

# 2

# The immortal germline

*Or: how do I get DNA samples?*

*“If you could use a big enough magnifying glass you would find that there is really only one kind of life on the Earth: the most central machinery in all organisms is built out of the same set of micro-components, the same set of small molecules.”*

Cairns-Smith, A.G. (1985) *Seven Clues to the Origin of Life*.  
Cambridge University Press

## What this chapter is about

The genomic information system is shared by all living things. It is worth learning a little of the history of the discovery of this genomic system because it serves to illustrate some of the important principles of heredity. Life relies upon the continuity of genetic information. This information is encoded in DNA, and copied from generation to generation. The practical upshot of these principles of heredity is that the DNA found in every living cell contains all the genetic information needed to construct the organism, as well as providing biologists with a wealth of information about biological history and evolutionary processes.

## Key concepts

*Evolutionary biology:* heredity

*Molecular evolution:* DNA structure

*Techniques:* DNA extraction



## → Unity of life

All life on Earth has a common ancestor. If you trace your family tree back far enough, you will find you are related to the rats in your attic, the fly on your window, the mould on your bread, the rice in your cupboard, even the bacterium in your gut (**Figure 2.1**). We know this because every organism on Earth uses the same basic system to carry the information that it needs to grow and reproduce. This genomic system consists of information stored in DNA, which is then transcribed into RNA, and then translated into proteins. This system is so intricate, with so many complex interlocking parts, that we can be sure that it was not invented separately in different biological lineages. So we are all descended from an ancient, simple life form that carried the same fundamental genomic system that we share today with all of the Earth's biodiversity.

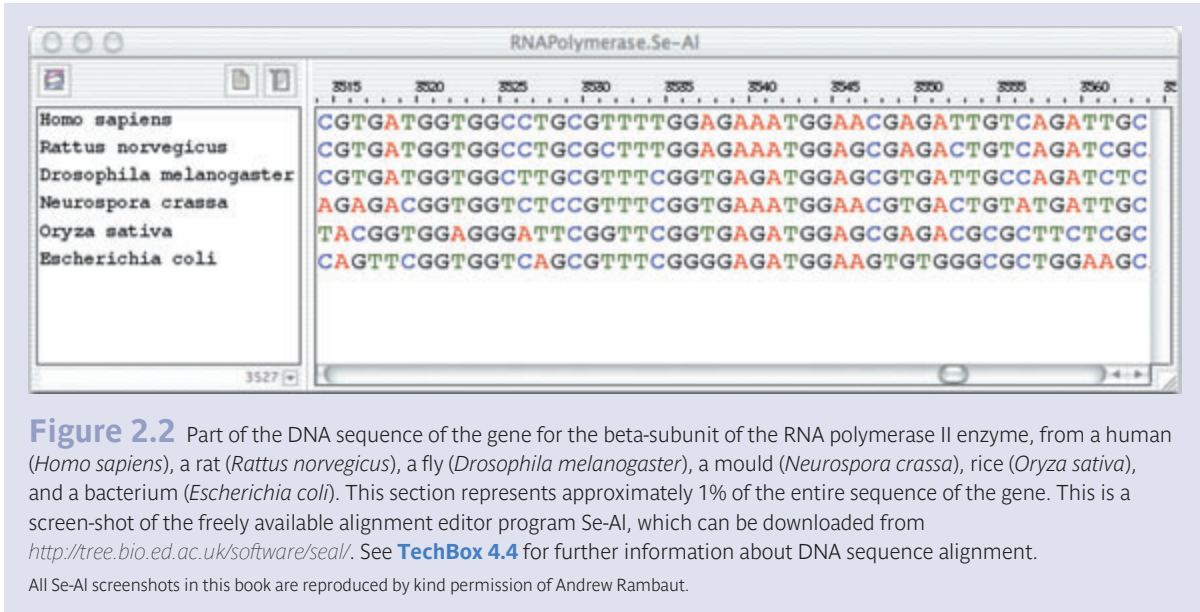
Not only is the genomic system of DNA, RNA, and proteins shared by lifeforms as different as rats, flies, mould, rice, and bacteria, but some of the actual information is also shared. In a sense this is not surprising, because all living things must share the instructions for constructing the genomic system itself. For example, all organisms must be able to convert the genomic information coded in DNA into RNA messages that provide the instructions needed to make a protein. One of the things they need in order to do this is a working copy of the enzyme RNA polymerase, which makes an RNA message from a DNA gene sequence. So all organisms must have a gene that codes for RNA polymerase.

In fact, eukaryotes (such as fungi, plants, and animals) have several different types of RNA polymerase



**Figure 2.1** Rice (*Oryza sativa*) is one of the most important crop plants in the world. It is often said that humans are, genetically speaking, 99% similar to chimpanzees. Perhaps more remarkably, we share about half our genes in common with plants. Remember that next time you eat a bowl of rice.

Reproduced by permission of Oliver Spalt, 2002.



enzymes, and each one is constructed from many different subunits, in some cases combining a dozen or more different proteins. For simplicity, let's concentrate on a gene for a single subunit of one of these RNA polymerase enzymes, the beta-subunit of RNA polymerase II. **Figure 2.2** shows part of the DNA sequence of this gene. Each horizontal line of letters represents the sequence of bases in the version of the RNA polymerase gene found in a particular species. In this alignment, we can compare the RNA polymerase II beta sequence from humans, rats, fruit flies, bread mould, rice, and bacteria. Although the exact DNA sequence is slightly different in each of these organisms, each of the versions provides the necessary instructions for making a working copy of the RNA polymerase gene.

This is just a small part of one gene. The genomes of most organisms contain thousands of genes. Every time the body manufactures a new molecule to help it grow or move or respond to environmental change, particular genes must be located in the genome, unwound, transcribed, and translated. Then the newly manufactured molecules must be folded, combined, and transported to where they are needed. The expression of a single gene requires the co-ordinated action of dozens of enzymes, the manufacture of a great number of spe-

cialized molecular building blocks, and the co-ordination of a large number of tasks in time and space within the cell. Yet this complex process is being continuously performed by every single living cell to produce thousands of proteins and other molecules, all in the right place, at the right time, in the right amounts. The beauty and complexity of the genomic system never ceases to amaze me.

In order to truly appreciate the wonder of the natural world, you need to gain some insight into the workings of the molecular level of organization that underlie all of the functions of the living world. This is important for two reasons. Firstly, an appreciation of molecular biology is the best way to bring home the complexity and intricacy of organisms. Secondly, a grasp of the biochemical basis of heredity is essential to understanding evolution, as it is at this level that mutations occur, substitutions accumulate, causing lineages to change and diverge over time. We are going to briefly consider the history of the discovery of the genomic system in order to review some of the key principles of heredity. These ideas are not just of historical interest, nor are they only relevant to those interested in genetics itself. These principles of heredity explain why we can use DNA as an information source in evolutionary biology and ecology.

## → Principles of heredity

Our knowledge of the genomic system is surprisingly recent. The basic principle of heredity – that offspring tend to resemble their parents – has long been observed by human societies. But the exact mechanics of inheritance were subtle and unknown. It had always been recognized that animals and plants tend to arise from parents of the same species, but it was commonly believed that in certain circumstances living beings could arise spontaneously, such as flies being generated from rotting meat, wasps arising from galls on plants, or bacteria forming *de novo* in chicken broth. Spontaneous generation was finally put to rest 150 years ago, when conclusive experiments in which potential parent organisms were carefully excluded from sterilized material ultimately convinced scientists of the importance of genetic continuity. Only living organisms contained the necessary information to make another organism; life cannot arise without a copy of this genetic information. But what was the material basis of genetic continuity? Did reproductive cells contain tiny preformed creatures that grew into new adults? If so, then how were traits from both mother and father inherited? And how could an organism such as a sponge reproduce by budding, where a small piece of its body could be induced to grow into a new individual? Somehow, cells must be able to transmit information to create a new individual.

One early theory of heredity, published by Pierre Louis Moreau de Maupertuis in 1745 in his natural history tract *Vénus Physique*, held that particles corresponding to all parts of the body were provided by the parents, and used to build the developing offspring. These particles could be altered to give rise to new hereditary types, and might even undergo isolation in different parts of the world to produce new species. The idea that all parts of the parent's body contribute information to the offspring is referred to as pangenesis. Theories of pangenesis have a long history, going back to the philosophers of ancient Greece, but one of the most famous proponents of pangenesis was the father of evolutionary biology, Charles Darwin.



**Darwin's theory of evolution is discussed in more detail in Chapter 5**

Darwin recognized the central role of heredity in evolutionary theory, devoting whole volumes to recording and interpreting observations of inheritance in the natural world (largely in domesticated animals). Darwin knew that variations arose continually in natural populations, and that many variations could be inherited, but he could only guess at the mechanism. Critics of natural selection pointed out that if the characteristics of the parents were blended to create their offspring, then any favourable new variant would be diluted each generation and eventually lost.

The lack of a clear mechanism for inheritance had been, in many ways, a stumbling block for the development of evolutionary theory. To fill this gap, Darwin developed his theory of pangenesis, speculating that all the body's cells produced particles, called gemmules, which carried information. Gemmules collected in the reproductive cells prior to fertilization, ensuring the offspring inherited all the information needed to make a functioning body. Because gemmules formed in the adult body, Darwin's theory of pangenesis specifically allowed for the inheritance of acquired characteristics – modifications of the body during an individual's lifetime could be inherited by its offspring. The theory of pangenesis was politely ignored by some of Darwin's contemporaries, and criticized by others. One strong critic was Francis Galton, one of the founders of modern statistics and leader of the early eugenics movement. Galton showed that blood transfusions did not appear to move hereditary information from one individual to another, as would be expected if gemmules were carried in the blood. Even 20 years after the publication of Darwin's classic work on evolution, *The Origin of Species*, the material basis of heredity was still unknown, despite being the focus of much study.

## The Weismannian barrier

“At the present time there is hardly any question in biology of more importance than this of the nature and causes of variability, and the reader will find in the present work an able discussion on the whole subject. . . . Whoever compares the discussions in this volume with those published twenty years ago on any branch of Natural History, will see how wide and rich a field for study has been opened up through the principle of Evolution; and such fields, without the light shed on them by this principle, would for long or for ever have remained barren.”

Charles Darwin, Foreword to *The Study of Heredity* by August Weismann (1880)

August Weismann transformed evolutionary biology by arguing forcefully against the inheritance of acquired characteristics. His arguments were largely made on theoretical grounds, by considering the implications of patterns of inheritance for the process of evolution. For example, how could the non-reproductive castes of social insects evolve if there was no way that a sterile worker could pass its bodily modifications, having no offspring of its own? How could mutations arise in organs that must be fully formed before use, therefore had no opportunity for acquiring new characteristics by use and disuse? And more importantly, how could information about the state of adult organs be translated into a form of inheritable instructions? (**Heroes 2:** August Weismann).

Weismann also argued on empirical grounds: despite widespread belief, there was simply no evidence that acquired characteristics could be inherited. For example, it was clear that human societies that practised male circumcision over many generations did not give rise to offspring that no longer had foreskins. Weismann carried out experiments that demonstrated that bodily modifications acquired in an individual's lifetime were not passed to their offspring. One of his most famous experiments was to dock the tails of mice, then breed from the tailless individuals, and dock the tails of their offspring. He continued this process through 21 generation of mice, docking the tails each

generation. But the tailless mice never produced tailless offspring. Although this experiment seems trivial, it is important to realize that this demonstration ran counter to the prevailing opinion of the times. Many animal breeders believed it was possible to produce a tailless breed in this way, by cutting the tails of individuals then breeding from them. But experiments such as this denied the universality of pangenesis, because the docked tail did not appear to contribute to the genetic information inherited by the offspring.

Thus Weismann argued persuasively for ‘hard inheritance’: genetic information was not added to throughout life, as the body grew and changed, but was set immutable from conception. The reproductive cells (germline) were not influenced by changes in other cells of the body (soma) and so heritable information was passed from one generation to the next largely unchanged. Weismann defined two key principles of heredity. Firstly, he suggested that the germline was effectively isolated from the soma. This meant that changes made to the body during a lifetime could not be passed to the gametes. Secondly, Weismann considered that the germline represented an unbroken chain of information passed from parents to offspring, and so on down through the generations. These two principles combine to give us our modern view of heredity: that genetic information is passed from generation to generation, essentially unaffected by changes to the body.

Although he made detailed studies of developmental biology, Weismann's conceptual advances were primarily theoretical, made by considering the implications of various models of heritability for evolution. At the same time, largely unknown to those in the scientific community debating heredity and evolution, breeding experiments were being carried out that would shed light on the nature of the immortal germline. These experiments would eventually be used to counter Darwin's critics, by demonstrating that hereditary information did not blend and dilute down the generations, but was passed on in discrete units that could be carried over many generations.

HEROES OF  
THE GENETIC  
REVOLUTION

## 2

## August Weismann

*“I have gradually become aware, that, after Darwin, Weismann was the greatest evolutionary biologist of the nineteenth century. Further, the problems he was concerned with are often the same problems that concern us today”*

Maynard Smith, J. (1989) *Weismann and Modern Biology*. Oxford Surveys in Evolutionary Biology, Volume 6, pages 1–12

**NAME**

August Freidrich Leopold Weismann

**BORN**

17 January 1834, Frankfurt am Main, Germany

**DIED**

5 November 1914, Freiburg, Germany

**KEY PUBLICATIONS  
(English translations)**

*Studies in the Theory of Descent* (1882) translated by R. Meldola. London, Simpson Low, Marston, Searle and Rivington.

*The Germ-Plasm: a theory of heredity* (1893) translated by W. N. Parker and H. Ronnfeldt, New York, Charles Scribener's Sons.

**FURTHER INFORMATION**

Facsimiles of Weismann's *Essays on Heredity* and *The Germ-Plasm* are freely available at [www.esp.org/books/chrono-1st.html](http://www.esp.org/books/chrono-1st.html)

**Figure Hero 2**

August Weismann.

Image from Conklin, E.G. (Oct–Dec., 1915) August Weismann. *Proceedings of the American Philosophical Society*, Volume 54, No. 220., pages iii–xii.

August Weismann was the first to be called a ‘neo-Darwinian’ (not intended as a compliment then, and, regrettably, often used in the same vein today). Like Darwin, much of his work was prescient, and it is surprising how many key ideas in modern evolutionary biology can be found in Weismann’s work, such as the role of sexual reproduction in generating variation, and a discussion of the cellular causes of ageing. In particular, Weismann’s careful observations of cell division led to the recognition of the role of chromosomes in heredity. Many of Weismann’s books, like Darwin’s, surprise modern readers with their vast catalogues of observations about the natural world. In Weismann’s case, his special interest was in the coloration of caterpillars and butterflies. While these intimate studies of butterflies may seem whimsical, they provided Weismann with abundant raw material for understanding developmental biology and genetics.

One of Weismann’s most important contributions was that he convincingly demonstrated that, counter to the prevailing viewpoint at the time, changes to the body during an individual’s lifetime are not a source of heritable variation. Because he was working in the late 1800s when the molecular basis of heredity was unknown, Weismann’s central theories are framed in terms of cell lines: the germline cells (which form sperm and eggs) are isolated from the somatic cells (all other cells in the body). We can now interpret Weismann’s principles in terms of DNA: the information in DNA is passed on exactly as it was inherited (bar the occasional mutation), because information about the state of the body is not recorded in an individual’s DNA during their lifetime.

The following extract from a contemporary review of Weismann’s work give some sense of the impact of his ideas, and the controversy and excitement surrounding the problem of heredity:

‘In spite of the difficulties involved in acceptance of Weismann’s view, however, it has been enthusiastically accepted in England by the younger Darwinian school. . . . The old school of Lamarck seemed dead; even the ideas of Herbert Spencer and of Darwin himself as to “use and disuse” began to be looked upon as antiquated and unphilosophical. . . . At the present moment a reaction has set in; the battle is raging fiercely. . . . Alike in Germany and in England, criticism and doubt as to Weismann’s premises are beginning to take place of the paean of exultation. . . . What is wanted now is some decisive experimental settlement of the question. Can it be shown that in any case a capacity or habit acquired beyond a doubt during the lifetime of the individual is transmissible to the off-spring? If that can be proved, Weismannism falls at once to the ground, and we revert to the primitive Darwinian and Spencerian problem.’

Allen, G. (1890) The new theory of heredity. *Review of Reviews*, Volume 1, pages 537–538

## → Discovery of the gene

“What [was] called for was a theory of heredity by which inheritance would be essentially discrete, discontinuous, and ensured by units that could be transmitted from generation to generation without losing their somatogenic qualities. Such is the gene.”

Jacques Monod (1974) On the molecular theory of evolution. In *Problems of Scientific Revolution*, D. Harre (Ed.), Oxford University Press.

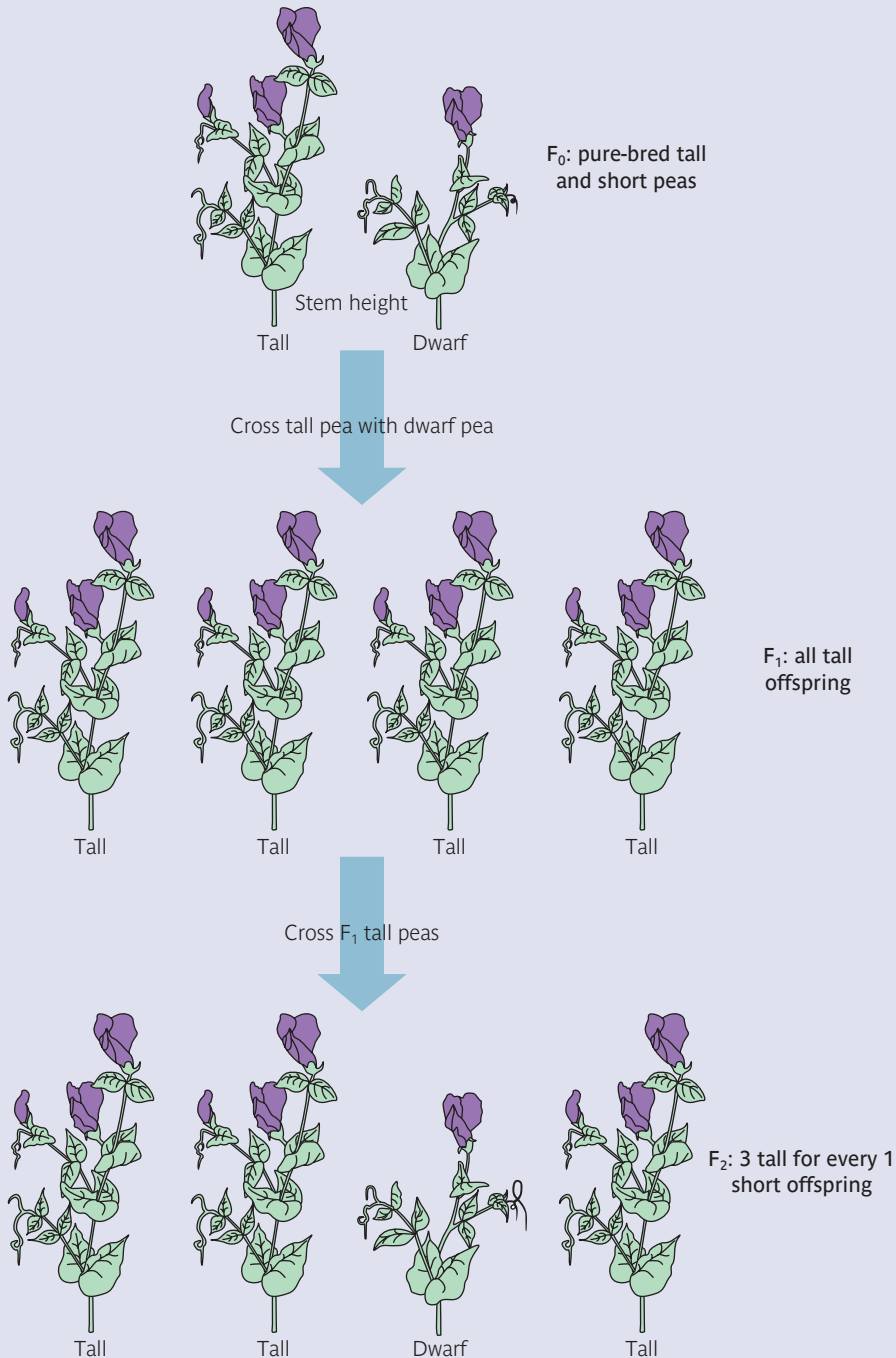
Gregor Mendel was a researcher and teacher at a monastery in Moravia (now in the Czech Republic), which had a thriving research programme in many aspects of natural science. For nearly a decade in the mid-1800s, Mendel conducted large-scale experiments in plant breeding. He systematically crossed 34 different pure-bred strains of peas and recorded the characteristics of over 10,000 individual plants. Through the pioneering application of statistical analysis to the problem of heredity, he was able to show that the heritable features of these pea strains were preserved down the generations. When Mendel crossed two varieties of peas, he found that offspring did not have a simple blend of their parents' characteristics. Instead of being intermediate between the two types, the offspring resembled one parent or the other. For example, a tall pea crossed with a short pea produced tall offspring, not offspring of medium height (Figure 2.3). But when these tall offspring were crossed together, they produced both tall and short plants. The variation in the parents' generation was not lost, because it reappeared in subsequent generations.

Furthermore, the proportion of offspring of each parental type varied in predictable frequencies. When tall and short plants were crossed, they produced all tall offspring. But when these tall offspring were crossed with each other, their offspring (the 'grandchildren' of the original tall by short cross) varied in height in predictable ratios: three tall offspring to every one short. Mendel had discovered that heritability was governed by discrete factors that were copied and combined down the generations, and did not disappear through interbreeding. He was therefore the first person to describe the action of the hereditary units, which he called 'factors' or 'elements', but are now known as genes.

But what were these inherited factors? Early geneticists and evolutionary biologists studied the behaviour of genes in great detail. They described patterns of inheritance, how different genes combined to produce particular traits, and how these heritable traits varied within populations. Yet they did not actually know what genes were made of, or where in the cell they were located. Chromosomes seemed a good candidate for the genetic material. Their ordered behaviour at cell division, with a copy of each chromosome going to each daughter cell, matched Mendel's description of the segregation and assortment of genetic factors. But chromosomes are made of both proteins and DNA. Which of these two types of molecules held the hereditary information?

DNA had been discovered in the 1860s by Friedrich Miescher. First, he collected cells, such as white blood cells taken from pus on bandages collected from a hospital. Then he used a number of protocols to lyse the cells and separate the cellular contents (see [TechBox 2.4](#), p. 51). When he isolated the central nuclei of the cells, he found them to be full of a phosphorous-rich material. He called this substance nuclein. Nuclein was found in every cell that Miescher tested, but it appeared to be inert: that is, it was non-reactive and didn't appear to have any special metabolic role. Miescher initially concluded it might simply be a way of storing phosphorous in cells, though later he began to suspect it had some kind of role in fertilization.

As knowledge of nuclein was refined over the next 60 years, it was renamed deoxyribonucleic acid (DNA). It was shown that DNA was found in chromosomes, and that it was made up of phosphates, sugars, and bases linked together in long chains. But relatively few scientists were interested in DNA, since it did not seem to do anything exciting. DNA was always in the same inert form, did not appear to do anything other than lie around in chromosomes, and had only four different units (the nucleotide bases). Proteins, on the other hand, existed in huge variety, did much of the important work in a cell, and were made up of over 20 different units (amino acids). Many scientists thought proteins were the obvious choice for storing the vast amount of information need to make even the simplest cell. But the problem with the protein theory of heredity



**Figure 2.3** Gregor Mendel crossed distinct pure-bred lines of peas and showed that, for certain traits, the first-generation ( $F_1$ ) offspring were not intermediate between the two parental types, but all resembled one parent. In the case illustrated here, crossing tall and dwarf varieties produced all tall offspring, no dwarfs. But the genetic information from the two parents was not lost. When Mendel crossed the first-generation ( $F_1$ ) offspring with each other, the second-generation ( $F_2$ ) offspring included both tall and dwarf plants. Mendel identified seven traits that varied discretely in this way, including wrinkly versus smooth seeds, yellow versus green peas, and purple versus white flowers. For more information see MendelWeb ([www.mendelweb.org](http://www.mendelweb.org)).

was not how information could be stored, but how it could be copied and passed to offspring. Could proteins be copied? And would a cell need to inherit a copy of every essential protein from its parent?

## DNA as genetic material

In the 1940s several experiments had suggested that it was DNA, not proteins, that carried genetic information. For example, Oswald Avery and colleagues showed that genetic information could be passed from one strain of the bacterium *Pneumococcus* to another. They used a series of experiments to show that the 'transforming factor' (genetic information) was preserved even when enzymes were used to remove all proteins, sugars, and RNAs. But if DNA was removed from the solution, then it could not transform cells. They concluded that it was the DNA that carried information from one cell to another. However, these experiments did not convince the majority of scientists

working on the molecular basis of heredity, most of whom continued to concentrate on proteins.

Conclusive proof was provided by Alfred Hershey and Martha Chase in 1952. Their 'blender experiment' (named for their innovative use of kitchen equipment) showed that DNA was responsible for genetic continuity. By attaching different radioactive labels to the protein and DNA in viruses, they could demonstrate that it was the DNA, not the protein, that transmitted the information needed to make a new virus (**TechBox 2.1**). But it was the publication of a single, one-page scientific article in the journal *Nature* the following year that put DNA at the heart of modern biology. This paper, in which James Watson and Francis Crick described the molecular structure of DNA, concludes with a sentence of elegant understatement: 'It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.' This is the statement that launched the genetic revolution (see **Heroes 4**: Francis Crick).

## → Structure of DNA

It had been known since the 1920s that DNA was made up of regular patterns of three kinds of molecular subunits: phosphates, sugars, and bases. But how were they connected together? Watson and Crick had seen the outstanding X-ray diffraction pictures of DNA taken by Rosalind Franklin. These pictures suggested that DNA was a helix, with long chains of linked phosphate and sugar molecules twisted around each other in a regular pattern. Importantly, Watson and Crick combined this insight with an earlier observation made by Erwin Chargaff that the four types of bases of DNA were curiously evenly mixed. The number of adenine bases was always equal to the number of thymines, and the number of guanines was the same as the number of cytosines. Watson and Crick realized that Chargaff's pairing rule – A matches T, G matches C – was the key to the structure of DNA.

Watson and Crick constructed a large model, cutting shapes from tin plate to represent the four nucleotide bases (A, T, G, and C). This model, with its the elegant spiral staircase of two intertwined strands of phos-

phates and sugars, connected by rungs of paired bases, is familiar to many biologists as the star of one of the most widely used publicity photographs in the history of biology (**Figure 2.4**, p. 40). The complementary pairing of bases between the double strands of the helix, A with T and G with C, meant that each strand was an exact complement of the other. One strand could act as a template for the other, providing a means of copying information. The answer was so obviously right that Francis Crick is said to have announced in their local pub that evening: 'We have just uncovered the secret of life!' (**TechBox 2.2**).

## From DNA to RNA to protein

The template-copying mechanism identified by Watson and Crick is the key to understanding not only the replication of DNA, but also the way that the information in DNA is used to build and operate living cells. The genome, made of DNA, is often described as a blueprint. It holds the instructions for making a cell, but it is not directly involved in construction. Instead, the genetic

TECHBOX  
2.1

## Hershey–Chase blender experiment

## KEYWORDS

viruses  
bacteriophage  
radioactive labels  
inheritance  
DNA  
proteins

## FURTHER INFORMATION

*The Race for DNA* is an account of the discovery of the structure of DNA, told through contemporary documents and interviews. <http://osulibrary.orst.edu/specialcollections/coll/pauling/dna/narrative/page1.html>

RELATED  
TECHBOXES

TB 2.2: DNA structure  
TB 2.4: DNA extraction

RELATED  
CASE STUDIES

CS 4.1: Glorious mud (using blenders to isolate DNA)  
CS 3.1: Viruses within (viral genomes)

“When asked what his idea of happiness would be, [Hershey] replied, ‘to have an experiment that works, and do it over and over again’.”

Hodgkin, J. (2001) Hershey and his heaven. *Nature Cell Biology*, Volume 3, page E77

The experiment that put DNA at the centre of molecular genetics was carried out by Alfred Hershey and Martha Chase in 1952. They used bacteriophage to test whether it was proteins or nucleic acids that carried hereditary information. Bacteriophage (otherwise known as ‘phage’) are viruses that parasitize bacteria. Viruses lack the necessary equipment to replicate their own genomes, so to reproduce they must parasitize the molecular machinery

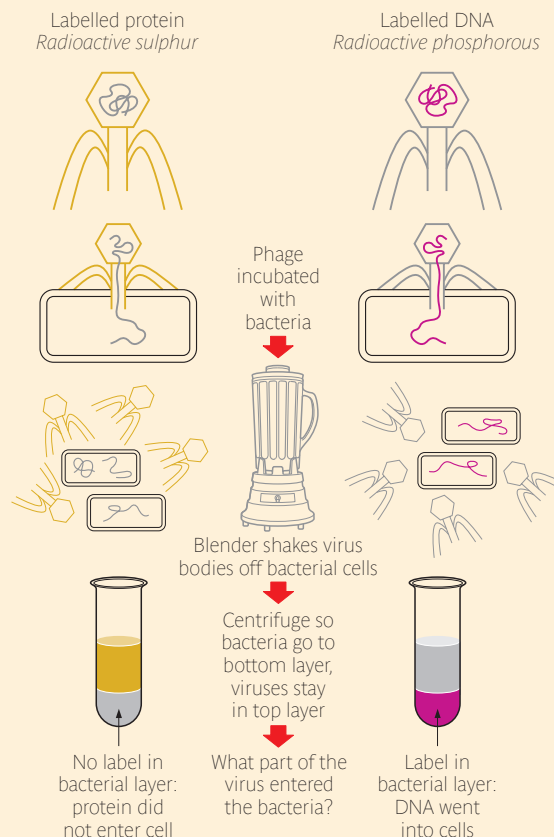


Figure TB2.1 The Hershey–Chase blender experiment.

of a living cell. A phage attaches to the host cell wall and injects its genome into the cell. The bacterial cell then makes copies of the virus genome, and the information in the virus genes is used to make the proteins that form the viral coat (i.e. the 'body' of the virus). The viral genomes are then packaged into the protein coats to form infectious virus particles.

Hershey and Chase labelled viral proteins with radioactive sulphur, and labelled viral DNA with radioactive phosphorous. They allowed these radioactive phages to infect bacteria, so that their genomes would be injected into the bacterial cells. Then they mixed the infected bacterial cells up in a Waring blender, an iconic 1950s domestic appliance. The blender separated the virus coats, which stayed outside the cell, from the genetic material which was injected into the bacteria. The radioactive labels allowed Hershey and Chase to show that the viral protein did not enter the bacterial cells, but the viral DNA did. DNA, not protein, was therefore the genetic material that carried the instructions for making new virus particles.

The effect of the Hershey–Chase experiment was to immediately convince both the leading scientists of the day (such as Linus Pauling) and less established researchers (such as James Watson and Francis Crick) that DNA was the key to understanding heredity. So the race began to discover the structure of DNA.

information stored in the DNA is expressed through the actions of RNA and protein molecules.

To illustrate this process of the conversion of genetic information from one form to another, let's consider the gene shown in **Figure 2.2**, p. 32. In humans, the gene that codes for the beta-subunit of RNA polymerase

II (given the acronym RPB2) is found on the short arm of chromosome 4. It takes 3525 bases of DNA to specify the amino sequence needed to make this protein. When the cell needs to make an RNA polymerase II enzyme, the RPB2 gene must be located. Then the DNA containing the gene is unwound from the chromosome, and the two strands of the double helix are unzipped to



**Figure 2.4** Francis Crick (right) and James Watson with the tin-plate model they built in 1953 (with the help of workshop technicians at Cambridge) to demonstrate their proposed helical DNA structure with complementary base-pairing. You can find out more about the construction of this model by listening to *A Twist of Life*, an entertaining radio programme narrated by Steve Jones, including interviews with many of the leading figures in the race for DNA. [www.bbc.co.uk/radio4/science/atwisttolife.shtml](http://www.bbc.co.uk/radio4/science/atwisttolife.shtml).

Reproduced by permission of A. Barrington Brown/SPL.

TECHBOX  
2.2

## DNA structure

## KEYWORDS

nucleotide

base

purine

pyrimidine

ribose

sugar

phosphate

5', 3'

double helix

phosphodiester bond

## FURTHER INFORMATION

Any good biochemistry text will explain DNA structure in more detail. For example, a very clear account can be found in Elliott, W.H. and Elliott, D.C. (2005) *Biochemistry and Molecular Biology*, Oxford University Press.

RELATED  
TECHBOXES

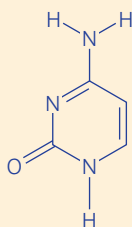
TB 4.1: DNA replication

TB 1.2: DNA sequencing

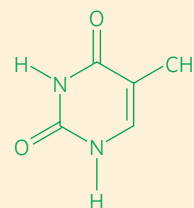
RELATED  
CASE STUDYCS 4.1: Glorious mud  
(DNA hybridization)

There are three basic subunits in DNA: bases, sugar, and phosphate.

**Bases:** The ring-shaped bases are the most charismatic of the DNA subunits. They come in four types which fit together in pairs. This pairing forms the basis of the information-carrying capacity of DNA. The bases are rings of oxygen, hydrogen, nitrogen, and carbon molecules. Two of the bases, the pyrimidines, are single rings.

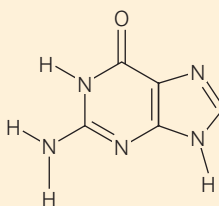


Cytosine (C)

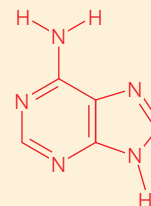


Thymine (T)

The other two bases, the purines, are made of double rings.

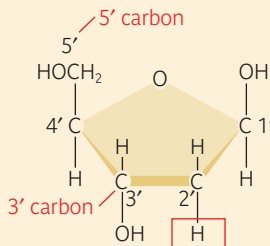


Guanine (G)

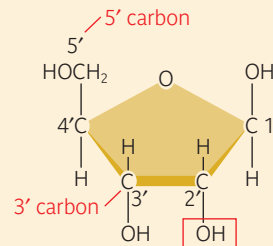


Adenine (A)

**Sugars:** Each base is joined to a sugar molecule, a 5-carbon (pentose) ring. In RNA, the sugar is ribose. In DNA, the sugar is a very similar molecule called deoxyribose (the 'deoxy' means that one hydroxyl (OH) group is missing from this form of ribose). Why do RNA and DNA have slightly different sugars? Deoxyribose makes DNA more chemically stable than RNA. It seems likely that RNA represents an earlier form of information storage, and DNA is the new improved version.

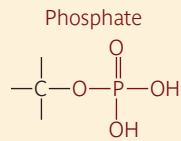


2-Deoxyribose

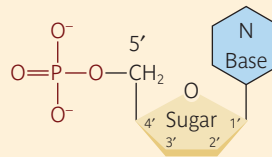


Ribose

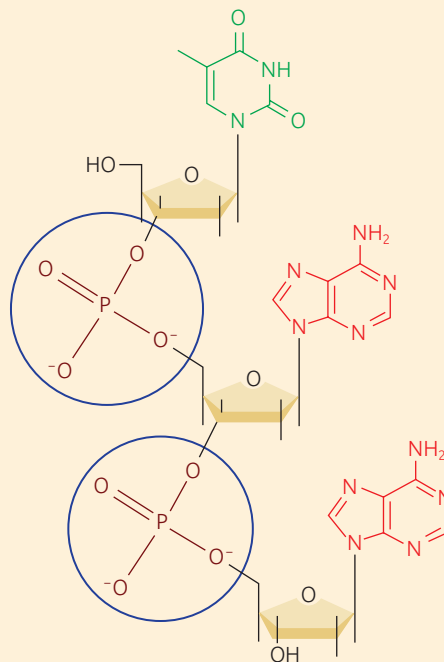
**Phosphate:** Phosphate molecules provide the structural ‘glue’ that holds the DNA backbone together, because they form phosphodiester bonds (strong covalent bonds) which link the phosphate and sugar molecules.



**Putting it all together:** The combination of base + sugar + phosphate is a nucleotide, the basic structural unit of DNA.

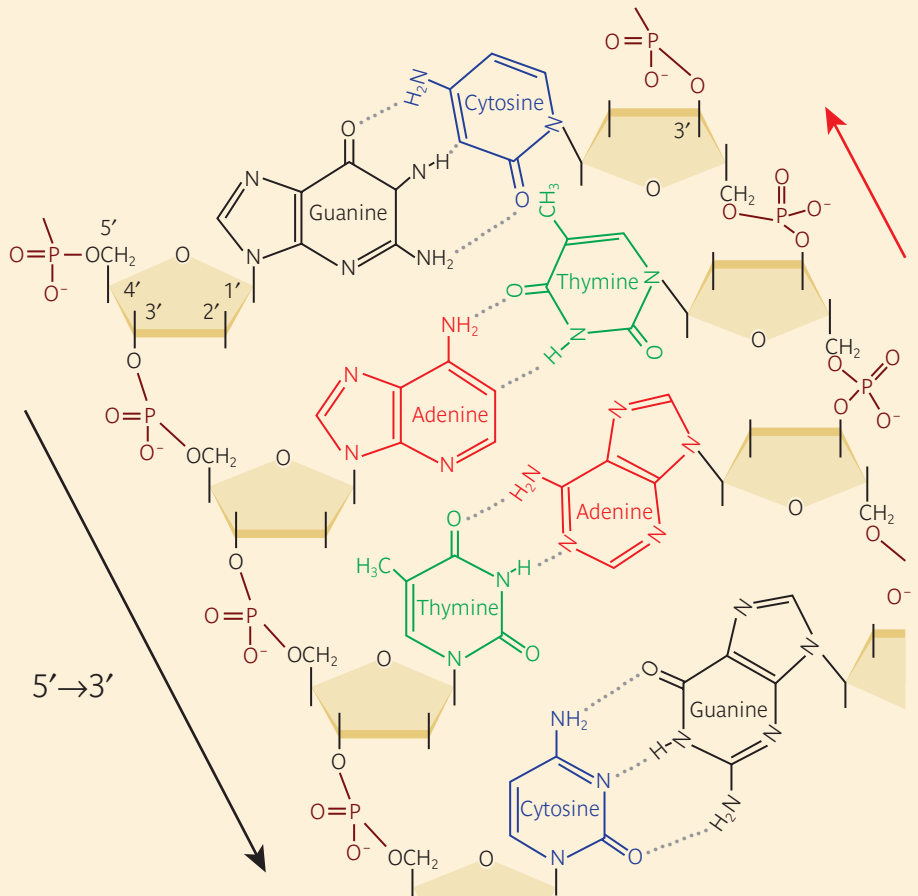


Nucleotides can form spontaneously under certain conditions, but linking them together into a DNA strand takes energy and specialized equipment (in the form of enzymes). A DNA polynucleotide strand is built by creating a phosphodiester bond that links the 3' carbon on the sugar of the growing chain with the phosphate attached to the 5' carbon of an incoming nucleotide. So the backbone of a polynucleotide strand is made of linked sugar–phosphate–sugar–phosphate, with one base joined to each sugar molecule.

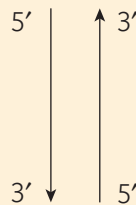


**Base pairing:** We have considered how the subunits of DNA – bases, sugars, and phosphates – link together to form a linear polynucleotide. Now we can consider how two strands fit together to make the famous double helix. If two polynucleotide strands face each other, the sugar–phosphate backbone runs down each side, and the bases stick into the middle,

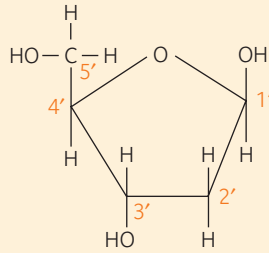
like the steps of a spiral staircase. Complementary pairs of bases can spontaneously form hydrogen bonds: three bonds between a C and a G, two between an A and a T. Each pair consists of one double-ring purine and one single-ring pyrimidine, so the complementary base pairs maintain an even 'step' width between the two sugar-phosphate strands.



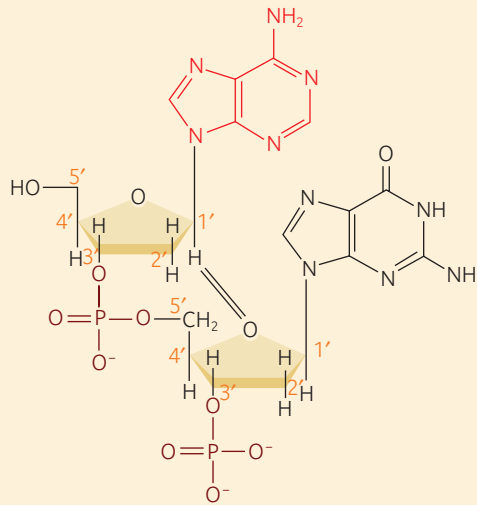
**What does 5' to 3' mean?** If you look closely at the double strand of DNA in the above figure, you can see that the strands are not mirror images of each other. They run antiparallel, which means that one strand is upside down with respect to the other.



This may be hard to see at first, but you can use the numbering on the sugar rings to spot the difference. Each ribose has five carbon atoms. The carbon attached to the base is 1' (pronounced 'one-prime'). Counting around the ring, the 5' (five-prime) carbon is the one attached to the phosphate group of the nucleotide.

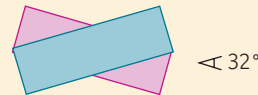
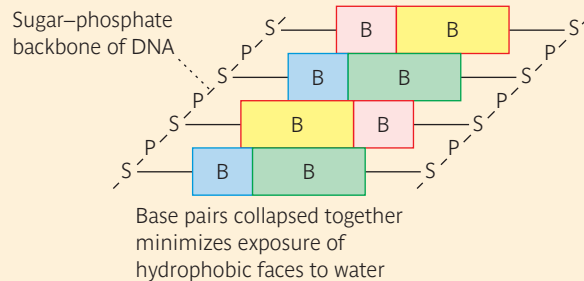
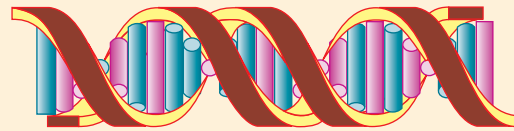


The 3' (three-prime) carbon is the one that forms the phosphodiester bond to link to the phosphate group on the neighbouring nucleotide.



So, looking at the double-stranded diagram on p. 43, if you follow the series of connections in the two polynucleotide strands from top to bottom, the left-hand DNA strand runs 5' to 3' (phosphate connected to 5' of sugar, which is connected at 3' to the next phosphate, and so on), but its matching sister strand runs 3' to 5' (phosphate connected to 3' of sugar, which is connected at 5' to the next phosphate). By convention, when the sequence of bases in DNA is written down, it is usually given from the 5' end and moving to the 3' end (of course, if you have the base sequence on one strand you can work out the other strand using the base-pairing rules). So the sequence of bases in this short section of DNA would be written 'GATC'. 5' to 3' will also be important when we look at DNA replication (see Chapter 4).

**Why is DNA a helix?** Nucleotide bases attract each other along the edges, forming hydrogen bonds between the base on one strand and its matching base on the other strand. So why doesn't DNA form a simple ladder, with sugar-phosphate uprights and straight base-pair rungs? The flat faces of the bases are hydrophobic, so they repel water. If DNA was a simple ladder, the gaps between the 'rungs' would leave the bases exposed to water molecules, making the whole structure unstable. But if the bases are stacked not directly on top of each other but offset slightly and rotated, base pairs can fit snugly on top of each other and minimize the destabilizing effect of water molecules. Because each base-pair 'rung' turns at 32° from the previous pair, the double-stranded DNA molecule makes a complete turn every ten base pairs.



Figures from Elliott, W.H. and Elliott, D.C. (2005) *Biochemistry and Molecular Biology*, Oxford University Press.

expose the base sequence of the gene. An existing RNA polymerase enzyme then uses the DNA template to make an RNA copy of the gene.

RNA is more or less the same as DNA but there are a number of differences. RNA is single-stranded (not double-stranded like DNA), uses a different sugar molecule in its backbone (ribose not deoxyribose), and one of the four bases is slightly different (instead of thymine (T) it uses uracil (U)). The RNA copy of the gene is made by matching each base on the exposed DNA strand to its complementary RNA base. Where there is an A in the gene (DNA), it is matched by a U in the message (RNA), a T in the gene is matched by an A in the message, a G with a C, and a C with a G. So the DNA sequence of the human RPB2 gene given in **Figure 2.2** begins 'CGTGATGGT. . .', but its complementary RNA would read 'GCACUACCA. . .' (**Figure 2.5**, p. 46).

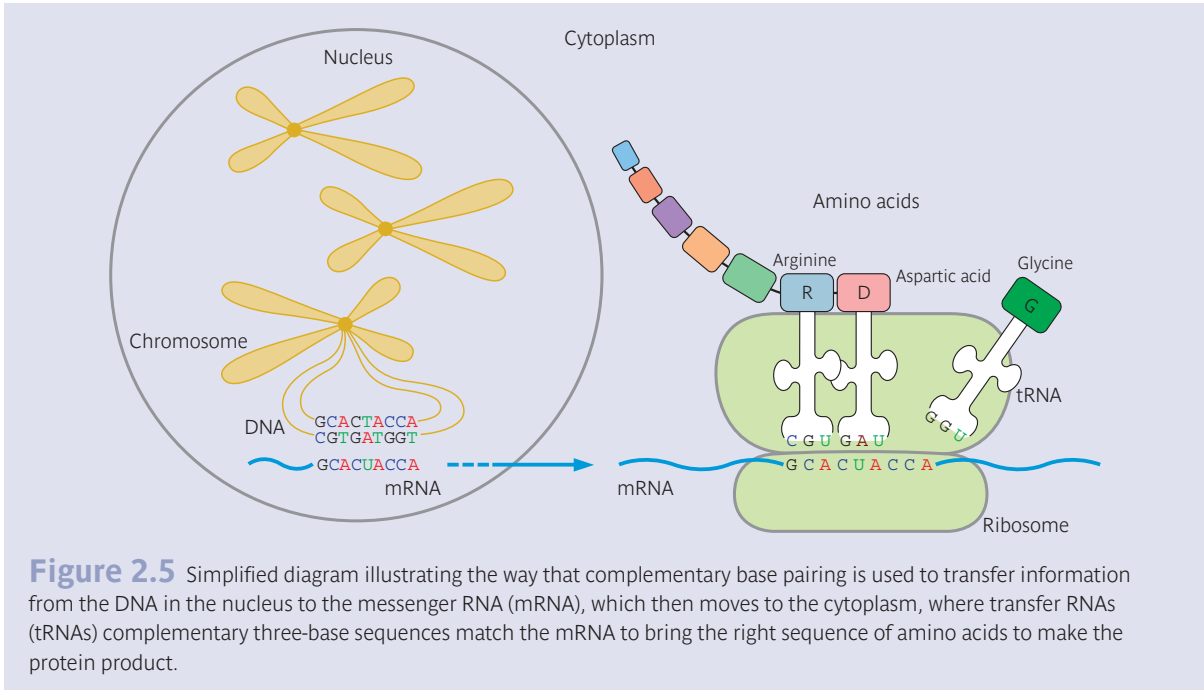


Complementary base pairing is also covered in the DNA replication chapter (Chapter 4)

DNA is stuck in the nucleus. But RNA can move from the nucleus to the cytoplasm. Because this RNA strand is made by complementary base pairing it contains the same information as the DNA strand it was copied from. This RNA strand is known as messenger RNA (mRNA) because it can take the information from the nucleus to the cytoplasm where it can be used to build useful things.

When messenger RNA leaves the nucleus, it is taken to a ribosome. Ribosomes are the workbenches of the genomic system, where information from the nucleus, transported in messenger RNA, is used to construct a protein. Here, complementary base pairing is used again to translate the information in the messenger RNA into the amino acid sequence of the protein product. The sequence of bases in the messenger RNA is matched to bases on transfer RNAs, each of which brings a specific amino acid to the ribosome.

There is a host of transfer RNAs (tRNAs) in the cytoplasm. Each tRNA has a particular three-base recognition

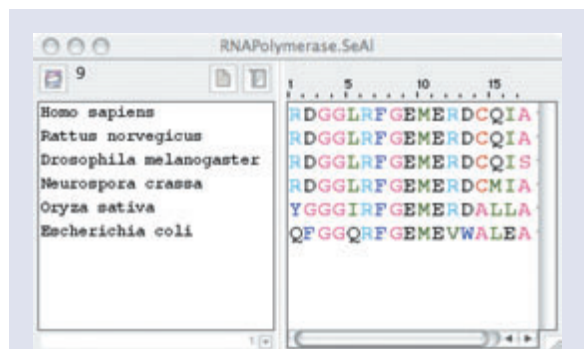


**Figure 2.5** Simplified diagram illustrating the way that complementary base pairing is used to transfer information from the DNA in the nucleus to the messenger RNA (mRNA), which then moves to the cytoplasm, where transfer RNAs (tRNAs) complementary three-base sequences match the mRNA to bring the right sequence of amino acids to make the protein product.

sequence and carries a specific amino acid. So the bases CGT in the *RPB2* gene are transcribed to GCA in the mRNA, which is matched by a tRNA with the recognition sequence CGU, and carries the amino acid arginine. Similarly, the next three letters in the gene are GAT, matching CUA in the mRNA and a tRNA with the sequence GAU, which carries the amino acid aspartic acid to join the protein. In this way the base sequence in the gene (DNA) determines the base sequence in the message (mRNA) that matches the recognition sequence of a particular tRNA which determines the sequence of amino acids in the protein.

**Figure 2.6** shows the DNA sequence given in **Figure 2.1** translated into amino acids (one amino acid for every three bases of DNA; **TechBox 2.3**).

The amino acids are attached to the growing peptide chain in the order specified by the gene. When all of the bases in the message have been 'read', the ribosome falls away from the message, releasing the chain of amino acids. The forces of attraction and repulsion between the amino acids in the chain cause parts of the sequence to spontaneously fold into energetically stable helices and sheets, which then twist around each other to form a



**Figure 2.6** Part of the amino acid sequence of the beta subunit of RNA polymerase II, from five different species. The amino acid sequence shown here was translated from the DNA sequence shown in **Figure 2.2**. Each letter stands for a single amino acid, so R is arginine, D is aspartic acid, G is glycine, and so on (see **TechBox 2.3**).

TECHBOX  
2.3

## Genetic code

## KEYWORDS

amino acid  
translation  
protein  
gene  
codon  
redundancy  
stop codon

## FURTHER INFORMATION

You can read the intricacies of the amino acid notation rules at [www.chem.qmul.ac.uk/iupac/AminoAcid/A2021.html](http://www.chem.qmul.ac.uk/iupac/AminoAcid/A2021.html)

A list of genetic codes can be found on the NCBI website: [www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi](http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi)

RELATED  
TECHBOXES

**TB 6.3:** Multiple alignment

**TB 2.2:** DNA structure

RELATED  
CASE STUDIES

**CS 3.1:** Viruses within (identifying protein coding sequences)

**CS 5.1:** Sweet and sour (identifying protein-coding sequences)

There are 20 different amino acids commonly found in proteins. But DNA only has four different nucleotides with which to specify all of the amino acids. So the nucleotide sequence in a protein-coding gene is read in triplets: three-letter ‘words’ that each specify a particular amino acid. These triplets of nucleotides are called codons. With the four-letter alphabet of DNA, there are 64 possible three-base codons, which is more than the number of amino acids, so many amino acids are represented by more than one codon. Having multiple codons specifying the same amino acid is known as the redundancy of the genetic code. This redundancy leads to some interesting patterns of sequence evolution (see Chapter 8).

There are a number of possible ‘start’ codons which specify the beginning of a protein (e.g. ATG, which also codes for the amino acid methionine). There are also several stop codons. The tRNAs that match stop codons do not carry an amino acid, so when a stop tRNA attaches to the messenger RNA, amino acid chain elongation stops and protein synthesis is finished. A mutation that causes a stop codon to occur in the middle of a gene will give rise to a truncated protein product which will probably be non-functional **Figure TB2.3b** gives all 64 possible codons with the corresponding amino acid for the ‘universal genetic code’. This is the code used in the majority of genomes. However, there are minor variants of this code (for example, mitochondrial genomes use a slightly different code; **Figure TB2.3a**).

Amino acids can be represented by their name (e.g. isoleucine), a three-letter abbreviation (Ile), or a single letter (I). The three-letter abbreviations are written with a capital and two lower-case letters: e.g. Thr for threonine or Asn for asparagine. When an international convention for symbols for the genetic code was established in 1968, there was an attempt to make the single-letter symbols of amino acids memorable using mnemonic associations. Some of the single letter codes are obvious, such as C for cysteine or M for methionine. Where more than one amino acid shared the same starting letter, it was given to the most commonly used amino acids: e.g. A for alanine but R for arginine. Other letter assignments are phonetic, like F for phenylalanine. And then the assignments start getting tenuous: for example, W for tryptophan because it’s a big letter for a big double-ring molecule. U and O weren’t assigned because they can be confused with V and zero, and J was left out because its not used in some languages.

**Echinoderm Mitochondrial**

UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	Stop	UGA	Trp
UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	Ile	ACA	Thr	AAA	Asn	AGA	Ser
AUG	Met	ACG	Thr	AAG	Lys	AGG	Ser
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

Done

**Mycoplasma**

UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	Stop	UGA	Trp
UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

Done

**Vertebrate Mitochondrial**

UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	Stop	UGA	Trp
UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	Met	ACA	Thr	AAA	Lys	AGA	Stop
AUG	Met	ACG	Thr	AAG	Lys	AGG	Stop
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

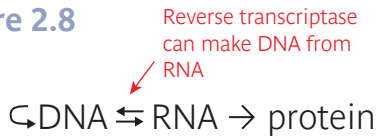
Done

**Figure TB2.3a** Spot the difference: three examples of variations on the genetic code, shown here as screenshots from the alignment editor Se-AL. This program allows you to select any of 14 different genetic codes then toggle between the nucleotide sequence and amino acid sequence of a gene. As an aside, mycoplasmas belong to a charmingly named class of bacteria, the mollicutes, which contain the smallest and simplest single-celled organisms. Mycoplasmas have the smallest genomes of any non-virus organism, with as few as 580,000 bases (only twice as large as the longest human gene). DNA analysis suggests that mollicutes are secondarily simple, having lost many characteristics such as a cell wall as they adapted to a parasitic lifestyle.

Reproduced by permission of Andrew Rambaut.



Figure 2.8

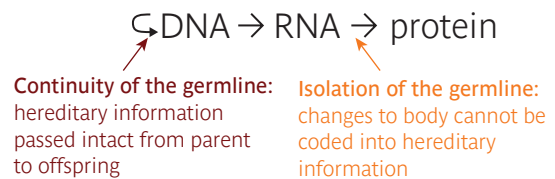


One of the important corollaries of the central dogma is this: the information in DNA and RNA is effectively interchangeable – DNA can be used to make a complementary RNA strand and vice versa (a fact exploited by many molecular technologies) – but the same is not true for the sequence of amino acids in a protein. Proteins don't have template copying. The sequence of amino acids in one protein cannot cause the formation of an identical copy of that sequence. Even prions, infectious proteins such as the one responsible for 'mad cow disease' (bovine spongiform encephalitis, or BSE) are thought to work by changing the conformation of existing prion proteins, rather than creating new copies of themselves. There is no known biochemical mechanism for translating the sequence of amino acids in a protein into the nucleotide code of a gene.

So the central dogma is a molecular statement of Weismann's barrier: information flows from the ger-

mline (DNA and RNA) to the soma (protein) but not back the other way (Figure 2.9). If the DNA sequence of a gene is changed, it may result in the formation of a protein with a novel sequence, and if that novel protein is advantageous, then carriers of that gene might be positively selected. But if a change is made directly to a protein, for example if the wrong amino acid is inserted as the protein is being constructed, then even if that change is advantageous, it cannot be coded back into the gene, so it is unlikely to be passed on down the generations.

Figure 2.9



Natural selection of genetic variants is discussed at more length in Chapter 5

## → Ubiquity of DNA

The template copying of DNA explains Weismann's principle of continuity: genetic information is copied from generation to generation because DNA can be faithfully replicated by complementary base pairing. The central dogma of molecular biology explains Weismann's theory of the isolation of the germline: changes to the body cannot be written back into an individual's DNA. We now know that the germline cells do not have to be physically isolated to preserve Weismann's barrier. The hereditary information is present in all cells, but the one-way flow of information from DNA and RNA to proteins means that the information in the genome is fundamentally unaffected by changes to the body.

The continuity and isolation of the genome make DNA a particularly handy molecule for biologists. If pangenesis was true, then each cell of the body could only provide information about the particular tissue it was drawn from. In that case, we could only read the com-

plete set of genetic information for an organism if we obtained sperm or eggs or an undifferentiated zygote, where all the hereditary information was collected together. As it is, every cell (with a few exceptions) contains an entire copy of the genome. This means that we can collect nearly any biological tissue and extract DNA from it. Incidentally, this fact also makes cloning possible, because each cell carries the instructions for the construction of the whole organism (see [TechBox 6.4](#)).

### How to get DNA samples

Typically, biologists working with DNA data will sample tissue from living individuals; for example, a seed, a blood sample, or, for the more unfortunate study animal, a piece of liver. Some types of specimen are easier to extract DNA from than others (see [TechBox 2.4](#)). Generally, the more biological material, and the fresher

## TECHBOX 2.4



# DNA extraction

### KEYWORDS

lysis  
purification  
nuclease  
proteinase  
DNAase  
centrifuge

### FURTHER INFORMATION

A number of websites provide simple instructions for extracting DNA at home or in the laboratory: for example <http://gslc.genetics.utah.edu/units/activities/extraction/>.



### RELATED TECHBOXES

**TB 4.2:** DNA amplification

**TB 6.4:** Cloning and conservation



### RELATED CASE STUDIES

**CS 2.1:** On the origin of faeces (DNA surveys)

**CS 3.1:** Glorious mud (DNA hybridization)



DNA is present in most biological samples. But before DNA can be sequenced, it must be isolated and purified. For some samples, DNA extraction is routine and reliable. For others, successful DNA extraction is an art that requires endless patience and a great deal of tinkering in the laboratory. The details of DNA-extraction techniques will vary from lab to lab, and different procedures will work best for particular samples. For this reason, there are a very large number of extraction protocols (which are essentially laboratory recipes) available on the internet, in scientific journals, and in laboratory manuals. However, all extraction protocols follow the same basic steps of cell lysis (to free the DNA), nuclease inactivation (to prevent DNA breakdown), and purification (to remove non-DNA molecules).

**1. Cell lysis:** First, sample tissues have to be broken up. For example, a leaf may be frozen in liquid nitrogen then pounded in a mortar and pestle, or a piece of liver might be pulverized in a blender. Then the crushed material is spun or strained to remove extraneous material, leaving just disassociated cells. A chemical (such as proteinase K) is added to the cells to burst the cell walls and release the DNA. Once cell lysis is complete, the digested material can be spun in a centrifuge, so that the cellular debris sinks to the bottom of the sample, permitting a liquid containing the DNA to be pipetted off the top. While it is possible to isolate organelle DNA (from mitochondria or chloroplasts) from nuclear DNA, it is more usual for all cellular DNA to be mixed together.

**2. Nuclease inactivation:** Enzymes that degrade DNA (DNases) are present in most biological samples. Various chemicals must be added to stop these enzymes from destroying the DNA in the sample, such as SDS (sodium dodecyl sulphate), EDTA (ethylenediaminetetra-acetate), and proteinase K.

**3. DNA purification:** The DNA solution contains other biological molecules such as lipids, polysaccharides, and proteins, which need to be removed. This is usually done by phenol-chloroform extraction or by running the solution through a column to separate the DNA from the other components. The DNA can then be precipitated out of solution using ethanol. The condensed DNA is usually extracted by a quick spin in the centrifuge which leaves the DNA in a pellet at the bottom of the test-tube. More dramatically, a glass rod or stick can be used to spool long strands of DNA out of the solution (**Figure TB2.4**).

**DNA extraction at home:** It is possible to carry out all of these steps using commonly available materials, for example using household detergent to lyse cells, contact lens cleaning solution to inactivate nucleases, and rubbing alcohol to purify the DNA. With a bit of experimentation you can produce clearly visible strands of DNA.

**Figure TB2.4** It is possible to produce visible white strands of DNA from a biological sample, as shown here in this photo, where DNA is being spooled out of the solution.

© Edward Kinsman/Science Photo Library.

it is, the easier it is to extract DNA. However, usually only a small sample is needed, such as a few grams of tissue or a few millilitres of blood. Sometimes DNA sampling is destructive, resulting in the demise of the sampled individual. For example, an entire beetle may be ground up for DNA extraction. Other organisms may survive being partially sampled, such as taking a leaf from a plant.

The harm done to study organisms by DNA sampling is an important ethical consideration. For example, it has been shown that the practice of toe-clipping amphibians – a common means of marking and taking tissue samples from captured individuals – can reduce their probability of survival. The distress caused to animals by being captured and handled to take a blood sample, or the destruction to habitat created by the search for elusive organisms, should not be underestimated. For many species, non-invasive methods of DNA sample collection are becoming widely advocated, such as using hair-traps, collecting faeces, or finding cast-off skin or feathers. Although these techniques may provide poorer-quality samples than destructive sampling, they can sometimes provide a practical way of collecting DNA and may even provide a rich source of data on individuals' movement and behaviour (see [Case Study 2.1](#)).

## DNA extraction

DNA extraction is generally best performed on fresh tissue, because DNA, like most biomolecules, degrades over time. DNA degradation can be reduced by preserving samples, particularly by freezing or immersion

in ethanol. But although fresh material is the easiest to work with, DNA samples have been successfully taken from Egyptian mummies, frozen mammoths, dehydrated penguins, pickled thylacines, carved whale teeth, preserved food, ancient timber, and 1000-year-old marine sediments. Even the last meal of Ötzi, the 'iceman' whose 5000-year-old frozen body was found in the European alps, has been determined through DNA analysis of his intestinal contents. Museum and herbarium specimens are proving particularly valuable for DNA analysis, although understandably many curators are not terribly keen on having pieces taken out of rare specimens. However, the inevitable decay of DNA means that there is little chance of recovering DNA from very old biological specimens.



**DNA from preserved specimens (ancient DNA) is discussed in [Case Studies 2.2](#) and [5.2](#)**

DNA extraction involves three basic steps: breaking up the cells to release the DNA, halting the action of enzymes that would destroy the DNA, then separating the DNA from the other cellular components. However, biological samples vary widely in their structure and contents, so most biologists find they have to fiddle with DNA-extraction protocols to get the best results for their samples. If a serious amount of fiddling in the lab is needed to make a particular sample yield useable DNA, then DNA extraction can seem like more of an art than a science, but the degree of elation when DNA is successfully extracted is often proportional to the amount of time spent trying to get it to work ([TechBox 2.4](#)).

## CASE STUDY

# 2.1



# On the origin of faeces: using DNA from scats to survey endangered lynxes

### KEYWORDS

DNA extraction  
species identification  
conservation  
contamination  
primers  
controls  
PCR



### RELATED TECHBOXES

**TB 2.4:** DNA extraction  
**TB 4.3:** Primer design



### RELATED CASE STUDIES

**CS 1.2:** DNA surveillance (monitoring endangered species)  
**CS 6.2:** Keeping the pieces (DNA and conservation)

Pires, A.E. and Fernandes, M.L. (2003) Last lynxes in Portugal? Molecular approaches in a pre-extinction scenario. *Conservation Genetics*, Volume 4, pages 525–532

“DNA methods on scats, or ‘molecular scatology’, are a relatively new means of gleaning information from animals that may be otherwise difficult to survey.”<sup>1</sup>

### Background

When a species becomes critically endangered, accurate estimates of the number of surviving individuals are crucial to developing effective conservation strategies. But, naturally, the more endangered an animal gets, the harder it is to spot. Indirect methods of surveying are often the only feasible approach. The presence of endangered animals can in some cases be established by finding scats (faeces). Scats can often be identified to species by characteristic shape and size, but this visual identification is not always reliable<sup>1,2</sup>.

### Aim

The Iberian lynx (*Lynx pardinus*), the most endangered cat species in the world, is close to extinction in the wild. There may be as few as 100 individuals left in the wild, in several isolated populations. In Portugal, the Iberian lynx are now so scarce that they are almost never seen. This study aimed to determine whether there are enough individuals remaining in the wild in Portugal to form a viable population.

### Methods

Potential lynx habitats were surveyed between 1997 and 2000, and a total of 104 putative lynx scats were collected. The tip of each scat was removed and placed in a plastic vial containing silica isolated with porous tissue, and the rest of the scat was frozen in plastic bags. Faeces are a relatively poor source of DNA, so choosing the best DNA-extraction method is critical to success. The authors tested a number of DNA-extraction methods (**TechBox 2.4**) and found that commercial DNA-extraction kits gave the best results. Following standard procedures for these kits (with some modifications), faecal material taken from the outside of the scats (the part most likely to have picked up cells on its way out of the lynx) was lysed, incubated with proteinase K, washed, and purified. Extractions were performed with blank controls to test for contamination. They used lynx-specific primers (sequences from mitochondrial genes known to match only lynx DNA, not other species: **TechBox 4.3**) in a PCR analysis to amplify any target DNA from the scats (**TechBox 4.2**). The PCR was also performed on scats from captive lynx to provide positive controls (to prove that the methods can amplify lynx DNA when it is present). Dog and wildcat scats were used to provide negative controls (to test whether the methods could give a false result in the absence of lynx DNA).



**Figure CS2.1a** The Iberian Lynx (*Lynx pardinus*) has the dubious distinction of being the most critically endangered cat species. Lynx populations were reduced by trophy hunting, trapping, and habitat loss. Development and road building have contributed to the fragmentation of the remaining lynxes into very small, isolated populations. With somewhere between 30 and 300 breeding females left in the wild, much hope is being placed in captive breeding programs to build up viable populations of Iberian lynx. Iberian lynx news, including latest counts of wild lynxes and announcements of cubs born in captivity, can be found at [www.iberianature.com/material/iberianlynxnews.htm](http://www.iberianature.com/material/iberianlynxnews.htm).

© John Cancalosi/ Auscape International.

### Results

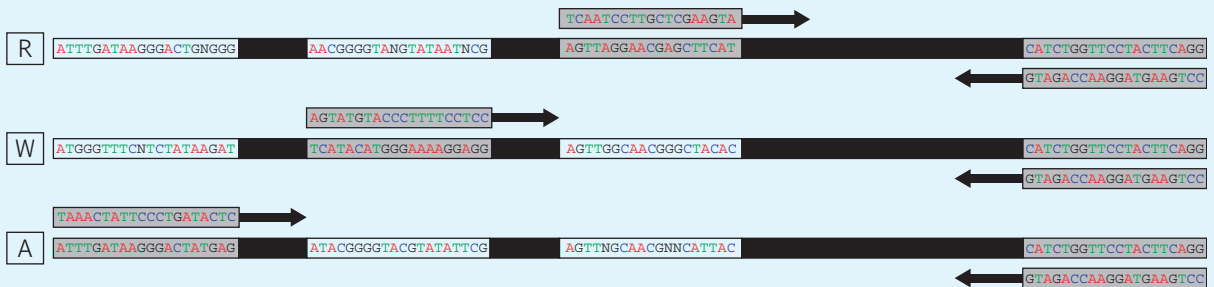
Of 104 scat samples, 95 were found to contain DNA. However, only two of these could be shown to be lynx DNA. This result provides the first conclusive proof that there have been lynx in the area in 1997. However, with only two positive samples, it is difficult to tell whether these individuals were visitors, or the last members of a population, or whether this survey method underestimates lynx numbers. Some samples showed evidence of cross-contamination, highlighting the importance of checks for contamination.

### Conclusions

The authors conclude that lynx have virtually disappeared from Portugal. If they are present, they are in such low numbers that they would be unable to maintain viable populations.

### Limitations

The problem of negative results – concluding absence of lynx due to failure to amplify lynx DNA – is an issue for remote surveying using DNA-poor samples such as faeces. Here, the authors used control samples, taken from captive lynx faeces, to demonstrate that their technique can detect lynx DNA from scats. Methods using multiple primers may be able to give rapid results and more protection against false negatives. For example, Dalén and



**Figure CS2.1b** Illustration of the concept of RCP-PCR. One of the three species-specific primers will react with the general primer. The resulting fragment size depends on whether red fox (R), wolverine (W), or arctic fox (A) DNA is present in the extract. Intraspecific variable sites in the template are shown as (N). From Dalén *et al.* (2004)<sup>3</sup>.

coworkers developed a *rapid classificatory protocol PCR* (RCP-PCR), where several different primers are added to the same sample: a universal primer that binds to all mammal species, and a number of species-specific primers, producing different-size fragments characteristic of each species (**Figure CS2.1b**)<sup>3</sup>.

### Future work

'Molecular scatology' has the potential not only to survey the presence or absence of particular species in a given area, but also to provide a picture of the activity of species in space and time. For example, DNA sampling from scats has been used to show how spatial overlap between arctic and red foxes changes between seasons<sup>4</sup>. DNA analysis of faeces can identify both predator and prey, revealing the species that made up the carnivores' most recent meals<sup>5</sup>, and even allowing individual differences in prey-preference to be studied<sup>6</sup>.

### References

1. Davison, A., Birks, J.D.S., Brookes, R.C., Braithwaite, T.C. and Messenger, J.E. (2002) On the origin of faeces: morphological versus molecular methods for surveying rare carnivores from their scats. *Journal of Zoology*, Volume 257, pages 141–143.
2. Pires, A.E. and Fernandes, M.L. (2003) Last lynxes in Portugal? Molecular approaches in a pre-extinction scenario. *Conservation Genetics*, Volume 4, pages 525–532.
3. Dalén, L., Götherström, A. and Angerbjörn, A. (2004) Identifying species from pieces of faeces. *Conservation Genetics*, Volume 5, pages 109–111.
4. Dalén, L., Elmhagen, B. and Angerbjörn, A. (2004) DNA analysis on fox faeces and competition induced niche shifts. *Molecular Ecology*, Volume 13, pages 2389–2392.
5. Deagle, B.E. *et al.* (2005) Molecular scatology as a tool to study diet: analysis of prey DNA in scats from captive Steller sea lions. *Molecular Ecology*, Volume 14, pages 1831–1842.
6. Fedriani, J.M. and Kohn, M.H. (2001) Genotyping faeces links individuals to their diet. *Ecology Letters*, Volume 4, pages 477–483.

CASE  
STUDY  
2.2

## More moa: using ancient DNA to catalogue diversity of extinct giant birds

### KEYWORDS

ancient DNA  
DNA extraction  
museum specimens  
genotyping  
phylogeny  
control region  
molecular clock



### RELATED TECHBOXES

**TB 7.3:** Bayesian inference  
**TB 8.4:** Molecular dating



### RELATED CASE STUDIES

**CS 6.1:** Barcoding nematodes  
(DNA taxonomy)  
**CS 8.1:** Same but different  
(defining species using DNA)

Baker, A.J., Huynen, L.J., Haddrath, O., Millar, C.D. and Lambert, D.M. (2005) Reconstructing the tempo and mode of evolution in an extinct clade of birds with ancient DNA: The giant moas of New Zealand. *Proceedings of the National Academy of Sciences USA*, Volume 102, pages 8257–8262

“ . . . Ancient DNA methods provide powerful tools for inferring the number of lineages, as well as the tempo and mode of evolution of entire extinct groups of animals. ”<sup>1</sup>

### Background

Found only in New Zealand, moa were a morphologically diverse family of birds belonging to the order Struthioniformes, which includes ostriches and emus and their allies (**Figure CS2.2a**). Currently six moa genera are recognized, ranging in size from the Giant Moa *Dinornis giganteus* (up to 250 kg, twice the size of an ostrich) to the relatively small Coastal Moa *Euryapteryx curtus* (20 kg, the size of a largish turkey). Moa went extinct not long after human settlement of New Zealand, less than 1000 years ago. However, moa bones are abundant, both in natural collections such as caves and swamps, and in middens (prehistoric rubbish dumps).

### Aim

Assignment of fossil remains to species is usually dependent on morphological similarity. Yet members of the same species may be morphologically very distinct, for example juveniles and adults, or males and females. Conversely, distinct non-interbreeding species may appear very similar from skeletal evidence alone. By sequencing DNA from a very large number of moa specimens, these biologists hoped to determine how many distinct lineages of moa had existed, and to explore reasons for the diversification of this endemic New Zealand group.

### Methods

Moa DNA samples were obtained by taking bone cores or shavings from 125 museum specimens. DNA was extracted from between 0.1 and 0.5 g of sampled bone using EDTA with proteinase K (**TechBox 2.4**). Samples were purified by phenol–chloroform extraction, then extracted DNA was amplified using PCR (**TechBox 4.2**). They checked for the presence of multiple bands that might indicate nuclear copies of mitochondrial genes. DNA was sequenced along both strands. DNA was amplified in laboratories in Canada and New Zealand. Identical sequences were obtained from specimens sequenced in both laboratories. Baker *et al.*<sup>1</sup> used 658 nucleotides of the mitochondrial control region to genetically type their specimens, because it has a rapid rate of molecular evolution, so is expected to differ between different species or populations. For a subsample of specimens, they used a longer alignment of 2184 nucleotides of mitochondrial genes to estimate a phylogeny and molecular dates. They used a Bayesian phylogenetic method (**TechBox 7.3**)



**Figure CS2.2a** Moa – large flightless birds from New Zealand – went extinct long before cameras were invented. This scene was posed in 1899 using a museum model in the botanic gardens being ‘hunted’ by medical students. One of the students (on the left) was Sir Peter Buck (Te Rangi Hiroa), who, amongst a great many other achievements, was the first Maori to qualify as a doctor. For more information on Te Rangi Hiroa see [www.nzedge.com/heroes/buck.html#FIRST](http://www.nzedge.com/heroes/buck.html#FIRST).

Image courtesy of Alexander Turnbull Library, New Zealand. Kehoe, E.L. Mock Moa Hunt – Photograph taken by Guy. 1899. Ref no:PACol-1308.

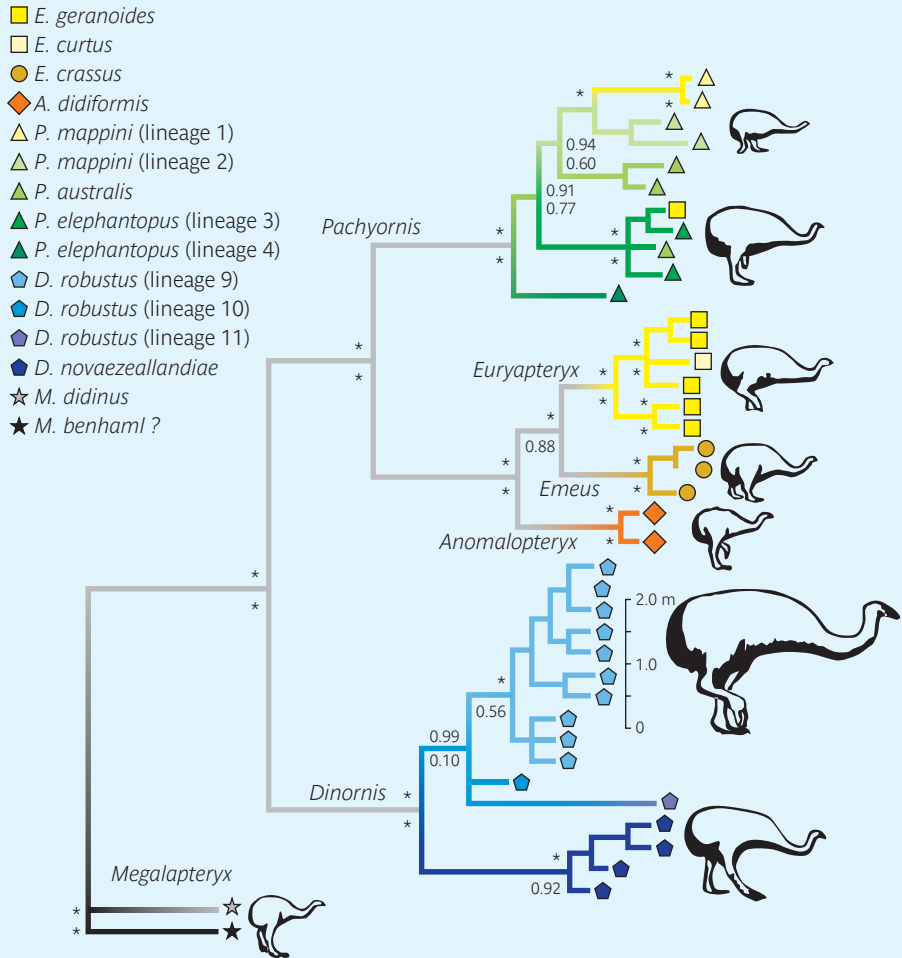
to estimate a phylogeny of the DNA samples, then used the genetic distance between samples (represented by branch lengths on the phylogeny in **Figure CS2.2b**), and the Bayesian probabilities on groupings, to judge which lineages represented distinct populations or species (**TechBox 6.2**).

### Results

Baker *et al.*<sup>1</sup> identified 14 distinct lineages of moa. Nine of these are currently recognized as species. The other five may be either geographically separate populations or newly recognized species. The phylogeny showed that some sequences did not group with the rest of the sequences for that species. For example, a sequence from a sample which had been labelled as a stout-legged moa *Euryapteryx geranoides* (yellow squares in **Figure CS2.2b**) is more similar to sequences from the heavy-footed moa *Pachyornis elephantopus* (dark green triangles). The authors suggest that as many as a third of the museum specimens had been incorrectly assigned to species.

### Conclusions

While other ratite lineages are relatively species-poor, the authors of this study suggest that moa diversity was as high as other classic island endemic radiations, such as Darwin’s



**Figure CS2.2b** Phylogeny (evolutionary tree) of moa, inferred from DNA extracted from moa bones. Moa bones are found in abundance in middens (prehistoric kitchen rubbish), suggesting moa were eaten by early colonists of New Zealand. A study that compared the species compositions of natural collections of bones in New Zealand (e.g. from animals that died from falling into caves) with those found in middens concluded that animals that were more often hunted and eaten were more likely to have gone extinct. From Baker A.J. *et al.* (2005) Reconstructing the tempo and mode of evolution in an extinct clade of birds with ancient DNA: The giant moas of New Zealand. *Proceedings of the National Academy of Sciences USA*, Volume 102, pages 8257–8262.

Reproduced by permission of Allan Baker, Royal Ontario Museum.

finches. On the basis of molecular dates, they propose that much of the diversification of moa lineages was relatively recent, possibly driven by geographic reshaping of New Zealand, 4–10 million years ago. Populations became isolated and ecologically specialized as mountains rose and islands separated (see Chapter 8 for a discussion of estimating evolutionary time from DNA sequence data).

### Limitations

Genetic difference is a continuous scale, and species definitions are a matter of opinion: what one biologist considers a separate species, another may consider a regional subtype. Some biologists dispute that molecular data are a valid way of identifying species, as morphologically identical populations may show consistent molecular differences (see Chapter 6 for a discussion of the use of molecular data to define species). Identification of misassigned specimens is a potentially valuable use of DNA data, but must be conducted with due recognition that DNA taxonomy is also prone to errors, through contamination, sequencing errors, or incorrect phylogenetic inference.

### Future work

DNA analysis could potentially result in changing the taxonomic assignment of many recently extinct taxa. For example, DNA data have been used to show that bones previously attributed to two co-occurring species of moa are actually from sexually dimorphic males and females of the same species (with the females twice the size of the males)<sup>2</sup>. Where abundant data are available, DNA analysis can provide a picture of the population biology of an extinct species. For example, DNA analysis has suggested that moa were present in much greater numbers than previously suspected<sup>3</sup>. The accuracy and precision of such estimates will be improved by increased sampling: more genes, longer sequences, more individuals, and broader geographic coverage.

### References

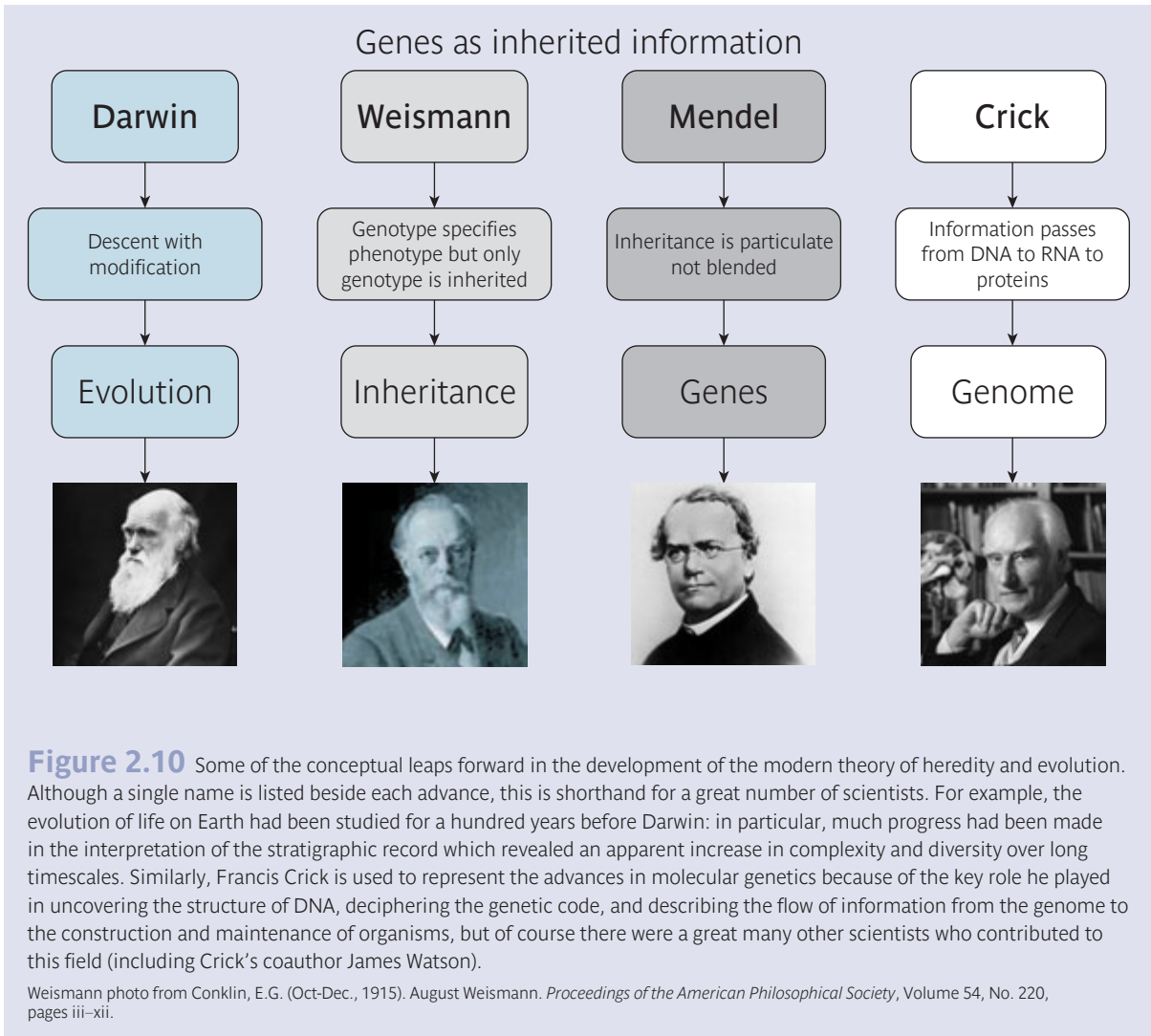
1. Baker, A.J., Huynen, L.J., Haddrath, O., Millar, C.D. and Lambert, D.M. (2005) Reconstructing the tempo and mode of evolution in an extinct clade of birds with ancient DNA: The giant moas of New Zealand. *Proceedings of the National Academy of Sciences USA*, Volume 102, pages 8257–8262.
2. Bunce, M. *et al.* (2003) Extreme reversed sexual size dimorphism in the extinct New Zealand moa *Dinornis*. *Nature*, Volume 425, pages 172–175.
3. Gemmel, N.J., Schwartz, M.K. and Roberston, B.C. (2004) Moa were many. *Proceedings of the Royal Society London B: Biology Letters Supplement*, Volume 271, pages S430–S432.



## Conclusions

The mechanisms of inheritance have been revealed in an astonishingly short period of time. From the point of view of someone interested in the role of DNA in evolutionary biology, there have been a number of notable leaps forward in understanding the nature of genetic information (**Figure 2.10**). It is convenient to think of these advances as connected to particular scientists, but of course, like all science, there were a great many other scientists who contributed to each of these fields.

Firstly, Charles Darwin noted that there was a great deal of heritable variation in natural populations, and the natural selection of successful variants provided a plausible mechanism for the evolution of adaptations. Without knowing about genes, Darwin clearly illustrated the role of heredity in the process of evolution by descent with modification. Secondly, August Weismann reasoned that the information needed to make an organism was passed from one generation to the next, essentially unaffected by bodily changes acquired during an individual's lifetime. Thirdly, Gregor Mendel showed through breeding experiments that this



variation was particulate: inheritance was controlled by genetic factors that were not blended but passed intact down the generations as discrete units of information. Fourthly, Francis Crick and others showed that complementary pairing of bases explains how DNA stores information, how it can be faithfully copied from one generation to the next, and how the genetic information is used to construct RNA molecules and proteins. This provided a molecular explanation for the observations of Darwin, Weismann, and Mendel: evolution acts on genes (information coded in DNA), which are then expressed in the body (through the action of RNA and proteins).

These scientists are four among many who have shown that understanding inheritance is the key to understanding evolution. But DNA is also incidentally a source of information about evolution. Genomes would not carry this information if every individual had exactly the same DNA sequence. It is because DNA sequences differ between individuals, between species, and

between evolutionary lineages that we can use DNA to understand evolutionary history and processes, as well as illuminating gene function itself. In the next chapter, we will consider the process of mutation, whereby the genome of an individual is permanently changed. Mutation creates heritable variation, which is the raw material of evolution. It also makes each one of us unique, a fact that is increasingly exploited in biology and medicine.

## Further information

A very readable account of the development of evolutionary biology can be found in:

**Young, D. (2007) *The Discovery of Evolution*, 2nd edn. Cambridge University Press.**

The development of ideas about heredity is outlined in:

**Cobb, M. (2006) Heredity before genetics: a review. *Nature Reviews Genetics*, Volume 7, pages 953–958.**

There are a number of excellent online resources covering the development of molecular genetics. It is worth exploring the excellent *DNA Interactive Timeline*, an animated history of genetics which includes many of the less-well known players:

[www.dnai.org/timeline/index.html](http://www.dnai.org/timeline/index.html)

Any biochemistry book will cover DNA and protein structure. The following provides a very accessible, biology-friendly account:

**Crowe, J., Bradshaw, T. and Monk, P. (2006) *Chemistry for the Biosciences: the essential concepts*. Oxford University Press.**

Animations of cellular processes such as DNA replication, transcription, and translation are available at:

[www.johnkyrk.com](http://www.johnkyrk.com)