

# De Novo Genome Assembly of the Electric Fish *Brachyhypopomus occidentalis* (Hypopomidae, Gymnotiformes)

Carlos F. Arias<sup>1,2,3,\*</sup>, Rebecca B. Dikow<sup>2</sup>, W. Owen McMillan<sup>3</sup>, and Luis F. De León<sup>1,3,4</sup>

<sup>1</sup>Department of Biology, University of Massachusetts, Boston, Massachusetts, USA

<sup>2</sup>Data Science Lab, Office of the Chief Information Officer, Smithsonian Institution, Washington, District of Columbia, USA

<sup>3</sup>Smithsonian Tropical Research Institute, Panamá, Panamá

<sup>4</sup>Instituto de Investigaciones Científicas y Servicios de Alta Tecnología (INDICASAT-AIP), Panamá, Panamá

\*Corresponding author: E-mail: [ariasc@si.edu](mailto:ariasc@si.edu).

Accepted: 21 September 2021

## Abstract

The bluntnose knifefish *Brachyhypopomus occidentalis* is a primary freshwater fish from north-western South America and Lower Central America. Like other Gymnotiformes, it has an electric organ that generates electric discharges used for both communication and electrolocation. We assembled a high-quality reference genome sequence of *B. occidentalis* by combining Oxford Nanopore and 10X Genomics linked-reads technologies. We also describe its demographic history in the context of the rise of the Isthmus of Panama. The size of the assembled genome is 540.3 Mb with an N50 scaffold length of 5.4 Mb, which includes 93.8% complete, 0.7% fragmented, and 5.5% of missing vertebrate/Actinoterigie Benchmarking Universal Single-Copy Orthologs. Repetitive elements account for 11.04% of the genome, and 34,347 protein-coding genes were predicted, of which 23,935 have been functionally annotated. Demographic analysis suggests a rapid effective population expansion between 3 and 5 Myr, corresponding to the final closure of the Isthmus of Panama (2.8–3.5 Myr). This event was followed by a sudden and constant population decline during the last 1 Myr, likely associated with strong shifts in both precipitation and sea level during the Pleistocene glacial-interglacial cycles. The de novo genome assembly of *B. occidentalis* will provide novel insights into the molecular basis of both electric signal productions and detection and will be fundamental for understanding the processes that have shaped the diversity of Neotropical freshwater environments.

**Key words:** Gymnotiformes, hybrid genome assembly, electric organ discharges (EODs), demographic history, ONT Nanopore, 10X Genomics linked-reads technologies.

## Significance

Gymnotiformes are a renowned group of Neotropical freshwater fish due to their ability to produce and detect electric organ discharges used for communication and electrolocation. Electric organs are believed to be associated with the diversification of the nearly 250 species of electric fishes of South America and Central America. Unfortunately, despite this extreme diversity, limited genomic resources are currently available for the group. Here, we present the genome of *Brachyhypopomus occidentalis* and describe its demographic history in the context of the rise of the Isthmus of Panama. Our high-quality genome assembly and annotation will provide new insights into the genes involved in electric signal production and detection. This genome assembly will facilitate future genome-wide studies to understand the patterns and processes driving diversification in Neotropical freshwater environments.

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

## Introduction

Neotropical electric fishes (Teleostei, Gymnotiformes) are a highly diverse group of nocturnal, primarily freshwater fishes with close to 250 recognized species that originated in South America (Crampton 2019). A prominent feature of Gymnotiformes is their ability to produce and detect electric organ discharges (EODs), which are used for communication and electrolocation (Hopkins 2009). The evolutionary history of EODs is particularly interesting because the electric organs that produce these discharges have evolved multiple times in independent fish lineages (Gallant et al. 2014). Moreover, within Gymnotiformes, there is species-specific variation in the type of electric signal, which depends on the shape and regularity of the discharge (Fugère and Krahe 2010). Despite the high diversity of Neotropical electric fishes, limited genomic resources are currently available for the group. In fact, to date, the only sequence of a Gymnotiform genome available is that of the electric eel (*Electrophorus electricus*; Gallant et al. 2014). Electric eels are unique among Gymnotiformes because they have evolved three distinct electric organs and can emit distinct high, middle, and low voltage discharges (Albert and Crampton 2006; Xu et al. 2021). Despite the historical interest in the electric eels, mainly because of their extremely high electric discharges, they represent a species-poor clade with only three species currently known to science (de Santana et al. 2019). Thus, generating genomic resources for more diverse Gymnotiformes with contrasting evolutionary and life histories, as well as different types of EODs is crucial to improving our understanding of the evolution of this hyperdiverse group of freshwater fishes.

Here, we assembled the genome of the Neotropical bluntnose knifefish *Brachyhypopomus occidentalis* (fig. 1A; Hypopomidae, Gymnotiformes; Regan, 1914) de novo. *Brachyhypopomus occidentalis* generates a low voltage, sexually dimorphic, biphasic pulse-type EOD (Hagedorn 1995) and is found in slow-flowing streams from north-western South America to the Sixaola River of Costa Rica (fig. 1A; Bussing 1976). This weakly electric fish is one of the few species of Gymnotiformes that have successfully colonized Lower Central America (LCA; Picq et al. 2014), representing an excellent study system to infer the processes (e.g., drift, gene flow, natural selection) shaping genome evolution during colonization of new habitats. Here, the generation of new genomic resources will provide a deeper insight into the population history of this range expansion, including testing for the role of selection and drift in the context of the formation of the Central American land bridge and also the processes that have shaped diversification in Neotropical freshwater environments.

## Results and Discussion

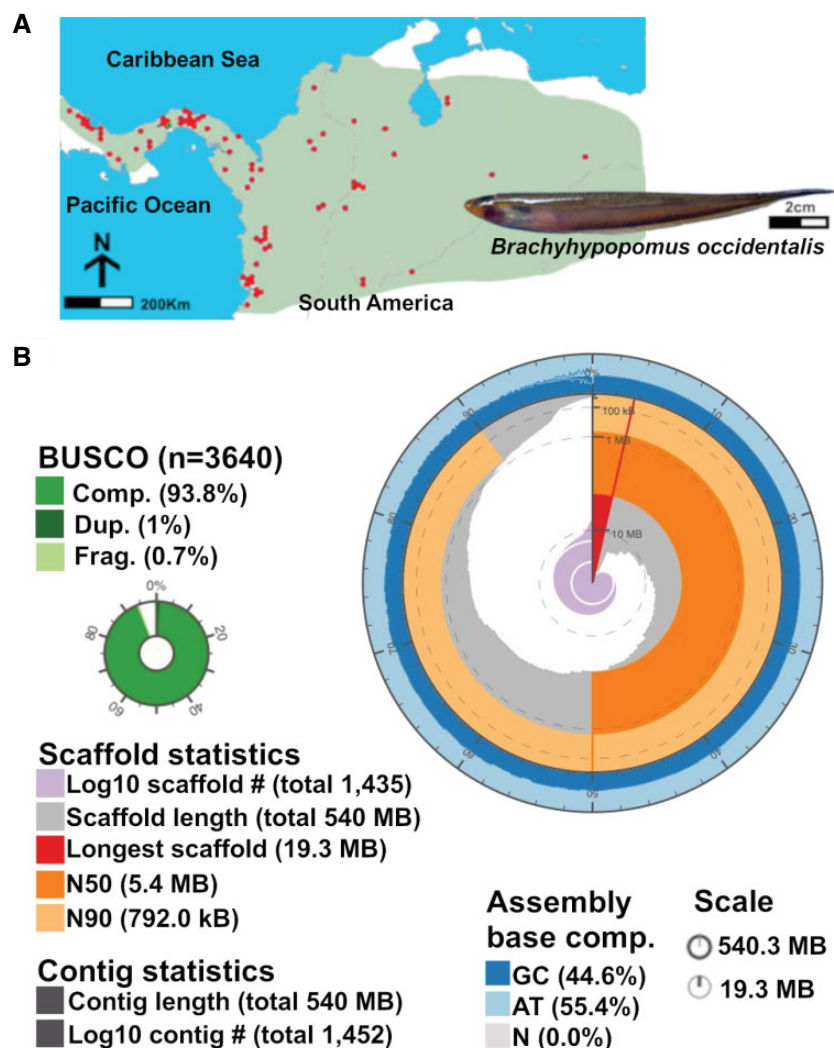
### Genome Assembly, Heterozygosity, and Size

We generated a total of ~4.7 million ONT reads, which constituted ~48 Gb of sequencing data with an average read

length of ~5.3 kb. The longest read was 136.08 kb, and N50 was 7.05 kb. In addition, we generated a total of ~651.6 million paired-end 10X reads, which produced ~120 Gb of sequencing data, with an average cleaned read length of 148.5 bp (supplementary table S1, Supplementary Material online). These data sets represent an approximate genome coverage of 46× and 160×, respectively, based on our final genome assembly. GenomeScope estimated a genome size of 647.8 Mbp with ~69.7% of unique content and a heterozygosity level of 0.32% (supplementary fig. S1, Supplementary Material online). Genome evaluated with Kraken2 resulted in the removal of 3.9% of the assembled scaffolds. Our final assembled genome was contained in 1,435 scaffolds with a total length of 540.3 Mb, a GC content of 44.6%, and contig and scaffold N50 of 5.4 and 5.1 Mb, respectively (fig. 1B; supplementary table S2, Supplementary Material online). Out of 3,640 Actinopterygii orthologs screened, we retrieved 3,414 complete (93.8%) and 27 fragmented (0.7%) Benchmarking Universal Single-Copy Orthologs (BUSCOs). A total of 99 BUSCOs (5.5%) were missing (fig. 1B).

### Genome Annotation

RepeatMasker estimated that 11.23% of the genome consisted of repetitive sequences, primarily LINES (2.2%), LTR elements (1.585%), DNA transposons (3.08%), and simple repeats (3.6%; supplementary table S3, Supplementary Material online). Repeat content was nearly identical to the electric eel (*E. electricus*), both in terms of total repeat content and the proportion of each type of repeat (supplementary table S3, Supplementary Material online). MAKER identified 25,023 gene models and 170,711 exons. Blast2GO assigned putative functional predictions to 80.4% (20,120) of the predicted protein-coding genes. In contrast, GeMoMa identified 34,347 gene models and 369,936 exons, where 69.7% (23,935) of the gene models had a functional annotation (supplementary table S4, Supplementary Material online). The number of coding sequences identified for *B. occidentalis* was similar to those of the electric eel (*E. electricus*) and within the range of those found in other fish species (supplementary table S4, Supplementary Material online). Furthermore, we explored if known physiological, developmental, and anatomical candidate genes involved in the evolution of EODs were present in our genome assembly. In particular, Gallant et al. (2014) and Wang and Yang (2021), using transcriptome data from three independent lineages of electric fish (Gymnotiformes, Siluriformes, Mormyroidea), identified 31 candidate genes associated with a variety of functions in EODs (supplementary table S5, Supplementary Material online). Our pipeline successfully annotated all 31 candidate genes (supplementary table S5, Supplementary Material online), suggesting that the genes involved in EODs production are highly conserved across electric fishes. However, variation also exists. An interesting case is the voltage-gated sodium channel gene



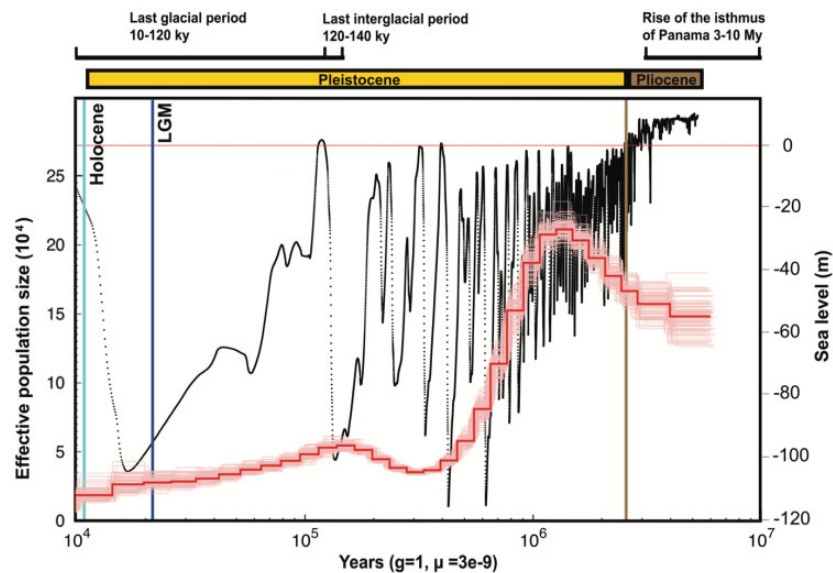
**FIG. 1.**—Distribution, and genome assembly statistics of the bluntnose knifefish. (A) Distribution (dark green) and georeferenced records (red dots; data download from [gbif.org](https://gbif.org) 2021) for *Brachyhyppopomus occidentalis* (Hypopomidae, Gymnotiformes; Regan, 1914). (B) Assembly statistic visualization (using script from <https://github.com/rjchallis/assembly-stats>) (Challis 2021) showing the genome scaffold N50 (dark orange), N90 (light orange), base composition (percentage of GC in dark blue, AT in light blue, and N in light grey), and completeness and quality of the genome implemented with BUSCO (in shades of green).

(*SCN4aa*), which is thought to have played a key role in the evolution of electric signal communication (Arnegard et al. 2010). Protein alignments between Gymnotiformes, Mormyroidea, other teleosts, and humans showed that the C-terminal domain of the *SCN4aa* gene is highly variable among species (Traeger et al. 2017). Overall, our assembled and annotated genome of *B. occidentalis* will facilitate functional comparative analyses, as well as the application of gene-editing techniques in electric fishes.

#### Ancestral Demographic Reconstructions

Pairwise Sequentially Markovian Coalescent (PSMC) estimates showed an increase in the effective population size for *B.*

*occidentalis* starting around 3–5 Myr. This population increase peaked between 1 and 2 Myr with a maximum of ~230,000 individuals. Afterward, there was a sudden population decrease during the last 0.5–1 Myr reaching a minimum of ~30,000 individuals. This event was followed by a slight increase in population size between 0.5 and 0.1 Myr, after which it has been declining slowly (fig. 2). Previous studies based on mtDNA time reconstructions have suggested that *B. occidentalis* colonization of the LCA occurred in two major waves. The first event likely happened during the late Miocene (8–12 Myr), whereas the second took place during the early Pliocene (3–7 Myr; Picq et al. 2014). The former event was supported by an early divergence of a population *B. occidentalis* from Bocas del Toro (Panama), which is likely a



**Fig. 2.**—Genome-wide demographic history in *Brachyhypopomus occidentalis* using PSMC. PSMC reconstruction of effective population size over time, estimated using a generation time of 1 year ( $g = 1$ ) and a mutation rate of  $\mu = 3 \times 10^{-9}$ , is shown as a solid red line. Red-shaded lines correspond to 100 bootstrap runs. The solid black line represents the global sea-level model for the last 5 Myr (de Boer et al. 2014), vertical brown line represents the beginning of the Pleistocene (2.6 Myr), vertical dark blue line denotes the last glacial maximum (LGM, 22 ky), vertical light blue line denotes the beginning of the Holocene (11 ky), and the orange horizontal line shows current sea level as 0 m.

remnant population of this first colonization episode. The latter event was supported by a clade composed of central Panama, western Panama, and Colombian populations. This clade contains short internodes, perhaps indicating a rapid expansion (Picq et al. 2014). Our genome assembly and demographic model comes from an individual collected in the Chagres drainage (Central Panama). Thus, our PSMC analysis may reflect the demographic history of the second colonization event. Similar to Picq et al. (2014), our PSMC results support a rapid population expansion in *B. occidentalis*, following the final closure of the Isthmus of Panama between 2.8 and 3.5 Myr (O’Dea et al. 2016)—a pattern that coincides with the colonization of LCA by many other freshwater fish (Reeves and Bermingham 2006; Aguilar et al. 2019).

The drastic geologic changes in LCA during the Pleistocene (11 ky to 2.6 Myr) have been accompanied by both pronounced climate and sea-level fluctuations, which can alter freshwater habitats, as well as species distribution. For instance, sea-level fluctuations can connect and disconnect drainages, facilitating dispersal, isolation, and new colonizations. In fact, during periods of low sea level during the Pleistocene, large portions of the eastern Pacific Ocean floor were exposed (Redwood 2020), likely facilitating the exchange of freshwater fishes between different watersheds. In contrast, during periods of high sea level, seawater intrusions likely isolated freshwater fish populations (Mondin et al. 2018). Consistent with these patterns, our data show an evident decline in population size during the last 1 Myr, suggesting that fluctuations associated with glacial-interglacial cycles

have strongly impacted *B. occidentalis* populations (fig. 2). A similar pattern was found in the electric fish *Sternopygus dardiensis* using complete mtDNA genomes. In this case, populations from eastern and western Panama showed strong divergence  $\sim 1.1$  Myr and very low genetic variation, consistent with a recent population decline (Aguilar et al. 2019). Changes in precipitation have also likely played an important role in the population fluctuations in *B. occidentalis*. Drier conditions during Pleistocene may have affected habitat conditions for *B. occidentalis*, resulting in low population sizes, higher local extinction rates, and restricted possibilities for dispersal. Indeed, dry conditions prevailed in LCA during Pleistocene (González et al. 2006; Piperno 2006). Thus, the combined effect of drier conditions and changes in sea level are likely important drivers of demographic history in *B. occidentalis* in LCA. Overall, our genome-wide demographic model for *B. occidentalis* suggests a complex paleogeographic history of colonization and extinctions in LCA. It also highlights the role of the rise of the Isthmus of Panama in shaping genomic diversity and structure Neotropical freshwater fishes.

This genome offers a unique opportunity for investigating the demographic history and the history of colonization and diversification of *B. occidentalis* in LCA. The assembled genome will also improve our understanding of the evolution of genes involved in both electric signal productions and detection in Neotropical electric fishes. Furthermore, *B. occidentalis* gene models and sequences will facilitate the manipulation of genes and gene products, enabling future research on the role of variability in the production of electric signal diversity.

## Materials and Methods

### Specimen, Library Construction, and Sequencing

A live field-caught individual of *B. occidentalis* was collected from the Juan Grande River on Pipeline road (Panama) and stored at  $-80^{\circ}\text{C}$  at the Smithsonian Tropical Research Institute (STRI; fig. 1A). High molecular weight (HMW) DNA was extracted from frozen muscle using a phenol:chloroform method. Genomic DNA sequencing was performed by combining two different sequencing technologies: Oxford Nanopore (ONT, Oxford, United Kingdom) and 10X Genomics linked-reads (10X Genomics Chromium platform; Zheng et al. 2016). ONT libraries were prepared using the ligation sequencing kit SQK-LSK109 (ONT). Sequencing was performed on a MinION-Mk1b (ONT) using SpotON flow cells (FLO-MIN106; ONT) in 48-hour sequencing runs controlled by the MinKNOW software (r.19.06.8, ONT). Base-calling was performed using Guppy (v.3.3.0; ONT). Porechop v.0.2.3 (Wick 2018) and NanoFilt v.2.5 (De Coster et al. 2018) were used to remove adapters and filter low-quality reads.

The same individual HMW DNA sample used for the ONT libraries was also used for 10X reads library preparation. The prepared library was sequenced on an Illumina HiSeq X Ten platform (Illumina) to yield  $2 \times 150$  bp paired-end sequences. Library preparation and sequencing were performed at Génome Québec InnovationCentre (Canada). Raw 10X reads were checked with FastQC v.0.11.8 (Andrews 2010) and adapters trimmed with Trimmomatic v.0.36 (Bolger et al. 2014). Genome size, heterozygosity, repeat, and duplication content were estimated on clean 10X reads. K-mer counting was performed with Jellyfish v.2.2.6 (Marçais and Kingsford 2011) by generating a k-mer frequency distribution of 21-mers. The resulting histogram was later processed by GenomeScope (Vurture et al. 2017).

### Hybrid De Novo Genome Assembly

A hybrid genome assembly pipeline was used to combine both Oxford Nanopore (ONT) and 10X reads technologies (supplementary fig. S2, Supplementary Material online). First, we assembled a draft genome from ONT long reads by running Wtdbg2 v.2.2 (Ruan and Li 2020). ONT reads were used to polish the contigs by mapping the reads with Minimap2 (Li 2018) and by running the wtpoa-cns consensus command from Wtdbg2 v.2.2. This was followed by a round of polishing with the 10X reads, which were mapped to the assembly with BWA (Li and Durbin 2009). Similar to the polishing with ONT reads a consensus assembly was obtained with the wtpoa-cns command from Wtdbg2 v.2.2. Medium-range scaffolding was performed with 10X reads by using Scaff10X v.4.2 (WTSI-HPAG 2020). A further round of polishing was performed with both ONT data with Racon v.1.4.20 (Vaser et al. 2017) and 10X reads with Pilon v1.23 (Walker et al. 2014). Finally, we used Kraken v.2 (Wood and Salzberg

2014) to eliminate bacterial, viral, and plasmid contamination. Summary statistics for our genome assembly were generated with Assembly\_Stats v.0.14 (Trizna 2020). Genome completeness was assessed through BUSCOs v.4.0.6 (Simão et al. 2015), by comparing 3,640 orthologs contained in the actinopterygii\_odb10 database.

### Genome Annotation

We used two different pipelines for genome annotation. First, we used the MAKER v.2.31.9 (Holt and Yandell 2011) pipeline to predict protein-coding genes and structural variation. MAKER was run in a total of three cycles; the initial cycle was based on three types of hints: transcript evidence, protein sequences, and repetitive elements. Transcript evidence was obtained by aligning the closely related transcriptome of *Brachyhypopomus gauderio* (supplementary table S6, Supplementary Material online) to our genome assembly using BLAT (Kent 2002). Protein evidence was obtained by downloading known protein sequences for three fish species: *Amphiprion ocellaris*, *Oreochromis niloticus*, and *E. electricus* (supplementary table S6, Supplementary Material online). Repetitive elements were predicted by running RepeatMasker open-4.0.6 (Smith et al. 2013) with the Teleostei database to identify repetitive elements in the genome and soft-mask the assembly. RepeatMasker.out was converted to GFF with RepeatMasker script rmOutToGFF3.pl. The second and final cycles included gene models trained with the first (and then second) cycle with ab initio gene predictors SNAP (Korf 2004) and Augustus (Stanke et al. 2006). From the MAKER output, we extracted both the protein and nucleotide sequences of the gene models identified, as well as the individual coding sequences, using the AUGUSTUS script getAnnoFasta.pl. Furthermore, to infer the putative function of these predicted proteins, we queried them against the nonredundant protein database using Blast v2.6.0 (Altschul et al. 1990). Finally, InterProScan v.5.26.65 (Jones et al. 2014) was used to examine protein domains and motifs present in the predicted protein sequences. All the results were processed and summarized in Blast2GO v.5.2.5 (Götz et al. 2008).

Second, we used the GeneModelMapper pipeline v.1.6.1 (GeMoMa) to further predict gene models (Keilwagen et al. 2019). GeMoMa is a homology-based gene prediction program that takes advantage of known protein-coding genes models in reference genomes to predict possible protein-coding genes in a target genome (Keilwagen et al. 2019). Here, we ran GeMoMa using annotations from three fish species as reference genomes: *A. ocellaris*, *O. niloticus*, and *E. electricus* (supplementary table S6, Supplementary Material online), and transcript evidence from the closely related taxa *B. gauderio* (supplementary table S6, Supplementary Material online). Transcripts were mapped to our draft genome with Minimap2 (Li 2018) and used as input for GeMoMa (Keilwagen et al. 2019).

## Demographic History

We estimated historical effective population size in *B. occidentalis* using the PSMC v.0.6.5 (Li and Durbin 2011). A consensus sequence was generated using the vcfutils.pl script (with the following parameters vcf2fq -d 10 -D 100). PSMC model estimates were run with the options -N30 -t30 -r5 -p "4 + 30\*2 + 4 + 6 + 10." Bootstrapping was conducted by randomly sampling with replacement 5-Mb sequence segments during 100 bootstrap replicates. The reconstructed population history was plotted assuming a generation time of 1 year (Hagedorn 1988) and mutation rate of  $3.5 \times 10^{-9}$  substitutions per site per year following Malinsky et al. (2018; calculated in the tropical freshwater cichlid radiation).

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

## Acknowledgments

The authors thank Marta Vargas from STRI for all her help in the laboratory. All computing was conducted by the Smithsonian Institution High-Performance Cluster (SI/HPC), Smithsonian Institution. <https://doi.org/10.25572/SIHPC>. This work was supported by the Secretaría Nacional de Ciencia y Tecnología e Innovación, SENACYT, Panamá, (Grant: FID16-116 to LFD). C.F.A. was funded by Postdoctoral Fellowship from UMASS Boston and SI OClO Biodiversity Genomics fellowship. L.F.D.L. was also funded by the Department of Biology and the Ronald E. McNair Post-Baccalaureate Achievement Program at UMASS Boston.

## Data Availability

All raw data, genome assembly and Annotation have been deposited at NCBI under BioProject PRJNA727273 and Biosample SAMN19011640 (*B. occidentalis* BaOcci\_JG\_B10).

## Literature Cited

- Aguilar C, et al. 2019. Tempo and mode of allopatric divergence in the weakly electric fish *Sternopygus dardiensis* in the Isthmus of Panama. *Sci Rep.* 9(1):18828.
- Albert JS, Crampton WGR. 2006. Electroreception and electrogenesis. In: Evans DH, Claiborne JB, editors. *The Physiology of Fishes*. Boca Raton (FL): CRC Press, Taylor & Francis. p. 429–470.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215(3):403–410.
- Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available from: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Accessed September 29, 2021.
- Arnegard ME, Zwickl DJ, Lu Y, Zakon HH. 2010. Old gene duplication facilitates origin and diversification of an innovative communication system—twice. *Proc Natl Acad Sci USA.* 107(51):22172–22177.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.
- Bussing WA. 1976. Geographic distribution of the San Juan ichthyofauna of Central America with remarks on its origin and ecology. In: Thorson TB, editor. *Investigations of the ichthyofauna of Nicaraguan Lakes*. Lincoln (NE): University of Nebraska-Lincoln. p. 157–175.
- Challis, R. 2021. Assembly-stats. Available from: <https://github.com/rjchallis/assembly-stats>. Accessed September 29, 2021.
- Crampton WGR. 2019. Electroreception, electrogenesis and electric signal evolution. *J Fish Biol.* 95(1):92–134.
- de Boer B, Lourens LJ, van de Wal RSW. 2014. Persistent 400,000-year variability of Antarctic ice volume and the carbon cycle is revealed throughout the Plio-Pleistocene. *Nat Commun.* 5:2999. doi: 10.1038/ncomms3999.
- De Coster W, D’Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34(15):2666–2669.
- de Santana CD, et al. 2019. Unexpected species diversity in electric eels with a description of the strongest living bioelectricity generator. *Nat Commun.* 10(1):4000.
- Fugère V, Krahe R. 2010. Electric signals and species recognition in the wave-type gymnotiform fish *Apteronotus leptorhynchus*. *J Exp Biol.* 213(2):225–236.
- Gallant JR, et al. 2014. Genomic basis for the convergent evolution of electric organs. *Science* 344(6191):1522–1525.
- GBIF.org. 2021. GBIF Occurrence Download. Available from: <https://doi.org/10.15468/dl.j7gvv>. Accessed September 29, 2021.
- González C, Urrego LE, Martínez JI. 2006. Late Quaternary vegetation and climate change in the Panama Basin: palynological evidence from marine cores ODP 677B and TR 163-38. *Palaeogeogr Palaeoclimatol Palaeoecol.* 234(1):62–80.
- Götz S, et al. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36(10):3420–3435.
- Hagedorn M. 1988. Ecology and behavior of a pulse-type electric fish, *Hypopomus occidentalis* (Gymnotiformes, Hypopomidae), in a freshwater stream in Panama. *Copeia* 1988(2):324–335.
- Hagedorn M. 1995. The electric fish *Hypopomus occidentalis* can rapidly modulate the amplitude and duration of its electric organ discharges. *Anim Behav.* 49(5):1409–1413.
- Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491.
- Hopkins C. 2009. Electrical perception and communication. *Encycl Neurosci.* 3:813–831.
- Jones P, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30(9):1236–1240.
- Keilwagen J, Hartung F, Grau J. 2019. GeMoMa: homology-based gene prediction utilizing intron position conservation and RNA-seq data. In: Kollmar M, editor. *Gene Prediction. Methods in Molecular Biology*. vol. 1962. New York (NY): Humana Press. p. 161–177. Available from: 10.1007/978-1-4939-9173-0\_9
- Kent WJ. 2002. BLAT—the BLAST-like alignment tool. *Genome Res.* 12(4):656–664.
- Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5:59.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754–1760.
- Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475(7357):493–496.
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34(18):3094–3100.
- Malinsky M, et al. 2018. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat Ecol Evol.* 2(12):1940–1955.
- Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27(6):764–770.

- Mondin LA, et al. 2018. Genetic pattern and demographic history of *Salminus brasiliensis*: population expansion in the Pantanal Region during the Pleistocene. *Front Genet.* 9:1.
- O'Dea A, et al. 2016. Formation of the Isthmus of Panama. *Sci Adv.* 2(8):e1600883.
- Picq S, Alda F, Krahe R, Bermingham E. 2014. Miocene and Pliocene colonization of the Central American Isthmus by the weakly electric fish *Brachyhyopomus occidentalis* (Hypopomidae, Gymnotiformes). *J Biogeogr.* 41(8):1520–1532.
- Piperno DR. 2006. Quaternary environmental history and agricultural impact on vegetation in Central America. *Ann Mo Bot Gard.* 93(2):274–296.
- Redwood SD. 2020. Late Pleistocene to Holocene sea level rise in the Gulf of Panama, Panama, and its influence on early human migration through the Isthmus. *Caribb J Earth Sci.* 51:15–31.
- Reeves RG, Bermingham E. 2006. Colonization, population expansion, and lineage turnover: phylogeography of Mesoamerican characiform fish. *Biol J Linn Soc.* 88(2):235–255.
- Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. *Nat Methods.* 17(2):155–158.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210–3212.
- Smith A, Hubley R, Green P. 2013. RepeatMasker Open-4.0. Available from: <https://www.repeatmasker.org/>. Accessed September 29, 2021.
- Stanke M, et al. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* 34(Web Server Issue):W435–W439.
- Traeger LL, Sabat G, Barrett-Wilt GA, Wells GB, Sussman MR. 2017. A tail of two voltages: proteomic comparison of the three electric organs of the electric eel. *Sci Adv.* 3(7):e1700523.
- Trizna M. 2020. *Assembly\_stats* 0.1.4. doi:10.5281/zenodo.3968775. Accessed September 29, 2021.
- Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 27(5):737–746.
- Vurture GW, et al. 2017. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 33(14):2202–2204.
- Walker BJ, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9(11):e112963.
- Wang Y, Yang L. 2021. Genomic evidence for convergent molecular adaptation in electric fishes. *Genome Biol Evol.* 13(3): evab038.
- Wick R. 2018. Porechop. Available from: <https://github.com/rwrick/Porechop>. Accessed September 29, 2021.
- Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 15(3):R46.
- WTSI-HPAG 2020. Scaff10X assembly tool. Available from: <https://github.com/wtsi-hpag/Scaff10X>. Accessed September 29, 2021.
- Xu J, Cui X, Zhang H. 2021. The third form electric organ discharge of electric eels. *Sci Rep.* 11(1):6193.
- Zheng GX, et al. 2016. Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nat Biotechnol.* 34(3):303–311.

**Associate editor:** Bonnie Fraser