DNA Fingerprinting fosters Genetic Research in Fisheries and Aquaculture

P. Jayasankar
Scientist (Senior Scale)
Central Marine Fisheries Research Institute
P B No. 1603,
Tatapuram, Kochi 682 014

The term 'Fingerprinting' was coined by Alec Jeffreys in 1985 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 their discovery, there has been ongoing progression in the technological development and breadth of application of genetic typing methodology. The technique (also known as DNA 'profiling') has been applied to such diverse biological phenomena 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. Useful relevant reviews are available (Wright, 1993; Waldman and Wirgin, 1993; Carvalho and Pitcher, 1994; Fergusson et al., 1995; O'Reilly and Wright, 1995) which highlight relative merits of DNA polymorphic markers over the other markers including protein (allozyme) in detecting genetic variations. Among the several advantages of DNA-level markers are 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 clearly shows that there are particular classes of sequences common to many organisms. Based on its function, structure, location etc., DNA can be classified as follows, though many of these characters may overlap (Park and Moran, 1995).

Coding and Non-coding DNA: Of the billions of nucleotides which constitute the total genomic DNA of eukaryotic organisms only 1% regulates or codes for essential proteins (that is precisely the major disadvantage of protein electrophoresis, which is incapable of assaying a majority of the genome). The non-coding portions, often referred to as 'junk DNA', are highly variable as they are not subject to selection pressures. However, some regulating function to the non-coding DNA sequences cannot be completely ruled out.

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 once in the haploid genome; the remainder, 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 equilibrium density centrifugation. This class of DNA may have 10-100 bp repeat units (mini-satellites) or only 1-4 bp repeat units (micro-stellites) and they form the corner stones of DNA profiling technology. 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.

Mitochondrial DNA (mtDNA): Mitochondrial DNA is haploid and predominantly inherited maternally. It is 16,000-20,000 bp in size and circular. In many organisms, the mt DNA seems to accumulate mutations more rapidly than do Scn DNA making them markers with greater variability and sensitivity to drift, and is 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). Some regions of the genome, such as cytochrome and ND genes exhibit greater variability. 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.

Nuclear DNA: The nuclear genome in bony fishes is about 0.3-4.0 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 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 generalised into tow categories: base substitutions (such as point mutation) and insertion/deletions. A substitution from a purine to purine or from a pyrimidine to pyrimidine is known as transition, whereas a substitution from a purine to pyrimidine and *vice versa*, is known as a transversion. Among closely related organisms transitions occur more frequently than do 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 (for example, VNTRs, (variable number of Tandem Repeats) resulting from variation in the copies of a basic unit, or core sequence.

Hybridisation of DNA molecules

By heating, the hydrogen bonds between the strands of DNA are broken (denaturation)

and as DNA cools, 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 hybridisation of DNA molecules depending on the degree of 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 hybridisation/annealing.

DNA techniques for studying genetic variation

Sequencing: The most direct strategy of studying polymorphisms 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 provides a highly reproducible and informative analysis of data and can be adapted at different levels of discriminatory potential by choosing appropriate regions of the genome. 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 $ 100, 000) and the cost prohibits their widespread use.

Restriction Fragment Length Polymorphisms (RFLPs): Restriction enzymes are endonucleases which occur in a variety of prokaryotes and their natural function is to destroy foreign DNA molecules by recognizing and cutting specific DNA sequence motifs typically consisting of four or six bases. Each enzyme has a particular recognition sequence, and the bacteria usually protect thier own DNA from being cut by methylating this sequence.

Analysis of RFLPs for evaluating DNA sequence variation is widely used, including fisheries field. Briefly, genomic DNA is extracted, digested with restriction enzymes (a large number of them are now available commercially) and separated by electrophoresis on a gel. This gel is Southern blotted on to a membrane, either nitrocellulose or nylon and specific fragments are made visible by hybridization with a labelled probe. RFLP probes are locus-specific and easy to screen co-dominant markers.

Nuclear RFLP probes are generated from a cDNA or genomic library of the investigated species, or a close relative. cDNA libraries are constructed by isolating mRNA and *in vitro* transcribing them into DNA by reverse transcriptase. These cDNA molecules are then cloned into a vector and used as probes in subsequent analyses. For the construction of a genomic library, total DNA is isolated, digested with a restriction enzyme and cloned into a vector. The choice of the restriction enzyme and the size of the cloned fragments determine the quality of the library. Since establishing a library requires considerable effort, nuclear RFLP studies are performed rarely for wild species.

There are two approaches for studying RFLPs in the cytoplasm. The first is to extract mtDNA separately from nDNA and digest them with restriction enzymes and electrophoresing on agarose or polyacrylamide gels. RFLPs are directly detected by ethidium bromide or silver staining. The second strategy is to isolate and digest the total DNA of the organism, followed by electrophoresis and southern blotting of the restriction fragments. The cytoplasmic DNA is then visualised by hyridisation with total mtDNA or specific mtDNA sequence labeled probe.

DNA Fingerprinting based on hybridisation: The technique of classical fingerprinting is methodologically derived from RFLP analysis and is mainly distinquished from the latter technique by the kind of hybridisation probe applied to reveal polymorphisms. Unlike in RFLP analysis, to obtain a typical DNA fingerprint, multilocus probes are used to create complex banding patterns by recognising multiple DNA loci simultaneously. Each of these loci is characterised by more or less regular arrays of tandemly repeated DNA motifs (mini-satellites and micro-satellites) that occur in different numbers at different loci. Several fish species have exhibited high degree of polymorphisms between related genotypes, which have been exploited for numerous studies in diverse areas of genome analysis. However, high complexity of profiles makes it difficult to identify both members of allelic pairs at individual loci. Thus it is almost impossible to estimate allele frequencies, necessary in many population analyses. A major practical advantage of multilocus fingerprinting, however is the availability of probes that cross hybridize in different taxas (Georges *et al.*, 1988).

Wright (1983) cautions the researchers to satisfy four conditions before using multilocus DNA fingerprinting in fisheries analyses, (1) the fingerprint profiles should be reproducible, (2) the diagnostic bands should be inherited in a Mendelian fashion, (3) variable loci should exhibit sufficient stability in the germ line and somatic tissue, and (4) restriction enzymes used should not be methylase sensitive.

Development of single locus profiling requires isolation and cloning of single VNTR loci. This approach is more efficient to estimate allelic frequencies in fish populations, though more laborious. O'Reilly and Wright (1995) have provided list of multilocus and single locus probes used in DNA fingerprinting of fishes. Mini-satellite probes 33.15 and 33.6 are used more frequently.

DNA Fingerprinting based on PCR: In 1985, a new technique was introduced which revolutionised the methodological repertoire of molecular biology, the Polymerase Chain Reaction (PCR). This technique allows us to amplify any DNA sequence of interest to high copy numbers, thereby circumventing the need of molecular cloning. In order to amplify a particular DNA sequence, two single-stranded oligonucleotide primers are designed, which are complementary to motifs on the template DNA. The primer sequences are so chosen as to allow base-specific binding to the template in reverse orientation. Addition of a thermostable DNA polymerase in a suitable buffer system and cycling programming of primer annealing, polymerisation and denaturation steps result in the exponential amplification of the sequence between the primer sites.

The primers may be specific (such as mini-satellite and micro-satellite sequences complementary to flanking regions of the desired locus), semi-specific (such as "Alu Repeats") or arbitrary (Random Amplified Polymorphic DNA primers).

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 McClleland, 1990, 1990; Williams *et al.*, 1990). Primers with ten nucleotides (10-mer) 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 primers) face one another on opposite strands. 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 polyacrylamide gel and detected by ethidium bromide or silver staining.

RAPDs have been analysed for a variety of genetic studies in *Oryzias latipes* (Kubota *et al.*, 1992; *Oreochromis* (Bardakci and Skibinski, 1994), *Danio* (Johnson *et al.*, 1994; Postlethwait *et al.*, 1994), *Dicentrarchus labrax* (Allegruci *et al.*, 1995), *Poecilia reticulata* (Fso *et al.*, 1995), Nile, *Mozambique* and *Aureus tilapia* (Dinesh *et al.*, 1996), *Penaeus monodon* (Garcia and Benzie, 1995) and *Penaeus vannamei* (Garcia *et al.*, 1996).

RAPDs of Scombroid fishes: I have analysed RAPD fingerprinting patters in Indian mackerel (*Rastrelliger kanagurta*) and king seer (*Scomberomorus commerson*). Mackerels were sampled from Mandapam, Mangalore and Fort Kochi fish landing centres while king seers were collected from Mandapam. Genomic DNA was extracted from ethanol-preserved/frozen muscle using a modified phenol chloroform extraction procedure. PCR amplification conditions using 10-mer random primers were optimised. Primers with the same nucleotide sequence generated different banding patterns in mackerel and seer while primers with different nucleotide sequences produced different patterns in the same template of these fishes (Fig.1). Of a total of 35 arbitrary primers from kits A, F and G (Operon technologies, Inc) tested, OPA 07 (GAAACGGGTG) yielded more polymorphic loci with high reproducibility. Better performance of the primers from kit A is attributed to their high G+C content. RAPDs in mackerel were highly sensitive to the concentrations of primer, $MgCl_2$ and the brand of Taq DNA polymerase enzyme. Same templates amplified with two brands of Taq (Amresco and Rama Biotechnologies) yielded different fingerprints (Figs.2 and 3).

Pair-wise comparison of RAPD loci was done to get Simple Matching Coefficient (SMC) whose value ranges from 0 to 1 and indicates degree of genetic similarity between two individuals. My data revealed maximum within-centre variability for Mandapam samples. Dendrographs generated from 1-SMC matrix (which is indicative of genetic distance) using Fitch-Margoliash programme version 3.56c of PHYLIP package did not show clear centre-specific clusters. However, based on the observations that mean SMC values were close to similarity in Mangalore and Fort Kochi samples and that maximum branch lengths in the 1-SMC dendrographs were Mangalore-Mandapam and Fort Kochi-Mandapam, restricted intermixing of mackerels between the east and west coasts could be assumed. Based on the present data, I suggest a combination of greater sample size, more number of informative arbitrary primers and a Taq polymerase which can yield more loci, to generate adequate data for interpreting genetic stock structure in fish populations.

Applying RAPD technique in genome analysis is advantageous mainly because of its simplicity and that DNA fragment with unknown sequence could be amplified and detected. The major disadvantage with this procedure lies in the fact that it cannot differentiate between homozygotes and heterozygotes as the markers are inherited in a dominant fashion. One approach to obviate this problem is to hybridize the RAPDs with labeled oligonucleotide motifs after Southern blotting and detecting by autoradiography or phosphor-imaging. I have successfully attempted hybridisation of vacuum blotted (a procedure related to Southern blotting) RAPD markers of mackerel with radioactively labeled micro-satellite probes, such as $(GT)_8$ and $(GA)_{12}$. Signals were developed by phosphor-imaging (Fig.4) which indicated the presence of micro-satellite complementary sequences among the RAPD loci scored by electrophoresis and ethidium bromide staining. I observed 25% positive hybridisation in one gel. Some of the loci not adequately represented in the gel were clearly detected in the phosphor-image. This technique is basically that of

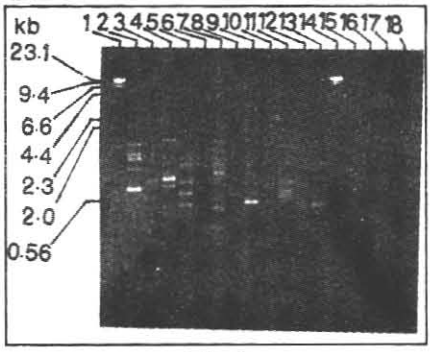

Fig.1

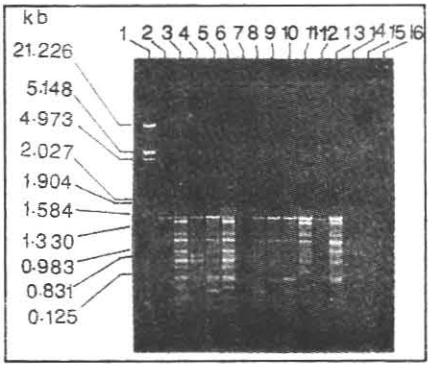

Fig.2

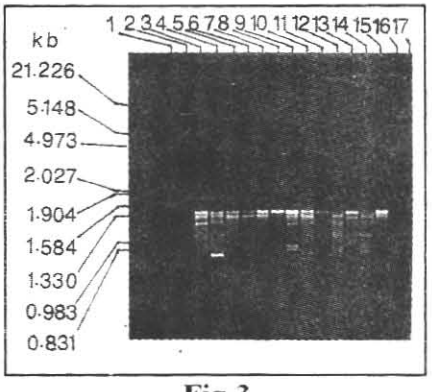

Fig.3

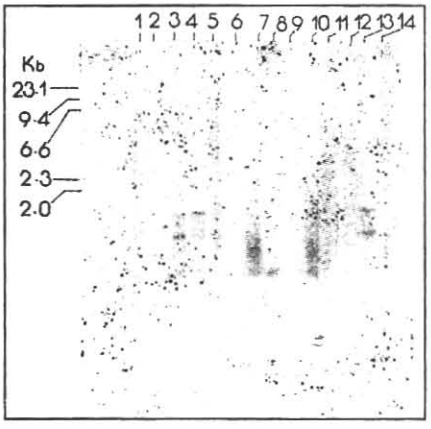

Fig.4

Cifarelli *et al.* (1995) who named it as Random Amplified Hybridisation Micro-satellites (RAHM). The loci which produced positive signals can be eluted from the gel and used as micro-satellite probes to detect genetic polymorphisms.

Micro-satellite VNTRs are already recognised as genetic markers for the future (O'Reilly and Wright, 1995; Wright and Bentzen, 1995). Properties, such as comparative ease of assay via PCR and accuracy of scoring alleles make micro-satellites attractive genetic variations. They can be easily isolated by standard cloning methodologies. Briefly, genomic DNA of the species under investigation is digested in separate reactions by restriction endonucleases that recognise 4-5 bp sequences (e.g.Pal I, Sau 3A I, etc). The DNA is then size fractionated by agarose gel electrophoresis and fragments of 300-600 bp are inserted in suitable vectors, such as pUC18 or M13. Following transformation of *E. coli* and plating on selective media, the size selected library is screened using radiolabelled oligonucleotides (nowadays non-radio active methods are also widely used) like $(GT)_8$ or $(GA)_{12}$. Inserts containing micro-satellites from positive clones are then sequenced, and flanking PCR primers designed.

Micro-satellite markers are particularly useful in species which show low levels of genetic variation and those which are inbred. Their co-dominant inheritance in Mendelian fashion is beneficial in pedigree studies. Cloning and characterisation of micro-satellite loci in temperate species, such as Atlantic cod, rainbow trout and Atlantic salmon have been reported. However, their applciation in tropical fishes is yet to make headway. High development costs, appearance of shadow/stutter bands and null alleles in the amplification products are the disadvantages of these markers.

Concluding remarks

The pace of innovation in DNA fingerprinting technology is breathtaking. Several DNA typing methods, particularly locus-specific typing strategies ("DNA Profiling") are being developed for the analysis of parentage, behavioural genetics, population definition and forensics. Novel techniques, such as Mini-satellite Variant Repeat Mapping, DNA typing with tetranucleotide repeats, multiplexing, etc will allow us to assay more loci, for more individuals, faster than ever before. However, our understanding of molecular biology of the highly polymorphic sequences is perhaps the tip of the iceberg (Wright, 1993).

DNA-level research in fish is still in its infancy especially with reference to the tropical species. High cost of equipments, enzymes, etc could be one reason for the hesitancy on the part of researchers from developing nations to jump into the DNA bandwagon. 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 (hybridisation-based fingerprinting) and 1.45 (PCR-based fingerprinting) per sample, per experiment, which do not include the cost for labour, laboratory equipment and space. Computer software for image analysis as well as for the evaluation and processing of fingerprinting data is expected to improve considerably which will greatly facilitate and accelerate genome mapping studies as well as the analysis of genetic diversity and relatedness within and between species, populations and individuals.

In India, some database is generated in the allozyme markers of few species of fishes. However, DNA-level research is yet to take shape. My work on RAPDs of Indian mackerel is one of the first contributions in this context. Agencies, such as the Department of Biotechnology (DBT) have to play a major role in funding projects related to molecular genetics in fish, shellfish and marine algae for the long term objectives of their conservation, profitable management in the wild and genetic improvement in aquaculture systems.

References

ALLEGRUCCI, G., A. CACCONE, S. CATANDELLA, J.R. POWELL and V. SBORDONI 1995. Acclimation of the European Seabass to freshwater: monitoring genetic changes by RAPD polymerase chain reaction to detect DNA polymorphisms. *Mar Biol.*, 121: 591-599.

BARDAKCI, F. and D.O.F. SKIBINSKI 1994. The application of the RAPD technique in tilapia fish: species and subspecies identification. *Heredity*, 73: 117-123.

CARVALHO, G.R. and T.J. PITCHER EDS 1994. Molecular genetics in Fisheries. *Rev. Fiush Biol. Fish.*, 4: 269-399.

CIFARELLI, R.A., M. GALLITELLI and F. CELLINI 1995. Random amplified hybridisation micro-satellites (RAHM): isolation of a new class of microsatellite-containing DNA clones. *Nucleic Acids Res.*, 23 (18): 3802-3803.

DINESH, K.R., T.M. LIM, W.K. CHAN and V.P.E. PHANG 1993. RAPD analysis: An efficient method of DNA fingerprinting in three species of tilapia. *Aquacult. Int.*, 4 (1) : 19-30.

FERGUSSON, A., J.B. TAGGART, P.A. PRODHOL, O. MCMEEL, C. THOMPSON, C. STONE, P. McGINNITY and R.A. HYNES 1995. The application of molecular markers to the study and conservation of fish populations, wth special reference to *Salmo. J. Fish Biol.*, 47 (Supplement A): 103-126.

FOO, C.L., K.R. DINESH, T.M. LIM, W.K. CHAN and V.P.E. PHANG 1995. Inheritance of RAPD markers in the guppy fish, *Poecilia reticulata. Zool. sci.*, 12: 535-541.

GARCIA, D.K., and J.A.H. BENZIE 1995. RAPD markers of potential use in penaeid prawn (*Penaeus monodon*) breeding programs. *Aquaculture*, 130: 137-144.

GARCIA, D.K., A.K. DHAR and A. ALCIVARWARREN 1996. Molecular analysis of a RAPD marker (B20) reveals two microsatellites and differential mRNA expression in *Penaeus vannamei*. *Mol. Mar. Biol. Biotechnol.*, 5(1): 71-83.

GEORGES, M., A.S. LEQUARRE, M. CASTELLI, R. HANSET and VASSART 1988. DNA fingerprinting in domestic animals using four different mini-satellite probes. *Cytogenetics and Cell genetics*, 47: 127-131.

JEFFREYS, A.J., V. WILSON and S.L. THEIN 1985. Hypervariable minisatellite regions in human DNA. *Nature*, 314: 67-74.

JHONSON, S.L., C.N. MIDSON, E.W BALLINGER and J.H. POSTLETHWAIT 1994. Identification of RAPD primers that reveal extensive polymorphisms between laboratory strains of zebrafish. *Genomics*, 19 : 152 - 156.

KUBOTA, Y., A. SHIMADA and A. SHIMA 1992. Detection of gamma ray-induced DNA damages in malformed dominant lethal embryos of Japanese medaka (*Oryzias latipes*) using AP-PCR fingerprinting. *Mutat. Res.*, 283: 263-270.

OHNO, S. 1974. Chordata I: Protochordata, Cyclostomata and Pisces. *Animal Cytogenetics*, 4: 1-92.

O'REILLY, P and J.M. WRIGHT 1995. The evolving technology of DNA fingerprinting and its applcation to fisheries and aquacul-

ture. *J. Fish Biol.*, 47 (Supplement A): 29-55.

OVENDEN, J.R., 1990. Mitochondrial DNA and marine stock assessment. *Aus. J. Mar. Freshwat. Res.*, 41: 835-853.

PARK, L.K. and P. MORAN 1995. Developments in molecular genetic techniques in fisheries. Molecular Genetics in Fisheries, G.R. Carvalho & T.J. Pitcher (Eds), Chapman & Hall, London: 1-27.

POSTLETHWAIT, J.H., S.L. JOHNSON, C.N. MIDSON, W.S. TALBOT, M. GATES, E.W. BALLINGER, D. AFRICA, R. ANDREWS, T. CARL, J.S. EISEN, S. HORNE, C.B. KIMMEL, M. HUTCHINSON, M. JOHNSON and A. RODRIGUES 1994. A genetic linkage map for the zebra fish. *Science*, 264: 699-703.

WALDMAN, J.R. and I. WIRGIN 1993. Use of DNA analyses in the managemnt of natural fish popualtions. In Molecular Environmental Biology, S.J. Garte (Ed), pp. 29-64, London: Lewis.

WEISING, K., H. NYBOM, K. WOLFF and W. MEYER 1995. DNA Fingerprinting in plant and fingi. cRC Press Inc. 322 pp.

WELSH, J. and M. McCLELLAND 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res., 18; 7213-7218.

WILLIAMS, G.K., A.R. KUBELIK, K.L. LIVAK, J.A. RAFALSKI and S.V. TINGEY 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.*, 18: 6531-6535.

WRIGHT, J.M. 1993. DNA fingerprinting of fishes. In *Biochemistry and Molecular Biology of Fishes*, Vol.2 P.W. Hochachka & T. Mommsen, Eds. pp. 58-91. Amssterdam: Elsevier.

WRIGHT, J.M. and P. BENTZEN 1995. Microsatellites: genetic markers for the future, Molecular Genetics in Fisheries, R. Carvalho and T.J. Pitcher (Eds), Chapman & Hall, London: 117-121.