

## Review Article

# Recombinant DNA technology and DNA sequencing

 **Mark A. J. Roberts**

Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom

**Correspondence:** Mark A.J. Roberts (mark.roberts@qmul.ac.uk)



DNA present in all our cells acts as a template by which cells are built. The human genome project, reading the code of the DNA within our cells, completed in 2003, is undoubtedly one of the great achievements of modern bioscience. Our ability to achieve this and to further understand and manipulate DNA has been tightly linked to our understanding of the bacterial and viral world. Outside of the science, the ability to understand and manipulate this code has far-reaching implications for society. In this article, we explore some of the basic techniques that enable us to read, copy and manipulate DNA sequences alongside a brief consideration of some of the implications for society.

This article is an updated version of the first Biochemical Society guide to recombinant DNA technology written by Peter Moore in 1994, and aims to reflect the current changes in this fast-moving field and think briefly about how these fit within society.

DNA is the basic store of information within cells and our ability to both understand and manipulate its content has been an important way by which we have come to understand our world. DNA itself is made up of nucleotides which are capable of hydrogen bonding to each other forming base pairs. The nucleotides link to each other via a phosphate group which link between carbons 5 and 3 on the sugar part of the nucleotide. DNA is made of up two such chains which run in the opposite direction (anti-parallel) to each other (Figure 1). The bases sit within the centre of this and pair with each other forming the typical double helical structure of DNA. These chains have a large number of phosphate groups which are negatively charged, giving DNA a large negative charge. For a further discussion on DNA, see the companion article in this issue [1].

Information is stored within DNA in the order of the bases and that information is used in the cell during the process of transcription. Because of the base pairing, each DNA strand is the opposite of its complement DNA and can be easily copied by using one strand as a template to synthesise the other, a mechanism termed semi-conservative. This mechanism results in a high level of accuracy, meaning information stored in DNA is very stable over time. For a further discussion of DNA replication and the semi-conservative mechanism, see the companion article in this issue [1].

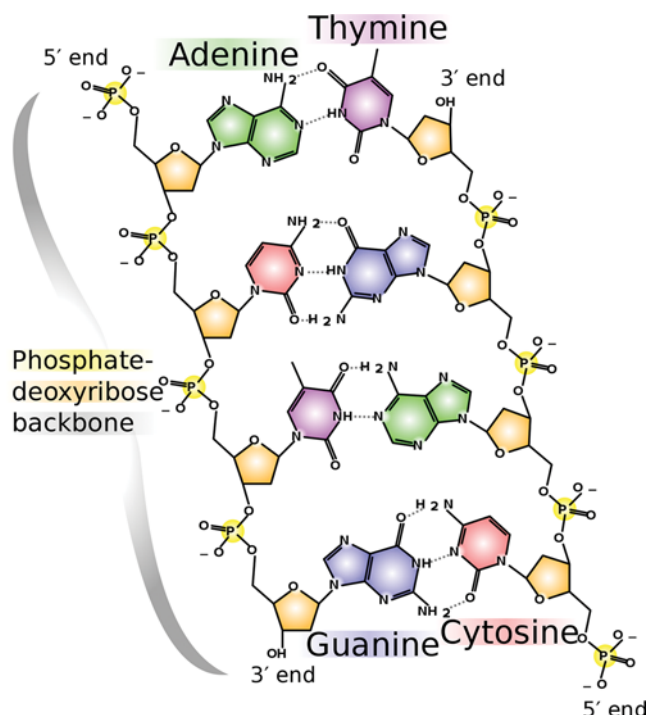
Our ability to understand and manipulate DNA has grown over time and has been linked to our increased understanding of how organisms use DNA and in particular the lifestyle of viruses and bacteria.

## DNA isolation

DNA can be isolated from cells through several simple steps. First, the cells are broken open using either a detergent or through force (e.g. sound waves) which breaks open the cell membrane. Then proteins from the cell are degraded using a protease enzyme, this ensures they do not precipitate with the DNA at the next step, and then the DNA turned to a solid precipitate by adding cold alcohol. You can then isolate the solid DNA using centrifugation to collect the now solid DNA at the bottom of a tube. A simple method for isolating DNA at home is described in the Supplementary Information.

Received: 02 May 2019  
Revised: 20 August 2019  
Accepted: 02 September 2019

Version of Record published:  
11 October 2019



**Figure 1.** The structure of DNA showing the phosphate backbone and the base pairing between the bases

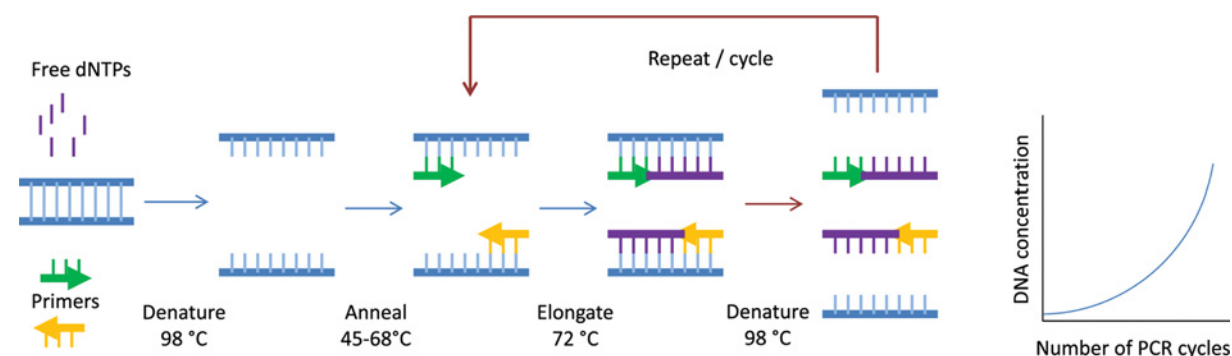
The 5' and 3' ends are shown along with the antiparallel nature of the two backbones. Image created by Madeleine Price Ball (CC0 licence).

## DNA amplification—the polymerase chain reaction

One of the first things we may wish to do when studying and manipulating with DNA is to amplify it from a particular source such that we have plenty of that DNA to work with. This is done using the polymerase chain reaction. The polymerase chain reaction was first developed in the 1980s and is an artificial way to copy and amplify DNA, though has many similarities to how cells copy DNA. This was developed at the biotech company, Cetus, by the biochemist Kary Mullis working with Saiki and Erlich. Mullis was awarded a Nobel Prize in 1993 for his work in developing PCR.

To perform PCR, you need a template DNA you wish to copy from, two primers to define which region of DNA you wish to copy. To this, you add free nucleotide tri-phosphates and a thermostable DNA polymerase. This is then placed in a PCR machine, which is essentially a programmable heat block. This cycles the reactants through three stages (Figure 2). Stage one heats the sample to 95–98°C and acts to separate all the DNA in the sample into single strands. In the second stage, the sample is cooled usually approximately 50–65°C, in this stage the primers anneal (bind) to the template DNA. The choice of temperature here is important to ensure you get the product you want, too high and your primers will not anneal, too low and there is an increased probability that your primers may bind elsewhere giving you unwanted products. Once this is done, the third and final step involved heating the sample to the optimal temperature for the DNA polymerase enzyme, which is usually between 72 and 74°C. In this stage, the polymerase is active and extends out new DNA from the primers, effectively copying the DNA. This is run in cycles and results in an exponential growth in the concentration of DNA with the products of each round becoming a further template for amplification. This exponential growth results in rapid amplification of the DNA of interest that is defined by the primers at each end.

The DNA polymerases used for PCR are isolated from organisms and are the same polymerases that are used by those organisms to replicate their DNA. These polymerases have a number of different properties, including the ability to proofread, their speed and different stabilities at different temperatures. When the technique was first developed, the polymerases were not thermostable, thus degraded after each round, so were replaced after each round of PCR. Then following from the purification of DNA polymerases from bacteria that live at high temperatures, thermostable polymerases, which did not denature during the high temperatures of PCR, became the norm. The first to be used was derived from the bacteria *Thermus aquaticus* and was thus called taq polymerase. This enzyme, while being thermostable, lacks the ability to proofread. Proofreading is the ability of DNA polymerases to check the last base



**Figure 2. The polymerase chain reaction**

Left: A PCR mix is made with primers, free deoxy-nucleotides (dNTPs), a DNA template and a DNA polymerase. This is heated to make the DNA single stranded (denature the DNA). The sample is then cooled to enable the primers to bind to the single-stranded template DNA. The sample is then heated to the optimal temperature for the DNA polymerase (typically 72°C for Taq), which enables the polymerase to extend from the primer, copying the DNA. This process is then repeated (or cycled through) resulting in the exponential increase of the concentration of DNA. Right: A graph showing the exponential increase in DNA concentration compared with each repeat (cycle) of PCR.

added was correct and to remove it if it is not. Taq polymerase lacks this ability and so introduces errors while copying DNA at a rate of 0.2% (for an average human gene this would result in approximately 34 bases being incorrect), clearly if we wish to use the DNA for further manipulation then accurate copying is important thus in more recent years a wider range of polymerases is available to the researcher, which include those with high accuracy. These polymerases are used where the sequence of the DNA is important for the downstream study or where you want to produce protein from the DNA.

As a technique, PCR enables us to produce large amounts of DNA for manipulation. This can be useful, in that it enables us to detect and work with small quantities of sample, for example those from a crime scene.

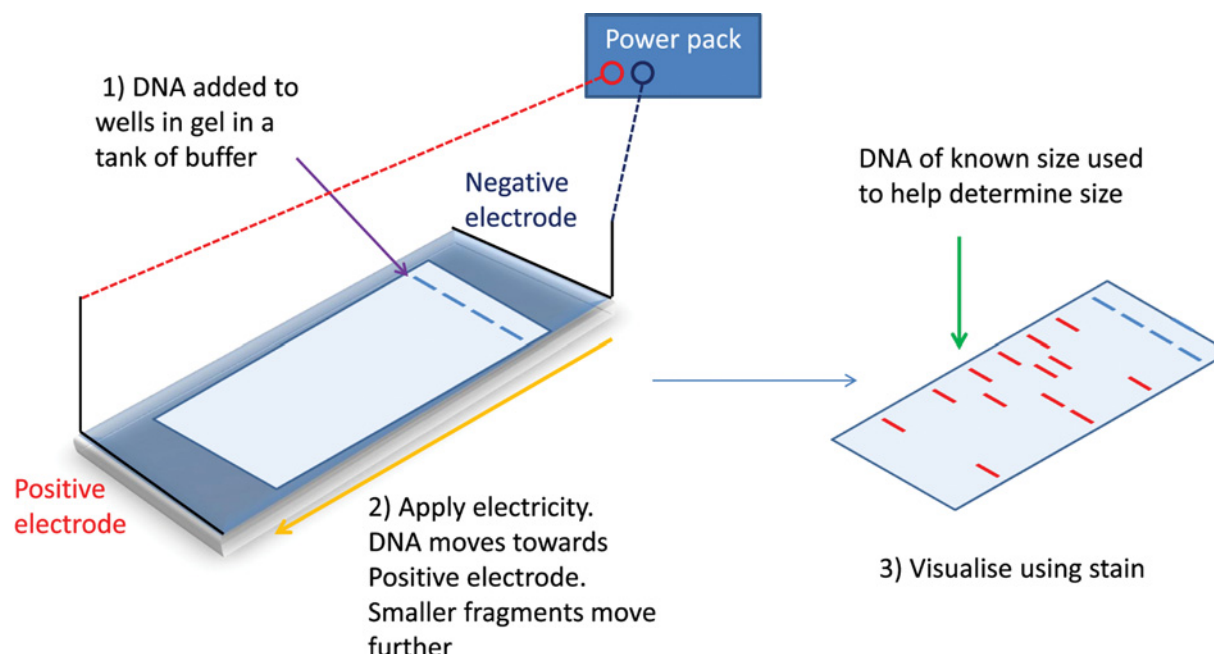
## Detecting DNA

DNA is invisible to the naked eye and so we need to visualise DNA. This can be done using several methods. One method is to make use of the fact that DNA absorbs UV light at 260 nm, we can thus measure the absorption of this to measure DNA. We can also make use of stains such as Ethidium Bromide, SYBR Green and SYBR safe. These stains bind to DNA and then we can visualise the stain to observe the DNA.

One method in which these are used is DNA electrophoresis. DNA separates DNA by size using a gel made of the seaweed derived agarose. The agarose is set to form a porous mesh, the size of the mesh being dependant on how much agarose is used. Typically 0.8–2% agarose gels are used, which also contain a stain such as the ones above to enable visualisation. The DNA to be separated is added to wells in the gel and a voltage applied across the gel by connecting the tank to a power source. As noted previously, DNA is negatively charged so moves towards the positive electrode, through the gel. The mesh of the gel impedes the DNA migration, linked to the length of the DNA, in that larger DNA fragments migrate more slowly (Figure 3). This separates the DNA by size. Exact lengths can then be measured by also running pieces of DNA of known sizes (a DNA ladder) alongside your sample. Once run, the DNA is visualised using the correct light source for the stain, with the DNA appearing as bands on the gel (Figure 3). This is the mechanism of how people's DNA fingerprints are measured in forensic science.

## Vectors—a means of storing DNA

Once DNA has been amplified, one thing that is often done is to place it and store it within a vector. In this context, a vector is a piece of DNA that is used to carry another. Vectors often have the ability to replicate within an organism and can include appropriate signals allowing the piece of DNA to be expressed, enabling further study. Vectors range from plasmids to artificial chromosomes (Table 1). Plasmids are small circular pieces of DNA and as a bare minimum they contain an origin of replication, enabling cellular machinery to replicate the plasmid and often some form of selectable marker, enabling you to select for organisms which contain it. This marker can either be a gene which confers antibiotic resistance or a gene that enables growth on media lacking a certain nutrient. Growing the bacteria with the antibiotic or in media lacking the nutrient enables you to ensure that all your population of cells contains the



**Figure 3. Gel electrophoresis**

A gel is made of agarose, which is melted and then set. DNA samples are then added to the wells in the gel. A voltage is then applied using a power pack via the electrodes and DNA, due to its negative charge, moves towards the positive electrode. The gel acts to impede progress, with larger pieces moving more slowly. Thus the DNA separates by size. The gel can be visualised by using a stain, which binds to DNA for example ethidium bromide or SYBRGreen and UV light. Measurement of exact size is possible if on the same gel some standards of known sizes are also run.

**Table 1 Common vectors used for DNA manipulation and the size of DNA that can be inserted**

Vector type	Insert size (kilobase pair)
Plasmid	0.5–8
Bacteriophage vector	9–25
Cosmid	30–45
Bacterial artificial chromosome	50–300
Yeast artificial chromosome	250–1000
Human artificial chromosome	6000–10000

vector. Vectors are replicated by the host machinery and thus are copied enabling the DNA to be stored. Vectors can also be stored outside of an organism in the freezer. Plasmids are versatile vectors which can be used in many bacterial species and eukaryotes such as yeast. Plasmids can also contain the necessary signals for the DNA they contain to be expressed as a protein and are thus called expression plasmids. The ability to produce protein from DNA contained in plasmids allows us to make recombinant protein which is useful both for biological study and also for using proteins as therapeutics, such as recombinant insulin.

## DNA manipulation joining DNA together

DNA can be placed into vectors using a variety of methods. The first method we will discuss is by using restriction endonucleases. These are enzymes, isolated from bacteria, which bacteria use to protect themselves from viral attack. These enzymes recognise and cut DNA at a specific DNA sequence. Depending on how they cut, some cut the DNA flat, cutting both strands at the same point, termed blunt ends, while others cut the backbone at different points resulting in small sections of single-strand DNA, termed sticky ends (Table 2). Once cut with a restriction enzymes, two pieces of DNA can be joined together by mixing them and adding the enzyme DNA ligase. DNA with complementary sticky ends produced by restriction enzymes join together easily with the base pairing of the sticky ends helping speed the process. Ligation of DNA with blunt ends is also possible but this occurs with a lower efficiency (i.e. less of the

**Table 2** Some commonly used restriction endonucleases along with the DNA recognition sites, where they cut and if the end produced is sticky or blunt

Enzyme	Recognition site (cut: /)	Ends produced
<i>AluI</i>	AG/CT TC/GA	Blunt
<i>BamHI</i>	G/GATC C C CTAG/G	Sticky
<i>EcoRI</i>	G/AATT C C TTAA/G	Sticky
<i>HindIII</i>	A/AGCT T T TCGA/A	Sticky
<i>NotI</i>	GC/GGCC GC CG CCGG/CG	Sticky
<i>Sau3AI</i>	/GATC CTAG/	Sticky
<i>SmaI</i>	CCC/GGG GGG/CCC	Blunt

DNA gets joined). Combinations of restriction endonucleases can be used in order to ensure that DNA pieces are joined together in a defined way.

In more recent years, DNA assembly techniques, based on PCR, have been developed and used to join pieces together. Good examples of these assembly techniques are splicing overlap extension PCR and Gibson assembly, the latter having been used to generate an entire bacterial chromosome from scratch. With these PCR-based techniques, DNA fragments are created with ends that overlap, through using primers that can bind to both sequences. With splicing overlap extension PCR, separate PCRs are done for each fragment. The products of the two reactions are then mixed with the products of the first reaction becoming ‘super primers’ for the second, joining the DNA together (Figure 4). Gibson assembly is similar but instead uses an exonuclease enzyme to cut back the one strand of the DNA, forming sticky ends akin to those formed in restriction enzymes. The DNA is then joined together by a DNA ligase (Figure 4). Both these techniques can be used to assemble large DNA fragments and to place DNA into vectors.

The ability to make an entire genome using techniques like Gibson assembly opens up the possibility of designer organisms, designing a whole bacterial or eukaryotic cell to do a specific biotechnological role.

## DNA sequencing—reading what is there

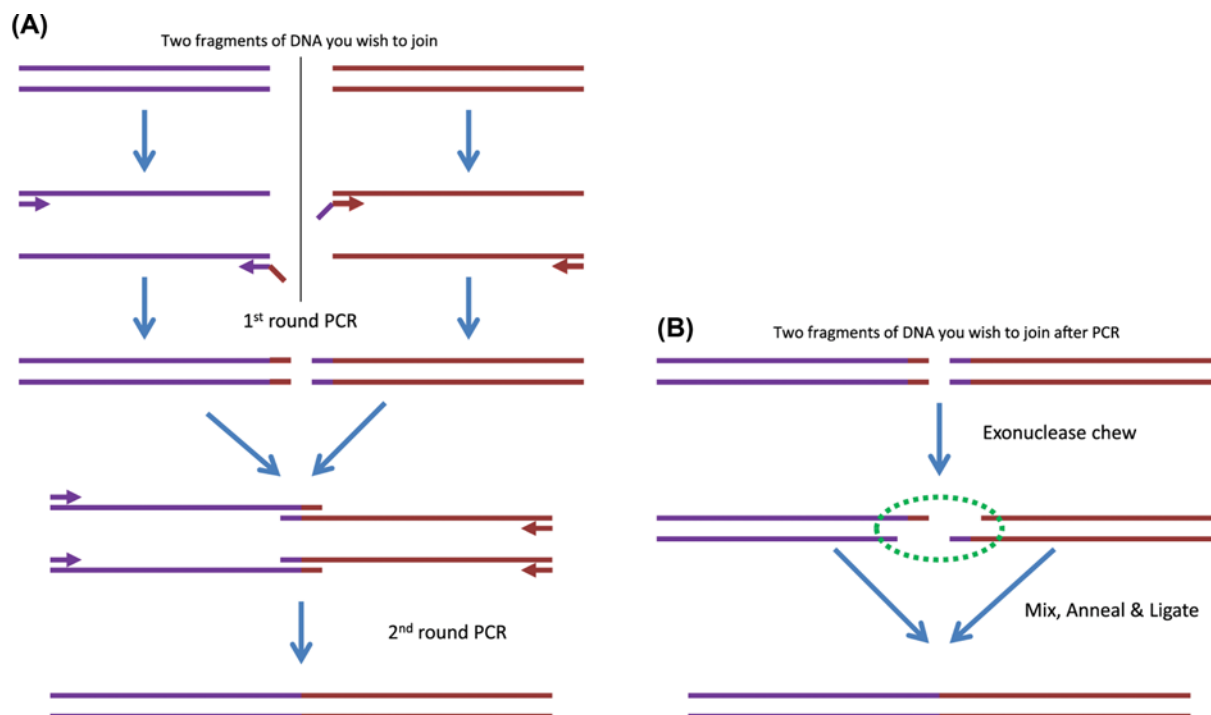
One of the greatest achievements of modern times is perhaps the human genome project. This was determined using the Sanger DNA sequencing method, a method based on PCR. To perform Sanger sequencing you require a primer, some nucleotide tri-phosphates (dNTP) and some labelled di-deoxy nucleotides (ddNTPs). A PCR is conducted from the primer and the DNA extended by the DNA polymerase.

In normal PCR, you extend by adding nucleotide tri-phosphates, within the sequencing reaction if a ddNTP gets incorporated then this cannot be extended further by PCR so the reaction terminates. If those di-deoxy are labelled, then the fragments produced can be separated by size using capillary electrophoresis (this works on the same principle as gel electrophoresis but on a smaller scale) and then this used to read the sequence. When the technique was first developed, the di-deoxynucleotides were labelled using radioactive phosphate, but more modern versions of the technique use different coloured fluorescent dyes to label the ddNTPs (Figure 5). This method allows you to sequence up to approximately 1000–1500 bases in length before accuracy is lost.

It is important to put the 1000 bases in context, in that a typical bacterial genome is 4 million base pairs and the human genome 3 billion base pairs, so to be able to sequence an entire genome, we need to extend this method and we do this by what is termed shotgun sequencing. The genome is split into small fragments either using restriction enzymes or by using sonication [sound waves] to break up the genome into small fragments. These small fragments are placed into vectors and then sequenced. Once sequenced the individual sequences are then assembled by using computer algorithms to put together overlapping segments in order to build up an entire genome.

While Sanger sequencing was used to generate the first human genome in 2003, since then a number of other DNA sequencing techniques have been developed. The first of these were termed next generation sequencing and include 454 and Solexa/Illumina sequencing. Like Sanger, both these techniques use DNA polymerase and added nucleotides to perform sequencing. Further, third generation techniques have been developed including ion torrent and nanopore sequencing.



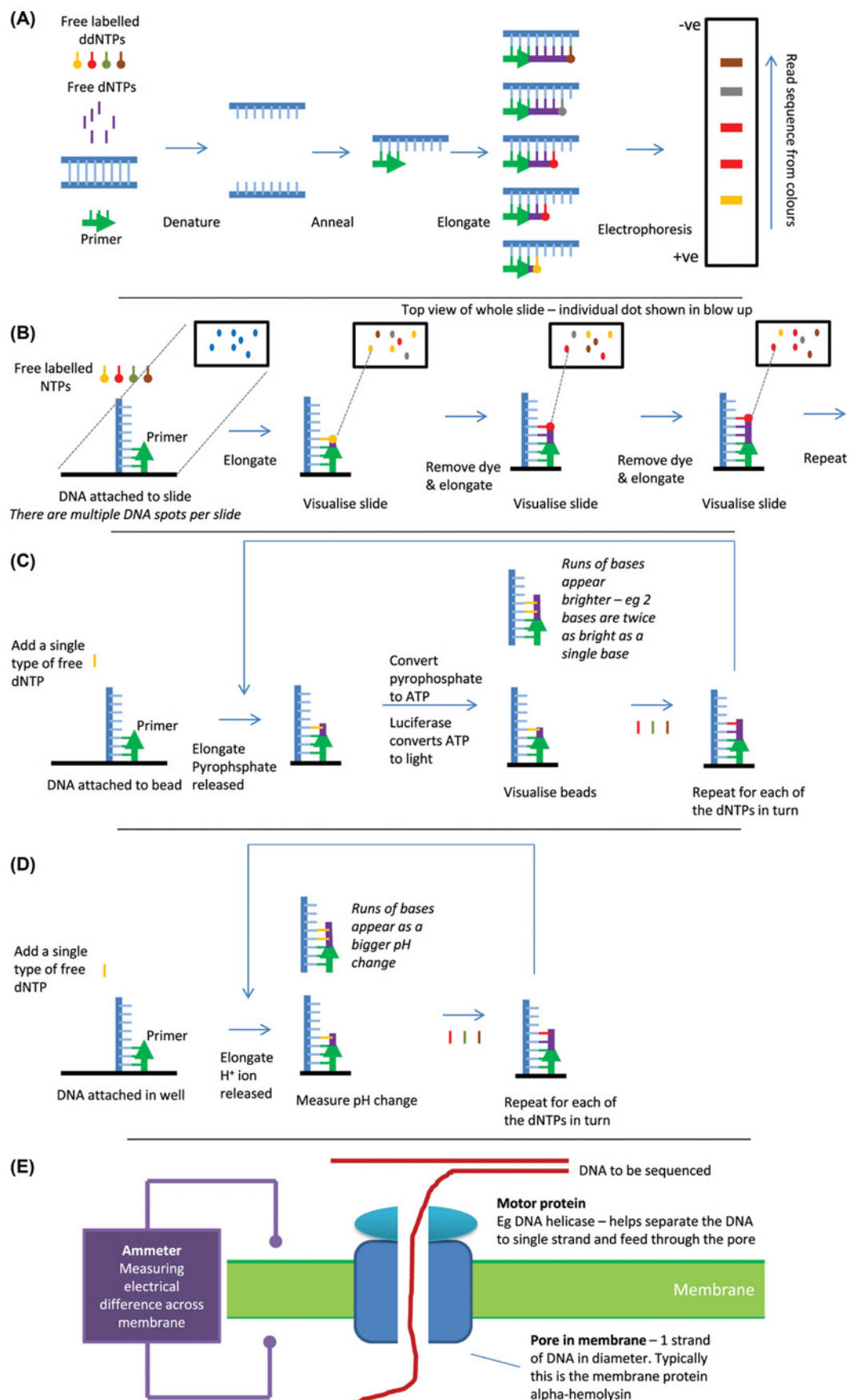


**Figure 4. DNA Assembly Techniques**

**(A)** Splicing overlap extension PCR—indicative diagram. Two separate PCR are done with some overlap between the middle primers. These amplify the two pieces of DNA you wish to splice together. The products of these two separate PCR are then mixed, and along with the outer primers, a further PCR is performed. The two products from the first reactions are complementary and thus anneal linking the two fragments together. **(B)** Gibson assembly. Two separate PCR are done with some overlap between the middle primers as per with splicing overlap extension PCR. These result in two PCR products, which have a region, which is complementary. An exonuclease enzyme is added which removes single strands from each end producing sticky ends, similar to those produced by restriction enzymes. The exposed ends can then be annealed and ligated. Both splicing overlap extension PCR and Gibson assembly can be scaled to join multiple fragments together.

To perform Solexa/Illumina sequencing, first the DNA is split into small fragments and attached to as a spot on a slide (Figure 5), with multiple spots on the slide. A PCR is then performed resulting in each spot containing multiple copies of the fragment, effectively amplifying the final signal. The DNA is split to form a single strand and primers and DNA polymerase added to perform the sequencing. Bases are then washed over the slide, with a fluorescent dye blocking the further extension. This means that each fragment is extended one base at a time, with a different fluorescent dye depending on the base added. Once the base is read using the dye, the dye removed, enabling extension and bases washed over again (Figure 5). The numerous spots on the slide make it possible to sequence a large number of fragments in parallel. The downside of this technique is the short read lengths of 50–150 bases, which can make the assembly of the sequences to make a larger one difficult for repetitive regions of DNA. However, if you have a reference sequence then it is possible to use this to aid computer assembly process.

In 454 sequencing, DNA is attached to beads and PCR is performed to amplify the DNA on the beads such that each bead is covered in one particular DNA sequence, which is then denatured to form single-stranded DNA. These beads are then transferred onto a slide containing wells that are one bead in size. A primer and DNA polymerase is added and nucleotides are then run over the slide in turn. If the DNA polymerase is able to add a base then a pyrophosphate molecule and a  $H^+$  ion is released [1]. The sequence is read by measuring this pyrophosphate that is released when the complimentary base is added at the end of the fragment as part of the action of DNA polymerase. The pyrophosphate is measured by enzymatically converting it to ATP. The ATP is then measured using the enzyme luciferase which uses ATP to produce light. The light produced is then measured electronically. Unlike Illumina and Sanger, if there is a run of the same base then multiple bases can be added at the same time and thus the level of the light emitted relates to the number of that particular nucleotide at that point; for example, CCC will be three times brighter than C. Read lengths from 454 can be as long as 700 bases.



### Figure 5. DNA sequencing techniques

A Sanger: dye-labelled (ddNTPs, lacking the 3' hydroxyl group) are added to a PCR. The ddNTP terminates the PCR at the point where it is added. The DNA is then separated by size through electrophoresis and the base at the end at each point can be read, depending on the dye label (colour); B Illumina: DNA is fixed to spots on a slide and amplified by PCR such that each spot contains multiple copies of one piece of DNA sequence. Once this is done, the DNA is made single stranded and deoxynucleotides (dNTPs), which are labelled with a dye such that they cannot be extended in a PCR (the 3' hydroxyl group is blocked), are washed over the slide, along with DNA polymerase. Those that bind are detected at each point on the slide, by visualising the dye. The dyes are then removed, and fresh-labelled deoxynucleotides (dNTPs) are added which enables the cycle to be repeated and the sequence read. C 454 sequencing: DNA is attached to individual beads and amplified by PCR. The beads are then placed onto a slide. Deoxynucleotides (dNTPs) are then washed over in turn with DNA polymerase present. If the dNTP is complementary, then pyrophosphate is released by the action of DNA polymerase. This pyrophosphate is detected through enzymatic conversion to ATP and then the ATP is detected by the enzyme luciferase, which converts ATP to light which can be measured. Once measured, a different dNTP is then washed over and the cycle repeated in turn to read the sequence. In 454 sequencing, if there is a run of the same base then these are read at once as they will produce a brighter light signal, e.g. CC will be twice as bright as C and CCC three times as bright. D Ion Torrent: DNA is attached to the base of a well, amplified and made single stranded as per other sequencing techniques. Deoxynucleotides (dNTPs) are then washed over in turn with DNA polymerase present. If the dNTP is complementary, then a hydrogen ion ( $H^+$ ) is released by the action of DNA polymerase. This hydrogen ion release is detected by measuring the pH change. Once measured, a different dNTP is then washed over and the cycle repeated in turn to read the sequence. Similar to 454 sequencing, if there is a run of the same base in ion torrent sequencing then these are read at once as they will produce a bigger change in pH. E Nanopore: The DNA is made single-stranded DNA, typically by using a helicase and then is threaded through a protein pore (typically  $\alpha$ -haemolysin) held within a membrane. As the DNA passes through the membrane, there is a change in the electrical properties measured across the membrane which is different for each base. These electrical changes are measured and used to read the order of the bases passing through the pore.

In order to produce genome sequences, next generation techniques are often used alongside Sanger sequencing, the latter being used to bridge areas with larger repetitive elements, which require a longer read length.

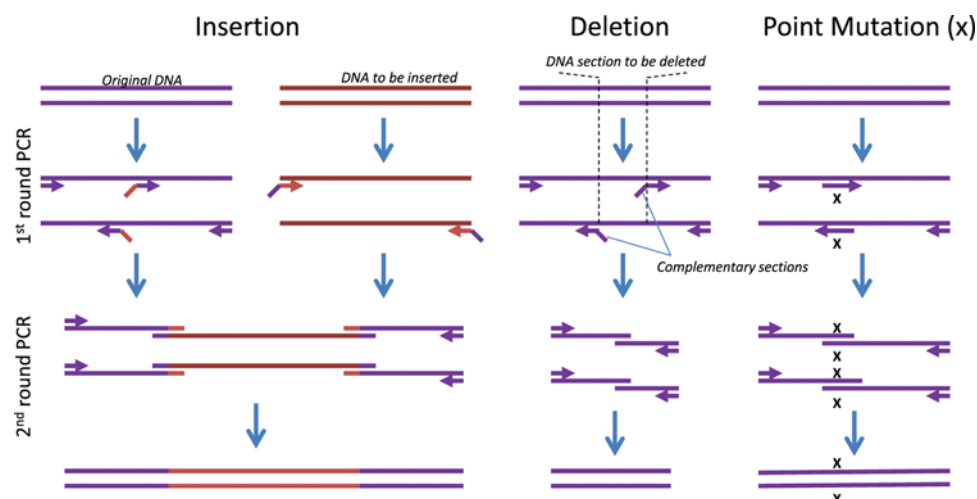
Third-generation sequencing techniques use a different approach which measures charge change when DNA is synthesised either directly in the case of ion torrent sequencing or across a membrane in the case of nanopore sequencing. In ion torrent sequencing DNA is immobilised within a microwell, with one DNA template per well. Again a primer is added along with DNA polymerase and nucleotides are then flooded over the slide in turn. Like 454 Ion torrent uses the chemistry of DNA synthesis, this time measuring the  $H^+$  ions which are produced as the DNA is elongated by DNA polymerase. The release of hydrogen ions is measured using a pH meter. Like 454, if there is a run of bases then multiple bases can be added at a single time and would be reflected in a larger pH change. Read lengths of approximately 400 base pairs have been achieved using this method.

Nanopore sequencing involves threading the DNA template through a pore in a membrane. This threading is done using an electric field, similar to the principle seen in electrophoresis earlier. To do this, a single-stranded DNA is passed through a pore in a membrane. Often this pore is made from the bacterial protein  $\alpha$ -haemolysin. The bases passing through are either measured by a change in the electrical properties, characteristic for each base (Figure 5) by measuring the electrical difference across the membrane. As part of the process, a helicase can be used to generate single-strand DNA. Nanopore sequencing offers the potential of long read lengths than other techniques and it also does not require a PCR step prior to analysis as required for other sequencing methods. Sequencing machines as small as a USB dongle have been developed which sequence DNA using the nanopore method, enabling DNA sequencing to be performed outside of the traditional lab setting.

## Changing the DNA sequence

For experimental reasons, you may wish to change the DNA sequence. Changing the DNA from the normal (termed wild-type) to something else is called a mutation and these can have positive and negative effects. This can be simply changing a single base from one to another, a point mutation, or much larger deletions and insertions of sections of DNA. This can be done randomly by using something that mutates DNA, for example, UV light or a chemical mutagen, which may be useful in some cases. These however make random changes and often the experimentalist wishes to make specific changes to the DNA sequence and this can be done in a number of ways which we will now explore.





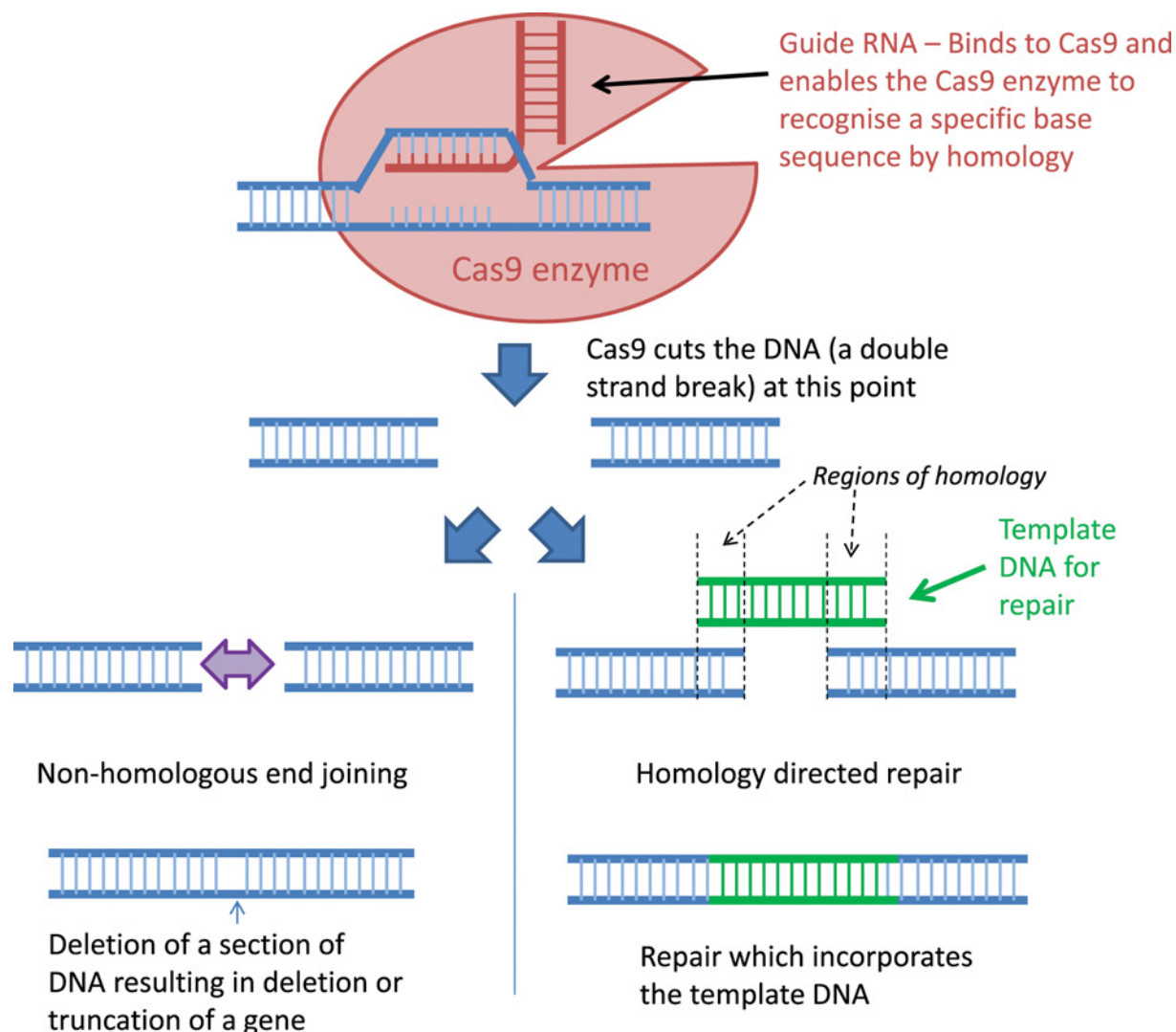
**Figure 6. Using splicing overlap extension PCR to create mutations**

(A) Insertion, (B) Deletion and (C) Point mutation, marked as X. In all cases, primers which overlap and contain matching sequences are used to join the DNA together as shown in more detail in Figure 4. For deletion and point mutation, the two first round PCRs are done in separate tubes and combined for the second round in the same way as they are for the insertion.

First, the simple way to do this is to perform PCRs with a mutation within the primer, which will result in the PCR product containing the mutation. While this will work for mutations at the end of a sequence, often there is a need to introduce mutations into the middle of a sequence or to make deletions or to join DNA together. This can also be done by PCR using splicing overlap extension PCR (Figure 4). When using this technique, the mutation you wish to perform is encoded by the middle primers, and this can be used to either make point mutations in the middle, delete portions of DNA or join portions of DNA together (Figure 6). In order to insert or delete DNA, the middle primers are used to join two complementary sections together, for a point mutation this is simply encoded in the middle primers. This technique uses two rounds (or stages) of PCR resulting in a piece of DNA, which is mutated in a specific way either by insertion, deletion or point mutation (Figure 6).

While these techniques allow you to create edited versions of the DNA for study techniques also are available to edit the genome *in situ*, based on the CRISPR-Cas9 system. Like restriction endonucleases, this pathway evolved as a mechanism for bacteria to avoid phage (viral) infection. Cas9 is an enzyme that cuts double-stranded DNA and it does so at a specific point. The Cas9 system uses a guide RNA to control where the Cas9 breaks the DNA. Within bacteria these guide RNAs are stored within the CRISPR region of their genomes, resulting in the name. They use these guide RNAs to degrade viral DNA thus preventing infection by the phage. By designing and synthesising guide RNAs, we can use the Cas9 enzyme to cut at a specific place in a genome (Figure 7). This ability is significant as it opens up the possibility of direct genome editing within a cell.

In order for the mutation to be made, once Cas9 has catalysed the specific break in the DNA, the DNA repair mechanism in the cell then kicks in to repair this break. DNA repair, especially of these double strand breaks is not always perfect, and the imperfect repair results in a mutation around the point where the DNA was cut by CRISPR. The result of this is that while the CRISPR-Cas9 system does mutate a specific gene, there is no control over the type of mutation that is caused as this depends on the DNA repair system. Two types of DNA repair are possible, the first (non-homologous end joining) results in the deletion of a piece of DNA (which in turn can result in deletion or truncation/shortening of a gene), but is not controlled, the second homology-directed repair, uses another piece of DNA as a guide to the repair and using this can result in very specific repair (Figure 7). In non-homologous end joining, the ends of the DNA are processed to make them compatible and the two ends joined by DNA ligase. This can result in the loss of genetic material. In homology-directed repair, a piece of DNA which shows homology to both sides of the break is used as a template for the repair. Thus depending of what DNA repair occurs depends on how specific the type of mutation is. The challenge of promoting one type of repair over another is currently a hot topic of research, with the aim of promoting homology-directed repair to enable specific genome editing.



**Figure 7. CRISPR-Cas9 genome editing**

The Cas9 enzyme is directed to the piece of DNA you wish to cut via a guide RNA. The enzyme cuts the DNA forming a double strand break. From this two possible things can occur. One option is non-homologous end joining where the DNA is joined back together. This can occur imprecisely and if so a mutation is introduced into the target gene. The second option is homology-directed repair, which uses a piece of DNA, which matches both sides of the break as a template for the repair. This process enables accurate DNA editing.

One further potential problem with this system is what is termed 'off target' effects. This refers to the ability of the guide RNA to bind elsewhere within a genome, this depends on their being similar sequences elsewhere in the genome. If it does then you could get mutations at this location in addition to your target mutation. Thus the guide RNA requires careful design.

## Recombinant DNA technology and society

Recombinant DNA technology and genetic modification are rarely out of the media spotlight be this through genetically modified crops, genetically modified mosquitos, the use of genome editing in humans or the role the DNA forensic technology is having in the world of crime. Recombinant technology is already in regular use in the production of medicines such as insulin and in the production of the anti-malarial drug artemisinin. While the potential benefits of the technology are huge, it must be noted that there are potential problems, both safety and ethical, which need to be taken into account prior to use. With the production of chemicals such as artemisinin, we also have the

interesting issue of making this in the developed world removes jobs from the developing world where the plant containing it is cultivated.

## Plant genetic modification

The genetic modification of plants both for food and other uses remains contentious. Modifications that have been introduced include herbicide resistance, allowing the use of herbicides to kill other plants, thus increase yield of your target crop, 'Golden Rice', which was engineered to produce  $\beta$ -carotene a precursor to vitamin A and strains of wheat which have been engineered to produce aphid alarm pheromones, with the aim of reducing aphid infestation. Plants have also been designed to produce vaccines.

Golden Rice was developed to provide a source of vitamin A for people with diets that lack vitamin A; akin to flour fortification with vitamin B1 in the U.K. Using these plants is not without problems both in terms of the potential issues with the release of a genetically modified organism, the interaction with other organisms, alongside potential conflicts of interest of agri-tech and farmers. For example, there is the question of interbreeding and should the genetically modified plants be able to breed. If they are able to breed then they could inter-breed with non-modified plants and produce unforeseen effects. If the genetically modified plants are made sterile then farmers are dependent on the seed manufacturer each year for seed rather than being able to derive them from a plant. Some genetically modified plants provide herbicide resistance, enabling the use of herbicides to kill unwanted plants, but this again increases farmers dependence on herbicides and increases their use in the environment. With these plants, there is also the danger that if the GM plants are able to breed with wild plants the gene enabling plants to be resistant to the herbicide may be released into wild plants and thus reduce the effectiveness of the herbicide in the wider environment.

In addition to this, there are also the ethical issues of food supplementation/fortification and modification. Some of these concerns stem from the potential for unforeseen interactions resulting from the modification or that the supplementation may not result in successfully curing the dietary deficiency.

## Genetic testing and the future

As can be seen from the companion article of genetic basis for disease [2], the ability to use these techniques to test for inherited genetic disorders is important and indeed is forming a bigger and bigger role in our healthcare. The 100000 genomes project, launched by the U.K. prime minister in 2012 is a large genomic study and has demonstrated the utility of genomic medicine, enabling the diagnosis and further study of a range of genetic diseases and in the study and treatment of cancer [3]. Other projects such as East London genes and health uses sequencing to find natural mutants within our populations for study [4]. By finding people who already have mutations in genes, who live naturally in our population, we can use cells derived from them, reducing our need to modify genes and further adding to our knowledge of how genes interact.

With the cost of genome sequencing constantly falling, it is likely that it will become the norm, as part of personalised healthcare. Within this people, genomes (or the genomes of their cancer cells/bacterial infection) are taken into account and used to define a specific treatment regime for them. Indeed this may also move into the realm of preventative medicine, but here we have the ethical issue that containing a gene may only give a chance of disease and at what point do you decide to act. A further discussion of this can be found in the article of genetic disease [2]. There is also the ethical issue of both privacy and ownership of genetic information.

Within the year, this article was published, we have already seen girls born in China who have had their genome edited using CRISPR-Cas9, carrying a mutation that may provide some immunity from HIV. They will pass this onto any children they have and the process around them does raise some challenging ethical issues around consent and who decides what we should modify amongst others. There is also the question in this case as it is not a full deletion if the truncated protein may have safety effects.

Overall, the impact of genetic understanding and modification on society throws up a challenging set of issues where both the potential benefit and potential harm to both the individuals involved and society must be weighed up.

## Summary

- The ability to sequence genomes in large numbers is continuing to change the questions we can answer and the new experiments we can design.
- Our ability to make precise DNA edits both for the purposes of study and application is constantly getting better and more accurate.
- This technology opens up the possibility of both personalised medicine and designer organisms and crops.
- These technological developments bring challenging ethical questions about their use and scope.

## Acknowledgments

I am grateful to Dr Ailsa Powell and Rob Maddock alongside Duncan McMillan and Charlotte Morgan from QMUL's Centre of the Cell for their helpful comments and suggestions in putting this article together. This article is an updated version of the first Biochemical Society guide to recombinant DNA technology written by Peter Moore in 1994.

## Competing Interests

The author declares that there are no competing interests associated with the manuscript.

## References

- 1 Lodge, J. and Minchin, S. Understanding Biochemistry: Structure and function of nucleic acids. *Essays Biochem.* **63**, 433–456, <https://doi.org/10.1042/EBC20180038>
- 2 Jackson, M., Marks, L., May, G.H.W. and Wilson, J.B. (2018) The genetic basis of disease. *Essays Biochem.* **62**, 643–723, <https://doi.org/10.1042/EBC20170053>
- 3 Genomics England. The 100,000 Genomes Project. <https://www.genomicsengland.co.uk/about-genomics-england/the-100000-genomes-project/> Accessed May 2019
- 4 Genes and Health. East London Genes and Health. <http://www.genesandhealth.org/> Accessed May 2019
- 5 Nuffield Council on Bioethics. Genome Editing and Human Reproduction: social and ethical issues. <http://nuffieldbioethics.org/project/genome-editing> Accessed May 2019

## Recommended reading

- Hofmann, A. and Clokie, S. (2018) *Wilson and Walker's Principles and Techniques of Biochemistry and Molecular Biology*, Cambridge University Press, Cambridge, U.K.
- Lodish, H.F., Berk, A., Kaiser, C.A., Krieger, M., Bretscher, A., Ploegh, H. et al. (2016) *Molecular cell biology*, WH Freeman, New York, U.S.
- Judson, H.F. (1996) *The eighth day of creation: makers of the revolution in biology*, Cold Spring Harbour Press, New York, U.S.
- Brown, T.A. (2016) *Gene cloning and DNA analysis: an introduction*, Wiley-Blackwell, New Jersey, U.S.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2008) *Molecular biology of the cell*, Garland Science, New York, U.S.
- Big Picture. Genes, Genomes and Health. <https://bigpictureeducation.com/genes-genomes-and-health> Accessed May 2019
- Golden Rice Humanitarian Board. The Golden Rice Project. <http://www.goldenrice.org/> Accessed August 2019
- World Health Organization. WHO report on genetically modified food. [https://www.who.int/foodsafety/areas\\_work/food-technology/faq-genetically-modified-food/en/](https://www.who.int/foodsafety/areas_work/food-technology/faq-genetically-modified-food/en/) Accessed May 2019
- The Biochemical Society. Sciberbrain activity on genetic engineering ethics. <http://sciberbrain.org/Advanced/Genetic-Engineering/Science> Accessed May 2019
- The Naked Scientists. Naked Scientist podcasts covering the whole area of genetics including both its application and ethics. <https://www.thenakedscientists.com/podcasts/naked-genetics> Accessed August 2019
- Jorgensen, Ellen. TED talk on using CRISPR gene editing. <https://www.youtube.com/watch?v=1BXYSgepx7Q> Accessed August 2019

## **Supplementary information**

### **DNA extraction at home**

This method can be used to extract DNA from pretty much any living thing. It works well on a variety of fruits and vegetables. For this example I use strawberries as they look and smell nice. Two methods are listed, the first working well on soft fruit, the second works well for all fruit and veg. In both methods pineapple juice is used. Within the method we are using the fact that pineapple juice contains protease enzymes to degrade the proteins with the mixture an alternative would be to use meat tenderiser.

#### **Zip lock bag method**

You will need: 1 zip lock bag

3-4 strawberries

Washing up liquid

100ml water

Table salt

Coffee filter

Cup

Pineapple Juice

Methylated Spirits or Surgical Spirit [available from pharmacies] – chilled in the fridge.

Take some [3-4] strawberries and place in a zip lock bag. Seal and place on a surface. Squash/gently smash the strawberries in the bag by hand.

Add 1 tsp of washing up liquid, 100ml of tap water and ½ a tsp of salt to the bag. Squash the strawberries further for about 1 minute.

Open the bag and pour the contents into the coffee filter placed over a cup. The liquid will then drip down into the cup leaving behind the strawberry debris.

Add roughly 5ml of pineapple juice to the liquid in the cup and mix. Leave 5-10 minutes

Add the total volume of chilled methylated spirits or surgical spirit to the cup. Do not mix.

You should now notice a white precipitate forming. This is DNA. You can scoop this out using a wooden stick or straw.

#### **Blender method**

You will need: 1 blender



3-4 strawberries

Washing up liquid

100ml water

Table salt

Coffee filter

Cup

Pineapple Juice

Methylated Spirits or Surgical Spirit [available from pharmacies] – chilled in the fridge.

Take some [3-4] strawberries and blend in the blender.

Add 1 tsp of washing up liquid, 100ml of tap water and  $\frac{1}{2}$  a tsp of salt to the blender. Leave with the strawberries for around 5-10 minutes.

Pour the contents into the coffee filter placed over a cup. The liquid will then drip down into the cup leaving behind the strawberry debris.

Add roughly 5ml volume of pineapple juice to the liquid in the cup and mix. Leave for 5-10 minutes.

Add the total volume of chilled methylated spirits or surgical spirit to the cup. Do not mix.

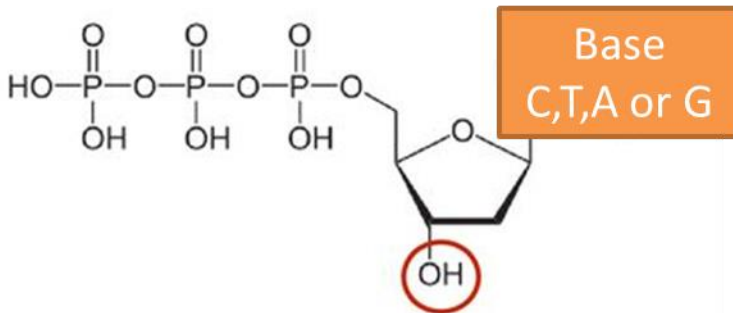
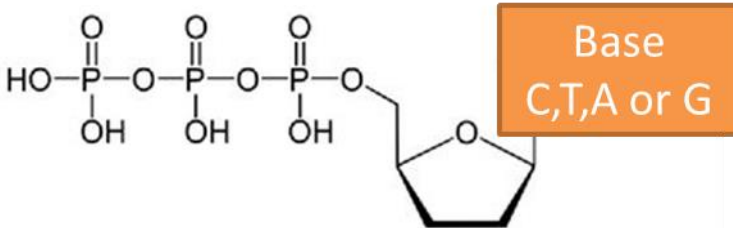
You should now notice a white precipitate forming. This is the DNA. You can scoop this out using a wooden stick or straw.

## **Gel electrophoresis at home**

It is possible to do simple electrophoresis on food dyes at home. The uses the same principles which are behind the DNA electrophoresis. To do this you will need a margarine/butter container or equivalent, some wire and 9V batteries to use as a power source. A detailed protocol with pictures can be found at

<https://www.instructables.com/id/Building-and-Running-a-Homemade-Agarose-Gel-Electr/>

## Glossary

Bacteriophage	Viruses that infect bacteria. Some of these result in the killing of the bacterial cell (lytic phages).
Base	A single nucleotide on one of the strands of the DNA sequence. A nucleotide is linked to the next nucleotide in the DNA sequence by a phosphate group.
Base Pair	Two nucleotides one on each strand of the DNA linked by hydrogen bonds. This is either adenine & thymine or cytosine and guanine within DNA.
Cas9	An enzyme that cuts DNA directed by a guide RNA. It is derived from bacteria where the guide RNAs are encoded in the CRISPR region.
Complementary	
Deoxynucleotide	<p>A nucleotide tri-phosphate which is the building block for DNA. This has a hydroxyl (OH) group on carbon 3 (3') of the sugar, circled in red.</p> 
Di-deoxynucleotide	<p>A nucleotide tri-phosphate which is lacking a hydroxyl (OH) group on carbon 3 (3') of the sugar. This means that it cannot have further bases attached to it, to extend the DNA chain.</p> 
Denature	In this context this means to split DNA into from double strand to single stranded DNA. This can be done by heat or chemical (eg: strong alkali) means.
DNA polymerase	An enzyme that copies DNA. Requires both a primer to tell it where to start and a DNA template to work from
DNA sequencing	A process by which the order of the bases in DNA is read.
Electrophoresis	Using a gel and an electrical circuit to separate a sample (in this case DNA) by size. In the case of DNA the negatively charged DNA moves towards the positive electrode with smaller pieces of DNA moving more quickly as they are impeded less by the gel.
Eukaryote	An organism containing cells which have a cell nucleus. This includes

	humans, plants and yeast but does not include bacteria.
Expressed	A DNA sequence that has been read and used to make protein
Exonuclease	An enzyme that breaks down DNA (or RNA) from one end. These work on a single strand of DNA (or RNA) and thus result in sticky ends. These enzymes play an important role in DNA repair and replication [ref]
Genome	The DNA in an organism that encodes for that organism. The human genome is made up of 46 chromosomes. Some bacterial genomes only have one chromosome.
Homology	Two pieces of DNA with the same sequence.
Marker	Something that labels your item of interest. In terms of a vector, this usually gives the cell containing it a unique property such as antibiotic resistance or the ability to produce a nutrient. This allows you to use this property to select or screen for cells containing the vector.
Nuclease	An enzyme that breaks cuts the backbone of DNA. Examples can include restriction endonucleases and endonucleases that only work on one strand.
Nucleotide triphosphate	The building block of DNA with three phosphates. Two of those phosphates are lost when this is used to form DNA.
PCR	Polymerase chain reaction, an artificial method of copying/amplifying a specific piece of DNA
Phage	See Bacteriophage
Plasmid	A circular piece of DNA found within cells. This can be used as a vector to carry a piece of DNA of interest.
Primer	A short piece of DNA (or RNA) typically 20-30base pairs in length
Restriction Endonuclease	An enzyme that cuts DNA at a specific site, termed a restriction site
Sonication	The use of ultrasound (high frequency sound waves) to break things apart.
SYBR Green / SYBR safe	Synthetic dyes which bind DNA and can be observed by illuminating the DNA with light of a specific wavelength
Thermostable	An enzyme that is able to work and that does not denature (fall apart) at high temperatures, typically this means above 37 degrees