Measurement of Genetic Variation in Fish II

polymorphisms can be identified based on the pattern of the size fragments on the agarose gel. RFLP analysis is particularly useful for mitochondrial DNA.

The fact that restriction enzymes will only cut DNA at specific sequences presents us with a simple way of identifying genetic variation caused by point mutations. Let's say we have a 2 kb length of DNA from an individual animal which can be amplified to a high copy number and we then incubate this amplified DNA in a microtube with a suite of restriction enzymes. The restriction enzymes will cut the DNA into a number of fragments that can then be easily sizeseparated on agarose gel and stained with ethidium bromide. In other individuals of the same species, point mutations will have altered the sequence at one or more restriction enzyme cut sites, or may have produced a cut site where one was not present before. This will result in different individuals producing variation in the size and number of fragments when their DNA is incubated with this suite of restriction enzymes. Genetic variation identified in this way is called restriction fragment length polymorphism (RFLP).

RFLP data from a sample of a population can be analysed in two ways. First, all the different fragment patterns detected on the gels are counted and the frequencies of each determined. These RFLP frequency data can then be compared between populations. Secondly, RFLP data can be analysed on the basis of the proportion of nucleotides that differ between individuals. Of course, the number of nucleotides actually sampled is limited by the number of restriction enzymes used and the number of bases each enzyme has in its cut site. Nevertheless, such data are of value in establishing relationships between populations, species or higher taxa.

Tandem Repeats

Variation in the sequence of DNA can occur at certain sites by a method which is not point mutation. Spread throughout the genome are regions called variable number tandem repeats (VNTR), also known as simple tandem repeats (STR) or simple sequence length polymorphisms (SSLPs), which contain tandem (i.e. linked in chains) repeats of DNA sequences. The sequences may be very short (from 1 to 10 bp) or much longer, but the key feature of these tandem repeats is that the number of repeats can vary between individuals. It is thought that increases or decreases in the number of the repeats occur during copying by recombination or replication slippage and that these processes are not only independent of point mutations, but also occur

at a much faster rates. Variation in the number of repeats at these satellite (repeated units 100 to 5000 bp), minisatellite (repeated units 5 to 100 bp) or microsatellite (repeated units 2 to 4 bp) loci can be very extensive in populations and provides a valuable tool for investigation of population genetic changes in the recent past. Microsatellite markers in particular are now used extensively for a number of reasons: because they are co-dominant (both alleles can be identified) and therefore can be analysed under the standard Hardy-Weinberg model; because, as 'junk DNA they can usually be considered to be free of selective pressures; because of the high number of both loci and alleles at each locus; and, not least, because automatic DNA sequencers can be used for automated genotyping at m.icrosatellite loci, vastly increasing the rate at which samples can be processed.

Isolation of Mitochondrial DNA

It is possible to separate mtDNA from the nuclear DNA by differential centrifugation. A buffered chemical solution is used to break up (lyse) the cells. The resulting cell lysate is then centrifuged at a speed that is high enough to sediment heavier material such as the nucleus and larger cell debris. The supernatant, which contains the mitochondria and other cell organelles, is removed and centrifuged again at a higher speed to sediment the mitochondria. Further purification can be achieved by density gradient centrifugation where material is centrifuged through a series of layered density gradients. Once separated, the mtDNA can be extracted from the mitochondria using the standard phenolchloroform extraction used for nuclear DNA.

Because mtDNA is a molecule of fixed length, and also because it is present in high copy number in cells, it is amenable to analysis without further amplification or preparation. Extracted mtDNA can be cut directly with restriction enzymes and the resulting fragments can be separated on an agarose gel and stained with ethidium bromide. Genetic variation between individuals is detected as sequence differences or as RFLPs . The pattern of mtDNA restriction fragments from an individual is called its haplotype and the frequencies of particular haplotypes in a population are used to determine differences between populations. The degree to which mutational changes have separated different haplotypes - the nucleotide divergence - can also be quantified and used for population genetic or systematic purposes.

Microsatellites

The genomes of animals and plants contain regions that consist of a series of repeated units of DNA - VNTR. One type of VN[illegible] -

microsatellites - consist of dinucleotide (e.g. CACACACA), trinucleotide (e.g. GTAGTAGTAGTA) or tetranucleotide (e.g. TAGCTAGCTAG CTAGC) repeats. A microsatellite sequence identified in the DNA of the common cockle *(Cerastoderma edule)*.

The number of repeated units contained within a particular microsatellite locus can vary within a population, and this produces variation in the length of the locus. This variation can be detected by amplifying the locus using PCR, followed by electrophoresis.

Isolation and identification of microsatellites in a species is done by first producing a library of recombinant clones containing fragments of DNA between 300 and 900 by in length. DNA is extracted from an individual of the species, cut with Ref and run out on an agarose gel against a size standard. Fragments of a size between 300 and 900 by are then extracted from the gel and these are used to make a clone library. This library is then screened using complementary repeat probes, for example (GT)n to identify (CA) repeat microsatellites. Insert DNA from positive clones is sequenced to confirm the existence of a microsatellite within the fragment of DNA and to determine the flanking sequences. Primers are designed' based on the flanking sequences and optimised for PCR. The microsatellite locus is then PCR-amplified from template DNA extracted from individual organisms. The PCR products are run on a high quality polyacrylamide gel which enables the detection of fragments that differ by a single base pair in length. Scoring of genotypes is by radiolabelling, silver staining or laser detection in automated sequencers. However, there is often an effect—called 'stutter'—where slippage errors in replication during PCR produce fragments with one or two more, or one or two fewer repeats. This is most commonly observed with dinucleotide microsatellites and a little care is required in interpretation Microsatellites are co-dominant, therefore both homozygous and heterozygous genotypes can be detected and microsatellite genotype and allele data can be analysed using the Hardy-Weinberg model.

DNA fingerprinting

DNA fingerprinting can be thought of as a combination of RFLP and VNTR. First, genomic DNA is cut with a particular suite of restriction enzymes and differently sized fragments are separated by electrophoresis. The DNA is transferred from the fragile gel to a nylon membrane by the technique known as Southern blotting and the membrane is then probed with a particular satellite repeat sequence which is common throughout the genome. Fragments that contain the

peat show up as a number of discrete bands after autoradiography. hese banding patterns are so variable as to be in practice unique to ch individual (the chances of a match between unrelated individuals e millions to one). Since the bands are inherited in a predictable shion, DNA fingerprinting is a very accurate way of determining rentage and the forensic uses of this method are now well known.

ndom Amplified Polymorphic DNA (RAPD)

The RAPD method is based on the principle that the shorter the gth of the primers which are used in PCR, the greater is the nce that non-target sequences will be amplified. Using a single 10- oligonucleotide as the sole primer, PCR is conducted on raw A and the resulting fragments, which come from annealing of the ners all across the genome, are separated on agarose gel. Variations veen individuals in the presence or absence of bands, reflect ational difference at the primer sites. RAPDs suffer from the ortant criticism (among others) that they are not entirely reliable repeatable.

Primers designed for use with PCR are usually 20-25 by in length nsure that they will be specific to the particular DNA sequence targeted. However, if a much shorter primer is used, say of 10 ngth, it is likely to anneal to many regions of the genome during nnealing step of PCR. If, by chance, the primer anneals to opposite ds of the DNA within a region up to about 3-4 kb in length, then R product spanning the annealing sites will be produced. Some RAPD primer sequences will produce no PCR product, while may produce a number of different size fragments. The PCR cts are run on agarose gel electrophoresis to identify any fragments ling to their size.

ow does this identify genetic polymorphisms? Consider the on where a point mutation is present in the sequence at one of of RAPD primer sites in some individuals in a population. f the DNA from these individuals will not produce the fragment in other individuals and therefore that band on the gel will be Thus, presence or absence of particular bands can be scored kers of genetic variation. This information of presence (+) or (–) of bands is called 'dominant' data. It is not possible to ish individuals that are homozygous (+/+) from those which rozygous (+/–) because both genotypes will produce a band on Allele frequencies, however, can be calculated based on the ion that the locus is in agreement with the Hardy-Weinberg

(H-W) model. The proportion of individuals scored for absence of the band (that is, -/- homozygotes) should equal the square of the (-) allele frequency.

When first developed, the RAPD technique seemed be a relatively quick and cheap method to obtain molecular genetic data that required no prior knowledge of DNA sequences. However, agreement with the H-W model has to be assumed and may not hold. Also, considerable care has to be taken to ensure that RAPD data are repeatable. Banding patterns are easily altered by subtle changes in the PCR conditions and some bands have been found which violate Mendelian inheritance. There is also the problem that it is difficult to identify bands that have been amplified from contaminating DNA (from parasites, gut contents, epibiota, etc.). For these reasons, the RAPD technique has fallen out of favour with many geneticists - though it is still widely used by those who work on clonal organisms as a way to identify particular clones.

Amplified Fragment Length Polymorphism (AFLP)

A method which amplifies randomly selected fragments of the genome much more reliably than RAPD is the AFLP method. AFLP is almost an inverse form of RFLP - the genomic DNA is cut into fragments with restriction enzymes and then just a few of those fragments are selectively amplified using special radiolabelled PCR primers. The resulting fragments are then visualised by separation on a polyacrylamide gel followed by autoradiography. The products of AFLP - a series of bands of different sizes - are similar to the products of RAPDs and variation between individuals is based on the presence or absence of bands. Variation is the result of point mutation differences in the DNA sequence at the PCR primer sites.

Like RAPDs, the AFLP method provides size fragment-markers from DNA of unknown sequence. Unlike RAPDs, AFLP markers are extremely reliable and reproducible. DNA extracted from an organism is first cut using a pair of restriction enzymes which leave cohesive ends - one a frequent (four-base) cutter, the other a rare (six-base) cutter. The majority of fragments produced will have two four-base-cut ends, the minority will have two six-base-cut ends and the remainder will have one end with a six-base cut, and the other with a four-base cut. Two oligonucleotide 'adapters' designed to attach to the cohesive cut ends of both the six-base and four-base fragment terminals are then ligated to the resulting fragments. These adapters are used as the basis for polymerase chain reaction (PCR) primers that are also designed

to overlap with the DNA fragment by one, two or three specified bases. The primer for the less frequent (six-base) cut adapter is radiolabelled before PCR amplification and polyacrylamide gel electrophoresis, so only those fragments which were cut by the six-base RE are visualised by autoradiography. By varying the number of specified bases on the primers the number of fragments amplified can be controlled. Typically 50-100 restriction fragments can be identified and separated in a single run of the AFLP method.

Variation Due to Building Blocks

So far we have considered genetic variation at the level of the DNA. However, DNA sequence variation, when transcribed, can give rise to differences in the resulting proteins. It is at this level that genetic variation begins to interact with the environment to affect the survivorship and reproduction of organisms and their genes.

Genetic variation at the level of proteins can be identified and quantified using electrophoresis to separate the different protein products of alleles followed by staining to visualise these protein products. It is possible to stain the gel to display all proteins, but it is more useful to take advantage of the substrate-specific catalytic abilities of the class of proteins known as enzymes. This involves using the specific substrate of an enzyme in a stain overlaid on the gel that will change colour where the substrate is altered by the enzyme. The position of any enzyme variants can therefore be located on the gel. These genetic variants of enzymes are known as allozymes and methods for detecting allozyme variation were first developed in the 1960s. Although nowadays regarded by some as an outdated method, the extensive allozyme data sets produced in the last 40 years of the twentieth century fundamentally shifted the ground upon which geneticists tread. From the practical point of view, allozymes enabled us to look at the genetics of natural populations of a whole range of aquatic organisms in a way that was never possible before. This is not to say that allozymes are of only historical interest - they are still a useful tool in answering many genetic questions, particularly given the equipment, time and money required to develop and use DNA techniques.

Although little used in fisheries or aquaculture research, the reader should be made aware of the technique of immunological testing which assesses the relationship between proteins on the basis of the relative strength of the antigen-antibody reaction that they will produce.

A comparison of the values of various molecular techniques to address different problems in fisheries and aquaculture. The high value

of VNTR methods in parentage and within species analysis is clear, but the extent to which allozyme data can be informative is also emphasised. Although RAPDs score quite highly for a number of approaches, there are real problems of repeatability and reliability with this method. AFLPs are now the recommended quick screening method for identifying quantitative trait loci.

Allozymes

Allozymes are the products of genetic variation at enzyme-encoding loci. They should not be confused with isozymes that are alternative forms of an enzyme produced at different loci. In order to identify and score allozymes, a small piece of tissue is obtained from an organism and is ground up with a buffer solution in a microtube. This grinding releases the soluble proteins from the cells. After centrifugation, these proteins present in the supernatant are subjected to starch or cellulose acetate gel electrophoresis that separates the proteins on the basis of charge and size. This technique has been called the 'find 'em and grind 'em' method. Following staining for a particular enzyme, individual samples are classified according to the alleles present at the enzyme locus being investigated. Alleles are best identified by their mobility relative to a known standard - i.e. a band that moves 90% of the distance of the standard is called allele 90 - but they are sometimes just labelled numerically or alphabetically.

Interpretation of stained gels

Where an individual is a homozygote for an allozyme variant a single stained band will be seen on the gel, while heterozygotes, which contain two different variants, will exhibit two bands on the gel. Unfortunately this simple expectation is complicated by the quaternary structure of the protein. Why heterozygotes for a dimeric enzyme exhibit three bands (two in the same positions as the homozygote bands + a hybrid band in between) and why heterozygotes for a tetrameric enzyme produce five bands (two in the same position as homozygotes + three hybrid bands) on a gel after electrophoresis.

Often, a particular enzyme will be coded for at more than one locus. This is common, for example, in salmonids where there has been a relatively recent evolutionary tetraploidisation event. In such cases, of course, the products of all the loci will stain on the gel. This can lead to difficulty in interpretation if the ranges of migration of the products of the loci overlap. The situation can get very tricky when there are duplicate loci for tetrameric enzymes whose products overlap. Note that additional hybrid bands will form between allelic

products of both loci making a very complicated banding pattern produced in an individual that is heterozygous at both loci. Sometimes the problem can be alleviated by using particular tissues because some loci are tissue specific. For example, there are several lactate dehydrogenase (Ldh) loci in the trout, but the Ldh-5 locus is only expressed in eye tissue. Therefore, you can just pop out a trout's eye and grind it up to score for Ldh-5.

Protein structure	Genotype	Protein subunits in the cells	Final proteins in the cell	Phenotype on gel
Monomeric	A	a	a	
	B	b	b	
Dimeric	A	a	aa	
			ab	
	B	b	bb	
Tetrameric	A	a	aaaa	
			aaab	
			aabb	
			abbb	
			bbbbb	

Figure 8.1. The effect of the final (quaternary) structure of proteins on the banding patterns of heterozygotes on electrophoretic gels.

Immunological detection of proteins

Vertebrates have fairly-well understood systems in their blood for fighting and removing non-self or foreign proteins such as those in viruses and bacteria. Each protein has sites on its surface that are called antigenic determinants. When a non-self protein is detected the blood produces antibodies (immunoglobulin molecules) specific to the antigenic determinants of that protein. However, changes in the DNA coding for a protein can change the amino acid sequence in the protein and this, in turn, can change the nature of the antigenic determinants. This feature of proteins can be used for a number of purposes: to assess relationships between proteins extracted from different individuals or different taxa; to test for the presence of specific disease proteins in an animal; to identify foreign proteins coded for by transgenic DNA in genetically modified organisms.

It is rather complicated, but the method works as follows. Let us assume we are dealing with a comparison of a protein between two related species, A and B. First, a protein from an individual of species A is injected into the blood stream of a rabbit. The rabbit's blood will respond to this foreign protein by forming antibodies specific to the protein. If a sample of the rabbit's blood serum (blood cells removed), called antiserum, is then mixed in a suitable medium with the original protein from species A, a reaction will take place; the antibodies will link to the antigenic determinant sites on the protein to form aggregations which will precipitate from solution. If the antiserum is mixed with protein from species B, the amount of aggregation will depend on how many of the antigenic determinant sites are common between the two species. If they are very similar, there will be much aggregation; if they are very different, there will be very little aggregation. So the principle is that the amount of aggregation is proportional to the difference between the two proteins.

How then is the amount of aggregation measured? It is done indirectly using a method known as complement fixation. In serum there are a group of proteins called collectively 'complement' which bind on to antigen - antibody complexes and the more of these complexes there are, the more complement is bound up or 'fixed'. In the test, a known amount of complement is added to the antigen - antibody mixture and after a specified period of time, the amount of unfixed complement is assessed. Free, unfixed complement has the ability to lyse specially sensitised sheep red blood cells allowing the release of haemoglobin. Following . this reaction, remaining blood cells are removed by centrifugation and the colour change produced by the released haemoglobin is measured in a spectrophotometer. So the colour change is inversely proportional to the amount of antigen-antibody aggregation strong colour means there has been a weak reaction between the antigen and the antibody.

Phenotypic Variation

There are very few examples of easily identifiable phenotypi variation in aquatic organisms controlled by single genes, or eve pairs of genes. The best examples are found in the colouring ornamental fish. It is interesting to consider just how lucky Greg Mendel (the nineteenth-century discoverer of the method of gene inheritance) was to have chosen to work with peas, which had a numb of easily identifiable characters (round or wrinkled seeds; tall or sh

TABLE 8.2 COMPARISON OF THE VALUE OF GENETIC TECHNIQUES IN ADDRESSING DIFFERENT TYPES OF GENETIC PROBLEMS IN AQUACULTURE AND FISHERIES.

	DNA sequencing	RFLP	VNTR	RAPD	AFLP	Alloz-ymes
Pedigree or parentage analysis	+	+	+ + +	+	+	+ +
Populations within species	+	+ + +	+ + +	+ +	+ +	+ + +
Genus and species level relationships	+ + +	+ +	+	+ +	+ +	+ + +
Quantitative trait loci	—	—	+ +	+ + +	+ + +	+
Cost	H	M	H	M	H	L/M
Tissue requirements	V	M	L	L	L	M
Codominant or dominant data	—	V	C	D	D	C
Neutral	V	V	Y	V	V	N

+ + + = highly informative, + + = informative, + = marginally informative, or constrained in some other way, - = not appropriate. H = high, M = moderate, L = low, V = variable, C = codominant data, D = dominant data, Y = yes, N = no.

plants) each controlled by single genes. Such easily identifiable single-gene phenotypes are rare in most organisms. Because of the extensive development of DNA and allozyme technologies, the search for single-gene phenotypic variation is now uncommon. However, it is important to realise that the visual identification of varieties can be of critical importance to fish farmers without access to a modern genetic laboratory. Phenotypic rarities can provide high-value niche markets, and understanding how to get them to breed true requires genetic knowledge.