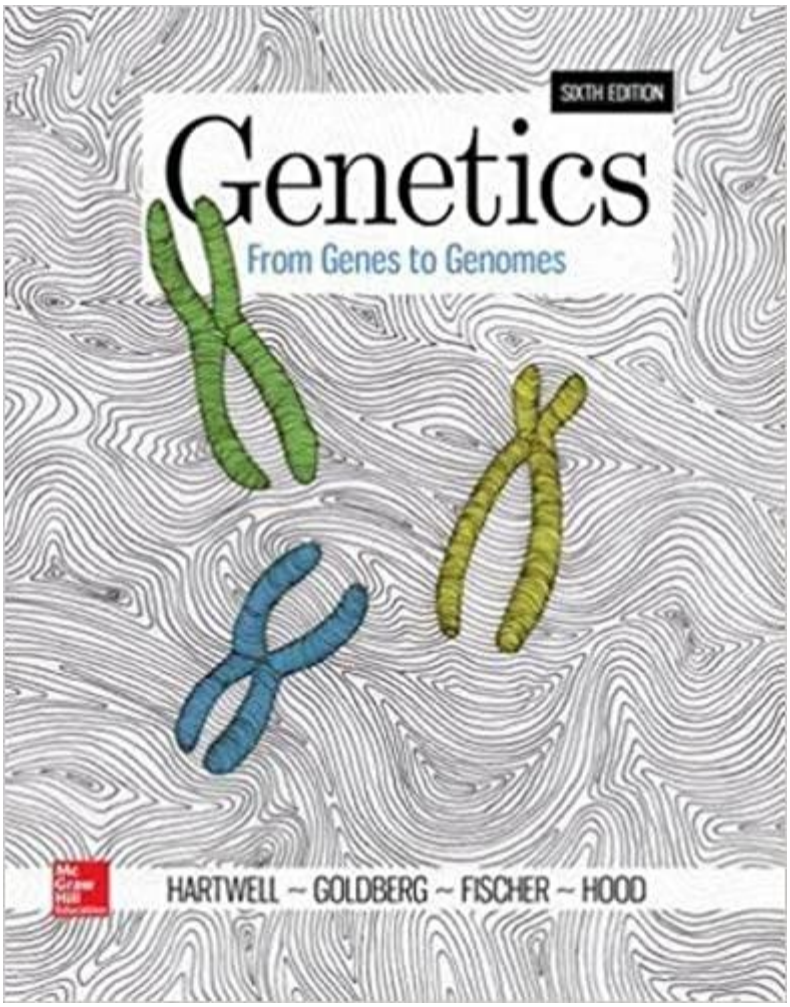
Solution manual for Genetics: From Genes to Genomes 6th Edition by Leland Hartwell



chapter 1

Genetics: The Study of Biological

Information

Synopsis

Chapter 1 is an introduction to the study of modern-day genetics. Genetics is the study of genes: how genes are segments of DNA molecules; how genes are inherited; and how genes direct an organism's characteristics. The most important insight from this chapter is that the basic function of most (but not all) genes is to direct the synthesis of (to **encode**) a particular type of protein.

Key terms

DNA – the macromolecular polymer that constitutes genes.

nucleotides – the chemical building blocks of DNA.

bases – components of nucleotides that are of four different types in DNA; abbreviated as A, G, C, and T.

base pair – DNA is double-stranded; two nucleotide polymers are held together by hydrogen bonds between A-T and G-C base pairs.

genes –segments of DNA that, in most cases, encode proteins.

chromosomes – large DNA molecules that can contain hundreds or thousands of genes.

genome – all the DNA, and thus all the genes, in an organism.

metabolism – the chemical reactions by which organisms use energy and matter to construct their bodies.

genetic code – the way that genes are "read" by the molecular machines that use genes to make proteins.

RNA – a polymer structurally similar to DNA that serves as a chemical intermediate in the pathway from genes to proteins.

proteins – linear polymers of amino acids that fold into complex three-dimensional shapes. Proteins constitute the structures of cells, and carry out the chemical reactions of metabolism.

amino acids – the chemical subunits of proteins. Twenty different common amino acids exist in proteins.

mutation – a heritable chemical change in the base sequence of DNA that enables evolution to occur.

evolution – the change in characteristics of populations of organisms over time due to the accumulation of mutations in genes.

model organisms – species used commonly for genetic analysis by scientists.

gene family – two or more genes with similar DNA sequences and similar functions that most likely arose from a single ancestral gene by a series of duplication and divergence events. A **gene superfamily** is a group of gene families that share a common ancestral gene.

exons and **introns** – the portions of genes that are used to make proteins (*exon*s) and the regions of DNA that separate them (*introns*).

prokaryotic cells – single-cell organisms like bacteria whose genomes are not enclosed within a membrane (not inside a nucleus).

eukaryotic cells – cells such as human cells whose genomes are within a nucleus (a membrane-enclosed organelle).

Human Genome Project – the effort to determine the DNA base sequence of every human chromosome and to analyze the genes making up the human genome.

Problem Solving

The first chapter of this book provides a broad overview of genetics. Chapter 1 covers a lot of ground, but only superficially. Don't worry if at this point you don't understand at a deep level all the information described – you will later on. However, you are likely familiar already (from introductory biology classes) with some of the fundamentals of what a gene is and how genes are used to make proteins. The problems in this chapter are meant to get you started in the habit of thinking like a geneticist – quantitatively, analytically, carefully, and logically.

Vocabulary

1.

a. complementarity hydrogen	4. G-C and A-T base pairing in DNA through bonds
b. nucleotide	11. subunit of the DNA macromolecule
c. chromosomes	7. DNA/protein structures that contain genes
d. protein	1. a linear polymer of amino acids that folds into a particular shape
e. genome	9. the entirety of an organism's hereditary information
f. gene	8. DNA information for a single function, such as production of a protein
g. uracil	12. the only one of the four bases in RNA not in DNA
h. exon	6. part of a gene that can contain protein coding information

i.	intron	2.	part of a gene that does not contain protein coding information
j.	DNA	10.	a double-stranded polymer of nucleotides that stores the inherited blueprint of an organism
k.	RNA	3.	a polymer of nucleotides that is an intermediary in the synthesis of proteins from instructions in DNA
l.	mutation	5.	alteration of DNA sequence

Section 1.1

2. The complementary strand of a DNA molecule is simply the strand with which the original DNA molecule forms base pairs. Remember two things: (1) The two strands of a double-stranded DNA molecule are oriented in the opposite direction with respect to each other (their 5′ and 3′ ends run in opposite directions); and (2) the base pairs are A-T and G-C. Therefore, the DNA strand complementary to the one shown is:

5′ AGCTTAATGCT 3′

3. **a.** If the 3 billion (3,000,000,000) base pairs of the human genome is divided into 23 chromosomes, the average size of a human chromosome is **3,000,000,000 base pairs/23 chromosomes ≈ 130,435,000 base pairs per chromosome.**

b. The human genome contains about 27,000 genes, and assuming they are spread evenly over the 23 chromosomes, on average there are **27,000 genes/23 chromosomes ≈ 1174 genes per chromosome.**

c. About half the DNA of the human genome contains genes, meaning that all the genes are found within 1.5 billion (1,500,000,000) base pairs. Therefore, on average there are **1,500,000,000 base pairs / 27,000 genes ≈ 55,555 base pairs per gene.**

Section 1.2

4. **a.** **Both.** Each protein is composed of a *string* of amino acids, and DNA is a *string* of nucleotides.

b. **DNA.** DNA is double-stranded through complementary base pairing of single strands with opposite orientations. A protein is a single strand of linked amino acids, and the strand folds into a particular shape.

c. **DNA.** Four different kinds of nucleotides – A, G, C, and T – are present in the DNA polymer. Twenty different common amino acids are present in almost all proteins.

d. **Proteins.** Twenty distinct amino acid subunits are the building blocks of almost all proteins. DNA is made up of only four different types of nucleotides.

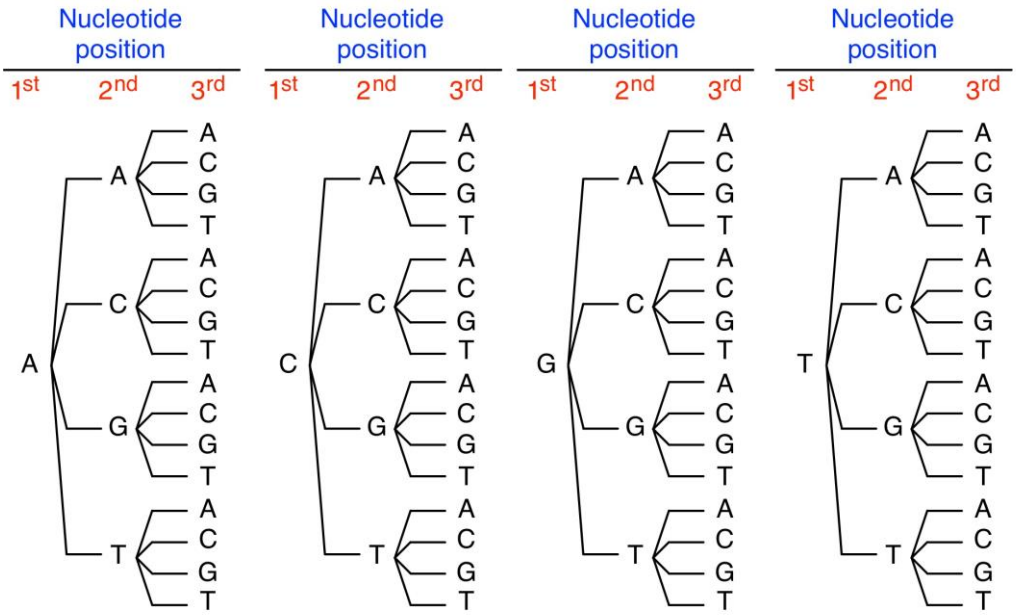
e. **Proteins.** Proteins are polymers of amino acids; DNA molecules are polymers of nucleotides.

f. **DNA.** DNA is a polymer of nucleotides; a protein is a polymer of amino acids.

g. **DNA.** Genes are segments of DNA; by using the genetic code, most genes encode proteins.

h. **Proteins.** Some proteins (*enzymes*) catalyze chemical reactions.

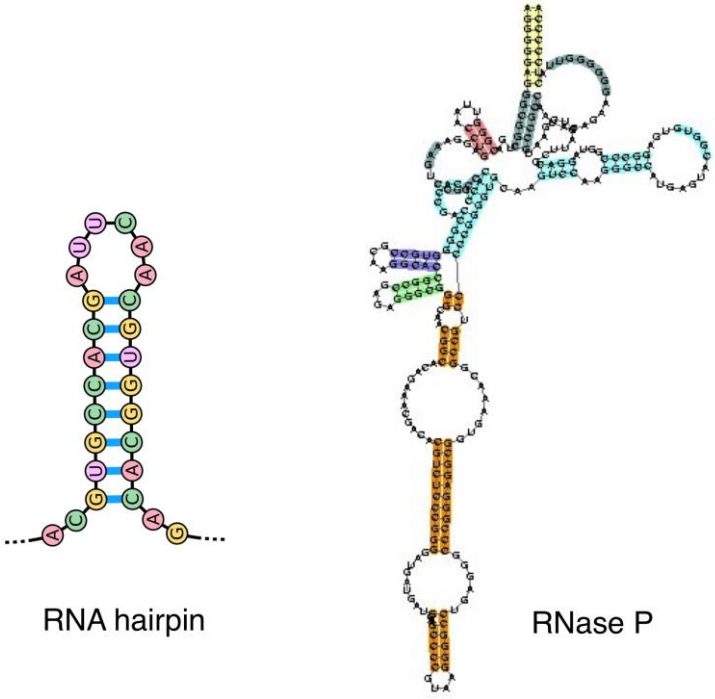
5. **a.** Each base in a single strand of a DNA molecule can be either an A, G, C or T. Therefore, a specific 100-nucleotide DNA strand could start with any one of the four nucleotides, the second nucleotide could be any one of the four nucleotides, etc. The number of different possible sequences increases by a factor of 4 at each successive step in the addition of a base (see the following figure). Thus, **the number of different possible sequences of a 100-nucleotide DNA strand is $4^{100} = \sim 1.6 \times 10^{60}$**. We need not consider the second, complementary strand of DNA, as its base sequence is determined by the sequence of the first strand.



b. Because each amino acid can be 1 of 20 different amino acids, by the same logic as in part (a), **the number of different 100-amino acid proteins is $20^{100} = \sim 1.3 \times 10^{130}$**.

6. **Regardless of the sequence of nucleotide pairs, DNA molecules all have the same three-dimensional structure** of a long double helix as seen from the outside. This unvarying structure does not lend itself to function as an enzyme, because the molecule is rigid and inflexible, and because the nucleotides on either strand are "tied up" by their pairing with the complementary nucleotides on the other strand. In contrast, many RNA molecules can fold back on themselves to form *RNA hairpins* (as seen in the accompanying figure on the *left)*, and some RNA molecules can form

intricate three-dimensional patterns of hairpins that allow these molecules to act as RNA enzymes, or *ribozymes* (such as RNase P in the accompanying figure at *right*).



https://en.wikipedia.org/wiki/Stem-loop

Rfam database, Public Domain, https://commons.wikimedia.org/w/index.php?curid=2078772

7. As shown in Fig. 1.5, three nucleotide pairs of DNA are required to specify one amino acid. The figure also shows that each nucleotide in the gene can be part of only a single triplet corresponding to a single amino acid. As a result, the smallest possible size of the parts of the human gene that specify lactate dehydrogenase using the genetic code would be 332 × 3 = **996 base pairs.** In fact, this gene is much larger because some of its nucleotides are found in *introns* that do not specify amino acids.

8. **a.** The genetic code table has U nucleotides but no T nucleotides; therefore, **the triplets are written in the form of RNA.**

b. The RNA triplet **AUG specifies Met** (the amino acid methionine), while the RNA triplet **UGG specifies Trp** (tryptophan). You should note that most of the 20 amino acids in the genetic code table correspond to more than one triplet. For example, four different triplets (CCU, CCC, CCA, and CCG) specify Pro (proline).

c. **The amino acid sequence of a protein allows you to infer many of the nucleotides in the RNA sequence. Some nucleotides will be ambiguous because most amino acids can be specified by more than one codon. The RNA nucleotides are copied from a gene according to base complementarity** (A in DNA pairs with U in RNA, C pairs with G, and T in DNA pairs with A in

RNA). **Thus, if we know some of the nucleotide sequence of the RNA, we can infer those unambiguous nucleotide sequences in the gene (the DNA).**

Section 1.3

9. Scientists think that all forms of life on earth have a common origin because **organisms as distant as humans and bacteria share the same genetic code, and many of their proteins are similar in amino acid sequence and biochemical function.**

10. Scientists study model organisms like yeast and fruit flies to understand universal biochemical pathways. Because of their common origin and because they have similar genes and proteins, all organisms share certain universal pathways. For example, many of the genes that help regulate cell division are similar in yeast and humans. Obviously, **scientists cannot perform experiments on humans, but researchers can manipulate organisms like yeast, fruit flies, and mice in the laboratory in many useful ways. Universal principles of biology may be learned from these model organisms because of the common origin of all life.**

11. **To detect proteins that have a common origin in different organisms, scientists use computer analysis of the DNA sequences of genomes to look for genes that encode proteins with large stretches of amino acids that are identical or similar. To assess whether related genes in different organisms have similar functions, scientists can generate mutations in the genes and see if the mutations have similar effects.** For example, suppose bacteria with a mutation in a particular gene are unable to grow because the cells cannot divide. If fruit flies with a mutation in a gene with related DNA sequences that encode a similar protein die as very young embryos with very few cells, you could conclude that the genes in each organism have a key function in cell division.

In some cases, you could go one step further by placing the normal fruit fly gene into the genome of the mutant bacterial cells (or the normal bacterial gene into the genome of the mutant fruit flies). If the mutant organisms with the gene from the other species could grow properly, you could then conclude that the genes from the different organisms do in fact encode proteins that fulfill the same biochemical role in cell division. Because bacteria and fruit flies are so distantly related to each other, this type of *gene rescue* experiment is successful only rarely. But for more closely related species (like fruit flies and yeast cells, both of which are eukaryotic organisms), such experiments have often demonstrated that genes from different species that have related DNA sequences also have similar gene function.

12. **The amino acids highlighted in *dark orange*** are exactly the same in all of these species and are likely to be the most important for cytochrome c protein function. The process of mutation is random and occurs all the time. The fact that these amino acids have been invariant over more than a billion years of evolution suggests that mutations in cytochrome c genes that alter the identity of these particular amino acids have been weeded out by natural selection. The reason is that cytochrome c is

essential for viability, while the protein products of the mutant genes would be either nonfunctional or have decreased ability to perform the important molecular role carried out by cytochrome c.

Section 1.4

13. Scientists think that new genes arise by duplication of an original gene and divergence by mutation because **the genomes of all organisms have gene families and superfamilies.** These gene families and superfamilies contain genes that encode proteins with similar amino acid sequences. The proteins in these families fold into similar three-dimensional structures and they perform related, though not necessarily identical, functions. The genomes of more complex organisms usually contain more members of the same gene/protein families that exist in the genomes of simpler organisms. **It is unlikely that all these gene/protein families arose anew in each organism.**

14. Genes have *exons* that include protein coding regions, and regions of DNA between the exons called *introns.* **Exons from different genes could be *shuffled* by chromosome rearrangements. Modules from different proteins could thus reassort to form new proteins with new functions.**

15. **A protein is likely to perform the same type of biochemical reaction in different cell types, but it will not necessarily interact with the same molecules in all kinds of cells.** For example, if a protein is a *kinase* (a kind of enzyme that adds a phosphate group to other molecules called *substrates*) it would probably be a kinase in all cells. However, the kinase might add a phosphate group to one substrate in one cell type but a different substrate in other kinds of cells. It is also possible that the same substrate could participate in different pathways in different cell types. Therefore, **a protein with a particular biochemical activity could function in the same or in different pathways in various cell types.**

Section 1.5

16. **a.** **Untrue**; the zebrafish that lacks a functional version of the gene is viable.

b. **True;** the zebrafish that lacks a functional version of the gene lacks stripes.

c. **Insufficient information;** no information is given as to why the stripes are absent in the mutant zebrafish and many explanations for this observation are possible. For example, the mutant zebrafish might be able to make the pigment but not be able to import the pigment into the subcellular structures within the striped areas where the pigment normally accumulates.

d. **Insufficient information;** the gene is not required for viability because the fish lacking a functional version of it are alive. However, no information is given about possible abnormalities in the mutant zebrafish other than a failure to form horizontal stripes. Perhaps these animals are more susceptible to certain diseases,

or their learning is impaired, or they may be subject other possible conditions that scientists apparently have not yet tested.

17. a. **The DNA sequence of the *WDR62* gene would have enabled scientists to predict the amino acid sequence of the protein it encodes. Conserved regions of amino acid sequence often reveal structural features indicative of the biochemical function of the protein.** In fact, *WDR62* is so named because the protein it encodes contains *WD repeats:* regions with similar amino acid sequences that are found in several proteins. These WD repeats allow the proteins that contain them to bind to other proteins.

b. **Knowing that the *WDR62* mutations cause microcephaly indicates that at the level of the organism, the gene and the protein it encodes are required for proper brain development.**

c. **If the mutant mice had a syndrome resembling that of people with microcephaly, then we would know for sure that *WDR62* is the microcephaly disease gene. These mice could also be used in various experiments to study the biochemical pathways in which the WDR62 protein participates,** as these pathways are likely to be similar in mice and humans and would be needed for proper brain development in both species.

18. a. ***RPE65* is needed for retinal function, but not for retinal development.** Note that the blind patients lacking functional RPE65 protein still had recognizable retinas into whose cells the normal gene was injected. These retinas must have had the potential to work properly, because patients with the injected gene were no longer blind.

b. Diseases like microcephaly are caused by the loss of gene functions required for development of an organ. **Providing a missing gene product after an organ has already developed improperly because of its absence likely will have no therapeutic effect.** The only exceptions would be rare cases in which gene therapy could somehow revamp the entire structure of a malformed organ.

Section 1.6

19. Different people may have very different perspectives about their interest in obtaining the DNA sequence of their genome. **Genome sequences may be helpful in identifying and treating diseases, in making reproductive decisions, and in providing clues about ancestry. However, presently, scientists can interpret only a small fraction of the information in a genome sequence, because many traits are influenced in very complicated ways by large networks of genes. Furthermore, some individuals may have excellent reasons for NOT wanting to learn about their genetic predispositions to certain traits.** For example, people whose parents have Huntington disease, a neurodegenerative condition that tends to affect people late in life, can learn for certain if they will develop the disease by analysis of the base sequence of a single gene in their genome. Some people may wish not to know they will eventually develop this disease because that knowledge

may affect their current quality of life. Your own perspectives about this issue may well change as your understanding of genetics increases.