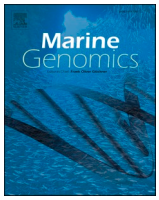
Marine Genomics 81 (2025) 101190

Contents lists available at ScienceDirect

Marine Genomics

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



Data Article

De novo assembly and annotation of the pantranscriptome of *Astyanax lacustris* on the liver and pituitary-gonadal axis

David Aciole Barbosa [a,1,*], Giovana Souza Branco [b,1], Aline Dal'Olio Gomes [c], Carlos Eduardo Tolussi [d], Marcela Muñoz-Peñuela [b], Bruno C. Araújo [e], Iuri Batista da Silva [f], Renata Guimarães Moreira [b], Luiz R. Nunes [g], Fabiano B. Menegidio [a]

[a] Núcleo Integrado de Biotecnologia, Universidade de Mogi das Cruzes (UMC), Brazil
[b] Laboratório de Metabolismo e Reprodução de Organismos Aquáticos, Departamento de Fisiologia, Instituto de Biociências, Universidade de São Paulo, Rua do Matão, Trav.14, n° 321, 05508-090 São Paulo, SP, Brazil
[c] Departamento de Biodiversidade, Universidade Estadual Paulista (UNESP), Instituto de Biociências, Rio Claro, SP, Brazil
[d] Departamento de Biologia, Universidade Federal do Ceará, Campus do Pici, Centro de Ciências, Bloco 909, 60440-900 Fortaleza, Ceará, Brazil
[e] Núcleo de Ciências Ambientais, Universidade de Mogi das Cruzes (UMC), Brazil
[f] Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil
[g] Centro de Ciências Naturais e Humanas, Universidade Federal do ABC, São Paulo, SP, Brazil

ARTICLE INFO

Keywords:
RNAseq
Reproduction
Hepatic
Testis
Ovaries
Pituitary

ABSTRACT

Astyanax lacustris is a model of laboratory native fish species. Reproductive studies of this species have already been performed. Nevertheless, there is a relative shortcoming of gene sequence information available in public databases, which hinder their use in more comprehensive investigations that employ sensitivity molecular biology techniques to assess gene expression profile for biomarker identification. In this data article, we report the first *de novo* transcriptome assembly of *A. lacustris* testicles, ovaries and male / female pituitary gland improving gene sequence data available for this fish species and transcriptome of male liver. Illumina sequencing generated 808,023,356 raw reads, filtered in 752,739,866 high-quality reads. Initially, a *de novo* assembly was filtered to include protein coding elements only in each tissue sample, which were merged in a final pantranscriptome (PAN) containing 109,232 contigs. The PAN was functionally annotated against a custom Actinopterygii proteins dataset and EggNOG terms with the aid of EnTAP, retrieving homology queries for about 90 % of all transcripts. Therefore, in this study we provide a PAN and a custom blast tool that can help discovery genomic information on metabolism pathways and their related genes in *A. lacustris*, enabling future research and molecular studies using this fish species as a model.

1. Introduction

Astyanax lacustris (Lütken, 1875) (Acestrorhamphidae: Acestrorhamphinae), is a South American teleost species previously classified as *Astyanax altiparanae* by Garutti and Britski (2000), synonymized by Lucena and Soares (2016), and popularly known as yellow tail tetra which was distinguished from other species of the genus *Astyanax* by presenting a horizontally oval, black humeral spot and two brown vertical bars, situated in the humeral region, characteristics that include it in the *Astyanax bimaculatus* group (Nakatani, 2001).

The yellow tail tetra that compose the *A. bimaculatus* group, including *A. lacustris*, present a wide geographical distribution, extending over practically the entire neotropical region (Lucena and Soares, 2016). They are small fish, well known for present high trophic opportunism (Bennemann et al., 2005; Peretti and Andrian, 2008) and for their ability to inhabit a variety of environments (rivers, streams, lakes, reservoirs) with different levels of preservation, including highly polluted (Alonso et al., 2019). Regarding their reproductive biology,

* Corresponding author at: Núcleo Integrado de Biotecnologia, Universidade de Mogi das Cruzes (UMC), Av. Dr. Cândido X. de Almeida e Souza, 200 - Centro Cívico, Mogi das Cruzes - SP, 08780-911, Brazil.
E-mail addresses: aciole.d@gmail.com (D. Aciole Barbosa), aline.d.gomes@unesp.br (A. Dal'Olio Gomes), brunoaraujo@umc.br (B.C. Araújo), iuribs@ufmg.br (I.B. da Silva), luiz.nunes@ufabc.edu.br (L.R. Nunes), fabianomenegidio@umc.br (F.B. Menegidio).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.margen.2025.101190
Received 6 February 2025; Received in revised form 2 April 2025; Accepted 7 April 2025
Available online 16 April 2025

they present early sexual maturation (Santos et al., 2016), the presence of sexual dimorphism (Siqueira-Silva et al., 2020) and multiple spawning (Cassel et al., 2017). Additionally, *A. lacustris* is widely important in neotropical aquaculture, commonly used in fish farming and human consumption and also as live bait for sport fishing (Porto-Foresti et al., 2005). Thus, biological characteristics and their ecological and economic importance have allowed its use as a model in many ecotoxicological studies (Abdalla et al., 2019; Assis et al., 2021; Branco et al., 2021; Godoi et al., 2020, 2024; Kida et al., 2016; Pinheiro et al., 2021) and about its reproduction in captivity (Brambila-Souza et al., 2021; Chehade et al., 2015; Machado-Evangelista et al., 2019).

To date, published studies regarding the reproductive biology of *A. lacustris* present various aspects such as early development (Adolfi et al., 2015; Santos et al., 2016), gametogenesis (Camargo et al., 2017; Cassel et al., 2017; Costa et al., 2014; Postingel Quirino et al., 2021; Rodrigues et al., 2015), regulation of pituitary hormones (Branco et al., 2019; Jesus et al., 2017) and cloning of hypothalamic-pituitary-gonadal axis genes (Chehade et al., 2020; Jesus et al., 2017). Despite the existence of these works, key gaps in the reproductive physiology and biology of this fish species may prevent the access of novel information, hence, obscuring the development of future studies and advances that can directly impact major interest fields including environment and aquaculture. Current scarcity of particular *A. lacustris* genomic and transcriptomic public information hinders the possibility of broadening genetic exploration methods. Provisioning of foundational omic data resources can aid in accelerating the investigation of structural and functional aspects of specific genes involved, for instance, with the endocrine and reproductive system.

In turn, the transcriptome analysis is an important tool that can be applied in the discovery of new genes, in the development of genetic markers, gene expression studies, among others (Leaver et al., 2015). The main objectives of this type of analysis are to investigate the functional elements of the organism's transcript sequences, and to obtain a global view of how gene expression changes under the influence of other variables (Mazurais et al., 2011). Transcriptomic analysis is a modern technique that stands out among the most reliable and fast for generating information about a given biological process. This is because the transcriptome allows access to all genes expression profiles in a given condition. The large amount of data generated by the transcriptome, often reaching billions of base pairs, can allow a deeper investigation of the physiological processes, including reproduction, facilitating, cheapening and complementing molecular analyzes, such as the quantification of gene expression.

Studies using molecular tools add several information to the endocrine disruption of aquatic organisms, among them the next-generation sequences, especially the transcriptome, which is the sequencing of the genes transcribed in a cell, tissue, or specific organism for a given stage of development, or physiological condition (Qian et al., 2014), it is not stable like the genome and varies according to the external environment (Wang et al., 2018).

We report here for the first time the *de novo* assembled pantranscriptome of *A. lacustris* pituitary, testicle, ovary and liver, highlighting a public custom blast database that can be further assessed to establish molecular signatures with different applications from reproductive biology and technology investigations, besides reproductive alterations against contaminants.

2. Data description

2.1. Sample collection, RNA extraction and sequencing

A. lacustris adults (3 males and 3 females) were provided by Grupo Votorantim Energia (formerly Estação de Hidrobiologia e Aquicultura de Paraibuna - CESP), Paraibuna, São Paulo (23°24′47"S, 45°36′7"W). Animals were safely transported to the laboratory facilities of the Ectotherms Lab at the Department of Physiology - Institute of Biosciences, University of São Paulo, USP and maintained in black boxes (350 L) for 7 days, in a 12-h light/12-h dark cycle and *ad libitum* access to commercial extruded feed (32 % crude protein) until 24 h before any manipulation.

The animals were anesthetized in benzocaine solution (0.1 g L^{-1}) and then submitted to biometrics, measuring the total length (cm) and body mass (g) (9.76 ± 1.31 cm, 10.38 ± 2.86 g, respectively). The spinal cord of each animal was sectioned and the head separated from the animal's body. The pituitary (female - PFL, male - PML), liver (LL), testis (TL) and ovaries (OL) were dissected and immediately frozen on dry ice for RNA extraction. The experimental procedure was approved by the Ethics Committee on Animal Use, CEUA, IB, USP, processes n° 275/2017 and n° 366/2020.

Total RNA was obtained using the PureLink® RNA Mini Kit (Code 12183018 A), according to the manufacturer's instructions and the concentration (ng/μl), purity and yield were assessed using a NanoDrop One spectrophotometer (ThermoFisher). Only the samples with Abs260 / Abs280 ratio greater than 1.8 were used, and the RNA integrity (RIN) was determined by the 2100 Bioanalyzer (Agilent Technologies, CA, USA), only samples with RIN between 8 and 10 were used. All purified RNA samples were kept at −80 °C. A total of 1 μg of RNA per sample was used as the input material for the RNA sample preparations.

Library construction and sequencing were carried out at the GENIAL Facility, at the Center for Research Support Facilities (CEFAP-USP), Institute of Biomedical Sciences of the University of São Paulo (USP). RNA extracted from tissue samples from all animals were equally diluted to 1.000 ng μL^{-1} and pooled for library construction using the TruSeq RNA Sample Preparation kit, according to the manufacturer's specifications (Illumina Inc., USA). Library fragments were purified with the AM Pure XP system (Beckman Coulter, Beverly, MA, USA) and library was quantified by Qubit 2.0 before subjected to paired-end sequencing (2 × 75 bp) in a NextSeq® equipment (Illumina, Inc., USA) according to the manufacturer's recommendations. MIxS data (Minimum Information about any Sequence) for this study is available in Table 1. Raw sequences have been submitted to the Sequence Read Archive (SRA) database, as detailed in item 3.

2.2. De novo transcriptome assembly and functional annotation

Raw fastq sequencing data was mainly processed (Fig. S0) in a public Galaxy server, available at https://usegalaxy.eu (Jalili et al., 2020). Initially, the quality of raw sequences was assessed using FastQC (Andrews, 2010) and MultiQC (Ewels et al., 2016). Fastp (Chen et al., 2018) was used to remove low-quality reads (Q < 30), adapters, and other contaminant sequences. Trinity platform (Haas et al., 2013) was then used for *de novo* transcriptome assembly of filtered reads, and assembly metrics were obtained using the TrinityStats script. Transfuse

Table 1
MIxS descriptors.

Item	Description
Investigation type	Eukaryote transcriptome
Project name	PRJNA1168055: *De novo* assembly and annotation of the Pantranscriptome of Astyanax lacustris on the liver and pituitary-gonadal axis
Geographic location (latitude and longitude)	23°24′47"S, 45°36′7"W
Geographic location (country and/or sea, region)	Brazil: Paraibuna, Sao Paulo
Collection date	2019–10; 2020–10
Broad-scale environmental context	aquatic biome ENVO_00002030
Local environmental context	freshwater river biome ENVO_01000253
Environmental medium	freshwater ENVO_00002011
Sequencing method	Illumina NextSeq 500
Assembly	*De novo* assembly with Trinity v2.9.1 (Usegalaxy.Eu)

(Boursnell, 2024) approach was applied to unify the tissue-derived assemblies in a PAN. Transcriptome completeness was finally assessed using the Benchmarking Universal Single-Copy Orthologs (BUSCO) software (Manni et al., 2021) based on actinopterygii_odb10 (Kriventseva et al., 2019).

Elements of the assembled transcriptome were functionally annotated with the aid of the Eukaryotic Non-Model Transcriptome Annotation Pipeline (EnTAP) v0.10.8 (Hart et al., 2020) in two moments; (i) similarity search using blastx with e-value $\leq$ e-5 and $\geq$ 50 % minimum coverage against the custom database FishProteinDB (Batista da Silva et al., 2023) and EggNOG database (Huerta-Cepas et al., 2016). FishProteinDB consists of 171,502 protein sequences of Hyperoartia, Myxini, Chondrichthyes, Actinopterygii, and Sarcopterygii species (excluding the Tetrapod clade) from RefSeq.

The EggNOG hits carried assignments for biological function to the putative *Astyanax* genes, identifying GO (Gene Ontology) (Ashburner et al., 2000; Gene Ontology Consortium, 2019) and KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa and Goto, 2000; Kanehisa et al., 2019) terms. The EnTAP functional annotation process was carried out using a Dugong container environment (Menegidio et al., 2018). Additional transcriptome filtering before NCBI submission was carried out with the aid of *subseq* from seqtk toolkit (Shen et al., 2016), based on sequence length (>200 nucleotides) and Univec Database (https://ftp.ncbi.nlm.nih.gov/pub/UniVec/).

Non-redundant Coding sequences (CDSs) provided by the Transdecoder after CD-HIT (0.9 sequence identity threshold) (Fu et al., 2012) for all tissues were then merged into a single assembly using Transfuse v0.5.0 (REF), a tool that can properly reconstruct and fuse multiple transcriptome sets. In this work, we consider a pantranscriptome as: the combined assembly constructed by Transfuse from the independent assembly data from all tissues of *A. lacustris* described above.

2.3. Structural and functional characterization of transcriptome

Sequencing of the cDNA libraries derived from *A. lacustris* material resulted in 808,023,356 raw reads. After high-quality reads selection and trimming, total sequences consisted of 752,739,866 reads (93.2 % of raw reads) (Table S1-S0), which were used for *de novo* transcriptome assembly, using Trinity v2.9.1 (Haas et al., 2013). General features of the *A. lacustris* transcriptome are summarized in Table S1-S1, showing total sequence counts of: PFL - Unigenes (U) = 20,487 and Isoforms (I) = 39,702; PML - U = 21,267; *I* = 39,386; OL - U = 18,947; *I* = 48,529; TL - U = 32,317; *I* = 106,456; LL - U = 14,089; *I* = 27,752. These initial transcriptomes exhibit average length [base pairs] of: PFL - U = 1159.85; *I* = 1262.33; PML - U = 8,81.86; *I* = 9,40.37; OL - U = 1262.38; *I* = 1345.87; TL - U = 1203.51; *I* = 1216.37; LL - U = 711.65; *I* = 750.57. The N50 metric identified for each dataset is: PFL - U = 1626; *I* = 1779; PML - U = 1164; *I* = 1242; OL - U = 1803; *I* = 1869; TL - U = 1773; *I* = 1743; LL - U = 846; *I* = 900. Completeness of the assemblies were pointed by BUSCO (Fig. 1A); PFL - 2747 (75.5 %); PML - 1650 (45.4 %); OL - 2818 (77.4 %); TL – 3197 (87.9 %); LL - 829 (22.8 %).

Functional annotations from FishProteinDB and EggNOG *via* EnTAP covered nearly 90 % of all sequences in each assembly: PFL - 93.38 %; PML - 93.23 %; OL - 93.41 %; TL - 89.13 %; LL - 90.93 %. Fig. 1B shows that, independently of the datasets' length, similar percentages of annotation were found along the multiple functional categories returned (protein domains, Gene Ontology-GO Biological Process, GO Molecular Function, GO Cellular Component and Kegg) (Table S1-S1).

Comparison of transcripts present in different tissues was conducted using BLAST Reciprocal Best Hits - RBH (Cock et al., 2015) megablast (70 % identity, 70 % coverage, manually removing duplicated pair hits). Considering that the number of non-shared transcripts in a tissue range from 25,505 (LL) to 99,168 (TL), few elements are shared among all tissues under comparisons, ranging from 3708 (PFL x OV) to 406 (OL x TL), as shown in Table S1-S2. This fact may reflect the importance and uniqueness of sequences contained in the individual assemblies that are then merged in the construction of a pantranscriptome.

The PAN derived from the independently assembled datasets presented 109,232 sequences, with 1216.37 bp average length, N50 = 1713, in a total of 132,866,035 assembled bases, while its completeness was found to be at 91.4 %. In total, 98,504 PAN elements (90.18 %) received functional annotations. Gene Ontology information was also attributed to PAN transcripts; 61,586 (56.4 %) displayed matches to GO Biological Processes, while 61,709 (56.5 %) correspond to GO Molecular Functions and 46,750 (42.80 %) were deemed as GO Cellular Components. Protein domain classification was applied to 89,835 of PAN sequences and, finally, 29,735 (27.2 %) transcripts were annotated into at least one KEGG pathway term (Fig. 1B-C, Table S1-S1).

The majority of hits with taxonomy information, based on sequence similarity, match *A. mexicanus* (38,040; 78.6 %), while the taxonomy scope from EggNOGG reveals around 76,099 hits (79.8 % of hits with taxonomy scope information) are directly mapped to fish (Fig. 1D-E).

All the relevant annotation data was made available in tabular files along with other intermediary outputs at OSF repository https://osf.io/5zmyp/.

2.4. Liver-pituitary-gonad relevant genes

Functional annotation of *A. lacustris* pantranscriptome provided means to find potentially relevant transcripts that may be involved in important reproduction and other general pathways in fish. A summarized collection of these transcripts and their respective annotations is consolidated in Table S1-S3.

The datasets presented herein contain functional terms associated, for instance, to: **(i) *oocyte*** (GATOR complex protein MIOS; H1 histone family, member O, oocyte-specific; missing oocyte meiosis regulator; oocyte zinc finger protein XlCOF6.1-like; oocyte-specific histone RNA stem-loop-binding protein 2-like; WEE2 oocyte meiosis inhibiting kinase); **(ii) *GnRH*** (chorionic gonadotropin, beta polypeptide 1; Gonadotropin Inhibitory Hormone Receptor 1 Like, partial; gonadotropin releasing hormone 2; gonadotropin subunit beta-2; Gonadotropin-releasing hormone; gonadotropin-releasing hormone II receptor-like; gonadotropin-releasing hormone II receptor-like; gonadotropin-releasing hormone II receptor-like, partial; luteinizing hormone; luteinizing hormone choriogonadotropin receptor), **(iii) *estrogen*** (estrogen receptor; estrogen receptor beta-1-like; estrogen receptor binding site associated antigen 9; estrogen receptor-like; estrogen related receptor beta; estrogen-related receptor alpha; estrogen-related receptor gamma a; estrogen-related receptor gamma b; estrogen-related receptor gamma-like; g protein-coupled estrogen receptor 1; greb1 growth regulating estrogen receptor binding 1; growth regulating estrogen receptor binding 1), **(iv) *progesterone*** (membrane-associated progesterone receptor component 2-like; progesterone immunomodulatory binding factor 1; progesterone receptor); and **(v) *TGF-beta*** (tgf-beta activated kinase 1 map3k7 binding protein 1; tgf-beta activated kinase 1 map3k7 binding protein 2; tgf-beta activated kinase 1 map3k7 binding protein 3; tgf-beta propeptide; tgf-beta receptor type-1; tgf-beta receptor type-2-like; tgf-beta-activated kinase 1 and map3k7-binding protein 1; tgf-beta-activated kinase 1 and map3k7-binding protein 3; tgfb-induced factor homeobox 1; tgfb-induced factor homeobox 2; transforming growth factor beta 1; transforming growth factor beta 1 induced transcript 1; transforming growth factor beta 2; transforming growth factor beta 3; transforming growth factor beta induced; transforming growth factor beta receptor 2; transforming growth factor beta receptor associated protein 1; transforming growth factor beta-1-like; transforming growth factor beta-2-like). We believe that further investigations making use of the data obtained in the present work can help elucidate more about these *A. lacustris* metabolic actors.

2.5. Astyanax lacustris Pantranscriptome BLAST database

To facilitate access to the functionally annotated transcript

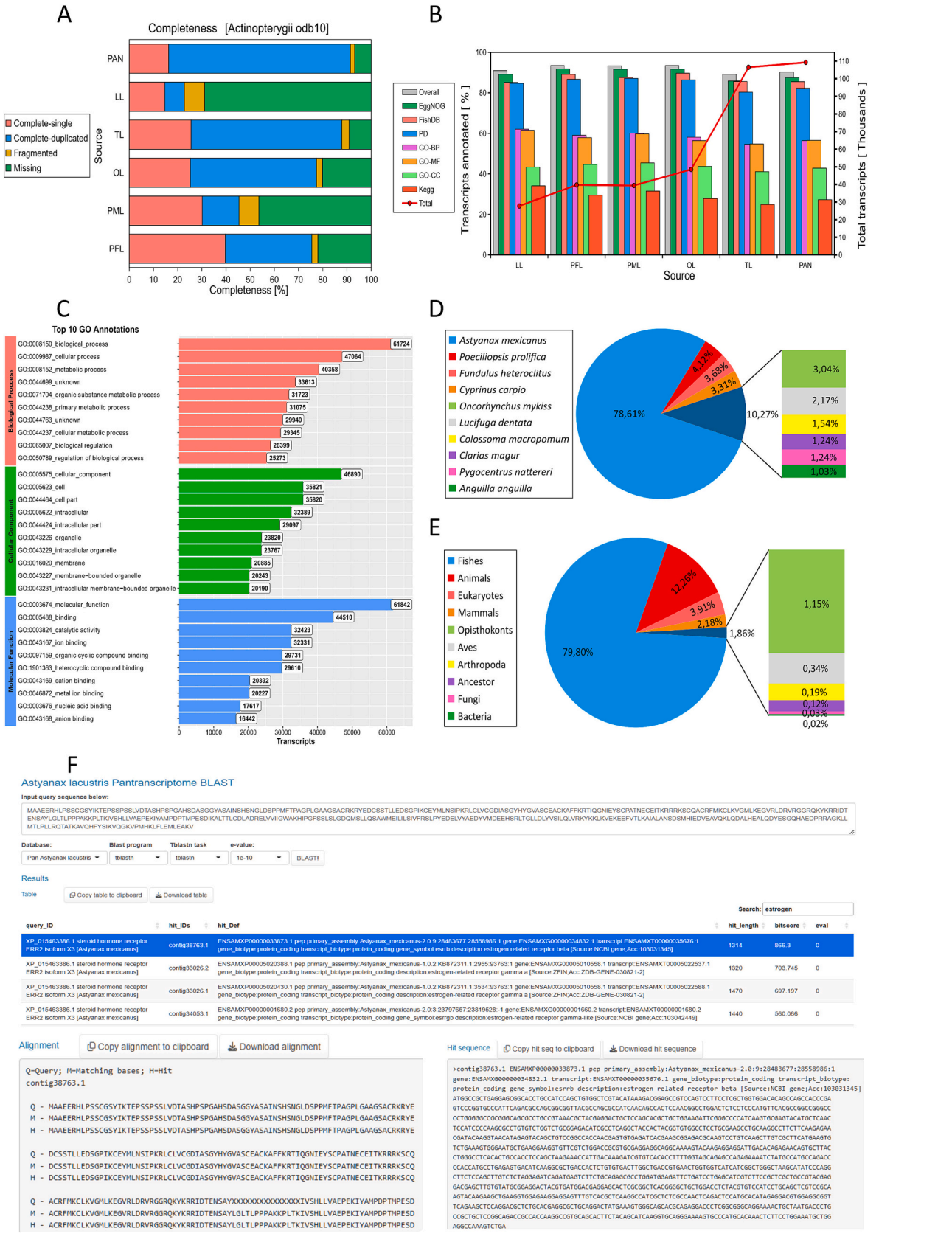


Fig. 1. A: Busco completeness based on actinopterygii_odb10. B: Functional annotation along diverse databases and transcriptome size. C: Gene ontology categories assigned. D: Taxonomy of hits by similarity. E: Taxonomy of hits by taxscope. F: Blast database interface and results.

sequences of the *A. lacustris* pantranscriptome, we provide a custom web database at https://aciole-d.shinyapps.io/astypan-blast based on R (R Core Team, 2024) and Shiny (Chang et al., 2024) template engine (Warner, 2024). This resource allows the search for sequence similarity of both DNA and protein queries through the blastn and tblastn programs from the BLAST software (Camacho et al., 2009), with options to set different task (blastn: blastn, blastn-short, dc-megablast, megablast, rmblastn; tblastn: tblastn, tblastn-fast) and e-value parameters. It also provides copy and direct download of blast results in the format of table, alignment and hit sequence (Fig. 1F).

To ensure that this database remains available in a continuous and reproducible way, we also provide the entire environment available in a Docker container (available at https://github.com/Aciole-David/panasty_lacus_blast and https://hub.docker.com/r/davidaciole/panasty_lacus_blast). These GitHub and Docker Hub repositories also contain download and install instructions, so users can use them to install fully functional mirrors in their own work environments. The Docker image can be downloaded for free and deployed in a local computer, with few infrastructure requirements, as described by Aciole Barbosa et al. (2019).

This study represents a significant milestone by offering the first pantranscriptome of *A. lacustris*, expanding the genomic resources available for this native fish species and enabling detailed exploration of genetic pathways related to reproduction and metabolism. The high functional annotation rate (90 % of transcripts) highlights the effectiveness of the bioinformatics pipeline employed herein, confirming the applicability of tools like Trinity and EnTAP for non-model organisms (Bauersachs and Wolf, 2012; Sudhagar et al., 2018).

The results revealed the presence of crucial genes related to the hypothalamic-pituitary-gonadal axis, such as GnRH, progesterone receptors, and components of the TGF-beta pathway. These findings align with previous studies in *A. altiparanae*, which demonstrated the importance of these genes in hormonal regulation and reproductive processes (Branco et al., 2019; Jesus et al., 2017). The comparison of transcripts with fish databases reinforces the functional proximity between *A. lacustris* and *A. mexicanus*, allowing ecological and adaptive inferences (O'Quin and Mcgaugh, 2016).

The creation of a BLAST database for *A. lacustris* constitutes a strategic tool, enabling future studies on environmental biomarkers and reproductive adaptations. Recent studies highlight the role of *Astyanax* as a bioindicator in ecotoxicology, particularly in response to exposure to contaminants such as endocrine disruptors (Godoi et al., 2024; Schulz and Martins-Junior, 2001). Furthermore, the inclusion of liver- and gonad-specific genes in the pantranscriptome allows for more detailed investigations into physiological responses to environmental and dietary stressors.

2.6. Conclusion

Our study built the first liver, pituitary-gonads pantranscriptome assembly of this important native species, providing an important tool for further research with *A. lacustris* and related organisms. The availability and deposition of the transcriptome sequence allows researchers to access novel gene sequences, contributing to gene expression assessments, consequently improving our knowledge regarding fish physiology and reproduction. In addition, the provided assembly and blast engine will be essential to aid in future projects involving this species.

3. Data accessibility

Sequencing raw data were deposited in the Sequence Read Archive (SRA) repository of the National Center for Biotechnology Information (NCBI), under accession numbers SRR30861699, SRR30861700, SRR30861701, SRR30861702, SRR30861703, SRR30861704, SRR30861705, SRR30861706, SRR30861707, associated to the BioProject numbers PRJNA1168055 and BioSamples numbers SAMN44023690, SAMN44023691, SAMN44023692, SAMN44023693, SAMN44023694, SAMN44023695, SAMN44023696, SAMN44023697, SAMN44023698. The Transcriptome Shotgun Assembly (TSA) project corresponding to the univec-filtered (REF UNIVEC), deduplicated assembly has been deposited at DDBJ/EMBL/GenBank under accession number GKZC00000000. The version described in this paper is the first version, GKZC01000000. Supplementary Table S1 is available from Figshare (doi:https://doi.org/10.6084/m9.figshare.27609957.v1). Additional data derived from this study (including all relevant intermediate data) are also available from the Open Science Framework (OSF) repository https://osf.io/5zmyp/.

CRediT authorship contribution statement

David Aciole Barbosa: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Giovana Souza Branco:** Writing – review & editing, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Aline Dal'Olio Gomes:** Methodology. **Carlos Eduardo Tolussi:** Methodology. **Marcela Muñoz-Peñuela:** Methodology. **Bruno C. Araújo:** Methodology. **Iuri Batista da Silva:** Validation, Software, Methodology. **Renata Guimarães Moreira:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Luiz R. Nunes:** Writing – review & editing, Software, Formal analysis. **Fabiano B. Menegidio:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Ethics approval

The experiments involving animals mentioned herein was approved by the Ethics Committee on Animal Use, CEUA, IB, USP, process n° 275/2017 and n° 366/2020.

Declaration of competing interest

The authors report no conflicts of interest. The authors alone are responsible for the content and the writing of the paper.

Acknowledgments

We are grateful to the Grupo Votorantim Energia (formerly Estação de Hidrobiologia e Aquicultura de Paraibuna – CESP, Paraibuna, SP, Brazil), especially to Danilo Caneppele, Leandro Celestino and employees of the pisciculture station, for all the help and donated fish. We thank the Core Facility for Scientific Research – University of São Paulo (CEFAP-USP/GENIAL- (Genome Investigation and Analysis Laboratory)) for the preparation of libraries and sequencing analysis. This work was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) [grants 2017/11530-1; 2022/11174-9; 2022/10856-9].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.margen.2025.101190.

Data availability

Data are available on cited repositories and additional information can be requested

References

Abdalla, R.P., Kida, B.M.S., Pinheiro, J.P.S., Oliveira, L.F., Martinez, C.B.F., Moreira, R. G., 2019. Exposure to aluminum, aluminum+manganese and acid pH triggers

different antioxidant responses in gills and liver of *Astyanax altiparanae* (Teleostei: Characiformes: Characidae) males. Comp. Biochem. Physiol. C Toxicol Pharmacol. 215, 33–40. https://doi.org/10.1016/j.cbpc.2018.09.004.

Aciole Barbosa, D., Menegidio, F.B., Alencar, V.C., Goncalves, R.S., Silva, J.D.F.S., Vilas Boas, R.O., Nunes, L.R., 2019. ParaDB: a manually curated database containing genomic annotation for the human pathogenic fungi *Paracoccidioides* spp. PLoS Negl. Trop. Dis. 13 (7), e0007576. https://doi.org/10.1371/journal.pntd.0007576.

Adolfi, M.C., Carreira, A.C.O., Jesus, L.W.O., Bogerd, J., Funes, R.M., Schartl, M., Sogayar, M.C., Borella, M.I., 2015. Molecular cloning and expression analysis of dmrt1 and sox9 during gonad development and male reproductive cycle in the Lambari fish, *Astyanax altiparanae*. Reprod. Biol. Endocrinol. 13 (2), 1–15. https://doi.org/10.1186/1477-7827-13-2.

Alonso, M.B., Carvalho, D.R., Alves, C.B.M., Moreira, M.Z., Pompeu, P.S., 2019. Changes in trophic characteristics of two fish species of *Astyanax* (Teleostei: Characidae) in response to aquatic pollution. Zoologia 36, 1–12. https://doi.org/10.3897/zoologia.36.e30445.

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc (accessed 01 November 2024).

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene ontology: tool for the unification of biology. Nat. Genet. 25 (1), 25–29. https://doi.org/10.1038/75556.

Assis, C.B., Branco, G.S., Moreira, R.G., Pinheiro, J.P.S., 2021. Aluminum at environmental concentrations affects the sperm quality of the freshwater teleost *Astyanax altiparanae*: an in vitro study. Comp. Biochem. Physiol. C Toxicol Pharmacol. 243, 108983. https://doi.org/10.1016/j.cbpc.2021.108983.

Batista da Silva, I., Aciole Barbosa, D., Kavalco, K.F., Nunes, L.R., Pasa, R., Menegidio, F. B., 2023. Discovery of putative long non-coding RNAs expressed in the eyes of *Astyanax mexicanus* (Actinopterygii: Characidae). Sci. Rep. 13 (1), 12051. https://doi.org/10.1038/s41598-023-34198-5.

Bauersachs, S.; Wolf, E. 2012. Transcriptome analyses of bovine, porcine, and equine endometrium during the pre-implantation phase. Anim. Reprod. Sci., 134, 84–94. http://doi:https://doi.org/10.1016/j.anireprosci.2012.08.015.

Bennemann, S.T., Gealh, A.M., Orsi, M.L., Souza, L.M., 2005. Ocorrência e ecologia trófica de quatro espécies de *Astyanax* (Characidae) em diferentes rios da bacia do rio Tibagi, Paraná, Brasil. Iheringia, Sér. Zool. 95 (3), 247–254. https://doi.org/10.1590/S0073-47212005000300004.

Boursnell, C., 2024. Transfuse. https://github.com/cboursnell/transfuse (accessed 01 November 2024).

Brambila-Souza, G., Mylonas, C.C., De Mello, P.H., Kuradomi, R.Y., Batlouni, S.R., Tolussi, C.E., Moreira, R.G., 2021. GnRHa implants and temperature modulate Lambari-do-rabo-amarelo, *Astyanax altiparanae* (Characiformes: Characidae) induced reproduction out of the reproductive season. Aquac. Res. 0, 1–11. https://doi.org/10.1111/are.15435.

Branco, G.S., Melo, A.G., Ricci, J.M.B., Digmayer, M., Jesus, L.W.O., Habibi, H.R., Nóbrega, R.H., 2019. Effects of GnRH and the dual regulatory actions of GnIH in the pituitary explants and brain slices of *Astyanax altiparanae* males. Gen. Comp. Endocrinol. 273, 209–217. https://doi.org/10.1016/j.ygcen.2018.08.006.

Branco, G.S., Moreira, R.G., Borella, M.I., Camargo, M.P., Muñoz-Peñuela, M., Gomes, A. D., Tolussi, C.E., 2021. Nonsteroidal anti-inflammatory drugs act as endocrine disruptors in *Astyanax lacustris* (Teleostei: Characidae) reproduction: an *ex vivo* approach. Aquat. Toxicol. 232, 105767. https://doi.org/10.1016/j.aquatox.2021.105767.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC bioinformatics 10, 1–9. https://doi.org/10.1186/1471-2105-10-421.

Camargo, M.P., Cassel, M., Jesus, L.W.O., Nóbrega, R.H., Borella, M.I., 2017. Characterization of undifferentiated spermatogonia and the spermatogonial niche in the Lambari fish *Astyanax altiparanae*. Theriogenology 96, 97–102. https://doi.org/10.1016/j.theriogenology.2017.03.027.

Cassel, M., Chehade, C., Branco, G.S., Caneppele, D., Romagosa, E., Borella, M.I., 2017. Ovarian development and reproductive profile of Astyanax altiparanae (Teleostei, Characidae) over a year and its application in fish farming. Theriogenology 98, 1–15. https://doi.org/10.1016/j.theriogenology.2017.04.044.

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., Mcpherson, J., Dipert, A., Borges, B., 2024. Shiny: web application framework for R. R package version 1.9.1.9000. https://github.com/rstudio/shiny (web archive link, 1 November 2024). https://shiny.posit.co/ (accessed 01 November 2024).

Chehade, C., Cassel, M., Borella, M.I., 2015. Induced reproduction in a migratory teleost species by water level drawdown. Neotrop. Ichthyol. 13, 205–212. https://doi.org/10.1590/1982-0224-20140028.

Chehade, C., Amaral, F.G., Branco, G.S., Cassel, M., Jesus, L.W.O., Costa, F.G., Bordin, S. A., Moreira, R.G., Borella, M.I., 2020. Molecular characterization of different preproGnRHs in *Astyanax altiparanae* (Characiformes): effects of GnRH on female reproduction. Mol. Reprod. Dev. 87, 720–734. https://doi.org/10.1002/mrd.23351.

Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34 (17), i884–i890.

Cock, P.J., Chilton, J.M., Grüning, B., Johnson, J.E., Soranzo, N., 2015. NCBI BLAST+ integrated into galaxy. Gigascience 4 (1) s13742–015.

Costa, F.G., Adolfi, M.C., Gomes, C.C., Jesus, L.W.O., Batlouni, S.R., Borella, M.I., 2014. Testes of *Astyanax altiparanae*: the Sertoli cell functions in a semicystic spermatogenesis. Micron 61, 20–27. https://doi.org/10.1016/j.micron.2014.02.004.

Ewels, P., Magnusson, M., Lundin, S., Käller, M., 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32 (19), 3047–3048.

Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28 (23), 3150–3152.

Garutti, V., Britski, H.A., 2000. Descrição de uma espécie nova de *Astyanax* (Teleostei: Characidae) da bacia do alto Paraná River e considerações sobre as demais espécies do gênero. Comun. Mus. Ciênc. Tecnol. PUCRS. Sér. Zool Porto Alegre. 13, 65–88.

Gene Ontology Consortium, 2019. The gene ontology resource: 20 years and still GOing strong. Nucleic Acids Res. 47 (D1), D330–D338.

Godoi, F.G.A., Muñoz-Peñuela, M., Gomes, A.D.O., Tolussi, C.E., Brambila-Souza, G., Branco, G.S., Lo Nostro, F.L., Moreira, R.G., 2020. Endocrine disruptive action of diclofenac and caffeine on *Astyanax altiparanae* males (Teleostei: Characiformes: Characidae). Comp. Biochem. Physiol. C Toxicol Pharmacol. 231, 108720. https://doi.org/10.1016/j.cbpc.2020.108720.

Godoi, F.G.A., Dias, M.A., Guerreiro, A.S., Branco, G.S., Montagner, C.C., Moreira, R.G., Lo Nostro, F.L., 2024. Physiological responses on the reproductive, metabolism and stress endpoints of *Astyanax lacustris* females (Teleostei: Characiformes) after diclofenac and ibuprofen exposure. Comp. Biochem. Physiol. C Toxicol Pharmacol. 278, 109846. https://doi.org/10.1016/j.cbpc.2024.109846.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. Nat. Protoc. 8 (8), 1494–1512.

Hart, A.J., Ginzburg, S., Xu, M., Fisher, C.R., Rahmatpour, N., Mitton, J.B., Paul, R., Wegrzyn, J.L., 2020. EnTAP: bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. Mol. Ecol. Resour. 20 (2), 591–604.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., Rattei, T., Mende, D.R., Sunagawa, S., Kuhn, M., Jensen, L.J., von Mering, C., Bork, P., 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 44 (D1), D286–D293.

Jalili, V., Afgan, E., Gu, Q., Clements, D., Blankenberg, D., Goecks, J., Taylor, J., Nekrutenko, A., 2020. The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. Nucleic Acids Res. 48 (W1), W395–W402.

Jesus, L.W.O., Bogerd, J., Vieceli, F.M., Branco, G.S., Camargo, M.P., Cassel, M., Moreira, R.G., Yan, C.Y.I., Borella, M.I., 2017. Gonadotropin subunits of the characiform *Astyanax altiparanae:* molecular characterization, spatiotemporal expression and their possible role on female reproductive dysfunction in captivity. Gen. Comp. Endocrinol. 246, 150–163. https://doi.org/10.1016/j.ygcen.2016.12.004.

Kanehisa, M., Goto, S., 2000. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28 (1), 27–30.

Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., Tanabe, M., 2019. New approach for understanding genome variations in KEGG. Nucleic Acids Res. 47 (D1), D590–D595.

Kida, B.M.S., Abdalla, R.P., Moreira, R.G., 2016. Effects of acidic water, aluminum, and manganese on testicular steroidogenesis in Astyanax altiparanae. Fish Physiol. Biochem. 42, 1347–1356. https://doi.org/10.1007/s10695-016-0222-6.

Kriventseva, E.V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F.A., Zdobnov, E.M., 2019. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Res. 47 (D1), D807–D811.

Leaver, M.J., Bautista, J.M., Björnsson, L.R., Jönsson, E., Krey, G., Tocher, D.R., Torstensen, B.E., 2015. Towards fish lipid nutrigenomics: current state and prospects for fin-fish aquaculture. Rev. Fish. Sci. 56, 1064–1262.

Lucena, C.A.S., Soares, H.G., 2016. Review of species of the *Astyanax bimaculatus* "caudal peduncle spot" subgroup sensu Garutti & Langeani (Characiformes, Characidae) from the rio La Plata and rio São Francisco drainages and coastal systems of southern Brazil and Uruguay. Zootaxa 4072 (1), 101–125. https://doi.org/10.11646/zootaxa.4072.1.5.

Lütken, C.F., 1875. Velha-Flodens Fiske: Et Bidrag til Brasiliens Ichthyologie [Peixes do rio das Velhas: uma contribuição para a ictiologia do Brasil]. In: Alves, C.B.M., Pompeu, P.S. (Eds.), Peixes do Rio das Velhas: passado e presente. SEGRAC, Belo Horizonte, pp. 23–164.

Machado-Evangelista, M., Sussel, F.R., Romagosa, E., 2019. Environmental manipulation on *Astyanax altiparanae* out-of-season spawning. Lat. Am. J. Aquat. Res. 47 (2), 292–302. https://doi.org/10.3856/vol47-issue2-fulltext-9.

Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., Zdobnov, E.M., 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol. Biol. Evol. 38 (10), 4647–4654.

Mazurais, D., Darias, M., Zambonino-Infante, J.L., Cahu, C.L., 2011. Transcriptomics for understanding marine fish larval development. Can. J. Zool. 89 (7), 599–611. https://doi.org/10.1139/z11-036.

Menegidio, F.B., Jabes, D.L., Costa de Oliveira, R., Nunes, L.R., 2018. Dugong: a Docker image, based on Ubuntu Linux, focused on reproducibility and replicability for bioinformatics analyses. Bioinformatics 34 (3), 514–515.

Nakatani, K., 2001. Ovos e larvas de peixes de água doce: desenvolvimento e manual de identificação. In: UEM Eletrobrás, Maringá.

O'Quin, K., Mcgaugh, S., 2016. Mapping the Genetic Basis of Troglomorphy in Astyanax: How Far we Have Come and where Do we Go from here?.

Peretti, D., Andrian, I.F., 2008. Feeding and morphological analysis of the digestive tract of four species of fish (*Astyanax altiparanae, Parauchenipterus galeatus, Serrasalmus marginatus and Hoplias aff. Malabaricus*) from the upper Paraná River floodplain, Brazil. Braz. J. Biol. 68 (3), 671–679.

Pinheiro, J.P.S., Lima, J., Assis, C.B., Branco, G.S., Gomes, A.D., Moreira, R.G., 2021. Paternal exposure to aluminum, acidity, and temperature affect fatty acid seminal profile, embryonic and larval development of *Astyanax altiparanae*. Chemosphere 266, 128935. https://doi.org/10.1016/j.chemosphere.2020.128935.

Porto-Foresti, F., Castilho-Almeida, R.B., Foresti, F., 2005. Biologia e criação do lambari-do-rabo-amarelo (*Astyanax altiparanae*). In: Baldisseroto, B., Gomes, L.C. (Eds.), Espécies nativas para piscicultura no Brasil - Editora UFSM, Santa Maria, pp. 101–116.

Postingel Quirino, P., Da Silva Rodrigues, M., Da Silva Cabral, E.M., Siqueira-Silva, D.H., Mori, R.H., Butzge, A.J., Nóbrega, R.H., Ninhaus-Silveira, A., Veríssimo-Silveira, R., 2021. The influence of increased water temperature on the duration of spermatogenesis in a neotropical fish, *Astyanax altiparanae* (Characiformes, Characidae). Fish Physiol. Biochem. 47 (3), 747–755. https://doi.org/10.1007/s10695-020-00869-7.

Qian, L., Qian, W., Snowdon, R.J., 2014. Sub-genomic selection patterns as a signature of breeding in allopolyploid Brassica napus genome. BMC Genomics 15, 1170.

R Core Team., 2024. _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (accessed 01 November 2024).

Rodrigues, M.S., Siqueira-Silva, D.H., Quirino, P.P., Ninhaus-Silveira, A., Veríssimo-Silveira, R., 2015. Spermatogenesis in the yellowtail tetra *Astyanax altiparanae*: a histological analyses with emphasis to spermatogonial and spermatid types. B. Inst. Pesca. 41, 697–705. https://doi.org/10.20950/1678-2305.2015v41nep697.

Santos, M.P., Yasui, G.S., Xavier, P.L.P., Adamov, N.S.M., Nascimento, N.F., Fujimoto, T., Senhorini, J., Nakaghi, L.S.O., 2016. Morphology of gametes, post-fertilization events and the effect of temperature on the embryonic development of *Astyanax altiparanae* (Teleostei, Characidae). Zygote 24, 795–807. https://doi.org/10.1017/S0967199416000101.

Schulz, U., Martins-Junior, H., 2001. *Astyanax fasciatus* as a bioindicator of water pollution in Rio dos Sinos, RS, Brazil. Braz. J. Biol. 61 (4), 615–622. https://doi.org/10.1590/S1519-69842001000400010.

Shen, W., Le, S., Li, Y., Hu, F., 2016. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS One 11 (10), e0163962. https://doi.org/10.1371/journal.pone.0163962.

Siqueira-Silva, D.H., Bertolini, R.M., Levy-Pereira, N., Nascimento, N.F., Senhorini, J.A., Piva, L.H., Ferraz, J.B.S., Yasui, G.S., 2020. Factors affecting secondary sex characteristics in the yellowtail tetra, *Astyanax altiparanae*. Fish Physiol. Biochem. 47, 737–746. https://doi.org/10.1007/s10695-020-00832-6.

Sudhagar, A., Kumar, G., El-Matbouli, M., 2018. Transcriptome analysis based on RNA-Seq in understanding pathogenic mechanisms of diseases and the immune system of fish: a comprehensive review. Int. J. Mol. Sci. 19. https://doi.org/10.3390/ijms19010245.

Wang, F., Liu, F., Chen, W., Xu, R., Wang, W., 2018. Effects of triclosan (TCS) on hormonal balance and genes of hypothalamus-pituitary-gonad axis of juvenile male Yellow River carp (*Cyprinus carpio*). Chemosphere 193, 695–701. https://doi.org/10.1016/j.chemosphere.2017.11.088.

Warner, J., 2024. Shiny_BLAST. 2024. https://github.com/ScientistJake/Shiny_BLAST (accessed 01 November 2024).