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 of the DNA, and thus all the genes, in a particular 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 also 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 take place

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

convergent evolution – the evolution of similar structures independently in the lineages leading to different species

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 **multigene family** is a large gene family; a **gene superfamily** is a group of gene families and multigene 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 all of the information given at a deep level – 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	4.	G-C and A-T base pairing in DNA through hydrogen 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 a protein
g.	uracil	12.	the one of the four bases in RNA that is not in DNA
h.	exon	6.	part of a gene that contains 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 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 ≈ 135,435,000 base pairs per chromosome.**

b. The human genome contains about 25,000 genes, and assuming that they are spread evenly over the 23 chromosomes, on average there are **25,000 genes/23 chromosomes ≈ 1087 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 / 25,000 genes ≈ 60,000 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 in 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. **Protein.** Twenty distinct amino acid subunits are the building blocks of almost all proteins. DNA is made up of only four different types of nucleotides.

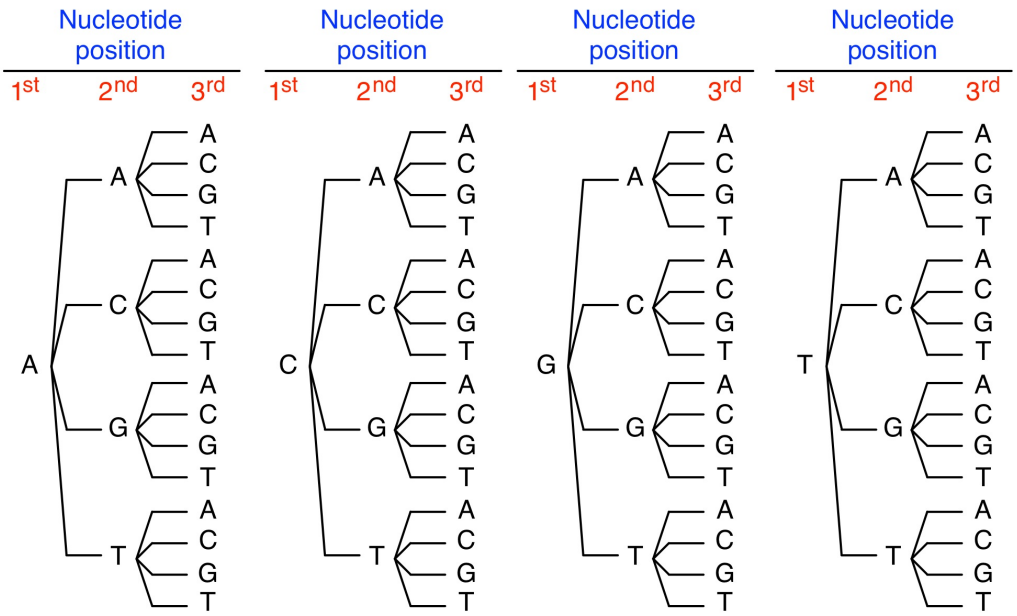
e. **Protein.** Proteins are polymers of amino acids; DNA is a polymer of nucleotides.

f. **DNA.** DNA is a polymer of nucleotides; proteins are polymers of amino acids.

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

h. **Protein.** Some proteins (*enzymes*) perform 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 $20^{100} = \sim 1.3 \times 10^{130}$**.

Section 1.3

6. 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.**

7. Scientists study model organisms like yeast and fruit flies in order 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.**

8. **To detect proteins in different organisms that have a common origin, 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 were able to 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 only rarely successful. 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.

Section 1.4

9. 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 with similar amino acid sequences; the proteins in these families fold into similar three-dimensional structures and they perform related 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 of these gene/protein families arose anew in each organism.

10. Genes have *exons* that include protein coding regions, and also 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.**

11. **A protein is likely to perform the same type of biochemical reaction in different cell types.** 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. Therefore, **a protein with a particular biochemical activity could function in the same or in different pathways in various cell types.**

Section 1.5

12. **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.

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.

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

c. **If the mutant mice had a syndrome similar to 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.

Section 1.6

14. Different people may have very different perspectives about their interest in obtaining the DNA sequence of their genome. Genome sequences may be helpful in treating diseases, in making reproductive decisions, and in providing clues about ancestry. At the present time, only a small fraction of the information in genome sequences can be interpreted by scientists because many traits are influenced in very complicated ways by large networks of genes. In some cases, individuals may have excellent reasons for NOT wanting to learn about their genetic predispositions to certain traits. For example, many people whose parents have Huntington disease, a neurodegenerative condition that tends to affect people late in life, can know for certain whether or not they will develop the disease by analysis of the base sequence of a single gene. 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.