

## Chapter 2

# Biological Background

This chapter is a brief introduction, for mathematicians, to genes, cells, and cancer. It provides general descriptions of the biological phenomena that are the subject of the mathematical applications developed in later chapters. More specific information relevant to each application is given at the beginning of the section containing the application. No knowledge of biology or chemistry is assumed beyond that learned in high school and forgotten due to disuse. Many biological details are omitted for lucidity. Readers familiar with the biological topics in this introduction may proceed directly to the later chapters.

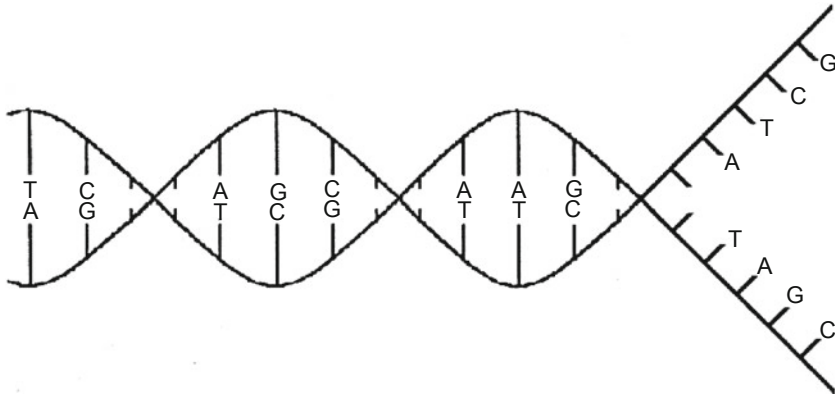
### 2.1 Genomes: Changes in DNA and Chromosomes

#### 2.1.1 *Genome*

The term “genome” refers to the molecules that function in the storage, expression, and inheritance of information in biological systems. The genome of humans and other organisms is dynamic. The number and sequence of its subunits can undergo rapid changes within a few generations.

#### 2.1.2 *DNA and Genes*

Deoxyribonucleic acid (DNA) is the chemical that is the primary genetic material in the genome of all cells. It is responsible for the storage and inheritance of genetic information. DNA is a polymer consisting of two long complementary strands (Fig. 2.1). Each strand contains a linear sequence of four monomer subunits called bases. The bases are abbreviated *A*, *T*, *G*, and *C*. Each *A* base on one strand pairs with a *T* base on the other strand, and each *G* base pairs with a *C* base. The total length of DNA in the genome of each mammalian cell is about  $3 \times 10^9$  bases. Genes are specific subsequences of DNA that code for proteins. A given gene is typically  $10^3$ – $10^4$



**Fig. 2.1** DNA structure. Pairs of complementary bases (*A* and *T*, *G* and *C*) hold together the double-stranded helix of DNA. The two strands are separated on the right, for replication, described further in Fig. 2.3. The sequence of bases in DNA is transcribed into a sequence of bases in RNA and translated into a sequence of bases in protein which functions to determine the observable traits of the organism. Mutational changes in the sequence of bases in DNA result in changes in observable traits which are inherited

bases long and occurs one time or only a few times in the genome. The mammalian genome contains approximately 21,000 different genes that code for proteins. The expression of the genetic information in DNA is accomplished by transcribing a sequence of bases in DNA into a sequence of bases in a related molecule, ribonucleic acid (RNA). The sequence of bases in RNA is then translated into a sequence of amino acids, the subunits of proteins. The proteins carry out catalytic and structural roles which result in the biological properties of cells and organs. So, the flow of information is usually as follows:

DNA → RNA → protein → phenotype,

in words :

gene → message → catalyst → observable trait.

### 2.1.3 *Mutation*

An alteration in the sequence of bases in DNA is referred to as a mutation. The mutation may be as small as a change in a single base or as large as a rearrangement of most of the DNA in a chromosome. A mutation in DNA may result in an altered sequence of amino acids in protein and/or an altered amount of protein. This may result in a change in the ability of a protein to function properly, resulting in altered properties of cells and organisms. Many mutations are deleterious, but others may

be advantageous or neutral. Examples of altered properties of cells containing mutations include misregulation of cell growth and division leading to malignant tumors, and new capability of mutant cells to grow in the presence of a drug that would kill normal cells. A multitype process model describing mutations as occurring in several possible steps, and an improved method of estimating mutation rates from experimental observations are given in Sect. 6.1.

### **2.1.4 Noncoding Sequences of DNA**

Genes account for a small portion of the genome of mammals. Only about 5–10% of the DNA codes for proteins, the remainder is referred to as the noncoding portion of DNA. The function of the noncoding DNA is only partially understood. Some noncoding regions specify the sequence of bases in RNA that is never translated into protein. Another small portion consists of special DNA sequences that are required to maintain the ends of DNA called telomeres. Maintenance and loss of telomere sequences are discussed as a Bellman–Harris process with denumerable type space in Sect. 7.7. Yet other noncoding sequences, centromeres, are required for the accurate segregation of the DNA into progeny cells. Fragments of DNA that do not contain centromeres distribute into progeny cells in uneven numbers. This is modeled as a single-type Galton–Watson branching process in Sect. 3.7 and as random walk with absorbing boundary in Sect. 7.4.

### **2.1.5 Repeated Sequences of DNA**

Much of the noncoding mammalian DNA consists of sequences which are repeated many times in the genome, most of unknown function. Some repeated sequences are tandemly distributed, others are dispersed throughout the genome. The length (number of bases) of a tandemly repeated unit may be as short as one base or longer than  $10^3$  bases. The number of repeated units may be as small as two or larger than  $10^2$ . An increase in the number of tandemly repeated units is referred to as amplification, and decrease as deamplification.

The emergence of periodicities of tandemly repeated sequences in DNA by recombination slippage, simulated by a discrete stochastic dynamical system, was discussed by Baggerly and Kimmel (1995). Repeats may also arise by other mechanisms as discussed in Sect. 3.8. See also Bat et al. (1997).

For example, *Alu* elements are DNA sequences of about 300 base pairs. They do not code for proteins. They appear only in mammals. The number of the *Alu*-repeated elements per genome has increased during the evolution of primates. The human genome contains more than one million *Alu* repeat elements dispersed throughout the genome, some with slightly different sequences. Amplification of *Alu*

repeat elements is modeled as a discrete-time Griffith and Pakes branching process, Sect. 7.10.1.

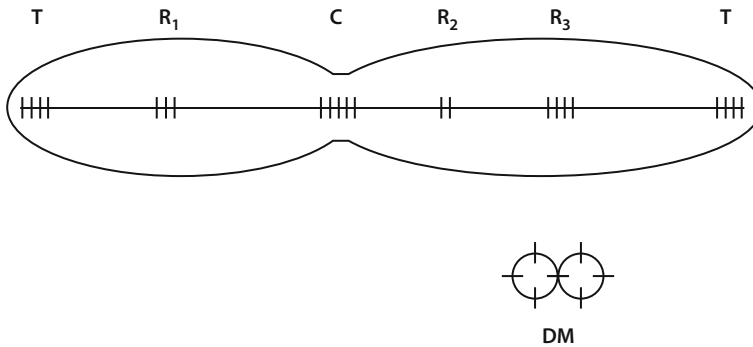
### **2.1.6 Gene Amplification**

Regions of DNA may undergo an increase in number (amplification) or decrease in number (deamplification). The regions of DNA that undergo amplification and deamplification may contain genes, or contain no genes. Amplification and deamplification of regions of DNA-containing genes can result in increases or decreases of amounts of proteins necessary for cell functions. Overproduction of rate-limiting proteins may confer new properties on cells. For example, if the protein is involved in cell proliferation, the cells with an increased amount of this protein may grow as malignant tumors. As another example, if the protein is the target of a toxic drug, then an increased amount of the protein may allow the cells to be resistant and grow in the presence of the drug. Models for gene amplification resulting in tumor cell growth and drug resistance are the subjects of Sects. 3.7, 7.4 and 7.5.

Some inherited human syndromes, such as predisposition to some cancers and neurological diseases, have been related to a rapid change in the numbers of copies of DNA sequences. An unusual aspect of these is the apparently explosive increase in numbers of copies of some sequences from one generation to another. This increase has been modeled as an iterated Galton–Watson process in Sect. 3.8. In contrast to these cases of concerted increases in gene copy numbers, there are situations in which the number of amplified genes is unstable and decreases. Unstable decreases in numbers of amplified genes are modeled as a subcritical Galton–Watson process in Sect. 3.7, and as a branching random walk with absorbing barrier in Sect. 7.4.

### **2.1.7 Chromosomes**

In human cells, the DNA of the genome is divided into 23 pieces of various lengths, each containing large numbers of different genes. Each piece of DNA is folded compactly and associated with proteins and RNA to form a chromosome (Fig. 2.2). In human cells, there are 23 pairs of chromosomes. Each chromosome contains one double-stranded piece of DNA from end to end. The ends of chromosomes are called telomeres. They have special sequences and structures that are necessary for the replication of the end of DNA and the maintenance of chromosomes. DNA in chromosomes replicates and then the products separate from each other in a process called mitosis. Special DNA sequences near the center of chromosomes (centromeres) function to segregate one of each pair of duplicated chromosomes into each of two progeny cells during cell division. This process assures that each progeny cell receives one copy of each chromosome and its associated DNA-containing genes. Fragments of DNA without centromeres may increase in number (replicate) to more



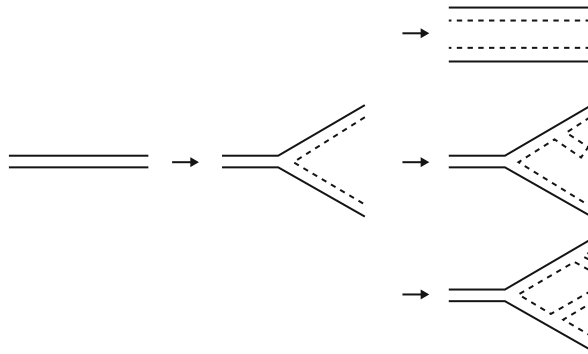
**Fig. 2.2** Chromosome. One double-stranded piece of DNA, represented here by a single horizontal line, extends from one end of a chromosome to the other. Several classes of repeated sequences of bases are represented. Telomere (*T*) repeats at each end of the chromosome function to maintain chromosome ends. Centromere (*C*) repeats function to separate chromosomes at mitosis and cell division. Other repeated (*R*) sequences of bases are dispersed throughout the chromosome. Some function to code for proteins (e.g., genes), others are noncoding sequences. Repeated sequences may exist as extrachromosomal elements, also called double minute (*DM*) chromosomes because of their appearance. They may replicate but are not partitioned evenly to progeny cells because they lack the centromeres of chromosomes. The number of repeated units may be variable

than two copies per cell but without centromeres there is no mechanism to distribute exactly equal numbers to each progeny cell.

### 2.1.8 DNA Replication

DNA replication occurs by a so-called semiconservative mechanism. Two complementary parental strands of DNA separate and each strand forms a template for the production of a new complementary progeny strand. Usually, replication is initiated by the local separation of two strands, the replication fork, and then proceeds along the DNA. The result is two double-stranded DNA molecules, each molecule containing one old strand and one complementary new strand (Fig. 2.3).

Two types of errors in DNA replication have been proposed to result in amplification of repeated DNA sequences, replication slippage, and replication reinitiation. Replication slippage may occur when repeated DNA sequences on one strand fold back on themselves forming a hairpin-like structure. This may cause slippage of the replication complex along one strand relative to the other and resulting in stuttering and repeated replication of a portion of the DNA sequence. The generation of unstable numbers of DNA repeats by replication slippage may contribute to the explosive increase in numbers of repeated sequences in certain cancers and inherited neurological diseases. A mathematical model describing amplification of repeats by replication slippage has been developed (Bat et al. 1997) but is not described as an application here because it is not a branching process. Replication reinitiation

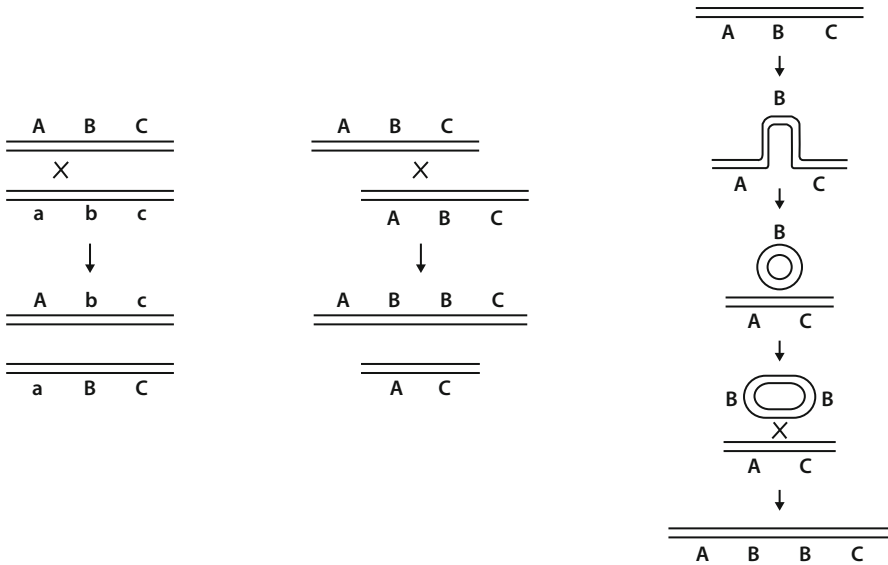


**Fig. 2.3** DNA replication. Double-stranded DNA (*left*) replicates by a semiconservative mechanism. The parental strands separate (*center*) and each codes for a new complementary strand. This results in two progeny double-stranded DNA molecules, each containing one old strand and one new strand (*right top*). Two types of errors in DNA replication may result in locally repeated regions (repeat sequences). These errors include slippage and fold back forming hairpin-like structures (*right center*), and replication reinitiation forming branches within branches (*right bottom*)

is another possible mechanism that may contribute to gene amplification. It is visualized as the start of a new replication fork before the previous replication fork has completed moving through the DNA. This leads to the formation of branched DNA structures. Gene amplification by replication reinitiation has been modeled in Sect. 3.8.

### 2.1.9 Recombination

Recombination is an exchange of pieces of DNA (Fig. 2.4). Recombination can result in new combinations of genes, and an increase or decrease in the numbers of genes. Recombination occurs during the formation of germ cells for sexual reproduction (meiosis) and the division of nonsexual somatic body cells (mitosis). If replicated parts of chromosomes, called chromatids, align properly before recombination, and exchange occurs, then new combinations of genes may occur and be segregated into sex cells. Sometimes parts of chromatids misalign before recombination. Such recombination with misalignment can result in an increase or decrease in the numbers of genes on chromosomes in either meiosis or mitosis. Recombination misalignment leading to gene amplification or deamplification is modeled as a Markov chain with denumerable infinity of states in Axelrod et al. (1994), and simulated as a discrete stochastic dynamical system in Baggerly and Kimmel (1995). Recombination within loops of DNA on the same chromosome may yield small fragments containing genes but not centromeres. When the acentric fragments replicate, amplified numbers of genes may be produced in tandem arrays. If these pieces of DNA recombine and reintegrate into a larger chromosomal piece containing a centromere, then the tandem arrays of amplified genes can become stabilized. This is modeled as a Galton–Watson process with denumerable type space in Sect. 7.5.

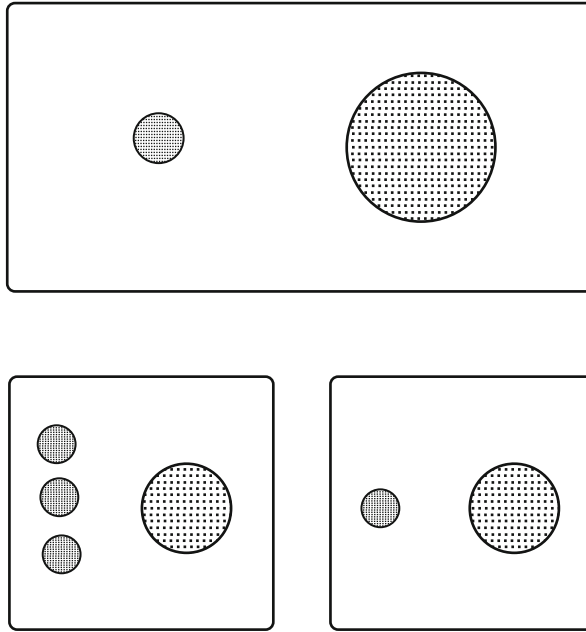


**Fig. 2.4** Recombination. New combinations and numbers of genes may be formed by rearrangement of pieces of DNA. Three examples are shown. Double-stranded DNA is represented by *double lines*, genes are represented by *letters*, exchange is represented by an *X*. *Left*, the DNA molecules exchange genes, *upper case* from the mother and *lower case* from the father, to produce new combinations of genes which may then be passed on to progeny. *Center*, the DNA strands slip and misalign before recombination producing one molecule with an increased number of a gene and another molecule with a decreased number of the gene. *Right*, one molecule of DNA undergoes exchange with itself producing a circular piece of DNA. If this piece replicates and then reintegrates, then the result may be an increased number of copies of a gene

## 2.2 Cells: Cell Cycle Kinetics and Cell Division

### 2.2.1 Cells as the Basic Units of Life

The basic structural unit of biological function and reproduction is the cell. Mammalian cells are in the range of  $20 \times 10^{-6}$  m of size although there are many cells of different functions and different shapes that are smaller or larger. The structure of the cell is a series of bag-like compartments with specialized functions. The “bags” are made up of membranes that function as barriers, and for transport of molecules. The innermost compartment is the nucleus which contains highly compacted DNA and accessory molecules for expression of genes. Outside of the nucleus is a series of compartments for the synthesis and degradation of molecules used for catalysis, structure, and energy generation. The outermost cell membrane, and its accessory molecules, also function as barriers and for transport, and in addition, for communication with other cells. Communication between cells can occur via small molecules that diffuse between cells such as hormones, or via molecules that become fixed to the surface of other cells, such as antigens which function in the immune system. A model for multivalent antigen binding as a multitype Galton–Watson process is given in Sect. 6.5.



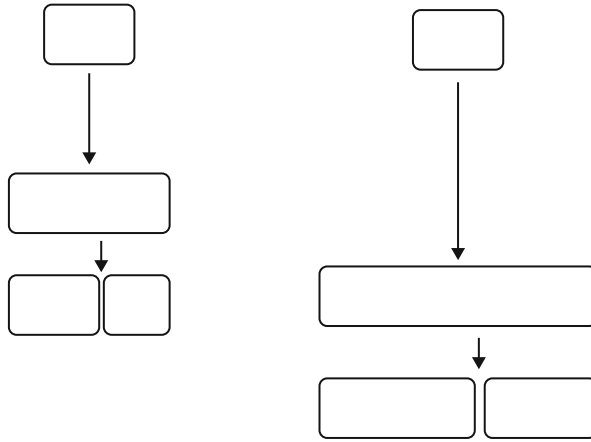
**Fig. 2.5** Cell division and partition of contents. During the growth of cells the amount of DNA in the nucleus (*large dark circle*) doubles and is partitioned evenly *into two daughter cells* at cell division. However, other cell constituents may not exactly double and may not partition evenly, resulting in cells with different numbers of these constituents. These constituents include extrachromosomal pieces of DNA, subcellular organelles, and intracellular parasites

### 2.2.2 Cell Growth, Division, and Death

In a multicellular organisms, adult tissue homeostasis is maintained. The number of new cells produced by cell division is balanced by the the number of differentiated cells that die and are removed. For instance, in the colon, some differentiated cells are constantly being removed and excreted with the feces. Other cells die by a conservative prograded cell death called apoptosis, in which pieces of the cells are engulfed by specialized neighboring cells. In the adult nervous system, dividing cells can give rise to several cell types, and a subset of newly born cells are culled, undergoing cell death via apoptosis. The production of several cell types and the death of some have been modeled as a multitype branching process in Sect. 6.2.

Cells grow in size and divide into two. The DNA in the nucleus exactly doubles in amount, is packaged into chromosomes, and is then partitioned evenly between two progeny cells at cell division. However, other processes are less exact. The size to which cells grow before they divide is not exactly the same for all cells of a given type, the lifetimes of cells at division are not exactly the same for all cells, and the non-DNA materials are not partitioned exactly between the two progeny cells (Figs. 2.5 and 2.6). The distribution of cell sizes and cell lifetimes may be stable over time for





**Fig. 2.6** Cell division and cell size. Cells may grow for different times and obtain different sizes before they divide. At cell division, cells may divide asymmetrically resulting in progeny sibling cells of different size

a population of one cell type, but differ for populations of cells of different types. Apparently, mechanisms exist to maintain these parameters within a population of cells of one type. Populations of cells with different values of parameters may differ in important characteristics, such as whether or not they are malignant. A Galton–Watson model describing the growth and division, and death and quiescence of cells is given in Sect. 3.2. Another model in the form of a Galton–Watson process with continuous type space is described in Sect. 7.8.1.

During development of multicellular organisms some cells divide into two cells which differ in shape and function. This situation is modeled as multitype branching process in Sects. 6.3 and 7.8.2. If fragments of DNA are not connected to chromosomes they may not exactly double in number before each cell division and may not partition exactly into the two progeny cells. Entities such as subcellular organelles or intracellular parasites can divide within dividing cells. An appropriate model for this is a Markov process model of infinitely many types. Such a model exhibits quasistationarity, as discussed in Sect. 7.6.

### 2.2.3 Stem Cells

Multicellular organisms are composed of many specialized cells that differ in function. Totipotent stem cells in the early embryo can yield all of the different specialized cells in a multicellular organism. Multipotent stem cells found in the adult can yield only a limited number of different kinds of cells. An example of the latter are the hematopoietic stem cells found in the bone marrow that yield the many specialized cells in the blood. The purpose of the bone marrow transplant procedure is to replenish active hematopoietic stem cells that are capable of yielding all of the

different specialized cells in the blood. Stem cells may be quiescent and may not divide, or if stimulated, may become active and divide. Active stem cells may divide symmetrically to produce two stem cells, or divide symmetrically to produce two differentiated cells, or divide asymmetrically to produce one stem cell and one differentiated cell. Stem cell division has been modeled as age-dependent and multitype branching processes in Sect. 7.8.1.

### 2.2.4 *Cell Cycle Kinetics*

The time period between cell birth and cell division is referred to as the cell cycle time. Several distinct events or phases can be distinguished during each cell's lifetime. The first event is the birth of two progeny cells at cell division, also called cytokinesis or mitosis, abbreviated  $M$ . The time gap between the birth of a new cell and the initiation of DNA synthesis is called gap one, abbreviated  $G_1$ . The period of DNA synthesis is abbreviated  $S$ . The time gap between  $S$  and the next mitosis is abbreviated  $G_2$ . After  $G_2$ , during the next  $M$  phase the cell divides to form two new cells. The sequence of phases  $M$ ,  $G_1$ ,  $S$ ,  $G_2$ , and  $M$  repeats in progeny cells of each subsequent generation, and thus the name cell cycle. For mammalian cells, a typical cell cycle time may be 12–24 h, or even longer. For a cell cycle time of 24 h, the duration for the cell cycle phases  $M$ ,  $G_1$ ,  $S$ , and  $G_2$  might be 0.5, 8, 12, and 3.5 h. The duration of the  $G_1$  phase is usually the most variable portion of the cell cycle. Cells which have longer cell cycle times, either because of genetics, environment or developmental fate, usually have equally extended  $G_1$  time periods, although important exceptions exist. The relative duration of the cell cycle phases in a growing population of cells can be inferred from the percentage of cells with different amounts of DNA, or from the rate at which cells accumulate in one phase of the cell cycle when blocked with a phase-specific drug (stathmokinesis). Cell cycle kinetics are modeled as a Bellman–Harris process in Sect. 5.4, and as a Markov time-continuous branching process in Sect. 4.2.

## 2.3 Cancer

### 2.3.1 *Cancer Cell Populations Are Immortal*

Cancer is a problem in persistent cell proliferation. Tumors are populations of cells that accumulate in abnormal numbers. The increased number of cells is due to an increased ratio of cell birth rate over cell death rate. Cancer cells do not necessarily grow faster than normal cells, but they are persistent. They may not stop dividing under conditions where normal cells would stop, and/or they may not die under conditions where normal cells would die. Normal cells in an adult seem to be capable of a finite number of divisions and then the lineage dies out. The process of a cell lineage losing proliferative potential is referred to as senescence. Some cells in

Branching Processes in Biology

Kimmel, M.; Axelrod, D.

2015, XX, 280 p. 69 illus., 3 illus. in color., Hardcover

ISBN: 978-1-4939-1558-3