

DATA NOTE

Open Access



# Draft genome of *Ompok bimaculatus* (Pabda fish)

Ruby Dhar<sup>1</sup>, Karthikeyan Pethusamy<sup>1</sup>, Sunil Singh<sup>1</sup>, Indrani Mukherjee<sup>1</sup>, Ashikh Seethy<sup>1</sup>, Bharoti Sengupta<sup>1</sup>, Tryambak Srivastava<sup>1</sup>, Sajib Sarkar<sup>1</sup>, Vikramjit Mandal<sup>2</sup>, Madhumita Karmakar<sup>2</sup>, Sandipan Gupta<sup>3</sup>, Arpita Ghosh<sup>2</sup> and Subhradip Karmakar<sup>1\*</sup>

## Abstract

**Objective:** Pabda (*Ompok bimaculatus*) is a freshwater catfish, largely available in Asian countries, especially in Bangladesh, India, Pakistan and Nepal. This fish is highly valued for its fabulous taste and high nutritional value and is very popular as a rich source of proteins, omega-3 and omega-6 fatty acids, vitamins and mineral for growing children, pregnant females and elders. We performed *de-novo* sequencing of *Ompok bimaculatus* using a hybrid approach and present here a draft assembly for this species for the first time.

**Data Description:** The genome of *Ompok bimaculatus* (Fig. 1: Table 1, Data file 3) from Ganges river, has been sequenced by hybrid approach using Illumina short reads and PacBio long reads followed by structural annotations. The draft genome assembly was found to be 718 Mb with N50 size of 81 kb. MAKER gene annotation tool predicted 21,371 genes.

**Keywords:** Butter catfish, *Ompok bimaculatus*, Siluridae, Whole genome sequence, NGS

## Objective

Pabda fish is a freshwater catfish, with its fabulous taste and high nutritional value. *Ompok bimaculatus* (Family: Siluridae), also known as Indian butter catfish or commonly known as pabda, has fascinated considerable attention in diversification due to its good taste, high nutritional value and soft bony structure [1–3]. Pabda is largely available in Asian countries, especially in Bangladesh, India, Pakistan and Nepal. Catfishes are an excellent source of omega-3 and omega-6 fatty acids, vitamins and mineral those are excellent for growing children, pregnant females and elders. Overexploitation for food is a major threat and has resulted in remarkable population decline [4, 5]. The fish is in declining phase due to lack of definite information on the biological aspects

accompanied by declining population owing to habitat loss, indiscriminate pesticide and weedicide use, loss of breeding grounds and overfishing valued for its unique taste has pushed the population of Pabda in IUCN red list of threatened species [3, 5].

Knowing the complete genome of this fish will help in better understanding of the genome organization, evolution as well as for conservation and farming applications [6]. This involves stepping up of breeding process, as well as to identify lineage specific changes that are critical for its adaptation besides knowing about the risk factors as well as its immune system that helps in its survival in the wild or in captivity.

## Data description

Fresh *Ompok bimaculatus* which were approximately 4 months old were freshly caught from the Ganges river in India and instantly used for DNA extraction. The taxonomic identification of this fish has been confirmed following standard taxonomic keys like studying

\*Correspondence: subhradip.k@aiims.edu

<sup>1</sup> Department of Biochemistry, All India Institute of Medical Sciences, New Delhi, India

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 [16]	Sequencing statistics	Table	<a href="https://doi.org/10.6084/m9.figshare.9827312.v1">https://doi.org/10.6084/m9.figshare.9827312.v1</a>
Data file 2 [17]	Genome assembly statistics	Table	<a href="https://doi.org/10.6084/m9.figshare.9830954.v1">https://doi.org/10.6084/m9.figshare.9830954.v1</a>
Data file 3 [18]	Fish Pictures	Figure 1	<a href="https://doi.org/10.6084/m9.figshare.9831155.v1">https://doi.org/10.6084/m9.figshare.9831155.v1</a>
Data file 4 [19]	Whole genome assembly data	FASTA	<a href="https://www.ncbi.nlm.nih.gov/bioproject/PRJNA552450">https://www.ncbi.nlm.nih.gov/bioproject/PRJNA552450</a>
Data file 5 [20]	Whole genome sequence	FASTA	<a href="https://www.ncbi.nlm.nih.gov/assembly/GCA_009108245.1">https://www.ncbi.nlm.nih.gov/assembly/GCA_009108245.1</a>

the fin formula and other standard morphological characteristics. Muscle tissue was dissected from this fish and high-molecular weight genomic DNAs was purified from one specimen using Qiagen Genomic-tip 100/G as per the manufacturer's instruction. The quality and quantity of the isolated genomic DNA was checked on NanoDrop followed by 0.8% Agarose gel.

Two pair-end libraries were constructed for Illumina sequencing and one SMRTbell library was constructed for Pacific Bioscience Sequel, single molecule, real time (SMRT, Single Molecule Real Time) sequencing platforms as mentioned in Table 1, Data file 1. The quality of the reads was checked using FastQC (Table 1) [7].

MaSuRCA (Maryland Super- Read Celera Assembler) v3.2.8 was used for hybrid de novo assembly [8] using both the Illumina and PacBio data. The genome assembly has been deposited in the NCBI GeneBank under the Bioproject ID: PRJNA552450 (Table 1, Data files 2, 4). The assembled genome size of *Ompok bimaculatus* is 718 Mb and approximately 72% of the genome has been assembled as per the in silico genome size estimation.

The BUSCO v3 [9] analysis revealing 85.7% completeness, indicating the genome to be of good quality. MAKER v3.0 pipeline [10] was used for structural annotation. GC content of the genome was determined to be 38.84%. RepeatMasker v4.0.9 was used with the latest version of Repbase database [11, 12], repeat elements identified were 7.87%. Altogether, 21,371 genes were predicted by the MAKER gene annotation pipeline using proteins from channel catfish. Out of the 21,371 genes, 20,923 were annotated using Diamond [13] (BlastX mode) against NCBI 'NR' database and 5589 genes were found to have GO (Gene Ontology) term assigned to them. The butter catfish genome was found to be comparable to *Ictalurus punctatus* (Channel catfish, 892 Mb genome and 27,156 genes) [14] and to the genome of *Pangasianodon hypophthalmus* (Striped catfish, 715 Mb genome and 24,083 genes) [15].

## Limitations

The number of the scaffolds containing N in the genome are 27 and a total of 3773 bases are positioned in this gap region. The assembled genome size of the Indian butter catfish is 718 MB as compared to in silico estimated genome size of 992 Mb.

## Abbreviations

BUSCO: Benchmarking Universal Single-Copy Orthologs; MaSuRCA: Maryland Super-Read Celera Assembler; IUCN: International Union for conservation of nature; PacBio: Pacific Bioscience; GO: Gene Ontology; SMRT: single molecule real time sequencing.

## Acknowledgements

The authors like to thank Joydeep Karmakar for sample collection and the entire department of Biochemistry, AllIMS Delhi for providing support, infrastructure and logistic to work.

## Authors' contributions

SK and RD conceptualized, designed the overall project with assistance from AS and BS. KP, SS collected the samples. SK, AG, VM, MK sequenced and assembled the genome and also has performed the structural and functional annotations. SK, SG, AG, VM, RD wrote the manuscript. IM, TS, BS, SS assisted in writing the draft. All authors reviewed the manuscript. All authors read and approved the final manuscript.

## Funding

This research is jointly funded by RD and SK.

## Availability of data materials

The data described in this Data note can be freely and openly accessed via figshare, Refer Table 1 for details and links. The genome assembly has been deposited in the NCBI GeneBank under the Bioproject ID: PRJNA552450 (Table 3).

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup> Department of Biochemistry, All India Institute of Medical Sciences, New Delhi, India. <sup>2</sup> Eurofins Genomics India Pvt Ltd, Bangalore, India. <sup>3</sup> Department of Industrial Fish and Fisheries, Bramhananda Keshab Chandra College, West Bengal State University, Kolkata, India.

Received: 28 October 2019 Accepted: 19 December 2019  
Published online: 26 December 2019

## References

- Banik S, Goswami P, Acharjee T, Malla S. Ompok pabda (Hamilton-Buchanan, 1822): an endangered catfish of Tripura, India: reproductive physiology related to freshwater lotic environment. *J Environ.* 2012;1(2):45–55.
- Biswas P, Jena AK, Saha H, Chowdhury TG. Induced breeding and seed production of Pabda: a species with potential for aquaculture diversification in northeast India. *World Aquac.* 2018;49(1):41–5.
- Rawat P, Biswas P, Jena AK, Patel AB, Pandey PK. Effect of dietary incorporation of natural attractants on growth and survival during seed rearing of Indian butter catfish, *Ompok bimaculatus*. *J Environ Biol.* 2019;40(4):661–7.
- Mishra SS, Acharjee SK, Chakraborty SK. Development of tools for assessing conservation categories of siluroid fishes of fresh water and brackish water wetlands of South West Bengal, India. *Environ Biol Fishes.* 2009;84(4):395–407.
- Ng, H.H., Tenzin, K. & Pal, M. 2010. *Ompok bimaculatus*. The IUCN Red List of Threatened Species. 2010. <http://dx.doi.org/10.2305/IUCN.UK.2010-4.RLTS.T166616A6248140.en>.
- Ryder OA. Conservation genomics: applying whole genome studies to species conservation efforts. *Cytogenet Genome Res.* 2005;108(1–3):6–15.
- FastQC program. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. The MaS-uRCA 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 Holt, Alvarado C, 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. 2015. pp 289–300. (<http://www.repeatmasker.org/faq.html>).
- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res.* 2005;110(1–4):462–7.
- Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* 2015;12(1):59.
- Liu Z, Liu S, Yao J, Bao L, Zhang J, Li Y, Jiang C, Sun L, Wang R, Zhang Y, Zhou T. The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts. *Nat Commun.* 2016;7:11757.
- Kim OT, Nguyen PT, Shoguchi E, Hisata K, Vo TT, Inoue J, Shinzato C, Le BT, Nishitsuji K, Kanda M, Nguyen VH. A draft genome of the striped catfish, *Pangasianodon hypophthalmus*, for comparative analysis of genes relevant to development and a resource for aquaculture improvement. *BMC Genomics.* 2018;19(1):733.
- Dhar R, Pethusamy K, Singh S, Mukherjee I, Seethy A, Sengupta B, et al. Table 2.doc. figshare. 2019. <https://doi.org/10.6084/m9.figshare.9827312.v1>.
- Dhar R, Pethusamy K, Singh S, Mukherjee I, Seethy A, Sengupta B, et al. Table 3.doc. figshare. 2019. <https://doi.org/10.6084/m9.figshare.9830954.v1>.
- Dhar R, Pethusamy K, Singh S, Mukherjee I, Seethy A, Sengupta B, et al. Ompok Picture. figshare. 2019. <https://doi.org/10.6084/m9.figshare.9831155.v1>.
- <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA552450>.
- [https://www.ncbi.nlm.nih.gov/assembly/GCA\\_009108245.1/](https://www.ncbi.nlm.nih.gov/assembly/GCA_009108245.1/).

## Publisher's Note

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

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](https://biomedcentral.com/submissions)

