Draft genome and multi-tissue transcriptome assemblies of the Neotropical leaf-frog _Phyllomedusa bahiana_

Pedro Andreádê, 1,2,∗ Mariana L. Lyra 1,4,5 Juliana Zina 6,7 Deivson F.O. Bastos 6,7 André E. Brunetti 6,7 Délio Baêta 1,2,7 Sandra Afonso 1,2,7 Tuliana O. Brunes 1,2,8 Pedro P.G. Taucce 1,2,9 Miguel Carneiro 1,2,8 Célio F.B. Haddad 2,3 Fernando Sequeira 1,2,∗

1CIBIO, Centro de Investigación em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, Vairão 4485-661, Portugal.
2BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, Vairão 4485-661, Portugal.
3Departamento de Biodiversidade and Centro de Aquicultura, Instituto de Biociências, Universidade Estadual Paulista (UNESP), Rio Claro 13506-900, Brazil.
4Present address: New York University Abu Dhabi, Saadiyat Island, PO Box 129188, Abu Dhabi, United Arab Emirates.
5Departamento de Ciências Biológicas, Universidade Estadual do Sudoeste da Bahia, Jequié 45206-190, Brazil.
6Laboratory of Evolutionary Genetics, Institute of Subtropical Biology, National University of Misiones (UNaM-CONICET) Posadas N3300LQH, Misiones, Argentina.
7Departamento de Zoologia, Instituto de Biociências, Universidade de São Paulo, São Paulo 05508-090, Brazil.
8Present address: Coordenação de Biodiversidade, Instituto Nacional de Pesquisas da Amazônia (INPA), Manaus, AM 69067-375, Brazil.

∗Corresponding author: CIBIO, Centro de Investigación em Biodiversidad y Recursos Genéticos, InBIO Laboratorio Associado, Campus de Vairão, Universidade do Porto, Vairão 4485-661, Portugal. Email: pandrade@cibio.up.pt; *Corresponding author: CIBIO, Centro de Investigación en Biodiversidad y Recursos Genéticos, InBIO Laboratorio Associado, Campus de Vairão, Universidade do Porto, Vairão 4485-661, Portugal. Email: fsequeira@cibio.up.pt

Abstract

Amphibians are increasingly threatened worldwide, but the availability of genomic resources that could be crucial for implementing informed conservation practices lags well behind that for other vertebrate groups. Here, we describe draft de novo genome, mitogenome, and transcriptome assemblies for the Neotropical leaf-frog _Phyllomedusa bahiana_ native to the Brazilian Atlantic Forest and Caatinga. We used a combination of PacBio long reads and Illumina sequencing to produce a 4.74-Gbp contig-level genome assembly, which has a contiguity comparable to other recent nonchromosome level assemblies. The assembled mitogenome comprises 16,239 bp and the gene content and arrangement are similar to other Neobatrachia. RNA-sequencing from 8 tissues resulted in a highly complete (86.3%) reference transcriptome. We further use whole-genome resequencing data from _P. bahiana_ and from its sister species _Phyllomedusa burmeisteri_, to demonstrate how our assembly can be used as a backbone for population genomics studies within the _P. burmeisteri_ species group. Our assemblies thus represent important additions to the catalog of genomic resources available from amphibians.

Keywords: amphibian; Anura; Phyllomedusidae; transcriptome; genome; mitogenome

Introduction

Amphibians are a highly species-rich group of vertebrates, and a large number of its species are affected by widespread, human-driven environmental changes that have impacted global biodiversity. According to the IUCN’s Red List, 41% out of 7,215 species assessed are thought to be endangered (IUCN 2021), and for a further ~2,200 species for which scarce data prevents a proper IUCN assessment, about half are estimated to be endangered as well (González-del-Pliego et al. 2019). Major threats include habitat loss, the spread of the fungal pathogen Batrachochytrium dendrobatidis, pesticides and climate change (Collins and Storfer 2003; Bishop et al. 2012; Lips 2016). While these threats are impacting amphibian diversity worldwide, effects seem to be particularly severe for Neotropical species (Stuart et al. 2004).

In this context, there is thus ample need to ramp up conservation efforts of amphibian groups worldwide and generate resources for researchers working on this taxonomic group. Genomic resources, in particular, can provide unique insights into the evolutionary history of species and the functional basis underlying adaptation to environmental stressors (Supple and Shapiro 2018). Apart from conservation concerns, generating genomic resources for amphibians is also highly valuable to increase our understanding of genome evolution in vertebrates. Amphibians are noteworthy for presenting highly variable genome sizes, which are mostly (but not only) a consequence of the evolution of their transposable element content (Liedtke et al. 2018; Herrick and Sclavi 2020). Despite these reasons, amphibians lag behind other major vertebrate taxa when considering the availability of genomic resources. For example, a quick search in NCBI’s Assembly database (https://www.ncbi.nlm.nih.gov/assembly; last accessed 2022 Jul 27) currently recovers 42 published amphibian assemblies, compared with more than 700 birds, 2,300 mammals, 1,300 teleosts, and 90 nonavian saurapsids.

Treefrogs (Anura, Arboranae) is a widespread group of amphibians across the whole American continent, Eurasia, North Africa, and the Australo-Papuan region, with a much higher

Received: August 03, 2022. Accepted: September 07, 2022
© The Author(s) 2022. Published by Oxford University Press on behalf of Genetics Society of America.
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.
number of species around its center of diversification in South America, where they are estimated to have originated in the Palaeocene (Duellman et al. 2016). They comprise 3 families (Hyliidae, Pelodyridae, and Phyllomedusidae), with estimated 1,035 species (Frost 2022), making treefrogs one of the more species-rich groups within the Anura. In spite of this, very few genomic resources have been generated for this group.

In this study, we present draft genome, mitogenome, and transcriptome assemblies of the Neotropical leaf-frog genomic resources have been generated for this group. This species is part of the P. burmeisteri species group (Faivovich et al. 2010), which includes 4 diploid and one tetraploid species. This group is essentially distributed along the Mata Atlântica (Brazilian Atlantic Forest), Caatinga (Brazilian dry forest), and Pampa (grassy plains) biomes of South America, mostly in Brazil, and are estimated to have diverged ca. 4.9 MYA (Fig. 1b; Brunes et al. 2010). Cytogenetic analyses indicate a conserved karyotype of 2n = 26 for diploid species of this group (Gruber et al. 2013; Barth et al. 2013, 2014), and a haploid genome size between 6.0 and 7.0 Gbp (Goin et al. 1968; Batistic et al. 1975; Liedke et al. 2018). Leaf frogs have a combination of traits that make them promising models to study a variety of biological processes. For example, they have a very high capacity to deal with temperature stress by reducing evaporative water loss considerably when compared with other anurans, a skill that is achieved through secreting lipids that are spread over the skin coupled with increased water retention by excreting uric acid instead of urea (Withers et al. 1984; Shoemaker et al. 1972). Another important aspect of the biology of leaf frogs is the abundant and complex mixture of more than 200 bioactive peptides that evolved for defense against predators and pathogens, and that are of interest in the pharmaceutical industry (Calderon et al. 2011; Bartels et al. 2019; Zandsalimi et al. 2020). These aspects underscore the opportunities that may arise from a more complete characterization of the molecular diversity in these amphibians at multiple levels, for which the establishment of genomic resources has the potential to be a decisive first step.

Materials and methods
Whole-genome long- and short-read sequencing
For the genome assembly, an adult male of P. bahiana was collected from the municipality of Jequêí (Bahia, Brazil, −13.987925−40.007450W; Fig. 1b). The individual was euthanized by applying 5% lidocaine to the skin. Heart tissue was collected and immediately frozen in −80°C. The specimen was fixed in 10% formalin and transferred to 70% ethanol for permanent storage at Museu de História Natural de Jequêí Coletão Herpetológica, Universidade Estadual do Sudoeste da BahiaM (Voucher: MHNJCH 1462).

High molecular weight DNA (fragment size >60 kb) was extracted from heart tissue using the MagAttract HMW DNA Kit (QIAGEN). DNA quantity and integrity was assessed using a NanoDrop instrument, Qubit dsDNA BR Assay Kit and Genomic DNA ScreenTape (Agilent). To assemble the genome of P. bahiana, we used long-read sequencing as the main source of data. PacBio libraries were prepared with the SMRTbell Express Template Prep Kit 2.0 and sequenced on a Sequel II instrument on 2 independent runs, each with a movie time of 30 h to generate CCS reads (performed at Novogene UK). Additionally, we also generated a 150-bp short-read library with a 10x Genomics Chromium linked-read protocol (Novogene UK), which was sequenced to an estimated depth of 53.7× with an Illumina NovaSeq instrument (assuming a genome size of 6.5 Gbp).

Genome assembly
To assemble our genome, we input the raw PacBio sub-read data, converted into fastq format using the software bamUtil v1.0.14 (https://github.com/statgen/bamUtil) into wtdbg2 v2.5 (Ruan and Li 2020) assuming a genome size of 6.5 Gb and maintaining other parameters as default. An additional assembly was performed by reducing the minimum length of alignment (−l, from 2,048 to 1,000), the minimum subread length (−L, from 5,000 to 4,000) and minimum depth (−X, from 50 to 30); however, results were qualitatively unchanged, so we retained the assembly that resulted from the run with default parameters. In parallel, we also attempted to generate a Chromium assembly with the linked-read data using Supernova v2.0 (Weissenfeld et al. 2017), but assembly statistics (e.g. scaffold N50, contig N50) were not satisfactory, so we discarded this assembly.

The sequence accuracy of assemblies based on long reads are typically error-prone, so following assembly we corrected the sequence using the Illumina Chromium whole-genome sequencing data. For polishing, we used NextPolish v1.3.1 (Hu et al. 2020) imposing a maximum depth of 100 and a single iteration of polishing (contrary to other tools, NextPolish does not require repeated polishing iterations). Due to the general lack of knowledge on amphibian genome evolution patterns from the lack of available reference genome sequences, repeat element detection and masking was done using a de novo method based on the detection of repeated k-mers, using the software Red v2.0 (Girgis 2015) using default parameters. This approach, while not identifying specific classes of repeat elements, is unconstrained by the comparison to a reference repeat library, which makes it likely more accurate at detecting true repeats. To perform a comparative analysis of repeat element abundance in anurans, we also carried out a k-mer-based repeat detection step using Red for other anuran genome assemblies retrieved from the NCBI assembly database and other sources (Table 1).

After an initial submission to NCBI’s GenBank database, an automated screening for adaptors sequences detected potential contamination (adaptor sequences NGBO09721.1 and NGBO1063) in 44 genomic regions (from 40 contigs). To correct this, we converted these adaptor sequences into Ns and used these to split the contigs using the script split.scaffolds to contigs.pl (Mads Albertsen, https://github.com/MadsAlbertsen).

To evaluate the quality of our reference assembly, we calculated common metrics using the script assemblathon_stats.pl (https://github.com/KorLab/Assemblathon; Bradnam et al. 2013). To assess genome completeness, we also quantified the number of highly conserved single-copy orthologs using BUSCO v5 (Simão et al. 2015). Gene prediction for this orthology search was performed using MetaEuk v4-a0f584d (Levy Karin et al. 2020), with the lineage dataset metazoa_odb10.

Mitochondrial genome sequencing and assembly
To assemble the mitochondrial genome (mitogenome) we used the forward read sequences from the Illumina sequencing data. The raw reads were trimmed for adapters using Trimmomatic v0.39 (Bolger et al. 2014) and all reads smaller than 50 bp were discarded. The mitogenome was assembled in 2 steps: we first used GetOrganelle v1.7.5 (Jin et al. 2020) under default settings to produce a draft of the P. bahiana mitogenome and then we used this draft as seed for an iterative mapping assembly using MITobim v1.9.1 (Hahn et al. 2013). Iterations were run until no additional reads could be incorporated into the assembly. We evaluated the assemblies for completeness and coverage by importing the
Fig. 1. a) Photo of Phyllomedusa bahiana (by Pedro P. G. Taucce). b) Map with the approximate distribution range of P. bahiana (blue shading), with indication on sampling location for the specimens used for reference genome and transcriptome assemblies (Jequié, Bahia, Brazil). Distribution map was adapted from Pombal and Haddad (1992), while the topographic map was retrieved from Natural Earth (https://www.naturalearthdata.com). c) Comparison between ours and other anuran assemblies regarding assembly size and contig N50 (natural logarithm transformed for ease of visualization). The position of P. bahiana is highlighted in blue (see Supplementary Figs. 1 and 2 for names of all species). d) Relationship between genome size (Gbp) and the proportion of repetitive sequence for all anuran species with published genomes. Repetitive sequence content was calculated for each genome using a de novo, unbiased repeat detection method. e) Frequency distribution of genome-wide values of nucleotide diversity (p) and Tajima’s D for P. bahiana and P. burmeisteri, based on pool-sequencing data. Each statistic was calculated in 10-kb nonoverlapping windows across the whole genome. For better visualization, a small number of windows with nucleotide diversity values above 3.0% was omitted.
mapping output from MITObim in Geneious vR11 (https://www.geneious.com). The final mitogenome annotation was carried out using MITOS v2 (Donath et al. 2019). The protein-coding regions were checked to confirm that no indels or stop codons were present.

### RNA-sequencing

For the transcriptome assembly, we collected a male Phyllomedusabahiana, euthanized it as previously described and harvested the following tissues: liver, heart, testicle, spleen, kidney, and skin (from the latter, 3 individual patches from the dorsum, inner thigh, and lateral trunk). All tissues were immediately placed in −80°C. Total RNA was isolated using the RNeasy Mini kit (QIAGEN) followed by DNAse digestion. RNA purity and concentration were assessed using a Nanodrop instrument prior to library preparation. cDNA was generated from DNA using the QIAamp DNA Micro Kit (QIAGEN). RNA purity was assessed using a Qubit dsDNA BR Assay Kit (ThermoFisher), DNA purity was assessed using

### Transcriptome assembly

RNAseq read quality was confirmed using FastQC v0.11.8. Since we intended to perform a general transcriptome assembly, for the following steps we merged reads from all tissues. Merged reads were corrected with RoScract v1.0.3.1 (Song and Florea 2015) to exclude read pairs with at least one unfixable read, and Trim Galore! v0.6.0 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore) to remove adapters, low quality bases (Phred score <3), and reads smaller than 36 bp after trimming. To remove contamination from ribosomal RNA, reads were mapped with Bowtie2 v2.3.5 (Langmead and Salzberg 2012) to the SSUParc and LSUParc fasta files from the database SILVA (Quast et al. 2013) (https://www.arb-silva.de/; downloaded April 2019) using the very-sensitive-local option. Reads that were positive hits were discarded. For assembly, we used SPAdes v3.15.2 (Prijibelski et al. 2020) with kmer sizes of 21, 33, 51, 77, 99, and 127. To reduce redundancy in the assembly, we clustered highly similar sequences (0.99% identity threshold) with CD-HIT v4.8.1 (Fu et al. 2012). We assessed completeness of the transcriptome assembly by comparing it to the metazoa_odb10 database with BUSCO, as performed for the genome assembly. To generate annotations for our transcripts, we used Smal v2 (Casimiro-Soriguer et al. 2017), using BLAST v2.2.29 (Camacho et al. 2009) to search (low complexity filter activated) against the whole UniProt database to retrieve gene names and GO (gene ontology) terms. To infer orthology and duplication events, we compared our de novo transcriptome with a set of published anuran transcriptomes (see Supplementary Table 2 for a list of species and accession codes for the datasets) using OrthoFinder v2.5.4 (Emms and Kelly 2019), employing FastTree (Price et al. 2010) for tree inference.

### Whole-genome pool-sequencing

To investigate the potential of our reference assembly to conduct population genomics studies, we carried out whole-genome sequencing based on pools of individuals. To do so we collected samples of Phyllomedusa bahiana from the Brazilian locality of Jequié, Bahia (n = 23) and allopatric populations from its sister species Phyllomedusa burmeisteri from different Brazilian municipalities (n = 18; Campinas, Queluz, São José do Rio Pardo, Nazaré Paulista, Atibaia and Patrocínio Paulista—all in São Paulo state). From each sample, we extracted DNA using the QIAamp DNA Micro Kit (QIAGEN). DNA concentration was quantified using a Qubit dsDNA BR Assay Kit (ThermoFisher), DNA purity was assessed using

### Table 1. Summary of the sequencing datasets of Phyllomedusa bahiana and P. burmeisteri used in this study, including accession codes for NCBI’s Sequence Read Archive (SRA).

| Dataset                              | Accession codes | Data type      | Sequencing instrument | Number of reads | Tissue type |
|--------------------------------------|-----------------|----------------|-----------------------|-----------------|-------------|
| P.bahiana long-read sequencing (library A) | SRR18363239     | PacBio CCS     | Sequel II             | 18,011,507      | Heart       |
| P.bahiana long-read sequencing (library B) | SRR18363238     | PacBio CCS     | Sequel II             | 10,641,819      | Heart       |
| P.bahiana Chromium sequencing        | SRR18363237     | Chromium linked reads (150-bp PE) | Illumina NovaSeq | 2,330,800,682   | Heart       |
| P.bahiana RNAseq (liver)             | SRR18363236     | Illumina RNAseq (100-bp PE) | Illumina NovaSeq | 65,108,950      | Liver       |
| P.bahiana RNAseq (heart)             | SRR18363235     | Illumina RNAseq (100-bp PE) | Illumina NovaSeq | 71,641,510      | Heart       |
| P.bahiana RNAseq (testicle)          | SRR18363234     | Illumina RNAseq (100-bp PE) | Illumina NovaSeq | 65,952,858      | Testicle    |
| P.bahiana RNAseq (spleen)            | SRR18363233     | Illumina RNAseq (100-bp PE) | Illumina NovaSeq | 62,894,824      | Spleen      |
| P.bahiana RNAseq (kidney)            | SRR18363232     | Illumina RNAseq (100-bp PE) | Illumina NovaSeq | 60,424,896      | Kidney      |
| P.bahiana RNAseq (skin-dorsum)       | SRR18363229     | Illumina RNAseq (100-bp PE) | Illumina NovaSeq | 55,109,830      | Skin (dorsum)|
| P.bahiana RNAseq (skin-thigh)        | SRR18363231     | Illumina RNAseq (100-bp PE) | Illumina NovaSeq | 67,226,174      | Skin (inner thigh) |
| P.bahiana RNAseq (skin-trunk)        | SRR18363230     | Illumina RNAseq (100-bp PE) | Illumina NovaSeq | 61,183,514      | Skin (lateral trunk) |
| P.bahiana WGS pool-seq               | SRR18363241     | Illumina WG pool-seq (150-bp PE) | Illumina NovaSeq | 742,662,820     | Muscle      |
| P.burmeisteri WGS pool-seq           | SRR18363240     | Illumina WG pool-seq (150-bp PE) | Illumina NovaSeq | 846,586,334     | Muscle      |

**Table 1.** Summary of the sequencing datasets of Phyllomedusa bahiana and P. burmeisteri used in this study, including accession codes for NCBI’s Sequence Read Archive (SRA).
NanoDrop spectrophotometry (ThermoFisher) and average fragment size assessed using agarose gel electrophoresis. After these assessments, we pooled DNA from each individual in equimolar concentrations. Illumina TruSeq PCR-free libraries from the 2 pools were prepared according to the manufacturer’s protocol, quantified by qPCR using the KAPA Library Quantification Kit (Roche) and sequenced using 2 × 150 bp reads in an Illumina NovaSeq instrument (Novogene UK).

Population genomics

We evaluated the quality of the sequencing runs using FastQC v0.11.8. We mapped the reads to the de novo reference assembly with BWA-MEM v1 (Li 2013) using default settings and calculated mapping statistics using SAMtools v1.9 (Li et al. 2009) and custom scripts. Pool-sequencing does not allow to extract individual genotypes, so we calculated statistics based on allele frequencies. SAMtools mpileup was used to gather per-site allele counts, from which allele frequencies were calculated and population genetics parameters derived using PoPoolation v1.2.2 (Kofler, Orozco-TerWengel, et al. 2011) and PoPoolation2 v1.201 (Kofler, Pandey, et al. 2011). Nucleotide diversity (θ, Nei 1987), Tajima’s D (Tajima 1989), and Fst (Hartl and Clark 2007) were calculated in nonoverlapping 10 kb windows, imposing a minimum coverage of 10, a maximum coverage of 100, a minimum base quality of 30, a minimum count of 2 for the alternate allele to consider a position as polymorphic, and a mapping rate of 70%. BUSCO completeness values of our assembly are in line with those of other amphibian species (Metcalfe and Casane 2013; Rogers et al. 2018; Li, Yu, et al. 2019). Considering these findings together with BUSCO results indicating low amounts of gene duplication, the expansion of repetitive elements is one likely contributing factor for the large genome size of P. bahiana, as has been suggested for other amphibian species (Metcalfe and Casane 2013; Rogers et al. 2018; Streicher et al. 2021a), since the assembly of repetitive sequences is highly complex (Tørresen et al. 2019).

The complete mitogenome assembly resulted in a total of 16,239 bp with mean coverage of 164×, containing the expected 2 ribosomal genes (12S and 16S rRNA), the 13 protein-coding genes, 22 tRNAs and the noncoding control region. The gene order arrangement and gene content are consistent with Phyllomedusa tomopterna and Pithecopus megacephalus, so far the only mitogenomes available for the Phyllomedusidae; they also agree with the pattern found for other Neobatrachia (Zhang et al. 2013; Lima et al. 2020).

Results and discussion

Genome and mitogenome assemblies

Our two PacBio runs generated a total of 28,653,326 subreads, with an average subread length of 14,046.7 bp (Table 1). Chromium Genome and mitogenome assemblies were processed with the BUSCO v1.2b (Kofler, Orozco-TerWengel, et al. 2011) and BUSCO2 v1.2.2 (Kofler, Orozco-TerWengel, et al. 2011). BUSCO analysis revealed that our genome was reasonably complete. A total of 954 single-copy orthologs from a metazoan database were searched against our genome, from which 671 (70.3%) were complete and single-copy. In spite of the large genome size of P. bahiana only 15 of the reference single-copy orthologs (1.6%) were duplicated. Fragmented orthologs account for a large portion of noncomplete BUSCO groups (154, or 16.1% of the total number), which is a likely consequence of a fragmented assembly. When comparing to other amphibian genome assemblies for which similar completeness analyses were performed, completeness values of our assembly are in line with those of other assemblies with reduced or no scaffolding (e.g. Hammond et al. 2017; Lamichhaney et al. 2021), which represents ~10–20% lower completeness than of highly scaffolded assemblies (e.g. Li, Ren, et al. 2019; Seidl et al. 2019; Li et al. 2020; Streicher et al. 2021a, 2021b; Wu et al. 2022).

To further test if the large genome size of P. bahiana is explained by either widespread duplications or expansions of repetitive genomic elements, we calculated the percentage of the total genome that is repetitive in our assembly and the other anuran genomes. As expected, the percentage of the genome that was found to be repetitive increased linearly and significantly with assembled genome size across all species (ordinary least squares regression, $r^2 = 0.23$, $P = 0.008$), with P. bahiana having the fifth largest assembly size and twelfth largest proportion of the genome that is repetitive (Fig. 1d; P. bahiana = 0.53; min. = 0.27; max. = 0.71). Considering these findings together with BUSCO results indicating low amounts of gene duplication, the expansion of repetitive elements is one likely contributing factor for the large genome size of P. bahiana, as has been suggested for other amphibian species (Metcalfe and Casane 2013; Rogers et al. 2018; Li, Yu, et al. 2019). It should, however, be taken into account that a more detailed characterization of the repeat element landscape of the genome is needed to draw definitive conclusions. If this is confirmed, it could explain the discrepancy between estimated genome sizes from flow cytometry and the final assembly size in other large anuran genomes (Edwards et al. 2018; Rogers et al. 2018; Streicher et al. 2021a, since the assembly of repetitive sequences is highly complex (Tørresen et al. 2019).

Transcriptome assembly and orthology inference

RNA-sequencing from multiple tissues generated a total of 509,542,356 raw reads (number of reads per library ranging between 55,109,630 and 71,641,510; Table 1). The assembly of an overall transcriptome from the merged dataset yielded a total of 168,692 transcripts, with a mean transcript length of 657 bp (longest transcript = 11,718 bp) and 43.47% GC content. The BUSCO search for single-copy orthologs indicated that 86.3% of the searched groups were complete and single-copy. Similarly to the genome assembly, we found little evidence of widespread genome duplication, since only 5 BUSCO groups (0.5% of the total) were duplicated. Of the total number of transcripts, we were able to annotate 35,598 transcripts (21.10%) with either putative orthologs from the Uniprot database, GO term annotations or both (the majority of cases). Comparison of orthologous transcripts between our assembled transcriptome and those of other anurans recovered a total of 2,746,416 independent transcripts, the majority of which (73.7%) were assigned to orthogroups. For P. bahiana, 75.5% of transcripts were placed in orthogroups, with 24.6% in species-specific orthogroups.

Population genomics

To estimate population genetics parameters from samples of P. bahiana and its sister species P. burmeisteri, we performed whole-genome sequencing of 2 pools of individuals, 1 pool from each species. We generated a total of 742,662,620 reads for the P. bahiana pool (~23.50× coverage) and 846,586,334 reads for P. burmeisteri (~26.78× coverage), which after read trimming, were mapped to the reference genome assembly with high mapping rates for both pools (93.98% and 95.78%, respectively). We observed in
both cases a substantial drop in the percentage of properly paired reads (74.08% and 72.81%), which likely reflects the fragmented nature of our current assembly.

Using this data, we next compared patterns of sequence variation in our 2 samples. Variant calling based on allele counts retrieved a total of 42,502,110 biallelic single nucleotide polymorphisms. Out of these, 6,388,773 (15.0%) are diagnostic mutations between the species (allele frequency difference = 1.0). Genome-wide FST between the 2 species was high (average ± SD; FST = 0.45 ± 0.15). We did not find evidence that genetic differentiation was clustered in any particular region of the genome (Supplementary Fig. 3). For example, when looking at counts of diagnostic mutations in nonoverlapping 10-kb windows for the largest 250 contigs, the windows with higher counts of diagnostic alleles were scattered across these contigs (and apart from these contigs, only 12 additional genomic windows had higher values, and none located in the same contig). A similar pattern was found for FST, which is consistent with these populations representing lineages with largely independent evolutionary histories. However, despite a deep divergence time between the taxa (~2.9 MY; Brunes et al. 2014), average genome-wide nucleotide diversity is similar between them (average ± SD; \( \pi_{\text{bahiana}} = 0.89 \pm 0.50; \ \pi_{\text{burmeisteri}} = 0.82 \pm 0.56; \ \text{Fig. 1e and Supplementary Fig. 4} \)). These values are relatively high when compared with most vertebrates (Romiguier et al. 2014). Tajima’s D was also similar (slightly negative) between the 2 species (average ± SD; TD\textsubscript{bahiana} = –0.26 ± 0.40; TD\textsubscript{burmeisteri} = –0.21 ± 0.49; Fig. 1e and Supplementary Fig. 5). These results are consistent with previous findings using nuclear markers (Brunes et al., 2014), and are suggestive of population expansion following a bottleneck, a pattern commonly found among many other amphibians from the Atlantic Forest (e.g. Carnaval et al. 2009; Brunes et al. 2015).

Arboranae represent around 10% of the total number of known amphibian species. Our de novo reference assemblies of the nuclear genome, mitogenome, and transcriptome of P. bahiana, which are among the first genomic data for the Arboranae, constitute important resources that can serve as the basis for future studies on the ecology and evolution of Neotropical leaf frogs. Future enhancements to our assemblies should focus on improving genome scaffolding and contiguity in order to achieve a chromosome-scale assembly. This should be achieved through the use of additional sources of data, such as chromosome conformation capture (HiC) and also by increasing the sequencing depth with long-reads (either PacBio or Oxford Nanopore) (Luo et al. 2021).

**Data availability**

Raw whole-genome sequencing data (PacBio long reads, Chromium linked reads and Illumina pool-sequencing) and RNA-sequencing data have been deposited in the Sequence Read Archive (www.ncbi.nlm.nih.gov/sra) under BioProject PRJNA71272. Fasta files for the assemblies have been deposited in GenBank, with the following accession codes: genome (JACDAL000000000), mitogenome (OM460708), and transcriptome (GJVR0000000). Supplemental material is available at G3 online.

**Acknowledgments**

Specimens were collected under permits numbers #22511, #76477-1, and #35068 issued by SISBio/ICMBio (Instituto Chico Mendes de Conservação da Biodiversidade). The access to the National System for the Management of Genetic Heritage and
Associated Traditional Knowledge was registered (SISGen #ACC8816). Victor M. Ghirotto helped with the fieldwork in Campinas.

**Funding**

This study was funded by the FEDER Funds through the Operational Competitiveness Factors Program—COMPETE and by National Funds through FCT (Fundação para a Ciência e Tecnologia) within the scope of the projects PTDC/BIA-EVL/28621/2017 and UIDB/50027/2020, and by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) within the scope of the project #2018/07586-4. The work was also co-funded by the project NORTE-01-0246-FEDER-000063, supported by Norte Portugal Regional Operational Programme (NORTE2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). PA was supported by FCT (within project PTDC/BIA-EVL/28621/2017 and research contract 2020.01405/CEECIND/CP1601/CT0011). MC was supported by FCT (CEECINST/00014/2018/CP1512/CT0002). CFBH is grateful to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), for a research fellowship. MLL, TOB, and PPGT were supported by postdoctoral fellowships of FAPESP (#2017/26162-8, #2016/03146-4, #2019/04076-8, respectively). DFOB was supported by FAPESP (grant# 441/2020) and FCT (within project PTDC/BIA-EVL/28621/2017). AEB was supported by Concurso Nacional de Investigaciones Científicas y Técnicas (CONICET—PIP 2021–2023, CONICET and PICT 2022–2024).

**Conflicts of interest**

None declared.

**Literature cited**

Bartels EJH, Dekker D, Amiche M. Dermaseptins, multifunctional antimicrobial peptides: a review of their pharmacology, effectiveness, mechanism of action, and possible future directions. Front Pharmacol. 2019;10:1421. doi: 10.3389/fphar.2019.01421.

Barth A, Souza VA, Solé M, Costa MA. Molecular cytogenetics of nucleolar organizer regions in *Phyllomedusa* and *Phasmatryla* species (Hyliidae, Phyllomedusinae): a cytotaxonomic contribution. Genet Mol Res. 2013;12(3):2400–2408.

Barth A, Vences M, Solé M, Costa MA. Molecular cytogenetics and phylogenetic analysis of Brazilian leaf frog species of the genera *Phyllomedusa* and *Phasmatryla* (Hyliidae: Phyllomedusinae). Can J Zool. 2014;92(9):795–802.

Batistic RF, Soma M, Beçak ML, Beçak W. Further studies on polyploid amphibians: a diploid population of *Phylomedusa burmeisteri*. J Hered. 1975;66(3):160–162.

Bishop PJ, Angulo A, Lewis JP, Moore RD, Rabb GB, Moreno JG. The Amphibian extinction crisis—what will it take to put the action into the Amphibian Conservation Action Plan? SAPIEN. 2012;5:2: 97–111.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2141–2142.

Bradnam KR, Fass JN, Alexandrov A, Baranay P, Bechner M, Birol I, Boisvert S, Chapman JA, Chaupis G, Chikhi R, et al. Assemblyloch 2: evaluating de novo methods of genome assembly in three vertebrate species. GigaScience. 2013;2(1):2047–217X-2-10.

Brunes TO, Sequeira F, Haddad CF, Alexandrino J. Gene and species trees of a Neotropical group of treefrogs: genetic diversification in the Brazilian Atlantic Forest and the origin of a polyploid species. Mol Phylogenet Evol. 2010;57(3):1120–1133.

Brunes TO, Alexandrino J, Baeta D, Zina J, Haddad CF, Sequeira F. Species limits, phylogeographic and hybridization patterns in Neotropical leaf frogs (*Phyllomedusinae*). Zool Scr. 2014;43(6):586–604.

Brunes TO, Thomé MTC, Alexandrino J, Haddad CFB, Sequeira F. Ancient divergence and recent population expansion in a leaf frog endemic to the southern Brazilian Atlantic Forest. Org Divers Evol. 2015;15(4):695–710.

Calderon LA, Alexandre de Almeida ES, Ciancaglini P, Stäbeli RG. Antimicrobial peptides from *Phylomedusa* frogs: from biocumulatory diversity to potential nanotechnologic medical applications. Amino Acids. 2011;40(1):29–49.

Carnacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10(1):1–9.

Carnaval AC, Hickerson MJ, Haddad CFB, Rodrigues MT, Moritz C. Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. Science. 2009;323(5915):785–789.

Casimiro-Soriguier CS, Muñoz-Mérida A, Pérez-Pulido AJ. Sma3s: a universal tool for easy functional annotation of proteomes and transcriptomes. Proteomics. 2017;17(12):1700071.

Collins JP, Storfer A. Global amphibian declines: sorting the hypotheses. Divers Distrib. 2003;9(2):89–98.

Donath A, Jühlung F, Al-Abrah M, Bernhart SH, Reinhardt F, Stadler PF, Middendorf M, Bernt M. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. Nucleic Acids Res. 2019;47(20):10543–10552.

Duellman WE, Marion AB, Hedges SB. Phylogenetics, classification, and biogeography of the treefrogs (Amphibia: Anura: Arboranae). Zootaxa. 2016;4104(1):1–109.

Edwards RJ, Tuipulotu DE, Amos TG, O’Meally D, Richardson MF, Russell TL, Vallinoto M, Carneiro M, Ferrand N, Wilkins MR, et al. Draft genome assembly of the invasive cane toad, *Rhinella marina*. GigaScience. 2018;7(9):gjy095.

Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20(1):1–14.

Faivovich J, Haddad CFB, Baêta D, Jungfer K-H, Álvares GFR, Brandão RA, Sheil C, Barrientos LS, Barrio-Amorós CL, Cruz CAG, et al. The phylogenetic relationships of the charismatic poster frogs, *Phyllomedusinae* (Anura, Hyliidae). Cladistics. 2010;26(3):227–261.

Frost DR. Amphibian Species of the World: An Online Reference. Version 6.1 Electronic Database. New York (NY): American Museum of Natural History; 2022. https://doi.org/10.5531/db.vz.0001 [accessed 2022 Mar 29]. https://amphibiansoftheworld.amnh.org/index.php.

Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the Amino Acids. 2011;40(1):29–49.

Girgis HZ. Red: an intelligent, rapid, accurate tool for detecting repeats de-novo on the genomic scale. BMC Bioinform. 2012;28(23):3150–3152.

González-del-Piñero P, Freckleton RP, Edwards DP, Koo MS, Scheffers BR, Pyron RA, Jetz W. Phylogenetic and trait-based prediction of extinction risk for data-deficient amphibians. Curr Biol. 2019;29(9):1557–1563.e3.

Gruber SL, Silva APZ, Haddad CFB, Kasahara S. Cytogenetic analysis of *Phylomedusa* distincta Lutz, 1950 (2n = 2x = 26), P. *tetraploidea* Pombal and Haddad, 1992 (2n = 4x = 52), and their natural triploid...
hybrids (2n= 3x= 39) (Anura, Hylidae, Phyllomedusinae). BMC Genet. 2013;14(1):75–13.

Hammond SA, Warren RL, Vandervalk BP, Kucuk E, Khan H, Gibb EA, Pandoh P, Kirk H, Zhao Y, Jones M, et al. The North American bullfrog draft genome provides insight into hormonal regulation of long noncoding RNA. Nat Commun. 2017;8(1):1–8.

Hahn C, Bachmann L, Chevreux B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Res. 2013;41(13):e129.

Hartl DL, Clark AG. Principles of Population Genetics. Sunderland (MA): Sinauer; 2007. p. 545.

Herrick J, Sclavi B. Genome evolution in amphibians. In: eLS. Chichester (UK): John Wiley & Sons; 2020. p. 1–10. https://onlinelibrary.wiley.com/doi/10.1002/9780470015902.a002896

Hu J, Fan J, Sun Z, Liu S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. Bioinformatics. 2020;36(7):2253–2255. 2020

IUCN. The IUCN Red List of Threatened Species. Version 2021-2; 2021 [accessed 2021 Nov 15].https://www.iucnredlist.org.

Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ. Comparative population genetic analysis of next generation sequencing data from pooled individuals. PLoS One. 2011;6(1):e15925.

Koller R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V, Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21(1):241.

Koller R, Pandey RV, Schlottere C. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics. 2011;27(24):3435–3436.

Lamichhane S, Catullo R, Keogh JS, Clulow S, Edwards SV, Ezaz T. A bird-like genome from a frog: mechanisms of genome size reduction in the omaru burrowing frog, Platypseudactylus omnus. PNAS. 2021;118(11):e2011649118.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–359.

Levy Karin E, Mirdita M, Söding J. MetaEuk—sensitive, high-throughput gene discovery, and annotation for large-scale eukaryotic metagenomics. Microbiome. 2020;8(1):1–15.

Li H. Aligning sequence reads, clone sequences and assembly contigs with SQA-MEM. arXiv preprint 1303.3997, 2013.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–2079.

Li Y, Ren Y, Zhang D, Jiang H, Wang Z, Li X, Rao D. Chromosome-level assembly of the mustache toad genome using third-generation sequencing data. Gigabyte. 2020;2020:1. https://doi.org/10.46471/gigabyte.2.

Liedtke HC, Gower DJ, Wilkinson M, Gomez-Mestre I. Macreovolutionary shift in the size of amphibian genomes and the role of life history and climate. Nat Ecol Evol. 2018;2(11):1792–1799.

Lima NGDS, Carmo AOD, Souza RCCD, Kalapothakis E, Eterovich PC. Complete mitochondrial genome sequence of the high altitude Brazilian treefrog Pithecopus megaecophalus (Anura, Phyllomedusidae). Mitochondrial DNA B Resour. 2020;5(1):388–389.

Lips KR. Overview of chytrid emergence and impacts on amphibians. Phil Trans R Soc B. 2016;371(1709):20150465.

Luo J, Wei Y, Lyu M, Wu Z, Liu X, Luo H, Yan C. A comprehensive review of scaffolding methods in genome assembly. Brief Bioinform. 2021;22(5):bbab033.

Metcalfe CF, Casane D. Accommodating the load: the transposable element content of very large genomes. Mob Genet Elements. 2013;3(2):e24775.

Nei M. Molecular Evolutionary Genetics. New York (NY): Columbia University Press; 1987.

Pombal JP Jr, Haddad CF. Espécies de Phyllomedusa do grupo burmeisteri do Brasil oriental, com descrição de uma espécie nova (Amphibia, Hylidae). Rev Bras Biol. 1992;2:217–229.

Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5(3):e9490.

Prijibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes de novo assembler. Curr Protoc Bioinformatics. 2020;70(1):e102.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glickner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41(Database Issue):D590–D596.

Rogers RL, Zhou L, Chu C, Márquez A, Corl A, Linderoth T, Freeborn L, MacManes MD, Xiong Z, Zheng J, et al. Genomic takeover by transposable elements in the strawberry poison frog. Mol Biol Evol. 2018;35(12):2913–2927.

Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, Chiari Y, Dernat R, Duret L, Faivre N, et al. Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature. 2014;515(7526):261–263.

Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. Nat Methods. 2020;17(2):155–158.

Seidl F, Levis NA, Schell R, Pfennig DW, Pfennig KS, Ehrenreich IM. Genome of Spea multiplicata, a rapidly developing, phenotypically plastic, and desert-adapted spadefoot toad. G3 (Bethesda). 2019;9(12):3909–3919.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–3212.

Shoemaker VH, Balding D, Ruibal R, McClanahan LL. Uricotelism and low evaporative water loss in a South American frog. Science. 1972;175(4025):1018–1020.

Song L, Florea L. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. GigaScience. 2015;4(1):s13742-015.

Streicher JW, Wellcome Sanger Institute Tree of Life Programme Darwin Tree of Life Consortium. The genome sequence of the common toad, Bufo bufo (Linnaeus, 1758). Wellcome Open Res. 2021a;6(281):1–12.

Streicher JW, Wellcome Sanger Institute Tree of Life Programme Darwin Tree of Life Consortium. The genome sequence of the common frog, Rana temporaria (Linnaeus 1758). Wellcome Open Res. 2021b;6(286):1–12.

Stuart SN, Chanson JS, Cox NA, Young BE, Rodrigues AS, Fischman DL, Waller RW. Status and trends of amphibian declines and extinctions worldwide. Science. 2004;306(5702):1783–1786.

Supple MA, Shapiro B. Conservation of biodiversity in the genomics era. Genome Biol. 2018;19(1):1–12.

Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;123(3):585–595.

Tätrenos OK, Star B, Mier P, Andrade-Navarro MA, Bateman A, Jaront P, Gruba A, Grynberg M, Kajava AV, Promponas VJ, et al. Tandem repeats lead to sequence assembly errors and impose multi-level
challenges for genome and protein databases. Nucleic Acids Res. 2019;47(21):10994–11006.

Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB. Direct determination of diploid genome sequences. Genome Res. 2017;27(5):757–767.

Withers PC, Hillman SS, Drewes RC. Evaporative water loss and skin lipids of anuran amphibians. J Exp Zool. 1984;232(1):11–17.

Wu W, Gao YD, Jiang DC, Lei J, Ren JL, Liao WB, Deng C, Wang Z, Hillis DM, Zhang Y-P, et al. Genomic adaptations for arboreal locomotion in Asian flying treefrogs. Proc Natl Acad Sci USA. 2022;119(13):e2116342119.

Zandsalimi F, Talaei S, Noormohammad Ahari M, Aghamiri S, Raee P, Roshanzamiri S, Yarian F, Bandehpour M, Zadeh ZZ. Antimicrobial peptides: a promising strategy for lung cancer drug discovery? Expert Opin Drug Discov. 2020;15(11):1343–1354.

Zhang P, Liang D, Mao RL, Hillis DM, Wake DB, Cannatella DC. Efficient sequencing of anuran mtDNAs and a mitogenomic exploration of the phylogeny and evolution of frogs. Mol Biol Evol. 2013;30(8):1899–1915.

Communicating editor: A. Kern