APPLICATION OF DNA FINGERPRINTING IN GENOME ANALYSIS OF FISHES

P. JAYASANKAR
Central Marine Fisheries Research Institute,
PO Box 1603, Tatapuram, Kochi - 682 014, Kerala, India

Alec Jeffreys in 1985 coined the term 'Fingerprinting' to describe the pattern of highly variable, individual-specific profiles that he and his colleagues saw in Southern blots of human genomic DNA hybridized to a tandemly arrayed repetitive DNA sequence. Since then there has been continuing effort for technological innovation in genetic typing methodology. This technique finds application in diversified fields such as tumour biology, genome mapping and ecological population and evolutionary biology.

DNA-level markers find several applications in fisheries research analysis and aquaculture. Species identification, evaluation of phylogeny, delineation of stock structure, measuring levels of genetic variation in wild and cultured stocks, conservation genetics, determination of breeding strategies, gene mapping, marker-assisted selection and assessment of genome manipulation techniques are among various uses of DNA markers in studies of natural and cultured fish populations. Relevant reviews include Wright (1983), Waldman and Wirgin (1993), Carvalho and Pitcher (1995), Fergusson et al. (1995) and O'Reilly and Wright (1995). Parker et al. (1998) provides a general review on molecular techniques in ecology. These studies highlight relative merits of DNA polymorphisms over the other conventional markers such as the requirement of only a small amount of tissue which could be ethanol-preserved/frozen for DNA extraction and availability of innumerable potential markers.

CATEGORIES OF DNA

Comparative examination of DNA sequences across divergent taxa explicitly shows that there are particular classes of sequences common to many organisms. DNA can be classified based on its function, structure, location, etc. However, many of these characters may overlap (Park and Moran, 1995).

Coding and non-coding DNA: Diploid organisms have two copies of each genetic region (locus) on homologous pairs of chromosomes. These two copies are called alleles, regardless of whether they represent coding or noncoding regions of the genome. Coding regions (exons) are often interspersed with more variable noncoding regions (introns or intergenic regions). Of the billions of nucleotides which constitute the total genomic DNA of eukaryotic organisms, only 1% regulates or codes for essential proteins.
Mendelian fashion often maternally. In many organisms, the genome. These classes of DNA may have repetitive DNA (long and short interspersed respectively) also occur multiple times (sometimes hundreds of thousands of times) throughout the genome. The mt genome contains over 30 genes, it is treated as a single locus in population genetic analyses, because of the absence of recombination in the mtDNA molecule.

Nuclear DNA: The nuclear genome in bony fishes is about 0.3-0.4 million base pairs in size (Ohno, 1974) and represents a wealth of genetic information that researchers in fish population genetics have only begun to exploit. Today many researchers are attempting to look at the sequence variation in the nuclear genome using various strategies: examining introns, pinpointing specific genes, looking at repetitive sequences, etc. Variation at the DNA level can be generalized into two categories: base substitution (such as point mutation) and insertion/deletions. A substitution from a purine to purine or from a pyrimidine to pyrimidine is known as a transversion. Among closely related organisms transitions occur more frequently than transversions. The actual insertion/deletion can be a single nucleotide, or it can be hundreds or thousands of nucleotides long. A common type of insertion/deletion mutation is copy number variation resulting from variation in the copies of a basic unit, or core sequence. VNTRs are typical examples of this type of mutation.

Non-repetitive and repetitive DNAs: About 70% of the eukaryotic genome consists of non-repetitive DNA (also referred to as 'single copy' or Scn DNA) whose sequence is present only in the haploid genome; the remainder are linked together as consecutive tandem repeats and the repetitive sites, due to their abundance of repeats and their slightly different base pair composition compared to bulk genomic DNA, form a separate 'satellite' band distinct from the main genomic band in Csl density gradient centrifugation. They are also known as Variable Number Tandem Repeats (VNTR). Large repetitive units of satellite DNA are often associated with heterochromatin near the centromere. Smaller regions occur throughout the genome. These classes of DNA may have 10-100 bp repeat units (minisatellites) or only 1-4 bp repeat units (microsatellites), interspersed repeated DNA (long and short interspersed repetitive elements, LINES and SINES, respectively) also occur multiple times (sometimes hundreds of thousands of times) throughout the genome, but constitute a smaller part of the genome than satellite DNAs. In terms of function, satellite DNA is considered to be non-coding one.

Mitochondrial DNA (mtDNA): Mitochondrial DNA (mtDNA) is haploid and predominantly inherited in a non-Mendelian fashion often maternally. It is 16-20 kb in size and circular. In many organisms, the mtDNA seems to accumulate mutations more rapidly than do ScnDNA, making them markers with greater variability and sensitivity to drift, and therefore more likely to show differences between populations/species. Certain marine species appear to have less variable mt genomes than freshwater or terrestrial species (Ovenden, 1990). The mitochondrial genome also includes a small non-coding region known as the Displacement loop (D-Loop), which serves as the origin or replication for the mitochondrial genome. In most animals, the D-Loop is much more variable than the rest of the mt genome and is therefore a very useful marker for the study of very recently divergent populations or species. Although the mt genome contains over 30 genes, it is treated as a single locus in population genetic analyses, because of the absence of recombination in the mtDNA molecule.

Extraction of DNA: DNA can be extracted from tissue samples that are fresh, frozen, dried, or stored in alcohol or buffers. Non-invasive DNA extraction procedures can be done from tissues, such as fin tips, scales, etc. to ensure that the specimen is not sacrificed. The extraction procedure begins with mechanical pulverization to separate cells and destroy cell membranes, while leaving the nucleus intact (cell cultures and tissues such as blood do not require mechanical pulverization). Tissue is then immersed in a solution containing a detergent that lyses the nuclear membrane, as well as a proteinase that denatures other proteins, especially nucleic acids intact. Proteins are separated from nucleic acids by extraction with organic compounds (usually phenol and chloroform), and the DNA is purified from the reagents in the extraction buffer by alcohol precipitation or dialysis. Separation of mtDNA from genomic DNA is a more extensive procedure involving layering of the digested tissue solution on a cesium chloride gradient which is subjected to high-speed centrifugation. Detailed extraction protocols have been developed (e.g. Sambrook et al., 1989). Most extraction procedures are completed in small volume (1.5 ml) microfuge tubes, facilitating population biological research that requires samples from large numbers of individuals. 

Digestion of DNA with restriction enzymes: Restriction enzymes cleave ('restrict') DNA at specific nucleotide recognition sites and generate DNA fragments that differ in size when mutations have created or destroyed restriction sites. This size variation is seen in markers known as "restriction fragment length polymorphisms" or RFLPs. A large number, more than 100, of enzymes are commercially available. Each recognizes and restricts a unique 4-, 5-, 6- or 8-base sequence. The smaller the recognition sequence of the enzyme, the smaller average size of the DNA fragments produced and the greater the number of different fragments generated by digestion. The choice of the restriction enzyme for digesting a genome depends on the size and number of DNA fragments desired, followed by trial and error to find the best enzyme for a particular system.

Enzymatic digestion of small molecules, such as mtDNA results in a small enough number of fragments that can be resolved by electrophoresis on an agarose gel. Allele frequencies can then be quantified by the presence or absence of restriction sites among individuals. However, large molecules, such as nDNA, once cleaved by restriction enzymes, would yield only a smear on agarose gel. In such a case, fragments of particular interest may be identified after electrophoresis by hybridization with a genetic probe (a fragment of DNA containing the sequence of interest). The presence of sequences complementary to the probe DNA is detected using Southern blot hybridization and autoradiography.

Hybridization of DNA molecules: By heating, the hydrogen bonds between the strands of DNA are broken (denaturation) and as DNA
coids, the complementary strands join (annealing). Under experimental conditions, hybrid DNA molecules can thus be formed between the two strands of DNA from two different individuals that have some fraction of unmatched base pairs; annealing of two non-identical strands is called hybridization. The stability of the association depends on factors like percent of mismatch, length of sequence, or temperature. Researchers can manipulate experimental conditions to promote or prevent hybridization of DNA molecules depending on the degree or complementarity. This has important implications for many of the methods used to study DNA. Southern blotting, PCR and DNA sequencing are all methods that depend upon DNA hybridization/annealing.

Southern blotting: This is a common method for identifying polymorphic fragments of DNA that differ in size because of the gain or loss of restriction sites. Usually, a radioactively labelled, single-stranded DNA “probe” of a known genetic region is used to identify the allelic fragments. First, DNA fragments are cleaved with a restriction enzyme, separated on an agarose gel, and treated with sodium hydroxide to separate complementary strands, making the DNA single-stranded. The DNA is then transferred from the gel by blotting onto a nitrocellulose/nylon filter, and is permanently fixed to the filter by baking or exposure to UV radiation. The filter can then be hybridized with a specific fragment of single-stranded DNA (the probe) that has been labelled either radioactively or by a nonradioactive labelling method. Hybridization allows one to determine the size of the fragments (based on their position after electrophoresis) that carry sequences complementary to those in the probe. The specific region of annealing appears as a "band" on X-ray film exposed to the filter, which has been hybridized with a radioactively labelled probe (nonradioactive methods are also available).

The probe can be obtained from various sources, including probes developed for another species. If sequence divergence is sufficiently small (usually less than 25%), such "heterologous" probes will allow the identification of known genes. Probes can also be obtained by choosing a restriction fragment from a genomic DNA library of the species under study, or by using synthetic oligonucleotides of known sequence, such as (CA), a common microsatellite marker. Hybridization conditions determine the "stringency" or degree of sequence similarity required between probe and substrate for hybridization to occur, and at lower stringency heterologous probes will hybridize with the fragment of interest despite minor sequence differences. This makes it possible to use probes that are isolated from different populations or even different species to survey allelic diversity in DNA fragments. Examples include minisatellite multilocus probes, such as 33.15 and 33.6 which are cloned in M13.

Genomic library construction: If probes are not available for a given application, one can often produce them from a sample of genomic DNA by creating a recombinant DNA library. This term refers to a collection of bacteria containing plasmids or phage vectors into which DNA fragments have been inserted. Usually, DNA fragments are generated by digestion of total DNA with appropriate restriction enzymes and cloning into a vector. The choice of the restriction enzyme and the size of the cloned fragments determine the quality of the library. Screening these fragments in various ways allows identification of specific regions of DNA representing genes of interest, loci linked to specific genes, or random genetic segments whose coding function is not known.

Sequencing: The most direct strategy of studying polymorphism at DNA level is the determination of the nucleotide sequence of a defined region and the alignment of this sequence to an orthologous region in the genome of another, more or less related organism. DNA sequencing has become a routine procedure since the development of the dideoxy chain termination method. In combination with PCR it provides a method of collecting highly reproducible, informative and precise data for short DNA sequences and can be adapted at different levels of discriminatory potential by choosing appropriate regions of the genome. Reagents are available for sequencing of DNA generated from phage vectors or asymmetric PCR reactions and stretches of 250-800 nucleotide bases can be sequenced in a single sequencing reaction. Currently, DNA sequencing is mainly applied for evaluating medium and long-distance relatedness in phylogeny, but sometimes it is also used in population studies. Technical difficulty and expense limit use of conventional sequencing methods. Fluorescence-based automated sequencing machines have made the technology easier and faster to perform. However, the sequencers are expensive (more than US $1,000,000) and the cost prohibits their widespread use.

Polymerase Chain Reaction (PCR): The isolation of thermally stable DNA polymerase from the hot springs bacteria Thermus aquaticus (Taq polymerase) led to an efficient means of amplifying short fragments of DNA using automated thermal cycler. PCR technique has revolutionized the methodological repertoire of molecular biology. This technique allows us to amplify DNA sequence of interest to high copy numbers, thereby circumventing the need of molecular cloning. The primers may be specific (such as minisatellite and microsatellite sequences complementary to flanking regions of the desired locus), semi-specific (such as "Alu Repeats") or arbitrary (Random Amplified Polymorphic DNA primers). Primers (generally 10-30 bases length) are combined with nanogram quantities of genomic DNA, plus free deoxynucleotides, a reaction buffer and Taq DNA polymerase.

During a series of heating and cooling cycles, the DNA is denatured into single-stranded molecules, the primers anneal to their complementary sequences on either side of the target region and the DNA polymerase replicates the region downstream from each primer. The amount of target DNA doubles with each cycle, until microgram quantities are present. PCR can significantly decrease the amount of time required to isolate a desired segment of the genome. Also, PCR allows DNA analysis to be performed from small tissue samples. Recently, an advanced technique of PCR has been developed under which larger fragments can be amplified without compromising the precision of amplification ("long PCR") (Cheng et al., 1994). Long PCR also eases concern about limited or partially degraded samples, because even if only a small fraction of DNA is intact, target segments from unbroken DNA can be amplified completely in one bout.
**APPROACHES IN DNA FINGERPRINTING**

DNA fingerprinting is done mainly by two approaches, namely based on hybridization and on PCR. Weising et al., (1995) have worked out 'near-realistic' costs for chemicals and disposables in the Netherlands as US$ 1.4-4.0 (RFLPs) 2.3-7.0 (hybridization - based DNA fingerprinting) and 1.45 (PCR-based DNA fingerprinting) per sample per experiment, which do not include the cost for labour, laboratory equipment and space. It has been already mentioned about the techniques associated with both the approaches. Following are some more details on the latter approach also incorporating my experience on genetic analysis of scombroid fishes.

**Random Amplified Polymorphic DNA (RAPD):** The technique involves amplification of anonymous DNA fragments using synthetic oligodeoxynucleotides and is based on standard PCR methodology (Welsh and McClelland, 1990; Williams et al., 1990). Primers with ten bases (decamer) and a GC content of at least 50% (Usually more than 60%) are generally used. The region between two primer sites will be amplified if the 5' ends of the annealing sites (or the 3' end of the annealing primer) face one another on opposite stands. Furthermore, the annealing sites (which are inverted repeats and mostly complementary to the primers) must be within 2.5-3.0 kb so that the intervening region can be amplified during routine PCR. The amplification products are resolved on an agarose or polycrylamide gel and detected by ethidium bromide or silver staining.

**RAPDs of scombroid fishes:** I have analyzed RAPD fingerprinting patterns in Indian mackerel (Rastrelliger kanagurta) and king seer (Scomberomorus commerson). Specimens of mackerel were collected from commercial gill net landings at Mandapam, Mangalore and Fort Kochi (9°17'-12°52'N, 74°52'-7°15'E) and those of king seer from Mandapam during 1995-96. Tissue was transported to the laboratory in one of the three media, such as crushed ice, lysate buffer or 95% ethanol. In the laboratory, all tissues were immediately stored at -70°C until DNA extraction. A modified phenol-chloroform extraction procedure (Jayasankar and Dharmalingam, 1997) was used to extract genomic DNA. DNA from 30 specimens of mackerel and 5 specimens of king seer were analyzed by the author.

**Standardized PCR amplification conditions in the study by the author:**

- *Master mix*: 100 mM Tris HCl pH 8.3, 50 mM KCl, 0.01% gelatin, 2.4 mM MgCl₂, 0.03 mM each of dATP, dTTP, dGTP and dCTP, 0.64 μM random primer (35 primers from kits A, F and G of Operon Technologies, Inc.) and 1 μg Taq DNA polymerase enzyme (either Amersco or Rama Biotechnologies).
- Thermal cycles: 1 initial cycle of 30 sec denaturation (94°C), 30 sec annealing (36°C) and 1 min extension (72°C) followed by 45 cycles of 30 sec denaturation, 30 sec annealing and 2 min extension. Final extension for 7 min.

In this study, OPA 07 (GAAACGGGGT) yielded more polymorphic loci with better reproducibility. Better performance of kit A primers is attributed to their high G+C content. Primers with the same nucleotide sequences produced varying band patterns and those with different sequences generated different banding patterns in mackerel and seerfish. RAPDs were highly sensitive to the concentrations of primer, MgCl₂ and the brand of Taq DNA polymerase enzyme.

**Conclusion**

The pace innovation in DNA fingerprinting technology is breathtaking. Several DNA typing methods, particularly locus-specific typing strategies ("DNA Profiling") are being developed for the stock structure in fish populations.

**RAHM of mackerel**

- Radioisotope labeling of microrolylated probe: 30 pmols of (GT)₃, were 5' end-labeled with 50 μCi of ³²p dATP. The unincorporated nucleotides were separated from the probe using Nuc Trap probe purification push column (Strategene).
- Hybridization: 0.1% SDS, 5 μl each of SSPE and Denhardt's solution and 64 μg of snuffe E. coli DNA. Temperature 45°C.
- Hybridized: 0.1% SDS, 5 μl each of SSPE and Denhardt's solution, 64 μg of snuffe E. coli DNA, 1.5 g of Dextran sulphate and 12 μl of purified probe. Temperature 45°C.
- Final washes: 3 X SSC and 0.3% SDS, 3 washes for 20 min at 37°C.
- The filter was wrapped in a cellophane and developed in a phosphorimage (Molecular Dynamics Inc.).

The author has observed 25% positive hybridization in one gel. Some of the loci which were not adequately represented in the gel were clearly detected in the phosphorimage. Signal intensity of bands depends both on the microsatellite motifs and the abundance of the particular loci.
analysis of parentage, behavioural genetics, population definition and forensics. Novel techniques, such as minisatellite variant repeat mapping, DNA typing with tetrabaseotide repeats, multiplexing, etc. will allow us to assay more loci, for more individuals, faster than ever before. DNA-level research in tropical fish is still in its infancy. High cost of equipments and expendables could be one reason for the hesitancy on the part of researchers from developing nations to jump into the DNA bandwagon. However, with the cost is likely to come down. Computer software for image analysis as well as for the evaluation and processing of fingerprinting data is expected to improve considerably. This would greatly facilitate and accelerate genome mapping studies as well as the analysis of genetic diversity and relatedness within and between species, populations and individuals.

REFERENCES


Spring Harbour Press, New York, USA.

Waldman, J.R. and J. Wirgin 1995. Use of DNA analyses in the management of natural fish populations. In : S.J. Gartie (Ed), Molecular Environmental Biology, pp. 29-64.


