Chronicle of Aquatic Science

TYPE: Popular Article
PUBLISHED: 17 June 2023
ISSUE: Volume 1, Issue 1
ARTICLE ID: CoAS_V1IS1_08

OPEN ACCESS

EDITED BY
Ankures Bhattacharya
West Bengal University of Animal and Fishery Sciences

REVIEWED BY
Dr. Surya Kanta Sau
West Bengal University of Animal and Fishery Sciences

Dr. Asik Ikbal
Department of Fisheries
Govt. of West Bengal

*CORRESPONDENCE
Gowhar Iqbal
drgowhariqbal@gmail.com

RECEIVED 5 May 2023
ACCEPTED 13 June 2023
PUBLISHED 17 June 2023

CITATION
Iqbal et al., (2023) Status of Whole Genome Sequencing in Fisheries and Aquaculture. Chronicle of Aquatic Science 1(1): 40-44

Status of Whole Genome Sequencing in Fisheries and Aquaculture

Gowhar Iqbal[1], Showkat Ahmad Dar[2] and Yaqoota Maqbool[3]

[1]Fish Genetics & Biotechnology Division- ICAR- Central Institute of Fisheries Education – Mumbai 400061

[2]College of Fisheries Kishanganj Bihar Animal Science University, Patna

[3]College of Fisheries Lembucherra, Central agricultural university (Imphal)

Whole genome sequencing (WGS), is the process of determining the entirety or nearly the entirety of the DNA sequence of an organism's genome at a single time in a short span of time. Whole genome sequencing of a given species is a fundamental tool to address some important issues in aquaculture, which are critical for sustainable and profitable aquaculture. The development of whole genome sequencing in aquaculture is very quick. However, no good correlation exists between an organism's body size or complexity and its genome size. In the case of *Labeo rohita* the DNA was sequenced using Illumina and Oxford Nanopore Technologies (ONT) platforms. The draft whole-genome sequences of over 24 species have been published, and they can be served as reference genomes for downstream studies on important issues related to aquaculture and fisheries. NGS has various applications in aquaculture. With the rapid advance of NGS technologies and bioinformatics analysis of large sequencing data, the development of whole genome sequencing in aquaculture is very quick.

Keywords
Whole genome, Aquaculture, NGS, Genome-Wide Association Study, Human Genome Project

Introduction

Whole genome sequencing (WGS), also known as full genome sequencing, complete genome sequencing, or entire genome sequencing, is the process of determining the entirety or nearly the entirety of the DNA sequence of an organism's genome at a single time (Gilissen, 2014). This entails sequencing all of an organism's chromosomal DNA as well as DNA contained in the mitochondria and, for plants, in the chloroplast. Whole genome sequencing has largely been used as a research tool but was introduced to clinics in 2014 (Van *et al.*, 2014). In the future of personalized medicine, whole genome sequence data may be an important tool to guide therapeutic intervention. The tool of gene sequencing at the SNP level is also used to pinpoint functional variants from association studies and improve the knowledge available to researchers interested in evolutionary biology, and hence may lay the foundation for predicting disease susceptibility and drug response (Mooney, 2014). The first virus to have its complete genome sequenced was the Bacteriophage MS2 in 1976. The first multicellular eukaryote, and animal, to have its whole genome sequenced was the nematode worm in 1998. Whole genome sequencing of a given species is a fundamental tool to address some important issues in aquaculture, which are critical for sustainable and profitable aquaculture. Genome size is the total amount of DNA contained within one copy of a single complete genome. It is typically measured in terms of mass in picograms (trillionths (10−12) of a gram, abbreviated pg) or less frequently in daltons, or as the total number of nucleotide base pairs, usually in megabases (millions of base pairs, abbreviated Mb or Mbp). Eukaryotic genomes are sequenced by several methods including Shotgun sequencing of short DNA fragments and sequencing of larger DNA clones from DNA libraries such as bacterial artificial chromosomes (BACs) and yeast artificial chromosomes (YACs). In 1999, the entire DNA sequence of human chromosome 22, the shortest human autosome, was published. By the year 2000, the second animal and second invertebrate (yet first insect) genome were sequenced – that of the fruit fly *Drosophila melanogaster* – a popular choice of a model organism in experimental research. The first plant genome – that of the model organism *Arabidopsis thaliana* – was also fully sequenced by 2000. By 2001, a draft of the entire human genome sequence was published. In 2004, the Human GenomeProject published an incomplete version of the human genome. In 2008, a group from Leiden, the Netherlands, reported the sequencing of the first female human genome (Yano *et al.*, 2016).

Relation between genome size and complexity of an organism

- There is no good correlation between the body size or complexity of an organism and the size of its genome. In prokaryotes (Archaea and Bacteria) there is, in general, a linear relationship between genome size and the number of genes.
- The smallest genomes are found in symbionts and parasites, as they undergo a gene degradation process during adaptation to their new lifestyle.
- However, in eukaryotes there is no correlation between genome size and the complexity of the organism. This is known as the C-value paradox.
- The largest genome is found in an amoeba, a one-cell organism, with 686,000 Mb, 200-fold larger than the human genome and 20,000-fold larger than the one found in yeast. Now we know that most excess DNA is repetitive DNA, apparently lacking a function (selfish DNA) and whose possible role in genome evolution is still unknown.

Cells used for sequencing

Almost any biological sample containing a full copy of the DNA—even a very small amount of DNA or ancient DNA-can provide the genetic material necessary for full genome sequencing. Such samples may include saliva, epithelial cells, bone marrow, hair (as long as the hair contains a hair follicle), seeds, plant leaves, or anything else that has DNA-containing cells. The genome sequence of a single cell selected from a mixed population of cells can be determined using techniques of single-cell genome sequencing.

Early techniques

- Sequencing of nearly an entire human genome was first accomplished in 2000 partly through the use of shotgun sequencing technology. While full genome shotgun sequencing for small (4000–7000 base pair) genomes was already in use in 1979, broader applications benefited from pairwise end sequencing, known colloquially as double-barrel shotgun sequencing.

- The first published description of the use of paired ends was in 1990 as part of the sequencing of the human HPRT locus, although the use of paired ends was limited to closing gaps after the application of a traditional shotgun sequencing approach.

- In 1995 the innovation of using fragments of varying sizes was introduced and demonstrated that a pure pairwise end-sequencing strategy would be possible on large targets.

- The strategy was subsequently adopted by The Institute for Genomic Research (TIGR) to sequence the entire genome of the bacterium *Haemophilus influenza* in 1995, and then by Celera Genomics to sequence the entire fruit fly genome in 2000, and subsequently the entire human genome. Applied Biosystems, now called Life Technologies, manufactured the automated capillary sequencers utilized by both Celera Genomics and The Human Genome Project.

Current techniques

While capillary sequencing was the first approach to successfully sequence a nearly full human genome, it is still too expensive and takes too long for commercial purposes. Since 2005 capillary sequencing has been progressively displaced by high-throughput (formerly "next-generation") sequencing technologies such as Illumina dye sequencing, pyrosequencing, and SMRT sequencing. All of these technologies continue to employ the basic shotgun strategy, namely, parallelization and template generation via genome fragmentation. Other technologies have emerged, including Nanopore technology. Though the sequencing accuracy of Nanopore technology is lower than those above, its read length is on average much longer. This generation of long reads is valuable, especially in de novo whole-genome sequencing applications.

Applications

Mutation frequencies: Whole genome sequencing has established the mutation frequency for whole human genomes. The mutation frequency in the whole genome between generations for humans (parent to child) is about 70 new mutations per generation. In cancer, mutation frequencies are much higher, due to genome instability.

Genome-wide association studies: In research, whole-genome sequencing can be used in a Genome-Wide Association Study (GWAS) – a project aiming to determine the genetic variant or variants associated with a disease or some other phenotype.

Diagnostic use:

- In 2009, Illumina released its first whole genome sequencers that were approved for clinical as opposed to research-only use and doctors at academic medical centers began quietly using them to try to diagnose what was wrong with people whom standard approaches had failed to help.

- Due to recent cost reductions whole genome sequencing has become a realistic application in DNA diagnostics.

Oncology:

- In this field, whole genome sequencing represents a great set of improvements and challenges to be faced by the scientific community, as it makes it possible to analyze, quantify and characterize circulating tumor DNA (ctDNA) in the bloodstream.

- This serves as a basis for early cancer diagnosis, treatment selection, and relapse monitoring, as well as for determining the mechanisms of resistance, metastasis, and phylogenetic patterns in the evolution of cancer.

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 (Yue *et al.*, 2012).

- For common carp (*Cyprinus carpio*, Cyprinidae), a whole-genome shotgun strategy and combining data from several NGS platforms were used.

- *Labeo rohita* (rohu) is one of several carp species that account for much of the protein in the diets of Bangladeshis. To provide insight into genes that can be used to increase the efficiency of the aquaculture of rohu, they sequenced the whole genome of *Labeo rohita*. Briefly, blood from a single male rohu was used to isolate DNA. The DNA was sequenced using Illumina and Oxford Nanopore Technologies (ONT) platforms.

Sequencing projects

- The Atlantic Salmon Genome Sequencing Project is funded by Canada, Norway, and Chile. A large team of scientists referred to as the International Cooperation to Sequence the Atlantic Salmon Genome (ICSASG) is leading the project.

- The Atlantic cod genome, which is about one-third the size of the Atlantic salmon genome, has been sequenced by a Norwegian consortium. The cod genome was sequenced with whole genome shotgun sequencing and paired-end sequencing using the second-generation technology of the Roche 454 FLX Titanium platform.

- The genome for Pacific oysters (*Crassostrea gigas*) is being sequenced by a collaboration among the Institute of Oceanology of the Chinese Academy of Sciences, Beijing Genomics Institute, and the international Oyster Genome Consortium involving many scientists in the United States.

- The oyster genome is being sequenced using the second-generation sequencing technology of the Illumina sequencing platform.

- The tilapia genome sequencing project was funded by the National Institutes of Health, and the work is being conducted by the Broad Institute at the Massachusetts Institute of Technology and Harvard University.

- The tilapia genome sequencing project was funded by the National Institutes of Health, and the work is being conducted by the Broad Institute at the Massachusetts Institute of Technology and Harvard University.

- Sequencing of the shrimp genome is ongoing. The International Shrimp Genome Sequencing Consortium is doing research on this.

Future work

With the rapid advance of NGS technologies and bioinformatics analysis of large sequencing data, the development of whole genome sequencing in aquaculture is very quick. The draft whole-genome sequences of over 24 species have been published and served as reference genomes for downstream studies on important issues related to aquaculture. The genome sequencings have identified a large number of SNPs and microsatellites.

Conclusion

The rapid advancement of NGS technologies and bioinformatics processing of massive sequencing data has accelerated the development of whole genome sequencing in aquaculture. Over 24 species' draught whole-genome sequences have been released, and they have used as reference genomes for subsequent investigations on significant aquaculture challenges. The genome sequencings revealed a vast number of SNPs and microsatellites that have been and are being used in a variety of research, including linkage mapping, QTL mapping, and GWAS for economically important features.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

Conflict of interest

The authors declare that the manuscript was formulated in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by

its manufacturer, is not guaranteed or endorsed by the publisher.

References

Yano, K; Yamamoto, E; Aya, K; Takeuchi, H; Lo, PC; Hu, L; Yamasaki, M; Yoshida, S; Kitano, H; Hirano, K; Matsuoka, M (2016). "Genome-wide association study using whole- genome sequencing rapidly identifies new genes influencing agronomic traits in rice".Nature Genetics. 48 (8)

Gilissen (2014). "Genome sequencing identifies major causes of severe intellectual disability". Nature. 511 (7509): 344–7.

van El, CG; Cornel, MC; Borry, P; Hastings, RJ; Fellmann, F; Hodgson, SV; Howard, HC; Cambon-Thomsen, A; Knoppers, BM; Meijers-Heijboer, H; Scheffer, H; Tranebjaerg, L; Dondorp, W; de Wert, GM (2013). "Whole-genome sequencing in health care. Recommendations of the European Society of Human Genetics". European Journal of Human Genetics.

Mooney, Sean (2014). "Progress towards the integration of pharmacogenomics in practice". Human Genetics. 134 (5)

Yue, G.H. and Wang, L., (2017). Current status of genome sequencing and its applications in aquaculture. Aquaculture, 468,.337-347.