Chapter 1

Genome-based technologies useful for aquaculture research and genetic improvement of aquaculture species

Zhanjiang Liu

The Fish Molecular Genetics and Biotechnology Laboratory, Department of Fisheries and Allied Aquacultures and Program of Cell and Molecular Biosciences, Aquatic Genomics Unit, Auburn University, Auburn, Alabama, 36849, USA

e-mail:zliu@acesag.auburn.edu

Table of Contents

References

Abstract

In spite of the late start of aquaculture genome research, significant progress has been made in aquaculture genomics. Many of the genomic resources, tools, and technology developed from genomics have wide applications for aquaculture research and genetic improvements of aquaculture species. The major genome technologies include DNA marker technologies, novel sequencing technologies, gene discovery technologies, genome mapping technologies, and technologies for analysis of genome expression. This chapter was written to provide basic information such as theories behind the technologies, principles of the technologies, and potential applications of each of these genome technologies relevant to aquaculture. In some cases, research and technology development trends are discussed such that the readers are aware of the direction of future research. For instance, single nucleotide polymorphisms (SNP) are clearly the markers of choice in the future even though microsatellites are currently predominant type of markers in aquaculture research and applications; the next generation of sequencing technologies will not only play a dominant role in sequencing, but also may replace the traditional genome expression analysis technologies such as microarrays.

Keywords: Genome, AFLP, Microsatellites, SNP, Marker, QTL, EST, next generation sequencing technology, microarray, linkage mapping, radiation hybrid, comparative mapping, sequence tag

Aquaculture, like animal husbandry, has a long history dating back some 3000 years when the Chinese began culturing carp. However, whereas genetic selection of animal livestock has been conducted for thousands of years, until very recently, most cultured aquatic animals were essentially wild. For thousands of years, aquaculture was limited to the harvesting of immature fish or shellfish and transferring them to an artificially created environment favorable to their growth. In spite of much earlier aquaculture practices in China, the documented artificial production of aquaculture seed was first introduced in 1733 when a German farmer successfully gathered fish eggs, fertilized them, and raised the fish that hatched. Initially this kind of aquaculture was limited to freshwater fish. In the 20th century, new techniques were developed to successfully breed saltwater species. Most brood stocks used in aquaculture were not genetically selected until the late 20th century. In the last 30-40 years, major progress has been made in the genetic enhancement programs of many aquaculture species by using traditional selective breeding techniques, and it is only within the last decade or so that large-scale genome research of aquaculture species started to take off.

Genomics, as a new field, has its own technologies and requires new sets of resources. Let me first clarify the term I used in the title of this chapter, genome-based technologies. The term is used to include technologies required to conduct genome research and technologies derived as a result of genome research for genetic improvement programs in aquaculture. As such, I will include in this chapter DNA marker technologies, sequencing technologies, genome mapping technologies, gene discovery technologies, and genome expression analysis technologies. I will briefly introduce these technologies and briefly discuss their applications in genome research and genetic improvements of aquaculture species. Readers with an interest in

technical details of these technologies are referred to a recently published book, *Aquaculture Genome Technologies* (Liu 2007).

1. DNA marker technologies

A. Historical perspectives

The genome compositions of each individual of the same species are similar but different at the level of DNA sequences and its encoding capacity, and thereby have different transcriptional activities, biological characteristics and performance. The entire task of DNA marker technologies is to provide the means to reveal such DNA level differences of genomes among individuals of the same species, as well as among various related taxa. Historically, these measurements relied on phenotypic or qualitative markers. Morphological differences such as body dimensions, size, and pigmentation are some examples of phenotypic markers. Genetic diversity measurements based on phenotypic markers are often indirect, and inferential through controlled breeding and performance studies (Okumus and Cifci 2003; Parker et al. 1998). Because these markers are polygenically inherited and have low heritability, they may not represent the true genetic differences (Smith and Chesser 1981). Only when the genetic basis for these phenotypic markers is known, can some of them be used to measure genetic diversity. Molecular markers including protein markers and DNA markers were developed to overcome problems associated with phenotypic markers.

Much before the discovery of DNA markers, allozyme markers were used to identify broodstocks in fish and other aquaculture species (Kucuktas and Liu, 2007). Allozymes are different allelic forms of the same enzymes encoded at the same locus (Hunter and Market 1957; Parker et al. 1998; May 2003). Genetic variations detected in allozymes may be the result of

point mutations, insertions or deletions (indels). Allozymes have had a wide range of applications in fisheries and aquaculture including population analysis, mixed stock analysis, and hybrid identification (May 2003). However, they are becoming a marker type of the past due to the limited number of loci that in turn prohibits genome-wide coverage for the analysis of complex traits (Kucuktas and Liu, 2007). In addition, mutation at the DNA level that causes a replacement of a similarly charged amino acid may not be detected by allozyme electrophoresis. Another drawback is that most commonly used tissues in allozyme electrophoresis are the muscle, liver, eye and heart, the collection of which is lethal.

Two specific technological advances, the discovery and application of restriction enzymes in 1973 and the development of DNA hybridization techniques in 1975, set the foundation for the development of the first type of DNA markers, restriction fragment length polymorphism (RFLP, for a recent review, see Liu 2007b). Restriction endonucleases cut DNA wherever their recognition sequences are encountered. Therefore, changes in the DNA sequence due to insertions/deletions (indels), base substitutions, or rearrangements involving the restriction sites can result in the gain, loss, or relocation of a restriction site. Digestion of DNA with restriction enzymes results in fragments whose number and size can vary among individuals, populations, and species. Two approaches are widely used for RFLP analysis. The first involves the use of Southern blot hybridization (Southern 1975), while the second involves the use of PCR. Traditionally, fragments were separated using Southern blot analysis, in which genomic DNA is digested, subjected to electrophoresis through an agarose gel, transferred to a solid support such as a piece of nylon membrane, and visualized by hybridization to specific probes. Most recent analysis replaces the tedious Southern blot analysis with techniques based on 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 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.

The major strength of RFLP markers is that they are codominant markers, i.e., both alleles in an individual are observed in the analysis. The major disadvantage of RFLP is the relatively low level of polymorphism. In addition, either sequence information (for PCR analysis) or a molecular probe (for Southern blot analysis) is required, making it difficult and time-consuming to develop markers in species lacking known molecular information. Due to these disadvantages, the application of RFLP markers in aquaculture and fisheries has been, and will be limited.

Mitochondrial genome evolves more rapidly than the nuclear genome. The rapid evolution of the mtDNA makes it highly polymorphic within a given species. The polymorphism is especially high in the control region (D-loop region), making the D-loop region highly useful in population genetic analysis. The analysis of mitochondrial markers is mostly RFLP analysis, or direct sequence analysis (Liu and Cordes 2004). Due to the high levels of polymorphism and the ease of mitochondrial DNA analysis, mtDNA has been widely used as markers in aquaculture and fisheries settings. However, mtDNA is maternally inherited in most cases, and this non-Mendelian inheritance greatly limits the applications of mtDNA for genome research. In addition, most aquaculture-related traits are controlled by nuclear genes. For most aquaculture finfish species, their nuclear genome is at the level of a billion base pairs while their mitochondrial genomes are usually tens of thousands of times smaller than the nuclear genome. Clearly, in spite of their usefulness for the identification of aquaculture stocks, mitochondrial

DNA markers will not be tremendously useful for aquaculture genome research and genetic improvement programs in aquaculture. However, some recent studies suggested that mitochondrial DNA could influence performance traits such as growth (Steele et al., 2008).

When the Human Genome Project was launched in the mid 1980s, the capacity and capabilities of available DNA marker technologies seriously limited genome research. Such severe limits put pressure to develop more efficient marker systems for analysis of complex traits and genome organizations. At the end of 1980s, the simple sequence repeats (SSR) or microsatellites were discovered; and they have since been used as one of the most preferred marker types because of their high levels of polymorphism, abundance, roughly even genome distribution, co-dominant inheritance, and small locus size that facilitate PCR-based genotyping (Tautz 1989). Because of the importance of microsatellites for aquaculture genome research and genetic improvement programs, I will provide a greater level of discussion below.

At the beginning of 1990s, efforts were also devoted to develop multi-loci, PCR-based fingerprinting techniques. Such efforts resulted in the development of two marker types that were highly popular for a while: RAPD (random amplified polymorphic DNA, Welsh and Mcclelland 1990; Williams et al. 1990) and AFLP (amplified fragment length polymorphism, Vos et al., 1995).

RAPD is a multi-locus DNA fingerprinting technique using PCR to randomly amplify anonymous segments of nuclear DNA with a single short PCR primer (8-10 bp in length) (for a recent review, see Liu 2007c). Because the primers are short, relatively low annealing temperatures (often 36-40ºC) must be used. 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. 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 insertions or deletions (indels) in the regions between the two close primer binding sites. The potential power for detection of polymorphism is reasonably high as compared to RFLP, but much lower than microsatellites; 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. The major advantages of RAPD markers are their applicability to all species regardless of known genetic, molecular or sequence information, relatively high level of polymorphic rates, simple procedure and a minimal requirement for both equipment and technical skills. RAPD has been widely used in genetic analysis of aquaculture species, but its further application in genome studies is limited by its lack of high reproducibility and reliability. In addition, RAPD is inherited as dominant markers and transfer of information with dominant markers among laboratories and across species is difficult.

B. Amplified Fragment length polymorphism (AFLP)

Alternatives of RAPD that overcome the major problems such as its low reproducibility were actively sought. AFLP (Vos et al. 1995) was the outcome of such efforts. AFLP is based on the selective amplification of a subset of genomic restriction fragments using PCR (for a recent review, see Liu 2007d). Genomic DNA is digested with restriction enzymes, and double-

stranded DNA adaptors with known sequences are ligated to the ends of the DNA fragments to generate primer-binding sites for amplification. The sequence of the adaptors and the adjacent restriction site serve as primer binding sites for subsequent amplification of the restriction fragments by PCR. Selective nucleotides extending into the restriction sites are added to the 3' ends of the PCR primers such that only a subset of the restriction fragments is recognized. Only restriction fragments in which the nucleotides flanking the restriction site match the selective nucleotides will be amplified. The subset of amplified fragments is then analyzed by denaturing polyacrylamide gel electrophoresis to generate the fingerprints.

AFLP analysis is an advanced form of RFLP. Therefore, the molecular basis for RFLP and AFLP are similar. First, any deletions and/or insertions between the two restriction enzymes, e.g., between *Eco* RI and *Mse* I that are most often used in AFLP analysis, will cause shifts of fragment sizes. Second, base substitution at the restriction sites will lead to loss of restriction sites and thus a size change. However, only base substitutions in all *Eco* RI sites and roughly 1/8 of *Mse* I sites are detected by AFLP since only the *Eco* RI primer is labeled and AFLP is designed to analyze only the *Eco* RI-*Mse* I fragments. Third, base substitutions leading to new restriction sites may also produce AFLP. Once again, gaining *Eco* RI sites always leads to production of AFLP, gaining *Mse* I sites must be within the *Eco* RI-*Mse* I fragments to produce new AFLP. In addition to the common mechanisms involved in polymorphism of RFLP and AFLP, AFLP also scans for any base substitutions at the first three bases immediately after the two restriction sites. Considering large numbers of restriction sites for the two enzymes (250,000 Eco RI sites and 500,000 Mse I sites immediately next to Eco RI sites for a typical fish genome with one billion base pairs), a complete AFLP scan would also examine over 2 million bases immediately adjacent to the restriction sites.

The potential power of AFLP in the study of genetic variation is enormous. In principle, any combination of a 6-bp cutter with a 4-bp cutter in the first step can be used to determine potential fragment length polymorphism. For each pair of restriction enzyme used in the analysis, e.g., *Eco* RI and *Mse* I, a total of approximately 500,000 *Eco*RI-*Mse* I fragments would exist for a genome with a size of 1×10^9 bp. Theoretically, 4096 primer combinations compose a complete genome-wide scan of the fragment length polymorphism using the two restriction enzymes if three bases are used for selective amplification. As hundreds of restriction endonucleases are commercially available, the total power of AFLP for analysis of genetic variation can not be exhausted. However, it is probably never necessary to perform such exhaustive analysis. Since over 100 loci can be analyzed by a single primer combination, a few primer combinations should display thousands of fingerprints. For genetic resource analysis, the number of primer combinations required for construction of phylogenetic trees/dendrograms depends on the level of polymorphism in the populations, but probably takes no more than five to ten primer combinations.

AFLP combines the strengths of RFLP and RAPD. It is a PCR-based approach requiring only a small amount of starting DNA; it does not require any prior genetic information or probes; and it overcomes the problem of low reproducibility inherent to RAPD. AFLP is capable of producing far greater numbers of polymorphic bands than RAPD in a single analysis, significantly reducing costs and making possible the genetic analysis of closely-related populations. It is particularly well adapted for stock identification because of the robust nature of its analysis. The other advantage of AFLP is its ability to reveal genetic conservation as well as genetic variation. In this regard, it is superior to microsatellites for applications in stock identification. Microsatellites often possess large numbers of alleles, too many to obtain a clear

picture with small numbers of samples. Identification of stocks using microsatellites, therefore, would require large sample sizes. For instance, if 10 fish are analyzed, each of the 10 fish may exhibit distinct genotypes at a few microsatellite loci, making it difficult to determine relatedness without any commonly conserved genotypes. In closely related populations, AFLP can readily reveal commonly shared bands which define the common roots in a phylogenetic tree, and polymorphic bands that define branches in the phylogenetic tree.

The major weakness of AFLP markers is their dominant nature of inheritance. Genetic information is limited with dominant markers because essentially only one allele is scored; and at the same time, since the true alternative allele is scored as a different locus, AFLP also inflate the number of loci under study. As dominant markers, information transfer across laboratories is difficult. In addition, AFLP is more technically demanding, requiring special equipment such as automated DNA sequencers for optimal operations.

AFLP has been widely used in aquaculture such as analysis of population structures, migration, hybrid identification, strain identification, parentage identification, genetic resource analysis, genetic diversity, reproduction contribution, and endangered species protection (Seki et al. 1999; Jorde et al. 1999; Sun et al. 1999; Cardoso et al. 2000; Chong et al. 2000; Kai et al. 2002; Mickett et al. 2003; Whitehead et al. 2003; Mock et al. 2004; Campbell and Bernatchez 2004; Simmons et al. 2006).

AFLP has also been widely used in genetic linkage analysis (Liu et al. 1998; 1999; Kocher et al. 1998; Griffiths and Orr, 1999; Agresti et al. 2000; Robison et al. 2001; Rogers et al. 2001; Liu et al. 2003; Li et al. 2003; William et al. 2005), and analysis of parental genetic contribution involving interspecific hybridization (Young et al. 2001) and meiogynogenesis (Felip et al. 2000). In a study of the black rockfish (*Sebastes inermis*), Kai et al. (2002) used

AFLP to distinguish three color morphotypes, in which diagnostic AFLP loci were identified as well as loci with significant frequency differences. In such reproductive isolated populations, it is likely that "fixed markers" of AFLP can be identified to serve as diagnostic markers. Fixed markers are associated most often with relatively less migratory, reproductive isolated populations (Kucuktas et al. 2002). With highly migratory fish species, fixed markers may not be available. However, distinct populations are readily differentiated by difference in allele frequencies. For instance, Chong et al. (2000) used AFLP for the analysis of five geographical populations of Malaysian river catfish (*Mystus nemurus*) and found that AFLP was more efficient for the differentiation of sub-populations and for the identification of genotypes within the populations than RAPD although similar clusters of the populations were concluded with either analysis.

In spite of its popularity, AFLP has two fundamental flaws that prohibit its wider applications in the future: the dominance inheritance and lack of information to link it to genome sequence information. In some cases, AFLP can be used as a rapid screening tool, and useful markers can then be converted to SCAR (sequence characterized amplified region) markers. However, genome scale applications of SCAR markers are unlikely.

C. Microsatellites

Microsatellites are simple sequence repeats (SSRs) of 1-6 base pairs. The advantages of microsatellites as molecular markers include their abundance in genomes, even distribution, small locus size facilitating PCR-based genotyping, co-dominant nature of Mendelian inheritance, and high levels of polymorphism (for a recent review, see Liu 2007e).

Microsatellites are highly abundant in various eukaryotic genomes including all aquaculture species studied to date. In most of the vertebrate genomes, microsatellites make up a few percent of the genome in terms of the involved base pairs, depending on the compactness of the genomes. Generally speaking, more compact genomes tend to contain smaller proportion of repeats including simple sequence repeats, but this generality is not always true. For example, the highly compact genome of Japanese pufferfish contains 1.29% of microsatellites, but its closely related *Tetraodon nigroviridis* genome contains 3.21% of microsatellites (Crollius et al. 2000). During a genomic sequencing survey of channel catfish, microsatellites were found to represent 2.58% of the catfish genome (Xu et al. 2006; Liu 2007g). In fugu, one microsatellite was found for every 1.87 kb of DNA. For comparison, in the human genome, one microsatellite was found for every 6 kb of DNA (Beckmann and Weber 1992). It is reasonable to predict that in most aquaculture fish species, one microsatellite should exist every 10 kb or less of the genomic sequences, on average.

Dinucleotide repeats are the most abundant forms of microsatellites. For instance, in channel catfish, 67.9% of all microsatellites are present in the form of dinucleotide repeats; 18.5% are present as trinucleotide repeats; and 13.5% as tetranucleotide repeats, excluding mononucleotide repeats which are not nearly as useful for molecular markers. Generally speaking, dinucleotide microsatellites are the most abundant, followed by tri-, or tetranucleotide repeats, but in some cases, tetranucleotide repeats can be more frequent than the trinucleotide repeats. Of the dinucleotide repeat types, $(CA)_n$ is the most common dinucleotide repeat type, followed by (AT)n, and then (CT)n (Toth et al. 2000; Xu et al. 2006). (CG)n type of repeats is relatively rare in the vertebrate genomes. Partially this is because the vertebrate genomes are often A/T-rich. Of the trinucleotide repeats and tetra-nucleotide repeats, relatively A/T-rich

repeat types are generally more abundant than G/C-rich repeat types. Microsatellites longer than tetranucleotide repeats (penta- and hexanucluotides) are much less abundant and therefore, are less important as molecular markers (Toth et al. 2000). It is important to point out that the definition of microsatellites limiting to repeats of six bases long are quite arbitrary. Technically speaking, repeats with seven bases or longer sequences are also microsatellites, but because they become rarer as the repeats are longer, they are less relevant as molecular markers.

Microsatellites are distributed in the genome on all chromosomes and all regions of the chromosome. They have been found inside gene coding regions (e.g. Liu et al. 2001), introns, and in the non-gene sequences (Toth et al. 2000). The best known examples of microsatellites within coding regions are those causing genetic diseases in humans, such as the CAG repeats that encode polyglutamine tract, resulting in mental retardation. In spite of their wide distribution in genes, microsatellites are predominantly located in noncoding regions (Metzgar et al. 2000). Only about 10-15% of microsatellites reside within coding regions (Moran 1993; van Lith and van Zutphen 1996; Edwards et al. 1998; Serapion et al. 2004). This distribution should be explained by negative selection against frameshift mutations in the translated sequences (Metzgar et al. 2000; Li et al. 2004). Because the majority of microsatellites exist in the form of dinucleotide repeats, any mutation by expansion or shrinking would cause frameshift of the protein encoding open frames if they reside within the coding region. This also explains why the majority of microsatellites residing within coding regions have been found to be trinucleotide repeats, though the presence of dinucleotide repeats and their mutations within the coding regions do occur.

Most microsatellite loci are relatively small, ranging from a few to a few hundred repeats. The relatively small size of microsatellite loci is important for PCR-facilitated genotyping.

Generally speaking, within a certain range, microsatellites containing a larger number of repeats tend to be more polymorphic, though polymorphism has been observed in microsatellites with as few as five repeats (Karsi et al., 2002). For practical applications, microsatellite loci must be amplified using PCR. For best separations of related alleles that often differ one another by as little as one repeat unit, it is desirable to have small PCR amplicons, most often within 200 bp. However, due to the repetitive nature of microsatellites, their flanking sequences can be quite a simple sequence as well, prohibiting the design of PCR primers for the amplification of microsatellite loci within a small size limit.

Microsatellites are highly polymorphic as a result of their hypermutability and thereby the accumulation of various forms in the population of a given species. Microsatellite polymorphism is based on size differences due to varying numbers of repeat units contained by alleles at a given locus. Microsatellite mutation rates have been reported as high as 10^{-2} per generation (Weber and Wong 1993; Crawford and Cuthbertson 1996; Ellegren 2000), which are several orders of magnitude greater than that of nonrepetitive DNA (10^{-9}; Li 1997). In several fish species, the mutation rates of microsatellites were reported to be at the level of 10^{-3} per locus per generation: 1.3×10^{-3} in common carp (Zhang et al., 2008), 2×10^{-3} in pipefish (Jones et al. 1999), $3.9–8.5 \times 10^{-3}$ in salmon (Steinberg et al. 2002) and 2×10^{-3} in dollar sunfish (MacKiewicz et al. 2002).

Microsatellites are inherited in a Mendelian fashion as codominant markers. This is one of the strengths of microsatellite markers in addition to their abundance, even genomic distribution, small locus size, and high polymorphism. Genotyping of microsatellite markers are

usually straightforward. However, due to the presence of null alleles (alleles that can not be amplified using the primers designed), complications do exist. As a result, caution should be exercised to assure the patterns of microsatellite genotypes fit the genetic model under application.

The disadvantage of microsatellites as markers include the requirement for existing molecular genetic information, a large amount of up front work for microsatellite development, and tedious and labor intensive nature of microsatellite primer design, testing, and optimization of PCR conditions. Each microsatellite locus has to be identified and its flanking region sequenced for the design of PCR primers. Technically, the simplest way to identify and characterize a large number of microsatellites is through the construction of microsatellite-enriched small-insert genomic libraries (Orstrander et al. 1992; Lyall et al. 1993; Kijas et al. 1994; Zane et al. 2002). In spite of the variation in techniques for the construction of microsatellite-enriched libraries, the enrichment techniques usually include selective hybridization of fragmented genomic DNA with a tandem repeat-containing oligonucleotide probe and further PCR amplification of the hybridization products. In spite of the simplicity in the construction of microsatellite-enriched libraries and thereby the identification and characterization of microsatellite markers, for a large genome project, the real need of direct microsatellite marker development may not be the wisest approach. Recent progress in sequencing technologies with the next generation of sequencers will allow large numbers of genomic sequence tags to be generated that would include numerous microsatellites. Microsatellites can be identified and sequenced directly from genome sequence surveys such as BAC end sequencing (Xu et al., 2006; Somridhivej et al., 2008), and from EST analysis from which many microsatellites can be developed into type I markers (Liu et al. 1999; Serapion et al.

2004). Caution has to be exercised, however, on microsatellites developed from ESTs. First, due to the presence of introns, one has to be careful not to design primers at the exon-intron boundaries. Second, the presence of introns would make allele sizes unpredictable. Finally, many microsatellites exist at the 5'- or 3'-UTR, making flanking sequences insufficient for the design of PCR primers. While introns are not a problem for microsatellites derived from BAC end sequencing, sequencing reactions often terminate immediately after the microsatellite repeats, which also makes flanking sequences insufficient for the design of PCR primers.

Microsatellites have recently become an extremely popular marker type in a wide variety of genetic investigations, as evidenced by the recent debut of the journal *Molecular Ecology Notes*, dedicated almost entirely to publishing primer and allele frequency data for newly-characterized microsatellite loci in a wide range of species. Over the past decade, microsatellite markers have been used extensively in fisheries research including studies of genome mapping, parentage, kinships, and stock structure. The major application of microsatellite markers is for the construction of genetic linkage and QTL maps. This is because of the high polymorphic rate of microsatellite markers. When a resource family is produced, the male and female fish parents are likely to be heterozygous in most microsatellite loci. The high polymorphism of microsatellites makes it possible to map many markers using a minimal number of resource families. There are other reasons for the popularity of microsatellites. One of these is because microsatellites are sequence-tagged markers that allow them to be used as probes for the integration of different maps including genetic linkage and physical maps. Communication using microsatellite markers across laboratories is easy, and the use of microsatellite across species borders is sometimes possible if the flanking sequences are conserved (FitzSimmons et al. 1995; Rico et al. 1996; Leclerc et al. 2000; Cairney et al. 2000). As a result, microsatellites

can be also used for comparative genome analysis. If microsatellites can be tagged to gene sequences, their potential for use in comparative mapping is greatly enhanced.

In spite of the popularity and great utilization of microsatellites, recent advances in molecular markers will have a major impact on the choice of DNA markers. In particular, the rapid progress in single nucleotide polymorphism (SNP) including its rapid identification and automation in genotyping make SNP the far more preferred marker system for genome studies as detailed below.

D. Single nucleotide polymorphism (SNP)

Single nucleotide polymorphism (SNP) describes polymorphisms caused by point mutations that give rise to different alleles containing alternative bases at a given nucleotide position within a locus (for a recent review, see Liu 2007f). Such sequence differences due to base substitutions have been well characterized since the beginning of DNA sequencing in 1977, but genotyping SNPs for large numbers of samples was not possible until several major technological advances in the late 1990s. SNPs are again becoming a focal point of molecular markers since they are the most abundant polymorphism in any organism, adaptable to automation, and reveal hidden polymorphism not detected with other markers and methods. SNP markers have been regarded by many as the markers of choice in the future.

Theoretically, a SNP within a locus can produce as many as four alleles, each containing one of four bases at the SNP site: A, T, C, and G. Practically, however, most SNPs are usually restricted to one of two alleles (quite often either the two pyrimidines C/T or the two purines

A/G) and have been regarded as bi-allelic. They are inherited as codominant markers in a Mendelian fashion.

In spite of its increasing popularity as the choice of markers for the future, SNP discovery is a daunting task. As defined by its definition, single nucleotide polymorphism discovery depends on sequencing. Several approaches have been used for the discovery of SNPs in humans and animals. Earlier efforts used approaches such as SSCP analysis (Gonen et al. 1999), heteroduplex analysis (Sorrentino et al. 1992), and direct DNA sequencing. However, several recently developed approaches provide greater efficiencies, two of which are described here in detail.

The first and the simplest one is to conduct deep sequencing of reduced representation libraries (Altshuler et al. 2000; Van Tassel et al., 2008). In this method, genomic DNA from multiple individuals is mixed and digested with restriction endonuclease and subjected to electrophoresis through an agarose gel. The idea is that the allelic fragments from these individuals (with potential SNPs) should all migrate to the same gel location. The gel slice is cut and the DNA extracted for the construction of the reduced representation library. The reduced representation library is deeply sequenced to generate coverage of 20-30X using next generation DNA sequencers. The generated sequences can then be assembled for the identification of SNPs. Solexa sequencing can be used for species in which a draft genome sequence exists where assembly of the short sequence tags can be achieved. Longer sequences would be needed if a draft genome sequence is not yet available. In most aquaculture species, draft genome sequences are not yet available, and therefore, 454 sequencing should be considered.

The second strategy involves data mining from EST projects, if EST libraries were constructed using multiple individuals (each individual contains two sets of chromosomes so it is

possible also to use just one individual, but greater level of polymorphism is provided by multiple individuals). This approach is realistic because EST resources already exist, or are to be developed for the majority of important aquaculture species. In addition, EST-derived SNPs are coming from genes and therefore are type I markers. Mapping of gene-associate SNPs would allow analysis of association of SNPs with traits for the discovery of the "causing SNPs" for the traits (Bader 2001; Marnellos 2003; Halldorsson et al. 2004; Stram 2004). However, this approach has major limitations. Because evolutionary restraint on mutations in coding regions, SNP rates are generally much lower in coding regions than in non-coding regions. The major problem of EST-derived SNPs could be related high sequence errors in EST sequencing resulting in pseudo-SNPs (Wang et al., 2008). In order to avoid pseudo-SNPs, two factors appeared to be crucial: the contig sizes and the lower sequence allele frequency. Contigs larger than four sequences with the minor sequence being represented for at least twice seem to provide a high level of SNP validation rates (Wang et al., 2008).

SNP genotyping requires special equipment. Many methods have been developed through the years to differentiate the alleles of SNPs. A lot of the earlier methods, in spite of being adaptable to individual laboratory situations, are not suitable for large-scale genome wide applications. These include direct sequencing, single base sequencing (reviewed by Cotton 1993), allele-specific oligonucleotide (ASO, Malmgren et al. 1996), heteroduplex analysis, denaturing gradient gel electrophoresis (DGGE, Cariello et al. 1988), single strand conformational polymorphism assays (SSCP, Suzuki et al. 1990), and ligation chain reaction (LCR, Kalin et al. 1992). Large-scale analysis of SNP markers, however, depends on the availability of expensive, cutting-edge equipment.

Several options are available for efficient genotyping using the state of the art equipment. Particularly popular are methods involving MALDI-TOF (Matrix-assisted laser desorption ionization - time of flight) mass spectrometry (Ross et al. 1998; Storm et al. 2003), and the Beadarray technology developed by Illumina. The latter, as it can be adapted for large scale genome studies, is becoming the most popular SNP genotyping method. The Beadarray technology is based on 3-micron silica beads that self assemble in microwells on either of two substrates: fiber optic bundles or planar silica slides. When randomly assembled on one of these two substrates, the beads have a uniform spacing of ~5.7 microns. Each bead is covered with hundreds of thousands of copies of a specific oligonucleotide that act as the capture sequences in one of Illumina's assays (Figure1). The manufacturing process includes a sequential hybridization of every single array element (bead with oligos). This process, called decoding, allows validation of every feature of every array to ensure that each array element is present and functional.

The allele discrimination at each SNP locus is achieved by using three oligos (P1, P2, and P3, Figure 2), of which P1 and P2 are allele-specific and are Cy3- and Cy5-labeled. P3 is locus-specific designed several bases downstream from the SNP site. Upon allele-specific extension and ligation, the artificial, allele-specific template is created for PCR using universal primers. If the template DNA is homozygous, either P1 or P2 will be extended to meet P3; if the template is heterozygous, both P1 and P2 will be extended to meet P3, allowing ligation to happen (Figure 2). P3 contains a unique address sequence that targets a particular bead type with complementary sequence to the address sequence. After downstream-processing, the single-stranded, dye-labeled DNAs are hybridized to their complement bead type through their unique address sequences. After hybridization, the Beadarray Reader is used to analyze fluorescence signal on the beadchip,

which is in turn analyzed using software for automated genotype clustering and calling (http://www.illumina.com/).

The Illumina's platform has been widely used for large-scale SNP analysis in the Human Genome Project. It is highly cost-effective. Among all the factors, the relatively low cost of the Illumina's genotyping platform is the key for selection of an SNP genotyping platform. Currently, Illumina offers two platforms for the Beadarray technology: the Golden Gate Platform for up to 7600 SNPs, and the iSelect platform for 7,600-60,000 SNPs. On the basis of cost per sample, the iSelect platform is the most efficient with a cost of several cents per sample.

In spite of its current low levels of application in aquaculture genome research, SNP markers should gain in popularity as more and more sequence information becomes available in aquaculture species. Equally important, once the genetic linkage maps are well constructed, genome scans for QTLs are expected to follow to study traits important to aquaculture, which then depends on the use of well-defined association analysis. As SNP markers are great markers for the analysis of trait-genotype associations, their application to aquaculture will become essential. It is clear that SNPs will become the major markers of choice for genome research and genetic improvement programs in aquaculture.

E. Trend of DNA marker technologies

DNA marker technologies become essential for aquaculture genetics research and the genetic improvement of aquaculture species. As a matter of fact, DNA markers, both the quality and quantity, have always been a limiting factor for in-depth genome research. Throughout the years, aquaculture geneticists have used various markers including allozyme markers, mitochondrial

markers, RFLP markers, RAPD, AFLP, microsatellites, and SNPs. The overall trend, however, has been driven by: 1) the need for large numbers of markers for high density coverage of the genomes, and 2) the need for sequence-tagged markers for comparative genome analysis. Such demands have driven aquaculture genetic research away from using systems that do not offer a great number of markers such as RFLP and allozyme markers, and away from anonymous dominant markers such as RAPD and AFLP. Microsatellites, being co-dominant and sequence-tagged, have recently become very popular. However, with the draft genome sequence very soon becoming available for major aquaculture species, microsatellites are not without limitations. Their genotyping can be multiplexed, but the extent of multiplexing is limited. Automation of microsatellite genotyping is limited thus prohibiting large-scale genome wide applications. Mapping of thousands of microsatellites to the genome is a lot of work, and analysis using tens or hundreds of thousands of microsatellites would be a daunting task, if not technically impossible for repeated analysis. This only leaves the SNP marker system to be viable. SNPs are the most abundant in genomes when compared to any other types of markers; SNPs are sequence-tagged and therefore would allow comparative mapping analysis; SNP genotyping is highly automated and therefore is adaptable to large-scale genome wide analysis. Therefore, it is clear that SNP markers are the choice marker of the future. In spite of the current lack of draft whole genome sequences for aquaculture species, it is anticipated that they will soon become available for major aquaculture species. In addition, the availability of next generation sequencing technologies (see below) makes it unnecessary to have the whole genome draft sequences in order to develop large number of SNP markers.

2. DNA sequencing technologies

Two independent DNA sequencing technologies were originally invented in 1977, and they have been referred to as Sanger's enzymatic method, and Maxam-Gilbert chemical method (Sanger et al., 1977; Maxam and Gilbert, 1977). In the last 30 years, most DNA has been sequenced by the enzymatic method as the chemical method never gained popularity because of its use of toxic chemicals in the reactions. However, the Sanger's method, once the golden standard, is rapidly losing ground as the next generation DNA sequencers are now emerging. As the principles and applications of original DNA sequencing technologies, especially those of the Sanger's DNA sequencing technology were well documented and most readers are familiar with them, I will focus this section on the next generation of DNA sequencing technologies. Readers who are interested in the basic principles of the traditional Sanger's sequencing are referred to a chapter in Aquaculture Genome Technologies (Liu, 2007i; 2007j).

Several major new sequencing platforms have been adopted recently and they are collectively referred to as the next generation DNA sequencers. A common feature among the new generation of sequencing procedures is the elimination of the need to clone DNA fragments and the subsequent amplification and purification of DNA templates prior to sequencing. Instead, sequence templates are handled in bulk, and massively parallel sequencing allows the generation of numerous sequences simultaneously. Readers need to know that many sequencing platforms are being developed, and this field is one of the most active areas in technology development. Here I will focus only on the principles of three sequencing platforms: the SOLiD sequencing platform, and the Solexa sequencing platform, and the 454 sequencing platform (Table 1).

A. The SOLiD sequencing platform

The SOLiD (Sequencing by Oligonucleotide Ligation and Detection) method utilizes ligation of fluorescently labeled 8-mer primers containing random bases at six of its eight positions and specific dinucleotides at the remaining two positions (the earliest version was the fourth and the fifth, but it can be the first and second, and I will use that for explaination here). The primers are fluorescently labeled with four specific dyes with each dye corresponding to four specific dinucleotides. Note that at each DNA base position, there are four base possibilities: A, C, G, or T. For dinucleotide at two consecutive base positions, a total of 16 dinucleotides (AA, AC, AG, AT, CA, CC, CG, CT, GA, GC, GG, GT, TA, TC, TG, and TT) should cover all possibilities. In order to make each of the four fluorescent dyes to uniquely represent one specific nucleotide, a two step decoding process is required. A random primer is ligated to the template only when the first and second nucleotides on the primer are complementary to those on the template. After visualizing the color, the fluorescent tag is removed by cleaving the primer between the fifth and sixth positions, removing bases 6, 7, and 8. The process is repeated; and in the second round of ligation, a new primer is ligated with the new first and second nucleotides on the primer being complementary to those on the template, counting from the end of the previously ligated primer after clevage (base 6 and 7 of the previous primer). Repeat this process and every first and second positions are recorded. Next, the system is reset to generate the recording for every n-1, n-2, n-3, and n-4 positions (Kate Marusina, http://www.genengnews.com/articles/chitem.aspx?aid=1946) such that each base included in the dinucleotide is "sequenced" twice to allow base calling. For example, the base at position 4 is

sequenced in the dinucleotide base 3 and base 4, and it is sequenced again in the dinucleotide base 4 and base 5.

As all next generation sequencing platforms, the SOLiD sequencing does not require cloning of genomic DNA. It starts with the creation of a "library" by ligation of two adaptors to sheared genomic DNA (Figure 3). Once the adapters are ligated to the library, emulsion PCR is conducted using the common primers to generate "bead clones" in which each contains a single nucleic acid species. Each bead is then attached to the surface of a flow cell (microscope slide) via 3' modifications to the DNA strands (Figure 4). Each microbead can be considered a separate sequencing reaction which is monitored simultaneously via sequential digital imaging.

SOLiD sequencing chemistry depends on specific ligation of a random primer to the existing primer only when the random primer harbors the specific dinucleotide that is complementary to the template DNA being sequenced. The actual base detection is no longer done by the polymerase-driven incorporation of labeled dideoxy terminators. Instead, SOLiD uses a mixture of labeled oligonucleotides and queries the input strand with ligase. In the early version, each oligo has degenerate nucleotide at positions 1-3 (N's at 3' first three bases), one of 16 specific dinucleotide at positions 4-5, and degenerate nucleotide at positions 6-8 that are fluorescently labeled (demonstrated in the figure 3 with the first and second nucleotide being specific, and the remaining bases being degenerate).. The sequencing reaction involves: 1) Hybridization and ligation of a specific oligo whose 4th & 5th bases match that of the template; 2) Detection of the specific fluor associated with the specific dinucleotide; 3) Cleavage of bases 6-8; 4) Repeat, this time querying the 9th & 10th bases. Seven cycles of ligation would allow putative nucleotide identities at positions 4 and 5, 9 and 10, 14 and 15, 19 and 20, 24 and 25, 29 and 30, and 34 and 35 to be recorded. After 7 cycles of this, a "reset" is performed in which the

initial primer and all ligated portions are melted from the template and discarded. Next a new initial primer is used that is N-1 in length. Repeating the initial cycling (steps 1-4) now generates an overlapping data set.   In this manner, ligations using primer N generate sequences for bases 4 and 5, 9 and 10, 14 and 15, 19 and 20, 24 and 25, 29 and 30, and 34 and 35; ligations using primer N-1 generate sequences for bases 3 and 4, 8 and 9, 13 and 14, 18 and 19, 23 and 24, 28 and 29, and 33 and 34, and so on. After use of primers N, N-1, N-2, N-3, and N-4, every base is "sequenced" twice using two primers (Table 2). Base calling is dependent on a two-step encoding procedure with known nucleotides in every fourth and fifth position of the primer (Figure 5). For example, the dinucleotides CA, AC, TG, and GT are all encoded by the green dye. Because each base is queried twice, it is possible, using the two colors, to determine which bases were at which positions. For instance, if the sequence is TCGAACGTA…blue label is detected for the first ligation reaction using primer N that determines base composition at base 4 and 5. There are still four possibilities with blue label: AA, CC, GG, and TT. However, in the first ligation reaction of primer N-1, yellow label is detected, which effectively calls the base at position 4 to be "A". Similarly, with the second ligation reaction using N-4 primer (that determines base compositions at base 5 and 6), green should be detected that effectively calls the base at position 5 to be "A" as well. Repeating this decoding process to include "sequencing" of each base with two primers will provide unambiguous base callings.

As with any other type of technology, SOLiD sequencing technology is making rapid improvements allowing greater throughput and more applications. Several areas are intensely being worked out to increase the efficiency and accuracy of the technology including: increasing the bead density, read lengths, and ability of multiplexing. Currently, SOLiD read length is around 35 bases, and it is anticipated to be 50 bases in 2009.

The SOLiD sequencing technology has wide applications including, but not limited to targeted resequencing, gene expression analysis by digital counting of sequence tags (e.g., Cloonan et al., 2008), microRNA discovery, chromatin immune-precipitation (ChIP), and whole genome sequencing. Due to its short read length, it is anticipated that its application to whole genome sequencing will be mostly applicable to whole genome re-sequencing. However, it is less amenable to *de novo* whole genome sequencing projects because the assembly of whole genome based on short sequence tags has proven to be a great challenge. Because the technology is quite new, there are no publications of using the SOLiD sequencing platform in aquaculture species, to the author's best knowledge.

B. The Solexa sequencing platform

Solexa sequencing platform depends on two of its core technologies: the Clonal Single Molecule Array™ technology that allows simultaneous analysis of hundreds of millions of individual molecules, and the reversible terminator technology that allows specific base calling based on sequencing by synthesis.   Solexa sequencing starts with sheering of genomic DNA to small segments to which different adaptor sequences are ligated to either end. Upon Binding single-stranded template DNA fragments with adaptors randomly to the inside surface of the flow cell channels, template DNA is amplified clonally through bridge PCR. DNA is sequenced by detection of fluorescently labeled dideoxynucleotide terminators. These specially created nucleotides, which also possess a reversible termination property, allow each cycle of the sequencing reaction to occur simultaneously in the presence of all four nucleotides (ddA, ddC, ddT, ddG). In the presence of all four nucleotides, the polymerase is able to select the correct

30

base to incorporate, with the natural competition between all four alternatives leading to higher accuracy than methods where only one nucleotide is present in the reaction mix at a time (which require the enzyme to reject an incorrect nucleotide). Sequences where a particular base is repeated one after another ("homopolymer repeats") are dealt with as for any other sequence and with high accuracy; this avoids the problems of measuring intensity and deducing how many bases were present in the repeat that are the cause of uncertainty seen with "one base per reaction" methods, as described with the 454 sequencing platform.

Solexa sequencing technology achieves an unparalleled data density with highly accurate results. A typical Solexa sequencing run generates millions of sequence tags with the capability to generate over a billion bases of DNA sequence per run. Currently, the read length is limited to about 35 bp. Like the SOLiD sequencing technology, the Solexa sequencing technology has wide applications including, but not limited to gene expression analysis by digital counting of sequence tags, microRNA discovery, epigenetic studies, chromatin immune-precipitation (ChIP), and whole genome sequencing (Bentley, 2006; Dolan and Denver, 2008; Glazov et al., 2008; Butler et al., 2008; Chen et al., 2008; Cokus et al., 2008; Hillier et al., 2008). However, due to its short read length, it is anticipated that its application to whole genome sequencing will be mostly applicable to whole genome re-sequencing. As Solexa technology is also quite new, its application in aquaculture species is still limited, but the potential is tremendous. One of the earliest applications of the Solexa technology was actually conducted in oysters. Dennis Hedgecock's group used the Solexa technology for the study of genes involved in heterosis of pacific oysters (Hedgecock et al., 2007).

C. The 454 Sequencing Platform:

The 454 Sequencing is based on pyrosequencing. During DNA synthesis, a pyrophosphate (PPi) is released when each base is incorporated. The released pyrophosphate can be converted to ATP that generates a fluorescent signal upon the actions of luciferase in the presence of its substrates. Measurement of the light signal after sequential injection of A, C, G, and T would allow the determination of the base composition. After each base addition, the whole sequencing reaction system is reset by cleaning out all existing ATP and nucleotides with apyrase. When mononucleotide repeats are encountered in the sequence, the pyrosequencing reaction continuously incorporates the repeated nucleotide until it reaches a different nucleotide. The light signal produced is proportional to the number of mononucleotides incorporated-up to 8 bases. Mononucleotide repeats greater than 8 bp cannot be accurately sequenced by pyrosequencing.

The 454 sequencing platform uses microfabricated high-density picolitre reactors (Margulies et al. 2005). No cloning is necessary for 454 sequencing. The clonal DNA used for sequencing is obtained by clonal PCR amplification of a single molecule in emulsified water-in-oil microreactors. Preparation of the DNA library consists of a few simple steps. Genomic DNA is fractionated into smaller fragments (300-500 base pairs) that are subsequently filled in to polished ends (blunted), allowing ligation of adaptors to the genomic DNA for PCR amplification. In order to prevent intermolecular ligation of the genomic DNA fragments, the DNA fragments are dephosphorylated. Short Adaptors (A and B) are then ligated onto the ends of the fragments. After ligation, the gap needs to be repaired, presumably using a DNA ligase. The adaptors provide priming sequences for both amplification and sequencing of the sample-library fragments. The two adaptors are different. Adaptor B contains a 5'-biotin tag that enables

immobilization of one strand of the library onto streptavidin coated beads. The non-biotinylated strand is released and used as a single-stranded template DNA library.

The single-stranded template DNA library is immobilized onto beads carrying short primers complementary to the Adaptor A sequences by base pairing. The key element here is attaining the correct proportion of beads to DNA molecules such that only one molecule is captured by each bead. The beads containing a single molecule of the single-stranded template are emulsified with the amplification reagents in a water-in-oil mixture. Each bead is captured within its own microreactor where PCR amplification occurs. This results in bead-immobilized, clonally amplified DNA fragments.

The single strand template DNA library beads are added to the DNA Bead Incubation Mix (containing DNA polymerase) and are layered with enzyme beads (containing sulfurylase and luciferase) onto the PicoTiterPlate device. The device is centrifuged to deposit the beads into the wells. The layer of enzyme beads ensures that the DNA beads remain positioned in the wells during the sequencing reaction. Due to the size of the wells in relation to the beads, only one bead containing a specific clonally amplified genomic DNA segment should be placed into each well of the PicoTiterPlate device. The loaded PicoTiterPlate device is placed into the "454 sequencer," the Genome Sequencer 20 Instrument or the FLX generation systems, which performs pyrosequencing-like reactions. Unlike a traditional pyrosequencing reaction, hundreds of thousands of beads, each with millions of copies of clonally amplified DNA, are sequenced in parallel. Each well of the PicoTiterPlate device is a separate pyrosequencing reaction. If a nucleotide complementary to the template strand is flowed into a well, the polymerase extends the existing DNA strand by adding a nucleotide(s). Addition of one (or more) nucleotide(s) results in a reaction that generates a light signal that is recorded by the CCD camera in the

instrument. The signal strength is proportional to the number of nucleotides incorporated in a single nucleotide flow. Typically, over 200,000 reads can be achieved in a single run. Assuming generation of 250 bp by a single reaction, each run should generate 50 million bp or more of sequence in several hours using a single instrument. This is approximately a 10X coverage of a bacterial genome!

The 454 sequencing platform holds much potential for its great applications such as gene expression profiling, epigenetic analysis, and whole genome sequencing (Patrick, 2007; Bekal et al., 2008; Vera et al., 2008; Hafner et al., 2008). The 454 sequencing platform is probably the most promising technology for *de novo* whole genome sequencing among the next generation sequencers as it produces sequences of approximately 250 bp that is significantly shorter than the traditional Sanger sequencing (800-1000 bp), but much longer those generated by SOLiD or Solexa sequencing platforms. The current major problems prohibiting its application for the sequencing of complex genomes are its relatively short sequencing reads and difficulties in accurate determination of homopolymeric runs in the DNA. The short reads complicate genome sequence assembly, while the inability to determine the number of bases within a long homopolymeric run prohibits accurate sequencing of genomes. These problems are more significant for complex genomes with high levels of repeat structure. However, the technologies' high throughput and low costs are very attractive, especially for aquaculture species. As the technology is perfected to minimize these drawbacks, the 454 sequencing platform will show even greater promise.

3. Gene discovery technologies

Performance and production traits are controlled by genes, environments, and gene-environment interactions. In order to gain detailed understanding of performance and production traits, understanding the genes in aquaculture genomes become essential. Sequencing of expressed sequence tags (ESTs) has been the primary approach for the discovery of genes in aquaculture species, although several other approaches are also available such as serial analysis of gene expression (SAGE). Recently, however, the adoption of several novel sequencing platforms using next generation sequencers has allowed generation of expressed sequence tags through de novo sequencing of whole transcriptomes.

A. Expressed sequence tags and gene discovery

ESTs are single pass sequences of random cDNA clones. They are partial cDNA sequences corresponding to mRNAs generated from randomly selected cDNA library clones (for recent reviews, see Liu 2006; 2007h). EST analysis has traditionally been conducted by sequencing random cDNA clones from cDNA libraries. Such an approach is efficient at initial stages of gene discovery, but has proven to be inefficient in the gene discovery of rarely expressed genes. The rate of gene discovery usually drops precipitously soon after reaching a level of several thousand ESTs. By using regular cDNA libraries, the most abundantly expressed genes would have been sequenced many times before the most rarely expressed genes are sequenced just once. Clearly, EST sequencing from non-normalized libraries is inefficient for gene discoveries of rarely expressed genes. Normalization decreases the prevalence of clones representing abundant transcripts and dramatically increases the efficiency of random sequencing and rare gene discovery.

Normalized cDNA libraries are cDNA libraries that have been equalized in representation to reduce the representation of abundantly expressed genes and to increase the representation of rarely expressed genes. While the details of how the subtraction is conducted may differ greatly, the basic principles behind normalization are the same, i.e. they all depend on the faster hybridization kinetics of abundantly expressed genes to form double-stranded complexes that can be removed by various means, whereas it takes a long time for the rarely expressed genes to reassociate.

Several strategies have been developed for the normalization of cDNA libraries, but the fundamental principles behind all the normalization procedures are the same. We have used a strategy utilizing the Evrogen TRIMMER DIRECT Kit (http://www.evrogen.com/p3_2.shtml). This system is specially developed to normalize cDNA enriched with full length sequences (Zhulidov et al. 2004). The method involves denaturation-reassociation of cDNA, degradation of ds-fraction formed by abundant transcripts, and PCR amplification of the equalized ss-DNA fraction. The key element of this method is degradation of ds-fraction formed during reassociation of cDNA using Duplex-Specific Nuclease (DSN) enzyme (Shagin et al. 2002). A number of specific features of DSN make it ideal for removing ds-DNA from complex mixtures of nucleic acids. DSN displays a strong preference for cleaving ds-DNA in both DNA-DNA and DNA-RNA hybrids, compared to ss-DNA and RNA, irrespective of the sequence length. Moreover, the enzyme remains stable over a wide range of temperatures and displays optimal activity at 55-65°C. Consequently, degradation of the ds DNA-containing fraction by this enzyme occurs at elevated temperatures, thereby decreasing loss of transcripts due to the formation of secondary structures and non-specific hybridization involving adapter sequences.

EST analysis is one of the most rapid approaches for gene discovery. A small collection of ESTs in a species without any genome information can result in the rapid identification of a large number of genes. Gene discovery and identification is, therefore, the primary function of EST analysis. Because of the exceptionally high gene discovery rate of the EST approach, EST analysis has been extremely popular. The EST database dbEST has been one of the fastest growing databases at NCBI. As of May 30, 2008, there are 52,858,766 entries in the NCBI's public EST database dbEST (dbEST release 053008, http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html).

Large EST resources are available for several major aquaculture species including Atlantic salmon, rainbow trout, catfish, oysters, and shrimps. ESTs provide the information and material basis for the development of microarrays for the analysis of genome expression as discussed below.

B. *de novo* sequencing of whole transcriptomes and gene discovery

As detailed under sequencing technologies, all next generation sequencing platforms have the ability to generate hundreds of thousands (the 454 platform) to millions of expressed sequence tags (SOLiD and Solexa sequencing platforms) using RNA as the starting material. Clearly, such sequencing projects not only allow many genes to be identified, and more importantly allow expression profiling through digital counting of sequence tags, as will be further discussed under genome expression analysis section below. As compared to traditional EST analysis, the next generation sequencing technologies can rapidly produce a large number of expressed sequence tags. They probably provide a much greater power in terms of expression profiling as accurate

estimation can be made based on the number of sequenced tags. In addition, relatively accurate counting of tags from various exons may also provide information concerning alternative splicing and alternative polyadenylation. However, it may be more difficult for the identification of short sequence tags for less conserved genes as there is no draft genome sequences are yet available from aquaculture species. This problem will soon be alleviated when the draft genome sequences become available. In comparison, the 454 sequencing platform perhaps provide a greater power for sequence identification because of its longer sequence reads, but at the expense of the number of sequence tags to be generated that would otherwise offer greater accuracy for expression profiling based on tag counting.

4. Genome mapping technologies

Genome mapping and sequencing is the core of structural genomics. While no whole genome draft sequences are yet available with aquaculture species, genomes of many aquaculture species have been subjected to mapping. Genome mapping can be classified into several categories based on resource needs, technology requirement, and principles used in mapping, i.e., genetic linkage mapping, QTL mapping, physical mapping, cytogenetic mapping, radiation hybrid mapping, and comparative mapping (Liu 2007a). While cytogenetic mapping is highly useful for the identification of chromosomes and physical mapping of genes to chromosomes, its resolution is low for large-scale genome wide mapping analysis. It is well covered in the chapter written by Dr. Gideon Hulata in this volume, so I will not cover it further here. Limited work has been conducted in comparative mapping of aquaculture species. A recent chapter in Aquaculture Genome Technologies written by Dr. Thomas Kocher provided excellent coverage

on comparative mapping and positional cloning (Lee and Kocher 2007). Interested readers are referred to the chapter for details on comparative mapping.

A. Genetic linkage mapping

Genetic linkage mapping is an old technology. It is purely based on co-segregation of markers within a well-defined segregating population that is often referred to as the reference family or mapping population. Genetic mapping requires two major resources: the segregating population, and the molecular markers. The segregating population can be F2 population, backcross progenies, or higher generation of intercrosses. In some cases, F1 can be used and treated as a pseudo backcross because heterozygous markers are segregating in F1 population. However, for best results, three generation pedigrees are needed to provide non-ambiguous linkage phases. The number of individuals used for mapping analysis depends on the desired resolution. A population of 100 individuals can resolve markers that are 1 cM away (1 out of 100 are recombinant). If higher resolution is needed to detect rare recombinants, larger numbers of the mapping population can be used.

Any molecular markers that are polymorphic and segregating in the mapping population in a Mendelian fashion can be used for linkage mapping. However, sequence tagged markers provide co-dominance and transferability across laboratories and possibly also across species borders allowing comparative mapping analysis (see below). The most often used marker type for linkage mapping is microsatellite, although SNP markers may soon dominate on linkage maps. As the advantages and disadvantages of various types of markers are detailed above, I will not repeat them here. Genetic linkage maps can be constructed upon analysis of marker

segregation data within the mapping population. This is achieved by use of various software packages.

Linkage maps have been constructed in over 30 major aquaculture species (Table 3). A recent review by Danzmann and Gharbi provided excellent details on linkage mapping, and summarized recent progress of linkage mapping in aquaculture species (Danzmann and Gharbi, 2007). It is clear, however, the marker densities on the genetic linkage maps of aquaculture species are too low currently to provide sufficient coverage for efficient QTL analysis. The use of more markers, in particular the SNPs to be developed, should very soon change the situation.

B. QTL mapping

Most performance and production traits are controlled by multiple genes, and therefore, simple Mendelian genetic analysis is not sufficient to provide answers as too how many genes are controlling the traits, and how they function. The multi-gene controlled traits are defined as quantitative traits and the loci controlling the traits are defined as quantitative trait loci (QTL). QTLs can be mapped genetically by correlation of segregating markers with the traits, and such a process is referred to as QTL mapping.

QTL mapping starts with creation of a population in which the traits and the markers are segregating. For instance, fish resistant to a particular disease can be crossed with susceptible fish to produce F1 fish. F2 fish can be produced from the F1 fish in which the disease trait is segregating. In a sense, QTL mapping is not any different from genetic linkage mapping except that the proper mapping population is needed in which the trait of interest is segregating along

with DNA markers. QTL mapping is the core of aquaculture genomics as the ultimate practical goal of aquaculture genomic research is to provide tools for genetic improvements.

Much progress has been made in QTL mapping in aquaculture species. A recent review by Korol et al. (2007) covered many of the examples of QTL mapping studies in aquaculture species, and interested readers are referred to this review. As a whole, QTL mapping in aquaculture species has fallen behind, and greater efforts should be devoted to this area.

The practical application of QTL mapping is marker-assisted selection. There are only few examples of marker-assisted selection in aquaculture species (e.g., Sakamoto et al., 2007). Before the wider applications of marker-assisted selection are developed, the trend is that selection will soon become whole genome-based (Meuwissen *et al.* 2001), as has already occurred in terrestrial livestock species. With whole genome selection, selection is performed on estimates of associations of phenotype with largest possible markers across the genome. This contrasts with the traditional marker-assisted selection which is based on a small number of significant markers, thus limiting overall effectiveness. The application of whole genome selection involves using "training data" to estimate "breeding values" of SNP haplotypes or alleles.

C. Radiation hybrid mapping

The concept of radiation hybrid mapping was initially derived from somatic cell hybrids. Back in the 1970s, technology was developed to fuse different types of cells to form hybrid cell lines. In 1990, Cox et al. resurrected the technology by fusing X-ray radiated cells (donor cells) with normal cells (recipient cells). As the chromosomes of the X-ray radiated cells were broken, the

cells can not survive by themselves. However, upon fusion with a recipient cell, broken chromosomal segments from the donor cells can be fused into the recipient cells. In order to have a selection marker for the hybrid cells, a drug resistant gene (e.g., neo) can be first inserted into the donor cell's genome. Upon fusion of the cells, the selection drug G418 can be applied to select for hybrid cells as the recipient cells do not have the drug resistance and should be killed by the antibiotic. Fusion cells containing the neo gene would be selected for growth. Clonal expansion of such fused hybrid cells would create a panel of radiation hybrid cell lines with each containing a different segment of the broken chromosomes, along with the chromosomal segments containing the neo gene. Because radiation causes random chromosomal breakage, various chromosomal segments contained within the panel of radiated cell lines would collectively cover the entire genome, with some portions of the chromosomal segments overlapping one another. Such radiation panels are used to map the genome through radiation hybrid mapping RH mapping).

RH mapping is based on co-retention of markers. A recent chapter in Aquaculture Genome Technologies written by Dr. Caird Rexroad provides an excellent explanation for RH mapping (Rexroad, 2007), in which he wrote: "RH mapping strategies are based on the concept that markers which are close together on chromosomes will frequently be co-retained in the same hybrids – the probability that irradiation will induce a chromosome break between two markers decreases as the physical distance between the two markers decreases". To provide adequate statistical support for mapping, marker retention frequencies - the percentage of times a marker is scored positive in a RH panel, is critical. Optimal retention is 20-50% (Walter & Goodfellow 1993). RH mapping is calculated based on the co-retention of markers in fragments across the panel. The estimated frequency of breakage between two markers is θ, which ranges from 0 to 1

and is analogous to recombination frequencies (r) used in genetic mapping. A θ value of 0 means two markers are always co-retained, a value of 1 means they are co-retained at random. This raw value is then included in multipoint analyses and transformed into centiRays (cR) - the RH map unit - using map functions similar to those of Haldane or Kosambi which are used in genetic map construction. Hence, observation of chromosome breaks between two markers in RH mapping is analogous to observing recombination between two markers in genetic mapping. In fact, the term "linkage" is often used in RH mapping. The frequency of chromosome breaks between two markers is not only due to their physical distance, but also to the intensity of the radiation used to create the panel. Siden and colleagues conducted experiments to observe the effects of different dosages of radiation on a segment of the human X chromosome (Siden et al. 1992). At 5000 rad 10% of the clones retained the entire chromosome arm, 40% had fragments of 3-30 MB, and 50% had fragments less than 3 MB. At 25,000 rad only 6% had fragments larger than 3 Mb. Therefore the radiation hybrid map-distance unit is annotated with a subscript stating the dosage used to create the panel in rads, i.e. cR_{3000}. Retention of multiple fragments from a single chromosome in a hybrid cell line complicates analyses; therefore $100 - 300$ cell lines must be scored for a panel to construct statistically significant maps.

RH mapping was initially created to map the non-polymorphic markers. Back in mid-1990s, polymorphic markers were limiting in most species. By fusing cells of the interest species to a rodent cell line, the background genes are usually not amplified by using PCR primers designed from the genes of the species of interest. Thus ESTs were mapped to the genome maps. Later, however, it was found that RH maps were critically important for guiding the whole genome assembly as they are essentially physical linkage maps.

RH mapping was mostly applied to mammalian genome mapping, but less so in aquatic species. However, it has been used for the zebrafish, an aquatic model organism used to study the genetics of development, growth, reproduction, and disease resistance (Kwok et al. 1999; Geisler et al. 1999; Hukriede et al. 1999). To date, the only aquaculture species in which RH maps have been constructed is the gilthead sea bream (Senger et al. 2006; Sarropoulou et al., 2007). One reason is that higher resolution can be achieved by bacterial artificial chromosome (BAC)-based physical maps. In addition, the need to map non-polymorphic markers is drastically reduced as now many polymorphic markers become available, especially the SNP markers. Therefore, the application of RH mapping in aquaculture species is limited.

D. BAC-based physical mapping

Bacterial artificial chromosome (BAC)-based physical maps are important for the understanding of genome structure and organization, and for position-based cloning of economically important genes. A well characterized physical map can often be an important foundation for whole genome sequencing. A BAC-based physical map would also allow exploitation of existing genomic information from map-rich species using comparative mapping, thus accelerating genome research in the species of interest. The first step of BAC-based physical mapping is the construction of large-insert BAC libraries.

BAC libraries are large insert genomic libraries. A recent chapter in Aquaculture Genome Technologies (He et al., 2007) provided excellent technical details on construction and characterization of BAC libraries, and related chapters by Davidson (2007) and Xu et al. (2007a) had a great coverage on details of physical mapping. Interested readers are referred to these

chapters. Briefly, large inserts contained in a BAC library were derived from multiple copies of

the genome broken randomly by partial restriction enzyme digest. Therefore, the BAC library

can be viewed as multiple genome copies broken randomly into segments that are overlapping

one another. Because the genomic segments originally from the same genome locations harbor

the same restriction sites, overlapping genome segments can be aligned by the presence of the

same sets of restriction fragments. A set of overlapping genome segments aligned by

overlapping restriction fingerprints are then defined as a contig [a contig (from *contiguous*) is a

set of overlapping DNA segments derived from a single genetic source], and many contigs make

up the entire genome, possibly with gaps (or short overlapping segments, not supporting

statistical overlapping status).

The efforts of making BAC contig-based physical maps in aquaculture species is a recent

event. As a result, physical maps have been only constructed in Atlantic salmon (Ng et al.,

2005), tilapia (Katagiri et al., 2005), and channel catfish (Quiniou et al., 2007; Xu et al., 2007b),

and a physical map is under construction in rainbow trout (Yniv Palti, USDA ARS, personal

communication). It is expected that physical maps will soon be constructed in many of the

important aquaculture species.

5. Genome expression analysis technologies

The development of high throughput technologies for global or genome-wide measurements of

gene expression requires the availability of genome resources. Most often, such genome

resources come in the form of whole gene sequencing or the availability of a large resource of

expressed sequence tags such that the major fraction of the transcriptome is represented. Once

the sequences representing all the genes of an organism or the vast majority of the genes of the

organism are known, global or genome-wide measurements of gene expression can be made using either microarray technologies or sequence tag-based technologies.

A. Microarray technology

While microarrays utilize several recent technological innovations, they are, at their core, simply a high density dot blot. There are two primary approaches to microarrays, differing in both their construction and their sample labeling. Spotted arrays are constructed by spotting long oligos or cDNAs using a printing robot, whereas *in situ* arrays are constructed by synthesizing short oligos directly onto the slide by photolithography (for a detailed review, see Peatman and Liu, 2007).

Spotted array technology encapsulates the printing of either PCR products or long oligos (60-70 mers). Traditionally referred to as cDNA arrays, spotted arrays are today just as likely to be long oligos, as the cost of synthesizing oligos continues to decline, and because the parallel PCR required to prepare for cDNA arrays is labor-intensive, costly, and requires having clones on hand. While these cDNA-associated difficulties can be overcome through hard work and collaboration among members of a species group, the printing of long oligos offers advantages in start-up time, the purity of commercial oligo synthesis, easier clone tracking, and the ability to utilize all available sequences in public genetic databases for array construction. Readers are referred to Whitfield et al.'s (2002) EST sequencing and microarray research on honey bee using spotted cDNAs; Rise et al. (2004a) and von Schalburg et al. (2005) describe considerations taken in construction of salmonid spotted cDNA arrays; and Zhao et al. (2005) reports validation of a porcine spotted oligo array. Operon Biotechnologies (http://www.operon.com/) is a leading provider of sets of synthetic oligos for microarray spotting, and their website provides an excellent resource for criteria used in gene selection and long oligo design. Additionally, the

Institute for Genomic Research (TIGR), well known for its EST indices, provides 70mer oligo predictions for genes in each of its indices that have been utilized by some groups (Zhao et al. 2005). Researchers should also decide in the design phase their array layout, feature duplication, and the controls to be spotted on the slide (Whitfield et al. 2002; Smyth et al. 2005).

A variety of microarray slides are available for printing, most are poly-L-lysine and amino silane-coated (see Hessner et al. 2004 for surface-chemistry comparisons). Telechem (http://www.arrayit.com/Products/Substrates/) and Erie Scientific (http://www.eriemicroarray.com/index.aspx) are leading providers of microarray slides. The actual robotic printing of microarrays is increasingly being outsourced to large university core labs or private companies which now have years of experience in the field. For groups that anticipate printing multiple array designs and batches and want increased printing flexibility, purchasing a spotting robot may be a good choice. Perkin Elmer (http://las.perkinelmer.com/) and Genomic Solutions (http://www.genomicsolutions.com) offer popular printing systems.

In situ array technology relies on photolithography for microarray construction (Lipshutz et al. 1999), a technique often used in computer chip fabrication. In contrast to spotting nucleotide products on the slide surface, oligonucleotides are synthesized directly on the surface of the array, one base at a time. To achieve sufficient feature densities, unique physical lithographic masks are created for each array design, to either block or allow light to reach the slide. In the places the mask does not cover, light deprotects, or converts a special protective group to a hydroxyl group. This allows the binding of single oligo at that specific site by its phosphate group. This oligo also bears a protective group that must be deprotected before an additional oligo can be coupled to it. Through repeated cycles of deprotection and coupling, 25-mer oligos are synthesized directly on the slide at densities currently as high as 1.3 million

features per array. Affymetrix (http://www.affymetrix.com/) is recognized as the developer and industry leader for *in situ* arrays. While their technology made genome-wide arrays a reality for model species and continues to expand the horizons of microarray research in biomedical fields, the technology has been prohibitively expensive for the smaller species groups including aquaculture species. Nimblegen Systems (http://www.nimblegen.com) has recently developed a "maskless" version of the Affymetrix technology that uses digital mirrors to achieve the same effect (Nuwaysir et al. 2002) at a significantly lower startup cost, now making *in situ* arrays a feasible choice for aquaculture genomic research.

The majority of array design considerations for *in situ* arrays overlap with those of spotted arrays. EST analysis, clustering, quality control, and probe selection are still necessary steps to arrive at the set of genes that will be synthesized on the array. The higher feature density allowed with *in situ* arrays means that more genes, duplicates, and/or controls may be included on the array, if desired. Because the per array cost is significantly higher for *in situ* arrays and project flexibility considerably less than for spotted arrays, researchers usually attempt to maximize the information that can be gained from each slide. Usually, desired sequences for the array are sent electronically to the company, which then carries out oligo probe selection (23-25-mers) and designs the array layout. Both Affymetrix and Nimblegen use a perfect match (PM) and mismatch (MM) system that accounts for the majority of the features on *in situ* arrays. Mismatch probes, as their name suggests, contain one or more mismatched nucleotides in the PM probe sequence and are used to detect and screen out false background fluorescence resulting from non-specific cross-hybridization. Commonly, 10 PM and 10 MM probes are synthesized for each gene included on the array, and are believed to significantly increase the accuracy and

sensitivity of gene expression detection (see Chen et al. 2005; Han et al. 2004; Irizarry et al. 2003 for more information on PM and MM probe theory).

Spotted and *in situ* microarrays differ not only in their array construction but also in the procedures used to label and hybridize experimental samples (probes in the traditional sense) to them. Both array platforms require that you start with RNA sources. The RNA is extracted from the samples of interest. Each RNA sample is reverse transcribed to cDNA, after quantification and quality-checking by spectrophotometer measurement and agarose gel electrophoresis. From this step, differences in the procedure arise between the two microarray platforms.

The cDNA samples for spotted arrays are labeled with two different fluorescent dyes, Cy3 and Cy5, which fluoresce "green" and "red" respectively under two different wavelengths of light (633 nm and 543 nm). The control sample is labeled with one dye and the treatment sample with the other. Dye assignments should be swapped in replicates to avoid dye-associated bias of hybridization (Churchill 2002). Dye labeling is most commonly done either directly or through indirect aminoallyl labeling (see Manduchi et al. 2002; Badiee et al. 2003 for a comparison of labeling methods). The two labeled samples are hybridized simultaneously in equal amounts to the same array for 16-20 hrs. The hybridized array, after washing to remove unhybridized probes, is scanned under a laser scanner (e.g. Molecular Devices/Axon Instruments' Axon 4000B) at both fluorescent wavelengths (or channels) for the two dyes. A digital image is acquired for both channels, and, by overlaying the two images, a fluorescent signal ratio for each array feature is obtained. This fluorescent signal ratio indicates gene expression levels. Using the Cy3/Cy5 labeling system, yellow spots indicate approximately equal levels of mRNA from both the control and treatment samples (equal signals from the green Cy3 and the red Cy5). Features that appear red or green have hybridized a majority of mRNA from only one sample.

Fluorescent intensity data for each feature are recorded, and the scanned image and data can be linked back to gene feature identities through programs such as Molecular Devices/Axon Instruments' GenePix Pro software. Background subtraction and normalization is customarily carried out at this point, followed by microarray analysis and validation of genes determined to be significantly differentially expressed after treatment.

For *in situ* arrays, the RNA samples are reverse transcribed using a T7 promoter oligo-dT primer. The resulting cDNA is converted to a double-stranded template by a second strand synthesis reaction. After purification, these double-stranded cDNA samples (again control and treatment) are converted by *in vitro* transcription to biotin-labeled cRNA using a T7 RNA polymerase. The cRNA from each sample is fragmented and hybridized to its own slide (note: no mixing of samples). Streptavidin-phycoerithryn is added as the fluorescent dye for both the control and treatment samples. To clarify, each biological sample for *in situ* arrays is hybridized to a *different* slide and labeled with a *single* dye. Differential expression is measured by comparing the fluorescent intensity measurement of a given gene on the control slide with a separate measurement for the same gene from the treatment slide. Labeling reactions and hybridizations of *in situ* arrays are commonly carried out by the array provider or core lab. Several groups have experimentally compared the precision and accuracy of the two platforms using the same biological samples. Their studies may prove helpful to those considering which system to implement in their own research (see Woo et al. 2004; Yauk et al. 2004; Meijer et al. 2005).

Microarray research has advanced dramatically in recent years in aquaculture or aquatic species (summarized in Table 4). However, the field is still in its infancy and distribution of resources remains uneven. A number of microarrays have been developed from variety of

aquaculture species that has led to the publication of a special issue in the Journal of Fish Biology devoted entirely to the description of microarrays in aquatic species (Table 4). Interested readers are referred to this Special Issue (Journal of Fish Biology, Volume 72, issue 9, 2008).

To-date, most published microarray studies have used PCR-amplified spotted cDNA clones to fabricate the array. However, as microarray research typically takes several years from its inception to reach publication, the recent trends toward spotted oligos and *in situ* microarrays may not be reflected in the aquaculture literature for several years. A well-designed microarray can be a valuable asset to an aquaculture species group, especially if the cost per slide can be minimized to the extent that researchers can integrate transcriptomic approaches into their already established research. Microarray studies are most successful when they are just one of several approaches used to answer biological questions. For example, salmonid researchers have implemented array technology in their study of reproductive development, toxicology, physiology, and repeat structures (von Schalburg et al. 2006; Tilton et al. 2005; Vornanen et al. 2005; Krasnov et al. 2005; Schiøtz et al., 2008; Jørgensen et al., 2008; Eichner et al., 2008; Young et al., 2008; Wynne et al., 2008; Gahr et al., 2008; Vanya et al., 2008; Roberge et al., 2007; Martin et al., 2006; Ewart et al., 2005; Rise et al., 2004b). In a similar effort, microarrays have been used to identify defense related genes in ESC-resistant blue catfish and ESC-susceptible channel catfish (Peatman et al., 2007; 2008).

Due to low funding levels and a relatively small research community, aquaculture genomics stands today where the model species did almost a decade ago. In the same way, microarray research in aquaculture species is only in its infancy. Like researchers of humans and mice ten years ago, we are currently using microarrays to accelerate gene expression analysis

under varied experimental conditions, to reveal novel functions in genes, and to discover

possible gene interactions and networking through cluster analysis. To find future directions for

microarray research in aquaculture species, we need only to observe microarray studies in model

species today. The future looks especially promising for using microarrays for SNP analysis and

QTL mapping to make tangible progress towards widespread marker-assisted selection (MAS) in

aquaculture. In particular, merging positional candidate genes with expression candidate genes

from microarray information may reveal QTL genes responsible for important performance traits

(see Drake et al. 2006).   Microarrays have, furthermore, evolved to allow studies of

metabolomics and proteomics that will be important in development of fish vaccines (Cretich et

al. 2006). A Veterinary Immune Reagent Network has already been established in the US toward

development of a set of antibodies for use in agricultural research including aquaculture

(http://www.avma.org/onlnews/javma/jun06/060615b.asp).   Microarrays are also being utilized

in livestock disease diagnostics, a use easily adapted for detection of outbreaks of aquaculture

pathogens (Schmitt and Henderson 2005; Baxi 2006). Much of the groundwork for practical

microarray research has already been laid. It is up to the aquaculture community to exploit and

adapt these advances for the advantage of their respective species.

    In spite of the bright outlook of microarray research, emerging next generation

sequencing technologies may soon replace, at least in part, the capacity of microarrays. This is

because sequence-based gene expression profiling can provide not only gene identities without

any ambiguity, but also more accurate assessment of genome expression based on sequence tag

counting.

B. Sequence tag-based technology

Tag profiling is a revolutionary approach to gene expression analysis that generates expression profiles for any transcript from any organism. Using Solexa or SOLiD sequencing technologies, millions of expressed sequence tags can be generated from a single run allowing gene expression profiles to be accurately characterized based on sequence tag counting. The major advantage of tag profiling is its high ability to identify, quantify, and annotate expressed genes on the level of the whole genome without prior sequence knowledge. Because sequence tag-based genome expression analysis does not require any existing genome resources, it is much more adaptable to aquaculture species where whole genome sequences are lacking.

C. Comparison of the microarray technology with tag or sequence-based technology

In microarray experiments, hybridization signal intensities are used to generate abundance measurements that correspond to the amount of target mRNA that has hybridized to a specific probe; and relative measurements are determined by a comparison of two samples. Tag-based technologies measure the expression level of a gene by counting the abundance of a specific transcript in a sample. This count provides an abundance measure of each gene's expression level within the sample. Recently, some studies have compared microarray and massively parallel signature sequencing (MPSS) technology. Their results suggested a moderate correlation between the two platforms. However, one platform often detects expression for some genes that are not measured by the other platform (Coughlan et al., 2004; Oudes et al., 2005; Liu et al., 2007), suggesting that using a combination of transcription profiling technologies would provide more complete coverage of gene expression measurements. Different technology platforms likely will provide significant differences in the measurements of gene expression.

Chen et al. (2007) recently found that RNA samples exhibited higher correlations within the technology platform used to measure RNA abundance rather than expected similarities due to the biological nature of the samples. In particular, the tag- or sequence-based platforms may be more variable in measuring RNA abundance than Affymetrix or Agilent microarray platforms. Therefore, comparison of RNA abundance across technology platforms requires exercise of caution. However, when relative expression between samples with different biological treatment is of issue, which most often is the most important question for aquaculture research settings, the samples were more closely clustered according to their biological nature than the technology platform (Chen et al., 2007).

Acknowledgments

References

Agresti J J, Seki S, Cnaani A, Poompuang S, Hallerman E M, Umiel N, Hulata G, Gall G a E &
    May B (2000), Breeding new strains of tilapia: development of an artificial center of
    origin and linkage map based on AFLP and microsatellite loci, *Aquaculture,* 185**,** 43-56.

Altshuler D, Pollara V J, Cowles C R, Van Etten W J, Baldwin J, Linton L & Lander E S (2000),
    An SNP map of the human genome generated by reduced representation shotgun
    sequencing, *Nature,* 407**,** 513-6.

Bader J S (2001), The relative power of SNPs and haplotype as genetic markers for association
    tests, *Pharmacogenomics,* 2**,** 11-24.

Badiee A, Eiken H G, Steen V M & Lovlie R (2003), Evaluation of five different cDNA labeling
    methods for microarrays using spike controls, *BMC Biotechnol,* 3**,** 23.

Bargelloni L, Franch R, Patarnello T, Tsalavouta M, Sarropoulou E, Magoulas A, Kotoulas G,
    Chatziplis D, Georgoudis A, Louro B & Power D M (2007), A genetic linkage map of the
    hermaphrodite teleost fish *Sparus aurata* L. *Aquaculture*, 272, Supplement 1, 243.

Baxi M K, Baxi S, Clavijo A, Burton K M & Deregt D (2006), Microarray-based detection and
    typing of foot-and-mouth disease virus, *Vet J,* 172**,** 473-81.

Beckman J S & Weber J L (1992), Survey of human and rat microsatellites, *Genomics,* 12**,** 627-
    31.

Bekal S, Craig J P, Hudson M E, Niblack T L, Domier L L & Lambert K N (2008), Genomic
    DNA sequence comparison between two inbred soybean cyst nematode biotypes
    facilitated by massively parallel 454 micro-bead sequencing, *Mol Genet Genomics*.

Bentley D R (2006), Whole-genome re-sequencing, *Curr Opin Genet Dev,* 16**,** 545-52.

Bonnet E, Fostier A & Bobe J (2007), Microarray-based analysis of fish egg quality after natural or controlled ovulation, *BMC Genomics*, 8, 55.

Butler J, Maccallum I, Kleber M, Shlyakhter I A, Belmonte M K, Lander E S, Nusbaum C & Jaffe D B (2008), ALLPATHS: de novo assembly of whole-genome shotgun microreads, *Genome Res,* 18**,** 810-20.

Cairney M, Taggart J B & Hoyheim B (2000), Characterization of microsatellite and minisatellite loci in Atlantic salmon (*Salmo salar* L.) and cross-species amplification in other salmonids, *Mol Ecol,* 9**,** 2175-8.

Campbell D & Bernatchez L (2004), Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes, *Mol Biol Evol,* 21**,** 945-56.

Cardoso S R, Eloy N B, Provan J, Cardoso M A & Ferreira P C (2000), Genetic differentiation of Euterpe edulis Mart. populations estimated by AFLP analysis, *Mol Ecol,* 9**,** 1753-60.

Cariello N F, Scott J K, Kat a G, Thilly W G & Keohavong P (1988), Resolution of a missense mutant in human genomic DNA by denaturing gradient gel electrophoresis and direct sequencing using in vitro DNA amplification: HPRT Munich, *Am J Hum Genet,* 42**,** 726-34.

Chen D T, Chen J J & Soong S J (2005), Probe rank approaches for gene selection in oligonucleotide arrays with a small number of replicates, *Bioinformatics,* 21**,** 2861-6.

Chen J, Agrawal V, Rattray M, West M A, St Clair D A, Michelmore R W, Coughlan S J & Meyers B C (2007), A comparison of microarray and MPSS technology platforms for expression analysis of *Arabidopsis*, *BMC Genomics,* 8**,** 414-26.

Chen W, Kalscheuer V, Tzschach A, Menzel C, Ullmann R, Schulz M H, Erdogan F, Li N, Kijas Z, Arkesteijn G, Pajares I L, Goetz-Sothmann M, Heinrich U, Rost I, Dufke A, Grasshoff U, Glaeser B, Vingron M & Ropers H H (2008), Mapping translocation breakpoints by next-generation sequencing, *Genome Res*.

Chistiakov D A, Hellemans B, Haley C S, Law a S, Tsigenopoulos C S, Kotoulas G, Bertotto D, Libertini A & Volckaert F A (2005), A microsatellite linkage map of the European sea bass *Dicentrarchus labrax* L, *Genetics,* 170**,** 1821-6.

Chong L K, Tan S G, Yusoff K & Siraj S S (2000), Identification and characterization of Malaysian river catfish, *Mystus nemurus* (C&V): RAPD and AFLP analysis, *Biochem Genet,* 38**,** 63-76.

Churchill G A (2002), Fundamentals of experimental design for cDNA microarrays, *Nat Genet,* 32 Suppl**,** 490-5.

Cloonan N, Forrest a R, Kolle G, Gardiner B B, Faulkner G J, Brown M K, Taylor D F, Steptoe a L, Wani S, Bethel G, Robertson a J, Perkins a C, Bruce S J, Lee C C, Ranade S S, Peckham H E, Manning J M, Mckernan K J & Grimmond S M (2008), Stem cell transcriptome profiling via massive-scale mRNA sequencing, *Nat Methods*.

Coimbra M R M, Kobayashi K, Koretsugu S, Hasegawa O, Ohara E, Ozaki A, Sakamoto T, Naruse K & Okamoto N (2003), A genetic linkage map of the Japanese flounder, *Paralichthys olivaceus*, *Aquaculture,* 220**,** 203-18.

Cokus S J, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild C D, Pradhan S, Nelson S F, Pellegrini M & Jacobsen S E (2008), Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning, *Nature,* 452**,** 215-9.

Cotton R G (1993), Current methods of mutation detection, *Mutat Res,* 285**,** 125-44.

Coughlan S J, Agrawal V & Meyers B (2004), A Comparison of Global Gene Expression Measurement Technologies in Arabidopsis thaliana, *Comparative and Functional Genomics,* 5**,** 245-52.

Crawford a M & Cuthbertson R P (1996), Mutations in sheep microsatellites, *Genome Res,* 6**,** 876-9.

Cretich M, Damin F, Pirri G & Chiari M (2006), Protein and peptide arrays: recent trends and new directions, *Biomol Eng,* 23**,** 77-88.

Crollius R H, Jaillon O, Dasilva C, Ozouf-Costaz C, Fizames C, Fischer C, Bouneau L, Billault A, Quetier F, Saurin W, Bernot A & Weissenbach J (2000), Characterization and repeat analysis of the compact genome of the freshwater pufferfish *Tetraodon nigroviridis*, *Genome Res*, 10, 939-49.

Danzmann R G & Gharbi K (2007), Linkage mapping in aquaculture species, in Liu Z, *Aquaculture Genome Technologies*, Blackwell Publishing, Ames, Iowa, 139-67.

Danzmann R G, Cairney M, Davidson W S, Ferguson M M, Gharbi K, Guyomard R, Holm L E, Leder E, Okamoto N, Ozaki A, Rexroad C E, 3rd, Sakamoto T, Taggart J B & Woram R A (2005), A comparative analysis of the rainbow trout genome with 2 other species of fish (Arctic charr and Atlantic salmon) within the tetraploid derivative Salmonidae family (subfamily: Salmoninae), *Genome,* 48**,** 1037-51.

Davidson W S (2007), Bacterial artificial chromosome libraries and BAC-based physical mapping of aquaculture genomes, in Liu Z, *Aquaculture Genome Technologies*, Blackwell Publishing, Ames, Iowa, 245-59.

de la Vega E, O'Leary N A, Shockey J E, Robalino J, Payne C, Browdy C L, Warr G W & Gross P S (2008), Anti-lipopolysaccharide factor in Litopenaeus vannamei (LvALF): a broad

spectrum antimicrobial peptide essential for shrimp immunity against bacterial and fungal infection, *Mol Immunol,* 45 (7), 1916-25.

Diab a M, Williams T D, Sabine V S, Chipman J K & George S G (2008), The GENIPOL European flounder Platichthys flesus L. toxicogenomics microarray: application for investigation of the response to furunculosis vaccination, *Journal of Fish Biology,* 72**,** 2154-69.

Dolan P C & Denver D R (2008), TileQC: a system for tile-based quality control of Solexa data, *BMC Bioinformatics,* 9**,** 250.

Douglas S E, Knickle L C, Williams J, Flight R M & Reith M E (2008), A first generation Atlantic halibut *Hippoglossus hippoglossus* (L.) microarray: application to developmental studies, *Journal of Fish Biology,* 72**,** 2391-406.

Drake T A, Schadt E E & Lusis a J (2006), Integrating genetic and gene expression data: application to cardiovascular and metabolic traits in mice, *Mamm Genome,* 17**,** 466-79.

Edwards Y J, Elgar G, Clark M S & Bishop M J (1998), The identification and characterization of microsatellites in the compact genome of the Japanese pufferfish, *Fugu rubripes*: perspectives in functional and comparative genomic analyses, *J Mol Biol,* 278**,** 843-54.

Eichner C, Frost P, Dysvik B, Jonassen I, Kristiansen B & Nilsen F (2008), Salmon louse (Lepeophtheirus salmonis) transcriptomes during post molting maturation and egg production, revealed using EST-sequencing and microarray analysis, *BMC Genomics,* 9**,** 126.

Ellegren H (2000), Microsatellite mutations in the germline: implications for evolutionary inference, *Trends Genet,* 16**,** 551-8.

Ewart K V, Belanger J C, Williams J, Karakach T, Penny S, Tsoi S C, Richards R C & Douglas S E (2005), Identification of genes differentially expressed in Atlantic salmon (*Salmo salar*) in response to infection by Aeromonas salmonicida using cDNA microarray technology, *Dev Comp Immunol*, 29**,** 333-47.

Felip A, Martinez-Rodriguez G, Piferrer F, Carrillo M & Zanuy S (2000), AFLP Analysis Confirms Exclusive Maternal Genomic Contribution of Meiogynogenetic Sea Bass (*Dicentrarchus labrax* L.), *Mar Biotechnol (NY),* 2**,** 301-6.

Fitzsimmons N N, Moritz C & Moore S S (1995), Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution, *Mol Biol Evol,* 12**,** 432-40.

Gahr S A, Vallejo R L, Weber G M, Shepherd B S, Silverstein J T & Rexroad C E, 3rd (2008), Effects of short-term growth hormone treatment on liver and muscle transcriptomes in rainbow trout (*Oncorhynchus mykiss*), *Physiol Genomics,* 32**,** 380-92.

Garcia-Reyero N, Griffitt R J, Liu L, Kroll K J, Farmerie W G, Barber D S & Denslow N D (2008), Construction of a robust microarray from a non-model species largemouth bass, *Micropterus salmoides* (Lacèpede), using pyrosequencing technology, *Journal of Fish Biology,* 72**,** 2354-76.

Geisler R, Rauch G J, Baier H, Van Bebber F, Bross L, Dekens M P, Finger K, Fricke C, Gates M A, Geiger H, Geiger-Rudolph S, Gilmour D, Glaser S, Gnugge L, Habeck H, Hingst K, Holley S, Keenan J, Kirn A, Knaut H, Lashkari D, Maderspacher F, Martyn U, Neuhauss S, Neumann C, Nicolson T, Pelegri F, Ray R, Rick J M, Roehl H, Roeser T, Schauerte H E, Schier a F, Schonberger U, Schonthaler H B, Schulte-Merker S, Seydler C, Talbot W S, Weiler C, Nusslein-Volhard C & Haffter P (1999), A radiation hybrid map of the zebrafish genome, *Nat Genet,* 23**,** 86-9.

Geoghegan F, Katsiadaki I, Williams T D & Chipman J K (2008), A cDNA microarray for the three-spined stickleback, *Gasterosteus aculeatus* L., and analysis of the interactive effects of oestradiol and dibenzanthracene exposures, *Journal of Fish Biology,* 72**,** 2133-53.

Gharbi K, Gautier A, Danzmann R G, Gharbi S, Sakamoto T, Hoyheim B, Taggart J B, Cairney M, Powell R, Krieg F, Okamoto N, Ferguson M M, Holm L E & Guyomard R (2006), A linkage map for brown trout (*Salmo trutta*): chromosome homeologies and comparative genome organization with other salmonid fish, *Genetics,* 172**,** 2405-19.

Gilbey J, Verspoor E, Mclay A & Houlihan D (2004), A microsatellite linkage map for Atlantic salmon (*Salmo salar*), Anim Genet, 35, 98-105

Glazov E A, Cottee P A, Barris W C, Moore R J, Dalrymple B P & Tizard M L (2008), A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach, *Genome Res,* 18**,** 957-64.

Gonen D, Veenstra-Vanderweele J, Yang Z, Leventhal B & Cook E H, Jr. (1999), High throughput fluorescent CE-SSCP SNP genotyping, *Mol Psychiatry,* 4**,** 339-43.

Gracey a Y (2008), The Gillichthys mirabilis Cooper array: a platform to investigate the molecular basis of phenotypic plasticity, *Journal of Fish Biology,* 72**,** 2118-2132.

Gracey a Y, Fraser E J, Li W, Fang Y, Taylor R R, Rogers J, Brass A & Cossins a R (2004), Coping with cold: An integrative, multitissue analysis of the transcriptome of a poikilothermic vertebrate, *Proc Natl Acad Sci U S A,* 101**,** 16970-5.

Griffiths R & Orr K (1999), The use of amplified fragment length polymorphism (AFLP) in the isolation of sex-specific markers, *Mol Ecol,* 8**,** 671-4.

Hafner M, Landgraf P, Ludwig J, Rice A, Ojo T, Lin C, Holoch D, Lim C & Tuschl T (2008),
Identification of microRNAs and other small regulatory RNAs using cDNA library
sequencing, *Methods,* 44**,** 3-12.

Halldorsson B V, Istrail S & De La Vega F M (2004), Optimal selection of SNP markers for
disease association studies, *Hum Hered,* 58**,** 190-202.

Han E S, Wu Y, Mccarter R, Nelson J F, Richardson A & Hilsenbeck S G (2004),
Reproducibility, sources of variability, pooling, and sample size: important
considerations for the design of high-density oligonucleotide array experiments, *J
Gerontol A Biol Sci Med Sci,* 59**,** 306-15.

He L, Du C, Li Y, Scheuring C & Zhang H B (2007), Construction of large-insert bacterial clone
libraries and their applications, in Liu Z, *Aquaculture Genome Technologies,* Blackwell
Publishing, Ames, Iowa, 215-44.

Hedgecock D, Lin J Z, Decola S, Haudenschild C D, Meyer E, Manahan D T & Bowen B
(2007), Transcriptomic analysis of growth heterosis in larval Pacific oysters (*Crassostrea
gigas*), *Proc Natl Acad Sci U S A,* 104**,** 2313-8.

Hessner M J, Meyer L, Tackes J, Muheisen S & Wang X (2004), Immobilized probe and glass
surface chemistry as variables in microarray fabrication, *BMC Genomics,* 5**,** 53.

Hillier L W, Marth G T, Quinlan a R, Dooling D, Fewell G, Barnett D, Fox P, Glasscock J I,
Hickenbotham M, Huang W, Magrini V J, Richt R J, Sander S N, Stewart D A,
Stromberg M, Tsung E F, Wylie T, Schedl T, Wilson R K & Mardis E R (2008), Whole-
genome sequencing and variant discovery in C. elegans, *Nat Methods,* 5**,** 183-8.

Hubert S & Hedgecock D (2004), Lingkage maps of microsatellite DNA markers for the Pacific
oyster *Crassostrea gigas*, Genetics, 168, 351-62.

Hukriede N A, Joly L, Tsang M, Miles J, Tellis P, Epstein J A, Barbazuk W B, Li F N, Paw B, Postlethwait J H, Hudson T J, Zon L I, Mcpherson J D, Chevrette M, Dawid I B, Johnson S L & Ekker M (1999), Radiation hybrid mapping of the zebrafish genome, *Proc Natl Acad Sci U S A,* 96**,** 9745-50.

Hunter R L & Markert C L (1957), Histochemical demonstration of enzymes separated by zone electrophoresis in starch gels, *Science,* 125**,** 1294-5.

Irizarry R A, Hobbs B, Collin F, Beazer-Barclay Y D, Antonellis K J, Scherf U & Speed T P (2003), Exploration, normalization, and summaries of high density oligonucleotide array probe level data, *Biostatistics,* 4**,** 249-64.

Jenny M J, Chapman R W, Mancia A, Chen Y A, Mckillen D J, Trent H, Lang P, Escoubas J M, Bachere E, Boulo V, Liu Z J, Gross P S, Cunningham C, Cupit P M, Tanguy A, Guo X, Moraga D, Boutet I, Huvet A, De Guise S, Almeida J S & Warr G W (2007), A cDNA microarray for *Crassostrea virginica* and *C. gigas*, *Mar Biotechnol (NY),* 9**,** 577-91.

Jones a G, Rosenqvist G, Berglund A & Avise J C (1999), Clustered microsatellite mutations in the pipefish Syngnathus typhle, *Genetics,* 152**,** 1057-63.

Jorde P E, Palm S & Ryman N (1999), Estimating genetic drift and effective population size from temporal shifts in dominant gene marker frequencies, *Mol Ecol,* 8**,** 1171-8.

Jørgensen SM, Afanasyev S, Krasnov A (2008), Gene expression analyses in Atlantic salmon challenged with infectious salmon anemia virus reveal differences between individuals with early, intermediate and late mortality. *BMC Genomics* 9, 179.

Ju Z, Karsi A, Kocabas A, Patterson A, Li P, Cao D, Dunham R & Liu Z (2000), Transcriptome analysis of channel catfish (*Ictalurus punctatus*): genes and expression profile from the brain, *Gene,* 261**,** 373-82.

Kai Y, Nakayama K & Nakabo T (2002), Genetic differences among three colour morphotypes of the black rockfish, *Sebastes inermis*, inferred from mtDNA and AFLP analyses, *Mol Ecol,* 11**,** 2591-8.

Kalin I, Shephard S & Candrian U (1992), Evaluation of the ligase chain reaction (LCR) for the detection of point mutations, *Mutat Res,* 283**,** 119-23.

Kane M D, Sringer J A, Iannotti N V, Gough E, Johns S M, Schlueter S D & Sepulveda M S (2008), Identification of development and tissue-specific gene expression in the fathead minnow *Pimephales promelas*, Rafinesque using computational and DNA microarray methods, *Journal of Fish Biology,* 72**,** 2341-53.

Karsi A, Cao D, Li P, Patterson A, Kocabas A, Feng J, Ju Z, Mickett K D & Liu Z (2002), Transcriptome analysis of channel catfish (*Ictalurus punctatus*): initial analysis of gene expression and microsatellite-containing cDNAs in the skin, *Gene,* 285**,** 157-68.

Katagiri T, Kidd C, Tomasino E, Davis J T, Wishon C, Stern J E, Carleton K L, Howe a E & Kocher T D (2005), A BAC-based physical map of the Nile tilapia genome, *BMC Genomics,* 6**,** 89.

Kijas J M, Fowler J C, Garbett C A & Thomas M R (1994), Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles, *Biotechniques,* 16**,** 656-60, 662.

Klaper R, Carter B J, Richter C A, Drevnick P E, Sandheinrich M B & Tillitt D E (2008), Use of a 15 k gene microarray to determine gene expression changes in response to acute and

chronic methylmercury exposure in the fathead minnow *Pimephales promelas* Rafinesque, *Journal of Fish Biology,* 72**,** 2207-80.

Kocher T D, Lee W J, Sobolewska H, Penman D & Mcandrew B (1998), A genetic linkage map of a cichlid fish, the tilapia (*Oreochromis niloticus*), *Genetics,* 148**,** 1225-32.

Korol A, Shirak A, Cnaani A & Hallerman E (2007), Detection and analysis of quantitative trait loci (QTLs) for economic traits in aquatic species, in Liu Z, *Aquaculture Genome Technologies*, Blackwell Publishing, Ames, Iowa, 169-97.

Krasnov A, Koskinen H, Afanasyev S & Molsa H (2005), Transcribed Tc1-like transposons in salmonid fish, *BMC Genomics,* 6**,** 107.

Kucuktas H and Liu Z (2007), Allozyme and mitochondrial markers, in Liu Z, *Aquaculture Genome Technologies*, Blackwell Publishing, Ames, Iowa,73-85

Kucuktas H, Wagner B K, Shopen R, Gibson M, Dunham, R A & Liu Z J (2002), Genetic analysis of Ozark Hellbenders (*Cryptobranchus alleganiensis bishopi*) utilizing RAPD markers, *Proc Ann Conf SEAFWA* 55, 126-37.

Kwok C, Critcher R & Schmitt K (1999), Construction and characterization of zebrafish whole genome radiation hybrids, *Methods Cell Biol,* 60**,** 287-302.

Leclerc D, Wirth T & Bernatchez L (2000), Isolation and characterization of microsatellite loci in the yellow perch (*Perca flavescens*), and cross- species amplification within the family Percidae, *Mol Ecol,* 9**,** 995-7.

Lee B-Y, Kocher T (2007), Comparative genomics and positional cloning. In: *Aquaculture Genome Technologies* (edited by Liu ZJ.), pp. 323-335, Blackwell Publishing, Ames, IA.

Lee B Y, Lee W J, Streelman J T, Carleton K L, Howe a E, Hulata G, Slettan A, Stern J E, Terai Y & Kocher T D (2005), A second-generation genetic linkage map of tilapia (*Oreochromis spp.*), *Genetics,* 170**,** 237-44.

Li B, Kadura I, Fu D J & Watson D E (2004), Genotyping with TaqMAMA, *Genomics,* 83**,** 311-20.

Li L & Guo X (2004), AFLP-based genetic linkage maps of the pacific oyster *Crassostrea gigas* Thunberg, *Mar Biotechnol (NY),* 6**,** 26-36.

Li L, Xiang J, Liu X, Zhang Y, Dong B & Zhang X (2005), Construction of AFLP-based genetic linkage map for Zhikong scallop, *Chlamys farreri* Jones et Preston and mapping of sex-linked markers, *Aquaculture,* 245**,** 63-73.

Li Y, Byrne K, Miggiano E, Whan V, Moore S, Keys S, Crocos P, Preston N & Lehnert S (2003), Genetic mapping of the kuruma prawn *Penaeus japonicus* using AFLP markers, *Aquaculture,* 219**,** 143-56.

Li Z, Li J, Wang Q, He Y & Liu P (2006), AFLP-based genetic linkage map of marine shrimp *Penaeus* (*Fenneropenaeus*) *chinensis*, *Aquaculture*, 261, 2, 463-72.

Lipshutz R J, Fodor S P, Gingeras T R & Lockhart D J (1999), High density synthetic oligonucleotide arrays, *Nat Genet,* 21**,** 20-4.

Liu Z J (2006), Transcriptome characterization through the generation and analysis of expressed sequence tags: Factors to consider for a successful EST project, *Israel Journal of Aquaculture-Bamidgeh* 58, 328-41.

Liu Z J (2007a), Aquaculture Genome Technologies, Blackwell Publishing, Oxford, UK.

Liu Z J (2007b), Marking the genome: Restriction fragment length polymorphism (RFLP). In: *Aquaculture Genome Technologies* (edited by Liu ZJ.), pp. 11-20, Blackwell Publishing, Ames, IA.

Liu Z J (2007c), Random amplified polymorphic DNA (RAPD). In: *Aquaculture Genome Technologies* (edited by Liu ZJ.), pp. 21-28, Blackwell Publishing, Ames, IA.

Liu Z J (2007d), Amplified fragment length polymorphism (AFLP). In: *Aquaculture Genome Technologies* (edited by Liu ZJ.), pp. 29-42, Blackwell Publishing, Ames, IA.

Liu Z J (2007e), Microsatellite markers and assessment of marker utility. In: *Aquaculture Genome Technologies* (edited by Liu ZJ.), pp. 43-58, Blackwell Publishing, Ames, IA.

Liu Z J (2007f), Single nucleotide polymorphism (SNP). In: *Aquaculture Genome Technologies* (edited by Liu ZJ.), pp. 59-72, Blackwell Publishing, Ames, IA.

Liu Z J (2007g), Genomescape: characterizing the repeat structure of the genome. In: *Aquaculture Genome Technologies* (edited by Liu ZJ.), pp. 277-290, Blackwell Publishing, Ames, IA.

Liu Z J (2007h), Transcriptome characterization through the analysis of expressed sequence tags. In: *Aquaculture Genome Technologies* (edited by Liu ZJ.), pp. 301-356, Blackwell Publishing, Ames, IA.

Liu Z J (2007i), DNA sequencing technologies. In: *Aquaculture Genome Technologies* (edited by Liu ZJ.), pp. 465-476, Blackwell Publishing, Ames, IA.

Liu Z J (2007j), Sequencing the genome. In: *Aquaculture Genome Technologies* (edited by Liu ZJ.), pp. 477-490, Blackwell Publishing, Ames, IA.

Liu F, Jenssen T K, Trimarchi J, Punzo C, Cepko C L, Ohno-Machado L, Hovig E & Kuo W P (2007), Comparison of hybridization-based and sequencing-based gene expression technologies on biological replicates, *BMC Genomics,* 8**,** 153.

Liu X, Liu X, Guo X, Gao Q, Zhao H & Zhang G (2006), A preliminary genetic linkage map of the Pacific abalone *Haliotis discus hannai* Ino, *Mar Biotechnol (NY),* 8**,** 386-97.

Liu Z J & Cordes J F (2004), DNA marker technologies and their applications in aquaculture genetics, *Aquaculture,* 238**,** 1-37.

Liu Z, Karsi A, Li P, Cao D & Dunham R (2003), An AFLP-based genetic linkage map of channel catfish (Ictalurus punctatus) constructed by using an interspecific hybrid resource family, *Genetics,* 165**,** 687-94.

Liu Z, Li P, Kucuktas H, Nichols A, Tan G, Zheng X, Argue B J & Dunham R A (1999), Development of amplified fragment length polymorphism (AFLP) markers suitable for genetic linkage mapping of catfish, *Trans Am Fish Soc*, 128, 317-27.

Liu Z, Li R W & Waldbieser G C (2008), Utilization of microarray technology for functional genomics in ictalurid catfish, *Journal of Fish Biology,* 72**,** 2377-90.

Liu Z, Nichols A, Li P & Dunham R A (1998), Inheritance and usefulness of AFLP markers in channel catfish (*Ictalurus punctatus*), blue catfish (*I. furcatus*), and their F1, F2, and backcross hybrids, *Mol Gen Genet,* 258**,** 260-8.

Liu Z, Li P, Kocabas A, Karsi A & Ju Z (2001), Microsatellite-containing genes from the channel catfish brain: evidence of trinucleotide repeat expansion in the coding region of nucleotide excision repair gene RAD23B, *Biochem Biophys Res Commun,* 289, 317-24.

Lyall J E, Brown G M, Furlong R A, Ferguson-Smith M A & Affara N A (1993), A method for creating chromosome-specific plasmid libraries enriched in clones containing [CA]n

microsatellite repeat sequences directly from flow-sorted chromosomes, *Nucleic Acids Res,* 21**,** 4641-2.

Mackiewicz M, Fletcher D E, Wilkins S D, Dewoody J A & Avise J C (2002), A genetic assessment of parentage in a natural population of dollar sunfish (*Lepomis marginatus*) based on microsatellite markers, *Mol Ecol,* 11**,** 1877-83.

Malmgren H, Gustavsson J, Tuvemo T & Dahl N (1996), Rapid detection of a mutation hot-spot in the human androgen receptor, *Clin Genet,* 50**,** 202-5.

Manduchi E, Scearce L M, Brestelli J E, Grant G R, Kaestner K H & Stoeckert C J, Jr. (2002), Comparison of different labeling methods for two-channel high-density microarray experiments, *Physiol Genomics,* 10**,** 169-79.

Margulies M, Egholm M, Altman W E, Attiya S, Bader J S, Bemben L A, Berka J, Braverman M S, Chen Y J, Chen Z, Dewell S B, Du L, Fierro J M, Gomes X V, Godwin B C, He W, Helgesen S, Ho C H, Irzyk G P, Jando S C, Alenquer M L, Jarvie T P, Jirage K B, Kim J B, Knight J R, Lanza J R, Leamon J H, Lefkowitz S M, Lei M, Li J, Lohman K L, Lu H, Makhijani V B, Mcdade K E, Mckenna M P, Myers E W, Nickerson E, Nobile J R, Plant R, Puc B P, Ronan M T, Roth G T, Sarkis G J, Simons J F, Simpson J W, Srinivasan M, Tartaro K R, Tomasz A, Vogt K A, Volkmer G A, Wang S H, Wang Y, Weiner M P, Yu P, Begley R F & Rothberg J M (2005), Genome sequencing in microfabricated high-density picolitre reactors, *Nature,* 437**,** 376-80.

Marnellos G (2003), High-throughput SNP analysis for genetic association studies, *Curr Opin Drug Discov Devel,* 6**,** 317-21.

Martin S A, Blaney S C, Houlihan D F & Secombes C J (2006), Transcriptome response following administration of a live bacterial vaccine in Atlantic salmon (*Salmo salar*), *Mol Immunol,* 43**,** 1900-11.

Maxam a M & Gilbert W (1977), A new method for sequencing DNA, *Proc Natl Acad Sci U S A,* 74**,** 560-4.

May B (2003), Allozyme variation, in Hallerman E M , *Population genetics: principles and applications for fisheries scientists*, American Fisheries Society, Bethesda, Maryland, 23-36.

Meijer a H, Verbeek F J, Salas-Vidal E, Corredor-Adamez M, Bussman J, Van Der Sar a M, Otto G W, Geisler R & Spaink H P (2005), Transcriptome profiling of adult zebrafish at the late stage of chronic tuberculosis due to *Mycobacterium marinum* infection, *Mol Immunol,* 42**,** 1185-203.

Metzgar D, Bytof J & Wills C (2000), Selection against frameshift mutations limits microsatellite expansion in coding DNA, *Genome Res,* 10**,** 72-80.

Mickett K, Morton C, Feng J, Li P, Simmons M, Cao D, Dunham R A & Liu Z (2003), Assessing genetic diversity of domestic populations of channel catfish (*Ictalurus punctatus*) in Alabama using AFLP markers, *Aquaculture,* 228**,** 91-105.

Mock K E, Brim-Box J C, Miller M P, Downing M E & Hoeh W R (2004), Genetic diversity and divergence among freshwater mussel (Anodonta) populations in the Bonneville Basin of Utah, *Mol Ecol,* 13**,** 1085-98.

Moen T, Hoyheim B, Munck H & Gomez-Raya L (2004), A linkage map of Atlantic salmon (*Salmo salar*) reveals an uncommonly large difference in recombination rate between the sexes, *Anim Genet,* 35**,** 81-92.

Moore S S, Whan V, Davis G P, Byrne K, Hetzel D J S & Preston N (1999), The development and application of genetic markers for the Kuruma prawn *Penaeus japonicus*, *Aquaculture,* 173**,** 19-32.

Moran C (1993), Microsatellite repeats in pig (*Sus domestica*) and chicken (*Gallus domesticus*) genomes, *J Hered,* 84**,** 274-80.

Morishima K, Arias-Rodriguez L, Nakayama I & Arai K (2007), A genetic linkage map of the loach *Misgurnus anguillicaudatus* (Teleostei: Cobitidae), *Aquaculture*, 272, Supplement 1, 291

Ng S H, Artieri C G, Bosdet I E, Chiu R, Danzmann R G, Davidson W S, Ferguson M M, Fjell C D, Hoyheim B, Jones S J, De Jong P J, Koop B F, Krzywinski M I, Lubieniecki K, Marra M A, Mitchell L A, Mathewson C, Osoegawa K, Parisotto S E, Phillips R B, Rise M L, Von Schalburg K R, Schein J E, Shin H, Siddiqui A, Thorsen J, Wye N, Yang G & Zhu B (2005), A physical map of the genome of Atlantic salmon, *Salmo salar, Genomics,* 86**,** 396-404.

Nichols K M, Young W P, Danzmann R G, Robison B D, Rexroad C, Noakes M, Phillips R B, Bentzen P, Spies I, Knudsen K, Allendorf F W, Cunningham B M, Brunelli J, Zhang H, Ristow S, Drew R, Brown K H, Wheeler P A & Thorgaard G H (2003), A consolidated linkage map for rainbow trout (*Oncorhynchus mykiss*), *Anim Genet,* 34**,** 102-15.

Ning Y, Liu X, Wang Z Y, Guo W, Li Y & Xie F (2007), A genetic map of large yellow croaker *Pseudosciaena crocea*, *Aquaculture*, 264, 16-26.

Nuwaysir E F, Huang W, Albert T J, Singh J, Nuwaysir K, Pitas A, Richmond T, Gorski T, Berg J P, Ballin J, McCormick M, Norton J, Pollock T, Sumwalt T, Butcher L, Porter D, Molla M, Hall C, Blattner F, Sussman M R, Wallace R L, Cerrina F & Green R D (2002), Gene

expression analysis using oligonucleotide arrays produced by maskless photolithography, *Genome Res*, 12, 1749-55.

Ohara E, Nishimura T, Nagakura Y, Sakamoto T, Mushiake K, Okamoto N (2005), Genetic linkage maps of two yellowtails (*Seriola quinqueradiata* and *Seriola lalandi*), *Aquaculture*, 244, 41-8.

Okumuş I & Çiftci Y (2003), Fish population genetics and molecular markers: II. Molecular markers and their applications in fisheries and aquaculture, Turkish, *J Fish Aquatic Sci*, 3, 51-79.

Olohan L A, Li W, Wulff T, Jarmer H, Gracey a Y & Cossins a R (2008), Detection of anoxia-responsive genes in cultured cells of the rainbow trout *Oncorhynchus mykiss* (Walbaum), using an optimized, genome-wide oligoarray, *Journal of Fish Biology,* 72**,** 2170-86.

Ostrander E A, Jong P M, Rine J & Duyk G (1992), Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences, *Proc Natl Acad Sci U S A,* 89**,** 3419-23.

Oudes a J, Roach J C, Walashek L S, Eichner L J, True L D, Vessella R L & Liu a Y (2005), Application of Affymetrix array and Massively Parallel Signature Sequencing for identification of genes involved in prostate cancer progression, *BMC Cancer,* 5**,** 86.

Parker P G, Snow A A, Schug M D, Booton G C & Fuerst P A (1998), What molecules can tell us about populations: choosing and using a molecular marker, *Ecology*, 79, 361-82.

Patrick K (2007), 454 life sciences: illuminating the future of genome sequencing and personalized medicine, *Yale J Biol Med,* 80**,** 191-4.

Peatman E & Liu Z J (2007), Microarray fundamentals: Basic principles and application in aquaculture. In: *Aquaculture Genome Technologies* (edited by Liu ZJ.), pp. 357-370, Blackwell Publishing, Ames, IA.

Peatman E, Baoprasertkul P, Terhune J, Xu P, Nandi S, Kucuktas H, Li P, Wang S, Somridhivej B, Dunham R & Liu Z J (2007), Expression analysis of the acute phase response in channel catfish (*Ictalurus punctatus*) after infection with a Gram-negative bacterium, *Developmental & Comparative Immunology,* 31**,** 1183-96.

Peatman E, Terhune J, Baoprasertkul P, Xu P, Nandi S, Wang S, Somridhivej B, Kucuktas H, Li P, Dunham R, Liu Z J (2008), Microarray analysis of gene expression in the blue catfish liver reveals early activation of the MHC class I pathway after infection with *Edwardsiella ictaluri*, *Molecular Immunology* 45, 553-66.

Pérez F, Erazo C, Zhinaula M, Volckaert F & Calderón J (2004), A sex-specific linkage map of the white shrimp Penaeus (Litopenaeus) vannamei based on AFLP markers, *Aquaculture,* 242**,** 105-18.

Poompuang S & Na-Nakorn U (2004), A preliminary genetic map of walking catfish (*Clarias macrocephalus*), *Aquaculture,* 232**,** 195-203.

Quiniou S M, Waldbieser G C & Duke M V (2007), A first generation BAC-based physical map of the channel catfish genome, *BMC Genomics,* 8**,** 40.

Reith M, Reid D, Smith C, Blanchard B & Martin-Robichaud D (2007), A genetic linkage map for Atlantic halibut (*Hippoglossus hippoglossus*), *Aquaculture,* 272, Supplement 1, 304.

Rexroad III C E (2007), Radiation hybrid mapping in aquatic species, Aquaculture Genome Technologies, Blackwell Publishing, Ames, Iowa, 313-322.

Rico C, Rico I & Hewitt G (1996), 470 million years of conservation of microsatellite loci

    among fish species, *Proc R Soc Lond B Biol Sci*, 263, 549–557.

Rise M L, Jones S R, Brown G D, Von Schalburg K R, Davidson W S & Koop B F (2004b),

    Microarray analyses identify molecular biomarkers of Atlantic salmon macrophage and

    hematopoietic kidney response to *Piscirickettsia salmonis* infection, *Physiol Genomics,*

    20**,** 21-35.

Rise M L, Von Schalburg K R, Brown G D, Mawer M A, Devlin R H, Kuipers N, Busby M,

    Beetz-Sargent M, Alberto R, Gibbs a R, Hunt P, Shukin R, Zeznik J A, Nelson C, Jones S

    R, Smailus D E, Jones S J, Schein J E, Marra M A, Butterfield Y S, Stott J M, Ng S H,

    Davidson W S & Koop B F (2004a), Development and application of a salmonid EST

    database and cDNA microarray: data mining and interspecific hybridization

    characteristics, *Genome Res,* 14**,** 478-90.

Robalino J, Almeida J S, Mckillen D, Colglazier J, Trent H F, 3rd, Chen Y A, Peck M E,

    Browdy C L, Chapman R W, Warr G W & Gross P S (2007), Insights into the immune

    transcriptome of the shrimp *Litopenaeus vannamei*: tissue-specific expression profiles

    and transcriptomic responses to immune challenge, *Physiol Genomics,* 29**,** 44-56.

Roberge C, Paez D J, Rossignol O, Guderley H, Dodson J & Bernatchez L (2007), Genome-wide

    survey of the gene expression response to saprolegniasis in Atlantic salmon, *Mol*

    Immunol, 44**,** 1374-83.

Robison B D, Wheeler P A, Sundin K, Sikka P & Thorgaard G H (2001), Composite interval

    mapping reveals a major locus influencing embryonic development rate in rainbow trout

    (*Oncorhynchus mykiss*), *J Hered,* 92**,** 16-22.

Rogers S M, Campbell D, Baird S J, Danzmann R G & Bernatchez L (2001), Combining the analyses of introgressive hybridisation and linkage mapping to investigate the genetic architecture of population divergence in the lake whitefish (*Coregonus clupeaformis*, Mitchill), *Genetica,* 111**,** 25-41.

Ross P, Hall L, Smirnov I & Haff L (1998), High level multiplex genotyping by MALDI-TOF mass spectrometry, *Nat Biotechnol,* 16**,** 1347-51.

Sakamoto T, Danzmann R G, Gharbi K, Howard P, Ozaki A, Khoo S K, Woram R A, Okamoto N, Ferguson M M, Holm L E, Guyomard R & Hoyheim B (2000), A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates, *Genetics,* 155**,** 1331-45.

Salem M, Kenney P B, Rexroad C E & Yao J (2008), Development of a 37 k high-density oligonucleotide microarray: a new tool for functional genome research in rainbow trout, *Journal of Fish Biology,* 72**,** 2187-206.

Sanger F, Nicklen S & Coulson a R (1977), DNA sequencing with chain-terminating inhibitors, *Proc Natl Acad Sci U S A,* 74**,** 5463-7.

Sarropoulou E, Franch R, Louro B, Power D M, Bargelloni L, Magoulas A, Senger F, Tsalavouta M, Patarnello T, Galibert F, Kotoulas G & Geisler R (2007), A gene-based radiation hybrid map of the gilthead sea bream *Sparus aurata* refines and exploits conserved synteny with *Tetraodon nigroviridis*, *BMC Genomics,* 8**,** 44.

Sarropoulou E, Kotoulas G, Power D M & Geisler R (2005), Gene expression profiling of gilthead sea bream during early development and detection of stress-related genes by the application of cDNA microarray technology, *Physiol Genomics,* 23**,** 182-91.

Schiøtz BL, Jørgensen SM, Rexroad C, Gjøen T, Krasnov A (2008), Transcriptomic analysis of responses to infectious salmon anemia virus infection in macrophage-like cells. Virus Res. Jun 3. [Epub ahead of print] PMID: 18534703 .

Schmitt B & Henderson L (2005), Diagnostic tools for animal diseases, *Rev Sci Tech,* 24**,** 243-50.

Seki S, Agresti J J, Gall GAE, Taniguchi N & May B (1999), AFLP analysis of genetic diversity in three populations of ayu *Plecoglossus altivelis, Fish Sci,* 65, 888-92.

Senger F, Priat C, Hitte C, Sarropoulou E, Franch R, Geisler R, Bargelloni L, Power D & Galibert F (2006), The first radiation hybrid map of a perch-like fish: the gilthead seabream (*Sparus aurata* L), *Genomics,* 87**,** 793-800.

Serapion J, Kucuktas H, Feng J & Liu Z (2004), Bioinformatic mining of type I microsatellites from expressed sequence tags of channel catfish (*Ictalurus punctatus*), *Mar Biotechnol (NY),* 6**,** 364-77.

Shagin D A, Rebrikov D V, Kozhemyako V B, Altshuler I M, Shcheglov a S, Zhulidov P A, Bogdanova E A, Staroverov D B, Rasskazov V A & Lukyanov S (2002), A novel method for SNP detection using a new duplex-specific nuclease from crab hepatopancreas, *Genome Res,* 12**,** 1935-42.

Shen X, Yang G, Liu Y, Liao M, Wang X, Zhu M, Song W, Zou G, Wei Q, Wang D & Chen D (2007), Construction of genetic linkage maps of guppy (*Poecilia reticulata*) based on AFLP and microsatellite DNA markers, *Aquaculture*, 271, 178-87.

Siden T S, Kumlien J, Schwartz C E & Rohme D (1992), Radiation fusion hybrids for human chromosomes 3 and X generated at various irradiation doses, *Somat Cell Mol Genet,* 18**,** 33-44.

Simmons M, Mickett K, Kucuktas H, Li P, Dunham R & Liu Z J (2006), Comparison of

 domestic and wild catfish populations provide no evidence for genetic impact,

 Aquaculture, 252, 133-46.

Smith M H & Chesser R K (1981), Rationale for conserving genetic variation of fish gene pools.

 In Ryman N, *Fish Gene Pools*, Ecological Bulletin, Stockholm, 34, 13-20.

Smyth G K, Michaud J & Scott H S (2005), Use of within-array replicate spots for assessing

 differential expression in microarray experiments, *Bioinformatics,* 21**,** 2067-75.

Somridhivej B, Wang S, Sha Z, Liu H, Quilang J, Xu P, Li P & Liu Z J (2008), Characterization,

 polymorphism assessment, and database construction for microsatellites from BAC end

 sequences of catfish: a resource for integration of linkage and physical maps,

 Aquaculture, 275, 76-80.

Sorrentino R, Potolicchio I, Ferrara G B & Tosi R (1992), A new approach to HLA-DPB1 typing

 combining DNA heteroduplex analysis with allele-specific amplification and enzyme

 restriction, *Immunogenetics,* 36**,** 248-54.

Southern E M (1975), Detection of specific sequences among DNA fragments separated by gel

 electrophoresis, *J Mol Biol,* 98**,** 503-17.

Steele C A , Wheeler P A & Thorgaard G H (2008), Mitochondrial And Maternal Effects On

 Growth In Clonal Rainbow, Abstract of Plant & Animal Genome Conference XVI,

 (http://www.intl-pag.org/16/abstracts/PAG16_P05p_619.html)

Steinberg E K, Lindner K R, Gallea J, Maxwell A, Meng J & Allendorf F W (2002), Rates and

 patterns of microsatellite mutations in pink salmon, *Mol Biol Evol,* 19**,** 1198-202.

Storm N, Darnhofer-Patel B, Van Den Boom D & Rodi C P (2003), MALDI-TOF mass

 spectrometry-based SNP genotyping, *Methods Mol Biol,* 212**,** 241-62.

Stram D O (2004), Tag SNP selection for association studies, *Genet Epidemiol,* 27**,** 365-74.

Sun X W & Liang L Q (2004), A genetic linkage map of common carp (*Cyprinus carpio* L.) And mapping of a locus associated with cold tolerance, *Aquaculture ,* 238, 165-172.

Sun Y, Song W, Zhong Y, Zhang R, Abatzopoulos T J & Chen R (1999), Diversity and genetic differentiation in *Artemia* species and populations detected by AFLP markers, *Int J Salt Lake Res*, 8, 341-50.

Suzuki Y, Orita M, Shiraishi M, Hayashi K & Sekiya T (1990), Detection of ras gene mutations in human lung cancers by single-strand conformation polymorphism analysis of polymerase chain reaction products, *Oncogene,* 5**,** 1037-43.

Taggart J B, Bron J E, Martin S a M, Seear P J, Hoyheim B, Talbot R, Carmichael S N, Villeneuve L a N, Sweeney G E, Houlihan D F, Secombes C J, Tocher D R & Teale a J (2008), A description of the origins, design and performance of the TRAITS-SGP Atlantic salmon Salmo salar L. cDNA microarray, *Journal of Fish Biology,* 72**,** 2071-94.

Tautz D (1989), Hypervariability of simple sequences as a general source for polymorphic DNA markers, *Nucleic Acids Res,* 17**,** 6463-71.

Tilton S C, Gerwick L G, Hendricks J D, Rosato C S, Corley-Smith G, Givan S A, Bailey G S, Bayne C J & Williams D E (2005), Use of a rainbow trout oligonucleotide microarray to determine transcriptional patterns in aflatoxin B1-induced hepatocellular carcinoma compared to adjacent liver, *Toxicol Sci,* 88**,** 319-30.

Toth G, Gaspari Z & Jurka J (2000), Microsatellites in different eukaryotic genomes: survey and analysis, *Genome Res,* 10**,** 967-81.

Van Lith H A & Van Zutphen L F (1996), Characterization of rabbit DNA microsatellites extracted from the EMBL nucleotide sequence database, *Anim Genet,* 27**,** 387-95.

Van Tassell C P, Smith T P L, Matukumalli L K, Taylor J F, Schnabel R D, Lawley C T, Haudenschild C D, Moore S S, Warren W C & Sonstegard T S (2008), SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries, *Nat Meth,* 5**,** 247-52.

Vanya Ewart K, Williams J, Richards R C, Gallant J W, Melville K & Douglas S E (2008), The early response of Atlantic salmon (Salmo salar) macrophages exposed in vitro to *Aeromonas salmonicida* cultured in broth and in fish, *Dev Comp Immunol,* 32**,** 380-90.

Vera J C, Wheat C W, Fescemyer H W, Frilander M J, Crawford D L, Hanski I & Marden J H (2008), Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing, *Mol Ecol,* 17**,** 1636-47.

Villeneuve D L, Knoebl I, Larkin P, Miracle a L, Carter B J, Denslow N D & Ankley G T (2008), Altered gene expression in the brain and liver of female fathead minnows *Pimephales promelas* Rafinesque exposed to fadrozole, *Journal of Fish Biology,* 72**,** 2281-340.

Volckaert F A M J, Batargias C, Chatziplis D, Chistiakov D, Guo Y, Haley C S, Hellemans B, Kotoulas G, Law A, Libertini A, McAndrew B J, Tsigenopoulos C & Whitaker H A (2007), A second generation linkage map of European sea bass, *Aquaculture*, 272, Supplement 1, 316.

Von Gersdorff Jorgensen L, Heinecke R D, Kania P & Buchmann K (2008), Occurrence of gyrodactylids on wild Atlantic salmon, *Salmo salar* L., in Danish rivers, *J Fish Dis*, 31, 127-34

Von Schalburg K R, Mccarthy S P, Rise M L, Hutson J C, Davidson W S & Koop B F (2006), Expression of morphogenic genes in mature ovarian and testicular tissues: potential stem-cell niche markers and patterning factors, *Mol Reprod Dev,* 73**,** 142-52.

Von Schalburg K R, Rise M L, Cooper G A, Brown G D, Gibbs a R, Nelson C C, Davidson W S & Koop B F (2005), Fish and chips: various methodologies demonstrate utility of a 16,006-gene salmonid microarray, *BMC Genomics,* 6**,** 126.

Vornanen M, Hassinen M, Koskinen H & Krasnov A (2005), Steady-state effects of temperature acclimation on the transcriptome of the rainbow trout heart, *Am J Physiol Regul Integr Comp Physiol,* 289**,** R1177-84.

Vos P, Hogers R, Bleeker M, Reijans M, Van De Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M & Et Al. (1995), AFLP: a new technique for DNA fingerprinting, *Nucleic Acids Res,* 23**,** 4407-14.

Waldbieser G C, Bosworth B G, Nonneman D J & Wolters W R (2001), A microsatellite-based genetic linkage map for channel catfish, Ictalurus punctatus, *Genetics,* 158**,** 727-34.

Walter M A & Goodfellow P N (1993), Radiation hybrids: irradiation and fusion gene transfer, *Trends Genet,* 9**,** 352-6.

Wang B, Li F, Luan W, Xie Y, Zhang C, Luo Z, Gui L, Yan H & Xiang J (2008), Comparison of Gene Expression Profiles of *Fenneropenaeus chinensis* Challenged with WSSV and Vibrio, *Mar Biotechnol (NY).*

Wang L L, Song L S, Chang Y Q, Xu W, Ni D J & Guo X M (2005), A preliminary genetic map of Zhikong scallop (*Chlamys farreri* Jones et Preston 1904), *Aquaculture Res*, 36, 643-53.

Wang S, Sha Z, Sonstegard T S, Liu H, Xu P, Somridhivej B, Peatman E, Kucuktas H, Liu Z J (2008), Quality assessment parameters for EST-derived SNPs from catfish, *BMC Genomics*, in press.

Watanabe T, Fujita H, Yamasaki K, Seki S & Taniguchi N (2004), Preliminary study on linkage mapping based on microsatellite DNA and AFLP markers using homozygous clonal fish in ayu (*Plecoglossus altivelis*), *Mar Biotechnol (NY),* 6**,** 327-34.

Weber J L & Wong C (1993), Mutation of human short tandem repeats, *Hum Mol Genet,* 2**,** 1123-8.

Welsh J & Mcclelland M (1990), Fingerprinting genomes using PCR with arbitrary primers, *Nucleic Acids Res,* 18**,** 7213-8.

Whitehead A, Anderson S L, Kuivila K M, Roach J L & May B (2003), Genetic variation among interconnected populations of *Catostomus occidentalis*: implications for distinguishing impacts of contaminants from biogeographical structuring, *Mol Ecol,* 12**,** 2817-33.

Whitfield C W, Band M R, Bonaldo M F, Kumar C G, Liu L, Pardinas J R, Robertson H M, Soares M B & Robinson G E (2002), Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee, *Genome Res,* 12**,** 555-66.

Williams D R, Li W, Hughes M A, Gonzalez S F, Vernon C, Vidal M C, Jeney Z, Jeney G, Dixon P, Mcandrew B, Bartfai R, Orban L, Trudeau V, Rogers J, Matthews L, Fraser E J, Gracey a Y & Cossins a R (2008), Genomic resources and microarrays for the common carp Cyprinus carpio L, *Journal of Fish Biology,* 72**,** 2095-117.

Williams J G, Kubelik a R, Livak K J, Rafalski J A & Tingey S V (1990), DNA polymorphisms amplified by arbitrary primers are useful as genetic markers, *Nucleic Acids Res,* 18**,** 6531-5.

Wilson K, Li Y, Whan V, Lehnert S, Byrne K, Moore S, Pongsomboon S, Tassanakajon A, Rosenberg G, Ballment E, Fayazi Z, Swan J, Kenway M & Benzie J (2002), Genetic mapping of the black tiger shrimp *Penaeus monodon* with amplified fragment length polymorphism, *Aquaculture,* 204**,** 297-309.

Woo Y, Affourtit J, Daigle S, Viale A, Johnson K, Naggert J & Churchill G (2004), A comparison of cDNA, oligonucleotide, and Affymetrix GeneChip gene expression microarray platforms, *J Biomol Tech,* 15**,** 276-84.

Woram R A, Mcgowan C, Stout J A, Gharbi K, Ferguson M M, Hoyheim B, Davidson E A, Davidson W S, Rexroad C & Danzmann R G (2004), A genetic linkage map for Arctic char (*Salvelinus alpinus*): evidence for higher recombination rates and segregation distortion in hybrid versus pure strain mapping parents, *Genome,* 47**,** 304-15.

Wynne J W, O'sullivan M G, Cook M T, Stone G, Nowak B F, Lovell D R & Elliott N G (2008), Transcriptome Analyses of Amoebic Gill Disease-affected Atlantic Salmon (*Salmo salar*) Tissues Reveal Localized Host Gene Suppression, *Mar Biotechnol (NY),* 10**,** 388-403.

Xu P, Wang S & Liu Z (2007a), Physical characterization of aquaculture genomes through BAC end sequencing, in Liu Z, *Aquaculture Genome Technologies*, Blackwell Publishing, Ames, Iowa, 261-74.

Xu P, Wang S, Liu L, Peatman E, Somridhivej B, Thimmapuram J, Gong G & Liu Z (2006), Channel catfish BAC-end sequences for marker development and assessment of syntenic conservation with other fish species, *Anim Genet,* 37**,** 321-6.

Xu P, Wang S, Liu L, Thorsen J, Kucuktas H & Liu Z (2007b), A BAC-based physical map of the channel catfish genome, *Genomics,* 90**,** 380-8.

Yauk C L, Berndt M L, Williams A & Douglas G R (2004), Comprehensive comparison of six microarray technologies, *Nucleic Acids Res,* 32**,** e124.

Young N D, Cooper G A, Nowak B F, Koop B F & Morrison R N (2008), Coordinated down-regulation of the antigen processing machinery in the gills of amoebic gill disease-affected Atlantic salmon (*Salmo salar* L.), *Mol Immunol,* 45**,** 2581-97.

Young W P, Ostberg C O, Keim P & Thorgaard G H (2001), Genetic characterization of hybridization and introgression between anadromous rainbow trout (*Oncorhynchus mykiss irideus*) and coastal cutthroat trout (O. clarki clarki), *Mol Ecol,* 10**,** 921-30.

Young W P, Wheeler P A, Coryell V H, Keim P & Thorgaard G H (1998), A detailed linkage map of rainbow trout produced using doubled haploids, *Genetics,* 148**,** 839-50.

Yu Z & Guo X (2003), Genetic linkage map of the eastern oyster Crassostrea virginica Gmelin, *Biol Bull,* 204**,** 327-38.

Zane L, Bargelloni L & Patarnello T (2002), Strategies for microsatellite isolation: a review, *Mol Ecol,* 11**,** 1-16.

Zhao S H, Recknor J, Lunney J K, Nettleton D, Kuhar D, Orley S & Tuggle C K (2005), Validation of a first-generation long-oligonucleotide microarray for transcriptional profiling in the pig, *Genomics,* 86**,** 618-25.

Zhang Y, Liang L, Jiang P, Li D, Lu C & Sun X (2008), Genome evolution trend of common carp (*Cyprinus carpio* L.) as revealed by the analysis of microsatellite loci in a gynogentic family, *J Genet Genomics,* 35**,** 97-103.

Zhou Z, Bao Z, Dong Y, Wang S, He C, Liu W, Wang L & Zhu F (2006), AFLP lingkage map of sea urchin constructed using an interspecific cross between *Strongylocentrous nudus* (Venus) and *S. intermedius* (Mars). *Aquaculture*, 259, 56-65.

Zhulidov P A, Bogdanova E A, Shcheglov a S, Vagner L L, Khaspekov G L, Kozhemyako V B, Matz M V, Meleshkevitch E, Moroz L L, Lukyanov S A & Shagin D A (2004), Simple cDNA normalization using kamchatka crab duplex-specific nuclease, *Nucleic Acids Res,* 32**,** e37.