A Review of PCR-Based DNA Fingerprinting Using Arbitrary Primers in Tropical Ornamental Fishes of South-East Asia

Gideon Khoo, Tit Meng Lim, and Violet P.E. Phang
Faculty of Science, Universiti Tunku Abdul Rahman, Perak, Malaysia
Department of Biological Sciences, National University of Singapore, Kent Ridge, Singapore
gideonkhoo@utar.edu.my

ABSTRACT
DNA fingerprinting has assumed an important role in aquaculture and fisheries management. Molecular genetic markers have been developed into powerful tools to analyze genetic diversity and relationships among different aquatic species. Besides the varied techniques using polymorphic DNA markers, Random Amplified Polymorphic DNA (RAPD) markers generated by Arbitrarily Primed Polymerase Chain Reaction (AP-PCR) have become ubiquitous and essential in aquaculture genetics. Single 9-20 base oligonucleotide primers with G+C contents ranging between 40-80% were randomly chosen for PCR to generate useful genomic fingerprints. DNA amplification products were resolved by denaturing polyacrylamide gel electrophoresis (dSDS-PAGE). Subsequent detection by silver staining could easily allow the visualization of fragments below 3.5 kb. The spectrum of amplified products varied with each primer-template combination but was consistent and reproducible under similar PCR conditions. The primers could detect polymorphisms in the absence of specific nucleotide sequence information of the fishes. Such DNA polymorphisms are especially useful for genotyping, taxonomic identification, population studies and genetic mapping. This approach is also applicable for assessing inbreeding levels and population structure within wild and cultivated fish and shrimp stocks. Here, we review the development, applications and advances of AP-PCR in elucidating DNA polymorphisms in tropical ornamental fishes. Genetic differences among species and strains, sexes, species- and phenotype-specific DNA polymorphisms, and applications of DNA marker-based approaches in fish breeding and management programmes are discussed.

Keywords: DNA fingerprinting, AP-PCR, ornamental fishes, linkage mapping.
1. Introduction

Genetic markers have provided valuable tools in various analyses ranging from phenotypic analysis to the positional cloning of genes. In general, genetic markers are specific “landmarks” on a chromosome that can be used for genome analysis (Kumar, 1999; Chistiakov et al., 2006). There are basically two types of genetic markers: morphological and molecular. Visible traits such as phenotypic colour patterns act as morphological markers as they can be monitored without specialized biochemical or molecular techniques. These traits are typically controlled by a single locus, and their alleles interact in a dominant-recessive manner. Nevertheless, the number of morphological markers is limited. Polymorphisms that are revealed at the protein level are known as biochemical markers, whereas those at the DNA level are DNA markers. Biochemical markers or protein markers such as isozymes and allozymes are the products of gene expression that have to be separated by electrophoresis to identify their alleles.

These markers reveal variations in gene sequence and function as co-dominant markers, but are few in number and subject to post-translational modifications. To understand the functional aspect of genes, it is essential to identify chromosomal segments that contain genetic markers. These markers, which include morphological, biochemical and nuclear DNA markers, always display Mendelian inheritance patterns. Morphological and protein-based markers are dependent on the expression of genes which, in turn, may be influenced by environmental conditions, epistasis, tissue specificity, developmental stages and age (Kumar, 1999; Chistiakov et al., 2006). In contrast, DNA markers are not governed by such external factors because any changes in them are due to natural variation or mutation within the gene sequence.

2. DNA Marker Technology

Polymorphic DNA markers may be produced by conventional hybridisation-based, or more recently, polymerase chain reaction (PCR)-based procedures (Jeffreys et al., 1985a, b, 1987; Jeffreys & Morton, 1987; Saiki et al., 1988; Park & Moran, 1994; Kumar, 1999; Chistiakov et al., 2006).

Hybridisation-based DNA markers include RFLPs (restriction fragment length polymorphisms) and VNTRs (variable number tandem repeats) where probes for genomic clones, cDNA clones, and mini- and microsatellite sequences are hybridised onto filters containing DNA that has been digested with restriction enzymes. RFLPs usually result from point mutations, translations, inversions and deletions while VNTRs are due to a difference in the number of repeats (Jeffreys et al., 1985a, b, 1987; Jeffreys & Morton, 1987; Kumar, 1999; Chistiakov et al., 2006).

PCR-based polymorphisms can be random or specific, depending on the type of primer, stringency of PCR conditions, and the method of fragment separation and detection. During PCR, oligonucleotide primers oriented with their 3’-ends pointing towards each other, copy each strand of denatured template DNA with the use of a polymerase enzyme that adds nucleotides starting from the 3’-ends of the primers which have recognised and annealed to the complementary sequences of the template DNA (Saiki et al., 1988; Williams et al., 1990; Welsh
The three steps of PCR: template denaturation, primer annealing and enzymatic extension, are repeated about 25-30 times to ensure a high level of amplification of the intervening regions (usually less than 3000 bp) specified by the primers.

The quantity of DNA is doubled during each cycle, resulting in an exponential increase in the amount of target DNA which is then separated and detected on a gel (Williams et al., 1990, 1993; Welsh & McClelland, 1990; Caetano-Anollés et al., 1991; Dinesh et al., 1993).

Multiple arbitrary amplicon profiling (MAAP) techniques, e.g., random amplified polymorphic DNA (RAPD) (Williams et al., 1990, 1993), arbitrarily primed PCR (AP-PCR) (Welsh & McClelland, 1990) and DNA amplification fingerprinting (DAF) (Caetano-Anollés et al., 1991) are the main sources of random PCR-based markers. These three strategies are similar in that they use at least one oligonucleotide primer to target sites randomly distributed in a genome, but they differ in primer length, amplification stringency and procedure used to resolve and detect the fragments (Saiki et al., 1988; Williams et al., 1990, 1993; Welsh & McClelland, 1990; Caetano-Anollés et al., 1991; Dinesh et al., 1993; Park & Moran, 1994; Kumar, 1999).

RAPD uses primers that are 9-10 nucleotides (mers or bp) in length followed by electrophoresis of PCR products on agarose gels and detection with ethidium bromide. Longer primers of 18-30 mers are utilised during AP-PCR but amplified products are resolved and visualised as for RAPDs. DAF employs very short primers of 5-8 bp, and the amplification products are separated on polyacrylamide gels (PAGE) containing 2-7 M urea and stained with silver (Caetano-Anollés et al., 1991; Bassam et al., 1991; Dinesh et al., 1993, 1995; Dinesh, 1995). For the three MAAP techniques, only one primer is required to amplify several bands that correspond to several loci. Hence, the number of primers that can be used is theoretically unlimited, and may provide potential coverage for the whole genome of an organism.

Another innovation in DNA marker technology is amplified fragment length polymorphism (AFLP) (Kumar, 1999; Mueller & Wolfenberger, 1999). During AFLP analysis, adaptors are ligated to the ends of restriction enzyme-digested DNA fragments followed by amplification with adaptor-homologous primers. Thousands of loci can be detected using this technique in which amplicons are resolved on denaturing polyacrylamide sequencing gels. A slight disadvantage of MAAP and AFLP markers is that they are dominantly expressed (presence against absence of a band) in most instances, making it almost impossible to differentiate heterozygous individuals from homozygous dominants (Williams et al., 1990, 1993; Welsh & McClelland, 1990; Caetano-Anollés et al., 1991).

Specific PCR-based markers are derived from primers of known sequence that are 18-30 bp in length, and amplifications with annealing temperatures of 50-70°C (Kumar, 1999; Chistiakov et al., 2006). Primers that amplify markers called sequence characterised amplified regions (SCARs) are synthesised based on the sequence at the ends of RAPD fragments (Paran & Michelmore, 1993). SCARs can be detected directly or after cleaving with restriction enzymes. Other specific PCR markers are short tandem repeats (STRs) and simple-sequence repeats (SSRs) where a terminally anchored primer specific to a certain repeat is used to amplify the
DNA between two opposing repeats of the same type. Polymorphism occurs when a repeat is not present in a particular genome, or the repeat has been modified by deletion or insertion. STRs and SSRs are especially useful because of their abundance, high level of polymorphism and ease of detection through automated systems. Single-stranded conformational polymorphisms (SSCPs) and simple sequence length polymorphisms (SSLPs) are produced by amplification of target sequences using radioisotope-labelled primers after which amplified products are denatured to a single-stranded form and electrophoresed on non-denaturing PAGE gels (Kumar, 1999; Chistiakov et al., 2006; Mutebi et al., 1997).

In addition, primers can be designed to match the nucleotide sequence of a probe to yield sequence tagged site (STS) markers (Laayouni et al., 2000), while expressed sequence tags (ESTs) have sequences corresponding to the ends of complementary DNA (cDNA) strands.

The development of DNA marker technology has seen the application of these markers in the construction of genetic linkage maps, comparative mapping analysis, tagging of economically important genes and quantitative trait loci (QTLs), marker-assisted selection (MAS), and map-based cloning (Park & Moran, 1994; Poompuang & Hallerman, 1997; Kumar, 1999; Chistiakov et al., 2006).

A genetic linkage map illustrates graphically the arrangement of loci, which include morphological, biochemical and DNA markers, along a chromosome. The recombination rates between two loci yield the map distance in map units or centiMorgans (cM) whereby 1% recombination = 1 map unit = 1 cM (Strickberger, 1990; Liu, 1997). Recombination frequencies, however, vary along the length of a chromosome as crossing-over near the centromeres is usually suppressed (Strickberger, 1990; Purdom, 1993). Of late, PCR-based DNA markers, for instance AP-PCR/RAPDs, AFLPs, STRs/SSRs, SSCP, SSLPs and STSs, are increasingly being used to construct linkage maps as an alternative to hybridisation-based markers (Postlethwait et al., 1994; Kazianis et al., 1996; Kocher et al., 1998; Liu et al., 1999; Ohtsuka et al., 1999; Agresti et al., 2000; Laayouni et al., 2000; Sakamoto et al., 2000; Li et al., 2006; Shen et al., 2007; Nomura et al., 2011).

An essential step during genetic linkage mapping is segregation analysis since parental lines should be genetically divergent to exhibit sufficient polymorphisms, but not so far apart as to cause sterility in the offspring (Foo et al., 1995; Garcia & Benzie, 1995; Elo et al., 1997; Stott et al., 1997; Liu et al., 1998). Genetic markers are initially identified in selected parental species or strains which are then crossed to produce segregating populations that may comprise F2s, backcrosses, recombinant inbred lines (RILs), or double haploids. Recombination frequencies for all pairwise comparisons between loci are estimated using the maximum likelihood method (Liu, 1997), and map units are calculated using Haldane’s (1919) or Kosambi’s (1944) mapping function. Computer programs such as MAPMAKER (Lander et al., 1987; Lincoln et al., 1992) and Map Manager (Manly & Olson, 1999; Manly et al., 2001) have been developed to facilitate linkage analysis and construction of genetic maps. Information acquired from conventional cytological, morphological and enzymatic data of several plant and animal species today is being integrated into linkage maps that have been developed on the basis of molecular markers.
Genome organisation and evolution of a species can also be investigated by comparative mapping analysis. Using this strategy, mapping information of one species is applied to predict the linkage relationships of closely related or distant species with homologous DNA markers and hybridisation probes found within conserved regions of their chromosomes (Postlethwait & Talbot, 1997; Beier, 1998; Kocher et al., 1998; Liu et al., 1999; Ohtsuka et al., 1999; Agresti et al., 2000; Laayouni et al., 2000; Sakamoto et al., 2000; Li et al., 2006; Shen et al., 2007; Nomura et al., 2011). Elucidation of conserved syntenic relationships across different species will allow consensus and comparative maps to be developed, and serve as additional sources of genetic markers.

Economically important traits, which include morphological characters, phenotypic colour patterns, growth rate, and resistance to diseases, parasites and stress, are controlled by a number of genes, each of which contributes to its final expression in a cultured species (Poompuang and Hallerman, 1997; Kumar, 1999). Genetic improvement of these traits, also termed as quantitative trait loci (QTLs), offers the potential to significantly increase yield and profitability.

Through genome mapping, it is possible to assign chromosomal positions to individual QTLs, determine the type and magnitude of gene effects, and ascertain the parent which contributes the positive allele of each QTL. Since QTLs can be inferred from the markers that are tightly linked to it on a highly saturated map, the inheritance of major genes of interest can be traced to each parent by marker-assisted selection (MAS). To carry out MAS, near-isogenic lines (NILs) or bulked segregation analysis (BSA) are usually applied (Michelmore et al., 1991; Reiter et al., 1992; Poompuang and Hallerman, 1997; Knight et al., 1999). NILs differ in the presence or absence of a target gene and a small region flanking the gene, hence revealing polymorphisms between NILs and the parental contributor of the gene (Kumar, 1999). BSA, in contrast, uses the DNA of a large number of selected individuals from a segregating population that has been pooled into two bulks. Based on the genotype of markers in the segregating population, one bulk is homozygous for a parental allele whiles the other for a second parental allele (Michelmore et al., 1991).

The order of markers relative to a gene of interest is determined by segregation analysis, and when a tight linkage is found, the inheritance of the gene can be traced through genetic crosses. Such information will enable the genomic composition of the chromosomes of a selected family, i.e., the broodstock parents and their offspring, to be established (Poompuang and Hallerman, 1997).

3. Applications in Ornamental Fish Genetics

In the ornamental fish industry, fish breeders have two major objectives (Fernando & Phang, 1985; Khoo, 2001). The first is the improvement of fish stocks to enhance their viability, reproductive fitness and adaptability to environmental changes and stress. The second is the need to continually develop, introduce and supply exotic and novel strains to captivate and sustain the interest of fish hobbyists and aquarists. The emphasis of ornamental fish breeding is towards the
development of fancy strains with more intense and uniform colours on the body and fins, different colour combinations, larger body size, and varied finnage patterns, lengths and shapes for the commercial market (Kirpichnikov, 1981). The inherent problems associated with ornamental fish breeding are due to small founder populations, limited effective population size of broodstocks, stringent selective breeding for a particular trait, intensive breeding over successive generations and lack of hybridisation with other strains, all of which cause a reduction of genetic variability and erosion of the gene pool in fish stocks (Kirpichnikov, 1981; Fernando & Phang, 1985; Carvalho, 1993; Khoo, 2001). This leads to inbreeding depression and genetically inferior fish stocks that are susceptible to diseases, and have low fecundity, fertility, fitness, viability, and higher incidences of morphological abnormalities and mortalities.

The conventional method of estimating genetic variability within plant and animal stocks is based on large-scale breeding experiments and comparisons of the degree of resemblance between relatives of a family in quantitative traits of economic importance such as growth rates. Such studies involve tremendous cost, space and labour, and are therefore very expensive to carry out and may take several years to complete. The generated data would give the heritability index ($h^2$), which is an estimate of genetic variability within a population (Kirpichnikov, 1981). Heritability estimates are, however, not very accurate because they are based on morphology, whereas molecular markers utilise variation at the DNA level to provide estimates of genetic variability.

PCR-based DNA fingerprinting techniques are extremely efficient in generating molecular markers that can be utilised to assess genetic variation within and among ornamental fish species, stocks and populations (Figs. 1 & 2) (Dinesh et al., 1993, 1995; Chan, 1995; Dinesh, 1995; Gan, 1997; Chen, 1999; Haridas, 1999; Koh et al., 1999; Khoo, 2001; Khoo et al., 2001, 2002). In contrast to the method of $h^2$ estimation, the measurement of genetic variability by DNA fingerprinting requires only small amounts of tissue samples from individuals of a population, and the whole procedure is completed in a laboratory within a short period of time without involving large-scale and long-term experiments in a farm. The foremost potential application of DNA fingerprinting in ornamental fishes is in efficient monitoring and genetic management of stocks, and in stock improvement programmes. Screening of DNA fingerprints of fishes can reveal the level of genetic variability and inbreeding within stocks as well as the genetic differences among fish stocks from various sources.

Our laboratory in the Department of Biological Sciences, National University of Singapore, has employed a combination of amplification, resolution and detection techniques for RAPD, AP-PCR and DAF to analyse genetic polymorphisms in various teleost species (Dinesh et al., 1993, 1995; Chan, 1995; Dinesh, 1995; Foo et al., 1995; Gan, 1997; Chen, 1999; Haridas, 1999; Koh et al., 1999; Khoo, 2001; Khoo et al., 2001, 2002). This combination of protocols, which Dinesh et al. (1993, 1995) and Dinesh (1995) termed as “RAPD fingerprinting”, comprises PCR amplification with primers of 9-30 bp, electrophoresis of amplified products on discontinuous sodium docedyl sulphate polyacrylamide gels (dSDS-PAGE) containing 7M urea followed by silver staining of separated DNA fragments. RAPD fingerprints of three tilapia species,
Oreochromis aureus (Aureus tilapia), O. mossambicus (Mozambique tilapia) and O. niloticus (Nile tilapia) are shown in Fig. 1. These fingerprints, generated using a single primer of 16 nucleotide length, show polymorphic DNA markers that could differentiate the three species while the rest of the markers are monomorphic. In order to generate more confirmatory data of fish stocks, it is always best to use at least five different primers.

In contrast to conventional agarose gel electrophoresis, dSDS-PAGE gel electrophoresis (Fig. 1), in combination with silver staining, provides much higher sensitivity in resolving and detecting RAPD fragments (Dinesh et al., 1993; Dinesh, 1995; Khoo, 2001; Khoo et al., 2001). This method is more powerful in revealing genetic variation of individuals and is, therefore, a better choice when working with highly homogeneous fish stocks or closely related fish strains. Polyacrylamide gel electrophoresis is, however, more tedious to perform than agarose gel electrophoresis. The RAPD profiles are subsequently analysed using freshly processed electrophoresis gels or black and white digitised photographs. Such fingerprints can be used to quantify the level of genetic variability or inbreeding. Data are scored manually on photographs or by using computer software linked to an imaging system. Pairwise comparisons of the RAPD profiles of two individuals represented on the same gel are conducted based on the presence or absence of RAPD bands. This information is used for the estimation of genetic distance ($D$) and genetic similarity ($S$), which provide a measure of the genetic variation among individuals (Chan, 1995; Dinesh et al., 1995; Dinesh, 1995; Gan, 1997; Chen, 1999; Koh et al., 1999; Khoo, 2001; Khoo et al., 2002).

![Fig.1. RAPD fingerprints of four individuals each of three species of tilapia generated by a 16-mer primer, 5'-TGCCTGTGGGAATCC-3' (Dinesh et al., 1995). Lanes M: λDNA/BstEII Digest marker (Sigma), 1-4: Oreochromis aureus, 5-8: O. mossambicus and 9-12: O. niloticus. Arrowheads indicate polymorphic markers among the three species. Open arrowheads show monomorphic bands shared by O. mossambicus and O. niloticus.](image_url)
Fig. 2. RAPD fingerprints generated for four wild forms and five cultivated varieties of the discus fish (*Symphysodon* spp.) using five arbitrary primers (NUSZG4, NUSZG6, NUSZG8, OPA-3 and OPA-4) (Koh et al., 1999). Lanes M: 100 bp DNA marker (Gibco), T: Turquoise, P: Pigeon, G: Ghost, C: Cobalt, S: Solid Red, L: Wild Blue, R: Wild Brown, N: Wild Green, H: Heckel and E: negative control without template DNA.

4. Methods of Estimating Genetic Variability

In our studies on genetic variation in tropical fish species, we have analysed the RAPD profiles of the tiger barb (*Barbus tetrazona*), tilapia (*Oreochromis* spp.), goldfish (*Carassius auratus*), marine clownfish (*Amphiprion ocellaris*), discus fish (*Symphysodon* spp.) and guppy (*Poecilia reticulata*) (Chan, 1995; Dinesh et al., 1995; Dinesh, 1995; Gan, 1997; Chen, 1999; Koh et al., 1999; Khoo, 2001; Khoo et al., 2002).

For all comparisons, pairwise genetic distances between individuals were calculated using RAPDPLOT, a FORTRAN program designed by Black (1997) to display genetic relationships. Nei and Li’s (1979) pairwise genetic similarity index ($S$) among individuals was then computed and converted by RAPDPLOT into a measure of genetic distance ($D$) according to the formula $D = 1 - S$ (Hillis & Moritz, 1990).

Genetic similarity index ($S$) is given as $S = 2N_{AB}/(N_A + N_B)$, where $N_{AB}$ is the number of bands shared in common by individuals A and B, and $N_A$ and $N_B$ the total number of bands for A and B, respectively (Nei & Li, 1979; Lynch, 1990). $S$ values may range from 0 when no bands are shared between the RAPD profiles of two individuals, to 1 when no differences are observed, i.e., they are identical. The converse holds true for $D$ values.
Genetic variability in populations and cultured strains may also be estimated from the percentage of polymorphic RAPD loci (%P) and Nei’s (1973) average gene diversity using POPGENE (Yeh et al., 1997), a population genetic analysis software. Nei’s (1973) average gene diversity, also known as Nei’s average heterozygosity (H), is a measurement of genetic variation for randomly mating populations.

Average heterozygosity (H) is defined as the average of heterozygosity (h) over all loci (Nei, 1987). h is given as $h = 1 - \frac{\sum_{i=1}^{m} x_i^2}{m}$, where $x_i$ is the population frequency of the $i$th allele at a particular locus and $m$, the number of alleles. Calculations for genetic identity and distances were also performed using Nei’s (1972) genetic distance and Nei’s (1978) unbiased genetic distance that is corrected for small sample sizes. Nei’s (1972) genetic distance is given as $D = -\ln[G_{XY}^{}/\sqrt{(G_X G_Y)}]$ where $G_X$, $G_Y$ and $G_{XY}$ are the means of $\Sigma p_i^2$, $\Sigma q_i^2$ and $\Sigma p_i q_i$, respectively, over all loci, and $p_i$, $q_i$, and $p_i q_i$ the frequencies of the $i$th allele in populations X and Y. Nei’s (1978) unbiased genetic distance is defined as $D' = -\ln[G'_{XY}^{}/\sqrt{(G'_X G'_Y)}]$ where $G'_X$, $G'_Y$ and $G'_{XY}$ are the averages of $G_X$, $G_Y$ and $G_{XY}$, corrected for small sample sizes, respectively.

Cluster analysis was performed on the pairwise genetic distance estimates using the unweighted pair-group method with arithmetic mean (UPGMA; Sneath & Sokal, 1973) and Neighbour-Joining (NJ; Saitou & Nei, 1987) algorithms, as employed by the PHYLIP phylogeny inference package (Felsenstein, 1999). NJ and UPGMA phylograms were subsequently plotted using the TreeView program (Page, 1996).

To assess the confidence level of each branch or node in the phylogram, the binary matrix was bootstrapped 100-1000 times to generate matrices of the genetic distances. The matrices were then clustered into pseudo-replicated trees, and a consensus tree was generated using PHYLIP and TreeView.

5. Species Differentiation and Genetic Diversity

The capability of RAPD fingerprinting to clearly differentiate fish species, especially closely related ones as well as strains and colour variants, provides for an important application in tropical fish species. Here we present an example from three species of tilapia (Fig. 1). According to Dinesh et al. (1995), genetic similarity estimates indicated high intraspecific SI values within the Nile tilapia (0.73), Mozambique tilapia (0.78) and Aureus tilapia (0.87) species, while interspecies SI values were average for Mozambique-Nile tilapia (0.59), Aureus-Nile tilapia (0.46) and Aureus-Mozambique tilapia (0.39) comparisons.

Efficient differentiation and classification of closely related species and strains using RAPD fingerprinting has proven to be very useful in ornamental species such as the discus fish. The discus is one of the most expensive and popular aquarium fishes, and the current classification of two species and four subspecies is debatable.

Using the UPGMA method to analyze RAPD fingerprinting data (Figs. 2 & 3), Koh et al. (1999) found that the Heckel discus (Symphysodon discus) was genetically the most divergent, and was 2.89 times further in mean genetic distance from the other three wild forms, namely, Wild Green (S. aequifasciata aequifasciata), Wild Brown (S. a. axelrodi) and Wild Blue (S. a.
Fig. 3 and Table 1 show that the cultivated varieties (Turquoise, Pigeon, Ghost, Cobalt and Solid Red) are 3.18 times genetically closer to the three *S. aequifasciata* wild forms (mean genetic distance = 0.033) than to *S. discus* (mean genetic distance = 0.105). Based on these observations, Koh et al. (1999) concluded that the cultivated varieties most likely originated from the *S. aequifasciata* wild form, and there was no molecular genetic basis for the phenotypic classification of the cultivated varieties.

Pairwise genetic distances within and among feral populations and cultured strains of the guppy (*P. reticulata*) that were analysed using UPGMA and NJ revealed distinct clustering of guppy individuals into their respective populations and strains (Fig. 4) (Chan, 1995; Chen, 1999; Khoo, 2001; Khoo et al., 2002). Percentage polymorphic loci in these guppies ranged from 54.96% to 68.70%, while average heterozygosity was 0.22-0.27.

Feral guppies sampled from isolated locales appeared to have high intrapopulation genetic similarity values of 0.78-0.85, which were comparable to those of the cultured strains (Table 2). In the study of genetic diversity within and among populations, when heterozygosity values are high, genetic similarity would be approximately 0.5 for first-degree relatives (parent-offspring and full-siblings), 0.25 for second degrees (first cousins), and so forth. Our results for the guppy and goldfish (range of $S$ values = 0.75-0.90) are typically in the range exhibited by small, isolated populations or highly inbred strains. Thus, high similarity coefficients generally reflect high homozygosity among individuals of each population or strain.

This is indicative of high inbreeding levels. Similar studies may be conducted to estimate the genetic variation of ornamental fish stocks in farms, especially those which have been established for many years without introduction of fresh stocks from other sources (Khoo, 2001). DNA fingerprinting may also be applied to estimate genetic variation among stocks from different farms as well as from different sources. It could also be used to determine the genetic constitution of founder brood stocks. Such genetic information would be useful in genetic improvement programmes aimed at rectifying gene pool deterioration in stocks, and to enhance the genetic diversity through outcrossing and hybridisation.

Tropical ornamental fish stocks should ideally be monitored annually, and managed to prevent inbreeding depression and genetic erosion.
Fig. 3. UPGMA phenogram showing the genetic relationships among all individuals of the four wild forms and five cultivated varieties of the discus fish (Symphysodon spp.) constructed using PHYLIP (ver. 3.57c) (Koh et al., 1999). One unit of the scale represents 0.01 mean genetic distance. (T: Turquoise, P: Pigeon, G: Ghost, C: Cobalt, S: Solid Red, L: Wild Blue, R: Wild Brown, N: Wild Green and H: Heckel).
Fig. 4. Neighbour-Joining cluster analysis of Hillis and Moritz’s (1990) genetic distance following 100 bootstrap replications of guppies (*Poecilia reticulata*) sampled from feral populations in Bukit Timah (BT), Nee Soon (NS), Tuas (TS), Mount Faber (MF), Kranji (KR), Lab.-Inbred (LI), and the Tuxedo (TX) and Green Variegated (GV) strains (Khoo et al., 2002).

Numbers at the nodes of the consensus tree indicate the percentage of 100 replications supporting each branch of the tree.
Table 1: Mean similarity indices (corrected for intra-wild form or intra-cultivated variety similarity [Lynch, 1990]) for pairwise comparisons within and among four wild forms and five cultivated varieties of the discus fish (*Symphysodon* spp.) using five arbitrary primers (Koh et al., 1999). (L: Wild Blue, R: Wild Brown, N: Wild Green, H: Heckel, T: Turquoise, P: Pigeon, G: Ghost, C: Cobalt and S: Solid Red).

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
<th>N</th>
<th>H</th>
<th>T</th>
<th>P</th>
<th>G</th>
<th>C</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>0.906</td>
<td>0.972</td>
<td>0.957</td>
<td>0.872</td>
<td>0.904</td>
<td>0.883</td>
<td>0.914</td>
<td>0.925</td>
<td>0.907</td>
</tr>
<tr>
<td>R</td>
<td></td>
<td>0.860</td>
<td>0.982</td>
<td>0.894</td>
<td>0.907</td>
<td>0.901</td>
<td>0.919</td>
<td>0.932</td>
<td>0.934</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
<td>0.838</td>
<td>0.860</td>
<td>0.893</td>
<td>0.877</td>
<td>0.906</td>
<td>0.930</td>
<td>0.911</td>
</tr>
<tr>
<td>H</td>
<td></td>
<td></td>
<td></td>
<td>0.910</td>
<td>0.821</td>
<td>0.807</td>
<td>0.799</td>
<td>0.825</td>
<td>0.834</td>
</tr>
<tr>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.820</td>
<td>0.975</td>
<td>0.953</td>
<td>0.966</td>
<td>0.946</td>
</tr>
<tr>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.834</td>
<td>0.971</td>
<td>0.953</td>
<td>0.957</td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.868</td>
<td>0.983</td>
<td>0.970</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.790</td>
<td>0.985</td>
</tr>
<tr>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.834</td>
</tr>
</tbody>
</table>

Table 2: Estimations among six feral populations and two domesticated strains of the guppy (*Poecilia reticulata*) using Nei’s (1972) genetic distance (below diagonal) and Nei’s (1978) unbiased genetic distance (above diagonal) (Khoo et al., 2002). (BT: Bukit Timah, NS: Nee Soon, TS: Tuas, MF: Mount Faber, KR: Kranji, LI: Lab.-Inbred, TX: Tuxedo strain, GV: Green Variegated strain).

<table>
<thead>
<tr>
<th></th>
<th>BT</th>
<th>NS</th>
<th>TS</th>
<th>MF</th>
<th>KR</th>
<th>LI</th>
<th>TX</th>
<th>GV</th>
</tr>
</thead>
<tbody>
<tr>
<td>BT</td>
<td>–</td>
<td>0.109</td>
<td>0.115</td>
<td>0.091</td>
<td>0.142</td>
<td>0.155</td>
<td>0.223</td>
<td>0.234</td>
</tr>
<tr>
<td>NS</td>
<td>0.125</td>
<td>–</td>
<td>0.122</td>
<td>0.068</td>
<td>0.120</td>
<td>0.124</td>
<td>0.162</td>
<td>0.187</td>
</tr>
<tr>
<td>TS</td>
<td>0.131</td>
<td>0.138</td>
<td>–</td>
<td>0.133</td>
<td>0.148</td>
<td>0.138</td>
<td>0.214</td>
<td>0.221</td>
</tr>
<tr>
<td>MF</td>
<td>0.107</td>
<td>0.085</td>
<td>0.148</td>
<td>–</td>
<td>0.096</td>
<td>0.129</td>
<td>0.166</td>
<td>0.203</td>
</tr>
<tr>
<td>KR</td>
<td>0.160</td>
<td>0.138</td>
<td>0.166</td>
<td>0.114</td>
<td>–</td>
<td>0.082</td>
<td>0.148</td>
<td>0.174</td>
</tr>
<tr>
<td>LI</td>
<td>0.172</td>
<td>0.141</td>
<td>0.155</td>
<td>0.146</td>
<td>0.101</td>
<td>–</td>
<td>0.181</td>
<td>0.209</td>
</tr>
<tr>
<td>TX</td>
<td>0.239</td>
<td>0.179</td>
<td>0.230</td>
<td>0.183</td>
<td>0.166</td>
<td>0.199</td>
<td>–</td>
<td>0.096</td>
</tr>
<tr>
<td>GV</td>
<td>0.249</td>
<td>0.202</td>
<td>0.236</td>
<td>0.218</td>
<td>0.191</td>
<td>0.225</td>
<td>0.112</td>
<td>–</td>
</tr>
</tbody>
</table>
6. Pedigree Analysis, Sex Differentiation and Linkage Mapping

The most important application of RAPD fingerprinting to ornamental fishes is the detection of DNA markers linked to genetic factors controlling economically important traits such as specific colours and form, sex (male or female), disease-resistance, etc. Such a study in ornamental fishes could assist in genetic marker-based breeding programmes to improve stocks. Although the colour and form of fishes are controlled by multiple genes, many ornamental fishes are extremely inbred and hence, the genes responsible for specific traits are fixed through continuous selective breeding for numerous generations. In our analyses of the tiger barb (Dinesh, 1995) and guppy (Foo et al., 1995; Khoo, 2001; Lim, 1999; Khoo et al., 2003), we demonstrated Mendelian inheritance of RAPD markers in ornamental fishes (Fig. 5). This technology has paved the way for a detailed pedigree analysis of high value ornamental fishes such as the dragon fish or arowana (Scleropages spp. and Osteoglossum spp.), discus and Siamese fighting fish (Betta splendens). Pedigree analysis is especially crucial for the Asian arowana which is listed as endangered by the 2011 IUCN Red List of Threatened Species. International trade in these fishes is controlled under the Convention on the International Trade in Endangered Species of Wild Flora and Fauna (CITES), which means that only fishes bred in captivity on a fish farm for at least two generations can be traded. DNA fingerprinting certification is the only foolproof method of proving that the offspring sold in the international markets are farm-bred and not captured from the wild. DNA markers that distinguish among different strains of the arowana and between sexes have also been established (Yue et al., 2003; Mohd-Shamsudin et al., 2011), allowing fish breeders to identify these characteristics in immature animals.

Sex determination in fishes has great significance in fish breeding because there are differences in growth rate, behaviour pattern, breeding time, body colour, or size between males and females of each cultured species (Khoo, 2001). For high value ornamental fishes like the arowana and discus, the inability to differentiate between mature males and females greatly hampers their culture, management and breeding programmes. At present, mature breeding arowanas are stocked in ponds without any knowledge of the sex ratio. Another importance of sex determination is that fish farmers may want to culture males and females separately, or to achieve a monosex culture depending on biological traits like growth rate, colour variations and other economic traits. Monosex culture is crucial where one sex is in greater demand than the other, e.g., the Siamese fighting fish and guppy, where the males fetch higher prices than females because of their highly striking and fascinating colours. The problem of sex identification in fishes poses a significant challenge for the development and application of breeding programmes for commercially important fish species (Ferreiro et al., 1989). Unfortunately, the genetic sex of many fish species such as arowana and discus, cannot be deduced from external morphology. The problem is also exacerbated when dealing with embryonic and juvenile forms, or when only small tissue samples are available. One effective solution is to exploit DNA markers to diagnose sex. Such markers are present in the genome whenever sex determination is genetically controlled.
Sex identification at the DNA level has been successful for some animals (Griffiths & Tiwari, 1993). In mammals, the difficulty of obtaining Y- or X-linked markers has been solved by the discovery of the SRY gene (sex-determining region Y) which is structurally conserved and Y-linked across all mammals. For non-mammals, including fish, no widely conserved sex-specific genes have yet been described, and molecular sex identification in such organisms has been based on junk DNA sequences (Devlin et al., 1991). In the Chinook salmon (Onchorhynchus tshawytscha), a Y-chromosomal DNA fragment (OtY1) capable of determining genetic sex has been isolated. PCR-based DNA fingerprinting has been successfully used to identify Y-linked DNA markers in mice (Wardell et al., 1993), and to detect sex-specific markers in birds (Griffiths & Tiwari, 1993; Ong & Vellayan, 2007). This suggests that it is possible to detect a sex-specific marker from any organism where one sex possesses a unique chromosome. The detection of male-linked DNA polymorphisms in the guppy has been attributed to the use of extremely inbred aquarium stocks in which the male-specific elements have presumably been fixed for generations (Hornaday et al., 1994). To date, RAPD markers of the guppy that are diagnostic for sex (male and female) (Fig. 5), phenotype (Bcp and Rdt) and population (feral and domesticated) have been isolated, cloned and sequenced (Chen, 1999; Haridas, 1999; Khoo, 2001; Khoo et al., 2003). These markers could be converted into probes, or used directly as probes for in situ hybridisation to chromosomes (Kumar, 1999; Laayouni et al., 2000; Chistiakov et al., 2006). Following hybridisation of the probes, these markers would be assigned to specific locations on the X- and Y-chromosomes, and autosomes. DNA sequences of these markers could also be employed as anchors for linkage mapping, physical mapping and chromosome walking, or converted into SCARs, STSs and ESTs (Michelmore et al., 1991; Paran & Michelmore, 1993; Kumar, 1999; Laayouni et al., 2000; Chistiakov et al., 2006; Shen et al., 2007).

Genetic linkage maps have been constructed for the guppy from independent reciprocal crosses between the ornamental Tuxedo strain and a highly inbred feral line (Fig. 5) (Lim, 1999; Khoo, 2001; Khoo et al., 2003). Segregation patterns of RAPD markers and phenotypic markers were investigated in F1 and F2 offspring, and linkages were identified using MAPMAKER (Lincoln et al., 1992) and Map Manager (Manly & Olson, 1999; Manly et al., 2001) for about 77% of the markers scored from the crosses (Fig. 6). Linkage maps of the reciprocal crosses spanned about 1900-2100 Kosambi centiMorgans (KcM) in 28 linkage groups, with an average marker resolution of 10 KcM. Genome length of the guppy was estimated to be approximately 4060-4410 KcM with an average physical distance of 166 kbp/KcM. Several RAPD markers were found to be closely linked or mapped onto the putative sex-determining region (SdR) of the guppy Y-chromosome (Figs. 5 & 6) (Khoo, 2001; Khoo et al., 2003). These primary linkage maps have served as the initial step toward the construction of a composite high-density map to facilitate map-based cloning, and marker-assisted selection of QTLs that are linked to economically important traits in the guppy, e.g., phenotypic colour patterns, morphological characters such as size and shape of body and fins, and resistance to diseases, parasitic infections and stress, onto the sex chromosomes and autosomes (Brummell et al., 2006; Shen et al., 2007; Tripathi et al., 2009a & b). This will facilitate the insertion of these genes into the guppy
genome, improve the genetic constitution of guppy stocks and enable novel exotic strains to be produced in future without sacrificing genetic diversity as a result of conventional selection and breeding practices.

Lastly, DNA fingerprinting may prove to be an invaluable tool for the conservation of genetic diversity in natural and domesticated ornamental fish populations in the near future. For example, in the determination of conservation sites of natural populations of cichlids, osteoglossids, poeciliids, barbs, rasboras and tetras in freshwater river systems, and coral reef fish populations like the clownfish, cardinalfish, triggerfish and parrotfish. Overfishing, pollution, environmental destruction and overexploitation of natural resources are the major causes of diminishing biodiversity and possible extinction of rare ornamental species. The conservation of natural genetic resources has become crucial and DNA fingerprinting technology certainly has a major role to play in the planning and implementation of genetic conservation programmes.

Fig. 5. Inheritance of RAPD markers in F1 and F2 offspring of the Wildtype ♂♂ × Tuxedo ♀♀ cross as revealed by primer GEN2-60-12 (Khoo, 2001; Khoo et al., 2003). Lanes M: Amplisize™ Molecular Ruler 50-2000 bp Ladder (Bio-Rad) and N: negative control without template DNA. Wildtype ♂♂ and Tuxedo ♀♀ parents (P) are designated as W and T, respectively. Full-sib F1 parents (PF1) that were crossed to obtain the F2 generation are denoted by ♂ (male) and ♀ (female), respectively. F1 and F2 progenies are shown as ♂ for males and ♀ for females. Putative Y-linked markers generated by primer GEN2-60-12 are indicated by arrows.
Fig. 6. Comparison of some representative linkage groups generated from the Tuxedo ♀♀ × Wildtype ♂♂ (LG1, 2, 5 and 11) and Wildtype ♂♂ × Tuxedo ♀♀ (LGI, II, V, XI) crosses to highlight the extent of integration of common genetic markers. The complete linkage maps of the guppy are displayed in Khoo (2001) and Khoo et al. (2003).

Acknowledgements

This work was supported by research grants from the National University of Singapore and European Commission.

References


