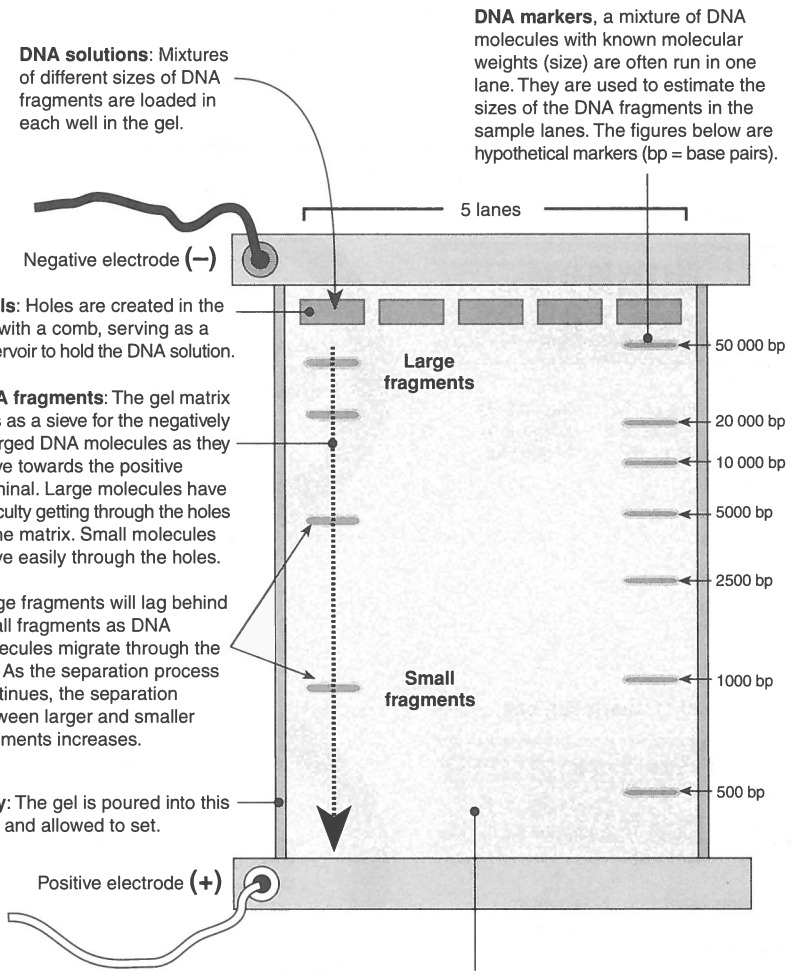
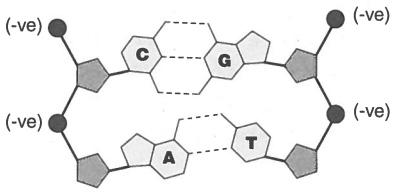
Gel Electrophoresis

Gel electrophoresis is a method that separates large molecules (including nucleic acids or proteins) on the basis of size, electric charge, and other physical properties. Such molecules possess a slight electric charge (see DNA below). To prepare DNA for gel electrophoresis the DNA is often cut up into smaller pieces. This is done by mixing DNA with restriction enzymes in controlled conditions for about an hour. Called **restriction digestion**, it produces a range of DNA fragments of different lengths. During electrophoresis, molecules are forced to move through the pores of a **gel** (a jelly-like material), when the electrical current is applied. Active electrodes at each end of the gel provide the driving force. The electrical current from one electrode repels the molecules while the other electrode simultaneously attracts the molecules. The frictional force of the gel resists the flow of the molecules, separating them by size. Their rate of migration through the gel depends on the strength of the electric field, size and shape of the molecules, and on the ionic strength and temperature of the buffer in which the molecules are moving. After staining, the separated molecules in each lane can be seen as a series of bands spread from one end of the gel to the other.

Analysing DNA using Gel Electrophoresis





DNA is negatively charged because the phosphates (black) that form part of the backbone of a DNA molecule have a negative charge.

Steps in the process of gel electrophoresis of DNA

1. A tray is prepared to hold the gel matrix.
2. A gel comb is used to create holes in the gel. The gel comb is placed in the tray.
3. Agarose gel powder is mixed with a buffer solution (the liquid used to carry the DNA in a stable form). The solution is heated until dissolved and poured into the tray and allowed to cool.
4. The gel tray is placed in an electrophoresis chamber and the chamber is filled with buffer, covering the gel. This allows the electric current from electrodes at either end of the gel to flow through the gel.
5. DNA samples are mixed with a "loading dye" to make the DNA sample visible. The dye also contains glycerol or sucrose to make the DNA sample heavy so that it will sink to the bottom of the well.
6. A safety cover is placed over the gel, electrodes are attached to a power supply and turned on.
7. When the dye marker has moved through the gel, the current is turned off and the gel is removed from the tray.
8. DNA molecules are made visible by staining the gel with **methylene blue** or ethidium bromide (which binds to DNA and fluoresces in UV light).

1. Explain the purpose of gel electrophoresis: _____

_____

_____

2. Describe the two forces that control the speed at which fragments pass through the gel:

   (a) _____

   (b) _____

3. Explain why the smallest fragments travel through the gel the fastest: _____

_____

_____

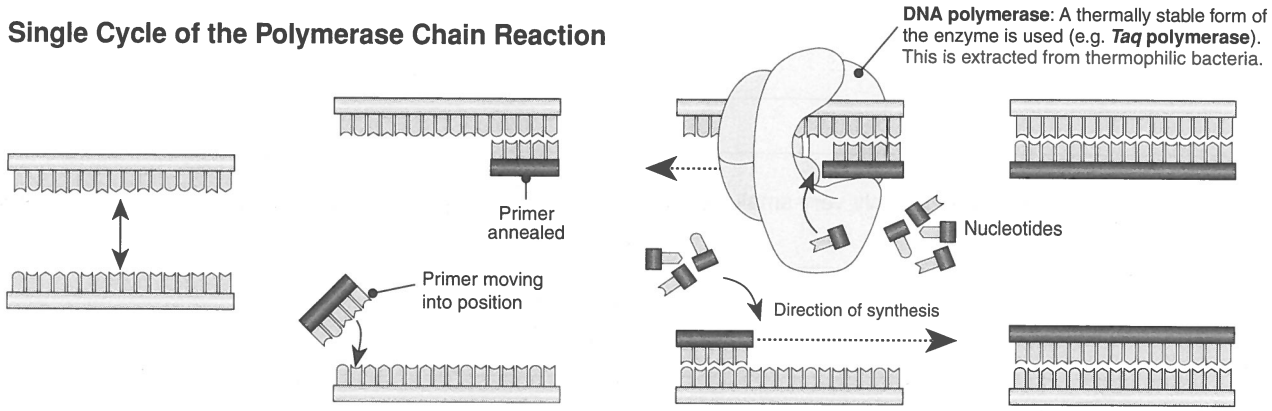
Related activities: Automated DNA Sequencing

Web links: DNA Extraction, Gel Electrophoresis

Polymerase Chain Reaction

Many procedures in DNA technology (such as DNA sequencing and DNA profiling) require substantial amounts of DNA to work with. Some samples, such as those from a crime scene or fragments of DNA from a long extinct organism, may be difficult to get in any quantity. The diagram below describes the laboratory process called **polymerase chain reaction** (**PCR**). Using this technique, vast quantities of DNA identical to trace samples can be created. This process is often termed **DNA amplification**. Although only one cycle of replication is shown below, following cycles replicate DNA at an exponential rate. PCR can be used to make billions of copies in only a few hours.

A Single Cycle of the Polymerase Chain Reaction



1. A DNA sample (called target DNA) is obtained. It is **denatured** (DNA strands are separated) by heating at 98°C for 5 minutes.
2. The sample is cooled to 60°C. Primers are **annealed** (bonded) to each DNA strand. In PCR, the primers are short strands of DNA; they provide the starting sequence for DNA extension.
3. Free nucleotides and the enzyme DNA polymerase are added. DNA polymerase binds to the primers and, using the free nucleotides, synthesises complementary strands of DNA.
4. After one cycle, there are now two copies of the original DNA.

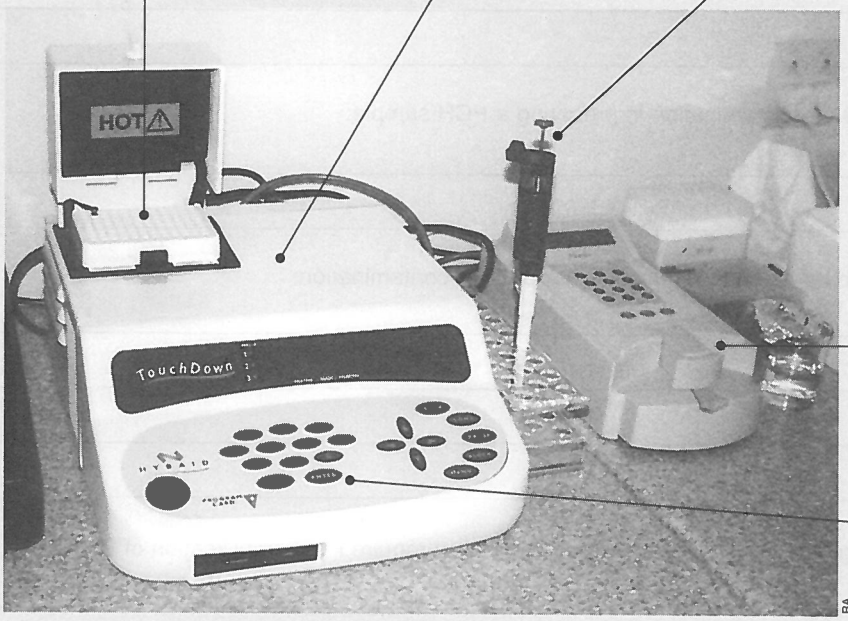
Repeat for about 25 cycles

Repeat cycle of heating and cooling until enough copies of the target DNA have been produced

Loading tray
Prepared samples in tiny PCR tubes are placed in the loading tray and the lid is closed.

Temperature control
Inside the machine are heating and refrigeration mechanisms to rapidly change the temperature.

Dispensing pipette
Pipettes with disposable tips are used to dispense DNA samples into the PCR tubes.



Thermal Cycler

Amplification of DNA can be carried out with simple-to-use machines called **thermal cyclers**. Once a DNA sample has been prepared, in just a few hours the amount of DNA can be increased billions of times. Thermal cyclers are in common use in the biology departments of universities, as well as other kinds of research and analytical laboratories. The one pictured on the left is typical of this modern piece of equipment.

DNA quantitation
The amount of DNA in a sample can be determined by placing a known volume in this quantitation machine. For many genetic engineering processes, a minimum amount of DNA is required.

Controls
The control panel allows a number of different PCR programmes to be stored in the machine's memory. Carrying out a PCR run usually just involves starting one of the stored programmes.

1. Explain the purpose of PCR: _____

_____

_____

_____

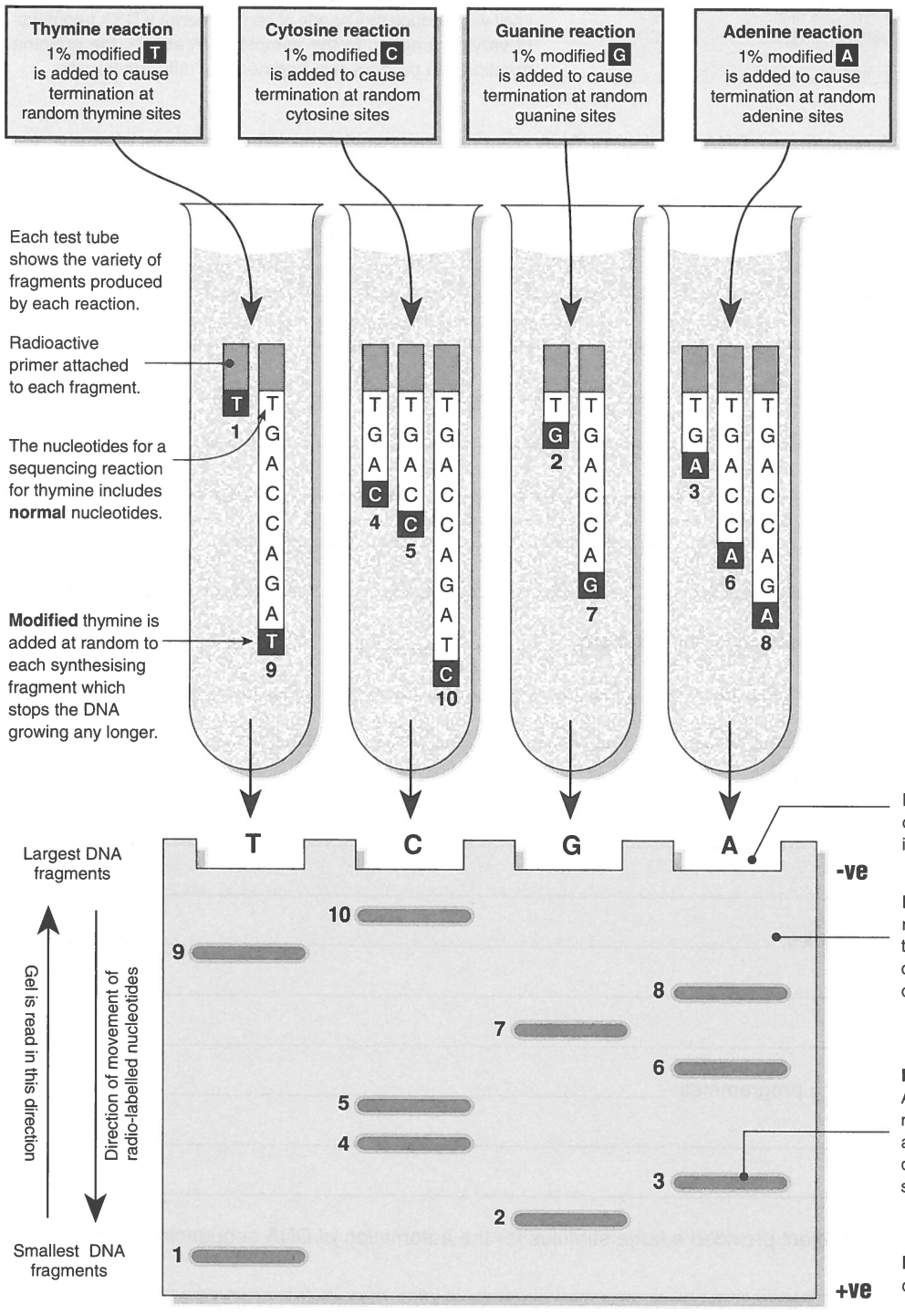
Manual DNA Sequencing

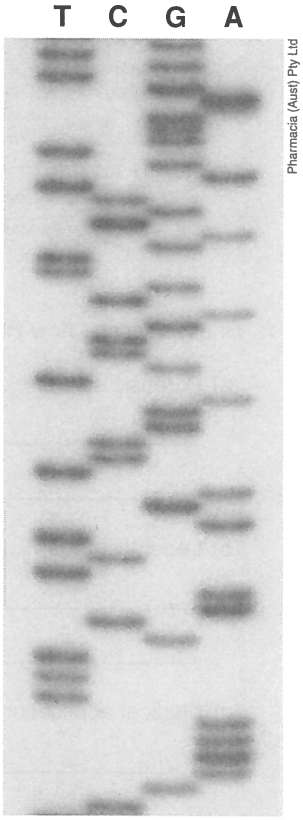
DNA sequencing techniques are used to determine the nucleotide (base) sequence of DNA. Two manual methods are in current use: the **Maxim-Gilbert** procedure and the most common method, the **Sanger** procedure (illustrated below). Both methods use a procedure called **electrophoresis**. The Sanger method is based on the premature termination of DNA synthesis resulting from the inclusion of specially modified nucleotides. DNA synthesis is initiated from a **primer** which is **radio-labelled** (contains a radioactive isotope that will appear on a photographic film called an **autoradiograph**). Four separate reactions are run, each containing a modified nucleotide mixed with its normal counterpart, as well as the three other normal nucleotides. When a modified nucleotide is added to the growing complementary DNA, synthesis stops. Each reaction yields a series of different sized fragments extending from the radioactive primer. The fragments from the four reactions are separated by electrophoresis and analysed by autoradiography to determine the DNA sequence.

The Sanger Method for DNA Sequencing

Four sequencing reactions

Using the same DNA sample to be sequenced (example used: **A C T G G T C T A G**), a separate sequencing reaction is carried out for each of the 4 bases: T, C, G, and A. In addition to the DNA sample, each reaction has normal (unaltered) copies of nucleotides: **T**, **C**, **G**, and **A**, plus a small quantity of one of the modified nucleotides:





A typical **autoradiograph** showing a DNA sequence. The unexposed film is laid in contact with the gel after it has run. Radioactivity from the clustered DNA fragments create the dark shadows (blobs). Each blob contains millions of fragments.

DNA samples: The four reactions containing DNA fragments are placed in separate wells at the top of the gel.

Electrophoresis gel: A jelly-like material that allows DNA fragments to move through it when an electric charge is applied. It is usually made of a material called *acrylamide*.

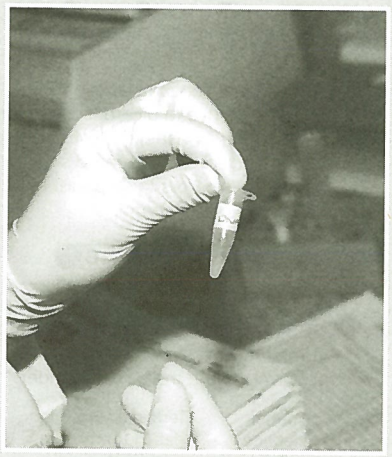
Radio-labelled DNA fragments: Attracted to the positive terminal, millions of DNA fragments of similar size and sequence move as a dark shadow down the gel. Larger pieces move more slowly and therefore do not travel as far.

Positive terminal: Attracts the fragments of DNA that are negatively charged.

Related activities: *Gel Electrophoresis*

Automated DNA Sequencing

The process of DNA sequencing can be automated using **gel electrophoresis** machines that can sequence up to 600 bases at a time. Automation improves the speed at which samples can be sequenced and has made large scale sequencing projects possible. Instead of using radio-labelled fragments, automated sequencing uses nucleotides labelled with **fluorescent dyes**. Another advantage is that the entire base sequence for a sample can be determined from a single lane on the gel (not four lanes as with manual sequencing). Computer software automatically interprets the data from the gel and produces a base sequence. Also see the weblink pdf download, which covers manual sequencing and provides questions relevant to both activities.

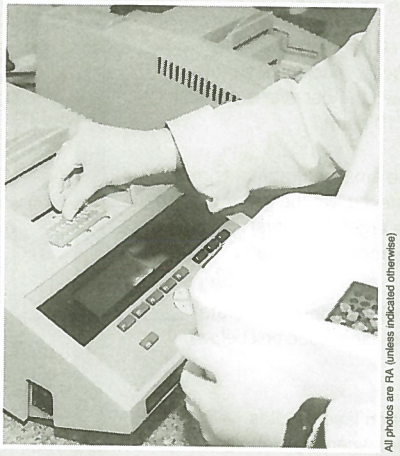


1. DNA sample arrives

Purified DNA samples may contain linear DNA or plasmids. The sample should contain about 1 x 10^{11} DNA molecules. The sample is checked to ensure that there is enough DNA present in the sample to work with.

2. Primer and reaction mix added

A **DNA primer** is added to the sample which provides a starting sequence for synthesis. Also added is the **sequencing reaction mix** containing the *polymerase enzyme* and free nucleotides, some which are labelled with dye.



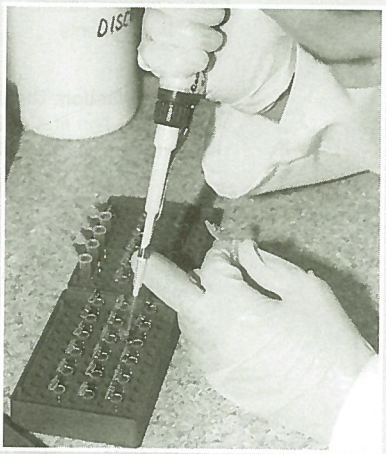
All photos are RA (unless indicated otherwise)

3. Create dye-labelled fragments

A **PCR** machine creates fragments of DNA complementary to the original template DNA. Each fragment is tagged with a fluorescent dye-labelled nucleotide. Running for 25 cycles, it creates 25 x10^{11} single-stranded DNA molecules.

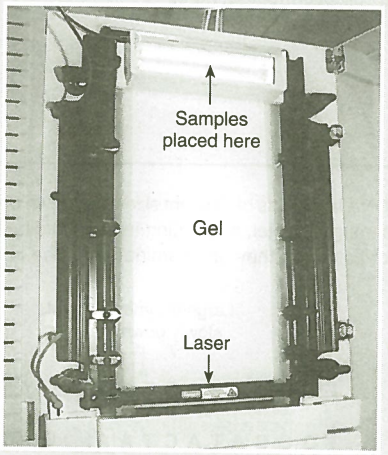
4. Centrifuge to create DNA pellet

The sample is chemically precipitated and centrifuged to settle the DNA fragments as a solid pellet at the bottom of the tube. Unused nucleotides, still in the liquid, are discarded.



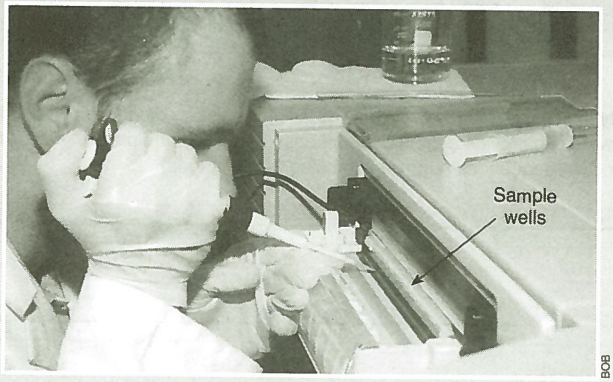
5. DNA pellet washed, buffer added

The pellet is washed with ethanol, dried, and a gel loading buffer is added. All that remains now is single stranded DNA with one dye-labelled nucleotide at the end of each molecule.



6. Acrylamide gel is loaded

The DNA sequencer is prepared by placing the gel (sandwiched between two sheets of glass) into position. A 36 channel 'comb' for receiving the samples is placed at the top of the gel.



7. Loading DNA samples onto gel

Different samples can be placed in each of the 36 wells (funnel shaped receptacles) above the gel. A control DNA sample of known sequence is applied to the first lane of the sequencer. If there are problems with the control sequence then results for all other lanes are considered invalid.

8. Running the DNA sequencer

Powerful computer software controls the activity of the DNA sequencer. The gel is left to run for up to 10 hours. During this time an argon laser is constantly scanning across the bottom of the gel to detect the passing of dye-labelled nucleotides attached to DNA fragments.

Related activities: *Polymerase Chain Reaction, The Human Genome Project, Genome Projects*

Web links: *Manual DNA Sequencing*

How a DNA Sequencer Operates

The gel is loaded following preparation of the samples and the gel (see steps 1-7 and box, right).

Comb with 36 lanes into which different samples can be placed.

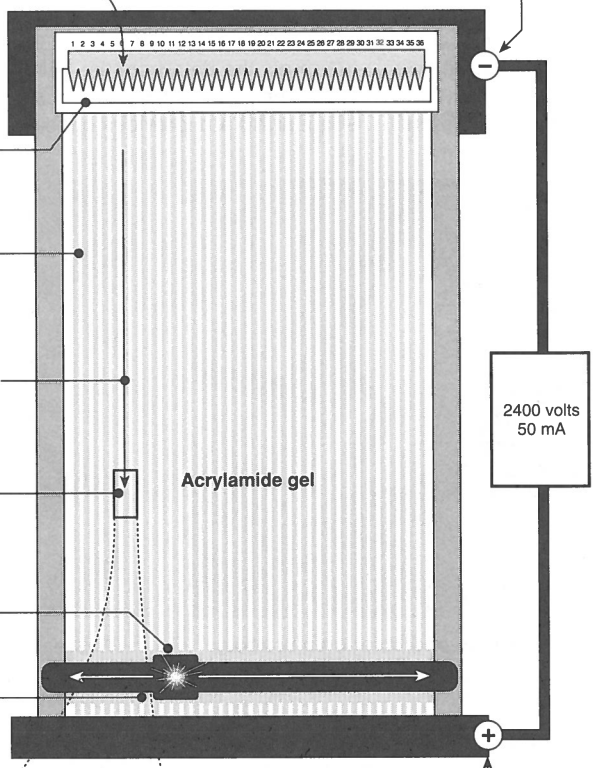
DNA fragments with dye-labelled nucleotides move down the gel over a period of 10 hours.

The smallest fragments move fastest down the gel and reach the argon laser first. Larger fragments arrive later.

DNA fragments separate into bands (see box below).

Argon laser excites fluorescent dye labels on nucleotides.

Lenses collect the emitted light and focus it into a spectrograph. An attached digital camera detects the light. See 'data collection' (below, right).



Creating the dye labelled fragments for gel electrophoresis is outlined in step 3. Key ingredients are:

(a) Original DNA template (the sample)

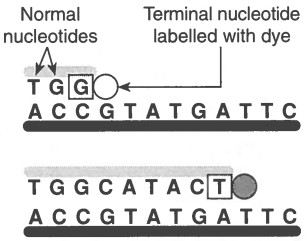
A C C G T A T G A T T C

(b) Many normal unlabelled nucleotides:

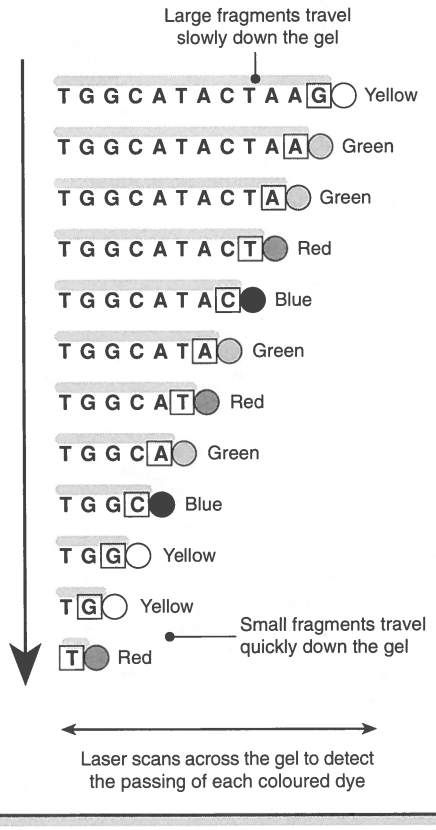
A T G C

(c) Terminal nucleotides labelled with fluorescent dye (a different colour for each of the 4 bases). The structure of the nucleotides is altered so they act as terminators to stop further synthesis of the strand:

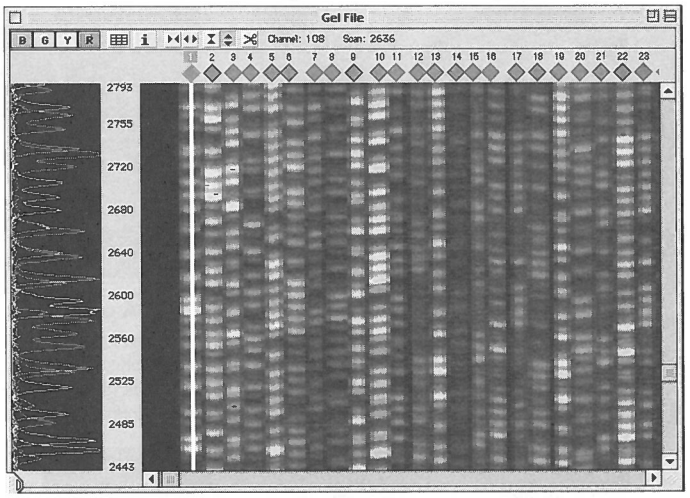
Two examples of synthesised DNA fragments are shown below. One is relatively short, the other is longer:



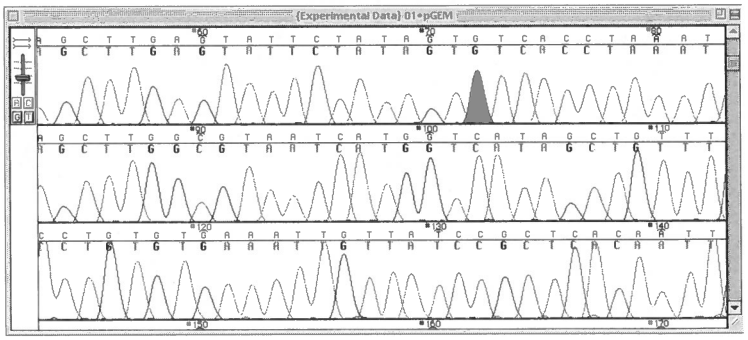
DNA fragments of different sizes are drawn down through the gel, separating into distinct bands of colour as they are illuminated by the laser:



Data collection: The data from the digital camera are collected by computer software. The first of 23 samples is highlighted below in lane 1 with base sequences appearing on the far left.



Data analysis: The data can be saved as a file for analysis by other computer software. Such software can provide a printout of the base sequence and perform comparisons with other DNA sequences (e.g. when looking for mutations).



Genome Analysis

Genome analysis involves determining the exact order of all of the millions of bases making up the DNA of an organism's **genome**. Genome analysis must also identify all the genes present, their correct and exact location in the base sequence, and the regions of DNA that control the activity of the genes. Chromosomes range in size from 50 million to 250 million bases; too large to handle for high resolution mapping and sequencing. They must be broken down into much shorter pieces, cloned, and sequenced. It is important to be able to assemble all of the sequenced fragments into a complete, continuous sequence for each chromosome. This is achieved by mapping known **genetic markers** at regular intervals along each chromosome. When DNA fragments are sequenced, the presence of DNA markers enables them to be correctly positioned in the overall sequence. A wide variety of applications are currently or potentially available for genome analysis, from the treatment and diagnosis of disease to ecological studies of diversity and phylogeny. Note that a small proportion of an organism's DNA occurs outside the nucleus, in the mitochondria, and is called the **mitochondrial genome** (mtDNA). Mitochondrial DNA is often targeted in studies of phylogeny because it is highly **conserved** (it codes for vital functions and changes very little over evolutionary time).

Mapping and Sequencing the Genome

1 Chromosome cut into large fragments:

A human chromosome, consisting of 50-280 million base-pairs, is cut randomly into large fragments with **restriction enzymes**. Each fragment is 150 000 - 1 million base pairs long.

2 Create a clone library:

Each large fragment is inserted into a separate vector (yeast or bacterial plasmid) using DNA **ligation**. The plasmids are then cloned in cells to produce many copies of the fragments; known as a **clone library** (each type of DNA fragment resides in a separate culture).

3 Map fragment with DNA markers:

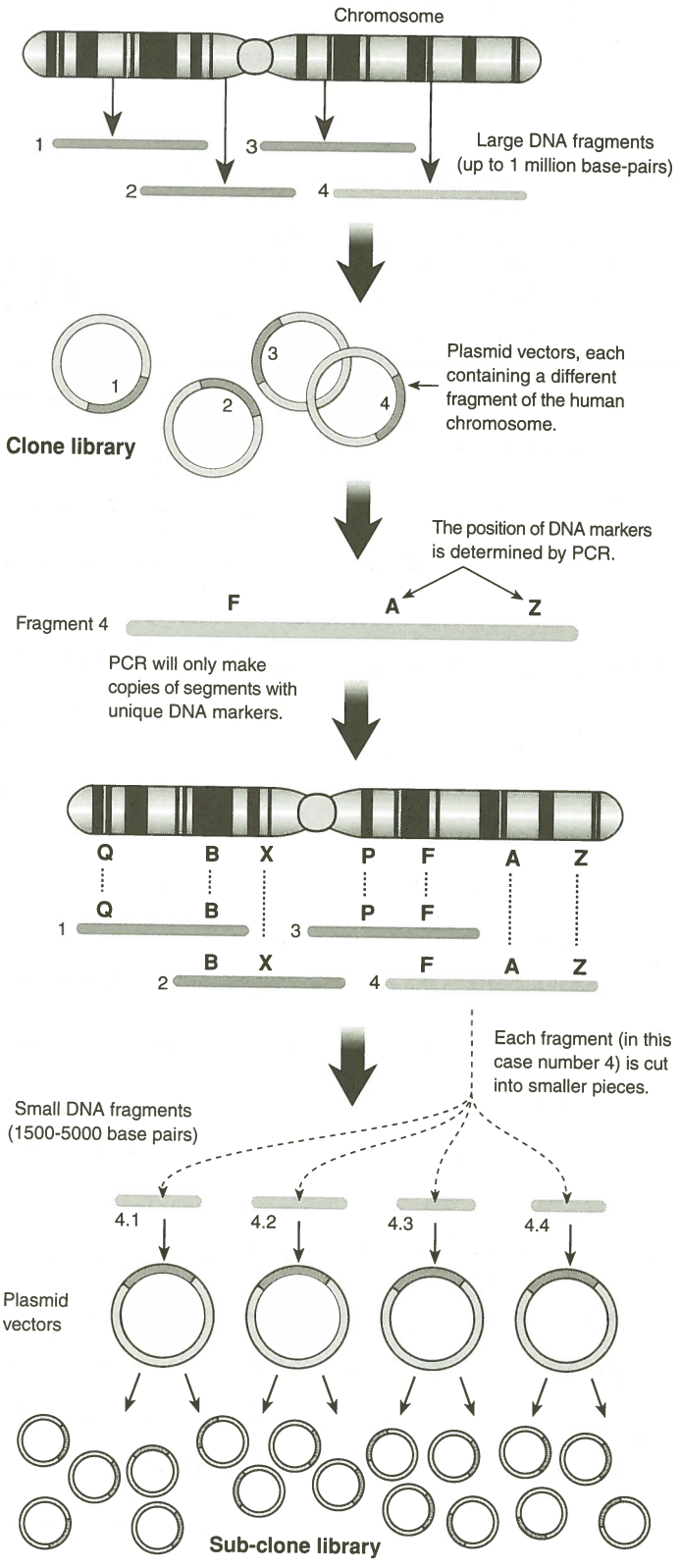
Each of the fragments are then mapped using **PCR** and **gel electrophoresis** to determine the position of **DNA markers** on each fragment.

4 Create a low resolution map of a chromosome:

Overlaps between the large fragments can be determined and a low resolution map of the chromosome can be built up. In reality, there would be many hundreds of large fragments created by the cutting up of a single chromosome (only four are shown here).

5 Each large fragment is cut into smaller pieces:

A large fragment is cut into smaller pieces (1500 - 5000 base pairs) with **restriction enzymes**. These smaller fragments are inserted into vectors using DNA **ligation** techniques. These vectors are then cloned to make copies; a **sub-clone library**.



6 Sequence fragments:

The small fragments are then sequenced using **gel electrophoresis** to find the exact order of the bases. Note: the examples on the right show only 15 base-pairs in each small fragment. In reality, they would be 1500 to 5000 base pairs in length.

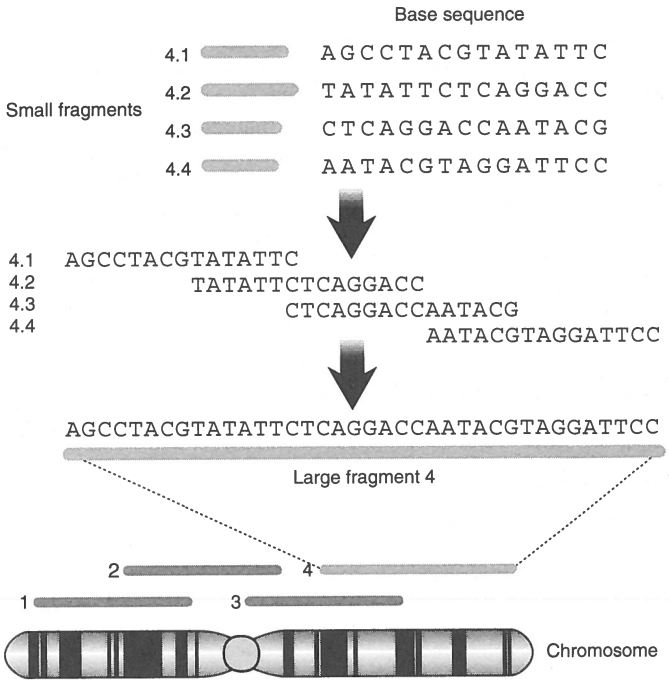
7 Overlapping sequences are assembled:

Overlapping sequences are assembled together using a computer to work out the sequence of the large fragment.

Steps 5-7 are repeated for all large fragments until the entire chromosome is sequenced.

8 Assemble sequenced fragments on chromosome:

The sequences of the large fragments are then assembled on the chromosome map to make a complete chromosome sequence.



1. Explain how the following two main components of genome analysis contribute to the overall process:

   (a) Genome **mapping**: _____

   _____

   (b) Genome **sequencing**: _____

   _____

2. Explain the difference between a **clone library** and a **sub-clone library**, identifying their role in genome analysis:

   _____

   _____

   _____

   _____

   _____

3. Describe the steps in the genome analysis process where the following techniques are used:

   (a) Restriction enzymes: _____

   _____

   (b) Ligation: _____

   _____

   (c) Polymerase Chain Reaction (PCR): _____

   _____

   (d) Gel electrophoresis: _____

   _____

4. Explain the role that DNA markers play in genome analysis: _____

   _____

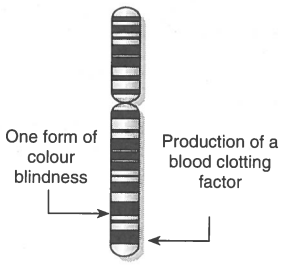
   _____

The Human Genome Project

The **Human Genome Project** (HGP) is a publicly funded venture involving many different organisations throughout the world. In 1998, Celera Genomics in the USA began a competing project, as a commercial venture, in a race to be the first to determine the human genome sequence. In 2000, both organisations reached the first draft stage, and the entire genome is now available as a high quality (golden standard) sequence. In addition to determining the order of bases in the human genome, genes are being identified, sequenced, and mapped (their specific chromosomal location identified). The next challenge is to assign functions to the identified genes. By identifying and studying the protein products of genes (a field known as **proteomics**), scientists can develop a better understanding of genetic disorders. Long term benefits of the HGP are both medical and non-medical (see next page). Many biotechnology companies have taken out patents on gene sequences. This practice is controversial because it restricts the use of the sequence information to the patent holders. Other genome sequencing projects have arisen as a result of the initiative to sequence the human one. In 2002 the International HapMap Project was started with the aim of developing a haplotype map (HapMap) of the human genome. Initially data was gathered from four populations with African, Asian and European ancestry and additional populations may be included as analysis of human genetic variation continues.

Gene Mapping

This process involves determining the precise position of a gene on a chromosome. Once the position is known, it can be shown on a diagram.

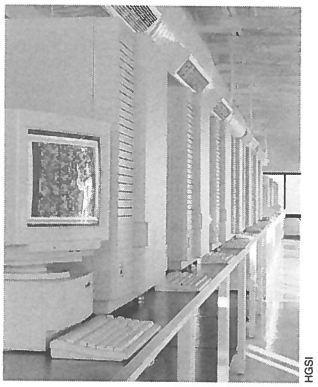


X chromosome

Equipment used for DNA Sequencing

Genesis Research and Development Corp, Auckland

Banks of PCR machines prepare DNA for the sequencing gel stage. The DNA is amplified and chemically tagged (to make the DNA fluoresce and enable visualisation on a gel).



HGSI

Banks of DNA sequencing gels and powerful computers are used to determine the base order in DNA.

Count of Mapped Genes

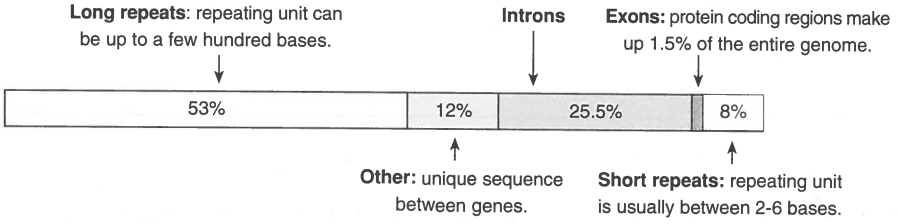
The length and number of mapped genes to date for each chromosome are tabulated below. The entire human genome contains approximately 20 000-25 000 genes.

Chromosome	Length (Mb)	No. of Mapped Genes
1	263	1873
2	255	1113
3	214	965
4	203	614
5	194	782
6	183	1217
7	171	995
8	155	591
9	145	804
10	144	872
11	144	1162
12	143	894
13	114	290
14	109	1013
15	106	510
16	98	658
17	92	1034
18	85	302
19	67	1129
20	72	599
21	50	386
22	56	501
X	164	1021
Y	59	122
	Total:	**19 447**

Data to March 2008 from gdb.org (now offline)

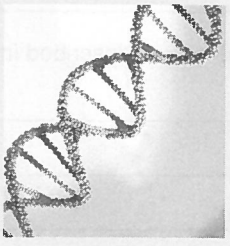
Composition of the Genome

About 97% of the genome does not code for protein and its function was largely unknown. Recent genomic analyses have revealed that some this DNA (the intronic DNA) codes for functional RNA molecules with important regulatory roles. Some of it is repeat sequence DNA, which means the same section (the repeating unit) of DNA sequence is present many times, often in close proximity. The length of a repeating unit varies from two to many hundred bases and may be present hundreds of times. Some repeat DNA can be difficult, or impossible, to sequence; a consequence of technical difficulties of working with sections of DNA with unusual chemistry. As a result of this, 8-10% of the human genome will probably remain unsequenced.



Qualities of DNA Sequence Data

The aim of the HGP was to produce a continuous block of sequence information for each chromosome. Initially the sequence information was obtained to draft quality, with an error rate of 1 in 1000 bases. The **Gold Standard sequence**, with an error rate of <1 per 100 000 bases, was completed in October 2004. Key results of the research are:



- The analysis suggests that there are perhaps only 20 000-25 000 protein-coding genes in our human genome.
- The number of gaps has been reduced 400-fold to only 341
- It covers 99% of the gene containing parts of the genome and is 99.999% accurate.
- The new sequence correctly identifies almost all known genes (99.74%).
- Its accuracy and completeness allows systematic searches for causes of disease.

Benefits and ethical issues arising from the Human Genome Project

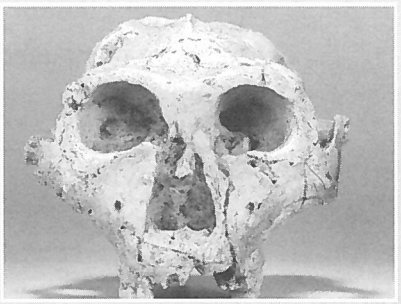
Medical benefits

- Improved **diagnosis** of disease and predisposition to disease by genetic testing.
- Better identification of disease carriers, through genetic testing.
- Better **drugs** can be designed using knowledge of protein structure (from gene sequence information) rather than by trial and error.
- Greater possibility of successfully using **gene therapy** to correct genetic disorders.

Couples can already have a limited range of genetic tests to determine the risk of having offspring with some disease-causing mutations.

Non-medical benefits

- Greater knowledge of **family relationships** through genetic testing, e.g. paternity testing in family courts.
- Advances **forensic science** through analysis of DNA at crime scenes.
- Improved knowledge of the evolutionary relationships between humans and other organisms, which will help to develop better, more accurate classification systems.



When DNA sequences are available for humans and their ancestors, comparative analysis may provide clues about human evolution.

Possible ethical issues

- It is unclear whether third parties, e.g. health insurers, have rights to genetic test results.
- If treatment is unavailable for a disease, genetic knowledge about it may have no use.
- Genetic tests are costly, and there is no easy answer as to who should pay for them.
- Genetic information is hereditary so knowledge of an individual's own genome has implications for members of their family.

Legislation is needed to ensure that there is no discrimination on the basis of genetic information, e.g. at work or for health insurance.

1. Briefly describe the objectives of the Human Genome Project (HGP) and the Human Genome Diversity Project (HGDP):

   HGP: _____

   _____

   HGDP: _____

2. Suggest a reason why indigenous peoples around the world are reluctant to provide DNA samples for the HGDP:

   _____

   _____

3. Describe two possible **benefits** of Human Genome Project (HGP):

   (a) Medical: _____

   _____

   (b) Non-medical: _____

   _____

4. Explain what is meant by **proteomics** and explain its significance to the HGP and the ongoing benefits arising from it:

   _____

   _____

   _____

   _____

5. Suggest two possible points of view for one of the **ethical issues** described in the list above:

   (a) _____

   _____

   (b) _____

   _____