Supplementary File S2. Draft assembly and annotation of genome of *Cichlidogyrus casuarinus*

DNA extraction and library preparation followed procedures previoulsy described(Leeming et al. 2023). In brief, genomic DNA was extracted using the Quick-DNA™ Miniprep Plus Kit (Zymo Research, Irvine, CA, USA) following the manufacturer's instructions with minor modifications, specifically, initial incubation overnight, and elution in 2 × 50 µL after 10 min incubation at room temperature each. DNA quantity was assessed using a Qubit 4.0 fluorometer and the Qubit dsDNA BR Assay. DNA integrity was assessed on an Agilent TapeStation system. Library preparation (Illumina Nextera XT, 550 bp insert size) and sequencing on the NovaSeq6000 (2× 150 bp) platform were outsourced (Macrogen Europe, The Netherlands).

Illumina reads were quality checked using FastQC and adapter and quality trimmed using Trimgalore v. 0.6.0 (https://github.com/FelixKrueger/TrimGalore; accessed 11 Apr 2024), which employs Cutadapt (Martin 2011) for adapter trimming. Subsequently, error correction was performed using correction module of SPAdes v3.14.0 (Prjibelski et al. 2020). Read pairs were merged using USEARCH v11.0.667_i86linux32 (Edgar 2010). Genome assemblies were performed using different sets of reads, specifically, trimmed, corrected, and merged (with trimmed and corrected reads), using SPAdes, AbySS v2.2.5 (Simpson et al. 2009), and Platanus v1.2.4 (Kajitani et al. 2014). The kmer lengths used for ABySS were selected using KmerGenie v1.7.051 (Chikhi & Medvedev 2014). The contiguity of all assemblies was assessed using Quast-LG v5.0.2 (Mikheenko et al. 2018). Assembly completeness was evaluated using BUSCO v. 5.2.1_cv1 (Manni et al. 2021) (dataset eukaryota_odb10) and potential host contamination was assessed using BlobTools v1.1.1 (Laetsch & Blaxter 2017). All assemblies were then compared with respect to contiguity (N50), BUSCO completeness and contamination. The assembly result obtained with Platanus using corrected reads merged with usearch was selected as the best for subsequent analyses based on these three criteria. The entire process (data trimming, correction, merging, assembly, assembly evaluation) was run through the workflow demogenas (https://github.com/chrishah/demogenas; accessed 11 Apr 2024) implemented with Snakemake (Köster & Rahmann 2012).

The draft genome of *Cichlidogyrus casuarinus* was annotated following a strategy previoulsy described (see Vorel et al. (2023)), with some modifications. In brief, core eukaryotic genes were identified in the final assembly using CEGMA v2.5 (Parra et al. 2007) and BUSCO v3.0.2 (Simão et al. 2015) (Metazoa dataset, odb9, 978 searched groups). The latter was run with the *optimize_augustus* option to train the AUGUSTUS v3.3.3 *ab initio* gene predictor (Stanke et al. 2006) in the process. Genes identified by CEGMA were used to train the SNAP v2006–07-28 (Korf 2004) *ab initio* gene predictor. Species-specific repeats were identified using RepeatModeler v1.0.10 (Flynn et al. 2020). RepeatMasker v4.0.7 (Smit et al. 2023) was then run to mask repetitive regions, using 1) the de novo library identified in the previous step, and 2) using a prebuilt repeat library (RepBaseRepeatMaskerEdition-20181026) with species set to *eukaryota*. *Ab-initio* gene predictor Genemark-ES (Lomsadze et al. 2005) (*gmes_petap.pl*) v4.69_lic was run on the repeat soft-masked genome. As protein evidence that would further inform downstream gene prediction, we concatenated the complete UniProt/Swiss-Prot protein database (Bateman et al. 2023) (release 2022_01) and 33 available protein complements of parasitic flatworms downloaded from the NCBI

GenBank (Sayers et al. 2022) and WormBase ParaSite databases (Howe et al. 2017) (accessed 2 Feb 2022). To remove redundancy in the reference protein set, it was clustered at 98% similarity using CD-HIT (Fu et al. 2012) v4.8.1. Further, gene prediction was performed in two passes: First, using MAKER2 (Holt & Yandell 2011) v2.31.10 on the repeat masked genome, based on the physical protein (see above), and using the gene models obtained with SNAP (see above). Gene models of the first MAKER pass (only genes with evidence score < 0.1) were used to retrain the AUGUSTUS and SNAP ab-initio predictors. In a second pass, MAKER2 was rerun combining all evidence and using AUGUSTUS, Genemark, and SNAP and their pre-trained models. Subsequently we ran the *predict* Funannotate v1.8.7 (https://github.com/nextgenusfs/funannotate, accessed 31 Jan 2023) with AUGUSTUS, SNAP, and GlimmerHMM (Fu et al. 2012), incorporating the gene models initially predicted with Genemark and predictions obtained via the two passes of MAKER (weight 2). The resulting set of gene predictions was functionally annotated using the annotation module of Funannotate, combining the results from InterProScan (Jones et al. 2014) v5.48–83.0 with a similarity search against databases UniProt/Swiss-Prot (release 2022_01), MEROPS(Rawlings et al. 2018) (database of proteolytic enzymes and inhibitors, release 12.0), and Phobius (Käll et al. 2007) using search tool DIAMOND (Buchfink et al. 2021) v2.0.7 (BLASTp algorithm) and with a search against the complete eggNOG 5.0 database (Huerta-Cepas et al. 2019) conducted with the eggNOG-mapper (Huerta-Cepas et al. 2017) (emapper.py) v. 1.0.3. The entire prediction and annotation process as described above was run through Annocomba (https://github.com/reslp/annocomba, accessed 31 Jan 2023), which uses the Snakemake workflow management system (Köster & Rahmann 2012).

References

Bateman A et al. 2023. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res*. 51:D523–D531. https://www.doi.org/10.1093/nar/gkac1052.

Buchfink B, Reuter K, Drost HG. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods*. 18:366–368. https://www.doi.org/10.1038/s41592-021-01101-x.

Chikhi R, Medvedev P. 2014. Informed and automated k-mer size selection for genome assembly. *Bioinformatics*. 30:31–37. https://www.doi.org/10.1093/bioinformatics/btt310.

Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 26:2460–2461. https://www.doi.org/10.1093/bioinformatics/btq461.

Flynn JM et al. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A*. 117:9451–9457. https://www.doi.org/10.1073/pnas.1921046117.

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 28:3150–3152. https://www.doi.org/10.1093/bioinformatics/bts565.

Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*. 12. https://www.doi.org/10.1186/1471-2105-12-491.

Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. 2017. WormBase ParaSite − a comprehensive resource for helminth genomics. *Mol Biochem Parasitol*. 215:2–10. https://www.doi.org/10.1016/j.molbiopara.2016.11.005.

Huerta-Cepas J et al. 2019. EggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res*. 47:D309–D314. https://www.doi.org/10.1093/nar/gky1085.

Huerta-Cepas J et al. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol Biol Evol*. 34:2115–2122. https://www.doi.org/10.1093/molbev/msx148.

Jones P et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 30:1236–1240. https://www.doi.org/10.1093/bioinformatics/btu031.

Kajitani R et al. 2014. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res*. 24:1384–1395. https://www.doi.org/10.1101/gr.170720.113.

Käll L, Krogh A, Sonnhammer ELL. 2007. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res*. 35:W429–W432. https://www.doi.org/10.1093/nar/gkm256.

Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics*. 5:59. https://www.doi.org/10.1186/1471-2105-5-59.

Köster J, Rahmann S. 2012. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*. 28:2520–2522. https://www.doi.org/10.1093/bioinformatics/bts480.

Laetsch DR, Blaxter ML. 2017. BlobTools: interrogation of genome assemblies. *F1000Res*. 6:1287. https://www.doi.org/10.12688/f1000research.12232.1.

Leeming SJ et al. 2023. Amended diagnosis, mitochondrial genome, and phylogenetic position of *Sphyranura euryceae* (Neodermata, Monogenea, Polystomatidae), a parasite of the Oklahoma salamander. *Parasite*. 30:27. https://www.doi.org/10.1051/parasite/2023025.

Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. 2005. Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res*. 33:6494–6506. https://www.doi.org/10.1093/nar/gki937.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 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*. 38:4647–4654. https://www.doi.org/10.1093/molbev/msab199.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 17:10–12. https://www.doi.org/10.14806/ej.17.1.200.

Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics*. 34:i142–i150. https://www.doi.org/10.1093/bioinformatics/bty266.

Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*. 23:1061–1067. https://www.doi.org/10.1093/bioinformatics/btm071.

Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes De Novo Assembler. *Curr Protoc Bioinformatics*. 70:e102. https://www.doi.org/10.1002/cpbi.102.

Rawlings ND et al. 2018. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res*. 46:D624–D632. https://www.doi.org/10.1093/nar/gkx1134.

Sayers EW et al. 2022. Database resources of the national center for biotechnology information. *Nucleic Acids Res*. 50:D20–D26. https://www.doi.org/10.1093/nar/gkab1112.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 31:3210–3212. https://www.doi.org/10.1093/bioinformatics/btv351.

Simpson JT et al. 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res*. 19:1117–1123. https://www.doi.org/10.1101/gr.089532.108.

Smit AFA, Hubley R, Green P. 2023. RepeatMasker Open-4.0. http://www.repeatmasker.org/ (Accessed January 31, 2023).

Stanke M et al. 2006. AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Res*. 34:W435–W439. https://www.doi.org/10.1093/nar/gkl200.

Vorel J, Kmentová N, Hahn C, Bureš P, Kašný M. 2023. An insight into the functional genomics and species classification of *Eudiplozoon nipponicum* (Monogenea, Diplozoidae), a haematophagous parasite of the common carp *Cyprinus carpio*. *BMC Genomics*. 24:363. https://www.doi.org/10.1186/s12864-023-09461-8.