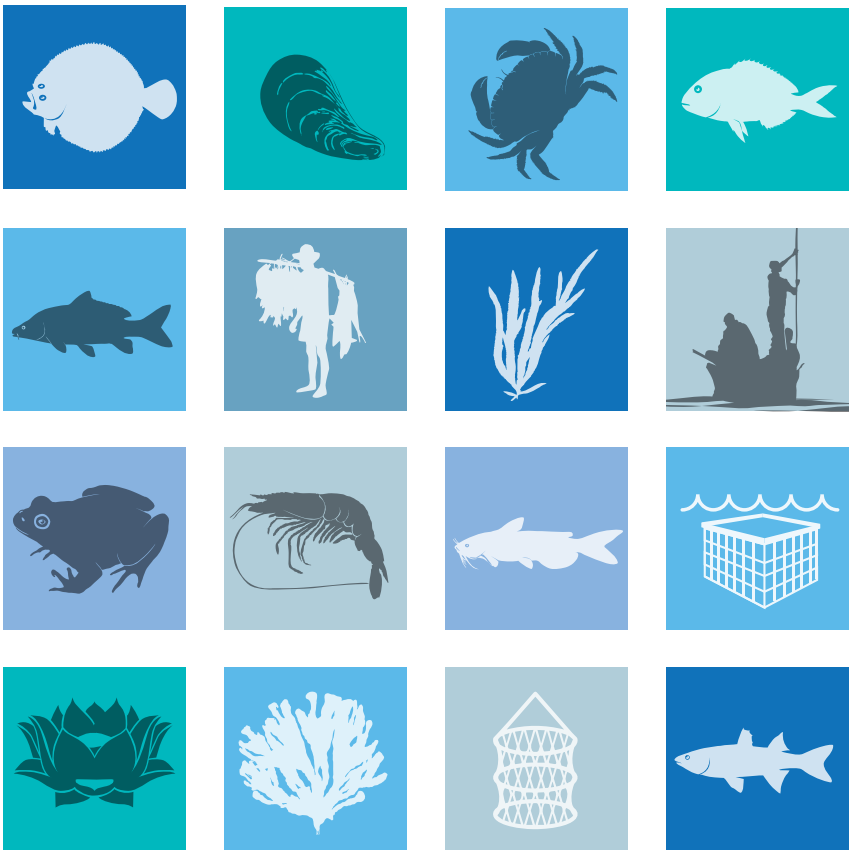
Food and Agriculture Organization of the United Nations

THEMATIC BACKGROUND STUDY

Genome-based biotechnologies in aquaculture



THEMATIC BACKGROUND STUDY

Genome-based biotechnologies in aquaculture

Food and Agriculture Organization of the United Nations
Rome, 2021

Required citation:
FAO. 2021. *Thematic Background Study – Genome-based biotechnologies in aquaculture.* Rome. https://doi.org/10.4060/cb7955en

ISBN 978-92-5-135464-3

Contents

Tables

Acronyms and abbreviations

AFLP	amplified fragment length polymorphism
ASE	allele-specific expression
BAC	bacterial artificial chromosome
BOLD	Barcode of Life Data System
Cas9	CRISPR-associated protein-9 nuclease
cDNA	complementary DNA
CNV	copy number variation
COI	cytochrome oxidase I
CRISPR	clustered regularly interspaced short palindromic repeats
CRISPR/Cas	clustered regularly interspaced short palindromic repeats-CRISPR associated DSB double-strand break
EST	expressed sequence tag
FDA	Food and Drug Administration
F1	first filial generation
F2	second hybrid generation
FPC	FingerPrintedContigs
GMO	genetically modified organism
GO	gene ontology
GWAS	genome-wide association study
HPC	high-performance computer clusters
IPR	Intellectual Property Rights
lncRNA	long non-coding RNA
INDELS	insertions or deletions
IPN	infectious pancreatic necrosis
LG	linkage group
MAS	marker-assisted selection or marker-aided selection miRNA microRNA
mRNA	messenger RNA
mtDNA	mitochondrial DNA
NCBI	National Center for Biotechnology Information
NHEJ	non-homologous end joining

ORF	open reading frame
QTL	quantitative trait locus
PAM	protospacer adjacent motif
PCR	polymerase chain reaction
RAD-seq	restriction site-associated DNA sequencing RAPD random amplified polymorphic DNA
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RNA-seq	ribonucleic acid sequencing
RH	radiation hybrid
sgRNA	single guide RNA
siRNA	short interfering RNA
SNP	single nucleotide polymorphism
TALEN	transcription activator-like effector nucleases
USDA	United States Department of Agriculture
ZFN	zinc finger nuclease

1. INTRODUCTION

1.1 History of genome research

The discussion of sequencing the entire human genome started approximately in the mid-1980s. Such a task was daunting at the time because of the limitations of sequencing technologies. The most efficient automated DNA sequencer was able to run 96 samples at a time, and approximately 500–800 base pairs of nucleotides could be read per sample. The sequencing reactions themselves were not fully automated and were expensive. With this technology, it was estimated that the project could take at least 15 years to complete. With many planning activities involving government and scientific communities with various workshops, the project was officially launched in 1990 with a single goal of sequencing and assembling the entire genome of a single human individual containing three billion base pairs of DNA. The project started in the United States of America, but became an international collaborative project involving the participation of universities and research centres in China, France, Germany, Japan, the United Kingdom of Great Britain and Northern Ireland, and the United States of America.

It should be noted that a parallel project was conducted by a private corporation, the Celera Corporation, which formally launched in 1998. It is widely believed that this parallel run of the project from the private sector significantly increased competition, allowing speeding up of the project as well as enhancement of the sequence quality.

1.2 Development of genomics as a new branch of science

Into the mid-phase of the 13-year human genome project, it was becoming evident that it would not be possible to put the human genome together with just DNA sequences generated with sequencing technologies. This was because the vast majority of the three billion base pairs of human DNA are repetitive in nature. These repetitive elements of the genome are different, but highly similar sequences. Computer software would not lend the capability to assemble all such sequences correctly into their proper place. Therefore, the demands on the development of additional technologies and research to meet this challenge were tremendously high. A new branch of science, genomics, was born. The scope of genomics is to apply the techniques of genetics and molecular biology for the understanding of genome structure, organization, expression and functions. Starting from its beginnings, genomics was meant to be a branch of science that uses principles and methods of multidisciplinarity, including those of biochemistry, molecular biology, genetics, statistics and informatics, to name a few. It is clear, as detailed below, that genomics is now the centre of life sciences. Just as the molecularization of biology in the past 40 years, now many scientific disciplines have been or are being genomicized. This is largely because genomics has led to major changes on how science is conducted. Rather than on the traditional process of proposing a hypothesis based on existing knowledge, testing it experimentally, and accepting or rejecting the hypothesis as the route of scientific research, genomics and its related large data sets allowed the change from hypothesis-driven research to data-driven research. All the data can be suddenly created, and it is the researchers' role to figure out the simple facts from the complex data.

As progress of the human genome project was being made, it became obvious that sequences themselves can be assembled by extension of the overlapping DNA sequences into longer contiguous sequences (contigs). However, the size of the contigs is limited by the presence of the repetitive elements. Thus, the human genome sequences could be assembled into hundreds of thousands of small contigs. Although the sequences within the contigs were known, the relationship or the positions of the contigs in the genome could not be determined by sequence assembly. This demanded the development of methods to place and order the contigs onto the chromosomes and eventually to the genome. This was the start of the development of genome mapping technologies, including genetic linkage mapping, physical mapping, radiation hybrid mapping, and optical mapping. Within the area of physical mapping, a few technologies were developed,

with the most popular being radiation hybrid mapping and large-insert libraries such as fingerprinting based on bacterial artificial chromosome (BAC) for mapping the physical genomic DNA. Similar to the situation of the human genome project, genome projects of aquaculture species require similar technological support to assemble billions of base pairs of DNA together correctly. These mapping technologies will be discussed in detail below.

1.3 Science demands the development of sequencing technologies

The sequencing difficulties and the involved costs have demanded rapid development of sequencing technologies. Principles for many sequencing technologies were known even at the beginning of the human genome project, but the practicalities were not possible because of the manufacturing limitations of minimizing the reactions, allowing many reactions to happen within a single machine simultaneously. Two lines of technological advances had the largest push on sequencing technologies. One was the advances in polymerase chain reaction, and the other the advances in nano technologies.

Genomic sciences have made drastic advances in the past ten years, largely because of the application of next-generation sequencing technologies. It is not just the high throughput that has revolutionized the way science is conducted, but the rapidly reducing cost for sequencing has also made the technologies applicable to all aspects of molecular biological research, as well as to all organisms including aquaculture and fisheries species. Twenty some years ago, Dr Francis Collins had a vision of achieving the sequencing of one genome with USD 1 000, and that is now possible. From the billion dollar human genome project, to the genome projects of agriculture animals with a budget of ten million dollars or so, down to a million dollars just a few years ago, to the current cost level of just tens of thousands of dollars for a *de novo sequencing* project, the potential for research using genomic approaches has become unlimited. Today, commercial services are available worldwide for genome sequencing projects, whether they are new sequencing projects for a species or resequencing projects for many individuals. The key issue is to achieve a balanced output of quality and quantity with minimal costs.

Rapid technological advances provide huge opportunities to apply modern genomics to enhance aquaculture production and performance traits. However, we are facing a number of new challenges, especially in the area of bioinformatics. This challenge may be paramount for aquaculture researchers and educators. Although aquaculture students are well educated in aquaculture, they may have no background in computer science or lack the knowledge of bioinformatics analysis of large data sets. The large data sets in tera-scales themselves pose great

computational challenges. Therefore, new ways of thinking are required in terms of education and training of the next generation of scientists. For instance, a few laboratories in the world may be sufficient for the production of data, but an infinite number of laboratories may be required for data analysis or bioinformatics data mining to link data with biology.

1.4 Aquaculture genomics, a historical review

Aquaculture is a relatively new but increasingly growing sector of agriculture. It is very important not only for economic interests but also for social and cultural significance. In many Asian countries, serving food fish at the dinner table is seen as highly prestigious. Food fish accounts for 20 percent of animal protein sources for the world population. The food fish industry involves a total of 144 million tonnes annual production with 44 million fishers and fish farmers and 2.1 million vessels, contributing USD 166 billion to the world economy and over USD 25 billion of international trade annually. Currently, aquaculture accounts for over 40 percent of food fish consumed in the world, and China is the only country where aquaculture produces more than 50 percent of consumed fish food (FAO, 2016).

Aquaculture genomics officially started in the 1990s, although related genome research was conducted in the 1980s. It was signified by the first Aquaculture Genomics Workshop held in 1997 in Dartmouth, Massachusetts, United States of America. This workshop targeted a group of six species for genome research in the United States: salmonids, catfish, tilapia, striped bass, oysters and shrimps.

As any other agricultural sector, sustained production requires research of basic biology, including growth, nutrition, reproduction, physiology, and genetics and genomics. One interesting observation is that all these research fields are being unified through the use of genomic technologies or, that is, they are being genomicized.

The availability of a draft whole-genome sequence significantly enhances genome research and applications of genome-based technologies for improving agricultural production and quality. As such, whole-genome sequences are available or nearly available for major livestock species, including cattle, swine, chicken and horse. Whole-genome sequencing used to be nearly utopian because of its involved huge costs. Take the human genome as an example; it took the major genome centres over a decade to complete the draft genome sequencing of the first human individual, from 1986 to 2000. The direct cost involved in the human genome sequencing was at the level of hundreds of millions of dollars. Along with the sequencing effort, related human genome research cost much more. Human genome related research probably spent in excess of USD 6 billion–USD 8 billion. Thereafter, the sequencing of the cattle genome cost over USD 60 million, much less expensive than the human genome sequencing, but yet is still at a cost level unthinkable for many aquaculture species. It is because of such huge costs that whole-genome sequencing used to be regarded as a milestone of a lifetime achievement.

The daunting task of whole-genome sequencing has now become possible for many species, including for many aquaculture species. With the second- and third-generation sequencing technologies, the cost of sequencing a genome with a size of one billion base pairs has reduced to manageable levels, usually about USD 100 000. With such a major reduction in costs, the sequencing rush is on the way for many species.

The rapid advances in sequencing technologies made science road maps and plans almost meaningless. In 2008, the United States Department of Agriculture (USDA) spent much energy developing the Blueprint for USDA Efforts in Agricultural Animal Genomics 2008–2017. Among many aquaculture species, the blueprint calls for the production of a draft genome sequence for catfish with only 6X genome coverage and a much lower coverage for several other aquaculture species, including rainbow trout, tilapia and shrimps. As short as two years later, the drastic reduction in sequencing costs made it possible to sequence the whole genome of many aquaculture species.

Genomes of at least two dozen aquaculture species have been sequenced or are now being sequenced. Of the six aquatic species groups included in the United States Animal Genome NRSP-8 Program, a whole genome has been sequenced for Nile tilapia, rainbow trout, Atlantic salmon, catfish, striped bass, oysters and shrimps. While it is extremely exciting for aquaculture geneticists to have the whole genome sequenced for many of the important aquaculture species, it is pivotally important to address some of the most important issues related to whole-genome sequencing. These include the issues related to effective assembly of whole-genome sequences and those related to the effective uses of whole-genome sequences. The usefulness of the whole-genome sequence relies on the efficient sequence assembly and adequate sequence annotation, which in turn depends on the availability of a number of genome resources, including transcriptome sequencing and assembly and characterization of the non-coding part of the genome.

This document provides the basic concept, the descriptions of the technologies, and their application or potential application to aquaculture species. It begins with a brief description of traditional genetic biotechnologies for aquaculture, followed with DNA marker technologies, genome-mapping technologies, genome sequencing technologies, transcriptome analysis of aquaculture species, understanding the non-coding portions of the genomes, genetic analysis technologies, genome-based genetic selection technologies, and genome editing technologies. At the end, the document discusses some challenges for the applications of genome-based technologies in aquaculture.

2. TRADITIONAL GENETIC BIOTECHNOLOGIES FOR AQUACULTURE

2.1 Selective breeding

The goal of selective breeding is to improve production and performance traits in farmed animals via the selection of individuals for desirable phenotypic characteristics. For aquaculture species, these traits may include faster growth, feed conversion efficiency, disease resistance, processing yields, low oxygen tolerance, stress tolerance, robustness, morphology and sexual maturation, among other traits of interest.

Aquaculture breeding programmes have a short history. Selective breeding programmes for strain development began in the late 1960s (Gjedrem and Baranski, 2010) with the development of Atlantic salmon breeding programmes in Norway. Domestication of many aquaculture species is still a recent event. Therefore, a number of techniques are effective, including strain selection, cross breeding, hybridization, and within-strain selection. However, the main goal of genetic improvement programmes for production traits in modern breeding schemes is within-strain selection. This is achieved within a well-managed, commercial programme of family and pedigree tracking combined with extensive trait measurements on selection candidates or their relatives (Gutierrez and Houston, 2016). Selective breeding programmes exist for various aquaculture species, such as Atlantic salmon, rainbow trout, Nile tilapia, common carp, grass carp, silver carp, crucian carp, rohu carp, yellowtail, sea bream, channel catfish, European seabass, turbot, Asian seabass, Pacific and eastern oyster, shrimps, scallops and pearl oysters, among 60 some species (Gjedrem and Baranski, 2010).

Selective breeding has proven to be very effective in enhancing the traits of agricultural plants and animals. For instance, the genetic gain has been greater than 12 percent per generation for growth rate and for disease resistance when challenge tests are applied (Gjedrem and Robinson, 2014). The main reasons for the large genetic gains observed for aquatic species are their relatively high fertility and the natural existence of broad genetic variation for economically important traits, both of which allow a very high selection intensity to be applied. However, the genetic improvement of species grown in aquaculture is far behind that achieved for animals and plants. Less than 10 percent of aquaculture production is based on genetically improved stocks (Gjedrem and Robinson, 2014). Reviews and books are available that cover selective breeding in great detail (e.g. Hulata 1995, 2001; Stickney, 1994; Gjedrem and Robinson, 2014; Gjedrem and Baranski, 2010; Gjedrem, Robinson and Rye, 2012), and therefore this subject area is not the focus of this review.

Advances in sequencing technology and genomics have significantly improved the tools available for the genetic improvement of livestock. In particular, the development of genetic markers and linkage maps has permitted great advances in the quantitative analyses of commercially important traits. Some of the genome- based technologies in relation to selection will be covered in various sections, such as the analysis of quantitative trait locus, marker-assisted selection and whole-genome selection.

2.2 Polyploidy

Polyploidy is lethal in mammals and birds (Chourrout *et al.,* 1986), but has led to the development of many productive plant varieties such as domesticated wheat. Triploid fish are viable (Thorgaard, Jazwin and Stier, 1981; Wolters, Libey and Chrisman, 1981; Chourrout, 1984; Cassani and Caton, 1986) and are usually sterile, while tetraploid fish are usually viable and fertile (Dunham, 2011).

Through chromosome manipulation, aquaculture organisms can be obtained with various levels of chromosome sets. A regular diploid organism has two sets of chromosomes. The polyploid state refers to individuals with extra sets of chromosomes (Dunham, 2011). Triploid fish have three sets of chromosomes, and tetraploid fish have four sets of chromosomes. In addition to full sets of chromosomal compositions, aneuploids have at least a diploid set of chromosomes with one or more additional chromosomes, but not a full complement to the set (Dunham, 2011).

Triploids are organisms with three sets of homologous chromosomes. They are found spontaneously in both wild and cultured populations and can be induced in many commercial species of fish and shellfish. Triploidy is induced by allowing normal fertilization and then forcing retention of the second polar body (Chourrout, 1980, 1984; Lou and Purdom, 1984). The second polar body is retained by applying temperature (hot shocks or cold shocks), hydrostatic pressure, anaesthetics or chemical shocks shortly after fertilization (Dunham, 2011). Additionally, in some cases such as in rainbow trout, triploidy can be induced by applying high pH and high calcium to either sperm or eggs (Ueda, Sato and Kobayashi, 1988). Many reviews and books are available covering the technical aspects of polyploidy, and therefore this paper will not cover this extensively.

The performance of triploid fish varies. Triploidy can affect growth, feed conversion efficiency, disease resistance and other traits. For growth, triploid fish can grow faster, at the similar rate, or slower. However, even for those that grow faster, this advantage is not obvious until sexual maturity. It is apparent in sterile fish that energy otherwise expended on development of gonads and gametes is converted into growth.

In addition to direct induction of triploid fish and shellfish such as oysters, the production of tetraploid fish has been a way for mass production of triploid fish (Chourrout *et al*., 1986). First-generation tetraploids can be produced by hydrostatic pressure treatment before the first cleavage and raised until the adult stage. With rainbow trout, survival and growth of tetraploids are severely depressed when compared with the diploid control (Chourrout *et al*., 1986). However, when the tetraploid male fish were mated with normal diploid female fish to produce triploid fish, the progenies were consistently normal with high survival rates. The progenies were found to be almost all triploids by karyology, which failed to detect a significant rate of aneuploidies. Therefore, the use of tetraploid as a means of producing triploid fish is a viable option, although the fertilizing ability of tetraploid males was always low, 0 to 97 percent of the control, with an average of 40 percent compared with the diploid controls (Chourrout *et al*., 1986).

A review by Piferrer et al. (2009) covers all the details for triploid production and the performance of triploid fish and shellfish as compared with the diploid counterparts. Interested readers are referred to this review article.

2.3 Gynogenesis

Gynogenesis is a form of all-female inheritance. In fish species, ultraviolet (UV) irradiation has been used to inactivate the sperm, and such UV-inactivated sperm are used to trigger gynogenetic development without contributing the paternal genome to the progeny. Practically, sperm from a closely related, but different species, are used to reduce the possibility of real fertilization in case the sperm are not completely inactivated (Arai, 2001; Suwa, Arai and Suzuki, 1994). The developing embryo is initially haploid, but the diploidy is recovered by inhibiting either the second polar body extrusion or the first cleavage. The gynogen produced by inhibiting the second polar body extrusion is called meiotic gynogen, while the gynogen produced by inhibiting the first cleavage is called mitotic gynogen. Mitotic gynogens are completely homozygous, while meiotic gynogens are partially heterozygous. The gene-centromere recombination makes the distal part of the chromosome from the centromere more heterozygous. Meiotic gynogens are easier to produce than the mitotic gynogens because the survival rate of mitotic gynogens is very low.

One of the practical goals of gynogenesis is the production of genetically identical populations, i.e. the clonal lines (Arai, 2001). The clonal lines may be important for breeding programmes, and they can certainly be important for genome research because they provide a homozygous template for genome sequencing, such as the doubled haploid of channel catfish (Waldbieser, Bosworth and Quiniou, 2010) that was used as the sequencing template for whole-genome sequencing (Liu *et al.,* 2016). Although clonal lines have been produced with aquaculture species such as ayu (Taniguchi, *et al*., 1996), amago salmon (Kobayashi *et al.,* 1994) and hirame (Yamamoto, 1999), their large-scale aquaculture has not been realistic. The major purpose of gynogen production has been for research.

2.4 Androgenesis

Androgenesis refers to all-paternal inheritance. Androgens can be produced by irradiating eggs and then doubling the paternal genome. Androgens are more difficult to produce than gynogens (Scheerer et al., 1986), presumably because of the extremely low survival rate of irradiated eggs. Diploidy can be recovered only by blockage of first cell cleavage (Dunham, 2011).

Like gynogenesis, it can be used to produce clonal populations or monosex populations for the purpose of breeding programmes or to elucidate sex-determining mechanisms. If the male is heterogametic, XX and YY androgens will result in equal proportion. If the male is the homogametic, then the androgens will be 100 percent ZZ and all male. YY individuals are viable. As a matter of fact, production of YY fish through androgenesis followed by regular mating with a normal XX female is a major way of producing all male populations in fish (Dunham, 2011; Parsons and Thorgaard, 1985). In many species, males grow faster, and use of monosex populations is of interest to aquaculturists.

2.5 Sex reversal

Sex dimorphism for growth is very common with aquaculture fish species. In some cases, males grow faster, while in other cases females grow faster. Apparently, the generation of a monosex population for faster growth is of interest to aquaculturists. In addition to chromosomal

manipulations such as polyploidy, gynogenesis and androgenesis, monosex populations can also be created by hormonal treatment. Although genotypic sex is established at the time of fertilization, the phenotypic sex is not determined until later in development. For instance, channel catfish phenotypic sex is determined at around 19 days after fertilization. Among various approaches for sex reversal, hormonal sex reversal is the most used approach for aquaculture.

The phenotypic sex can be altered by administration of oestrogens or androgens during the critical period of sex determination. Several androgens, most of them derivatives of testosterone, have been used to produce monosex male populations (Yamazaki, 1983; Dunham, 1990). For instance, 17-methyltestosterone (Dunham, 1990) is widely used for sex reversal in fish. Several estrogenic compounds have been used to produce monosex female populations, of which 3-estradiol is the most commonly used hormone for feminization (Yamazaki, 1983; Dunham, 1990). The hormonal treatment can be applied by bath soaking (Donaldson and Hunter, 1982; Yamazaki, 1983), in feed (Shelton, Rodriquez-Guerrero and Lopez-Macias, 1981), or through implants (Boney *et al.,* 1984), depending on the developmental and culture characteristics of the species (Dunham, 2011). Readers interested in the technical details for the production of monosex population through sex reversal are referred to Dunham (2011).

2.6 Gene transfer

Gene transfer is a process to transfer one or a few foreign gene(s) into an organism. However, the foreign gene can be from other organisms or from the organism itself. The concept is relatively straightforward: if the functions of a gene are well known, then the gene can be potentially transferred into the organism to deliver the functions. For instance, the growth hormone gene was well studied and known for its functions for promoting growth. The transfer of the growth hormone into an organism should promote growth. This simple concept was demonstrated when transgenic mice with the growth hormone gene grew much larger, up to 2.5 times larger than its non-transgenic controls (Palmiter *et al., 1982).*

The first successful gene transfer in fish was demonstrated in 1985 when Zhu, He and Chen transferred the human growth hormone gene in goldfish. Since then, transgenic fish have been produced with various aquaculture species, including rainbow trout (Chourrout *et al*., 1986), channel catfish (Dunham *et al.,* 1987), Nile tilapia (Brem *et al*., 1988) and northern pike (Gross *et al*., 1992).

A number of techniques were developed for transferring the genes of interest into fish, including microinjection (Zhu *et al*., 1985) and electroporation (Inoue *et al*., 1990; Powers *et al*., 1992). Either method works well. However, transgenic technologies suffer from several major lines of shortcomings: the doses of gene transfer cannot be controlled. Most often, multiple copies, often in the form of tandem head to tail arrays are integrated into the genome; the integration sites are random, and such sites can be within a functional gene; and the pleiotropic effect of genes cannot be controlled. The major concern of transgenic fish was the use of foreign DNA, including the promoters and the regulatory DNA sequences such as poly A signals. Liu *et al*. (1990) introduced the concept of all-fish vectors that contain all sequences from fish for transfer into fish.

Transgenic fish have generally performed the way scientists were expecting. For instance, significantly enhanced growth rates were observed with transgenic fish with growth hormone

genes. This has been demonstrated in goldfish, channel catfish, northern pike, Atlantic salmon, rainbow trout, Nile tilapia and common carp, among many other species (reviewed by Dunham, 2011), although failure to observe enhanced growth was also reported (Guyomard *et al.,* 1989; Penman *et al*., 1991). The best example of growth enhancement was demonstrated with transgenic salmon.

In addition to the growth trait, gene transfer was used to improve several other traits, including cold tolerance (e.g. Fletcher and Davies, 1991) and disease resistance (e.g. Anderson, Mourich and Leong, 1996). For instance, transgenic expression of viral coat protein genes, or antisense of viral early phase genes, has been explored in shrimp (Ahanger *et al*., 2014). Attempts for improving disease resistance were made by expression transgenes of lytic peptides, and organisms containing these genes exhibited enhanced disease resistance (Dunham, 2011). Antimicrobial peptide genes were also used to enhance disease resistance. For instance, transgenic fish expressing cecropins were found to be more resistant against bacterial diseases (Dunham *et al*., 2002).

In addition to the purpose of enhancing performance traits of aquaculture species, fish have been considered for the production of pharmaceuticals as biological factories (Dunham, 2011). Fish have potential advantages as bioreactors compared with mammals. These advantages include a short generation interval, low cost of maintenance of the animals, easy maintenance, large numbers of individuals, high-density culture, and mammalian viruses and prions that are not found in fish.

Public acceptance of transgenic fish has been relatively low because of two lines of concerns: food safety concerns; and ecological safety concerns. The question if it is safe to consume transgenic fish has been one major questions from consumers. As aquaculture species have aquatic living environments, tracking of transgenic aquatic animals is more difficult, and therefore the concerns over ecological safety have been serious. However, as reviews and books for gene transfer with aquaculture species are widely available, interested readers are referred to existing resources on the details of transgenic work with aquaculture species.

One commonality shared by all the traditional genetic technologies is that researchers are operating in a black box without the knowledge of the genome, how the genome works, and how the genetic changes will affect the expression and functions of the genome. Modern genetic technologies depend on the knowledge of the genome. The next sections will be cover how the genomes are being studied and the biotechnologies developed for the understanding of genomes, transcriptomes, and gene networks and functions. Using genome information, genome-based technologies have been developed and are continuously being developed for genetic gains in aquaculture species.

3. DNA MARKER TECHNOLOGIES

3.1 History of DNA marker technologies

The development of molecular markers has been one of the major efforts in the first decade of genome research with aquaculture species (Liu and Cordes, 2004). In the early days of aquaculture genome research, most of the efforts were devoted to markers that were readily available without the availability of genome information or resources. This included the development of random amplified polymorphic DNA (RAPD) and amplified fragment length polymorphism (AFLP) markers (Bardakci and Skibinski, 1994; Naish *et al.,* 1995; Liu *et al.,* 1998; Liu *et al.,* 1999; Bagley, Anderson and May, 2001). Although these efforts certainly contributed to the development of genome programmes at that time, these dominant markers are not very useful in terms of polymorphic DNA markers in the long term. Since 1997, great effort has been devoted to microsatellite development in aquaculture species (e.g. Nielsen, Crow and Founatin, 1999; Sakamoto *et al.,* 2000; Rexroad *et al.,* 2001; Nichols *et al.,* 2003; Palti, Danzmann and Rexroad, 2003; Gilbey *et al.,* 2004). Several approaches were adopted, including construction of microsatellite- enriched genomic libraries, targeted sequencing (Serapion, *et al.,* 2004), and identification of microsatellites through data mining using genome resources such as expressed sequence tags (ESTs) or BAC-end sequences (Serapion *et al.,* 2004; Wang *et al.,* 2010; Xu and Guo, 2006; Li *et al.,* 2007; Xu *et al.,* 2006; Liu *et al.,* 2009). Large numbers of useful microsatellites have been obtained through data mining of genome sequence surveys such as BAC-end sequences and EST data mining. However, it is apparent that such huge efforts can be avoided now with the application of next-generation sequencing technologies (see below).

The effort of marker development is shifted to the identification of single nucleotide polymorphisms (SNPs) over time (He *et al.,* 2003; Liu *et al.,* 2011; Sun *et al., 2014;* Wang *et al.,* 2010; Van Bers *et al.,* 2012; Xu *et al.,* 2012; Kongchum *et al.,* 2010; Ryynänen and Primmer, 2006; Lorenz *et al.,* 2010; Andreassen, Lunner and Høyheim, 2010; Gomez-Uchida *et al.,* 2011; Harwood and Phillips; 2011; Hauser *et al.,* 2011; Karlsson *et al.,* 2011; Freamo *et al.,* 2011; Yáñez *et al.,* 2016). SNPs are now the markers of choice for genetic studies because they are the most abundant genetic variations widely distributed in the genome and are generally bi-allelic polymorphisms that are amenable to automated genotyping. SNPs are efficient for genome-wide association studies (GWAS) because linkage disequilibrium can be detected with high- density SNPs when dealing with complex traits (Liu *et al.,* 2014). Once again, the huge efforts devoted in SNP discoveries can now be easily achieved through next-generation sequencing.

3.2 Genomic variations as the basis of polymorphism

Several types of genetic variations can be found by comparing the genomes of individuals within a population with the reference genome sequence of their species: deletions due to the loss of one or more DNA bases; insertions due to the gain of one or more DNA bases; base substitutions; inversions of DNA segments; rearrangements of multiple DNA segments; and copy number variations caused by insertions, deletions, duplications or multiplications of a DNA segment(s), including whole genes.

The most widespread genomic variation among individuals within a population is base substitution. Such base substitution along the DNA chain is defined as SNPs. The second most widespread genomic variation among individuals within a population is probably deletions and insertions. A deletion mutation and an insertion mutation can be viewed as the same phenomenon depending on what is used as the reference, and therefore they are together referred to as indels. The molecular basis for microsatellite marker polymorphism is the insertion or deletion of the microsatellite sequence because when the microsatellite repeats, expands or shrinks, the individuals differ in the number of repeats.

Inversion of a DNA segment in its orientation can be quite widespread in the genome, but this type of variation has not been well studied and probably will not be very useful for large-scale genomic studies.

Copy number variation (CNV) owing to insertions, deletions and duplication or multiplication of a DNA segment is widespread, and this type of genomic variation has recently caught the attention of genome researchers. CNV can involve large or small genome segments that are duplicated or multiplied in one genome while not in another. Such copy number variations can involve genes or just genomic segments that do not harbour genes. Obviously, when genes are involved, the duplicated or multiplied genes can affect genome expression activities. The significance of CNV has caught much attention recently, and CNV could potentially be used for whole-genome selection programmes upon identification of correlation or causation of certain genome segments with performance traits. The importance of CNV in teleost fish is further signified by the fact that teleost fish have an additional round of genome duplication followed with random gene loss, thereby resulting in various CNV situations involving various genes.

3.3 Allozyme markers

Allozymes are protein products of genes that are encoded by a single gene locus (Kucuktas and Liu, 2007). Since they represent genes of known function, they are considered as Type I markers (Liu and Cordes, 2004). The term isozyme refers to multiple biochemical forms of an enzyme having identical substrate specificity (or same catalytic activities) within the same organism. Allozymes, or allelic isozymes, are the different allelic forms of the same enzymes encoded at the same locus (Hunter and Markert, 1957; Parker *et al.,* 1998; May, 2003). Strictly speaking, allozymes represent different allelic forms of the same gene and isozymes represent different genes whose products catalyse the same reaction. However, the two terms are usually used interchangeably. It is believed that isozymes usually form as a result of gene duplication; however, there may be other events (hybridization, polyploidization) that lead to the formation of isozymes.

The most common use of allozyme electrophoresis is to detect genetic variation in natural populations. In the past 30 years, a large amount of allelic frequency data has been collected from many fish species for management purposes. Although the use of allozyme data in aquaculture appears to be limited compared with population studies in fisheries, the aquaculture industry has long used this information for its development because aquaculture and fisheries cannot be separated from each other (Dunham, 2004). Allozyme electrophoresis in aquaculture is used for stock identification, parentage analysis, inbreeding analysis and limited genetic mapping (Liu and Cordes, 2004). However, because the number of allozymes are limited (approximately 100), the broad application of allozyme markers for genomic and genetic research is limited.

3.4 Restriction fragment length polymorphism markers

Restriction fragment length polymorphism (RFLP) markers (Botstein *et al.,* 1980) were the very first generation of markers. RFLP was the most popular approach for analysis of genetic variation during the entire 1980s. As indicated by its name, RFLP is based on DNA fragment length differences after digesting genomic DNA with one or more restriction enzymes. In its original format, genomic DNA is digested by one or more restriction enzymes, separated on an agarose gel, and then followed with a procedure called Southern blot (Southern, 1975) to visualize the fragment length differences (Liu, Liu and Zhang, 2007). After sequencing and polymerase chain reaction (PCR) technologies, RFLP are nowadays analysed using PCR amplification, or PCR amplification followed by restriction analysis depending on if the size is the same.

RFLP is able to detect only large shifts in DNA fragment sizes. Therefore, it can detect only insertions and deletions of large sizes, and the gain or loss of restriction sites. It is unable to detect the vast majority of point mutations, nor deletions or insertions involving just a few base pairs because of its low resolution using agarose gel electrophoresis. As a result, polymorphic rates are low at most loci. The efforts involved in RFLP marker development have been enormous. RFLP attempts to detect genetic variation one locus at a time. The low polymorphic rates, when coupled with expensive and laborious processes, have made application of RFLP limited. It should be particularly noted that RFLP requires previous genetic information, such as the availability of probes or sequence information, information often not available for many fish or other aquaculture species at that time. Although sequence information is widely available now with many aquaculture species, the limited use of RFLP makes it sort of obsolete.

3.5 Mitochondrial DNA markers

The mitochondrial genome evolves more rapidly than the nuclear genome. The rapid evolution of mitochondrial DNA (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 (Liu, 2007c).

Mitochondrial DNA is maternally inherited for the most part, but there are reports of paternal leakage during fertilization (Birky, Fuerst and Maruyama, 1989). Because of the high levels of polymorphism and the ease of mitochondrial DNA analysis, mtDNA has been widely used as markers in aquaculture and fisheries settings. The non-Mendelian inheritance, however, greatly limits the applications of mtDNA for genome research. Nevertheless, as an identification tool often used in aquaculture, mtDNA can be used as a supplemental tool for aquaculture genomic research.

Mitochondrial markers can be analysed using PCR or using restriction fragment length polymorphism (see RFLP markers), except that the target molecule is mtDNA rather than nuclear genomic DNA (Liu and Cordes, 2004). The high levels of polymorphism, the maternal inheritance, and the relatively small size of mtDNA make the RFLP analysis using mtDNA one of the easiest methods for many population studies (Okumuş and Çiftci, 2003; Liu and Cordes, 2004; Billington, 2003). The RFLP polymorphism detected in mtDNA is usually caused by a gain or loss of restriction sites. For example, striped bass (*Morone saxatilis)* exhibits a site loss or gain at the Xba I restriction site, causing an RFLP polymorphism that could easily be detected with PCR amplification of the polymorphic region followed by Xba I restriction digestion. However, polymorphism could also

be caused by insertions or deletions leading to a length variation of mtDNA (Ravago, Monje and Juinio-Meñez, 2002). In this case, electrophoresis of PCR products in the region should provide information on mtDNA haplotypes (Billington, 2003).

Analysis of mtDNA polymorphism has become a useful genetic tool for studies of genetic divergence within and among populations (Avise, 1995). Because mtDNA shows considerable variations among individuals, it is regarded as an effective marker for population structure and geographic variations. Distinct mtDNA lineages are detected in many freshwater fishes in different parts of their species ranges. Since only half of the population (assuming 1:1 sex ratio) pass on their mtDNA to their offspring, the effective population size for mtDNA is smaller than that of nuclear DNA (Harrison, 1989).

Mitochondrial DNA markers have been used extensively to analyse genetic variations in several different aquaculture species, including striped bass (Wirgin and Maceda, 1991; Garber and Sullivan, 2006); channel catfish (Waldbieser, Bilodeau and Nonneman, 2003); walleye (Merker and Woodroff, 1996); salmonids (Nielsen, Hansen and Mensberg, 1998; Crespi and Fulton, 2004); red snapper (Pruett, Saillant and Gold, 2005); and bluegill (Chapman, 1989). Data analysis in mtDNA studies include determining the number of mtDNA haplotypes and calculating the haplotype frequencies and nucleotide diversity.

There are two major drawbacks of the mtDNA markers. One is the non-Mendelian inheritance of mtDNA, and the other is the proportion of the total genomic variation one can observe with mtDNA alone. Additionally, mtDNA markers are subject to the similar problems that exist for other DNA-based markers. For example, in back mutation cases, nucleotide sites that have already undergone substitution are returned to their original state, mutations taking place at the same site on the mtDNA in independent lineages and the unparallel rate of heterogeneity at the same region (Liu and Cordes, 2004) all can place limitations to the validity of using mtDNA for genetic studies.

3.6 DNA barcoding

DNA barcoding involves the amplification and sequencing of a short universal molecular tag of approximately 650 base pairs from the 5′ region of the mitochondrial cytochrome oxidase I (COI) gene (Hebert *et al.,* 2003; Tavares and Baker, 2008). DNA barcoding using COI has been widely employed in various biological fields with proven ability to differentiate closely related species in studies ranging from forensic sciences (Dawnay *et al.,* 2007) to molecular systematics (Hardman, 2005). Importantly, community-based efforts to develop extensive DNA barcode libraries, most notably the Barcode of Life Data System (BOLD, www.boldsystems.org/index.php/databases), has led to the adoption of DNA barcoding technology as the gold standard for species identification and has greatly expanded the power of the technique. The BOLD system provides detailed information of COI-sequenced species, including the origin and current location of voucher specimens (Ratnasingham and Hebert, 2007). Out of almost 30 000 fish species estimated in the world, barcodes for more than 10 000 fish species are currently recorded in the BOLD system. These COI barcodes are gathered from several sources, including the Fish Barcode of Life Initiative (FISH-BOL, 2010) and the Marine Barcode of Life (MarBOL, www.marinebarcoding.org). However, for many species, BOLD barcodes are gleaned from uncurated GenBank records and require additional validation before use.

DNA barcoding is most useful for species identification. It is simple and highly reliable. For aquaculture purposes, the most important application of DNA barcoding is for the protection of consumer interests against mislabelling of food fish. Mislabelling of global food fish products is a growing concern, particularly in western nations that import a large percentage of fish and shellfish products. As food fish choices expand, consumers are faced with identifying products often prepared and processed in ways that remove differentiating anatomical and morphological features. Food fish exporters/importers, processors, wholesalers and restaurateurs may knowingly or unwittingly substitute one species for another as food travels through the supply chain from pond to plate (Wong *et al.,* 2011). Food fish fraud may involve the labelling of a lower-market-value species as a higher-market-value species to realize a larger profit. In the case of international trade, food fish may be mislabelled to avoid protective tariffs, or transhipped or commingled with products from a third country to avoid import duties and/or Food and Drug Administration (FDA) import alerts. Cases of such mislabelling are not uncommon in the media and scientific literature (Wong and Hanner, 2008; Hsieh, 1998; Marko *et al.,* 2004; Miller and Mariani, 2010; Wong *et al., 2011).*

In spite of the value of DNA barcodes for species identification, it is not very useful for genetic and genomic studies because it is maternally inherited as a part of mitochondria.

3.7 RAPD markers

Random amplified polymorphic DNA (RAPD) is a PCR-based multi-locus DNA fingerprinting technique. The RAPD procedure was first developed in 1990 (Welsh and McClelland, 1990; Williams *et al.,* 1990) using PCR to randomly amplify anonymous segments of nuclear DNA with a single short PCR primer (8–10 base pairs in length). Because the primers are short and relatively low annealing temperatures (often 36–40° C) are used, the likelihood of amplifying multiple products is great, with each product presumably representing a different locus. Once different bands are amplified from related species, population or individuals, RAPD markers are produced. RAPD markers thus are differentially amplified bands using a short PCR primer from random genome sites (Liu *et al.,* 1998b, 1999b; Liu, 2007b). Because most of the nuclear genome in vertebrates is noncoding, it is presumed that most of the amplified loci will be selectively neutral. Genetic variation and divergence within and between the taxa of interest are assessed by the presence or absence of each product, which is dictated by changes in the DNA sequence at each locus. RAPD polymorphisms can occur due to base substitutions at the primer binding sites, or to insertions or deletions (indels) in the regions between the sites. The potential power for detection of polymorphism is relatively high; typically, 5–20 bands can be produced using a given primer, and multiple sets of random primers can be used to scan the entire genome for differential RAPD bands. Because each band is considered a bi-allelic locus (presence or absence of an amplified product), polymorphic information content values for RAPDs fall below those for microsatellites and SNPs, and RAPDs may not be as informative as AFLPs because fewer loci are generated simultaneously. However, because of its relatively high level of polymorphic rates, its simple procedure and a minimal requirement for both equipment and technical skills, RAPD has been widely used in genetic analysis, including that of aquaculture species.

The most important applications of RAPD are for species identification, hybrid identification, strain differentiation and, to a much lesser extent, for genetic analysis such as mapping. RAPD markers were widely used in the 1990s, but over time they have become less popular because

they are dominant markers. One of the downsides of dominant markers is that they are less informative than codomintant markers. In fact, dominant markers detect only a single allele for a particular locus and, consequently, do not allow to distinguish homozygous from heterozygous genotypes. RAPD also have poor reproducibility between experiments and laboratories (Liu, 2007b).

3.8 Amplified fragment length polymorphism markers

Amplified fragment length polymorphism (AFLP) is based on the selective amplification of a subset of genomic restriction fragments using PCR (Liu, 2007c). 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 subsets of amplified fragments are then analysed by denaturing polyacrylamide gel electrophoresis to generate the fingerprints.

AFLP analysis can be viewed as an advanced form of RFLP. Therefore, the molecular basis for RFLP and AFLP is similar. First, any deletions and/or insertions between the sites of the two restriction enzymes (e.g. between Eco RI and Mse I, the most commonly used enzymes 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. Third, base substitutions leading to new restriction sites may also produce AFLP. In addition, AFLP also scans for any base substitutions at the first three bases immediately after the two restriction sites. Considering the 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.

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 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.

AFLP has been widely used in aquaculture, such as for analysis on population structures, migration, hybrid identification, strain identification, parentage identification, reproduction contribution, and endangered species conservation (Seki *et al.,* 1999; Jorde, Palm and Ryman, 1999; Sun *et al.,* 1999; Cardoso *et al.,* 2000; Chong *et al.,* 2000; Kai, Nakayama and Nakabo, 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 in the early days of aquaculture genomics (Liu *et al., 1998a, 1999a;* 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; Felip *et al.,* 2005), and analysis of parental genetic contribution involving interspecific hybridization (Young *et al.,* 2001) and meiogynogenesis (Felip *et al*., 2000).

The major weakness of AFLP markers is their dominant nature. Dominant markers, have also a limited experimental reproducibility, makinginformation transfer across laboratories difficult. In addition, AFLP is technically demanding, requiring special equipment such as automated DNA sequencers for optimal operations. The dominant nature of AFLP fundamentally limits its broad applications for genetic analysis. 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. Now with the advances of next-generation sequencing, the fundamental principles of AFLP have been adopted in restriction site-associated DNA sequencing technologies (see below).

3.9 Microsatellite markers

Microsatellites are simple sequence repeats of 1–6 base pairs. They are highly abundant in various eukaryotic genomes, including all aquaculture species studied to date. In most fish genomes, the frequency of detecting microsatellites is approximately one microsatellite per 2–10 kb of DNA. Microsatellite polymorphism is based on size differences due to varying numbers of repeat units characterizing the alleles of a given locus.

Dinucleotide repeats, particularly AC and AG repeats, are the most abundant forms of microsatellites. CG repeats are relatively rare in vertebrate genomes. Partially this is because the vertebrate genomes are often A/T-rich. Among the trinucleotide and tetra-nucleotide repeats, A/T-rich repeat types are generally more abundant than G/C-rich repeat types. Microsatellites consisting of repeats longer than 4 nucleotides (penta- and hexanucleotides) are much less abundant (Toth, Gaspari and Jurka, 2000).

Microsatellites are distributed in the genome on all chromosomes and chromosome regions. They have been found within both coding (e.g. Liu *et al.,* 2001) and noncoding DNA sequences (Toth, Gaspari and Jurka, 2000) even if they are usually more commonly found in noncoding regions (Metzgar, Bytof and Wills, 2000). Only about 10–15 percent 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 and Guo, 2004).

Microsatellites can be viewed as special cases of insertions or deletions. An addition of a dinucleotide microsatellite repeat can be viewed as an insertion of two base pairs into the genome. They are perhaps the most abundant type of insertions and deletions. 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 from one another by as little as one repeat unit, it is desirable to have small PCR amplicons, usually within 200 base pairs. However, due to the

repetitive nature of microsatellites, their flanking sequences can contain quite simple sequences 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. Microsatellite mutation rates have been reported as high as 10^{-2} per generation (Weber and Wong, 1993; Crawford and Cuthbertson, 1996; Ellegren, 2000), which is several orders of magnitude greater than that of nonrepetitive DNA (in the range of 10^{-9} per locus per generation; 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 level of polymorphism. Genotyping of microsatellite markers is usually straightforward. However, due to the presence of null alleles (alleles that cannot 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 disadvantages of microsatellites as markers include the requirement for existing molecular genetic information, a large amount of up-front work for microsatellite development, and the tedious and labour-intensive nature of microsatellite primer design, testing and optimization of PCR conditions.

However, with the advances of next-generation sequencing technologies, the work necessary to discover microsatellites has been minimized. Today, numerous microsatellites can be identified by data mining from genome sequences generated by next-generation sequencing.

Microsatellites have been an extremely popular marker type in a wide variety of genetic investigations. Over the past decade, microsatellite markers have been used extensively in fisheries and aquaculture research, including studies of genome mapping, parentage, kinships and genetic structure of stocks. The major application of microsatellite markers is for the construction of genetic linkage and quantitative trait locus (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. Microsatellite marker protocols can be easily exchanged between laboratories due to the high reproducibility of results, and the use of the same microsatellites for closely related species is sometimes possible if the flanking sequences are conserved (FitzSimmons, Moritz and Moore, 1995; Rico, Rico and Hewitt, 1996; Leclerc, Wirth and Bernatchez, 2000; Cairney, Taggart and Hoyheim, 2000). As a result, microsatellites can also be used for comparative genome analysis. If microsatellites can be tagged to gene sequences, their potential for use in comparative mapping is greatly enhanced (Kucuktas *et al.,* 2009; Ninwichian *et al.,* 2012b).

In spite of the popularity and great utilization of microsatellites, several major limitations of microsatellites restrict them to rise to the top of all marker systems: despite being very abundant, the development of hundreds of thousands or millions of microsatellite markers is practically almost impossible; automation has not been possible for microsatellite genotyping; multiplexing has been limited to about a dozen of loci, at the most; and for the most part, microsatellites can be just associated with traits, but are not usually the causes of the phenotypic variations.

On top of these limitations 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, including its rapid identification and automation in genotyping, make SNP the far more preferred marker system for genome studies.

3.10 SNP markers

Single nucleotide polymorphism (SNP) is a polymorphism caused by a single-base mutation (more often from A to G, or from C to T) that gives rise to different alleles containing alternative bases at a given nucleotide position within a locus. Such sequence differences due to base substitutions have been well characterized since the beginning of DNA sequencing in 1977, but the ability to identify them and to genotype them rapidly in large numbers of samples was not possible until several major technological advances in the late 1990s, and especially after the adoption of next-generation sequencing after 2006.

SNP detection depends on sequencing. Next-generation sequencing technologies, in particular, allows the indetnifcation of large numbers of SNPs scattered throughout the genome. The large-scale sequence data enabled efficient and effective identification of SNPs from genomes of various organisms. For example, using Illumina-based ribonucleic acid sequencing technology, large numbers of SNP markers have been identified from catfish (Liu *et al*., 2011), yielding large sets of gene-associated SNPs within channel catfish and blue catfish. Moreover, whole-genome resequencing of four major channel catfish aquaculture populations and one wild population identified a total of 8.4 million putative SNPs (Sun *et al*., 2014). On average, there is one SNP every 93 base pairs in the catfish genome (Liu et al., 2016). The abundant and high-quality SNPs were used for the development of high-density SNP arrays (Liu *et al*, 2014), enabling large-scale genotyping of genetic markers for GWAS (Geng *et al*, 2015), high-density linkage mapping (Li *et al*., 2015), fine QTL mapping, haplotype analysis, and whole genome-based selection. Similarly, genome-wide SNPs have been identified from common carp (Xu *et al*, 2012), rainbow trout (Palti *et al*, 2015a), pearl oyster (Jones *et al*, 2013), and Atlantic salmon (Yáñez *et al*., 2016).

High throughput and efficient SNP genotyping depends on the development of SNP arrays. The Illumina BeadArray and the Affymetrix SNP, are among the major thechnologies used for SNP identification.

The Illumina BeadArray technology: Illumina offers two viable options for aquaculture researchers interested in SNP genotyping, the GoldenGate assay and the iSelect HD Custom BeadChip. GoldenGate assays rely on allele-specific primer extension for SNP calling. In the GoldenGate assay, DNA samples are first bound to paramagnetic particles. Three oligonucleotides are designed for each SNP locus: two allele-specific oligos and a locus-specific oligo that hybridizes several bases downstream from the SNP site and which contains a bead-specific address. Following hybridization between genomic DNA and assay oligonucleotides, the template-primer

complex is extended with DNA polymerase. Only when extension happens, the allele-specific primer is brought in proximity with the locus-specific primer for ligation. The ligation joins the appropriate allele-specific product (genotype) with the locus-specific primer (address) to form a full-length product that serves as a template for PCR using Cy3- and Cy5-labelled allele-specific primers. The single-stranded, dye-labelled DNAs are hybridized to their complement bead type contained on a BeadChip through their locus-specific primer address, the fluorescent signal captured and SNP is called.

The Illumina iSelect BeadChip uses a related technique, single-base extension (SBE) for SNP calling. In this approach, a two-step allele detection strategy is employed. Amplified, fragmented genomic DNA is first hybridized to bead-bound 50-mer oligos, providing locus specificity. Then SBE is carried out, allowing for the incorporation of a fluorescently labelled dideoxynucleotide for assay readout and SNP calling.

The Affymetrix Axiom genotyping technology: Although the Illumina BeadArray technology is efficient, it has limitations with the densities and the number of SNPs that can be analysed with a reasonable cost. This limitation is overcome by Affymetrix Axiom technology that allows a huge number of SNPs to be analysed, up to millions.

Several SNP arrays have been developed for aquaculture species (Table 1). Because of the high densities used, most of these were constructed using the Affymetrix technology (Table 1). For instance, a catfish 250K SNP array using Affymetrix Axiom genotyping technology has recently been developed (Liu, *et al.,* 2014), which has been used for genome-wide association studies, high-density linkage mapping, fine-scale QTL mapping, haplotype analysis, and whole genome-based selection. More recently, a 690K SNP array is being constructed for catfish.

In addition to SNP arrays, various other methods are available for SNP genotyping. These include Sequenom's MassARRAY, Beckman Coulter's SNPstream platform, the Applied Biosystems OpenArray system and Fluidigm's Dynamic Array (EP1/BioMark system), among many other methods (Liu *et al., 2011).*

TABLE 1
Development of high-density SNP arrays in aquaculture species

Species	SNP array technology	SNP array density	References
Atlantic salmon	Illumina iSelect technology	15K	Gidskehaug *et al*., 2011
Atlantic salmon	Affymetrix Axiom technology	286K	Houston *et al*., 2014
Catfish	Affymetrix Axiom technology	250K	Liu *et al*., 2014
Catfish	Affymetrix Axiom technology	690K	Authors' unpublished data
Common carp	Affymetrix Axiom technology	250K	Xu *et al*., 2014a
Rainbow trout	Affymetrix Axiom technology	57K	Palti *et al*., 2015a

3.11 Restriction site-associated DNA sequencing markers

RAD sequencing, or simply RAD-seq, refers to a method called restriction site-associated DNA sequencing that can identify and score thousands of genetic markers randomly distributed across the target genome from a group of individuals using next-generation sequencing. RAD-seq works

by first fragmenting the target genome using a restriction enzyme. After digestion, a series of molecular processing steps transform the DNA into a fragment library suitable for sequencing on a next-generation sequencing platform. Sequence data are then analysed to identify and score genetic variations in the samples or population of interest. The variations identified mostly are SNP markers in nature. RAD-seq simultaneously identify and genotype SNPs in the samples. RAD-seq is widely used for a variety of molecular genetic studies, including identification of genetic variants (SNPs), phylogenetic analysis, germplasm assessment, analysis of population structure, linkage and QTL mapping, and GWAS analysis.

RAD -seq has been broadly used in aquaculture species. With aquaculture species, RAD-seq has been used for mapping QTLs (Houston *et al*., 2012; Gagnaire *et al*., 2013; Liu *et al.,* 2015; Palti *et al.,* 2015b; for a recent review, see Yue *et al*., 2014). Palaiokostas *et al.* (2013) used RAD-seq to map the sex-linked markers in Atlantic halibut. RAD markers were used for linkage mapping in several aquaculture species, including Atlantic salmon (Gonen *et al.,* 2014), sea urchins (Zhou *et al*., 2015), pearl oysters (Li and He, 2014) and Japanese flounder (Shao *et al.,* 2015), and also for conservation genetic studies (Ogden *et al*., 2013; Huete-Pérez and Quezada, 2013).

4. GENOME MAPPING TECHNOLOGIES

The genomes of aquaculture fish vary from several hundreds of millions of base pairs to several billion base pairs. It is very difficult to study such large genomes without first breaking them into smaller pieces, and then sorting out their relationships, which is the task of genome mapping. There are two distinctive types of mapping methods: genetic linkage mapping and physical mapping which produce genetic linkage maps and physical maps, respectively. While both maps are a collection of genetic markers and gene loci, genetic maps' distances are based on the genetic linkage information and recombination rate between markers, while physical maps use actual physical distances of DNA, usually measured in the number of base pairs.

4.1 Genetic linkage mapping of aquaculture genomes

Researchers begin a genetic map by collecting samples of blood or tissue from family members. The DNA is then isolated from the samples and analysed for association of marker patterns. When markers are inherited mostly together, they are considered linked physically on the same chromosome. When they are always linked together, they are located at the same genetic locus. The tightness of their linkage depends on the distance between the markers. The farther they are, the more likely a recombination can happen between them. Based on the frequency of their recombination, the genetic distances are assigned. The unit of genetic distance is the centiMorgan (cM). A cM corresponds to one percent probability that two loci on a chromosome will be separated through crossing over.

The first steps of building a genetic map are the development of genetic markers and a mapping population. The closer the two markers are on the chromosome, the more likely they are to be passed on to the next generation together; therefore, the co-segregation patterns of all markers can be used to reconstruct their order. The genotypes of each genetic marker are recorded for both parents and in each individual in the progenies. The quality of the genetic maps is largely dependent upon these two factors: the number of genetic markers on the map and the size of the mapping population. The two factors are interlinked, as a larger mapping population could increase the resolution of the map and prevent the map being saturated with stacked markers.

In genetic mapping, any sequence feature that can be faithfully distinguished from the two parents can be used as a genetic marker. Genes, in this regard, are represented by traits that can be faithfully distinguished between two parents. Their linkage with other genetic markers is calculated in the same way as if they are common markers, and the actual gene loci are then bracketed in a region between the two nearest neighbouring markers. The entire process is then repeated by looking at more markers.

All markers that are more or less co-segregated belong to the same linkage group (LG). The number of linkage groups is equal to the number of chromosomes. Thus, a genetic linkage map is composed of polymorphic markers that are assigned to LGs, and their distances within the LG are defined by the recombination fraction among the markers.

Several population types are suitable for genetic linkage mapping. Most often, second hybrid generation (F_2) or higher generation intercrosses are good because in F_2 populations, markers, as well as the traits, are segregating. In a similar fashion, backcross progenies can be used for genetic linkage mapping as well. For loci that are heterozygous in the parents, they would be segregating in first filial generation (F_1) already, and therefore F_1 populations have also been used for genetic linkage mapping.

All types of polymorphic markers can be used for genetic linkage mapping. However, in order to make a linkage map that has a high density of markers, thousands or tens of thousands of markers are needed. In this regard, microsatellites and SNP markers are more appropriate because they are the most abundant types of markers in the genomes. Genetic linkage maps have been constructed in many aquaculture species, and some examples are shown in Table 2, although the marker density and resolution differ greatly. Once again, the marker density provides the level of genome coverage, while the resolution is defined by the number of samples used in the genetic analysis.

TABLE 2
Examples of genetic linkage maps in aquaculture species

Species	Number and type of markers	References
Asian seabass	790 microsatellites and SNPs	Wang *et al*., 2011
Atlantic salmon	5 650 SNPs	Lien *et al*., 2011
Brown trout	288 microsatellites, 13 allozymes	Gharbi *et al*., 2006
Catfish	54 342 SNPs	Li *et al*., 2015
Common carp	732 microsatellites	Zhang *et al*., 2013
Eastern oyster	282 AFLPs	Yu and Ximing, 2003
European seabass	Microsatellites	Chistiakov *et al*., 2005
Grass carp	279 microsatellites and SNPs	Xia *et al*., 2010a
Japanese flounder	1 375 microsatellites	Castaño-Sánchez *et al*., 2010
Pacific oyster	1 166 SNPs and microsatellites	Hedgecock *et al*., 2015
Rainbow trout	2 226 microsatellites and SNPs	Guyomard *et al*., 2012
Scallop	169 microsatellites	Li *et al*., 2012b
Sea bream	321 microsatellites, ESTs and SNP markers	Tsigenopoulos *et al*., 2014
Shrimp	3 959 SNPs	Baranski *et al*., 2014
Tilapia	525 microsatellites	Lee *et al*., 2005
Yellowtail	217 microsatellites	Ohara *et al*., 2005

4.2 Physical mapping of aquaculture genomes

Since actual base-pair distances are generally hard or impossible to directly measure, physical maps are constructed by first shattering the genome into smaller pieces. By characterizing each single piece and assembling them back together, the overlapping path or tiling path of these small fragments would allow researchers to infer physical distances between genomic features. The fragmentation of the genome can be achieved by restriction enzyme cutting or by physically shattering the genome by processes like sonication. Once cut, the DNA fragments are separated

by electrophoresis. The resulting pattern of DNA migration (i.e. its genetic fingerprint) is used to identify what stretch of DNA is in the clone. By analysing the fingerprints, contigs are assembled by automated (FingerPrintedContigs), or by manual means (Pathfinders) into overlapping DNA stretches.

Physical mapping starts with the cloning of large genomic DNA segments into cloning vectors such as bacterial artificial chromosome (BAC) vectors. The genomic segments are approximately 150–200 kb long. Thus, a genome of one billion base pairs would require 5 000–7 000 BAC clones. It takes an average of 6–10X genome coverage to make sure that the whole genome is covered at least once everywhere. These segments are generated by random partial restriction digestions, and therefore analysis of multiple clones (6–10 times) from the same genomic location would allow the determination of their positional relationship. This is usually done by restriction fingerprinting.

A physical map is an ordered set of DNA fragments that aims to cover the entire genome. BACs are the preferred building blocks of physical maps, and in today's context a physical map comprises a set of ordered, overlapping BAC clones. The goal is to identify the smallest number of BACs required to represent the genome (i.e. the minimal tiling path). Such an order is established by placing overlapping restriction patterns one to another to extend to the whole genome. In practical situations, however, contigs would break upon the gaps that are present in the BAC library.

Fingerprinting of BAC clones is conducted by fluorescence-based fingerprinting methods. In general, methods that produce a greater number of bands can detect overlaps more efficiently than methods that produce fewer bands. Labelling the fragments with different colours increases the information content. The procedure of Ding *et al.* (2001) gave the largest number of informative fragments followed by the SNaPshot labelling method (Luo *et al*., 2003).

After fingerprinting, the first stage in constructing a physical map from fingerprint data is to transform the fragment sizes and related information into a data set that can be recognized by the automated physical map assembly software, FingerPrintedContigs (FPC) (Soderlund *et al.,* 2000). FPC considers fragments to be shared by two BAC clones if they have the same size within a given tolerance. The probability that two fragments are shared by chance between clones is calculated as a Sulston cut-off score (Sulston *et al.,* 1988). During the assembly process in FPC, clones are binned together if they satisfy a user-defined cut-off value for fingerprint similarity based on the Sulston score. Automated assemblies are usually performed at high stringency to avoid false inclusions of clones into the same contigs.

Physical gaps result from segments of the genome that are not present in the BAC libraries. This can be mitigated by increasing the number of genome equivalents (usually to at least tenfold) and by using libraries constructed of different restriction enzymes.

Physical maps were constructed for a limited number of aquaculture fish and shellfish species (Table 3). Recent advances in the generation of long sequencing reads may reduce the demands of physical mapping. However, physical maps can provide an independent validation for the whole-genome reference sequences.

TABLE 3
Examples of physical maps constructed from aquaculture species

Species with physical maps	References
Atlantic salmon	Ng *et al*., 2005
Tilapia	Katagiri *et al*., 2005
Channel catfish	Xu *et al*., 2007
Rainbow trout	Palti *et al*., 2009
Common carp	Xu *et al*., 2011
Asian seabass	Xia *et al*., 2010b
Scallop	Zhang *et al*., 2011a

4.3 Radiation hybrid mapping

Irradiation and fusion gene transfer technology has been used for a long time. In the early 1990s, Cox *et al.* resurrected irradiation and fusion gene transfer technology by using a somatic cell hybrid containing only human chromosome 21 as a donor cell line (Cox *et al.,* 1990). They were able to establish a linear order of DNA markers by observing their co-retention in hybrid cell lines. Radiation hybrid mapping strategies are based on the concept that markers that 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 increases as the physical distance between the two markers increases. To provide adequate statistical support for mapping marker retention frequencies, the percentage of times a marker is scored positive in a radiation hybrid (RH) panel is critical. RH mapping is calculated based on the co-retention of markers in fragments across the hybrid panel cell lines. 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 similarly to the use of centiMorgans in linkage mapping. Hence, the observation of chromosome breaks between two markers in RH mapping is analogous to observing recombination between two markers in genetic mapping. To simplify this procedure, Walter *et al.* (1994) reported the development of whole-genome radiation hybrid panels. The benefit of this strategy is that screening all markers on a single panel of ~100 radiation hybrid cell lines can produce high-resolution maps of all chromosomes.

Very different from linkage mapping where polymorphic markers must be used, RH mapping does not require polymorphic markers, but just known DNA sequences of the species for PCR amplification. It prefers the use of gene markers because gene sequences are more likely to be conserved and more likely to be unique in the genome. Because RH mapping defines linkage as co-retention of markers in the same cell after irradiation and fusion, use of gene markers reduce the complexities in PCR amplification.

Radiation hybrid mapping has been used for various mammalian species, but less so for aquaculture species. With aquatic and aquaculture species, it has been only used for just a few species, such as zebrafish, European seabass and gilthead seabream (Dahm *et al.,* 2006; Senger *et al., 2006;* Sarropoulou *et al.,* 2007; Guyon *et al.,* 2010).

4.4 Optical mapping

Optical mapping is a physical mapping method for constructing high-resolution restriction maps of a whole genome from single, fluorescently stained molecules of DNA (Schwartz *et al.,* 1993). In principle, a single DNA molecule can be digested by restriction endonuclease, stained with fluorescence dye, and then the fluorescence is captured optically. The physical lengths of the restriction fragments are recorded. Such ordered, whole-genome coverage of restriction fingerprints is then referred to as optical maps. The advantage of this mapping technology is the total void of cloning or gel electrophoresis.

Although the principles of optical mapping have been established for more than two decades, it has not been widely used until recently. This is largely due to recent technological advances in nanotechnology and the ability to optically capture the fluorescence from a single molecule of DNA. Optical mapping is now mostly used to validate the whole-genome reference sequence assembly (Dong *et al.,* 2013; Reslewic *et al*., 2005; Zhou *et al.,* 2007). Optical mapping has not been used in aquaculture species yet, according to the author's knowledge.

4.5 Integration of physical and linkage maps

Although BAC fingerprinting places BACs into contigs, it neither gives the order of these contigs relative to one another nor their relationships to the genome. The utility of the physical map is greatly enhanced if it can be integrated with other genomic resources such as a high-density linkage map. Identifying contigs that correspond to genetic map assignments provides access to candidate gene regions for QTL and hence the raw material for gene-assisted selection protocols.

One way to integrate physical maps with genetic linkage maps is to genetically map polymorphic markers anchored to known physical map contigs. Such markers can be developed from BAC-end sequencing. This approach was used for map integration of several aquaculture species such as catfish and carp (Li *et al., 2015;* Zhao *et al.,* 2013).

5. GENOME SEQUENCING TECHNOLOGIES

5.1 First-generation DNA sequencers

The double helix DNA structure was revealed in 1953. It took the next 24 years, until 1977, for the world to develop technologies to sequence DNA. Although the Maxam-Gilbert chemical method was invented at the same time and the group shared the Nobel Prize with Frederick Sanger's group, it became rapidly obsolete because of its use of toxic chemicals in the sequencing reactions. The development of Sanger's chain termination, or the dideoxy sequencing technique (Sanger, Nicklen and Coulson, 1977) marks the theoretical maturation of sequencing technologies. The chain-termination technique makes use of chemical analogues of the deoxyribonucleotides, the dideoxyribonucleotides that will terminate the synthesizing DNA chain upon their addition. Sanger sequencing became the most common technology used to sequence DNA for many years, until early 2000. Automated sequencers, most typical of the ABI sequencers based on Sanger sequencing, were developed in the early 1980s and they became the first generation of DNA sequencers.

The first-generation DNA sequencers, such as ABI 3700 or 3730, can sequence 96 samples per run with a read length of 500–800 base pairs per sample. Thus, the output per run was approximately 60 000 base pairs. The cost was very high in today's standards.

5.2 Second-generation (the next generation) sequencers

The pressure to produce more, faster and cheaper DNA sequences that provide basic information pivotal for scientific research and personalized medicine has been the force for the development of DNA sequencing technologies. Starting with the capability to sequence the first base of DNA (Maxam and Gilbert, 1977; Sanger, Nicklen and Coulson, 1977), DNA sequencing technologies have had revolutionary advances reflected in their progress, from manual sequencing to automation, from sequencing a single template to mass parallel sequencing of billions of reads, and from costing about USD 15 per base to less than just a few pennies per million base pairs. A recent review by Heather and Chain (2016) provides a good coverage of the history of DNA sequencing technologies.

The so-called next-generation sequencing started with the launch of the Roche 454 Genome Sequencer FLX System in 2005, followed by the Solexa (now Illumina) sequencing platform, commercialized in 2006. The GS FLX System based on sequencing-by-synthesis (pyrosequencing) technology was developed by 454 Life Sciences as the first next-generation sequencing platform available on the market (Margulies *et al.,* 2005).

Although several aquaculture genomes were initially sequenced using the 454 technology, such as Atlantic salmon, Atlantic cod, rainbow trout, crucian carp, scallops and catfish (Vera *et al.,* 2008; Salem *et al.,* 2010; Hou *et al.,* 2011; Star *et al.*, 2011; Liao *et al.,* 2013), some of the fundamental issues associated with the pyrosequencing technology made the 454 system more vulnerable to sequencing errors. In addition, lower output compared with several other platforms made it rapidly obsolete.

5.3 Illumina sequencers

The Solexa sequencing platform was commercialized in 2006. The principle is based on sequencing-by-synthesis chemistry. Its efficiency and capabilities are powered by massive parallel sequencing of hundreds of millions of templates simultaneously.

In 2008, Illumina introduced an upgrade, the Genome Analyzer II. It offered a powerful combination of the cBot and Paired-End module. cBot is a revolutionary automated system that creates clonal clusters from single molecule DNA templates, preparing them for sequencing by synthesis on the Genome Analyzer. For Genome Analyzer II, the run time was highly decreased and the output per paired-end run can reach 45– 50 Gb. Compared with Sanger sequencing, the Illumina system can produce more data at a reduced time and cost. In 2010, Illumina launched its HiSeq series of sequencers, whose capacity was enhanced to over 400 million reads per lane, or over 3 billion reads per run. Currently, Illumina carries a series of platforms of MiniSeq system (maximum output 7.5 Gb), MiSeq (maximum output 15 Gb), NextSeq series (maximum output 120 Gb), HiSeq series (maximum output 1 500 Gb), and HiSeq X series (maximum output 1 800 Gb). Clearly, the drastic increase of output allowed almost a proportional drop in costs.

In addition to the Illumina system, other systems offer similar output and sequencing quality. These include the Applied Biosystems SOLiD System, which is based on a sequencing-by-ligation technology. The advantage of this platform is its very high sequencing accuracy. However, it appears that Illumina technology is the dominant technology in today's marketplace among all second-generation DNA sequencers.

5.4 Third-generation DNA sequencers

There is no clear definition and clear borderline between the second- and the third-generation sequencers. However, it is generally accepted that the third-generation sequencing technologies are marked by single molecule sequencing (SMS) and real-time sequencing (Heather and Chain, 2016).

The first SMS technology was developed in the early 2000s (Braslavsky *et al.*, 2003; Harris et al., 2008). Helicos was the first company that worked on third-generation sequencing, but it closed in 2012. Currently, the most widely used third-generation technology is the single molecule real-time platform from Pacific Biosciences (van Dijk *et al*., 2014). With the new P6-C4 chemistry, the PacBio sequencers can generate 50 000–100 000 reads per flow cell. This sequencing platform provides long reads of up to 40 kb, with an average of 10–15 kb. Therefore, each flow cell can generate 500 million to 1 billion bases. The long reads provide tremendous advantage for de novo genome sequencing.

One major disadvantage of the PacBio sequencing is its high error rate of almost 10 percent. However, the errors occur randomly along the bases of DNA. High genome coverage would allow generation of consensus sequences. For instance, if the sequences have a 10X coverage, the errors to occur at any given base becomes 1 × 1010. Recently, PacBio has produced a base-correction software, allowing the generation of error-corrected consensus. In some cases, Illumina sequences can be used to correct the sequence errors in PacBio sequences.

In addition to the PacBio sequencing, there are several other third-generation sequencing platforms. Of these, the most particularly promising is the Nanopore DNA sequencing. However, at this time, these platforms are less frequently used in the marketplace.

5.5 Application of next-generation sequencing

Next-generation sequencing technologies are applied in a variety of areas. Some areas of applications include the following: de novo genome sequencing, whole-genome resequencing; marker development for the identification of microsatellites or SNP markers; transcriptome sequencing for the analysis of genome level expression profiling and identification of differentially expressed genes or co-induced genes; large-scale analysis of epigenetic regulation, such as DNA methylation, by deep sequencing of bisulfite-treated DNA; and genome-wide mapping of DNA-protein interactions by deep sequencing of DNA fragments pulled down by chromatin immunoprecipitation.

5.6 Genome sequencing in aquaculture species

With the advances of the sequencing technologies, rapid progress has been and is being made with whole-genome sequencing with aquaculture species. Genome sequencing projects have been initiated with at least 30 fish and shellfish species. These projects included the sequencing of the genomes for several species of carps such as the common carp (Xu et al., 2014b), grass carp (Wang et al., 2015) and rohu carp (personal communication, unpublished), and the sequencing of the genomes of Nile tilapia, channel catfish, European seabass, tongue sole, large yellow croaker, rainbow trout, and the Atlantic salmon, among many others (Table 4). Quite a few shellfish genomes have also been sequenced. These include the Pacific oyster (Zhang et al., 2012), eastern oyster, abalone, shrimp and scallops.

Most of the aquaculture species genomes have been sequenced using the Illumina technology, and in some cases supplemented with third-generation sequencing technologies such as PacBio sequencing. The assembly qualities vary. Four parameters are often used to assess the quality of the genome assemblies: the completeness, the accuracy, the contiguity and the connectivity (Table 4). In addition, the proportion of the genome anchored to the chromosomes is also another important parameter. Completeness refers to the percentage of the genome currentlysequenced; the accuracy refers to the correctness of the assembly. Because of the large genome sizes and the complexities of the genome sequences, errors may exist in some genome assemblies. Some of the genome assemblies are of poor quality, including many mistakes, and a good example is the Pacific oyster's genome assembly, which was reported to include a high percentage of mistakes with its scaffolds (Hedgecock et al., 2015). The contiguity refers to the median length of the contiguous sequences as expressed as N50 length of the contigs. Connectivity refers to the distribution of scaffold sizes as expressed by N50 scaffold size (Table 4).

TABLE 4
Whole-genome sequencing of aquaculture species. Comparison of genome assembly continuity and connectivity of fish and shellfish species. The species are sorted based on the scaffold L50.

Species	Contig L50 (Kb)	Scaffold L50 (Mb)	Percentage on chromosome	Sequencing platform	Total size (Mb)	References
Three-spined stickleback	83.2	10.8	86.9	9.0X Sanger	463	Jones *et al*., 2012
Channel catfish	77.2	7.73	97.2	Illumina, PacBio	783	Liu *et al*., 2016
Grass carp	40.8	6.46	64.0	132X Illumina	900.5	Wang *et al*., 2015
European seabass	53.2	5.09	86.0	30X	675.4	Tine *et al*., 2014
Turbot	31.2	4.3	-	Illumina	544	Figueras *et al*., 2016
Nile tilapia	29.3	2.80	70.9	269X Illumina	1 010	Brawand *et al*., 2014
Zebrafish	25.0	1.55	96.5	7.5X Sanger, Illumina	1 410	Howe *et al*., 2013
Medaka	9.8	1.41	89.7	10.6X Sanger	700.4	Kasahara *et al*., 2007
Common carp	68.4	1.0	51.8	454, Illumina, SOLiD	1690	Xu *et al*., 2014b
Tongue sole	26.5	0.87	93.3	Illumina	477	Chen *et al*., 2014
Atlantic cod	7.1	0.69	44.1	454	753	Star *et al*., 2011
Large yellow croacker	25.7	0.50	-	76X Illumina	644	Wu *et al*., 2014
Rainbow trout	7.7	0.38	54.0	70X Illumina	1 900	Berthelot *et al*., 2014
Fugu	16.5	0.05-0.1	-	5.6X Sanger	332.5	Aparicio *et al*., 2002

5.7 Genome annotation of aquaculture species

Annotation is the process by which the information from raw DNA sequences is added to the genome databases. Strictly speaking, genome annotation is the process of interpreting the genome sequence and it includes the identification of: protein encoding genes; gene structure (e.g. the organization of exons, introns, promoter sequences and regulatory elements); epigenetic information; and polymorphism information such as the positions of SNPs. However, for aquaculture species, the identification of genes, especially protein coding genes, is usually the first step of the genome annotation.

Genomes can be annotated by computer software predictions or by evidence-based approaches. Many software packages are available, but the most used are Fgenesh, and the Augustus. These software packages predict genes based on sequence similarities and the presence of open reading frames for proteins. Evidence-based annotation is based on the identification of transcripts detected from the cells of the organism. In the past, expressed sequence tag (EST) analysis was the major source of the transcript evidence. Recently, the application of ribonucleic acid sequencing (RNA-seq) has allowed rapid progress for transcriptome analysis, as detailed below.

6. TRANSCRIPTOME ANALYSIS OF AQUACULTURE SPECIES

Transcriptome refers to the complete composition of RNAs of an organism. Years ago, EST analysis was the major approach for transcriptome sequencing. Recently, RNA-seq using next-generation sequencing has allowed the most rapid progress.

6.1 Development of expressed sequence tag resources of aquaculture species

EST are single-pass sequences of random complementary DNA (cDNA) clones from cDNA libraries. They are traditionally generated using Sanger sequencing and therefore the resultant sequences are approximately 500 to 800 base pairs in length. Several years ago, because sequencing was relatively cheap, large numbers of ESTs can now be generated at a reasonably low cost from either the 5′ or 3′ end of a cDNA clone to get an insight into transcriptionally active regions. ESTs were used as a primary resource for human gene discovery (Adams *et al.,* 1991). Thereafter, there has been an exponential growth in the generation and accumulation of EST data in public databases for various organisms, with approximately 74 million ESTs now available in these databases (http://www.ncbi.nlm.nih.gov/dbEST; March 2016 release number 130101, all species).

EST analysis is an effective genomic approach for rapid identification of expressed genes and has been widely used in genome-wide gene expression studies in various tissues, developmental stages and under different environmental conditions (Franco *et al.,* 1995). In addition, the availability of cDNA sequences has accelerated further molecular characterization of genes of interest and provided sequence information for microarray construction and genome annotation (Rise *et al.,* 2004).

Gene expression analysis plays an important role in identifying differentially expressed genes under different environmental conditions and gene expression regulation, shedding light on gene functions. EST analysis has been demonstrated effective for detection of differential expression and regulation of certain genes. Without normalization or subtraction in library construction, the number of the sequenced ESTs for a given gene reflected the abundance of the gene expression at the corresponding scenario (e.g. environmental conditions, developmental stages, treatments).

Direct EST sequencing is inefficient in discovery of rarely expressed genes. To solve this problem, the method to construct normalized cDNA libraries was developed (Bonaldo, Lennon and Soares, 1996; Soares *et al.,* 1994). The basic principle is using hybridization to reduce redundant genes and increase the representation of rarely expressed genes. However, with the adoption of RNA-seq technology, as described below, it is unlikely that additional EST analysis will be conducted because of the low efficiency as compared with RNA-seq.

Initial annotation of ESTs can be conducted by simple sequence similarity comparisons. Further annotation analysis can be carried out after obtaining the consensus sequences (putative unigenes), such as determination of gene identity based on homology search, open reading

frame identification, gene ontology annotation and gene-enrichment analysis (e.g. Nakaya *et al., 2007).*

In order to assign gene identity to contigs and singletons, homology search is widely used. Such an approach is especially helpful for newly studied species. BLAST is the most widely used programme to obtain high throughput EST analysis and annotation results. The BLAST package provides different flavours of algorithms for sequence similarity searching. BLASTX is used to search against protein database by translated consensus EST sequences, while BLASTN is used to search against nucleotide sequence databases. The National Center for Biotechnology Information (NCBI), ENSEMBL and Swiss-Prot are three important databases for BLAST search. For instance, the Swiss-Prot database has a fully manually curated and annotated unigene database, Uniprot, which can be used for identifying putative function for unigene by BLASTX. The NCBI provides a dbEST database that can be used to search novel transcripts by BLASTN. The dbEST is a main EST resource database, including ESTs for over 200 aquaculture species. The ENSEMBL database can provide chromosome location information of genes, which is a useful tool for comparative genome analysis. However, the BLAST sequence similarity comparison provides only information on sequence homology and it is not recommended to enetirely rely on BLAST for gene identification. In-depth phylogenetic analysis and/or orthology analysis is needed to determine the identities of genes.

For a greater level of annotation, the open reading frame (ORF) is identified to determine the full or portion of a coding region in the unigene. The unigene with a full ORF usually represents a full-length cDNA. There are some useful tools for ORF detection. For example, ESTScan (Iseli, Jongeneel and Bucher, 1999) can extract coding regions from low-quality ESTs and correct frame shift errors. OrfPredictor (Min *et al*., 2005) is another programme for identification of protein-coding sequences from ESTs through predicting the most probable coding regions from all six translation frames.

Gene ontology (GO) annotation can provide descriptions of gene products behaving in a cellular context. Gene functions are placed into three categories: biological processes, cellular components and molecular functions. Consensus sequences can be linked to GO terms and assigned a possible function by Blast2GO (Conesa *et al.,* 2005).

Gene ontology enrichment analysis is to cluster most relevant GO terms associated with certain biological pathways. GOEAST (Zheng and Wang, 2008), Ontologizer (Bauer *et al.,* 2008), GeneTrail (Backes *et al.,* 2007), and DAVID functional annotation tool (Huang, Sherman and Lempicki, 2009) are useful tools for these analyses.

EST analysis is an efficient approach for gene discovery and gene identification. For instance, between 2001 and 2007, catfish ESTs increased from 10 000 to 44 000 and the putative genes number increased

from 5 905 to 25 000 (Li *et al., 2007).* In Pacific oyster (*Crassostrea gigas),* 40 845 high-quality ESTs represented 29 745 unique transcribed sequences (Fleury *et al.,* 2009); gilthead sea bream (*Sparus auratus),* 30 000 ESTs represented 18 196 putative unigenes (Louro *et al.,* 2010). Currently, there are over 180 aquaculture species having more than 100 ESTs in dbEST, with approximately a dozen species having over 10 000 ESTs (Table 5).

EST analysis can provide comparisons of gene expression profiling in different tissues and conditions. For instance, in a recent study with rainbow trout (*Oncorhynchus mykiss),* Kondo *et al. (2011)* sequenced over 30 000 ESTs from rainbow trout adipose tissue. These ESTs were used to search adipokine-related genes. The result showed that none of them encoded adipokine and PPARγgene, which play important roles in mammalian adipocytes. Further qRT-PCR results confirmed EST analysis results, that is, rainbow trout adiponectin transcripts were weakly detected in adipose tissue but strongly detected in muscle, suggesting the difference of energy metabolism between fish and mammals (Kondo *et al.,* 2011). Chini *et al.* (2008) constructed normalized cDNA libraries from liver, ovary and testis in bluefin tuna (*Thunnus thynnus),* identifying several sequences with known functions in other organisms, but not previously described in this species. Also, sequences were described being expressed in one, two or more tissue libraries. Similarly, Zou *et al. (2011)* constructed normalized cDNA libraries from testis, ovary and mixed organs of mud crab (*Scylla paramamosain).* Through EST analysis, sex-specific transcripts were identified.

TABLE 5
Some examples of aquatic species with major EST resources (>10 000) using zebrafish as a reference

Species	Number of ESTs
Danio rerio (zebrafish)	1 488 275
Ciona intestinalis (ciona)	1 205 674
Xenopus laevis (African clawed frog)	677 911
Oryzias latipes (Japanese medaka)	666 891
Salmo salar (Atlantic salmon)	498 245
Ictalurus punctatus (channel catfish)	354 516
Oncorhynchus mykiss (rainbow trout)	287 564
Crassostrea gigas (Pacific oyster)	206 388
Litopenaeus vannamei (white shrimp)	161 248
Ictalurus furcatus (blue catfish)	139 475
Oreochromis niloticus (Nile tilapia)	120 991
Petromyzon marinus (sea lamprey)	120 731
Sparus aurata (gilthead seabream)	79 216

EST resources provide sequence information for microarray development. For instance, in a recent study, Booman *et al.* (2011) developed a large-scale oligonucleotide microarray platform containing 20 000 features (20K), which was used to study immune response of the Atlantic cod spleen with stimulation of formalin-killed atypical Aeromonas salmonicida (Booman *et al.,* 2011). Similarly, oligo microarray for gilthead sea bream (*Sparus aurata)* was developed based on ESTs, and the microarray was used to identify 1 050 differentially expressed genes between two developmental stages (Ferraresso *et al*., 2008).

Although EST analysis has been important for transcriptome characterization, it is now becoming expensive, relative to several of the recently developed approaches, as detailed below. However, EST resources still have a great value to serve as reference for RNA-seq analysis. It has been found that ESTs are useful for high-*quality* reference-guided assembly of next-generation sequencer-generated short reads (Liu *et al., 2011).*

6.2 RNA-seq technologies

RNA-seq is a technology to sequence transcriptomes using next-generation sequencing technologies. It has been widely used for analysis of gene expression profiling and identification of differentially expressed genes. RNA-seq can be done with several sequencing platforms, including the Illumina sequencing platform, ABI Solid Sequencing, and less efficiently, the Life Science's 454 sequencing. Of these, the Illumina HiSeq sequencers are the most popular because of their very high throughput and accuracy of sequencing reads.

RNA-seq analysis starts with RNA samples. Before RNA-seq, considerations need to be made to allow statistical analysis of the results with proper biological and technical replica. For instance, a control sample can be compared with the infected samples at various times after infection. In order to allow statistical analysis of the results, a minimal of three biological replications is required for each condition (treatment).

The biological issues of RNA-seq are actually extremely simple. The researchers are interested in which genes are expressed, how much they are expressed, and how the samples in different conditions compare. Before answering these questions, the immediate task is to assemble the short reads into reference transcriptome. In general, one of the two types of assembly methods can be used for the assembly of RNA-seq sequences, depending on the existing genome resources. If a reference genome sequence is available, reference-guided assembly methods can be used. In contrast, *de novo RNA-seq* assembly methods must be used in the absence of a reference genome sequence.

RNA-seq has been extensively used for the identification of gene-associated markers. In catfish, hundreds of thousands of gene-associated SNPs have been identified by deep sequencing of RNA from many individuals of both channel catfish and blue catfish, which will be used in the development of high-density catfish SNP chips for genome-wide association studies (Liu *et al.,* 2011). In a study to understand the adaptive divergence between dwarf and normal lake whitefish species, the 454 sequencing was used with the aim to generate a set of SNP markers; 89 SNPs showed pronounced allele frequency differences between sympatric normal and dwarf whitefish (Renaut, Nolte and Bernatchez, 2010).

RNA-seq data sets are being deposited to the NCBI's Sequence Read Archive (SRA) at www.ncbi.nlm.nih.gov/sra. The RNA-seq data sets can be searched, retrieved and downloaded from the SRA database. For instance, searching of SRA using catfish as the keyword resulted in 109 records, as of March 2016. Researchers can download any of the records for additional analysis. Such data sets are most often used for meta-analysis.

6.3 Analysis of differentially expressed genes using RNA-seq

RNA-seq has many applications. However, the two most common analyses using RNA-seq are the identification of differentially expressed genes during development, under a specific physiological condition, or after certain treatments. The basic principle for the identification of differentially expressed genes is to calculate the ratio of counts of short reads aligned to each gene before and after treatment. For instance, if 50 RNA-seq reads align to a certain gene after infection while only 5 RNA-seq reads align to the same gene before infection, then the infection induced tenfold the expression of this gene. This is to demonstrate the concept. In practice, however, detailed normalization and statistical analysis must be conducted.

With RNA-seq, the higher the expression level a gene has, the more reads are sequenced from this gene. However, the read count is correlated with the size of the gene exons. Therefore, the sizes of the gene must be normalized to compare their expression. Reads per kb per million (RPKM) has been widely used as a reasonable normalizer. RPKM is the number of reads per kb long of a transcript per 1 000 000 reads of RNA-seq sequences. RPKM = 109 × C/NL, where: C is the total number of reads mapped onto the gene; N is the total number of mapped reads; and L is the sum of the genes in base pairs. For instance, the RPKM of a 2 kb transcript with 3 000 alignments in a sample of 10 million of mapped reads is calculated as:

RPKM = 109 × (3 000 / 2 000) × 10 000 000 = 150

Once all the read counts are normalized into RPKM, expression levels among genes and before and after treatment can be calculated.

6.4 Analysis of co-expressed genes using RNA-seq data sets

In addition to the identification of differentially expressed genes, RNA-seq can be used to determine correlated or coordinated expression. Such analysis can be quite insightful. For instance, a single treatment such as high temperature treatment may induce a common set of genes, and these genes may all have similar functions in dealing with the heat shocks.

The correlation of expression patterns among various genes, referred to as co-expression, can be revealed by network analysis. An increasing number of studies have demonstrated associated behaviour of genes with related biological functions (Carter *et al.,* 2004; Rocke and Durbin, 2001). Given that most biological processes cannot be carried out by a single gene, analysis of co-expressed genes from RNA-seq data sets may be quite informative.

6.5 Gene ontology, enrichment analysis and pathway analysis

Gene ontology (GO) is a controlled vocabulary term to describe gene characteristics in terms of their localization and function. Transcriptome information can be analysed as to what genes are enriched after a certain treatment. This can be accomplished by sequential analysis of GO, followed by enrichment analysis to determine which GO terms are enriched after the treatment, and then by pathway analysis to determine what gene pathways these enriched genes are involved in. Such analysis can provide functional insights into the induced or suppressed expression patterns.

Gene ontology was initially developed by researchers studying the genome of three model organisms: *Drosophila melanogaster (fruit* fly), *Mus musculus (mouse)* and *Saccharomyces cerevisiae* (yeast) in 1998 (Gene Ontology Consortium, 2010). Now, databases for many other model organisms have joined the Gene Ontology Consortium and made contributions to this project (Gene Ontology Consortium, 2015). The GO project provides three structured ontologies that describe gene products in terms of their biological processes, cellular components and molecular functions in a species independent manner. Several software packages have been developed for GO analysis with the most popular being Blast2GO.

Gene set enrichment analysis is a method to identify classes of genes or proteins that are over-represented in a set of genes or proteins (Subramanian *et al.,* 2005). The method uses statistical approaches to identify significantly enriched or depleted groups of genes. The principal foundation of enrichment analysis is that the gene set should have a higher chance to be selected if its underlying biological process is abnormal under a given condition (Huang, Sherman and Lempicki, 2009). Multiple software packages have been developed for gene enrichment analysis. Gene set enrichment analysis is the most popular approach for enrichment analysis. Following the enrichment analysis, gene pathway analysis can be conducted to determine what pathway is operating under the condition of analysis.

6.6 Analysis of allele-specific expression

A diploid organism has two sets of chromosomes and thereby two alleles at a given locus. Allele-specific expression (ASE) refers to the phenomenon that the two alleles are not equally expressed, up to the exclusive expression of only one of the two alleles. A number of recent studies have demonstrated that allele-specific gene expression is common (Lo *et al.,* 2003; Yang *et al.,* 2003; Chen *et al.,* 2016). Interest in the existence of ASE in non-imprinted autosomal genes has increased with awareness of the important role that variation in non-coding DNA sequences can play in determining phenotypic diversity (Knight, 2004).

A number of approaches have been used for the detection of ASE. In early studies, ASE was detected by a single-base extension of a primer adjacent to the variable single nucleotide polymorphism (Carrel and Willard, 2005; Cowles *et al.,* 2002). Several recent studies applied a variety of technologies to scale up the tested genes (Guo *et al.,* 2008), of which the array-based approach was the most widely used. Several array-based ASE studies have been published in the past decade (e.g. Bjornsson *et al.,* 2008; Daelemans *et al.,* 2010). Owing to rapidly increasing throughput and decreasing costs, next-generation sequencing is rapidly replacing array-based technology for functional genomic assays (Rozowsky *et al., 2011).* In addition, the ability to resolve single-base differences, digital quantification, and comprehensive genome-wide coverage provides information on the abundance and the allelic biases in transcripts or regulatory DNA, which otherwise could not be achieved using hybridization-based arrays (Wood *et al.,* 2015). RNA-seq technology using high-throughput sequencing platforms allows for relatively unbiased measurements of expression levels across the entire length of a transcript. This technology has several advantages, including the ability to detect transcription of unannotated exons, to measure both overall and exon-specific expression levels and to assay allele-specific expression (Pickrell *et al., 2010). Notably,* RNA-seq is the only technology that provides concurrent allelic and total expression data.

Measuring ASE is vital to better understanding global mechanisms of genetic variations. ASE analysis has been widely employed in mammals, insects and plant systems (Bell *et al.,* 2013; Combes *et al.,* 2015; Gregg *et al.,* 2010; Serre *et al.,* 2008; Shi *et al.,* 2012; Springer and Stupar, 2007; Wittkopp, Haerum and Clark, 2008). In aquaculture species, despite many reports regarding expression of specific genes, very little is known about ASE (Murata, Oda and Mitani, 2012; Shen *et al.,* 2012). Recently, Chen *et al.* (2016) reported that ASE is highly enriched with ribosomal protein genes. With more and more application of RNA-seq in aquaculture species, it is inevitable that it will be used for the analysis of ASE.

7. UNDERSTANDING THE NON-CODING PORTIONS OF THE GENOMES

During the first two decades of molecular biology research in the 1970s and 1980s, it was widely believed that only a small fraction, 1–5 percent of the genome, was transcribed. Part of the reason for this was due to technological limitations for the detection of transcripts expressed at low levels. However, this notion was challenged by the discovery of new classes of regulatory non-coding RNAs (ncRNAs). As such, the term "transcript" is now used not only to refer to protein-encoding messenger RNA (mRNA), like in the original usage, but also to various transcriptional products that cover almost the entire genome. Thus, the concept of pervasive transcription evolved to include various non-coding RNAs (Mercer, Dinger and Mattick, 2009) in addition to the traditional mRNA, rRNA (ribosomal RNA) and tRNA (transfer RNA). The proportion of such non-coding RNAs may vary among species but there is increasing awareness that it is significantly represented. In humans, for instance, it was found that only one-fifth of the transcription across the genome is associated with protein-coding genes (Kapranov *et al.,* 2007).

Various types of non-coding RNAs have been identified, including long non-coding RNAs (lncRNAs), microRNAs (miRNAs), short interfering RNAs (siRNAs), Piwi-interacting RNAs (piRNAs), and small nucleolar RNAs (snoRNAs) (Kapranov *et al.,* 2007).

7.1 Long non-coding RNAs

The lncRNAs are non-coding RNAs whose sizes are greater than 200 bases. Such classification is arbitrary, but is based on practical considerations, including the separation of RNAs in experimental protocols.

The functions of non-coding RNAs are being unravelled. Recent research has indicated that lncRNAs could be involved in a number of functions (Huarte and Rinn, 2010; Pauli *et al.,* 2012), including: interacting with and modulating the activity of the chromatin modifying machinery (Huarte *et al.,* 2010; Nagano *et al.,* 2008; Rinn *et al.,* 2007; Tian, Sun and Lee, 2010); serving as the decoys in the sequestration of miRNAs (Poliseno *et al., 2010),* transcription factors (Hung *et al.,* 2011), or other proteins (Tripathi *et al.,* 2010); and serving as precursors for the generation of sRNAs (Kapranov *et al.,* 2007). In addition to the co-expression of lncRNAs and mRNAs, the co-localized expression of lncRNA and protein coding genes were also observed (Ponjavic *et al.,* 2009), suggesting their cooperative actions and/or sharing of cis-regulatory elements in the transcription process. In many instances, the act of lncRNA transcription alone is sufficient to regulate the expression of nearby genes (e.g. Martens, Laprade and Winston, 2004; Petruk *et al.,* 2006; Wilusz and Sharp, 2013), or distant genes through modification of chromatin complexes (e.g. Tsai *et al.,* 2010), or binding to transcription elongation factors (Yang, Froberg and Lee, 2014).

Analysis of lncRNA in aquaculture species is very limited. A few studies are being conducted in rainbow trout and catfish (Al-Tobasei, Paneru and Salem, 2016; authors' unpublished data), but these investigations are still in the stage of infancy.

7.2 MicroRNAs and their target genes

Several distinct classes of small non-coding RNAs, including miRNA, siRNA, piRNA, and repeat- associated short interfering RNA (rasiRNA), have been identified. These molecules are typically ~18– 40 nucleotides in length and play profound roles in many cellular processes. Among the non-coding RNAs, the functions for the small regulatory non-coding RNAs such as miRNA are probably the best studied (Bartel, 2009). The miRNAs, with a length of ~22 nucleotides, play critical roles in post-transcriptional regulation of gene expression.

Several studies have been conducted for the analysis of microRNA in aquaculture species. These include analysis in tilapia (Yan *et al.,* 2012a, 2012b), sea cucumber (Li *et al.,* 2012), Atlantic cod (Johansen *et al.,* 2011), and channel catfish (Barozai *et al.,* 2012). However, functional analysis of their target genes in aquaculture species is rarely existent.

7.3 ENCODE project and FAANG project

After the assembly of the human genome sequence, in 2003, the National Human Genome Research Institute in the United States of America launched the ENCODE (Encyclopedia of DNA Elements) project. The objective of ENCODE was to obtain a complete list of the functional elements of the human genome, including the elements that act at the protein and RNA level, as well as the regulatory elements fortranscription, post-transcriptional regulation, translation and replication. This type of project has been extended to annotate the animal genomes, including those of aquaculture species, and this project is called FAANG (Functional Annotation of ANimal Genomes, www.faang.org/plan).

8. GENETIC ANALYSIS TECHNOLOGIES

8.1 Traits important for aquaculture

The practical purpose of aquaculture genomics and genetics studies is to reveal the genetic basis of performance and production traits and use such information for genetic breeding programmes. In the aquaculture sector, domestication is a very recent event for many species and, therefore, many farmed aquatic organisms are still genetically and phenotypically similar to their wild relatives.

Many traits are important for aquaculture production. These include growth rates, feed conversion efficiency, disease resistance, low oxygen tolerance, stress tolerance, processing yields, sexual maturation time, robustness, body conformation and reproductive traits. In catfish for example, phenotypic data sets have been produced for various traits that are important for aquaculture. Such data sets, along with genetic pedigrees, are important resources for genetic and QTL mapping as well as genome-wide association studies.

8.2 Quantitative trait locus mapping in aquaculture species

The fundamental goal of aquaculture genomics in the practical sense is to understand the basis for performance and production traits. Because most aquaculture traits are complex and likely controlled by multiple genes, QTL mapping is the core of applied aquaculture genomics. In recent years, great efforts and good progress have been made in this area. QTL analyses have been conducted in several dozen aquaculture species and some of these examples are summarized in Table 6. The studied traits include growth rate, disease resistance, sex maturation time, body conformation, fat content, response to stress, swimming abilities, salinity tolerance, muscle traits, osmoregulation capacities and smoltification, among other traits. Of these, the largest amount of efforts has been devoted to QTL mapping of growth traits and disease resistance.

TABLE 6
Examples of QTL studies in aquaculture species

Species	Traits	Reference
Arctic charr	Body weight, condition factor and age of sexual maturation	Moghadam *et al.*, 2007; Küttner *et al.*, 2011
Arctic charr	Salinity tolerance	Norman *et al.*, 2011
Asian seabass	Growth	Wang *et al.*, 2006; 2011
Asian seabass	Resistance against viral nervous necrosis disease	Liu *et al.*, 2015a
Asian seabass	Omega-3 fatty acids contents	Xia *et al.*, 2014
Atlantic salmon	Body weight and condition factor	Reid *et al.*, 2005
Atlantic salmon	Adaptive traits	Boulding *et al.*, 2008
Atlantic salmon	Growth	Baranski, Moen and Våge, 2010

(cont.)

Species	Traits	Reference
Atlantic salmon	Resistance against infectious pancreatic necrosis virus	Gheyas *et al*., 2010; Houston *et al*., 2010; Houston *et al*., 2008; Moen *et al*., 2009
Atlantic salmon	Resistance against infectious salmon anaemia	Moen *et al*., 2007
Atlantic salmon	Flesh colour	Baranski, Moen and Våge, 2010
Atlantic salmon	Life history	Vasemägi *et al*., 2010
Atlantic salmon	Resistance to pancreas disease	Gonen *et al*., 2015
Atlantic salmon	Late sexual maturation	Gutierrez *et al*., 2014
Blacklip abalone	Growth	Baranski *et al*., 2008
Catfish	Columnaris disease resistance	Geng *et al*., 2015
Catfish	Head size and shape	Authors' unpublished data
Clam	Clam metrix	Lu *et al*., 2013
Coho salmon	Hatch timing, weight, length and growth	McClelland and Naish, 2010
Common carp	Muscle fiber-related	Zhang *et al*., 2011b
Common carp	Growth	Boulton *et al*., 2011
Common carp	Body weight, body length and body thickness	Laghari *et al*., 2015
Common carp	Swimming ability	Laghari *et al*., 2014
Eastern oyster	Disease resistance	Yu and Guo, 2006
European seabass	Growth	Louro *et al*., 2016
European seabass	Body weight, morphometric traits and stress response	Massault *et al*., 2010
Gilthead sea bream	Sex determination and body growth	Loukovitis *et al*., 2011
Gilthead sea bream	Resistance to fish pasteurellosis	Massault *et al*., 2011
Gilthead seabream	Skeletal deformities	Negrín-Báez *et al*., 2015
Japanese flounder	*Vibrio anguillarum* resistance	Wang *et al*., 2014
Kelp grouper	Growth	Kessuwan *et al*., 2016
Large yellow croaker	Growth	Ye *et al*., 2014
Pacific abalone	Growth	Liu *et al*., 2007
Pacific oyster	Growth	Guo *et al*., 2012
Pacific oyster	Resistance against summer mortality	Sauvage *et al*., 2010
Pacific oyster	Growth	Guo *et al*., 2012
Rainbow trout	Upper thermal tolerance	Jackson *et al*., 1998; Danzmann, Jackson and Ferguson, 1999; Perry *et al*., 2001; Perry *et al*., 2005
Rainbow trout	Life history	Leder, Danzmann and Ferguson, 2006
Rainbow trout	Spawning time	O'Malley *et al*., 2003; Colihueque *et al*., 2010
Rainbow trout	Osmoregulation capacities	Le Bras *et al*., 2011
Rainbow trout	Development rate	Robison *et al*., 2001; Easton *et al*., 2011
Rainbow trout	Whirling disease resistance	Baerwald *et al*., 2011

(cont.)

Species	Traits	Reference
Rainbow trout	Growth	Wringe *et al*., 2010
Rainbow trout	Smoltification	Nichols *et al*., 2008
Rainbow trout	Bacterial cold-water disease resistance	Vallejo *et al*., 2014a; Palti *et al*., 2015b
Rainbow trout	*Flavobacterium psychrophilum* resistance	Vallejo *et al*., 2014b
Rainbow trout	Osmoregulation capacity	Le Bras *et al*., 2011
Rainbow trout	Cortisol response to crowding	Liu *et al*., 2015b
Rainbow trout	Response to crowding stress	Rexroad *et al*., 2013
Tilapia	Sex determination	Cnaani *et al*., 2007; Shirak *et al*., 2006
Turbot	Growth	Sánchez-Molano *et al*., 2011
Turbot	*Aeromonas* resistance	Rodríguez-Ramilo *et al*., 2011
Turbot	Resistance against *Philasterides*	Rodríguez-Ramilo *et al*., 2013

8.3 Quantitative trait locus analysis of disease resistance

Disease resistance is among the most important traits for aquaculture species and great efforts have been made in mapping QTLs controlling this trait.

The QTL effect for disease resistance varies. In some cases, major QTLs have been found that suggested a single gene or just a few genes may be operating and, in these situations, the mapped QTL is very useful for marker-assisted selection. For instance, several QTLs were mapped in Atlantic salmon for resistance against the infectious pancreatic necrosis (IPN) virus. One QTL on linkage group (LG) 21 can explain 25 percent of the observed within-family variance in the overall data set, while the QTL on LGs 26 and 19 were estimated to explain 18 and 9 percent of the variance, respectively, suggesting that these QTLs are major.

In other situations, multiple QTLs were mapped with each having a relatively small effect. Most disease resistance QTLs have a relatively small effect, suggesting many genes are involved in the resistance. In addition, it may also suggest that the phenotypic evaluation is difficult, and the environment effect may be large such that the percentage of phenotypic variation explained by the QTL is small. For example, in Asian seabass, viral nervous necrosis disease causes mass mortality in mariculture. Using 149 microsatellites, Liu *et al.* (2015a) successfully mapped several QTLs located in different LGs. These results suggested that viral nervous necrosis resistance in Asian seabass is controlled by many loci with small effects. The QTL with the largest effect accounted for only 2–4 percent of the phenotypic variations. In Atlantic salmon, a QTL was mapped to LG 8 that accounted for 6 percent of the phenotypic variation (Moen *et al., 2007).*

8.4 Genome-wide association studies of performance traits

Genome-wide association study (GWAS) is another method for mapping genes involved in performance traits. GWAS is different from QTL mapping in several aspects: QTL analysis requires

genetically structured families with pedigree information, while GWAS typically uses genetically unrelated individuals; and QTL mapping relies on the detection of genetic linkage, while GWAS depends on the detection linkage disequilibrium between the trait under study and the related markers.

GWAS has been extensively used for genetic analysis of genetic diseases in humans. It has also been used for aquaculture species. For instance, Geng *et al.* (2015) used GWAS to identify genes associated with disease resistance against columnaris disease in catfish (Geng *et al.,* 2015). In another study, Tosh (2014) used GWAS to identify genes involved in harvest weight. Most recently, GWAS was used to identify associated markers with fillet yield in rainbow trout (Gonzalez-Pena, 2016). However, the application of GWAS in aquaculture species is still at the earliest stage.

9. GENOME-BASED GENETIC SELECTION TECHNOLOGIES

9.1 Marker-assisted selection

Marker-assisted selection is also called marker-aided selection (MAS) and is a process whereby a selection decision is made based on the genotypes of DNA markers. MAS is especially useful for traits that are difficult to measure, lethal to measure, exhibit low heritability, and/or are expressed late in development. Its implementation requires information of DNA markers that are tightly linked to QTL for traits of interest based on QTL mapping or association studies (Lande and Thompson, 1990). Ideally, the DNA markers should be the causative mutation underlying the phenotypic variation. However, for practical purposes, it would not make a difference if the DNA marker is always linked with the trait of interest even if the marker variation is only correlated, but not the cause of the phenotypic difference. In order to implement MAS, QTLs need to be mapped and validated within the breeding populations. MAS has been applied mostly with plants and livestock animal species but less with aquaculture species, although a few good examples exist for application in aquaculture (Ozaki *et al., 2012).*

The best example of MAS in aquaculture species is perhaps the situation of Japanese flounder. A microsatellite locus, Poli9-8TUF, was mapped near the major QTL for resistance to lymphocystis disease. Additional analysis indicated that the disease resistance was controlled by a single gene, and the resistance allele was dominant. Based on the marker linkage information, Fuji *et al.* (2007) developed a new population of Japanese flounder using MAS with the marker Poli9-8TUF. They selected a female homozygous for the favourable allele (B-favourable) and a male with a higher growth rate and good body shape, but without the resistant allele as parents. In the females, the marker Poli9-8TUF is tightly linked to the QTL for resistance to lymphocystis disease; therefore, a female was selected as the linkage disequilibrium-resistant parent. The B-favourable allele was transmitted from the mother to the progeny. All the progeny are heterozygotes with the resistance allele, and the progeny was entirely resistant to lymphocystis disease, while the control group without B-favourable alleles showed incidences of 4.5 and 6.3 percent of mortality due to lymphocystis disease. These results clearly demonstrate that MAS is an efficient strategy for breeding. MAS lymphocystis disease-resistant flounder had a market penetration rate of 35 percent in Japan in 2012 (Ozaki *et al., 2012).*

Another good example of MAS is the selection of IPN resistance. In salmon, IPN is a major problem. One major QTL was mapped to linkage group 21, which accounts for 29 percent and 83 percent of the phenotypic and genetic variances, respectively. Three microsatellite markers were tightly linked to the QTL, and these markers have been used for the selection of IPN resistance (Moen *et al., 2009).*

It should be noted that MAS refers to marker-assisted selection, not marker selection. This means that markers are used to supplement the routine selective breeding programmes rather than replacing them. Although MAS is theoretically very sound and attractive, little is known about the economic benefits gained from MAS in aquaculture species, apart from the above cases where the phenotypes were controlled by a single gene rather than by many genes. Information of this nature is important because the additional genetic gains depend on the magnitude of the

allelic effects, and thus the marginal increase should offset the costs of applying the technology (e.g. genotyping and labour costs).

9.2 Sex markers and their applications

In aquaculture settings, it is often important to know the sex information of fish and shellfish species. In addition to biological interest for sex determination and regulation, sex is also of interest to aquaculturists in relation to sexual dimorphism. In most aquaculture species, one sex grows significantly faster than the other . For instance, females of soles, eels and many other species grow much faster than the males. In contrast, males grow much faster in tilapia and catfish, among many other species. In addition to growth rate, sex also affects body shape, colouration and carcass composition (Beardmore, 2001; Cnaani, 2009).

Sex-linked markers have been mapped for many aquaculture species. For example, AFLP and microsatellite markers have been identified to be linked with sex in various aquaculture species, including common carp (Chen *et al.,* 2009), tilapia (Lee, Penman and Kocher, 2003), catfish (Ninwichian *et al.,* 2012a, 2012b), Zhikong scallops (Li *et al.,* 2005), half-smooth tongue sole (Chen *et al.,* 2007), white shrimps (Pérez *et al*.,

2004), kuruma prawns (Li *et al.,* 2003), and rainbow trout (Felip *et al.,* 2005). These sex-linked markers have been useful for the identification of sex in the absence of phenotypic information.

9.3 Genome selection

Recent advances of genome analysis, including the availability of a large number of polymorphic markers, highly efficient genotyping platforms such as SNP arrays and the application of next-generation sequencing technologies, have allowed the mapping of dense markers across the entire genome, which in turn enables an estimation of the genetic merit of every chromosome fragment contributing variation in a population with phenotypic observations. Not only can the merit of every chromosomal segment be estimated, but also all the traits of interest can be estimated simultaneously. Whole-genome selection is based on such abilities of estimating the value of every chromosome fragment contributing variation in a population with phenotypic observations (training); the results of training can be used to predict the merit of new animals that have genotypes but are not included in the training data set.

Genomic selection was first proposed by Meuwissen *et al.* (2001). Since then, it has gained a tremendous level of attention in the animal genetics community. Genomic selection is a form of marker-assisted selection in which genetic markers covering the whole genome are used so that all quantitative trait loci are in linkage disequilibrium with at least one marker. Compared with MAS, genomic selection uses the estimated effect of many loci across the entire genome at once, not just the small number of linked loci as done with MAS, or individual locus as in GWAS.

Although genomic selection has been successfully used in dairy cow and beef cattle and other livestock species (Hayes *et al.,* 2009), its use in aquaculture species has been slow (Ragavendran and Muir, 2011; Sonesson, 2011).

10. GENOME EDITING TECHNOLOGIES

10.1 Zinc finger nuclease and history of genome editing technologies

Genome editing refers to the ability to make specific changes at targeted genomic sites. The history for the development of genome editing technologies has been well summarized by Nemudryi *et al.* (2014). The zinc finger nuclease (ZFN) technology was developed in 1996. A zinc finger protein domain coupled with the FokI endonuclease domain was demonstrated to act as a site-specific nuclease cutting DNA at strictly defined sites *in vitro (Kim,* Cha and Chandrasegaran, 1996). This chimeric protein has a modular structure, with each zinc finger domain recognizing one nucleotide triplet. This method was used for genome editing of cultured cells of both plants and animals, including pluripotent stem cells (e.g. Bibikova *et al.,* 2002; Townsend *et al.,* 2009; Provasi *et al.,* 2012). However, the ZFN-based technology has several disadvantages. It is complex to use, has a high cost, and the cleavage site is not accurate.

Active searches for better methods have been a major focus of research in the last several years. That has led to the development of new genome editing technologies, such as the transcription activator-like effector nucleases (TALENs) and the clustered regularly interspaced short palindromic repeats (CRISPR) -associated (CRISPR/Cas). These new genome editing technologies overcome the disadvantages of ZFN technology and have become very efficient for the modification of genomes through genome editing.

10.2 TALEN

Transcription activator-like effector nucleases (TALEN) are restriction enzymes that can be engineered to cut specific sequences of DNA. They are made by fusing a transcription activator-like effector DNA-binding domain to a DNA cleavage domain. Transcription activator-like effectors can be engineered to bind practically any desired DNA sequence so, when combined with the cleavage domain (a nuclease), the DNA can be cut at specific locations (Boch et al., 2009). The restriction enzymes, when being introduced into cells, can be used for target gene or genome editing. Along with zinc finger nucleases and CRISPR-associated protein-9 nuclease (Cas9) proteins, TALEN is becoming a prominent tool in the field of genome editing

Genome editing starts with efficient generation of a double-strand break (DSB) in the target DNA. DSBs are repaired either by homologous recombination, or, in the absence of a homologous repair template, via non-homologous end joining (NHEJ) which causes small insertions or deletions (INDELS) as the broken ends are ligated together. The creation of INDELS is exploited as a convenient method for a gene knockout. Both TALEN and CRISPR can edit DNA through either NHEJ or homologous recombination. TALEN is comprised of a pair of DNA binding proteins fused to the FokI nuclease, while CRISPR is a complex between the Cas9 nuclease and a target-specific single guide RNA (sgRNA).

TALEN and CRISPR/Cas9 endonucleases have provided ways to significantly improve genome editing efficiency. These endonucleases make a DSB at a predetermined DNA sequence and

trigger natural DNA repair processes such as NHEJ or homologous recombination with a donor DNA template. Among these existing approaches, RNA-guided CRISPR/Cas9 is the most user-friendly and versatile system.

10.3 CRISPR/Cas9

CRISPR consist of bacterial DNA containing short palindrome repeats that are regularly spaced. CRISPR-associated (Cas) proteins are nucleases that cut DNA. CRISPR is a mechanism of bacterial immunity against invading viruses or plasmids. Its mechanism of actions is similar to eukaryotic RNA interference pathways.

This system was first used as a genome editing system in 2012. Since then, CRISPR/Cas9 has been used for modification of genomes in many species, including zebrafish (Hwang *et al.,* 2013; Jao, Wente and Chen, 2013). The CRISPR works by three components: CRISPR RNA (crRNA), which binds the target DNA and guides cleavage, and the trans-activating RNA (tracrRNA), which base pairs with the crRNA and enables the Cas9-crRNA complex to locate the targeted DNA and the Cas9 nuclease. In a typical situation, the Cas9 nuclease is transiently expressed using a promoter suitable to the cells for which genome editing is for, while the RNA components are co-transfected or injected with the expression construct of Cas9. The Cas9 is an enzyme that cuts DNA, and CRISPR is a collection of DNA sequences that tells Cas9 where to cut.

10.4 Comparison of TALEN and CRISPR/Cas9

TALEN and CRISPR differ in four aspects: specificity; target selection; efficiency; and simplicity of construction. CRISPR achieves its specificity through the single guide RNA (sgRNA), which is an artificial fusion of two naturally occurring short RNAs (Jinek *et al.,* 2012). The sgRNA directs the Cas9 nuclease to a 20-nucleotide target site on the chromosome which must be immediately followed by an N-G-G trinucleotide known as the protospacer adjacent motif, or PAM. The sgRNA hybridizes with the strand opposite to the PAM site, and Cas9 nuclease cuts the DNA. In this process, sgRNAs can tolerate up to five mismatches to guide mutagenesis to off-target sites (Fu *et al.,* 2013).

Compared with CRISPR, TALEN has higher levels of specificity. A TALEN pair (each ~18 base pair long, total 36 base pairs) must bind on opposite sides of the target site, separated by a spacer ranging from 14 to 20 nucleotides. The target to match the 18 base pairs is expected to be unique, thus providing high specificity. In terms of target selection, both TALEN and CRISPR are quite flexible, but TALEN is more flexible. TALEN and CRISPR have a comparable efficiency, but a slightly higher efficiency (up to 70 percent) was reported for CRISPR. TALEN may be sensitive to cytosine methylation within CpG dinucleotides. Among all the features, CRISPR is simple to design and use compared with TALEN or ZFN technologies. Therefore, CRISPR is gaining its popularity in applications to various species systems. For each target site, all that is needed for CRISPR is to programme a 20-nucleotide genomic target site into the overall sgRNA. Plasmid construction is straightforward and simple. For editing experiments, the sgRNA is co-expressed with the reusable Cas9 nuclease. In contrast, TALEN construction involves re-engineering a new protein for each target.

11. CHALLENGES FOR THE APPLICATIONS OF GENOME-BASED TECHNOLOGIES IN AQUACULTURE

As discussed above, genome-based technologies include DNA marker technologies, genome mapping technologies, sequencing technologies and genome editing technologies. To a certain extent, these technologies have been used in aquaculture species, but the potential for their applications is tremendous.

The most practically important genome-based technologies are probably genome editing technologies involving ZFN, TALEN or CRISPR/Cas9. Using such technologies allows precise knockout of genes, gene modification, or targeted gene insertion (Lauth et al., 2012). Such technologies have been used in cattle, pigs, rabbits and catfish, as well as in zebrafish (Doyon, 2008; Meng, 2008; Hauschild, 2011; Yu, 2011; Flisikowska, 2011; Dong, 2011).

With the rapid advances of genomic research and the development and applications of genome-based technologies, several are the challenges that need to be faced. Some of these challenges are presented in the following sections.

11.1 Decoupling of genomics with breeding programmes

In the past two decades, huge progress has been made in the areas of genomic research and development of genome-based technologies. Applications of such technologies have made large strides with plants and livestock animal species; however, applications of genomic technologies in aquaculture species have been limited. This was partly due to the uncoupling of genome research with breeding programmes. With the exception of Atlantic salmon and perhaps rainbow trout, where private corporations run major breeding programmes in which genomic information and genome-based technologies are more commonly used, the use of genomic information and genome-based technologies is very limited in other aquaculture species. A detailed examination of the situations indicated that lack of major breeding companies, public or private, may be a major bottleneck for breeding programmes of many aquaculture species, including various species such as common carp, grass carp, crucian carp, black carp, tilapia, catfish and various shellfish species. In shrimps, private or public institutions carrying out breeding programmes do exist, but their efforts have been focused on the production of specific pathogen-free stocks rather than breeding for disease resistance using genomic information or genome-based technologies. Therefore, more resources should be allocated from the government and other stakeholders to support the breeding programmes of aquaculture species. This is not only an issue of aquaculture production, but also in terms of environmental sustainability because a more efficient use of aquatic genetic resources in aquaculture would help to reduce the pressure on wild stocks. Great progress will be made only when the application of genomic research and technologies will become more routinely integrated in breeding programmes.

11.2 Bioinformatics challenges

Genomics relies on the ability to analyse large data sets through bioinformatic data mining, data analysis, data sharing, meta-analysis and data re-analysis. Of all the expertise, bioinformatics

is key to the success of genome research. The development and application of genome-based technologies depend on the ability to analyse the impact of such applications on economics. Students of biology usually have limited background in informatics, and informatics students have limited background in biology. Training scientists with combined expertise is truly a great challenge. This challenge is paramount for aquaculture because aquaculture workers tend to have even less understanding of informatics.

This raises a question as to what kind of scientists should be trained today as the next generation of scientists. Decades ago, graduate students spent most of their graduate school studies on hands-on experiments. Data analysis was a minor part of the graduate life. However, the data analysis component has recenty become more time demanding. For genomics students, they may have to spend most of their time doing data analysis using supercomputers. Yet, these students should also be trained in order to have a good understanding of biological aspects of the organisms they are working on. Keeping a balance between these things is a significant challenge.

11.3 Computational limitations

With next-generation sequencers, terabytes of data can be readily generated in any of the simple experimentations. The key is the ability to analyse such large data sets. In most cases, this must be handled using supercomputers, or high-performance computer clusters (HPC). This raises two lines of challenges: the first is that HPC computers mostly use Unix or Linux platforms that are less user-friendly and demand the users to have basic knowledge of command lines; the second is that the purchasing of HPC computers is very costly, most often costing over US$1 million. Even more difficult is the maintenance and update of the HPCs as they become sort of obsolete after a few years of service. This challenge is difficult to manage and it is apparent that such a challenge would be almost impossible for some developing countries. Though the opportunities to use cloud computing are available, it requires a certain level of infrastructure and information technology resources to be able to undertake bioinformatic analyses.

11.4 Funding challenges and unbalanced research advances

Although funding limitation is a universal problem, it is even a greater problem for aquaculture because many aquaculture species represent minor commodities compared with major plant or animal species. In fact, funding availability is most often dictated by the commodity's importance.

The big investments usually required for genomic research and development lead also to a huge imbalance in the geographical distribution of genomic studies in the world, especially with regard to aquaculture species. The major research projects of aquaculture genomics are in a dozen countries, such as Brazil, Canada, Chile, China, Greece, Japan, India, Norway, Singapore, Spain and the United States of America. Genome projects, however, are rare in Africa. In a way, the geographical distribution of fish genome projects is correlated with geographical distribution of the application of genome-based technologies in aquaculture species.

11.5 Challenges of working with small-scale farmers

It is relatively easy to deal with species for which the industries are vertically integrated. As such, decisions can be made by the top management dealing with various problems and issues in a

systems approach, including genetics, nutrition, culture technologies, harvest, processing and marketing. With aquaculture species, the situation vary greatly from species to species. The best scenario is perhaps that of Atlantic salmon. In northern Europe, such as in Norway, breeding companies hold broodstocks that are continuously genetically improved with traditional selective breeding or with genomic selection. For many other species, however, breeding programmes may exist but the systematic selection using genome-based technologies have not been practised. This issue is more serious in small-sale aquaculture. In some cases, aquaculture is practised in a more traditional way, with the broodstock or the seed often captured from the wild. No selection using the principles of genetics is involved in these contexts, making the adoption of genome-based technologies a remote scenario.

11.6 The push for a fast return on investment

Research, especially basic research, takes time and there is always a gap between research and application. This is particularly true with genome sciences, where it can take decades to generate enough information for possible applications. This is well-known by the scientific community, but not always understood by the public and governments. Governments are often anxious to produce quick results. While that is understandable with public pressure for return of investment, it often backfires if application is sought prematurely. For instance, applying the marker-assisted selection before the achievement of a comprehensive understanding of major QTLs can cause a failure of the breeding programme.

11.7 Ethical, legislative and regulatory issues

The new line of genomic technologies, especially genome editing technologies, brings risks and ethical challenges, as outlined in FAO's Statement on Biotechnology (2000). While there is little controversy about many aspects of biotechnology and its application, genetically modified organisms (GMOs) have become the target of a very intensive and, at times, emotionally charged debate. It is generally well recognized that genetic engineering has the potential to help increase production and productivity. However, there is concern about the potential risks posed by certain aspects of biotechnology. These risks fall into two basic categories: the effects on human and animal health and the environmental consequences. Caution must be exercised in order to reduce the risks of transferring toxins from one life form to another, creating new toxins or transferring allergenic compounds from one species to another, which could result in unexpected allergic reactions. Risks to the environment include the possibility of outcrossing which could lead, for example, to the development of wild relatives with increased resistance to diseases or environmental stresses, upsetting the ecosystem balance.

There is a strong case to support a science-based evaluation system that would objectively determine the benefits and risks of each individual GMO. This calls for a cautious case-by-case approach to address legitimate concerns for the biosafety of each product or process prior to its release. The possible effects on biodiversity, the environment and food safety need to be evaluated, and the extent to which the benefits of the product or process outweigh its risks assessed. The evaluation process should also take into consideration the experience gained by national regulatory authorities in clearing such products. Careful monitoring of the post-release effects of these products and processes is also essential to ensure their continued safety to human

beings, animals and the environment. For instance, the TALEN or CRISPR technologies allow the introduction or knockout of any gene, without much difficulty, in any fish or shellfish species. The altered genome is able to pass on the genetic material to future generations. While it is clear that the genome editing technologies are different from gene transfer technologies, it is widely believed that government agencies should be regulating any commercial products generated using gene-editing technologies. At this time, it is not clear how stringent such regulation should be because the scientific community is still at the early stages of discussion.

An important aspect concerning the regulation of biotechnology is the issue of Intellectual Property Rights (IPR), defined as the right to control the commercial exploitation of the projected subject matter for a specific period (FAO 2016). Different forms of IPR exist, such as copyrights or patents, each having different requirements. It was noted that minimum standards for protecting IPR are set by the World Trade Organization Agreement on Trade-Related Aspects of Intellectual Property Rights for signatory countries. The use of IPR in agricultural biotechnology has been controversial, especially in developing countries, and have been criticized for a variety of reasons including: conflicts with farmers' traditional practices to reuse seed; excessively broad patent claims; patentability of genetic material and plant varieties; uncertainty regarding the scope of research use; high transaction costs; pricing of improved varieties (high because it has to cover costs of licensing of IPR); and appropriation of traditional knowledge and sovereign genetic resources. In many instances, it is not the IPR per se that impede the diffusion of agricultural biotechnologies in developing countries, but other confounding issues involved. IPR are legal tools to arrange a licensing agreement; there are other legal tools available to handle conflicts raised within IPR (e.g. consumer protection legislation), which should be activated whenever needed. National biotechnology policies and legislation should encompass aspects related not only to the safety of biotechnology products but also to their ownership.

Levels of regulation on agricultural biotechnology, specifically genetically engineered organisms, are diverse between countries. In most developed countries genetically engineered organisms are well regulated by various government agencies. For example, in the United States of America both the FDA and the USDA regulate GMOs. Within the USDA's Animal and Plant Health Inspection Service, there is the Biotechnology Regulatory Services (https://www.aphis.usda.gov/aphis/ourfocus/biotechnology). The FDA is responsible for ensuring the safety and proper labelling of all plant- and animal-derived food and feed, including those developed through genetic engineering. A typical issue that GMOs have to go through is well illustrated by the decade-long fight over the marketing of transgenic salmon by AquaBounty Technologies.

While similar regulatory agencies exist in many countries, regulation may vary across the world. International regulation of biotechnology is well described by Matthias Herdegen (http://legal.un.org/avl/pdf/ls/Herdegen_slideshow.pdf). Concerns for international regulation are focused on risk, free trade and development perspectives in the field of human rights, environment protection and international trade law. For environment protection, the United Nations Convention on Biological Diversity and the Cartagena Protocol on Biosafety are guiding documents for actions. However, with the advanced genome editing technologies, it is probably necessary to have stronger international regulations, as well as those regulations placed by governments in various countries.

The increased convergence between biological and other sciences, higher investment requirements, the higher profile of intellectual property and biosafety issues, the changed role

of the private sector (both in the development of the technologies and the technology delivery systems) are all aspects that should be clearly present in an effective policy development process. It is important to consider also that the development of new biotechnologies is progressing extremely quickly, and the state-of-the-art is changing faster and faster. As such, flexible and forward-looking regulatory and legislative frameworks will increasingly be needed.

11.8 Public perception

There is a clear need for a common understanding on vocabulary and definitions of the terminology used for agriculture biotechnologies (e.g. what is meant by genetic modification, genetic engineering or a GMO?) for an informed discussion. This common understanding is crucial for public perception (FAO 2016).

Public perception and acceptance have always been an issue to date with genetically modified organisms (GMOs). However, in most cases, GMOs have been created by genetic engineering. Aquaculture products produced using genome editing technologies can encounter issues of public perception. Although the exact intensity and the nature of the concerns are unknown at present, it is expected that some levels of negative public perception may exist for aquaculture products created using genome editing technologies. As such, the level of public acceptance for products created by genome editing technologies may be lower than those produced using traditional breeding techniques.

11.9 Technology transfer

Current investment in biotechnological research tends to be concentrated in the private sector and oriented towards agriculture in higher-income countries where there is purchasing power for its products. In view of the potential contribution of biotechnologies for increasing food supply and overcoming food insecurity and vulnerability, efforts should be made to ensure that developing countries, in general, and resource-poor farmers, in particular, benefit more from biotechnological research, while continuing to have access to a diversity of sources of genetic material.

Although technologies are different, it appears that the technology transfer part of the equation has not been changed by the development of genome-based technologies. It will all depend on the patents and other protective measures placed on a particular product or technology.

Concluding remarks

The historical Human Genome Project was officially launched in 1990, and the project lasted more than a dozen years when the human genome was sequenced and assembled in 2004. This multi-billion dollar project, being started with a simple objective of just the sequence of the entire DNA composition or the genome of a single person, ended with massive achievements in biology, ranging from technology development, to methodology breakthrough, to the emergence of the systems biology that takes an entirely different approach changing the traditional way of doing science by proposing a working hypothesis, to experimentally demonstrate the processes, to drawing conclusions proving or denying the original hypothesis. Genomic sciences have made drastic advances in the past ten years, largely because of the application of the next-generation sequencing technologies. It is not just the high throughput that has revolutionized the way science is conducted, but also the rapidly reducing cost for sequencing has made the technologies applicable to all aspects of molecular biological research as well as to all organisms, including aquaculture and fisheries species. The potential for research using genomic approaches has become impressive. Today, commercial services are available worldwide, the key issue is to achieve a balanced output of quality and quantity in a cost-effective way.

Rapid technological advances provide huge opportunities to apply modern genomics to enhance aquaculture production. It is now not too difficult to map and sequence a genome of an aquaculture species. However, understanding the genes controlling economically important traits takes tremendous levels of additional research. Once the important genes for performance and production traits are identified, they can be selected through marker-assisted selection or genomic selection. These genes can be modified to destroy a "bad" gene, edited to have the beneficial allele, or inserted in the genome at exact locations. Such powerful technologies open a huge potential for genetic enhancement of aquaculture species.

However, we are facing a number of new challenges, especially in the area of bioinformatics. This challenge may be paramount for aquaculture researchers and educators. Aquaculture students may be well educated with aquaculture but may have limited background in computer science, or not advanced enough for the bioinformatics analysis of large data sets. The large data sets in tera-scales themselves pose great computational challenges. This requires new ways of thinking in terms of education and training of the next generation of scientists. For instance, few laboratories in the world may be sufficient for data production but many laboratories may be required for data analysis or bioinformatics data mining. In addition, working with small family-based farmers will pose tremendous levels of challenges when applying expensive and novel technologies. Consumer acceptance and public perception is also an issue to consider, and government and international regulations may play a significant role in increasing public acceptance as well as in expanding the application of genome-based technologies beyond research.

References

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B. & Moreno, R.F. 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, 252(5013): 1651–1656.

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(1–2): 43–56.

Ahanger, S., Sandaka, S., Ananad, D., Mani, M.K., Kondadhasula, R., Reddy, C.S., Marappan, M., Valappil, R.K., Majumdar, K.C. & Mishra, R.K. 2014. Protection of shrimp *Penaeus monodon* from WSSV infection using antisense constructs. *Marine biotechnology*, 16(1): 63–73.

Al-Tobasei, R., Paneru, B. & Salem, M. 2016. Genome-wide discovery of long non-coding RNAs in rainbow trout. *PLOS ONE*, 11(2): e0148940.

Anderson, E.D., Mourich, D.V. & Leong, J.A.C. 1996. Gene expression in rainbow trout (*Oncorhynchus mykiss*) following intramuscular injection of DNA. *Molecular Marine Biology and Biotechnology*, 5(2): 105–113.

Andreassen, R., Lunner, S. & Høyheim, B. 2010. Targeted SNP discovery in Atlantic salmon (*Salmo salar*) genes using a 3'UTR-primed SNP detection approach. *BMC Genomics*, 11(1): 706.

Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A. & Gelpke, M.D.S. 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science*, 297(5585): 1301–1310.

Arai, K. 2001. Genetic improvement of aquaculture finfish species by chromosome manipulation techniques in Japan. *Aquaculture*, 197(1): 205–228.

Avise, J.C. 1995. Mitochondrial DNA polymorphism and a connection between genetics and demography of relevance to conservation. *Conservation Biology*, 9(3): 686–690.

Backes, C., Keller, A., Kuentzer, J., Kneissl, B., Comtesse, N., Elnakady, Y.A., Müller, R., Meese, E. & Lenhof, H.P. 2007. GeneTrail– advanced gene set enrichment analysis. *Nucleic Acids Research*, 35(suppl_2): W186–W192.

Baerwald, M., Petersen, J., Hedrick, R., Schisler, G. & May, B. 2011. A major effect quantitative trait locus for whirling disease resistance identified in rainbow trout (*Oncorhynchus mykiss*). *Heredity*, 106(6): 920–926.

Bagley, M.J., Anderson, S.L. & May, B. 2001. Choice of methodology for assessing genetic impacts of environmental stressors: polymorphism and reproducibility of RAPD and AFLP fingerprints. *Ecotoxicology*, 10(4): 239–244.

Baranski, M., Gopikrishna, G., Robinson, N.A., Katneni, V.K., Shekhar, M.S., Shanmugakarthik, J., Jothivel, S. *et al.* 2014. The development of a high density linkage map for black tiger shrimp (*Penaeus monodon*) based on cSNPs. *PLOS ONE*, 9(1): e85413.

Baranski, M., Moen, T. & Våge, D.I. 2010. Research mapping of quantitative trait loci for flesh colour and growth traits in Atlantic salmon (*Salmo salar*). *Genetics Selection Evolution*, 42(1): 17.

Baranski, M., Rourke, M., Loughnan, S., Hayes, B., Austin, C. & Robinson, N. 2008. Detection of QTL for growth rate in the blacklip abalone (*Haliotis rubra* Leach) using selective DNA pooling. *Animal Genetics*, 39(6): 606–614.

Bardakci, F. & Skibinski, D.O. 1994. Application of the RAPD technique in tilapia fish: species and subspecies identification. *Heredity*, 73(2): 117–23.

Barozai, M.Y.K. 2012. The MicroRNAs and their targets in the channel catfish (*Ictalurus punctatus*). *Molecular Biology Reports*, 39(9): 8867–8872.

Bartel, D.P. 2009. MicroRNAs: target recognition and regulatory functions. *Cell*, 136(2): 215–233.

Bauer, S., Grossmann, S., Vingron, M. & Robinson, P.N. 2008. Ontologizer 2.0 – a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinformatics*, 24(14): 1650–1651.

Beardmore, J., Mair, G.C. & Lewis, R.I. 2001. Monosex male production in finfish as exemplified by tilapia: applications, problems, and prospects. *In*: Lee, C.S., & Donaldson, E.M., eds. *Reproductive biotechnology in finfish aquaculture*, pp. 283– 301. Amsterdam, The Netherlands, Elselvier.

Bell, G.D., Kane, N.C., Rieseberg, L.H. & Adams, K.L. 2013. RNA-seq analysis of allele-specific expression, hybrid effects, and regulatory divergence in hybrids compared with their parents from natural populations. *Genome biology and evolution*, 5(7): 1309–1323.

Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., Bento, P., *et al*. 2014. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature communications*, 5(1): 1–10.

Bibikova, M., Golic, M., Golic, K.G. & Carroll, D. 2002. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics*, 161(3): 1169–1175.

Billington, N. 2003. *Mitochondrial DNA*. *Population genetics: principles and applications for fisheries scientists*. pp. 59–100. American Fisheries Society. Bethesda, Maryland, USA.

Birky, C., Fuerst, P. & Maruyama, T. 1989. Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics*, 121(3): 613–627.

Bjornsson, H.T., Albert, T.J., Ladd-Acosta, C.M., Green, R.D., Rongione, M.A., Middle, C.M., Irizarry, R.A., Broman, K.W. & Feinberg, A.P. 2008. SNP-specific array-based allele-specific expression analysis. *Genome research*, 18(5): 771–779.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. & Bonas, U. 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, 326(5959): 1509–1512.

Bonaldo, M.D.F., Lennon, G. & Soares, M.B. 1996. Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Research*, 6(9): 791–806.

Boney, S.E., Shelton, W.L., Yang, S.L. & Wilken, L.O. 1984. Sex reversal and breeding of grass carp. *Transactions of the American Fisheries Society*, 113(3): 348–353.

Booman, M., Borza, T., Feng, C.Y., Hori, T.S., Higgins, B., Culf, A., Léger, D. *et al*. 2011. Development and experimental validation of a 20K Atlantic cod (*Gadus morhua*) oligonucleotide microarray based on a collection of over 150,000 ESTs. *Marine Biotechnology*, 13(4): 733–750.

Botstein, D., White, R.L., Skolnick, M. & Davis, R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, 32(3): 314.

Boulton, K., Massault, C., Houston, R.D., de Koning, D.J., Haley, C.S., Bovenhus, 2011. QTL affecting morphometric traits and stress response in the gilthead seabream (*Sparus aurata*). *Aquaculture*, 319(1): 58–66.

Boulding, E., Culling, M., Glebe, B., Berg, P., Lien, S. & Moen, T. 2008. Conservation genomics of Atlantic salmon: SNPs associated with QTLs for adaptive traits in parr from four trans-Atlantic backcrosses. *Heredity*, 101(4): 381–391.

Braslavsky, I., Hebert, B., Kartalov, E. & Quake, S.R. 2003. Sequence information can be obtained from single DNA molecules. *Proceedings of the National Academy of Sciences*, 100(7): 3960–3964.

Brawand, D., Wagner, C.E., Li, Y.I., Malinsky, M., Keller, I., Fan, S., Simakov, O., *et al.* 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, 513(7518): 375–381.

Brem, G., Brenig, B., Hörstgen-Schwark, G. & Winnacker, E.-L. 1988. Gene transfer in tilapia (*Oreochromis niloticus*). *Aquaculture*, 68(3): 209–219.

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. *Molecular Ecology*, 9(12), 2175–2178.

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. *Molecular Biology and Evolution*, 21(5): 945–956.

Cardoso, S.R.S., Eloy, N.B., Provan, J., Cardoso, M.A. & Ferreira, P.C.G. 2000. Genetic differentiation of *Euterpe edulis* Mart. Populations estimated by AFLP analysis. *Moleuclar Ecology*, 9(11): 1753–1760

Carrel, L. & Willard, H.F. 2005. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature*, 434(7031): 400–404.

Carter, S.L., Brechbühler, C.M., Griffin, M. & Bond, A.T. 2004. Gene co-expression network topology provides a framework for molecular characterization of cellular state. *Bioinformatics*, 20(14): 2242–2250.

Cassani, J.R. & Caton, W.E. 1986. Growth comparisons of diploid and triploid grass carp under varying conditions. *The Progressive Fish-Culturist*, 48(3): 184–187.

Castaño-Sánchez, C., Fuji, K., Ozaki, A., Hasegawa, O., Sakamoto, T., Morishima, K., Nakayama, I., *et al.* 2010. A second generation genetic linkage map of Japanese flounder (*Paralichthys olivaceus*). *BMC Genomics*, 11(1): 554.

Chapman, R.W. 1989. Mitochondrial and nuclear gene dynamics of introduced populations of *Lepomis macrochirus*. *Genetics*, 123(2): 399–404.

Chen, J., Wang Y., Yue, Y., Xia, X., Du, Q., & Chang, Z. 2009. A novel male-specific DNA sequence in the common carp, *Cyprinus carpio*. *Molecular and Cellular Probes*, 23(5): 235–239.

Chen, S.L., Li, J., Deng, S.P., Tian, Y.S., Wang, Q.Y., Zhuang, Z.M., Sha, Z.X. & Xu, J.Y. 2007. Isolation of female-specific AFLP markers and molecular identification of genetic sex in half-smooth tongue sole (*Cynoglossus semilaevis*). *Marine Biotechnology*, 9(2): 273–280.

Chen, A., Wang, R., Liu, S., Peatman, E., Sun, L., Bao, L., Jiang, C., *et al.* 2016. Ribosomal protein genes are highly enriched among genes with allele-specific expression in the interspecific F1 hybrid catfish. *Molecular Genetics and Genomics*, 291(3): 1083–1093.

Chen, S., Zhang, G., Shao, C., Huang, Q., Liu, G., Zhang, P., Song, W., *et al* 2014. Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nature Genetics*, 46(3): 253–260.

Chini, V., Cattaneo, A.G., Rossi, F., Bernardini, G., Terova, G., Saroglia, M. & Gornati, R. 2008. Genes expressed in blue fin tuna (*Thunnus thynnus*) liver and gonads. *Gene*, 410(1): 207–213.

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(4): 1821–1826.

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. *Biochemical Genetics*, 38(3–4): 63–76.

Chourrout, D. 1980. Thermal induction of diploid gynogenesis and triploidy in the eggs of the rainbow trout (*Salmo gairdneri* Richardson). *Reproduction Nutrition Development*, 20(3A): 727–733.

Chourrout, D. 1984. Pressure-induced retention of second polar body and suppression of first cleavage in rainbow trout: production of all-triploids, all-tetraploids, and heterozygous and homozygous diploid gynogenetics. *Aquaculture*, 36(1–2): 111–126.

Chourrout, D., Chevassus, B., Krieg, F., Happe, A., Burger, G. & Renard, P. 1986. Production of second generation triploid and tetraploid rainbow trout by mating tetraploid males and diploid females – potential of tetraploid fish. *Theoretical and Applied Genetics*, 72(2): 193–206.

Cnaani, A. 2009. Sexual development in fish, practical applications for aquaculture. *Sexual Development*, 3(2–3): 164–175.

Cnaani, A., Lee, B.-Y., Ozouf-Costaz, C., Bonillo, C., Baroiller, J., D'Cotta, H. & Kocher, T. Mapping of sox2 and sox14 in tilapia (*Oreochromis* spp.). *Sexual Development*, 1(3): 207–210.

Colihueque, N., Cárdenas, R., Ramírez, L., Estay, F. & Araneda, C. 2010. Analysis of the association between spawning time QTL markers and the biannual spawning behavior in rainbow trout (*Oncorhynchus mykiss*). *Genetics and Molecular Biology*, 33(3): 578–582.

Combes, M.C., Hueber, Y., Dereeper, A., Rialle, S., Herrera, J.C. & Lashermes, P. 2015. Regulatory divergence between parental alleles determines gene expression patterns in hybrids. *Genome Biology and Evolution*, 7(4): 1110–1121.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M. & Robles, M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21(18): 3674–3676.

Cowles, C.R., Hirschhorn, J.N., Altshuler, D. & Lander, E.S. 2002. Detection of regulatory variation in mouse genes. *Nature Genetics*, 32(3): 432–437.

Cox, D.R., Burmeister, M., Price, E.R., Kim, S. & Myers, R.M. 1990. Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science*, 250(4978): 245–250.

Crawford, A.M. & Cuthbertson, R.P. 1996. Mutations in sheep microsatellites. *Genome research*, 6(9): 876–879.

Crespi, B.J. & Fulton, M.J. 2004. Molecular systematics of Salmonidae: combined nuclear data yields a robust phylogeny. *Molecular Phylogenetics and Evolution*, 31(2): 658–679.

Daelemans, C., Ritchie, M.E., Smits, G., Abu-Amero, S., Sudbery, I.M., Forrest, M.S., Campino, S., *et al.* 2010. High-throughput analysis of candidate imprinted genes and allele-specific gene expression in the human term placenta. *BMC Genetics*, 11(1): 1.

Dahm, R. & Geisler, R. 2006. Learning from small fry: the zebrafish as a genetic model organism for aquaculture fish species. *Marine Biotechnology*, 8(4): 329–345.

Danzmann, R.G., Jackson, T.R. & Ferguson, M.M. 1999. Epistasis in allelic expression at upper temperature tolerance QTL in rainbow trout. *Aquaculture*, 173(1–4): 45–58.

Dawnay, N.R., Ogden, R., McEwing, R., Carvalho, G.R. & Thorpe, R.S. 2007. Validation of the barcoding gene COI for use in forensic genetic species identification. *Forensic Science*, 173(1): 1–6.

Ding, Y., Johnson, M.D., Chen, W.Q., Wong, D., Chen, Y.J. *et al.* 2001) Five-color-based high- information-content fingerprinting of bacterial artificial chromosome clones using Type IIS restriction endonucleases. *Genomics*, 74(2): 142–154.

Donaldson, E.M. & Hunter, G.A. 1982. Sex control in fish with particular reference to salmonids. *Canadian Journal of Fisheries and Aquatic Sciences*, 39(1): 99–110.

Dong, Z. 2011. Heritable targeted inactivation of myostatin gene in yellow catfish (*Pelteobagrus fulvidraco*) using engineered zinc finger nucleases. *PLOS ONE*, 6(12): e28897.

Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., Tosser-Klopp, G., *et al.* 2013. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nature Biotechnology*, 31(2): 135–141.

Doyon, Y., McCammon, J.M., Miller, J.C., Faraji, F., Ngo, C., Katibah, G.E., Amora, R. *et al.* 2008. Heritable targeted gene disruption in zebrafish using designed zinc finger nucleases. *Nature biotechnology*, 26(6): 702–708.

Dunham, R.A. 1990. Production and use of monosex or sterile fishes in aquaculture. *Reviews in Aquatic Sciences*, 2(1): 1–17.

Dunham, R. 2004. *Aquaculture and fisheries biotechnology*. CABI Publishing, Alabama, USA.

Dunham, R.A. 2011. *Aquaculture and fisheries biotechnology: genetic approaches*. CABI Publishing, Alabama, USA.

Dunham, R.A. Eash, J., Askins, J. & Townes, T.M. 1987. Transfer of the metallothionein-human growth hormone fusion gene into channel catfish. *Transactions of the American Fisheries Society*, 116(1): 87–91.

Dunham, R.A., Warr, G.W., Nichols, A., Duncan, P.L., Argue, B., Middleton, D. & Kucuktas, H. 2002. Enhanced bacterial disease resistance of transgenic channel catfish *Ictalurus punctatus* possessing cecropin genes. *Marine Biotechnology*, 4(3): 338–344.

Easton, A., Moghadam, H., Danzmann, R. & Ferguson, M. 2011. The genetic architecture of embryonic developmental rate and genetic covariation with age at maturation in rainbow trout *Oncorhynchus mykiss. Journal of Fish Biology*, 78(2): 602–623.

Edwards, Y.J.K., 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. *Journal of molecular biology*, 278(4): 843–854.

Ellegren, H. 2000. Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in genetics*, 16(12): 551–558.

FAO. 2005–2016. National Aquaculture Sector Overview. China. National Aquaculture Sector Overview Fact Sheets. **Text by Shuping, C.** In: *FAO Fisheries and Aquaculture Department* [online]. Rome. Updated 1 February 2005. http://www.fao.org/fishery/countrysector/naso_china/en

FAO. 2000. FAO Statement on Biotechnology on the occasion of the "Codex Alimentarius Ad Hoc Intergovernmental Task Force on Foods Derived from Biotechnology. [online] . Accessed 25 January 2017. http://www.fao.org/biotech/fao-statement-on-biotechnology/en/

FAO. 2016. Proceedings of the FAO International Symposium on the Role of Agricultural Biotechnologies in Sustainable Food Systems and Nutrition. Rome. 192–206. http://www.fao.org/3/a-i5922e.pdf

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.). *Marine Biotechnology*, 2(3): 301–306.

Felip, A., Young, W.P., Wheeler, P.A. & Thorgaard, G.H. 2005. An AFLP-based approach for the identification of sex-linked markers in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*, 247(1–4): 35–43.

Ferraresso, S., Vitulo, N., Mininni, A.N., Romualdi, C., Cardazzo, B., Negrisolo, E., Reinhardt, R., Canario, A.V., Patarnello, T. & Bargelloni, L. 2008. Development and validation of a gene expression oligo microarray for the gilthead sea bream (*Sparus aurata*). *BMC Genomics*, 9(1): 1.

Figueras, A., Robledo, D., Corvelo, A., Hermida, M., Pereiro, P., Rubiolo, J.A., Gómez-Garrido, J., *et al.* 2016. Whole genome sequencing of turbot (*Scophthalmus maximus*; Pleuronectiformes): a fish adapted to demersal life. *DNA research*, 23(3): 181–192.

FISH-BOL. 2010. Fish barcode of life initiative campaign: species of siluriformes yet to be barcoded. Available at: www.fishbol.org/progress_list_2.php?region=7&level=order_name&type=Siluriformes&acti on=pdf&list=not. Accessed 2 Aug 2010.

FitzSimmons, N.N., Moritz, C. & Moore, S.S. 1995. Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. *Molecular biology and evolution*, 12(3): 432–440.

Fletcher, G.L. & Davies, P.L. 1991. Transgenic fish for aquaculture. *In Genetic engineering*, pp. 331–370. USA, Springer.

Fleury, E., Huvet, A., Lelong, C., De Lorgeril, J., Boulo, V., Gueguen, Y., Bachère, *et al.* 2009. Generation and analysis of a 29 745 unique expressed sequence tags from the Pacific oyster (*Crassostrea gigas*) assembled into a publicly accessible database: the GigasDatabase. *BMC Genomics*, 10(1): 1.

Flisikowska, T. 2011. Efficient immunoglobulin gene disruption and targeted replacement in rabbit using zinc finger nucleases. *PLOS ONE*, 6(6): e21045.

Franco, G.R., Adams, M.D., Soares, M.B., Simpson, A.J., Venter, J.C. & Pena, S.D. 1995. Identification of new *Schistosoma mansoni* genes by the EST strategy using a directional cDNA library. *Gene*, 152(2): 141–147.

Freamo, H., O'Reilly, P., Berg, P.R., Lien, S. & Boulding, E.G. 2011. Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs. *Molecular Ecology Resources*, 11(1): 254–267.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K. & Sander, J.D. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology*, 31(9): 822–826.

Fuji, K., Hasegawa, O., Honda, K., Kumasaka, K., Sakamoto, T. & Okamoto, N. 2007. Marker-assisted breeding of a lymphocystis disease-resistant Japanese flounder (*Paralichthys olivaceus*). *Aquaculture*, 272(1–4): 291–295.

Gagnaire, P. A., Normandeau, E., Pavey, S. A., & Bernatchez, L. 2013. Mapping phenotypic, expression and transmission ratio distortion QTL using RAD markers in the Lake Whitefish (*Coregonus clupeaformis*). *Molecular Ecology*, 22(11): 3036–3048.

Garber, A.F. & Sullivan, C.V. 2006. Selective breeding for the hybrid striped bass *(Morone chrysops*, *Rafinesque* × *M. saxatilis*, Walbaum) industry: status and perspectives. *Aquaculture Research*, 37(4): 319–338.

Gene Ontology Consortium. 2010. The gene ontology in 2010: extensions and refinements. *Nucleic Acids Research*, 38 (Suppl_1): D331–D335.

Gene Ontology Consortium. 2015. Gene Ontology Consortium: Going forward. *Nucleic Acids Research*, 43(D1): D1049–D1056.

GWAS analyses in catfish reveal the functional hubs of related areas within major QWLs for columnaris disease resistance. Plant and Animal Genome XIII Conference. *Plant and Animal Genome,* 10–14 January 2015. San Diego, CA, USA.

Geng, X., Sha, J., Liu, S., Bao, L., Zhang, J., Wang, R., Yao, J., *et al.* 2015. A genome-wide association study in catfish reveals the presence of functional hubs of related genes within QTLs for columnaris disease resistance. *BMC Genomics*, 16(1): 1.

Gharbi, K., Gautier, A., Danzmann, R.G., Gharbi, S., Sakamoto, T., Høyheim, B., Taggart, J.B., *et al.* 2006. A linkage map for brown trout (*Salmo trutta*): chromosome homeologies and comparative genome organization with other salmonid fish. *Genetics*, 172(4): 2405–2419.

Gheyas, A., Haley, C., Guy, D., Hamilton, A., Tinch, A., Mota, J. 2010. Effect of a major QTL affecting IPN resistance on production traits in Atlantic salmon. *Animal Genetics*, 41: 666–668.

Gidskehaug, L., Kent, M., Hayes, B.J. & Lien, S. 2011. Genotype calling and mapping of multisite variants using an Atlantic salmon iSelect SNP array. *Bioinformatics*, 27: 303–310.

Gilbey, J., Verspoor, E., McLay, A. & Houlihan, D. 2004. A microsatellite linkage map for Atlantic salmon (*Salmo salar*). *Animal genetics*, 35(2): 98–105.

Gjedrem, T. & Baranski, M. 2010. *Selective breeding in aquaculture: an introduction*. Vol. 10. Springer Science and Business Media.

Gjedrem, T. & Robinson, N. 2014. Advances by selective breeding for aquatic species: a review. *Agricultural Sciences*, 5(12): 1152–1158.

Gjedrem, T., Robinson, N. & Rye, M. 2012. The importance of selective breeding in aquaculture to meet future demands for animal protein: a review. *Aquaculture*, 350–353: 117–129.

Gomez-Uchida, D., Seeb, J.E., Smith, M.J., Habicht, C., Quinn, T.P. & Seeb, L.W. 2011. Single nucleotide polymorphisms unravel hierarchical divergence and signatures of selection among Alaskan sockeye salmon (*Oncorhynchus nerka*) populations. *BMC Evolutionary Biology*, 11(1): 48.

Gonen, S., Baranski, M., Thorland, I., Norris, A., Grove, H., Arnesen, P., Bakke, H., Lien, S., Bishop, S.C. & Houston, R.D. 2015. Mapping and validation of a major QTL affecting resistance to pancreas disease (salmonid alphavirus) in Atlantic salmon (*Salmo salar*). *Heredity*, 115(5): 405–414.

Gonen, S., Lowe, N.R., Cezard, T., Gharbi, K., Bishop, S.C. & Houston, R.D. 2014. Linkage maps of the Atlantic salmon (*Salmo salar*) genome derived from RAD sequencing. BMC Genomics, 15(1): 1.

Gonzalez-Pena, D. 2016. *Genome-wide association study for identifying genome loci that affect fillet yield in rainbow trout* (Oncorhynchus mykiss). Plant and Animal Genome XXIV Conference. Plant and Animal Genome. 8–13 January 2016.

Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G.P., Haig, D. & Dulac, C. 2010. High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science*, 329(5992): 643–648.

Griffiths, R. & Orr, K. 1999. The use of amplified fragment length polymorphism (AFLP) in the isolation of sex-specific markers. *Molecular ecology*, 8(4): 671–674.

Gross, M., Schneider, J., Moav, N., Moav, B., Alvarez, C., Steve, H., Myster, S. *et al.* 1992. Molecular analysis and growth evaluation of northern pike (*Esox lucius*) microinjected with growth hormone genes *Aquaculture*, 103(3–4): 253–273.

Guo, M., Yang, S., Rupe, M., Hu, B., Bickel, D.R., Arthur, L. & Smith, O. 2008. Genome-wide allele-specific expression analysis using massively parallel signature sequencing (MPSS™) reveals cis-and trans-effects on gene expression in maize hybrid meristem tissue. *Plant Molecular Biology*, 66(5): 551–563.

Guo, X., Li, Q., Wang, Q.Z. & Kong, L.F. 2012. Genetic mapping and QTL analysis of growth-related traits in the Pacific oyster. *Marine Biotechnology*, 14(2): 218–226.

Gutierrez, A.P. & Houston, R.D. 2016. Quantitative trait locus mapping in aquaculture species: principles and practice. *In* Z. Liu, ed. *Bioinformatics in Aquaculture.* Chichester, UK, John Wiley & Sons.

Gutierrez, A.P., Lubieniecki, K.P., Fukui, S., Withler, R.E., Swift, B. & Davidson, W.S. 2014. Detection of quantitative trait loci (QTL) related to grilsing and late sexual maturation in Atlantic salmon (*Salmo salar*). *Marine Biotechnology*, 16(1): 103–110.

Gutierrez, A.P., Lubieniecki, K.P., Davidson, E.A., Lien, S., Kent, M.P., Fukui, S., Withler, R.E., Swift, B. and Davidson, W.S. 2012. Genetic mapping of quantitative trait loci (QTL) for body-weight in Atlantic salmon (*Salmo salar*) using a 6.5 K SNP array. *Aquaculture*, 358–359: 61–70.

Guyomard, R., Boussaha, M., Krieg, F., Hervet, C. & Quillet, E. 2012. A synthetic rainbow trout linkage map provides new insights into the salmonid whole genome duplication and the conservation of synteny among teleosts. *BMC Genetics*, 13(1): 15.

Guyomard, R., Chourrout, D., Leroux, C., Houdebine, L.M. & Pourrain, F. 1989. Integration and germline transmission of foreign genes microinjected into fertilized trout eggs. *Biochimie*, 71(7): 857–863.

Guyon, R., Senger, F., Rakotomanga, M., Sadequi, N., Volckaert, F.A., Hitte, C. & Galibert, F. 2010. A radiation hybrid map of the European sea bass (*Dicentrarchus labrax*) based on 1581 markers: synteny analysis with model fish genomes. *Genomics*, 96(4): 228–238.

Hardman, M. 2005. The phylogenetic relationships among non-diplomystid catfishes as inferred from mitochondrial cytochrome *b* sequences; the search for the ictalurid sister taxon (Otophysi: Siluriformes). *Molecular Phylogenetics and Evolution*, 37: 700–720.

Harris, T.D., Buzby, P.R., Babcock, H., Beer, E., Bowers, J., Braslavsky, I., Causey, M. *et al.* 2008. Single-molecule DNA sequencing of a viral genome. *Science*, 320(5872): 106–109.

Harrison, R.G. 1989. Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends in Ecology & Evolution*, 4(1): 6–11.

Harwood, A.S. & Phillips, R.B. 2011. A suite of twelve single nucleotide polymorphism markers for detecting introgression between cutthroat and rainbow trout. *Molecular ecology resources*, 11(2): 382–385.

Hauschild, J. 2011. Efficient generation of a biallelic knockout in pigs using zinc-finger nucleases. *Proceedings of the National Academy of Sciences*, 108(29): 12013–12017.

Hauser, L., Baird, M., Hilborn, R., Seeb, L.W. & Seeb, J.E. 2011. An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. *Molecular ecology resources*, 11, Suppl_1: 150–61.

Hayes, B.J., Bowman, P.J., Chamberlain, A.J. & Goddard, M.E. 2009. Invited review: genomic selection in dairy cattle: progress and challenges. *Journal of dairy science*, 92(2): 433–443.

He, C., Chen, L., Simmons, M., Li, P., Kim, S. & Liu, Z.J. 2003. Putative SNP discovery in interspecific hybrids of catfish by comparative EST analysis. *Animal Genetics*, 34(6): 445–448.

Heather, J.M. & Chain, B. 2016. The sequence of sequencers: the history of sequencing DNA. *Genomics*, 107: 1–8.

Hebert, P.D.N., Cywinska, A., Ball, S.L. & de Waard, J.R. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London, Series B: Biological Sciences.* 270(1512): 313–321.

Hedgecock, D., Shin, G., Gracey, A.Y., Van Den Berg, D. & Samanta, M.P. 2015. Second-generation linkage maps for the Pacific oyster *Crassostrea gigas* reveal errors in assembly of genome scaffolds. G3: Genes – Genomes. *Genetics*, 5(10): 2007–2019.

Hou, R., Bao, Z., Wang, S., Su, H., Li, Y., Du, H., Hu, J., Wang, S. & Hu, X. 2011. Transcriptome sequencing and *de novo* analysis for Yesso scallop (*Patinopecten yessoensis*) using 454 GS FLX. *PLOS ONE*, 6(6): e21560.

Hou, R., Bao, Z., Wang, S., Su, H., Li, Y., Du, H., Hu, J., Wang, S. & Hu, X. 2012. Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon. *BMC Genomics*, 13(1): 1.

Houston, R.D., Haley, C.S., Hamilton, A., Guy, D.R., Tinch, A.E., Taggart, J.B., McAndrew, B.J. & Bishop, S.C. 2008. Major quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*). *Genetics*, 178(2): 1109–1115.

Houston, R.D., Haley, C.S., Hamilton, A., Guy, D.R., Mota-Velasco, J.C., Gheyas, A.A., Tinch, A.E., *et al.* 2010. The susceptibility of Atlantic salmon fry to freshwater infectious pancreatic necrosis is largely explained by a major QTL. *Heredity*, 105(3): 318–327.

Houston, R.D., Taggart, J.B., Cézard, T., Bekaert, M., Lowe, N.R., Downing, A., Talbot, R., *et al.* 2014. Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). *BMC Genomics*, 15(1): 1–13.

Howe, K., Clark, M.D., Torroja, C.F., Torrance, J., Berthelot, C., Muffato, M., Collins, J.E., Humphray, S., McLaren, K., Matthews, L. & McLaren, S. 2013. The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, 496(7446): 498–503.

Hsieh, Y.H.P. 1998. Species substitution of restaurant fish entrees. *Journal of Food Quality*, 21(1): 1–11.

Huang, D.W., Sherman, B.T. & Lempicki, R.A. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1): 44–57.

Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M.J., Kenzelmann-Broz, D., Khalil, A.M., *et al.* 2010. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell*, 142(3): 409–419.

Huarte, M. & Rinn, J.L. 2010. Large non-coding RNAs: missing links in cancer? *Human Molecular Genetics*, 19(R2): R152–R161.

Huete-Pérez, J.A. & Quezada, F. 2013. Genomic approaches in marine biodiversity and aquaculture. *Biological Research*, 46(4): 353–361.

Hulata, G. 1995. A review of genetic improvement of the common carp (*Cyprinus carpio* L.) and other cyprinids by crossbreeding, hybridization and selection. *Aquaculture*, 129(1): 143–155.

Hulata, G. 2001. Genetic manipulations in aquaculture: a review of stock improvement by classical and modern technologies. *Genetica*, 111(1–3): 155–173.

Hung, T., Wang, Y., Lin, M.F., Koegel, A.K., Kotake, Y., Grant, G.D., Horlings, H.M., Shah, N., Umbricht, C. & Wang, P. 2011, Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nature Genetics*, 43(7): 621–629.

Hunter, R.L. & Markert, C.L. 1957. Histochemical demonstration of enzymes separated by zone electrophoresis in starch gels. *Science*, 125(3261): 1294–1295.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.J. & Joung, J.K. 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnology*, 31(3): 227–229.

Inoue, K., Yamashita, S., Hata, J.I., Kabeno, S., Asada, S., Nagahisa, E. & Fujita, T. 1990. Electroporation as a new technique for producing transgenic fish. Cell differentiation and development, 29(2): 123–128.

Iseli, C., Jongeneel, C.V. & Bucher, P. 1999. August. ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *ISMB*, Vol. 99: 138–148.

Jackson, T.R., Ferguson, M.M., Danzmann, R.G., Fishback, A.G., Ihssen, P.E., O'Connell, M. & Crease, T.J. 1998. Identification of two QTL influencing upper temperature tolerance in three rainbow trout (*Oncorhynchus mykiss*) half-sib families. *Heredity*, 80(2): 143–151.

Jao, L.E., Wente, S.R. & Chen, W. 2013. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proceedings of the National Academy of Sciences*, 110(34): 13904–13909.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. & Charpentier, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096): 816–821.

Johansen, S.D., Karlsen, B.O., Furmanek, T., Andreassen, M., Jørgensen, T.E., Bizuayehu, T.T., Breines, R., Emblem, Å., Kettunen, P., Luukko, K. & Edvardsen, R.B. 2011. RNA deep sequencing of the Atlantic cod transcriptome. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, 6(1): 18–22.

Jones, A.G., Rosenqvist, E., Berglund, A. & Avise, J.C. 1999. Clustered microsatellite mutations in the pipefish *Syngnathus typhle*. *Genetics*, 152(3): 1057–1063.

Jones, D.B., Jerry, D.R., Forêt, S., Konovalov, D.A. & Zenger, K.R. 2013. Genome-wide SNP validation and mantle tissue transcriptome analysis in the silver-lipped pearl oyster, *Pinctada maxima*. *Marine Biotechnology*, 15(6): 647–58.

Jones, F.C., Grabherr, M.G., Chan, Y.F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., *et al.* 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484(7392): 55–61.

Jorde, P.E., Palm, S. & Ryman, N. 1999. Estimating genetic drift and effective population size from temporal shifts in dominant gene marker frequencies. *Molecular Ecology*, 8(7): 1171–1178.

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. *Molecular Ecology*, 11(12): 2591–2598.

Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermüller, J. & Hofacker, I.L. 2007. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*, 316: 1484–1488.

Karlsson, S., Moen, T., Lien, S., Glover, K.A. & Hindar, K. 2011. Generic genetic differences between farmed and wild Atlantic salmon identified from a 7K SNP-chip. *Molecular ecology resources*, 11, Suppl_1: 247–53.

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 micro satellite-containing cDNAs in the skin. *Gene*, 285(1–2): 157–168.

Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Qu, W., Ahsan, B., Yamada, T.,*et al.* 2007. The medaka draft genome and insights into vertebrate genome evolution. *Nature*, 447(7145): 714–719.

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(1): 1.

Kessuwan, K., Kubota, S., Liu, Q., Sano, M., Okamoto, N., Sakamoto, T., Yamashita, H., Nakamura, Y. & Ozaki, A. 2016. Detection of growth-related quantitative trait loci and high-resolution genetic linkage maps using simple sequence repeat markers in the kelp grouper (*Epinephelus bruneus*). *Marine Biotechnology*, 18(1): 57–84.

Kim, Y.G., Cha, J. & Chandrasegaran, S. 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences*, 93(3): 1156–1160.

Knight, J.C. 2004. Allele-specific gene expression uncovered. *Trends in Genetics*, 20(3): 113–116.

Kobayashi, T., Ide, A., Hiasa, T., Fushiki, S. & Ueno, K. 1994. Production of cloned amago salmon *Oncorhynchus rhodurus*. *Fisheries Science*, 60(3): 275–281.

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(3): 1225–1232.

Kondo, H., Suga, R., Suda, S., Nozaki, R., Hirono, I., Nagasaka, R., Kaneko, G., Ushio, H. & Watabe, S. 2011. EST analysis on adipose tissue of rainbow trout *Oncorhynchus mykiss* and tissue distribution of adiponectin. *Gene*, 485(1): 40–45.

Kongchum, P., Palti, Y., Hallerman, E.M., Hulata, G. & David, L. 2010. SNP discovery and development of genetic markers for mapping innate immune response genes in common carp (*Cyprinus carpio*). *Fish Shellfish Immunology*, 29(2): 356–361.

Kucuktas, H. & Liu, Z. 2007. Allozyme and mitochondrial markers. *In* Z. Liu, ed. *Aquaculture genome technologies*, pp. 73–85. Ames, IA, USA, Blackwell Publishing.

Kucuktas, H., Wang, S., Li, P., He, C., Xu, P., Sha, Z., Liu, H., Jiang, Y., Baoprasertkul, P., Somridhivej, B. & Wang, Y. 2009. Construction of genetic linkage maps and comparative genome analysis of catfish using gene-associated markers. *Genetics*, 181(4): 1649–1660.

Küttner, E., Moghadam, H.K., Skúlason, S., Danzmann, R.G. & Ferguson, M.M. 2011. Genetic architecture of body weight, condition factor and age of sexual maturation in Icelandic Arctic charr (*Salvelinus alpinus*). *Molecular Genetics and Genomics*, 286(1): 67–79.

Laghari, M.Y., Lashari, P., Zhang, X., Xu, P., Narejo, N.T., Xin, B., Zhang, Y. and Sun, X. 2015. QTL mapping for economically important traits of common carp (*Cyprinus carpio* L.). *Journal of Applied Genetics*, 56(1): 65–75.

Laghari, M.Y., Lashari, P., Zhang, X., Xu, P., Narejo, N.T., Liu, Y., Mehboob, S., Al-Ghanim, K., Zhang, Y. & Sun, X. 2014. Mapping QTLs for swimming ability related traits in *Cyprinus carpio* L. *Marine biotechnology*, 16(6): 629–37.

Lande, R. & Thompson, R. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics*, 124(3): 743–756.

Lauth, X., Buchanan, J. & Hansen, K. 2012. Efficient targeted genome editing for finfish aquaculture and other industries. *Industrial Biotechnology*, 8(3): 90–93.

Le Bras, Y., Dechamp, N., Krieg, F., Filangi, O., Guyomard, R., Boussaha, M., Bovenhuis, H., Pottinger, T.G., Prunet, P. & Le Roy, P. 2011. Detection of QTL with effects on osmoregulation capacities in the rainbow trout (*Oncorhynchus mykiss*). *BMC Genetics*, 12(1): 46.

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. *molecular ecology* 9(7): 995–997.

Leder, E., Danzmann, R. & Ferguson, M. 2006. The candidate gene, Clock, localizes to a strong spawning time quantitative trait locus region in rainbow trout. *Journal of Heredity*, 97(1): 74–80.

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(1): 237–244.

Lee, B.Y., Penman, D.J. & Kocher, T.D. 2003. Identification of a sex-determining region in Nile tilapia (*Oreochromis niloticus*) using bulked segregant analysis. *Animal Genetics*, 34(5): 379–383.

Li, W.H. 1997. Genome organization and evolution. *In* W.H. Li, ed. *Molecular evolution*. Sunderland, MA, USA. Sinauer Associates, Inc.

Li, C., Feng, W., Qiu, L., Xia, C., Su, X., Jin, C., Zhou, T., Zeng, Y. & Li, T. 2012. Characterization of skin ulceration syndrome associated microRNAs in sea cucumber *Apostichopus japonicus* by deep sequencing. *Fish & Shellfish Immunology*, 33(2): 436–441.

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(1–4): 63–73.

Li, L. & Guo, X.M. 2004. AFLP-based genetic linkage maps of the Pacific oyster *Crassostrea gigas* Thunberg. *Marine biotechnology*, 6(1): 26–36.

Li, Y.T., 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(1–4): 143–156.

Li, P., Peatman, E., Wang, S., Feng, J., He, C., Baoprasertkul, P., Xu, P., *et al.* 2007. Towards the ictalurid catfish transcriptome: generation and analysis of 31,215 catfish ESTs. *BMC Genomics*, 8(1): 177.

Li, Y., Liu, S., Qin, Z., Waldbieser, G., Wang, R., Sun, L., Bao, L., Danzmann, R.G., Dunham, R. & Liu, Z. 2015. Construction of a high-density, high-resolution genetic map and its integration with BAC-based physical map in channel catfish. *DNA Research*, 22(1): 39–52.

Li, Y. & He, M. 2014. Genetic mapping and QTL analysis of growth-related traits in *Pinctada fucata* using restriction-site associated DNA sequencing. *PLOS ONE*, 9(11): p.e111707.

Liao, X., Cheng, L., Xu, P., Lu, G., Wachholtz, M., Sun, X. & Chen, S. 2013. Transcriptome analysis of crucian carp (*Carassius auratus*), an important aquaculture and hypoxia-tolerant species. *PLOS ONE*, 8(4): e62308.

Lien, S., Gidskehaug, L., Moen, T., Hayes, B.J., Berg, P.R., Davidson, W.S., Omholt, S.W. & Kent, M.P., 2011. A dense SNP-based linkage map for Atlantic salmon (*Salmo salar*) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. *BMC Genomics*, 12(1): 615.

Liu, H., Jiang, Y., Wang, S., Ninwichian, P., Somridhivej, B., Xu, P., Abernathy, J., Kucuktas, H. & Liu, Z. 2009. Comparative analysis of catfish BAC end sequences with the zebrafish genome. *BMC Genomics*, 10(1): 592.

Li, H., Liu, X. & Zhang, G. 2012. A consensus microsatellite-based linkage map for the hermaphroditic bay scallop (*Argopecten irradians*) and its application in size-related QTL analysis. 7(10): eE46926.

Liu, Z.J., Moav, B., Faras, A., Guise, K., Kapuscinski, A. & Hackett, P. 1990. Development of expression vectors for transgenic fish. *Nature Biotechnology*, 8(12): 1268–1272.

Liu, Z.J., Li, P., Argue, B. & Dunham, R. 1998. Inheritance of RAPD markers in channel catfish (*Ictalurus punctatus*), blue catfish (*I. furcatus*) and their F1, F2 and backcross hybrids. *Animal Genetics,* 29(1): 58–62.

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. *Molecular and General Genetics*, 258(3): 260–268.

Liu, Z.J., Li, P., Argue, B.J. & Dunham, R.A. 1999. Random amplified polymorphic DNA markers: usefulness for gene mapping and analysis of genetic variation of catfish. *Aquaculture*, 174(1–2): 59–68.

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. *Transactions of the American Fisheries Society,* 128(2): 317–327.

Liu, Z.J., Li, P., Kocabas, A., Karsi, A. & Ju, Z.L. 2001. Microsatellite-containing genes from the channel catfish brain: evidence of trinucleotide repeat expansion in the coding region of nucleotide excision repair gene RAD23B. *Biochemical and biophysical research communications*, 289(2): 317–324.

Liu, Z.J., Karsi, A., Li, P., Cao, D., Ju, Z. & Dunham, R. 2003. An AFLP-based genetic linkage map of channel catfish (*Ictalurus punctatus*) constructed by using an interspecific hybrid resource family. *Genetics*, 165(2): 687–694.

Liu, Z.J. & Cordes, J.F. 2004. DNA marker technologies and their applications in aquaculture genetics. *Aquaculture*, 238(1–4): 1–37.

Liu, X., Liu, X., & Zhang, G. 2007. Identification of quantitative trait loci for growth- related traits in Pacific abalone *Haliotis discus hannai* Ino. *Aquaculture Research*, 38(8): 789–797.

Liu, Z.J. 2007a. Marking the genome: restriction fragment length polymorphism (RFLP). *In* Z.J. Liu, ed. *Aquaculture genome technologies,* pp. 11–20. Ames, IA, USA, Blackwell Publishing.

Liu, Z.J. 2007b. Random amplified polymorphic DNA (RAPD). *In* Z.J. Liu, ed. *Aquaculture genome technologies,* pp. 21–28. Ames, IA, USA, Blackwell Publishing.

Liu, Z.J. 2007c. Amplified fragment length polymorphism (AFLP). *In* Z.J. Liu, ed. *Aquaculture genome technologies,* pp. 29–42. Ames, IA, USA, Blackwell Publishing.

Liu, Z.J. 2007d. Single nucleotide polymorphism (SNP). *In* Z.J. Liu, ed. *Aquaculture genome technologies,* pp. 59–72. Ames, IA, USA, Blackwell Publishing.

Liu, S., Zhou, Z., Lu, J., Sun, F., Wang, S., Liu, H., Jiang, Y. *et al.* 2011. Generation of genome-scale gene-associated SNPs in catfish for the construction of a high-density SNP array. *BMC Genomics*, 12(1): 53.

Liu, S., Sun, L., Li, Y., Sun, F., Jiang, Y., Zhang, Y., Zhang, J., *et al.* 2014. Development of the catfish 250K SNP array for genome-wide association studies. *BMC Research Notes*, 7(1): 135.

Liu, P., Wang, L., Wan, Z.Y., Ye, B.Q., Huang, S., Wong, S.M. & Yue, G.H. 2015a. Mapping QTL for resistance against viral nervous necrosis disease in Asian seabass. *Marine Biotechnology*, 18(1): 107–116.

Liu, S., Vallejo, R.L., Gao, G., Palti, Y., Weber, G.M., Hernandez, A. & Rexroad III, C.E. 2015b. Identification of single-nucleotide polymorphism markers associated with cortisol response to crowding in rainbow trout. *Marine Biotechnology*, 17(3): 328–37.

Lo, H.S., Wang, Z., Hu, Y., Yang, H.H., Gere, S., Buetow, K.H. & Lee, M.P. 2003. Allelic variation in gene expression is common in the human genome. *Genome Research*, 13(8): 1855–1862.

Lorenz, S., Brenna-Hansen, S., Moen, T., Roseth, A., Davidson, W.S., Omholt, S.W. & Lien, S. 2010. BAC-based upgrading and physical integration of a genetic SNP map in Atlantic salmon. *Animal Genetics*, 41(1): 48–54.

Lou, Y.D. & Purdom, C.E. 1984. Polyploidy induced by hydrostatic pressure in rainbow trout, *Salmo gairdneri* Richardson. *Journal of Fish Biology*, 25(3): 345–351.

Loukovitis, D., Sarropoulou, E., Tsigenopoulos, C.S., Batargias, C., Magoulas, A., Apostolidis, A.P., Chatziplis, D. & Kotoulas, G. 2011. Quantitative trait loci involved in sex determination and body growth in the gilthead sea bream (*Sparus aurata* L.) through targeted genome scan. *PLOS ONE*, 6(1): e16599.

Louro, B., Passos, A.L.S., Souche, E.L., Tsigenopoulos, C., Beck, A., Lagnel, J., Bonhomme, F., *et al.* 2010. Gilthead sea bream (*Sparus auratus*) and European sea bass (*Dicentrarchus labrax*) expressed sequence tags: characterization, tissue-specific expression and gene markers. *Marine Genomics*, 3(3): 179–191.

Louro, B., Kuhl, H., Tine, M., De Koning, D.J., Batargias, C., Volckaert, F.A., Reinhardt, R., *et al.* 2016. Characterization and refinement of growth related quantitative trait loci in European sea bass (*Dicentrarchus labrax*) using a comparative approach. *Aquaculture*, 455: 88–21.

Lu, X., Wang, H., Liu, B. & Xiang, J. 2013. Three EST-SSR markers associated with QTL for the growth of the clam *Meretrix meretrix* revealed by selective genotyping. *Marine Biotechnology* 15(1): 16–25.

Luo, M.C., Thomas, C., You, F.M., Hsiao, J., Ouyang, S., Buell, C.R., Malandro, M., McGuire, P.E., Anderson, O.D. & Dvorak, J. 2003. High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. *Genomics*, 82(3): 378–389.

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. *Molecular Ecology*, 11(9): 1877–1883.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., *et al.* 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057): 376–380.

Marko, P.B., Lee, S.C., Rice, A.M., Gramling, J.M., Fitzhenry, T.M., McAlister, J.S., Harper, G.R. & Moran, A.L. 2004. Fisheries: mislabelling of a depleted reef fish. *Nature*, 430(6997): 309–310.

Martens, J.A., Laprade, L. & Winston, F. 2004. Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene. *Nature*, 429(6991): 571–574.

Massault, C., Hellemans, B., Louro, B., Batargias, C., Van Houdt, J., Canario, A., Volckaert, F., Bovenhuis, H., Haley, C. & De Koning, D. 2010. QTL for body weight, morphometric traits and stress response in European sea bass *Dicentrarchus labrax*. *Animal Genetics*, 41(4): 337–345.

Massault, C., Franch, R., Haley, C., De Koning, D.J., Bovenhuis, H., Pellizzari, C., Patarnello, T. & Bargelloni, L. 2011. Quantitative trait loci for resistance to fish pasteurellosis in gilthead sea bream (*Sparus aurata*). *Animal Genetics*, 42(2): 191–203.

Maxam, A.M. & Gilbert, W. 1977. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences*, 74(2): 560–564.

May, B. 2003. Allozyme variation. *In* E.M. Hallerman, ed. *Population genetics: principles and applications for fisheries scientists*, pp. 23–36. Bethesda, Maryland, USA, American Fisheries Society.

McClelland, E. & Naish, K. 2010. Quantitative trait locus analysis of hatch timing, weight, length and growth rate in coho salmon, *Oncorhynchus kisutch*. *Heredity*, 105(6): 562–573.

Meng, X. 2008. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nature biotechnology*, 26(6): 695–701.

Mercer, T.R., Dinger, M.E. & Mattick, J.S. 2009. Long non-coding RNAs: insights into functions. *Nature Reviews Genetics*, 10(3): 155–159.

Merker, R. & Woodruff, R. 1996. Molecular evidence for divergent breeding groups of walleye (*Stizostedion vitreum*) in tributaries to western Lake Erie. *Journal of Great Lakes Research*, 22(2): 280–288.

Metzgar, D., Bytof, J. & Wills, C. 2000. Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome research*, 10(1): 72–80

Meuwissen, T.H.E., Hayes, B.J. & Goddard, M.E. 2001. Prediction of total genetic value using genome wide dense marker maps. *Genetics*, 157(4): 1819–1829.

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(1–4): 91–105

Miller, D.D. & Mariani, S. 2010. Smoke, mirrors, and mislabelled cod: poor transparency in the European food fish industry. *Frontiers in Ecology and the Environment*, 8(10): 517–521.

Min, X.J., Butler, G., Storms, R. & Tsang, A. 2005. OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Research*, 33(suppl_2): W677–W680.

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. *Molecular Ecology* 13: 1085–1098.

Moen, T., Baranski, M., Sonesson, A.K. & Kjøglum, S. 2009. Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*): population-level associations between markers and trait. *BMC Genomics*, 10(1): 368.

Moen, T., Sonesson, A.K., Hayes, B., Lien, S., Munck, H. & Meuwissen, T.H. 2007. Mapping of a quantitative trait locus for resistance against infectious salmon anaemia in Atlantic salmon (*Salmo salar*): comparing survival analysis with analysis on affected/resistant data. *BMC Genetics*, 8(1): 1.

Moghadam, H.K., Poissant, J., Fotherby, H., Haidle, L., Ferguson, M.M. & Danzmann, R.G. 2007. Quantitative trait loci for body weight, condition factor and age at sexual maturation in Arctic charr (*Salvelinus alpinus*): comparative analysis with rainbow trout (*Oncorhynchus mykiss*) and Atlantic salmon (*Salmo salar*). *Molecular Genetics and Genomics*, 277(6): 647–661.

Moran, C. 1993. Microsatellite repeats in pig (*Sus domesticus*) and chicken (*Gallus domesticus*) genomes. *Journal of Heredity*, 84(4): 274–280.

Murata, Y., Oda, S. & Mitani, H. 2012. Allelic expression changes in Medaka (*Oryzias latipes*) hybrids between inbred strains derived from genetically distant populations. *PLOS ONE*, 7(5): e36875.

Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R. & Fraser, P. 2008. The air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science*, 322(5908): 1717–1720.

Naish, K.A., Warren, M., Bardakci, F., Skibinski, D.O., Carvalho, G.R. & Mair, G.C. 1995. Multilocus DNA fingerprinting and RAPD reveal similar genetic relationships between strains of *Oreochromis niloticus* (Pisces: Cichlidae) *Molecular Ecology*. 4(2): 271–274.

Nakaya, H.I., Amaral, P.P., Louro, R., Lopes, A., Fachel, A.A., Moreira, Y.B., El-Jundi, T.A., Da Silva, A.M., Reis, E.M. & Verjovski-Almeida, S. 2007. Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. *Genome Biology*, 8(3): R43.

Negrín-Báez, D., Navarro, A., Afonso, J.M., Ginés, R. & Zamorano, M.J. 2015. Detection of QTL associated with three skeletal deformities in gilthead seabream (*Sparus aurata* L.): Lordosis, vertebral fusion and jaw abnormality. *Aquaculture*, 448: 123–127.

Nemudryi, A.A., Valetdinova, K.R., Medvedev, S.P. & Zakian, S.M. 2014. TALEN and CRISPR/Cas genome editing systems: tools of discovery. *Acta Naturae*, 6(3): 19–40.

Ng, S.H., Artieri, C.G., Bosdet, I.E., Chiu, R., Danzmann, R.G., Davidson, W.S., Ferguson, M.M., *et al.* 2005. A physical map of the genome of Atlantic salmon, *Salmo salar*. *Genomics*, 86(4): 396–404.

Nichols, K.M., Edo, A.F., Wheeler, P.A. & Thorgaard, G.H. 2008. The genetic basis of smoltification-related traits in *Oncorhynchus mykiss*. *Genetics*, 179(3): 1559–1575.

Nichols, K.M., Young, W.P., Danzmann, R.G., Robison, B.D., Rexroad, C., Noakes, M., Phillips, R.B., *et al.* 2003. A consolidated linkage map for rainbow trout (*Oncorhynchus mykiss*). *Animal genetics*, 34(2): 102–115.

Nielsen, E.E., Hansen, M.M. & Mensberg, K.L. 1998. Improved primer sequences for the mitochondrial ND1, ND3/4 and ND5/6 segments in salmonid fishes: application to RFLP analysis of Atlantic salmon. *Journal of Fish Biology*, 53(1): 216–220.

Nielsen, J.L., Crow, K.D. & Fountain, M.C. 1999. Microsatellite diversity and conservation of a relic trout population: McCloud river redband trout. *Molecular Ecology*, 8: S129–S142.

Ninwichian, P., Peatman, E., Perera, D., Liu, S., Kucuktas, H., Dunham, R. & Liu, Z. 2012a. Identification of a sex-linked marker for channel catfish. *Animal Genetics*, 43(4): 476–477.

Ninwichian, P., Peatman, E., Liu, H., Kucuktas, H., Somridhivej, B., Liu, S., Li, P., *et al.* 2012b. Second-generation genetic linkage map of catfish and its integration with the BAC-based physical map. *G3: Genes, Genomes, Genetics*, 2(10): 1233–1241.

Norman, J.D., Danzmann, R.G., Glebe, B. & Ferguson, M.M. 2011. The genetic basis of salinity tolerance traits in Arctic charr (*Salvelinus alpinus*). *BMC Genetics*, 12(1): 81.

Ogden, R., Gharbi, K., Mugue, N., Martinsohn, J., Senn, H., Davey, J.W., Pourkazemi, M., *et al.* 2013. Sturgeon conservation genomics: SNP discovery and validation using RAD sequencing. *Molecular Ecology*, 22(11): 3112–3123.

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(1–4): 41–48.

Okumuş, Ý. & Çiftci, Y. 2003. Fish population genetics and molecular markers: II-molecular markers and their applications in fisheries and aquaculture. *Turkish Journal of Fisheries and Aquatic Sciences*, 3(1): 51–79.

O'malley, K., Sakamoto, T., Danzmann, R. & Ferguson, M. 2003. Quantitative trait loci for spawning date and body weight in rainbow trout: testing for conserved effects across ancestrally duplicated chromosomes. *Journal of Heredity*, 94(4): 273–284.

Ozaki, A., Araki, K., Okamoto, H., Okauchi, M., Mushiake, K., Yoshida, K., Tsuzaki, T., Fuji, K., Sakamoto, T. & Okamoto, N. 2012. Progress of DNA marker-assisted breeding in maricultured finfish. *Bulletin of Fisheries Research Agency*, 35: 31–37.

Palaiokostas, C., Bekaert, M., Davie, A., Cowan, M.E., Oral, M., Taggart, J.B., Gharbi, K., McAndrew, B.J., Penman, D.J. & Migaud, H. 2013. Mapping the sex determination locus in the Atlantic halibut (*Hippoglossus hippoglossus*) using RAD sequencing. *BMC Genomics*, 14(1): 566.

Palmiter, R.D., Brinster, R.L., Hammer, R.E., Trumbauer, M.E., Rosenfeld, M.G., Birnberg, N.C. & Evans, R.M. 1982. Dramatic growth of mice that develop from eggs microinjected with metallothionein-growth hormone fusion genes. *Nature*, 300(5893): 611–615.

Palti, Y., Danzmann, R.G. & Rexroad III, C.E. 2003. Characterization and mapping of 19 polymorphic microsatellite markers for rainbow trout (*Oncorhynchus mykiss*). *Animal genetics,* 34(2): 153–156.

Palti, Y., Luo, M.C., Hu, Y., Genet, C., You, F.M., Vallejo, R.L., Thorgaard, G.H., Wheeler, P.A. and Rexroad, C.E. 2009. A first generation BAC-based physical map of the rainbow trout genome. *BMC Genomics*, 10(1): 1.

Palti, Y., Gao, G., Liu, S., Kent, M.P., Lien, S., Miller, M.R., Rexroad III, C.E. & Moen, T. 2015a. The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. *Molecular ecology resources*, 15(3): 662–672.

Palti, Y., Vallejo, R.L., Gao, G., Liu, S., Hernandez, A.G., Rexroad III, C.E. & Wiens, G.D. 2015b. Detection and validation of QTL affecting bacterial cold water disease resistance in rainbow trout using restriction-site associated DNA sequencing. *PLOS ONE*, 10(9): p.e0138435.

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: 361–382

Parsons, J.E. & Thorgaard, G.H. 1985. Production of androgenetic diploid rainbow trout. *Journal of Heredity*, 76(3): 177–181.

Pauli, A., Valen, E., Lin, M.F., Garber, M., Vastenhouw, N.L., Levin, J.Z., Fan, L., Sandelin, A., Rinn, J.L., Regev, A. & Schier, A.F. 2012. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Research*, 22(3): 577–591.

Penman, D.J., Iyengar, A., Beeching, A.J., Rahman, A., Sulaiman, Z. & Maclean, N. 1991. Patterns of transgene inheritance in rainbow trout (*Oncorhynchus mykiss*). *Molecular Reproduction and Development*, 30(3): 201–206.

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(1–4): 105–118.

Perry, G.M., Danzmann, R.G., Ferguson, M.M. & Gibson, J.P. 2001. Quantitative trait loci for upper thermal tolerance in outbred strains of rainbow trout (*Oncorhynchus mykiss*). *Heredity*, 86(3): 333–341.

Perry, G.M., Martyniuk, C.M., Ferguson, M.M. & Danzmann, R.G. 2005. Genetic parameters for upper thermal tolerance and growth-related traits in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*, 250(1–2): 120–128.

Petruk, S., Sedkov, Y., Riley, K.M., Hodgson, J., Schweisguth, F., Hirose, S., Jaynes, J.B., Brock, H.W. & Mazo, A. 2006. Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference. *Cell*, 127(6): 1209–1221.

Pickrell, J.K., Marioni, J.C., Pai, A.A., Degner, J.F., Engelhardt, B.E., Nkadori, E., Veyrieras, J.B., Stephens, M., Gilad, Y. & Pritchard, J.K. 2010. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature*, 464(7289): 768–772.

Piferrer, F., Beaumont, A., Falguière, J.C., Flajšhans, M., Haffray, P. & Colombo, L. 2009. Polyploid fish and shellfish: production, biology and applications to aquaculture for performance improvement and genetic containment. *Aquaculture*, 293(3–4): 125–156.

Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W.J. & Pandolfi, P.P. 2010. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature*, 465(7301): 1033–1038.

Ponjavic, J., Oliver, P.L., Lunter, G. & Ponting, C.P. 2009. Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. *PLOS Genetics*, 5(8): e1000617.

Powers, D.A., Chen, T.T., Dunham, R.A. & Murray, J.A.H. 1992. *Transgenic fish*. *Transgenesis: applications of gene transfer*, pp. 233–249.

Provasi, E., Genovese, P., Lombardo, A., Magnani, Z., Liu, P.Q., Reik, A., Chu, V., Paschon, D.E., Zhang, L. & Kuball, J. 2012. Editing T cell specificity towards leukemia by zinc finger nucleases and lentiviral gene transfer. *Nature medicine*, 18(5): 807–815.

Pruett, C.L., Saillant, E. & Gold, J.R. 2005. Historical population demography of red snapper (*Lutjanus campechanus*) from the northern Gulf of Mexico based on analysis of sequences of mitochondrial DNA. *Marine Biology*, 147(3): 593–602.

Ragavendran, A. & Muir, W.M. 2011. Genomic selection in aquaculture: methods and practical considerations. *In* Liu, ed. *Next generation sequencing and whole genome selection in aquaculture*, pp. 165–183. Ames, USA, Wiley and Blackwell Publishing.

Ratnasingham, S. & Hebert, P.D. 2007. BOLD: The Barcode of Life Data System. *Molecular ecology notes*, 7(3): 355–364.

Ravago, R.G., Monje, V.D. & Juinio-Meñez, M.A. 2002. Length and sequence variability in mitochondrial control region of the milkfish, *Chanos chanos*. *Marine Biotechnology*, 4(1): 40–50.

Reid, D., Szanto, A., Glebe, B., Danzmann, R. & Ferguson, M. 2005. QTL for body weight and condition factor in Atlantic salmon (*Salmo salar*): comparative analysis with rainbow trout (*Oncorhynchus mykiss*) and Arctic charr (*Salvelinus alpinus*). *Heredity*, 94(2): 166–172.

Renaut, S., Nolte, A.W. & Bernatchez, L. 2010. Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. Salmonidae). Molecular Ecology, 19: 115–131.

Reslewic, S., Zhou, S., Place, M., Zhang, Y., Briska, A., Goldstein, S., Churas, C., *et al.* 2005. Whole-genome shotgun optical mapping of *Rhodospirillum rubrum*. *Applied and Environmental Microbiology*, 71(9): 5511–5522.

Rexroad, C.E., Coleman, R.L., Martin, A.M., Hershberger, W.K. & Killefer, J. 2001. Thirty-five polymorphic microsatellite markers for rainbow trout (*Oncorhynchus mykiss*). *Animal Genetics*, 32(5): 317–319.

Rexroad, C.E., Vallejo, R.L., Liu, S., Palti, Y. & Weber, G.M. 2013. Quantitative trait loci affecting response to crowding stress in an F(2) generation of rainbow trout produced through phenotypic selection. *Marine Biotechnology*, 15(5): 613–27.

Rico, C., Rico, I. & Hewitt, G. 1996. 470 million years of conservation of microsatellite loci among fish species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 263(1370): 549–557

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J. & Segal, E. 2007. Functional demarcation of active and silent chromatin domains in human< i> HOX</i> loci by noncoding RNAs. *Cell*, 129(7): 1311–1323.

Rise, M.L., Von Schalburg, K.R., Brown, G.D., Mawer, M.A., Devlin, R.H., Kuipers, N., Busby, M., *et al.* 2004. Development and application of a salmonid EST database and cDNA microarray: data mining and interspecific hybridization characteristics. *Genome Research*, 14(3): 478–490.

Robison, B., Wheeler, P., Sundin, K., Sikka, P. & Thorgaard, G. 2001. Composite interval mapping reveals a major locus influencing embryonic development rate in rainbow trout (*Oncorhynchus mykiss*). *Journal of Heredity*, 92(1): 16–22.

Rocke, D.M. & Durbin, B. 2001. A model for measurement error for gene expression arrays. *Journal of Computational Biology*, 8(6): 557–569.

Rodríguez-Ramilo, S.T., Toro, M.A., Bouza, C., Hermida, M., Pardo, B.G., Cabaleiro, S., Martínez, P. & Fernández, J. 2011. QTL detection for *Aeromonas salmonicida* resistance related traits in turbot (*Scophthalmus maximus*). *BMC Genomics*, 12(1): 1.

Rodríguez-Ramil, S.T., Fernández, J., Toro, M.A., Bouza, C., Hermida, M., Fernández, C., Pardo, B.G., Cabaleiro, S. & Martínez, P 2013. Uncovering QTL for resistance and survival time to *Philasterides dicentrarchi* in turbot (*Scophthalmus maximus*). *Animal Genetics*, 44(2): 149–157.

Rodríguez-Ramilo, S.T., De La Herrán, R., Ruiz-Rejón, C., Hermida, M., Fernández, C., Pereiro, P., Figueras, A., *et al.* 2014. Identification of quantitative trait loci associated with resistance to viral haemorrhagic septicaemia (VHS) in turbot (*Scophthalmus maximus*): a comparison between bacterium, parasite and virus diseases. *Marine Biotechnology*, 16(3): 265–276.

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(1–3): 25–41

Rozowsky, J., Abyzov, A., Wang, J., Alves, P., Raha, D., Harmanci, A., Leng, J., Bjornson, R., Kong,Y., Kitabayashi, N. & Bhardwaj, N. 2011. AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Molecular Systems Biology*, 7(1): 522.

Ryynänen, H.J. & Primmer, C.R. 2006. Single nucleotide polymorphism (SNP) discovery in duplicated genomes: intron-primed exon-crossing (IPEC) as a strategy for avoiding amplification of duplicated loci in Atlantic salmon (*Salmo salar*) and other salmonid fishes. *BMC Genomics*, 7(1): 192.

Sakamoto, T., Danzmann, R.G., Gharbi, K., Howard, P., Ozaki, A., Khoo, S.K., Woram, R.A., *et al.*. 2000. A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. *Genetics*, 155(3): 1331–1345.

Salem, M., Rexroad, C.E., Wang, J., Thorgaard, G.H. & Yao, J. 2010. Characterization of the rainbow trout transcriptome using Sanger and 454- pyrosequencing approaches. *BMC Genomics*, 11(1): 564.

Sánchez-Molano, E., Cerna, A., Toro, M.A., Bouza, C., Hermida, M., Pardo, B.G., Cabaleiro, S., Fernández, J. & Martínez, P. 2011. Detection of growth-related QTL in turbot (*Scophthalmus maximus*). *BMC Genomics*, 12(1): 473.

Sanger, F., Nicklen, S. & Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors.

Proceedings of the National Academy of Sciences, 74(12): 5463–5467.

Sarropoulou, E., Franch, R., Louro, B., Power, D.M., Bargelloni, L., Magoulas, A., Senger, *et al.* 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(1): 1.

Sauvage, C., Boudry, P., De Koning, D.J., Haley, C.S., Heurtebise, S. & Lapègue, S. 2010. QTL for resistance to summer mortality and OsHV- 1 load in the Pacific oyster (*Crassostrea gigas*). *Animal genetics*, 41(4): 390–399.

Scheerer, P.D., Thorgaard, G.H., Allendorf, F.W. & Knudsen, K.L. 1986. Androgenetic rainbow trout produced from inbred and outbred sperm sources show similar survival. *Aquaculture*, 57(1): 289–298.

Schwartz, D.C., Li, X., Hernandez, L.I., Ramnarain, S.P., Huff, E.J. & Wang, Y.K. 1993. Ordered restriction maps of *Saccharomyces cerevisiae* chromosomes constructed by optical mapping. *Science*, 262(5130): 110–114.

Seki, S., Agresti, J.J., Gall, G.A.E., Taniguchi, N. & May, B. 1999. AFLP analysis of genetic diversity in three populations of ayu *Plecoglossus altivelis*. *Fisheries Science*, 65(6): 888–892

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(6): 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*). *Marine Biotechnology*, 6(4): 364–377.

Serre, D., Gurd, S., Ge, B., Sladek, R., Sinnett, D., Harmsen, E., Bibikova, M., *et al.* 2008. Differential allelic expression in the human genome: a robust approach to identify genetic and epigenetic cis-acting mechanisms regulating gene expression. *PLOS Genetics*, 4(2): e1000006.

Shao, C., Niu, Y., Rastas, P., Liu, Y., Xie, Z., Li, H., Wang, L., *et al.* 2015. Genome-wide SNP identification for the construction of a high-resolution genetic map of Japanese flounder (*Paralichthys olivaceus*): applications to QTL mapping of *Vibrio anguillarum* disease resistance and comparative genomic analysis. *DNA Research*, 22(2): 161–170.

Shelton, W.L., Rodriguez-Guerrero, D. & Lopez-Macias, J. 1981. Factors affecting androgen sex reversal of *Tilapia aurea*. *Aquaculture*, 25(1): 59–65.

Shen, Y., Catchen, J., Garcia, T., Amores, A., Beldorth, I., Wagner, J., Zhang, Z., Postlethwait, J., Warren, W. & Schartl, M. 2012. Identification of transcriptome SNPs between Xiphophorus lines and species for assessing allele specific gene expression within F 1 interspecies hybrids. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, 155(1): 102–108.

Shi, X., Ng, D.W., Zhang, C., Comai, L., Ye, W. & Chen, Z.J. 2012. Cis-and trans-regulatory divergence between progenitor species determines gene-expression novelty in *Arabidopsis* allopolyploids. *Nature*, *Communications*, 3(1): 950.

Shirak, A., Seroussi, E., Cnaani, A., Howe, A.E., Domokhovsky, R., Zilberman, N., Kocher, T.D., Hulata, G. & Ron, M. 2006. Amh and Dmrta2 genes map to tilapia (*Oreochromis* spp.) linkage group 23 within quantitative trait locus regions for sex determination. *Genetics*, 174(3): 1573–1581.

Simmons, M., Mickett, K., Kucuktas, H., Li, P., Dunham, R. & Liu, Z.J. 2006. Comparison of domestic and wild channel catfish (*Ictalurus punctatus*) populations provides no evidence for genetic impact. *Aquaculture*, 252(2–4): 133–146.

Smith, M.H, & Chessar, R.K. 1981. Rationale for conserving genetic-variation of fish gene pools. *Ecological Bulletins,* 13–20.

Soares, M.B., Bonaldo, M.F., Jelene, P., Su, L., Lawton, L. & Efstratiadis, A. 1994. Construction and characterization of a normalized cDNA library. *Proceedings of the National Academy of Sciences*, 91(20): 9228–9232.

Soderlund, C., Humphray, S., Dunham, A. & French, L. 2000. Contigs built with fingerprints, markers, and FPC V4.7. *Genome Research*, 10(11): 1772–1787.

Sonesson, A.K. 2011. Genomic selection for aquaculture: principles and procedures. *In* Z Liu, ed. *Next generation sequencing and whole genome selection in aquaculture*, pp. 151–163. Ames, IA, USA. Wiley-Blackwell.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology*, 98(3): 503–517.

Springer, N.M. & Stupar, R.M. 2007. Allele-specific expression patterns reveal biases and embryo- specific parent-of-origin effects in hybrid maize. *Plant Cell*, 19(8): 2391–2402.

Star, B., Nederbragt, A.J., Jentoft, S., Grimholt, U., Malmstrøm, M., Gregers, T.F., Rounge, T.B., *et al.* 2011. The genome sequence of Atlantic cod reveals a unique immune system. *Nature*, 477(7363): 207–210.

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. *Molecular Biology and Evolution*, 19(7): 1198–1202.

Stickney, R.R. 1994. *Principles of aquaculture*. New York, USA, John Wiley and Sons, Inc.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., *et al.* 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43): 15545–15550.

Sulston, J., Mallett, F., Staden, R., Durbin, R., Horsnell, T. & Coulson, A. 1988. Software for genome mapping by fingerprinting techniques. *Bioinformatics*, 4(1): 125–132.

Sun, L., Liu, S., Wang, R., Jiang, Y., Zhang, Y., Zhang, J., Bao, L., *et al.* 2014. Identification and analysis of genome-wide SNPs provide insight into signatures of selection and domestication in channel catfish (*Ictalurus punctatus*). *PLOS ONE*, 9(10): p.e109666.

Sun, Y., Song, W.Q., Zhong, Y.C., Zhang, R.S., Abatzopoulos, T.J. & Chen, R.Y. 1999. Diversity and genetic differentiation in Artemia species and populations detected by AFLP markers. *International journal of salt lake research*, 8(4): 341–350.

Suwa, M., Arai, K. & Suzuki, R. 1994. Suppression of the first cleavage and cytogenetic studies on the gynogenetic loach. *Fisheries Science*, 60(6): 673–681.

Taniguchi, N., Yamasaki, M., Takagi, M. & Tsujimura, A. 1996. Genetic and environmental variances of body size and morphological traits in communally reared clonal lines from gynogenetic diploid ayu, *Plecoglossus altivelis*. *Aquaculture*, 140(4): 333–341.

Tavares, E.S. & Baker, A.J. 2008. *Single mitochondrial gene barcodes reliably identify sister-species in diverse clades of birds. BMC Evolutionary Biology*, 8(1): 81.

Thorgaard, G.H., Jazwin, M.E. & Stier, A.R. 1981. Polyploidy induced by heat shock in rainbow trout. *Transactions of the American Fisheries Society*, 110(4): 546–550.

Tian, D., Sun, S. & Lee, J.T. 2010. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell*, 143(3): 390-403.

Tine, M., Kuhl, H., Gagnaire, P.A., Louro, B., Desmarais, E., Martins, R.S., Hecht, J., *et al.* 2014. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Communications*, 5(1): 1–10.

Tosh, J. 2014. *Genome-wide association analysis of harvest weight in a North American Atlantic salmon population.* Proceedings, 10th World Congress on Genetics Applied to Livestock Production.

Toth, G., Gaspari, Z. & Jurka, J. 2000. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Research*, 10(7): 967–981.

Townsend, J.A., Wright, D.A., Winfrei, R.J., Fu, F., Maeder, M.L., Joung, J.K. & Voytas, D.F. 2009. High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature*, 459(7245): 442–445.

Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A. & Bubulya, P.A. 2010. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Molecular Cell*, 39(6): 925–938.

Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E. & Chang, H.Y. 2010. Long noncoding RNA as modular scaffold of histone modification complexes. *Science*, 329(5992): 689–693.

Tsigenopoulos, C.S., Louro, B., Chatziplis, D., Lagnel, J., Vogiatzi, E., Loukovitis, D., Franch, R., *et al.* 2014. Second generation genetic linkage map for the gilthead sea bream *Sparus aurata* L. *Marine Genomics*, 18(Part A): 77–82.

Ueda, T., Sato, R. & Kobayashi, J. 1988. Triploid rainbow-trout induced by high-ph high-calcium. *Nippon Suisan Gakkaishi*, 54(11): 2045–2045.

Vallejo, R.L., Palti, Y., Liu, S., Evenhuis, J.P., Gao, G., Rexroad, C.E. & Wiens, G.D. 2014a. Detection of QTL in rainbow trout affecting survival when challenged with *Flavobacterium psychrophilum*. *Marine Biotechnology*, 16(3): 349–360.

Vallejo, R.L., Palti, Y., Liu, S., Marancik, D.P. and Wiens, G.D. 2014b. Validation of linked QTL for bacterial cold water disease resistance and spleen size on rainbow trout chromosome Omy19. *Aquaculture*, 432: 139–143.

Van Bers, N.E., Crooijmans, R.P., Groenen, M.A., Dibbits, B.W. & Komen, J. 2012. SNP marker detection and genotyping in tilapia. *Molecular ecology resources*, 12(5): 932–941.

Van Dijk, E.L., Auger, H., Jaszczyszyn, Y. & Thermes, C 2014. Ten years of next-generation sequencing technology. *Trends in Genetics*, 30(9): 418–426.

Van Lith, H.A. & Van Zutphen, L.F. 1996. Characterization of rabbit DNA microsatellites extracted from the EMBL nucleotide sequence database. *Animal genetics*, 27(6): 387–395.

Vasemägi, A., Gross, R., Palm, D., Paaver, T. & Primmer, C.R. 2010. Discovery and application of insertion-deletion (INDEL) polymorphisms for QTL mapping of early life-history traits in Atlantic salmon. *BMC Genomics*, 11(1): 156.

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. *Molecular Ecology*, 17(7): 1636–1647.

Viñas, A., Taboada, X., Vale, L., Robledo, D., Hermida, M., Vera, M. & Martínez, P. 2012. Mapping of DNA sex-specific markers and genes related to sex differentiation in turbot (*Scophthalmus maximus*). *Marine Biotechnology*, 14(5): 655–663.

Waldbieser, G.C., Bosworth, B.G. & Quiniou, S.M. 2010. Production of viable homozygous, doubled haploid channel catfish (*Ictalurus punctatus*). *Marine Biotechnology*, 12(4): 380–385.

Waldbieser, G.C., Bilodeau, A.L. & Nonneman, D.J. 2003. Complete sequence and characterization of the channel catfish mitochondrial genome. *DNA Sequence*, 14(4): 265–277.

Walter, M.A., Spillett, D.J., Thomas, P., Weissenbach, J. & Goodfellow, P.N. 1994. A method for constructing radiation hybrid maps of whole genomes. *Nature Genetics*, 7(1): 22–28.

Wang, C.M., Lo, L.C., Zhu, Z.Y. & Yue, G.H. 2006. A genome scan for quantitative trait loci affecting growth-related traits in an F1 family of Asian seabass (*Lates calcarifer*). *BMC Genomics*, 7(1): 1–13.

Wang, S., Sha, Z., Sonstegard, T.S., Liu, H., Xu, P., Somridhivej, B., Peatman, E., Kucutas, H. & Liu, Z. 2008. Quality assessment parameters for EST-derived SNPs from catfish. *BMC genomics*, 9(1): 450.

Wang, S., Peatman, E., Abernathy, J., Waldbieser, G., Lindquist, E., Richardson, P., Lucas, S., Wang, M., Li, P. & Thimmapuram, J. 2010. Assembly of 500,000 inter-specific catfish expressed sequence tags and large scale gene-associated marker development for whole genome association studies. *Genome Biology*, 11(1): 1–14.

Wang, C.M., Bai, Z.Y., He, X.P., Lin, G., Xia, J.H., Sun, F., Lo, L.C., Feng, F., Zhu, Z.Y. & Yue, G.H. 2011. A high-resolution linkage map for comparative genome analysis and QTL fine mapping in Asian seabass, *Lates calcarifer*. *BMC Genomics*. 12(1): 174.

Wang, L., Fan, C., Liu, Y., Zhang, Y., Liu, S., Sun, D., Deng, H., *et al.* 2014. A genome scan for quantitative trait loci associated with *Vibrio anguillarum* infection resistance in Japanese flounder (*Paralichthys olivaceus*) by bulked segregant analysis. *Marine Biotechnology*, 16(5): 513–521.

Wang, Y., Lu, Y., Zhang, Y., Ning, Z., Li, Y., Zhao, Q., Lu, H., *et al.* 2015. The draft genome of the grass carp (*Ctenopharyngodon idellus*) provides insights into its evolution and vegetarian adaptation. *Nature Genetics*, 47(6): 625–631.

Weber, J.L. & Wong, C. 1993. Mutation of human short tandem repeats. *Human molecular genetics*, 2(8): 1123–1128.

Welsh, J. & McClelland, M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research*, 18(24): 7213–7218.

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. *Molecular ecology*, 12(10): 2817–2833.

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 Research*, 18(22): 6531–6535.

Wilusz, J.E. & Sharp, P.A. 2013. A circuitous route to noncoding RNA. *Science*, 340(6131): 440–441.

Wirgin, I. & Maceda, L. 1991. Development and use of striped bass-specific RFLP probes. *Journal of Fish Biology*, 39: 159–167.

Wittkopp, P.J., Haerum, B.K. & Clark, A.G. 2008. Regulatory changes underlying expression differences within and between *Drosophila* species. *Nature genetics*, 40(3): 346–350.

Wolters, W.R., Libey, G.S. & Chrisman, C.L. 1981. Induction of triploidy in channel catfish. *Transactions of the American Fisheries Society*, 110(2): 310–312.

Wong, E.H.K. & Hanner, R.H. 2008. DNA barcoding detects market substitution in North American food fish. *Food Research International*, 41(8): 828–837.

Wong, L.L., Peatman, E., Lu, J., Kucuktas, H., He, S., Zhou, C., Na-Nakorn, U. & Liu, Z. 2011. DNA barcoding of catfish: species authentication and phylogenetic assessment. *PLOS ONE*, 6(3): e17812.

Wood, D.L., Nones, K., Steptoe, A., Christ, A., Harliwong, I., Newell, F., Bruxner, T.J., Miller, D., Cloonan, N. & Grimmond, S.M. 2015. Recommendations for accurate resolution of gene and isoform allele-specific expression in RNA-Seq data. *PLOS ONE*, 10(5): e0126911.

Wringe, B.F., Devlin, R.H., Ferguson, M.M., Moghadam, H.K., Sakhrani, D. & Danzmann, R.G. 2010. Growth-related quantitative trait loci in domestic and wild rainbow trout (*Oncorhynchus mykiss*). *BMC Genetics*, 11(1): 1–16.

Wu, C., Zhang, D., Kan, M., Lv, Z., Zhu, A., Su, Y., Zhou, D., *et al.* 2014. The draft genome of the large yellow croaker reveals well-developed innate immunity. *Nature Communications*, 5(1), 1–7.

Xia, J.H., Liu, F., Zhu, Z.Y., Fu, J., Feng, J., Li, J., & Yue, G.H. 2010a. A consensus linkage map of the grass carp (*Ctenopharyngodon idella*) based on microsatellites and SNPs. *BMC Genomics*. 11(135).

Xia, J.H., Feng, F., Lin, G., Wang, C.M. & Yue, G.H.. 2010b. A first generation BAC-based physical map of the Asian seabass (*Lates calcarifer*). *PLOS ONE*, 5(8): e11974.

Xia, J.H., Lin, G., He, X., Yunping, B., Liu, P., Liu, F., Sun, F., Tu, R. & Yue, G.H. 2014. Mapping quantitative trait loci for omega-3 fatty acids in Asian seabass. *Marine biotechnology*, 16(1): 1–9.

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. *Animal genetics*, 37(6): 321–326.

Xu, P., Wang, S., Liu, L., Thorsen, J., Kucuktas, H. & Liu, Z. 2007. A BAC-based physical map of the channel catfish genome. *Genomics*, 90(3): 380–388.

Xu, P., Wang, J., Wang, J., Cui, R., Li, Y., Zhao, Z., Ji, P., Zhang, Y., Li, J. & Sun, X. 2011. Generation of the first BAC-based physical map of the common carp genome. *BMC Genomics*, 12(1): 537.

Xu, J., Ji, P., Zhao, Z., Zhang, Y., Feng, J., Wang, J., Li, J., *et al.* 2012. Genome-wide SNP discovery from transcriptome of four common carp strains. *PLOS ONE*, 7(10): e48140.

Xu, J., Zhao, Z., Zhang, X., Zheng, X., Li, J., Jiang, Y., Kuang, Y., *et al.* 2014a. Development and evaluation of the first high- throughput SNP array for common carp (*Cyprinus carpio*). *BMC Genomics*, 15(1): 307.

Xu, P., Zhang, X., Wang, X., Li, J., Liu, G., Kuang, Y., Xu, J., Zheng, X., Ren, L., Wang, G. & Zhang, Y. 2014b. Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. *Nature Genetics*, 46(11): 1212–1219.

Yamamoto, E. 1999. Studies on sex-manipulation and production of cloned populations in hirame, *Paralichthys olivaceus* (Temminck et Schlegel). *Aquaculture*, 173(1): 235–246.

Yamazaki, F. 1983. Sex control and manipulation in fish. *Aquaculture*, 33(1–4): 329–354.

Yan, B., Zhao, L.H., Guo, J.T. & Zhao, J.L. 2012a. miR-429 regulation of osmotic stress transcription factor 1 (OSTF1) in tilapia during osmotic stress. *Biochemical and Biophysical Research Communications*, 426(3): 294–298.

Yan, B., Guo, J.T., Zhao, L.H. & Zhao, J.L. 2012b. microRNA expression signature in skeletal muscle of Nile tilapia. *Aquaculture,* 364: 240–246.

Yáñez, J.M., Naswa, S., López, M.E., Bassini, L., Correa, K., Gilbey, J., Bernatchez, L., *et al.* 2016. Genome-wide single nucleotide polymorphism (SNP) discovery in Atlantic salmon (*Salmo salar*): validation in wild and farmed American and European populations. *Molecular ecology resources*, 16(4), 1002–1001.

Yang, L., Froberg, J.E. & Lee, J.T. 2014. Long noncoding RNAs: fresh perspectives into the RNA world. *Trends in Biochemical Sciences*, 39(1): 35–43.

Yang, W.S., Tsou, P.L., Lee, W.J., Tseng, D.L., Chen, C.L., Peng, C.C., Lee, K.C., *et al.* 2003. Allele-specific differential expression of a common adiponectin gene polymorphism related to obesity. *Journal of Molecular Medicine*, 81(7): 428–434.

Ye, H., Liu, Y., Liu, X., Wang, X. & Wang, Z. 2014. Genetic mapping and QTL analysis of growth traits in the large yellow croaker *Larimichthys crocea*. *Marine Biotechnology*, 16(6): 729–738.

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*). *Molecular Ecology,* 10(4): 921–930.

Yu, S. 2011. Highly efficient modification of beta-lactoglobulin (BLG) gene via zinc-finger nucleases in cattle. *Cell Research*, 21(11): 1638–1640.

Yu, Z. & Guo, X. 2006. Identification and mapping of disease-resistance QTLs in the eastern oyster, *Crassostrea virginica* Gmelin. *Aquaculture*, 254(1–4): 160–170.

Yu, Z. & Ximing, G. 2003. Genetic linkage map of the eastern oyster *Crassostrea virginica* Gmelin. *The Biological Bulletin*, 204(3): 327–338.

Yue, G.H. 2014. Recent advances of genome mapping and marker-assisted selection in aquaculture. *Fish and fisheries*, 15(3): 376–396.

Zhan, A., Hu, J., Hu, X., Hui, M., Wang, M., Peng, W., Huang, X., Wang, S., Lu, W. & Sun, C. 2009. Construction of microsatellite-based linkage maps and identification of size-related quantitative trait loci for Zikhong scallop (*Chlamys farreri*). *Animal Genetics*, 40(6): 821–831.

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. *Journal of Genetics and Genomics*, 35(2): 97–103.

Zhang, X., Zhao, C., Huang, C., Duan, H., Huan, P., Liu, C., Zhang, X., *et al.* 2011a. A BAC-based physical map of Zhikong scallop (*Chlamys farreri* Jones et Preston). *PLOS ONE*, 6(11): e27612.

Zhang, Y., Xu, P., Lu, C., Kuang, Y., Zhang, X., Cao, D., Li, C, *et al.*. 2011b. Genetic linkage mapping and analysis of muscle fiber-related QTLs in common carp (*Cyprinus carpio* L.). *Marine Biotechnology*, 13(3): 376–392.

Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., Yang, P., *et al.* *2012*. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature*, 490(7418): 49–54.

Zhang, X., Zhang, Y., Zheng, X., Kuang, Y., Zhao, Z., Zhao, L., Li, C. *et al.* 2013. A consensus linkage map provides insights on genome character and evolution in common carp (*Cyprinus carpio* L.). *Marine Biotechnology*, 15(3): 275–312.

Zhao, L., Zhang, Y., Ji, P., Zhang, X., Zhao, Z., Hou, G., Huo, L., *et al.* 2013. A dense genetic linkage map for common carp and its integration with a BAC-based physical map. *PLOS ONE*, 8(5): e63928.

Zheng, Q. & Wang, X.J. 2008. GOEAST: a web-based software toolkit for gene ontology enrichment analysis. *Nucleic Acids Research*, 36(suppl_2): W358–W363.

Zhou, S., Bechner, M.C., Place, M., Churas, C.P., Pape, L., Leong, S.A., Runnheim, R., *et al.* 2007. Validation of rice genome sequence by optical mapping. *BMC Genomics*, 8(1): 1.

Zhou, Z., Liu, S., Dong, Y., Gao, S., Chen, Z., Jiang, J., Yang, A., *et al.* 2015. High-density genetic mapping with interspecific hybrids of two sea urchins, *Strongylocentrotus nudus* and *S. intermedius*, by RAD sequencing. PLOS ONE, 10(9): e0138585.

Zhu, Z., He, L. & Chen, S. 1985. Novel gene transfer into the fertilized eggs of gold fish (*Carassius auratus* L. 1758). *Journal of Applied Ichthyology*, 1(1): 31–34.

Zou, Z., Zhang, Z., Wang, Y., Han, K., Fu, M., Lin, P. & Xiwei, J. 2011. EST analysis on the gonad development related organs and microarray screen for differentially expressed genes in mature ovary and testis of *Scylla paramamosain*. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, 6(2): 150–157.

Genome technologies have been developed to study genome structure, organization, expression and function, and to select and modify genomes of interest to increase benefits to humans. Of these genome technologies, DNA marker technologies have been intensely used to map the genome to understand genome structure and organization. These DNA marker technologies include restriction fragment length polymorphism markers; mitochondrial DNA markers; DNA barcoding; random amplified polymorphic DNA markers; amplified fragment length polymorphism markers; microsatellite markers; single nucleotide polymorphism (SNP) markers; and restriction site-associated DNA sequencing markers (SNP markers per se). Although these marker systems have been used at various levels for various purposes, the microsatellite markers and SNP markers are currently the most important. Various genome-mapping technologies have been developed, including both genetic mapping and physical mapping methods. Genetic mapping is based on recombination during meiosis, while physical mapping is based on fingerprints of DNA segments. Although several variations of physical mapping methods are available, such as radiation hybrid mapping and optical mapping, the most popular physical mapping method is the bacterial artificial chromosome based fingerprinting.

ISBN 978-92-5-135464-3
9 789251 354643
CB7955EN/1/12.21