



ELSEVIER

Contents lists available at ScienceDirect

Data in Brief

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



Data article

Mitochondrial transcripts and associated heteroplasmies of *Ancistrus* spp. (Siluriformes: Loricariidae)



Daniel A. Moreira^a, Carolina Furtado^b, Thiago E. Parente^{a,*}

^a Laboratório de Toxicologia Ambiental, DCB, Escola Nacional de Saúde Pública – ENSP, Fundação Oswaldo Cruz – FIOCRUZ, Rio de Janeiro, Brasil

^b Divisão de Genética, Instituto Nacional do Cancer – INCA, Rio de Janeiro, Brasil

ARTICLE INFO

Article history:

Received 9 September 2015

Accepted 10 September 2015

Available online 22 October 2015

ABSTRACT

This data-set complements our paper entitled “The use of transcriptomic next-generation sequencing data to assembly mitochondrial genomes of *Ancistrus* spp. (Loricariidae)” [6]. Here, we present the nucleotide sequences of each transcript used for mitogenomes assembly, as well as tables presenting the location of each transcript in the mitogenomes; the frequency, location and codon position of the detected heteroplasmic sites; and the start/stop codons usage, UTR, CDS and poliA-tail length for each protein coding gene. Readers are referred to the paper cited above for data interpretation and discussion.

© 2015 Published by Elsevier Inc. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

Specifications table

Subject area *Biology*

More specific subject area *Biodiversity*

DOI of original article: <http://dx.doi.org/10.1016/j.gene.2015.08.059>

* Corresponding author.

E-mail address: parente@ensp.fiocruz.br (T.E. Parente).

<http://dx.doi.org/10.1016/j.dib.2015.09.010>

2352-3409/© 2015 Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Type of data	Table, text (.fasta) file
How data was acquired	Illumina HiSeq2500
Data format	Analyzed
Experimental factors	Raw reads were assembled by Trinity
Experimental features	Mitochondrial transcripts were identified by BLASTN against two other Loricariidae mitogenomes
Data source location	Fishes were donated by a fish exporter and sampled in the Amazon region
Data accessibility	Genbank KP960567, KP960568, KP960569

Value of the data

- These are the first sequences of long mitochondrial transcripts from *Ancistrus* spp.
 - This data will aid primer design for other Loricariidae species in phylogenetic studies.
 - Heteroplasmies position and frequency can be compared to other species.
 - Start/stop codons usage, UTR, CDS and poliA-tail length for protein coding gene are described.
-

1. Data

Nucleotide sequences of each transcript used to assemble the mitogenomes of three *Ancistrus* spp. fish described by Moreira et al. [6] are available in fasta format (supplemental material). Table 1 shows the number and the length of each of those mitochondrial transcripts, as well as the maximum number of supporting reads per individual nucleotide and the total sum of supporting reads for each transcript. The position of heteroplasmic sites is shown at Table 2, along with the read counts of each nucleotide, the gene category and the codon position of each heteroplasmic site. Specifically for protein-coding genes, Table 3 shows the lengths of 5'UTR, 3'UTR, ORF and Poly-A tail. In addition, Table 3 also shows the initiation and the termination codon, as well as the number of Adenines added to complete the stop codons, if needed.

2. Experimental design, materials and methods

Fish sampling, RNA extraction, Illumina library preparation and sequencing, raw reads processing and transcriptome assemble are described elsewhere [6]. Each transcriptome was subjected to a BLASTN against the complete mitogenome of the closest related species, whose mitogenome is publically available, *Pterygoplichthys disjunctivus* (GI: 339506171) [7]. The transcripts aligned to the reference mitogenome were used for the mitogenomes assembly. The selected transcripts were edited according to the information of strand orientation given by the BLASTN result, and aligned by SeaView using the built-in CLUSTAL alignment algorithm and the reference mitogenome [2]. A CONTIG sequence was generated using the sequence information of just the transcripts of each individual fish. The sequence of the CONTIG was then manually checked for inconsistencies and gaps. The mitogenomes were annotated using the web-based services MitoFish and MITOS [1,3] and features in protein-coding genes were analyzed according to Temperley et al. [9]. In order to estimate the support of each base in the mitogenomes, Bowtie v. 1.0.0 was used to align the reads of each fish on its own assembled mitogenome, and this mapping was viewed using the Integrated Genome Viewer (IGV) or the Tablet [10,4,5,8]. Heteroplasmic sites were detected using IGV, setting the software to show positions in which the frequency of the second most frequent base was equal to or higher than 10% and the total reads number were higher than 100.

Acknowledgments

This data-set was generated with the financial support of the Partnerships for Enhanced Engagement in Science (PEER) initiative from the United States Agency for International Development [Grant no. PGA-2000003446] to TP.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.dib.2015.09.010>.

References

- [1] M. Bernt, A. Donath, F. Jühling, F. Externbrink, C. Florentz, G. Fritzsch, J. Pütz, M. Middendorf, P.F. Stadler, MITOS: improved de novo metazoan mitochondrial genome annotation, *Mol. Phylogenet. Evol.* 69 (2013) 313–319.
- [2] M. Gouy, S. Guindon, O. Gascuel, SeaView Version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building, *Mol. Biol. Evol.* 27 (2010) 221–224.
- [3] W. Iwasaki, T. Fukunaga, R. Isagozawa, K. Yamada, Y. Maeda, T.P. Satoh, T. Sado, K. Mabuchi, H. Takeshima, M. Miya, M. Nishida, Mitofish and mitoannotator: a mitochondrial genome database of fish with an accurate and automatic annotation pipeline, *Mol. Biol. Evol.* 30 (2013) 2531–2540.
- [4] B. Langmead, C. Trapnell, M. Pop, S.L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, *Genome Biol.* 10 (2009) R25.
- [5] I. Milne, M. Bayer, L. Cardle, P. Shaw, G. Stephen, F. Wright, D. Marshall, Tablet – next generation sequence assembly visualization, *Bioinformatics* 26 (2009) 401–402.
- [6] D.A. Moreira, C. Furtado, T.E. Parente, The use of transcriptomic next-generation sequencing data to assembly mitochondrial genomes of *Ancistrus* spp (Loricariidae), *Gene* 573 (2015) 171–175.
- [7] M. Nakatani, M. Miya, K. Mabuchi, K. Saitoh, M. Nishida, Evolutionary history of Otophysi (Teleostei), a major clade of the modern freshwater fishes: Pangaeen origin and Mesozoic radiation, *BMC Evol. Biol.* 11 (2011) 177.
- [8] J.T. Robinson, H. Thorvaldsdóttir, W. Winckler, M. Guttman, E.S. Lander, G. Getz, J.P. Mesirov, Integrative genomics viewer, *Nat. Biotechnol.* 29 (2011) 24–26.
- [9] R.J. Temperley, M. Wydro, R.N. Lightowlers, Z.M. Chrzanowska-Lightowlers, Human mitochondrial mRNAs-like members of all families, similar but different, *Biochim. Biophys. Acta – Bioenerg.* 1797 (2010) 1081–1085.
- [10] H. Thorvaldsdóttir, J.T. Robinson, J.P. Mesirov, Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration, *Brief Bioinform.* 14 (2013) 178–192.