This chapter is intended to enable students to:

■ recognise that the genome of an organism is the total base sequence of its genetic material, DNA
■ gain knowledge of the Human Genome Project
■ give examples of the actual and potential uses of genomics research
■ gain knowledge and understanding of the concept of genes and the history of their elucidation
■ identify the difference between genes and their alleles.

It is hard to overstate the importance of reading our own instruction book and that’s what the Human Genome Project is all about.

Francis Collins
What is a genome?

In the journal Science in 1910 (vol. 32, p. 120) Thomas Hunt Morgan (1866–1945), an American geneticist, published the first experimental evidence that genes are located on chromosomes. Morgan put into place the first piece of a giant jigsaw by showing the gene for white eye colour in the fruit fly (Drosophila melanogaster) is located on its X chromosome. At that time, no one knew just what a gene was, but locating genes on chromosomes showed that they were physical entities.

Morgan would have been amazed to know that just 90 years later the genetic jigsaw of his fruit flies was complete. In the 24 March 2000 issue of the same Science journal the base sequence of the entire DNA of the Drosophila chromosomes was published (see figure 13.2). This base sequence is expressed using just four letters: A, T, C and G. The base sequence contains the complete set of genetic instructions of the fruit fly and is called its genome.

The genome of an organism is its complete set of genetic instructions, encoded in DNA. For humans, the haploid genome consists of the DNA of the haploid set (22) of autosomes plus the sex chromosomes. The human genome has been variously described as ‘the book of life’, ‘humanity in chemical language’ and the ‘instructions to make a human’.

Similarly, the genomes of other eukaryotes (animals, plants, fungi and protists) are the DNA of the haploid sets of their chromosomes. When we refer to the genome of an eukaryotic organism, such as the chimp genome or the rice genome, we are speaking about the nuclear DNA. We can also talk about the genomes of those cell organelles that contain DNA, such as the mitochondrial genome or the chloroplast genome.

The genomes of prokaryotes (bacteria and archaea) comprise the DNA of their single circular chromosome that carries the genetic instructions of each species. The genomes of viruses consist of their entire genetic instructions encoded in one DNA molecule, or, in the case of retroviruses, in one RNA molecule.

The field of study of genomes is called genomics. Each genome is the sum total of an organism’s DNA and is expressed as the base sequence of the haploid set of chromosomes.

The Human Genome Project

Start counting! One, two, three . . . one million . . . three billion. To reach that three billion counting at one number per second would take about 95 years. Three billion is the approximate number of bases in the total DNA of a haploid set of human chromosomes, the human genome. Every human egg and sperm cell has a copy of the haploid genome and every nucleated somatic cell in the human body has a copy of the diploid human genome.

The international cooperative and publicly funded research project known as the Human Genome Project (HGP) began in 1990 (see figure 13.3). Its aim was to sequence the human genome, to store the sequence in accessible databases and to map all the human genes. When a genome is sequenced, it means that the precise order or sequence of bases in the DNA of the genome has been identified, with the error rate in the final sequence being less than one base in
every 10,000 bases. Research groups in several countries, principally the United States, United Kingdom, France, Germany and Japan, undertook the sequencing of individual human chromosomes. The HGP was ‘big science’ that could only be undertaken by involving many scientists from different laboratories around the world, working cooperatively.

In December 1999, chromosome 22 became the first human chromosome to be sequenced and it was found to have just 51 million bases and carry 431 genes. Six months later, in May 2000, chromosome 21 was also sequenced and its 48 million bases were found to contain 225 genes, far fewer than expected. In contrast chromosome 1, the largest human chromosome, has almost 250 million bases with at least 2100 genes.

In June 2000 the first working draft of the human genome was announced (see figure 13.4). In February 2001, a draft sequence that covered more than 90 per cent of the human genome was published in the journal *Nature*. The HGP formally ended in April 2003 when the final version of the human genome sequence was completed.

In revealing the sequence of the entire genetic instructions that make us human, the HGP stands as one of the greatest scientific explorations ever carried out. In announcing the first draft of the human genome in June 2000, US President Clinton described this achievement as being ‘the most important and most wondrous map ever produced by humankind’.
The first human genome sequence was completed using DNA from anonymous donors from diverse populations. Since then, complete sequences of the genomes of hundreds of persons have been completed and published, including that of James Watson, the co-discoverer of the double helical structure of DNA. **Sequencing is the process of identifying the order of the bases in DNA.** Technological advances, such as the development of new generation DNA sequencers, have made the process of sequencing genomes both quicker and cheaper (see figure 13.5). To complete a full human genome sequence in 2015 costs a few thousand dollars — a lot cheaper than the billions of dollars, and much faster than the 13 years, required to produce the first human genome.

**FIGURE 13.5** A new generation high-speed DNA sequencer that rapidly reads the sequence of DNA input to this system

Benefits resulting from the HGP include positive impacts on medicine and other areas of human biology:

- **Diagnosis.** Data from the HGP are being used to provide improved and more accurate diagnoses of inherited disorders due to single genes. In 2013, researchers identified a mutated gene, CALR, that is present in 40 per cent of people with certain chronic blood cancers. This discovery will enable such people to be diagnosed more quickly and accurately.

- **Treatment.** Data from the HGP are being used to identify the products of genes and infer how mutant alleles produce the undesirable effects of inherited disorders. Such understanding will generate new treatments for inherited disorders as well as improve accuracy of diagnosis.

- **Prevention.** Data from the HGP are identifying genetic factors that predispose some people to disabling conditions, such as strokes and cancers. This knowledge is expected to identify people at risk and develop treatments to reduce the incidence of these conditions. It is of interest to note that in a recent survey of nearly 7000 people, 98 per cent of them stated that they would want to be told if a researcher found that their genomes contained an indicator of a serious treatable or a serious preventable disease (see www.sanger.ac.uk/about/press/2015/150429.html)

- **Human biology.** Data from the HGP provide better understanding of the genetic control of normal human development; for example, the Deciphering Developmental Disorders (DDD) project, a 5-year research project that began in 2011 in the United Kingdom, is collecting genomic
data from a large sample of children with multiple malformations or with significantly delayed physical or mental development. The researchers will use this genomic information to identify any genetic bases for these conditions with the intention of achieving improvements in the diagnosis and the care of such children. In December 2014, researchers on the DDD project announced the identification of the first 12 genes found to cause these rare but serious developmental disorders.

- **Human evolution.** Mitochondrial DNA sequence data are providing new insights into aspects of human evolution and the prehistoric migrations of human groups. The sequencing in December 2013 of the genome of a Neanderthal female (*Homo neanderthalensis*) provided the first genomic information about a first cousin to modern humans (*Homo sapiens*).

While the HGP has officially ended, ongoing research continues including deciphering the functions of the various parts of the human genome, identifying how and when genes function and how they are controlled. Because of the enormous amount of data involved, computers are needed for the storage, retrieval and analysis of genomic data.

Many projects have built on the HGP database, such as the DDD project mentioned above. Other projects include:

- **the 1000 Genomes Project.** This project aims to sequence the genomes of at least a thousand people from around the world in order to develop a new map of the human genome that will highlight biomedically relevant DNA variations. These variations may be in protein-coding genes, but they may also be in regions of the genome that include gene promoters and enhancers which control whether genes are switched on or off and, if switched on, the level of gene activity.

- **the Cancer Genome Anatomy Project.** This project (see figure 13.6) aims to determine the patterns of gene expression of normal, precancer and cancer cells, including the identification of genomic changes seen in various tumours as compared with normal cells. This information is leading to improvements in the early detection, diagnosis and treatment of patients with cancer.

- **the International HapMap Project.** The aim of this project is to develop a haplotype map of the human genome that will identify the common patterns of human DNA sequence variation and to identify genes affecting health, disease and responses to drugs.

- **the Encyclopedia of DNA Elements (ENCODE) Project.** This project involves many hundreds of scientists worldwide working collaboratively. Its aims are to identify the function of all the DNA of the human genome. To date, scientists have identified biochemical functions for about 80 per cent of the genome.

Apart from large-scale studies, genomic information can be used at an individual level where, along with a person’s family history and information about their lifestyle, the risk for a particular disease can be estimated. In addition, people differ in their responses to various pharmaceutical drugs. As genomic information becomes more widely available, this is expected to lead to decisions about aspects of medical care becoming more personalised, for example, about which drugs to prescribe and what dosages to administer.

**What does the HGP tell us?**

Every person — apart from identical twins — has a unique genome, but the genomes of all members of the human species *Homo sapiens* share many similarities. It is estimated that the genomes of two unrelated people will on average have about three million differences. Sounds like a lot, but it is not as
this means they differ by only one base pair in a thousand, or 0.1 per cent. Put another way, the genomes of two unrelated people are 99.9 per cent the same (see figure 13.7).

![Figure 13.7](a) Each person in this group has a unique genome but, as humans, their genomes are 99.9 per cent similar. How many of the 3 billion bases in the human genome are represented by a difference of 0.1 per cent? (b) A colour-coded computer rendering of the base sequence of a fragment of DNA of the human genome. It is not possible to fit the complete 3 billion bases on a single screen display. How many different colours are needed to represent the bases involved?

Many of the differences between the genomes of different people are single-base differences in the DNA sequences of their genomes (see figure 13.8). When a particular variant appears in at least 1 per cent of the population, it is called a **single nucleotide polymorphism** (SNP, pronounced ‘snip’). When SNPs occur in a protein-coding region of DNA, they can alter the function of the protein and result in an inherited disorder. However, most SNPs occur in the noncoding region of the DNA.

![Figure 13.8](Matching parts of the DNA sequence in the genome of two people. There is just a single difference (shown arrowed) in one of the DNA strands (and of course in the complementary strand). Differences of this type between genomes are known as SNPs.)

| Person 1 | - - C C T T G C G T A | A T C C G - - - |
| - - G G A A C G C A T | T A G G C - - - |
| Person 2 | - - C C T T G C G T A | C T C C G - - - |
| - - G G A A C G C A T | G A G G C - - - |

Other differences that exist between people’s genomes include the number of repeats of short DNA sequences, typically 3 to 5 nucleotides long, known as **short tandem repeats (STRs)** — this difference has important forensic applications.

The HGP revealed various facts about the human genome, for example:
- **How big?** About 3 000 000 000 bp (3 billion base pairs) organised as a DNA double helix of the chromosomes. The full print-out of the 3 billion plus base sequence of the entire human genome fills 116 volumes, each with about 1000 pages and printed in very small type. The base sequence of chromosome 1, the largest human chromosome, fills nine volumes, while the base sequences of the smallest chromosomes, 21 and 22, and the Y chromosome fill just two volumes each. Figure 13.9 shows these volumes at the Wellcome Collection in London.
**Figure 13.9** (a) The print-out of the HGP in the Wellcome Collection’s ‘Medicine Now’ exhibition contains 116 volumes, which have the complete set of genetic instructions to make a human being. In what language are these books written? (b) One page from the chromosome 6 volume. Note the very small typeface.

- **How many genes?** There are about 21,000 genes located on the DNA of the 22 non-homologous autosomes and the X and Y sex chromosomes. Possible genes are identified by computer scans of a chromosomal DNA sequence that looks for a start signal (TAC) followed by a long sequence before a stop signal is reached. Such sequences are called open reading frames (ORFs) and indicate possible genes.

- **What do these genes do?** The majority of the genes on a chromosome are protein-coding genes that are transcribed into messenger RNA (mRNA) molecules; these in turn are translated into polypeptide chains. Surprisingly, these protein-coding genes constitute less than 2 per cent of the total genome! In addition, there are genes that are transcribed into other kinds of RNA, including ribosomal RNA (rRNA), which forms part of the ribosomes, and transfer RNA (tRNA), which is involved in the synthesis of proteins in a cell. Together, these two kinds of gene make up just a small percentage of the human genome. This means that the bulk of the human genome is noncoding DNA.

- **What are the functions of the noncoding DNA?** The function of much of the noncoding DNA is not known. However, some of the DNA sequences in the genome are regulatory elements, including so-called promoters and enhancers that govern when a gene is active and its level of expression. Other noncoding DNA is present in the telomeres at the chromosomal ends and these are important in maintaining chromosome structure. Yet other noncoding DNA comprises multiple repeats of short DNA sequences.

**Sequencing of other genomes**

By the end of 2014, the complete genomes of more than 15,000 different species had been sequenced, either completely or as permanent drafts. Scientific journals regularly include papers on the complete sequencing of the genome of various species; for example, in 2008 the completion of the platypus genome was announced (refer to figure 13.1).
Table 13.1 lists a few key species whose genomes have been sequenced. In addition, the genomes of the mitochondria of several species, including the human mitochondrial genome (16568 bp), have been sequenced, as have the chloroplast genomes of several plant species. GOLD (Genomics OnLine DataBase) identifies all genomes that have been completely sequenced as well as those for which the sequence is incomplete.

### Table 13.1 Dates of complete sequencing of genomes of a virus and species from the three domains (Bacteria, Archaea and Eukarya)

<table>
<thead>
<tr>
<th>Organism</th>
<th>Date published</th>
<th>Size of genome (base pairs, bp)</th>
<th>Estimated number of coding genes</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>virus phiX174 (virus)</td>
<td>Apr. 1993</td>
<td>5 386</td>
<td>11</td>
<td>first genome sequenced</td>
</tr>
<tr>
<td><em>Haemophilus influenza</em> (bacterium)</td>
<td>July 1995</td>
<td>1 830 000</td>
<td>1 850</td>
<td>first bacterium</td>
</tr>
<tr>
<td><em>Saccharomyces cerevisiae</em> (brewers' yeast)</td>
<td>Apr. 1996</td>
<td>12 069 000</td>
<td>6 294</td>
<td>first eukaryote and first fungus</td>
</tr>
<tr>
<td><em>Methanococcus janaschii</em> (archaean found at hydrothermal vents)</td>
<td>Aug. 1998</td>
<td>1 700 000</td>
<td>1 738</td>
<td>first archaean</td>
</tr>
<tr>
<td><em>Caenorhabditis elegans</em> (nematode worm)</td>
<td>Dec. 1998</td>
<td>97 000 000</td>
<td>19 099</td>
<td>first animal</td>
</tr>
<tr>
<td><em>Arabidopsis thaliana</em> (thale cress)</td>
<td>Dec. 2000</td>
<td>115 000 000</td>
<td>25 498</td>
<td>first plant</td>
</tr>
<tr>
<td><em>Equus caballus</em> (horse)</td>
<td>Nov. 2009</td>
<td>2 689 000</td>
<td>20 322</td>
<td>thoroughbred</td>
</tr>
<tr>
<td><em>Yersinia pestis</em></td>
<td>Aug. 2011</td>
<td>4 553</td>
<td></td>
<td>cause of Black Death plague</td>
</tr>
<tr>
<td><em>Brassica rapa</em> (Chinese cabbage)</td>
<td>Aug. 2011</td>
<td>485 000</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Anolis carolinensis</em> (green anole lizard)</td>
<td>Aug. 2011</td>
<td>2 200 000</td>
<td>16 533</td>
<td>see figure 13.10a</td>
</tr>
<tr>
<td><em>Anopheles gambiae</em> (mosquito)</td>
<td>2002</td>
<td>278 268 413</td>
<td>12 843</td>
<td>main vector of malaria in sub-Saharan Africa</td>
</tr>
<tr>
<td><em>Gallus gallus</em> (chicken)</td>
<td>2004</td>
<td>1 072 544 763</td>
<td>15 508</td>
<td></td>
</tr>
<tr>
<td><em>Canis familiaris</em> (dog)</td>
<td>2005</td>
<td>2 392 715 236</td>
<td>19 856</td>
<td></td>
</tr>
<tr>
<td><em>Felis catus</em> (domestic cat)</td>
<td>2007</td>
<td>2 365 745 914</td>
<td>19 493</td>
<td></td>
</tr>
<tr>
<td><em>Oryctolagus cuniculus</em> (rabbit)</td>
<td>Nov. 2009</td>
<td>2 604 023 284</td>
<td>19 293</td>
<td></td>
</tr>
<tr>
<td><em>Pongo pygmaeus</em> (orangutan)</td>
<td>2011</td>
<td>3 109 347 532</td>
<td>20 424</td>
<td>one of the five great apes</td>
</tr>
<tr>
<td><em>Macropus eugenii</em> (tammar wallaby)</td>
<td>Aug. 2011</td>
<td>2 549 429 531</td>
<td>15 290</td>
<td>see box on page xxx</td>
</tr>
<tr>
<td><em>Sarcophilus harrisii</em> (Tasmanian devil)</td>
<td>Feb. 2011</td>
<td>2 931 556 433</td>
<td>18 788</td>
<td></td>
</tr>
<tr>
<td><em>Escherichia coli</em> 0111 (bacterium)</td>
<td>2011</td>
<td>5 766 081</td>
<td>5 407</td>
<td>10 000th genome in GOLD database</td>
</tr>
<tr>
<td><em>Falco peregrines</em> (Peregrine falcon)</td>
<td>2013</td>
<td>1 200 000 000</td>
<td>16 263</td>
<td>a top predator in some ecosystems</td>
</tr>
<tr>
<td><em>Panthera tigris</em> (Amur (Siberian) tiger)</td>
<td>Sep. 2013</td>
<td>~2 440 000 000</td>
<td>20 226</td>
<td>first entire genome of the endangered Amur tiger (see figure 13.10b)</td>
</tr>
<tr>
<td><em>Phascolarctos cinereus</em> (koala)</td>
<td>Apr. 2013</td>
<td>approx 15 000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Many research groups around the world continue to collaborate on identifying the genomes of key species. Among these are the Koala Genome Consortium, a joint initiative of the Australian Museum and the Queensland University of Technology. Another group is the Open Tiger Genome Project (see figure 13.11), which has many collaborators, including major institutions in Korea.

Comparative genomics

The availability of the complete genome sequences for an increasing number of species has created a new field of study known as **comparative genomics**. Comparing the genomes of various species will elucidate how various features have evolved and how the genomes of closely related species differ.

In August 2005, the sequencing of the genome of the chimp (Pan troglodytes) (see figure 13.12) was completed. Comparisons between the chimp genome and the human genome can identify the genes that control the distinctive features of primates, such as high brain-to-body-mass ratios, and the genes that are unique to the human species.

**ODD FACT**

The bacterium *Carsonella ruddii* has the smallest number of genes of any living organism — just 182 genes — and a genome of just 160,000 base pairs. *C. ruddii* lives as an endoparasite in the cells of a sap-eating insect.
Comparative genomics provides data to assist research into medicine, ecology and biodiversity. It is also a powerful tool in exploring evolution; for example, comparative genomics can provide evidence of the occurrence of processes, such as gene duplication, where a second copy of a gene appears in a genome, and horizontal gene transfer, where a new gene is acquired by one species as a result of the transfer of DNA from a second species.

Comparative genomics also reveals which genes are conserved in different species. Figure 13.13 shows a map of the B2 cat chromosome alongside human chromosome 6. The genes that are conserved in the two species are indicated by the dotted lines. These conserved genes have key functions so that natural selection favours their retention during evolution.

Comparative genomics has become a powerful tool in exploring the degree of evolutionary relationships between different species and enabling a more precise identification of their degree of relatedness. For example, the genomes of the big cats (lions, tigers, leopards) are being compared to identify regions of similarity and regions of difference that provide clues to their evolutionary history. In addition, knowledge emerging from comparative genomics may provide clues to improved strategies for conservation of endangered and rare species.

New fields of research are emerging

The field of genomics involves far more than producing a sequence of nucleotides or base pairs for different species. From genome sequences, researchers are producing maps that identify the order of genes and their relative positions on the chromosomes.
Genomics is raising questions about the how, where and when of gene action. Exploring these questions is generating new fields of research including (see figure 13.14):

- **transcriptomics**, which seeks to identify all the mRNA transcripts produced by each cell type
- **proteomics**, which seeks to identify all the different proteins produced by each cell type
- **metabolomics**, which seeks to identify all the metabolites present in each cell type.

Another recent field to emerge is that of **microbiomics**. The **microbiome** is composed of the total collection of genes (genomes) of the several thousand different kinds of microbe that colonise the human body, including the gut, throat and skin. The US National Institutes of Health has sponsored the Human Microbiome Project. This project is sampling and sequencing, not individual microbes but the microbial communities living at different sites on and in the bodies of 300 healthy young adults. This project will enable researchers to identify differences in the microbiomes of persons as they change from good health to a state of disease. In addition, the microbiome is being investigated in various chronic conditions, such as the microbiome of the lung in persons with chronic lung conditions and the microbiome of the gut in persons and with an inflammatory condition of the gut, known as Crohn’s disease.

In the next section, we will briefly examine DNA and its constituent bases. It is these bases that form the genome of an organism.

---

**BIOLOGIST AT WORK**

**Dr Sue Forrest — molecular geneticist**

Dr Sue Forrest is a molecular geneticist and the CEO of the Australian Genome Research Facility, a major national research facility with nodes in Brisbane, Melbourne, Adelaide, Perth and Sydney. In her previous position she spent a total of 13 years at the Murdoch Children’s Research Institute at the Royal Children’s Hospital in Parkville. There she headed the Gene Discovery Group for 5 years, developing methodologies for the discovery of the genes responsible for common human diseases and, prior to that, ran the DNA Diagnostic laboratory for 8 years.

My interest in genetics developed right from my first introduction to this fascinating area in first-year Biology as part of my Bachelor of Science at Melbourne University. Following completion of my Honours degree, majoring in Biochemistry and Genetics, I headed overseas to study for my DPhil at Oxford University where I was fortunate to work with Professor Kay Davies. We cloned the gene dystrophin in 1987. This gene, when mutated, results in Duchenne muscular dystrophy and was one of the first disease-causing genes to be cloned in the late 1980s.

A fantastic event in genetic history occurred in 2003 when the sequence of the human genome was announced as completed. The first major outcome was that there were only about 21,000 genes in the human genome, compared with the 100,000 originally predicted, requiring new ideas about gene structure and function to be developed. Since the Human Genome Project was instigated, there is far more information about genes and their sequences on the internet and much of the research is now done as ‘computer cloning’ rather than actual laboratory bench work! The challenge now is to determine the functions of the genes in the human genome and how they are regulated.

During this finalisation of the Human Genome Project in 2001, I was offered the position of Scientific Director of the Australian Genome Research Facility (AGRF) followed by Director/CEO in 2003. AGRF is partly funded by the federal...
government to provide access to state-of-the-art genetic tools and technologies that can be used by researchers across the whole biological spectrum. Thus, moving to this position dramatically opened my eyes to the vast array of molecular biology and genetic research in all different species, from microbes through to animals, that was occurring within Australia and around the world.

Australia did not play a major role in the sequencing of the human genome, but through a unique collaboration between the US National Institutes of Health and the Australian Genome Research Facility, funded by the state government of Victoria, the genetic sequence of the tammar wallaby was determined. The project took 3½ years and was completed in 2009. This sequence has assisted with defining which regions of the genome share sequence between human and wallaby, thereby indicating that they are likely to have a significant function. Such sequences could be involved in regulating gene expression as an example. Also, much biological research has been done in Australia on the tammar wallaby demonstrating novel properties of lactation, development and reproduction that will be unravelled using the genetic sequence.

What next? In the last few years, the technology available for sequencing DNA has become faster and cheaper. The ability to sequence a human genome for $1000 is just about here! The challenge now is to understand the ethical, legal and social issues that surround the availability of human genetic information. It is an exciting time in genetics and I certainly would never have predicted in the early 1980s that, 30 years later, I could be reading my own genome sequence!

**Key Ideas**

- A genome is the complete set of genetic instructions for an organism.
- The haploid human genome consists of the base sequence of the DNA of the 22 autosomes plus the sex chromosomes.
- The Human Genome Project (HGP) produced the first working draft of the complete human genome.
- The genomes of members of the human species have a high level of similarity.
- Many benefits and new fields of research have flowed from the HGP.

**Quick Check**

1. Identify whether each of the following statements is true or false.
   - a. A complete genome can be expressed using just four letters.
   - b. The human genome (haploid) contains about 3 million base pairs.
   - c. The first human chromosome to be sequenced was the number-21 chromosome.
   - d. The genomes of unrelated persons would be expected to show a high degree of difference.
   - e. The first eukaryotic organism to be fully sequenced was a nematode worm.

2. Briefly identify two benefits of the HGP.

3. What is meant by the following terms:
   - a. comparative genomics
   - b. proteomics
   - c. microbiome?
DNA and its bases

As we have seen in the previous section, genomics is concerned with the sum total of DNA in the haploid set of chromosomes of an organism.

Genomes are the complete sequence of bases present in the DNA of an organism. A genome can be shown in several ways: as a long sequence using just four letters — A, T, C and G — to denote the four bases in DNA (see figure 13.16), or expressed as a series of coloured bands, using just four different colours to denote the four bases — typically, green for adenine (A), red for thymine (T), yellow for cytosine (C) and blue for guanine (G) (refer to figure 13.7b).

Nucleotides: building blocks of DNA

Complex structures are built from one or more building blocks (sub-units) that are organised in a regular manner: insect eyes are built from ommatidia, walls are built of bricks, fences are built of palings. The genetic material DNA is a complex molecule built of many basic building blocks called nucleotides. (Other complex molecules that are made of basic building blocks include proteins that are built of amino acids.)

The nucleotide sub-units in DNA are assembled head-to-tail forming a chain. Four different kinds of nucleotides are found in DNA and they are normally distinguished by the letters A, C, G and T.

Each nucleotide has a sugar (deoxyribose) part, a phosphate part and an N-containing base. The sugar and the phosphate parts are the same in all four nucleotides. However, the different nucleotides vary in the bases they contain.

Note that the letters A, T, C and G that are used to label the four different kinds of nucleotides come from the names of the bases they contain. Figure 13.17 shows various representations of the four nucleotides.

When many nucleotides join to form a chain, a bond forms between the sugar of one nucleotide and the phosphate group of the next nucleotide, and so on (see figure 13.18). So, one chain of nucleotides runs from ‘head-to-tail’ with a phosphate group at the ‘head’ end (also known as the 5′ [5 prime] end) and a sugar molecule at the ‘tail’ end (also known as the 3′ [3 prime] end).

So DNA is built from nucleotides joined to form a chain. However, the question remained: how is a typical DNA molecule arranged in three-dimensional space? For example, does it consist of one nucleotide chain coiled into a ball? Does it contain more than one nucleotide chain?

**FIGURE 13.16** Part of the sequence of bases in the DNA that makes up the genome of an organism. Would this image look significantly different if this organism were a plant rather than an animal?

**FIGURE 13.17** Four different representations of the nucleotides that are the sub-units of DNA. Can you identify the phosphate, the sugar and the base in each nucleotide?

**ODD FACT**

Two of the bases, cytosine and thymine, belong to the class of chemical compounds called pyrimidines. The other two, adenine and guanine, are larger and belong to the class of chemical compounds called purines.
Early analysis of DNA

The relative proportions of the different nucleotides in DNA from various organisms can be identified. Table 13.2 shows a summary of experimental results obtained by various scientists in the late 1940s and early 1950s.

**TABLE 13.2** Approximate values for the total amounts of the four kinds of nucleotides in DNA samples from different organisms. What predictions would you make about DNA from calf liver cells?

<table>
<thead>
<tr>
<th>Source of DNA</th>
<th>A</th>
<th>T</th>
<th>C</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>calf thymus</td>
<td>1.7</td>
<td>1.6</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>yeast cells</td>
<td>1.8</td>
<td>1.9</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>tubercle bacteria</td>
<td>1.1</td>
<td>1.0</td>
<td>2.6</td>
<td>2.6</td>
</tr>
<tr>
<td>herring sperm</td>
<td>1.1</td>
<td>1.1</td>
<td>0.9</td>
<td>0.9</td>
</tr>
</tbody>
</table>

These figures indicate a possible pattern. It appears that in DNA the proportions of A and T are about equal and also that the proportions of C and G are about equal. This idea, known as **Chargaff’s rule**, was an important observation that later contributed to understanding the 3D structure of DNA.

**DNA forms a double helix**

In 1953, James Watson (1928– ) and Francis Crick (1916–2004) (see figure 13.19) announced that they had identified the 3D structure of DNA as being two nucleotide chains arranged to form a **double helix**. Their work built on the work of other scientists, in particular, the pioneering X-ray crystallography work of Rosalind Franklin (1920–58) and Maurice Wilkins (1916–2004).

The key features of the double helix model of DNA (see figure 13.20) are:

- Each DNA molecule consists of two nucleotide chains.
- The chains run in opposite directions and are said to be ‘anti-parallel’.
- The sugar–phosphate backbones of the two chains are on the outside of the DNA double helix and they coil around each other in a regular manner to form a molecule with a constant diameter.
- The nucleotide bases (A, T, C and G) are arranged so that they point to the inside of the DNA molecule.
- The bases in one chain pair with the bases in the second chain in a very specific way; there is pairing only between A and T and between C and G.

Weak **hydrogen bonds** form between the base pairs.
FIGURE 13.20 Two representations of DNA. (a) Double helix (side view) showing the sugar-phosphate backbones on the outside of the DNA, with the sugars (deoxyribose) shown in yellow and the phosphates in pink. The inner ‘rungs’ are formed by base pairing of A with T and C with G. (b) Pairing of the nucleotide bases of the inner rungs of the DNA double helix. Here we see the complementary pairing between A and T and between C and G. Notice that the size of an A–T pair is the same as the size of a C–G pair. This observation was an important clue in solving the structure of DNA.

The base pairs between the two strands, namely, A with T and C with G, are said to be complementary base pairs (see figure 13.21). This complementary double helix structure for DNA fits with the known properties of the genetic material including the facts that DNA:

- can act as a template for its own replication
- contains genetic instructions
- can undergo change or mutation.

FIGURE 13.21 Chemical formulae for the complementary bases in DNA. Note the two hydrogen bonds in the A–T pair and the three hydrogen bonds in the C–G pair (shown in red).

The box on page xxx contains excerpts from James Watson’s personal account of the discovery of the DNA double helix (The Double Helix, Atheneum, New York, 1968). Reading these excerpts may help you understand how scientists work and realise that they spend time thinking about problems and assessing alternative ideas, not just doing experiments.

A double helical DNA molecule contains complementary base pairs: A–T, T–A, C–G and G–C. If the order along one DNA chain is known, the sequence of bases in the second chain can be inferred. So, if one DNA chain has the base sequence G G T A C G T A . . . , the sequence in the complementary chain must be C C A T G C A T . . .

Double helical DNA consists of two nucleotide chains held together by weak hydrogen bonds between the complementary nucleotides. These bonds are easily broken with a low input of energy. If a solution of DNA is heated to 90 °C for 2 minutes, each double helical DNA molecule separates (a process called dissociation) to form two single chains of DNA (see figure 13.22). This heating does not break the strong sugar-phosphate bonds that join nucleotides.

ODD FACT

In 1962 the Nobel Prize was awarded jointly to Watson, Crick and Wilkins for their work in discovering the structure of DNA. The other person who played a decisive role in this discovery was Rosalind Franklin. She died of cancer in 1958 at the age of 37. The rules governing the Nobel Prize do not permit an award to be made to a person after death.
into one chain. If this solution is then allowed to cool, the complementary regions of the chains pair, the hydrogen bonds re-form and the DNA returns to a double-stranded helix form (re-association) (see figure 13.22). Pairing between complementary DNA chains or parts of chains from different sources is referred to as hybridisation.

**How is DNA packed into a chromosome?**

The total length of DNA in a diploid human cell is more than 2 metres. Yet, a human cell is typically only about 30 micrometres (μm) in diameter. Seems strange, if not impossible, that this amount of DNA could fit into one cell, let alone just the nucleus of a cell.

The width of a DNA double helix is about 2 nanometres (or just 0.002 μm). The DNA double helix becomes tightly coiled around proteins known as histones and these coils become further coiled into super coils called nucleosomes (see figure 13.23). Further coiling enables the DNA to be organised into chromosomes.
Finding the double helix

In the 1950s, in addition to James Watson and Francis Crick from Cambridge University, other scientists were trying to discover the structure of DNA. These included Rosalind (Rosy) Franklin and Maurice Wilkins, also at Cambridge, who identified the X-ray diffraction patterns of crystalline DNA, and Linus Pauling in the United States.

Possibly three chains?

Watson wrote:

Superficially, the X-ray data were compatible with two, three, or four strands.

At first, Watson and Crick tried a three-chain model, with the sugar–phosphate backbones at the inside of the model. Watson wrote:

... we decided upon models in which the sugar–phosphate backbone was in the centre of the molecule. Only in that way would it be possible to obtain a structure regular enough to give the crystalline diffraction patterns observed by Rosy and Maurice.

Our first few minutes with the models, though, were not joyous... After tea, however, a shape began to emerge which brought back our spirits. Three chains twisted about each other... Admittedly, a few of the atomic contacts were still too close for comfort, but, after all, the fiddling had just begun.

Eventually, Watson and Crick realised that this three-chain model had major faults. Watson wrote:

A fresh start would be necessary to get the problem rolling again.

Chargaff's rule provides a clue

A new avenue of exploration was raised by the ratio of the four bases in DNA:

The moment was thus appropriate to think seriously about some curious regularities in DNA chemistry... the number of adenine (A) molecules was very similar to the number of thymine (T) molecules, while the number of guanine (G) molecules was very close to the number of cytosine (C) molecules.

Back in my rooms I lit the coal fire... With my fingers too cold to write legibly I huddled next to the fireplace, daydreaming about how several DNA chains could fold together in a pretty and hopefully scientific way.

At that time, Linus Pauling, an American scientist, developed his model for the structure of DNA. He sent the details of his model through his son Peter to Watson and Crick, who at first were disappointed to think that they had been beaten to the answer. Watson commented:

... my stomach sank in apprehension... Seeing that neither Francis nor I could bear any further suspense, he [Peter] quickly told us that the model was a three-chain helix with the sugar–phosphate backbone in the centre. This sounded so suspiciously like our aborted effort of last year...

In fact, Pauling’s model was incorrect. The critical breakthrough came when Rosalind Franklin prepared X-ray diffraction patterns of a different form of DNA (the B-form) (see figure 13.24). Watson wrote:

The instant I saw the picture my mouth fell open and my pulse began to race. The pattern was unbelievably simpler than those obtained previously. Moreover, the black cross of reflections which dominated the picture could arise only from a helical structure...
He later recalled:

Then as the train jerked towards Cambridge, I tried to decide between the two- and three-chain models. Thus by the time I had cycled back to college and climbed over the back gate, I had decided to build two-chain models. Francis would have to agree. Even though he was a physicist, he knew that important biological objects come in pairs.

**Fitting two chains together**

Over the next months, Watson and Crick tried to build a two-chain model, but they were still working with the incorrect idea that the sugar-phosphate backbones were in the centre of the molecule and the bases on the outside. Watson said:

...for a day and a half I tried to find a suitable two-chain model with the backbone in the centre... Though I kept insisting that we should keep the backbone in the centre, I knew none of my reasons held water.

But the real stumbling block was the bases. As long as they were outside, we did not have to consider them. If they were pushed inside, the frightful problem existed of how to pack together two or more chains with irregular sequences of bases.

Slowly, the correct model started to evolve:

The next morning, however, as I took apart a particularly repulsive backbone-centred molecule, I decided that no harm could come from spending a few days building backbone-out models. This meant temporarily ignoring the bases... There was no difficulty in twisting an externally situated backbone into a shape compatible with the X-ray evidence.

The sugar-phosphate backbones were arranged on the outside of the model; they posed no further problem, but the nagging problem of what to do with the bases still remained:

I went ahead spending most evenings at the films, vaguely dreaming that at any moment the answer would suddenly hit me... Even during good films I found it almost impossible to forget the bases. Thus, unless some very special trick existed, randomly twisting two polynucleotide chains around one another should result in a mess.

Another clue was recognised:

Thus, conceivably the crux of the matter was a rule governing hydrogen bonding between bases.

Watson first considered the pairing of identical bases, that is, A with A, T with T, and so on:

I thus started wondering whether each DNA molecule consisted of two chains with identical base sequences held together by hydrogen bonds between pairs of identical bases.

For over two hours I happily lay awake with pairs of adenine residues whirling in front of my eyes. Only for brief moments did fear shoot through me that an idea this good could be wrong.

However, the answer was very close. Watson wrote:

...so I spent the rest of the afternoon cutting accurate representations of the bases out of stiff cardboard... the following morning, I quickly cleared away the papers from my desk top so that I would have a large, flat surface on which to form pairs of bases held together by hydrogen bonds... Suddenly I became aware that an adenine-thymine pair held together by two hydrogen bonds was identical in shape to a guanine-cytosine held together by at least two hydrogen bonds. All the hydrogen bonds seemed to form naturally; no fudging was required...

The DNA double helix structure was at last identified. The pairing between the bases in DNA involves hydrogen bonding between complementary bases, not identical bases. This model with A-T and C-G pairs between the chains made sense in terms of Chargaff’s rule that the number of As was about equal to those of T, and that the number of Gs was about the same as those of C. This model also gave a clue as to how DNA could be replicated:

Chargaff’s rules then suddenly stood out as a consequence of a double-helical structure for DNA... Given the base sequence of one chain, that of its partner was automatically determined. Conceptually, it was thus very easy to visualise how a single chain could be a template for the synthesis of a chain with the complementary sequence.

Watson and Crick announced their double-helix model in the short article that briefly outlines a discovery that ranks as one of the major discoveries of the twentieth century:


Use the [Francis Crick weblink in your eBookPLUS](#) to read what Francis Crick thought.
KEY IDEAS

- DNA is a macromolecule built of nucleotide sub-units.
- Each nucleotide contains a sugar (deoxyribose), a phosphate group and a base.
- Nucleotides differ only in their bases, which may be one of four different bases.
- DNA normally exists as a double helix, with the two chains stabilised by weak hydrogen bonds.
- In a DNA double helix, each base along one chain pairs with its complementary base in the other chain.
- Double helical DNA can be reversibly dissociated into two single DNA chains by heating and the chains then re-associate on cooling.
- Chargaff’s rule states that in double helical DNA the proportions of A and T are equal as are the proportions of C and G.

QUICK CHECK

4 Identify whether each of the following statements is true or false.
   a DNA is composed of four different nucleotides.
   b Nucleotides differ in their sugar component.
   c Treating double helical DNA at 90 °C causes irreversible denaturation of the molecule.
   d Complementary base pairing occurs between bases in the two chains of a DNA double helix.

5 Consider part of a DNA chain that has the nucleotide sequence . . . T T A G G A C . . . Which of the following is part of the complementary strand?
   b . . . T T A G G A C . . .
   c . . . A A T C C T G . . .

6 The relative proportion of the G base in DNA from human gut cells was found to be 1.4 and that of T was 0.9. What other valid conclusions may be drawn?

Solving the puzzle: the nature of genes

DNA (deoxyribonucleic acid) is now widely known to be the raw material of genes. The term DNA now appears in newspapers: ‘DNA found at the crime scene . . .’ People talk about attributes as ‘being in their DNA’. However, the nature of genes as segments of DNA with particular functions was not known to your grandparents.

The story of genes begins in a monastery garden in 1856 where the patterns of inheritance of ‘factors’ was identified. The term ‘gene’ was first used to refer to these factors in 1909. However, the chemical nature of genes was not experimentally demonstrated until the 1920s, and it was not until the 1940s that their chemical nature was shown to be DNA. The structure of DNA, that underpins key properties of genes, such as their ability to self-replicate and to mutate, was not elucidated until 1953.

Let’s look briefly at the emergence of genes, firstly as Mendel’s factors and finally as specific segments of double helical DNA controlling a specific function.
Factors in the monastery garden

In the summer of 1856, visitors to the monastery of St Thomas in the town of Brno, in what is now the Czech Republic, would have seen monks at work and prayer. Visitors may have noticed one monk examining flowers on pea plants in the vegetable garden near the monastery kitchen.

Figure 13.25 shows the typical structure of a pea flower. Under normal conditions, pea plants are self-fertilising, that is, pollen from one flower fertilises the ovules of the same flower. However, this monk was carrying out a procedure to prevent self-fertilisation. Using forceps, he carefully removed the stamens from flower buds on one pea plant and dusted pollen that he had collected from another pea plant onto the stigma of the first plant. In doing this, he was artificially crossing the pea plants (see figure 13.26).

Later, the monk wrote about his procedure for an artificial cross as follows: 'For this purpose, the bud is opened before it is perfectly developed, the keel is removed and each stamen carefully extracted by means of forceps, after which the stigma can once be dusted over with the foreign pollen.'

At another time, this monk could be seen in another section of the vegetable garden where he recorded the characteristics of mature pea plants in his notebook. Later, with others assisting him, the monk sat at a table where he shelled peas, sorted them into groups of different colours and shapes and counted the numbers in the various groups.

Who was this quiet monk? He was Gregor Mendel (1822–1884) (see figure 13.27a). Growing up on a farm, the young Mendel would have noticed variation in the offspring of farm animals. Years later in the monastery, Mendel turned his attention to edible pea plants (Pisum sativum) and examined the inheritance of variation in seven different traits of this species (see figure 10.28). He also used other plant species, such as beans, and experimented with bees.
**FIGURE 13.27** (a) Gregor Mendel and (b) the monastery gardens in which he carried out his plant-breeding experiments. Mendel stopped his genetic experiments in 1871 after being elected Abbot in 1868. He died of Bright’s disease.

**FIGURE 13.28** Variations in pea plants used by Mendel in his experiments. Dominant traits are underlined. Which peas can be readily eaten, pods and all? (Modern allele symbols are shown.)

<table>
<thead>
<tr>
<th>Trait</th>
<th>Variations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stem length</td>
<td>tall Le</td>
</tr>
<tr>
<td></td>
<td>short le</td>
</tr>
<tr>
<td>Seed (cotyledon) colour</td>
<td>yellow I</td>
</tr>
<tr>
<td></td>
<td>green i</td>
</tr>
<tr>
<td>Seed (cotyledon) shape</td>
<td>round R</td>
</tr>
<tr>
<td></td>
<td>wrinkled r</td>
</tr>
<tr>
<td>Seed coat colour</td>
<td>grey A</td>
</tr>
<tr>
<td></td>
<td>white a</td>
</tr>
<tr>
<td>Pod texture</td>
<td>inflated V</td>
</tr>
<tr>
<td></td>
<td>constricted v</td>
</tr>
<tr>
<td>Unripe pod colour</td>
<td>green Gp</td>
</tr>
<tr>
<td></td>
<td>yellow gp</td>
</tr>
<tr>
<td>Flower position</td>
<td>axial Fa</td>
</tr>
<tr>
<td></td>
<td>terminal fa</td>
</tr>
</tbody>
</table>

Constricted pods lack a hard inner pod lining so that the seed outlines can be seen (as in snow peas); inflated pods have a tough parchment-like lining.

In the axial arrangement, flowers can arise along the entire length of the stem; in the terminal arrangement, flowers are bunched at the top of the stem.
Why Mendel did it
What inspired Mendel to carry out his experiments is not known as many of his personal papers were burned by the abbot who succeeded him. Fortunately, Mendel’s official papers and the notebooks that held the records of his experimental results were safely stored in the monastery archives.

Just one first-hand account of Mendel exists and it is that of a horticulturalist named Eichling who visited Mendel at the Brno monastery in 1878. Recalling this visit years later in 1942, Eichling wrote that Mendel gave him lunch and showed him the monastery garden. Mendel told Eichling that he had ‘reshaped [the green peas] in height as well as in type of fruit.’ In response to Eichling’s question of how he had done that, Mendel answered: ‘It is just a little trick, but there is a long story connected with it which would take too long to tell.’

How Mendel did it
For 8 years from 1856 to 1864, Mendel carefully carried out breeding experiments with varieties of pea plants. Features of his experimental crosses that led to his success included the following:

• One trait at a time. Mendel initially examined variation in only one trait at a time. He set up crosses between plants that differed in just one trait, such as pod colour. Such crosses are termed monohybrid crosses. After he had recognised the pattern of inheritance of single traits, Mendel studied crosses of plants differing in two traits. Such crosses are termed dihybrid crosses. In contrast, other plant breeders tried to study variation in many traits at once and were confounded by all the variation that they observed in the offspring.
• Known history of parents. For his starting point (the P generation), Mendel used plants that were known to be pure breeding.
• Recording parentage. Unlike other plant breeders, Mendel kept careful records of the parents of every offspring.
• Counting offspring. Unlike other plant breeders, Mendel counted the appearance and numbers of different kinds of offspring produced in each generation. He also repeated his experimental crosses, obtaining large numbers of offspring so that average ratios could be determined.

Oodles of peas
Mendel’s choice of pea plants for his breeding experiments meant that he was able to obtain relatively large numbers of offspring from even a single cross. Every pea in a pod on a pea plant is a single offspring and each pea ‘baby’ will grow into a mature plant (see figure 13.29).

In all, Mendel produced thousands of offspring from his pea plant crosses over eight years. Large numbers of offspring allow regularities to be recognised and valid averages to be identified. If only small numbers of offspring are obtained, regularities may not be seen and averages may be biased by chance events. Numbers do matter! For example, imagine that you have four coins and that one of them is double-headed. Would you be absolutely confident that you could identify the double-headed coin on the basis of the result of tossing each coin just once? What about ten tosses? Likewise, when Mendel was examining various outcomes from his crosses, such as green pods or yellow pods, he obtained large numbers because he wanted to ‘ascertain their statistical relations.’

Mendel’s model of inheritance
Mendel developed a model to explain the patterns of inheritance of the pea variations and to enable predictions to be made about the outcome of crosses (see figure 13.30). Mendel’s model was built on several assumptions:

1. Each trait was controlled by a pair of inherited factors. For example, the trait ‘seed colour’ was assumed to be controlled by a pair of factors, with one producing ‘yellow’ and the other producing ‘green.’
2. For each trait, individual plants had two factors that could be identical or different. Plants with two identical factors (such as ‘long’ and ‘long’) were referred to as pure breeding, while plants with different factors (such as ‘long’ and ‘short’) were called hybrids.

3. Each factor was a discrete particle that retained its identity across generations. This idea challenged the commonly held view that inheritance was a blending process in which factors lost their identity (see figure 13.31).

4. The character that was expressed in the F1 hybrid plants was **dominant**, while the hidden character in the hybrid was **recessive**. For example, green pod colour is dominant and yellow pod colour is recessive.
5. During gamete formation, the members of each pair of factors separated to different gametes, with one factor per gamete. This is the principle of segregation of alleles or Mendel’s first law.

6. In separating, members of one pair of factors behaved independently of members of other pairs of factors. This is the principle of independent assortment or Mendel’s second law.

7. The results of a particular cross were the same, regardless of which plant was used as the male parent and which as the female parent.

Response to Mendel’s results

Mendel’s results and his model of inheritance were first reported at a meeting of a local scientific society in Brno in 1865 and were published in its journal in 1866. More than one hundred copies of this journal were distributed, including two copies that were sent to London. In addition, Mendel sent reprints of his paper to several scientists, including Carl Nägeli, one of the leading European biologists of that time.

Mendel’s results were, however, ignored. Why? The scientific community failed to recognise the significance of Mendel’s findings, perhaps because he was not a professional plant breeder or biologist. Nägeli published a book on heredity in 1884 that made no reference to Mendel. In that book, Nägeli commented on the appearance of a long-haired kitten in the litter from two crossbred shorthaired cats, but could not account for this observation. Nägeli did not realise that a monk from the monastery in Brno could easily have explained the occurrence of this long-haired kitten!

In the years following the publication of Charles Darwin’s *The Origin of Species* in 1859, it is claimed that the attention of scientists moved to evolution and to the differences *between* species. As a result, there was a decline in interest in the work of plant and animal breeders who were concerned with differences *within* species. In this climate, few biologists would have been interested in the plant-breeding experiments of an obscure monk in a monastery in Brno. Mendel’s explanatory model was ignored for more than 30 years.

Mendel’s model was rediscovered in 1900 by three biologists working independently. The biologists were de Vries (1848–1935), a Dutch plant breeder; Correns (1864–1935), an Austrian botanist; and Tschermak (1871–1962), a German botanist. After its rediscovery, biologists in Europe and America demonstrated that Mendel’s model applied to inheritance in many plants and
animals. By the end of the first decade of the twentieth century, Mendelian principles had been found to apply to many organisms, including:

- nettles (*Urtica pilulifera*) — serrated leaf margin dominant to entire
- wheat (*Triticum* sp.) — late ripening dominant to early ripening
- stocks (*Matthiola* sp.) — coloured dominant to white
- mice (*Mus musculus*) — coloured coat dominant to albino
- rabbits (*Oryctolagus cuniculus*) — Angora (long) fur dominant to short fur
- cattle (*Bos taurus*) — polled (hornless) dominant to horned
- poultry (*Gallus gallus*) — brown eggs dominant to white eggs
- sheep (*Ovis aries*) — white wool dominant to black.

The Mendelian model of inheritance was soon universally accepted as the basis of inheritance in plant and animal species.

**Identifying Mendel’s factors**

Early in the twentieth century, a gene was simply ‘something’ that was inferred to be present in a gamete and ‘something’ that acted in an unknown way to produce a particular phenotype. However, nothing was known about the nature of genes or how genes acted to produce particular phenotypes. The writings of biologists show that the nature of genes (or factors) was a mystery:

> Beyond their existence in the gamete and their mode of transmission we make no suggestion as to the nature of these factors.

—Punnett, 1911

> ... the material is termed for convenience a factor or a gene, terms which do not imply any knowledge as to the nature of the substance causing characters to appear...

—Cutler, 1923

A commonly held, but incorrect, view at that time was that genes were probably made of protein. However, over the first half of the twentieth century, several experiments revealed what genes were made of.

Two biologists played important roles in elucidating (1) that genes were made of a chemical substance and (2) that chemical substance of genes was DNA. They were Frederick Griffith and Oswald Avery.

**The clue from Griffith’s changing bacteria**

Imagine a substance that can change mild-mannered and harmless organisms into disease-causing killer organisms. In 1928, a British biologist, Frederick Griffith (1877–1941), extracted such a substance. Just as pea plants show variations, pneumococci bacteria also show variations (see table 13.3).

<table>
<thead>
<tr>
<th>Strain</th>
<th>Distinctive feature</th>
<th>Disease-causing or not</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘smooth’ type</td>
<td>external capsule present outside cell wall</td>
<td>cause pneumonia</td>
</tr>
<tr>
<td>‘rough’ type</td>
<td>external capsule absent</td>
<td>harmless</td>
</tr>
</tbody>
</table>

Griffith carried out experiments using two kinds of pneumococci bacteria (see figure 13.32). He found that:

- injection of living smooth type into mice caused them to die from pneumonia
- injection of living rough type left the mice healthy.

Griffith also killed smooth bacteria by heating and extracted the contents of these dead cells. When mice received an injection of this material, they remained healthy. These results supported the conclusion that pneumonia was caused by living smooth pneumococci bacteria.
Griffith then mixed the contents from dead smooth cells with living rough cells. He injected mice with these treated rough cells and found that they died from pneumonia. When the dead mice were examined, living smooth cells were found, even though no living smooth cells had been injected. How had this happened?

The deadly smooth bacteria found in the mice had formerly been harmless rough bacteria. The harmless rough bacteria had been changed or transformed by ‘something’ in the contents of the smooth cells. This change agent caused the harmless rough bacteria to produce an external capsule and become the deadly smooth type of bacteria. This ‘something’ became known as the transforming factor. It was concluded that the transforming substance was equivalent to the substance of the genetic material itself.

Griffith's experiments demonstrated that genetic material was a chemical substance (see figure 13.33). But, what was it?

**Avery's answer to the puzzle**

In 1943, Oswald Avery (1877–1955) and his co-workers identified the chemical nature of the transforming factor. Avery (see figure 13.34) obtained extracts of dead, smooth pneumococci bacteria and treated this material with various enzymes that could destroy lipids, proteins and carbohydrates. He found that the extracts could still transform harmless rough bacteria to deadly smooth bacteria (see table 13.4). However, when the extracts were treated with enzymes that destroy deoxyribonucleic acid (DNA), the extract could no longer
transform other bacteria. These experiments gave the first clue to the identity of the genetic material and supported the conclusion that genes were made of the chemical compound DNA.

**TABLE 13.4** Outcomes of treatments of transforming factor from dead smooth pneumococci bacteria with various agents. Which treatment destroyed the transforming factor?

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>protein-destroying enzymes</td>
<td>ability to transform rough to smooth remained</td>
</tr>
<tr>
<td>lipid-destroying enzymes</td>
<td>ability to transform rough to smooth remained</td>
</tr>
<tr>
<td>carbohydrate-destroying enzymes</td>
<td>ability to transform rough to smooth remained</td>
</tr>
<tr>
<td>DNA-destroying enzymes</td>
<td>transformation ability destroyed</td>
</tr>
</tbody>
</table>

Later, Avery extracted the contents from smooth bacteria, separated and purified the various components until he had a highly purified sample of the transforming factor. When this was identified, it was found to be DNA.

The momentous discovery of Avery and his co-workers was not accepted immediately by the entire scientific community. Some biologists doubted the validity of Avery’s conclusion. Even books published some years after Avery announced his discovery include cautious statements about the identity of genetic material, such as:

... the present experiments ‘strongly suggest’ rather than prove that genes are pure DNA...  

—1957

New scientific discoveries are not always rapidly accepted by the entire scientific community. If scientists hold strongly competing alternative views, they may not readily accept new findings that disagree with their views.

It is now universally accepted that genes are made of the chemical compound DNA. DNA belongs to the class of chemical substances called nucleic acids. Figure 13.35 shows one bacterial cell (centre) that has been treated to release its genetic material — strands of the nucleic acid DNA.
The impact of Avery's experimental discovery that genes were made of the chemical substance deoxyribonucleic acid (DNA) was far reaching. Contrast the definitions of a gene taken from textbooks published before Avery's discovery with later definitions. A gene was variously defined as:

- a name for the thing in a germ cell that makes the germ cell develop a particular characteristic, such as tallness as opposed to dwarfishness (1911)
- the hypothetical unit in a germ cell that determines the production of a particular characteristic in the individual derived from that germ cell (1921)
- a hypothetical unit in the chromatin of a cell that has a specific influence on certain characteristics (1934)
- a segment of a DNA molecule that can copy itself and pass on to other generations the directions it contains (1966)
- a locatable region of genomic sequence, corresponding to a unit of inheritance, that is associated with regulatory regions, transcribed regions and/or other functional sequence regions (2011).

**KEY IDEAS**

- Mendel carried out carefully controlled experiments in which he focused on the inheritance of variations in one characteristic at a time.
- Mendel developed a model of inheritance that built on a set of assumptions that explained observed results and allowed predictions to be made.
- Mendel's model of inheritance was ignored by the scientific community but was rediscovered in 1900 independently by three biologists.
- After its rediscovery, Mendel's model was soon found to apply to many other species.
- Griffith's discovery of the transforming factor in bacteria provided evidence that the genetic material was a chemical substance.
- Avery's experiments provided the evidence that the genetic material was composed of DNA.

**QUICK CHECK**

7. Identify whether each of the following statements is true or false.
   a. Mendel's model assumed that parental characters behaved as discrete entities or particles that did not blend.
   b. Pea plants normally undergo cross-fertilisation.
   c. Mendel's model of inheritance applies only to plants.
   d. The first human condition shown to behave in accord with Mendel's model was brachydactyly (abnormally short fingers).

8. List two assumptions of Mendel's explanatory model for inheritance.

9. What impact did Mendel's model have on the scientific community at the time it was first reported?

10. What was the significance of Griffith's discovery of a transforming factor?

11. What contribution did Avery make to the debate about the nature of genes?

**Looking at genes**

Genes are the basic units of heredity that transmit information in the form of discrete sequences of bases in DNA to the next generation. Let's look at genes by asking a series of questions.

**How many genes?**

Earlier in this chapter we saw that the total number of protein-coding genes in the human genome is about 21,000. (The most recent estimate is 20,678 protein-coding genes.)
What is the function of genes?

Many genes are protein-coding genes. The coded information present in the base sequence of these genes is translated (expressed) as a polypeptide chain or protein. Each protein-coding gene controls a specific characteristic or trait; for example, Mendel experimented with seven genes, each of which control different traits in pea plants (refer to figure 13.27).

In addition to these protein-coding genes, other genes present in the human genome are expressed only as RNA, such as ribosomal RNA (rRNA) and transfer RNA (tRNA). Because these genes do not code for proteins they are denoted as ncRNA genes (where the ‘nc’ means non-coding for protein).

How are genes named?

Genes can be named after the functions they control, such as ‘the gene controlling Rhesus blood type.’ For convenience, genes are given shorter identifiers. Just as a young boy named Benjamin James McDonald is called Ben, so scientists have developed a shorthand scheme for naming genes. In this scheme, genes are usually given a name consisting of a group of up to five characters (capital letters or numbers), with the first character always being a letter.

Examples of human genes include the ABO gene that controls ABO blood type, the BRCA1 gene that controls DNA repair and is related to increased risk of various forms of cancer and the CFTR gene that controls the chloride ion channel protein in the plasma membrane.

Where are genes located in cells?

Almost all human genes are present in the DNA of the cell nucleus (see figure 13.36). Genes are transmitted to the next generation in gametes — eggs and sperm — that are produced by the process of meiosis (refer to chapter 11), so that each gamete contains the haploid human genome. (In chapter 14, we will explore the location of genes more precisely on chromosomes.)
In addition to the genes in the nuclear DNA, a small number of genes are present in the DNA of mitochondria (mtDNA). The mtDNA consists of 16,568 base pairs (see figure 13.37). This DNA includes 13 protein-coding genes that are involved in cellular respiration — not surprising, since key steps of cellular respiration occur in the mitochondria. In addition, mtDNA contains 24 ncRNA genes that code for both rRNA and tRNA.

**FIGURE 13.37** Map of the double-stranded circular molecule of human mitochondrial DNA (mtDNA) showing the various groups of genes coded for by the mtDNA.

**How do genes differ?**

Every genetic instruction can be shown as a sequence of bases (written as As, Ts, Cs and Gs) in the nucleotides that form the DNA of the gene. The genetic material of all organisms is DNA and the structure of DNA is identical, regardless of whether it comes from wheat, jellyfish, wombats, bacteria, insects or people. In all organisms, genes are built of the same alphabet of four letters, namely the A, T, C and G of the bases in the nucleotide sub-units of DNA. So, the genetic instruction kit of the white shark and that of an oak tree and that of a person consists of thousands of different instructions, each consisting of DNA with different base sequences.

Each of the 21,000 or so protein-coding genes in the human genome carries a different genetic instruction. One gene may have the code for the instruction ‘make the nail protein keratin’; another gene may have the code for the instruction ‘make the enzyme cytochrome oxidase’, yet another gene may have the code for the instruction ‘make the alpha chain of haemoglobin, the oxygen-carrying molecule of red blood cells’. These different genetic instructions differ in the base sequences of the genes concerned.

Table 13.5 shows the base sequence from segments of different genes from various organisms: a mallard duck (*Anas platyrhynchos*), a *Bacillus* bacterium, a corn plant (*Zea mays*) and a human being (*Homo sapiens*). Could you pick the human gene? Using logic only, it is not possible; the gene sequences share many similarities because the genetic language of all living organisms is written in the same language that is based on an ‘alphabet’ of four letters (A, T, C and G) that denote the bases in the nucleotide sub-units of DNA.
**How much DNA is in a gene?**

An average gene consists of about 3000 base pairs. Genes, however, vary markedly in size. The longest human gene is the **DMD** gene that encodes the muscle protein dystrophin and is 2,220,223 nucleotides long. An error in this gene is the cause of the inherited disorder Duchenne muscular dystrophy. Among the shortest human genes is a gene that encodes a histone protein and it consists of about 500 nucleotides.

**KEY IDEAS**

- The information in DNA is present as a sequence of bases.
- Different genes consist of DNA with different bases sequences.
- Genes may be grouped into protein-coding genes and noncoding RNA genes.
- DNA is located almost exclusively in the nucleus.
- Genes average about 3000 nucleotides in length, but considerable variation about that average exists.

**QUICK CHECK**

12. Identify whether each of the following statements is true or false.
   a. The **ABO** gene and the **CFTR** gene are both human genes and so could have identical base sequences.
   b. The genetic material of plants differs from that of animals.
   c. Noncoding RNA genes are expressed as RNA products, not as proteins.
   d. The majority of genes in mtDNA are ncRNA genes.

13. If you saw the base sequence of part of a gene, could you identify if it came from a dog or from a flea? Briefly explain.

14. If the human **ABO** gene and the **CFTR** gene were compared:
   a. in what way would they be similar
   b. in what way would they differ?

**Alleles: particular forms of a gene**

A gene that controls one function can exist in different forms or variants that are called **alleles** of that gene. One gene can have several alleles and each is identified in terms of its specific action. The **ABO** gene has three different alleles. Alleles of one gene are commonly represented by variations of one letter of the alphabet (see table 13.6).
### Table 13.6 Common alleles of selected genes, represented by variations of one letter of the alphabet

<table>
<thead>
<tr>
<th>Gene and its function</th>
<th>Chromosomal location</th>
<th>Common alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABO</strong> encode A B H antigens</td>
<td>9</td>
<td>I&lt;sup&gt;A&lt;/sup&gt; produces antigen A, I&lt;sup&gt;B&lt;/sup&gt; produces antigen B, i produces neither antigen</td>
</tr>
<tr>
<td><strong>CFTR</strong> encode trans-membrane conductance regulator</td>
<td>7</td>
<td>C normal secretions, c abnormal secretions (cystic fibrosis)</td>
</tr>
<tr>
<td><strong>CBD</strong> encode green-sensitive pigment</td>
<td>X</td>
<td>V produces green sensitive pigment, v lacks pigment (colour vision defect)</td>
</tr>
<tr>
<td><strong>DMD</strong> encode muscle protein (dystrophin)</td>
<td>X</td>
<td>M produces normal muscle protein, m produces abnormal muscle protein</td>
</tr>
<tr>
<td><strong>EL1</strong> encode red blood cell membrane protein</td>
<td>1</td>
<td>E elliptical red blood cells, e usual shape</td>
</tr>
<tr>
<td><strong>F8</strong> encode factor VIII blood clotting protein</td>
<td>X</td>
<td>H produces factor VIII, h no factor VIII (haemophilia)</td>
</tr>
<tr>
<td><strong>HBB</strong> encode beta chains of haemoglobin</td>
<td>11</td>
<td>T produces beta chains, t beta chains missing (thalassaemia)</td>
</tr>
<tr>
<td><strong>LDLR</strong> encode low density lipoprotein receptor</td>
<td>19</td>
<td>B abnormally high cholesterol level, b normal range</td>
</tr>
<tr>
<td><strong>PHA</strong> encode phe hydroxylase enzyme</td>
<td>12</td>
<td>P produces normal enzyme, p enzyme absent (PKU)</td>
</tr>
<tr>
<td><strong>RHD</strong> encode Rhesus D antigen</td>
<td>1</td>
<td>D Rhesus positive, d Rhesus negative</td>
</tr>
<tr>
<td><strong>TYR</strong> encode tyrosinase enzyme</td>
<td>11</td>
<td>A produces tyrosinase enzyme, a no enzyme (albinism) (see figure 13.38)</td>
</tr>
</tbody>
</table>

**How many alleles?**

In terms of the phenotypic expression of genes, the common situation for most genes is that they have two alleles, as indicated in Table 13.6. This is not always the case, as can be seen for the **ABO** gene, which has three common alleles, I<sup>A</sup>, I<sup>B</sup> and i. When three or more alleles exist for a gene, the gene is said to have multiple alleles. Table 13.7 shows some examples of multiple alleles. Some chromosomal regions, known as short tandem repeats (STRs), are hypervariable and each has multiple alleles, ranging from 10 to 40 different alleles for the various STRs.
### Table 13.7 Multiple alleles of selected genes in various organisms. Not every allele is shown in each case, for example, there are more than 12 multiple alleles for the *Drosophila* eye colour gene in this table.

<table>
<thead>
<tr>
<th>Gene function</th>
<th>Multiple alleles and their action</th>
</tr>
</thead>
<tbody>
<tr>
<td>controls human ABO blood type</td>
<td>$I^A$ antigen A present $I^B$ antigen B present $i$ neither antigen present</td>
</tr>
<tr>
<td>controls white spotting in dogs</td>
<td>$S$ white spots absent $s_i$ Irish spotting (as in collies) (see figure 13.39a) $sp$ piebald spotting (as in fox terriers) (see figure 13.39b) $se$ produces extreme spotting (as in Samoyeds and Maltese terriers)</td>
</tr>
<tr>
<td>controls pigment levels in cats</td>
<td>$C$ intense pigment (as in a black cat, see figure 13.40a) $cb$ Burmese dilution (see figure 13.40b) $cs$ Siamese dilution (see figure 13.40c)</td>
</tr>
<tr>
<td>controls colour intensity in rabbits</td>
<td>$C$ intense pigment (as in solid black) $cch$ Chinchilla colouring (white fur with black tips) $ch$ Himalayan colouring (colour on ears, nose, feet and tail only) $c$ albino (white fur and pink eyes)</td>
</tr>
<tr>
<td>controls white markings in cattle</td>
<td>$S$ white band around middle (as in Galloways) $sh$ Hereford type spotting $sc$ solid colour with no spots (as in Belmont Reds) $s$ Friesian type spotting</td>
</tr>
<tr>
<td>controls eye colour in fruit fly (<em>Drosophila</em> sp.)</td>
<td>$w+$ red eye $wa$ apricot $wh$ honey $wp$ pearl $wi$ ivory $w$ white</td>
</tr>
</tbody>
</table>
Alleles of different genes may also be seen in plants. Figure 13.28 on page xxx shows the pea plants (*Pisum sativum*) that were the focus of Mendel’s experiments. Note that the term ‘trait’ in this figure was appropriate in the time of Mendel but today the term used would be gene. Likewise, the ‘variations’ would today be referred to as alleles.

Table 13.8 identifies some alleles of genes in some other plant species. Note that the gene controls a general function, such as flower colour, but its alleles produce specific expressions of that function, such as purple and white. Figure 13.41a shows the smooth and wrinkled kernel textures in corn (*Zea mays*). These differences in texture are due to the difference in sugar levels in the kernels. Kernels with a high sugar content take up more water and swell more than kernels with a high starch content. As the kernels dry out, the greater loss of water from the sugary kernels causes them to become wrinkled.

**TABLE 13.8 Alleles of some genes in plants**

<table>
<thead>
<tr>
<th>Gene function</th>
<th>Alleles and their action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flower colour in delphinium</td>
<td>P purple</td>
</tr>
<tr>
<td></td>
<td>p white</td>
</tr>
<tr>
<td>Kernel colour in corn</td>
<td>Pr purple</td>
</tr>
<tr>
<td></td>
<td>pr yellow</td>
</tr>
<tr>
<td>Kernel texture in corn</td>
<td>Su smooth (starchy)</td>
</tr>
<tr>
<td></td>
<td>su wrinkled (sugary)</td>
</tr>
<tr>
<td>Mature fruit colour in capsicum</td>
<td>R red</td>
</tr>
<tr>
<td></td>
<td>r yellow</td>
</tr>
</tbody>
</table>
Figure 13.41b shows the purple and yellow kernels in corn, while figure 13.42 shows some of the mature fruit colours in capsicum.

**FIGURE 13.41** Corn showing (a) wrinkled (starchy) and smooth (sugary) kernels and (b) purple and yellow kernels.

**FIGURE 13.42** Various colours in mature fruit of capsicum (*Capsicum annum*).

---

**KEY IDEAS**

- One gene can exist in a number of different forms called alleles.
- A gene controls a general function and its alleles produce specific expressions of that function.
- A gene controls a general trait and its alleles act to produce specific expressions of the trait.
- Each allele of a gene can be identified by its specific action.
- Some genes can be seen to have two alleles but in other cases multiple alleles exist.

---

**QUICK CHECK**

15 Fill in the following gaps with the term gene or allele.
   a. The _____ that produces short fur length in cats
   b. The _____ controlling fur length in cats

16 Refer to table 13.6. What is the phenotype and sex of a person with the genotype *PP*, *MM*?

17 A cob of corn consists of many individual cobs that are the offspring of a pair of parents. In one particular cob it is seen that some of the cobs are smooth and swollen but a smaller number are wrinkled and shrunken. This variation is due to the action of a single gene with two alleles. Using table 13.8, suggest which alleles of this gene might give rise to these two phenotypes.
As part of the Deciphering Developmental Disorders (DDD) project, a survey was carried out on almost 7000 people to identify whether or not they would wish to be informed on a range of matters in the event that a researcher was to make a chance discovery or a so-called ‘incidental finding’ relating to their genomes. Figure 13.43 shows the results of this survey, published in April 2015.

The survey posed questions related to the following:
1. **Life-threat, can be prevented** — conditions that are life-threatening and can be prevented
2. **Carrier** — tells if a person is a carrier of a condition that could be relevant to their children
3. **Medications** — demonstrates how a person might respond to different medications or drugs (e.g. statins, anti-depressants)
4. **Useful later in life** — information that is not immediately relevant but could be useful later in life (e.g. relating to a very late onset cancer or predisposition to strokes)
5. **Ancestry** — tells about the person’s ancestry
6. **Life-threat, cannot be prevented** — conditions that are life-threatening and cannot be prevented
7. **Not serious health importance** — is not likely to be of serious health importance (e.g. mild eyesight problems)
8. **Uncertain** — information that is uncertain and cannot be interpreted at the moment

1. Identify the four groups of participants who participated in this survey.
2. What was the general view in regard to receiving information about conditions that are life-threatening and can be prevented?
3. Did attitudes change if these life-threatening conditions could not be prevented?
4. What about conditions that are serious (but not life-threatening) and cannot be prevented?
5. What about conditions that are serious (but not life-threatening) and can be prevented?
6. Which three categories had the highest proportions of participants wanting to receive that information?
7. Which two categories had the lowest proportions of participants wanting to receive that information?
8. Some categories show very little difference in the responses between the different groups of participants. Which groups of participant show the most marked difference in their response to the various situations?
9. Which might be your own responses to incidental information of various kinds arising from your genome?

**Weblink**
Welcome Trust Sanger Institute data

**Figure 13.43** Results of the DDD project survey show that most people would prefer to know about their genetic information.
Chapter review

Key words
adenine (A)  dominant  ‘junk’ DNA  re-association
alleles  double helix  met abolomics  recessive
base pairs  enhancers  microbiome  retrovirus
deoxyribose  gene  microbiomics  ribosomal RNA (rRNA)
dihybrid  gene duplication  monohybrid  RNA (ribonucleic acid)
dissociation  genome  ncRNA gene  short tandem repeats (STRs)
(DNA) deoxyribonucleic acid  guanine (G)  single nucleotide polymorphism
transscriptomics
DNA sequencers  cytosine (C)  histone  thymine (T)
dna
hybridisation
DNA adenosine
hybridisation
base sequence
Chargaff’s rule
comparative genomics
complementary base pairs
cytosine (C)
deoxyribose
dihybrid
dissociation
(DNA) deoxyribonucleic acid
DNA sequencers

Questions
1 Making connections ➔ Use at least eight of the key words above to prepare a concept map on the Human Genome Project. You may add other concepts that you wish.
2 Using the web to access information ➔ Use the weblink History of genomics in your eBookPLUS and answer the following questions.

(a) Who discovered the four different bases (A, T, C and G) that are present in nucleotides, the building blocks of DNA? When? (Uracil is found in another nucleic acid, ribonucleic acid (RNA).)
(b) Who was the first person to identify DNA? When? What did he call it?

3 Developing explanations ➔ Suggest explanations for the following:
(a) The statement that ‘Genes are made of DNA’ is absent from textbooks published before the mid-1940s.

(b) In a DNA double helix, the number of adenine molecules can be used to predict the number of thymine molecules.
(c) In a single strand of DNA, the number of adenine molecules cannot be used to predict the number of thymine molecules.
(d) In a DNA double helix, the ratio (A + C)/(T + G) is equal to 1.

4 Evaluating information ➔ Refer to pages xxx–x relating to the discovery of the double helix structure for DNA. Using Watson and Crick as examples of scientists, identify the following statements as true or false and briefly explain your choice.

(a) Scientists might spend more time planning experiments than doing them.
(b) Scientists start investigations without reference to the work of other scientists.
(c) Discoveries occur only in laboratories as a result of experiments.
(d) Unplanned inspiration can play a role in scientific discoveries.

5 Demonstrating knowledge and understanding ➔

(a) What is Chargaff’s rule?
(b) After demonstrating that the proportions of A and T and of C and G were about equal in the DNA from many organisms, Chargaff and his co-workers wrote:

A comparison of the molar proportions reveals certain striking, but perhaps meaningless, regularities.

—1949

Does this fact suggest that scientists will be aware of the importance of facts or regularities that they discover? Explain.

(c) The 1953 paper describing Watson and Crick’s 3D model of DNA appeared in the journal Nature, vol. 171, p. 737. Visit a library or search the internet and locate this article. Does this article suggest that major discoveries can be explained only in long articles that are difficult to understand?
6 Applying principles ➔ Consider part of the DNA coding strand of a gene from a flowering gum that includes the base sequence below.

1  5  10  15
ATG AGT CGC GAT CGT

(a) Write the DNA sequence in the corresponding portion of the complementary strand.

7 Interpreting data ➔ Samples of DNA were analysed and the following proportions of bases were found on the nucleotide sub-units:

<table>
<thead>
<tr>
<th>DNA</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>1.2</td>
<td>1.2</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>Sample 2</td>
<td>0.8</td>
<td>0.8</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Sample 3</td>
<td>1.0</td>
<td>0.7</td>
<td>0.7</td>
<td>1.0</td>
</tr>
</tbody>
</table>

(a) Identify which, if any, of the samples could be double helical DNA.
(b) Identify which, if any, could not be double helical DNA.
(c) Briefly explain the reason for your answers in parts (a) and (b).

8 Demonstrating understanding ➔ The following is part of the nucleotide sequence of one chain in a DNA double helix:


(a) Identify the base sequence of the complementary strand.
(b) What holds the two chains of DNA together in a double helix?

9 Suggest explanations in biological terms for the following observations:
(a) The first genome to be sequenced was that of the virus phiX174.
(b) Genomic studies indicate that some protein-coding genes have been conserved across many organisms from diverse classes.

10 Making distinctions ➔ Which of the following entries refer to a gene? Which refer to the particular alleles of a gene?
(a) the . . . that controls eye colour in humans
(b) the . . . that produces blue eye colour in humans
(c) the . . . that produces non-blue eye colour.

11 Discussion question ➔ The following statement appeared in November 2014:

Ultimately, the goal is for all of us to have our genomes sequenced and available as a medical reference for our clinical care.


Another statement is as follows:

It may be decades before interactions between genes, behavior and environment are understood well enough to provide substantial utility to warrant individualized recommendations based on genomic profiles. Furthermore, behavior change interventions that take advantage of some of the more unique aspects of genetic risk information are in their infancy.

—Kurt D Christensen and Robert C Green, ‘How could disclosing incidental information from whole-genome sequencing affect patient behavior?’, vol. 10, no. 4, 10.2217/pme.13.24

Discuss with your classmates your thoughts about these statements. Identify any positive perspectives that you as a group identify in favour of the statements and any negative standpoints that are identified against the statements.