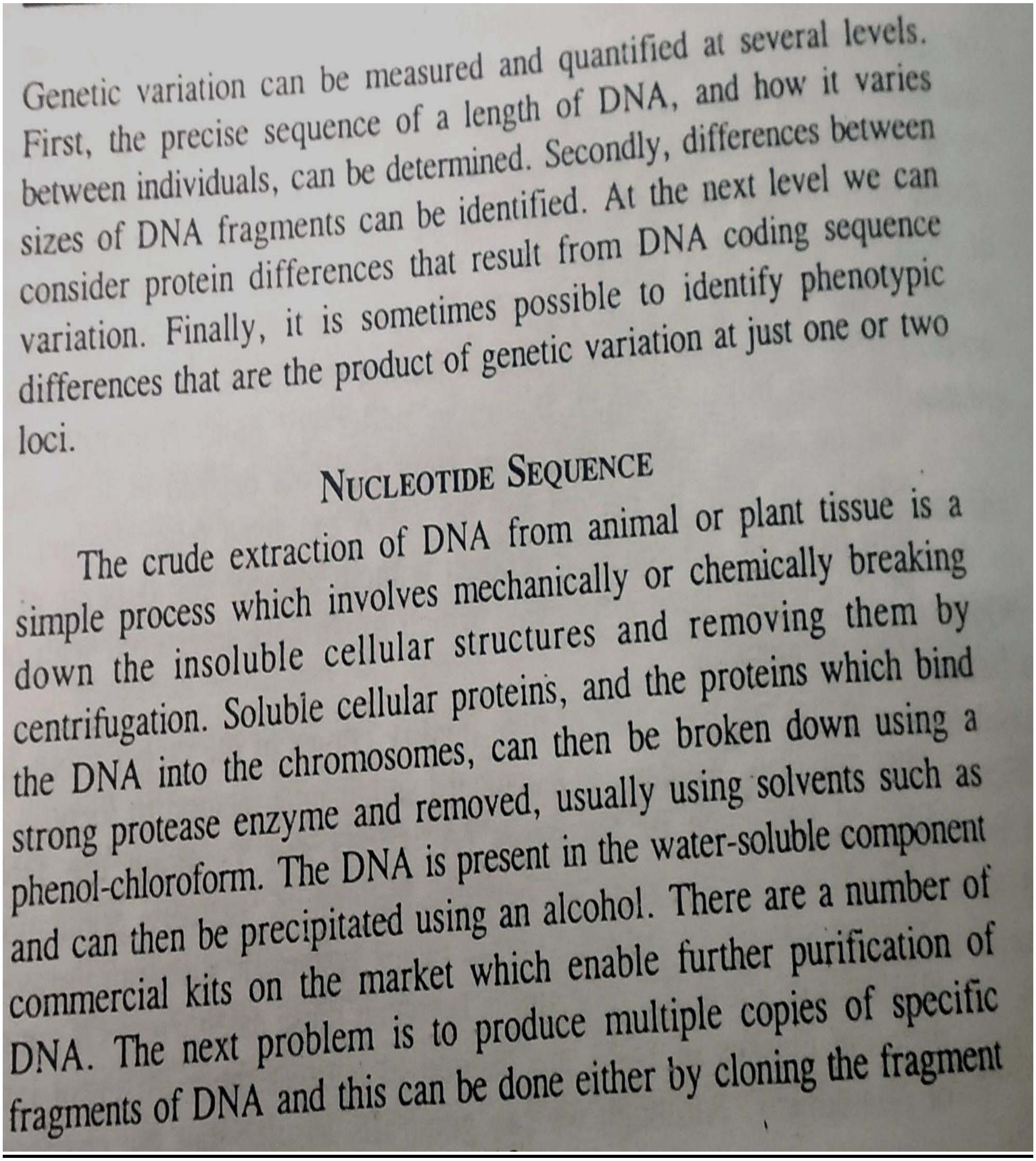
Measurement of Genetic Variation in Fish I



Genetic variation can be measured and quantified at several levels. First, the precise sequence of a length of DNA, and how it varies between individuals, can be determined. Secondly, differences between sizes of DNA fragments can be identified. At the next level we can consider protein differences that result from DNA coding sequence variation. Finally, it is sometimes possible to identify phenotypic differences that are the product of genetic variation at just one or two loci.

Nucleotide Sequence

The crude extraction of DNA from animal or plant tissue is a simple process which involves mechanically or chemically breaking down the insoluble cellular structures and removing them by centrifugation. Soluble cellular proteins, and the proteins which bind the DNA into the chromosomes, can then be broken down using a strong protease enzyme and removed, usually using solvents such as phenol-chloroform. The DNA is present in the water-soluble component and can then be precipitated using an alcohol. There are a number of commercial kits on the market which enable further purification of DNA. The next problem is to produce multiple copies of specific fragments of DNA and this can be done either by cloning the fragment or by the use of the polymerase chain reaction (PCR). In the process of cloning, the target DNA is inserted into a vector molecule which is taken up or inserted into host cells. Subsequent rapid replication of these host cells and the vector molecules inside them results in the production of millions of copies of the target DNA. As far as most DNA markers are concerned, cloning is usually only needed during the development phase—once the DNA sequences flanking the markers have been found from the cloned fragments, PCR can be used to produce millions of copies of the target sequence within a few hours. The PCR method relies on the fact that double-stranded DNA becomes denatured and separates into single strands when heated above 90°C. Once denatured, the temperature is lowered to a predetermined annealing temperature which allows short manufactured lengths of single-stranded DNA of known sequence (primers), designed to be complementary to the regions flanking the target DNA, to attach (anneal) to these flanking regions. Raising the temperature to 72°C in the presence of a DNA polymerase enzyme and the building blocks of DNA results in two copies of the double-stranded target DNA. Each time the cycle is repeated the number of copies is doubled and, since each cycle takes only a minute or two, millions of copies can be produced within a few hours by this method.

The sizes of pieces of DNA produced from cloning or PCR can be determined by subjecting the DNA to electrophoresis alongside known size-standards. Since electrophoresis separates fragments based on their sizes, it can be used to purify DNA fragments. For example, the results of a PCR reaction can be run (electrophoresed) on an agarose gel. The DNA is stained during or after electrophoresis with ethidium bromide which fluoresces under UV light. Hopefully there will be a nice bright band of the right size, our desired PCR product, which can then be cut out and the DNA extracted from the gel. We thus have the desired PCR product without leftover components of the PCR reaction, such as primers, which might have interfered with later DNA sequencing.

When we have lots of high-quality copies of the target DNA in a pure solution, we can use a standard sequencing method to identify the precise sequence of the bases (A, C, G and T) along the DNA. Comparison of the sequence between individuals, between populations, between species or between higher order systematic divisions, provides information about the relatedness between these categories. Of course, different classes of DNA are needed to address these different levels of relatedness. Although we can generally assume that the chances of [redacted] a point mutation occurring are the same anywhere along the DNA molecules that make up the genome of a particular species, the important question is what the consequences might be of such a point mutation.

Let us first consider a mutation within the coded part (exon) of a gene that codes for an enzyme. We might expect such DNA mutations to have important effects. However, the mutation could occur at the third base of a codon and, because of the redundancy of the genetic code, will be unlikely to change the amino acid coded for. Alternatively, it could change one of the amino acids in the enzyme produced, but even this may not have any effect on the ability of the enzyme to carry out its cellular biochemical function. Nevertheless, *some* mutations within the exon of an enzyme gene are bound to have a deleterious effect such that individuals carrying that mutation produce an ineffective enzyme and are less likely to survive. Exceptionally, a mutation might be advantageous and improve performance of an enzyme. So enzyme exon DNA sequences are free to change slowly over evolutionary time, at a rate that is considerably less than the rate of mutation, and the rate varies between different enzymes depending partly on the specificity of their biochemical task in the cell.

What about DNA sequences which form part of an intron? These sequences are not translated into a protein product and so we would expect changes to have neither deleterious nor advantageous effects. Mutations at non-coding sites are effectively neutral and therefore are likely to accumulate without constraint over evolutionary time.

Finally, let us consider sequences that code not for proteins, but for the very RNA molecules which are involved in the process of translation of the DNA code. Here, almost every letter of the code is critical to the functioning of the RNA product and almost any mutation will render it non-functional. The strongly deleterious effect on any individual subjected to such a mutation means that the rate of evolutionary change of these parts of the DNA molecule is extremely slow. Such DNA is said to be highly conserved.

It follows from the three examples above that some regions of DNA are valuable for identifying evolutionary changes far back in time, while others will detect more recent changes.

Production of Identical Fragments

It perhaps should come as no surprise to discover that DNA is a very tough molecule and can withstand considerable stresses during its extraction. However, for accurate DNA analysis, long, unbroken molecules are required, and care is required to reduce shearing of the molecules during preparation. Once long, high molecular weight DNA molecules have been extracted and purified they can be cut into fragments using restriction endonucleases (REs) that are enzymes purified from bacteria. One class (type II) of these enzymes have the useful property of only cutting the DNA molecule at particular points in the sequence, each enzyme having its own recognition sequence of four or more bases. For example, a restriction endonuclease isolated from the bacterium *Escherichia coli*, named *EcoRI*, cuts DNA only where the hexanucleotide 5'-GAATTC-3' occurs. The cut is uneven, producing an overlap on each end making them 'sticky' or 'cohesive'. Other restriction endonucleases, such as *Alul*, make blunt ended cuts.

TABLE 8.1. RECOGNITION SEQUENCES AND TYPE OF END SEQUENCE OF THREE COMMONLY USED RESTRICTION ENDONUCLEASES.

Restriction endonu--clease	*Recognition sequence*	*End sequences*		*Type of end*
EcoRI	5'-GAATTC-3'	5'-G	AATTC-3'	Sticky
	3'-CI'IAAG-5'	3'-C1'IAA	G-5'	
Alul	5'-AGCT-3'	5'-AG	CT-3'	Blunt
	3'-TCGA-5'	3'-TC	GA-5'	
Hinfl	5'-GANTC-3'	5'-G	ANTC-3'	Sticky
	3'-CTNAG-5'	3'-CrNA	G-5'	

G = guanine, A = adenine, C = cytosine, T = thymine, N = a nucleotide.

Once DNA has been cut into fragments, the fragments can 'pasted' into a vector using the enzyme DNA ligase. There are number of vectors available depending on such factors as the size the fragments to be cloned, the host organism (bacteria, yeast, plar mammals) and whether one wishes to express (i.e. transcribe translate) the genes on the cloned fragments. However, by far most common cloning system for purposes relevant to aquacultu where we tend to be probing for particular genes or marker sequen is to use a modified form of the bacterium *Escherishia coli* as host. There are two principle vectors used to get DNA into a bacteri one a virus (bacteriophage or just phage) that infects the bacter and the other a plasmid, which is a circular DNA molecule occur as a natural inclusion in many bacteria. Some labs still prefer ph because the infectious particles naturally contain extractable single-stranded DNA that they find gives good sequencing results. However, most labs prefer to be able to sequence the DNA in both directions (which can not be achieved with only one strand) and find that plasmid vectors are less likely to 'chew up' the inserted DNA. There are many variants of plasmid and a very common and well-behaved one is pUC19 ('p' for plasmid, 'UC' for the University of California, where the plasmid was created, '19' to show that it was the nineteenth such plasmid created there).

DNA is extracted from an organism and then cut by incubation of the DNA with a restriction enzyme. The cut fragments are then mixed in solution with the enzyme DNA ligase and the vector, in this case a plasmid, which has been previously cut with the same or a compatible restriction enzyme (plasmids and other vectors have been engineered to contain a polycloning site, which contains the recognition sequences for many different restriction enzymes). Many of the DNA fragments become ligated into plasmid molecules and the vector, plus its included DNA, is then inserted into a special form of *E. coli*. This 'competent' *E. coli* takes in the vector when subjected to a shock of some kind, usually heat.

The plasmid vector contains the gene sequence for resistance to an antibiotic (e.g. ampicillin, chloramphenicol). The *E. coli* used has no resistance of its own. The antibiotic is added to the agar plates so only bacterial clones that include the plasmid will grow on the plates.

The *E. coli* cells are spread very thinly over the agar plates so that each transformed cell can form a separate colony when allowed to replicate overnight at 37°C. As well as the bacterial multiplication, the plasmid replicates within each bacterial cell, thereby producing millions of copies of the included DNA in bacterial clones.

A second plasmid gene is employed to identify those colonies that contain non-recombinant plasmids, that is, bacteria which took up plasmids which had self-ligated and had no added, recombinant, DNA. The plasmid used has a gene for B-galactosidase, but the plasmid's cut site is in the middle of this gene. Therefore plasmids that have self-ligated will still have an active B-galactosidase gene, while plasmids that contain recombinant DNA will not. Using the substrate X-gal i the agar plates, which produces a blue product on reaction with B galactosidase, enables blue-coloured colonies (non-recombinant DNA and white colonies (recombinant DNA) to be identified.

If desired, white colonies containing recombinant DNA can individually picked from the plates using a sterile toothpick and maintained in a 'DNA library' of clones (such DNA libraries are commercially available for some species). However, most researchers probe for the required genomic DNA sequence on the original transformed colonies. This is done by carefully laying a nylon membrane onto the plate so that some of each colony is transferred to the membrane which is then carefully peeled off. While the membrane is on the plate their relative positions are marked, for example by puncturing both with a red-hot needle, so that they can be accurately lined up again later. The membranes are treated to break down the bacterial cell walls and to separate the strands of the DNA (denaturation). The DNA is then fixed to the membrane using heat or ultraviolet light. The membranes are then probed for the sequence of interest. This is done by hybridising the plasmid DNA with a labelled probe, a short sequence of single-stranded DNA complementary to the sequence of interest. Probes are generally radioactively labelled, though fluorescent labels are available. Hybridisation involves exposing the nylon membranes to the labelled probe at a temperature high enough to melt all but a very good DNA match. The probe DNA thus becomes annealed only to the target DNA, carrying its label with it. The radiolabel is visualised by autoradiography - exposure of the membranes to X-ray film. The needle holes on the membranes can be marked to show up on the film, so that the original agar plates with their re-grown bacterial colonies can be lined up with the autoradiograph and those clones which gave a positive radiolabel signal can be identified and isolated for sequencing or further analysis.

DNA Sequence and PCR

The PCR technique makes millions of copies of a particular target DNA sequence. The whole amplification process takes place in microtubes or in microwells in plastic plates in a small thermal-cycling machine on the bench. Each microtube contains a number of ingredients together with the template DNA that is to be copied from. Millions of copies of a pair of primers - short single-strand sequences of DNA each complementary to one end of the target DNA sequence - are included. A thermostable DNA polymerase enzyme (e.g. *Taq* polymerase, derived from the bacterium *Thermus quaticus*, a resident of hot springs) is present together with the four deoxynucleotide triphosphates (dATP, dCTP, dGTP and dTTP, collectively dNTPs) in a buffer. Using *Taq poly*merase, the maximum length of the target DNA is effectively around 3-4 kb, because longer fragments cannot be

[redacted]

successfully amplified and inaccuracies begin to accumulate to unacceptable levels. There are special DNA polymerases available for those who need long and accurate PCR replication.

There are three stages to PCR - denaturation, primer annealing and polymerization - each one lasting only about a minute and each operating at a different temperature:

The denaturation step: the contents of the microtubes are heated to above 90°C to separate the two strands of the template DNA.

The primer annealing step: the temperature is decreased rapidly to a predetermined annealing temperature, usually around 55°C, to allow the primers to 'sit down', that is, to become annealed to their complementary sequences on the template DNA.

The polymerisation step: The temperature is increased to 72°C, the temperature at which the *Taq* polymerase is most active, to enable the synthesis of new DNA in the 3' direction away from the primers.

These steps are then repeated:

The denaturation step (2): the mixture is again heated to about 94°C to denature all the newly built molecules, and any other parts of the template DNA which have become annealed by chance.

The primer annealing step (2): primers anneal as in the first cycle, but this time some will anneal to the newly manufactured strands of DNA.

The polymerisation step (2): new synthesis of molecules takes place and results in some molecules which have one strand of the precise length defined by the primer sequences at each end.

The denaturation step (3): the strands of DNA are again separated ready for the annealing step.

From this point on, the number of newly synthesised molecules of the precise length specified by the two primers increases exponentially at each new cycle. Usually 20 to 40 cycles are used and the resulting PCR product should consist of a very high copy number of the target sequence together with a small amount of original and fragmented DNA.

The above describes the theory. In practice, the PCR method is extremely sensitive to small variables. For each pair of primers, there is an optimum annealing temperature and optimum Mg^{2+} concentration. The slightest contamination of the template DNA with proteins or other material can often inhibit PCR amplification. *Taq* polymerase and buffers from different manufacturers have slightly different characteristics and may require re-optimisation. And, of course, while temperatures are changing between steps, all the ingredients are free to interact in the most unpredictable way. In spite of these considerations, the PCR method has become routine in the laboratory and provides a simple and effective means of producing high copy numbers of specific DNA sequences.

Electrophoresis

Electrophoresis is used to separate molecules by size. It works on the principle that charged molecules, such as proteins or DNA, will be drawn through a slab of gel when a current is passed across it. A number of different gel types can be used, such as hydrolysed starch, cellulose acetate, agarose or polyacrylamide. Polyacrylamide gels are normally oriented vertically while other gel types are usually positioned horizontally. Polyacrylamide gel electrophoresis is sometimes known by the acronym PAGE. In vertical polyacrylamide gels, samples are placed at the top of the gel and ate separated from one another by a comb-like structure, or by spacers. In horizontal gel systems, samples are inserted into slots in the gel close to, or at, one end. An example of a horizontal starch gel apparatus.

The strength of gels can be adjusted to make the pore size similar to the size of the molecules being separated so that some sieving effect can take place in addition to the electrical charge dragging the molecules through the gel. Because passing an electrical current through water changes the pH, the solution used to make electrical connection with the gel is always buffered.

Once the current has been run for sufficient time to separate the fastermigrating from the slower-migrating molecules, electrophoresis is stopped and the gels are prepared for visualisation of the resulting bands of protein or DNA. For proteins a general non-specific protein stain can be used, though in the case of enzymes the positions of the different bands (allozymes) on the gel are identified using substrate-specific stains. High concentrations of DNA are usually stained with ethidium bromide, which fluoresces under UV light, but lower concentrations require the more sensitive silver staining method. For very small quantities of DNA the most sensitive staining method is to use radiolabelling. In radiolabelling, a radioactive isotope of an element, such as sulphur-35 (^{35}S) or potassium-32 (^{32}P), is incorporated into the DNA before electrophoresis, then the gel is dried and placed adjacent to a sheet of film (the autoradiograph negative) and the radioactive decay of the element exposes the negative at the point of the signal.

Automated DNA analysis machines use chemo-luminescent stains that an be read by a laser. This removes the risks of working with adioisotopes and, by virtue of different-coloured stains, enables more DNA sequences to be obtained from a single gel. DNA can be adiolabelled after electrophoresis, but this requires it to be transferred rom the fragile gel to a more robust membrane by a technique known s Southern blotting before hybridisation with single-stranded DNA omplementary to the sequence of interest.

Various standards can be run on gels alongside samples for comparison. Dyes and samples from individuals of known genotypes are run on protein and enzyme gels, while DNA bands of known sizes in base pairs, bp, or kilobases, kb) are used as molecular size standards n DNA electrophoresis. The sizes of DNA fragments run on sequencing gels are found by comparison with a known sequence that is run alongside.

Sequencing DNA

Sequencing of DNA is now a highly automated and, in some cases, roboticised procedure. All well-provisioned large genetic laboratories will have their own sequencers and there are a number of commercial companies that provide a relatively cheap sequencing service to nstitutions such as marine stations or aquaculture institutions where genetic study is usually only a small part of their activities.

DNA sequencing uses DNA polymerase enzymes, such as those used in PCR, to copy the DNA strand but with two added twists. The first is that one of the dNTPs are fluorescently or radiolabelled, so that the copies can be visualised. The second trick is that we sabotage the copying process. We do this by introducing a small proportion of dideoxynucleotides (ddNTPs) along with the dNTPs. Like dNTPs, the polymerase joins ddNTPs to the new DNA strand, but unlike dNTPs they lack the bond which would enable another dNTP to be joined after them, so they stop the copying process. Sequencing one piece of DNA involves carrying out four separate reactions for Adenine, Cytosine, Guanine and Thymine using ddATP, ddCTP, ddGTP and ddTTP respectively. The proportion of ddNTP to dNTP is balanced so that copy strands are produced of many different lengths, from those that only extend a few bases from the sequencing primer to strand hundreds of bases long. But each will end in a ddNTP. So when w run the four reactions out on a sequencing gel, the ddATP reactio will produce bands of many lengths, but we will know that each bar shows the length of a DNA fragment which ends with the nucleoti

Adenine. Imagine that the sequence has Adenine occurring at the 2nd, 5th, 6th, 9th, 12th, 13th, etc. positions after a 20-base sequencing primer. In that case the A series will contain molecules of 22, 25, 26, 29, 32, 33, etc. bases long. Similarly, the ddCTP, ddGTP and ddTTP reactions will consist of molecules of lengths specific to the positions of the bases Cytosine, Guanine and Thymine, respectively, along the DNA. These four series of molecules are run in four lanes, side by side, down a high resolution polyacrylamide gel. The sequence of the DNA then can be read from these four ACGT lanes from the bottom of the gel upwards.

Sequence data have now been obtained from a great range of organisms and this information is collected together in DNA databases such as the one at EMBL in Europe and GenBank in the USA. Scientists have free access to these databases and powerful computer programs are available to analyse new sequences and to compare them with all other available sequences on the databases. This field of bioinformatics is rapidly expanding.

FRAGMENT SIZE VARIATION

At the beginning of this chapter we said that genetic variation can be measured and quantified at several levels. We have shown how we can determine the precise sequence of a length of DNA, and how it varies between individuals. Now we shall progress to see how differences between sizes of DNA fragments can be identified and used to address particular genetic questions. Techniques that fall into this category include those known by the acronyms RFLP, VNTR, DNA fingerprinting, RAPD and AFLP. Of these, VNTR markers (microsatellites in particular) have come to the fore in recent years as being the most generally useful, though the others all have their place in answering particular genetic questions.

Restriction Fragment Length Polymorphisms (RFLPs)

We can make good use of fragments of DNA as genetic markers without going through the procedure of sequencing them. If we have a high copy number of a particular fragment produced by the cloning method or from the PCR machine, this can be incubated with a number of different restriction endonucleases (REs) which will cleave it into a number of lengths depending on the position of the RE recognition sites. The various lengths produced can be separated by size and stained on an agarose gel. The same piece of DNA from different individuals will produce different sets of restricted fragments if there have been point mutations affecting the RE recognition sequences. In this way,