

Laboratory Manual For SCI200 Genetics at Roxbury Community College

Nikolaus Sucher

Contents

Welcome	11
Acknowledgements	13
1 Course Description	15
1.1 Instructional Objectives:	15
1.2 Student Learning Outcomes:	15
2 How to do well in this class	17
Before the lab	17
During the lab	17
After the lab	17
3 Lab Safety	19
3.1 General rules	19
3.2 Class dissections	20
3.3 Clothing	20
3.4 Handling chemicals	20
3.5 Glassware	21
3.6 Heating substances	21
3.7 Handling microbiology materials	21
Classical Genetics	23
3.8 Review Questions	23
4 Mitosis and Meiosis	25
4.1 View Prepared Slides	26
4.2 Preparing an Onion root tip squash	26
4.3 Review Questions	28
5 Chromosomes And Karyotypes	29
5.1 Chromosomes	29
5.2 Karyotypes	31
5.3 Observations on karyotypes	33
5.4 Human karyotype	34
5.5 Chromosome abnormalities	34
5.6 Chromosome banding	36
5.7 Determine the karyotype	39
5.8 Review Questions	39
6 Mendelian Genetics: Mendel in His Own Words	41
6.1 Experimental Procedures	41

7 Mendelian Genetics: A Practical Example	43
7.1 Monohybrid Cross (Experiment 1)	43
7.2 Punnett Square	44
7.3 Dihybrid Cross (Experiment 2)	46
7.4 Pearson's Chi-Squared (χ^2) And Goodness Of Fit Test	48
7.5 A Look At Mendel's Monohybrid Cross Data	50
7.6 Review Questions	50
Molecular Genetics	53
8 Molecular Laboratory Skills	55
8.1 Exercise 1: using graduated (serological) macropipettes	56
8.2 Exercise 2: using air displacement micropipettes	57
8.3 Exercise 3: using a microcentrifuge	60
8.4 Review Questions	61
9 DNA Extraction and Polymerase Chain Reaction	63
9.1 DNA	63
9.2 Polymerase chain reaction	64
9.3 DNA Extraction	65
9.4 Experimental Procedures	65
9.5 GAPDH Polymerase Chain Reaction	66
9.6 Experimental Procedures	66
9.7 Review Questions	67
10 Exonuclease I Digestion And Nested PCR	69
10.1 Exonuclease I Digestion	69
10.2 Experimental Procedures	69
10.3 Nested polymerase chain reaction	69
10.4 Experimental Procedures	70
10.5 Review Questions	70
11 Agarose Gel Electrophoresis, Ligation and Transformation	71
11.1 Agarose gel electrophoresis	71
11.2 Experimental Procedures	71
11.3 PCR Product Purification	72
11.4 Experimental Procedures	72
11.5 Ligation	73
11.6 Experimental Procedures	73
11.7 Transformation	73
11.8 Experimental Procedures	74
11.9 Review Questions	75
12 Plasmid Purification And Restriction Digest	77
12.1 Plasmid Purification	77
12.2 Plasmid	77
12.3 Experimental Procedures	78
12.4 Restriction Digest	79
12.5 Experimental Procedures	79
13 DNA sequencing	81
13.1 Experimental Procedures	82
13.2 Review Questions	82
14 DNA sequence analysis	83

CONTENTS	5
-----------------	----------

14.1 Experimental procedures	85
14.2 BLAST search	86
14.3 Experimental procedures	86
14.4 Review Questions	87
15 The Human Genome	93
15.1 The Human Genome Project	93
15.2 Public-Private Competition Around The Human Genome Project	95
15.3 Genome Annotation	96
15.4 International HapMap Project	97
15.5 The Single Nucleotide Polymorphism Database (dbSNP) of Nucleotide Sequence Variation .	98
15.6 The 1000 Genomes Project	98
15.7 Personal genomics	100
15.8 Viewing The Human Genome	100
15.9 Review Questions	102

List of Tables

7.1	Monohybrid Cross.	46
7.2	Mendel's Laws Of Inheritance.	47
7.3	Dihybrid Cross.	47
7.4	Mendel's monohybrid cross data for Pearson's chi-squared test.	52
7.5	Pearson's chi-squared test results for Mendel's monohybrid cross data.	52
8.1	Standard micropipette volumes and colors.	55
8.2	Experimental determination of graduated pipette volumes.	57
8.3	Experimental determination of micropipette volumes.	59
9.1	Plant material record.	65
9.2	PCR protocol for GAPDH amplification.	67
10.1	PCR protocol for second round (nested) GAPDH amplification.	70
12.1	Composition of restriction digest master mix.	79
13.1	DNA Sequencing Reactions	82
15.1	Human Genome Assembly GRCh38.p12 ¹ (Release date: 2017-12-21). Length of DNA sequence of each human chromosome. Total lengths are calculated by summing the length of the sequenced bases and estimated gaps. Ungapped lengths are calculated by summing the length of sequenced bases only. 'Ns' are excluded.	94
15.2	Human Genome Assembly GRCh38.p12 (Release date: 2017-12-21). GC content (%) and number of proteins, ribosomal RNA (rRNA), transfer RNA (tRNA), other RNA, genes and pseudogenes on each chromosome. Retrieved on August 10, 2018 at https://www.ncbi.nlm.nih.gov/genome/?term=txid9606%5Borgn%5D	96
15.3	Gene counts based on the Ensembl human genome annotations ² of the primary assembly of GRCh38.p12 (Genome Reference Consortium Human Build 38).	97
15.4	Populations represented in the 1000 genomes collection.	99
15.5	The types and numbers of sites with variations in the autosomes and X chromosome in the genomes of 2504 people examined in the 1000 genomes project (phase 3 v5a).	99

¹<https://www.ncbi.nlm.nih.gov/grc/human/data>

²http://useast.ensembl.org/Homo_sapiens/Info/Annotation

List of Figures

4.1	Onion root tip	26
4.2	Fish blastodisc	27
4.3	Several different phases of mitosis are visible in this onion root tip spread.	27
5.1	(ref:chromosome)	30
5.2	(ref:karyo)	32
5.3	(ref:Turner)	35
5.4	(ref:Klinefelter)	36
5.5	(ref:Down)	37
5.6	(refs:sky)	39
7.1	Punnett square for homozygous cross.	44
7.2	Punnett square for heterozygous cross.	45
7.3	Monohybrid cross	45
7.4	Dihybrid cross	46
7.5	χ^2 probability density function for one degree of freedom.	49
7.6	χ^2 probability density function for three degrees of freedom.	50
7.7	Results of Mendel's monohybrid cross experiments with peas.	51
8.1	(ref:bulb)	57
8.2	Schematic of an air displacement pipette. The “Digital volume indicator” is a dial display that indicates the digits (i.e. unrelated to electronic). The components vary between brands and different volume sizes have different components, for example the piston in a P2 is needle-like and can be separated with ease from the piston assembly whereas in a P10 it is drum-like and over 1 cm in diameter and is enclosed in plastic. By Squidonius [CC BY-SA 3.0] (https://creativecommons.org/licenses/by-sa/3.0/), from Wikimedia Commons.	58
8.3	Graphical representation of the results obtained by the students performing Experiments 1 and 2 in the Genetics course in the spring term 2019.	59
14.1	The Sanger (chain-termination) method for DNA sequencing. (1) A primer is annealed to a sequence, (2) Reagents are added to the primer and template, including: DNA polymerase, dNTPs, and a small amount of all four dideoxynucleotides (ddNTPs) labeled with fluorophores. During primer elongation, the random insertion of a ddNTP instead of a dNTP terminates synthesis of the chain because DNA polymerase cannot react with the missing hydroxyl. This produces all possible lengths of chains. (3) The products are separated on a single lane capillary gel, where the resulting bands are read by a imaging system. (4) This produces several hundred thousand nucleotides a day, data which require storage and subsequent computational analysis. By Estevezj [CC BY-SA 3.0](https://creativecommons.org/licenses/by-sa/3.0/), from [Wikimedia Commons.](https://commons.wikimedia.org/wiki/File:Sanger-sequencing.svg)	84
14.2	DNA sequence chromatogram and base calls viewed with SnapGene software.	85
14.3	The web page of the U.S. National Library of Medicine.	86

14.4 The Basic Local Alignment Search Tool (BLAST) start page.	87
14.5 Nucleotide BLAST (BLASTN) sequence entry form.	88
14.6 After you have pasted your sequence into the sequence entry field, click the BLAST button at the lower left of the page.	88
14.7 The BLASTN query status updated page.	89
14.8 The BLASTN results page.	89
14.9 Graphic summary of the BLASTN search results.	90
14.10 List of sequences producing significant alignments with your query sequence.	90
14.11 Alignment of the best retrieved sequence with the query sequence.	91
15.1 Principle component analysis (PCA) of genetic variation in the 1000 genomes project reveals population stratification. PCA is based on 473,964 autosomal variants with a minor allele frequency greater than 10% in 2504 people from 5 superpopulations.	101
15.2 (ref:genomebrowser)	102

Welcome

This is the **Laboratory Manual** for the Genetics course (SCI200) at RCC.



This work is licensed under the Creative Commons Attribution-Share Alike 3.0 Unported³ United States License.

³<https://creativecommons.org/licenses/by-sa/3.0/deed.en>

Acknowledgements

The creation of this manual was greatly facilitated and owes a major debt to Wikipedia⁴ and its large number of voluntary contributors. I very liberally copied from many Wikipedia pages and then remixed, edited, adapted and added text. With your continued support and help this manual can only get better over time. I urge you to email me with your criticisms and suggestions at nsucher@rcc.mass.edu⁵. This manual is made available as an open educational resource under Creative Commons Attribution-Share Alike 3.0 Unported⁶ United States License for others to do as I did and improve and adapt to specific requirements.

⁴<https://www.wikipedia.org>

⁵<mailto:nsucher@rcc.mass.edu>

⁶<https://creativecommons.org/licenses/by-sa/3.0/deed.en>

Chapter 1

Course Description

1.1 Instructional Objectives:

At the end of the course, students will be able to

- contrast the role of mitosis and meiosis in cellular reproduction in eukaryotes.
- describe how the concept of the “gene” has originated and evolved over time.
- identify eukaryotic chromosomes and describe their structure and function.
- explain how DNA stores information, read the genetic code, and describe how it is translated and transmitted to the offspring.
- define genetic mutations and describe how they happen and how they can be prevented and give examples of how they influence the function of cells and can lead to disease.
- outline how we can isolate, sequence, manipulate and move DNA between cells and organisms.
- define the terms genome and genomics, evaluate genomics-related information and discuss the far-reaching consequences for our society of the use of genomics.
- outline molecular mechanisms of gene expression.
- give examples of complex traits and discuss current models of their genetic basis.
- outline the role of genes in the development of organisms and give some specific examples.
- discuss how variation in the types and frequencies of “genes” in organisms and populations forms the substrate for natural selection and evolution.

1.2 Student Learning Outcomes:

At successful completion of the course, students will be able to

- describe the molecular and cellular mechanisms explaining why children resemble their parents to a certain degree yet are also distinctly different both from them and (if applicable) their own siblings.
- give examples and evaluate the strength of the relationship between specific genes and health disorders.
- demonstrate “information literacy” and critically analyze reports describing genetics-related research.
- discuss the science underlying purported or real medical and scientific breakthroughs or doomsday scenarios related to the manipulation of cells and organisms at the level of DNA.
- articulate and appraise the personal and societal implications of genetic testing and genetic engineering.
- embark on the next level of training or start work in the life sciences or related fields.

Chapter 2

How to do well in this class

Before the lab

- Know what's coming up. Each week, look at the schedule to know what lab is coming up.
- Be prepared. Read the chapter in the manual **before** you come to the lab.
- In your own words, explain to yourself or someone else what the upcoming lab is about.

During the lab

- Read the lab instructions carefully.
- Follow the instructions carefully.
- Make sure that you know what you are doing and why you are doing it.
- Don't be afraid to ask your instructor for help or clarifications.

After the lab

- Ask yourself: what was this lab about? What did we do? How did we do it?
- In your own words, tell yourself or someone else what you learned in the lab.

Chapter 3

Lab Safety

The laboratory classes are hands-on. Some classes require the use of hazardous chemicals and materials. Safety in the classroom is the #1 priority for students and faculty. To ensure a safe science laboratory, a list of rules must be followed at all times.

Please watch the RCC safety video¹



Prior to your participation in the science lab course, you must read this safety document and sign and return the acknowledgment and agreement page.

3.1 General rules

1. NO FOOD, BEVERAGES, GUM, in the labs. Cell phone usage is also prohibited in the lab.
2. Conduct yourself in a responsible manner at all times in the lab. Horseplay, pranks and practical jokes are prohibited and will not be tolerated. If you participate in inappropriate behavior the INSTRUCTOR HAS THE RIGHT TO ASK YOU TO LEAVE THE lab.
3. Students cannot be in the lab without an instructor present.
4. Read all lab procedures, precautions, and equipment instructions thoroughly before each lab. Follow all written and verbal instructions carefully. Perform only those experiments authorized by the instructor. If during the lab you don't understand, stop and ask the instructor before proceeding. Never do anything in the lab that is outside of your instructors directions or that is not in your lab procedure.
5. Do not begin lab activities; touch any chemicals or equipment until you are instructed to do so.
6. Work areas should be kept organized and clean at all times. Only necessary items (lab notebook, worksheets, etc.) should be on your workbench. Backpacks and purses must be stored under the benches or against the walls. CLEAN ALL OF YOUR WORK SURFACES AND EQUIPMENT AT THE END OF THE EXPERIMENT. Safely dispose of waste in its proper container and place glassware in the grey bins by the sink. DO NOT STACK GLASSWARE. If the bin is full ask the instructor for another bin.
7. Keep aisles clear. Push lab stools under the lab benches when not in use.
8. Know where the safety equipment is and how to use it. This includes the first aid kit, eyewash station, safety shower, fire extinguisher, and fire blanket. Know the location of the fire alarm, and emergency phone. In the event of a fire drill during lab time containers must be closed, gas valves off, fume hood, and all electrical equipment must be turned off.

¹<https://youtu.be/NxcsyTv7stQ>

9. NEVER DISPOSE OF ANYTHING IN THE SINK. All materials are to be disposed of in the proper hazardous waste containers with the assistance of the instructor. All waste containers must be closed and placed inside a secondary containment bin.
10. As classes in these labs use toxic chemicals, keep your hands away from your face, eyes and mouth while working in the lab. Always wash your hands thoroughly with warm water and soap before leaving the lab to prevent injury or illness. This is part of proper lab procedure.
11. Students are not permitted in the prep room areas (between the lab rooms).
12. Handle all living organisms used for lab experiments in a respectful, humane manner.
13. Microscopes must be properly cleaned, the electrical cords properly wrapped, and returned to their places with their protective covers.

Disposal of all hazardous waste is ONLY to be handled by the instructor and in a manner consistent with federal, state and local hazardous waste disposal regulations. Organic solvents are never to be disposed of down the sink; receptacles will be provided as needed for their collection. All hazardous chemical substances must be placed in the appropriate type of container and labeled with chemical, name and date, sealed and placed upright in a gray plastic bin.

3.2 Class dissections

14. Preserved biological specimens should be treated with respect and disposed of in a clear plastic bag and placed inside the hazardous waste drum located in the classroom. This container must be sealed at the end of each dissection.
15. When using sharp objects always carry the tips pointed down and away from you. When dissecting, cut away from your body. Grasp the instrument only by the handle. Never try to catch falling sharp instruments or glassware. When you are finished dissecting, wash and dry your instruments and dissecting pan before storing them in the proper location. Do not leave any instruments in the sink.

3.3 Clothing

16. Students must wear lab goggles when using chemicals, or using heat. NO EXCEPTIONS! Lab coats are mandatory for Anatomy and Physiology, Biology, Microbiology, Chemistry, and Biotechnology labs, except for when lecturing and working on dry lab (for example, looking at models or prepared slides) activities.
17. Gloves should be worn when handling solutions, solids, specimens, etc.
18. Proper dress should always be observed in the lab. Long hair must be tied back. Loose or baggy clothing (especially sleeves), dangling jewelry, hats, shorts, short skirts, bare mid riffs, high heels, sleeveless shirts, and open toed or open heeled shoes and sandals are prohibited in the lab. Failure to comply may result in expulsion from class.

3.4 Handling chemicals

18. Always work in a well-ventilated area. Use the fume hood when working with volatile substances or poisonous vapors, or any chemical with an odor.
19. Never smell a chemical by sniffing. Use your hand to wave the chemical towards your nose.
20. DO NOT TASTE, TOUCH or smell anything unless instructed to do so. You should wear a lab apron and gloves at all times. Your instructor will tell you the proper gloves to wear depending on the chemical being used.
21. CHECK EACH LABEL TWICE before removing any of its contents. Take only what is needed of each chemical. NEVER return unused chemicals to their original container; put it in the waste container.
22. Do not use your fingers to transfer solid chemicals. Use a scoop or spatula.

23. Use a rubber bulb, pipette, or pi-pump when transferring liquid chemicals. NEVER USE YOUR MOUTH TO PIPETTE!
24. When transferring reagents hold the containers away from your body while working on the bench.
25. Acids must be handled with extreme care. You will be shown the proper method for diluting strong acids. ALWAYS ADD ACID TO WATER, swirl or invert the solution and be careful of the heat produced particularly with sulfuric acid.
26. Handle hazardous liquids over a pan to contain spills.
27. Never handle flammable liquids anywhere near an open flame or heat source.
28. Be careful when transporting chemicals across the lab. Hold securely and walk carefully.
29. NEVER POUR CHEMICALS INTO SINK. Waste should be disposed of in the proper hazardous waste container provided.

3.5 Glassware

30. Never handle broken glass with bare hands. Use a brush and a dustpan to clean up broken glass. Place uncontaminated broken glass in the white and blue broken glass receptacles. Contaminated trash goes in the biohazard bin.
31. Fill wash bottles only with distilled water and use only as intended, e.g.; rinsing glassware, adding water to a container.
32. Never use chipped, cracked or dirty glassware to avoid shattering.
33. Never immerse hot glassware in cold water. It may shatter.
34. Never place dirty glassware with the clean glassware. All dirty glassware should be placed in gray wash bins. DO NOT STACK DIRTY GLASSWARE IN BINS.

3.6 Heating substances

35. Exercise extreme caution when using a gas burner. Be careful to keep hair, loose clothing and hands away from flames at all times. Wear safety goggles. Do not put any substances into the flame unless specifically instructed to do so. Never reach over an exposed flame. The instructor will provide a demonstration of the proper way to operate a Bunsen burner. Never leave a lighted burner or hot plate unattended. Always turn the burner or hot plate off when not in use.
36. Do not point the open end of a test tube being heated at yourself or anyone else. Never look into a container that's being heated.
37. Heated metals and glass remain hot for a very long time. They should be set aside to cool and picked up with caution. Use tongs or heat protective gloves.

3.7 Handling microbiology materials

38. Please be aware that micro labs include work with pathogenic organisms. Be alert. Conduct yourself in a responsible manner at all times.
39. If you spill anything notify your instructor immediately. There are special procedures to be followed for spills containing microorganisms.
40. A lab coat must be worn during lab activities. Lab coats/aprons may never leave the lab. If you must leave the lab during a class then your lab coat must be removed.
41. Gloves must be worn at all times when working with bacteria. Gloves need to be disposed of in the biohazard waste container.
42. All contaminated waste must be disposed of in the biohazard container. Do not overfill biohazard containers.

43. You must spray down your lab bench with Lysol after each lab. Do not wipe with paper towels. The bench surface must remain wet for at least five minutes for the Lysol to destroy any micro organisms.
44. Wash your hands thoroughly before and after each lab as well as before you leave the lab for any reason.
45. Dispose of contaminated broken glass in the biohazard bin. Please wrap the broken glass in paper towels before disposal so that the broken glass doesn't cut through the bag. Dispose of uncontaminated glass in the white and blue cardboard glass boxes.

Classical Genetics

Genetics² is, generally, the study of genes, genetic variation, and heredity. In the first four lab sessions of this course, we will study classical genetics³. This is the oldest form of genetics that began with Gregor Mendel's⁴ experiments that formulated and defined a fundamental biological concept now referred to as Mendelian Inheritance. In the second part of this course, we will learn and apply some of the fundamental techniques of molecular genetics⁵.

Classical genetics is the part of genetics that is about the transmission of genetic traits via the acts of reproduction⁶. The process by which characteristics are passed down from parents to their offspring is called heredity. In the sense of classical genetics, variation is known as the lack of resemblance in related individuals and can be categorized as discontinuous or continuous. The phenotype⁷ is a general term that defines an individual's visible, physical traits. The genotype⁸ of an offspring is known as its genetic makeup. Mendel was the first to systematically investigate how certain well-defined traits are transmitted from parents to offspring. Mendel's conclusions were largely ignored by the vast majority of scientists at the time. In 1900, however, his work was "re-discovered" by three European scientists, Hugo de Vries⁹, Carl Correns¹⁰, and Erich von Tschermak¹¹. In 1905, Wilhelm Johannsen¹² introduced the term *gene* and William Bateson¹³ the term *genetics*. Our understanding of what a *gene*¹⁴ is has undergone quite a bit of change. Currently, genes are considered to be pieces of DNA that contain information for synthesis of ribonucleic acids (RNAs) that can be directly functional or serve as the intermediate template for a protein that performs a function.

3.8 Review Questions

1. What is the subject of the field of genetics?
2. What does the term phenotype denote?
3. What does the term genotype denote?
4. What is a gene?

²<https://en.wikipedia.org/wiki/Genetics>

³https://en.wikipedia.org/wiki/Classical_genetics

⁴https://en.wikipedia.org/wiki/Gregor_Mendel

⁵https://en.wikipedia.org/wiki/Molecular_genetics

⁶<https://en.wikipedia.org/wiki/Reproduction>

⁷<https://en.wikipedia.org/wiki/Phenotype>

⁸<https://en.wikipedia.org/wiki/Genotype>

⁹https://en.wikipedia.org/wiki/Hugo_de_Vries

¹⁰https://en.wikipedia.org/wiki/Carl_Correns

¹¹https://en.wikipedia.org/wiki/Erich_von_Tschermak

¹²https://en.wikipedia.org/wiki/Wilhelm_Johannsen

¹³https://en.wikipedia.org/wiki/William_Bateson

¹⁴https://en.wikipedia.org/wiki/Gene#Discovery_of_discrete_inherited_units

Chapter 4

Mitosis and Meiosis

Mitosis¹ is the part of the cell cycle when replicated chromosomes are separated into two new nuclei. In general, mitosis (division of the nucleus) is preceded by the S stage of interphase (during which the DNA is replicated) and is often accompanied or followed by cytokinesis, which divides the cytoplasm, organelles and cell membrane into two new cells containing roughly equal shares of these cellular components. Mitosis and cytokinesis together define the mitotic (M) phase of an animal cell cycle (the division of the mother cell into two daughter cells genetically identical to each other). The process of mitosis is divided into stages corresponding to the completion of one set of activities and the start of the next. These stages are prophase, pro-metaphase, metaphase, anaphase, and telophase.

Meiosis² is a specialized type of cell division that reduces the chromosome number by half, creating four haploid cells, each genetically distinct from the parent cell that gave rise to them. This process occurs in all sexually reproducing single-celled and multicellular eukaryotes, including animals, plants, and fungi. Errors in meiosis resulting in aneuploidy are the leading known cause of miscarriage and the most frequent genetic cause of developmental disabilities. In meiosis, DNA replication is followed by two rounds of cell division to produce four daughter cells, each with half the number of chromosomes as the original parent cell. The two meiotic divisions are known as Meiosis I and Meiosis II. Before meiosis begins, during S phase of the cell cycle, the DNA of each chromosome is replicated so that it consists of two identical sister chromatids, which remain held together through sister chromatid cohesion. This S-phase can be referred to as “premeiotic S-phase” or “meiotic S-phase”. Immediately following DNA replication, meiotic cells enter a prolonged G2-like stage known as meiotic prophase. During this time, homologous chromosomes pair with each other and undergo genetic recombination, a programmed process in which DNA is cut and then repaired, which allows them to exchange some of their genetic information. A subset of recombination events results in crossovers, which create physical links known as chiasmata (singular: chiasma, for the Greek letter Chi (X)) between the homologous chromosomes. In most organisms, these links are essential to direct each pair of homologous chromosomes to segregate away from each other during Meiosis I, resulting in two haploid cells that have half the number of chromosomes as the parent cell. During Meiosis II, the cohesion between sister chromatids is released and they segregate from one another, as during mitosis. In some cases, all four of the meiotic products form gametes such as sperm, spores, or pollen. In female animals, three of the four meiotic products are typically eliminated by extrusion into polar bodies, and only one cell develops to produce an ovum. Because the number of chromosomes is halved during meiosis, gametes can fuse (i.e. fertilization) to form a diploid zygote that contains two copies of each chromosome, one from each parent. Thus, alternating cycles of meiosis and fertilization enable sexual reproduction, with successive generations maintaining the same number of chromosomes. For example, diploid human cells contain 23 pairs of chromosomes including 1 pair of sex chromosomes (46 total), half of maternal origin and half of paternal origin. Meiosis produces haploid gametes (ova or sperm) that contain one set of 23 chromosomes. When two gametes (an egg and a sperm) fuse, the resulting zygote is once again diploid, with the mother and father each contributing 23

¹<https://en.wikipedia.org/wiki/Mitosis>

²<https://en.wikipedia.org/wiki/Meiosis>

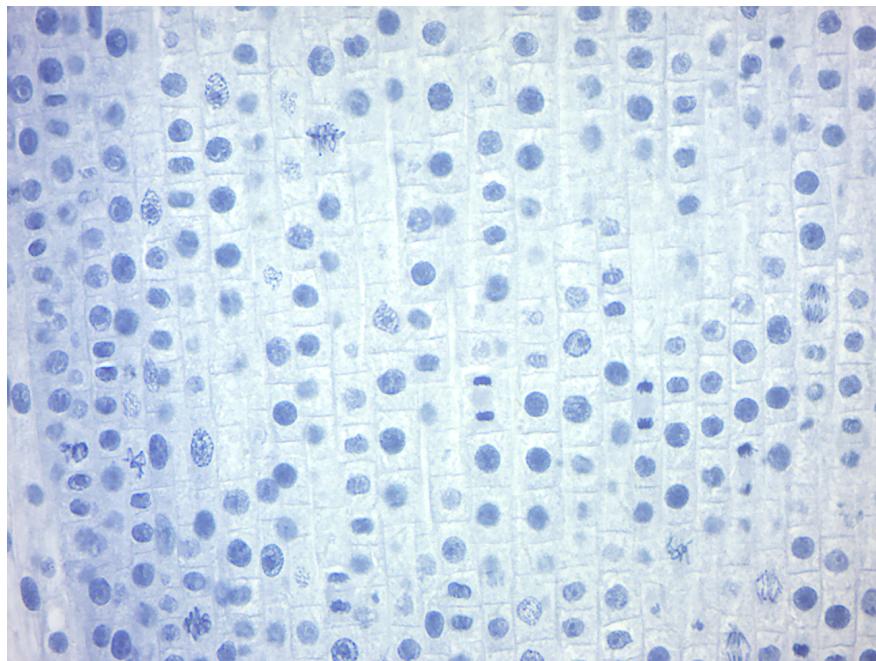


Figure 4.1: Onion root tip

chromosomes. This same pattern, but not the same number of chromosomes, occurs in all organisms that utilize meiosis.

4.1 View Prepared Slides

1. View the onion root tip and observe the different stages of mitosis (Figure 4.1).
2. View the fish blastodisc and observe the different stages of mitosis (Figure 4.2).

4.2 Preparing an Onion root tip squash

4.2.1 Experimental procedures

1. Obtain an onion bulb that shows some roots.
2. Cut off a root tip and place it on a clean slide.
3. Cut off 1mm to 2mm of the root tip and throw away the upper portion of the root.
4. Cover the root tip with four drops of 1 N HCl and warm the slide over an alcohol burner flame for 1 minute. Do not boil.
5. Blot off the excess HCl and cover the root tip with 0.5% aqueous toluidine blue.
6. Again, pass the slide through the alcohol burner flame for 1 minute without boiling.
7. Blot off the excess stain, add a drop of fresh stain, and apply a coverslip.
8. Cover the slide with a paper towel and carefully squash the coverslip firmly with your thumb.
9. Examine the slide for the stages of mitosis as well as interphase and cytokinesis.

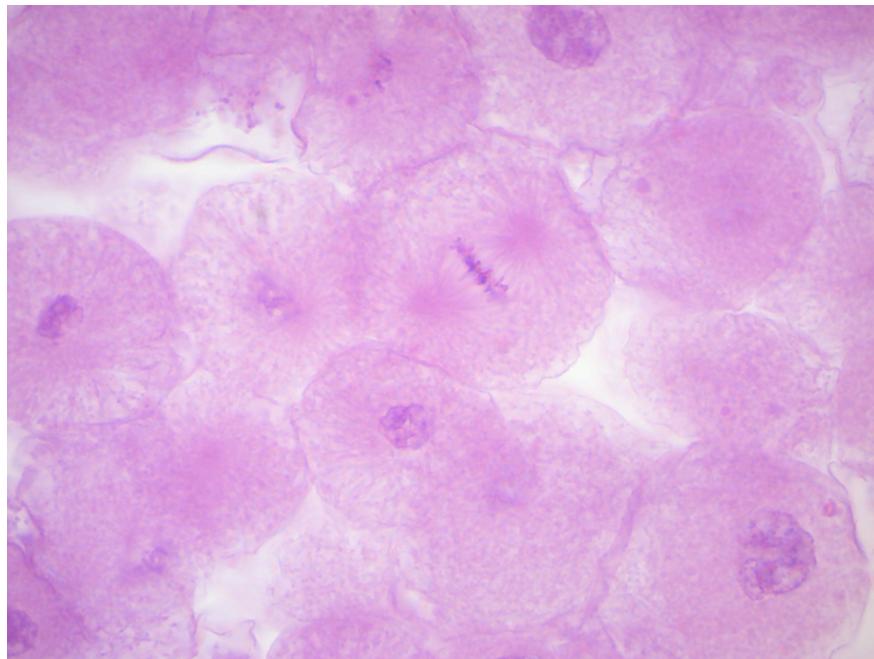


Figure 4.2: Fish blastodisc

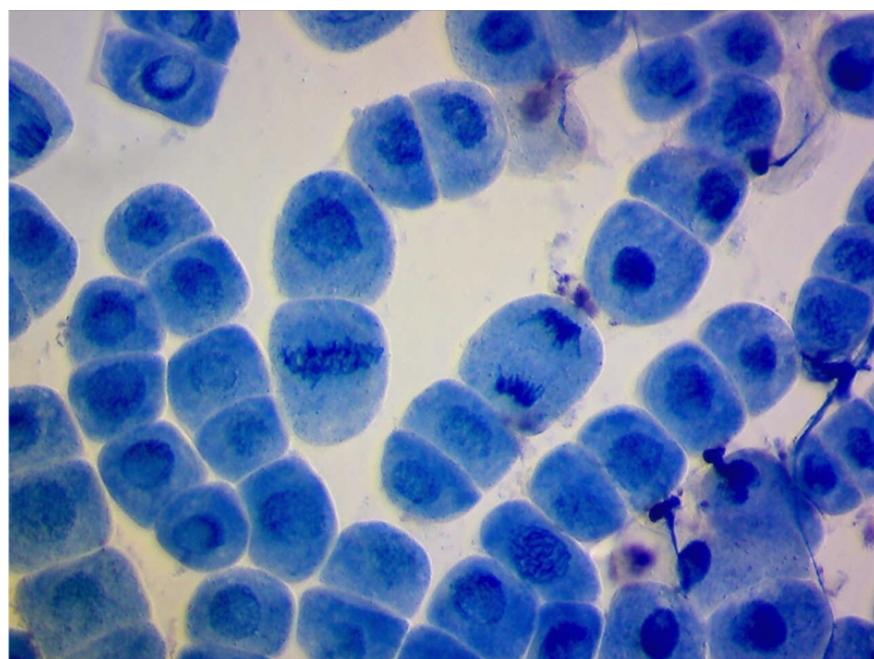


Figure 4.3: Several different phases of mitosis are visible in this onion root tip spread.

4.3 Review Questions

1. What is mitosis and what is its outcome?
2. What is meiosis and what is its outcome?
3. What is homologous recombination and what is its outcome?
4. What do the terms haploid and diploid mean?
5. Are you a haploid or a diploid organism?
6. What are gametes?
7. What is a zygote?

Chapter 5

Chromosomes And Karyotypes

5.1 Chromosomes

A chromosome¹ is a DNA molecule with part or all of the genetic material (genome) of an organism. Most eukaryotic chromosomes include packaging proteins which, aided by chaperone proteins, bind to and condense the DNA molecule to prevent it from becoming an unmanageable tangle. The word chromosome comes from the Greek *χρώμα*, “colour” and *σῶμα*, “body”), describing their strong staining by particular dyes. The term was coined by Heinrich Wilhelm Gottfried von Waldeyer-Hartz², referring to the term chromatin, which was introduced by Walther Flemming³.

Chromosomes are normally visible under a light microscope only when the cell is undergoing the metaphase of cell division (where all chromosomes are aligned in the center of the cell in their condensed form). Before this happens, every chromosome is copied once (S phase), and the copy is joined to the original by a centromere, resulting either in an X-shaped structure (Figure 5.2) if the centromere is located in the middle of the chromosome or a two-arm structure if the centromere is located near one of the ends. The original chromosome and the copy are now called sister chromatids. During metaphase the X-shape structure is called a metaphase chromosome. In this highly condensed form chromosomes are easiest to distinguish and study. In animal cells, chromosomes reach their highest compaction level in anaphase during segregation.

(ref:chromosome) Diagram of a replicated and condensed metaphase eukaryotic chromosome. (1) Chromatid – one of the two identical parts of the chromosome after S phase. (2) Centromere – the point where the two chromatids touch. (3) Short (p) arm. (4) Long (q) arm.⁴ By derivative work: Tryphon (talk)*Chromosome-upright.png: Original version: Magnus Manske, this version with upright chromosome: User:Dietzel65 [GFDL (<http://www.gnu.org/copyleft/fdl.html>) or CC-BY-SA-3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)], via Wikimedia Commons

Chromosomal recombination during meiosis and subsequent sexual reproduction play a significant role in genetic diversity. If these structures are manipulated incorrectly, through processes known as chromosomal instability and translocation, the cell may undergo mitotic catastrophe and die. Mutations in the cell can allow it to inappropriately evade apoptosis and lead to the progression of cancer.

Some use the term chromosome in a wider sense, to refer to the individualized portions of chromatin in cells, either visible or not under light microscopy. Others use the concept in a narrower sense, to refer to the individualized portions of chromatin during cell division, visible under light microscopy due to high condensation.

¹<https://en.wikipedia.org/wiki/Chromosome>

²https://en.wikipedia.org/wiki/Heinrich_Wilhelm_Gottfried_von_Waldeyer-Hartz

³https://en.wikipedia.org/wiki/Walther_Flemming

⁴<https://commons.wikimedia.org/wiki/File:Chromosome.svg>

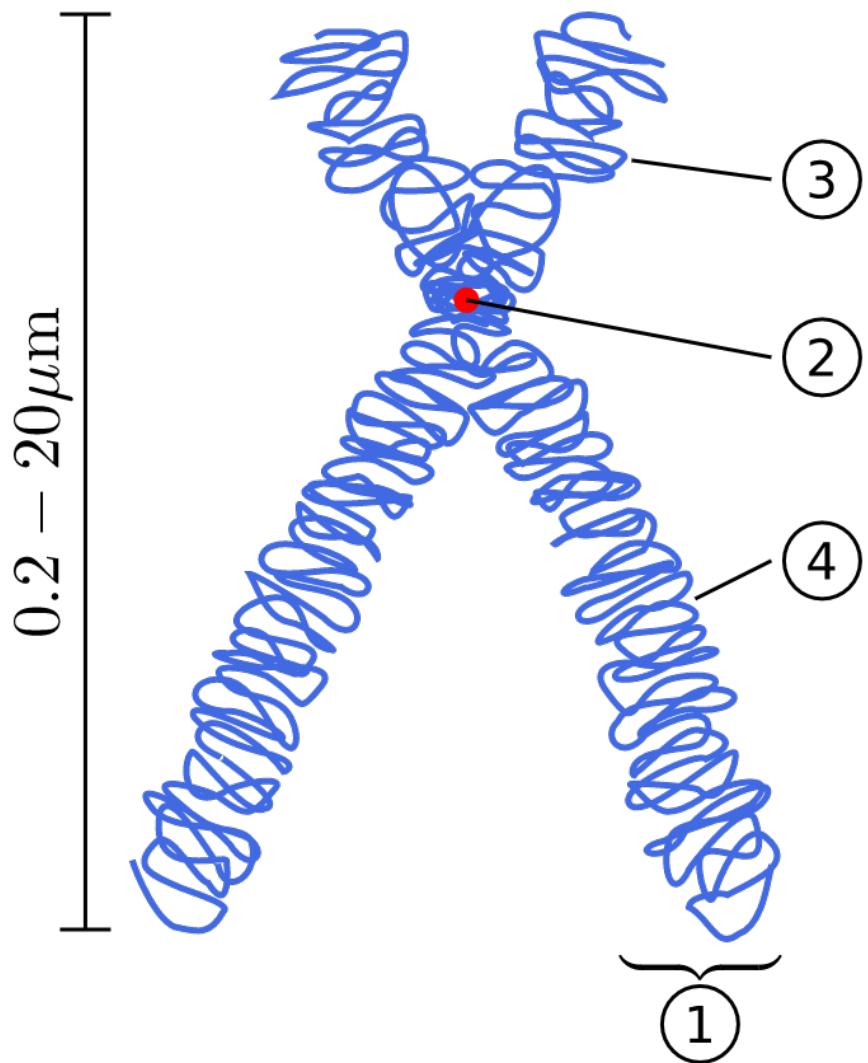


Figure 5.1: (ref:chromosome)

The prokaryotes – bacteria and archaea – typically have a single circular chromosome, but many variations exist. The chromosomes of most bacteria can range in size from only 130,000 base pairs in the endosymbiotic bacteria *Candidatus Hodgkinia cicadicola* to more than 14,000,000 base pairs in the soil-dwelling bacterium *Sorangium cellulosum*. Spirochaetes of the genus *Borrelia* are a notable exception to this arrangement, with bacteria such as *Borrelia burgdorferi*, the cause of Lyme disease, containing a single linear chromosome.

Chromosomes in eukaryotes are composed of chromatin fiber. Chromatin fiber is made of nucleosomes (histone octamers with part of a DNA strand attached to and wrapped around it). Chromatin fibers are packaged by proteins into a condensed structure called chromatin. Chromatin contains the vast majority of DNA and a small amount inherited maternally, can be found in the mitochondria. Chromatin is present in most cells, with a few exceptions, for example, red blood cells.

Chromatin allows the very long DNA molecules to fit into the cell nucleus. During cell division chromatin condenses further to form microscopically visible chromosomes. The structure of chromosomes varies through the cell cycle. During cellular division chromosomes are replicated, divided, and passed successfully to their daughter cells so as to ensure the genetic diversity and survival of their progeny. Chromosomes may exist as either duplicated or unduplicated. Unduplicated chromosomes are single double helixes, whereas duplicated chromosomes contain two identical copies (called chromatids or sister chromatids) joined by a centromere. Eukaryotes possess multiple large linear chromosomes contained in the cells' nucleus. Each chromosome has one centromere, with one or two arms projecting from the centromere, although, under most circumstances, these arms are not visible as such. In addition, most eukaryotes have a small circular mitochondrial genome, and some eukaryotes may have additional small circular or linear cytoplasmic chromosomes. Eukaryotes (cells with nuclei such as those found in plants, fungi, and animals) possess multiple large linear chromosomes contained in the cell's nucleus. Each chromosome has one centromere, with one or two arms projecting from the centromere, although, under most circumstances, these arms are not visible as such. In addition, most eukaryotes have a small circular mitochondrial genome, and some eukaryotes may have additional small circular or linear cytoplasmic chromosomes.

In the nuclear chromosomes of eukaryotes, the uncondensed DNA exists in a semi-ordered structure, where it is wrapped around histones (structural proteins), forming a composite material called chromatin.

Chromosomes in humans are divided into two types: autosomes (body chromosome(s)) and allosome (sex chromosome(s)). Certain genetic traits are linked to a person's sex and are passed on through the sex chromosomes. The autosomes contain the rest of the genetic hereditary information. All act in the same way during cell division. Human cells have 23 pairs of chromosomes (22 pairs of autosomes and one pair of sex chromosomes), giving a total of 46 per cell. In addition to these, human cells have many hundreds of copies of the mitochondrial genome.

5.2 Karyotypes

A karyotype⁵ is the number and appearance of chromosomes in the nucleus of a eukaryotic cell. The term is also used for the complete set of chromosomes in a species or in an individual organism and for a test that detects this complement or measures the number. The name was coined by the German anatomist, Heinrich von Waldeyer in 1888. It is New Latin from Ancient Greek κάρυον, "kernel", "seed", or "nucleus", and τύπος, "general form").

Karyotypes describe the chromosome count of an organism and what these chromosomes look like under a light microscope. Attention is paid to their length, the position of the centromeres, banding pattern, any differences between the sex chromosomes, and any other physical characteristics. The preparation and study of karyotypes is part of cytogenetics.

The study of whole sets of chromosomes is sometimes known as karyology. The chromosomes are depicted (by rearranging a photomicrograph) in a standard format known as a karyogram or idiogram: in pairs, ordered by size and position of centromere for chromosomes of the same size as shown in Figure 5.2.

⁵<https://en.wikipedia.org/wiki/Karyotype>



Figure 5.2: (ref:karyo)

(ref:karyo) Karyogram of human male using Giemsa staining⁶

The basic number of chromosomes in the somatic cells of an individual or a species is called the somatic number and is designated $2n$. In the germ-line (the sex cells) the chromosome number is n (humans: $n = 23$). Thus, in humans $2n = 46$.

So, in normal diploid organisms, autosomal chromosomes are present in two copies. There may, or may not, be sex chromosomes. Polyploid cells have multiple copies of chromosomes and haploid cells have single copies.

The study of karyotypes is important for cell biology and genetics, and the results may be used in evolutionary biology (karyosystematics) and medicine. Karyotypes can be used for many purposes; such as to study chromosomal aberrations, cellular function, taxonomic relationships, and to gather information about past evolutionary events.

The German scientists Matthias Jakob Schleiden⁷, Rudolf Virchow⁸ and Otto Bütschli⁹ were among the first scientists who recognized the structures now familiar as chromosomes. Chromosomes were first observed in plant cells by Carl Wilhelm von Nägeli¹⁰ in 1842. Their behavior in animal (salamander) cells was described by Walther Flemming, the discoverer of mitosis, in 1882.

The next stage took place after the development of genetics in the early 20th century, when it was appreciated that chromosomes (that can be observed by karyotype) were the carrier of genes.

In his famous textbook *The Cell in Development and Heredity*, Wilson linked together the independent work of Theodor Boveri¹¹ and Walter Sutton¹² (both around 1902) by naming the chromosome theory of

⁶https://commons.wikimedia.org/wiki/File:NHGRI_human_male_karyotype.png

⁷https://en.wikipedia.org/wiki/Matthias_Jakob_Schleiden

⁸https://en.wikipedia.org/wiki/Rudolf_Virchow

⁹https://en.wikipedia.org/wiki/Otto_Bütschli

¹⁰https://en.wikipedia.org/wiki/Carl_Nägeli

¹¹https://en.wikipedia.org/wiki/Theodor_Boveri

¹²https://en.wikipedia.org/wiki/Walter_Sutton

inheritance the Boveri–Sutton chromosome theory¹³ (the names are sometimes reversed).

Investigation into the human karyotype took many years to settle the most basic question: how many chromosomes does a normal diploid human cell contain? In 1912, Hans von Winiwarter reported 47 chromosomes in spermatogonia and 48 in oogonia, concluding an XX/XO sex determination mechanism. The number of human chromosomes was published in 1923 by Theophilus Painter. By inspection through the microscope, he counted 24 pairs, which would mean 48 chromosomes. His error was copied by others and it was not until 1956 that the true number, 46, was determined by Indonesia-born cytogeneticist Joe Hin Tjio¹⁴. He was responsible for finding a new approach that settled the issue:

- Using cells in tissue culture
- Pretreating cells in a hypotonic solution, which swells them and spreads the chromosomes
- Arresting mitosis in metaphase by a solution of colchicine
- Squashing the preparation on the slide forcing the chromosomes into a single plane
- Cutting up a photomicrograph and arranging the result into an indisputable karyogram.

The work took place in 1955, and was published in 1956. The karyotype of humans includes only 46 chromosomes. Rather interestingly, the great apes have 48 chromosomes. Human chromosome 2 is now known to be a result of an end-to-end fusion of two ancestral ape chromosomes.

5.3 Observations on karyotypes

5.3.1 Staining

The study of karyotypes is made possible by staining. Usually, a suitable dye, such as Giemsa, is applied after cells have been arrested during cell division by a solution of colchicine usually in metaphase or prometaphase when most condensed. In order for the Giemsa stain to adhere correctly, all chromosomal proteins must be digested and removed. For humans, white blood cells are used most frequently because they are easily induced to divide and grow in tissue culture. Sometimes observations may be made on non-dividing (interphase) cells. The sex of an unborn fetus can be determined by observation of interphase cells (see amniotic centesis and Barr body).

5.3.2 Observations

Six different characteristics of karyotypes are usually observed and compared:

1. Differences in absolute sizes of chromosomes. Chromosomes can vary in absolute size by as much as twenty-fold between genera of the same family. For example, the legumes *Lotus tenuis* and *Vicia faba* each have six pairs of chromosomes, yet *V. faba* chromosomes are many times larger. These differences probably reflect different amounts of DNA duplication.
2. Differences in the position of centromeres. These differences probably came about through translocations.
3. Differences in relative size of chromosomes. These differences probably arose from segmental interchange of unequal lengths.
4. Differences in basic number of chromosomes. These differences could have resulted from successive unequal translocations which removed all the essential genetic material from a chromosome, permitting its loss without penalty to the organism (the dislocation hypothesis) or through fusion. Humans have one pair fewer chromosomes than the great apes. Human chromosome 2 appears to have resulted from the fusion of two ancestral chromosomes, and many of the genes of those two original chromosomes have been translocated to other chromosomes.

¹³https://en.wikipedia.org/wiki/Boveri–Sutton_chromosome_theory

¹⁴https://en.wikipedia.org/wiki/Joe_Hin_Tjio

5. Differences in number and position of satellites. Satellites are small bodies attached to a chromosome by a thin thread.
6. Differences in degree and distribution of heterochromatic regions. Heterochromatin stains darker than euchromatin. Heterochromatin is packed tighter. Heterochromatin consists mainly of genetically inactive and repetitive DNA sequences as well as containing a larger amount of Adenine-Thymine pairs. Euchromatin is usually under active transcription and stains much lighter as it has less affinity for the giemsa stain. Euchromatin regions contain larger amounts of Guanine-Cytosine pairs. The staining technique using giemsa staining is called G banding and therefore produces the typical “G-Bands”.

A full account of a karyotype may therefore include the number, type, shape and banding of the chromosomes, as well as other cytogenetic information.

Variation is often found:

- between the sexes,
- between the germ-line and soma (between gametes and the rest of the body),
- between members of a population (chromosome polymorphism), in geographic specialization, and in mosaics or otherwise abnormal individuals. Human karyotype.

5.4 Human karyotype

The normal human karyotypes contain 22 pairs of autosomal chromosomes and one pair of sex chromosomes (allosomes; Figure 5.2). Normal karyotypes for females contain two X chromosomes and are denoted 46,XX; males have both an X and a Y chromosome denoted 46,XY. Any variation from the standard karyotype may lead to developmental abnormalities.

5.4.1 Ploidy

Ploidy¹⁵ is the number of complete sets of chromosomes in a cell.

5.4.2 Aneuploidy

Aneuploidy¹⁶ is the condition in which the chromosome number in the cells is not the typical number for the species. This would give rise to a chromosome abnormality such as an extra chromosome or one or more chromosomes lost. Abnormalities in chromosome number usually cause a defect in development. Down syndrome¹⁷ (Figure 5.5) and Turner syndrome¹⁸ are examples of this.

5.5 Chromosome abnormalities

Chromosome abnormalities can be numerical, as in the presence of extra or missing chromosomes, or structural, as in derivative chromosome, translocations, inversions, large-scale deletions or duplications. Numerical abnormalities, also known as aneuploidy, often occur as a result of nondisjunction during meiosis in the formation of a gamete; trisomies, in which three copies of a chromosome are present instead of the usual two, are common numerical abnormalities. Structural abnormalities often arise from errors in homologous recombination. Both types of abnormalities can occur in gametes and therefore will be present in all cells of an affected person's body, or they can occur during mitosis and give rise to a genetic mosaic individual who has some normal and some abnormal cells.

¹⁵<https://en.wikipedia.org/wiki/Ploidy>

¹⁶<https://en.wikipedia.org/wiki/Aneuploidy>

¹⁷https://en.wikipedia.org/wiki/Down_syndrome

¹⁸https://en.wikipedia.org/wiki/Turner_syndrome

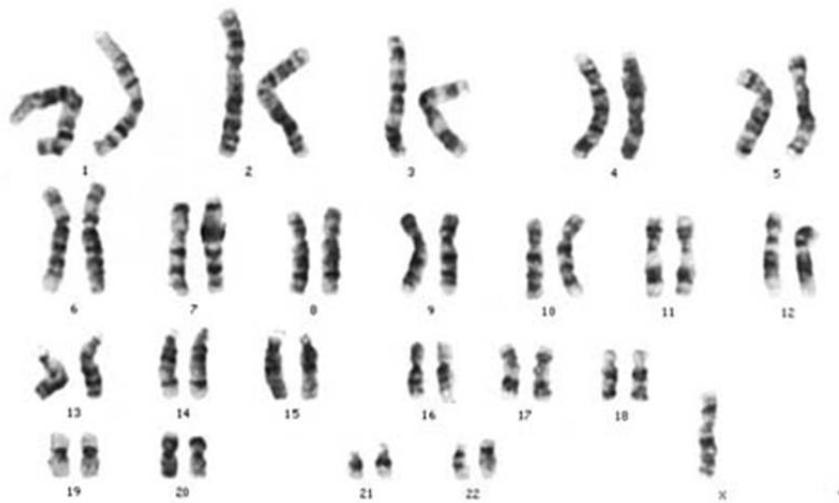


Figure 5.3: (ref:Turner)

5.5.1 Chromosomal abnormalities that lead to disease in humans include

- Turner syndrome (Figure 5.3) results from a single X chromosome (45,X or 45,X0).
- Klinefelter syndrome¹⁹ (Figure 5.4), the most common male chromosomal disease, otherwise known as 47,XXY, is caused by an extra X chromosome.
- Edwards syndrome²⁰ is caused by trisomy (three copies) of chromosome 18.
- Down syndrome (Figure 5.5), a common chromosomal disease, is caused by trisomy of chromosome 21.
- Patau syndrome²¹ is caused by trisomy of chromosome 13.
- Trisomy 9²², believed to be the 4th most common trisomy, has many long lived affected individuals but only in a form other than a full trisomy, such as trisomy 9p syndrome or mosaic trisomy 9. They often function quite well, but tend to have trouble with speech.
- Also documented are trisomy 8 and trisomy 16, although they generally do not survive to birth.

(ref:Turner) 45,X karyotype, showing an unpaired X at the lower right²³

(ref:Klinefelter) Klinefelter syndrome karyotype (XXY)²⁴

(ref:Down) Karyotype for trisomy Down syndrome: notice the three copies of chromosome 21²⁵

Some disorders arise from loss of just a piece of one chromosome, including

- Cri du chat²⁶ (cry of the cat), from a truncated short arm on chromosome 5. The name comes from the babies' distinctive cry, caused by abnormal formation of the larynx.
- 1p36 Deletion syndrome, from the loss of part of the short arm of chromosome 1.
- Angelman syndrome – 50% of cases have a segment of the long arm of chromosome 15 missing; a deletion of the maternal genes, example of imprinting disorder.
- Prader-Willi syndrome – 50% of cases have a segment of the long arm of chromosome 15 missing; a deletion of the paternal genes, example of imprinting disorder.

¹⁹https://en.wikipedia.org/wiki/Klinefelter_syndrome

²⁰https://en.wikipedia.org/wiki/Edwards_syndrome

²¹https://en.wikipedia.org/wiki/Patau_syndrome

²²https://en.wikipedia.org/wiki/Trisomy_9

²³<https://commons.wikimedia.org/wiki/File:45,X.jpg>

²⁴https://commons.wikimedia.org/wiki/File:Human_chromosomesXXY01.png

²⁵https://commons.wikimedia.org/wiki/File:Down_Syndrome_Karyotype.png

²⁶https://en.wikipedia.org/wiki/Cri_du_chat_syndrome

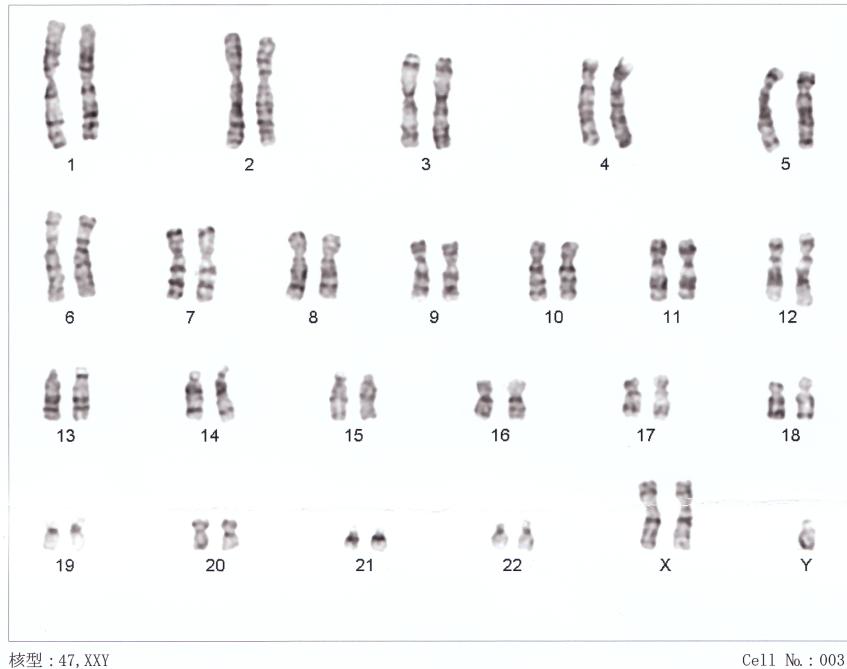


Figure 5.4: (ref.Klinefelter)

- Chromosomal abnormalities can also occur in cancerous cells of an otherwise genetically normal individual; one well-documented example is the Philadelphia chromosome, a translocation mutation commonly associated with chronic myelogenous leukemia and less often with acute lymphoblastic leukemia.

5.6 Chromosome banding

Chromosomes display a banded pattern when treated with some stains. Bands are alternating light and dark stripes that appear along the lengths of chromosomes. Unique banding patterns are used to identify chromosomes and to diagnose chromosomal aberrations, including chromosome breakage, loss, duplication, translocation or inverted segments. A range of different chromosome treatments produce a range of banding patterns: G-bands, R-bands, C-bands, Q-bands, T-bands and NOR-bands.

Cytogenetics employs several techniques to visualize different aspects of chromosomes.

5.6.1 Types of banding

- G-banding is obtained with Giemsa stain following digestion of chromosomes with trypsin. It yields a series of lightly and darkly stained bands — the dark regions tend to be heterochromatic, late-replicating and AT rich. The light regions tend to be euchromatic, early-replicating and GC rich. This method will normally produce 300–400 bands in a normal, human genome.
- R-banding is the reverse of G-banding (the R stands for “reverse”). The dark regions are euchromatic (guanine-cytosine rich regions) and the bright regions are heterochromatic (thymine-adenine rich regions).
- C-banding: Giemsa binds to constitutive heterochromatin, so it stains centromeres. The name is derived from centromeric or constitutive heterochromatin. The preparations undergo alkaline denaturation prior to staining leading to an almost complete depurination of the DNA. After washing the probe the remaining DNA is renatured again and stained with Giemsa solution consisting of methylene azure,

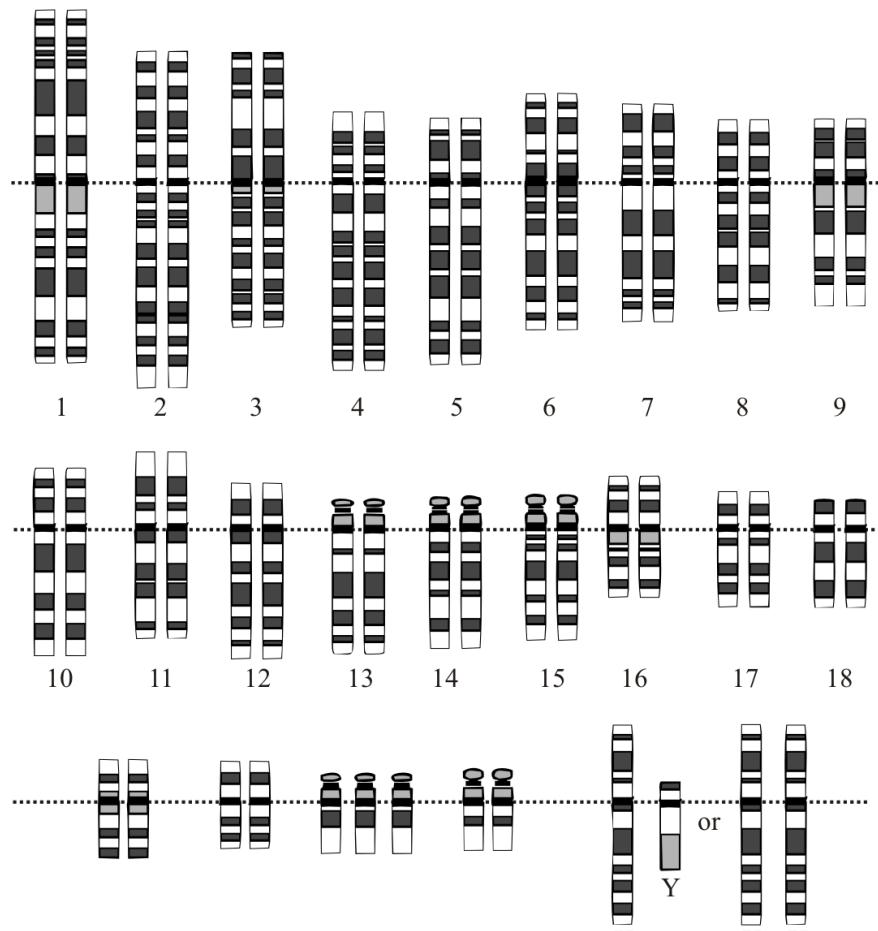


Figure 5.5: (ref:Down)

methylene violet, methylene blue, and eosin. Heterochromatin binds a lot of the dye, while the rest of the chromosomes absorb only little of it. The C-bonding proved to be especially well-suited for the characterization of plant chromosomes.

- Q-banding is a fluorescent pattern obtained using quinacrine for staining. The pattern of bands is very similar to that seen in G-banding. They can be recognized by a yellow fluorescence of differing intensity. Most part of the stained DNA is heterochromatin. Quinacrin (atebrin) binds both regions rich in AT and in GC, but only the AT-quinacrin-complex fluoresces. Since regions rich in AT are more common in heterochromatin than in euchromatin, these regions are labelled preferentially. The different intensities of the single bands mirror the different contents of AT. Other fluorochromes like DAPI or Hoechst 33258 lead also to characteristic, reproducible patterns. Each of them produces its specific pattern. In other words: the properties of the bonds and the specificity of the fluorochromes are not exclusively based on their affinity to regions rich in AT. Rather, the distribution of AT and the association of AT with other molecules like histones, for example, influences the binding properties of the fluorochromes.
- T-banding: visualize telomeres. Silver staining: Silver nitrate stains the nucleolar organization region-associated protein. This yields a dark region where the silver is deposited, denoting the activity of rRNA genes within the NOR.

5.6.2 Classic karyotype cytogenetics

In the “classic” (depicted) karyotype, a dye, often Giemsa (G-banding), less frequently mepacrine (quinacrine), is used to stain bands on the chromosomes (Figure 5.2). Giemsa is specific for the phosphate groups of DNA. Quinacrine binds to the adenine-thymine-rich regions. Each chromosome has a characteristic banding pattern that helps to identify them; both chromosomes in a pair will have the same banding pattern.

Karyotypes are arranged with the short arm of the chromosome on top, and the long arm on the bottom. Some karyotypes call the short and long arms p and q, respectively. In addition, the differently stained regions and sub-regions are given numerical designations from proximal to distal on the chromosome arms. For example, Cri du chat syndrome involves a deletion on the short arm of chromosome 5. It is written as 46,XX,5p-. The critical region for this syndrome is deletion of p15.2 (the locus on the chromosome), which is written as 46,XX,del(5)(p15.2).

5.6.3 Multicolor FISH (mFISH) and spectral karyotype (SKY technique)

Multicolor FISH and the older spectral karyotyping are molecular cytogenetic techniques used to simultaneously visualize all the pairs of chromosomes in an organism in different colors (Figure 5.6). Fluorescently labeled probes for each chromosome are made by labeling chromosome-specific DNA with different fluorophores. Because there are a limited number of spectrally distinct fluorophores, a combinatorial labeling method is used to generate many different colors. Fluorophore combinations are captured and analyzed by a fluorescence microscope using up to 7 narrow-banded fluorescence filters or, in the case of spectral karyotyping, by using an interferometer attached to a fluorescence microscope. In the case of an mFISH image, every combination of fluorochromes from the resulting original images is replaced by a pseudo color in a dedicated image analysis software. Thus, chromosomes or chromosome sections can be visualized and identified, allowing for the analysis of chromosomal rearrangements. In the case of spectral karyotyping, image processing software assigns a pseudo color to each spectrally different combination, allowing the visualization of the individually colored chromosomes. Multicolor FISH is used to identify structural chromosome aberrations in cancer cells and other disease conditions when Giemsa banding or other techniques are not accurate enough.

(ref:sky) Spectral human karyotype²⁷

²⁷<https://en.wikipedia.org/wiki/File:Spectralkaryotype98-300.jpg>



Figure 5.6: (ref:sky)

5.7 Determine the karyotype

5.7.1 Experimental procedures

1. Obtain a photomicrograph of a metaphase chromosome spread.
2. Use a blank sheet of paper as your data page on which you will assemble the karyogram.
3. Write the code associated with the photomicrograph on the top left corner of the data page.
4. Count the chromosomes on the photomicrograph. As you count, use a pencil to write the number next to each counted chromosome. Write the number of chromosomes on the top right corner of the data page. A number other than 46 is an indication of an aneuploidy.
5. Using scissors carefully cut out (one by one) the chromosomes of the photomicrograph. One by one, lay out the cut-out chromosomes on the data page.
6. Rearrange the cut out chromosomes sorted by size and matching banding patterns (as shown in Figure 5.2).
7. Make sure that the number of chromosomes on the data page is equal to the number of chromosomes that you wrote on the top right corner of the data page.
8. When you are confident that you have correctly identified each chromosome, use a glue stick to glue them onto the data page.
9. Determine the type of sex chromosomes on the data page.
10. Write down the karyotype.
11. Write down the suggested diagnosis.

5.8 Review Questions

1. What are chromosomes?
2. How many chromosomes are found in somatic cells in humans?
3. How many chromosomes are found in human gametes?
4. What is a karyotype?
5. What are examples of abnormal karyotypes in humans?
6. What is Down syndrome?
7. What is Turner syndrome?
8. What is Klinefelter syndrome?

Chapter 6

Mendelian Genetics: Mendel in His Own Words

The principles of Mendelian inheritance were named for and first derived by Gregor Johann Mendel¹, a nineteenth-century monk who formulated his ideas after conducting hybridization experiments with pea plants² (*Pisum sativum*) he had planted in the garden of his monastery in Moravia³. Between 1856 and 1863, Mendel cultivated and tested some 5,000 pea plants. From these experiments, he induced two generalizations which later became known as Mendel's Principles of Heredity or Mendelian inheritance (Table 7.2). He described these principles in a two-part paper, *Versuche über Pflanzen-Hybriden* (Experiments on Plant Hybridization), that he read to the Natural History Society of Brno⁴ on 8 February and 8 March 1865, and which was published in 1866. Mendel's conclusions were largely ignored by the vast majority of scientists at the time. In 1900, however, his work was “re-discovered” by three European scientists, Hugo de Vries⁵, Carl Correns⁶, and Erich von Tschermark⁷.

In this laboratory session, we will read an English translation⁸ of Mendel's famous paper⁹. If you prefer to work with a printed copy, you can download a PDF version of the English translation¹⁰ and print it before coming to the lab.

6.1 Experimental Procedures

1. Right-click on the link to the English translation¹¹ to open a new browser tab showing an HTML version of Mendel's paper.
2. Read the first section entitled “INTRODUCTORY REMARKS”.
3. Notice that some terms are highlighted indicating that they are clickable. Clicking on a link will bring you to a glossary that gives some explanation of the term. You can go back to the main text by clicking on the back button of your browsers.
4. After you have read the “INTRODUCTORY REMARKS”, answer the following questions:
 - What is the background of Mendel's experiments?

¹https://en.wikipedia.org/wiki/Gregor_Mendel

²<https://en.wikipedia.org/wiki/Pea>

³<https://en.wikipedia.org/wiki/Moravia>

⁴<https://en.wikipedia.org/wiki/Brno>

⁵https://en.wikipedia.org/wiki/Hugo_de_Vries

⁶https://en.wikipedia.org/wiki/Carl_Correns

⁷https://en.wikipedia.org/wiki/Erich_von_Tschermark

⁸<http://www.mendelweb.org/Mendel.html>

⁹<http://www.mendelweb.org/MWGerText.html>

¹⁰papers/bateson-mendel-3-peas.pdf

¹¹<http://www.mendelweb.org/Mendel.html>

- What set Mendel out to do?
 - How long did it take Mendel to conduct the experiments ?
5. Now, read the second section of the paper entitled “SELECTION OF THE EXPERIMENTAL PLANTS”.
6. Next, answer the following questions:
- According to Mendel, what determines whether or not any experiment has any value?
 - What criteria did Mendel use to select the plants for his experiments?
 - Which plants did Mendel choose for his experiments?
 - How many varieties did Mendel obtain originally and how many of those did he select and cultivate for the entire length of his experiments?
 - What is the definition of “species” that Mendel quotes? What does Mendel think of the applicability of this definition in relation to his experiments?
7. Continue and read the section entitled “DIVISION AND ARRANGEMENT OF THE EXPERIMENTS”.
8. Answer the following questions:
- According to Mendel, what was the object(ive) of the experiment?
 - How many differentiating characters did Mendel investigate?
 - How many plants were in total used in these experiments?
 - How many fertilizations were performed in all plants?
 - What is the meaning of “reciprocal crossings”?
9. Read the next section entitled “THE FORMS OF THE HYBRIDS”.
10. Answer the following questions:
- What did Mendel observe when he compared the characters of the hybrids to those of their parents? . Which characters did he call dominant? Which did he term recessive?
 - What “interesting” fact about the transmission of the dominant characters from parent to hybrid does Mendel mention in a separate paragraph?
 - List the pairs of dominant and recessive characters observed in these experiments.
11. Continue reading the next section entitled “THE FIRST GENERATION FROM THE HYBRIDS”.
12. Answer the following questions:
- What does Mendel observe in the first generation obtained by crossing the hybrids?
 - What did Mendel **not** observe? Why do you think does he mention and even emphasize a negative observation?
 - How did Mendel quantitatively analyze the data? Which numbers are observed and which number is computed from the observed numbers? According to Mendel, what factors complicated the quantitative analysis and how did Mendel deal with them?
 - Mendel explains that the observation of a dominant character in the first generation from the hybrids can have a “double signification”? What does he mean and how can the two “significations” be distinguished?
13. Read the next section entitled “THE SECOND GENERATION FROM THE HYBRIDS”.
14. Answer the following questions:
- What is the difference between offspring derived from first generation hybrid parents with dominant characters and offspring derived from first generation parents with recessive characters?
 - What are the quantitative relationships observed?
15. Skip one section and continue reading *the first three paragraphs* of the section entitled “THE OFFSPRING OF HYBRIDS IN WHICH SEVERAL DIFFERENTIATING CHARACTERS ARE ASSOCIATED”.
16. Answer the following questions:
- What was the task Mendel set himself in the experiments described in this section?
 - What experiments did Mendel conduct?
 - What was the observed outcome of these experiments?

Chapter 7

Mendelian Genetics: A Practical Example

In this laboratory session, we will use maize *Zea mays* subsp. *mays*, from Spanish: maíz after Taíno mahiz), also known as corn to study Mendelian inheritance¹. This cereal grain was first domesticated by indigenous peoples in southern Mexico about 10,000 years ago. The leafy stalk of the plant produces separate pollen and ovuliferous inflorescences or ears, which are fruits, yielding kernels or seeds. Maize has become a staple food in many parts of the world, with total production surpassing that of wheat or rice. However, not all of this maize is consumed directly by humans. Some of the maize production is used for corn ethanol, animal feed and other maize products, such as corn starch and corn syrup. The six major types of corn are dent corn, flint corn, pod corn, popcorn, flour corn, and sweet corn.

7.1 Monohybrid Cross (Experiment 1)

A monohybrid cross² is a mating between two individuals with different variations at one genetic trait of interest. For example, the gene for seed color in corn exists in two forms, one for purple and the other for yellow. These alternative “forms” are called alleles. For each biological trait, an organism inherits two alleles, one from each parent. These alleles may be the same or different. An organism that has two identical alleles for a gene is said to be homozygous for that gene (and is called a homozygote). An organism that has two different alleles for a gene is said be heterozygous for that gene (and is called a heterozygote).

The genotype of an individual is made up of the many alleles it possesses. An individual’s physical appearance, or phenotype, is determined by its alleles as well as by its environment. The presence of an allele does not mean that the trait will be expressed in the individual that possesses it. If the two alleles of an inherited pair differ (the heterozygous condition), then one determines the organism’s appearance and is called the dominant allele; the other has no noticeable effect on the organism’s appearance and is called the recessive allele. Thus, in the example above the dominant purple kernel allele (designated R) will hide the phenotypic effects of the recessive yellow kernel allele (designated r). This is known as the Law of Dominance but it is not a transmission law: it concerns the expression of the genotype. Upper case letters are used to represent dominant alleles whereas lowercase letters are used to represent recessive alleles.

The character(s) being studied in a monohybrid cross are governed by two or multiple alleles of a single gene. A cross between two parents possessing a pair of different alleles is known as monohybrid cross. To carry out such a cross, each parent is chosen to be homozygous or true breeding for a given trait (allele).

¹https://en.wikipedia.org/wiki/Mendelian_inheritance

²https://en.wikipedia.org/wiki/Monohybrid_cross

	Yellow	
Purple	r	r
R	Rr	Rr
R	Rr	Rr

Figure 7.1: Punnett square for homozygous cross.

When a cross satisfies the conditions for a monohybrid cross, it is usually detected by a characteristic distribution of second-generation (F2) offspring that is sometimes called the monohybrid ratio.

Generally, the monohybrid cross is used to determine the dominance relationship between two alleles. The cross begins with the parental (P) generation. One parent is homozygous for one allele, and the other parent is homozygous for the other allele. The offspring make up the first filial (F1) generation. Every member of the F1 generation is heterozygous and the phenotype of the F1 generation expresses the dominant trait. Crossing two members of the F1 generation produces the second filial (F2) generation. Probability theory predicts that three quarters of the F2 generation will have the dominant allele's phenotype. And the remaining quarter of the F2s will have the recessive allele's phenotype. This predicted 3:1 phenotypic ratio assumes Mendelian inheritance.

In the first experiment, we will study the result obtained from a monohybrid cross. A strain of corn producing pure purple kernels (RR) is crossed with a strain producing pure yellow kernels (rr). Purple is dominant with the resulting F1 ears all bearing purple kernels. These plants that are heterozygous for a single trait are called monohybrids. When the F1 is self-pollinated, the resulting F2 ears bear both purple and yellow kernels (Figure 7.3).

7.2 Punnett Square

The Punnett square³ (Figures 7.1 and 7.2) is a visual representation of Mendelian inheritance and used to predict an outcome of a particular cross or breeding experiment. It is named after Reginald C. Punnett⁴, who devised the approach. In our first experiment, both parents are homozygous, one carrying two copies of the dominant allele (R), the other two copies of the recessive (r) allele. Each parent can only make gametes that have either the R (purple) or r (yellow) allele. The Punnett square for the parental cross is shown in Figure 7.1

The squares containing the single letters represent the possible gametes. The squares with two letters represent the zygotes resulting from the combination of the respective gametes. It can be easily seen that all offspring will be heterozygous (Rr) and therefore purple. The Punnett square for the F1 cross is depicted in Figure 7.2

³https://en.wikipedia.org/wiki/Punnett_square

⁴https://en.wikipedia.org/wiki/Reginald_Punnett

		Maternal gametes	
		R	r
Paternal gametes	R	RR	Rr
	r	Rr	rr

Figure 7.2: Punnett square for heterozygous cross.



Figure 7.3: Monohybrid cross



Figure 7.4: Dihybrid cross

7.2.1 Experimental Procedures

1. Count the number of purple and yellow kernels on one row of the F2 ear without removing the kernels.
2. Determine the ratio of purple to yellow.
3. Now tabulate the numbers obtained by each of your class mates in Table 7.1 and add these figures to get a total.
4. Using the total numbers, determine a ratio of purple to yellow.

Table 7.1: Monohybrid Cross.

Row #	purple	yellow
1		
2		
3		
4		
5		
6		
7		
8		
9		
Total		

7.3 Dihybrid Cross (Experiment 2)

In the second experiment, we will study the result obtained from a dihybrid cross⁵. A dihybrid cross is a cross between two different lines (varieties, strains) that differ in two observed traits. In the name “Dihybrid cross”, the “di” indicates that there are two traits involved (in our example designated R and Su), the “hybrid” means that each trait has two different alleles (in our example R and r, or Su and su), and “cross” means that there are two individuals who are combining or “crossing” their genetic information. In our example, a pure strain of corn producing purple-starchy kernels (RR SuSu) is crossed with a pure strain producing yellow-sweet (rr susu). The starchy seeds are smooth, the sweet seeds are wrinkled. The resulting F1 ears all bear purple-starchy (smooth) kernels. Plants that are heterozygous for two traits are called dihybrids. When the F1 is self-pollinated, the resulting F2 generation contains various combinations (Figure 7.4).

The rules of meiosis, as they apply to the dihybrid, are codified in Mendel’s first law and Mendel’s second law, which are also called the Law of Segregation and the Law of Independent Assortment, respectively (Table

⁵https://en.wikipedia.org/wiki/Dihybrid_cross

7.2). For genes on separate chromosomes, each allele pair showed independent segregation. If the first filial generation (F1 generation) produces four identical offspring, the second filial generation, which occurs by crossing the members of the first filial generation, shows a phenotypic (appearance) ratio of **9:3:3:1**, where:

- the **9** represents the proportion of individuals displaying both dominant traits
- the first **3** represents the individuals displaying the first dominant trait and the second recessive trait
- the second **3** represents those displaying the first recessive trait and second dominant trait
- the **1** represents the homozygous, displaying both recessive traits.

Table 7.2: Mendel's Laws Of Inheritance.

Law	Definition
Law of Segregation	During gamete formation, the alleles for each gene segregate from each other so that each gamete carries only one allele for each gene.
Law of Independent Assortment	Genes for different traits independently assort during the formation of gametes.
Law of Dominance	Some alleles are dominant while others are recessive; an organism with at least one dominant allele will display the effect of the dominant allele.

7.3.1 Experimental Procedures

1. Carefully count the number of kernels of each phenotype appearing on a row of F2 ear. Tabulate the results and determine the totals and total ratios in Table 7.3.

Table 7.3: Dihybrid Cross.

Row #	purple and starchy (smooth)	purple and sweet (wrinkled)	yellow and starchy (smooth)	yellow and sweet (wrinkled)	ratio
1					
2					

Row #	purple and starchy (smooth)	purple and sweet (wrinkled)	yellow and starchy (smooth)	yellow and sweet (wrinkled)	ratio
3					
4					
5					
6					
7					
9					
Total					

7.4 Pearson's Chi-Squared (χ^2) And Goodness Of Fit Test

Now that we have obtained our data, we can ask how well the Mendelian model of inheritance actually fits our data. We will use Pearson's χ^2 test as a measure of goodness of fit. The goodness of fit⁶ of a statistical model describes how well it fits a set of observations. Measures of goodness of fit typically summarize the discrepancy between observed values and the values expected under the model in question. Pearson's chi-squared test is a statistical test applied to sets of categorical data (like in our case, where we simply observe if a kernel is purple or yellow) to evaluate how likely it is that any observed difference between the sets arose by chance. It tests a null hypothesis stating that the frequency distribution of certain events observed in a sample is consistent with a particular theoretical distribution. The events considered must be mutually exclusive and have total probability 1. In our case, the “events” correspond to observing either purple or yellow kernels. The theoretical distribution of these events is 3:1 (we expect to observe three purple kernels for every yellow one) as predicted by Mendel's laws of inheritance. Chi-squared (χ^2) is the sum of the squared differences between the observed and expected outcome frequencies (that is, counts of observations) divided by the expectation:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

where:

O_i = number (frequency) of observations of type i

E_i = expected number (frequency) of observations of type i according to the model

In our experiment, i is two, since we have two types of observations (purple and yellow). Let's calculate χ^2 using our data first from the monohybrid and then the dihybrid cross.

The χ^2 statistic can be used to calculate a p-value by comparing the value of the statistic to a chi-squared distribution⁷ with k degrees of freedom. The number of degrees of freedom is equal to the number of categories minus 1 (in our case, we have two categories and therefore $k = 1$). The p-value tells us how likely it would be to make the observations that we made, assuming the null hypothesis is true. The null hypothesis in our case is that the ratio of purple to yellow kernels is in fact 3:1. A low p-value, below the chosen significance level, indicates statistical significance, i.e., sufficient evidence to reject the null hypothesis. A significance level of 0.05 is often used as the cutoff between significant and not-significant results.

7.4.1 Calculating χ^2 Using The Monohybrid Cross Data

1. Add the numbers of observations in each row of the “Purple” column. This number corresponds to O_{purple} .

⁶https://en.wikipedia.org/wiki/Goodness_of_fit

⁷https://en.wikipedia.org/wiki/Chi-squared_distribution

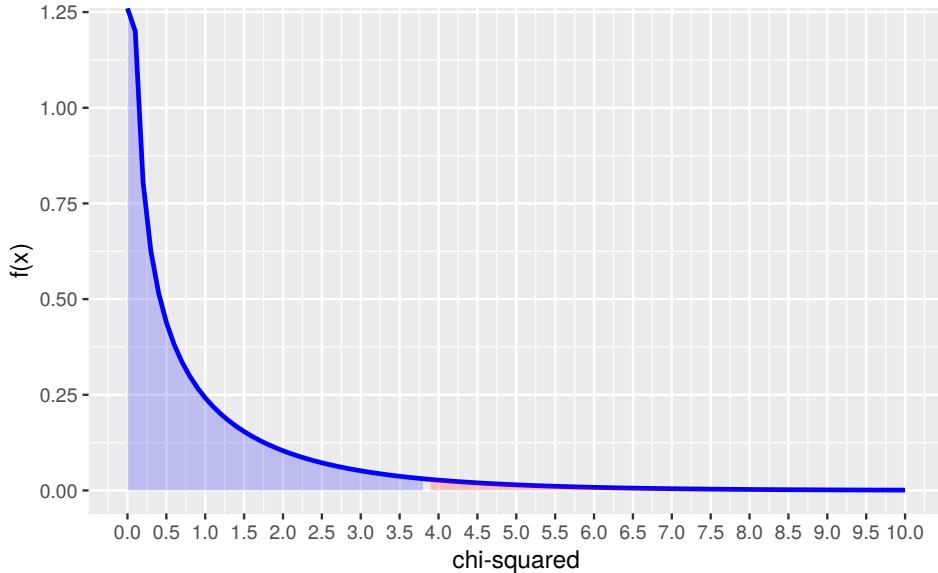


Figure 7.5: χ^2 probability density function for one degree of freedom.

2. Add the numbers of observations in each row of the “Yellow” column. This number corresponds to O_{yellow} .
3. Add both numbers. This is the total number of observations.
4. Multiply the total number of observations by 0.75 (the expected fraction of purple kernels). The result is E_{purple} .
5. Multiply the total number of observations by 0.25 (the expected fraction of yellow kernels). The result is E_{yellow} .
6. Subtract E_{purple} from O_{purple} , square the result and divide by E_{purple} . The result is χ^2_{purple} .
7. Subtract E_{yellow} from O_{yellow} , square the result and divide by E_{yellow} . The result is χ^2_{yellow} .
8. Obviously, our data do not correspond 100% to the prediction. If they did, χ^2 would be 0. The next step, then, is to see how likely it is that the difference between observation and model are simply due to chance, i.e. due to random errors that can never be completely avoided.

7.4.2 Testing Goodness Of Fit

1. State the null hypothesis (H_0). Our H_0 is that kernels with purple color will be 3 times more frequent than kernels with yellow color.
2. Select a desired level of confidence (significance level, p-value or alpha level) for the result of the test. It is common in many scientific investigations to use $p < 0.05$. This essentially means that we will reject the null hypothesis (H_0), i.e. that our observations are consistent with a 3:1 ratio, if the p-value in our experiments is less than 0.05. In other words, a p-value less than 0.05 would indicate that there is less than a 5% (1 in 20) chance that we would have observed the numbers we did, if the Mendelian model of inheritance is in fact applicable in our case.
3. Determine the p-value corresponding to our χ^2 value. The blue curve shown in Figure 7.5 is called the probability density function of χ^2 with one degree of freedom. The light blue shaded area under the curve between $\chi^2 = 0$ and $\chi^2 = 3.84$ corresponds to a probability of 95% of observing a χ^2 value in that range if H_0 is true. χ^2 values greater than 3.84 indicate a less than 5% chance of observing the values we did, if H_0 is true (light red shaded area).
4. Reject H_0 if χ^2 is greater than 3.84.

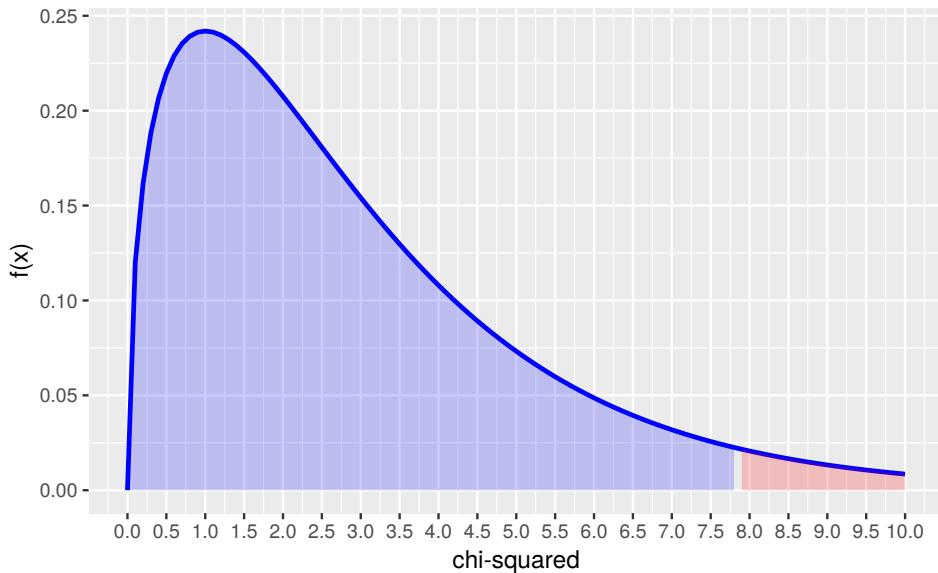


Figure 7.6: χ^2 probability density function for three degrees of freedom.

7.4.3 Calculate χ^2 For Your Dihybrid Cross Data

1. Calculate the χ^2 value for your dihybrid cross data.
2. State the null hypothesis (H_0).
3. Select $p < 0.05$ as your level of confidence, i. e. you will reject the null hypothesis (H_0) if the p-value in your experiment is less than 0.05. The blue curve shown in Figure 7.6 is the probability density function of χ^2 with three degrees of freedom. The light blue shaded area under the curve between $\chi^2 = 0$ and $\chi^2 = 7.81$ corresponds to a probability of 95% of observing a χ^2 value in that range if H_0 is true. χ^2 values greater than 7.81 indicate a less than 5% chance of observing the values we did, if H_0 is true (light red shaded area).
4. Reject H_0 if χ^2 is greater than 7.81.
5. Calculate the goodness of fit of your data to the outcome predicted by Mendel's laws of inheritance applied to a dihybrid cross.
6. What is the p-value corresponding to your χ^2 value?
7. Do you accept or reject your null hypothesis?

7.5 A Look At Mendel's Monohybrid Cross Data

Compare your χ^2 and p-values with Mendel's results (Fig. 7.7) from his monohybrid cross experiments in peas (Chapter 6). Table 7.4 summarizes the experimental results and Table 7.5 lists the calculated χ^2 and p-values.

7.6 Review Questions

2. What is an allele?
3. What are dominant and recessive alleles?
4. What is the genotype of an organism?
5. What is a trait?
6. What is the phenotype of an organism?

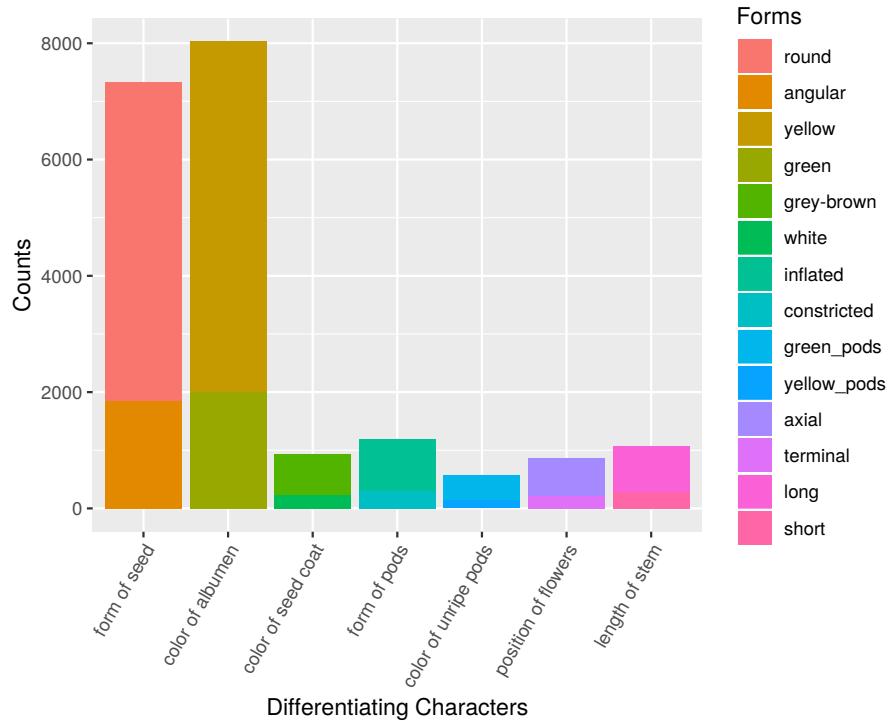


Figure 7.7: Results of Mendel's monohybrid cross experiments with peas.

7. What is the genotype of the F1 generation of the monohybrid cross?
8. What is the phenotype of the F1 generation monohybrid cross?
9. What are the possible maternal and paternal genotypes of the F1 gametes monohybrid cross?
10. What is the genotype of the parents of the dihybrid cross?
11. What are the phenotypes of the parents of the dihybrid cross?
12. What are the possible genotypes of the parent gametes of the dihybrid cross?
13. What is the genotype of the F1 generation of the dihybrid cross?
14. What is the phenotype of the F1 generation dihybrid cross?
15. What are the possible maternal and paternal genotypes of the F1 gametes dihybrid cross?

Table 7.4: Mendel's monohybrid cross data for Pearson's chi-squared test.

experiment	differentiating_character	forms	counts	character
1	form of seed	round	5474	dominant
1	form of seed	angular	1850	recessive
2	color of albumen	yellow	6022	dominant
2	color of albumen	green	2001	recessive
3	color of seed coat	grey-brown	705	dominant
3	color of seed coat	white	224	recessive
4	form of pods	inflated	882	dominant
4	form of pods	constricted	299	recessive
5	color of unripe pods	green_pods	428	dominant
5	color of unripe pods	yellow_pods	152	recessive
6	position of flowers	axial	651	dominant
6	position of flowers	terminal	207	recessive
7	length of stem	long	787	dominant
7	length of stem	short	277	recessive

Table 7.5: Pearson's chi-squared test results for Mendel's monohybrid cross data.

Experiment	chi_squared	p-value
1	0.26	0.61
2	0.01	0.90
3	0.39	0.53
4	0.06	0.80
5	0.45	0.50
6	0.35	0.55
7	0.61	0.44

Molecular Genetics

In the remaining 8 laboratory sessions of this course, we will clone a portion of a glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene from plants, insert this gene fragment into a plasmid vector, and analyze the sequences of resulting clones using bioinformatics.

Specifically, we will:

1. Learn and practice some essential skills for the molecular laboratory (Chapter 8)
2. Identify the plant or plants to be studied, extract the gDNA, and amplify a portion of the GAPC gene using PCR (Chapter 9).
3. Assess the results of PCR using agarose gel electrophoresis, perform exonuclease I digestion to remove the first round PCR primers, perform nested PCR (Chapter 10).
4. Ligate (insert) the GAPC gene fragment into a plasmid vector and transform bacteria with the plasmid (Chapter 11).
5. Isolate the plasmid from the bacteria and confirm the presence of the insert by restriction enzyme digestion (Chapter 12).
6. Prepare plasmid for DNA sequencing and send to a facility for sequencing (Chapter 13).
7. Obtain the sequence of the cloned GAPC gene fragment and analyze the cloned gene using bioinformatics (Chapter 14).
8. Look at the human genome and known variations that distinguish different populations and variations that have been linked to increased risk for certain diseases (Chapter 15).

While we will chose and provide the plants, the reagents used in the experiments will come in the form of ready made kits. In the early days of molecular biology, researchers had to prepare all required materials themselves including the laborious purification of enzymes. Today, virtually any reagent that is required can be purchased “off-the-shelf”, prepackaged and ready to use from a great number of suppliers. One of the oldest companies specialized in supplying materials for life science research laboratories are the BioRad Laboratories Inc.⁸, which developed the curriculum for these experiments and supplies the materials under Bio-Rad ExplorerTM Cloning and Sequencing Explorer Series.

As our hands-on-time in the laboratory is limited, the RCC laboratory technicians will help us by performing some behind the scenes work between the lab sessions.

Glyceraldehyde 3-phosphate dehydrogenase⁹ (abbreviated as GAPDH or less commonly as G3PDH) (EC 1.2.1.12) is an enzyme of ~37kDa that catalyzes the sixth step of glycolysis and thus serves to break down glucose for energy and carbon molecules. In addition to this long established metabolic function, GAPDH has recently been implicated in several non-metabolic processes, including transcription activation, initiation of apoptosis, ER to Golgi vesicle shuttling, and fast axonal, or axoplasmic transport.

Because the GAPDH gene is often stably and constitutively expressed at high levels in most tissues and cells, it is considered a housekeeping gene. For this reason, GAPDH is commonly used by biological researchers as a loading control for western blot and as a control for qPCR. Plants have multiple GAPDH genes; the specific ones that have been selected for cloning are the GAPC genes.

⁸https://en.wikipedia.org/wiki/Bio-Rad_Laboratories

⁹https://en.wikipedia.org/wiki/Glyceraldehyde_3-phosphate_dehydrogenase

Chapter 8

Molecular Laboratory Skills

In this laboratory you will learn how to handle small amounts of liquids using pipettes. A pipette¹ is a laboratory tool commonly used in biology, chemistry, and medicine to draw up, transport and dispense a measured volume of liquid. Pipettes come in several designs for various purposes with differing levels of accuracy and precision, from single piece glass pipettes to more complex adjustable or electronic pipettes. Micropipettes are used to handle small (1 - 1000 μL) or even very small ($< 1 \mu\text{L}$) amounts of liquids. Macropipettes are used to handle larger volumes. Accurate pipetting is crucial to the success of molecular and biochemical experiments. Mistakes during pipetting may cause experiments to fail or to make them irreproducible.

Commonly, we will use two main types of pipettes:

- graduated pipettes
- air displacement micropipettes

Graduated pipettes² or serological pipettes are a type macropipette made of glass or plastic consisting of a long tube with a series of graduations, as on a graduated cylinder or burette, to indicate different calibrated volumes. Today we use manual (e.g. a rubber bulb (Figure 8.1)) or automatic propipettors to generate negative or positive pressure to draw up and dispense the liquids.

(ref:bulb) A rubber bulb propipettor.³

Air displacement micropipettes⁴ are a type of adjustable micropipette that deliver a measured volume of liquid; depending on size, it could be between about 0.1 μL to 1000 μL (1 ml). These pipettes require disposable tips that come in contact with the fluid. Micropipettes (also referred to as pipettors) come in several standard sizes with disposable plastic tips that are labeled in different colors (Table 8.1).

Table 8.1: Standard micropipette volumes and colors.

Name	Min. volume (μL)	Max. volume (μL)	Color	tip size (μL)
P2	0.2	2	white	10
P10	1	10	white	10
P20	2	20	yellow	200
P100	20	100	yellow	200
P200	50	200	yellow	200
P1000	200	1000	blue	1000
P5000	1000	5000	green	5000

¹<https://en.wikipedia.org/wiki/Pipette>

²https://en.wikipedia.org/wiki/Graduated_pipette

³<https://commons.wikimedia.org/wiki/File:Pipetteerballon.png>

⁴https://en.wikipedia.org/wiki/Air_displacement_pipette

Name	Min. volume (μL)	Max. volume (μL)	Color	tip size (μL)
P10000	1000	10000	red	10000

These pipettes operate by piston-driven air displacement (Figure 8.2) . A vacuum is generated by the vertical travel of a metal or ceramic piston within an airtight sleeve. As the piston moves upward, driven by the depression of the plunger, a vacuum is created in the space left vacant by the piston. The liquid around the tip moves into this vacuum (along with the air in the tip) and can then be transported and released as necessary. These pipettes are capable of being very precise and accurate. However, since they rely on air displacement, they are subject to inaccuracies caused by the changing environment, particularly temperature and user technique. For these reasons this equipment must be carefully maintained and calibrated, and users must be trained to exercise correct and consistent technique.

 **The recommendation is for using a pipette whose size is nearest to the volume being worked with.**

All adjustable pipettes should be calibrated and checked periodically. The simplest way to accomplish this is to pipette a specific volume of a solution with a known density (remember $\text{density} = \frac{\text{mass (g)}}{\text{volume (ml)}}$). The density of distilled water is 1.0 g/ml. Therefore, the accuracy of the pipette or your pipetting skills (if you use an already calibrated pipette as in these experiments) can be checked by determining the mass of a specific volume of water using an analytical balance. We will use this technique in the following two pipetting exercises.

8.1 Exercise 1: using graduated (serological) macropipettes

8.1.1 Experimental procedures

1. Obtain a small beaker and fill it with distilled water.
2. Obtain a weigh boat.
3. Turn on the microbalance.
4. Put the empty weigh boat on the microbalance and set the scale to zero.
5. Obtain a graduated pipette.
6. Notice that the tapered delivery tip at one end of the pipette.
7. Always hold the pipette upright and straight.
8. Remove the pipette from the plastic wrapper by pulling apart the wrapper at the top (blunt) end of the pipette.
9. Attach the rubber bulb or mechanical pipetter to the other (blunt) end of the pipette.
10. Dip the pipette into the water in the small beaker. Don't touch the bottom of the beaker.
11. Draw up the water until the meniscus reaches the position on the scale printed on the pipette corresponding to the desired volume.
12. Dispense the water into a weigh boat.
13. Determine the mass of the dispensed water and record your result in Table 8.2.
14. Empty the weigh boat, put it on the balance and set the balance to zero again.
15. Repeat this procedure twice for each pipette size and volume listed in Table 8.2.
16. After you have completed all measurements, calculate the average for each triplicate set, and the percentage error (relative to the average value) for each set of measurements and enter your results in Table 8.2.

$$\text{Error}(\%) = \frac{\text{expected mass} - \text{experimentally determined mass}}{\text{expected mass}} \times 100\%$$



Figure 8.1: (ref:bulb)

Your results should be within $\pm 5\%$ of the correct volume set on the dial of the pipette.

Table 8.2: Experimental determination of graduated pipette volumes.

Name	Volume (ml)	Trial 1	Trial 2	Trial 3	Average	Error (%)
1 ml	0.1					
1 ml	0.5					
1 ml	1					
5 ml	0.5					
5 ml	1					
5 ml	5					
10 ml	1					
10 ml	5					
10 ml	10					
20 ml	5					
20 ml	10					
20 ml	20					

8.2 Exercise 2: using air displacement micropipettes

8.2.1 Experimental procedures

1. Your laboratory instructor will demonstrate the use of the micropipette before you begin your own experiment.
2. Obtain a micropipette.
3. Identify the main parts of the micropipette (Figure 8.2)
 - Plunger button
 - Tip ejector button
 - Volume adjustment dial
 - Digital volume indicator Shaft
 - Attachment point for a disposable tip
4. Set the pipetting volume. Do not adjust the volume beyond the maximum setting! This will damage the pipettor.
5. Attach a tip to the pipette.
6. Hold the pipette vertically and press the plunger button to the first stop.
7. Dip the tip into the beaker with water and slowly release it, moving the plunger button to the home position.
8. Press the plunger button all the way to the second stop to dispense the sample.
9. Now, carefully draw up and dispense the volumes of distilled water listed in Table 8.3 with the indicated

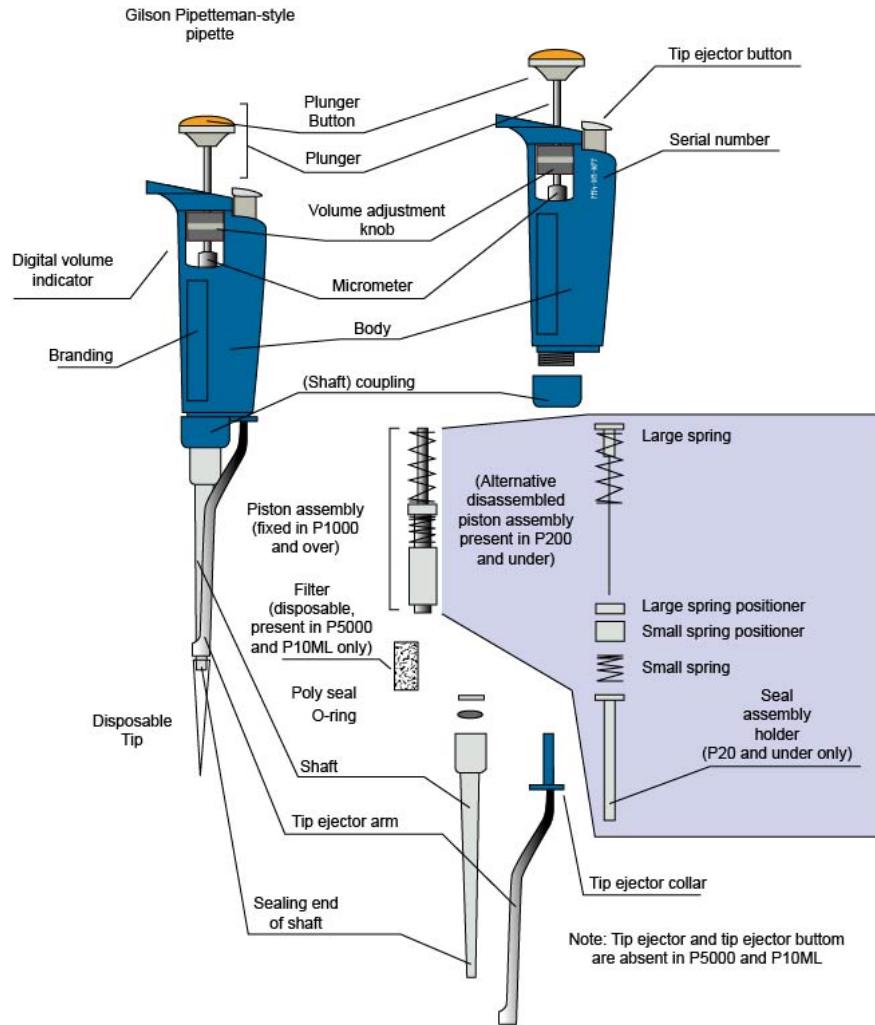


Figure 8.2: Schematic of an air displacement pipette. The “Digital volume indicator” is a dial display that indicates the digits (i.e. unrelated to electronic). The components vary between brands and different volume sizes have different components, for example the piston in a P2 is needle-like and can be separated with ease from the piston assembly whereas in a P10 it is drum-like and over 1 cm in diameter and is enclosed in plastic. By Squidonius [CC BY-SA 3.0] (<https://creativecommons.org/licenses/by-sa/3.0/>), from Wikimedia Commons.

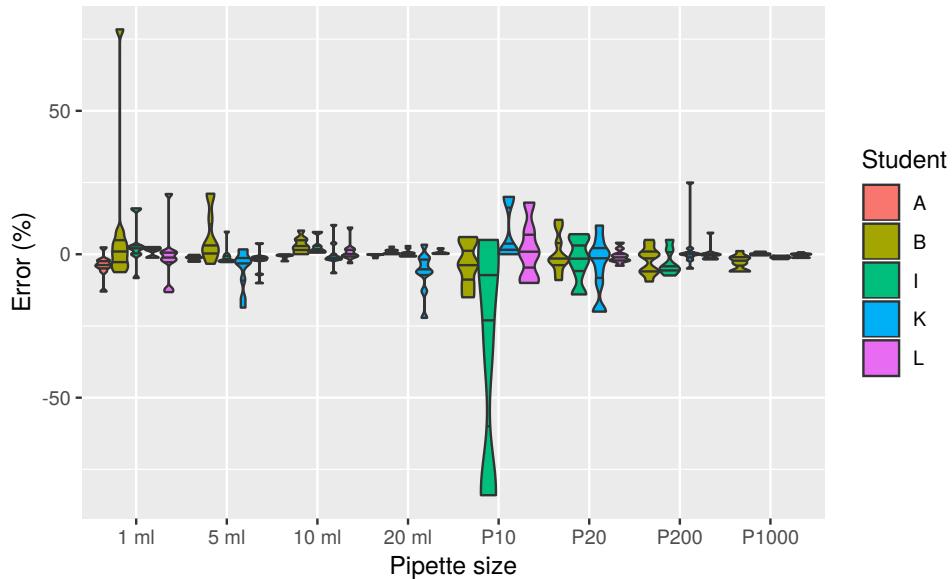


Figure 8.3: Graphical representation of the results obtained by the students performing Experiments 1 and 2 in the Genetics course in the spring term 2019.

$$\text{Error}(\%) = \frac{\text{expected mass} - \text{experimentally determined mass}}{\text{expected mass}} \times 100\%$$

Your results should be within $\pm 5\%$ of the correct volume set on the dial of the pipette.

Table 8.3: Experimental determination of micropipette volumes.

Name	Volume (μL)	Trial 1	Trial 2	Trial 3	Average	Error (%)
P10	5					
P10	10					
P20	5					
P20	10					
P20	20					
P200	20					
P200	50					
P200	200					
P1000	200					
P1000	500					
P1000	1000					

Obtain an empty microcentrifuge tube and perform the following pipetting exercises transferring distilled water from a beaker to and from a microcentrifuge tube:

- Use a fresh tip each time.
- Do not touch the tube with the pipette, only with the tip.
- Large volume:
 1. Add 265 μL to a tube
 2. Add 435 μL to the same tube

3. Remove 615 μL from the tube
 4. Remove 32 μL from the tube
 - Small volume:
 1. Add 5.3 μL to a tube
 2. Add 15.5 μL to the same tube
 3. Remove 12.6 μL from the tube
 4. Remove 7.2 μL from the tube
 - Very small volume:
 1. Add 1.3 μL
 2. Add 3.8 μL
 3. Remove 4.2 μL
 4. Remove 0.7 μL
1. Hold the pipette almost vertically in the palm of your right hand (if you are right handed).
 2. Depress plunger to first stop and hold. Dip tip 2-4 mm into the beaker with distilled water.
 3. Gently release the plunger. If you release too quickly, you will create aerosols (small droplets) which will contaminate the pipette.
 4. Wait for a second or so and confirm that all the liquid has been taken up.
 5. Touch the pipette tip to the inside wall of the beaker to dislodge any remaining droplets adhering to the outside of the tip.
 6. Check that there is no air space at the very end of the tip.
 7. Put the tube containing the sample in the tube rack and pick up the tube that you want to add the sample to. Hold the tube between your thumb and index finger of your left hand.
 8. To expel sample into the tube, touch the tip to wall of the tube and slowly depress the plunger to the first, and then to the second stop to expel the liquid.
 9. Keeping the plunger at the second stop, slide the tip out of the fluid, along the tube wall and out of the tube.
 10. Look for the small droplet of liquid in the tube to make sure that it was transferred.

8.3 Exercise 3: using a microcentrifuge

A laboratory centrifuge⁵ is a piece of laboratory equipment, driven by a motor, which spins liquid samples at high speed to separate heavy materials in a substance from lighter materials in the same substance. The heavier particles will clump at the bottom and form a pellet, while the solution above the pellet is referred to as the supernatant. Centrifuge speeds are measured in RPM (revolutions per minute) or RCF (relative centrifugal force; often referred to as ‘g-force’).

Like all other centrifuges, laboratory centrifuges work by the sedimentation principle, where the centripetal acceleration is used to separate substances of greater and lesser density.

Microcentrifuges sit on the bench top and are used to spin small (2 ml or less) liquid samples at high speeds (up to tens of thousands times g-force). Centrifugation of small samples is important for many biological applications, such as pelleting nucleic acids or proteins from solution. Generally, we also use “microcentrifuges” to gather the typically very small volumes of liquids that we work with into the bottom of the tube.

All centrifuges must be balanced or they will be unstable and will not spin at the correct speed. To balance a centrifuge, tubes are symmetrically arranged around the center. An odd number of tubes can be balanced with empty tubes or “balance” tubes filled with water (always label the ALL tubes appropriately). It is a good practice to always place the tubes in the centrifuge such that the hinges face outward.

⁵https://en.wikipedia.org/wiki/Laboratory_centrifuge

This way, you will know that the pellet forms directly below the hinge of the microcentrifuge tube which helps when the pellet is small and difficult to see.

8.3.1 Experimental procedures

1. Obtain an empty microcentrifuge tube.
2. Label your tube.
3. Transfer 200 μL of distilled water to the microcentrifuge tube.
4. One by one, put your tubes into the centrifuge with the hinge facing outward.
5. Make sure that the centrifuge is properly balanced.
6. Spin the tubes at 10,000 rpm for 30 seconds.
7. Open the microcentrifuge and remove your tube.
8. Discard your tube and clean up your bench.

8.4 Review Questions

1. What type of pipette do you choose for pipetting small volumes?
2. How do you determine what size pipette to use?
3. Why is it important to balance centrifuges?
4. How do you balance a centrifuge when loading it?

Chapter 9

DNA Extraction and Polymerase Chain Reaction

9.1 DNA

Deoxyribonucleic acid (DNA) is a thread-like chain of nucleotides carrying the genetic instructions used in the growth, development, functioning and reproduction of all known living organisms and many viruses. DNA and ribonucleic acid (RNA) are nucleic acids; alongside proteins, lipids and complex carbohydrates (polysaccharides), they are one of the four major types of macromolecules that are essential for all known forms of life. Most DNA molecules consist of two biopolymer strands coiled around each other to form a double helix.

The two DNA strands are called polynucleotides since they are composed of simpler monomer units called nucleotides. Each nucleotide is composed of one of four nitrogen-containing nucleobases (cytosine [C], guanine [G], adenine [A] or thymine [T]), a sugar called deoxyribose, and a phosphate group. The nucleotides are joined to one another in a chain by covalent bonds between the sugar of one nucleotide and the phosphate of the next, resulting in an alternating sugar-phosphate backbone. The nitrogenous bases of the two separate polynucleotide strands are bound together, according to base pairing rules (A with T and C with G), with hydrogen bonds to make double-stranded DNA.

The complementary nitrogenous bases are divided into two groups, pyrimidines and purines. In a DNA molecule, the pyrimidines are thymine and cytosine, the purines are adenine and guanine.

DNA stores biological information. The DNA backbone is resistant to cleavage, and both strands of the double-stranded structure store the same biological information. This information is replicated as and when the two strands separate. A large part of DNA (more than 98% for humans) is non-coding, meaning that these sections do not serve as patterns for protein sequences.

The two strands of DNA run in opposite directions to each other and are thus antiparallel. Attached to each sugar is one of four types of nucleobases (informally, bases). It is the sequence of these four nucleobases along the backbone that encodes biological information. RNA strands are created using DNA strands as a template in a process called transcription. Under the genetic code, these RNA strands are translated to specify the sequence of amino acids within proteins in a process called translation.

Within eukaryotic cells, DNA is organized into long structures called chromosomes. During cell division these chromosomes are duplicated in the process of DNA replication, providing each cell its own complete set of chromosomes. Eukaryotic organisms (animals, plants, fungi and protists) store most of their DNA inside the cell nucleus and some of their DNA in organelles, such as mitochondria or chloroplasts. In contrast prokaryotes (bacteria and archaea) store their DNA only in the cytoplasm. Within the eukaryotic chromosomes, chromatin proteins such as histones compact and organize DNA. These compact structures guide the

interactions between DNA and other proteins, helping control which parts of the DNA are transcribed.

DNA was first isolated by Friedrich Miescher¹ in 1869. Its molecular structure was first identified by James Watson² and Francis Crick³ at the Cavendish Laboratory within the University of Cambridge in 1953, whose model-building efforts were guided by X-ray diffraction data acquired by Raymond Gosling⁴, who was a post-graduate student of Rosalind Franklin⁵.

DNA isolation is a process of purification of DNA from sample using a combination of physical and chemical methods. Currently it is a routine procedure in molecular biology.

9.2 Polymerase chain reaction

Polymerase chain reaction⁶ (PCR) is a technique used in molecular biology to amplify a single copy or a few copies of a segment of DNA across several orders of magnitude, generating thousands to millions of copies of a particular DNA sequence. PCR was developed in 1983 by Kary Mullis⁷ when he was an employee of the Cetus Corporation⁸, one of the first biotech companies. He won the Nobel Prize in Chemistry in 1993 for the invention of PCR. It is an easy, cheap, and reliable way to repeatedly replicate a focused segment of DNA, a concept which is applicable to numerous fields in modern biology and related sciences. PCR is probably the most widely used technique in molecular biology and has revolutionized the field.

PCR is now a common and often indispensable technique used in clinical and research laboratories for a broad variety of applications. These include DNA cloning for sequencing, gene cloning and manipulation, gene mutagenesis; construction of DNA-based phylogenies, or functional analysis of genes; diagnosis and monitoring of hereditary diseases; amplification of ancient DNA; analysis of genetic fingerprints for DNA profiling (for example, in forensic science and parentage testing); and detection of pathogens in nucleic acid tests for the diagnosis of infectious diseases. In 1993, Mullis was awarded the Nobel Prize in Chemistry along with Michael Smith for his work on PCR.

The vast majority of PCR methods rely on thermal cycling, which involves exposing the reactants to cycles of repeated heating and cooling, permitting different temperature-dependent reactions—specifically, DNA melting and enzyme-driven DNA replication—to quickly proceed many times in sequence. Primers (short DNA fragments) containing sequences complementary to the target region, along with a DNA polymerase⁹ (e.g. Taq polymerase), after which the method is named, enable selective and repeated amplification. As PCR progresses, the DNA generated is itself used as a template for replication, setting in motion a chain reaction in which the original DNA template is exponentially amplified. The simplicity of the basic principle underlying PCR means it can be extensively modified to perform a wide array of genetic manipulations. PCR is not generally considered to be a recombinant DNA method, as it does not involve cutting and pasting DNA, only amplification of existing sequences.

Almost all PCR applications employ a heat-stable DNA polymerase, such as Taq polymerase, an enzyme originally isolated from the thermophilic bacterium *Thermus aquaticus*. If heat-susceptible DNA polymerase is used, it will denature every cycle at the denaturation step. Before the use of Taq polymerase, DNA polymerase had to be manually added every cycle, which was a tedious and costly process. This DNA polymerase enzymatically assembles a new DNA strand from free nucleotides, the building blocks of DNA, by using single-stranded DNA as a template and DNA oligonucleotides (the primers mentioned above) to initiate DNA synthesis.

¹https://en.wikipedia.org/wiki/Friedrich_Miescher

²https://en.wikipedia.org/wiki/James_Watson

³https://en.wikipedia.org/wiki/Francis_Crick

⁴https://en.wikipedia.org/wiki/Raymond_Gosling

⁵https://en.wikipedia.org/wiki/Rosalind_Franklin

⁶https://en.wikipedia.org/wiki/Polymerase_chain_reaction

⁷https://en.wikipedia.org/wiki/Kary_Mullis

⁸https://en.wikipedia.org/wiki/Cetus_Corporation

⁹https://en.wikipedia.org/wiki/DNA_polymerase

In the first step, the two strands of the DNA double helix are physically separated at a high temperature in a process called DNA melting. In the second step, the temperature is lowered and the two DNA strands become templates for DNA polymerase to selectively amplify the target DNA. The selectivity of PCR results from the use of primers that are complementary to sequence around the DNA region targeted for amplification under specific thermal cycling conditions.

The PCR, like recombinant DNA technology, has had an enormous impact in both basic and diagnostic aspects of molecular biology because it can produce large amounts of a specific DNA fragment from small amounts of a complex template. Recombinant DNA¹⁰ techniques create molecular clones by conferring on a specific sequence the ability to replicate by inserting it into a vector and introducing the vector into a host cell. PCR represents a form of “in vitro cloning” that can generate, as well as modify, DNA fragments of defined length and sequence in a simple automated reaction. In addition to its many applications in basic molecular biological research, PCR promises to play a critical role in the identification of medically important sequences as well as an important diagnostic one in their detection.

9.3 DNA Extraction

9.4 Experimental Procedures

1. Obtain two microcentrifuge tubes and label them with your initials.
2. Add 200 μ l of lysis buffer into each tube.
3. Weigh 50–100 mg of plant tissue. Record the weight of the tissue.

Table 9.1: Plant material record.

Plant Name	Part of Plant (Leaf, Root, etc.)	Weight (mg)

4. Chop the plant tissue into small (1-2 mm in diameter) pieces using a razor blade.
5. Add the plant material to the microcentrifuge tubes with the lysis buffer.
6. Using a micropesle, carefully grind the plant tissue into fine particles (until no more chunks are visible). Be careful not to spill the lysis buffer. If any material gets stuck at the bottom of the tube use a clean pipet tip to dislodge it and continue grinding. You may have to continue grinding for several minutes before the leaves have been completely homogenized.
7. Add an additional 500 μ l of lysis buffer.
8. Grind until the lysate is homogeneous.
9. Close the microcentrifuge tube.
10. Spin for 5 min at full speed in a microcentrifuge at room temperature. Make sure to place the tubes in the rotor so that the microcentrifuge is balanced.
11. While tubes are centrifuging, label two microcentrifuge tubes for each extract.
12. Add 500 μ l of 70% ethanol to each tube.
13. Retrieve your tubes from the microcentrifuge.
14. Carefully remove 400 μ l of supernatant without disturbing the pellet and add it to the new tubes containing the 500 μ l of 70% ethanol. Avoid transferring any solid plant material to the ethanol; if

¹⁰https://en.wikipedia.org/wiki/Recombinant_DNA

necessary, re-centrifuge the lysate. Mix the lysate and ethanol thoroughly by pipetting up and down. Cap the tubes.

!
If a precipitate is visible, spin the microcentrifuge tube again for 5 min in a microcentrifuge at full speed to pellet the precipitated material and use the supernatant for the next stages.

15. Label the top of a mini column for each sample with your initials. Place the columns in 2 ml cap-less collection tubes.
16. Remove 800 μ l of cleared plant lysate from your microcentrifuge tube and add it to the column.

!
Be sure that the lysate that is added to the column is clear and does not contain any precipitate. If necessary, centrifuge the lysate again to pellet any plant material prior to loading the lysate on the column.

17. Place the cap-less collection tubes containing the columns into the microcentrifuge. Ensure that the centrifuge is balanced. Spin for 1 min at full speed at room temperature.
18. Discard the flowthrough from the collection tubes.

!
If all of the supernatant does not pass through column, centrifuge again for 1 min. If there is any supernatant left in the column, carefully remove and discard it. Do not disrupt the bed of the column.

19. Add 700 μ l of wash buffer to each column.
20. Spin at full speed at room temperature for 1 min. Discard the flowthrough.
21. Repeat steps 19 and 20 two more times for a total of 3 washes.
22. After the final wash step, discard the flow-through and put the columns back into the collection tubes. Centrifuge the columns for 2 min at full speed in the microcentrifuge at room temperature. This step removes the residual ethanol which will interfere with PCR if present.
23. Label a fresh microcentrifuge tube and transfer the column to the new tube.
24. Obtain a microcentrifuge tube and fill it with double distilled water from the bottle in the 70 °C heat water bath.
25. Immediately add 80 μ l of 70 °C sterile water to the bed of each column, making certain that the water wets the column bed and let it sit for 1 min.
26. Place the column in the microcentrifuge tube into the microcentrifuge. Orient the loose cap of the microcentrifuge tube downward, toward the center of the rotor, to minimize friction and damage to the cap during centrifugation.
27. Spin at full speed in the microcentrifuge at room temperature for 2 min.
28. Remove the spin column from the microcentrifuge tube.

9.5 GAPDH Polymerase Chain Reaction

9.6 Experimental Procedures

1. Obtain a PCR tube.
2. Label the tube with your initials.
3. Add 20 μ l of 2 \times MMIP (master mix with initial primers) into each PCR tube.
4. Add 15 μ l of sterile water to each tube.
5. Add 5 μ l of the DNA template to each tube and gently pipet up and down to mix reagents.
6. Place your PCR tubes into the thermal cycler.

7. The PCR reaction will run for the next several hours using the program listed in Table ??.
8. The laboratory technicians will remove the PCR tubes from the thermocycler after the PCR has finished and store the tubes at -20 °C until we continue our experiments next week.

Table 9.2: PCR protocol for GAPDH amplification.

PCR Step	Temperature	Time	Number of cycles
Initial denaturation	95 °C	5 min	1
Denaturation	95 °C	1 min	40
Annealing	52 °C	1 min	40
Extension	72 °C	2 min	40
Final extension	72 °C	6 min	1
Hold	15 °C	hold	

9.7 Review Questions

1. What is PCR?
2. What are the components of a PCR?
3. What is a thermocycler (PCR machine)?

Chapter 10

Exonuclease I Digestion And Nested PCR

10.1 Exonuclease I Digestion

Exonucleases are enzymes that work by cleaving nucleotides one at a time from the end (exo) of a polynucleotide chain. A hydrolyzing reaction that breaks phosphodiester bonds at either the 3' or the 5' end occurs. Its close relative is the endonuclease, which cleaves phosphodiester bonds in the middle (endo) of a polynucleotide chain. Exonuclease I (Phosphodiesterase I) hydrolytically removes 5'-nucleotides successively from the 3'-hydroxy termini of 3'-hydroxy-terminated oligonucleotides. We use exonuclease I to digest the oligonucleotide primers from the first PCR reaction so that they will not allow for PCR amplification in the second PCR. After the initial PCR primers have been digested, exonuclease I also needs to be inactivated before it is introduced into fresh PCR reactions to prevent it digesting the nested PCR primers.

10.2 Experimental Procedures

1. Add 1 μ l of exonuclease I to each PCR tube containing the amplified gDNA.
2. Mix well by pipetting up and down.
3. Incubate at 37 °C for 15 min.
4. Incubate at 80 °C for 15 min to heat-inactivate the exonuclease I enzyme.

10.3 Nested polymerase chain reaction

Nested polymerase chain reaction (Nested PCR) is a modification of polymerase chain reaction intended to reduce non-specific binding in products due to the amplification of unexpected primer binding sites.

Nested polymerase chain reaction involves two sets of primers, used in two successive runs of polymerase chain reaction, the second set intended to amplify a secondary target within the first run product. This allows amplification for a low number of runs in the first round, limiting non-specific products. The second nested primer set should only amplify the intended product from the first round of amplification and not non-specific product. This allows running more total cycles while minimizing non-specific products. This is useful for very rare templates or PCR with high background.

10.4 Experimental Procedures

1. Label a microcentrifuge tube for each of the two exonuclease-treated first-round PCR tubes.
2. To dilute each first-round PCR sample 50× (to 1/50 the original concentration), add 98 μ l of sterile water into each of the labeled microcentrifuge tubes.
3. Using a fresh tip each time, pipet 2 μ l of each first-round PCR into the corresponding microcentrifuge tube containing 98 μ l water. Close the cap.
4. Vortex or flick the tube with your finger to mix. Spin briefly in a microfuge to collect the liquid at the bottom of the tube.
5. Label 4 fresh PCR tubes with numbers 1 to 4 and your initials.
6. Place them in PCR tube adaptors on ice.
7. Pipet 20 μ l of 2 \times MMNP into each of the three PCR tubes.
8. Using a fresh tip, pipet 20 μ l of the each diluted first-round PCR into tube #1 and #2, respectively. Close the cap.
9. Use a fresh pipet tip to add 20 μ l of plasmid DNA (positive control) to tube #3. Gently pipet up and down to mix the reagents. Recap the tubes.
10. Use a fresh pipet tip to add 20 μ l of plasmid DNA (positive control) to tube #3.
11. Use a fresh pipet tip to add 20 μ l of sterile water (negative control) to PCR tube #4.
12. Place the PCR tubes into the thermal cycler. The PCR will run for the next several hours using the Nested GAPDH PCR program listed in Table 10.1.

Table 10.1: PCR protocol for second round (nested) GAPDH amplification.

PCR Step	Temperature	Time	Number of cycles
Initial denaturation	95 °C	5 min	1
Denaturation	95 °C	1 min	40
Annealing	52 °C	1 min	40
Extension	72 °C	2 min	40
Final extension	72 °C	6 min	1
Hold	15 °C	hold	

10.5 Review Questions

1. What is an exonuclease?
2. What is nested PCR?

Chapter 11

Agarose Gel Electrophoresis, Ligation and Transformation

11.1 Agarose gel electrophoresis

Electrophoresis (from the Greek “electrophoron” meaning “to bear electrons”) is the motion of dispersed particles relative to a fluid under the influence of a spatially uniform electric field.

Agarose gel electrophoresis is a method of gel electrophoresis used in biochemistry, molecular biology, genetics, and clinical chemistry to separate a mixed population of macromolecules such as DNA or proteins in a matrix of agarose, one of the two main components of agar. The proteins may be separated by charge and/or size, and the DNA and RNA fragments by length. Biomolecules are separated by applying an electric field to move the charged molecules through an agarose matrix, and the biomolecules are separated by size in the agarose gel matrix.

Agarose gels are easy to cast and are particularly suitable for separating DNA of size range most often encountered in laboratories. The separated DNA is visualized using a stain, most commonly under UV light, and the DNA fragments can be extracted from the gel with relative ease. Most agarose gels used are made using 0.5–3% agarose which is dissolved in a suitable electrophoresis buffer.

Agarose gel is easy to cast, has relatively fewer charged groups, and is particularly suitable for separating DNA of size range most often encountered in laboratories, which accounts for the popularity of its use. The separated DNA may be viewed with stain, most commonly under UV light, and the DNA fragments can be extracted from the gel with relative ease. Most agarose gels used are between 0.7–2% dissolved in a suitable electrophoresis buffer.

11.2 Experimental Procedures

 Take special care when pipetting very small volumes. Make sure only the soft stop of the pipet is used, when pulling up reagents even though it may feel like a very small movement. Also, look at the end of the pipet tip to be sure that the correct volume of reagent is in the tip. After adding the reagent to the tube, be sure that the pipet tip is empty. Never reuse a pipet tip.

2. Get the Erlenmeyer flask containing 1 g of agarose powder
3. Add 100 ml of 1× TAE (Tris base, acetic acid, EDTA) running buffer.
4. Add 10 μ l of SybrGreenTM dye (10,000× stock solution).

5. Heat in the microwave at full power for 1 minute.
6. Swirl to make sure that all powder has dissolved, and the solution is clear.
7. Add the comb into the comb slot.
8. Pour the solution onto the gel tray in the gel box.
9. Label a microcentrifuge tube for each PCR sample.
10. Transfer 20 μ l of each of first-round PCR reaction into the labeled microcentrifuge tube. Add 4 μ l of loading dye and stain into the tube. Pipette up and down to mix.
11. To assess the success of the nested PCR round, pipet 5 μ l of each nested PCR into a microcentrifuge tube and mix it with 1 μ l of 6 \times loading dye and stain. A smaller volume of the nested PCR is loaded because nested PCR is usually more efficient and produces very intense bands that can obscure bands in adjacent wells if samples are overloaded.
12. The gel will need about 30 minutes to solidify.
13. Start setting up the PCR product purification.

11.3 PCR Product Purification

11.4 Experimental Procedures

1. Label a PCR Kleen spin column and a cap-less collection tube with your initials.
2. Label a capped microcentrifuge tube with your initials, “purified PCR product,” and the plant name.
3. Resuspend the beads in the PCR Kleen spin column by vortexing. Return the beads to the bottom of the column with a sharp downward flick.
4. Snap the bottom off the spin column, remove the cap, and place the column in the cap-less collection tube. Discard cap and bottom of column.
5. Place the spin column, still in the cap-less collection tube, into the microcentrifuge. Make sure that your tube is placed in the rotor with another group’s or with a balance tube so that the microcentrifuge is balanced. Centrifuge columns at 735 \times g for 2 min. Do not use the top speed of the microcentrifuge for this step.
6. Move the spin column to the labeled microcentrifuge tube. Discard the flowthrough and the collection tube.
7. Pipet 30 μ l of the nested PCR product onto the top of the column bed in the spin column, without disturbing the resin (or column bed). Save 5 μ l of this original (not yet purified) yellow PCR sample and mix it with 1 μ l of 6 \times loading dye and stain for gel electrophoresis.
8. Place the column in the labeled microcentrifuge tube into the microcentrifuge. It is best to orient the cap of the microcentrifuge tube downward, toward the center of the rotor, to minimize friction and damage to the cap during centrifugation. Centrifuge at 735 \times g for 2 min. Make sure that another group’s sample counterbalances the microcentrifuge.
9. Remove the spin column from the microcentrifuge tube and discard the column. Cap the microcentrifuge tube, which now contains the purified PCR product. Store at 4 °C for up to 2 weeks or at -20 °C long term.
10. Add 1 μ l of 6 \times loading dye and stain into 5 μ l of the purified sample.
11. Electrophorese 5 μ l of this mix next to 5 μ l of the not purified sample on a 1% agarose gel to visualize differences between the original and the purified sample.
12. Check if the agarose gel has solidified. If so, proceed to the next step or wait until the gel is ready.
13. Pour electrophoresis running buffer into the chamber until it just covers the gel by 1–2 mm.
14. Load 10 μ l of the 500 bp molecular weight ruler into the first well.
15. Load 20 μ l from each microcentrifuge tube containing an initial PCR with added loading dye. Load 5 μ l from each tube containing a nested PCR with added loading dye into the wells of the gel according to your plan.
16. Connect your electrophoresis chamber to the power supply and turn on the power. Run the gel at 100 V for 30 min.
17. While the gel is running, start setting up the ligation reactions.

11.5 Ligation

In molecular biology, ligation¹ is the joining of two nucleic acid fragments through the action of an enzyme. It is an essential laboratory procedure in the molecular cloning of DNA whereby DNA fragments are joined together to create recombinant DNA molecules, such as when a foreign DNA fragment is inserted into a plasmid. The ends of DNA fragments are joined together by the formation of phosphodiester bonds between the 3'-hydroxyl of one DNA terminus with the 5'-phosphoryl of another.

11.6 Experimental Procedures

1. Label a microcentrifuge tube with your initials and “l” for “ligation.”
2. Briefly centrifuge the stock tubes containing the 2× ligation reaction buffer and proofreading polymerase in a microfuge to force contents to bottom of tubes.
3. Add 5.0 μ l 2× ligation reaction buffer to the labeled microcentrifuge tube.
4. Add 1.0 μ l purified PCR product
5. Add 2.5 μ l sterile water
6. Add 0.5 μ l proofreading polymerase
7. Close the cap and mix well. Centrifuge briefly in a microfuge to collect the contents at the bottom of the tube.
8. Incubate the tube at 70 °C for 5 min. 70 °C is the optimal temperature for the proofreading polymerase to blunt the PCR fragment.
9. Cool the tube on ice for 2 min. This re-condenses water vapor to maintain the reaction volume. Centrifuge the tube briefly to collect the contents at the bottom of the tube. Maintain the tube at room temperature.
10. Add 0.5 μ l T4 DNA ligase.
11. Add 0.5 μ l pJet1.2 blunted vector.
12. Close the cap and mix well.
13. Centrifuge briefly in a microcentrifuge to collect the contents at the bottom of the tube.
14. Incubate the tube at room temperature for 5–10 min.
15. Pipet 5 μ l of the ligation reaction into a fresh microcentrifuge tube labeled with your initials and “t” for “transformation” and store it on ice until needed for the transformation.
16. Take the gel out of the electrophoresis box to visualize the DNA bands and acquire an image of your gel.

11.7 Transformation

In molecular biology, transformation² is the genetic alteration of a cell resulting from the direct uptake and incorporation of exogenous genetic material from its surroundings through the cell membrane(s). For transformation to take place, the recipient bacteria must be in a state of competence, which might occur in nature as a time-limited response to environmental conditions such as starvation and cell density, and may also be induced in a laboratory.

Transformation is one of three processes for horizontal gene transfer, in which exogenous genetic material passes from one bacterium to another, the other two being conjugation (transfer of genetic material between two bacterial cells in direct contact) and transduction (injection of foreign DNA by a bacteriophage virus into the host bacterium). In transformation, the genetic material passes through the intervening medium, and uptake is completely dependent on the recipient bacterium.

¹[https://en.wikipedia.org/wiki/Ligation_\(molecular_biology\)](https://en.wikipedia.org/wiki/Ligation_(molecular_biology))

²[https://en.wikipedia.org/wiki/Transformation_\(genetics\)](https://en.wikipedia.org/wiki/Transformation_(genetics))

“Transformation” may also be used to describe the insertion of new genetic material into nonbacterial cells, including animal and plant cells; however, because “transformation” has a special meaning in relation to animal cells, indicating progression to a cancerous state, the process is usually called “transfection”. Transformation in bacteria was first demonstrated in 1928 by the British bacteriologist Frederick Griffith³. Griffith was interested in determining whether injections of heat-killed bacteria could be used to vaccinate mice against pneumonia. However, he discovered that a non-virulent strain of *Streptococcus pneumonia* could be made virulent after being exposed to heat-killed virulent strains. Griffith hypothesized that some “transforming principle” from the heat-killed strain was responsible for making the harmless strain virulent. In 1944 this “transforming principle” was identified as being genetic by Oswald Avery⁴, Colin MacLeod⁵, and Maclyn McCarty⁶. They isolated DNA from a virulent strain of *S. pneumoniae* and using just this DNA were able to make a harmless strain virulent. They called this uptake and incorporation of DNA by bacteria “transformation”. The results of Avery et al.’s experiments were at first skeptically received by the scientific community and it was not until the development of genetic markers and the discovery of other methods of genetic transfer (conjugation in 1947 and transduction in 1953) by Joshua Lederberg⁷ that Avery’s experiments were accepted.

11.8 Experimental Procedures

1. Add 1.5 ml C-growth medium into a 15 ml culture tube. Label tube with your initials and warm it at 37 °C for at least 10 min.
2. Label 2 LB amp IPTG agar plates with your initials (on the bottom of the plate, not the lid). Also label one of the plates “pGAP” for the control plasmid and the other for your ligation (pJET + your plant name).
3. Pipet 150 µl of fresh *E. coli* starter culture (inoculated one day prior) into the prewarmed C-growth medium and place in a 37 °C incubator or water bath for 20–40 min shaking at 275 rpm.
4. Label a 1.5 microcentrifuge tube with your initials and “competent cells.”
5. Prepare transformation buffer by combining 250 µl of transformation reagent A and 250 µl of transformation reagent B into a tube labeled “TF buffer” and mix thoroughly with a vortex mixer. Keep on ice until use.
6. After bacteria have grown in C-growth medium for 20–40 min at 37 °C with shaking, transfer the entire culture to the tube labeled “competent cells” by decanting. Do not put the actively growing cell culture on ice at this step.
7. Centrifuge the bacterial culture in a microcentrifuge at top speed for 1 min. Make sure that the microcentrifuge is balanced. Immediately put the pelleted bacterial culture on ice.
8. Locate the pellet of bacteria at the bottom of the tube. Remove the culture supernatant, avoiding the pellet, using a 1,000 µl pipet. Keep the cells on ice.
9. Pipet 300 µl of ice-cold transformation buffer into the microcentrifuge tube containing the bacterial pellet. Resuspend the pellet by gently pipetting up and down in the solution above the pellet with a 1,000 µl pipet, and gradually wear away the pellet from the bottom of the tube. Make sure that the bacteria are fully resuspended, with no clumps. Avoid removing the cells from the ice bucket for more than a few seconds.
10. Incubate the resuspended bacteria on ice for 5 min.
11. Centrifuge the bacteria in a microcentrifuge for 1 min, then place back in ice bucket immediately.

 Ensure that the bacteria are on ice immediately prior to and immediately following centrifugation. If the centrifuge is not close to the lab bench, take the entire ice bucket to the microcentrifuge so that the bacteria are only out of the ice bucket for 1 minute. Use a refrigerated microcentrifuge, if available.

³https://en.wikipedia.org/wiki/Frederick_Griffith

⁴https://en.wikipedia.org/wiki/Oswald_Avery

⁵https://en.wikipedia.org/wiki/Colin_Munro_MacLeod

⁶https://en.wikipedia.org/wiki/Maclyn_McCarty

⁷https://en.wikipedia.org/wiki/Joshua_Lederberg

12. Remove the supernatant from the pellet using a 1,000 μ l pipet.



After this step, it is very important to keep the bacteria on ice as much as possible during this procedure. Transformation efficiency will be severely compromised if the cells warm up. It is very important to treat the bacteria extremely gently during this procedure — the bacteria are very fragile and your transformation efficiency will be compromised unless you are very gentle.

8. Locate the pellet of bacteria at the bottom of the tube. Remove the culture
9. Pipet 120 μ l of ice-cold transformation buffer onto the pellet and resuspend by gently pipetting up and down with a 200 μ l pipet. Be sure that bacteria are fully resuspended with no clumps. Avoid removing the cells from the ice bucket for more than a few seconds.
10. Incubate the resuspended bacteria on ice for 5 min. The cells are now competent for transformation.
11. Label one microcentrifuge tube with your initials and “pGAP T” (for pGAP transformation) and another microcentrifuge tube with your initials and “T” (referred to below as the “plant T” tube).
12. Pipet 1 μ l of control pGAP plasmid into the microcentrifuge tube labeled “pGAP T.”
13. Using a fresh tip, pipet 50 μ l of competent bacteria directly into the ice-cold “plant T” tube containing 5 μ l of your ligation, and gently pipet up and down 2 times to mix.
14. Incubate the transformations for 10 min on ice.
15. Retrieve the warm LB amp IPTG agar plates from the 37 °C incubator.
16. Pipet the entire volume of each transformation onto the corresponding labeled LB amp IPTG plate and, using an inoculation loop or a sterile spreader, very gently spread the bacteria around the plate — remember that the bacteria are still very fragile! Once the plate is covered, stop spreading. Do not spread for more than 10 sec. It is vital that the LB amp IPTG plates are warm at this step to ensure sufficiently high transformation efficiency. This is the heat shock for the transformation. Spreading the plate until it is dry will also reduce transformation efficiency.
17. Once the volume is absorbed in the agar, cover and place the LB amp IPTG plates upside down and incubate them overnight at 37 °C.

11.9 Review Questions

1. What is agarose gel electrophoresis?
2. What is the overall electrical charge of DNA?
3. What is a ligation?
4. What is transformation?

Chapter 12

Plasmid Purification And Restriction Digest

12.1 Plasmid Purification

A plasmid preparation is a method of DNA extraction and purification for plasmid DNA¹. Many methods have been developed to purify plasmid DNA from bacteria. These methods invariably involve three steps

- Growth of the bacterial culture
- Harvesting and lysis of the bacteria
- Purification of plasmid DNA

12.2 Plasmid

A plasmid is a small DNA molecule within a cell that is physically separated from a chromosomal DNA and can replicate independently. They are most commonly found as small circular, double-stranded DNA molecules in bacteria; however, plasmids are sometimes present in archaea and eukaryotic organisms. In nature, plasmids often carry genes that may benefit the survival of the organism, for example antibiotic resistance. While the chromosomes are big and contain all the essential genetic information for living under normal conditions, plasmids usually are very small and contain only additional genes that may be useful to the organism under certain situations or particular conditions. Artificial plasmids are widely used as vectors in molecular cloning, serving to drive the replication of recombinant DNA sequences within host organisms. In the laboratory, plasmids may be introduced into a cell via transformation.

Plasmids are considered replicons, units of DNA capable of replicating autonomously within a suitable host. However, plasmids, like viruses, are not generally classified as life. Plasmids are transmitted from one bacterium to another (even of another species) mostly through conjugation. This host-to-host transfer of genetic material is one mechanism of horizontal gene transfer, and plasmids are considered part of the mobilome. Unlike viruses (which encase their genetic material in a protective protein coat called a capsid), plasmids are “naked” DNA and do not encode genes necessary to encase the genetic material for transfer to a new host. However, some classes of plasmids encode the conjugative “sex” pilus necessary for their own transfer. The size of the plasmid varies from 1 to over 200 kbp, and the number of identical plasmids in a single cell can range anywhere from one to thousands under some circumstances.

The relationship between microbes and plasmid DNA is neither parasitic nor mutualistic, because each implies the presence of an independent species living in a detrimental or commensal state with the host

¹<https://en.wikipedia.org/wiki/Plasmid>

organism. Rather, plasmids provide a mechanism for horizontal gene transfer within a population of microbes and typically provide a selective advantage under a given environmental state. Plasmids may carry genes that provide resistance to naturally occurring antibiotics in a competitive environmental niche, or the proteins produced may act as toxins under similar circumstances, or allow the organism to utilize particular organic compounds that would be advantageous when nutrients are scarce.

The term plasmid was introduced in 1952 by the American molecular biologist Joshua Lederberg to refer to “any extrachromosomal hereditary determinant.” The term’s early usage included any bacterial genetic material that exists extrachromosomally for at least part of its replication cycle, but because that description includes bacterial viruses, the notion of plasmid was refined over time to comprise genetic elements that reproduce autonomously. Later in 1968, it was decided that the term plasmid should be adopted as the term for extrachromosomal genetic element, and to distinguish it from viruses, the definition was narrowed to genetic elements that exist exclusively or predominantly outside of the chromosome and can replicate autonomously.

12.3 Experimental Procedures

1. Label one cap-less collection tube, one plasmid mini column, and two microcentrifuge tubes for each miniprep culture with your initial and “M” for “miniprep”.
2. Place each column in the appropriate cap-less collection tube.
3. Transfer 1.5 ml of each miniprep culture into one of the appropriately labeled microcentrifuge tubes by pipetting or decanting.
4. Centrifuge the microcentrifuge tubes for 1 min at top speed ($>12,000 \times g$) to pellet the bacteria. Make sure that the microcentrifuge is balanced.
5. Locate the bacterial pellet and remove the supernatant from each tube using a 1,000 μl pipet; avoid touching or extracting the pellet. Dispose of the supernatant in the provided waste bottles in the fume hood.
6. Add the remaining 1.5 ml of the appropriate miniprep culture to the correct microcentrifuge tube containing the bacterial pellet, centrifuge for 1 min and remove supernatant.
7. Add 250 μl of resuspension solution to each tube. Resuspend bacterial pellet by pipetting up and down or vortexing. Use a fresh tip each time. Ensure that no clumps of bacteria remain.
8. Pipet 250 μl of lysis solution into each tube and mix by gently inverting 6–8 times. Do not pipet or vortex this lysate or you risk shearing (fragmenting) the bacterial gDNA, which could contaminate your plasmid preparation.
9. Wait for 5 min, then pipet 350 μl of neutralization solution into each tube and mix by gently inverting 6–8 times. Do not pipet or vortex this lysate. A white precipitate should form.
10. Centrifuge the tubes for 5 min at top speed. Make sure the microcentrifuge is balanced.
11. Pipet supernatant from the centrifuged tubes onto the appropriately labeled column. Avoid transferring any precipitate. If necessary, re-centrifuge the microcentrifuge tubes. Discard the empty microcentrifuge tubes.
12. Centrifuge the columns in the cap-less collection tubes for 1 min at top speed.
13. Discard the flowthrough from the collection tube and place the column back in the collection tube.
14. Pipet 750 μl of wash solution onto each column (ensure that the wash solution has had ethanol added to it prior to use).
15. Centrifuge columns in the cap-less collection tubes for 1 min at top speed.
16. Discard the flowthrough from the collection tube.
17. Place the columns back into collection tubes and centrifuge for an additional 1 min to dry the column.
18. Transfer contents of each column to the appropriately labeled capped “miniprep DNA” microcentrifuge tube and pipet 100 μl of elution solution onto the column.
19. Let the elution solution be absorbed into the column for 2 min.
20. Place the column in the microcentrifuge tube into the centrifuge. It is best to orient the cap of the microcentrifuge tube downward, toward the center of the rotor, to minimize friction and damage to the cap during centrifugation.

21. Centrifuge the columns for 2 min.
22. Discard the columns and cap the tubes containing the eluted sample.
23. Proceed to setting up the restriction digest reactions.

12.4 Restriction Digest

12.5 Experimental Procedures

1. Label one microcentrifuge tube with your initials and “master mix.”
2. Prepare a 2× master mix for BgI II restriction digestion reactions according to Table 12.1 using stock reagents from the common workstation. Use a fresh tip for each reagent.
3. Label a microcentrifuge tube for each plasmid miniprep.
4. Prepare digestion reactions by adding 10 μ l of the BgI II master mix and 10 μ l of each plasmid DNA in the labeled microcentrifuge tubes.
5. Mix tube contents and spin briefly in a microcentrifuge to collect the contents at the bottom of the tube.
6. Incubate reactions at 37 °C for 1 hr. Set timer to sound an alarm after 1 hour.
7. While the restriction digest is incubating, prepare an agarose gel.
8. Get the Erlenmeyer flask containing 1 g of agarose powder
9. Add 100 ml of 1× TAE (Tris base, acetic acid, EDTA) running buffer.
10. Add 10 μ l of SybrGreen™ dye (10,000× stock solution).
11. Heat in the microwave at full power for 1 minute.
12. Swirl to make sure that all powder has dissolved, and the solution is clear.
13. Add the comb into the comb slot.
14. Pour the solution onto the gel tray in the gel box.
15. The gel will need about 30 minutes to solidify.
16. When the timer indicates that the restriction digest has been incubating for 1 hour, combine 5 μ l of miniprep DNA with 15 μ l of sterile water.
17. Briefly centrifuge samples to force the contents to the bottom of the tube.
18. Add 3.5 μ l of 6× loading dye and stain to each of the digested and undigested samples.
19. Load 20 μ l of each sample and 10 μ l of the 500 bp molecular weight ruler according to your plan on the 1% agarose gel.
20. Connect the electrophoresis chamber to the power supply and turn on the power. Run the gel at 100 V for 30 min.
21. Visualize your bands and acquire an image of your gel. Paste the image into your lab notebook and label it accordingly.

Table 12.1: Composition of restriction digest master mix.

Reagent	Volume
10× BgI II reaction buffer	2 μ l
Sterile water	7 μ l
BgI II enzyme	1 μ l
Total	10 μ l

Chapter 13

DNA sequencing

DNA sequencing is the process of determining the precise order of nucleotides within a DNA molecule. It includes any method or technology that is used to determine the order of the four bases—adenine, guanine, cytosine, and thymine—in a strand of DNA. The advent of rapid DNA sequencing methods has greatly accelerated biological and medical research and discovery.

Knowledge of DNA sequences has become indispensable for basic biological research, and in numerous applied fields such as medical diagnosis, biotechnology, forensic biology, virology and biological systematics. The rapid speed of sequencing attained with modern DNA sequencing technology has been instrumental in the sequencing of complete DNA sequences, or genomes of numerous types and species of life, including the human genome and other complete DNA sequences of many animal, plant, and microbial species.

The first DNA sequences were obtained in the early 1970s by academic researchers using laborious methods based on two-dimensional chromatography. Frederick Sanger¹ published a method for “DNA sequencing with chain-terminating inhibitors” in 1977. Walter Gilbert² and Allan Maxam³ at Harvard also developed sequencing methods, including one for “DNA sequencing by chemical degradation”. The chain-termination method developed by Frederick Sanger and coworkers in 1977 soon became the method of choice, owing to its relative ease and reliability. When invented, the chain-terminator method used fewer toxic chemicals and lower amounts of radioactivity than the Maxam and Gilbert method. Because of its comparative ease, the Sanger method was soon automated and was the method used in the first generation of DNA sequencers.

Sanger sequencing is the method which prevailed from the 1980s until the mid-2000s. Over that period, great advances were made in the technique, such as fluorescent labelling, capillary electrophoresis, and general automation. Advancements in sequencing were aided by the concurrent development of recombinant DNA technology, allowing DNA samples to be isolated from sources other than viruses. Following the development of fluorescence-based sequencing methods with a DNA sequencer⁴, DNA sequencing has become easier and orders of magnitude faster. Sanger sequencing and automated DNA sequencers will be used to determine the sequences of our GAPDH clones and be described in more detail in the subsequent chapter (Chapter 13).

The high demand for low-cost sequencing has driven the development of high-throughput (formerly “next-generation”) sequencing technologies that parallelize the sequencing process, producing thousands or millions of sequences concurrently. High-throughput sequencing technologies are intended to lower the cost of DNA sequencing beyond what is possible with standard dye-terminator methods. In ultra-high-throughput sequencing as many as 500,000 sequencing-by-synthesis operations may be run in parallel. High-throughput sequencing applies to genome sequencing, genome resequencing, transcriptome profiling (RNA-Seq), DNA-protein interactions (ChIP-sequencing), and epigenome characterization.

¹https://en.wikipedia.org/wiki/Frederick_Sanger

²https://en.wikipedia.org/wiki/Walter_Gilbert

³https://en.wikipedia.org/wiki/Allan_Maxam

⁴https://en.wikipedia.org/wiki/DNA_sequencer

In this laboratory session we will set up sequencing reactions. We will not perform the actual sequencing ourselves, however. Instead, we will mail the plates with the reactions to a sequencing facility. The facility will notify us when the sequencing data are available for download from their web site. In the next laboratory session (Chapter 14), we will use bioinformatics software to analyze the data.

13.1 Experimental Procedures

Each team will be assigned a group of wells on a 96-well plate for setting up the sequencing reactions. The rows and columns of the 96-well plate are by letters and numbers. The top left well is designated A1 and the bottom right well is H12.

1. Choose two of your plasmid minipreps to sequence.
2. Plan your experiment. You will prepare four sequencing samples for each of your two plasmid minipreps. You will combine each plasmid miniprep with each of the four different sequencing primers — two forward primers and two reverse primers — to ensure complete coverage of the insert. pJET SEQ F and pJET SEQ R anneal to the pJET1.2 vector, while GAP SEQ F and GAP SEQ R are degenerate primers homologous to internal GACG sequences. Enter in Table 13.1 below which plasmid was combined with which primer and added into each well. When you name your sequences (sequence wells), make sure you use the same names when submitting samples to the sequencing facility.
3. In your microcentrifuge tubes, combine 10 μ l of the miniprep DNA with 1 μ l of sequencing primer. Pipet up and down to mix.
4. Pipet 10 μ l of the plasmid/primer mixtures into the assigned wells of the 96-well plate. Write down the barcode number from the 96-well plate.
5. Once the entire class has added samples to the plate, seal the plate using the sealing film provided. Ensure a secure seal by rubbing extensively over the top of the plate with a gloved finger. It is essential to seal the plate completely so that the samples are not lost during transit.
6. Go to www.operon.com/bio-rad and fill out the sequencing request form.
7. The laboratory technicians will mail the plate to the sequencing facility.

Table 13.1: DNA Sequencing Reactions

Well Identifier	Plasmid Name	Sequencing Primers and Color

13.2 Review Questions

1. What is DNA sequencing?
2. What is Sanger sequencing?
3. What are dNTPs?

Chapter 14

DNA sequence analysis

In this laboratory session, we will analyze the DNA sequence data from our cloned GAPDH genes generated by the DNA sequencing facility. We will learn how to use various offline and online bioinformatics software tools. Bioinformatics¹ is an interdisciplinary field that develops methods and software tools for understanding biological data. As an interdisciplinary field of science, bioinformatics combines biology, computer science, mathematics and statistics to analyze and interpret biological data. Bioinformatics has been used for *in silico* analyses of biological queries using mathematical and statistical techniques.

In the last laboratory session (Chapter 13), we set up the DNA sequencing reactions that the laboratory technicians mailed to the sequencing facility. There, DNA sequencing was performed using the chain termination method which was invented by Frederick Sanger and colleagues in the 1970s. This sequencing method is commonly referred to as Sanger sequencing² (Figure 14.1). Frederick Sanger received a Nobel Prize for his invention.

The chain-termination method requires a single-stranded DNA template, a DNA primer, a DNA polymerase, normal deoxynucleosidetriphosphates (dNTPs), and modified di-deoxynucleotidetriphosphates (ddNTPs). These chain-terminating nucleotides lack a 3'-OH group required for the formation of a phosphodiester bond between two nucleotides. Therefore, the DNA polymerase ceases extension of DNA whenever a modified ddNTP is incorporated. The ddNTPs thus **terminate** the elongation of the newly copied DNA strand. Originally, the ddNTPs were radioactively labelled and four separate sequencing reactions, one for each of the four ddNTPs, had to be set up. Following several rounds of template DNA extension, the resulting DNA fragments were heat denatured and separated by size using gel electrophoresis. Each of the four sequencing reactions was loaded in separate lanes (lanes A, T, G, C). The DNA bands were then visualized by autoradiography. The DNA sequence could be read off the X-ray film by reading from the shortest (at the bottom of the gel) to the longest (at the top of the gel) fragment across the four lanes of the gel.

Since its invention, Sanger sequencing has been improved by several modifications. For example, in cycle sequencing a thermostable DNA polymerase is used. This polymerase can be heated to 95 °C and still retain activity and the sequencing reaction can be repeated multiple times in the same tube by heating the mixture to denature the DNA and then allowing it to cool to anneal the primers and polymerize new strands (similar to PCR). Thus, less DNA is needed than for conventional sequencing reactions. Moreover, today, fluorescently labeled ddNTPs are used. Each ddNTP is labelled with a different fluorescent dye (emitting light at different wavelengths; e.g. blue, green, yellow and red). The labelled DNA fragments are separated using capillary electrophoresis. A laser is used to excite the fluorescence of the labeled ddNTPs and a video camera records the color signal which is digitized and stored on a computer as a digital chromatogram. The chromatogram is analyzed using computer software which generates a graph (also referred to as a trace) of the intensity of each color against electrophoresis running time resulting in overlapping peaks and troughs

¹<https://en.wikipedia.org/wiki/Bioinformatics>

²https://en.wikipedia.org/wiki/Sanger_sequencing

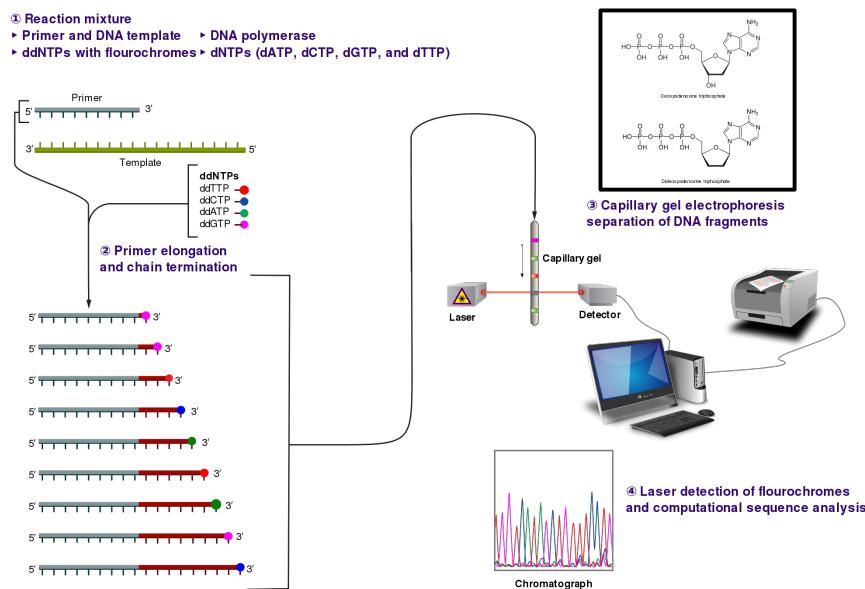


Figure 14.1: The Sanger (chain-termination) method for DNA sequencing. (1) A primer is annealed to a sequence, (2) Reagents are added to the primer and template, including: DNA polymerase, dNTPs, and a small amount of all four dideoxynucleotides (ddNTPs) labeled with fluorophores. During primer elongation, the random insertion of a ddNTP instead of a dNTP terminates synthesis of the chain because DNA polymerase cannot react with the missing hydroxyl. This produces all possible lengths of chains. (3) The products are separated on a single lane capillary gel, where the resulting bands are read by a imaging system. (4) This produces several hundred thousand nucleotides a day, data which require storage and subsequent computational analysis. By Estevezj [CC BY-SA 3.0](<https://creativecommons.org/licenses/by-sa/3.0/>), from [Wikimedia Commons.](<https://commons.wikimedia.org/wiki/File:Sanger-sequencing.svg>)



Figure 14.2: DNA sequence chromatogram and base calls viewed with SnapGene software.

of different color. The peaks are used to assign a corresponding sequence letter (A, T, C, G or N in case the software cannot unequivocally decide which base to call). The plotted peak intensities and associated base calls are saved in a sequencing data file with the extension “.ab1”. Part of a sequencing trace with associated base calls is shown in Fig. 14.2. A base is considered to be of high quality when its identity is unambiguous. A high-quality region of sequence has evenly spaced peaks that do only overlap at their base and has signal intensity in the proper range for the detection software. In today’s laboratory session, we will analyze the data files containing the results of our sequencing project.

14.1 Experimental procedures

1. Open a web browser on your computer. If you are reading this in your browser, right click on the highlighted link to open the SnapGene³ web site in a new tab. Otherwise enter the link in your browser manually.
2. Download SnapGene by clicking on the button corresponding to the operating system of your computer.
3. Once your download has finished, install the program.
4. After the installation has completed, double click on the program icon and start the program.
5. A program window will open. Click on the “Open” icon and navigate to the folder where the sequence files are stored on your computer.
6. Click on the sequence file that ends with “.ab1”. Click open.
7. A new window will open showing the sequence chromatogram trace and associated base calls (Figure 14.2).
8. Locate two buttons with arrows pointing to the right and left in the top right corner of the SnapGene Viewer window.
9. Click on the left button (arrow points to the right). This will bring you to the 5'-end of the sequence.
10. Click on the right button (arrow points to the left) to go to the 3'-end of the sequence.
11. Look for a number next to the right button. It indicates the length of the sequence.
12. Using your trackpad (or computer mouse) scan slowly from the 5' to the 3' end of the sequence and look at the chromatogram trace.
13. Are the peaks sharp and clearly separated from each other with little overlap?
14. Was the software able to call the bases or are there any (or many) “N” labels above the trace?

³http://www.snapgene.com/products/snapgene_viewer/

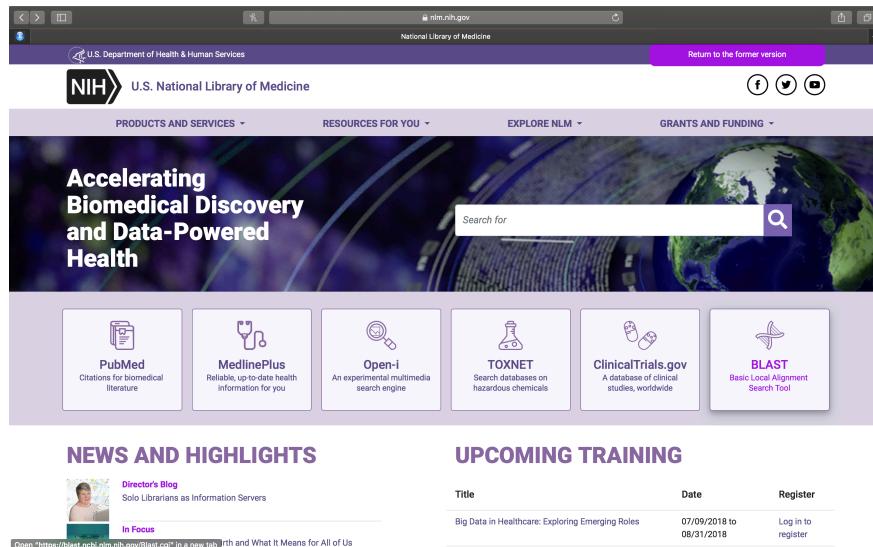


Figure 14.3: The web page of the U.S. National Library of Medicine.

14.2 BLAST search

Now that we have inspected our DNA sequences and ascertained their quality, we will use the BLAST (Basic Local Alignment Tool) program to compare it with all DNA sequences stored in Genbank. GenBank⁴ is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences. GenBank is part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI. These three organizations exchange data on a daily basis.

There are several ways to search and retrieve data from GenBank:

1. Search GenBank for sequence identifiers and annotations with Entrez Nucleotide, which is divided into three divisions:
 - CoreNucleotide (the main collection)
 - dbEST (Expressed Sequence Tags)
 - dbGSS (Genome Survey Sequences).
2. Search and align GenBank sequences to a query sequence using BLAST (Basic Local Alignment Search Tool).

We will use BLAST to search te CoreNucleotide main collection.

14.3 Experimental procedures

1. Open a web browser on your computer. If you are reading this in your browser, right click on the highlighted link to open the web site of the U.S. National Library of Medicine⁵ in a new tab (Fig. 14.3). Otherwise enter the link in your browser manually.
2. On the U.S. National Library of Medicine web page (Fig. 14.3), click on the right most square with the DNA icon and BLAST (Basic Local Alignment Tool) written on it.
3. On the newly opened page (Fig. 14.4), you will see a row of squares displaying various titles. Click on the left most square that has Nucleotide Blast written on it.

⁴<https://www.ncbi.nlm.nih.gov/genbank/>

⁵<https://www.nlm.nih.gov>

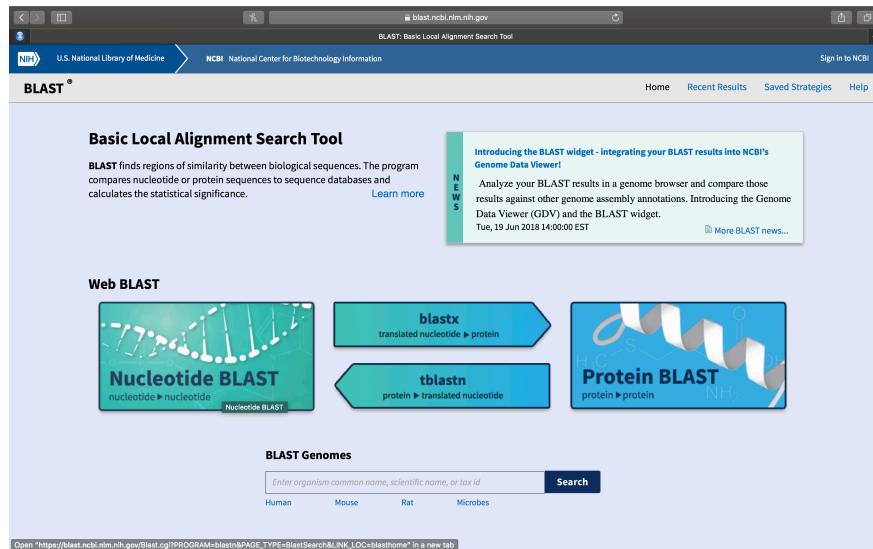


Figure 14.4: The Basic Local Alignment Search Tool (BLAST) start page.

4. A new page will open. Paste your DNA sequence into the white box on the top left (Fig. 14.5).
5. Leave the default values unchanged and click on the oval blue button at the lower left that has BLAST written on it (Fig. 14.6). This will upload your sequence (referred to from now on as the **Query** sequence) to the server where it will be compared to all sequences on record.
6. A new page will open that will be updated every 2 seconds while your query is being processed (Fig. 14.7).
7. Once the search has completed, the BLASTN results page will open (Fig. 14.8).
8. Click on the “+” sign in front of “Graphic Summary” (Fig. 14.8). A graphical summary of your results will be shown (Fig. 14.9. The light green line on top represents your query sequence, below are shown any retrieved sequences that align with your sequence.
9. Click on the “+” sign in front of “Description” (Fig. 14.10). A list of sequences that align with your query sequence will be shown (Fig. 14.10).
10. Click on the first sequence. A new page will show the alignment of the query sequence with the retrieved sequence (Fig. 14.11).
11. Which sequence matches yours? How good is the alignment?

14.4 Review Questions

1. What are the ingredients of a sequencing reaction based on the chain-termination method (Sanger sequencing)?
2. What is GenBank?
3. What is BLAST?

The screenshot shows the 'Standard Nucleotide BLAST' interface. The 'Enter Query Sequence' field contains a DNA sequence: ATAAAGAAGCTCTTTTACGTTAACAGATTTTTCATGAATAAGAAGGGCTAA... (the sequence continues). The 'Program Selection' section has 'Optimize for' set to 'Highly similar sequences (megablast)'.

Figure 14.5: Nucleotide BLAST (BLASTN) sequence entry form.

This screenshot is identical to Figure 14.5, except the 'Enter Query Sequence' field now contains a long DNA sequence: ATAAAGAAGCTCTTTTACGTTAACAGATTTTTCATGAATAAGAAGGGCTAA... (the sequence continues).

Figure 14.6: After you have pasted your sequence into the sequence entry field, click the BLAST button at the lower left of the page.

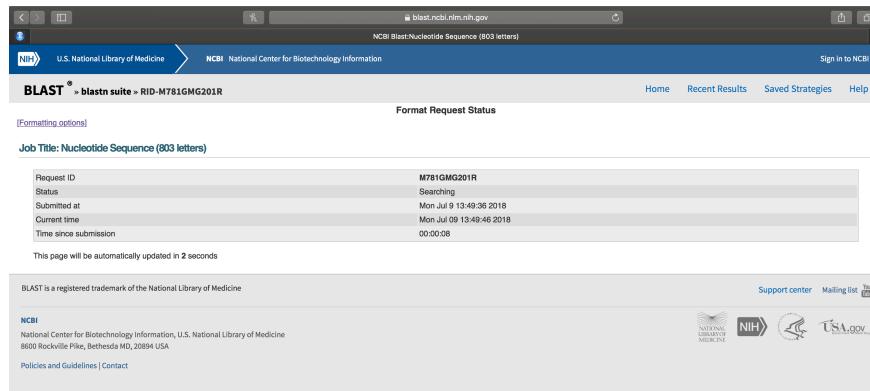


Figure 14.7: The BLASTN query status updated page.

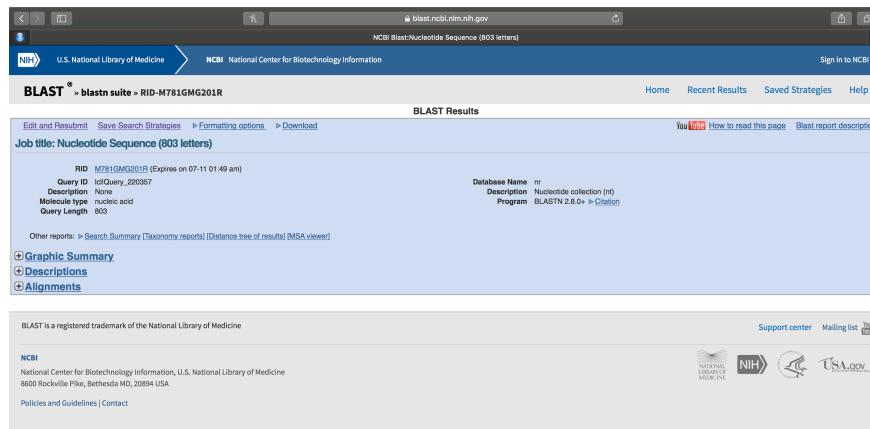


Figure 14.8: The BLASTN results page.



Figure 14.9: Graphic summary of the BLASTN search results.

Sequences producing significant alignments:						
Selected:		Alignments		Descriptions		
		Description	Max score	Total score	Query cover	E value
<input type="checkbox"/>	Equisetum arvense matruse K gene, partial cds; plastid	1456 1456 100% 0 0 99% JX929393.1				
<input type="checkbox"/>	Equisetum arvense plastid, complete genome	1443 1443 100% 0 0 99% JN680380.1				
<input type="checkbox"/>	Equisetum arvense chloroplast, complete genome	1443 1443 100% 0 0 99% GU191334.1				
<input type="checkbox"/>	Equisetum arvense tRNA-Lys (tmK) gene, intron; and MatK (matK) gene, complete cds; chloroplast	1443 1443 100% 0 0 99% AY348561.1				
<input type="checkbox"/>	Equisetum arvense matruse K gene, partial cds; plastid	1421 1421 97% 0 0 99% JX929262.1				
<input type="checkbox"/>	Equisetum arvense plastid partial matK gene for matruse K, isolate 26085	1402 1402 95% 0 0 99% AM883547.1				
<input type="checkbox"/>	Equisetum arvense plastid partial matK gene for matruse K, isolate 26084	1389 1389 95% 0 0 99% AM883548.1				
<input type="checkbox"/>	Equisetum x dyce plastid partial matK gene for matruse K, isolate 26093	1354 1354 95% 0 0 98% AM883545.1				
<input type="checkbox"/>	Equisetum arvense matruse K-like gene, partial sequence; plastid	1345 1345 100% 0 0 97% JX929264.1				
<input type="checkbox"/>	Equisetum palustre plastid partial matK gene for matruse K, isolate 17761	1206 1206 95% 0 0 99% AM883541.1				
<input type="checkbox"/>	Equisetum palustre plastid partial matK gene for matruse K, isolate 11638	1177 1177 93% 0 0 99% AM883538.1				
<input type="checkbox"/>	Equisetum x fontquieri plastid partial matK gene for matruse K, isolate 26093	1140 1140 95% 0 0 93% AM883555.1				
<input type="checkbox"/>	Equisetum sylvaticum plastid partial matK gene for matruse K, isolate 26091	1136 1136 95% 0 0 93% AM883563.1				
<input type="checkbox"/>	Equisetum palustre plastid partial matK gene for matruse K, isolate 11641	1133 1133 90% 0 0 95% AM883539.1				
<input type="checkbox"/>	Equisetum telmateia plastid partial matK gene for matruse K, isolate 11642	1127 1127 94% 0 0 93% AM883540.1				
<input type="checkbox"/>	Equisetum sylvaticum plastid partial matK gene for matruse K, isolate 11640	1107 1107 88% 0 0 94% AM883538.1				
<input type="checkbox"/>	Equisetum scirpoides plastid partial matK gene for matruse K, isolate 26090	985 985 94% 0 0 90% AM883552.1				
<input type="checkbox"/>	Equisetum scirpoides plastid partial matK gene for matruse K, isolate 10593	974 974 93% 0 0 90% AM883534.1				
<input type="checkbox"/>	Equisetum hyemale chloroplast, complete genome	955 955 99% 0 0 88% KQ117177.1				
<input type="checkbox"/>	Equisetum hyemale chloroplast tRNA-Lys (tmK) gene (partial); matK gene and tmK-cobA rGS (partial); isolate F018	955 955 99% 0 0 88% HF585136.1				
<input type="checkbox"/>	Equisetum giganteum plastid partial matK gene for matruse K, isolate 26087	948 948 95% 0 0 89% AM883549.1				
<input type="checkbox"/>	Equisetum bogotense tRNA-Lys (tmK-LysU) gene, partial sequence; and matruse K (matK) gene, complete cds; plastid	933 933 100% 0 0 99% AM883550.1				
<input type="checkbox"/>	Equisetum remesennense tmK gene, partial sequence; and matruse K (matK) gene, complete sequence	933 933 99% 0 0 99% AM883551.1				

Figure 14.10: List of sequences producing significant alignments with your query sequence.

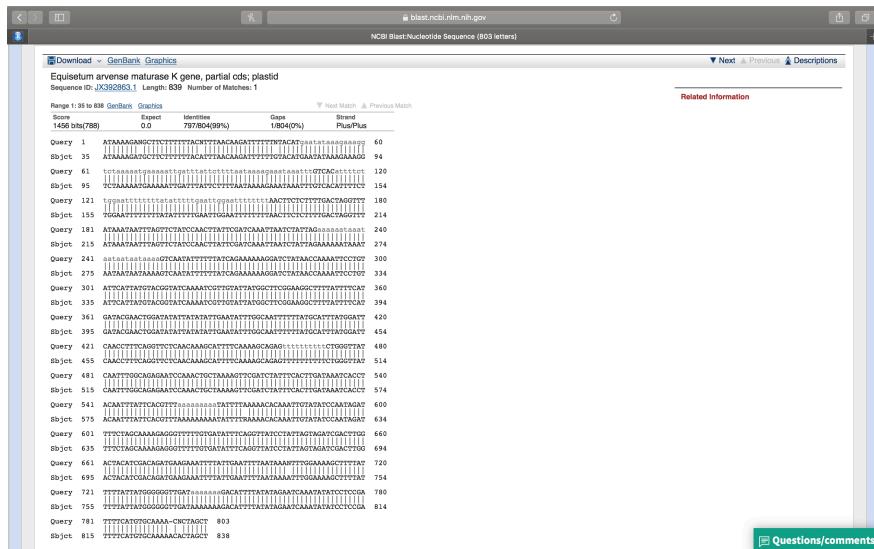


Figure 14.11: Alignment of the best retrieved sequence with the query sequence.

Chapter 15

The Human Genome

In this lab session, we will take a closer look at the human genome and known variations that distinguish different populations and variations that have been linked to increased risk for certain diseases.

The human genome¹ is the complete set of nucleic acid sequences for humans, encoded as DNA within the 23 chromosome pairs in cell nuclei and in a small DNA molecule found within mitochondria. The human genome includes both protein-coding DNA genes and noncoding DNA. The haploid human genome in egg and sperm cells consists of more than three billion DNA base pairs (Table 15.1), while the diploid genome in somatic cells has twice the DNA content. The Human Genome Project (HGP) produced the first (almost) complete sequence of the human genome, with the first draft sequence and initial analysis being published in 2001.

15.1 The Human Genome Project

The Human Genome Project² was an international scientific research project with the goal of determining the sequence of nucleotide base pairs that make up human DNA, and of identifying and mapping all of the genes of the human genome from both a physical and a functional standpoint. After the idea was picked up in 1984 by the US government when the planning started, the project formally launched in 1990 and was declared complete in 2003. Funding came from the US government through the National Institutes of Health (NIH) as well as numerous other groups from around the world. A parallel project was conducted outside government by the Celera Corporation, or Celera Genomics, which was formally launched in 1998. Most of the government-sponsored sequencing was performed in twenty universities and research centers in the United States, the United Kingdom, Japan, France, Germany, Spain and China.

The Human Genome Project originally aimed to map the nucleotides contained in a human haploid reference genome (more than three billion). The project was not able to sequence all the DNA found in human cells. It sequenced only euchromatic regions of the genome, which make up 92% of the human genome. The other regions, called heterochromatic, are found in centromeres and telomeres, and were not sequenced under the project. An initial rough draft of the human genome was available in June 2000 and by February 2001 a working draft had been completed and published followed by the final sequencing mapping of the human genome on April 14, 2003. Although this was reported to cover 99% of the euchromatic human genome with 99.99% accuracy, a major quality assessment of the human genome sequence was published on May 27, 2004 indicating over 92% of sampling exceeded 99.99% accuracy which was within the intended goal. Further analyses and papers on the HGP continue to occur.

¹https://en.wikipedia.org/wiki/Human_genome

²https://en.wikipedia.org/wiki/Human_Genome_Project

The most recent official version of the human genome sequence is the Dec. 2013 (GRCh38/hg38) assembly of the human genome (hg38, GRCh38 Genome Reference Consortium Human Reference 38 (GCA_000001405.15)). The Dec. 2013 human reference sequence (GRCh38) was produced by the Genome Reference Consortium³.

An assembly is a set of chromosomes, unlocalized and unplaced (random) sequences and alternate loci used to represent an organism's genome. Most current assemblies are a haploid representation of an organism's genome, although some loci may be represented more than once. The human genome reference assembly has been obtained from multiple individuals. The haploid assembly does not represent a single haplotype, but rather a mixture of haplotypes. As sequencing technology evolves, it is anticipated that diploid sequences representing an individual's genome will become available.

A haplotype (haplo: from Ancient Greek (haplos, single, simple)) is a contiguous section of closely linked segments of DNA within the larger genome that tend to be inherited together as a unit on a single chromosome. Haplotypes have no defined size and can refer to anything from a few closely linked loci up to an entire chromosome. The term is also used to describe groups of single-nucleotide polymorphisms (SNPs) that are statistically associated.

The term 'haplogroup' refers to the SNP/unique-event polymorphism (UEP) mutations that represent the clade to which a collection of particular human haplotypes belong. (Clade here refers to a set of haplotypes sharing a common ancestor.) A haplogroup is a group of similar haplotypes that share a common ancestor with a single-nucleotide polymorphism mutation. Mitochondrial DNA passes along a maternal lineage that can date back thousands of years. Similarly, the Y-chromosome passes along the paternal lineage.

A chromosome assembly represents a relatively complete pseudo-molecule assembled from smaller sequences (components) that represent a biological chromosome. Relatively complete implies that some gaps may still be present in the assembly (e.g. there are still gaps in the human genome assembly), but independent measures suggest that most of the sequence is represented by sequenced bases. An unlocalized sequence is a sequence found in an assembly that is associated with a specific chromosome but cannot be ordered or oriented on that chromosome. An unplaced sequence is a sequence found in an assembly that is not associated with any chromosome.

Assemblies are built from components, which in turn are joined to form contigs, which are used to build scaffolds (definitions of these and other relevant terms can be found on the GRC web site⁴).

Table 15.1: Human Genome Assembly GRCh38.p12⁵ (Release date: 2017-12-21). Length of DNA sequence of each human chromosome. Total lengths are calculated by summing the length of the sequenced bases and estimated gaps. Ungapped lengths are calculated by summing the length of sequenced bases only. 'Ns' are excluded.

Chromosome	Total length	Ungapped length
1	248,956,422	231,223,641
2	242,193,529	240,863,511
3	198,295,559	198,255,541
4	190,214,555	189,962,376
5	181,538,259	181,358,067
6	170,805,979	170,078,524
7	159,345,973	158,970,135
8	145,138,636	144,768,136
9	138,394,717	122,084,564
10	133,797,422	133,263,006
11	135,086,622	134,634,058
12	133,275,309	133,137,821
13	114,364,328	97,983,128

³<http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/>

⁴<https://www.ncbi.nlm.nih.gov/grc/help/definitions/#ALTERNATE>

Chromosome	Total length	Ungapped length
14	107,043,718	91,660,769
15	101,991,189	85,089,576
16	90,338,345	83,378,703
17	83,257,441	83,481,871
18	80,373,285	80,089,650
19	58,617,616	58,440,758
20	64,444,167	63,944,268
21	46,709,983	40,088,623
22	50,818,468	40,181,019
X	156,040,895	154,893,034
Y	57,227,415	26,452,288
MT	16,569	
Unplaced	4,485,509	4,328,403
Genome	3,099,734,149	2,948,611,470

HGP scientists used white blood cells from the blood of two male and two female donors (randomly selected from 20 of each) each donor yielding a separate DNA library. One of these libraries (RP11) was used considerably more than others, due to quality considerations. More than 70% of the reference genome produced by the public HGP came from RP11, a single anonymous male donor from Buffalo, New York (code name RP11).

The genome was broken into smaller pieces; approximately 150,000 base pairs in length. These pieces were then ligated into a type of vector known as “bacterial artificial chromosomes”, or BACs, which are derived from bacterial chromosomes which have been genetically engineered. The vectors containing the genes can be introduced into bacteria where they are copied by the bacterial DNA replication machinery. Each of these pieces was then sequenced separately as a small “shotgun” project and then assembled. The larger, 150,000 base pairs go together to create chromosomes. This is known as the “hierarchical shotgun” approach, because the genome is first broken into relatively large chunks, which are then mapped to chromosomes before being selected for sequencing.

The “genome” of any given individual is unique; mapping the “human genome” involved sequencing a small number of individuals and then assembling these together to get a complete sequence for each chromosome. Therefore, the finished human genome is a mosaic, not representing any one individual. In fact, we now know that there is variation in the genomes of individual cells in any one human. The implications of this recent finding for human health will keep scientists busy for a long time.

15.2 Public-Private Competition Around The Human Genome Project

In 1998, a privately funded quest to sequence the human genome was launched by the American researcher Craig Venter⁶, and his firm Celera Genomics⁷. Venter was a scientist at the NIH during the early 1990s when the project was initiated. The \$300,000,000 Celera effort was intended to proceed at a faster pace and at a fraction of the cost of the roughly \$3 billion publicly funded project. The Celera approach was able to proceed at a much more rapid rate, and at a lower cost than the public project because it relied upon data made available by the publicly funded project.

Celera used a technique called whole genome shotgun sequencing, employing pairwise end sequencing, which had been used to sequence bacterial genomes of up to six million base pairs in length, but not for anything nearly as large as the three billion base pair human genome.

⁶https://en.wikipedia.org/wiki/Craig_Venter

⁷https://en.wikipedia.org/wiki/Celera_Corporation

Celera initially announced that it would seek patent protection on “only 200-300” genes, but later amended this to seeking “intellectual property protection” on “fully-characterized important structures” amounting to 100-300 targets. The firm eventually filed preliminary (“place-holder”) patent applications on 6,500 whole or partial genes. Celera also promised to publish their findings in accordance with the terms of the 1996 “Bermuda Statement”, by releasing new data annually (the HGP released its new data daily), although, unlike the publicly funded project, they would not permit free redistribution or scientific use of the data. The publicly funded competitors were compelled to release the first draft of the human genome before Celera for this reason. On July 7, 2000, the UCSC Genome Bioinformatics Group released a first working draft on the web. The scientific community downloaded about 500 GB of information from the UCSC genome server in the first 24 hours of free and unrestricted access.

In March 2000, US President Clinton announced that the genome sequence could not be patented, and should be made freely available to all researchers. The statement sent Celera’s stock plummeting and dragged down the biotechnology-heavy Nasdaq. The biotechnology sector lost about \$50 billion in market capitalization in two days.

Although the working draft was announced in June 2000, it was not until February 2001 that Celera and the HGP scientists published details of their drafts. Special issues of Nature (which published the publicly funded project’s scientific paper)[43] and Science (which published Celera’s paper) described the methods used to produce the draft sequence and offered analysis of the sequence. These drafts covered about 83% of the genome (90% of the euchromatic regions with 150,000 gaps and the order and orientation of many segments not yet established). In February 2001, at the time of the joint publications, press releases announced that the project had been completed by both groups. Improved drafts were announced in 2003 and 2005, filling in to approximately 92% of the sequence currently.

15.3 Genome Annotation

The process of identifying the boundaries between genes and other features in a raw DNA sequence is called genome annotation and is in the domain of bioinformatics. Annotation of genes (coding sequences, CDS) is provided by multiple public resources, using different methods, and resulting in information that is similar but not always identical. The Consensus CDS (CCDS) project⁸ is a collaborative effort to identify a core set of protein coding regions that are consistently annotated and of high quality and support convergence towards a standard set of gene annotations. While expert biologists make the best annotators, their work proceeds slowly, and computer programs are increasingly used to meet the high-throughput demands of genome sequencing projects. Beginning in 2008, a new technology known as RNA-seq⁹ was introduced that allowed scientists to directly sequence the messenger RNA in cells. This replaced previous methods of annotation, which relied on inherent properties of the DNA sequence, with direct measurement, which was much more accurate. Today, annotation of the human genome and other genomes relies primarily on deep sequencing of the transcripts in every human tissue using RNA-seq. These experiments have revealed that over 90% of genes contain at least one and usually several alternative splice variants, in which the exons are combined in different ways to produce 2 or more gene products from the same locus.

Table 15.2: Human Genome Assembly GRCh38.p12 (Release date: 2017-12-21). GC content (%) and number of proteins, ribosomal RNA (rRNA), transfer RNA (tRNA), other RNA, genes and pseudogenes on each chromosome. Retrieved on August 10, 2018 at <https://www.ncbi.nlm.nih.gov/genome/?term=txid9606%5Borgn%5D>

Name	GC (%)	Protein	rRNA	tRNA	Other RNA	Gene	Pseudogene
Chr 1	42.3	11,321	17	90	4,457	5,109	1,386
Chr 2	40.3	8,291	-	7	3,728	3,871	1,181
Chr 3	39.7	7,150	-	4	2,782	2,990	900

⁸<https://www.ncbi.nlm.nih.gov/projects/CCDS/CcdsBrowse.cgi>

⁹<https://en.wikipedia.org/wiki/RNA-Seq>

Name	GC (%)	Protein	rRNA	tRNA	Other RNA	Gene	Pseudogene
Chr 4	38.3	4,599	-	1	2,193	2,441	803
Chr 5	39.5	4,729	-	17	2,194	2,592	778
Chr 6	39.6	5,522	-	138	2,453	3,005	882
Chr 7	40.7	5,112	-	22	2,330	2,792	911
Chr 8	40.2	4,199	-	4	2,011	2,165	671
Chr 9	42.3	4,699	-	3	2,222	2,270	706
Chr 10	41.6	5,429	-	3	2,133	2,179	640
Chr 11	41.6	6,394	-	13	2,336	2,924	829
Chr 12	40.8	5,975	-	9	2,457	2,526	691
Chr 13	40.2	2,056	-	4	1,243	1,385	475
Chr 14	42.2	3,501	-	18	1,704	2,065	585
Chr 15	43.4	3,623	-	9	1,810	1,824	554
Chr 16	45.1	4,625	-	27	1,761	1,938	469
Chr 17	45.3	6,226	-	33	2,243	2,450	556
Chr 18	39.8	2,029	-	1	996	984	295
Chr 19	47.9	6,750	-	6	1,877	2,499	523
Chr 20	43.9	2,904	-	-	1,308	1,358	338
Chr 21	42.2	1,297	12	1	707	77	207
Chr 22	47.7	2,582	-	-	1,014	1,189	354
Chr X	39.6	3,801	-	4	1,265	2,186	875
Chr Y	45.4	324	-	-	311	580	392
MT	44.4	13	2	22	-	37	-
Un	44.3	6,143	17	161	3,437	6,543	1,878

Table 15.3: Gene counts based on the Ensembl human genome annotations¹⁰ of the primary assembly of GRCh38.p12 (Genome Reference Consortium Human Build 38).

Coding sequence type	Count
Coding genes	20,376 (incl 612 readthrough)
Non coding genes	22,305
Small non coding genes	5,363
Long non coding genes	14,720 (incl 256 readthrough)
Misc non coding genes	2,222
Pseudogenes	14,692 (incl 7 readthrough)
Gene transcripts	203,903

15.4 International HapMap Project

The International HapMap Project¹¹ was an organization that aimed to develop a haplotype map (HapMap) of the human genome, to describe the common patterns of human genetic variation. HapMap is used to find genetic variants affecting health, disease and responses to drugs and environmental factors. The information produced by the project was made freely available for research.

Four populations were selected for inclusion in the HapMap: 30 adult-and-both-parents Yoruba trios from Ibadan, Nigeria (YRI), 30 trios of Utah residents of northern and western European ancestry (CEU), 44 unrelated Japanese individuals from Tokyo, Japan (JPT) and 45 unrelated Han Chinese individuals from Beijing, China (CHB).

¹¹https://en.wikipedia.org/wiki/International_HapMap_Project

All samples were collected through a community engagement process with appropriate informed consent. The community engagement process was designed to identify and attempt to respond to culturally specific concerns and give participating communities input into the informed consent and sample collection processes.

In phase III, 11 global ancestry groups were assembled: ASW (African ancestry in Southwest USA); CEU (Utah residents with Northern and Western European ancestry from the CEPH collection); CHB (Han Chinese in Beijing, China); CHD (Chinese in Metropolitan Denver, Colorado); GIH (Gujarati Indians in Houston, Texas); JPT (Japanese in Tokyo, Japan); LWK (Luhya in Webuye, Kenya); MEX (Mexican ancestry in Los Angeles, California); MKK (Maasai in Kinyawa, Kenya); TSI (Tuscans in Italy); YRI (Yoruba in Ibadan, Nigeria).

Through this research millions of SNPs were discovered and many GWAS studies used this dataset in research for disease association. This project was a stepping stone for the 1000 genomes project which utilizes many of the same populations.

15.5 The Single Nucleotide Polymorphism Database (dbSNP) of Nucleotide Sequence Variation

The Single Nucleotide Polymorphism database¹² (dbSNP) is a public-domain archive for a broad collection of simple genetic polymorphisms. This collection of polymorphisms includes single-base nucleotide substitutions (also known as single nucleotide polymorphisms or SNPs), small-scale multi-base deletions or insertions (also called deletion insertion polymorphisms or DIPs), and retroposable element insertions and microsatellite repeat variations (also called short tandem repeats or STRs). Please note that in this chapter, you can substitute any class of variation for the term SNP. Each dbSNP entry includes the sequence context of the polymorphism (i.e., the surrounding sequence), the occurrence frequency of the polymorphism (by population or individual), and the experimental method(s), protocols, and conditions used to assay the variation.

The Reference SNP cluster ID (rsid) is an accession number used to refer to specific SNPs in the database.

15.6 The 1000 Genomes Project

The 1000 Genomes Project¹³ launched in January 2008, is an international research effort to establish by far the most detailed catalogue of human genetic variation (Table 15.4). Scientists planned to sequence the genomes of at least one thousand anonymous participants from a number of different ethnic groups within the following three years, using newly developed technologies which were faster and less expensive. In 2015, two papers in the journal Nature reported results and the completion of the project and opportunities for future research. Many rare variations, restricted to closely related groups, were identified, and eight structural-variation classes were analyzed.

The project unites multidisciplinary research teams from institutes around the world, including China, Italy, Japan, Kenya, Nigeria, Peru, the United Kingdom, and the United States. The team have been contributing to the enormous sequence dataset and to a refined human genome map, which are freely accessible through public databases to the scientific community and the general public alike. The International Genome Sample Resource¹⁴ (IGSR) was established at the European Bioinformatics Institute (EMBL-EBI¹⁵) in January 2015. The resource was established with three main aims¹⁶, to:

1. Ensure maximal usefulness and relevance of the existing 1000 Genomes data resources
2. Extend the resource for the existing populations

¹²<https://www.ncbi.nlm.nih.gov/books/NBK21088/>

¹³https://en.wikipedia.org/wiki/1000_Genomes_Project

¹⁴<http://www.internationalgenome.org/home>

¹⁵<https://www.ebi.ac.uk>

¹⁶http://www.internationalgenome.org/sample_collection_principles

3. Expand the resource to new populations

By providing an overview of all human genetic variation, the consortium has generated a valuable tool for all fields of biological science, especially in the disciplines of genetics, medicine, pharmacology, biochemistry, and bioinformatics.

Table 15.4: Populations represented in the 1000 genomes collection.

Superpopulation	Description	Population	Samples
AFR	African Ancestry in Southwest US	ASW	112
AFR	African Caribbean in Barbados	ACB	123
AFR	Esan in Nigeria	ESN	173
AFR	Gambian in Western Division, The Gambia - Fula	GWF	100
AFR	Gambian in Western Division, The Gambia - Mandinka	GWD	280
AFR	Gambian in Western Division, The Gambia - Wolof	GWW	100
AFR	Luhya in Webuye, Kenya	LWK	116
AFR	Mende in Sierra Leone	MSL	128
AFR	Yoruba in Ibadan, Nigeria	YRI	186
AMR	Colombian in Medellin, Colombia	CLM	148
AMR	Mexican Ancestry in Los Angeles, California	MXL	107
AMR	Peruvian in Lima, Peru	PEL	130
AMR	Puerto Rican in Puerto Rico	PUR	150
EAS	Chinese Dai in Xishuangbanna, China	CDX	109
EAS	Han Chinese in Beijing, China	CHB	112
EAS	Han Chinese South	CHS	171
EAS	Japanese in Tokyo, Japan	JPT	105
EAS	Kinh in Ho Chi Minh City, Vietnam	KHV	124
EUR	British in England and Scotland	GBR	107
EUR	Finnish in Finland	FIN	105
EUR	Iberian populations in Spain	IBS	162
EUR	Toscani in Italy	TSI	112
EUR	Utah residents (CEPH) with Northern and Western European ancestry	CEU	183
SAS	Bengali in Bangladesh	BEB	144
SAS	Gujarati Indian in Houston, TX	GIH	113
SAS	Indian Telugu in the UK	ITU	118
SAS	Punjabi in Lahore, Pakistan	PJL	158
SAS	Sri Lankan Tamil in the UK	STU	128

Some basic statistics about the variant sites on the autosomes (chromosomes 1 to 22) and the X chromosome in phase 3 release version 5a from Feb. 20th, 2015 are listed in Table 15.5. The numbering of chromosome locations is based on Genome Reference Consortium Human Build 37 patch release 13¹⁷ (GRCh37.p13).

Table 15.5: The types and numbers of sites with variations in the autosomes and X chromosome in the genomes of 2504 people examined in the 1000 genomes project (phase 3 v5a).

Type	Autosomes	X chromosome
SNPs	78136341	3246232
indels	3135424	227112
others	58671	2040

¹⁷https://www.ebi.ac.uk/ena/data/view/GCA_000001405.14

Type	Autosomes	X chromosome
multiallelic sites	416023	30994
multiallelic SNP sites	259370	1505
Total	81271745	3468087

15.7 Personal genomics

Personal genomics¹⁸ or consumer genetics is the branch of genomics concerned with the sequencing, analysis and interpretation of the genome of an individual. The genotyping stage employs different techniques, including single-nucleotide polymorphism (SNP) analysis chips (typically 0.02% of the genome), or partial or full genome sequencing. Once the genotypes are known, the individual's variations can be compared with the published literature to determine likelihood of trait expression, ancestry inference and disease risk.

Automated high-throughput sequencers have increased the speed and reduced the cost of sequencing, making it possible to offer genetic testing to consumers for less than \$1,000. The emerging market of direct-to-consumer genome sequencing services has brought new questions about both the medical efficacy and the ethical dilemmas associated with widespread knowledge of individual genetic information. Companies like Ancestry¹⁹ and 23andMe²⁰, however, do not sequence your DNA but perform “genotyping” using DNA microarrays (“genotyping chips”) to determine SNPs at hundreds of thousands of locations in your genome. Ancestry, for example state that they examine some 700,000 SNPs.

Starting in 2005 as a pilot experiment with 10 individuals, the Harvard Personal Genome Project²¹ (Harvard PGP) pioneered a new form of genomics research. The main goal of the project is to allow scientists to connect human genetic information (human DNA sequence, gene expression, associated microbial sequence data, etc) with human trait information (medical information, biospecimens and physical traits) and environmental exposures.

PGP participants consent to publicly share their genomic and trait data in a free and open manner to be used for unimpeded research and other scientific, patient care and commercial purposes worldwide. Consistent with this consent, the project organizers seek to lower as many barriers as possible to access PGP data and cells to empower and engage the scientific community to drive new knowledge about human biology. The project now has over 5,000 participants.

15.8 Viewing The Human Genome

15.8.1 Experimental Procedures

1. Open a web browser and go to the Ensemble genome browser²² (Fig. 15.2). This site provides a data set based on the December 2013 Homo sapiens high coverage assembly GRCh38 from the Genome Reference Consortium.

(ref:genomebrowser) The Ensembl genome browser web page²³.

¹⁸https://en.wikipedia.org/wiki/Personal_genomics

¹⁹<https://www.ancestry.com>

²⁰<https://www.23andme.com>

²¹<https://pgp.med.harvard.edu>

²²http://useast.ensembl.org/Homo_sapiens/Info/Index

²³http://useast.ensembl.org/Homo_sapiens/Info/Index

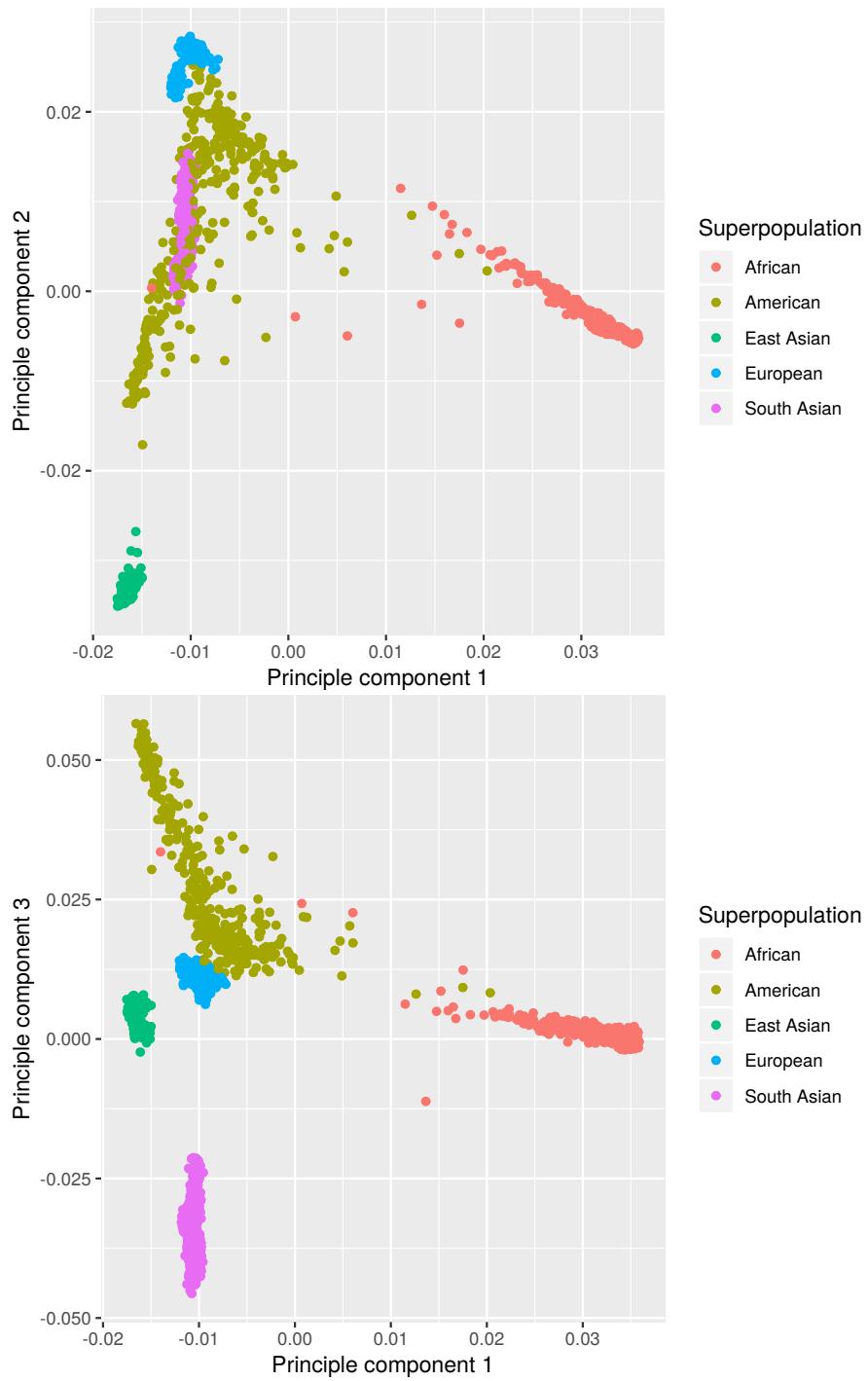


Figure 15.1: Principle component analysis (PCA) of genetic variation in the 1000 genomes project reveals population stratification. PCA is based on 473,964 autosomal variants with a minor allele frequency greater than 10% in 2504 people from 5 superpopulations.

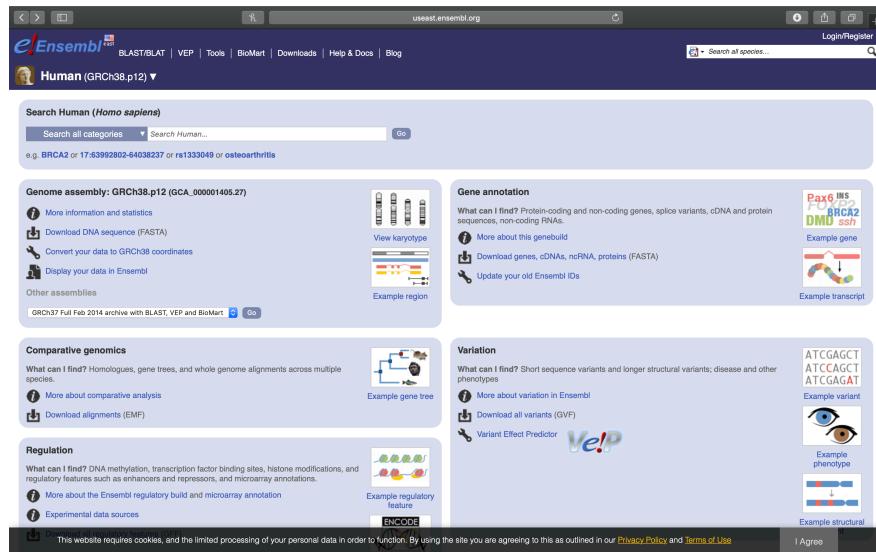


Figure 15.2: (ref:genomebrowser)

15.9 Review Questions

1. What is the length of the human genome (in base pairs)?
2. What is a genome assembly?
3. What is genome annotation?
4. How many genes are in the human genome?