

DATA NOTE

Open Access



Genome of *Tenualosa ilisha* from the river Padma, Bangladesh

Avizit Das¹, Peter Ianakiev², Abdul Baten^{3,4}, Rifath Nehleen¹, Tasneem Ehsan¹, Oly Ahmed¹, Mohammad Riazul Islam¹, M. Niamul Naser⁵, Mong Sano Marma⁶ and Haseena Khan^{1*}

Abstract

Objective: Hilsa shad (*Tenualosa ilisha*), is a popular fish of Bangladesh belonging to the Clupeidae family. An anadromous species, like the salmon and many other migratory fish, it is a unique species that lives in the sea and travels to freshwater rivers for spawning. During its entire life, *Tenualosa ilisha* migrates both from sea to freshwater and vice versa.

Data description: The genome of *Tenualosa ilisha* collected from the river Padma of Rajshahi, Bangladesh has been sequenced and its de novo hybrid assembly and structural annotations are being reported here. Illumina and PacBio sequencing platforms were used for high depth sequencing and the draft genome assembly was found to be 816 MB with N50 size of 188 kb. MAKER gene annotation tool predicted 31,254 gene models. Benchmarking Universal Single-Copy Orthologs refer 95% completeness of the assembled genome.

Keywords: Hilsa shad, *Tenualosa ilisha*, Clupeidae, Whole genome sequence, NGS platform

Objective

Hilsa shad known as ilish in Bangladesh is popular for its taste and the texture of its flesh. This species of fish belongs to the shad in Clupeidae family. In addition to the Bay of Bengal and riverine Bangladesh (the Padma, Jamuna, Meghna, and other coastal rivers of Bangladesh), this fish is also found in the Persian Gulf, Mediterranean Sea, Arabian Sea and China Sea [1]. Fisheries, a part of the Bangladesh's cultural heritage, have played an important role on its socioeconomic development in terms of protein supply, generation of employment and earning of foreign currency. According to the FAO, in 2018 Bangladesh ranked 3rd in the world in inland fish production. Hilsa (*Tenualosa ilisha*), is the most popular among the 650 or so marine and inland fish found in Bangladesh. It contributes to 11% of total fish production and 1% to the national GDP, 3.00% of the total export earnings and about 2.5 million people in Bangladesh are directly dependent on Hilsa in providing for their

families [2, 3]. At present more than 60% of global Hilsa catch is reported from Bangladesh, 20–25% from Myanmar, 15–20% from India and 5–10% from other countries (e.g., Iraq, Kuwait, Malaysia, Thailand and Pakistan). The recent Hilsa production of Bangladesh is about half a million metric ton [4]. In spite of such importance Hilsa is still lacks molecular genomic information. Significance of this data for the improvement in sustainability and maintenance of diversity of this fish cannot therefore be overemphasized.

Data description

Fresh *Tenualosa ilisha* samples from the river Padma at Rajshahi were collected and instantly preserved on dry ice. White and red muscles of the fish were used for DNA extraction. A modified SDS (Sodium Dodecyl Sulfate) method [5], optimized in our lab was used for DNA extraction (detailed methodology in Data File 1, Table 1).

Pair end library with an insert size of around 300 bp was constructed for Illumina sequencing using NEB Neb-Next Ultra II DNA kit (detailed methodology in Data File 1, Table 1) Genomic DNA was sequenced by Illumina HiSeq 4000 and Pacific Bioscience Sequel, single molecule, real time (SMRT, Single Molecule Real Time)

*Correspondence: haseena@du.ac.bd

¹ Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka 1000, Bangladesh

Full list of author information is available at the end of the article



Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	DNA isolation and library preparation methodology	.docs file	https://figshare.com/s/467b8b670149f1a0617c
Data file 2	Whole genome assembly data	FASTA	NCBI GeneBank (Accession numbers: GCA_003651195.1) (http://identifiers.org/ncbi/insdc:gca:gca_003651195.1)
Data file 3	Whole genome sequence	FASTA	NCBI GeneBank (Accession numbers: QYSC01000001–QYSC01124209) (http://identifiers.org/ncbi/insdc:QYSC01000001)
Data file 4	Annotation data file	.tsv	https://figshare.com/s/270b54d9d076ef5e5901

sequencing platforms. The quality of the reads were checked using FastQC [6]. MaSuRCA (Maryland Super-Read Celera Assembler) ver 3.2.6 was used for hybrid de novo assembly [7] using both the Illumina and PacBio data. The genome assembly data has been deposited in the NCBI GeneBank under the Accession numbers GCA_003651195.1 (Data file 2; Table 1). Illumina only data generated a fragmented assembly and showed 91% BUSCO [8] completeness. Addition of 15.7 Gbp data from PacBio significantly improved the quality and contiguity of the genome. Compared to Illumina only, N50 improved from 13 Kb (kilo base pair) to 188 Kb. Similarly, the total number of scaffolds reduced from 475,121 to 124,209. The assembled genome size of *Tenuulosa ilisha* Padma Bangladesh is now 816 Mb (Mega base pair) and approximately 82% of the genome has been assembled. The BUSCO analysis revealing 95% completeness as well as significantly lower number of scaffolds and considerably better N50 indicates the genome to be of high-quality. The genome sequence data has been deposited in the NCBI GeneBank under the Accession numbers QYSC01000001–QYSC01124209 (Data file 3; Table 1). MAKER ver 3.0 pipeline [9] was used for structural annotation. GC content of the genome was determined to be 43.61%. RepeatMasker and Repeatmodeler using the latest version of repbase database [10–12] identified 27.27% repeat elements. Altogether, 31,254 gene models were predicted using the MAKER gene annotation pipeline based on both de novo and reference based predictions using genes/proteins from other fish species (Atlantic herring, carp, salmon, zebrafish). Out of the 31,254 genes, 24,648 were annotated using InterProScan [13] and 16,078 genes were found to have at least 1 GO (Gene Ontology) term assigned to them (Data file 4, Table 1). The Hilsa genome was found to be comparable to the Atlantic herring (807 Mb genome and 28,335 genes) [14] and to the genome of the common carp (1.8 Gb and 52,000 genes) [15].

Limitations

The number of the regions unassembled in the genome is 4605 and the total number of bases positioned in this gap is 2,268,925 bp.

Abbreviations

BUSCO: Benchmarking Universal Single-Copy Orthologs; PacBio: Pacific Bioscience; Gbp: giga base pair; Mb: mega base pair; Kb: kilo base pair; bp: base pair; GO: gene ontology; SDS: sodium dodecyl sulfate; EDTA: ethylenediaminetetraacetic acid; qPCR: quantitative polymerase chain reaction; SMRT: single molecule real time sequencing; MaSuRCA: Maryland Super-Read Celera Assembler; EST: expressed sequenced tag; SNAP: Semi-HMM-based Nucleic Acid Parser.

Authors' contributions

HK and MSM initiated the project. HK, MSM, MRI, PI, MNN and AD designed the overall project. HK and MRI led the project. AD and OA collected the samples with the help of MNN. AD and OA extracted the DNA. PI sequenced the *Tenuulosa ilisha* Padma BD genome. AB assembled the genome and performed the structural and functional annotations. TE and RN performed the repeat and GC content analysis. HK, MRI and AD wrote the manuscript. HK, MRI, MSM, PI, AB, MNN, AD, OA, TE, RN reviewed the manuscript. All authors read and approved the final manuscript.

Author details

¹ Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka 1000, Bangladesh. ² Hera Biosciences LLC, Medford, MA, USA. ³ AgResearch, Grasslands Research Centre, Palmerston North 4442, New Zealand. ⁴ Southern Cross Plant Science, Southern Cross University, Lismore, NSW, Australia. ⁵ Department of Zoology, University of Dhaka, Dhaka 1000, Bangladesh. ⁶ Qiagen Sciences, Waltham, MA, USA.

Acknowledgements

Authors concede the support of Hera Biosciences for the sequencing service and that of Southern Cross University, Lismore, Australia for the computational support.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Data availability

The genome sequence data has been available at DDBJ/ENA/GenBank under the Accession numbers QYSC01000001–QYSC01124209 and the assembled genome at GCA_003651195.1. The version described in this paper is the first version, QYSC00000000.1.

Ethics approval and consent to participate

The experiments mentioned in this study have been approved by the institutional review committee of University of Dhaka.

Funding

This study did not receive any formal funding.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 16 October 2018 Accepted: 19 December 2018

Published online: 22 December 2018

References

- Naser MN. Hilsa Shad (*Tenualosa ilisha*)—the iconic fish of the Bengal Delta. In: Nishat B, Mandal S, Pangare G, editors. Conserving Ilish, securing livelihoods: Bangladesh—India initiatives. India: IWA, Academic Foundation; 2018. p. 37–52.
- Sarker JM, Uddin AMMB, Patwary SAM, Tanmay MH, Rahman F, et al. Livelihood status of Hilsa (*Tenualosa ilisha*) fishermen of greater Noakhali regions of Bangladesh. Fish Aquac J. 2016;7:168. <https://doi.org/10.4172/2150-3508.1000168>.
- FRSS. Yearbook of fisheries statistics of Bangladesh. Dhaka: Dept of Fisheries, GoB; 2017. p. 116.
- Ahsan D, Naser N, Bhoumik U, Hazra S, Bhattacharya SB. Migration, spawning patterns and conservation of Hilsa shad (*Tenualosa ilisha*) in Bangladesh and India. 2016; Academic Foundation.
- Kumar R, Singh PJ, Nagpure NS, Kushwaha B, Srivastava SK, Lakra WS. A non-invasive technique for rapid extraction of DNA from fish scales. Indian J Exp Biol. 2007;45(11):992–7.
- FastQC program. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Accessed 12 Jan 2017.
- Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. The MaSuRCA genome assembler. Bioinformatics. 2013;29(21):2669–77.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–2.
- Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, Holt C, Alvarado AS, Yandell M. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 2008;18(1):188–96.
- Smit A, Hubley R, Green P. RepeatMasker open-4.0. 2013–2015. Seattle, WA, USA: Institute for Systems Biology; 2015. <http://www.repeatmasker.org/faq.html>.
- Smit A, Hubley R. RepeatModeler open 1.0. Seattle, WA, USA: Institute for Systems Biology; 2008. <http://www.repeatmasker.org/RepeatModeler/>.
- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase update, a database of eukaryotic repetitive elements. Cytogen Genome Res. 2005;110(1–4):462–7.
- Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. InterProScan: protein domains identifier. Nucleic Acids Res. 2005;33(Suppl 2):W116–20.
- Martinez Barrio A, Lamichhaney S, Fan G, Rafati N, Pettersson M, Zhang H, Dainat J, Ekman D, Höppner M, Jern P, Martin M, Nystedt B, Liu X, Chen W, Liang X, Shi C, Fu Y, Ma K, Zhan X, Feng C, Gustafson U, Rubin CJ, SällmanAlmén M, Blass M, Casini M, Folkvord A, Laikre L, Ryman N, Ming-Yuen Lee S, Xu X, Andersson L. The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing. Elife. 2016;3(5):e12081.
- Xu P, Zhang X, Wang X, Li J, Liu G, Kuang Y, Xu J, Zheng X, Ren L, Wang G, Zhang Y, et al. Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. Nat Genet. 2014;46(11):1212.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

