The Pharma Innovation Journal 2021; SP-10(7): 348-351

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2021; SP-10(7): 348-351
© 2021 TPI
www.thepharmajournal.com
Received: 22-04-2021
Accepted: 10-06-2021

Lakshman Sahoo
Fish Genetics and Biotechnology Division, ICAR-CIFA, Kausalyaganga, Bhubaneswar, Odisha, India

Susanta Kumar Dash
Department of Animal Breeding and Genetics, College of Veterinary Science and Animal Husbandry, Odisha University of Agriculture and Technology, Bhubaneswar, Odisha, India

Corresponding Author:
Lakshman Sahoo
Fish Genetics and Biotechnology Division, ICAR-CIFA, Kausalyaganga, Bhubaneswar, Odisha, India

Next generation sequencing technologies and its applications in aquaculture

Lakshman Sahoo and Susanta Kumar Dash

Abstract
Sequencing of the DNA of an organism leads to better understanding of genes and their physiological role. With the advent of NGS technologies it is now feasible to sequence whole genome sequencing and re- sequencing at faster rate in turn can be subsequently used for comparing aquatic genomes for studying functional genome diversity employing various bioinformatics tools. Massive parallel sequencing methods made it practically possible whole genome sequencing and re-sequencing of large number of aquatic genomes at much higher speed and relatively lower cost compared to traditional sequencing methods. The major gap in the genomic approaches to aquaculture improvement is in the application of genomic information for development of improved aquatic strains through next-generation sequencing technologies. The most effective effort to fulfill the gap is to integrate various research disciplines which form core components of molecular aquaculture breeding. The integration of various approaches required knowledge of whole genome organization, strong statistical knowledge to estimate the gene/genetic effects, good experience in molecular biology techniques and traditional breeding methodologies.

Keywords: Aquatic, breeding, genome, sequencing, technology

Introduction
With worldwide stagnation or decline in capture fisheries due to overexploitation and habitat degradation, aquaculture will be central to meeting fish demand, which will continue to increase with burgeoning population growth and increasing urbanization. Aquaculture is one of the fastest growing food producing sectors in the world. It plays an important role in economy providing employment opportunity and rural livelihood. The employment opportunity in aquaculture has grown rapidly over the past few decades, increasing more than threefold from 13 million people in 1970 to 41 million in 2004. Among all protein sources, fish is the cheapest one, with great potential to meet the ever-increasing demand. Fish comprises a nutritionally important part of many people's diet, is a vital source of protein and micronutrient and improves the quality of protein in largely vegetable and starch based diets by providing essential amino acids in developing countries. Sustainable aquaculture production requires genetically improved strains/varieties/stocks.

DNA sequences of any organism comprise the blue print of that organism. Process of DNA sequencing is the determination of the particular sequence of nucleotides. Technique of DNA sequencing pioneered by Sanger revolutionized molecular biology raising numerous fields such as forensic biology, biotechnology and more. Sequencing of the DNA of an organism leads to better understanding of genes and their physiological role. Whole genome sequencing of zebrafish in 2002 enabled positional cloning of many genes affecting embryogenesis, behaviour, physiology, and health and disease further annotated reference genome has enabled the generation of accurate whole-exome enrichment reagents, which are accelerating both positional cloning projects and new genome-wide mutation discovery efforts (Howe *et al*., 2013: Kettleborough *et al*., 2013: Varshney *et al*., 2013) [9, 14, 20]. Surfacing of high throughput sequencing technology known as next generation sequencing (NGS) technology also known as massive parallel or multiplex sequencing, is the key element in identifying many novel genes and understanding the genetics of various qualitative and quantitative traits in aquaculture important fish species. In comparison to Sanger sequencing method NGS methods are more economical, informative, and readily applicable for the ultimate transition from empirical practices to precision genetic improvement programs. Recently several such non Sanger ultra-high throughput sequencing platforms are came into sight and commercialized altogether known as second generation sequencing technologies. Massive parallel sequencing methods made it practically possible whole genome sequencing and re-sequencing of large number of

aquatic genomes at much higher speed and relatively lower cost compared to traditional sequencing methods. At present commercially available NGS platforms includes the followings

1. Roche 454 GS-FLX
2. Solexa Illumina Genome Analyzer
3. Life Technologies ABI SOLiD
4. Ion Torrent
5. Helicos Biosciences Heliscope
6. Pacific Biosciences PacBio

Roche 454 SG-FLX

Illumina Genoe Analyzer

ABI SOLiD

Ion Torrent

Heliscope

PacBio

Applications of NGS technologies

Genetic diversity within a population is maintained by sexual reproduction as it mixes up the parent's genetic material through recombination, leading to offspring with unique genetic blue prints. Population of an organism having high genetic diversity has a greater chance of surviving and flourishing than a population with limited genetic variability. Genetic diversity or variability is the currency upon which genetic improvement program relies. Traditionally information on genetic diversity present in the breeding materials were harnessed through morphological observations and recently through the use of molecular markers such as random amplified polymorphic DNA (RAPD), amplified length fragment length polymorphism (AFLP), restriction fragment length polymorphism (RFLP) and simple sequence repeats (SSRs). Genotypic and environmental interaction leads to the phenotypic variation in a population and the magnitude of phenotypic variability differs under different environmental conditions. However, information on genetic diversity based on morphological observations is prone to biased by the environment and does not reveal the complete variation present in the aquatic species. At the same time molecular markers like microsatellites from the genic or genomic regions represent allelic variation but their functional role is unknown. With the advent of NGS technologies it is now feasible to sequence whole genome sequencing and re-sequencing at faster rate in turn can be subsequently used for comparing aquatic genomes for studying functional genome diversity employing various bioinformatics tools. Development of massive parallel sequencing techniques revolutionized genetics by changing the nature of genetic experimentation leading to genomic selection with the help of genome wide Single nucleotide polymorphism (SNP) markers. In combination with appropriate computational algorithm immensely increases the possibility to answer questions about mutational spectrum of an organism, from single base to copy number polymorphism, on a genome wide scale, is likely to radically alter our understanding of model organisms.

a. Marker discovery

Molecular markers are the heritable polymorphisms that can be detected in individuals of one or more populations are the centre of the modern genetics. In recent times biochemical markers have been replaced by DNA based molecular markers as techniques for DNA analysis in many organisms. Traditionally development of DNA based molecular markers is time consuming and labor intensive. With the introduction of NGS technology methods for both isolation and analysis of molecular markers have been improved in terms of number, time and labor.

1. SSR marker development

Simply sequence repeats (SSRs) or microsatellite markers are the most informative and versatile genetic markers consisting of tandem DNA repeats of 2 to 6 base pairs (Brooker *et al.*, 1994; Weber and May, 1989) [4, 21]. Microsatellites are the markers of choice for various genetic studies owing to their

advantages such as high level of polymorphism, co dominance, even distribution in the genome and easy analysis using PCR (Jerry *et al*., 2006) [11]. Large number of SSR marker development within short span of time is possible with the introduction different NGS platforms in the market e. g.in rohu (Sahu *et al.* 2012) [18], Atlantic cod (Carlsson *et al*., 2013) [5] and dusky kob (Mirimin *et al.*, 2013) [16] etc.

2. SNP discovery for genome studies

SNPs are the result of base variation among individuals at any site of the genome. The benefit of the SNPs in comparison to other markers is that they are profuse throughout the entire genome (~ 3 X 10^7 in human genome), highly polymorphic and amenable to automation. Due to this now a days, SNP markers have gained lot of interest in the scientific and breeding community. Employing NGS platforms and reduced representation libraries (RRL) methods it is possible to isolate large number of SNP markers in the non-model aquatic organism. For example in Atlantic cod just from 2% of the genome more than 25000 SNPs were isolated (Carlsson *et al*., 2013) [5]. The high frequency of SNPs discovered reveals the need of the NGS technologies in fast discovery of SNP's throughout the genome of individuals which can be used for genome wide association studies of economically important traits such as body shape and fin position in Atlantic salmon (Boulding *et al*., 2008) [3], grisling and late sexual maturation in Atlantic salmon (Gutierrez *et al*., 2013) [6], and for growth in Asian seabass (Xia *et al*., 2013) [22].

3. RAD tags

Segregation of DNA sequence polymorphisms by most of the organisms disrupts restriction sites. Restriction site associated DNA (RAD) tags are a genome wide representation of a particular restriction enzyme by short tags. Therefore RAD tags serves as genetic markers spread at high density throughout the genome. In the early days of development RAD tags are used in conjunction with low cost microarray genotyping resources (Miller *et al*, 2007) [15]. With the emergence of massive parallel, next generation sequencing technologies at the same time reduction in sequencing cost results in the integration of the short read sequencing with RAD genotyping (Baird *et al.*, 2008) [2]. RAD sequencing is an efficient protocol in combination with high throughput sequencing platforms can be used to develop large number of SNPs by sequencing a large set of restriction fragments. The identified SNPs could be useful to produce dense linkage maps with the help of high-throughput genotyping technologies. Recently RAD tags for a number of non-model aquatic species have been developed like stickleback (Hohenlohe *et al.* 2010) [8], spotted gar (Amores *et al*., 2011) [1], rainbow trout (Hale *et al*., 2013) [7] and gudgeons (Kakioka *et al*., 2013) [13].

b. QTL mapping and marker assisted selection

Traditionally selection is based on phenotypic characters. With the advent of molecular biology techniques various DNA markers have been isolated in order to support selective breeding programs. Quantitative trait locus (QTL) is a region of the genome that is associated with an effect on a quantitative trait. DNA markers that tightly linked to QTL governing traits of economic importance can be used as a supplement to phenotypic selection. Due to the higher cost involved in sequencing of the genes/genomes, identification of DNA variation at SNP level and subsequent mapping have not been exploited on a larger scale in aquatic species. Recently due to development of high throughput NGS technologies simultaneously reduction in sequencing cost SNPs based markers are expected to replace the SSR markers. In particular, NGS technologies will be useful for mining of SNP variation in parental genotypes for its use in gene mapping. Recently QTL for disease resistance against *A. hydrophila* in *Labeo rohita* (rohu) has been detected through the application of next generation sequencing technologies (Unpublished data).

c. Transcriptome analysis and functional genomics

Knowledge on when, where and to what extend a gene is expressed and under what circumstances its expression is affected will be helpful to recognize gene function. The identification and quantification of mRNA under different conditions have long been used by the scientists for gene expression studies on genome wide scale. Various protocols are followed for the analysis of transcript profiles like differential display, cDNA-AFLP, microarray and serial analysis of gene expression (SAGE). RNA analysis through cDNA sequencing using massive parallel sequencing revolutionized transcriptomics. This development eliminated several challenges posed by the earlier technologies including limited dynamic range of detections. With the advent of high throughput NGS technologies, genome wide gene expression analysis and mapping will be cost effective and facilitates wider use of functional genomics. Transcriptome profiling using NGS platforms also known as RNA-seq have been undertaken in many non-model aquatic organisms like rohu (Robinson *et al*., 2012) [17], rainbow trout (Salem *et al*., 2010) [19], nile tilapia (Huang *et al*., 2011) [10] and common carp (Ji *et al*., 2012) [12].

Future prospective

The future aquaculture genomics research will certainly derive benefit from the recent development of next-generation sequencing technologies. The major gap in the genomic approaches to aquaculture improvement is in the application of genomic information for development of improved aquatic strains. The most effective effort to fulfill the gap is to integrate various research disciplines which form core components of molecular aquaculture breeding. The integration of various approaches required knowledge of whole genome organization, strong statistical knowledge to estimate the gene/genetic effects, good experience in molecular biology techniques and traditional breeding methodologies. These integrated approaches will revolutionize the aquaculture improvement programs in future.

References

1. Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH. Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an out group for the teleost genome duplication. Genetics 2011;188(4):799-808.
2. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, *et al*. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PloS one 2008;3(10):e3376.
3. Boulding EG, Culling M, Glebe B, Berg PR, Lien S, Moen T. Conservation genomics of Atlantic salmon: SNPs associated with QTLs for adaptive traits in parr

from four trans-Atlantic back crosses. Heredity 2008;101(4):381-391.
4. Brooker AL, Cook D, Bentzen P, Wright JM, Doyle RW. Organization of microsatellites differs between mammals and cold-water teleost fishes. Canadian Journal of Fisheries and Aquatic Sciences 1994;51(9):1959-1966.
5. Carlsson J, Gauthier DT, Carlsson JEL, Coughlan JP, Dillane E, Fitzgerald RD, *et al*. Rapid, economical single-nucleotide polymorphism and microsatellite discovery based on de novo assembly of a reduced representation genome in a non-model organism: a case study of Atlantic cod Gadus morhua. Journal of fish biology 2013.
6. Gutierrez AP, Lubieniecki KP, Fukui S, Withler RE, Swift B, Davidson WS. Detection of Quantitative Trait Loci (QTL) Related to Grilsing and Late Sexual Maturation in Atlantic salmon (Salmo salar). Marine Biotechnology 2013, 1-8.
7. Hale MC, Thrower FP, Berntson EA, Miller MR, Nichols KM. Evaluating adaptive divergence between migratory and non-migratory ecotypes of a salmonid fish, Oncorhynchus mykiss, G3: Genes Genomes Genetics 2013;3(8):1273-1285.
8. Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genetics 2010;6(2):e1000862.
9. Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M *et al*. The zebrafish reference genome sequence and its relationship to the human genome. Nature 2013.
10. Huang BF, Sun YL, Wang DS. Transcriptome Analysis of Artificially Induced Sex Reversal in the Nile Tilapia. Indian Journal of Science and Technology 2011;4(S8):78-79.
11. Jerry DR, Evans BS, Kenway M, Wilson K. Development of a microsatellite DNA parentage marker suite for black tiger shrimp (*Penaeus monodon*). Aquaculture 2006;255(1):542-547.
12. Ji P, Liu G, Xu J, Wang X, Li J, Zhao Z, *et al*. Characterization of common carp transcriptome: sequencing, de novo assembly, annotation and comparative genomics. PloS one 2012;7(4):e35152.
13. Kakioka R, Kokita T, Kumada H, Watanabe K, Okuda N. A RAD-based linkage map and comparative genomics in the gudgeons (genus Gnathopogon, Cyprinidae). BMC Genomics 2013;14(1):32.
14. Kettleborough RN, Busch-Nentwich EM, Harvey SA, Dooley CM, de Bruijn E, van Eeden F, *et al*. A systematic genome-wide analysis of zebrafish protein-coding gene function. Nature 2013.
15. Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Research 2007;17(2):240-248.
16. Mirimin L, Guajardo JR, Vervalle J, Bester-Van der Merwe A, Kerwath S, Macey B, *et al*. Isolation and validation of microsatellite markers from a depleted South African sciaenid species, the dusky kob (*Argyrosomus japonicus*), by means of the FIASCO/454 approach. Conservation Genetics Resources 1-4.
17. Robinson N, Sahoo PK, Baranski M, Mahapatra KD, Saha JN, Das S, *et al*. Expressed sequences and polymorphisms in rohu carp (*Labeo rohita*, Hamilton) revealedbym RNA-seq. Marine Biotechnology 2012;14(5):620-633.
18. Sahu BP, Patel A, Sahoo L, Das P, Meher P, Jayasankar P. Rapid and cost effective development of SSR markers using next generation sequencing in Indian major carp, *Labeo rohita* (Hamilton, 1822). Indian Journal of Fisheries 2012;59(3).
19. Salem M, Rexroad C, Wang J, Thorgaard G, Yao J. Characterization of the rainbow trout transcriptome using Sanger and 454-pyrosequencing approaches. BMC genomics Genomics 2010;11(1):564.
20. Varshney GK, Lu J, Gildea DE, Huang H, Pei W, Yang Z, *et al*. A large-scale zebrafish gene knockout resource for the genome-wide study of gene function. Genome Research 2013;23(4):727-735.
21. Weber JL, May PE. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. American journal of human genetics 1989;44(3):388.
22. Xia JH, Lin G, He X, Liu P, Liu F, Sun F, *et al*. Whole genome scanning and association mapping identified a significant association between growth and a SNP in the IFABP-a gene of the Asian seabass. BMC genomics Genomics 2013;14(1):295.