Aquaculture Genome Technologies
Zhanjiang (John) Liu

Chapter 2
Restriction Fragment Length Polymorphism (RFLP)

Zhanjiang Liu

Restriction fragment length polymorphism (RFLP) markers (Botstein et al. 1980) were regarded as the first shot in the genome revolution (Dodgson et al. 1997), marking the start of an entirely different era in the biological sciences. RFLP was the most popular approach for analysis of genetic variation during the entire 1980s. As indicated by its name, RFLP is based on deoxyribonucleic acid (DNA) fragment length differences after digesting genomic DNA with one or more restriction enzymes. Most typically, genomic DNA is digested by one or more restriction enzymes and separated on an agarose gel. To adapt to further handling, the DNA in the gel must be transferred to a solid support such as nitrocellulose or nylon membranes. The specific DNA locus with a potential fragment length difference is characterized by hybridization to a probe, a radioactively labeled DNA or ribonucleic acid (RNA) molecule with sequence similarities to the locus of interest. After hybridization, the nonspecific probes must be washed away leaving only hybridized probes to the specific locus. The membrane is then exposed to a piece of X-ray film for autoradiography to visualize the DNA bands. In spite of its popularity, RFLP is able to detect only large shifts in DNA fragment sizes. Therefore, it can detect only insertions and deletions of large sizes, and the gain or loss of restriction sites. It is unable to detect the vast majority of point mutations and deletions or insertions involving small-sized segments because of its low resolution using agarose gel electrophoresis. As a result, polymorphic rates are low at most loci. The efforts involved in RFLP marker development have been enormous. RFLP attempts to detect genetic variation one locus at a time. The low polymorphic rates, when coupled with expensive and laborious processes, have made application of RFLP limited. It should be particularly noted that RFLP requires previous genetic information, such as the availability of probes or sequence information; information often not available for many fish or other aquaculture species.

In this chapter, technology advances leading to the development of RFLP, the principles and molecular basis of RFLP, inheritance of RFLP, power of RFLP, strengths and weaknesses of RFLP, and applications of RFLP for aquaculture genomics research will be summerized.

Technology Advances Leading to the Development of RFLP

Two specific technological advances—the discovery and application of restriction enzymes and the development of DNA hybridization—set the foundation for RFLP. To comprehend and appreciate the principles of RFLP, it is necessary to have a good

understanding of restriction enzymes and their applications, as well as that of DNA hybridizations.

Restriction enzymes are also called restriction endonucleases, proteins produced by bacteria that cleave DNA at specific sites along the molecule. In the bacterial cell, restriction enzymes cleave foreign DNA, thus eliminating infecting organisms. Restriction enzymes can be isolated from bacterial cells and used in the laboratory to manipulate fragments of DNA. They are indispensable tools of recombinant DNA technology or genetic engineering, as well as genomics.

It is generally believed that the biological function of restriction enzymes is to protect cells from foreign DNA. A bacterium uses a restriction enzyme to defend against bacterial viruses called bacteriophages, or phages. When a phage infects a bacterium, it inserts its DNA into the bacterial cell so that it might be replicated. The restriction enzyme prevents replication of the phage DNA by cutting it into many pieces. Restriction enzymes were named for their ability to restrict, or limit, the number of strains of bacteriophage that can infect a bacterium. An obvious question that often arises is why the restriction enzymes do not digest bacterial DNA. The answer is that the bacteria also harbor a set of defense weaponry containing so-called restriction enzyme modification systems. Usually, organisms that make restriction enzymes also make a companion modification enzyme (DNA methyltransferase) that protects their own DNA from cleavage. These enzymes recognize the same DNA sequence as the restriction enzyme they accompany, but instead of cleaving the sequence, they disguise it by methylating one of the bases in each DNA strand.

To date, more than 10,000 bacteria species have been screened for the existence of restriction enzymes; more than 2,500 restriction enzymes have been found with more than 250 distinct specificities. Occasionally enzymes with novel DNA sequence specificities are still found, but most now prove to be duplicates (isoschizomers) of already discovered specificities.

There are three classes of restriction enzymes, designated Type I, II, and III (Table 2.1). Type I and III enzymes are similar in that both restriction and methylase activities are carried out by one large enzyme complex, in contrast to the Type II system, in which the restriction enzyme is independent of its methylase. Type II restriction enzymes also differ from the other types in that they cleave DNA at specific sites within the recognition site; the others cleave DNA randomly, sometimes hundreds of bases from the recognition sequence. Type II restriction enzymes are endonucleases that cut DNA at specific sites, and are most useful for molecular biology research.

Table 2.1. Classification of restriction enzymes and their characteristics.

Type I	Type II (93%)	Type III
Restriction-methylase on the same subunit	Homo-dimers, methylase on a separate subunit	Restriction-methylase on the same subunit
ATP-dependent	Mg^{++} dependent	ATP-dependent
Binds to DNA recognition site and cuts DNA randomly—any DNA as long as it comes in contact	Recognize symmetric DNA sequences and cleave within the sequences	Cut the DNA at the recognition site and then dissociate from the DNA

Each restriction enzyme recognizes a short, specific sequence of nucleotide bases. These regions are called recognition sequences and are randomly distributed throughout the DNA. Different bacterial species make restriction enzymes that recognize different nucleotide sequences. Generally speaking, Type II restriction enzymes recognition sites are palindromes. A palindrome read from both sides yields the same sequence of characters (e.g., 121, IFFI, ABA). However, for a DNA sequence, a palindrome refers to reading the sequence from both strands from 5′-3′. For instance, the *Eco*R1 site is 5′-GAATTC-3′; and its complementary strand should also read 5′-GAATTC-3′. Thus, most 4–8 base pair palindromes are likely restriction sites. There are numerous commercial suppliers of restriction enzymes, such as New England Biolabs, Amersham Pharmacia Biotech, Qiagen, Promega, Invitrogen, and Stratagene, to name a few.

Restriction enzymes are named by using the first letter of the genus name and the first two letters of the species name from which they were isolated. Often, additional letters are used to designate the strains from which they were derived, or the chronological order in which the enzyme was isolated from the species. For example, the enzyme *Eco*RI is produced by *Escherichia coli* strain RY13; *Pst* I was isolated from *Providencia stuartii*; *Hind* III was isolated from *Haemophilus influenza*, and *Not* I was isolated from *Norcardia otitidis-caviarum*.

The odds or frequency of restriction enzymes digesting DNA depends on their recognition sequences. The shorter the recognition sequences, the higher the cutting frequency. Restriction enzymes have recognition sequences of 4, 6, or 8 base pairs. Examples of 4-base pair (bp) cutters are *Taq* 1, *Hpa* II, *Msp* I; examples of 6-bp cutters are *Eco*R1, *Hin*d III, *Bam* H1, *Pst* I, *Sal* I; and examples of 8-bp cutters are *Not* I and *Sfi* I. To date, many 4-bp cutters and 6-bp cutters are available, but the number of 8-bp cutters is limited. In addition to these 4-, 6-, and 8-bp cutters, some restriction enzymes have interrupted or ambiguous recognition sequences. For instance, *Acc* I has a recognition sequence of GT(at/gc)AC; *Bgl* I has a recognition sequence of GCCNNNNNGGC; and *Afl* III has a recognition sequence of ACPuPyGT. Restriction enzymes with 4-bp recognition sequences digest DNA at a frequency of one per 4^4 = 256 bp; restriction enzymes with 6-bp recognition sequences digest DNA at a frequency of one per 4^6 = 4,096 bp; restriction enzymes with 8-bp recognition sequences digest DNA at a frequency of one per 4^8 = 65,536 bp. When genomic DNA is digested with 4-, 6-, or 8-bp cutters, a smear should result except that the average size of the 8-bp cutter is the largest centered at approximately 65 kb; the average of the 4-bp cutter is the smallest centered at approximately 256 bp.

Three types of ends can be produced by Type II restriction enzymes including 3′-overhang (protruding), 5′-overhang, and blunt-ended molecules. These are important for the selection of restriction enzymes for cloning, filling-in labeling, or other operations. Proper planning should be made for the most efficient use of restriction endonucleases. In addition, some restriction enzymes do not digest DNA efficiently when the recognition sites are located close to the end of DNA. This is particularly important when incorporating restriction sites into PCR primers for cloning. For more information concerning this, readers are referred to an excellent list from New England Biolabs (http://www.neb.com/nebecomm/default.asp). With more than 250 commercially available and more than 2,000 total, considerations have to be made based on cutting frequency, what types of end they produce, ease of use, and economic considerations. Sources with patent rights and cloned products can be much cheaper than other sources.

In the early 1970s, the discovery of restriction enzymes offered biologists a great tool to cleave huge DNA into smaller pieces for analysis. At the same time, another line of technological advances, the establishment of principles of molecular hybridization using molecular probes, set the foundation for RFLP. The revolution brought about by molecular biology depended heavily on nucleic acid hybridization procedures. These techniques are used extensively in the research laboratory for detecting specific nucleotide sequences in DNA and RNA and are increasingly being applied in medicine for diagnosing diseases. All of the hybridization techniques started with a simple hybridization technique called Southern blot (Southern 1975). A Southern blot is a method in molecular biology of enhancing the result of an agarose gel electrophoresis by marking specific DNA sequences. The method is named after its inventor, the British biologist Edwin Southern. This caused other blot methods to be named as plays on Southern's name (for example, western blot, northern blot, southwestern blot, etc.). All of these blotting techniques require the use of molecular probes.

A probe refers to the agent that is used to detect the presence of a molecule in the sample. For Southern blot, the probe is a DNA sequence that is used to detect the presence of a complementary sequence by hybridization with a DNA sample. Probes are needed to screen for a gene of interest, to determine genomic structure and gene copy numbers, to analyze gene expression, or to validate allelic amplification in PCR.

Probes can consist of DNA, RNA, or antibodies. For DNA, the probes can be double-stranded or single-stranded. The probes can be continuously labeled to make very hot probes or can be end-labeled to trace the segments. Two methods are most frequently used to make continuously labeled probes: (1) Nick translation and (2) Random primer labeling (Sambrook et al. 1989). In the nick translation procedure, double-stranded DNA is nicked with a limited concentration of DNase I. The nicked ds-DNA is a perfect substrate for DNA polymerase I. DNA polymerase I has two major activities: 5′ to 3′ exonuclease activity and 5′ to 3′ polymerase activity. DNA polymerase I makes the new strand DNA with labeled dNTP while degrading the old strand of the DNA. In the random primer labeling procedure, DNA templates are heat-denatured and annealed to short random primers (hexomers), creating a perfect template for Klenow polymerase that makes the new strand with labeled dNTP. DNA synthesis continues until it reaches the next primer.

End-labeled probes can be made by labeling at the 3′ by filling-in reactions using a polymerase or by labeling at the 5′ by using polynucleotide kinase (Sambrook et al. 1989). Probes can be labeled in various other ways. For additional reading, readers are referred to Sambrook and others (1989), or *Current Protocols in Molecular Biology* edited by Fred M. Ausubel, Roger Brent, Robert E. Kingston, David D. Moore, J.G. Seidman, John A. Smith, and Kevin Struhl (2003).

Principles and Molecular Basis of RFLP

The molecular basis of RFLP is summarized in Figure 2.1.

Restriction endonucleases cut DNA wherever their recognition sequences are encountered so that changes in the DNA sequence due to indels, base substitutions, or rearrangements involving the restriction sites can result in the gain, loss, or relocation of a restriction site (Figure 2.1). Digestion of DNA with restriction enzymes

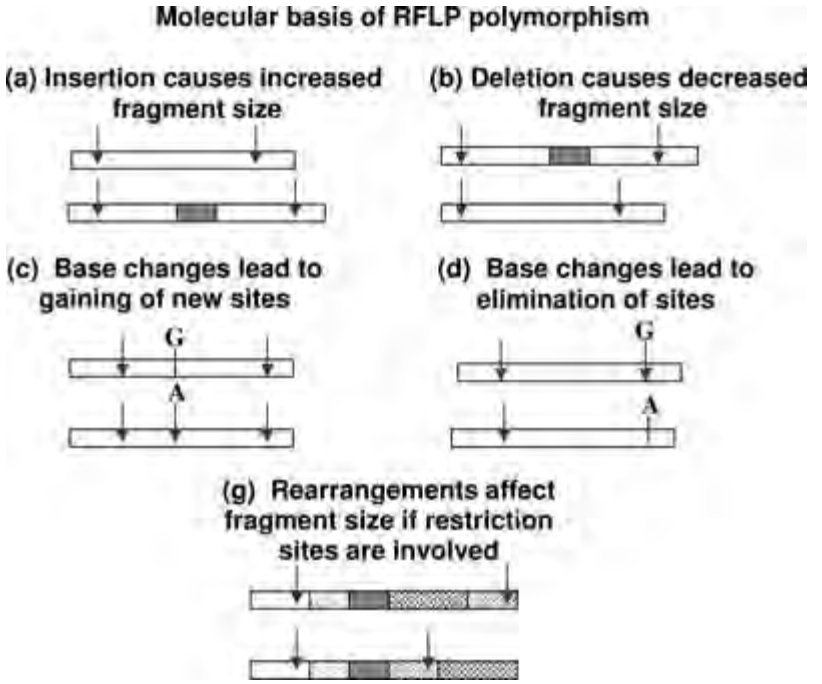

Figure 2.1. Molecular basis of RFLP polymorphism.

results in fragments whose number and size can vary among individuals, populations, and species. As RFLP analysis usually uses agarose gels, only large-size variations can be resolved. In terms of molecular basis, deletion and insertion between restriction sites within the locus of interest generates RFLP; base substitutions at restriction sites within the locus of interest leads to the loss of restriction sites and thus generating larger restriction fragments. Alternatively, base substitutions may lead to the generation of new restriction sites. For instance, the first base of AAATTC (not a restriction site) can mutate to G leading to GAATTC (now a site for *Eco*R1). In cases of rearrangements, the rearranged segments must involve the restriction enzyme sites under consideration to generate RFLP (Figure 2.1).

Two approaches are widely used for RFLP analysis. The first involves the use of hybridization, and the second involves the use of PCR. Traditionally, fragments were separated using Southern blot analysis (Southern 1975), in which genomic DNA is digested, subjected to electrophoresis through an agarose gel, transferred to a membrane, and visualized by hybridization to specific probes. Most recent analysis replaces the tedious Southern blot analysis with techniques based on the polymerase chain reaction (PCR). If flanking sequences are known for a locus, the segment containing the RFLP region is amplified via PCR. If the length polymorphism is caused by a relatively large (>approximately 100 bps depending on the size of the undigested PCR product) deletion or insertion, gel electrophoresis of the PCR products should reveal the size difference. However, if the length polymorphism is caused by base substitution at a restriction site, PCR products must be digested with a restriction enzyme to reveal the RFLP. With the increasing number of 'universal' primers available in the literature,

a researcher can target DNA regions that are either relatively conserved or rapidly evolving, depending on the amount of variation observed and the taxonomic level under examination. Also, PCR products can be digested with restriction enzymes and visualized by simple staining with ethidium bromide due to the increased amount of DNA produced by the PCR method. If the size shift is small, polyacrylamide gels or sequencing gels should be considered rather than agarose gels.

Inheritance of RFLP Markers

RFLP markers are inherited in a Mendelian fashion as codominant markers (Figure 2.2). Both alleles are expressed in molecular phenotypes (here, bands on gels). In the case of an individual heterozygous for two allelic RFLP patterns on alternative chromosomes, the phenotype includes both of the patterns (Figure 2.2). The codominance mode of inheritance is a strength of RFLP markers. In the mapping population, polymorphic RFLP bands segregate in a Mendelian fashion (Figure 2.3).

Differentiating Power of RFLP and Its Strengths and Weaknesses

The potential power of RFLP markers in revealing genetic variation is relatively low compared to more recently developed markers and techniques such as amplified fragment length polymorphism (AFLP) or microsatellites. Indels and rearrangements of regions containing restriction sites are perhaps widespread in the genomes of most species, but the chances of such an event happening within any given locus under study should be rare. Similarly, in a given genome of 10^9 base pairs, approximately 250,000 restriction sites should exist for any restriction enzyme with a 6-bp recognition sequence (that accounts for 1.5×10^6 bp or 0.15% of the entire genome). Base

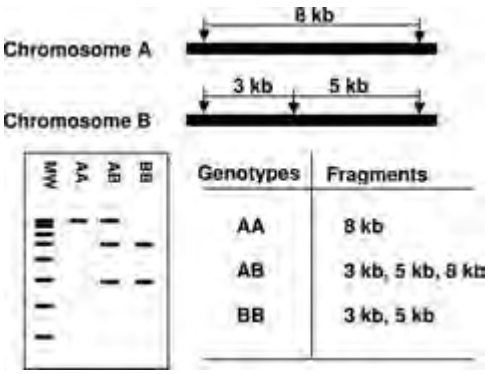

Figure 2.2. Codominant mode of inheritance of RFLP markers. In the example, a base substitution within the 8 kb fragment leads to the gaining of a new restriction site. For homozygous AA, one band of 8 kb should be generated; for homozygous BB, two bands of 3 kb and 5 kb should be generated; for heterozygous AB, three bands of 8 kb (from allele A), 3 kb and 5 kb (both from allele B) should be generated.

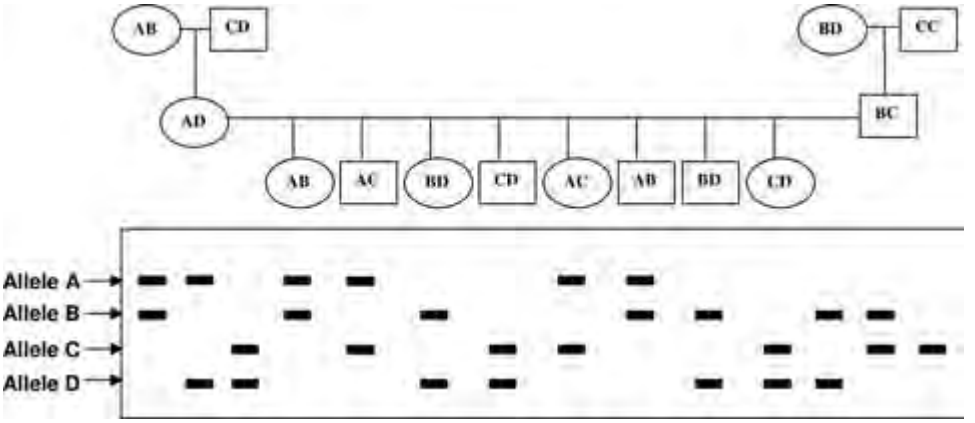

Figure 2.3. Segregation of RFLP markers highlighting codominant mode of inheritance. The first pair of grandparents is both heterozygous (AB and CD) and the second pair of grandparents is homozygous (BB and CC). When the daughter of the first grandparents (AD) mates the son from the second grandparents (BC), four types of segregation are possible: AB, AC, BD, and CD.

substitutions within these restriction sites must be widespread as well, but again, the chances that such base substitutions occur within the locus under study would be relatively small.

The major strength of RFLP markers is that they are codominant markers (i.e., both alleles in an individual are observed in the analysis). Because the size difference is often large, scoring is relatively easy. The major disadvantage of RFLP is the relatively low level of polymorphism. In addition, either sequence information (for PCR analysis) or probes (for Southern blot analysis) are required, making it difficult and time-consuming to develop markers in species lacking known molecular information.

Applications of RFLP in Aquaculture Genomics

RFLP markers are one of the most popular markers used in genetic studies. A search of the PUBMED database using RFLP as a key word led to the generation of 30,000 citations in early 2006. However, much of the popularity of RFLP markers was during earlier decades. Its popularity is reduced now due to availability of other more efficient marker systems. In spite of the popularity of RFLP markers, even in the earlier decades, its application in aquaculture genetics research was limited. (For a recent review, see Liu and Cordes 2004.) In most cases, RFLP markers have been used to differentiate species (Chow et al. 2006, Klinbunga et al. 2005), strains, or populations (Docker et al. 2003, Ohara et al. 2003, Zhang et al. 2005, Aranishi 2005, Sellos et al. 2003, Papakostas et al. 2006, Lehoczky et al. 2005, Apte et al. 2003). Of these studies using RFLP markers, many of them involved the use of mitochondrial DNA or the 16S rDNA (see Chapter 7 as well as Mamuris et al. 2001, Klinbunga et al. 2001, Lopez-Pinon et al. 2002, de los Angeles et al. 2005), which are not highly useful for genomic studies. In consideration of the

availability of several other more efficient marker systems and the relative difficulties involved in the development of RFLP from nuclear genes, the anticipated use of RFLP markers and their significance for aquaculture genome research is limited.

Acknowledgments

Research in my laboratory is supported by grants from the United States Department of Agriculture (USDA) National Research Initiative (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 United States Agency for International Development (USAID), and the Binational Agricultural Research and Development Fund (BARD).

References

Apte S, B Star, and JPA Gardner. 2003. A comparison of genetic diversity between cultured and wild populations, and a test for genetic introgression in the New Zealand greenshell mussel *Perna canaliculus* (Gmelin 1791). Aquaculture, 219, pp. 193–220.

Aranishi F. 2005. Rapid PCR-RFLP method for discrimination of imported and domestic mackerel. Mar Biotechnol, 7, pp. 571–575.

Ausubel FM, R Brent, RE Kingston, DD Moore, JG Seidman, JA Smith, and K Struhl. 2003. Current Protocols in Molecular Biology, Wisley InterScience, New York.

Bossier P, X Wang, F Catania, S Dooms, G Van Stappen, E Naessens, and P Sorgeloos. 2004. An RFLP database for authentication of commercial cyst samples of the brine shrimp *Artemia* spp. (International Study on *Artemia* LXX). Aquaculture, 231, pp. 93–112.

Botstein D, RL White, M Skolnick, and RW Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet, 32, pp. 314–331.

Bouchon D, C Souty-Grosset, and R Raimond. 1994. Mitochondrial DNA variation and markers of species identity in two penaeid shrimp species: *Penaeus monodon* Fabricius and *P. japonicus* Bate. Aquaculture, 127, pp. 131–144.

Chow S, N Suzuki, H Imai, and T Yoshimura. 2006. Molecular Species Identification of Spiny Lobster *Phyllosoma* Larvae of the Genus *Panulirus* from the Northwestern Pacific. Mar Biotechnol, Mar 29 (Epub ahead of print).

Cunningham CO. 2002. Molecular diagnosis of fish and shellfish diseases: present status and potential use in disease control. Aquaculture, 206, pp. 19–55.

de los Angeles Barriga-Sosa I, MY Pérez-Ramírez, F Soto-Aguirre, M Castillo-Rivera, and JL Arredondo-Figueroa. 2005. Inter-specific variation of the mitochondrial r16S gene among silversides, "Peces Blancos" (*Atherinopsidae: Menidiinae*) and its utilization for species identification. Aquaculture, 250, pp. 637–651.

Docker MF, A Dale, and DD Heath. 2003. Erosion of interspecific reproductive barriers resulting from hatchery supplementation of rainbow trout sympatric with cutthroat trout. Mol Ecol, 12, pp. 3515–3521.

Dodgson JB, HH Cheng, and R Okimoto. 1997. DNA marker technology: a revolution in animal genetics. Poult Sci, 76, pp. 1108–1114.

Gross R, K Kohlmann, and P Kersten. 2002. PCR-RFLP analysis of the mitochondrial ND-3/4 and ND-5/6 gene polymorphisms in the European and East Asian subspecies of common carp (*Cyprinus carpio* L.). Aquaculture, 204, pp. 507–516.

Gross R and J Nilsson. 1999. Restriction fragment length polymorphism at the growth hormone 1 gene in Atlantic salmon (*Salmo salar* L.) and its association with weight among the offspring of a hatchery stock. Aquaculture, 173, pp. 73–80.

Hansen MM, K-LD Mensberg, G Rasmussen, and V Simonsen. 1997. Genetic variation within and among Danish brown trout (*Salmo trutta* L.) hatchery strains, assessed by PCR-RFLP analysis of mitochondrial DNA segments. Aquaculture, 153, pp. 15–29.

Klinbunga S, B Khamnamtong, N Puanglarp, P Jarayabhand, W Yoosukh, and P Menasveta. 2005. Molecular taxonomy of cupped oysters (*Crassostrea, Saccostrea*, and *Striostrea*) in Thailand based on COI, 16S, and 18S rDNA polymorphism. Mar Biotechnol, 7, pp. 306–317.

Klinbunga S, D Siludjai, W Wudthijinda, A Tassanakajon, P Jarayabhand, and P Menasveta. 2001. Genetic heterogeneity of the giant tiger shrimp (*Penaeus monodon*) in Thailand revealed by RAPD and mitochondrial DNA RFLP analyses. Mar Biotechnol, 3, pp. 428–438.

Lehoczky I, Z Jeney, I Magyary, C Hancz, and K Kohlmann. 2005. Preliminary data on genetic variability and purity of common carp (*Cyprinus carpio* L.) strains kept at the live gene bank at Research Institute for Fisheries, Aquaculture and Irrigation (HAKI) Szarvas, Hungary. Aquaculture, 247, pp. 45–49.

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

Lopez-Pinon MJ, A Insua, and J Mendez. 2002. Identification of four scallop species using PCR and restriction analysis of the ribosomal DNA internal transcribed spacer region. Mar Biotechnol, 4, pp. 495–502.

Mamuris Z, C Stamatis, KA Moutou, AP Apostolidis, and C Triantaphyllidis. 2001. RFLP analysis of mitochondrial DNA to evaluate genetic variation in striped red mullet (*Mullus surmuletus* L.) and red mullet (*Mullus barbatus* L.) populations. Mar Biotechnol, 3, pp. 264–274.

Morán P, AM Pendás, and E García-Vázquez. 1996. Mitochondrial DNA variation in wild and hatchery brown trout (*Salmo trutta* L.) populations from Spain. Aquaculture, 141, pp. 59–65.

Naish KA, M Warren, F Bardakci, DOF Skibinski, and GR Carvalho. 1995. Use of DNA fingerprinting, RAPD and RAPD/RFLP markers for estimating variation between aquacultural strains of Tilapia (*Oreochromis niloticus*). Aquaculture, 137, pp. 48–49.

Ohara K, T Ariyoshi, E Sumida, and N Taniguchi. 2003. Clonal diversity in the Japanese silver crucian carp, *Carassius langsdorfii* inferred from genetic markers. Zool Sci, 20, pp. 797–804.

Palti Y, JE Parsons, and GH Thorgaard. 1999. Identification of candidate DNA markers associated with IHN virus resistance in backcrosses of rainbow (*Oncorhynchus mykiss*) and cutthroat trout (*O. clarki*). Aquaculture, 173, pp. 81–94.

Papakostas S, S Dooms, A Triantafyllidis, D Deloof, I Kappas, K Dierckens, T De Wolf, P Bossier, O Vadstein, S Kui, et al. 2006. Evaluation of DNA methodologies in identifying Brachionus species used in European hatcheries. Aquaculture, In Press.

Sambrook J, EF Frisch, and T Maniatis. 1989. Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sellos D, J Moal, L Degremont, A Huvet, JY Daniel, S Nicoulaud, P Boudry, JF Samain, and A Van Wormhoudt. 2003. Structure of amylase genes in populations of Pacific Cupped oyster (*Crassostrea gigas*): tissue expression and allelic polymorphism. Mar Biotechnol, 5, pp. 360–372.

Southern E. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol, 98, pp. 503–517.

Zhang Q, SK Allen Jr, and KS Reece. 2005. Genetic variation in wild and hatchery stocks of Suminoe Oyster (*Crassostrea ariakensis*) assessed by PCR-RFLP and microsatellite markers. Mar Biotechnol, 7, pp. 588–599.