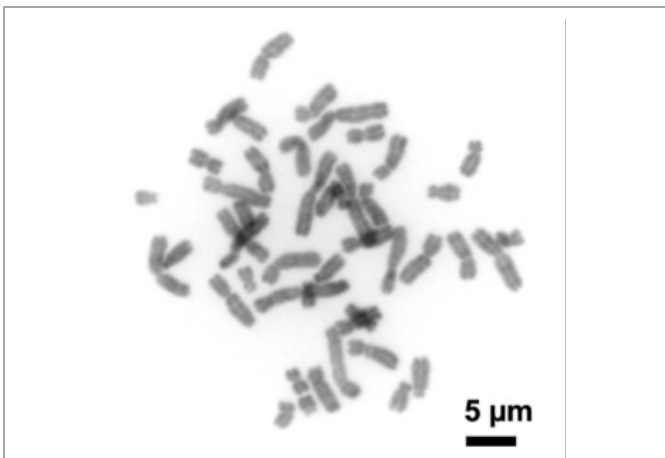
Figure 2.
A human metaphase chromosome spread. This image shows the 46 chromosomes that came from a single cell. There will be dozens of collections of chromosomes like this over the entire slide.
(Wikipedia-Steffen Dietzel-CC BY-SA 3.0)

This karyogram is made with G-banded chromosomes. The extra chromosome is indeed number 21 so this person has a **karyotype** of 47,XY,+21.

This table summarizes the terms introduced in this section. While these are the definitions proposed by the International Standing Committee on Human Cytogenetic Nomenclature, some people use the terms interchangeably.

Bright field microscopy has its limitations though - it only works with mitotic chromosomes and many chromosome rearrangements are either too subtle or too complex for even a skilled cytogeneticist to discern. Even with a more powerful phase contrast or DIC microscope there are limits to what can be seen using G-banding. There is a more powerful technique, one based upon hybridization probes, the topic of the next section.
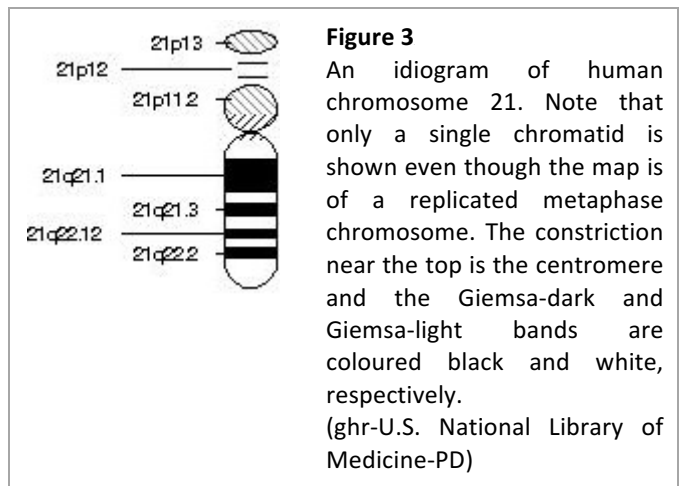
Figure 3
An idiogram of human chromosome 21. Note that only a single chromatid is shown even though the map is of a replicated metaphase chromosome. The constriction near the top is the centromere and the Giemsa-dark and Giemsa-light bands are coloured black and white, respectively.
(ghr-U.S. National Library of Medicine-PD)
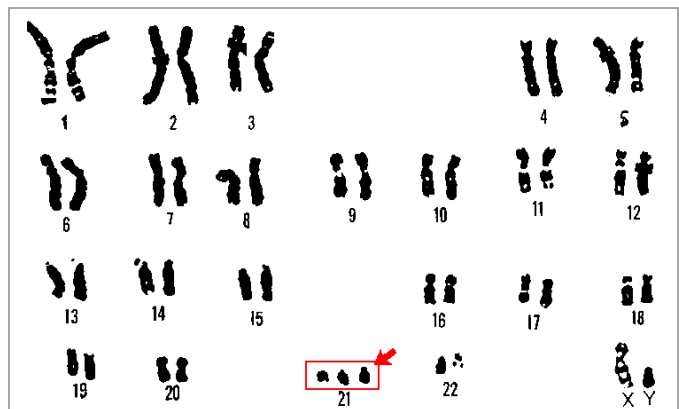
Figure 4.
A karyogram from a male with Down syndrome / trisomy-21. (Wikipedia- U.S. Department of Energy Human-PD)

Term	Definition
metaphase chromosome spread	a picture of all the chromosomes from a cell as they appear on the slide
karyogram	a picture of all the chromosomes from a cell rearranged into the standard pattern
idiogram	a map of one or more chromosomes showing its Giemsa banding pattern
karyotype	a written description of a person's chromosome composition.

2. HYBRIDIZATION PROBES

2.1. WHAT HYBRIDIZATION PROBES ARE

If you have a large amount of DNA, for example an entire chromosome or a large collection of restriction fragments, how could you identify a single gene? One method is to use a **hybridization probe**. These are a collection of short pieces of single stranded DNA that can bind to the target gene. Both the probe and target need to be single stranded so they can pair using complementary base pairing (As with Ts and Gs with Cs). In short, the procedure is to denature the target DNA, add the probe, and then detect where the probe has stuck (**Figure 5**). This section will discuss making probes and using them. Molecular geneticists use hybridization probes in different protocols. Later in this chapter we will cover **fluorescence *in situ* hybridization** (**FISH**) while Chapter 34 covers Southern blotting.

2.2. MAKING HYBRIDIZATION PROBES WITH A RANDOM PRIME LABELLING REACTION

There are a few methods to make hybridization probes but the most common is a random prime labelling reaction. Essentially, we use DNA Polymerases to make our probe for us in a microcentrifuge tube. The reaction contains the following:

- Template DNAs. This can be a **plasmid**, a **PCR product**, or a **BAC** (bacteria artificial chromosome). The template DNA must be denatured to make it single stranded first.

- DNA Polymerases. *E. coli* DNA Pol works well.
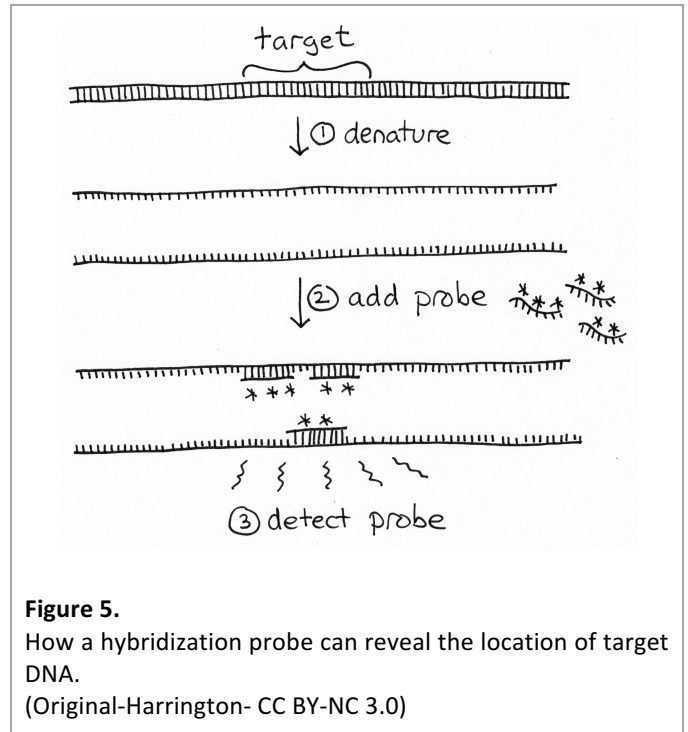
Figure 5.
How a hybridization probe can reveal the location of target DNA.
(Original-Harrington- CC BY-NC 3.0)

- Random oligonucleotides ("oligos") These are millions of different pieces of single stranded DNA. They are made synthetically and purchased as a mixture.

- Regular DNA nucleotides. For example dATP, dGTP, and dTTP.

- Labelled DNA nucleotides. For example [Cy3]-dCTP

You have to provide the template DNA but each of the other components can be purchased from Biotechnology companies. The only expensive component are the **labelled nucleotides**; of which there are many types (**Figure 6**). Radioactive ones are the best way to do Southern blotting and fluorescent ones are at the heart of FISH.

Figure 6.
Two examples of labelled nucleotides.
(Original-Harrington- CC BY-NC 3.0)

Figure 7.
A random prime labelling reaction can use a BAC to make a hybridization probe. The probe will hybridize to the chromosome region originally used to make the BAC.
(Original-Harrington- CC BY-NC 3.0)
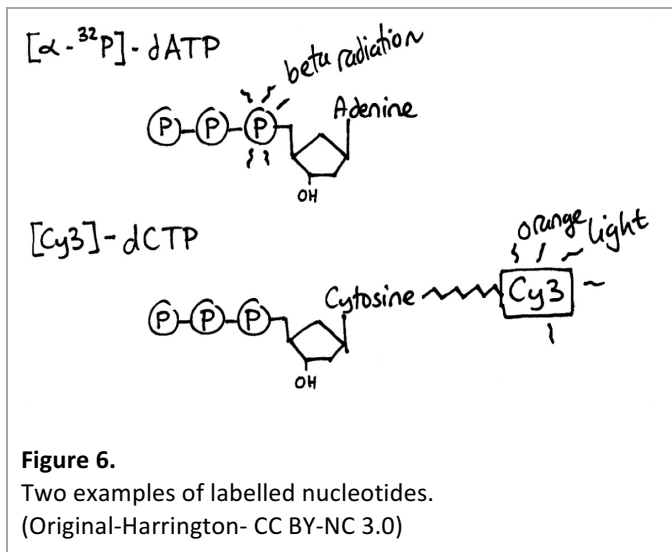
The reaction starts when the mixture is placed at 37°C. The oligos bind randomly to the denatured template DNA and act as primers. From these primers the DNA Pols make new DNA strands. Because some of the nucleotides are labelled the new DNA strands are labelled as well. After an hour or so the reaction is halted. When it is time to use the probe it is denatured. The easiest way to denature it is 5 minutes at 100°C and 5 minutes on ice. What we call the probe are the new DNA strands, each is about 100 nucleotides long, each contains several labelled nucleotides, and each is complementary to the DNA used as the template.

2.3. USING HYBRIDIZATION PROBES
Figure 7 summarizes the random prime labelling reaction. How we make our probe will depend upon what we are trying to detect. The typical sizes of the various template DNAs we can use are:

- PCR product - up to 5000 bp
- plasmid - inserts can be up to 15 000 bp
- BAC - inserts can be up to 350 000 bp

If we want to detect a single gene a hybridization probe made from a plasmid or PCR product will suffice. But if we want to detect a large chromosome region we have to use BACs. Sometimes several BACs are used if we want to detect an entire chromosome. BACs for humans and other commonly studied organisms can be purchased from Biotechnology companies
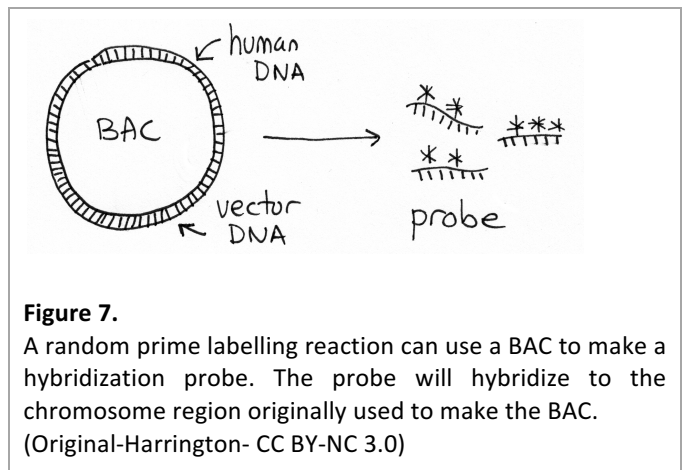
This raises a very important point. No matter what the *template* DNA was, the *probe* DNA will be short pieces of single stranded DNA, typically only 100 nucleotides long. Think of a probe as a cloud of tiny DNA molecules that can stick along the length of a much larger target region. Chapter 34 describes how hybridization probes are used in Southern blotting while the next section describes their use in FISH.

3. FLUORESCENCE IN SITU HYBRIDIZATION (FISH)

3.1. HOW FISH WORKS
The solution to the lack of resolution in G-banding is 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 Chapters 30 & 34). 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 coated 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 as a large blue sphere. In either case the results are observed with either a standard or confocal fluorescence microscope.

3.2. 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 maternal 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, factor is present the woman may choose to undergo an amniocentesis test. In this test, some amniotic fluid is withdrawn so that the fetal cells within it can be examined. **Figure 8** shows a positive result for trisomy-21. This diagram is based upon actual results. The colours are:

- Blue. The DNA has been stained with DAPI.
- Red. This hybridization probe binds to the centromere of chromosome 21.
- Green. This hybridization probe binds to the centromere of the X chromosome.

Based upon the available information this fetus has two X chromosomes and three chromosome 21s and therefore has a karyotype of 47,XX,+21.
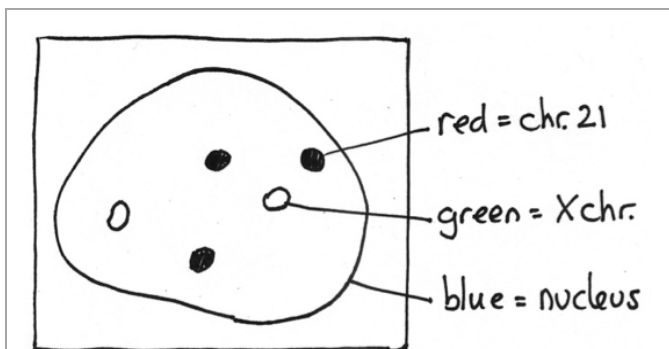
Figure 8.
Confirmation of Down syndrome in a fetus using amniocentesis and FISH.
Based upon: Antonarakis, S. E. et al. 2004. Chromosome 21 and Down syndrome: From genomics to pathophysiology. Nature Reviews Genetics 5:725-738 PubMed ID: 15510164. (Original-Harrington- CC BY-NC 3.0)

3.3. 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** shows 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). There are three colours shown in the diagram:

- Blue. The DNA has been stained with DAPI .
- Green. This hybridization probe binds within the short arm of chromosome 5. This region is absent in people with Cri-du-chat syndrome.
- Red. This is hybridization probe binds within the long arm of chromosome 5. It is used to identify chromosome 5.

The results show both chromosome 5s have intact long arms but one is missing part of its short arm. This child has the karyotype 46,XX,del(5), indicative of Cri-du-chat syndrome.
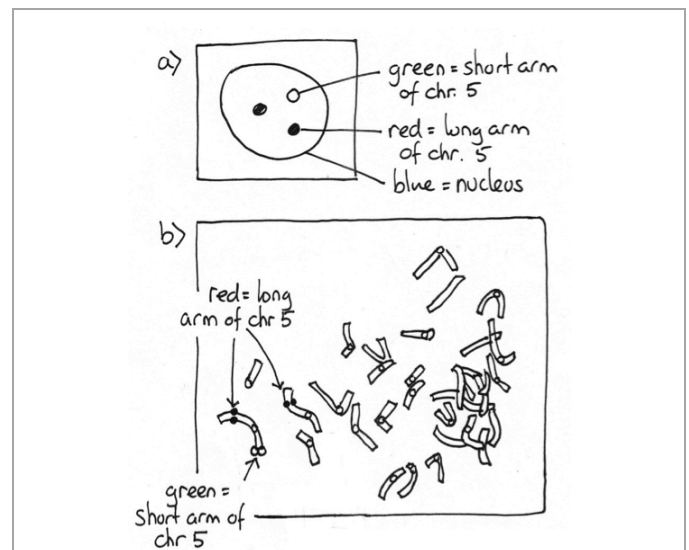
Figure 9.
Confirmation of Cri-du-chat syndrome in a child using interphase cells, metaphase cells, and FISH. Based upon: Fang J.-S. et al. 2008. Cytogenetic and molecular characterization of three-generation family with chromosome 5p terminal deletion. Clinical Genetics 73:585-590 PubMed ID: 18400035 (Original-Harrington- CC BY-NC 3.0)

3.4. 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 DNA sequences (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:

- Human chromosomes can be observed in either metaphase chromosome spreads or within intact nuclei.

- DNA stains such as Giemsa and DAPI bind to DNA non-specifically.

- Hybridization probes bind to DNA at specific target sites. They are collections of DNA molecules that are (i) short, (ii) single stranded, (iii) contain a labelled nucleotide, and (iv) are complementary to a target region.

- Chromosomes can be prepared with either Giemsa staining or G-banding and then observed with a visible light microscope.

- Chromosomes can be prepared with both DAPI staining and fluorescently labelled hybridization probes and then observed with a fluorescence microscope.

- Chromosome number or structural abnormalities can be recognized in a metaphase chromosome spread, an intact nucleus, or a karyogram diagram. They can be summarized in a karyotype statement.

KEY TERMS:

Giemsa	hybridization probe
Giemsa staining	fluorescence *in situ* hybridization (FISH)
G-banding	plasmid
metaphase chromosome spread	PCR product
idiogram	bacteria artificial chromosome (BAC)
karyogram	labelled nucleotide
karyotype	DAPI

QUESTIONS:

1) Giemsa and DAPI are both used to label DNA. Why can't we use only Giemsa or only DAPI in human cytogenetics?

2) What changes would you have to make to the karyogram in **Figure 4** to make it show the chromosomes from the patient with Cri-du chat syndrome described in **Figure 9**?

3) What are the similarities and differences between a PCR reaction and a random prime labelling reaction?

4) In nucleotide triphosphates, the phosphates are named alpha, beta, and gamma. In **Figure 6** why is it the alpha phosphate that is radioactive?

5) What would **Figure 8** look like if it also showed metaphase chromosomes from another cell?

6) Some men have an extra Y-chromosome. What is their karyotype? Describe as many ways as you can to detect this chromosome abnormality.

7) Some women have an extra X-chromosome. What is their karyotype? Describe as many ways as you can to detect this chromosome abnormality.

8) Design a FISH based experiment to find out if someone is a 47,XXX female or a 47,XYY male.

Chapter 33 – DNA Sequencing
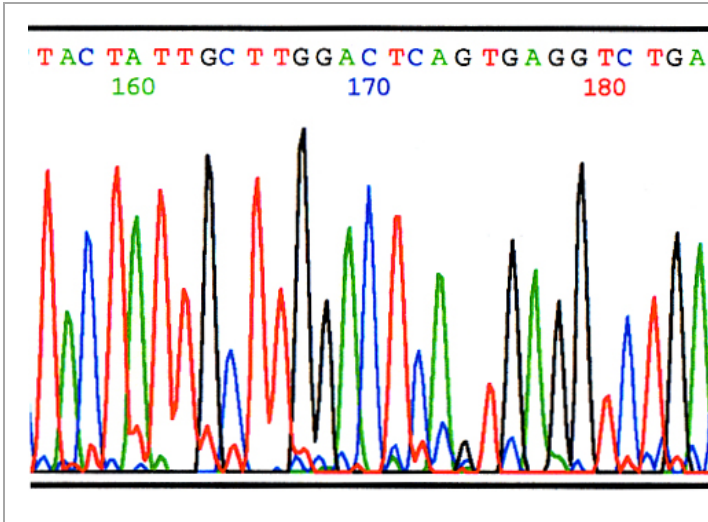
Figure 1.
Output from an automated Sanger DNA sequencer.
(Original-Harrington- CC BY-NC 3.0)

Introduction

DNA sequencing determines the order of nucleotide bases for a DNA molecule. These DNA molecules could be as small as a single restriction fragment, an entire gene, or as large as an organism's entire genome. Most DNA sequencing at the University of Alberta is done by the Molecular Biology Service Unit (MBSU). They use three machines: (1) an Applied Biosystems ABI 3730, (2) an Illumina MiSeq, and (3) an Illumina NextSeq 500, each has its own advantages and purposes. The 3730 uses an older technology called automated Sanger sequencing, while the Illumina machines perform next-generation DNA sequencing. This chapter will cover how these machines work and what they are used for.

1. Automated Sanger DNA Sequencing

1.1. Historical context
DNA sequencing has had a long history. Beginning in the 1970s there have been many methods and improvements. Some dates that stand out are:

- 1977 - Frederick Sanger invents a popular method, later called manual Sanger sequencing.

- 1986 - Leroy Hood improves upon this method to invent **automated Sanger sequencing**.

- 1987 - **Applied Biosystems** begins selling a machine to perform automated Sanger sequencing, their ABI 370.

- 1995 to 2003 - Using ABI 370s, ABI 377s, and similar machines scientists in the US, UK, and other countries sequenced the human genome.

- 2002 - Applied Biosystems begins selling the **ABI 3730** (**Figure 2**) which became the most popular way to do automated Sanger sequencing and remains so to this day.

Figure 2.
The Applied Biosytems ABI 3730 in the MBSU (Molecular Biology Service Unit, Biological Sciences Department, U. of Alberta).
(Original-Harrington- CC BY-NC 3.0)
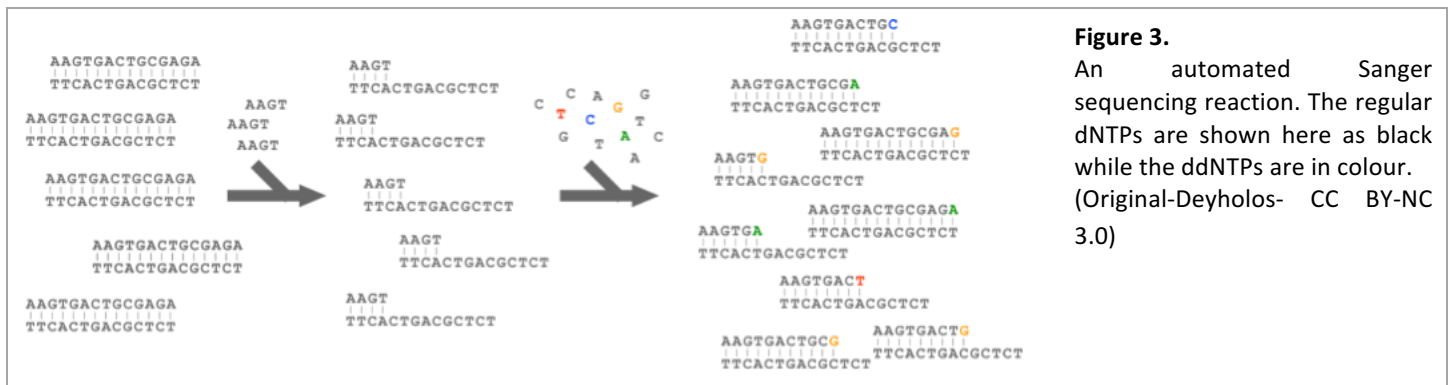
Figure 3.
An automated Sanger sequencing reaction. The regular dNTPs are shown here as black while the ddNTPs are in colour. (Original-Deyholos- CC BY-NC 3.0)

1.2. HOW AUTOMATED SANGER DNA SEQUENCING WORKS

Recall that **DNA Polymerases** incorporate nucleotides (**dNTPs**) into a growing strand of DNA, based on the sequence of a template strand. DNA Polymerases add a new base only to the 3'-OH group of an existing strand of

DNA; this is why primers are required in natural DNA synthesis and in techniques such as PCR. Automated Sanger sequencing relies on the random incorporation of modified nucleotides called **dideoxy nucleotides (ddNTPs**, **Figure 3**).

These lack a 3'-OH group and therefore cannot serve as an attachment site for the addition of the next nucleotide. After a ddNTP is incorporated into a strand of DNA, no further elongation can occur. The ddNTPs are labelled with one of four fluorescent dyes, each specific for one the four nucleotide bases (**Figure 4**).

To sequence a DNA fragment, you need many copies of that fragment (**Figure 5**). Unlike PCR, Automated Sanger sequencing does not amplify the target sequence and only one **primer** is used. This primer is hybridized to the denatured template DNA, and determines where on the template strand the sequencing reaction will begin. A mixture of **regular dNTPs**, **fluorescently-labelled ddNTPs**, and DNA Polymerase is added to a tube containing the primer-template hybrid. The DNA Polymerase will then synthesize a new strand of DNA until a fluorescently-labelled ddNTP nucleotide is incorporated, at which point extension is terminated. Because the reaction contains millions of template molecules,
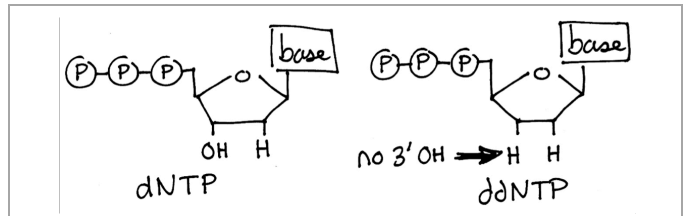
Figure 4.
ddNTPs are synthetic DNA nucleotides that lack a 3' hydroxyl group. If a DNA Polymerase uses one it can't continue.
(Original-Harrington- CC BY-NC 3.0)

ddATP---**GREEN**   ddGTP---**YELLOW**

ddCTP---**BLUE**   ddTTP---**RED**
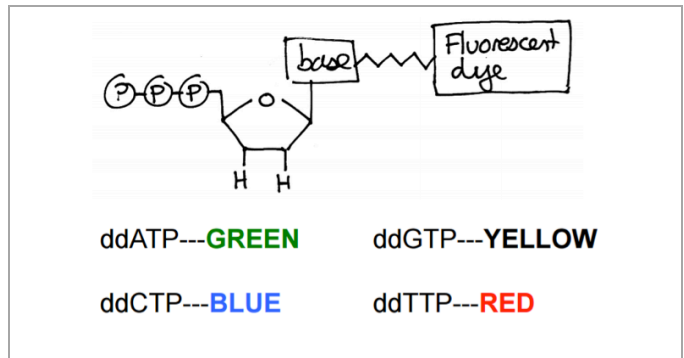
Figure 5.
Each ddNTP used in a sequencing reaction is attached to a different fluorescent molecule using a long chain of carbons. Note that black ink is used to represent yellow fluorescence.
(Original-Harrington- CC BY-NC 3.0)

a sufficient number of shorter molecules is synthesized, each ending in a fluorescent label that corresponds to the last base incorporated. The newly synthesized strands can be denatured from the template, and then separated electrophoretically based on their length (number of bases). The ABI machine is used for this electrophoresis step. While the original, old ABI 370 used a slab gel similar to the ones used in

Figure 6.
Fluorescently labeled products can be separated by capillary electrophoresis, generating a chromatogram from which the sequence can be read.
(Wikimedia-Abizar-PD)
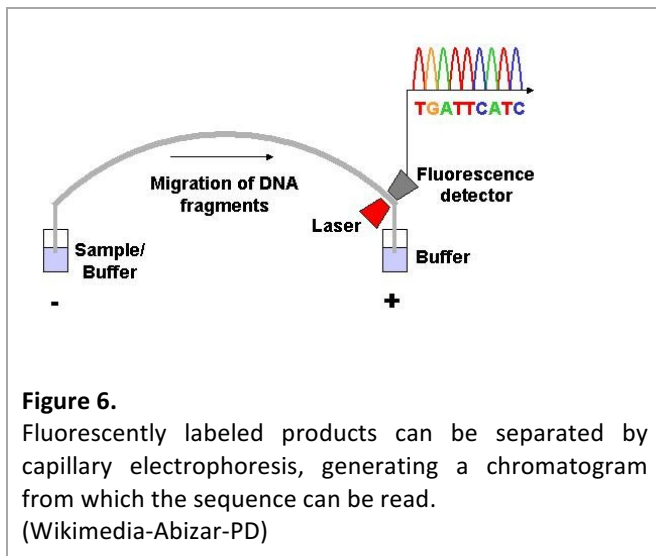
undergraduate labs, the newer ABI 3730 uses **capillary tube electrophoresis** (**Figure 6**). In this machine each sample travels through its own tube. Near the end of the tube is a laser, which excites any fluorescent dyes moving past and a detector that collects any emitted light. As each DNA molecule moves past the laser/detector it emits a specific colour. Because there will be several molecules with the same length and same colour the result appears as a peak of colour. A computer monitoring the results can add the sequence information to the colours since red = T and so on. In this way, the DNA sequence can be read simply from the order of the colors in successive peaks.

The results from a sequencing reaction are presented as a **chromatograph**. While **Figure 6** only shows 9 peaks, a successful sequencing reaction will generate about 700 nucleotides worth of data. The figure shows the results from a single tube but in fact there can be 48 or 96 tubes in total. Thus, in a single "run" an ABI 3730 machine can sequence up to 67,000 bp of DNA.

1.3. Using automated Sanger DNA sequencing to sequence a plasmid

Making a new recombinant plasmid takes time and money. You will want to confirm that it has the DNA sequence it should before you use it for important experiments. A simple way to find out is to sequence it. Let's say you have put a 3.0 kb

insert into a pBluescript II plasmid and right now the recombinant plasmid is in *E. coli* cells (**Figure 7**). The first step is to isolate plasmid DNA from some of the cells with a mini-prep protocol. This will be the template DNA. The primer will be oligonucleotides complementary to the pBluescript II vector adjacent to the insert. The sequencing reaction will tell you the sequence of the insert DNA within the plasmid.

1.4. Using automated Sanger DNA sequencing to sequence a gene

If you suspect that an organism has a mutation in a specific gene you can use automated Sanger sequencing to find out (**Figure 8**). Let's say you have a mouse strain and you think it has a mutation in a gene you are studying. As before, the first step is to isolate DNA. However, we can't sequence this DNA directly. Amongst all of the genomic DNA there just aren't enough copies of the gene to serve as the template DNA. To overcome this limitation a PCR reaction is used to amplify the gene sequence in question. Then we sequence the PCR product. Depending upon how large the gene is it may take several PCR products and several sequencing reactions to get the whole sequence.

Figure 7
Using automated Sanger sequencing to sequence a plasmid.
(Original – Harrington- CC BY-NC 3.0)

Figure 8.
Using automated Sanger sequencing to sequence a gene. The site of a point mutation is shown here as 'm'.
(Original –Harrington- CC BY-NC 3.0)
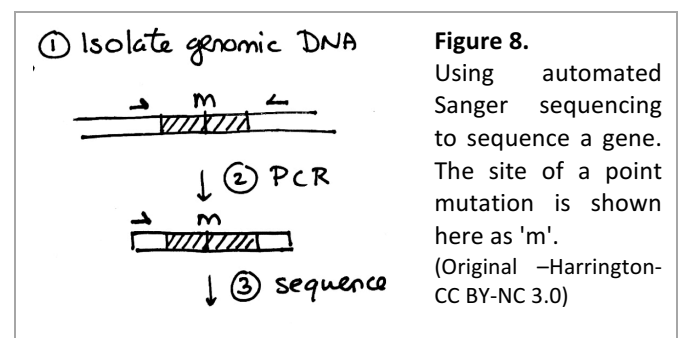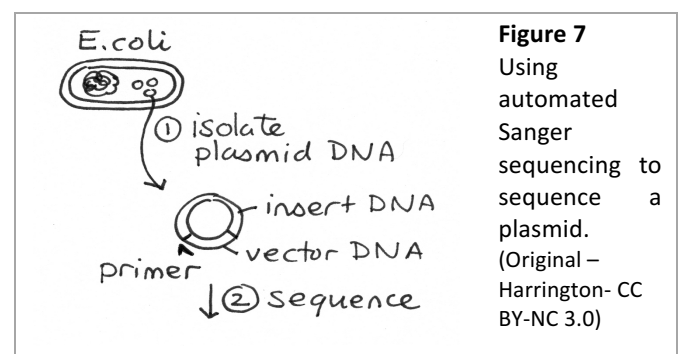
2. NEXT-GENERATION DNA SEQUENCING

2.1. HISTORICAL CONTEXT

Sequencing a single gene or plasmid with an ABI 3730 is quick and inexpensive. But sequencing a whole genome this way would be very slow and very expensive. There are two reasons for this.

The first is that automated Sanger sequencing requires many copies of the template DNA. A sample of purified plasmid DNA or purified PCR product has millions of copies of the target region. But a sample of genomic DNA has only a few copies of a specific target region. For many years, the only way to sequence an organism was to isolate its genomic DNA, break the DNA into large pieces, and then clone these pieces into BAC (bacteria artificial chromosome) vectors. The BAC clones would then have to be sequenced one by one. Most of the 13 years and millions of dollars it took to sequence the human genome was spent making and organizing these BAC clones.

The second limitation of automated Sanger sequencing is that each reaction can only generate 700 nucleotides worth of data. It took literally millions of independent sequencing reactions to sequence the human genome.

Beginning in the late 1990s scientists realized that there was a need for a machine that could sequence genomic DNA directly and with a single reaction. In several instances a technology was invented in a university lab, developed in a small biotechnology company, and then purchased by a larger biotech company. An example of this is:

- 1996 - Swedish scientists invent a completely new way to sequence DNA called pyrosequencing. It is clever but very labour intensive.

- 2000 - An American inventor and entrepreneur, Jonathan Rothberg, refines their technique into automated pyrosequencing.

- 2004 - His company, 454 Life Sciences, markets the first so called **next-generation sequencing** machine.

- 2007 - The largest biotech company in the world, Roche, buys 454 Life Sciences.

- 2015 - 454, now a subsidiary of Roche, continues to develop and sell next-generation machines.

In 2015, there are several choices for next-generation sequencing. For a few hundred thousand dollars you can purchase a GS FLX (made by 454/Roche), an ABI 5500 (Applied Biosystems), or an Ion Proton (Life Technologies). Each is a fancy looking machine that uses a unique and proprietary technology.

2.2. HOW NEXT GENERATION DNA SEQUENCING WORKS

As mentioned in the introduction to this chapter, the MBSU recently purchased two next generation machines: an **Illumina MiSeq** and an **Illumina NextSeq 500** (**Figure 9**).

Both use a similar workflow (**Figure 10**). The scientist has to isolate genomic DNA from an organism (step 1) and then use a kit to break it into small fragments (step 2). The scientist then loads the fragments into the machine and turns it on. Once inside, the DNA fragments are isolated from each other (step 3), amplified in place (step 4), and finally sequenced (step 5). The technology is called **sequencing by synthesis**. Illumina has made animated movies of what happens within their machines:

www.youtube.com/watch?v=HMyCqWhwB8E
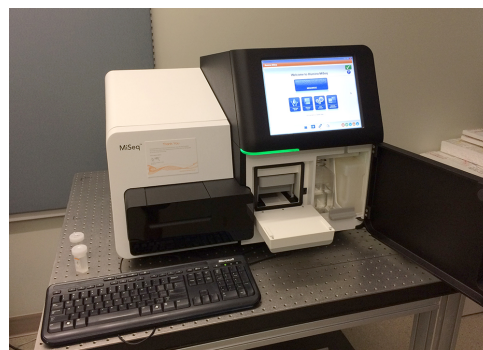
Figure 9.
The Illumina MiSeq in the MBSU (Biological Sciences Department, U. of Alberta). The door has been opened to show where the DNA sample and reaction mixtures are loaded. (Original-Harrington- CC BY-NC 3.0)
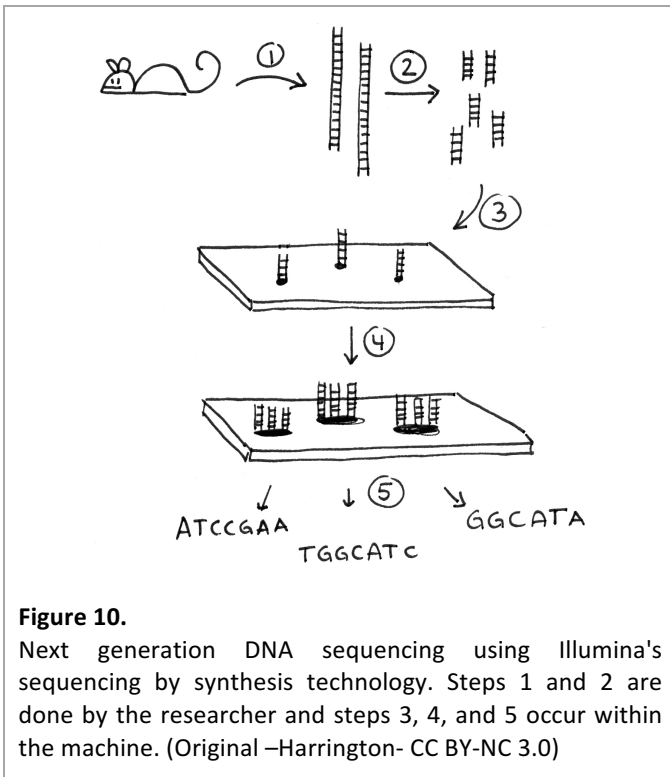
Figure 10.
Next generation DNA sequencing using Illumina's sequencing by synthesis technology. Steps 1 and 2 are done by the researcher and steps 3, 4, and 5 occur within the machine. (Original –Harrington- CC BY-NC 3.0)

Table 1. Comparison between different sequencing machines.

Machine	ABI 3730	Illumina MiSeq	Illumina NextSeq 500
DNA	plasmid or PCR product	genomic DNA	genomic DNA
Technology	automated Sanger	sequencing by synthesis	sequencing by synthesis
Data generated	700 bp	540 Mb to 15 Gb	16 Gb to 120 Gb
Price	$4.75	$1,250 - $1,850	$2,050 - $5,150

The output is just raw sequence data, there are no chromatograms. Powerful software is needed for **sequence assembly**, the process of joining these small pieces of sequence data into a continuous sequence (**Figure 11**). Ultimately there will be one sequence for each of the organism's chromosomes.

2.3. COMPARISON BETWEEN DNA SEQUENCING METHODS

Scientists all over the world now have a choice between automated Sanger sequencing and next-generation sequencing. For example, at the University of Alberta your choices at the MBSU are shown in **Table 1**.

Recall that DNA is measured in base pairs where:

1 kilobase (kb) = 1 000 base pairs (bp)
1 Megabase (Mb) = 1 000 000 bp
1 Gigabase (Gb) = 1 000 000 000 bp

Let's say you wanted to sequence a 2,000 bp long PCR product. You could do this with three sequencing reactions in the ABI 3730 or a single run in the Illumina MiSeq. The first method would cost $15 and the second would cost $1,250. Even though one machine is a decade older it is still the way to go! If you did use the MiSeq you'd end up sequencing the same PCR product over and over. It wouldn't produce any more data.
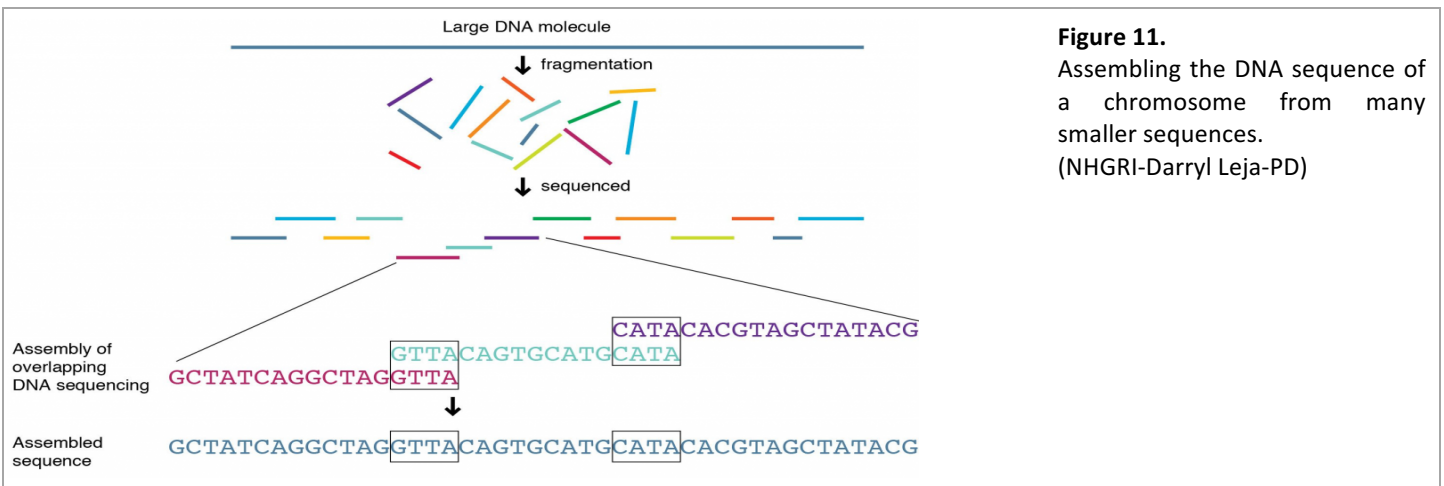
Figure 11.
Assembling the DNA sequence of a chromosome from many smaller sequences.
(NHGRI-Darryl Leja-PD)

On the other hand, let's say you wanted to sequence your own DNA. Even if you don't consider the time and cost of making the BAC clones it would still cost millions of dollars to do all of the sequencing reactions in the ABI 3730. Conversely the MBSU could use their NextSeq 500 and have everything done in two days for about $4 000. Each of your 46 chromosomes would be sequenced about 30 times each. A more expensive machine, the Illumina HiSeq, can sequence human DNA for about $1 000 a person.

2.4. USING NEXT-GENERATION DNA SEQUENCING TO SEQUENCE HUMANS

Even though we know the *average* human DNA sequence, each of us is unique. There are two reasons why human DNA continues to be sequenced. (**Table 2**)

2.5. USING NEXT-GENERATION DNA SEQUENCING TO SEQUENCE OTHER ORGANISMS

Next-generation sequencing has made it feasible to sequence anything. Here are just a few examples. (**Table 3**)

Table 2. Using next-generation sequencing to sequence humans.

Use of next-generation sequencing	Description
Personalized genomics	If we sequence a person's DNA it can reveal information about their susceptibility to disease and their response to various medical treatments.
Tumour cell sequencing	If a person has cancer it is now possible to sequence individual cancer cells. This has revolutionized how physicians help their patients. Instead of treatments based upon the location of tumours, treatments can now be designed around the genetic defects that lead to the cells becoming cancerous in the first place.

Table 3. Using next-generation sequencing to sequence other organisms.

Use of next-generation sequencing	Description
De novo **sequencing**	This is when an organism is sequenced for the first time. For example, in 2014 researchers in Sierra Leone sequenced 99 Ebola virus genomes from 78 patients. They identified changes in the virus that caused the recent outbreak.
Metagenomics	This is when the entire collection of DNA in an environment is sequenced to determine which species are present. This technique has been used to show that a person's gut microbes vary with their diet.
RNA Seq	This is when RNAs from a tissue or organ are isolated, copied into DNA molecules, and then sequenced. This reveals which genes were active in the cell, tissue, or organ.

_____

Sᴜᴍᴍᴀʀʏ:

- Automated Sanger sequencing became commonplace in 1987 and is still used today. It is used to sequence plasmids and PCR products. The most popular machines are Applied Biosystem's ABI 3730s.

- Next-generation sequencing began in 2004 as a better way to sequence whole genomes. There are several competing technologies, for example Illumina's MiSeq machine and its sequencing by synthesis technology.

- Sequencing centres offer both types of sequencing today.

- Sequencing any DNA molecules, large or small, is now fast and inexpensive.

Kᴇʏ ᴛᴇʀᴍs:

automated Sanger sequencing	Ilumina MiSeq
Applied Biosystems ABI 3730	Illumina NextSeq 500
DNA Polymerases	sequencing by synthesis
primer	sequence assembly
regular dNTPs	personalized genomics
fluorescently-labelled ddNTPs	tumour cell sequencing
capillary tube electrophoresis	*de novo* sequencing
chromatogram	metagenomics
next-generation sequencing	RNA Seq

STUDY QUESTIONS:

1) What would the chromatogram look like if you set up an automated Sanger sequencing reaction with only template, primers, polymerase, and fluorescent ddNTPs?

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

3) An alternative name for automated Sanger sequencing is dye-terminator sequencing. Why is this term appropriate?

4) Ten years ago, it would have cost $100,000,000 to sequence your DNA. Today it would cost as little as $1,000. Why did the cost go down so much?

5) Why haven't next-generation machines completely replaced the first generation of automated DNA sequencers?

6) True or false: Automated pyrosequencing and sequencing by synthesis are both considered next-generation DNA sequencing technologies.

CHAPTER 34 – SOUTHERN/NORTHERN/WESTERN BLOTS

Figure 1.
Agarose gel being placed on a Southern blotting transfer set up. The DNA in the gel will be transferred to a membrane, placed on the gel, via the movement of a buffer solution.
(WikimediaCommons-National Cancer Institute-PD)

INTRODUCTION

The separation of DNA, RNA, and polypeptides base on size is a useful biochemical technique that uses migration through a gel to fractionate these macromolecules. For example, bands of DNA in an electrophoretic gel form if many 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 very large number of different sized fragments in the digest and thus it will appear as a continuous smear of DNA, rather than distinct bands, on a lane in a gel. With the genomic DNA 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 DNA sequence probe of "**Southern Blot**". In this chapter, we will describe Southern blots, as well as other blotting techniques, such as **Northern Blots** (RNA) and **Western Blots** (protein), that use similar principles to detect those macromolecules. (The Eastern and SouthWestern blots will not be described here.)

1. SOUTHERN BLOT (DNA)

A **Southern blot** (also called a **Southern Transfer** because it more accurately describes the procedure) is named after Ed Southern, who invented it in the mid-1970s. This blotting method is used to identify specific DNA fragments, size-separated by gel electrophoresis, that cross hybridize with a labeled probe (often radio-active). For example, the presence/absence of a particular sized restriction fragment can be identified in a sample of genomic DNA digested with a specific restriction enzyme.

There are multiple steps in the Southern Blot procedure (**Figure 2**). In the first step, DNA is digested with restriction enzymes and separated by agarose gel electrophoresis. Then a sheet of a nylon derivative, Nitro-cellulose, or similar material (**membrane**) is laid under the gel (**Figure 1**). The DNA, in its separated position (bands or smear), is then transferred to the membrane by drawing a buffer solution out of the gel, in a process called **blotting**. At this point the blotted DNA is usually covalently attached to the membrane by briefly exposure to UV light or drying. The transfer to a 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 an alkali solution to **denature** (double stranded made single stranded) the attached DNA, and this is then neutralized. The membrane is added to a **hybridization** solution, containing a small amount of **labeled** single-stranded **probe** DNA that is complementary to a sequence target molecule on the membrane. This probe DNA is labeled using either fluorescent or radioactive molecules. If the hybridization is performed properly, the probe DNA will form a stable duplex only with those DNA molecules on the membrane to which it is complementary. Then, the unhybridized probe is **washed** off leaving the hybridized radioactive or fluorescent signal bound. This remaining signal will appear in a distinct band when appropriately detected (fluorescent or radioactive). The band represents the presence of a particular DNA sequence within the mixture of DNA fragments that is complementary to the probe sequence

The probe's specificity comes via the sequence specific hybridization (requires complementarity). However, variation in hybridization temperature and washing solutions can alter the **stringency** of

.

the probe's hybridization. At maximum stringency (higher temperature, low salt) hybridization conditions, probes will only hybridize efficiently with target sequences that are perfectly complementary (maximum number of hydrogen bonds). At lower stringency (lower temperature, higher salt), probes will be able to hybridize and detect sequences to which they do not match exactly, but have some **mismatch** along 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. One advantage is that Southern blots are able to detecting fragments larger than those normally amplified by PCR. Also, the long DNA probes can detect fragments that may be relatively dis-similar to the original sequence. Applications of Southern blotting will be discussed further in the context of molecular markers in a subsequent chapter. Southern blotting was invented long before PCR, but PCR has replaced blotting in many applications because of its simplicity, speed, and convenience.

Figure 2.
A diagram of Southern blotting. Genomic DNA that has been digested with a restriction enzyme is separated on an agarose gel, and then the DNA is transferred from the gel to a nylon membrane (grey sheet) by blotting. The DNA is immobilized on the membrane, and 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 fragment 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 BY-NC 3.0)

2. NORTHERN BLOT (RNA)

Following the development of the Southern blot, other types of blotting techniques were invented. The **Northern blot** is much like the Southern Blot, but involves the size separation of single stranded RNA in gels similar to that of DNA.

First, RNA molecules are extracted and isolated from a tissue sample or cells. RNA samples are loaded on a lane and separated by size using agarose gel electrophoresis. Because we wish to determine the native length of the RNA transcripts (and because RNA is single stranded) no DNA restriction enzymes are ever used with RNA. As with DNA, the gel is fragile and probes cannot enter the gel matrix, the samples are blotted to a membrane with a positive charge (nucleic acids have negative charge). In the hybridization step, single stranded DNA or RNA probes are used in order to detect the RNA of interest. Furthermore, transfer buffer often contains formamide that has an ability to lower the temperature of probe hybridization temperature since high temperature might damage the RNA.

Most RNA is single stranded and can fold into various conformations thorough intra-molecular base pairing, so the electrophoresis separation is more haphazard and the bands are often less sharp, compared to that of double stranded DNA. Using northern blot, one can observe the size (quality) and amount (quantity) of transcription of a gene. Thus a pattern of gene expression can be defined by comparing samples from different tissues.

3. WESTERN BLOT (PROTEIN)

In a **Western blot,** polypeptides are size separated on an acrylamide gel before transferring to a membrane. Acrylamide is used because it separates the smaller polypeptide molecules better. The membrane is then probed, not with DNA or RNA, but with an antibody that specifically binds to an antigenic site on the target protein (**primary antibody**). The unbound antibody is washed away and the bound antibody is then detected by a **secondary antibody** with some fluorescent or
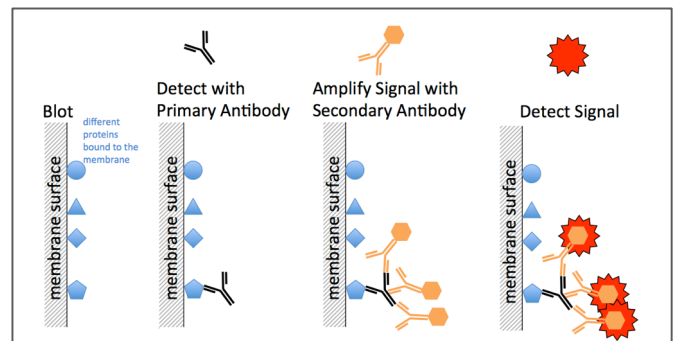
Figure 3.
Antibodies binding to the target protein on a Western Blot. Proteins on a membrane surface (left) can bind the primary antibody (centre-left). Then, secondary antibodies can bind to the primary (centre-right). Finally, the secondary antibodies can be detected by various means (right). This will provide protein specific localization to produce a band on a Western Blot. (Original-J. Locke-CC BY-NC 3.0)]
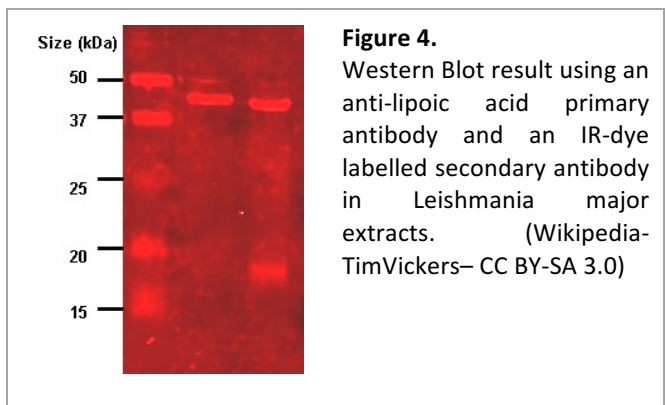
Figure 4.
Western Blot result using an anti-lipoic acid primary antibody and an IR-dye labelled secondary antibody in Leishmania major extracts. (Wikipedia-TimVickers– CC BY-SA 3.0)

colour production detection system. Western blots can also give bands proportional to the size and amount of the target protein. See (**Figure 3.Figure 4**)

One application of western blot application is HIV test. First, cells that may be infected by HIV virus are extracted and cellular proteins are separated that might contain the viral protein. They are then run on the electrophoresis gel and are transferred to a membrane. Antibodies that will bind to HIV viral proteins are added to the membrane to check the presence of the viral proteins.

A comparison of all three blotting methods is shown in **Figure 5.**
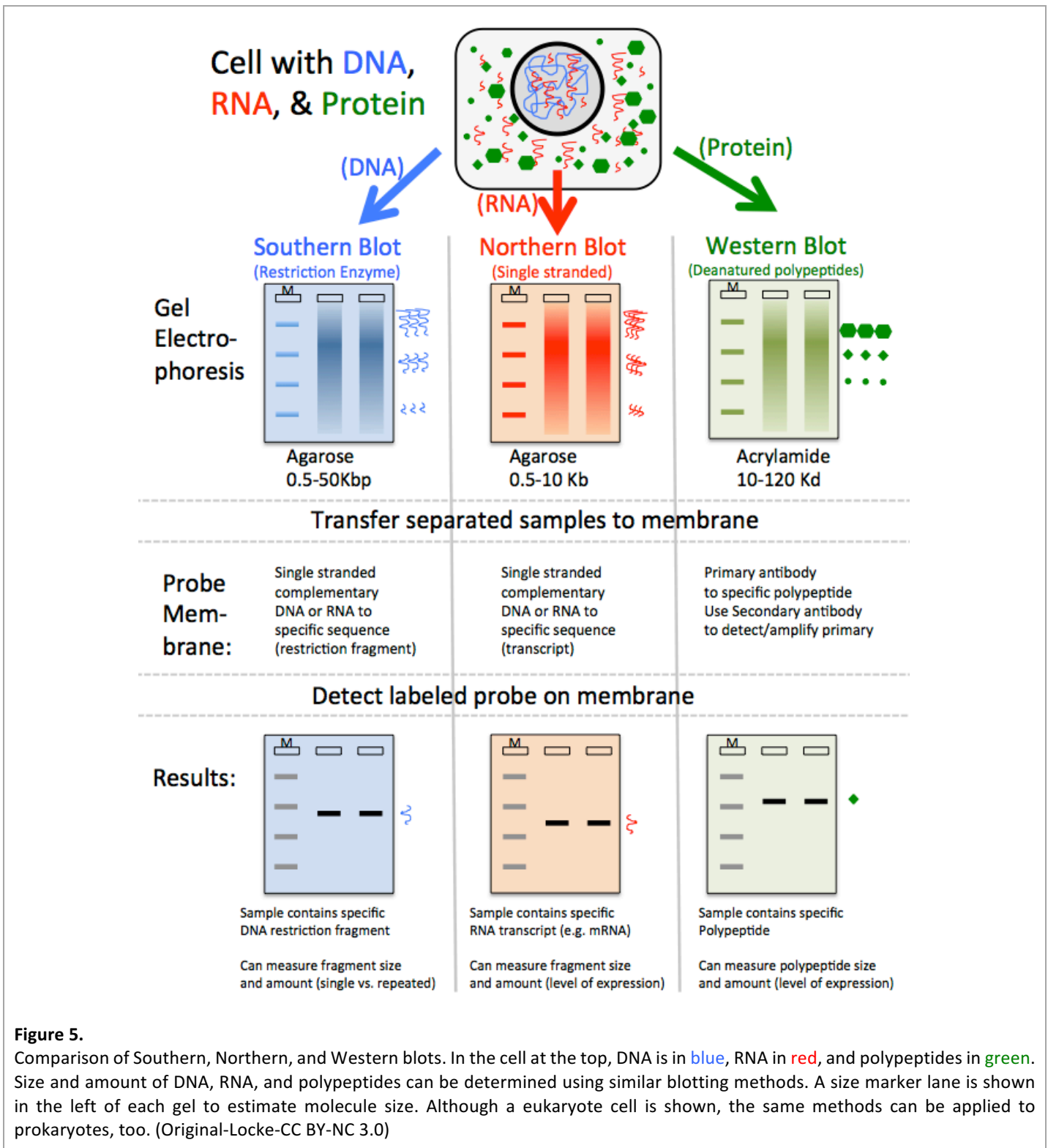
Figure 5.
Comparison of Southern, Northern, and Western blots. In the cell at the top, DNA is in blue, RNA in red, and polypeptides in green. Size and amount of DNA, RNA, and polypeptides can be determined using similar blotting methods. A size marker lane is shown in the left of each gel to estimate molecule size. Although a eukaryote cell is shown, the same methods can be applied to prokaryotes, too. (Original-Locke-CC BY-NC 3.0)

_____

SUMMARY:

- Southern blotting involves detecting the presence of DNA fragments, such as those from total genomic DNA digested with a restriction enzyme, separated by agarose gel electrophoresis and transferred to a membrane that is then probed with a labeled nucleic acid probe.

- With Northern Blots, the same principle in Southern blotting is used to detect single stranded RNA of interest

- Western Blots are also similar but use acrylamide gels to separate proteins and the membrane is probed with antibodies to detect the molecule of interest.

KEY TERMS:

Southern Blot	hybridization
Southern transfer	probe
Northern Blot	washing
Western Blot	stringency
membrane	mismatch
blotting	primary antibody
denaturation	secondary antibody

STUDY QUESTIONS:

1) 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 – it's dominant.
   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 (low stringency)?

2) Refer to question 1.
   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?

3) 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.
   a) How can you use your PCR fragment and genomic DNA from a dog to find this out?
   b) Do you think dogs have more or less of these genes?

CHAPTER 35 – DNA VARIATION STUDIED WITH SOUTHERN BLOTS

Figure 1.
Map of human migration according to mitochondrial DNA sequencing. Variations are inherited and can be detected using Southern Blots (this Chapter), PCR (Chapter 36), and/or microarrays (Chapter 37). The letters indicate the different haplogroups, while the colours and numbers represent thousands of years before present for the establishment of these populations.
(Wikipedia- Avsa – CC BY-SA 3.0)

INTRODUCTION

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 have some somatic differences). In fact, the genomic sequences of almost any two, unrelated people differ at millions of nucleotide positions dispersed throughout their genomes. 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 appearance, behavior, health, and other characteristics depend in part on these polymorphisms.

Most of these nucleotide differences, however, have no effect at all. They have no effect on gene sequences or expression, because they occur in regions of DNA that neither encode proteins, nor regulate the expression of genes. Nevertheless, these polymorphisms are very useful because they can be used as **molecular markers,** which can be mapped just like typical genetic markers. Molecular markers are more numerous and can be used in medicine, forensics, ecology, agriculture, evolution, and many other fields. In most situations, molecular markers obey the same rules of inheritance that we have already described, and so can be used to create detailed genetic maps with which to identify gene/disease locations though genetic linkage.

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, as discussed above, 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 hemophilia 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 are a particularly useful type of polymorphism for many areas of genetics research. Mutations of DNA sequences can arise in many ways.

2. MOLECULAR MARKERS – SNPS AND VNTRS

2.1. ORIGINS OF SINGLE BASE PAIR POLYMORPHISMS

Polymorphisms can be single base pair differences between or among individuals in a population (**Figure 2**). These are referred to as Single Nucleotide Polymorphisms (SNP) or "SNiPs". They are distributed randomly throughout the genome. SNPs occur about once in every 300 base pairs, on average, which means there are roughly 10 million SNPs in the human genome. SNPs are usually identified by DNA sequencing of multiple individuals and comparing the sequence to find the different bases.

Figure 2.
Single Nucleotide Polymorphisms (SNP). The variant region is marked in blue, and each variant sequence is arbitrarily assigned one of two allele labels. (Original-Deyholos-CC:AN)
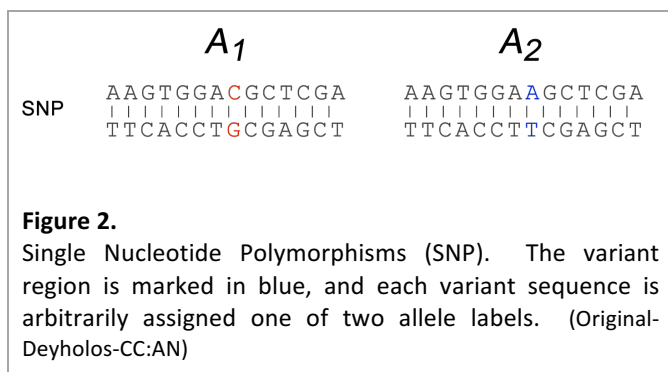
2.2. ORIGINS OF REPETITIVE SEQUENCE POLYMORPHISM

Some of the sequence changes occur during DNA replication, resulting in an insertion, deletion, or substitution of one or a few nucleotides. These replication errors occur most frequently at sequences where tandem repeats already exist. If they are relatively short repeats (1-5 base pairs)

then the expansion and contraction of the number of repeats are called Simple Sequence Repeat (SST) polymorphisms or Short Tandem Repeats (STR). If they are longer (10-50 base pairs) then they are called Variable Number Tandem Repeats (VNTR) (**Figure 3**). The difference in names here is just a matter of length.

Because of the tandem nature of these sequences and their propensity for addition/deletion, the number of repeats is typically very variable in a population of individuals. The number of repeats defines an allele, so there will be many alleles in a population and these loci will be highly polymorphic in a population. This leads to a high degree of heterozygosity in the population, which is good for genetic mapping of these markers.

Figure 3.
Simple Sequence Repeat (SST) polymorphisms or Variable Number Tandem Repeat (VNTR) polymorphism. Each red box represents a repeat. The variant region is marked in blue (increased number of repeats), and each variant sequence is arbitrarily assigned one of two allele labels. (Original-Deyholos-CC:AN)
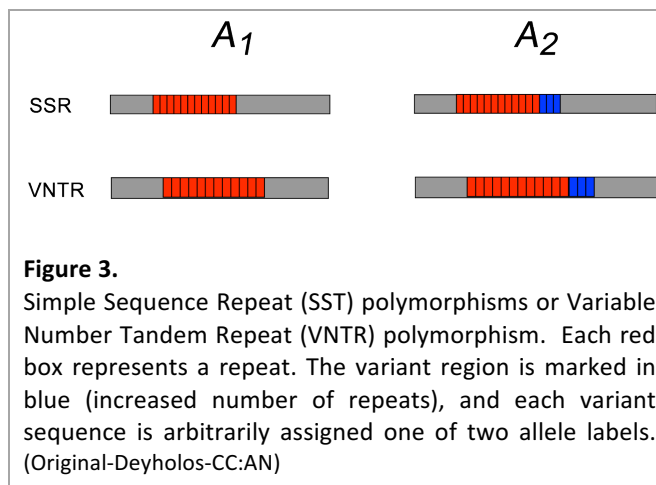
2.3. CLASSIFICATION AND DETECTION OF REPETITIVE SEQUENCE POLYMORPHISM

Repetitive Sequence Polymorphism 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. In both cases, because two or more alternative versions of the DNA sequence exist, and we can detect them, 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 4**). An **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.

3. RESTRICTION FRAGMENT LENGTH POLYMORPHISM (RFLP)

Another form of Molecular Marker is the **Restriction Fragment Length Polymorphism (RFLP)**. This polymorphism takes advantage of differences in the length of restriction enzyme (RE) fragments (**Figure 4**).

Figure 4.
Restriction Fragment Length Polymorphism (RFLP). The variant sequence is marked in blue, and each variant sequence is arbitrarily assigned one of two allele labels. (Original-Deyholos-CC:AN)
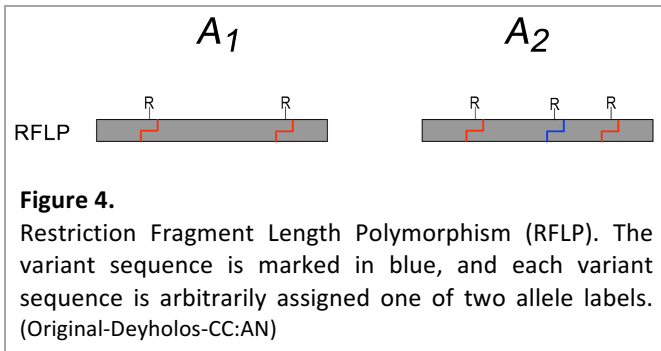
Here, the change in DNA sequence introduces or abolishes a restriction enzyme site. This will change the length of a restriction enzyme fragment that can be detected by Southern Blot of that genomic DNA. While the loss or gain of a RE site is the typical cause of RFLPs, other changes in RE
.

fragment length can be caused by the insertion of mobile genetic elements such as transposons, (inserted more or less randomly into chromosomal DNA) or to DNA deletions or duplications.

4. CONSTRUCTION OF GENETIC LINKAGE MAPS

In classical Mendelian genetics, two loci can be mapped relative to one another – they will either assort independently (unliked) or will be linked and the frequency of recombination will determine their distance apart. Molecular markers can be used in the same manner, both with each other and in combination with classic Mendelian markers, too.

By calculating the recombination frequency between pairs of molecular markers, a map of each chromosome can be generated for almost any organism. These maps are calculated using the same mapping techniques described previously for genes, however, the high density and ease with which molecular markers can be genotyped makes them more useful than the "old-style" visual phenotype method for constructing genetic maps. These more detailed maps are useful in further studies, including map-based cloning of protein coding genes that were identified by mutation, or for disease loci.

Figure 4 diagrammatically shows a set of hypothetical results of parentals and F2 progeny for a mapping cross. This type of experiment is needed to map the relative distance between two loci, A and B, which are part of a series of loci along a chromosome. Then these loci can be used to test for linkage to disease or other traits in a genome.
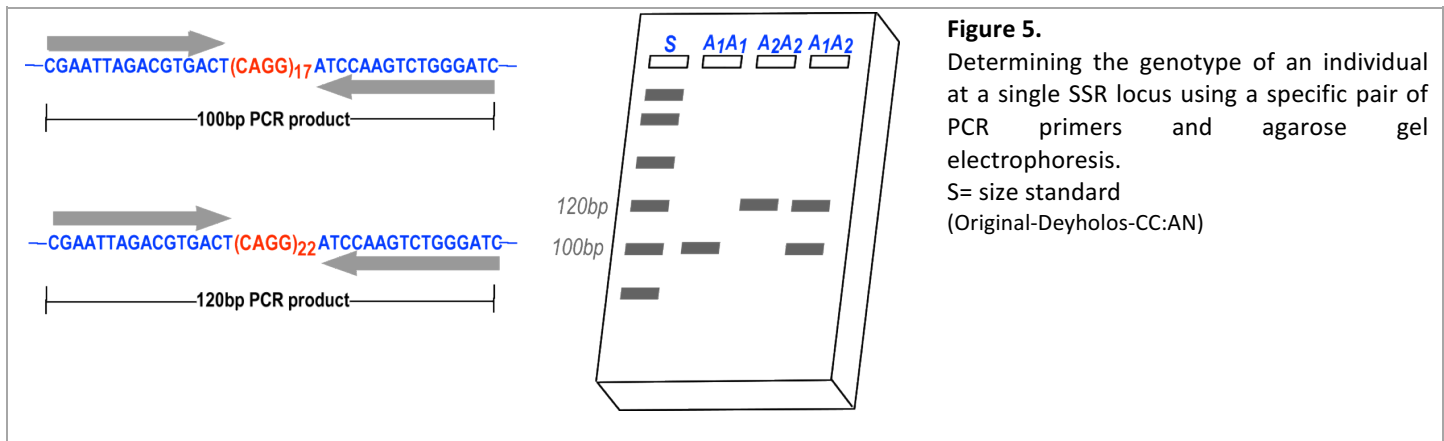
Figure 5.
Determining the genotype of an individual at a single SSR locus using a specific pair of PCR primers and agarose gel electrophoresis.
S= size standard
(Original-Deyholos-CC:AN)
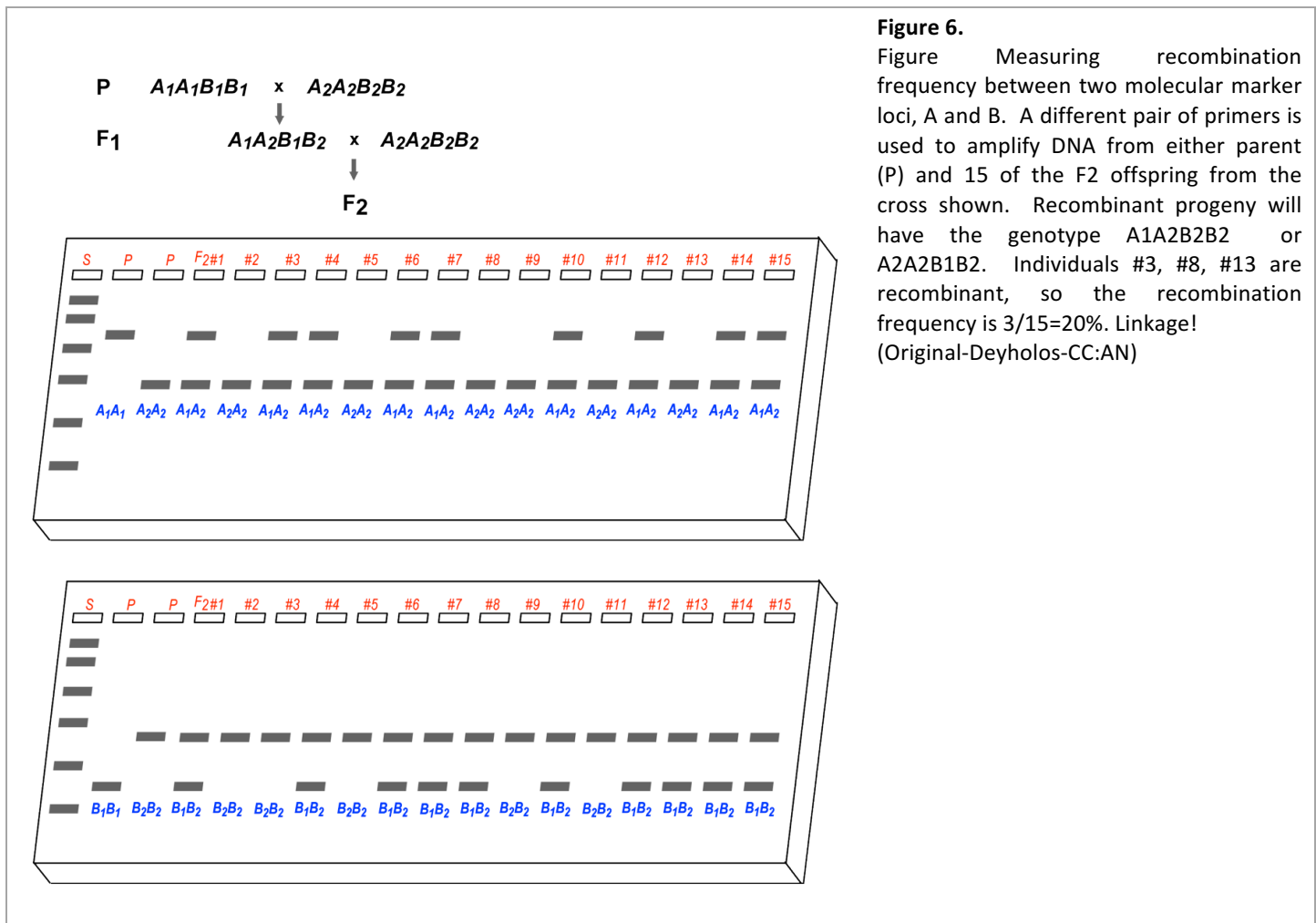
Figure 6.
Figure Measuring recombination frequency between two molecular marker loci, A and B. A different pair of primers is used to amplify DNA from either parent (P) and 15 of the F2 offspring from the cross shown. Recombinant progeny will have the genotype A1A2B2B2 or A2A2B1B2. Individuals #3, #8, #13 are recombinant, so the recombination frequency is 3/15=20%. Linkage!
(Original-Deyholos-CC:AN)

5. APPLICATIONS OF MOLECULAR MARKERS

Several characteristics of molecular markers make them useful to geneticists.
1. They are frequent throughout the genome
2. They are retained in the population (not selected for or against).
3. They are mostly phenotypically neutral.
4. It is relatively easy to find markers that differ between two individuals.
5. There is no worrying about gene interactions or other influences that make it difficult to infer genotype from phenotype.
6. It possible to study hundreds of loci
7. Molecular marker can be detected in any tissue or developmental stage.

8. The same type of assay can be used to score molecular phenotypes at millions of different loci.

9. The loci are co-dominant (both alleles are visible) so both can be tracked in pedigrees.

Thus, the neutrality, high density, high degree of polymorphism, co-dominance, and ease of detection of molecular markers led to their wide-spread adoption in many areas of genetic 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:

5.1. DNA FINGERPRINTING

Just like real finger prints on your fingers, the determination of alleles at genetic loci can be used to make a "**DNA fingerprint**". This is done by determining the allelic genotypes at multiple molecular marker loci to make a composite genotype. Then, by comparison, one can determine the similarity between two DNA samples. If marker genotypes differ, then clearly the DNAs are from different sources. However, if they don't differ, then they could come from the same source. But, there is a possibility that they came from different sources with both having the same genotype at the markers by chance alone. 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 the blood sample found on a weapon and the blood sample from a particular suspect are indistinguishable. Similarly, that cat hair on the suspect's clothing came from a particular cat at the home of a crime scene.

DNA fingerprinting is also useful in paternity testing and in commercial applications such as verification of species of origin of certain foods and herbal products.
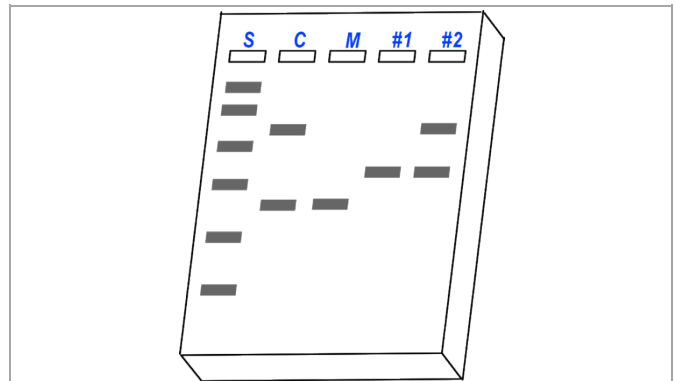
Figure 7.
Paternity testing. Given the molecular phenotype of the child (C) and mother (M), only one of the possible fathers (#2) has alleles that are consistent with the child's phenotype. (Original-Deyholos-CC:AN)

5.2. POPULATION STUDIES

Population equilibrium - As described in Chapter 38, 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.

Ancestry - Molecular markers can also be used by anthropologists to study migration events in human ancestry. There are many commercial businesses available that will genotype people and determine their deep genetic heritage for ~$100-$200. This can be examined through the maternal line via sequencing their mitochondrial genome and through the paternal line via genotyping their Y-chromosome.

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 and his relatives (the haplogroup C, although the specific group varies, depending on the source).

5.3. 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 (non-afflicted) 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. This is covered in more detail in Chapter 37.

Molecular markers may also be used in a similar way in agriculture to track desired traits in crops or livestock. 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.

_____

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.

KEY TERMS:

molecular marker   SNP
repetitive DNA   RFLP
SSR   neutral mutation
SSLP
VNTR

STUDY QUESTIONS:

For the next few questions, suppose that you have a 1.0kbp fragment from the human genome. You are told it contains only unique sequence (no repeated DNA sequences such as transposable elements or *Alu* sequences). If you label this fragment and use it to probe a Southern blot containing human genomic DNA (one person) digested with *EcoRI, HindIII,* and *BamHI* in lanes 2, 3, and 4, respectively. Lane 1 contains a size marker.

1) Will the probe only show one band per lane in DNA from the individual if they are homozygous for the region being probed?

2) What if the individual is heterozygous for this region?

3) What if you examined 100 different individual's genomic DNA in a similar manner, would they all be expected to have the same pattern?

4) What would you expect if the probe was not unique, but instead had an *Alu* repeat within the 1.0kbp fragment?

5)

CHAPTER 36 - DNA VARIATION STUDIED WITH PCR
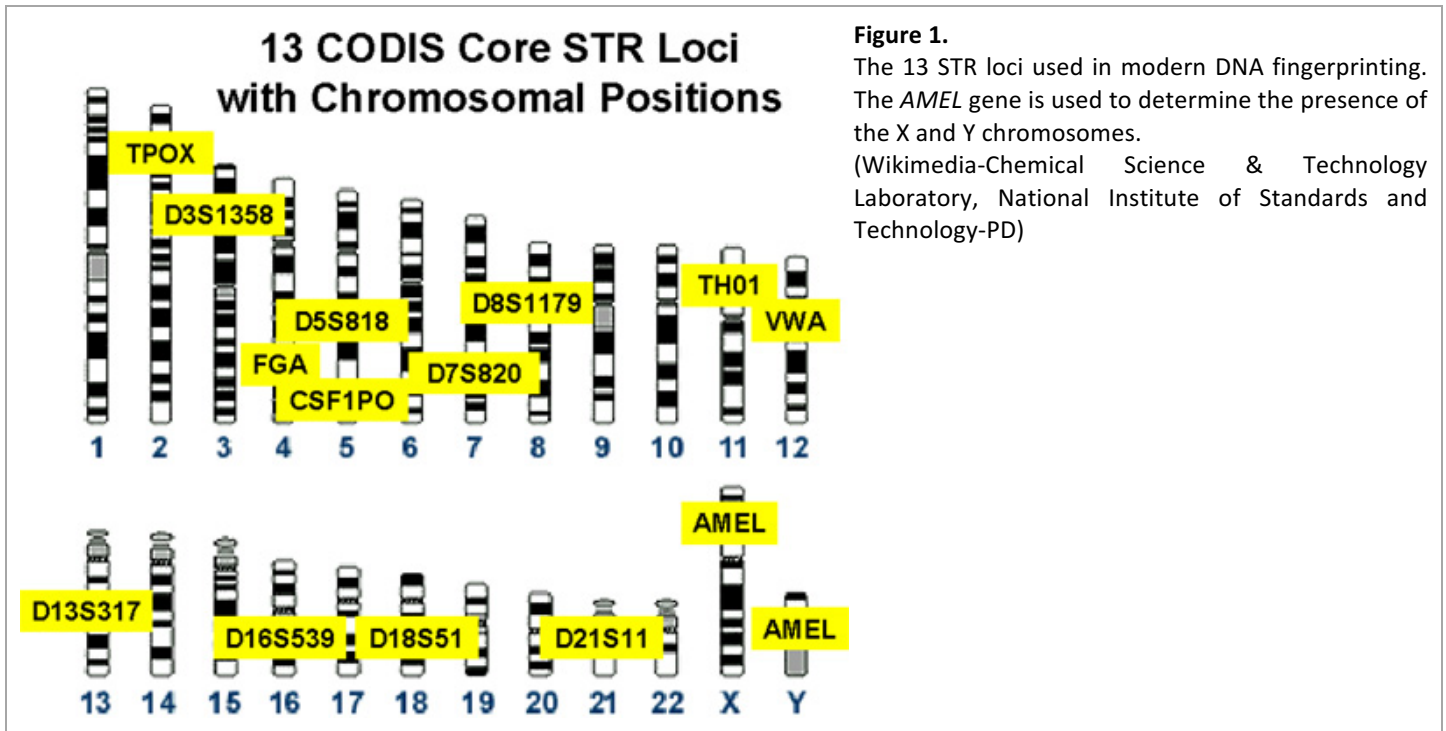
Figure 1.
The 13 STR loci used in modern DNA fingerprinting. The *AMEL* gene is used to determine the presence of the X and Y chromosomes.
(Wikimedia-Chemical Science & Technology Laboratory, National Institute of Standards and Technology-PD)

INTRODUCTION

Modern day DNA fingerprinting is based upon harmless DNA variations present in our chromosomes (**Figure 1**). This chapter will look at what these variations are, how they can be detected using two PCR based methods, and how DNA fingerprinting is used in forensics and paternity testing.

1. SHORT TANDEM REPEATS (STRs)

Here is a short sequence of human DNA within intron 6 of the *CSF* gene on chromosome 5:

––AGATAGATAGATAGATAGATAGATAGAT––
––TCTATCTATCTATCTATCTATCTATCTA––

Note that it has the sequence AGAT repeated seven times. This, and sequences like it, are called **short tandem repeats (STRs)**.

The DNA replication machinery occasionally makes errors at STRs (see Chapter 11). They may expand and end up to eight or more repeats, or they may contract and end up with six or fewer. Because these changes have been happening for tens, if not hundreds of thousands of years, different people in

the population have different numbers of this repeat. Some people have as few as five repeats while others have as many as sixteen. People may be homozygous, and have the same number of repeats on their maternal and paternal chromosomes, or be heterozygous, and have two different numbers. Note that in this case none of these changes affect a person's health, the *CSF* gene is functional no matter how many repeats are present.

This repeat array is named CSF1PO and is described at an online STR database.

http://www.cstl.nist.gov/strbase/str_CSF1P O.htm

CSF1PO is not important for human health but it does play an important role in DNA fingerprinting. Consider what it represents:

- It has twelve distinct alleles (five repeats, six repeats, etc.).

- The chance that two people would have the same genotype (e.g. both being a 7/12 repeat heterozygote) is very small.

- Using **PCR** it is easy to determine the alleles that each person has.

The rest of this chapter will discuss how we can determine which alleles a person has and how we can use these results for DNA fingerprinting.

2. DETECTING STRs WITH PCR AND AGAROSE GEL ELECTROPHORESIS

Let's say we have a person who is a 7/12 heterozygote at the CSF1P0 site. We could find out by isolating their genomic DNA, amplifying the region using standard PCR, and then running the PCR products on an agarose gel. We would get two bands, the faster moving band would be the smaller PCR products from the 7 allele and the slower moving band would be the products made from the 12 allele.

Figure 2 shows testing for a typical STR. These results show the people were genotypically 17/17, 22/22, and 17/22. Other people would have many other possibilities with different numbers of repeats per allele and different combinations of alleles per person.
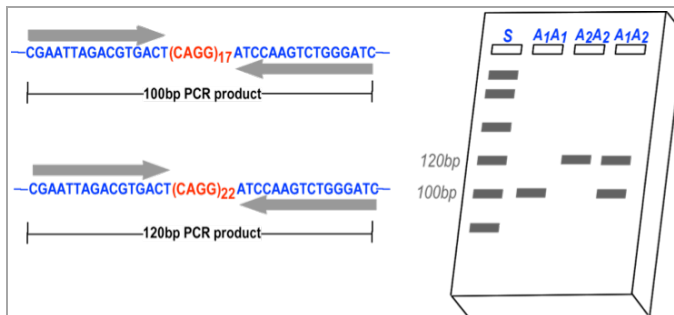
Figure 2.
Determining the genotype of an individual at a single STR using a specific pair of PCR primers and agarose gel electrophoresis. S = size standard.
(Original-Deyholos-CC BY-NC 3.0)

3. DETECTING STRs WITH PCR AND CAPILLARY TUBE ELECTROPHORESIS

The method described above is relatively simple to do, but very labour intensive. The gel has to be poured, loaded, run, and photographed by a technician. We can get the same information using a more automated process, one using capillary tube electrophoresis. Here, one of the two PCR primers already has a **fluorescent dye** (**Figure 3**) attached.

The PCR reaction occurs normally resulting in the PCR products all being fluorescently labelled. They can then be loaded into a **capillary tube electrophoresis** machine (**Figure 4**).

As before, DNA migrates through a gel material towards the positive electrode, but this time the gel is contained within a thin tube. Near the end of a tube is a laser to excite the fluorescent dye and a detector to record fluorescence. A computer can then monitor the tube for the appearance of any fluorescent signal. For an STRs we would get a single peak if a person was homozygous and two peaks if a person was heterozygous
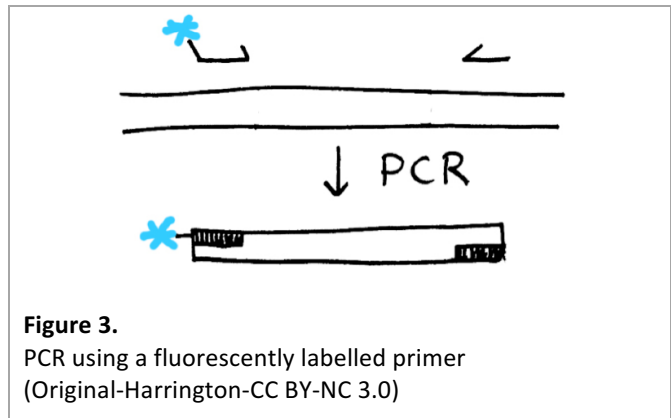
Figure 3.
PCR using a fluorescently labelled primer
(Original-Harrington-CC BY-NC 3.0)
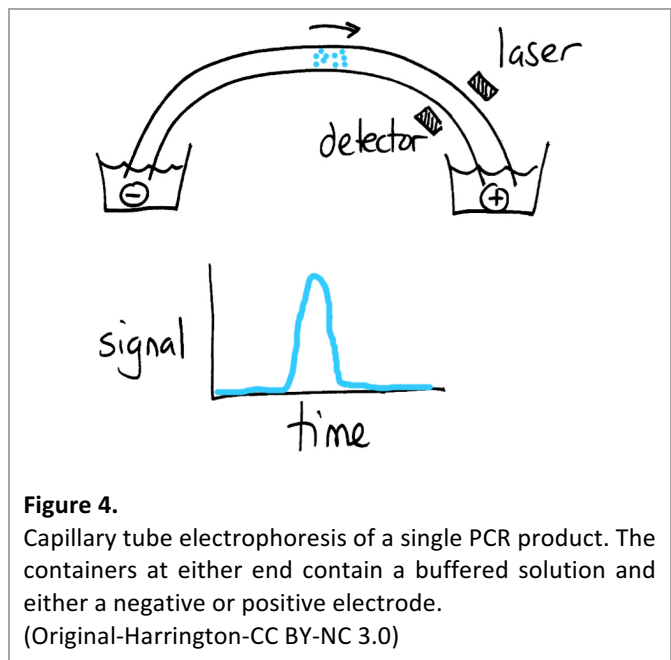
Figure 4.
Capillary tube electrophoresis of a single PCR product. The containers at either end contain a buffered solution and either a negative or positive electrode.
(Original-Harrington-CC BY-NC 3.0)

There are two other similarities between capillary tube electrophoresis and slab gel electrophoresis. Just as a slab gels contain several lanes, the machines used for capillary tube electrophoresis have several tubes, up to 96 in fact. This allows scientists to run many samples simultaneously. Another similarity is the need for molecular weight markers. In both systems these are pieces of DNA of known length that are used to estimate the sizes of the DNA molecules in each sample. In the case of capillary tube electrophoresis these DNA molecules are attached to a different coloured fluorescent dye. The computer uses them to estimate the size of the PCR products.

4. MODERN DNA FINGERPRINTING

4.1. OVERVIEW
DNA fingerprinting, as its name suggests, is the ability to produce a unique collection of data for every person, using their DNA. It was invented by the U.S. Federal Bureau of Investigation in the 1980s and is now done using the technique presented above. The procedure is sometimes called CODIS (COmbined DNA Index System) named after the software that converts the peaks into numerical data.

In Canada and the United States, the test is for 13 autosomal STRs. These are shown in **Figure 1**. This figure also shows the *AMEL* gene. The allele on the X chromosome is shorter than the allele on the Y chromosome. This difference can be detected with its own PCR reaction. If a person is XX they will only have the shorter allele while if a person is XY they will have two sizes of PCR products. The sum total of the PCR reactions produces a collection of data known as a **DNA profile**. For example, a person's DNA profile might begin:

STR	Chromosome	Result
CSF1P0	5	7/12
D8S1179	8	6/6
D21S11	21	9/10
etc.	etc.	etc.

The chance that another person, even a close relative, has the exact same profile is astronomically small. Identical twins, triplets, etc. will have the same DNA profile.

4.2. FORENSICS
Forensics is the process of gathering data that can be used in a court of law. Because DNA profiles are virtually unique to a person they can be used to match a person to a DNA sample recovered at a crime scene. In the example below only Suspect 2 matches the DNA at the crime scene; we can exclude suspects 1 and 3. (**Table 1**) Note what DNA profiling has done, it has made it easier to exclude a suspect than it is to convict someone.

The technician who does the analysis often has to appear in court to explain the results to the jury. In Canada if a person is convicted of a serious crime their DNA profile will then be stored at the National DNA data bank in Ottawa.

4.3. PATERNITY TESTING
The other common use of DNA fingerprinting is **paternity testing**. If we have a DNA profile for a child and their biological mother, we can identify who the biological father could be. Remember, the power of this type of test is that of exclusion. If the potential father lacks the alleles present in the child, then he cannot be the father. The large number of STRs and alleles makes it possible to exclude essentially everyone except the real father.

Table 1. DNA profiling of suspect 1, 2 and 3.

STR	DNA at crime scene	Suspect 1	Suspect 2	Suspect 3
CSF1PO	7/12	8/11	7/12	7/15
D8S1179	6/6	9/15	6/6	9/12
D21S11	9/10	4/5	9/10	4/9

For example, consider the situation below. Every child STR allele that isn't from the mother must have come from the father. For example, the child's CSF1PO 7 allele must be maternal (it lacks the "10" allele) so their 12 allele must be paternal. This means the real father must have at least one 12 allele, this is found in fathers #2 and #3.

Potential father#1 lacks this allele and thus must be excluded. When we apply this thinking to all of the STRs, only Potential father 3 could be the child's biological father, the other two males are excluded. Remember, there are 13 STRs, each with many alleles, so this method is powerful enough to exclude all but the real father.

Table 2. DNA profiling of child, mother and potential fathers.

STR	Child	Mother	Potential fathers		
			#1	#2	#3
CSF1PO	7/12	7/10	7/10	12/14	12/13
D8S1179	6/6	6/8	6/9	12/12	5/6
D21S11	9/10	10/11	5/5	9/16	9/9

_____

SUMMARY:

- Short tandem repeats (STRs) are easy to detect polymorphisms in human chromosomes and most are harmless in that they don't affect the phenotype.

- STRs alleles can be detected with standard agarose gel electrophoresis. The size of the band or bands reveals the number of repeats present in a particular STR.

- STRs can be detected more efficiently with capillary tube electrophoresis. The location of the peak or peaks is identified with laser illumination of fluorescent tagged primers and the number of repeats present in a particular STR can be determined by computer.

- DNA profiles are virtually unique to an individual. They are used in modern day DNA fingerprinting and paternity testing.

KEY TERMS:

short tandem repeat (STR)
PCR
agarose gel electrophoresis
fluorescent dye
capillary tube electrophoresis

DNA fingerprinting
DNA profile
forensics
paternity testing

QUESTIONS:

1) When CSF1PO is amplified with a standard set of PCR primers the 7 allele is 325 bp long and the 6 allele is 321 bp. How long would the 12 allele be?
2) What results would you expect from a person who is a 7/12 heterozygote at CSF1PO using:
   a) agarose gel electrophoresis
   b) capillary tube electrophoresis
3) In response to a 2015 terrorist attack, Kuwait has made DNA testing of its population (both citizens and foreign residents) mandatory. Do you think Canada should adopt this policy?
4) With regards to Table#2 showing paternity testing, label the child's alleles in red for maternal and blue for paternal. Assume that potential father 3 is indeed the child's biological father.
5) Again using the same table, what if male #3 had a brother with the following DNA profile:

STR	#3 brother
CSF1PO	12/15
D8S1179	4/6
D21S11	9/10

Could he be excluded as the father, genetically speaking?
6) The STR repeat number is very variable from person to person.
   a) How do we know they are relatively stable within an individual?
   b) If they weren't, would they be a good tool for DNA profiling?
7) 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

8) 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 (*E₁*) or a 90bp (*E₂*) 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?

9) Based on the genotypes you recorded in Question 8, can you determine which of the individuals could be a parent of individual #1?
10) At the bottom of this page, part of the DNA sequence of a chromosome is shown. 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

```
TAAAGGAATCAATTACTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTTCTTAGTTGTTTAAGTTTTAAGTTGTGA
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ATTTCCTTAGTTAATGAAGACACACACACACACACACACACACAAGAATCAACAAATTCAAAATTCAACACT
```

Chapter 37 - DNA Variation Studied with Microarrays

Figure 1.
A DNA microarray slide showing the array of test spots on the slide.
(Flickr- Argonne Laboratory-CC BY SA 2.0)

Introduction

Microarrays have revolutionized many types of genetics (**Figure 1**). Projects that used to take years can now be done in weeks if not days. This chapter will look at two of these techniques. One is used to identify genes that are responsible for human diseases. The other is used to reveal whether a person has mutations in any of these genes.

1. Single Nucleotide Polymorphisms (SNPs)

1.1. What are SNPs?

Here is a little section of human DNA from within an intron of the *INPP5B* gene on chromosome 1:

$$--\text{TCCTC}\textbf{T}\text{CCAGC}--$$

$$--\text{AGGAG}\textbf{A}\text{GGTCG}--$$

$$--\text{TCCTC}\textbf{A}\text{CCAGC}--$$

$$--\text{AGGAG}\textbf{T}\text{GGTCG}--$$

Most people have the sequence on the top but some people have the sequence on the bottom. The only difference is the single base pair shown in **bold**. This TA to AT base pair substitution mutation likely originated as a single event in the human population tens of thousands of years ago. The original allele is called the **ancestral allele** while the new allele is called the **minor allele**. These, and

changes like it, are called **single nucleotide polymorphisms** (**SNPs**). Most SNPs, including this one, do not affect the expression of genes they are near or within. Their sequence has no effect on the phenotype.

This SNP is named rs16824514 and is described at an online SNP database:

http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=16824514

rs16824514 is not important for human health, but it does provide an example of what SNPs are:

- It has two alleles, in this case T = ancestral and A = minor.

- People can have three possible genotypes, in this case TT, TA, and AA.

- It is easy to determine which alleles a person using one of various methods.

The rest of this chapter will discuss how we can determine which alleles of a SNP a person has and what this can tell us. But before then, where do SNPs come from?

1.2. How are SNPs discovered?

Most human SNPs were identified during DNA sequencing projects. Research teams deliberately chose people from different ethnic groups to sequence. See Chapter 35, Figure 1, for the spread

of humans across the world. For example, they may have sequenced people from countries X, Y, and Z, each in a different continent. Over the past thousands of years, mutations have been happening in people from each population. A few of the base pair substitutions would have become common and could even have replaced the original ancestral allele. When we sequence a present day members of these diverse populations, we have a good chance of getting a person that has the new allele. When we compare this person's genomic sequence to people from other populations, those people will still have the ancestral allele. This difference at a single place on the chromosome is our SNP. It will be entered into the SNP database as 'reference SNP' followed by a number (rs####).

There are methods other than sequencing to detect SNPs. The detection of SNPs is useful for other organisms that have either never been sequenced or have only been sequenced once. Even though this chapter only discusses the uses of SNPs in human biology and health, they are useful tools used by biologists working in many different organisms to understand basic questions in biology, genetics, and evolution.

2. MICROARRAY TECHNOLOGY

2.1. OVERVIEW
There are many ways to detect SNPs. Over the years, Southern blots, PCR, genome sequencing, and other techniques have been used. Most of these methods are rarely used on a large scale today because they are too labour intensive, expensive, time consuming, or some combination thereof. This Chapter will focus on a single method, chosen because it is how most SNP detection is done in 2015. It makes use of a specific type of microarray, a genotyping microarray made by a biotechnology company called Illumina, Inc. in San Diego, California, USA.

Microarrays are a technology used to quantify DNA or RNA molecules. There are many, many types. Some are sold by biotechnology companies, while others are custom made by scientists themselves. They go by different names: microarrays, DNA chips, lab-on-a-chips, biochips, gene arrays, and

others. What they have in common is shown in **Figure 2.** A microarray is a piece of glass or silicon with short pieces of single stranded DNA stuck to its surface. There are thousands of spots of these **oligonucleotides** (**oligos**). Within each spot all of the oligos have the same specific sequence.
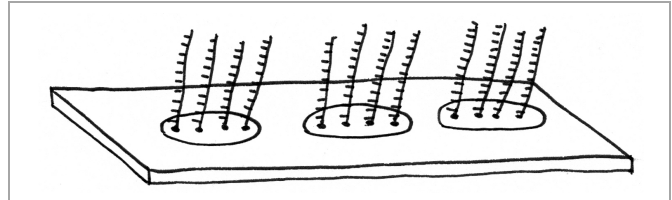
Figure 2.
A glass slide based microarray. Actual microarrays have hundreds of thousands of these spots, each with thousands of identical oligonucleotides attached.
(Original – Harrington – CC BY SA 4.0)

To use a microarray, you need to prepare the DNA or RNA sample first (Step 1 in **Figure 3**). DNA samples are broken into smaller fragments, made single stranded, and covalently attached to a fluorescent dye. RNA molecules just need to be attached to the fluorescent molecules. Next you pour the labelled sample onto the microarray (Step 2). If a piece of labelled nucleic acid is complementary to the oligos in a spot they will hybridize. After all of the unhybridized sample is washed off, the spot will fluoresce. The microarray is then put into a **microarray reader** (Step 3) to take a digital photograph of the fluorescent spots on the surface (Step 4). The spots with fluorescence indicate the presence of complementary sequence in the sample and the level of fluorescence indicates the amount.

2.2. GENOTYPING MICROARRAYS
Have a look at the DNA sequences at the beginning of the chapter. How could we design a microarray to detect which of these alleles a person has? Simply put, this microarray would need to have two spots for each SNP, one for each allele. One will have oligonucleotides that match the ancestral allele and one will have oligos that match the minor allele (**Figure 4**). To use this microarray we would need to isolate genomic DNA from a person,
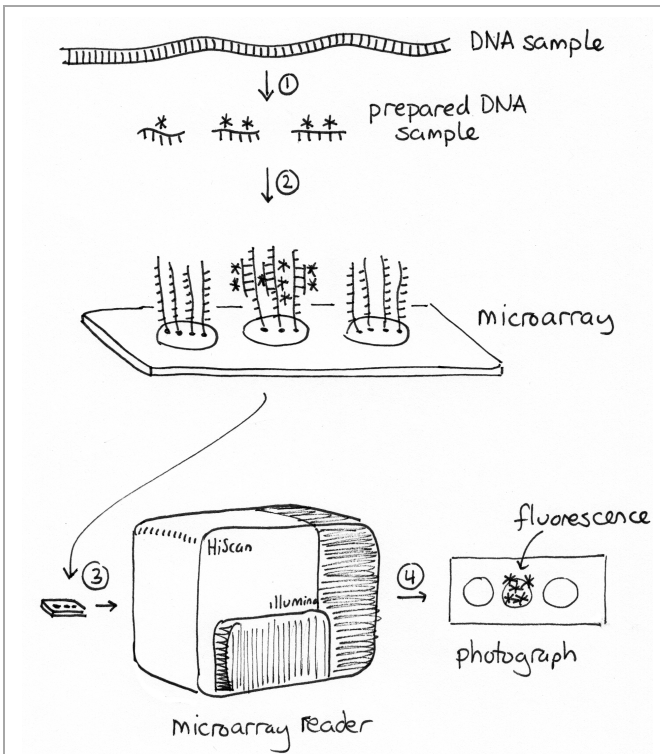
Figure 3.
How microarrays are used and read.
(Original – Harrington – CC BY SA 4.0)
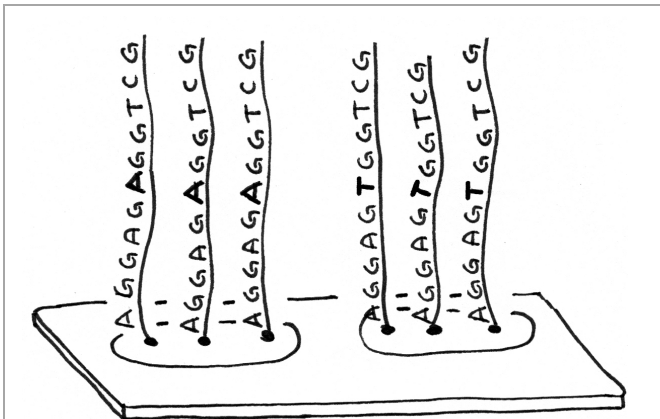
Figure 4.
Two spots to detect alleles of one SNP on a genotyping microarray.
(Original – Harrington – CC BY SA 4.0)

then process it into short, single stranded lengths, and fluorescently label those pieces. Next, we would need to inject it onto the microarray, let it hybridize, and wash away the unhybridized probe. In the above example if a person has the ancestral allele some of their DNA sample will bind to the

spot on the left. Likewise, if a person has the minor allele some of their DNA will bind to the spot on the right. There are therefore three possible results:

Table 1. Results from genotyping microarray.

Result	Interpretation
left spot fluoresces	the person is homozygous for the ancestral allele (TT)
right spot fluoresces	the person is homozygous for the minor allele (AA)
both spots fluoresce	the person is heterozygous (TA)

These two spots only take up a tiny portion of the surface of the microarray so there can be many, many other pairs of oligos to detect other SNPS. This type of microarray is called a **genotyping microarray** because it determines a person's genotype at many SNPs all at once.

Illumina, Inc. is a biotechnology company that makes popular genotyping microarrays and readers. For example, their HumanOmniExpress BeadChip microarray can test 12 people at a time for 730 000 SNPs. The process is somewhat different than shown in **Figure 3** but the data produced is the same. This microarray works with two microarray readers, a larger HiScan™ or a smaller iScan™. The microarray readers are expensive but if one is continuously used a laboratory can test up to 1400 people a week. The rest of this chapter will discuss two procedures that can be done with these microarrays: one looks for new disease causing mutations and the other reveals whether a person has any of these mutations.

3. GENOME-WIDE ASSOCIATION STUDIES (GWAS)

Some human diseases are caused by mutant alleles of single genes. But how can scientists identify which of our ~20,000 gene(s) is/are responsible?

Consider a type of neuronal degeneration called Huntington disease. It took researchers Nancy Wexler and James Gusella ten years to discover that the gene responsible for this disease was on chromosome 4. Their teams used a technique called restriction fragment length polymorphism (RFLP) mapping. This Southern blotting-based approach has been replaced with much faster microarray-based methods.

Using genotyping microarrays to discover genes is called **SNP mapping** or **genome-wide association studies** (**GWAS**). We need large DNA samples from two groups of people, those that have a mutation causing a disease (or a phenotype of interest) and those who do not. Each person's DNA is isolated and then genotyped using a genotyping microarray. A powerful statistical test is then used to study the results. What the software is looking for is correlations; are any of the SNPs correlated with the disease phenotype (**Figure 5**)?
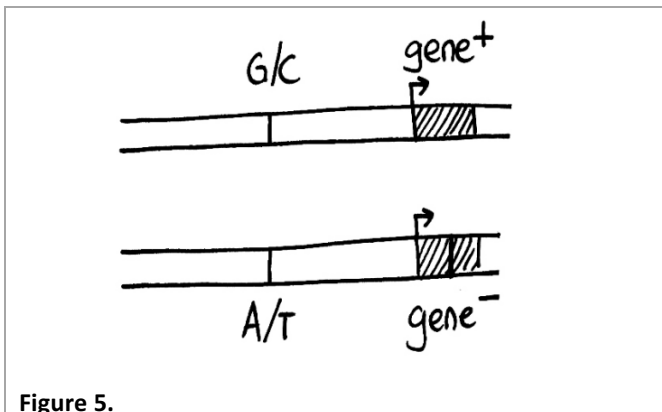
Figure 5.
SNPs can be used to determine the *location* of mutations that affect our health. The G/C vs A/T SNP is located near the gene of interest. The G/C form will tend to associate with the gene+ allele, while the A/T form will tend to associate with the gene- allele.
(Original – Harrington – CC BY SA 4.0)

For example, does everyone with the mutant gene have an A/T base pair allele of a specific SNP, while everyone with the functioning gene have a G/C base pair allele? If so, then the gene should be near this SNP. Because we know where the location of the SNP, is we now know where the approximate location of the gene is.

The results are displayed in a graph (**Figure 6**). The x-axis is the location of the SNPs, in this case all the

SNPs on chromosome 4. On the y-axis is the probability that each SNP is close to the disease-associated gene. For most SNPs the probability is low.
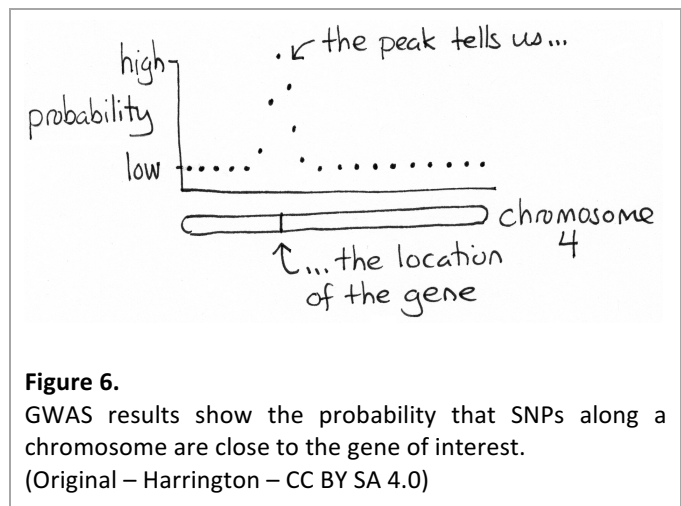
Figure 6.
GWAS results show the probability that SNPs along a chromosome are close to the gene of interest.
(Original – Harrington – CC BY SA 4.0)

GWAS is very effective at identifying genes when there is only one gene that can mutate to cause the disease. When there are two or more genes the results are harder to interpret.

For example, let's say that there are two genes that can mutate to cause the same disease. In our group of people with the disease some have mutations in the first gene and some have mutations in the second gene. SNPs near the first gene will not be close to the second gene, and vice versa. This dilutes the association between SNPs and the disease. GWAS is not very useful when it comes to multifactorial diseases. These are diseases such as cancer and heart disease, where there are mutations in many genes that can be responsible.

4. DIRECT TO CONSUMER DNA TESTING

Direct to consumer DNA testing, as the name suggests, is a form of genetic testing that does not involve a physician ordering the test, or helping the person to understand the results. The results go directly to the consumer and they are left to interpret their genotypes, usually with the assistance of a web site provided by the testing company. There are over a dozen companies offering this service worldwide.

In the case of 23andMe, for example, a person provides a DNA sample by spitting into a tube and

pays ~$200 to have it tested. Many people think the DNA is being sequenced but this isn't possible for this price. The cost to sequence a person's DNA in 2015 is about $1000+. Instead what 23andMe does is to load the DNA samples they receive into genotyping microarrays. In fact, they use the Illumina microarray described earlier in this chapter.

What they are looking for is SNPs known to be next to described genes. These are genes previously identified as being medically important.

For example, in **Figure 7** if a person is heterozygous or homozygous for the A/T base pair SNP allele they have a higher probability of having the mutant allele of the gene. Conversely if they only have the G/C base pair alleles they probably don't. 23andMe reports back a person's genotype as GG, GA, or AA and describes the likelihood of a person having this disease or condition. While this looks like a medical diagnostic test, 23andMe argues it isn't looking for the disease causing mutations directly and therefore they are not offering a diagnostic test.

23andMe and other such companies also offers heredity testing. Other SNPs on the same microarray are known to be correlated with different ethnic groups. For example, if a minor allele is common in English people and the person is homozygous for this allele chances are this person is English. If a person is heterozygous it means one of their parents is likely to be English and the other not. If a person only has the ancestral alleles it means neither of their parents is likely to be English. There are enough SNPs used for most people to learn where their ancestors came from.

Additionally, many companies offer testing for the percentage of your genome that came from Neanderthals. Neanderthals interbred with humans around 60,000 years ago and many people of European, Asian, Australian, and Native American origin have retained a few percent of their genome within their own. The human and Neanderthal genomes can be distinguished by SNPs.

Figure 7.
SNPs can be used to determine the *presence* of mutations that affect our health. The SNP that was originally used to discover this gene is now being used to test for it.
(Original – Harrington – CC BY SA 4.0)

_____

SUMMARY:

- Single nucleotide polymorphisms (SNPs) are harmless and easy to detect polymorphisms in human chromosomes

- SNPs can be detected with genotyping microarrays and microarray readers. These microarrays have pairs of oligonucleotide spots, one for the ancestral allele and one for the minor allele. Where a person's DNA hybridizes reveals their genotype at a particular SNP.

- Genome-wide association studies (GWAS) use SNPs and genotyping microarrays to determine the location of mutations that affect our health.

- Direct to consumer DNA testing uses SNPs and genotyping microarrays to determine the presence of allelic forms that show linkage to genes that may affect our health, predict our ancestry, and estimate the percentage of Neanderthal sequences we have.

KEY TERMS:

ancestral allele
minor allele
single nucleotide polymorphism (SNP)
microarray
microarray reader

genotyping microarray
SNP mapping
genome-wide association study (GWAS)
direct to consumer DNA testing

QUESTIONS:

1) Why do most SNPs only have two alleles?
2) When using a genotyping microarray why is it important that only perfectly matched DNA molecules be able to hybridize?
3) Could GWAS be used to find out why some people have blue eyes and other people don't?
4) Assuming the gene responsible for blue eye colour is known, could direct to consumer testing predict whether a person has blue eyes or not?
5) What do these GWAS results mean?
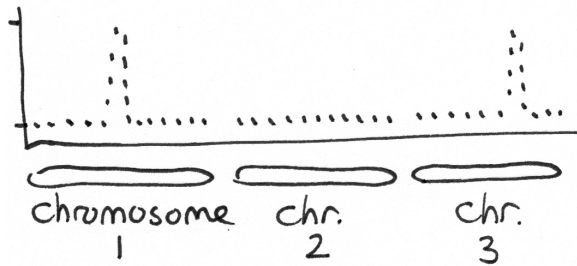
Notes:

CHAPTER 38 – POPULATION GENETICS

Figure 1.
Population genetics is important in ecology, evolution, and even in our daily lives since disease risks can be calculated using population genetics.
(Flickr-Zach Stern- CC BY-NC-ND 2.0)

INTRODUCTION

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.

1. ALLELE FREQUENCIES MAY 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

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**:

$$p^2 + 2pq + q^2 = 1$$

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 **p² + 2pq + q²=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 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 equilibrium can be an indication that one of the alleles affects the reproductive success of organism, for example through natural selection or **assortative mating**.

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:

$q2 = 0.04\% = 0.0004$

$q= \sqrt{0.0004} = 0.02$

$p= 1-q = 0.98$

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

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.

Table 1. Conditions for the Hardy-Weinberg equilibrium

- **Random mating**: Individuals of all genotypes mate together with equal frequency. Alternatively, **assortative mating**, in which certain genotypes preferentially mate together, is a type of non-random mating.

- No **natural selection**: All genotypes have equal fitness. None are selectively removed by selection.

- No **migration**: Individuals do not leave or enter the population.

- No mutation: The allele frequencies do not change due to mutation.

- Large population: Random sampling effects in mating (i.e. genetic drift) are insignificant in large populations.

_____

Summary:

- 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:

population	Hardy-Weinberg equilibrium
p / q	random mating
p+q=1	natural selection
Hardy-Weinberg formula	migration
$p^2 + 2pq + q^2=1$	assortative mating

STUDY QUESTIONS:

1) 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.

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

genotype	number of individuals
AA	432
Aa	676
aa	92

3) 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?

4) Propose an explanation for the deviation between the genotypic frequencies calculated in question 3 and those observed in the table in question 2.

CHAPTER 39 – EVOLUTION OF GENE EXPRESSION

Figure 1.
The stickleback is an example of an organism in which researchers have identified mutations that cause changes in the regulation of gene expression. These mutations confer a selective advantage in some environments. Natural selection acts on mutations that alter gene expression patterns, as well as those changing the coding regions of genes.
(Wikipedia-K. Szabolcs- CC BY-SA 3.0)

INTRODUCTION

Mutations can occur in both *cis*-elements (promoter, enhancers, etc.) and in the genes that code for *trans*-factors (transcription 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 of the Stickleback fish (**Figure 1**).

1. BASICS OF DEVELOPMENT

1.1. ONE CELL ➔ MULTICELLULAR ORGANISM

In **multicellular organisms**, the original, single **zygote** cell, which arises from the union of two gametes, replicates via mitosis to produce all the cells in the body. Thus, all the cells have essentially the same genotype. As these cells grow and divide they express different sets of genes in a programmed manner. Thus organismal development requires that the cells differentiated into various cell types and tissues, which ultimately become organs within the organism. This differentiation is the result of specific sets of genes being expressed in different cell types.

1.2. DETERMINATION AND DIFFERENTIATION – CELL FATE COMES FROM GENE EXPRESSION

In the process of development cells become more specialized. This specialization is the result of

specific genes being expressed to make or cause the cell to progress down specific developmental pathways. A cell can become "**determined**" (destined to go down one pathway – a cell fate) through the expression of one or more genes. Usually these genes are transcription factors that regulate the transcription of other genes. An example has been presented in Chapter 21 with the TDF-Y expression dictating the development of the gonad into a testis.

Once a cell has become determined, then it will **differentiate** into a specific cell type by expressing the genes for that cell type and turning off others that are not expressed in that type. For example a myoblast cell (muscle cell precursor) will begin the expression of actin, myosin, and other protein that for a muscle cell.

The **fate** of a cell is first determined and then the cell differentiates into that cell type. This typically happens as a cell goes through a division where each of the daughter cells goes on to become different cell types. There is a series of choices with each cell division. This can lead to the variety of cell and tissue types found in multicellular organism.

2. VARIATION IN GENE EXPRESSION AND EVOLUTION

Genetic variation arises from random mutation in DNA sequences. Mutations can occur in three categories of DNA: (1) **intergenic sequences**, (2) **gene coding sequences**, and (3) **gene regulatory sequences**. If a change (mutation) occurs it may be selected for in a population because it has an

advantage (selection). Others mutations may just increase in frequency in a population through chance alone (drift).

2.1. INTERGENIC SEQUENCES

Mutations in intergenic sequences (regions between genes) have no effect on gene expression or phenotype, so there is no selection for/against. Consequently, these mutations are useful for markers in genetic mapping or DNA finger printing. (Random genetic drift can cause **fixation**, where all members of a species have the same DNA sequence.) The result is that evolution occurs via random mutation and fixation by random drift, with no selection for or against these sequences.

2.2. GENE CODING SEQUENCES

Mutation in a gene coding sequence can change the effectiveness of a gene product (RNA or protein), which can consequently affect the phenotype. This type of change doesn't alter the gene's transcription, but natural selection may act for/against this new phenotype. The result is that evolution occurs via random mutation and selection against for/against the function of the gene's product.

2.3. GENE'S REGULATORY SEQUENCES

This is the most interesting type. Mutations in regulatory sequences do not change the product from the gene, just the pattern of transcription (time, place,). In other words, the time, place (tissue), level, and response to environment of expression are changed. **Regulatory mutations** can affect many traits and characteristics at once (**pleiotropic**) or create new and/or novel patterns of expression. This might result in a new function in an organism (e.g. neomorph). With this type of mutation, evolution occurs via random mutation

and selection for/against the novel expression pattern, not altered gene products.

Mutations in regulatory sequences are a key to understanding how evolution produces new patterns of expression and new phenotypes. Regulatory sequences have to be identified experimentally and shown to act combinatorial to regulate transcription. Most genes have multiple independent regulatory sites adjacent to the transcribed sequence that bind trans-acting factors to modulate expression.

3. EXAMPLE 1: DROSOPHILA *YELLOW* GENE

The *yellow* gene of Drosophila provides an example of the modular nature of enhancers (regulatory sequences). This gene encodes an enzyme in the pathway that produces a dark pigment in the insect's exoskeleton. Null mutants have a yellow cuticle rather than the wild type darker pigmented cuticle. This gene is called the gene "yellow" because it is named after their mutant phenotype.

Figure 2 shows five distinct enhancer elements that drive transcription of *yellow* (left, 5' up stream - wing, body, mouth parts; intron – bristles, claws). Each binds a different, tissue-specific transcription factor to enhance transcription of $yellow^+$ transcript (and thus express the protein) 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 part 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.
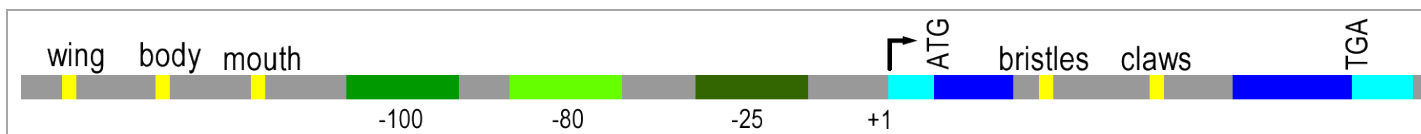
Figure 2.
Tissue-specific cis-regulatory elements within a simplified representation of the yellow gene of Drosophila.
(Origianl-Deyholos-CC BY-NC 3.0)

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 overall expression profile (transcription level, tissue specific, temporal specific) is a total combination of all the various enhancer and silencer elements that act on that gene.

4. EXAMPLE 2: *PITX* EXPRESSION IN STICKLEBACK FISH

The three-spined **stickleback** (*Gasterosteus aculeatus*; **Figure 1**) provides a classic example of natural selection that involves a mutation in a *cis*-regulatory element.

Background: Members of this species occur in one of two forms: (1) populations that inhabit deep, open water and have a spiny pelvic fin that is thought to deter 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 are a disadvantage because they allow predatory insects like dragon fly larvae in the sediment to grasp the stickleback.

Researchers compared gene sequences of individuals from both deep and shallow water environments as shown in **Figure 4**. They observed that in embryos from the deep-water population, a gene called **Pitx** (*paired-like homeodomain 1*

encodes a member of the RIEG/PITX homeobox family) 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*.

5. EXAMPLE 3: 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 (**Figure 3**). 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**.

Figure 3.
A tetramer of human hemoglobin, type $\alpha 2 \beta 2$. The α chains are labeled red, and the β chains are labeled blue. Heme groups are green. (
(Wikipedia- Zephyris- CC BY-SA 3.0)

Figure 4.
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.
(Original-Deyholos-CC BY-NC 3.0)

The composition of hemoglobin tetramers changes during development (**Figure 5**). From early childhood onward, most tetramers are of the type **α₂β₂**, 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 α₂δ₂, which has δ globin instead of the more common β globin. Other tetrameric combinations predominate before birth: ζ₂ε₂ is most abundant in embryos, and α₂γ₂ 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. Each of these globin polypeptides is encoded by a different gene. In humans, globin genes are located in clusters on two chromosomes (**Figure 6**). 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.
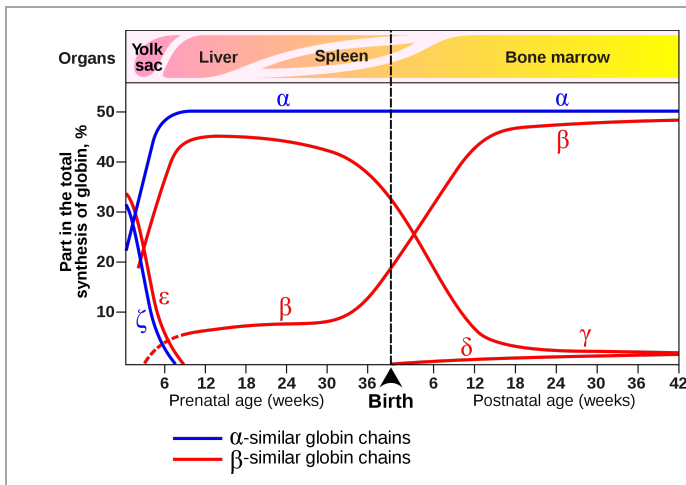
Figure 5.
Expression of globin genes during prenatal and postnatal development in humans. The organs in which globin genes are primarily expressed at each developmental stage are also indicated.

Data: Wood, W.G. 1976 Br. Med. Bull. 32, 282
Original: (Wikipedia-Furfur- CC BY-SA 3.0)
Derivative work/Translation: (Wikipedia-Leonid2- CC BY-SA 3.0)
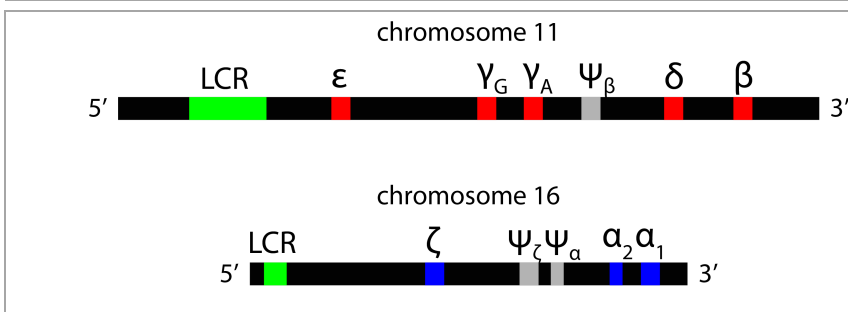
Figure 6.
Fragments of human chromosome 11 and human chromosome 16 on which are located clusters of β-like and α-like goblin genes, respectively. Additional globin genes (θ, μ) have also been described by some researchers, but are not shown here.
(Wikipedia – Modified by Kang-CC BY-NC 3.0)

_____

SUMMARY:

- Development of a single cell zygote to a multicellular organism involves the sequential expression of genes so that determination and differentiation can take place and the cells can form the variety of types found in the adult organism.

- Mutations can occur in intergenic, gene coding, or gene regulatory sequences. Changes in regulatory sequences can lead to altered gene expression including new developmental times or tissue locations.

- The Drosophila *yellow* gene is an example of mutations in gene regulatory sequences.

- Stickleback fish provide an example of recent evolutionary events in which mutation of an enhancer produced a change in morphology with a selective advantage (evolution).

- Expression of the various human globin genes, which generate hemoglobin, is an example of gene expression changes over developmental time. The family of globin genes arose via gene duplication. Not all duplications produce functional genes, some are pseudo-genes.

KEY TERMS:

multicellular organisms
zygote
determination
differentiation
cell fate
intergenic sequences
gene coding sequences
gene regulatory sequences
regulatory mutations

pleiotropic
stickleback
Pitx
Primordium
post-translational modification
hemoglobin/heme/globin
gene duplication
pseudogene
gene families

STUDY QUESTIONS:

1) 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 expected if a homozygous wild-type fish from a deep-water population mated with a homozygous wild-type fish from a shallow-water population?

2) The modular nature of transcription enhancer elements can easily be seen in the *yellow* gene of Drosophila. Suppose that there was a mutant that had a deletion of the three distal enhancer elements (wing, body, mouth – See **Figure 2.**). There was another, different mutation that resulted in a stop codon early in the protein coding sequence.

   a) What would the phenotype of the homozygote deletion mutant be?

   b) What would the phenotype of the homozygote stop codon mutant be?

   c) What would the phenotype of the heterozygote be?

   d) Suppose the heterozygote phenotype was wild type. How might that occur?

CHAPTER 40 – TRANSGENIC ORGANISMS

Figure 1.
Two transgenic mice expressing enhanced green fluorescent protein (eGFP) under UV-illumination flanking one plain NOD/SCID mouse from the non-transgenic parental line.

(Wikimedia-Moen *et. al* (2012)-CC BY 2.0)

INTRODUCTION

The addition of new genetic material to single cell organisms has been possible for decades. Recall R. Griffith's 1928 experiments with smooth and rough pneumococcus (Chapter 1). However, the routine transformation of bacteria with plasmids began in the early 1970s. The ability to transfer DNA (genes) into complex, multicellular organisms is more recent and usually called transfection when dealing with cells and it also began in the 1970s. This genetic technology has opened up whole new avenues of research as well as new possibilities for commercial gain and health improvement.

1. MODEL ORGANISMS FACILITATE GENETIC ADVANCES

1.1. 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 (**Figure 2**). 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 organisms 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 there are some practical difficulties working with mice, such as cost, slow reproductive time, and ethical considerations.
- The zebrafish (***Danio rerio***) has more recently been developed by researchers as a genetic model for vertebrates. Unlike mice, zebrafish embryos develop quickly and externally to their mothers, and are transparent, making it easier to study the development of internal structures and organs.
- 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.

1.2. 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 2.
Some of the most important genetic model organisms in use today. Clockwise from top left: yeast, roundworm, Arabidopsis, zebrafish, mouse, fruit fly)

Clockwise from top left: (Wikipedia- Masur-PD) / (Wikipedia- Kbradnam- CC BY-SA 2.5) / (Wikimedia- Frost Museum- CC BY 2.0) / (Flickr-Max Westby-CC BY-NC-SA 2.0) / (Wikimedia- Rama- CC BY-SA 2.0 FR) / (Flickr- tohru.murakami- CC BY-NC 2.0)

2. What are Transgenic Organisms?

Transgenic organisms contain foreign DNA that has been introduced using biotechnology. Foreign DNA (the **transgene**) is defined here as DNA from another species, or else recombinant DNA from the same species that has been manipulated in the laboratory then reintroduced. The terms transgenic organism and genetically modified organism (**GMO**) are generally synonymous. (Note that any mutant is technically "genetically modified", but the "GMO" term is usually not used to refer organism derived via classical mutant and breeding techniques.)

The process of creating transgenic organisms or cells to become whole organisms with a permanent change to their germline has been called either **transformation** or **transfection**. (Unfortunately, both words have other meanings. Transformation also refers to the process of mammalian cell becoming cancerous, while transfection also refers to the process of introducing DNA into cells in culture, either bacterial or eukaryote, for a temporary use, not germ line changes.)

Transgenic organisms are useful in several areas. (1) They are important research tools, and are often used when exploring a gene's function. (2) Transgenesis is also related to the medical practice of gene therapy, in which DNA is transferred into a patient's cells to treat disease.   (3) Transgenic organisms are widespread in agriculture. Approximately 90% of canola, cotton, corn, soybean, and sugar beets grown in North America are transgenic. No other transgenic livestock or crops (except some squash, papaya, and alfalfa) are currently (2014) produced in North America, although many are being researched.

3. Making a Transgenic Cell

To make a transgenic cell, DNA must first be transferred across the cell membrane, (and, if present, across the cell wall), without destroying the cell. In some cases, **naked DNA** (meaning plasmid or linear DNA that is not bound to any type of **carrier**) may be transferred into the cell by adding DNA to the medium and temporarily increasing the porosity of the cell membrane, for example by **electroporation**. When working with larger cells, naked DNA can also be **microinjected** into a cell using a specialized needle.   Other methods use **vectors** to transport DNA across the membrane. Note that the word "vector" as used here refers to any type of carrier, and not just plasmid vectors.   Vectors for transformation/transfection include **vesicles** made of lipids or other polymers that surround DNA; various types of particles that carry DNA on their surface; and infectious viruses and bacteria that naturally transfer their own DNA into a host cell, but which have been engineered to transfer any DNA molecule of interest. Usually the foreign DNA is a complete expression unit that includes its own cis-regulators (e.g. promoter) as well as the gene that is to be transcribed.

When the objective of an experiment is to produce a **stable** (i.e. heritable) transgenic eukaryote, the foreign DNA must be incorporated into the host's chromosomes. For this to occur, the foreign DNA must enter the host's nucleus, and recombine with host DNA (a chromosome).   In some species, the foreign DNA is inserted at a random location, probably wherever strand breakage and non-homologous end joining happen to occur. In other species, the foreign DNA can be targeted to a particular locus, by flanking the foreign DNA with DNA that is homologous to the host's DNA at that locus. The foreign DNA is then incorporated into the host's chromosomes through homologous recombination.

4. Detection of Transgenes and Their Products

Furthermore, to produce multicellular organisms in which all cells are transgenic and the transgene is stably inherited, the cell that was originally transformed must be either a gamete or must develop into tissues that produce gametes. Transgenic gametes can eventually be mated to produce homozygous, transgenic offspring. The presence of the transgene in the offspring is typically confirmed using PCR or Southern blotting (see other chapters), and the expression of the transgene can be measured using reverse-

transcription PCR (RT-PCR), Northern (RNA) blotting, and Western (protein) blotting.

The rate of transcription of a transgene is highly dependent on its insertion site (i.e. **position effects**). That is, the same transgene may be expressed at a high level when inserted at one location, while not expressed at all at a different location. It is the state of the chromatin at the insertion site that influences the level and pattern of expression. Therefore, researchers often generate several independently transformed/transfected lines with the same transgene, and then screen for the lines showing the highest expression. It is also good practice to sequence the transgenic locus from a newly generated transgenic organism, since errors (truncations, rearrangements, and other mutations) can be introduced during transformation/transfection.

5. PRODUCING A TRANSGENIC PLANT

The most common method for producing transgenic plants is **Agrobacterium-mediated transformation** (**Figure 3**). *Agrobacterium tumifaciens* is a soil bacterium that, as part of its natural pathogenesis, injects its own tumor-inducing (T_i) plasmid into cells of a host plant. The natural T_i plasmid encodes growth-promoting genes that cause a gall (i.e. tumor) to form on the plant, which also provides an environment for the pathogen to proliferate. Molecular biologists have engineered the T_i plasmid by removing the tumor-inducing genes and adding restriction sites that make it convenient to insert any DNA of interest. This engineered version is called a **T-DNA** (transfer-DNA) plasmid; the bacterium transfers a linear fragment of this plasmid that includes the conserved "left-border (LB)", and right-border (RB)" DNA sequences, and anything in between them (up to about 10 kb). The linear T-DNA fragment is transported into the nucleus, where it recombines with the host-DNA, probably wherever random breakages occur in the host's chromosomes. In Arabidopsis and a few other species, flowers can simply be dipped in a suspension of Agrobacterium, and ~1% of the resulting seeds will be transformed.
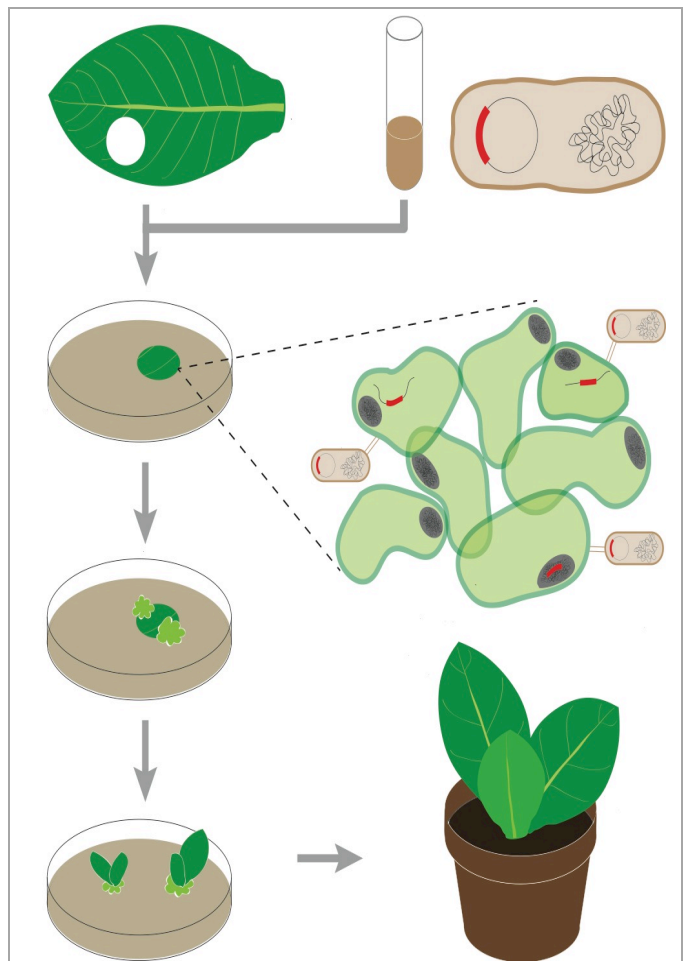
Figure 3.
Production of a transgenic plant using Agrobacterium-mediated transformation. The bacterium has been transformed with a T-DNA plasmid that contains the transgene and a selectable marker that confers resistance to a herbicide or antibiotic. A bacterial culture and plant tissue (e.g. a leaf punch) are co-cultured on growth medium in a Petri dish. Some of the plant cells will become infected by the bacterium, which will transfer the T-DNA into the plant cytoplasm. In some cases the transgene will become integrated into the chromosomal DNA of a plant cell. In the presence of certain combinations of hormones, the plant cells will dedifferentiate into a mass of cells called callus. The presence of a selective agent (e.g. herbicide or antibiotic) in the growth medium prevents untransformed cells from dividing. Therefore, each callus ideally consists only of transgenic plant cells. The resistant calli are transferred to media with other combinations of hormones that promote organogenesis, i.e. differentiation of callus cells into shoots and then roots. The regenerated transgenic plants are transferred to soil. Their seeds can be harvested and tested to ensure that the transgene is stably inherited. (Original-Deyholos-CC BY-NC 3.0)
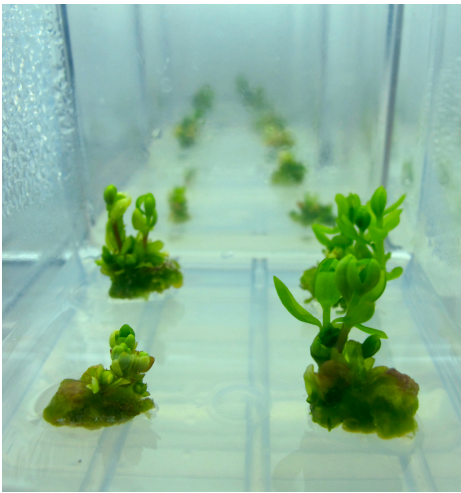
Figure 4.
Organogenesis of flax shoots from calli.
(Original-J. McDill- CC BY-NC-SA 2.0)

In most other plant species, cells are induced by hormones to form a mass of undifferentiated tissues called a **callus**. The Agrobacterium is applied to a callus and a few cells are transformed, which can then be induced by other hormones to regenerate whole plants (**Figure 4**).

Some plant species are resistant (i.e. "**recalcitrant**") to transformation by Agrobacterium. In these situations, other techniques must be used such as **particle bombardment**, whereby DNA is non-covalently attached to small metallic particles, which are accelerated by compressed air into callus tissue, from which complete transgenic plants can sometimes be regenerated. In all transformation methods, the presence of a selectable marker (e.g. a gene that confers antibiotic resistance or herbicide resistance) is useful for distinguishing transgenic cells from non-transgenic cells at an early stage of the transformation process.

6. PRODUCING A TRANSGENIC MOUSE
In a commonly used method for producing a transgenic mouse, **stem cells** are removed from a mouse embryo, and a transgenic DNA construct is transferred into the stem cells using electroporation, and some of this transgenic DNA enters the nucleus, where it may undergo homologous recombination (**Figure 5**). The transgenic DNA construct contains DNA

homologous to either side of a locus that is to be targeted for replacement. If the objective of the experiment is simply to delete ("**knock-out**") the targeted locus, the host's DNA can simply be replaced by selectable marker, as shown. It is also possible to replace the host's DNA at this locus with a different version of the same gene, or a completely different gene, depending on how the transgenic construct is made. Cells that have been transfected and express the selectable marker (i.e. resistance to the antibiotic neomycin resistance, **neoR**, in this example) are distinguished from unsuccessfully transfected cells by their ability to survive in the presence of the selective agent (e.g. an antibiotic). Transfected cells are then injected into early stage embryos, and then are transferred to a foster mother. The resulting pups are chimeras, meaning that only some of their cells are transgenic. Some of the chimeras will produce gametes that are transgenic, which when mated with a wild-type gamete, will produce mice that are hemizygous for the transgene. Unlike the chimeras, these hemizygotes carry the transgene in all of their cells. Through further breeding, mice that are homozygous for the transgene can be obtained.

7. HUMAN GENE THERAPY
Many different strategies for human gene therapy are under development. In theory, either the **germline** or **somatic** cells may be targeted for transfection, but most research has focused on somatic cell transfection, because of risks and ethical issues associated with germline transformation. Gene therapy approaches may be further classified as either *ex vivo* or *in vivo*, with the former meaning that cells (e.g. stem cells) are transfected in isolation before being introduced to the body, where they replace defective cells. *Ex vivo* gene therapies for several blood disorders (e.g. immunodeficiencies, thalassemias) are undergoing clinical trials. For *in vivo* therapies, the transfection occurs within the patient. The objective may be either stable integration, or **non-integrative** transfection. As described above, stable transfection involves integration into the host genome. In the clinical context, stable integration

may not be necessary, and carries with it higher risk of inducing mutations in either the transgene or host genome). In contrast, transient transfection does not involve integration into the host genome and the transgene may therefore be delivered to the cell as either RNA or DNA. Advantages of RNA delivery include that no promoter is needed to drive expression of the transgene. Besides mRNA transgenes, which could provide a functional version of a mutant protein, there is great interest in delivery of **siRNA** (small-inhibitory RNAs), which can be used to silence specific genes in the host cell's genome.

Vectors for *in vivo* gene therapy must be capable of delivering DNA or RNA to a large proportion of the targeted cells, without inducing a significant immune response, or having any toxic effects. Ideally, the vectors should also have high specificity for the targeted cell type. Vectors based on viruses (e.g. **lentiviruses**) are being developed for in both *in vivo* and *ex vivo* gene therapies. Other, non-viral vectors (e.g. vesicles and nanoparticles) are also being developed for gene therapy as well.
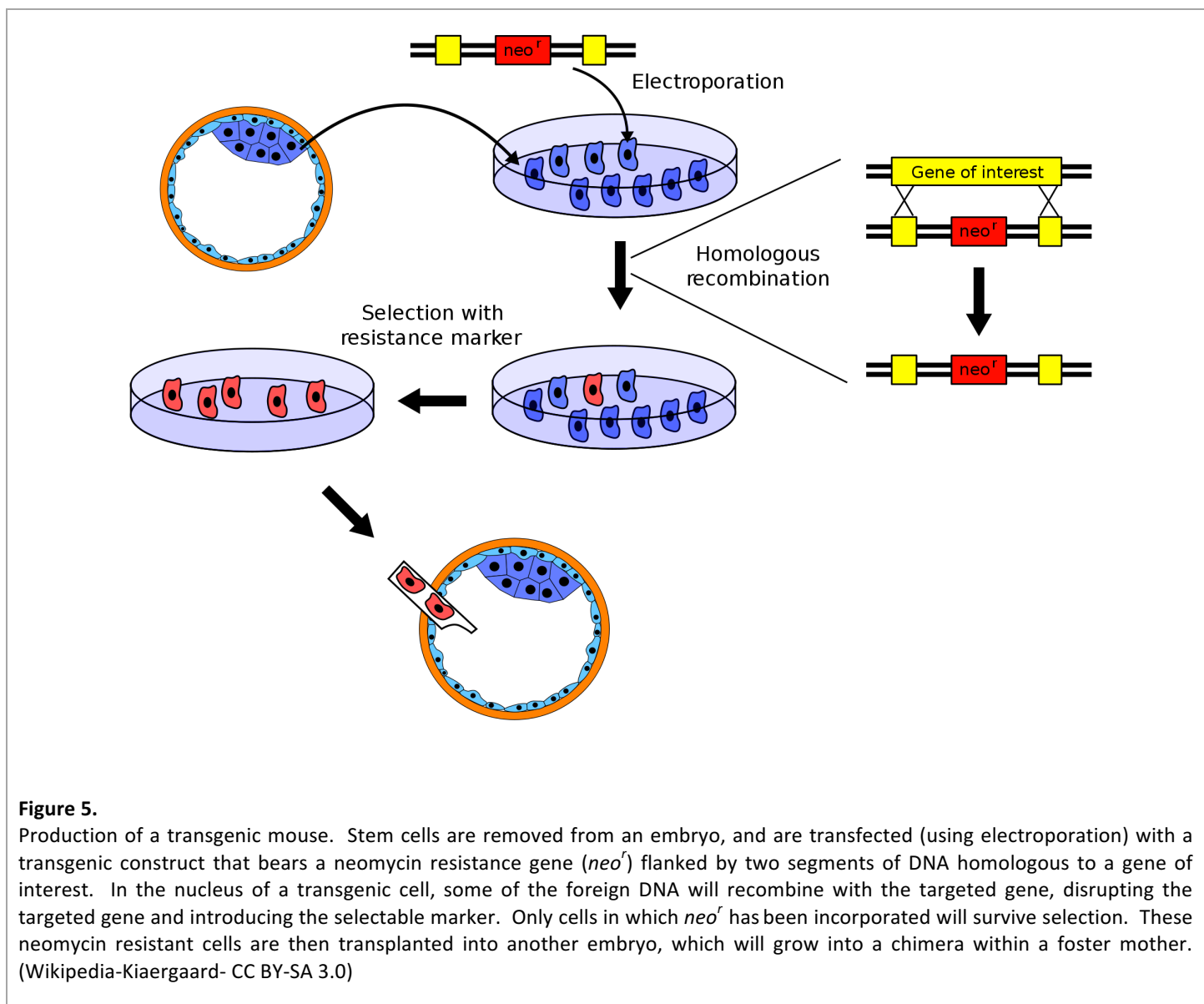
Figure 5.
Production of a transgenic mouse. Stem cells are removed from an embryo, and are transfected (using electroporation) with a transgenic construct that bears a neomycin resistance gene (neo^r) flanked by two segments of DNA homologous to a gene of interest. In the nucleus of a transgenic cell, some of the foreign DNA will recombine with the targeted gene, disrupting the targeted gene and introducing the selectable marker. Only cells in which neo^r has been incorporated will survive selection. These neomycin resistant cells are then transplanted into another embryo, which will grow into a chimera within a foster mother. (Wikipedia-Kiaergaard- CC BY-SA 3.0)

8. CRISPR-Cas9 TECHNOLOGY

To understand how genes work, geneticists need to modify them (mutation) in order to see how the changes affect the phenotype. This will identify the function(s) of the gene. In the past, random mutagenesis, followed by screening and selection, permitted researchers to identify mutations that affected gene function. While useful, this method has limitations. It produces only a small, very limited set of sequence changes to a gene. Much more could be learned if researchers could define the change first, then see the affect on the phenotype. A new method(s) is needed for this.

There have been several technologies regarding targeted **genome editing**. These include ZFNs (Zinc Finger Nuclease) and TALENs (Transcription Activator-like Effectors Nucleases), which depend on protein-DNA interaction. However, engineering specific protein sequences to bind to specific DNA sequences can be time consuming and expensive. Recently, a new method of genome editing tool has been introduced called "**CRISPR-Cas9 system**". It can (1) be easily customized to specific DNA sequences, (2) target with high precision, and (3) target many genes simultaneously.

8.1. WHAT IS CRISPR-CAS9?

The CRISPR system is a combination of two bacterial systems. It is an adaptive bacterial and archaeal immune system that fights against foreign plasmids or viruses. After a bacterium survives a viral infection, it "remembers" the viral DNA sequence by incorporating that sequence into its own genome, and fights back when it is re-infected. **CRISPR** (pronounced *crisper*) is an acronym for Clustered Regularly Interspaced Short Palindromic Repeats. In other words, it is a locus on bacterial or archaeal chromosome that has a cluster of many short sequences that are interspaced repeatedly. The CRISPR locus is composed of (1) CRISPR RNA direct repeats and (2) spacers. Now, there are three main components to the CRISPR-Cas9 system (**Figure 6**):

(1) CRISPR-associated (*cas*) genes

The *cas* genes encode enzymes that control the integration of foreign DNA into its own genome and the defense against bacteriophages. ***Cas9*** gene from *Streptococcus pyogenes* encodes for a nuclease that has the ability to denature foreign DNA.

(2) CRISPR RNA (crRNA) array

This crRNA array is responsible for the target specificity and has two parts to it: (1) CRISPR RNA **direct repeats** and interspaced (2) variable **spacers**. These sequences are not protein coding genes, they only transcribe RNA molecules. Transcription of this array results in the crRNA, which is modified into mature crRNA.

(3) trans-activating crRNA (tracrRNA)

Trans-activating crRNA (tracrRNA) base pairs with crRNA and recruits the Cas9 protein. This crRNA-tracrRNA-Cas9 Protein complex is what we call a "RNA guided endonuclease" that can target and cut specific sites on DNA.

Scientists have fused crRNA and tracrRNA together to form one chimeric RNA molecule called **sgRNA** (single-guided RNA) so that they can increase the efficiency when performing experiments.
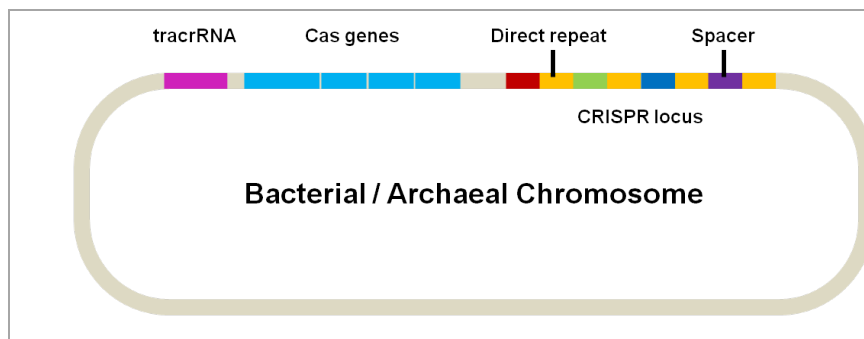
Figure 6.
General representation of CRISPR locus and related genes in bacterial / archaeal chromosome. Note that direct repeats are in yellow colour, and spacers have various colors since they have different sequences which are from multiple origins. (Original-Kang-CC BY-NC 3.0)

8.2. HOW CRISPR SYSTEM WORKS

After a bacterium survives a viral infection, it attains an immune memory so that it can fight back when the same virus infects the cell; this is where the CRISPR-Cas9 system comes in. There are three main components in this adaptive immune system (**Figure 7** and **Figure 8**):

(1) Adaption

First, virus injects its DNA to the host bacterial cell. Then, the host bacterial cell detects the foreign DNA, and integrates short fragment of the viral DNA into the CRISPR locus. This insert is called a **spacer** and the original sequence on the viral genome is called a **protospacer**.

(2) crRNA biogenesis

After the cell is re-infected with the same virus, CRISPR sequence is transcribed into a

premature **crRNA**. This crRNA will contain various spacer and direct repeat sequences, but after modification the mature crRNA will only have the matching spacer or guide RNA to the invading virus. The mature crRNA will form a complex with tracrRNA and Cas 9.

(3) Interference

The crRNA-tracrRNA-Cas9 complex binds to the **protospacer adjacent motif** (**PAM**) sequence first that is located to the target DNA sequence (protospacer) and opens the dsDNA and the guide RNA will base pair with the target sequence. The cas9 protein will leave a double stranded cut to the DNA.

Now, the host bacterial DNA do not have the PAM sequence but the foreign, viral DNA does. Thus, the CRISPR system does not cleave its own DNA.
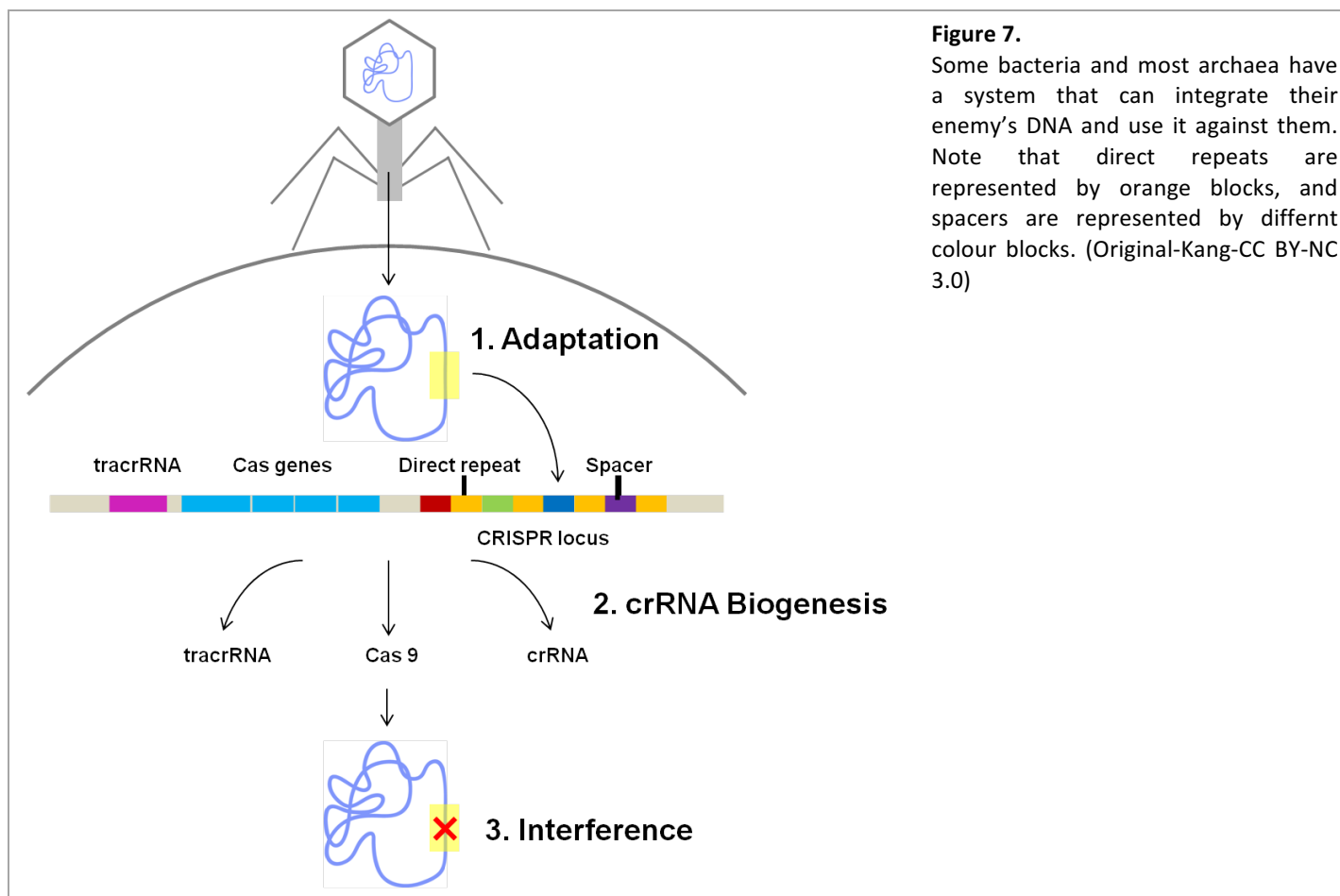
Figure 7.
Some bacteria and most archaea have a system that can integrate their enemy's DNA and use it against them. Note that direct repeats are represented by orange blocks, and spacers are represented by differnt colour blocks. (Original-Kang-CC BY-NC 3.0)

Figure 8.
crRNA will base pair with tracrRNA, and this will recruit Cas9 protein. Therefor,Cas9 protein is guided by the RNA molecule and leaves a double stranded break. (Original-Kang-CC BY-NC 3.0)
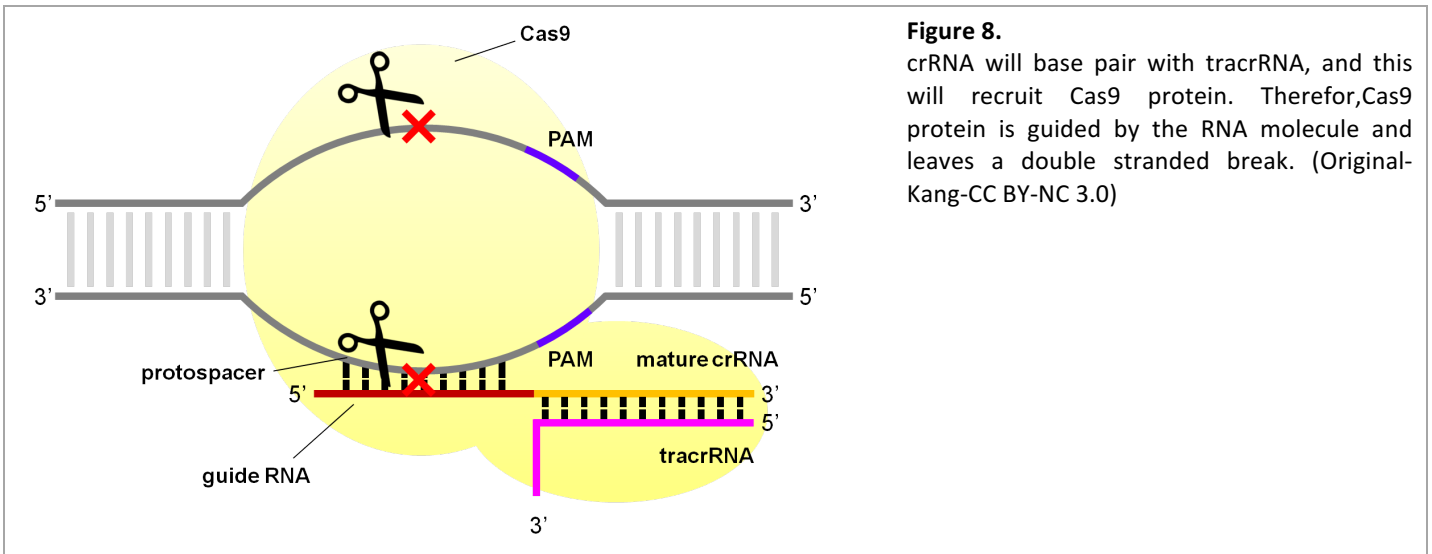
8.3. USING CRISPR-CAS9 FOR GENOME EDITING

What if we can engineer the guide RNA in such a way that it will find a sequence of interest on any DNA and edit it? This would be powerful, programmable genome editing technology. "

We can customize the "search and cut" system that CRISPR-Cas9 complex uses by changing the guide RNA. Then, this system will search for its new target and make a cut on the DNA at that sequence, creating a break. After this "attack" on the target DNA, the cell tries to repair the break and it may or may not be successful. We can use this DNA repair mechanism of the host cell so that we can induce mutations on a gene or even exchange the gene with a new one. Here are two possible ways for repairing the break (**Figure 9**):

(1) HDR (homology-directed repair)

For this type of repair mechanism, single stranded DNA donor molecule is flanked in the break. This way, new strand can be built upon this template DNA. Therefore, we can introduce a new version of the gene.

(2) NHEJ (non homologous end joining)

This repair mechanism does not require a template DNA. It joins the two ends of the DNA and connects them. This process is error-prone, and insertion-deletion (**indel**) mutations can occur which will lead to frameshift mutations.

Figure 9.
The DNA repairing enzymes tries to fix the break by either (1) HDR (homology-directed repair) or (2) NHEJ (non-homologous end joining.). NHEJ is error-prone and can leave a scar such as indel.
(Original-Kang-CC BY-NC 3.0)
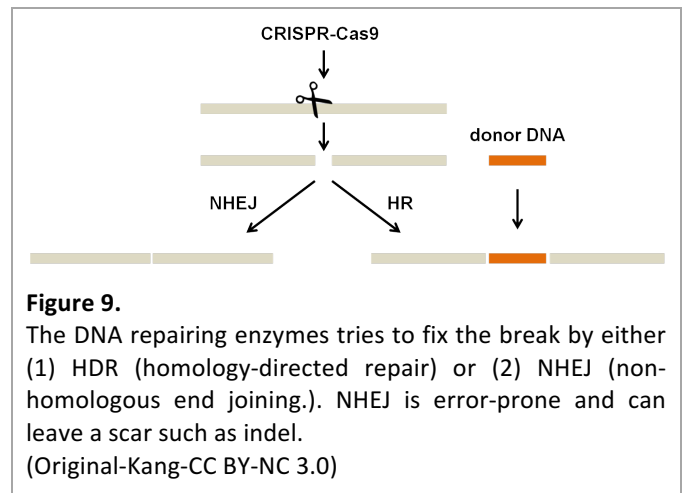
Note that we can also tweak the CRISPR system and use it for other purposes. For example, we can mutate few amino acids of the Cas9 protein so that it won't make any cuts. Instead, it can be accessorized with regulatory elements and can activate or repress certain gene expressions.

Note that this new technology is not perfect:
(1) we still have to solve off-target effects, which is CRISPR-Cas9 cutting at undesired, non-target sequences, and
(2) we need to find a way to effectively deliver all the RNAs to all of the target cells, if this is to be used as a means of gene therapy.

Nevertheless, CRISPR technology is just in its infancy and current undergraduate students will be improving and expanding this methodology in the future. It is the PCR of the current age.

_____

SUMMARY:

- There are a variety of model organisms that are used for genetic experimentation because they have advantages for various aspects of research.
- Research on one model organism can be applied to others. This permits genetic knowledge of model organisms to be transferred to humans, farm animals or crop plants.
- Transgenic organisms contain foreign DNA that has been introduced using biotechnology to make genetically modified organisms.
- Crispr-Cas9 is a RNA guided endonuclease that can find a specific sequence on DNA, make a cut, and the repair mechanism of the host cell will either introduce a mutation or integrate a new copy.

KEY TERMS:

model organisms
Escherichia coli
Sacchromyces cerevisiae
Caenorhabditis elegans
Drosophila melanogaster
Mus musculus
Danio rerio
Arabadopis thaliana
transgenic organisms
transgene
GMO
transformation
transfection
naked DNA
naked DNA
carrier
electroporation
microinjection
vector

vesicles
stable
position effects
Agrobacterium-mediated
transformation
Ti plasmid
T-DNA
callus
recalcitrant
particle bombardment
stem cells
knock-out
neo-R
germline
somatic
ex vivo
in vivo
non-integrative
siRNA

lentiviruses
genome editing
CRISPR-Cas9 system
cas
CAS9
crRNA
direct repeat
spacer
tracrRNA
sgRNA
adaptation
protospacer
crRNA Biogenesis
PAM
HDR
NHEJ
indel

Study Questions:

1) **a)** List the characteristics of an ideal model organism.

b) Which model organism can be used most efficiently to identify genes related to:

i)   eye development

ii) skeletal development

iii) photosynthesis

iii) cell division

iv) cell differentiation

v)   cancer

Notes:

CHAPTER 41– CANCER GENETICS
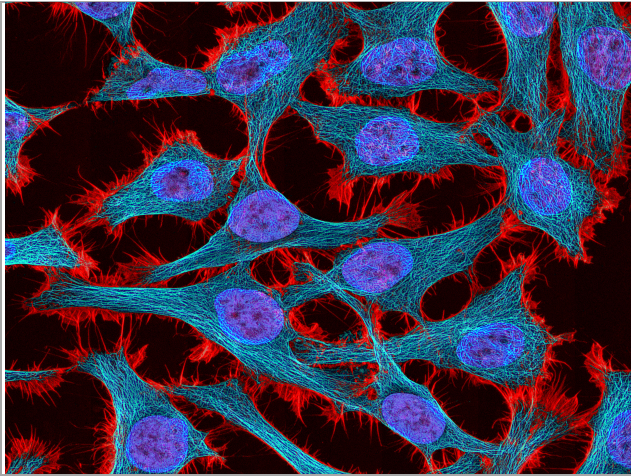
Figure 1.
Fluorescent image of HeLa cells stained for actin binding toxin phalloidin (red), microtubules (cyan) and cell nuclei (blue). HeLa cells are a line of immortal cultured cells derived from a cervical cancer taken from Henrietta Lacks in 1951.
(Wikipedia-CFCF, NIH-PD)

INTRODUCTION

Cancer is a group of diseases that exhibit uncontrolled cell 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 normally control cell division and limit abnormal growth have been disrupted, usually by the accumulation of several mutations in specific genes. 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.

1. CLASSIFICATION OF CANCERS

Cancers can be classified based on the tissues they resemble and thus in which they originate. For example, **Sarcomas** are cancers that originate in mesenchymal cells, such as bone, cartilage, fat, or muscle.

Carcinomas originate in epithelial cells (both inside the body and on its surface) and are the most common types of cancer (~85%). This includes glandular tissues (e.g. breast, prostate). 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.

Lymphomas arise from hematopoietic (blood forming) cells. This includes leukemia, the most common type of cancer in children.

2. CANCER CELL BIOLOGY

Cancer is a progressive disease that usually begins with increased frequency of cell division (**Figure 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. This increased cell division and abnormal tissue organization is called **dysplasia**. Eventually a tumour develops, which can grow rapidly and expand into adjacent tissues. As cellular damage accumulates and additional growth control mechanisms are lost, some cells may break free of the primary tumour, pass into the blood or lymph system, and be transported to another organ, where they develop into new tumours (**Figure 3**). The early detection of tumours is important so that they can be treated or removed before the onset of metastasis, but note that not all usually considered life threatening. In contrast, **malignant** tumours become invasive, and ultimately result in cancer.
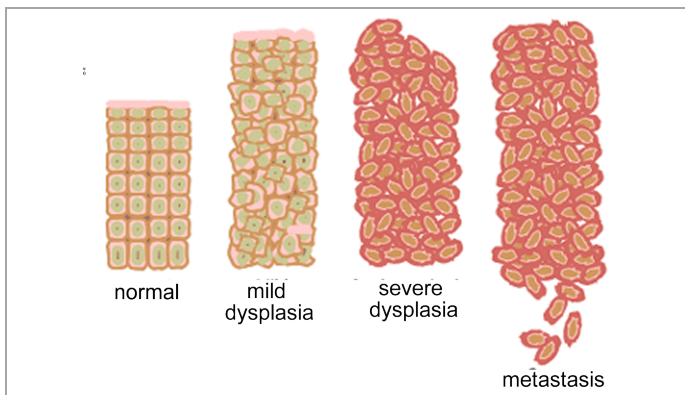
Figure 2.
Progressive increases in cell division and abnormal cell morphology associated with cancer. (Wikipedia-NIH-PD)
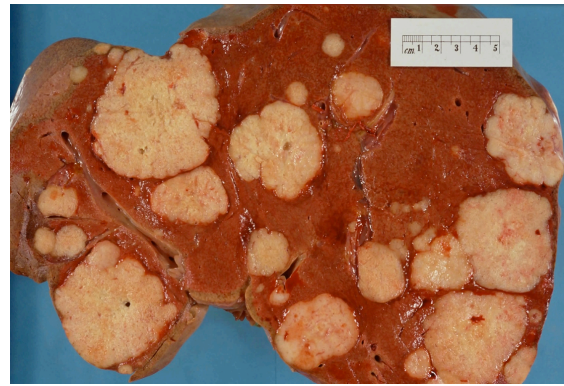
Figure 3.
Secondary tumours (white) develop in the liver from cells of a metastatic pancreatic cancer. (Wikipedia- J. Hayman-PD)

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 tumour. 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.

Table 1
Ten hallmarks of Cancer (Hanahan and Weinberg, 2000; Hanahan 2011)

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 1. In this chapter, we will focus on the first two hallmarks, namely growth signal autonomy and insensitivity to anti-growth signals.

4. MUTAGENS AND CARCINOGENS

A **carcinogen** is any agent that directly increases the incidence of cancer, while a **mutagen** is an agent that increases the incidence of mutations. 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.

4.1. HUMAN PAPILLOMA VIRUS (HPV)
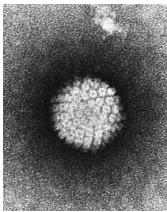
Figure 4.
Electron micrograph of HPV. (Laboratory of Tumor Virus Biology,-Unknown-PD)
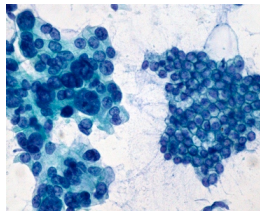
Figure 5.
Dysplastic (left) and normal (right) cells from a Pap smear.
(Flickr-Ed Uthman-CC BY 2.0)

Human Papilloma Virus (**HPV**, **Figure 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 tumour 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 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 develops cancer. Although most HPV-related cancers are cervical, HPV infection can also lead to cancer in other tissues, in both women and men.

4.2. IONIZING RADIATION

Ionizing 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 are required before cancer develops.

4.3. CHEMICAL CARCINOGENS

Chemical carcinogens (Table 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.

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

Table 2
Some classes of chemical carcinogens (Pecorino 2008)

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.

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.

Figure 6.
Structure of the ras protein.
(Wikipedia-Mark"AbsturZ"-PD)
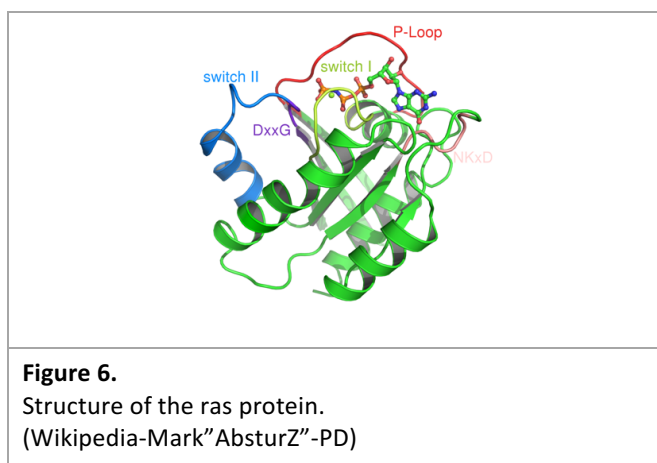
ras is an example of a proto-oncogene (**Figure 6**). *ras* acts as a switch within signal transduction pathways, including the regulation of cell division. When a receptor protein receives a signal for cell division, the receptor activates *ras*, which in turn activates other signaling components, ultimately leading to activation of genes involved in cell division. Certain mutations of the *ras* sequence cause 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 7**.
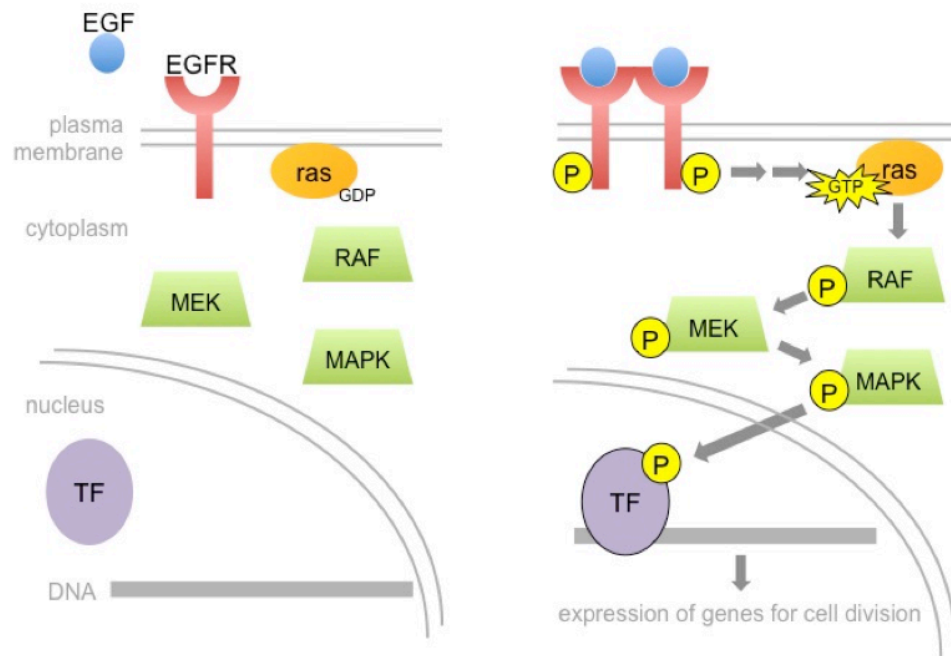
Figure 7.
Simplified representation of the epidermal growth factor (EGF) signaling pathway. In the panel on the left, the components are shown in their inactive forms, prior to stimulation of the pathway. The components include the soluble ligand, EGF, its receptor (EGFR, a tyrosine kinase), ras (a G 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 phosphorylated 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)

6. TUMOUR SUPPRESSOR GENES

More than 30 genes are classified as **tumour 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 tumour suppressors it is loss-of-function mutations that contribute to the progression of cancer. This means that tumour 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 tumour suppressor genes are more likely than oncogenes to be inherited. An example is the tumour 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 tumour suppressor genes.

An important tumour suppressor gene is a transcription factor named **p53** (**Figure 8**). 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 tumour suppressing functions, including DNA repair, cell cycle arrest, and apoptosis. Over 50% of human tumours 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).
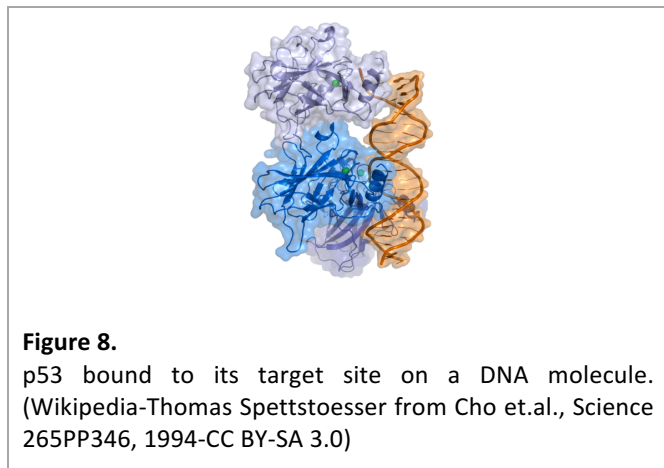
Figure 8.
p53 bound to its target site on a DNA molecule. (Wikipedia-Thomas Spettstoesser from Cho et.al., Science 265PP346, 1994-CC BY-SA 3.0)

7. GLEEVEC™ (IMATINIB) - THE "POSTER BOY" OF GENETIC RESEARCH LEADING TO A CANCER TREATMENT

7.1. CHRONIC MYELOGENOUS LEUKEMIA (CML)

Chronic myelogenous leukemia (**CML**) is a type of cancer of white blood cells, myeloid cells, which are mutated and proliferate uncontrollably through three stages (chronic, accelerated, and blast crisis) and lead eventually to death. Cytogenetics showed the myeloid cells of CML patients usually also have a consistent chromosome translocation (the mutant event) between the long arms of chromosomes 9 and 22, t(9:22)(q34;q11). It is also known as the **Philadelphia chromosome** (Ph$^+$). This translocation involves breaks in two genes, *c-abl* and *bcr*, on chromosomes 9 and 22, respectively. The fusion of the translocation breaks result in a chimeric gene, called ***bcr-abl***, that contains exons 1 and/or 2 from *bcr* (this varies from patient to patient) and 2-11 from *abl* and it produces a chimeric protein (BCR-ABL or p185$^{bcr-abl}$) that is transcribed like *bcr* and contains *abl* enzyme sequences. This chimeric protein has a tyrosine-kinase from the *abl* gene sequences that is unique to the CML mutant cell. The consistent,

unregulated expression of this gene and its kinase product causes activation of a variety of intracellular signaling pathways, promoting the uncontrolled proliferative and survival properties of CML cells (the cancer). Thus, the BCR-ABL tyrosine kinase enzyme exists only in cancer cells (and not in healthy cells) and a drug that inhibits this activity could be used to target and prevent the uncontrolled growth of the cancerous CML cells.

7.2. INHIBITING THE BCR-ABL TYROSINE KINASE ACTIVITY

Knowing that the kinase activity was the key to treatment, pharmaceutical companies screened chemical libraries of potential kinase inhibitory compounds. After initially finding low potency inhibitors, a relationship between structure and activity suggested other compounds that were optimized to inhibit the BCR-ABL tyrosine kinase activity. The lead compound was STI571, now called **Gleevec™** or imatinib (**Figure 9**).
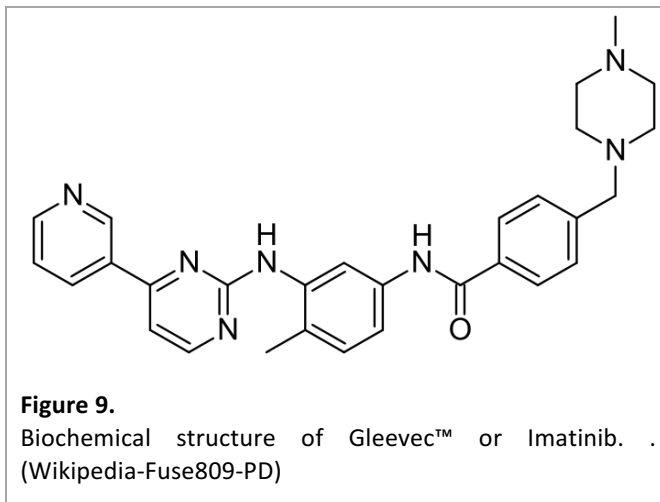
Figure 9.
Biochemical structure of Gleevec™ or Imatinib. . (Wikipedia-Fuse809-PD)

This drug was shown to inhibit the BCR-ABL tyrosine kinase activity and to inhibit CML cell proliferation *in vitro* and *in vivo*. Gleevec™ works via targeted therapy—only the kinase activity in cancer cells was targeted and thereby killed through the drug's action. In this regard, Gleevec™ was one of the first cancer therapies to show the potential for this type of targeted action. It was dependent upon the genetic identification of the cause and protein target and is often cited as a

paradigm for genetic research in cancer therapeutics.

7.3. CAUTION

This is a simplified presentation of the CML/cancer targeting by the drug Gleevec™. There are many more details than are presented here. This story represents as a model of finding a drug for each type of cancer, rather than the one, single "magic bullet" that kills all cancers. Remember, there are always complexities in this type of research-to-treatment process, such as patient genetic and environmental variation that leads to differences in drug metabolism, uptake, and binding. Also, changes in drug dose, mutation of the *bcr-abl* gene, and other events can affect the effectiveness of the treatment and the relapse rate. Biological systems are extremely complex and difficult to modulate in the specific, targeted manner necessary to treat cancer ideally.

Remember, the drug, Gleevec™, is not a cure, but only a treatment. It prevents the uncontrolled proliferation of the CML cells, but doesn't kill them directly. The arrested cells will die eventually, but there is always a small pool of CML cells that will proliferate if the drug is discontinued. While sustained use of this expensive drug may be financially beneficial to the pharmaceutical companies, it is certainly not the ideal situation for the patient.

_____

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 hyper activated regulators of cell division, and are often derived from gain-of-function mutations in proto-oncogenes.

- Tumour 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.

- Genetic research into cancer can provide enzyme targets for drug investigation and potential treatment. e.g. Gleevec™

KEY TERMS:

metastasis
sarcoma
carcinoma
squamous cell carcinoma
basal cell carcinoma
melanoma
lymphoma
dysplasia
benign
malignant
carcinogen
HPV
oncogene
ionizing

epidemiology
proto-oncogene
receptor
signal transduction
ras
tumour suppressors
apoptosis
BRC1A
p53
phosphorylation
CML
Philadelphia chromosome
bcr-abl
Gleevec™

STUDY QUESTIONS:

1) Why do oncogenes tend to be dominant, but mutations in tumour suppressors tend to be recessive?

2) What tumour suppressing functions are controlled by p53? How can a single gene affect so many different biological pathways?

3) Are all carcinogens mutagens? Are all mutagens carcinogens? Explain why or why not.

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?

5) Do all women with HPV get cancer? Why or why not?

6) Do all women with mutations in *BRCA1* get cancer? Why or why not?

Answers

for
Open Genetics Lectures
Chapter Questions

Fall 2017 Version

CHAPTER QUESTION – ANSWERS

CHAPTER 01 – ANSWERS

1) The genetic mechanism could be either they blend together like paint (they could not be separated again), they could be particulate, like DNA and genes, or it could be something else.
   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 exact 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.
2) 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.
3) The experiments shown in Figure 3 show that DNA is <u>necessary</u> for transformation, (since removing the DNA by nuclease treatment removes the competency for transformation). However, this does not demonstrate that only 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, while R Strain DNA could not.
4) An analogy for the particulate model could be mixing sterile broth (recessive) with inoculated broth (dominant). What others can you think of?

CHAPTER 02 – ANSWERS

1) **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.
   a) **b)** Avery and colleagues performed experiments, while Watson and Crick mostly analyzed the data of others and used that to build models.
   b) 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.
2) 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).
3) **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.
   a) The complementarity of the bases and the redundant nature of the strands.
   b) The order of the bases.
4) **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.
5) The ends of eukaryote chromosomes, telomeres, are in a constant state of flux, in that they potentially change with each replication cycle. Telomerase can add repeats to the end, while the lack of telomerase activity at the end of a chromosome can result in the loss of repeats.

CHAPTER 03 – ANSWERS

1) Mutant strain #1 has a mutation in gene B (but genes A and C should be functional).
   Mutant strain #2 is in gene C (but genes A and B should be functional).
   Mutant strain #3 is in gene A (but genes B and C should be functional).
2) Even prototrophs cannot produce the vitamin biotin, so it must be added for any strain to grow. Wild type strains also lack the enzyme(s) for this biochemical pathway. Biotin is present in Complete Medium.
3) No, we now know that genes also encode tRNA, rRNA, and a variety of other functional RNAs.
4)
   a. Changes in many amino acids do not cause a change in function. A specific amino acid is not required at that site for function to occur.
   b. Changes in many amino acids can cause a minor loss in function. A specific amino acid at a site may be required for optimal function to occur.
   c. Changes in some amino acids can cause a complete loss in function. Many specific amino acid are required at specific sites for any function to occur (e.g. the active site within an enzyme).
   d. Any one of the amino acids changed in part (c) can result in a complete loss of function.
5) No, the gene can be transcribed into an mRNA and translated into a polypeptide, but the polypeptide is not functional because of a change in an amino acid.
6) Chain A has ~268, while chain B has 450. The entire enzyme has ~ 4 chains, two A and two B (a heterodimer).
7) row 1 orange, orange, orange
   row 2 white, orange, orange
   row 3 yellow, yellow, orange
   row 4 white, yellow, orange

CHAPTER 04 – ANSWERS

1)
   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).
2) There are many correct answers for this question. Here is one.

	1	2	3	4	5	
1	-					Mutant in locus (1,2)
2	-	-				Mutant in locus (1,2)
3	+	+	-			Mutant in locus (3)

4	+	+	+	-		Mutant in locus (4)
5	+	+	-	-	-	Mutant in locus (3 and 4)

3) The auxotrophic strain is mutant in one gene. This gene has both a *HindIII* and *XhoI* site within its sequence, but not an *EcoRI* site. Thus, the *EcoRI* library could contain a restriction fragment with an entire, intact gene, while the two other enzymes would break the gene into two fragments that would not be cloned together as a functional gene.

4) No. *E. coli* cells do not normally import proteins from their environments, thus none of the Enzyme A proteins would enter the cells to affect a rescue. If the product of Enzyme A was added, then it could rescue the strain, but only if the product could be taken up by the cells.

CHAPTER 05 – ANSWERS

1) Gene: a hereditary unit, that occupies a specific position (Locus) within the genome or chromosome, and has at least one specific effect on the phenotype of the organism, and can mutate into various allelic forms, and can recombine with other such units. (From A Dictionary of Genetics, King & Stansfield, Third edition, 1985.)

2) No, not all DNA in a genome is part of a gene, or is needed. Not all DNA has a function. Some (most?) DNA in higher eukaryote genomes consists of non-functional DNA, also known as "junk" DNA, or "garbage DNA and has no known function.

3) No. RNA transcripts also include tRNA, rRNA, as well as a whole plethora of other RNAs that function as RNA molecules and don't (can't) be translated into polypeptides.

4) The UTRs (untranslated regions) are the regions at the 5' and 3' ends of the mRNA transcript, outside the transcribed sequence (start to stop) that is not translated. These sequences often have regulating elements (short sequence stretches) that alter the mRNA's translation.

5) No. The origin of replication in *E. coli* (*oriC*) is an example. The centromeres and telomeres are also sequences that are not transcribed, but their change or loss can have an effect on the phenotype.

CHAPTER 06 – ANSWERS

1) All are low glucose – lactose is as specified
Legend:
+++   Lots of β-galactosidase activity (100%)
++   Moderate β-galactosidase activity (10-20%)
+   Basal β-galactosidase activity (~≤1%)
-     No β-galactosidase activity (0%)

--	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)

2) All are low glucose – lactose is as specified
   Legend
   +++   Lots of β-galactosidase activity (100%)
   ++   Moderate β-galactosidase activity (10-20%)
   +   Basal β-galactosidase activity (~≤1%)
   -      No β-galactosidase activity (0%)

+++   a. I^+, O^+, Z^+, Y^+ / I^+, O^-, Z, Y *(high lactose)*
--   b. I^+, O^+, Z^+, Y^+ / I^+, O^-, Z, Y (no lactose)
+++   c. I^+, O^+, Z, Y^+ / I^+, O^-, Z^+, Y^+ (high lactose)
++   d. I^+, O^+, Z, Y^+ / I^+, O^-, Z^+, Y^+ (no lactose)
+++   e. I^+, O^+, Z, Y^+ / I^-, O^+, Z^+, Y^+ (high lactose)
--   f. I^+, O^+, Z, Y^+ / I^-, O^+, Z^+, Y^+ (no lactose)
+++   g. I^-, O^+, Z^+, Y^+ / I^+, O^+, Z, Y^+ (high lactose)
--   h. I^-, O^+, Z^+, Y^+ / I^+, O^+, Z, Y^+ (no lactose)
+++   i. I^+, 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)

3) You could demonstrate this with just $I^+O^cZ/I^+O^+Z^+$. The fact that this does not have constitutive β-galactosidase expression shows that the operator only acts on the same piece of DNA on which it is located. There are also other possible answers.

4) 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 shows 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.

5) For all of these, the answer is the same: The l*ac* operon would remain inducible by lactose, but only up to a basal level of expression, even in the absence of glucose

---

CHAPTER 07 – ANSWERS

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

2) 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.

3) 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 08 – ANSWERS

1) Various roles could include varying abilities to bind O_2. Greater need as an embryo and fetus, less so for an adult. This could be tested by obtaining the blood from those stages and determining the blood's ability to bind O_2.

2) The two main types (α-globins and β-globins) are both derived by gene duplication and evolution from a common ancestral gene.

3) Cartoon:

α-globin and β-globin, two of each type

4) Look it up. If you're too lazy to do a search, see:
   http://www.bloodjournal.org/content/125/24/3694?sso-checked=true

5) It is caused by the LCR keeping the gamma genes on after a person is born rather than switching to the beta and delta genes (see Figure 5). A person will usually be unaware they have this condition.

CHAPTER 09 – ANSWERS

1) Heterozygous people have the lactose persistence phenotype. As infants both alleles are active. As adults only the LP allele remains on but it continues to supply the intestinal epithelial cells with Lactases. LP alleles are dominant to the original allele because the LP allele's phenotype is the one seen when both alleles are together.

2) The gene's alternative symbols are *LAC* and *LPH*. It is important to have a single symbol to make searches of published journal articles possible. It is also a guide to authors of future articles, texts, and lectures.

3) If a person has a Lactase persistence phenotype they will break down the lactose and import the glucose (and galactose). Soon most of the glucoses will enter their circulatory system and cause a noticeable increase in blood glucose levels. The increase is noticeable after 15 minutes and peaks at 45 minutes. If a person has a non-persistence phenotype none of this will happen because the lactose will pass through their digestive tract (although some of it will be consumed by gut bacteria).

4) An Insulin protein begins with a ER signal sequence. Once their Ribosome has been delivered to the ER it can be removed. The Ribosome continues synthesizing the Insulin protein and feeding it into the ER lumen. Insulin proteins do not have a stop transfer sequence so the entire protein will be released into the ER lumen. Insulin proteins do have a pro sequence, it is removed once the protein has taken on its proper shape. When a transport vesicle carrying the Insulin fuses with the plasma membrane the protein is release from the cell and can enter the blood.

5) The major difference is *E. coli* cells imports lactose and then hydrolyse it while human intestinal cells do these steps in the order hydrolyse first and import second. The major similarities are in the proteins required: Lac Permease (*E. coli*) and SGLT1 (humans) are both carbohydrate transporters located in the

plasma membrane while Beta-Galactosidase (*E. coli*) and Lactase (humans) are both enzymes that hydrolyze the dissaccharide lactose.

6) There are enzymes to hydrolyse each dissaccharide and transport proteins to import the resulting monosaccharides. These proteins happen to be named Lactase, Sucrase, Maltase, SGLT1 (imports glucose and galactose), and GLUT5 (imports fructose).

7) Firstly, the excess lactose upsets the osmotic balance in the large intestine. Water enters the gut from the tissues leading to diarrhea and dehydration. Secondly, the lactose will be used as food by bacteria living in the large intestine. When they use the lactose to make ATP they expel carbon dioxide, hydrogen, and other gases which causes cramping and flatulence.

CHAPTER 10 - ANSWERS

1) A geneticist would use these as white = phenotype, *white* = gene, and WHITE = protein.
2) The world is 19 times brighter for these flies. Without the optical insulation provided by the pigments light from all directions strikes all of the photoreceptors. The flies are unable to make sense of the information their eyes send to their brains.
3) The flies would be unable to make either transporter and would have white eyes as a result.
4) Yes. Most noticeable is the male flies have difficulty performing the mating dance that leads to sex with female flies.

CHAPTER 11 – ANSWERS

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.
2) Misreading of bases during replication can lead to substitution and can be caused by things like tautomerism, DNA alkylating agents, and irradiation.
3) Looping out of DNA on the template strand during replication; strand breakage, due to radiation and other mutagens; and (discussed in other chapters) chromosomal aberrations such as deletions and translocations.
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.
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.
6) See Chapter 10.
7) Class I. see Figure 9 on Transposable Elements.

CHAPTER 12 – ANSWERS

1)

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).

2)

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.
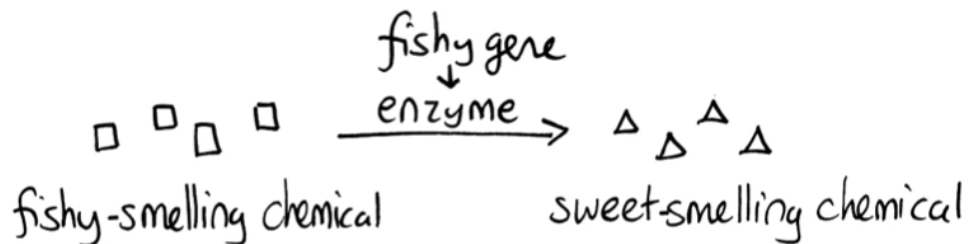
Note that although we show this biochemical pathway as leading from the fishy-smelling chemical to the sweet-smelling chemical in one step, it is likely that there are many other enzymes that act after the fishy enzyme to make the final, sweet-smelling product. In either case, blocking the pathway at 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:
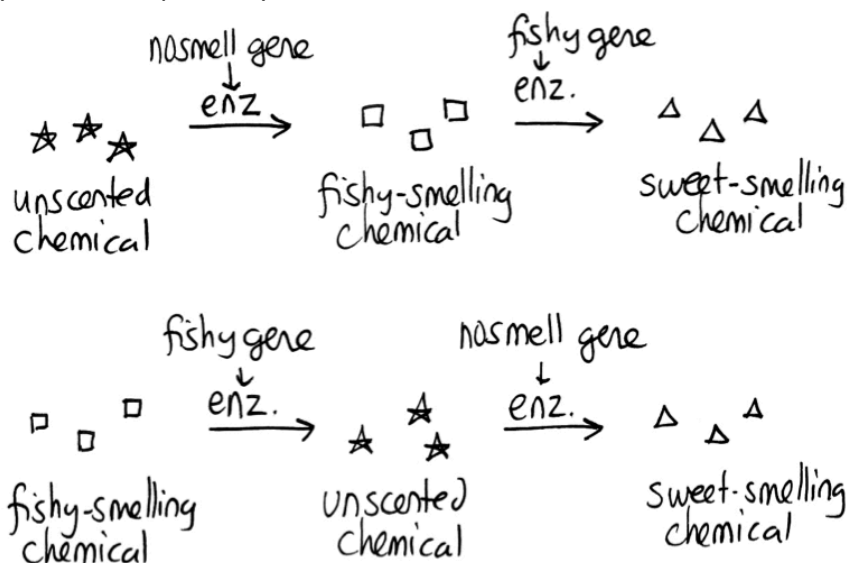
Alternatively, *nosmell* may not be part of the biosynthetic pathway for the sweet smelling chemical at all. It is possible that the normal function of this gene is to transport the sweet-smelling chemical into the cells from which it is released into the air, or maybe it is required for the development 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.

3)
   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)
   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).

CHAPTER 13 – ANSWERS

1) These are four common terms that are often used interchangeably by novice students, but do have distinctly different meanings and uses. (1) gene = general term for a segment of nucleic acid that is responsible for one or more phenotypes (2) locus = the position of a gene along a chromosome, (3) allele = the form (DNA sequence) of a gene at a locus, (4) transcription unit = the segment of DNA that is transcribed into RNA (often mRNA in the case of a protein coding gene).

2) Form (1) RR (red) x rr (white) gives Rr (red progeny). "R" is dominant to "r".

   Form (2) r^+r^+ (red) x r^-r^- (white) gives r^+r^- (red progeny). "r^+" is dominant to "r^-".

   For pink progeny, the symbols are the same, only "R" or "r^+" is semi-dominant to "r" or "r^-".

3) 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.

4) case 1 co-dominance

      case 2 incomplete-dominance

      case 3 incomplete penetrance

      case 4 pleiotropy

      case 5 haplo-sufficiency

      case 6 haplo-insufficiency

      case 7 broad (variable) expressivity

5) Mutant#1 = hypomorph

    Mutant#2 = hypermorph

    Mutant#3 = amorph

    Mutant#4 = neomorph

    Mutant#5 = antimorph

CHAPTER 14 – ANSWERS

1) No. Since chromosomes vary greatly in size, the number of chromosomes does not correlate with the total DNA content. For reasons discussed in Chapter 5 and this chapter, the number of genes does not correlate closely to DNA content either.

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

3)

    a. Only one (except for holocentric chromosomes, not discussed in this chapter).

    b. The two centromeres might get pulled towards opposite poles at mitosis/meiosis resulting in chromosome breakage.

    c. It would not segregate properly at mitosis or meiosis, leading to aneuploidy. In order to segregate correctly, there would have to be another way to control its movement at mitosis and meiosis.

4)

    a. At the end of G_1, 16 chromosomes with 1 chromatid each.

    b. At the end of S, 16 chromosomes with 2 chromatids each.

    c. At the end of G_2, 16 chromosomes with 2 chromatids each.

    d. At the end of mitosis, 16 chromosomes with 1 chromatid each.

5)

    a. There is little correlation between any of these, with the exception that larger genomes _tend_ to have more genes.

    b. The C-value paradox can be explained by genomes having different amounts of non-coding DNA between genes and within genes as introns.

    c. If we define "organismal complexity" as the size of the genome (or number of cells/organism), then larger, more complex organism _tend_ to have more genes although not always and not in a direct, linear, proportioned manner. Also, those with larger genomes _tend_ to have greater distances between genes.

CHAPTER 15 - ANSWERS:

1)

    a) Red blood cells do not have chromosomes. They are terminally differentiated and have expelled their nucleus.

    b) First, it is difficult to collect cells in anaphase. Second, in anaphase there would be twice as many chromosomes, which would make identifying them much harder.

2) Yes. Males being 46,XY have slightly less DNA than 46,XX females, but still have the same number of chromosomes.
3)
   a) True.
   b) True.
   c) False, only females have a paternal X chromosome.
   d) True.
   e) False, only males have a paternal Y chromosome.
   f) False, no one has a maternal Y chromosome. Females don't have Y-chromosomes
   g) False, typically no one has a paternal mitochondrial chromosome. Mitochondria are maternally inherited. However, there are rare cases of inheritance of paternal mitochondria.
   h) True.
4) Centromeres function as a "chromosome's handle". Each needs one handle but it doesn't matter where along the chromosome it is.
5) If there was only one ori in the middle of the chromosome it would take too long for the replication forks to reach the ends of the chromosome. Even with thousands of ori's per chromosome, it still takes 8 hours to replicate our DNA.
6) Chromatin is the material from which chromosomes are made (mostly DNA + protein). DNA is a component of both chromatin and chromosomes.
7) Top left: Histone proteins; top right: Histone and Cohesin proteins, bottom right: Histones, Cohesins, Condensins, and Kinetochore proteins; bottom left: Histone, Condensin, and Kinetochore proteins.
8)
   a) DNA Polymerases are found inside the nucleus and the mitochondria.
   b) RNA Polymerases are found inside the nucleus and the mitochondria
   c) Ribosomes are found free in the cytosol, on the surface of the rough ER, and inside the mitochondria. (Some have been found in the nucleus, too.)
9)
   a) The *F8* gene could work on an autosome. Its mRNAs would still leave the nucleus to be translated in the cytosol.
   b) The *SRY* gene would not work normally on an autosome because then females would have the gene as well as males and thus females would become males.
   c) The *MT-CO1* gene would not work on an autosome (nuclear) because the genetic code is different in the nucleus (vs. mitochondrion). The protein must be translated inside the mitochondria to be the correct amino acid sequence. If it were translated in the cytosol the amino acid sequence would be different, and thus likely not work normally.

CHAPTER 16 – ANSWERS

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, not white.
2) Your choice……
3) There is a maximum of two alleles for a normal autosomal locus from a diploid individual. In the whole population there can be essentially an unlimited number of different alleles; the limit being determined by the population size.
4)

    a. In the F1 generation, the genotype of all individuals will be Ww and all of the dogs will have wirey hair.

    b. In the F2 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.

5) 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.

6) 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 in the parent cell.

CHAPTER 17 – ANSWERS

1)

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

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

2) No, it's not necessary to write out a Punnett square in a true square 2x2 or 4x4, etc. For simplicity you can remove the duplicate gametes, and you will still get the same ratio. It isn't incorrect to write it out fully

though. For the Punnett square on the right **Figure 7**, you can simplify it as:

♀ ⁄ ♂	R;Y	R;y
r;y	R/r Y/y	R/r y/y

3) The "9" would increase, both "3" would decrease, and the "1" would increase.
4) Two classes, the <u>parentals</u>, would increase, while two classes would decrease, the <u>recombinants</u>.

CHAPTER 18 – ANSWERS

1) Crossovers can be observed cytologically directly under the microscope as chiasmata.
   Recombination is defined genetically as the frequency calculated from the observed phenotypic proportions in the progeny.
   Crossovers lead to recombination when they are detected using genetic marker loci. Not all crossovers result in recombination – some can't be detected because no visible markers are recombined.
   Some recombinants involve crossovers, but not all recombinants result from crossovers.
   Crossovers between non-sister chromatids can result in recombination, while crossovers between sister chromatids, which have identical alleles, will not show any recombination.
   When there are two crossovers between the loci being scored for recombination, the result will appear to be parental, not recombinant.
   Recombination can occur without crossovers when marker loci are on different chromosomes, which then assort independently.
2) The use of pure breeding lines allows the researcher to be sure that he/she is working with homozygous (known) 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.
3) This tight linkage would suggest that individuals with the earlobe phenotype would likely carry 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.
4)
   a. It assumes that the loci are completely unlinked.
   b. The expected ratio would be all partentals and no recombinants. For example, 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.
5)
   a. Parental: *CcEe* and *ccee;* Recombinant: *Ccee* and *ccEe*.
   b. Parental: *Ccee* and *ccEe;* Recombinant: *CcEe* and *ccee*.
6) a)- Let *WwYy* be the genotype of a purple-flowered (*W*), green seeded (*Y*) dihybrid. The cross is *WwYy* × *wwyy.* Half of the progeny will have yellow seeds whether the loci are linked or not. You cannot tell if they are linked or not given only this information.
   b)- You need to know the proportion of the seeds that are white or purple flowered, and in what frequencies they appear with the white and purple flowers, e.g. what the frequencies of the four classes

are. This would help you to know about the linkage between the two loci – unliked, or what degree of linkage.

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

AaBb 135   Aabb 430   aaBb 390   aabb 120

(135 + 120)/(135+120+390+430)= 24%

8) See section 3.3. Syntenic is the term for genes found on the same chromosome. Linked genes are always found on the same chromosome, and so are always syntenic. If the genes are sufficiently far enough away on the same chromosome, crossover events will make the two genes assort independently, so they won't appear linked. Therefore, in this latter situation, these genes are syntenic, but not linked.

CHAPTER 19 – ANSWERS

1) 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)

2) 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)

3) 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*).

4) 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%.

5)
   a. 4cM
   b. Random sampling effects; the same reason that many human families do not have an equal number of boys and girls.

6) 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)

A is fur color locus   B is tail length locus   C is behaviour locus

fur (A)	tail (B)	behavior (C)	Freq.		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
```
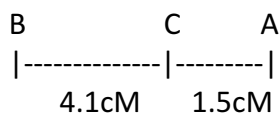
Pairwise recombination frequencies are as follows (calculations are shown below):

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 20 – ANSWERS

1) It depends on the chromosomal location of the disease locus. If the gene is autosomal, the probability is 50%. If it is sex-linked, that is on the X-chromosome, it would be 100%. If it is Y-linked, then 0%. In both situations the probability would decrease if the penetrance was less than 100%.

2)

3)

CHAPTER 21 – ANSWERS

1)

2) 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.

CHAPTER 22 – ANSWERS

1) Co-dominance
2) Note that a semicolon is used to separate genes on different chromosomes.

Phenotype	Genotype(s)
a) entirely black	$O^B / O^B ; s / s$ $O^B / Y ; s / s$
b) entirely orange	$O^0 / O^0 ; s / s$ $O^0 / Y ; s / s$
c) black and white	$O^B / O^B ; S / \_$ $O^B / Y ; S / \_$
d) orange and white	$O^0 / O^0 ; S / \_$ $O^0 / Y ; S / \_$
e) orange and black (tortoiseshell)	$O^0 / O^B ; s / s$
f) orange, black, and white (calico)	$O^0 / O^B ; S / \_$

3) People with hemophilia A use injections of purified Factor VIII proteins (made through the use of recombinant, cloned Factor VIII gene). It can be delivered on demand (to control existing bleeding) or regularly (to limit damage to joints).

Genotype: F8⁺/F8⁺ F8⁺/F8⁻ F8⁻/F8⁻ F8⁺/Y F8⁻/Y

Phenotype: [stick figures] ♀ ♀ ♀ ♂ ♂

🍶 = recombinant Factor VIII proteins

CHAPTER 23 – ANSWERS

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.

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).

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.

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) 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.

CHAPTER 24 – ANSWERS

   1)

a. As in Figure 4 homologous chromosomes pair during prophase I of meiosis. The shaded boxes are regions of sequence similarity, for example Alu transposable elements. A crossover occurs between two of the Alu elements on the same chromatid leading to a chromosomal inversion.
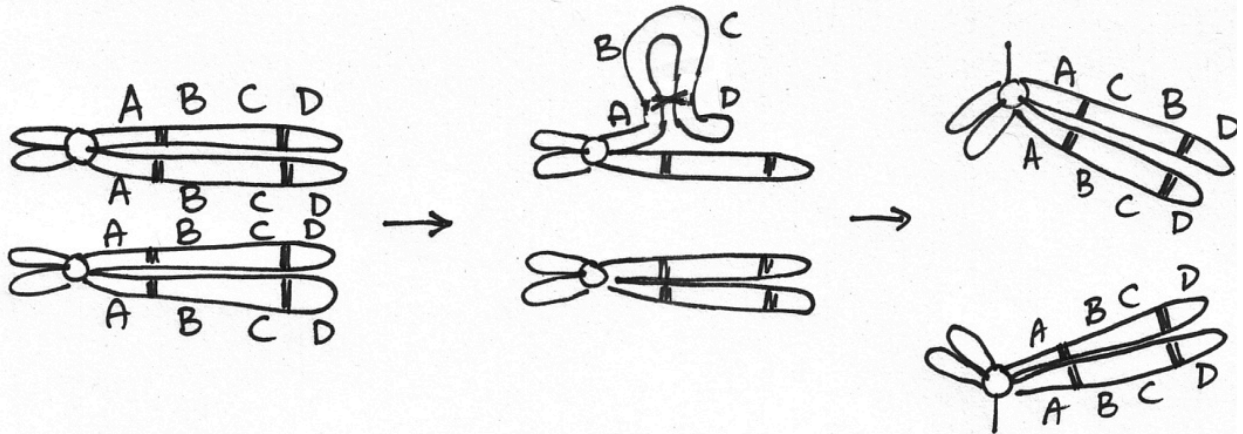
b.
A crossover occurs between Alu elements on different chromosomes leading to a chromosomal translocation. Note that the homologous chromosomes are not shown in this figure for simplicity.
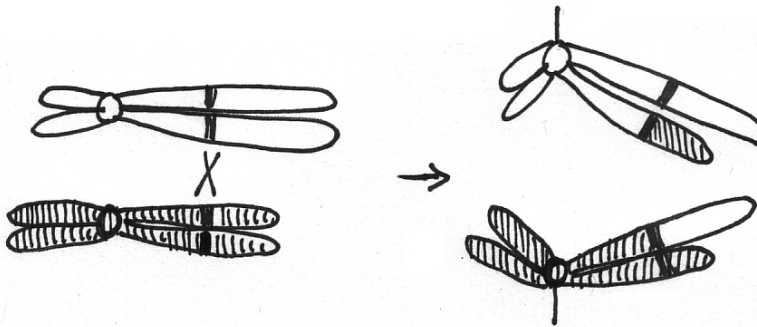
2) Gamma rays are efficient at causing double strand DNA breaks, which are then more likely to rejoin and produce a deletion.

3) First, obtain permission from the person (and ethical approval from the appropriate oversight board or committee). Next, isolate some white blood cells, place the cells on a slide, denature the DNA, hybridize with fluorescent nucleic acid probes specific for the X chromosome or the Y chromosome, observe the results with a fluorescence microscope. If they are XXX, there should be three X signals corresponding to the three X-chromosomes and no Y-chromosome signals. If they are XYY, there should be one X signal and two Y signals in each cell nucleus.

CHAPTER 25 - ANSWERS

1) 2n=6x=42

2)
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.

3)
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.

4) Maternal chromosomes are black and paternal chromosomes are grey.

a)

5)

a)

6)
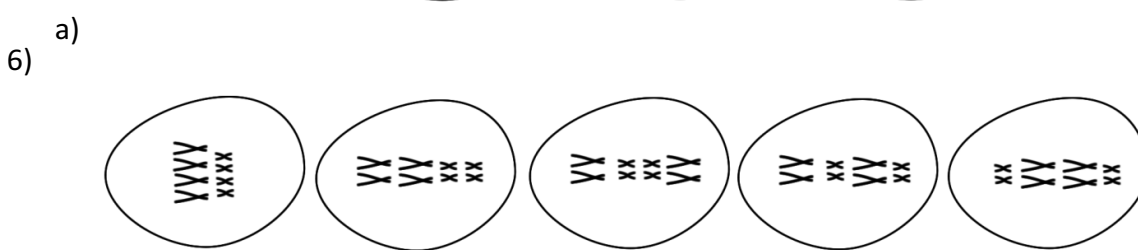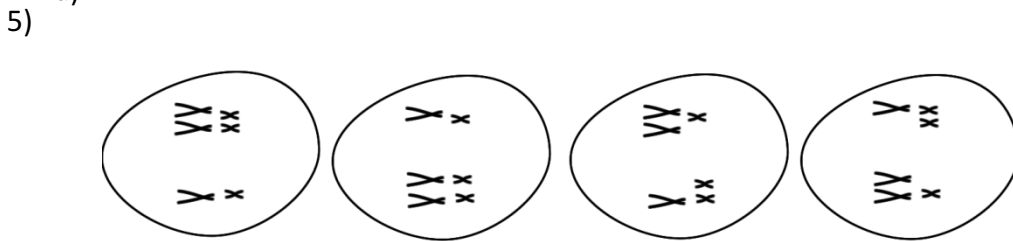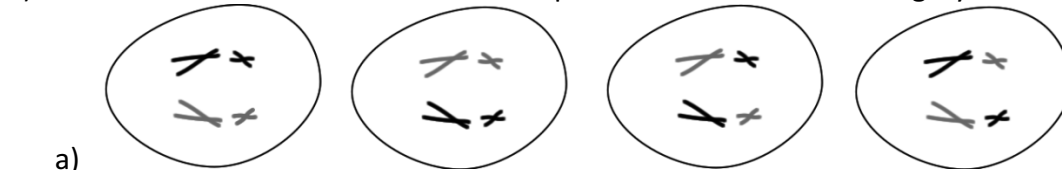
a)

7) As in Figure 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.
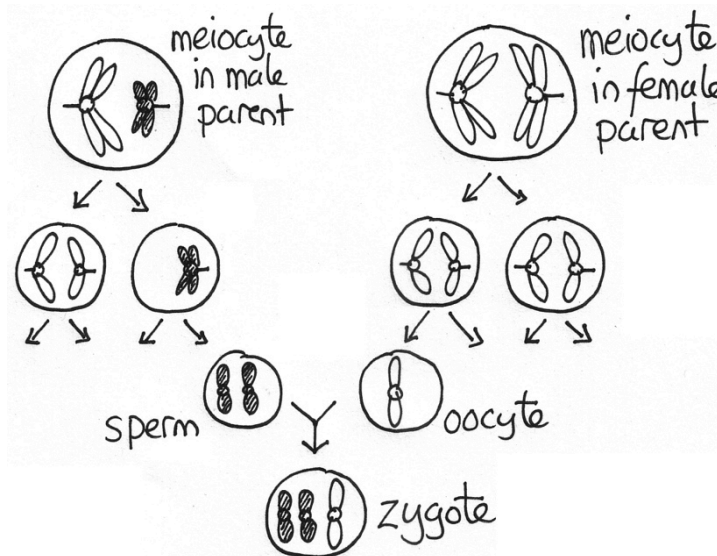
a)

8)

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) Having a shortage of key proteins is usually more detrimental than having an excess.

10) 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.

CHAPTER 26 – ANSWERS

1)
- **a)** case 1 co-dominance
- **b)** case 2 incomplete-dominance
- **c)** case 3 incomplete penetrance
- **d)** case 4 pleiotropy
- **e)** case 5 haplosufficiency
- **f)** case 6 haploinsufficiency
- **g)** case 7 broad (variable) expressivity

2) 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

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. 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   **i) 12 red : 4 blue**   **j) 12 red : 4 blue**

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

	BD	Bd	bD	bd
BD	BBDD	BBDd	BbDD	BbDd
Bd	BBDd	BBdd	BbDd	Bbdd
bD	BbDD	BbDd	bbDD	bbDd
bd	BbDd	Bbdd	bbDd	bbdd

	AD	Ad	aD	ad
AD	AADD	AADd	AaDD	AaDd
Ad	AADd	AAdd	AaDd	Aadd
aD	AaDD	AaDd	aaDD	aaDd
ad	AaDd	Aadd	aaDd	aadd

4)

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

5) Epistasis is demonstrated when the phenotype for a mutant in one locus is prevented from being expressed by a mutant at another locus. In this case, we would expect a homozygous mutant at one locus (e.g. D) to be the same phenotype as a homozygous mutant in both loci (e.g. D and A, or D and B).
So, the following situations from questions 2-4 demonstrated epistasis:

Q#2: No epistasis can be determined from the phenotypes (even though we know from the pathway provided that D is downstream of A and B). There are only two possible phenotypes. So even though the D locus might be epistatic to A and B, one cannot see this interaction because the product of both A and B (compound 3) is colourless.

Q#3: The phenotypes show that D is epistatic to A and B:
- *aadd* looks like *AAdd* or *Aadd*; dd prevents the expression of the A or a alleles.
- *bbdd* looks like *BBdd* or *Bbdd*: dd prevents the expression of the B or b alleles.
Note: that the triple mutant aabbdd would be colourless (white).

Q#4: The phenotypes show that D is epistatic to A, because aadd looks like AAdd or Aadd.
With bbdd, the difference between bbdd (green), Bbdd (blue), and BBdd (blue) is apparent, Thus, the phenotypes do not provide evidence for epistasis between B and D.

6) 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*

7) Conduct an enhancer/suppressor screen (which can also result in the identification of revertants, as well)

start with a population of mutants
gg

expose the population to a mutagen (e.g. feed them EMS)

g g* — some plants will be mutated again in the *g* gene (the allele with two mutations is marked *g**)

Aa gg — some plants will be mutated in a gene other than *g* (e.g., *a*, where wild-type is *A*)

gg — some plants will not have any additional mutations at all
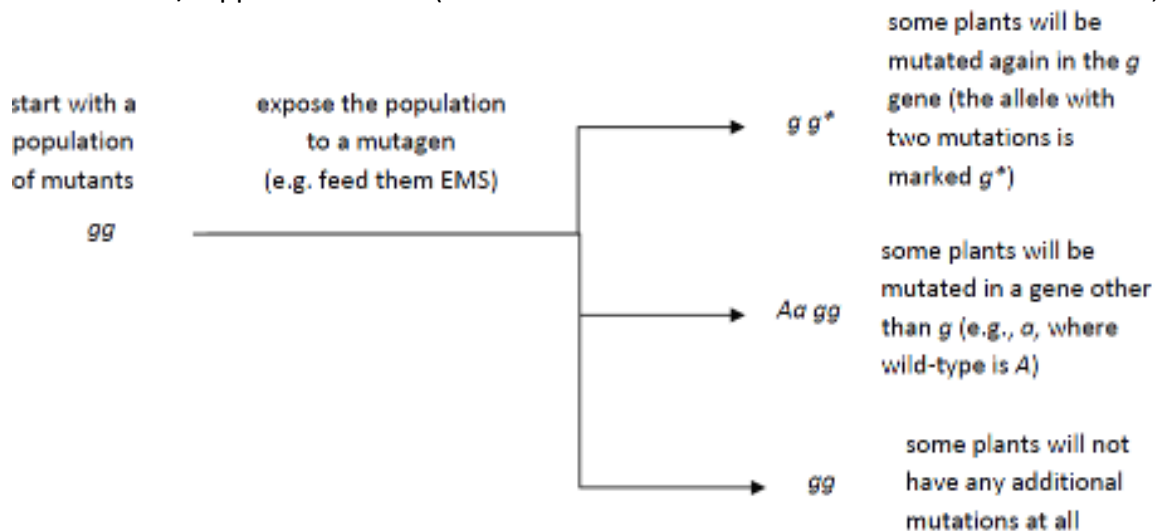
allow the plants to self-pollinate in order to make any new, recessive mutations homozygous

Parental: *gg** x *gg**

1 : 2 : 1

F$_1$: *gg* *g*g* *g*g**

for some mutations, individuals of this genotype will identify revertants of allele *g*

Aa gg x *Aa gg*

1 : 2 : 1

AA gg *Aa gg* (*aa gg*)

for some mutations, individuals of this genotype will identify recessive enhancers or suppressors of allele *g*; for other mutations, mutants in gene *a* could have an unrelated phenotype
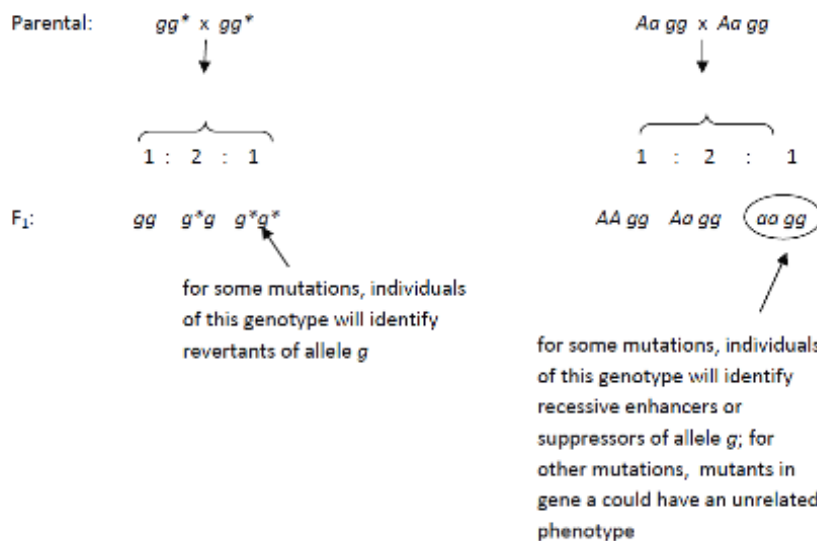
8) Depending which amino acids were altered, and how they were altered, a second mutation in *g*g** could either have no effect (in which case the phenotype would be the same as *gg*), or it could possibly cause a reversion of the phenotype to wild-type, so that *g*g** and *GG* have the same phenotype.

9) 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.

10) 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.

11) 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.

CHAPTER 27 – ANSWERS

1) 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%.

2)
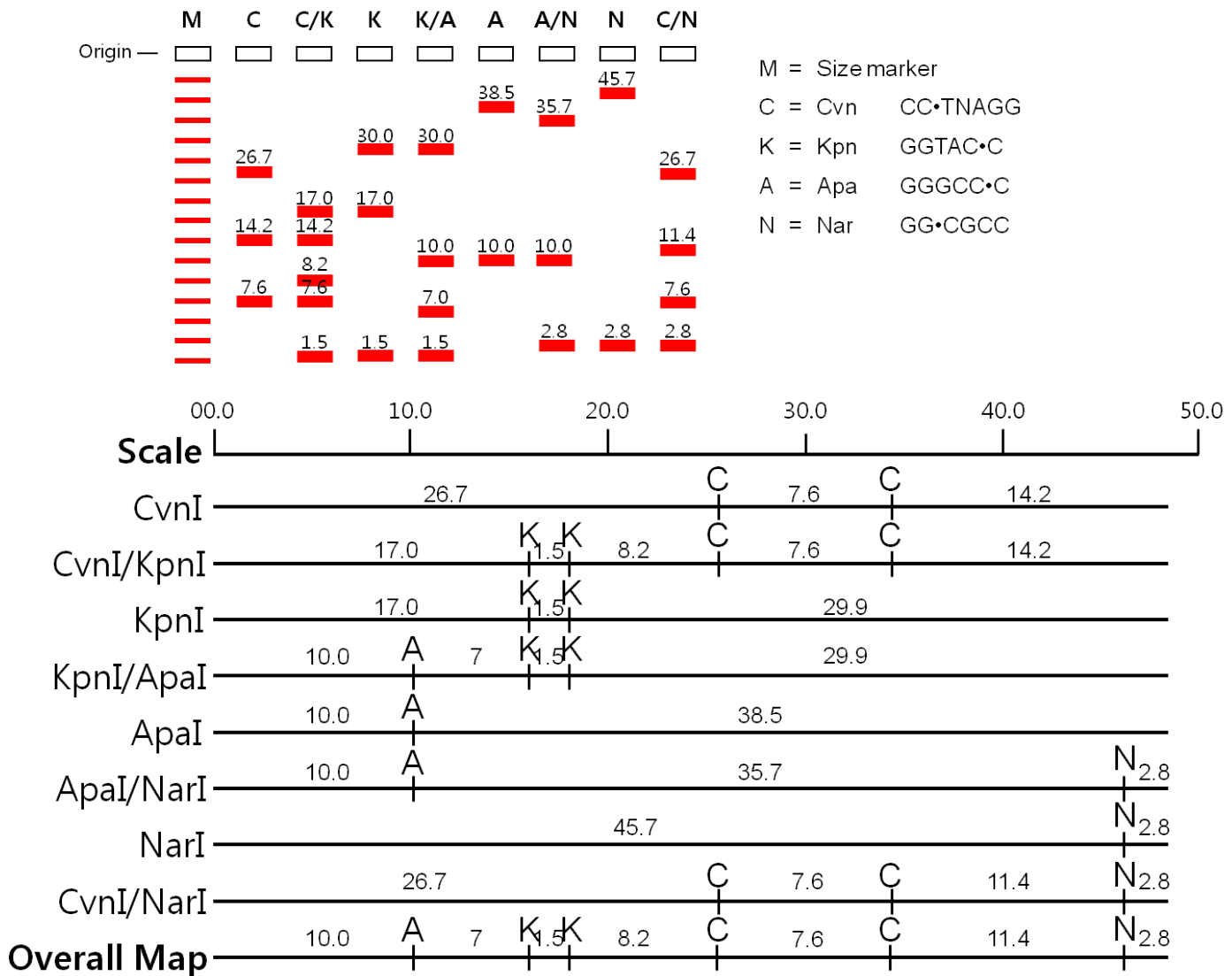
CHAPTER 28 – ANSWERS

1)
a) There will be a 6kb band (the insert) and a 3kb band (the plasmid vector)
b) There would be a single 9kb band.

c) There would only be a 3kb band, which represents three fragments: the plasmid, and both insert fragments. All are the same size = 3.0 kb, so they will appear on the gel as a single band.

2) The complementary, sticky ends of the insert and plasmid vector may anneal together, but the non-functional ligase will not be able to covalently link the insert and vector together. Thus, the annealed DNA fragments will not be stable enough to be transformed, and thus unable to replicate -> no transformants should be expected.

3) In electrophoresis, the main force driving moving of the molecules is the electrical force: that is, the amount of charge per mass. A duplex DNA molecule will have about twice as much charge per length as single stranded RNA, but it will have about twice as much mass. The difference being the deoxyribonucleic acid (DNA) will have one less Oxygen per base compared to ribonucleic acid (RNA) – DNA is 4% less than RNA per base on average. Thus for duplex DNA vs duplex RNA of the same sequence, DNA should run slightly faster. However, single stranded RNA (and DNA) is capable of intramolecular base pairing, which would dramatically reduce the molecule's length and thus increase its velocity through the gel matrix. Thus, a single stranded RNA will migrate faster than a double stranded DNA molecule (of the same length) in a typical agarose gel. The difference in velocity would be determined primarily by the amount and type of intramolecular base pairing.

CHAPTER 29 – ANSWERS

1) 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 without the fear of infecting with a whole virus.

2) 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 maintain the plasmid in the *E. coli* because the retention of the plasmid is dependent upon the antibiotic resistance selectable marker.

CHAPTER 30 - ANSWERS:

1) Use each sequence in an online BLAST search (e.g. http://blast.ncbi.nlm.nih.gov/Blast.cgi). See which is a 100% match to Drosophila and which is a 100% match to mouse.

2) a) The next three nucleotides are CC(ATCG). The third position can be any base.

a) The sequence options need to be described. Met is fixed at ATG. Lysine had two alternatives (AAA and AAG). Asparagine has two (AAT and AAC). Glutamic acid has two (GAA and GAG). Proline has four possible codons. Thus 2 x 2 x 2x 4 = 32 possible sequences.

CHAPTER 31 – ANSWERS

1) You would need to know that although HIV is an RNA-virus, you should be able to detect the DNA pro-virus in infected white blood cells. You would have to extract DNA from white blood cells, then use HIV-specific primers to detect 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 primer sequences 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.

2) 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.

3) At the end of a successful PCR amplification reaction will be polymerase, remaining dNTPs, and primers, as well as the original template, and PCR amplification products. By far the most abundant will be the amplification products, flanked at both ends by the primer sequences. These products will be the only thing observed on the gel, since the template and other PCR products are present in much lower abundance, too low to be seen as evidenced by the absence of bands in the negative control lane.

CHAPTER 32 - ANSWERS:

1) While both bind to DNA Giemsa is used with visible light microscopes because it absorbs green, yellow and red light (and thus appears purple) while DAPI is used with fluorescence microscopes because it absorbs UV light and emits blue light.

2) There would be only two chromosome 21s, there would be two X chromosomes and no Y, and one of the chromosome 5s would be shorter.

3) Both are *in vitro* DNA replication reactions using template DNA, primers, nucleotides, and DNA Polymerases. The differences are PCR uses only two sequence-specific primers while labelling reactions use millions of non-specific, PCR uses only regular nucleotides while labelling reactions include a labelled one, PCR reactions must use a heat resistant DNA Pol while labelling reactions don't require one. PCR reactions also go through about 30 cycles of (denature, anneal, and extend) while labelling reactions use a single cycle.

4) When nucleotides are incorporated into the growing strand the beta and gamma phosphates are discarded. The only way to make 32-phosphate labelled DNA is to use nucleotides with the alpha-phosphate be the radioactive isotope.

5) The results would be similar to Figure 9. 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.

6) These men are 47,XYY. This situation doesn't affect their health or fertility so it isn't a deleterious condition and it doesn't have a name. It can be detected with any of the techniques discussed in this chapter: Giemsa staining, G-banding, or FISH with hybridization probes that bind to the sex chromosomes.

7) These women are 47,XXX. This situation doesn't affect their health or fertility so it isn't a deleterious condition and it doesn't have a name. It can be detected with any of the techniques discussed in this chapter.

8) First, obtain permission from the person (and ethical approval from the appropriate oversight board or committee). Next, isolate some white blood cells, place the cells on a slide, denature the DNA, hybridize with fluorescent nucleic acid probes specific for the X chromosome or the Y chromosome, observe the results with a fluorescence microscope. If they are XXX, there should be three X signals corresponding to the three X-chromosomes and no Y-chromosome signals. If they are XYY, there should be one X signal and two Y signals in each cell nucleus.

CHAPTER 33 – ANSWERS

1) If only fluorescently-labelled ddNTPs (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.

2) You could extract raw, naked DNA from seawater in various different places in the world, then sequence all of this DNA, and build a database with the sequences. 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.

3) It is a reference to the fluorescently-labelled dideoxy nucleotides that are at the heart of the procedure.

4) Before 2008 sequencing a genome meant making BAC clones and millions of sequencing reactions. Today it is done with no cloning step and one reaction.

5) The first generation of automated sequencing machines, such as the ABI 3730, are still less expensive and more targeted when it comes to sequencing plasmids and PCR products. Many times a researcher only wants to know the sequence of a short stretch of DNA, not the whole genome.

6) True. Some other next-generation technologies are sequencing by ligation (used in the Applied Biosystems machines) and ion torrent sequencing (used in Life Technologies machines).

CHAPTER 34 – ANSWERS

1)
   e) 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 heterozygous for the deletion (affected with the cancer) will have two bands: one at the normal position and one at a lower position (200 bp lower) on the gel. Those without the deletion will have only one band, at the normal position.
   f) You would probably get hybridization to extra bands, or even just a big smear, since the probe would hybridize non-specifically to other bands in the genome.

2)
   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. Normal cells 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.

3)
   a. 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.
   b. More? Do the test.

CHAPTER 35 – ANSWERS

1) An individual homozygous for this region would have the same DNA sequence on both homologs and thus the same restriction site locations on both maternal and paternal chromosomes. Each restriction enzyme digestion should produce the same set of fragments from both homologs. Different enzymes should give different sized fragments. The probe would be expected to hybridize with the complementary sequences in these fragments. While the three lanes (E, H, and B) could have only a single band hybridizing in each,

there is the possibility that the 1.0 kbp fragment could contain one (or more) of these restriction enzyme sites and thus the probe would hybridize to two (or more) fragments per lane. Thus at least one fragment is expected, but more per lane could occur.

2) If the individual was heterozygous, with some sequence variation between the two homologous chromosomes, then one might expect more than one band per lane due to restriction site polymorphism leading to restriction fragment length polymorphism.

3) If 100 individuals were examined in this way, polymorphisms in restriction fragment length would likely be identified, although not for sure. Some regions of DNA are more variable in the population than others.

4) If the probe fragment had repeated DNA sequences (e.g. Alu repeats), then very many fragments would hybridize and the signal would not resolve individual band, but be a smear of signal paralleling the distribution of digested DNA fragments in each lane.

CHAPTER 36 - ANSWERS:

1) 345 bp
2) Here are the results. S = size standard, R = results
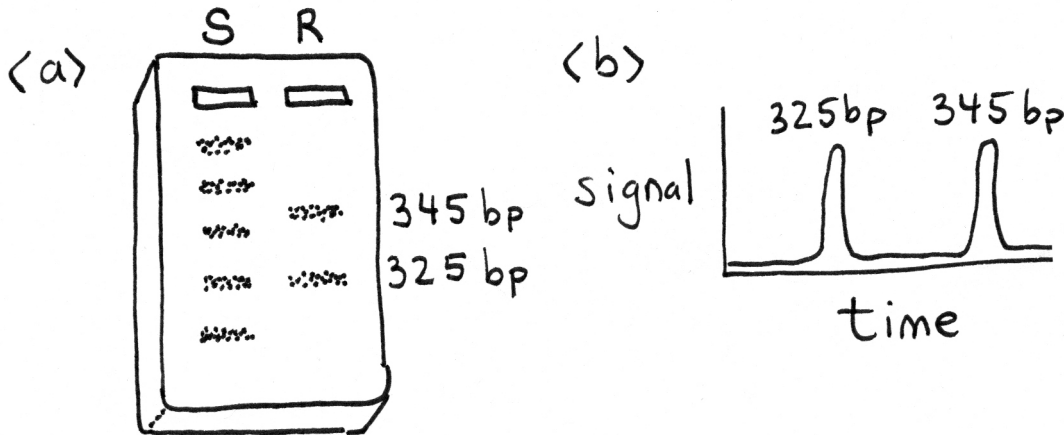
a)

3) No. This policy is not cost-effective, and would violate various constitutional rights. Thus, it is unlikely to happen. Most democratic countries only store DNA profiles for people accused of a crime (US) or convicted of a crime (Canada).

4) For CSF1P0 the 7 allele is maternal and the 12 allele is paternal; for D8S1179 we can't tell which allele is which; for D21S11 the 9 allele is paternal and the 10 allele is maternal.

STR	Child	Mother	Potential fathers		
			#1	#2	#3
CSF1PO	7/12	7/10	7/10	12/14	12/13
D8S1179	6/6	6/8	6/9	12/12	5/6
D21S11	9/10	10/11	5/5	9/16	9/9

5) No. He has all the paternal alleles present in the child. He would have to be excluded based on other evidence. This might come from other STRs being tested, and/or he wasn't "involved" with the mother.

6) a)- We know they are stable within an individual because we can test DNA samples from various tissue types and times during their lifetime and they all give the same DNA profile for all the individual's samples. They are invariant for one person.

b)- No, because there would not be a standard for a single person. This might show up as differences in DNA profiles if DNA samples are obtained from different tissues or at different times over their lifetime.

There are, however, situations where a person's DNA does change. The length of telomeres in cells is known to change during the lifetime of a person. The age of a person can be estimated from their length.

7)

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

8)

#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 $B_2B_3E_2E_2$

9) #3 and #6 cannot be a parent, since neither #3 and #6 have any alleles in common with #1 at locus E.

10)   **a)** the region of the fragment that is most likely to be polymorphic

```
i)   TAAAGGAATCAATTACTTCTGTGTGTGTGTGTGTGTGTGTGTGTTCTTAGTTGTTTAAGTTTTAAGTTGTGA
ii)  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
iii)    ATTTCCTTAGTTAATGAAGACACACACACACACACACACACACAAGAATCAACAAATTCAAAATTCAACACT
```

b) any simple sequence repeats

```
i)   TAAAGGAATCAATTACTTCTGTGTGTGTGTGTGTGTGTGTGTGTTCTTAGTTGTTTAAGTTTTAAGTTGTGA
ii)  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
iii)    ATTTCCTTAGTTAATGAAGACACACACACACACACACACACACAAGAATCAACAAATTCAAAATTCAACACT
```

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.
```
i)   TAAAGGAATCAATTACTTCTGTGTGTGTGTGTGTGTGTGTGTGTTCTTAGTTGTTTAAGTTTTAAGTTGTGA
ii)  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
iii)    ATTTCCTTAGTTAATGAAGACACACACACACACACACACACACAAGAATCAACAAATTCAAAATTCAACACT
```

CHAPTER 37: ANSWERS:

1) Most SNPs are the result of a single base pair substitution mutation that happened once during human evolution. The chance is very small that a different mutation could happen at the same site in a different person and also become prevalent.
2) If the conditions are not stringent enough the labelled DNA will attach to both oligos no matter what a person's genotype is. It would appear that the person is heterozygous for every single SNP, a very unlikely situation.
3) Yes. We would need to obtain DNA samples from many people with blue eyes and many people with brown or other eye colours. In fact, most people with blue eyes have this phenotype because they have mutations in two genes on chromosome 15 called *OCA2* and *HERC2*.
4) Yes. In fact, 23andme looks for a SNP near *HERC2* for this purpose. In Europeans most people with the GG genotype of the SNP have blue eyes while people with the AG and AA genotypes have brown eyes.

5) For this disease or phenotype there are two genes responsible, one is on chromosome 1 and the other is on chromosome 3.

CHAPTER 38 – ANSWERS

1)
   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

2) 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.

3) 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 × 1200 = 2(0.6)(0.4) × 1200=576) and 192 *aa* individuals (q^2 × 1200 = 0.4^2 × 1200 = 192), if the population was at Hardy-Weinberg equilibrium with 432 *AA* individuals.

4) The actual population appears to have more heterozygotes and fewer 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 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 Q3 and Q4. In question 3, 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 4, 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 39 – ANSWERS

1) These fish would all be heterozygotes and thus have spines like the deep-water population. The presence of the enhancers element (deep water) would be dominant to the absence of the element (shallow water).
2)

a) Yellow wings, body, and mouth parts, but normal bristles and claws.
b) All yellow, no normal colour.
c) Same as "a".
d) The enhancer elements on the stop codon allele might cross regulate the transcription unit on the deletion allele. That is, wild type enhancers on one allele drive a wild type transcript on the other allele. (Note: this is the case. See Morris et al. 1999 Genetics **151:** 633–651.)

CHAPTER 40 – ANSWERS

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

CHAPTER 41 – ANSWERS

1) Oncogenes usually arise from gain-of-function mutations, which tend to be haplosufficient. Mutations in tumour suppressors are usually loss-of-function mutations, which tend to be haploinsufficient.
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.
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.
4) Was the dose fed to the rats relevant? Were similar effects seen in other organisms? Do epidemiological studies support these conclusions? Could the results be replicated by a different research group? What was the proposed mechanism for this increased incidence?
5) Cancer results from an accumulation of mutations that activate cell division and disable tumour suppression. HPV infection alone does 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.
6) Same for BRCA1 mutations.

END OF ANSWERS

Notes:

Other Courses in Genetics

BIOL 207
Molecular Genetics
and Heredity

GENET 270
Foundations of
Molecular Genetics

GENET 301
Organization of
Simple Genomes

GENET 302
Organization of
Complex Genomes

GENET 304
Gene Expression and
Its Regulation

GENET 305
Genetic Analysis

GENET 364
Plant Gentics

GENET 390
Gene Manipulation

GENET 412
Genetic Control of
Animal Development

GENET 418
Human Genetics

GENET 422
Current Topics in
Developmental Genetics

GENET 424
Ethical Issues
in Genetics

GENET 375
Introduction to Molecular
Genetics Techniques

GENET 420 *6
Research Techniques
in Molecular Genetics