Aquaculture Genome Technologies
Zhanjiang (John) Liu

Chapter 3
Randomly Amplified Polymorphic DNA (RAPD)

Zhanjiang Liu

Random amplification of polymorphic DNA (RAPD) is a polymerase chain reaction (PCR)-based multilocus DNA fingerprinting technique. The RAPD procedure was first developed in 1990 (Welsh and McClelland 1990, Williams et al. 1990) using PCR to randomly amplify anonymous segments of nuclear DNA with a single short PCR primer (8–10 base pairs [bp] in length). Because the primers are short and relatively low annealing temperatures (often 36–40°C) are used, the likelihood of amplifying multiple products is pretty good, with each product presumably representing a different locus. Once different bands are amplified from related species, population, or individuals, RAPD markers are produced. RAPD markers thus are differentially amplified bands using a short PCR primer from random genome sites. Because most of the nuclear genome in vertebrates is noncoding, it is presumed that most of the amplified loci will be selectively neutral. Genetic variation and divergence within and between the taxa of interest are assessed by the presence or absence of each product, which is dictated by changes in the DNA sequence at each locus. RAPD polymorphisms can occur due to base substitutions at the primer binding sites or to indels in the regions between the sites. The potential power for detection of polymorphism is relatively high; typically, 5–20 bands can be produced using a given primer, and multiple sets of random primers can be used to scan the entire genome for differential RAPD bands. Because each band is considered a bi-allelic locus (presence or absence of an amplified product), polymorphic information content (PIC) values for RAPDs fall below those for microsatellites and single nucleotide polymorphisms (SNP), and RAPDs may not be as informative as amplified fragment length polymorphisms (AFLP) because fewer loci are generated simultaneously. However, because of its relatively high level of polymorphic rates, its simple procedure, and a minimal requirement for both equipment and technical skills, RAPD has been widely used in genetic analysis, including that of aquaculture species.

In this chapter, technology advances leading to the development of RAPD, the principles and molecular basis of RAPD, inheritance of RAPD markers, the power of RAPD analysis, strengths and weakness of RAPD, and applications of RAPD in aquaculture genomics research will be summerized.

Technology Advances Leading to the Development of RAPD

RAPD is a PCR-based fingerprinting technique. The invention of PCR in the mid-1980s revolutionized the entire life sciences, earning a Nobel Prize in 1993 for its inventor, Dr. Kary B. Mullis. Understanding how PCR works is fundamentally

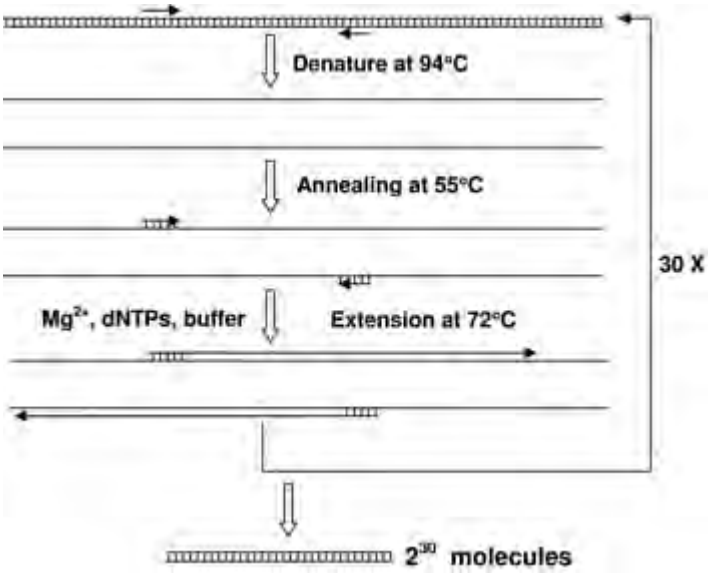

Figure 3.1. Principles and procedures of PCR.

important for appreciation of the principles of RAPD. Hence, I will briefly review the principles of PCR technology. PCR reactions start with double-stranded DNA templates. The first step is the denaturation of template DNA by heating it at 94°C; the second step is to anneal PCR primers to the templates. This step requires optimized temperatures according to the primers. Two factors significantly influence the fidelity of PCR—the length of the PCR primers and the annealing temperature. Generally speaking, the longer the PCR primer and the higher the annealing temperature, the higher the fidelity of the PCR reactions. Most often, however, PCR primers longer than 17 bp, and with an annealing temperature above 55°C, are sufficient to produce reasonably high fidelity for PCR. The third step of PCR is the extension of the annealed primers to synthesize new DNA. Once the extension is complete to the end of the template, PCR finishes its first cycle, and the original single molecule has been copied into two molecules. Let the process repeat 30 or more times, and one DNA molecule can be amplified into 2^{30} or more molecules (Figure 3.1).

Principles of RAPD

It is clear that in order to have exponential amplification, PCR requires two primers. Now we have genomic DNA, for which no sequence information is available. How can we conduct PCR reactions to produce genomic fingerprints revealing polymorphism? RAPD procedures are based on a fundamental understanding of the annealing process. At a given base position, any DNA has four possibilities of bases: A, C, G, or T. Therefore, if the primers are short enough, there would be numerous binding sites for them in genomic DNA. The odds for a perfect binding site to exist for a 10-base primer are once every 4^{10} base pairs (i.e., approximately once every million base pairs).

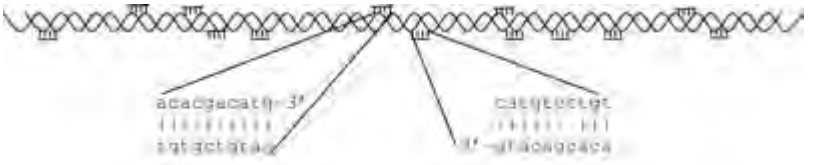

Figure 3.2. Schematic presentation of RAPD primer binding to genomic DNA. Short RAPD primers find their perfect and/or subperfect sites, anneal to genomic DNA, and amplify segments of genomic DNA when they are annealed close enough (generally <2,000 bp) on opposite strands of DNA.

In most eukaryotic organisms, such as fish, their genomes are at the billion base pairs range. There should be 1,000 perfect binding sites on each strand of the genome. However, the binding sites do not have to be perfect to initiate PCR if the annealing temperature is low enough. For instance, unless the last base at the 3′ is mismatched, when 9 out of 10 bases of the PCR primer have perfect matches to the template, PCR is likely to proceed if the annealing temperature is low. The possibility of subperfect binding greatly increases the number of binding sites in the genome from which a PCR reaction may proceed. The only exception is when the mismatches occur at the 3′ end of the primer. Therefore, there should be a large number of binding sites in a large genome for a short primer. However, PCR reactions are often limited to a certain length. Therefore, the short primers must bind to both strands of DNA close enough (within several kilobases [kb]) to produce a RAPD band. Using this principle, Welsh and McClelland (1990) and Williams and others (1990) used a single short PCR primer of 10 bases and conducted the special PCR reaction at 36°C, leading to the generation of PCR products using a single random short primer (Figure 3.2).

Based on the fundamental principles of RAPD, the technique can be regarded as a method of creating genomic fingerprints from species for which little is known about the target sequence to be amplified using arbitrarily primed PCR (AP-PCR). The creators of RAPD solved the dilemma of how to create a PCR primer without sequence information by using arbitrary short primers that increase the odds of finding suitable binding sites. The short primers require low temperatures for annealing.

Molecular Basis of RAPD Polymorphism

All mechanisms that led to the differential amplification of RAPD bands account for the molecular basis of RAPD polymorphism. First, RAPD depends on primer binding at adjacent sites on opposite strands of DNA. Any base substitutions at the primer binding sites may knock out primer binding and PCR amplification, thus leading to the loss of a RAPD band. Inversely, any base substitutions at a site originally with a sequence similar to the primer binding sites can lead to the generation of new primer binding sites. Once newly generated primer binding sites are close enough to another primer binding site on opposite strands of DNA, a RAPD band can be generated,

leading to polymorphism. Obviously, deletions and insertions within the RAPD bands would lead to either shorter or longer RAPD bands, producing polymorphism.

Inheritance of RAPD Markers

RAPD markers are inherited as Mendelian markers in a dominant fashion. As dominant markers, RAPD are scored as present/absent. Dominance means that one dose is enough, and therefore, a RAPD band is produced by dominant homozygotes as well as heterozygotes, though band intensity may differ. In spite of the theoretical validity of differentiating the dominant homozygotes from heterozygotes, variations in PCR efficiency make scoring of band intensities difficult. As a result, attempting to distinguish homozygous dominant from heterozygous individuals is not generally recommended. Also, it is difficult to determine whether bands represent different loci or alternative alleles of a single locus so that the number of loci under study can be erroneously assessed. This is especially true if the RAPD is caused by deletion or insertion within the locus rather than at the primer binding sites. As a result, the number of loci of RAPD markers can be inflated up to twofold.

As dominant markers, the alternative allele of a RAPD band is the absence of the band. Even though sometimes it is possible to determine alternative alleles by examination of the presence of alternative phases of RAPD bands, the exact nature of alternative RAPD bands must be verified by hybridization or by sequencing before calling them alternative alleles. As dominant markers, the number of RAPD bands seen in the F1 generation should equal the sum of the bands seen in the parents, assuming parental homozygosity at each locus; polymorphic RAPD then segregate in a 3:1 ratio in F2 populations, as shown in Figure 3.3 (Liu et al. 1998, 1999). If the RAPD bands are heterozygous in the parents, they segregate in a 1:1 fashion in F1 populations.

The Differentiating Power of RAPD

The RAPD approach is based on the fact that short oligonucleotide primers can bind to DNA with predicted odds. For instance, every 1 million (4^{10}) bp should contain one

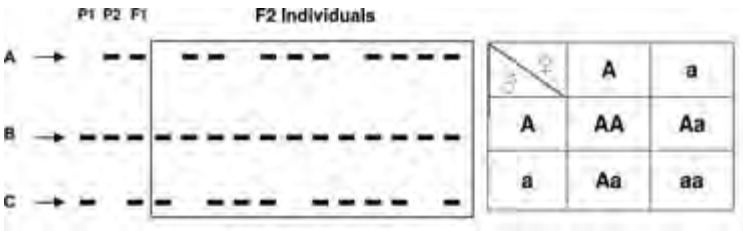

Figure 3.3. Inheritance of dominant markers. Here RAPD band A is inherited from parent 2 and band C is inherited from parent 1; both band A and C are heterozygous in F1. RAPD band B is inherited from both parents and thus are homozygous in F1. Heterozygous bands of F1 segregate in a 3:1 ratio among F2 individuals. Figure was modified from Liu and Cordes (2004).

sequence that matches with a primer of 10 nucleotides long. Therefore, a genome of 1 billion base pairs should contain 1,000 perfect binding sites for the 10-bp primer on each of its two strands of DNA. The 2,000 perfect binding sites plus many more subperfect binding sites (with 8–9 out of 10 nucleotides) would make it possible to amplify DNA using a single arbitrary short primer. The conditions for this special PCR reaction follow:

- The annealing temperature must be low because of the short primer.
- The short primer must bind to the opposite strands of DNA with its 3′ ends facing each other.
- The two binding sites must be close enough to allow a successful PCR reaction using *Taq* DNA polymerase, which often travels only several kilobases.

Generally, all these conditions can be met and often multiple bands can be amplified. Any deletion/insertion existing between the two successful primers would produce a polymorphic band. Additionally, base substitutions at primer binding sites can also cause gain or loss of amplified bands. Because about a dozen bands can be analyzed simultaneously and genome sequence information is not required, RAPD rapidly gained popularity for analysis of genetic variation in the 1990s.

Most often 5–20 bands can be amplified by using a single RAPD primer. Theoretically, primers with equal length should be equally efficient for generating RAPD bands, but G/C-rich RAPD primers were reported to produce more bands than A/T-rich primers, presumably due to stronger annealing of G/C-rich primers (Kubelik and Szabo 1995). Closely related species from which hybrids can be made often exhibit high levels of RAPD polymorphism; reproductively isolated populations often exhibit a reasonable level of RAPD polymorphism so that RAPD can be used to differentiate various strains, lines, and populations. RAPD tends to exhibit low levels of polymorphism among individuals of the same population, and thus are not ideal markers for parentage analysis, for which microsatellite markers are optimal.

Strengths and Weaknesses of RAPD

RAPDs have all of the advantages of a PCR-based marker, with the added benefit that primers are commercially available and do not require prior knowledge of the target DNA sequence or gene organization. Multilocus amplifications can be separated electrophoretically on agarose gels and stained with ethidium bromide, although higher resolution of bands has been achieved with discontinuous polyacrylamide gel electrophoresis (dPAGE) and silver staining (Dinesh et al. 1995), a somewhat costlier and more labor-intensive method. Other advantages of RAPDs are the ease with which a large number of loci and individuals can be screened.

The major weakness of RAPD is its low reproducibility due to the use of low annealing temperatures. However, if one stays conservative and scores only highly reproducible strong bands, this problem can be minimized. In our own experience, we have not encountered too much trouble with reproducibility. However, if one pushes to maximize the number of RAPD bands, then many very weak bands may not be reproduced, leading to a lack of reliability. Because of this reproducibility problem, there are reports that many RAPD bands do not follow Mendelian inheritance, though homozygous status was incorrectly assumed in cases. The second major

weakness of RAPD is its dominant mode of inheritance. Because of the dominant nature of inheritance, RAPD lack the ability to distinguish between dominant homozygotes and heterozygotes. In addition, the presence of paralogous PCR products (amplified from different DNA regions that have the same lengths and thus appear to be a single locus), limit the use of this marker system. These difficulties have limited the application of this marker in fisheries and aquaculture sciences (Wirgin and Waldman 1994).

Applications of RAPD Markers in Aquaculture Genome Research

RAPD markers have been widely used for species and strain identification in fish (Partis and Wells 1996, Liu et al. 1998, 1999) and mollusks (Klinbunga et al. 2000, Crossland et al. 1993), analysis of population structure in black tiger shrimp (Tassanakajon et al. 1998) and marine algae (van Oppen et al. 1996), analysis of genetic impact of environmental stressors (Bagley et al. 2001), and analysis of genetic diversity (Wolfus et al. 1997, Hirschfeld et al. 1999, Yue et al. 2002).

In addition to identification of species, strains, lines, and populations, RAPD markers have been extensively used in the model fish species such as zebrafish (Johnson et al. 1994). RAPD markers have also been used in many linkage-mapping studies in fish species (Table 3.1). However, as more efficient and reliable marker systems such as AFLP emerged, the use of RAPD markers in genome research declined rapidly. Due to the intrinsic problems as discussed under its weaknesses, the use of RAPD for future genome characterization of aquaculture species should be limited. Its coupled usage with codominant markers, such as microsatellites, may provide more reliable information. In closed aquaculture systems where the number of founders of the broodstock population is limited, RAPD may provide some rapid ways for association analysis of traits with markers. After the initial identification of the RAPD markers, it is highly recommended that the marker be converted into SCAR markers (sequence characterized amplified region) for further analysis. In spite of very limited uses of RAPD for long-term genome research, it is a useful marker system for rapid hybrid identification and strain identification commonly encountered in aquaculture breeding operations.

Table 3.1. Some examples of the use of RAPD markers for the construction of linkage maps in aquaculture or fish species. Note that most of these efforts were made earlier, and linkage mapping using RAPD markers is not highly recommended.

Species	Common name	References
Cyprinus carpio	Common carp	Sun and Liang 2004
Danio rerio	Zebrafish	Postlethwait et al. 1994, Mohideen et al. 2000
Oryzias latipes	Medaka	Ohtsuka et al. 1999
Oncorhynchus mykiss	Rainbow trout	Sakamoto et al. 2000
Astyanax mexicanus	Cave fish	Borowsky and Wilkens 2002
Xiphophorus sp.		Kazianis et al. 1996
Poecilia reticulata	Guppy	Khoo et al. 2003

Acknowledgments

Research in my laboratory is supported by grants from the USDA NRI Animal Genome and Genetic Mechanisms Program, the USDA NRI Basic Genome Reagents and Tools Program, the Mississippi-Alabama Sea Grant Consortium, the Alabama Department of Conservation, the USAID, and the BARD.

References

Bagley MJ, SL Anderson, and B May. 2001. Choice of methodology for assessing genetic impacts of environmental stressors: polymorphism and reproducibility of RAPD and AFLP fingerprints. Ecotoxicology, 10, pp. 239–244.

Borowsky R and H Wilkens. 2002. Mapping a cave fish genome: polygenic systems and regressive evolution. J Hered, 93, pp. 19–21.

Crossland S, D Coates, J Grahame, and PJ Mill. 1993. Use of random amplified polymorphic DNAs (RAPDs) in separating two sibling species of *Littorina*. Mar Ecol Prog Ser, 96, pp. 301–305.

Dinesh KR, WK Chan, TM Lim, and VPE Phang. 1995. RAPD markers in fishes: an evaluation of resolution and reproducibility. Asia-Pac J Mol Biol Biotechnol, 3, pp. 112–118.

Hirschfeld D, AK Dhar, K Rask, and A Alcivar-Warren. 1999. Genetic diversity in the eastern oyster (*Crassostrea virginica*) from Massachusetts using RAPD technique. J Shellfish Res, 18, pp. 121–125.

Johnson SL, CN Midson, EW Ballinger, and JH Postlethwait. 1994. Identification of RAPD primers that reveal extensive polymorphisms between laboratory strains of zebrafish. Genomics, 19, pp. 152–156.

Kazianis S, DC Morizot, BB McEntire, RS Nairn, and RL Borowsky. 1996. Genetic mapping in *Xiphophorus* hybrid fish: assignment of 43 AP-PCR/RAPD and isozyme markers to multipoint linkage groups. Genome Res, 6, pp. 280–289.

Khoo G, MH Lim, H Suresh, DK Gan, KF Lim, F Chen, WK Chan, TM Lim, and VP Phang. 2003. Genetic linkage maps of the guppy (*Poecilia reticulata*): assignment of RAPD markers to multipoint linkage groups. Mar Biotechnol, 5, pp. 279–293.

Klinbunga S, P Ampayup, A Tassanakajon, P Jarayabhand, and W Yoosukh. 2000. Development of species-specific markers of the tropical oyster (*Crassostrea belcheri*) in Thailand. Mar Biotechnol, 2, pp. 476–484.

Kubelik AR and LJ Szabo. 1995. High-GC primers are useful in RAPD analysis of fungi. Curr Genet, 28, pp. 384–389.

Liu ZJ and J Cordes. 2004. DNA marker technologies and their applications in aquaculture genetics. Aquaculture, 238, pp. 1–37.

Liu ZJ, P Li, B Argue, and R Dunham. 1998. Inheritance of RAPD markers in channel catfish (*Ictalurus punctatus*), blue catfish (*I. furcatus*) and their F1, F2 and backcross hybrids. Anim Genet, 29, pp. 58–62.

Liu ZJ, P Li, BJ Argue, and RA Dunham. 1999. Random amplified polymorphic DNA markers: usefulness for gene mapping and analysis of genetic variation of catfish. Aquaculture, 174, pp. 59–68.

Mohideen MA, JL Moore, and KC Cheng. 2000. Centromere-linked microsatellite markers for linkage groups 3, 4, 6, 7, 13, and 20 of zebrafish (*Danio rerio*). Genomics, 67, pp. 102–106.

Ohtsuka M, S Makino, K Yoda, H Wada, K Naruse, H Mitani, A Shima, K Ozato, M Kimura, and H Inoko. 1999. Construction of a linkage map of the medaka (*Oryzias latipes*) and mapping of the *Da* mutant locus defective in dorsoventral patterning. Genome Res, 9, pp. 1277–1287.

Partis L and RJ Wells. 1996. Identification of fish species using random amplified polymorphic DNA (RAPD). Mol Cell Probes, 10, pp. 435–441.

Postlethwait JH, SL Johnson, CN Midson, WS Talbot, M Gates, EW Ballinger, D Africa, R Andrews, T Carl, JS Eisen, et al. 1994. A genetic linkage map for the zebrafish. Science, 264, pp. 699–703.

Sakamoto T, RG Danzmann, K Gharbi, P Howard, A Ozaki, SK Khoo, RA Woram, N Okamoto, MM Ferguson, LE Holm, R Guyomard, and B Hoyheim. 2000. A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. Genetics, 155, pp. 1331–1345.

Sun X and L Liang. 2004. A genetic linkage map of common carp (*Cyprinus carpio* L.) and mapping of a locus associated with cold tolerance. Aquaculture, 238, pp. 165–172.

Tassanakajon A, S Pongsomboon, P Jarayabhand, S Klinbunga, and VV Boonsaeng. 1998. Genetic structure in wild populations of black tiger shrimp (*Penaeus monodon*) using randomly amplified polymorphic DNA analysis. J Mar Biotechnol, 6, pp. 249–254.

van Oppen MJH, H Klerk, JL Olsen, and WT Stam. 1996. Hidden diversity in marine algae: some examples of genetic variation below the species level. J Mar Biol Assoc UK, 76, pp. 239–242.

Welsh J and M McClelland. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucl Acids Res, 18, pp. 7213–7218.

Williams JGK, AR Kubelik, KJ Livak, JA Rafalski, and SV Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl Acids Res, 18, pp. 6531–6535.

Wirgin II and JR Waldman. 1994. What DNA can do for you. Fisheries, 19, pp. 16–27.

Wolfus GM, DK Garcia, and A Alcivar-Warren. 1997. Application of the microsatellite techniques for analyzing genetic diversity in shrimp breeding programs. Aquaculture, 152, pp. 35–47.

Yue G, Y Li, F Chen, S Cho, LC Lim, and L Orban. 2002. Comparison of three DNA marker systems for assessing genetic diversity in Asian arowana (*Scleropages formosus*). Electrophoresis, 23, pp. 1025–1032.