Aquaculture 468 (2017) 337–347

Contents lists available at ScienceDirect

Aquaculture

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

Review

Current status of genome sequencing and its applications in aquaculture

G.H. Yue *, L. Wang

Molecular Population Genetics and Breeding Group, Temasek Life Sciences Laboratory, 1 Research Link, National University of Singapore, 117604, Singapore

ARTICLE INFO

Article history:
Received 1 August 2016
Received in revised form 22 October 2016
Accepted 25 October 2016
Available online 26 October 2016

Keywords:
Fish
Genome
GWAS
NGS
SNP

ABSTRACT

Aquaculture is the fastest-growing food production sector in agriculture, with great potential to meet projected protein needs of human beings. Aquaculture is facing several challenges, including lack of a sufficient number of genetically improved species, lack of species-specific feeds, high mortality due to diseases and pollution of ecosystems. The rapid development of sequencing technologies has revolutionized biological sciences, and supplied necessary tools to tackle these challenges in aquaculture and thus ensure its sustainability and profitability. So far, draft genomes have been published in over 24 aquaculture species, and used to address important issues related to aquaculture. We briefly review the advances of next generation sequencing technologies, and summarize the status of whole genome sequencing and its general applications (i.e. establishing reference genomes and discovering DNA markers) and specific applications in tackling some important issues (e.g. breeding, diseases, sex determination & maturation) related to aquaculture. For sequencing a new genome, we recommend the use of 100–200× short reads using Illumina and 50–60× long reads with PacBio sequencing technologies. For identification of a large number of SNPs, resequencing pooled DNA samples from different populations is the most cost-effective way. We also discuss the challenges and future directions of whole genome sequencing in aquaculture.

Contents

1. Introduction

There are over 30,000 teleost fish species on the earth, representing the largest group of vertebrates (Nelson, 2006), and over 1 million shellfish species. Among them, around 400 species have been cultured (Gjedrem and Robinson, 2014). Fish aquaculture is one of the most sustainable sources of protein for humans. In the last few decades, aquaculture is the fastest- growing sector in agriculture. However, aquaculture of fishes is facing several serious challenges. Most cultured species are still in their wild state, and have not been genetically improved through breeding. Although there are several species-specific feeds, for most cultured species, there is no species-specific feed. There are many fish diseases caused by pathogens, including viruses, bacteria and parasites, leading to huge economic losses in aquaculture (Meyer and Schnick, 1989). The economic losses in aquaculture due to diseases world-wide was over several billion dollars/year. Aquatic toxicants have caused serious problems in aquaculture (Dunier and Siwicki,

---

* Corresponding author.
E-mail address: genhua@tll.org.sg (G.H. Yue).

http://dx.doi.org/10.1016/j.aquaculture.2016.10.036
0044-8486/© 2016 Published by Elsevier B.V.

1993), and in turn, aquaculture resulted in environmental problems (Cao et al., 2007). To make aquaculture sustainable and profitable, it is essential to tackle these challenges with novel ideas and technologies.

Since 2005, the advances in next-generation sequencing (NGS) technologies have been substantial (Metzker, 2010). Outputs of next generation sequencing platforms have increased from less than one Gb/run in 2007 to 1500 Gb/run in 2016, while cost per base has been reduced by >1000 folds. NGS technologies have revolutionized biological science by enabling fast and efficient discovery of genetic variation within and among individuals. NGS is increasingly applied to the field of fish biology (Liu, 2011). In recent years, NGS technologies have been applied to aquaculture fish species to tackle challenges in aquaculture. Novel genomic tools and resources are beginning to help identify genetic markers associated with traits of commercial interest (Yue, 2014), to unravel the molecular basis of different physiological processes (Xia et al., 2011; Yasuike et al., 2016), to study the eukaryotic and/or prokaryotic biodiversity of environmental samples (Hajibabaei et al., 2011), to characterize the innate immune system (Ao et al., 2015), to identify antimicrobial peptides and beneficial bacteria as probiotics (Ghanbari et al., 2015), to define inter-relationships between genomes and genes and to increase our understanding of how dietary constituents influence metabolism (Bonacic et al., 2016). However, in comparison to the applications of NGS technologies in model organisms and livestock, those in aquaculture lag far behind mainly because of the diversity of fish species that are cultivated, and fewer funding opportunities for aquaculture species.

In this paper, we review the recent advances of whole genome sequencing, and its applications in important issues of aquaculture. We also briefly discuss the challenges and future directions of application of whole genome sequencing in aquaculture.

2. Advances of next-generation sequencing (NGS) technologies

After Sanger invented the chain-terminating dideoxynucleotide analogues (Sanger et al., 1977) and Maxam and Gilbert (1977) developed chemical degradation methodologies, DNA sequencing was rapidly adopted in genetic studies (Barnhart, 1989). The coming of the next-generation sequencing (NGS) technologies since 2005 has brought revolutionary benefits to biological studies by reducing cost per base and increasing yield by several orders of magnitude (Metzker, 2010). Technological advances and intensified competition will definitely push more accurate, lower-cost, faster and more user friendly sequencing into application (Goodwin et al., 2016). Regarding NGS technology development, readers may refer to a very good current review (Goodwin et al., 2016). As NGS costs decrease and bioinformatics analysis pipelines improve, the utility of whole genome sequencing is ever expanding in aquaculture.

3. Current status of whole genome sequencing in aquaculture

Aquaculture genome sequencing projects started in the early 2000s in the USA, some EU countries, China and other counties. For sequencing the complete genomes of aquaculture species, different strategies were used. For example, for sequencing the pacific oyster (*Crassostrea gigas*, *Ostreidae*), a cost-effective fosmid-pooling strategy was used (Zhang et al., 2012). For common carp (*Cyprinus carpio*, *Cyprinidae*), a whole-genome shotgun strategy and combining data from several NGS platforms were applied (Xu et al., 2014). Recently, people have used more and more short-read sequencing (Ao et al., 2015) in combination with very long sequencing (e.g. SMRT sequencing) (Roberts et al., 2013; Vij et al., 2016). Currently, the draft genome sequences of over 24 aquaculture species (e.g. Pacific oyster, common carp, grass carp, European sea bass) have been published (Table 1). However, in shrimp species, no genome sequence has been reported, although efforts are being made to sequence shrimp species, such as pacific white shrimp *Litopenaeus vannamei* (Yu et al., 2015) and marine shrimp *Penaeus monodon* (Huang et al., 2011). The reason for this may be due to high contents of repeat sequences in the genomes of shrimp species (Huang et al., 2011). Among these 24 sequenced species, Atlantic cod, a large, cold-adapted teleost, is an important fishery species (Star et al., 2011).The Pacific oyster, belonging to the phylum *Mollusca*, is an important shellfish species for food (Zhang et al., 2012). The common carp is one of the most important freshwater food-fish species and accounts for 10% of the world's freshwater aquaculture production (Xu et al., 2014). Nile tilapia (*Oreochromis niloticus*, Cichlidae), a cichlid species, is a good representative species to understand the molecular mechanisms underlying cichlid phenotypic diversity (Brawand et al., 2014). Here we briefly summarize the genome sequences and the major findings of genome sequencing in these four species.

Atlantic cod is a cold-water marine fish species, which sustains long-standing commercial fisheries and aquaculture. Its genome was sequenced by exclusively using 454 shotgun sequencing and paired-end libraries (Star et al., 2011). The assembled genome was 629–750 Mb in size, containing over 13,000 contigs and over 3800 scaffolds. The genome annotation identified 22,154 protein-coding genes. One major finding of the cod genome sequencing is that Atlantic cod has lost the genes for MHCII, CD4 and invariant chain (Ii), which are essential in immune responses. In addition, a highly expanded number of MHCI genes and a unique composition of Toll-like receptor (TLR) families in the cod genome were found. The data of the cod genome indicate how the immune system of Atlantic cod has evolved compensatory mechanisms in both adaptive and innate immunity in the absence of MHCII. This finding challenges the fundamental assumptions about the evolution of the adaptive immune system. However, this result needs to be confirmed by sequencing more individuals in the species or closely related species. Researchers at the Centre for Ecological and Evolutionary Synthesis (CEES), University of Oslo, used long sequence reads generated with the PacBio sequencer to produce a substantially improved reference genome for the Atlantic cod. The improved genome assembly release 86 is now available in http://www.ensembl.org/Gadus_morhua/Info/Index.

The Pacific oyster is one of the most important marine shellfishes for human consumption. It has been cultured in many countries around the world, with an annual production of 0.55 million tons. The oyster genome was sequenced by using short reads and a fosmid-pooling strategy (Zhang et al., 2012). Due to a high content of repetitive DNA sequences in its genome, the assembly of the genome was extremely complicated in 2012. The final assembly comprised 559 Mb, with a contig N50 size of 19.4 Kb and a scaffold N50 size of 401 kb. Over 90% of the assembly was covered by the longest 1670 scaffolds. The oyster genome contains a high percentage of repetitive sequences and is highly polymorphic. Some transposable elements still actively shape variation. The oyster genome assembly is available at http://www.oysterdb.com/FrontHomeAction.do?method=home. Additional efforts are being made to improve the assembly.

The common carp is one of the most important freshwater food fish species in the world. Its genome was sequenced by combining data from several next-generation sequencing platforms (i.e. Roche 454 and Illumina sequencing) (Xu et al., 2014). The draft genome of domesticated *C. carpio* is 1.69 Gb and contains 52,610 protein-coding genes. It covers about 92.3% of its genome. 2503 large scaffolds contributed to 90% of the assembly. The GC content of the common carp genome is 37%. Re-sequencing of the genomes of 33 individuals from populations from different regions in the world found a single origin for common carp in the *C. carpio haematopterus* and *C. carpio carpio* subspecies. The draft genome supplies a useful resource for better molecular studies and improved genome-assisted breeding of *C. carpio* and other closely related species. More detailed info about the common carp genome and its downstream applications can be found in http://www.carpbase.org/. Additional efforts are being made to improve the reference genome.