dna: molecule of the century

If you were to nominate a 'molecule of the 20th century', one of the best candidates would surely be DNA. Although the science of genetics began around 1900, when the work of Mendel was rediscovered, it wasn't until the 1950s that techniques were developed allowing the structure and function of genes to be probed. The last quarter century has seen a revolution in molecular biology, bringing new knowledge that casts light on the essential unity of life on Earth.

DNA, genes and chromosomes

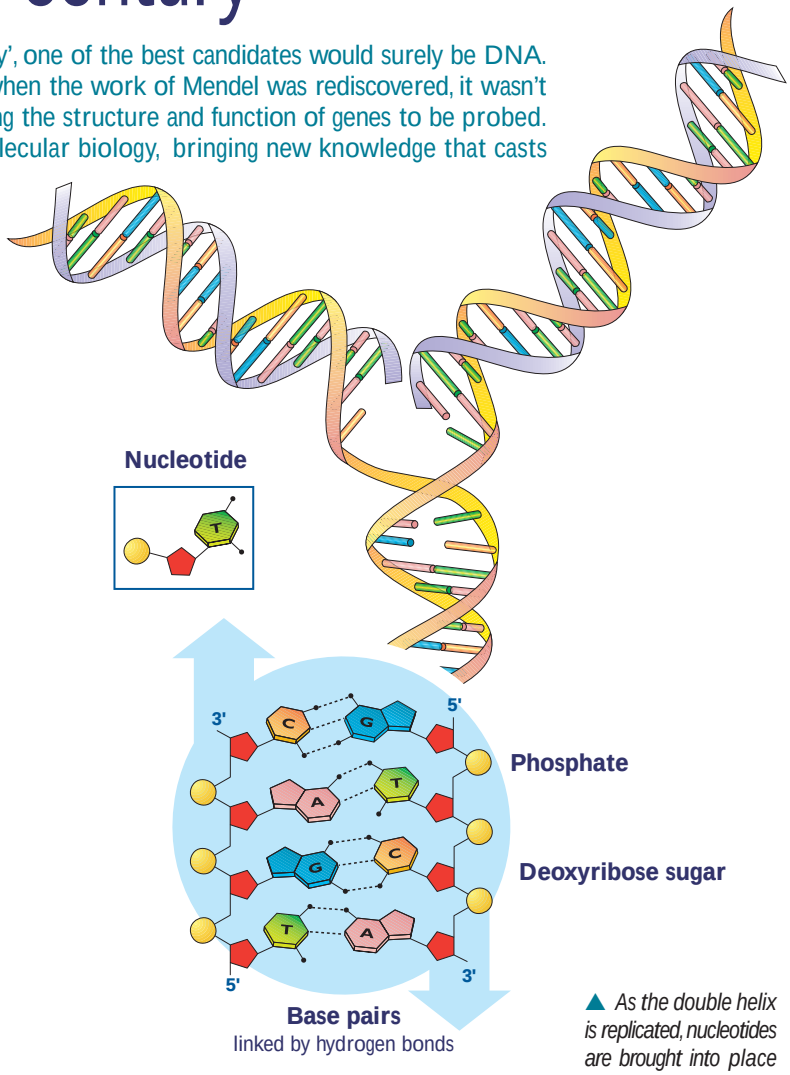
DNA (*deoxyribonucleic acid*), faithfully copied from one *cell* division to the next and passed on through the generations, forms a living link back to some of the first life on our planet. The DNA in our cells is both a record of where we've come from and a predictor of how we might develop, both as individuals and in subsequent generations. Although our environment plays a vital rôle in shaping development, DNA determines all inherited characteristics and is therefore of fundamental importance in understanding biology.

Most DNA is packaged, in fungal (including yeast), plant and animal cells, into *chromosomes* within the nucleus of each cell. Some DNA is also found outside the nucleus: in the *mitochondria* (which release energy for cellular activities); and within the *chloroplasts* (sites of photosynthesis) of plant cells. In bacteria, most of the DNA occurs in a single circular chromosome, although small rings of DNA, called *plasmids*, may also be present.

The *double helix* of DNA can be likened to a twisted rope ladder. The two intertwined helices are chains made from sugar and phosphate molecules linked together alternately. Attached to each sugar molecule is a nitrogenous '*base*'. There are four different bases in DNA: adenine (A); thymine (T); cytosine (C) and guanine (G). Relatively strong *hydrogen bonds* between the bases join the two strands of the double helix together like the rungs of a ladder. A always pairs with T, and C always pairs with G. This '*base pairing*' mechanism ensures identical replication of DNA strands during cell division.

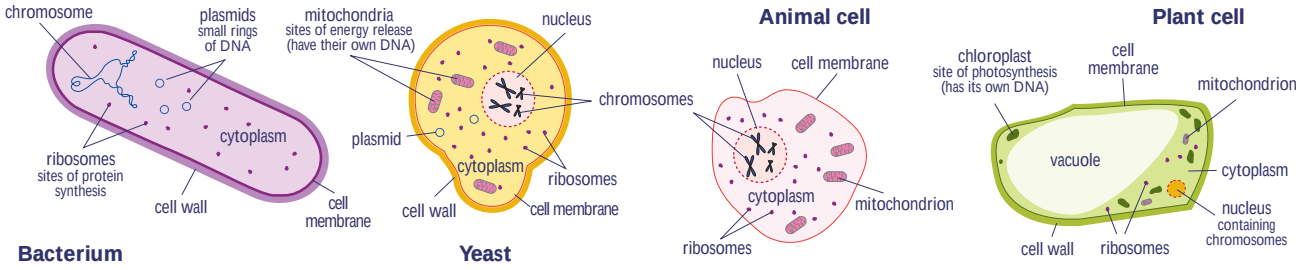
A *gene* is a length of DNA that determines the structure of all or part of a specific *protein*. How does this happen? Proteins are comprised of chains of *amino acids*. The sequence of bases on one DNA strand indirectly spells out the amino acids that are needed to make these chains. A short chain is called a *polypeptide*. A typical protein might be made up of several polypeptide chains linked together and folded into a precise three-dimensional structure.

Each amino acid in a polypeptide is encoded by three consecutive bases in the DNA. The code determining which amino acid each triplet of bases *translates* into — *the Genetic Code* — is virtually the same in all living things.



▲ *As the double helix is replicated, nucleotides are brought into place alongside those of the existing DNA strands. In this way each new double helix consists of a strand from the old molecule (shown here in yellow) and a completely new one (mauve). This type of replication is called 'semi-conservative'.*

cellular components



Also encoded within the DNA are sequences that regulate protein production. These are needed because although most cells of a particular organism contain the same DNA, only certain proteins are made at any one time or in any particular type of cell: that is, only certain genes are *expressed*.

Because the genetic code is universal, it is possible to transfer working genes between organisms — even between completely different species. The process of directly and deliberately transferring, removing or altering genetic information is commonly called *genetic modification* (or *genetic engineering*).

Protein synthesis

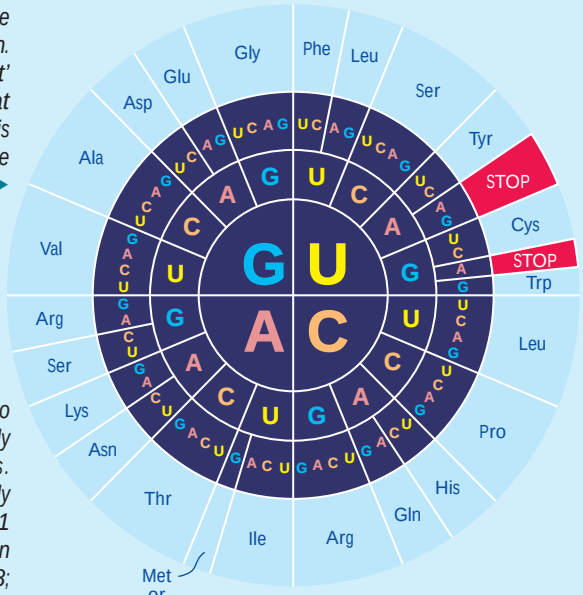
RNA — the genetic go-between

Most of a cell's DNA lies within its one or more chromosomes. However, proteins are not made there, but at special structures called *ribosomes*. Therefore before a particular protein can be made, a copy of the appropriate genetic instructions must first be *transcribed* from the DNA then ferried to the ribosomes. The copied instructions are made from *mRNA* (*messenger ribonucleic acid*), and assembled by an enzyme called *RNA polymerase*. This mRNA is virtually a 'mirror image' of the sequence of bases on one DNA strand, according to the base pairing rules.

On arrival at the ribosomes, the base sequence within the mRNA directs the construction of proteins from amino acids. A sequence of three adjacent bases (a *codon*) in the mRNA molecule determines each amino acid in a polypeptide chain. The amino acids are brought into place by *tRNA* (*transfer RNA*) molecules, which have *anticodons* mirroring the codons in the mRNA.

The Genetic Code

The Genetic Code indicates the amino acids that are encoded by each group of three mRNA bases or codon. There are also 'start' and 'stop' codons that control the synthesis of each polypeptide chain.



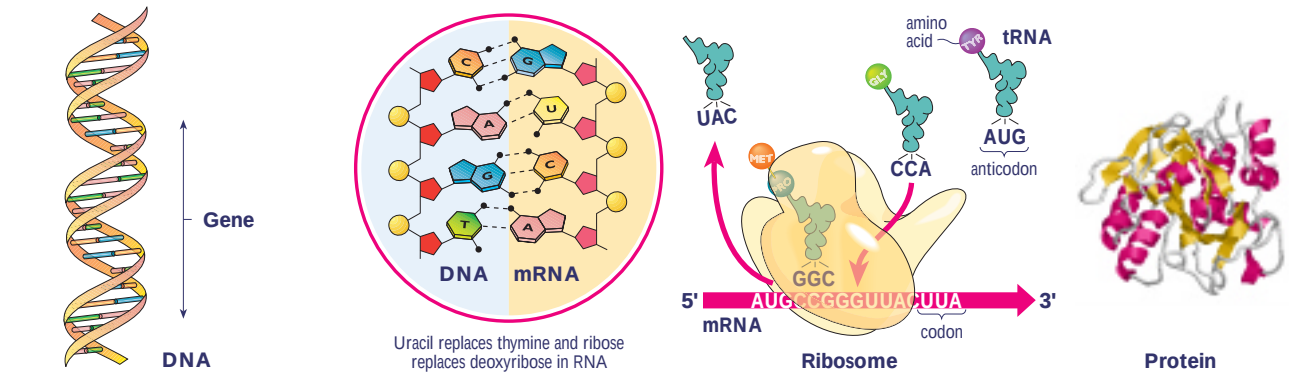
To decipher this chart, start at the centre and read outwards (5'→3').

There are 20 amino acids that commonly occur in proteins. Proteins vary greatly in size: insulin has 51 amino acids; human hæmoglobin has 578; urease (an enzyme) has 4 500.

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

*The sequence of amino acids in a protein is called its **primary structure**. The way in which each amino acid chain is twisted (e.g., to form coiled α-helices and flat β-pleated sheets) is its **secondary structure**. To show this here, the α-helices are coloured magenta; the β-sheets are yellow, with the arrows pointing to the C-terminus of the polypeptide chain. The amino acid chain(s) fold in on themselves to give a precise shape to the protein — its **tertiary structure**.*

protein synthesis



Genes are made of DNA
DNA determines hereditary characteristics by directing the formation of proteins.
A stretch of DNA carrying the instructions required to make all or part of a particular protein is called a gene.

DNA's message is transcribed to mRNA
DNA operates through an intermediary, RNA. The sequence of bases on the DNA molecule acts as a template for the production of messenger RNA (mRNA).

Amino acids are strung together at the ribosomes
The mRNA binds to the ribosomes, where it directs the production of the protein it encodes. Three mRNA bases in a row translate into one amino acid to be added to the protein.

The amino acid chain folds to form a protein
The amino acid chain(s) from which proteins are formed fold into precise three-dimensional structures, contributing to the protein's properties and function.

PDB ID: 1SVN

Enzymes — precise molecular tools

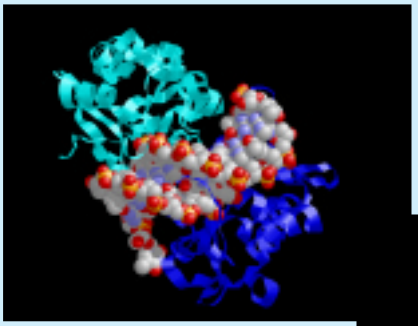
Restricting DNA

Special *enzymes* obtained from bacteria are an essential tool of the molecular biologist. In nature, these enzymes help bacteria to fend off viral attack by precisely dissecting the 'foreign' DNA of invading viruses. In this way, the proliferation of the *viruses* is restricted. *Restriction enzymes* (as they are known) 'recognise' and cut DNA molecules at specific locations. Many hundreds of different restriction enzymes have been isolated from microbes and are available commercially. These restriction enzymes allow almost any section of DNA, and so any given gene, to be isolated.

Sticky ends

Restriction enzymes leave blunt or staggered ('sticky') ends on the DNA sequences they cut. The end of one DNA molecule with a sticky end will readily link (by hydrogen bonding) to that of another cut with the same enzyme. To join two DNA molecules permanently, however, it is necessary to form stronger chemical bonds along the DNA's sugar-phosphate backbone. An enzyme called *DNA ligase* can do this job.

The function of restriction enzymes and ligase in cutting and pasting DNA molecules is obvious, but the genetic engineer's tool kit would be incomplete without one or two other enzymes.



▲ *Restriction enzyme BamHI, attached to a short length of DNA. Here, the DNA is shown as a space-filling model, while the two polypeptide chains of the restriction enzyme are depicted as blue ribbons.*
NDB ID: PDE020

▼ *The restriction enzyme BamHI, attached to a DNA molecule. This view, down the axis of the DNA, shows the symmetry of the enzyme. For clarity, the DNA is shown as a wireframe structure.*

Source microorganism STRAIN Name	Recognition site (5'→3')
Bacillus amyloliquefaciens H	
*Bam*HI	G↓GATCC
Escherichia coli RY13	
*Eco*RI	G↓AATTC
Bacillus globigii	
*Bgl*II	A↓GATCT
Haemophilus influenzae Rd	
*Hind*III	A↓AGCTT
Arthrobacter luteus	
*Alu*I	AGC↓T
Thermus aquaticus	
*Taq*I	T↓CGA
Haemophilus aegyptius	
*Hae*III	GG↓CC

◀ *Restriction enzymes bind to specific sequences of bases in double-stranded DNA molecules. They cut the DNA at or near these recognition sites, leaving fragments with blunt or staggered ends. The enzymes take their names from the various microbes from which they are obtained.*

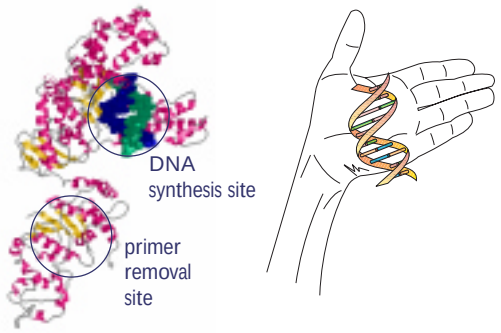
◀ *DNA ligase. ATP provides the energy for this enzyme to work.*
PDB ID: 1A01

DNA from RNA

The enzyme *reverse transcriptase* can be used to assemble a single strand of *complementary* DNA alongside a corresponding piece of mRNA. This is the reverse of the usual situation where DNA dictates the formation of RNA molecules.

Reverse transcriptase can be useful because cells that are making a particular protein will have many more copies of that protein's mRNA in them than the number of copies of the gene itself. It is therefore often easier to search for genetic information amongst these small mRNA molecules rather than along the entire length of the cell's chromosomal DNA.

the most accurate enzyme



PDB ID: 1TAU

The computer image above represents DNA polymerase from the hot-water bacterium *Thermus aquaticus* (*Taq*).

The enzyme is shaped like a hand, with the DNA under replication being held in the 'palm'. A small fragment of DNA is shown here with the newly-formed strand in green and the complementary template strand (against which it is being assembled) in blue.

Unusually, this enzyme has *two* active sites; the one in the lower part of the molecule removes the short 'primers', which are needed to start replication from the finished DNA molecule. DNA polymerases from other organisms go a step further, however, and have a *third* active site, which 'proof reads' the new DNA for errors as it is formed. Human DNA polymerase copies DNA so accurately that less than one error occurs in every billion bases.

Taq DNA polymerase is stable at high temperatures and is used in the polymerase chain reaction (see page 10).

DNA polymerase

Once a desired length of mRNA has been isolated, and the corresponding DNA constructed, a second enzyme — *DNA polymerase* — can be used to assemble a double-stranded helix, using the first DNA strand as a template. All living things make DNA polymerases, and a cell often has several different varieties of the enzyme: some for helping to replicate DNA during cell division and others that conduct day-to-day repair and maintenance.

In humans, several complex proteins act together to ensure that DNA is faithfully copied down the generations; in many other organisms, one multi-functional enzyme does it all.

Gene machines

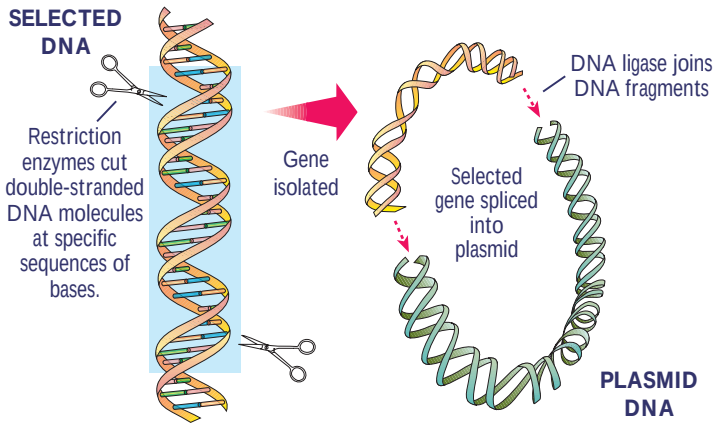
By the judicious use of restriction and other enzymes, molecular biologists are able to assemble DNA molecules that contain one or more genes of interest. Where a particular piece of DNA is difficult to isolate, it is sometimes possible to make it artificially using a *DNA synthesizer*. Under computer control, these machines string together the biochemical building blocks needed to make short stretches of DNA. Of course, to programme the synthesizer it is necessary to know the sequence of bases present in the desired gene; this too can be determined automatically using a *DNA sequencer* (see *Cycle sequencing*, page 13).

Plasmids as vectors

The construction of a suitable DNA molecule is only part of the story. Once this has been done it is necessary to move that molecule into a cell in which it may be expressed and duplicated so that it passes from one cell division to the next. For microorganisms, one of the most successful methods involves the use of plasmids as a vehicle (a *vector*) for transferring genes.

These small rings of DNA carry a limited set of genes and normally constitute only a tiny fraction of the cell's total DNA. Like the DNA of chromosomes, that of plasmids can be cut with restriction enzymes and additional DNA pasted into it. The result is a ring of *recombinant DNA*, which can be put into a bacterium. Specialised plasmids can be used to ferry genes from bacteria into yeast cells or even into plants.

recombinant dna



Recombinant DNA molecules, like the plasmid shown here, are made by cutting and pasting selected sequences of DNA with restriction enzymes and DNA ligase, respectively. Such recombinant molecules can then be introduced into microbes, plants or animals by a variety of methods.

Changing an organism's genetic make-up by adding 'naked' DNA such as plasmids is called *transformation*. Transformation is a relatively rare event in nature — most cells cannot take up DNA from their environment without first being made 'competent'. This can be done in several ways.

A popular method is to treat the cells with chemicals that produce temporary pores through the cell membrane, then to heat shock the bacteria in the presence of the DNA that is to be introduced. Because both the DNA and cell membrane are electrically uncharged in the circumstances used, the DNA can pass through the pores.

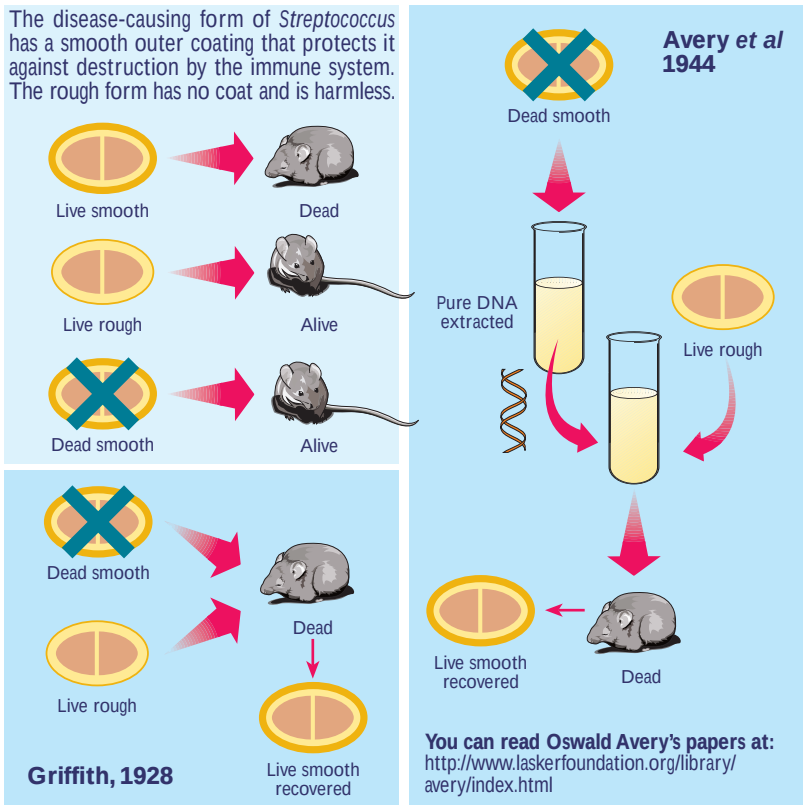
the search for the genetic material

Fred Griffith. Photo courtesy Society for General Microbiology

In 1928, a British medical officer at the Ministry of Health made a remarkable discovery. Fred Griffith found that one strain of *Streptococcus pneumoniae* (then known as *Pneumococcus*) could be converted into another by an unknown, non-living material. The nature of this 'transforming principle' remained a mystery for the next 15 years.

Oswald Avery and his colleagues in the USA carried out meticulous experiments for more than a decade to reveal the identity of the mystery material. Unfortunately, even though their findings (published in 1944) look conclusive to a modern observer, some scientists at the time argued against the Avery team's conclusion that DNA was the molecule of inheritance. Its molecular structure was thought to be too simple to carry the genetic message.

Not until 1952 — just a year before the publication of Watson and Crick's famous 'double helix' letter to *Nature* — were the majority of scientists convinced that DNA was the primary genetic material. It was then that Alfred Hershey and Martha Chase (then an undergraduate) proved, using radiolabelled DNA and protein, that the infective component of a bacteriophage was nucleic acid, and not the structurally more complex proteins.



Genetically modified plants

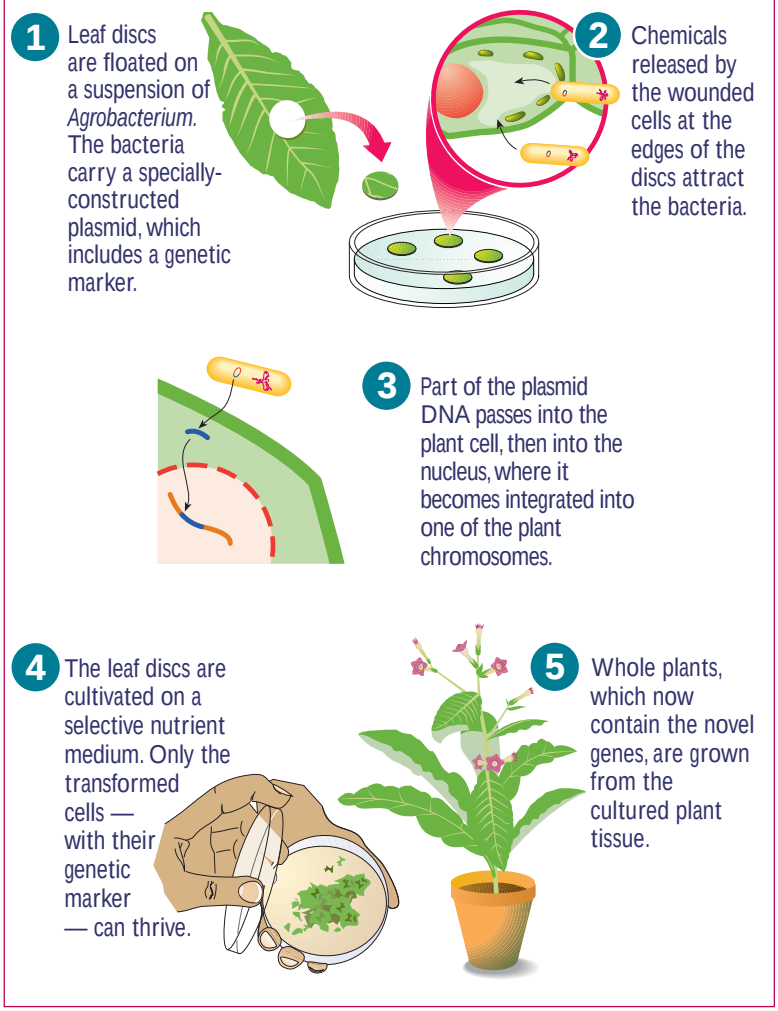
Nature's engineer

A genetic transfer system that is used for a wide variety of plants is the plant tumour-inducing plasmid (*Ti-plasmid*), found in the soil bacterium *Agrobacterium tumefaciens*. Through its plasmid, *Agrobacterium* has the ability to naturally engineer plant cells so that they grow into galls, the cells of which produce special compounds that the bacteria need to thrive.

Molecular biologists use 'disarmed' (non-tumour-inducing) versions of this plasmid to introduce genes of their choice into plants. As every cell carries a complete copy of all the plant's genes in its chromosomes, it is possible to re-grow an entire plant from a single modified cell. Special plasmids have now been constructed to help transfer fairly large genes into plants.

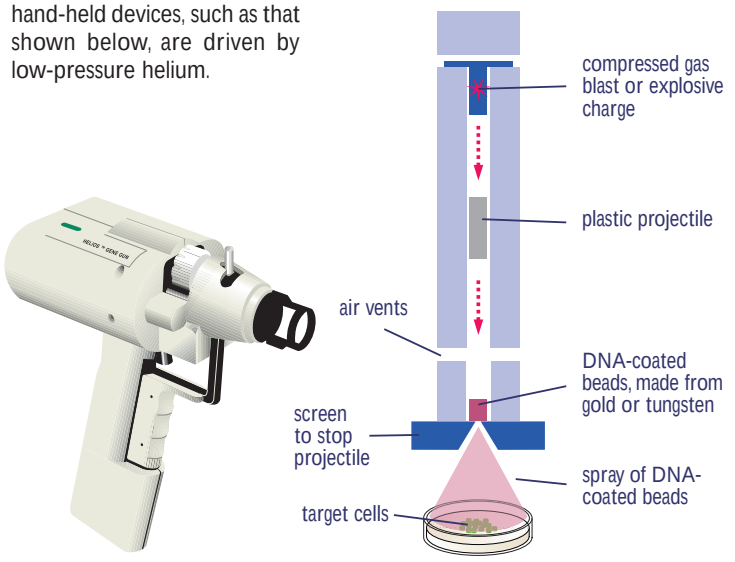
The first species to be modified in this way was tobacco, chosen because it and related plants (*e.g., tomatoes and potatoes*) are easy to propagate by tissue culture. *Agrobacterium* has been especially useful when working with trees which, because they are slow-growing and large, are difficult to improve by conventional plant breeding. Apricot, plum, apple and walnut trees have all been genetically modified with the help of *Agrobacterium*. Unfortunately, most mono-cotyledons (including several important cereal crops) are resistant to infection by *Agrobacterium*. Alternative methods have to be found for these sorts of plants.

transformation with agrobacterium



particle bombardment

Particle bombardment is used to genetically modify a wide variety of species. The first 'gene guns' were housed in heavy metal boxes and used explosive cartridges. Safer, hand-held devices, such as that shown below, are driven by low-pressure helium.



Blasted DNA

A procedure called *particle bombardment* or *ballistic impregnation* has achieved success with several important crops, including rice, wheat and soya.

With this method, the DNA to be introduced into the plant cells is first stuck onto minute tungsten or gold beads. Using a 'gene gun', the DNA-coated particles are fired into soft plant tissue *e.g., callus cells from tissue culture*. This introduces novel DNA into the plant cells, some of which, by chance, integrates into the plant chromosomes.

The cells are grown into miniature plantlets by cultivating them on a special medium that contains appropriate plant growth substances and a selective agent (see *Marker genes*, page 10) that permits only the modified cells to grow. These plantlets are eventually grown up by conventional means in soil.

Screening is necessary to select those plants in which the introduced genes are expressed. Careful checks are also made to ensure that the expression of other genes in the host plant is not affected by the novel introduction.

Efficient electroporation

DNA can also be introduced into the thin-walled tubes that develop from pollen grains by subjecting them to micro- or millisecond pulses of a strong electric field. This technique, known as *electroporation*, causes pores to appear momentarily in the pollen tubes, through which DNA from a surrounding solution can enter. Seeds that develop from ovules fertilised with such pollen carry the introduced genes.

Electroporation also works with plant cells from which the cell wall has been removed by enzyme treatment. From these naked plant cells (called *protoplasts*), whole plants can be regenerated by tissue culture. Entire plant cells can also be treated in this way, albeit with less success than with protoplasts. Electroporation is also used routinely in research to transfer recombinant DNA molecules into a range of microorganisms. It is the most efficient genetic transformation method so far for bacteria.

Genetically modified animals

The DNA of animals can also be modified, but it is necessary to introduce genes at an early stage of development if they are to be present in all the cells of a mature animal and be passed on to its offspring. DNA can be injected into newly-fertilised egg cells through a very fine glass pipette. Only a small proportion of the injected eggs take up the new genes; those that do are transferred into the uterus of a suitable foster mother.

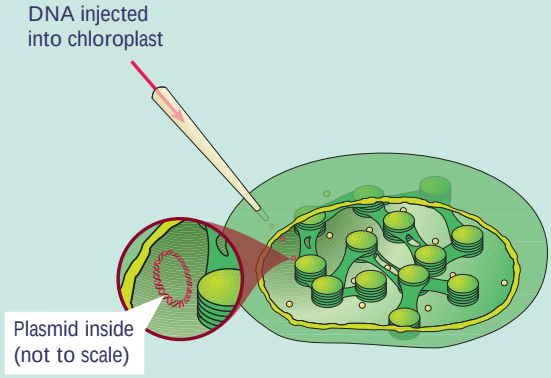
This crude method of genetic modification works for mice, cows, pigs, sheep and goats. Micro-injection can also be used to introduce new genes into fish eggs, but, for obvious reasons, it is not suitable for birds' eggs. However, specially-modified viruses have been used to introduce, for example, disease-resistance into chickens. The viruses are inserted through the shell of the egg. Similar viral vectors can also introduce genes into mammalian cells.

Chloroplast transformation

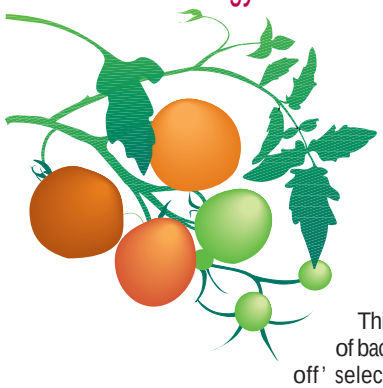
One of the major public concerns about genetically-modified organisms is that they might transfer their introduced genes into other, non-modified organisms with unpredictable and possibly undesirable consequences. Scientists are aware of this problem and have proposed several methods of preventing it. A promising development in plants is the genetic modification of chloroplasts.

Chloroplasts carry their own DNA, which is passed on (in most species) *via* the ovules alone; chloroplast DNA does not end up in pollen grains. If the DNA of chloroplasts could be modified, therefore, it might be possible to prevent the unwanted 'spread' of novel genes *via* pollen.

For the first time in 1999, hollow glass needles with tips a mere 0.1 μm in diameter were used, under microscope guidance, to inject DNA directly into chloroplasts. Such fine tips do not damage the chloroplast membrane, but very high pressures are required to push the DNA from the needle. Controlling the flow of DNA under high pressure is a significant practical problem. It was overcome by sealing a liquid metal alloy into the needle, behind the DNA. When the alloy was heated, it expanded, forcing the DNA out in a controllable manner. Researchers have also used same technique to inject DNA directly into individual bacteria and the nuclei of eukaryotic cells. It remains to be seen how useful this experimental method will become.

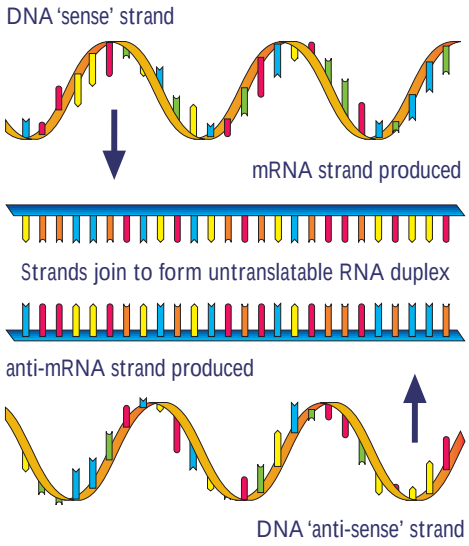


antisense technology

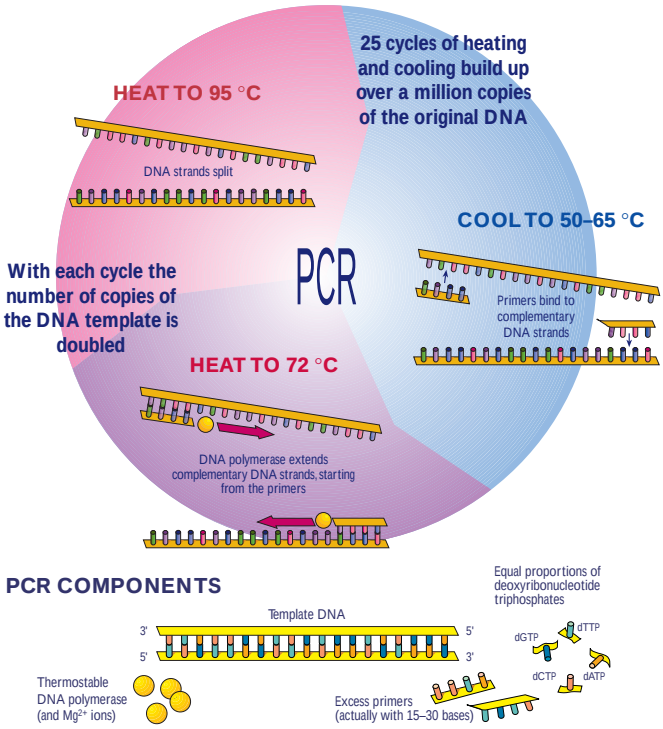


The sequence of bases along only one strand of the DNA double helix directs the production of proteins. This strand is called the 'sense' strand. Antisense RNA is an RNA sequence transcribed from the 'wrong' (antisense) strand of the DNA. When antisense and sense RNA molecules meet, they combine, preventing the production of proteins encoded by their sequences of bases.

This process, which occurs naturally in the control of bacterial genes, can be used by biologists to 'switch off' selected genes. This is done by inserting an appropriate piece of inverted DNA into the sense strand of the cell's DNA. Antisense RNA molecules are then formed from the inverted section when the DNA is transcribed. A diverse range of experimental crops has been produced using this technology, including for a short time in the USA, slow-ripening tomatoes (in which a pectinase gene was 'silenced') although these were never sold in Europe.



the polymerase chain reaction



The components

- Two primers are made that flank the specific stretch of DNA to be copied. These are generally 15–30 bases long, and are complementary to the 5' end of each DNA strand.
- Total (double-stranded) DNA is extracted from the sample under study.
- The DNA and excess primers are mixed.
- Equal proportions of very pure deoxyribonucleoside triphosphates (dCTP, dATP, dGTP, dTTP) are added.
- Heat-stable *Taq* DNA polymerase (originally obtained from the hot-water bacterium *Thermus aquaticus*, but now produced by genetically-modified organisms) is added to the reaction mix, together with Mg^{2+} ions that are required as a co-factor by the enzyme.

The process

- The double-stranded DNA is split into two single-stranded templates by heating it to 95–98 °C.
- The mixture is cooled to 50–65 °C.The primers bind (hybridise) to the complementary strands of DNA.
- Heating to 72 °C (the optimum temperature for *Taq* DNA polymerase) encourages the synthesis of new DNA strands alongside the templates.
- The mixture is heated again to separate the newly-formed DNA strands from their partners.

The cycle of cooling and heating is repeated. With each cycle (lasting about 2 minutes) the number of copies of the template DNA is doubled. After 25 cycles more than a million copies of the DNA will have been made.

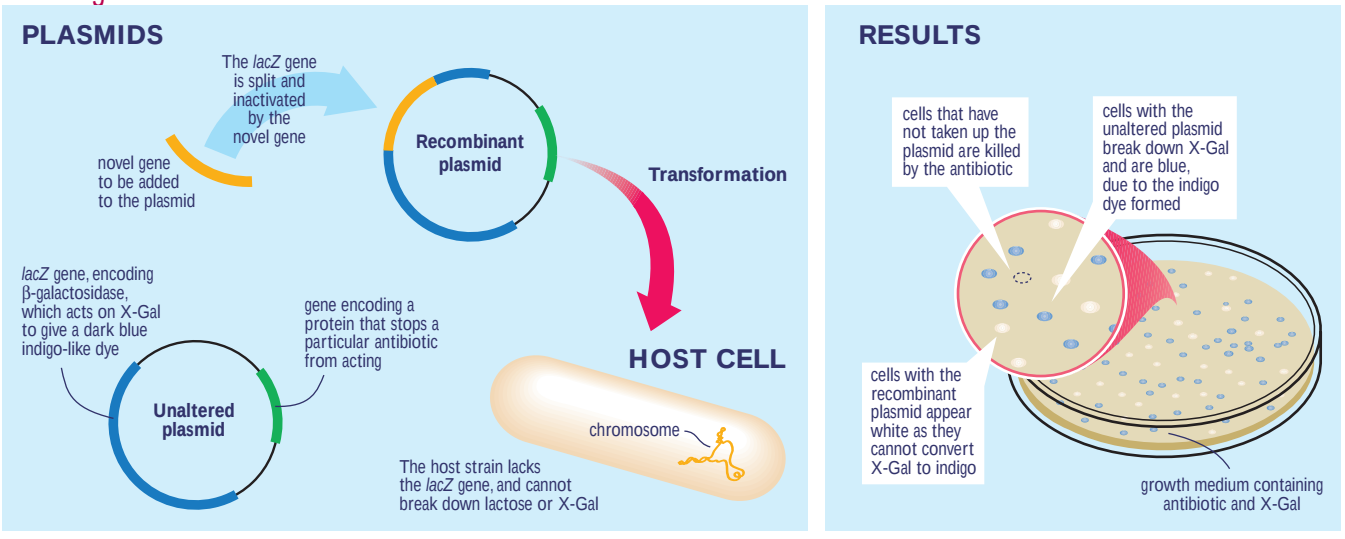
Marker genes

Whatever method is used for genetic modification, only a small proportion of treated cells will take up DNA in a stable and heritable manner. Screening is therefore necessary to discover which cells have done so. Some genes confer upon their hosts the ability to overcome the harmful effects of specific heavy metals, herbicides or antibiotics. Such genes can be used as '*markers*' to identify those cells that have taken up DNA. This is done by coupling the marker gene to the other genes that are to be introduced. When the transformed cells are cultivated on a growth medium which contains the toxic substance (*e.g., antibiotic or herbicide*), only those that have taken up DNA (and can, as a consequence, degrade or otherwise avoid the effects of the toxin) will thrive.

PCR — the genetic photocopier

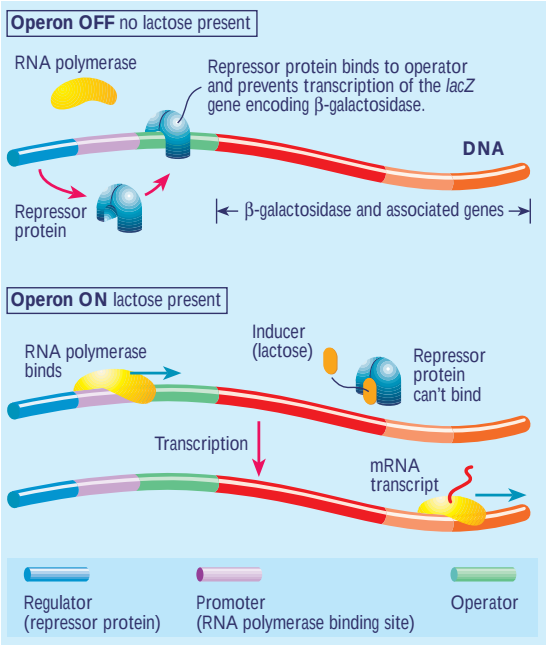
Other methods for identifying transferred genes include the PCR (*polymerase chain reaction*, see box). The PCR has been likened to a 'genetic photocopier'. From a very small amount of biological material, millions of copies of a specific section of its DNA can be made quickly. The PCR lies behind many of the spectacular successes of forensic genetic fingerprinting, where criminals have been identified from the DNA in just a few drops of blood, a couple of cells on a cigarette butt or even on the back of a licked postage stamp. This extremely sensitive method is also used, for example, to test food products for the presence of ingredients that have been derived from genetically-modified sources.

marker genes: blue-white selection



Switched-on genes

The ways in which the expression of genes is regulated have been studied for many years. One controversial method of controlling the expression and 'spread' of novel genes is to regulate them by an external stimulus. For example, it is possible to produce plants that require the application of a specific chemical to grow, or for their introduced genes to be expressed. If the chemical is no longer applied, the plant dies or its novel genes are 'switched off'. Such mechanisms might be used to ensure that, say, insecticides were only produced by modified plants when required, or that modified plants could not stray beyond the limits of sprayed fields. However, fears that similar techniques might be used to increase dependency upon sprays have forced the major agrochemical producers to agree that they will not develop such technologies.



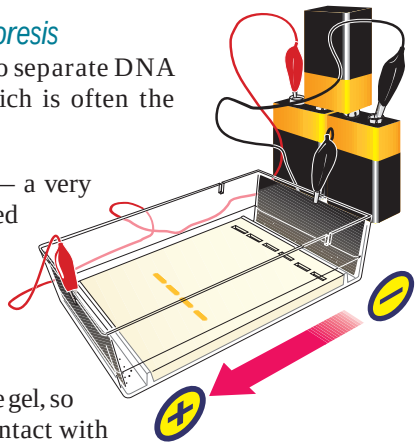
▲ *The lac operon, which was proposed in the late 1950s, explains how E. coli produces the enzyme β-galactosidase only in the presence of lactose or an inducer such as IPTG.*

Genetic analysis

Sorting by size — gel electrophoresis

Gel electrophoresis can be used to separate DNA fragments of different sizes, which is often the first step in their analysis.

First, a gel is cast from *agarose* — a very pure form of agar, which is obtained from seaweed. At one end of the slab of gel are several small wells, made by the teeth of a comb that was placed in the gel before it set.



▲ *Gel electrophoresis is used to separate DNA fragments of different sizes. This is often the first step in isolating and characterising genes, or mapping their relative positions on chromosomes.*

A buffer solution is poured over the gel, so that it fills the wells and makes contact with electrodes at each end of the gel. Ions in the buffer solution conduct electricity. The buffer also stops the gel from drying out. The invisible DNA fragments are mixed with a small volume of *loading dye*. This dye is dissolved in a dense sugar solution, so that when it is added to the wells, it sinks to the bottom, taking the DNA with it.

A current is applied to the electrodes, setting up an electrical field across the gel. Phosphate groups give the DNA fragments a negative electrical charge, so that the DNA migrates through the gel towards the positive electrode. Small fragments move quickly through the porous gel — larger fragments travel more slowly. In this way the pieces of DNA are separated by size. The loading dye also moves through the gel, so that the progress of the electrophoresis can be seen.

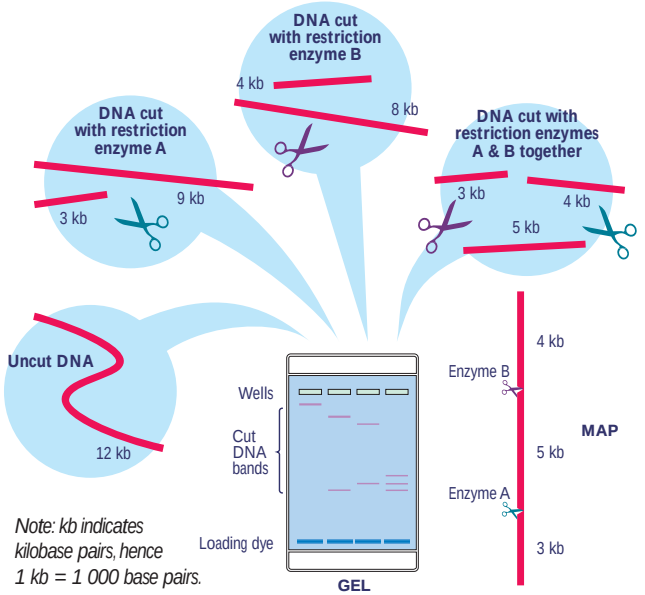
After electrophoresis, the gel is stained to reveal the DNA, either as a smear (many fragments of a wide range of sizes) or bands (each band is comprised of numerous DNA fragments of a similar size). Within a smear, specific bands can by highlighted using *probes* which bind to particular sequences of DNA (or RNA).

restriction site mapping

Construction of a restriction map

This diagram illustrates the principle of restriction site mapping. A 12 kb length of DNA can be cut with two restriction enzymes, A and B. Where do the restriction sites lie, relative to one another? Fragments of DNA produced by cutting with the individual enzymes and both enzymes together are run on a gel.

- *Enzyme A* cuts at one end of the DNA molecule while *Enzyme B* cuts towards the other end.
- The fragments could not be arranged in the sequence 3→4→5 kb because *Enzyme B* produces two pieces of 4 kb and 8 kb, and it is not possible to produce an 8 kb fragment unless the 3 kb piece is next to the 5 kb piece.
- For this reason, the fragments must be arranged in the sequence 3→5→4 kb.



Note: kb indicates kilobase pairs, hence 1 kb = 1 000 base pairs.

linkage maps

In 1913, the very first genetic map was published. It had been devised by Alfred Sturtevant, an undergraduate student at Colombia University (USA).

Alfred realised that the frequency with which characteristics tended to be inherited together might indicate the relative positions of their genes on a chromosome.

Working overnight with data from numerous fruit fly crosses, Sturtevant devised a 'linkage map' showing the relative positions of six fruit fly genes on a chromosome.

You can read Alfred Sturtevant's original paper at:
Ⓦ www.esp.org

1869
J. Friedrich Miescher discovers DNA in the nuclei of cells

1928
Fred Griffith observes bacterial transformation

1938
William Astbury takes X-ray photos of DNA

1944
Oswald Avery and his colleagues show that DNA is the genetic material

1950
Erwin Chargaff finds that A : T and G : C bases in DNA occur in ratios of 1:1

Maurice Wilkins and Ray Gosling obtain X-ray photographs of crystalline DNA

1951
Rosalind Franklin takes X-ray photos of hydrated DNA, revealing a helical structure

Esther Lederberg discovers phage λ

1952
Alfred Hershey and Martha Chase confirm that DNA is the genetic material

1953
James Watson and Francis Crick propose a double helical structure for DNA

1958
Matthew Meselson and Frank Stahl show that DNA replication is semi-conservative

1960
Sydney Brenner and François Jacob discover and explain the rôle of mRNA

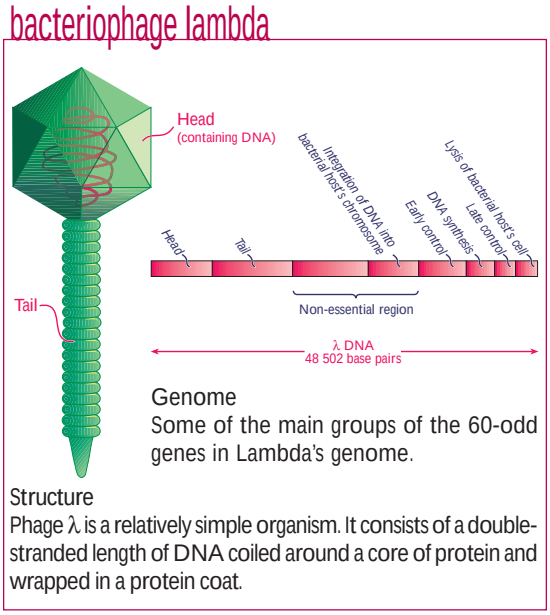
Early 1960s
Plasmids carrying antibiotic resistance are discovered by Japanese researchers

1966
Marshall Nirenberg, H. Gobind Khorana and their co-workers 'crack' the genetic code

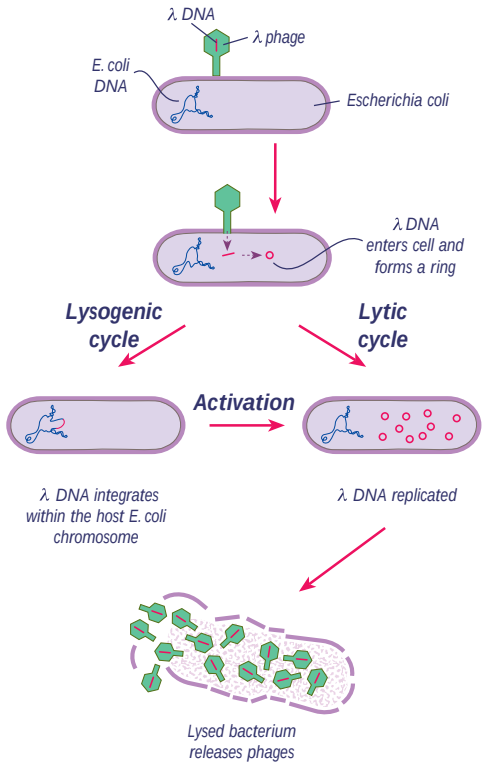
Bacteriophage lambda

Bacteriophages (literally: 'bacteria-eaters') are viruses that invade bacteria. To reproduce, bacteriophages (or phages) must take over the molecular machinery of their bacterial hosts. The phage lambda (λ) preys upon *Escherichia coli*. Lambda enjoys a choice of life styles — it can either multiply within its host and destroy it (the lytic cycle), or the λ DNA can be inserted into the bacterial chromosome and remain dormant there for several generations (the lysogenic cycle). An environmental trigger *e.g., ultraviolet light* activates the lytic cycle.

Bacteriophage λ was the first large virus to have its entire genetic make-up (*genome*) mapped and sequenced. It is 48 502 base pairs long. Within this genome are genes that code for the virus's protein coat, bursting (lysis) of the bacterial cell, integration of λ DNA into the host's chromosome and so on. The order in which these genes are activated is important. For example, it would be of little benefit to the virus if the host bacterial cell was broken open (lysed) before new virus particles had been assembled. Consequently, λ has evolved an elaborate system of gene regulation that has been studied in great detail.



Relatively little of the λ genome is required to package DNA and deliver it into bacterial cells. About 20 000 base pairs can be deleted from its central region and replaced with a similarly-sized fragment of DNA from another organism, without affecting the phage's viability. Several specially-constructed forms of λ, with differing amounts of DNA removed and replaced, are used by molecular biologists to ferry new genes into bacteria. Such vectors are far more efficient at introducing DNA into bacteria than plasmids are. This is hardly surprising because, after all, bacteriophages have evolved to do precisely this job.



▲ *Infection of E. coli by bacteriophage lambda (λ) showing the lytic and lysogenic cycles.*

Mapping and sequencing genomes

The ultimate aim of the *Human Genome Project*, a massive international effort, is to map the positions of the 30–100 000 genes (the exact figure is unknown) and determine the sequence of bases in the entire human genome. This should lead to the ability to diagnose human disease, and in the longer term to devise effective treatments and possibly cures for a wide range of conditions.

Much of this work is now highly automated, with a third of the human genome being sequenced in the UK at the Sanger Centre, near Cambridge. A draft of the genome is to be published in 2000 and the project should be completed by 2002–3.

▲ *Automated sequencing machines are able to determine 450000 bases per day. The interpretation of the vast amount of sequence data produced means that bioinformatics is of increasing importance.*

estimated genome sizes

Organism	Megabases	Made of
Viruses		
• HIV	0.0092–0.0103	ssRNA
• Adenovirus	0.0359	dsDNA
• Herpes Simplex	0.150	dsDNA
• Bacteriophage λ	0.48502	dsDNA
Bacteria		
• *Escherichia coli*	4.7	dsDNA
• *Bacillus subtilis*	4.2	dsDNA
Fungi		
• Baker's yeast	13.5	dsDNA
Nematodes	75–620	
• *Caenorhabditis elegans*	80	dsDNA
Insects	47–12 000	
• Fruit fly	120	dsDNA
• House fly	840	dsDNA
Birds	1 100–1 900	
• Chicken	1 125	dsDNA
Mammals	2 800–5 200	
• Human	3 300	dsDNA
• House mouse	3 300	dsDNA
Flowering plants	95–120 000	
• *Arabidopsis thaliana*	70	dsDNA
• Tomato	700	dsDNA
• Tobacco	3 500	dsDNA
• Maize	15 000	dsDNA

Except for HIV, these are all figures for haploid genomes

How long is a genome?
There is, roughly, one DNA base per Angstrom of the double helix. An Angstrom is 10^{-10} metres, thus it is easy to estimate the physical size of an organism's DNA. For example:

Organism	Bases	Length
E. coli	4.7 Mb	0.47 mm
Baker's yeast	13.5 Mb	1.35 mm
Human	3 300 Mb	330 mm

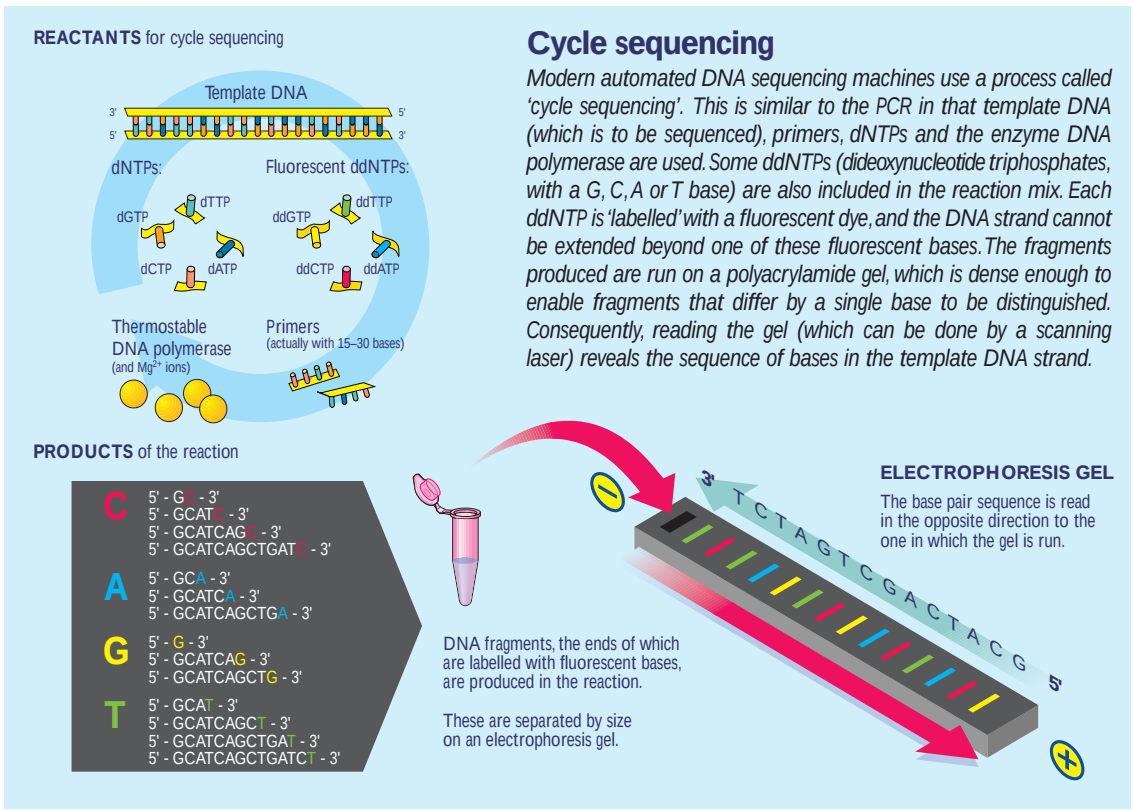
Other genomes are also being sequenced, particularly those of economically important plants and animals, organisms that are used in research (such as the fruit fly and the nematode worm *Caenorhabditis elegans*), and many pathogens.

Making use of genetic data

While crop biotechnology or medical applications frequently make the headlines, molecular genetics has led to important discoveries in almost all the biological sciences. For example, in the last few years, the evolutionary relationships of the flowering plants have been completely re-thought based on genetic data. An early application of genetic fingerprinting revealed the hitherto unsuspected mating habits of birds, which proved of great interest to zoologists and conservationists. Archæologists and forensic scientists now make routine use of DNA evidence. Perhaps the most far-reaching insights will emerge from the human genome project, which has already cast light on our origins, our nature and the fundamental unity of all life on Earth.

It is notable that while the early 20th century geneticists (who had almost no knowledge of the genotype) frequently sought to explain the differences between people, often to reinforce their own prejudices, those of the late 20th century have shown how similar we all are. The genes of humans differ by less than 0.1%.

If this new knowledge is to be applied wisely, it is essential that more people understand the basic science and technology of modern genetics and think carefully about the wider implications of such work.



Cycle sequencing

Modern automated DNA sequencing machines use a process called 'cycle sequencing'. This is similar to the PCR in that template DNA (which is to be sequenced), primers, dNTPs and the enzyme DNA polymerase are used. Some ddNTPs (dideoxynucleotide triphosphates, with a G, C, A or T base) are also included in the reaction mix. Each ddNTP is 'labelled' with a fluorescent dye, and the DNA strand cannot be extended beyond one of these fluorescent bases. The fragments produced are run on a polyacrylamide gel, which is dense enough to enable fragments that differ by a single base to be distinguished. Consequently, reading the gel (which can be done by a scanning laser) reveals the sequence of bases in the template DNA strand.

1967
Weiss and Richardson isolate DNA ligase

1970
Hamilton Smith and colleagues isolate and characterise the first sequence-specific restriction enzyme (*Hin*dIII)

Morton Mandel and A. Higa demonstrate artificial transformation

1972
Janet Mertz and Ron Davis cut and paste DNA molecules

1973
Stanley Cohen and Herbert Boyer put recombinant DNA into bacteria

1977
Fred Sanger sequences ΦX174 virus DNA

1982
The first genetically-modified plants are produced (tobacco)

1983
Kary Mullis conceives of the PCR

Fred Sanger and his colleagues publish the DNA sequence of bacteriophage λ

1986
DNA sequencing is automated

1987
Start of the Human Genome Project

1995
Haemophilus influenzae's genome is sequenced

1996
Saccharomyces cerevisiae's genome is sequenced

1998
Caenorhabditis elegans is sequenced

1999
Human chromosome 22 is sequenced

2000
The fruit fly, *Drosophila melanogaster*, is sequenced

A draft sequence of the entire human genome is published