Fisheries Research 186 (2017) 11–22

Contents lists available at ScienceDirect

Fisheries Research

journal homepage: www.elsevier.com/locate/fishres

Review

Applications of next-generation sequencing in fisheries research: A review

Girish Kumar*, Martin Kocour

University of South Bohemia in Ceske Budejovice, Faculty of Fisheries and Protection of Waters, South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, Zátiší 728/II, 389 25 Vodňany, Czech Republic

ARTICLE INFO

Article history:
Received 16 February 2016
Received in revised form 2 July 2016
Accepted 20 July 2016
Handled by Dr. J Viñas

Keywords:
Aquaculture
Evolution
Genome
Molecular marker
Population genetics
Traceability

ABSTRACT

The rapid advancement and decreasing costs of next-generation sequencing (NGS) technologies have revolutionized the field of genomics and beyond. NGS has led to astonishing number of scientific advances and has offered a paradigm shift from gene to genome-wide research in the field of fisheries and aquaculture. Researchers in fish biology have used NGS to discover novel genetic markers for population and traceability studies. Additionally, NGS technologies have been utilized for ecotoxicological applications, genome-wide characterization to find the genomic regions associated with commercially important traits, profiling of messenger-RNAs and micro-RNAs to study the control of biological processes, and to learn more about evolutionary questions. In recent years, great efforts have been made on sequencing the genomes of economically important aquaculture species. These efforts have shown astounding potential to bring enormous change in genetic and biological research in fisheries and aquaculture. Nevertheless, genomic information for most of the fish species is still lacking and results of some studies using NGS were presented without keeping proper sampling and/or experimental design. Future studies should be spread on other non-model species, but with proper sampling respecting aim of such studies. Only then, it is possible to understand well genetic and biological significance of investigated species for fisheries and aquaculture. This review summarizes various applications of next-generation sequencing that has been used in fishery research up to now.

Contents

* Corresponding author.
E-mail address: girishkumar.nio@gmail.com (G. Kumar).

http://dx.doi.org/10.1016/j.fishres.2016.07.021
0165-7836/

1. Introduction

Sequencing is the process of deciphering the precise order of nucleotides in a polymer of nucleic acids (DNA or RNA). Sanger's method of DNA sequencing remained the method of choice for more than two decades and was used for sequencing of complete genomes of numerous species, including the human genome. However, this method suffers from inherent limitations of low speed, high costs and labour intensity. To overcome these limitations, new and improved technologies were introduced in 2005 (Margulies et al., 2005; Shendure et al., 2005). For the new technologies various terms were adapted and used. One of the first terms used was next-generation sequencing (NGS). NGS techniques were further evolved into second-generation sequencing, third-generation sequencing and newly fourth-generation sequencing techniques (Ku and Roukos, 2013). Moreover, other terms as high-throughput sequencing, massively parallel sequencing (technologies that are capable of processing millions of sequence reads in parallel fashion in very short time-frames), clonal sequencing (clonal amplification of target DNA in order to generate sufficient signal for detection during the sequencing run) are often used. Sometimes the terms are used as synonyms, while in the other cases they represent different approaches. However, for sake of simplicity, in this review we use original term – next-generation sequencing (NGS) for all sequencing techniques other than those based on Sanger method. Since NGS introduction, the technologies have revolutionised biological science and driven a massive acceleration in research and development. The development of NGS technologies have enabled the researchers to generate large amount of sequence data simultaneously at relatively low costs compared to the traditional Sanger's sequencing method. This ability of NGS has offered a shift from gene to genome-wide research across scientific disciplines and thus allowed researchers to ask virtually any question of the genome, transcriptome, and epigenome of any organism. The scientific community has used NGS in numerous basic and applied fields of science, such as animal and plant breeding, drug discovery, biotechnology, forensic science, biological systematics and evolutionary biology. The growing accessibility to high-throughput sequencing technologies and the concurrent development of innovative bioinformatics tools to analyze the sequence data have made NGS as an indispensable and universal tool for biological research. This can be evidenced by an explosion in large number of scientific publications using NGS.

In fisheries science NGS technologies have been used to learn more about genome-wide and transcriptome-wide control of biological processes, identification of novel markers for population structure, traceability, phylogenetics, genetic mapping, ecotoxicological applications and investigation of association of loci with traits affected by selection. The use of NGS has enabled the researchers to gain high degree of resolution than using traditional genetic markers and thus unlocking the information never possible before. Despite the promising results obtained using NGS technologies, it has been utilized in fishery research on limited number of fish species. Keeping this in mind, the aim of the review is to discuss the biological insights already gained from NGS technologies for a variety of fish species in order to exploit full potential of this technique in future.

2. Sequencing platforms

This section compares the various platforms of NGS technologies (Table 1). The automated Sanger's method of DNA sequencing is known as 'first generation technology' while newer methods are referred as 'next-generation sequencing'. The NGS technologies are essentially grouped into second (2G), third (3G) and fourth generation (4G) approaches. Second generation technologies are based on sequencing by synthesis (SBS) or sequencing by ligation (SBL), including pyrosequencing and reversible chain termination (RCT). They remove the *in vivo* bacterial cloning stage of the Sanger methodology by using either emulsion PCR (emPCR) or 'bridge PCR' for target amplification. Third generation technologies are differentiated by deploying a single molecule template approach and removing the copy error and bias associated with PCR amplification. They also avoid the cyclic array approach and thereby enable further massive parallelisation. Fourth generation is an *in situ* sequencing method where DNA of individual cells in a histological section is sequenced. It exploits 2G NGS chemistry to read nucleic acid composition directly in fixed cells and tissues (Ke et al., 2013; Lee et al., 2014). Details of the NGS platforms and their advantages and limitations have been reviewed previously (see *e.g.* Metzker, 2010; Niedringhaus et al., 2011; Liu et al., 2012).

3. Applications of NGS

Advent of NGS technologies have allowed the genome-wide and/or transcriptome-wide detection and characterization of genetic markers, microsatellites and single nucleotide polymorphisms (SNPs), in various non-model fish species for which genomic resources are still limited or absent. For example, using NGS techniques polymorphic microsatellite markers have been developed in large number of marine fishes (Reid et al., 2012; Yin et al., 2012; Slattery et al., 2012; Taguchi et al., 2013; Schultz et al., 2013; Fernandez-Silva et al., 2013; Barnes et al., 2014; Bayona-Vasquez et al., 2015; AlMomin et al., 2015; Santos et al., 2015) and freshwater fishes (Carvalho et al., 2011; Wang et al., 2012; Luo et al., 2012; Sahu et al., 2012; Xu et al., 2013; Rodriguez-Zarate et al., 2014; Yu et al., 2014; Sahu et al., 2014; Villanova et al., 2015). Similarly, several studies have successfully used NGS for SNP discovery (Seeb et al., 2011; Shen et al., 2012; Houston et al., 2012; Carlsson et al., 2013; Montes et al., 2013; Houston et al., 2014; Xiao et al., 2015). In recent years, great efforts have been made on sequencing the whole genome of fish species. According to NCBI database (http://www.ncbi.nlm.nih.gov) genome of twelve fish species has been fully sequenced and assembled (Table 2). One of the important aspects of a whole genome sequence is that it might allow identification of genes responsible for superior performance traits. Such genes can be used for selective breeding programs using marker-assisted selection. Furthermore, whole genome sequence also helps to discover genetic variation in form of SNPs which is one of the fundamental reasons why individuals of same species performs differently from one another. The knowledge of SNPs also allows researchers to address the challenges concerning conservation of wild stocks and sustainability of aquaculture operations.

3.1. Genome-wide association studies, linkage maps, and QTLs

Genome-wide association studies (GWAS) are a relatively new approach in fisheries science to identify the genes associated with performance and production traits. This approach involves scanning of markers (typically SNPs) across the genomes to find the genomic variation associated with commercially important traits such as growth and disease resistance. Despite the promising results obtained in terrestrial livestock and human medical research, relatively few GWAS have been undertaken in aquaculture species (Tsai et al., 2015a). Nevertheless, decreasing costs of NGS and associated technological genotyping advancement has led to GWAS in many aquaculture species. For example, four quantitative trait loci (QTL) associated with columnaris disease resistance in catfish (Geng et al., 2015) and nine QTL in Japanese flounder (*Paralichthys olivaceus*) genome for *Vibrio anguillarum* disease