Chapter 1
What is genetic variation?

Have you ever seen someone who looks and sounds exactly like you? Have you ever seen your 'spitting image'? Unless you are one of a pair of monozygotic twins (twins produced by the division of a single egg), you will not have done so. It is commonly accepted that all humans are different – indeed all humans there have ever been were unique and were different from all humans living today. If this is true for *Homo sapiens*, is it also true for other sexually reproducing organisms? The answer is yes. Every salmon (*Salmo salar*) is different from every other salmon that has ever lived. Every mussel (*Mytilus edulis*) is different from every other mussel that has ever lived. This uniqueness of individuals within a species is the consequence of two factors: one is deoxyribose nucleic acid (DNA) and the other is sexual reproduction. These two factors produce and maintain the genetic diversity within a species, and an understanding of this is fundamental to our ability to sustainably exploit species of plants and animals. Without sex, the reproductive process can produce genetically identical individuals, as in plants where vegetative reproduction is common, and there are biotechnological methods that can produce clones (see Chapter 7).

Deoxyribose nucleic acid

The discovery by Watson and Crick of the structure of DNA in 1953 was a landmark in our understanding of how genetic information passes from generation to generation and how it codes for the amino acids of proteins. In the half century since then, the fields of molecular biology and genetics have become inextricably linked and developments, particularly over the past 25 years, have opened up the potential of DNA biotechnology and genomics.

The structure of DNA enables it to carry the information for a cell to reproduce itself. It is a polymeric molecule, that is, made up of a chain of subunits, consisting of chains of nucleotide monomers. Each nucleotide contains a base, along with a sugar (deoxyribose) and a phosphate group (Figure 1.1). There are four individual bases: adenine, guanine, thymine and cytosine, and they are usually referred to by their first letter abbreviations, A, G, T and C. Two of the bases, A and G, have a double-ring structure and are known as purines. The other two bases, T and C, are pyrimidines with a single carbon–nitrogen ring.

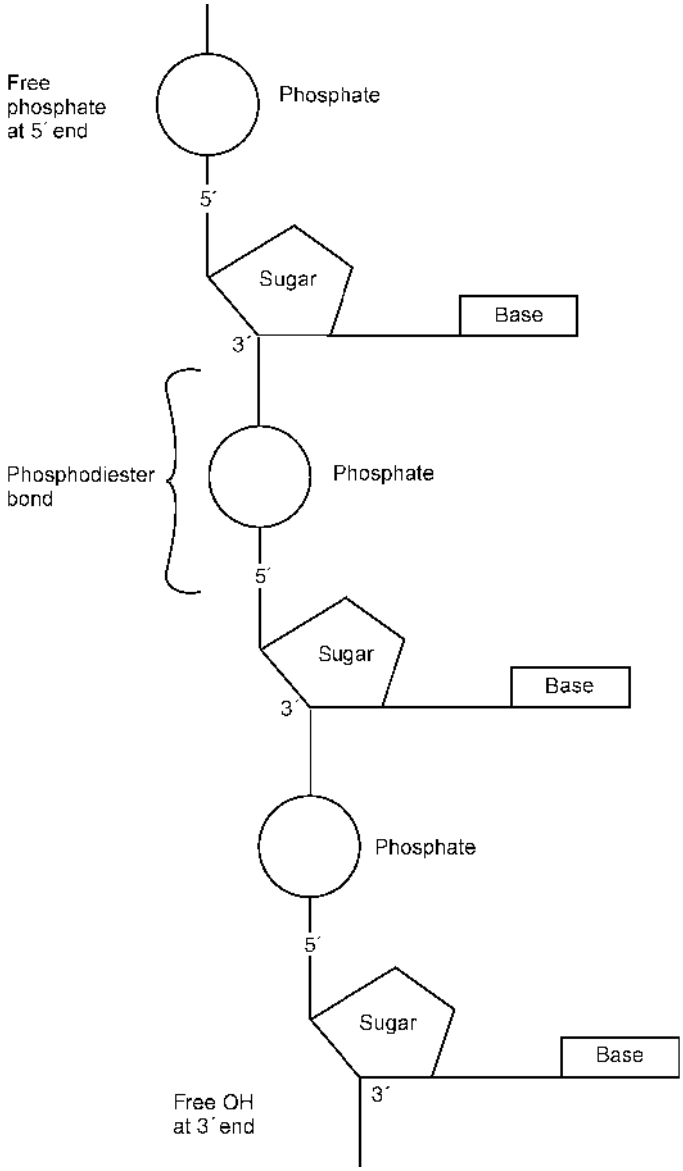

Figure 1.1 The structure of DNA. Each nucleotide consists of a sugar, a phosphate and a base. Nucleotides are joined by a phosphodiester bond between the 5′ of one ribose sugar and the 3′ of the next. The chain of nucleotides therefore has a 3′ and a 5′ end.

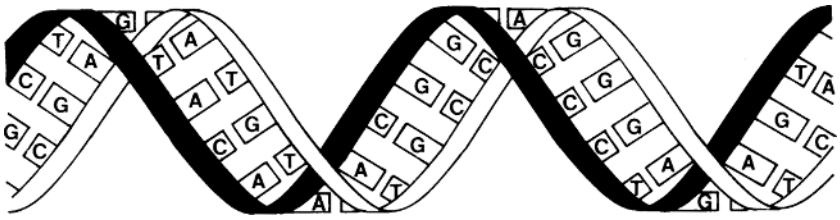

Figure 1.2 The structure of DNA. Two polynucleotide strands are wrapped around each other in the form of a double helix. Complementary base pairing occurs between the two strands such that guanine (G) always bonds with cytosine (C), and adenine (A) always bonds with thymine (T). (Modified from Utter, F., Aebersold, P. & Winans, G. (1987) Interpreting genetic variation detected by electrophoresis. In: *Population Genetics and Fishery Management* (eds N. Ryman & F. Utter), pp. 21–45. University of Washington, Washington, DC, USA, with permission from Washington Sea Grant Program, University of Washington.)

Each nucleotide is a single unit that joins with neighbouring nucleotides in a linear fashion to make up a polynucleotide chain. Particular carbon atoms in the 5-carbon structure of deoxyribose are referred to by numbers, 1′ (pronounced prime) to 5′. The link between nucleotides is formed when the 5′ of one bonds to the 3′ of the next via a phosphodiester bond (Figure 1.1). It is the sequence of the four bases in a polynucleotide chain, which acts as the code for genetic information.

The complete DNA molecule actually consists of two polynucleotide chains, or *strands*, wrapped around each other in the form of a double helix. The sugar + phosphate backbones are at the outside of the molecule while the bases point towards the middle of the structure; the two strands of the molecule run in opposite directions (Figure 1.2).

The functional beauty of the DNA molecule is a result of complementary base pairing where G can only bond with C, and A can only bond with T, at the middle of the molecule. It means that the two strands are complementary such that the base sequence of one strand predicts and determines the base sequence of the other strand. Because one strand predicts the other, it can be used to replicate the sequence.

The replication process produces daughter molecules, each of which has one parental strand and one copied strand. This is called semi-conservative replication. Replication of DNA takes place every time a cell divides. The cell's entire DNA is progressively unwound, revealing short single-stranded regions which can be copied by DNA polymerase enzymes. Unwinding does not begin at the ends of the molecule, but at points called replication origins, and it then proceeds from these points along the DNA. The new strands of DNA being synthesised during replication are always synthesised in the 5′ to 3′ direction. This means that as the original strands separate, one new strand can be continuously synthesised against its copy strand (the leading strand), while the other has to be synthesised intermittently in short lengths as enough copy strand (the lagging strand) becomes available (Figure 1.3).

Considering the enormous numbers of bases and coded information in the DNA of a cell, replication needs to be extremely accurate. Even a very small incidence of mistakes in copying would result in the loss of important genetic information within

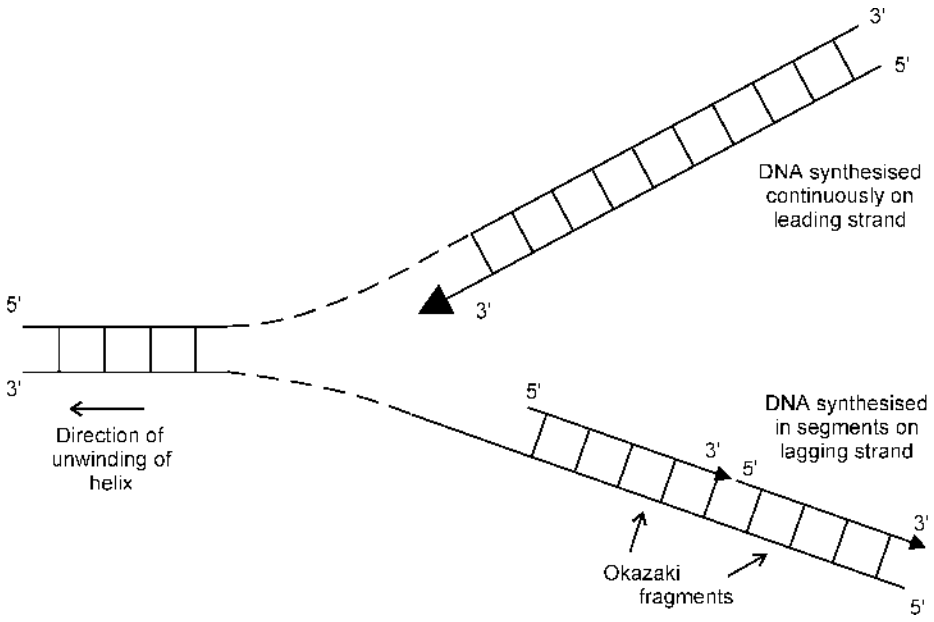

Figure 1.3 Replication of DNA. As the DNA double helix unwinds, new DNA is synthesised continuously in a 5′ to 3′ direction on the leading strand and in 5′ to 3′ directed segments (Okazaki fragments) on the lagging strand.

a few cell divisions. However, during the replication process, various proofreading activities take place and almost all errors are corrected by removing the incorrect base and inserting the correct one. In spite of proofreading, a few errors are inevitable when such high numbers of bases are to be copied and it is estimated that about one in every 3 billion bases is incorrectly inserted. Such errors are called point mutations, and they can also be induced at much higher rates by certain chemicals and radioactivity. Although there are very few of them at each genome replication event, they are nevertheless the fundamental source of variation which fuels the process of evolution. Without such errors, no genetic change at the DNA level would take place, preventing any adaptation of organisms in a changing environment, but with too many errors, daughter cells would too often be non-viable and the organism carrying that DNA would soon become extinct.

Functional sequences, i.e. coding DNA, represent only a small fraction of the total genome, for example around 3% in humans. The rest is made up of what has been called ‘junk DNA’. Whether all of it is really ‘junk’ is not known, and several hypotheses have been proposed to explain how junk DNA might have a function, notably in terms of regulation of gene expression. Some of this junk DNA consists of pseudogenes, genes that for some reason or another have become non-functional. Yet other parts of non-coding DNA consist of dispersed or clustered repeated sequences of varying length, from 1 base pair (bp) to thousands of bases (kilobases, kb) in length, known as ‘repetitive elements’. Junk DNA also includes transposons that encode proteins with no clear value

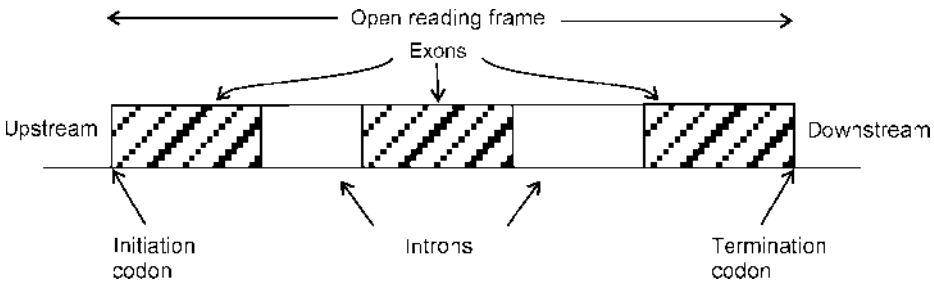

Figure 1.4 Generalised structure of a gene. The open reading frame for a gene begins with an upstream initiation codon and ends with a downstream termination codon. Many genes have a region, or regions, of non-coding DNA within them. These introns are spliced out of the mRNA during transcription so that only the codons within the exons are translated into amino acids.

to their host genome. The dispersed repeated sequences occur as copies spread across the genome and can be categorised as long or short interspersed nuclear elements, long terminal repeats and DNA transposons. The clustered repeated sequences, where the repeated sequence occurs in tandem copies, are classed as satellites, minisatellites or microsatellites, depending on the length of the repeat unit, and these have turned out to be useful genetic markers, as will be explained in later chapters. Between them, these repeated elements can constitute up to 40% of the genome.

A gene is a unit of information which is held as a code in a discreet segment of DNA. This code specifies the amino acid sequence of a protein or the sequence of an RNA molecule. Scientists were surprised to discover quite early on that the sequence information for a single gene was often not continuous along the DNA, but was interspersed with pieces of non-coding sequence. The coding parts of a gene sequence are exons, and the non-coding parts are introns (Figure 1.4). Before a gene can be expressed, the DNA that encodes it has to be transcribed into ribose nucleic acid (RNA).

Ribose nucleic acid

The structure of RNA is similar to that of DNA except that (a) the sugar is ribose instead of deoxyribose, (b) in the place of thymine, a similarly structured base called uracil (U) is present and (c) the molecule consists of only a single polynucleotide strand. RNA molecules are produced by the process of transcription of the linear sequence of bases in DNA and are then used in the translation of that sequence into a chain of amino acids that go to make up a protein. The type of RNA transcribed from the sequence is called messenger RNA (mRNA), and translation of this sequence into a string of amino acids is undertaken by ribosomal RNA (rRNA) and transfer RNA (tRNA) molecules.

During transcription of the DNA, an RNA copy is made of one of the strands of DNA (Figure 1.5). The two strands of DNA are called the template strand and the non-template strand. Other names for the non-template strand are the sense (+) strand or the coding strand. The RNA is synthesised by RNA polymerase enzymes using the

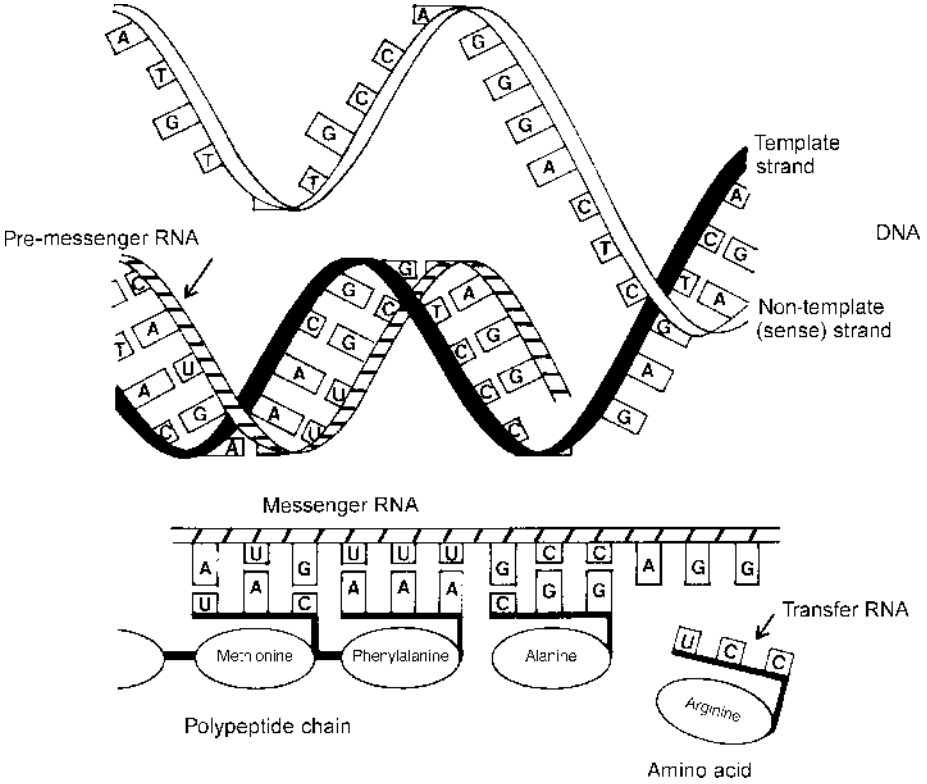

Figure 1.5 Transcription and translation of DNA. Introns are removed from pre-messenger RNA before translation takes place. The polypeptide chain is formed from amino acids coded for by the mRNA and brought together by tRNA. (Modified from Utter, F., Aebersold, P. & Winans, G. (1987) Interpreting genetic variation detected by electrophoresis. In: *Population Genetics and Fishery Management* (eds N. Ryman & F. Utter), pp. 21–45. University of Washington, Washington, DC, USA, with permission from Washington Sea Grant Program, University of Washington.)

template strand and is therefore a copy of the non-template (sense or coding) strand of DNA. Because this RNA is a direct copy of the DNA, it contains both the coding (exons) and the non-coding sequences (introns) of the gene. Introns are removed from this pre-messenger RNA, and the subsequent molecule is the final mRNA. The mRNA molecules are transported from the nucleus into the cytoplasm where the message is translated into a sequence of amino acids by rRNA in bodies known as ribosomes. Amino acids are brought to the ribosomes by tRNA molecules, each specifying a particular amino acid (Figure 1.5), and synthesised, in the presence of rRNA, into a linear sequence.

The detailed mechanics and biochemistry of the processes of transcription and translation are outside the scope of this book, but can be found in most standard genetic texts. The term ‘gene expression’ encompasses both of these processes and their various controlling steps. ‘Transcriptomics’ is the study of the transcriptome, the set of RNA

transcripts produced by the genome of a cell, an organ or a whole organism at any one time. For the purposes of this book, the reader need only appreciate the key concept that a coding sequence of bases in DNA leads, by a direct copying process involving RNA, to the production of a sequence of amino acids, the building blocks of proteins. This is what has been called the 'central dogma': information is transferred from DNA to RNA to protein. Some exceptions to the central dogma occur, such as retroviruses and prions, but they are outside the scope of this book.

What is the genetic code?

How are the four bases (A, C, G and T) in DNA organised to provide an unambiguous code for the 20 amino acids present in proteins? The 'words' of the code consist of three bases. There are $4^3 = 64$ possible combinations of the four bases into a triplet code, and it is these 64 triplet codons which define the 20 amino acids. Because there are more than 20 codons, the genetic code has some redundancy – most amino acids are coded for by more than one codon. The codons are written using the symbol U, for uracil (in mRNA), rather than T, for thymine (in DNA). Three codons (UAA, UAG and UGA) do not encode amino acids but act as signals for protein synthesis to stop and are called termination codons or stop codons. The triplet AUG codes for methionine (formyl methionine in bacteria and mitochondria) and is the signal for protein synthesis to start. It is thus the initiation codon which sets the reading frame. The amino acid sequence of all proteins therefore starts with methionine, but this is sometimes removed later. Details of the amino acids encoded by the various codons are given in Table 1.1. Note that the redundancy of the code is not random. In particular, the first two bases of the codons for an amino acid are usually the same. It is generally only the third base which varies. Different organisms often show particular preferences for one of the several codons that encode the same given amino acid and this is called codon bias. How these preferences arise within a genome is a debated question.

Protein structure

Proteins have many tasks. Some form the structure of tissues, others – the enzymes – act as specific catalysts of biochemical reactions, and yet other proteins, such as hormones, have a regulatory function. By their very nature, proteins are bound to be highly complex molecules, but it is possible to categorise their structure into four basic levels. The primary structure of a protein is the linear sequence of the chain of amino acids (the polypeptide chain) and this, as we have seen already, is directly related to the sequence of bases in the DNA which codes for it. Although most amino acids are pH neutral, two are negatively charged and two positively charged. In addition, some are hydrophilic (attracted to water) and others hydrophobic (repelled by water). Thus, secondary structure of a protein is based on characteristic patterns produced by the properties and interactions of particular types of amino acids within the chain. One such secondary structure

Table 1.1 The genetic code showing the amino acids coded by the 64 triplet combinations of the four bases. The bases down the left-hand side represent the first position in the reading frame, the bases along the top indicate the second position and the bases down the right-hand side show the third position.

	Second base				
First base	**U**	**C**	**A**	**G**	**Third base**
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop	Stop	A
	Leu	Ser	Stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Ala, Alanine; Arg, Arginine; Asn, Asparagine; Asp, Aspartic acid; Cys, Cysteine; Glu, Glutamic acid; Gln, Glutamine; Gly, Glycine; His, Histidine; Ile, Isoleucine; Leu, Leucine; Lys, Lysine; Met, Methionine; Phe, Phenylanaline; Pro, Proline; Ser, Serine; Thr, Threonine; Trp, Tryptophan; Tyr, Tyrosine; Val, Valine.

is an alpha-helix, another is a pleated sheet. The tertiary structure is dependent on how these secondary structures become folded in three dimensions. Therefore the DNA code, through the linear relationship of the various amino acids, dictates both the secondary and tertiary structures. This is an important point because it reveals that point mutations in the DNA coding for a particular protein can have far-reaching consequences on the final size, shape and overall charge of that protein.

Many proteins are composed of two or more polypeptide chains (subunits), and the subunits making up a protein may be identical or they may be different. The generic name for proteins with more than a single subunit is oligomers. This is the level of quaternary structure of proteins and it enables larger proteins to be produced without requiring a very long gene sequence in the DNA. It also allows greater functionality in proteins by combining different activities within a single molecule. Proteins with a single subunit are called monomers, those with two subunits are dimers and those with four are tetramers. For example, glucose phosphate isomerase, an enzyme involved in the production of energy from the breakdown of carbohydrates, is a dimer, with two subunits coded by the same gene, while haemoglobin, which carries oxygen around in the blood, is a tetrameric molecule consisting of two alpha-globin and two beta-globin chains, each coded by different genes.

So what about chromosomes?

In fish and shellfish, as with all other eukaryote organisms, the DNA molecules in the nucleus are combined with proteins, mainly histones, to make chromosomes. Each chromosome represents a single DNA molecule. Chromosomes are usually only clearly visible and identifiable when cells are dividing, at which time the chromosomes have already divided into daughter chromatids. However, the daughter chromatids retain connection to each other at a position called the centromere, or primary constriction, and this is the last part of the chromosome to divide. The position of the centromere on the chromosome can be central (metacentric), between the centre and one end (submetacentric), very close to one end (acrocentric), or terminal (telocentric). The number of chromosomes, their lengths and the positions of their centromeres are unique to each species and these characters are used as descriptors for the species karyotype (Figure 1.6). Chromosomes themselves mutate and evolve (Box 1.1), and before the advent of allozyme markers, some geneticists spent much of their time squinting down microscopes following the

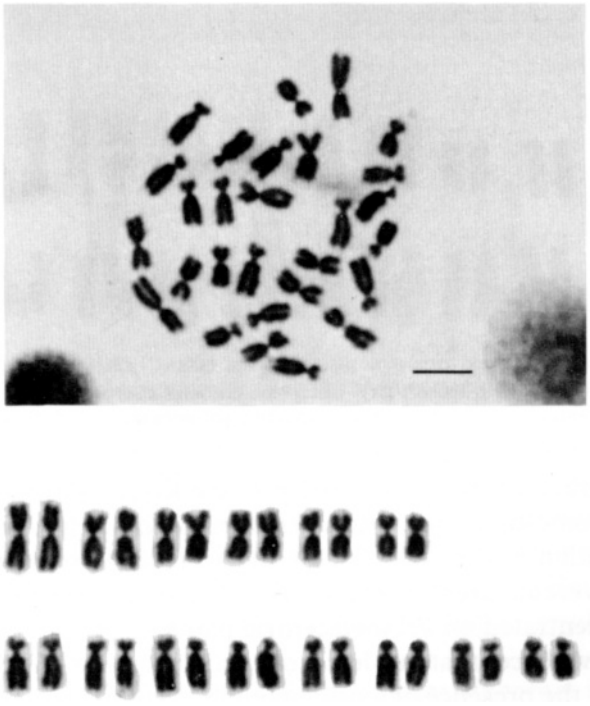

Figure 1.6 Metaphase chromosome spread and karyotype of *Mytilus edulis*. There are six metacentric and eight submetacentric pairs of chromosomes. Haploid number $N = 14$, diploid number $2N = 28$. Scale bar = 5 μm. (Reproduced with permission from Dixon, D.R. & Flavell, N. (1986) A comparative study of the chromosomes of *Mytilus edulis* and *Mytilus galloprovincialis*. *Journal of the Marine Biological Association, UK*, **66**, 219–228, Cambridge University Press.)

inheritance of chromosomal rearrangements. Nowadays, chromosomal variation is assessed for aquaculture and fisheries purposes, mainly in relation to interspecies hybridisations.

Box 1.1 Genetic variation at the level of the chromosomes

Most chromosome rearrangements arise, as do point mutations, as a result of mistakes during the replication of the DNA molecule. Such rearrangements, however, involve long segments of DNA, rather than single bases.

Chromosome deletions occur when the DNA strand breaks but fails to mend. Fragments of chromosome produced in this way that do not contain a centromere (acentric fragments) will be lost during subsequent cell divisions.

Chromosome duplications provide an extra copy of a block of DNA that may contain complete gene sequences. As might be expected, duplications are less harmful than deletions, and when duplications contain complete gene sequences, evolution can occur independently on both the new and the old sequences to produce ultimately divergent roles for the two genes. This is the principal process for the evolution of new genes.

Sometimes a fragment of one chromosome can become exchanged with a fragment of another non-homologous chromosome and such an exchange is called chromosomal translocation.

A chromosomal inversion occurs where a fragment of chromosome breaks off and reattaches to its original position in reversed orientation. The inverted fragment may have contained the centromere (pericentric inversion) or it may not (paracentric inversion).

There is one further type of chromosomal rearrangement which needs a mention. This involves fusion or fission of the centromere. Two telocentric or acrocentric chromosomes may fuse at their centromeres to produce a single bi-armed chromosome and this is called a Robertsonian translocation. Alternatively, a bi-armed chromosome can break at the centromere to produce two telocentric chromosomes. These types of chromosomal rearrangements may explain much of the variation in chromosome number between species.

Chromosome variations are no longer used as markers in population genetic studies, but they play an important role in evolution of species. Before allozyme electrophoresis provided geneticists with access to individual genes for study, many geneticists spent their time looking at the structure of chromosomes during meiosis. The structure of the paired chromosomes (bivalents, Figure 1.7) observed during meiosis reflects chromosomal rearrangements – chromosomes with translocations can only pair up by forming chains or rings, inversions produce loops in the bivalent, etc. Banding patterns on chromosomes shown up by particular stains (e.g. G-banding, produced by Giemsa stain) or by hybridisation of fluoresent probes (e.g. fluorescent in situ hybridisation) can also be used to characterise chromosomes and their rearrangements. Although used in early studies of heritable variation, chromosomal rearrangements are usually deleterious and often result in a non-viable gamete or zygote.

Chromosomal rearrangements may explain much of the variation in chromosome number between species and examination of karyotypes between closely related species is important when considering artificial hybridisations between them. Supernumerary or B chromosomes can be observed in many plant and animal species. They are mainly or entirely heterochromatic and largely non-coding. B chromosomes increase (1) asymmetric chiasma distribution, (2) crossing over and recombination frequencies and (3) unpaired chromosomes at meiosis. B chromosomes have tendency to accumulate in meiotic cell products, resulting in an increase of B number over generations, which is counterbalanced by selection against infertility resulting from increased unpaired chromosomes.

Polyploidy is a condition where individuals have more than two copies of each chromosome. For example, triploids have three sets of chromosomes and tetraploids have four. Polyploidy

occurs naturally in some plants (e.g. wheat, which is hexaploid), and a tetraploidisation event has occurred in the evolutionary history of salmonids. Triploidy can be artificially induced in normally diploid species for aquacultural purposes as will be seen in Chapter 7.

Somatic aneuploidy is an alteration in the number of chromosomes in a proportion of the somatic cells due to abnormalities that arise during mitosis. Although its causal factors are still unknown, the loss of one or more chromosomes from the diploid condition (hypodiploidy) has been reported in several bivalve species and is known to be increased by exposure to pollutants and also to be correlated with growth. A genetic basis of somatic aneuploidy has been suggested in cupped oysters.

Almost all fish and shellfish are diploids (but see Chapter 7). Diploids have two complete sets of DNA instructions, so that each chromosome is just one of a homologous pair. Normal cell division – mitosis – combines division with replication because prior to cell division each chromosome replicates (into two chromatids) and one copy passes into each daughter cell. Diploidy is thus maintained. During the process of sexual reproduction, a specialised cell division called meiosis takes place, which produces daughter cells, the gametes, which have only a single set of chromosomes. Gametes of diploid organisms are therefore haploid. Similarly, gametes of tetraploid organisms are diploid. We will be looking in more detail at the process of meiosis later in this chapter.

If an organism inherits the same version of a gene from both parents, it is said to be homozygous. If the two versions are different, the organism is heterozygous. Each version of a particular gene is called an allele; the two alleles possessed by a diploid organism at each locus (position on the chromosome, plural loci) make up its genotype for that locus.

In many organisms, there is a special pair of chromosomes which defines the sex of their carriers. For example, in the XX–XY system present in humans and some fish, females have a pair of identical sex chromosomes (the X chromosomes) while males have one X chromosome and a reduced size Y chromosome. In birds and also some fish, the ZW sex-determination system is observed: males are the homogametic sex (ZZ), while females are heterogametic (ZW). In that system, the Z chromosome is larger and has more genes, like the X chromosome in the XY system. The other chromosome pairs are called autosomes. In many shellfish, there are no identifiable sex chromosomes, and in the case of certain molluscs such as oysters, individuals change their sex during their lives. In some fish, sex can be influenced by environmental factors such as temperature or social interactions and sex chromosomes are often not sufficiently different to be identified in the karyotype.

How does sexual reproduction produce variation?

The key feature of sexual reproduction is the production of haploid gametes through the process of meiosis and the uniting of these gametes to produce a new diploid generation. The process of meiosis shuffles the genetic material in such a way that none of the haploid chromosome sets in the gametes are identical to either of the haploid sets present in the parent from which they are derived. To see how this happens, we must

look at particular stages of the process of meiosis. Here we give a brief outline of the behaviour of the chromosomes during meiosis, emphasising the genetic consequences rather than describing each stage in detail (Figure 1.7). The process of meiosis actually consists of two cell divisions: meiosis I and meiosis II. The full details of meiosis are given in all standard genetic texts.

Meiosis I begins long before the chromosomes become clearly visible. The chromosomes are initially very thin and uncontracted, but become progressively more contracted and more visible during the prophase stage. During this stage, homologous pairs of chromosomes come to lie closely together, and at the same time, each chromosome in each pair divides into chromatids that remain attached to one another at the centromere. So each pair of chromosomes consists of four chromatids. Such pairs of chromosomes at this stage are called bivalents. It is during this time, while the chromosome pairs are adhered closely together, that the process of recombination or crossing over occurs. Recombination involves the interaction between two ordinary (double-stranded) DNA molecules. The effect is that both molecules break, but the ends rejoin to the 'wrong' molecule. From the genetic point of view, each chromatid in a bivalent can be considered to be effectively a single DNA molecule and the recombination interaction takes place between chromatids that derive from different chromosomes of the pair – the non-sister chromatids (as opposed to the sister chromatids which are derived from an individual chromosome). The place where a recombination event is located on a bivalent is visible as a chiasma. This can take place in every pair of autosomes, but one chiasma formation reduces the probability of another one in an adjacent region of the chromosome, probably because of physical limitation of the chromatids to bend back upon themselves. This tendency of one chiasma to interfere with another is called interference. Interference is complete when there is only one chiasma per chromosome arm. Negative interference is where one chiasma increases the likelihood of adjacent ones.

At metaphase of meiosis I, the bivalents lie across the equator of the spindle with their centromeres attached to the arms of the spindle. Meiosis I is a reduction division (Figure 1.7). During anaphase, each pair of chromosomes is separated so that one of the pair goes into one daughter cell and the other into the other daughter cell. So at the start of meiosis I the cell contains four copies of the genetic information, but after division each daughter cell contains only two copies of the genetic material.

Meiosis II is effectively a mitotic division where the two chromatids from each chromosome separate into daughter cells (Figure 1.7). Therefore, each diploid cell that enters into meiosis produces four haploid gametes. Some genetic texts suggest that a cell can be regarded as 'tetraploid' when it enters meiosis I because it has four copies of the DNA.

What is the actual effect of recombination across the genome? Take a species such as the European flat oyster *Ostrea edulis*, for example, which has ten pairs of chromosomes. In all cells of the body, including the germ cells which will undergo meiosis, one of each pair of chromosomes will have come from the female and the other from the male parent. Envisage one of the chromosomes as a linear arrangement of genes along a single molecule of DNA. The other chromosome of the homologous pair will also consist of that same linear arrangement of genes along its length but will have come from a different parent. It has the same genes, in the same order, but has a different ancestry. That ancestry

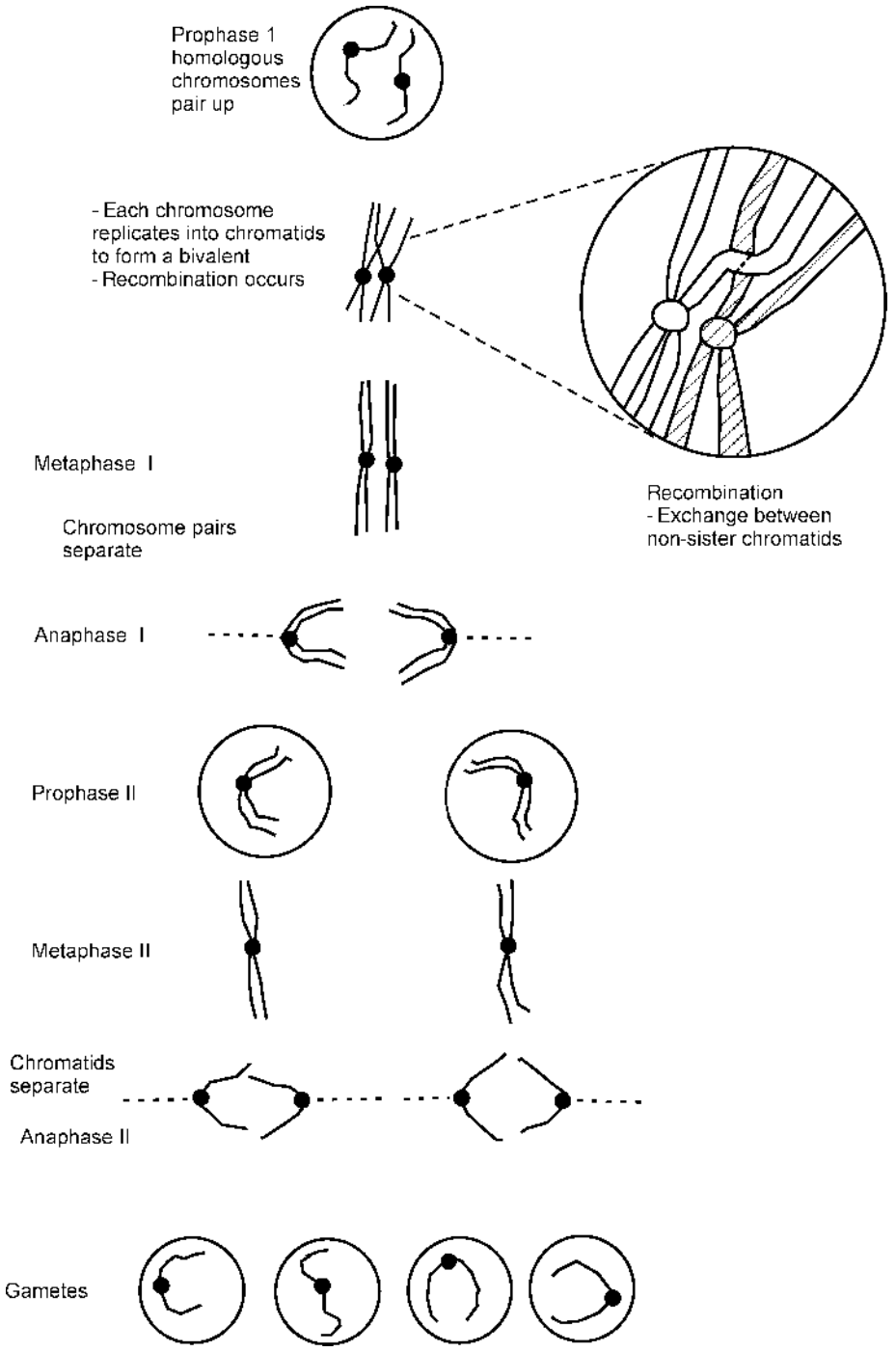

Figure 1.7 The process of meiosis. Recombination takes place during prophase of meiosis I.

will have provided it with different variations at many of its genes compared with the other chromosome of the pair. In early meiosis each chromosome has replicated itself, so there are two DNA copies (chromatids) of each chromosome. Recombination occurs between non-sister chromatids such that a stretch of DNA from one chromatid becomes exchanged for the equivalent stretch from the other chromatid. The resulting chromatid DNA molecules that have undergone recombination are therefore different from either of the parental ones. Any chromatids which have not been involved in a recombination event, of course, remain unaltered. The frequency of recombination events between two genes will determine genetic distance between those genes, and we will show how this enables mapping of the genome in Chapter 6.

Now note that this process can take place in all of the pairs of autosomes in that germ cell during that division. Then consider that this is just one germ cell among the millions of germ cells in the gonad of the oyster. All the other germ cells are also undergoing a meiotic division during which recombination is taking place in all the pairs of chromosomes. The precise position along the DNA molecules (chromatids) at which recombination events take place is (to some extent) random and will generally be different in each dividing germ cell. So it is easy to understand why the ten DNA molecules (chromosomes) in an oyster gamete are going to be different from any of the ten parental DNA molecules (chromosomes) that were present in the germ cell before meiosis. It can also be understood why the genetic make-up (the ten DNA molecules) of every gamete is likely to be different from every other gamete.

It can be seen that very extensive shuffling of the genome is achieved by recombination. However, this is not the only reshuffling that takes place during meiosis. Consider the ten pairs of chromosomes in the germ cell, with one from each pair derived from the male parent of the oyster, and the other from the female parent. These can be indicated as M1, M2, M3 up to M10 (for the male-derived chromosomes) and F1, F2, F3 up to F10 (for the female). Each of the daughter cells following meiosis I will contain ten chromosomes, but they will be a random mixture of F and M chromosomes as illustrated in Figure 1.8. For ten chromosomes, there are $2^{10} = 1024$ possible combinations. This is called independent assortment of chromosomes.

Therefore, shuffling of the genome takes place by two processes in meiosis – recombination and independent assortment – and these processes probably ensure that no two gametes are ever likely to be identical to either parental chromosome set, nor to one another.

The final part of the process of sexual reproduction, syngamy – the fusion of male and female gametes at fertilisation to form a zygote – further increases the genetic variation of offspring from their parents. Considering all these factors, it is not at all surprising that no two individuals in a sexually reproducing species are identical. The only exception is if an already-fertilised egg (a zygote) divides to produce two separate cells, both of which develop independently into normal embryos. These are monozygotic twins and are effectively genetic clones of one another. The original zygote from which they arose would still have been different from any other zygote, or individual, in that species.

Although the chromosomal behaviour during meiosis is the same in both males and females, there is an important difference in the production of spermatozoa and eggs

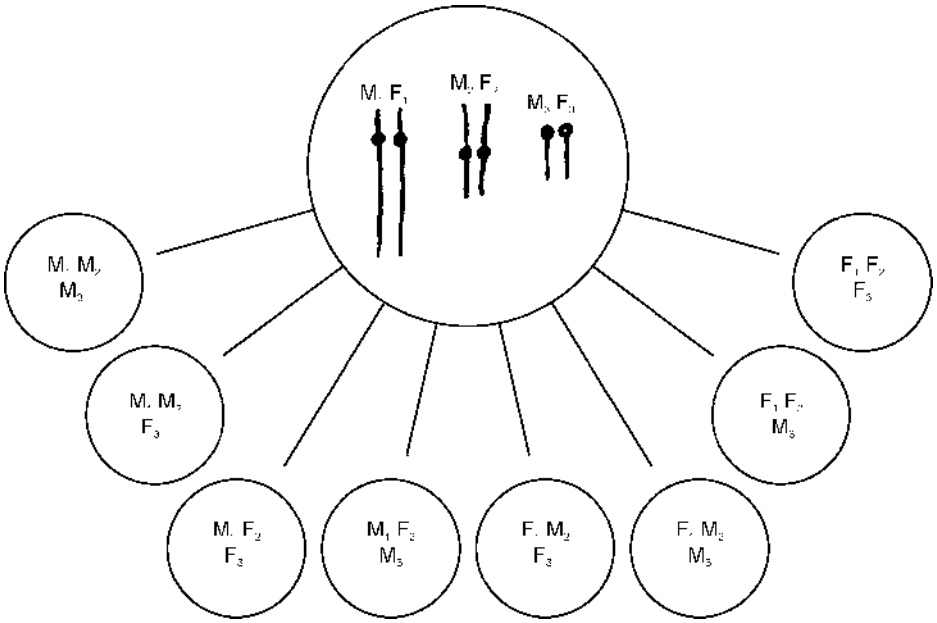

Figure 1.8 How independent assortment of chromosomes at the end of meiosis I creates variation. Three pairs of chromosomes are indicated: M1, M2 and M3 are the male parental chromosomes, and F1, F2 and F3 are the female parental chromosomes. Independent assortment gives eight possible combinations of chromosomes in daughter cells.

(ova). In males, four spermatozoa are produced from each germ cell. In females, only one egg (ovum) is produced from each germ cell or primary oocyte (Figure 1.9). One of the two cells produced during meiosis I is large, and the other is very small, so small that effectively it is really just the chromosomal material with little or no cytoplasm. This small cell – the first polar body – usually does not undergo meiosis II. The large cell – the secondary oocyte – again divides unequally during meiosis II to produce the large ovum and the small second polar body.

Agricultural animals are mostly mammals or birds in which the meiotic divisions take place inside the body of the animal or inside a shell and so are not easily accessible. However, in most fish species, only meiosis I takes place before spawning and the secondary oocytes are released into the water. Meiosis II occurs only when spermatozoa have become attached. In bivalve molluscan shellfish, the oocytes are spawned at metaphase of meiosis I and further development is dependent on the attachment of spermatozoa. This feature of fish and certain shellfish enables chromosome set manipulation to be simply engineered in such species for the production of polyploids (Box 1.1). Nevertheless, in certain fish, in crustacea and in brooding bivalves, eggs are not directly accessible during the meiotic divisions. The production of polyploids will be discussed in Chapter 7.

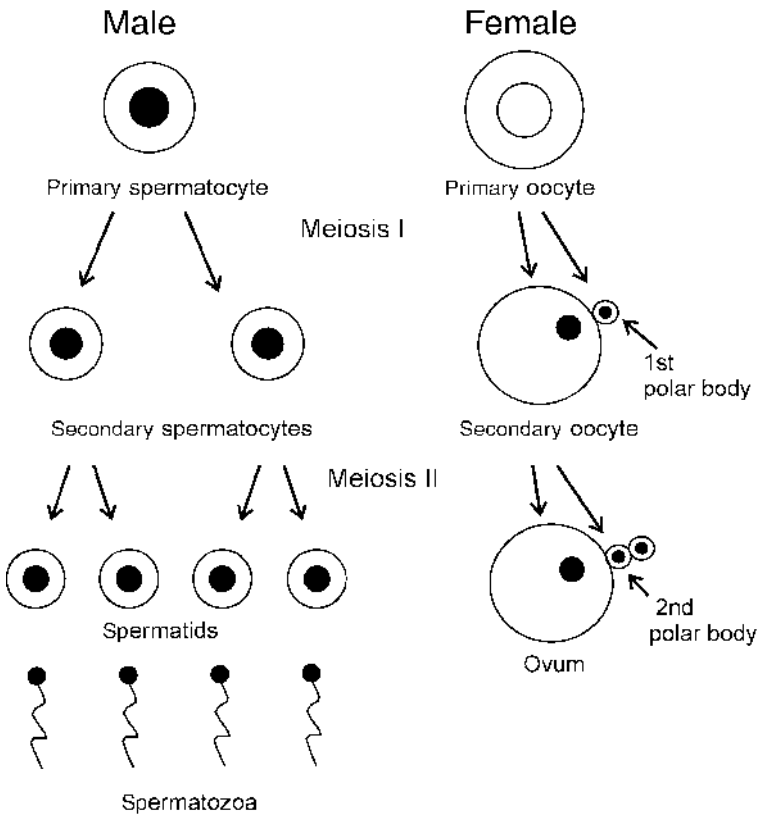

Figure 1.9 The difference in meiotic products in males and females. In males each primary spermatocyte produces four spermatozoa, while in females each primary oocyte produces a single ovum.

In addition to variation at the DNA level and the shuffling of the genes during meiosis, genetic variation occurs at the level of the chromosomes. Such variations are not very useful as genetic markers in aquaculture species, but chromosomal rearrangements have played an important role in evolution (Box 1.1).

Mitochondrial and chloroplast DNA

So far we have considered the DNA present in the nucleus, which is organised into chromosomes. There is actually more DNA in the cell – extra-chromosomal genes, contained within energy-generating organelles, mitochondria, of which there may be several hundred in each cell. In plants, the photosynthetic organelles – chloroplasts – also contain DNA.

Animal mitochondrial DNA (mtDNA) is normally present as a circular molecule of around 16 kb in length, and there are around ten copies of the DNA in each mitochondrion in humans. Unlike the chromosomal DNA, there is no meiosis and replication appears to be a simple copying process, though recent research does point to there being some form of recombination during mtDNA replication.

Because there are large numbers of mitochondria in an egg, but very few in a spermatozoon, it is hardly surprising to find that the mtDNA present in a sexually reproduced offspring is usually inherited entirely from its mother (strict maternal inheritance). This maternal-only inheritance of mtDNA is the situation in almost all animals.

However, one exception to this rule occurs in an important aquaculture species, the mussel *Mytilus* spp., which has a form of bi-parental inheritance of mtDNA. Females have an F-type of mtDNA in every body cell, while males have both the F-type and an M-type mtDNA in most cells of the body. The M-type is highly concentrated in the male gonad and is thought to be the only mtDNA present in the spermatozoa. These M-type mtDNA molecules present in a spermatozoon enter the egg, and, in some way which is not yet fully understood, the M-type remains in the egg after fertilisation and is eliminated in individuals destined to become female, but retained and preferentially replicated in individuals destined to become males. This unusual arrangement has been named 'doubly uniparental inheritance'. Doubly uniparental inheritance has recently been detected in other bivalves besides mussels such as the manila clam (*Tapes philippinarum*) and may be more widespread than currently thought.

The complete sequence of the mitochondrial genome is now known for quite a number of vertebrates and invertebrates, and the order of the genes within the circular genome is different in every phylum so far studied. Because fish mtDNA has been extensively used in phylogenetic studies, we will use this molecule as an example. The mitochondrial genome of fish contains 13 genes coding for proteins, 2 genes coding for rRNA (the small 12S and the large 16S rRNA), 22 genes coding for tRNA molecules and 1 non-coding section of DNA which acts as the initiation site for mtDNA replication and RNA transcription. This is called the control region (Figure 1.10).

In contrast to the nuclear genome, the mitochondrial genes of animals are very efficient and have no introns. In addition, there is virtually no 'junk DNA' or repetitive sequences in the mitochondrial genome, although the control region does often vary in length due to tandem repeats. Exceptions to this general rule are the scallops, many species of which exhibit several large (up to 1.4 kb) repeated sequences within the mtDNA genome which can consequently extend to beyond 30 kb in length.

For reasons which are not fully understood, the rate of mutation in animal mtDNA is higher than that in the nuclear DNA (about 5–10 times higher). This means that the rate of evolution is greater in mtDNA than in nuclear DNA, and this feature is of importance to us when we are looking for genetic markers which will reflect changes in the more recent past.

Chloroplasts of photosynthetic organisms also contain DNA molecules, having originated from endosymbiotic cyanobacteria. Each chloroplast can contain up to 100 copies of this small genome composed of circular molecules. It is commonly maternally inherited. The chloroplast genome was the first plant genome characterised because of its

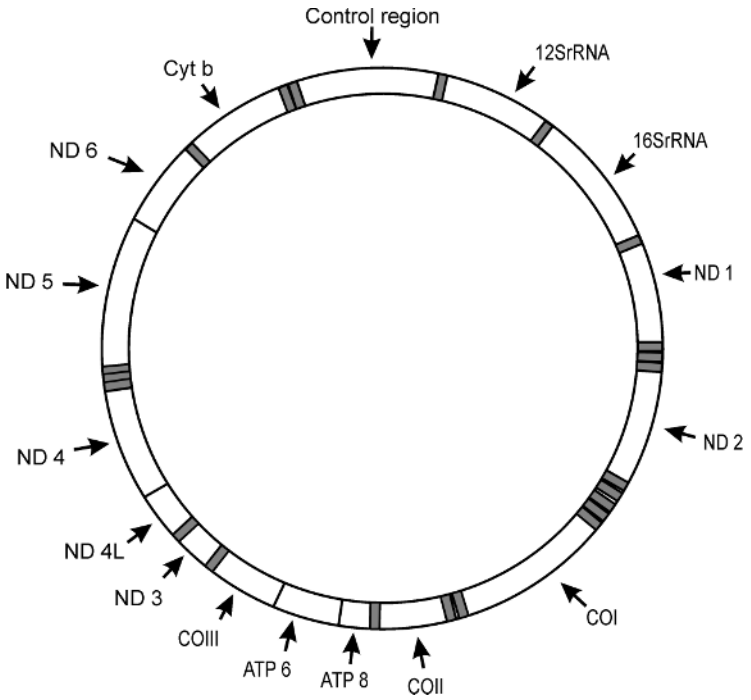

Figure 1.10 The generalised mitochondrial genome of fish. The position of the 13 protein coding genes is indicated on the molecule. They are seven subunits of the enzyme NADH dehydrogenase (ND 1, 2, 3, 4, 4L, 5, 6), cytochrome *b* (Cytb), three subunits of cytochrome *c* (COI, II, III) and two subunits of the enzyme adenosine triphosphate synthetase (ATP6 and ATP8). 12SrRNA, 12S rRNA; 16SrRNA, 16S rRNA. The shaded segments indicate the positions of the tRNA genes.

small size and ease of isolation. The complete chloroplast genome sequences of a number of plants and algae have been determined, revealing massive transfer of genes from ancestral organelles to the nucleus. This might explain why all chloroplast functions require imported proteins encoded by the nuclear genome.

Further reading

Avise, J.C. (1994) *Molecular Markers, Natural History and Evolution*. Chapman & Hall, London.

Brown, T.A. (1999) *Genomes*. Bios Scientific Publishers, Oxford.

Majerus, M., Amos, W. & Hurst, G. (1996) *Evolution: The Four Billion Year War*. Longman, New York.

Turner, P.C., McClennan, A.G., Bates, A.D. & White, M.R.H. (1997) *Instant Notes in Molecular Biology*. Bios Scientific Publishers, Oxford.

Winter, P.C., Hickey, G.I. & Fletcher, H.L. (1998) *Instant Notes in Genetics*. Bios Scientific Publishers, Oxford.