A draft genome assembly for the dart-poison frog *Phyllobates terribilis*

Roberto Márquez[1,*,†], Denis Jacob Machado[2,3,4,*,†], Reyhaneh Nouri[2,3], Kerry L. Gendreau[1] Daniel Janies[2,3], Ralph A. Saporito[5], Marcus R. Kronforst[6], Taran Grant[4,†]

1 Department of Biological Sciences, Virginia Tech, Blacksburg, VA. USA
2 enter for Computational Intelligence to Predict Health and Environmental Risks, University of North Carolina at Charlotte, Charlotte, NC. USA
3 Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC. USA
4 Laboratório de Anfíbios, Departamento de Zoologia, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil
5 Department of Biology, John Carroll University, University Heights, OH. USA
6 Department of Ecology and Evolution, University of Chicago. Chicago, IL. USA
*** Co-first authors.**

† Correspondence: rmarquezp@vt.edu, dmachado@charlotte.edu, taran.grant@ib.usp.br

Abstract

Dendrobatid poison frogs have become well established as model systems in several fields of biology. Nevertheless, the development of molecular and genetic resources for these frogs has been hindered by their large, highly repetitive genomes, which have proven difficult to assemble. Here we present a draft assembly for *Phyllobates terribilis* (12.6Gb), generated using a combination of sequencing platforms and bioinformatic approaches. Similar to other poison frog sequencing efforts, we recovered a highly fragmented assembly, likely due to the genome's large size and very high repetitive content, which we estimate to be ≈ 88%. Despite the assembly's low contiguity, we were able to annotate multiple members of three gene sets of interest (voltage-gated sodium channels and *Notch* and *Wnt* signaling pathways), demonstrating the usefulness of our assembly to the amphibian research community.

1 Introduction

The ability to generate genome assemblies is now available for virtually any organism from which DNA of a reasonable quality can be obtained. This has led to an explosion of sequencing efforts across the tree of life [e.g. 1–5]. In vertebrates, these efforts have led to significant comparative genomic coverage in some groups [3]. For instance, high-quality

assemblies are available for at least one species in more than 92% of bird families [6]. In amphibians, however, this is not the case. Despite considerable efforts, as of April 2024 assemblies for only 121 species (1.4%) from 31 families (40%) were available in NCBI's database [7]. However, only 63 of them have N50 values $\geq$ 1Kb, accounting for 0.7% of amphibian species (36% of families). The slow progress in amphibian genomics is due in large part to the technical challenges posed by the large, highly repetitive genomes of many amphibian species [7–10].

Among amphibians the genomes of poison frogs in the family Dendrobatidae have been notoriously challenging to assemble, given their often large sizes (up to 13Gb [11]), as well their high content and widespread distribution of repetitive elements [10, 12]. Over the past few decades, several features of this family's biology, such as a wide variety of parental care strategies [13], multiple cases of exuberant intraspecific variation in coloration [e.g. 14–16], and the recurrent evolution of alkaloid-based chemical defense coupled with aposematic coloration [17–19], have garnered the attention of many ecologists and evolutionary biologists. However, the functional and molecular mechanisms underlying the majority of these phenotypes and their evolution remain largely unknown. The development of molecular tools for functional genetic and evolutionary studies in poison frogs has, nonetheless, been slow, in part due to the challenges involved in genome assembly. Currently, assemblies are only available for four closely-related species: *Oophaga pumilio* [12], *Oophaga sylvatica*, *Dendrobates tinctorius* [20], and *Ranitomeya imitator* [21]. An assembly for the more distantly related *Allobates femoralis* (family Aromobatidae), has recently become available.

Here we present a draft assembly for a fifth dendrobatid species, *Phyllobates terribilis*, whose genome was recently estimated to be around 12Gb in size [11]. Frogs in the genus *Phyllobates* are well known for secreting batrachotoxins (BTXs [22, 23]), a group of potent neurotoxins that target voltage-gated sodium channels [24–26], as well as their dynamically evolving warning colorations [23, 27]. Further, the phylogenetic position of the genus *Phyllobates* as sister to a radiation of more than 50 chemically defended, aposematic species [19] make this clade crucial to understanding the evolutionary origin of alkaloid-based chemical defense in dendrobatids. These features make *Phyllobates* frogs well-suited models for studies in multiple fields that can certainly benefit from improved genomic resources, such as neurotoxin resistance and physiology [28–30], chemical ecology [31–33], and evolutionary genetics [27, 34].

2 Methods

2.1 Animal subjects

All animals used for tissue sampling were acquired from the pet trade (Josh's Frogs, Owosso, MI. USA) or from laboratory colonies maintained at the University of Chicago. Individuals were euthanized through either an overdose of topical benzocaine followed by pithing, or progressive cooling and flash-freezing in liquid nitrogen. Thigh muscle, tadpole tail muscle, or liver samples were then dissected in phosphate-buffered saline (PBS), and

either stored at -80°C, or immediately used for DNA extraction as detailed below. Frozen samples were processed within 6 months of collection. All animal work was approved by the University of Chicago and John Carroll University's Institutional Animal Care and Use Committees (UChicago protocol # 72416; JCU protocol # 1400).

2.2 DNA sequencing

We combined multiple sequencing strategies to produce a hybrid genome assembly. First, we generated paired-end (PE) and mate-paired (MP) Illumina libraries, as detailed by Jacob Machado et al. [35]. Briefly, DNA was extracted from thigh muscle with Qiagen MagAttract HMW DNA extraction kits (cat. no. 67563), and PE and MP libraries were prepared using the TruSeq Nano DNA and Nextera Mate-pair kits, respectively. DNA for MP libraries was gel size-selected to insert sizes of 3, 5, 8, and 10Kb prior to library preparation. All libraries were sequenced on HiSeq 2500 instruments (RRID:SCR_016383) with 125bp or 150bp reads.

Next, we used the Pacific Biosciences Sequel platform (RRID:SCR_017989) to generate long single-molecule real-time (SMRT) sequencing reads. High-molecular weight DNA was extracted from ∼70mg of tadpole tail muscle using a Qiagen DNeasy column, and eluted in 50μl Qiagen AE buffer previously heated to 60°C. After QC on an Agilent TapeStation (RRID:SCR_014994), one large-insert (10kb) library was prepared and sequenced in 16 SMRT 1M cells at the Duke University Sequencing and Genomic Technologies Shared Resource.

Finally, we produced an *in-vivo* Hi-C library using the Proximo Hi-C Animal kit (Phase Genomics). The starting material was a mix of ∼50mg tail muscle and ∼20mg liver tissue from a single tadpole, and the manufacturer's protocol was followed with no modifications. The resulting library was size-selected for 300–700bp fragments using SPRI magnetic beads (Sera-Mag), and sequenced on an Illumina HiSeq 4000 (RRID:SCR_016386) with 100bp reads. All Illumina reads were quality-trimmed using Trimmomatic v. 0.39 (RRID:SCR_011848) [36], except for MaSuRCA v. 3.3.4 runs (see below). PacBio reads were converted to fasta format, filtered for sequences longer than 100bp using BamTools (RRID:SCR_015987) [37], and error-corrected based on the Illumina paired-end reads with FMLRC [38]. Table 1 contains details and accession information for all sequence data used in this project.

2.3 Assembly pipeline

As an first approach to generate a starting assembly, we used WTDBG v. 2.3 (RRID:SCR_017225) [39] with the PacBio reads as input. We used a k-mer size of 19bp, and in order to account for our sequencing coverage, increased the k-mer sampling rate to 1/2 (-S 2) and retained contained reads for alignment (-A flag). This, however, resulted in a much smaller assembly than expected (1.96Mb), likely due to the lower than ideal coverage of our PacBio data. We therefore used MaSuRCA v. 3.3.4 (RRID:SCR_010691) [40, 41] to produce a starting assembly from the Illumina paired-end and mate-paired reads. Reads were not quality-trimmed, following developer guidelines. K-mer size for De Brujin graphs

Table 1. Sequence data used in the *Phyllobates terribilis* assembly. Numbers correspond to raw data. Estimated coverage was calculated as the product of read length and number of reads divided by the estimated genome size (12.8Gb [11]). PE: paired-end, MP: mate-paired, $\bar{x}$: mean, SD: standard deviation.

Library	Read Length	No. Reads	Estimated Coverage	SRA Accession
Illumina PE	150bp	263,828,699	6.28X	SRR27279919
Illumina PE	125bp	210,072,773	4.17X	SRR27279918
Total PE		**473,901,472**	**10.45X**	
Illumina MP - 3Kb	150bp	240,701,531	5.64X	SRR27279914
Illumina MP - 3Kb	125bp	126,584,808	2.47X	SRR27279916
Illumina MP - 5Kb	150bp	247,545,514	5.80X	SRR27279913
Illumina MP - 5Kb	125bp	109,366,991	2.14X	SRR27279915
Illumina MP - 8Kb	150bp	252,654,530	5.92X	SRR27279912
Illumina MP - 10Kb	125bp	86,657,163	1.69X	SRR27279917
Total MP		**1,063,510,537**	**23.67X**	
Illumina *in-vivo* Hi-C	100bp	919,642,651	14.37X	SRR27279911
P. bicolor RNAseq	100bp	256,624,922	NA	SRR12232938
PacBio Sequel	$\bar{x}$: 5,143bp SD: 5008bp	11,945,937	4.80X	SRR27279910

was determined automatically, and the coverage of mate-paired libraries was limited to 300X. The MaSuRCA assembly was further scaffolded and gap-filled through a series of complementary approaches. First we downloaded RNAseq data from a closely related species, *Phyllobates bicolor*, generated in a previous study (SRA accession SRX8741407, BioProject PRJNA645960 [27]), quality trimmed it as detailed above, and used P_RNA_scaffolder [42] for RNAseq-guided scaffolding. *P.hyllobates terribilis* and *P. bicolor* shared a common ancestor roughly 2 million years ago [27]. We then polished the resulting assembly using the Illumina PE reads with Nextpolish (RRID:SCR_025232) [43], and ran four iterations of the following pipeline: HiC-based scaffolding with SALSA 2 (RRID:SCR_022013) [44, 45], PacBio-based gap-filling and scaffolding with LR_gapcloser (RRID:SCR_016194) [46] and RAILS [47], and polishing with NextPolish (based on Illumina PE). At the end of the fourth iteration we ran an additional three rounds of NextPolish. Read alignments for all steps were done with bwa v.0.7.17 (RRID:SCR_010910) [48], and alignment files were handled with samtools v. 1.11 [49]. Figure 1 summarizes our assembly pipeline.

Assembly contiguity statistics were calculated using quast v. 5.1 (RRID:SCR_001228) [50], base error rates (QV) were evaluated with GAEP [51], and completeness was assessed using BUSCO v. 5.3.2 (RRID:SCR_015008) [52, 53]. QV scores were calculated based on read mapping and called genotypes following Rhie et al. [4], using all available Illumina reads (except RNAseq). Mate-paired and HiC reads were not paired for mapping, and genotypes were called using bcftools (RRID:SCR_005227) [54]. BUSCO used the tetrapoda_odb10 gene set, and was run under default parameters. Contiguity, QV, and completeness statistics evaluated at multiple points of our assembly pipeline are presented in Figs S1-2. Upon upload to NCBI, contigs showing signs of possible

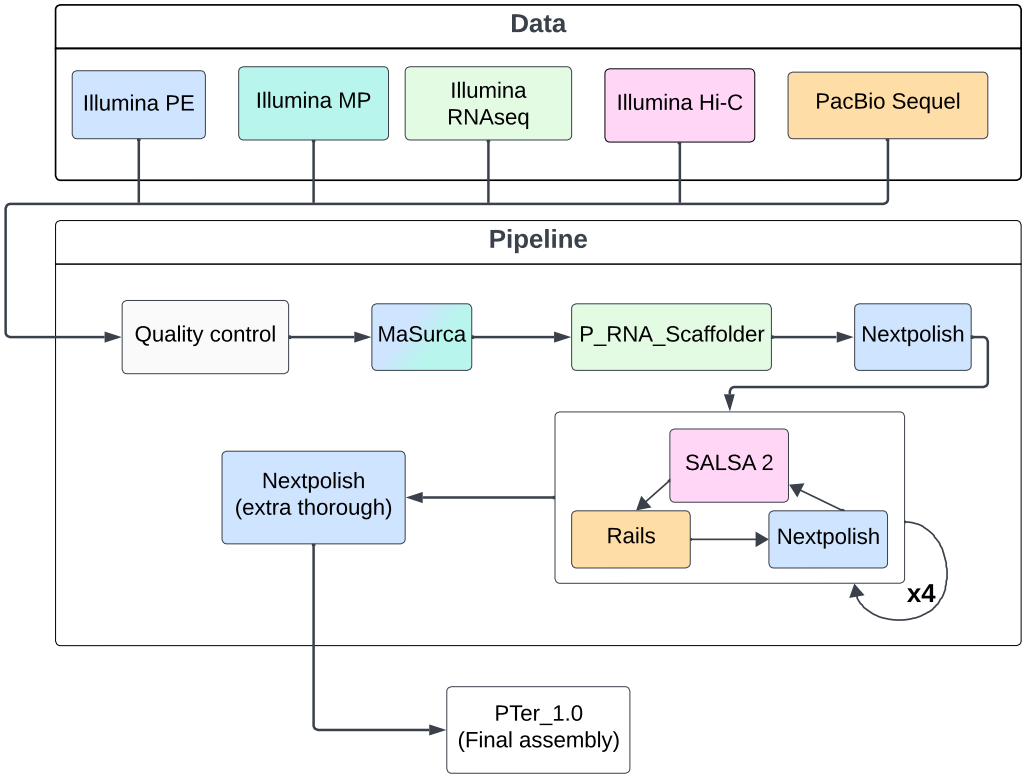

Figure 1. Pipeline used to generate a reference genome assembly for *Phyllobates terribilis*. Process boxes are filled with the same color as their input data.

contamination or shorter than 200bp were removed.

To gain insight on the degree of missassembly due to repetitive content (e.g. multiple repeats being collapsed into a single sequence), we estimated copy numbers for regions annotated as repetitive by RepeatMasker (details below) using DepthKopy [55], which uses sequencing depth at complete BUSCO genes as the expectation for single-copy regions, to then estimate copy number at other regions of the assembly. In addition, we examined copy number variation in non-overlapping 5Kb windows along the four longest scaffolds (17.4Mb), of which 8.62Mb (49.5%) were annotated as repetitive (see Repeat assembly and masking section below). Finally, to test for inflated coverage in repetitive regions, we compared the copy number of regions annotated as repetitive and single-copy BUSCO genes using negative binomial regression, as implemented in the R package MASS [56]. DepthKopy outputs copy number estimates on a continuous scale, so they were rounded to integers to match the discrete nature of the negative binomial distribution.

2.4 Repeat assembly and masking

We used our assembly and raw paired-end reads to characterize the composition and distribution of repetitive elements in the *P. terribilis* genome. First, we used REPdenovo v. 0.1.0 [57, 58] to identify reads originating from repetitive elements, assemble consensus sequences for these repeats, and estimate their copy number based on read depth. Next, we used RepeatModeler v. 1.0.11 (RRID:SCR_015027) [59] to generate a species-specific repeat library for *P. terribilis*. We merged the repeat sequences from REPdenovo and RepeatModeler with those available in RepBase (2018 version; [60]) and the curated Dfam sequences distributed with RepeatMasker 4.0.8 [61], and used RepeatMasker to annotate and mask them on the assembly. The repeat-masked assembly was used in all subsequent annotation steps. Consensus repetitive sequences identified by REPdenovo were annotated by blasting [62] against the combined Dfam and RepBase nucleotide and protein databases (E-value$\leq 10^{-3}$), and retaining the hit with the highest bit score. The contribution of each consensus element to the genome-wide repetitive content was estimated by multiplying the length of the element by its mean coverage, and dividing this value by the total amount of sequence in the paired-end reads used as input. Finally, to assess how well the REPdenovo sequences were incorporated into our assembly, we queried them against the (unmasked) scaffolds using blastn (RRID:SCR_001598) (E-value$\leq 10^{-10}$).

2.5 Gene prediction and annotation

We generated gene structure predictions using BRAKER v. 2.15 (RRID:SCR_018964) [63], based on a database of known proteins derived from UniProt's SwissProt and NCBI's RefSeq, and the *P. bicolor* RNAseq reads (see Table 1) aligned to the repeat-masked assembly. A transcriptome assembly generated previously from these reads [27] contained 80.8% of genes in the BUSCO tetrapoda_odb10 gene set (76.9% complete, 3.9% fragmented). The known protein database was generated by concatenating the SwissProt and RefSeq proteins for chordates (taxon code 7711), and removing duplicate or nested sequences, as well as those with duplicate headers, shorter than 32 amino acids, or marked as "partial" or "low quality". The filtered database consisted of 768,857 amino acid sequences. RNAseq reads were aligned using STAR 2.7.9a [64], and BRAKER2 was run on a single core to avoid parallelization problems associated with fragmented assemblies. To evaluate the extent of gene representation and fragmentation in our assembly and annotation we ran BUSCO on the resulting protein sequences as detailed above, and blasted the aforementioned *P. bicolor* transcriptome [27] against our assembly and annotation. Gene representation was assessed based on the proportion of transcripts with blast hits (E-value$\leq 10^{-10}$), and completeness as the proportion of these transcripts that had query coverages of at least 75% (qcovs statistic from BLAST).

We then annotated the BRAKER2 gene sequence predictions by aligning and comparing them with multiple protein function and gene ontology (GO) databases. First, we used DIAMOND (RRID:SCR_016071) [65] to query gene predictions against the NCBI's non-redundant (NR) database, and InterProScan (RRID:SCR_005829) [66] to generate an initial prediction of protein function. We then used Blast2Go (RRID:SCR_005828) [67] to perform GO mapping and annotation based on the DIAMOND and InterProScan results,

and extracted GO subsets (i.e. GO slims) based on the "generic GO subset" list available at https://geneontology.github.io/docs/download-ontology. Gene predictions and annotations are available in the GigaDB repository associated with this paper.

2.6 Targeted gene annotation

In addition to a general annotation of genes in our assembly, we evaluated its applicability through a systematic search for three gene sets of particular interest: The voltage-gated sodium channels, which have been frequently studied in the context of toxin resistance in poison frogs (e.g. [12, 29, 30, 68]) and other animals [e.g. 69–72], and genes involved in the *Notch* and *Wnt* signaling pathways, which exhibit a high degree of conservation across the animal kingdom [73–76]. The *Notch* pathway plays a pivotal role in regulating a spectrum of fundamental cellular processes, including differentiation, fate specification, proliferation, programmed cell death, and tissue patterning, and has been extensively characterized across a variety of species at both embryonic and post-embryonic stages [74–80]. The *Notch* and *Wnt* pathways interact in diverse molecular, cellular, and developmental contexts, which have also received substantial attention [81–83]. A comprehensive enumeration of selected genes pertinent to the *Notch* and *Wnt* signaling pathways is presented in Figure 2.

Our goal here was to showcase how, regardless of its suboptimal state (see Results and discussion), our assembly can still be used to annotate genes of interest. We note, however, that our approach collapses closely related paralogous genes with conserved structure and function into a single gene name since establishing orthology for these genes across species requires detailed annotation beyond the scope of this paper. For example, some gene names like "*Wnt*" represent a collection of paralogs, which vary in number and identity across species. With this in mind, the results of our targeted annotation must be interpreted bearing in mind that we cannot distinguish between finding some or all members of a closely related gene family. Moreover, it is important to emphasize that our targeted annotation approach was designed specifically to validate the presence of genes of interest rather than assess conventional metrics of genome completeness or gene model quality. Unlike traditional comparative analyses with well-annotated reference genomes (e.g., *Xenopus tropicalis*), our methodology focuses on gene presence validation through multiple independent methods. Many of the gene families we analyzed (such as voltage-gated sodium channels and Wnt signaling components) exhibit substantial variability in domain structure and sequence length, even among well-assembled genomes, making standardized length and completeness comparisons challenging. This inherent complexity in these gene families means that attempting to quantify completeness or provide consistent domain composition metrics would require extensive computational processing beyond the scope of this manuscript. This multi-method validation approach provides increased confidence in our annotations despite the fragmented nature of the assembly.

We employed three different sequence search algorithms to query a set of reference sequences for our genes of interest against the *P. terribilis* BRAKER2 gene predictions and assembly scaffolds, and then used NCBI's Conserved Domain Database (CDD; [84]) to identify conserved domains in the resulting hits in order to confirm that they represented actual matches to the target proteins. The reference database was generated from mouse,

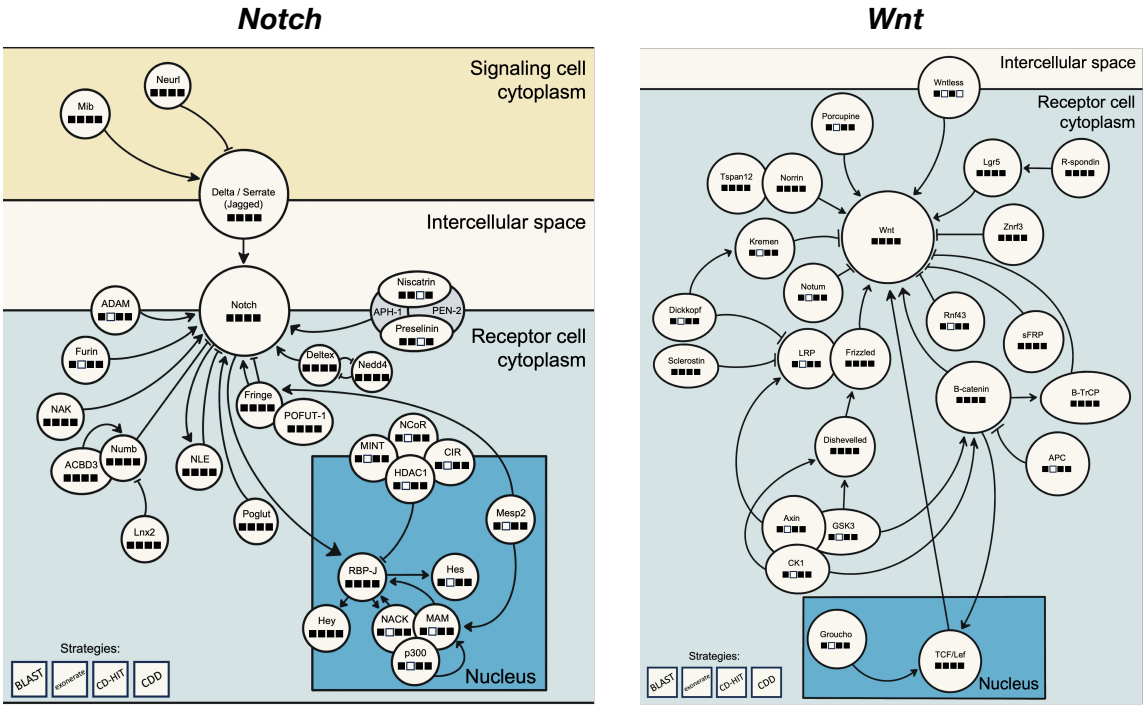

Figure 2. Simplified diagrams of the *Notch* and *Wnt* pathways, whose genes we targeted for systematic annotation. Boxes below gene names indicate whether a protein product was found using each of the three different search strategies employed, and whether they fulfilled the conserved domain requirement (see Targeted gene annotation section for details). Some gene names, like "*Wnt*," group proteins with similar structures and functions encoded by different closely related genes. Filled boxes denote a positive result.

human, and amphibian sequences on RefSeq, Uniprot, Xenbase, and previous publications [12], and was filtered as detailed above (see Gene prediction and Annotation section). Initial sequence searches were conducted using tblastn (RRID:SCR_011822) [62], exonerate (RRID:SCR_016088) [85], and (RRID:SCR_007105) [86]. The resulting hits were then queried against the CDD using NCBI's CD-Search online portal with an E-value cutoff of 0.01, and adjusting scores for sequence composition. Genes with conserved domains that departed from their putative function (i.e. *Notch* or *Wnt* signaling, or voltage-gated sodium channel) were discarded. The quality of a hit in our targeted annotation approach should be assessed by the number of independent sources confirming the presence of a gene (or gene family) in our assembly.

3 Results and Discussion

3.1 Genome assembly and annotation

The final assembly, **PTer_1.0**, spans 4.24Gb, and has a scaffold N50 of 11,957bp, L50 of 93,411 fragments, and GC content of 42.26%. Scaffolds range in length from 63 to 5,200,876

Table 2. Contiguity and completeness statistics for the *P. terribilis* assembly and annotation.

	Length	**N50**	**L50**	**QV**	**GC %**
Contigs	2.21Mb	1.06kb	601,531	28.05	42.26%
Scaffolds	4.24Mb	11.96kb	93,411	28.02	42.26%
BUSCO	**Complete SC**	**Complete Dupl.**	**Fragmented**	**Missing**	
Assembly - Unmasked	18.5%	0.4%	21.1%	60.0%	
Assembly - Masked	17.4%	0.3%	20.9%	61.4%	
Annotation	14.0%	0.3%	23.4%	61.3%	

bp. The final BRAKER2 annotation contains 45,969 protein-coding sequences with an average length of 211 amino acids (SD: 217.5aa). BUSCO assessment of found 40% of genes (18.9% complete, 21.1% fragmented) in the unmasked genome assembly and 38.7% of genes in the annotation (15.3% complete, 23.4% fragmented). Additional assembly statistics are available in Table 2, and statistics from intermediate steps in our pipeline are available as supplementary material (Fig S1-2). The fragmentation in our assembly is probably due to a combination of factors. First of all, the highly repetitive nature of the *P. terribilis* genome is bound to generate assembly issues. Beyond this, its very large size means that our sequencing effort, despite being considerable, resulted in suboptimal depth, especially in terms of long-read data. Future assembly attempts should incorporate higher long read depth, as well as longer reads that are able to transverse entire repetitive regions.

As with other poison frog assemblies based primarily on short-read data, the size of our assembled sequence is considerably smaller than genome size estimates based on read depth or DNA quantification. The current *Oophaga pumilio* reference assembly on GenBank (accession GCA_009801035.1 [12]) is 3.5Gb, while the genome size for this species has been estimated at 4.3-4.7Gb based on Feulgen staining [11, 87]. Our *P. terribilis* assembly is 4.2Gb, while a genome size estimate based on BUSCO gene read depth (generated with DepthSizer (RRID:SCR_021232) [55]) was between 10.1-17.6Gb, and the average DNA content in *P. terribilis* nuclei has been estimated at 12.88pg using flow cytometry [11], which equates to 12.6Gb. Despite this discordance, the vast majority of Illumina PE (96.3%) and PacBio (95.1%) reads successfully mapped to our assembly. However, we found pronounced variation in copy numbers across the four largest scaffolds, with some 5Kb windows reaching values above 1,000. Repetitive regions annotated by RepeatMasker had significantly higher copy numbers than single-copy BUSCO genes (negative binomial regression: $\beta = 0.712, z = 4.914, p = 8.94 \times 10^{-7}$; Fig. 3). In view of these results, and considering the high repetitive content of this and other dendrobatid genomes [10, 12, 20, 21], we consider it likely that the discrepancy between genome size and assembly size is due to multiple repetitive regions of the genome being collapsed into single sequences during assembly. This may also cause breaks in the assembly, which would explain the low contiguity. The fact that we also find a large discordance in repetitive element content and composition between the assembly and raw reads (see the Repetitive Content Section below) further supports these hypotheses, and highlights the challenge repetitive regions pose for dendrobatid frog genome assembly.

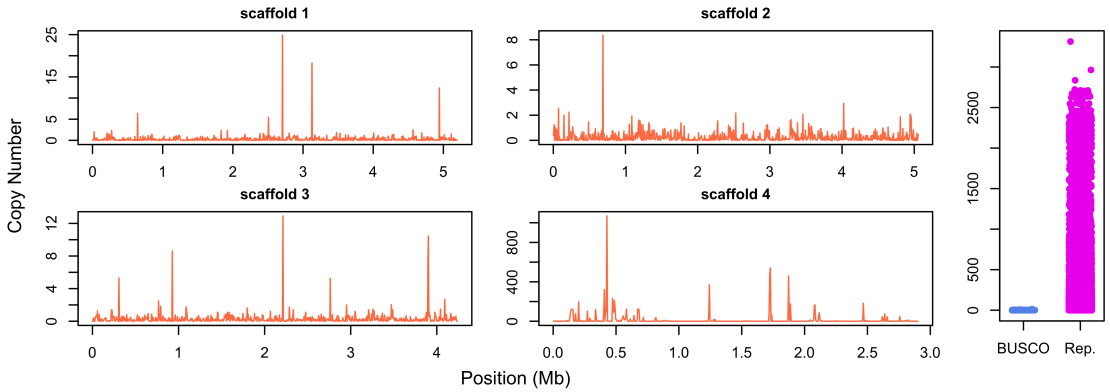

Figure 3. Copy number variation across the four longest scaffolds of our assembly (left and center panels) and between silge-copy BUSCO genes and repetitive regions annotated by RepeatMasker (right panel), estimated using DepthKopy [55].

3.2 Targeted gene annotation

Despite its fragmentation, our assembly contained most of the *P. bicolor* transcripts, and all the target genes (or gene families) in our three sets of interest. Ninety nine percent of the *P. bicolor* transcripts had BLAST hits against the genome, and 96% against the predicted protein sequences from our annotation. However, only 25% (genome) and 4.4% (annotation) of these hits spanned at least 75% of the query transcript within the same scaffold/peptide. In the same vein, all voltage-gated sodium channels and at least one member of the gene families involved in the *Wnt*/*Notch* pathways were represented in our assembly. Their sequences contained conserved domains concordant with their putative function, indicating that these sequences likely represent true members of their respective gene families (Fig. 2). The strength of our annotation approach lies in the use of multiple independent validation methods (tblastn, exonerate, and CD-HIT, followed by conserved domain confirmation), which provides robust evidence for gene presence even in a fragmented assembly context. The identification of characteristic conserved domains confirms their identity and functional relevance.

Given the low completeness of transcript BLAST hits and the BUSCO statistics reported above, we anticipate that many of the gene families annotated are likely represented as fragments rather than complete sequences, and that at least some gene families will have missing genes. Without further detailed annotation we cannot reliably evaluate how many unique orthologs within a gene family were annotated. Finally, the results above suggest that the low BUSCO scores obtained are at least partially due to fragmentation hindering BUSCO's ability to annotate genes, rather than the absence of at least partial or fragmented sequences of these genes in the assembly. In any case, these results highlight that, while suboptimal, our assembly remains a valuable resource for genetic research in dendrobatids and vertebrates in general.

3.3 Repetitive content

Repetitive elements identified by RepeatMasker make up 1.49Gb (35.17%) of the assembly, with the majority (1.39Gb; 32.9% of the assembly) remaining unclassified after annotation. REPdenovo, on the other hand, identified 58.35Gb of repetitive sequence in our 65.83Gb of PE reads (88.6%), suggesting a much higher repetitive content in the *P. terribilis* genome than what is currently in our assembly (Table 3). In addition to repetitive element collapsion during assembly, this discordance is likely to be exacerbated by the high number of gaps in the assembly (Table 2), which may have precluded the discovery of repetitive elements by RepeatMasker. With this in mind, and considering that the most available assemblies of dendrobatid frogs are over 70% repetitive (Table 3 [10, 20, 21]), and that the *P. terribilis* genome is considerably larger than both of these species [11], 88% is likely a closer estimate of the true repetitive content.

REPdenovo assembled the repeat-containing reads into 126,314 consensus sequences, of which 114,283 (90%) had BLAST hits against our assembly spanning at least 95% of the repeat with $\geq 98\%$ identity. This, again, indicates that most repetitive elements were indeed incorporated into the assembly, but were collapsed into a small subset of sequences due to their high degree of similarity. Blasting to the RepBase and Dfam protein libraries allowed us to annotate 22,952 of the REPdenovo repetitive elements, which together accounted for 17.12Gb of the 58.35Gb identified as repetitive. Among these, DNA, LTR, and LINE elements were most prevalent , similar to other poison frog assemblies (Table 3 [12, 20, 21]).

Our finding that a considerable portion of repetitive elements was unclassifiable through comparisons with commonly used repeat databases has also been obtained by several other amphibian genome assembly efforts [10]. Although the fact that amphibian repetitive elements are not well represented in the databases used could in part explain this result, recent studies (e.g. [21, 88]) as well as our own ongoing work using recently generated amphibian repetitive element libraries have found similar results, suggesting this may not be the case. Whether some unique feature of anuran repetitive elements, such as novel TE families, is behind the challenges with their classification remains to be determined. In-depth investigation of these unannotated sequences should, in any case, improve our ability to annotate and understand the evolution of repetitive elements in eukaryotic genomes.

4 Concluding Remarks

With the rapid accumulation of genome assemblies for non-traditional model species, comparative genomics has gained ground as a powerful approach across the biological sciences. Our draft genome assembly for *P. terribilis* will contribute to research efforts in a variety of fields, including systematics, phenotypic evolution, chemical ecology, developmental biology, molecular physiology, and sensory ecology. Despite a considerable multi-platform sequencing effort, our assembly remains highly fragmented, likely due to its large size and the rampant proliferation of repetitive elements, which comprise as much as 88% of the genomic sequence. Although its current level of fragmentation is certainly an

Table 3. Composition of repetitive elements identified from the assembly by RepeatMasker and from pair-ended reads by REPdenovo. RepeatMasker values are based on an assembly length of 4.24Gb. Sequence lengths and percentages from REPdenovo correspond to values across the 65.83Gb of paired-end reads used as input. Assuming the PE reads represent an unbiased sampling from the genome, percentages can be interpreted as estimates of the genome-wide repeat composition. Repeat contents and assembly sizes for two closely related species, *Allobates femoralis* and *Oophaga sylvatica*, are included, as reported in Table S4 by Kosch et al. [7].

Class	**Subclass**	**Repeat Masker**		**REPdenovo**		**Other species (percents)**	
		Mb in genome	Percent	Gb in reads	Percent	*A. femoralis* (5.3Gb)	*O. sylvatica* (5.2Gb)
DNA		49.5	1.17%	8.49	12.9%	11.86%	12.71%
	TcMar-Tc1	0.63	0.01%	4.78	7.3%	8.10%	6.82%
	hAT-Ac	0.18	<0.01%	3.38	5.1%	2.80%	4.91%
	Other	48.7	1.16%	0.33	0.5%	0.96%	0.98%
LTR		19.6	0.46%	5.24	8.0%	7.07%	24.46%
	Gypsy	<0.01	<0.01%	4.35	6.6%	5.82%	21.82%
	Other	19.6	0.46%	0.89	1.4%	1.25%	2.64%
LINE		29.1	0.69%	2.76	4.2%	8.06%	6.08%
	CR1	6.60	0.16%	0.83	1.3%	3.44%	3.35%
	L1	2.84	0.07%	0.96	1.5%	2.87%	0.51%
	Other	19.66	0.46%	0.97	1.5%	1.75%	2.73%
Other		<0.01	<0.01%	0.62	0.95%	1.94%	2.99%
Unclassified		1,393.1	32.85%	41.23	62.6%	44.59%	35.39%
Total Repetitive		1,491.3	35.17%	58.35	88.6%	73.52%	81.63%

obstacle for some studies, such as those that rely on scoring features across long contiguous stretches of DNA (e.g. ancestry tracts, runs of homozygosity), our analyses suggest that, even in its current form, our assembly represents a valuable resource. As DNA sequencing technologies continue to improve and become more cost-efficient, we are confident that this work will constitute a key stepping stone towards chromosome-level assemblies for highly repetitive amphibian genomes.

5 Data Availability

Our assembly and sequencing data are available under NCBI BioProject no. PRJNA1054463. Raw reads have SRA accessions SRR27279910-SRR27279919, as detailed in Table 1. The final assembly, PTer_1.0 has been assigned WGS accession JBBPXS01, and is available under GenBank accession **GCA_045270155.1**. The assembly (repeat-masked and unmasked), gene predictions and annotations, and REPdenovo and RepeatMasker output files, along with the code used to generate the assembly, are available in the GigaDB repository associated to this manuscript [89].

5.1 Funding

This research was funded by the São Paulo Research Foundation (grant nos. 2012/10000-5, 2013/05958-8, 2015/18654-2, and 2018/15425-0), the Brazilian Conselho Nacional de Desenvolvimento Científico e Tecnológico (grant no. 314480/2021-8), the US National Science Foundation (grant nos. DEB-1702014 and IOS-1827333), the US National Institute of General Medical Sciences (grant no. R35GM131828), and the University of Chicago's Hinds Fund. Computations were supported in part by Advanced Research Computing at Virginia Tech, and the Center for Research Informatics at the University of Chicago. We acknowledge funding and logistical support from several entities of the University of North Carolina at Charlotte including: The Bioinformatics Services Division, the Department of Bioinformatics and Genomics, the Bioinformatics Research Center, University Research Computing, the College of Computing and Informatics. RM was supported by the Michigan Society of Fellows.

6 Acknowledgements

We are grateful to Gabriel Massami Izumi de Freitas and Esdras Matheus Gomes da Silva for technical assistance, and to Carrie Olson-Manning and Adam Stuckert for advice with PacBio bioinformatics.

References

1. Robinson GE, Hackett KJ, Purcell-Miramontes M, Brown SJ, Evans JD, Goldsmith MR, et al. Creating a buzz about insect genomes. Science. 2011;331(6023):1386–1386. doi:10.1126/science.331.6023.1386.

2. Hamilton JP, Robin Buell C. Advances in plant genome sequencing. Plant Journal. 2012;70(1):177–190. doi:10.1111/j.1365-313X.2012.04894.x.

3. Cheng S, Melkonian M, Smith SA, Brockington S, Archibald JM, Delaux PM, et al. 10KP: A phylodiverse genome sequencing plan. GigaScience. 2018;7(3):1–9. doi:10.1093/gigascience/giy013.

4. Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, et al. Towards complete and error-free genome assemblies of all vertebrate species. Nature. 2021;592(7856):737–746. doi:10.1038/s41586-021-03451-0.

5. Lewin HA, Richards S, Aiden EL, Allende ML, Archibald JM, Bálint M, et al. The Earth BioGenome Project 2020: Starting the clock. Proceedings of the National Academy of Sciences. 2022;119(4):e2115635118. doi:10.1073/pnas.2115635118.

6. Feng S, Stiller J, Deng Y, Armstrong J, Fang Q, Reeve AH, et al. Dense sampling of bird diversity increases power of comparative genomics. Nature. 2020;587(7833):252–257. doi:10.1038/s41586-020-2873-9.

7. Kosch TA, Torres-Sánchez M, Liedtke HC, Summers K, Yun MH, Crawford AJ, et al. The Amphibian Genomics Consortium: advancing genomic and genetic resources for amphibian research and conservation. BMC Genomics. 2024;25(1):1025. doi:10.1186/s12864-024-10899-7.

8. Sun YB, Zhang Y, Wang K. Perspectives on studying molecular adaptations of amphibians in the genomic era. Zoological Research. 2020;41(4):351–364. doi:10.24272/j.issn.2095-8137.2020.046.

9. Funk WC, Zamudio KR, Crawford AJ. Advancing understanding of amphibian evolution, ecology, behavior, and conservation with massively parallel sequencing. In: Hohenlohe PA, Rajora OP, editors. Population Genomics: Wildlife. Cham: Springer International Publishing; 2021. p. 211–254. Available from: https://doi.org/10.1007/13836_2018_61.

10. Kosch TA, Crawford AJ, Lockridge Mueller R, Wollenberg Valero KC, Power ML, Rodríguez A, et al. Comparative analysis of amphibian genomes: An emerging resource for basic and applied research. Molecular Ecology Resources. 2025;25(1):e14025. doi:https://doi.org/10.1111/1755-0998.14025.

11. Douglas TE, Márquez R, Holmes VR, Johnston JS, Tarvin RD. Genome size evolution and phenotypic correlates in the poison frog family Dendrobatidae. Evolution. 2025;In Press:qpaf011. doi:10.1093/evolut/qpaf011.

12. Rogers RL, Zhou L, Chu C, Márquez R, Corl A, Linderoth T, et al. Genomic takeover by transposable elements in the Strawberry poison frog. Molecular Biology and Evolution. 2018;35(12):2913–2927. doi:10.1093/molbev/msy185.

13. Summers K, Tumulty J. Parental care, sexual selection, and mating systems in neotropical poison frogs. In: Macedo RH, Machado G, editors. Sexual Selection. San Diego: Academic Press; 2014. p. 289–320. Available from: https://doi.org/10.1016/B978-0-12-416028-6.00011-6.

14. Silverstone PA. A revision of the poison-arrow frogs of the genus *Dendrobates* Wagler. Natural History Museum of Los Angeles County, Science Bulletin. 1975;21:1–55.

15. Myers CW, Daly JW. Preliminary evaluation of skin toxins and vocalizations in taxonomic and evolutionary studies of poison-dart frogs (Dendrobatidae). Bulletin of the American Museum of Natural History. 1976;157:173–262.

16. Hoogmoed MS, Avila-Pires TCS. Inventory of color polymorphism in populations of *Dendrobates galactonotus* (Anura: Dendrobatidae), a poison frog endemic to Brazil. Phyllomedusa. 2012;11:95–115.

17. Vences M, Kosuch J, Boistel R, Haddad CFB, Marca EL, Lötters S, et al. Convergent evolution of aposematic coloration in Neotropical poison frogs: A molecular phylogenetic perspective. Organisms Diversity & Evolution. 2003;3:215–226.

18. Santos JC, Coloma LA, Cannatella DC. Multiple, recurring origins of aposematism and diet specialization in poison frogs. Proceedings of the National Academy of Sciences of the United States of America. 2003;100:12792–12797. doi:10.1073/pnas.2133521100.

19. Grant T, Rada M, Anganoy-Criollo M, Batista A, Dias PH, Jeckel AM, et al. Phylogenetic systematics of Dart-Poison frogs and their relatives revisited (Anura: Dendrobatoidea). South American Journal of Herpetology. 2017;12(s1):S1–S90. doi:10.2994/SAJH-D-17-00017.1.

20. Dittrich C, Hoelzl F, Smith S, Fouilloux CA, Parker DJ, O'Connell LA, et al. Genome Assembly of the Dyeing Poison Frog Provides Insights into the Dynamics of Transposable Element and Genome-Size Evolution. Genome Biology and Evolution. 2024;16(6):evae109. doi:10.1093/gbe/evae109.

21. Stuckert AMM, Chouteau M, McClure M, LaPolice TM, Linderoth T, Nielsen R, et al. The genomics of mimicry: Gene expression throughout development provides insights into convergent and divergent phenotypes in a Müllerian mimicry system. Molecular Ecology. 2024;33(14):e17438. doi:https://doi.org/10.1111/mec.17438.

22. Märki F, Witkop B. The venom of the Colombian arrow poison frog *Phyllobates bicolor*. Experientia. 1963;19:329–338. doi:10.1007/bf02152303.

23. Myers CW, Daly JW, Malkin B. A dangerously toxic new frog (*Phyllobates*) used by Emberá Indians of western Colombia, with discussion of blowgun fabrication and dart poisoning. Bulletin of the American Museum of Natural History. 1978;161:307–366.

24. Albuquerque EX, Warnick JE, Sansone FM, Daly JW. The pharmacology of batrachotoxin. V. A comparative study of membrane properties and the effect of batrachotoxin on sartorius muscles of the frogs *Phyllobates aurotaenia* and *Rana pipiens*. The Journal of Pharmacology and Experimental Therapeutics. 1973;184:129–315.

25. Warnick JE, Alburquerque EX, Onur R, Jansson SE, Daly J, Tokuyama T, et al. The pharmacology of batrachotoxin. VII. Structure-activity relationships and the effects of pH. The Journal of Pharmacology and Experimental Therapeutics. 1976;193:232–245.

26. Wang SY, Mitchell J, Tikhonov DB, Zhorov BS, Wang GK. How batrachotoxin modifies the sodium channel permeation pathway: Computer modeling and site-directed mutagenesis. Molecular Pharmacology. 2006;69:788–795. doi:10.1124/mol.105.018200.

27. Márquez R, Linderoth TP, Mejía-Vargas D, Nielsen R, Amézquita A, Kronforst MR. Divergence, gene flow, and the origin of leapfrog geographic distributions: The history of colour pattern variation in *Phyllobates* poison-dart frogs. Molecular Ecology. 2020;29(19):3702–3719. doi:10.1111/mec.15598.

28. Daly JW, Myers CW, Warnick JE, Albuquerque EX. Levels of batrachotoxin and lack of sensitivity to its action in poison-dart frogs (*Phyllobates*). Science. 1980;208:1383–1385. doi:10.1126/science.6246586.

29. Márquez R, Ramírez-Castañeda V, Amézquita A. Does batrachotoxin autoresistance coevolve with toxicity in *Phyllobates* poison-dart frogs? Evolution. 2019;73(2):390–400. doi:10.1111/evo.13672.

30. Abderemane-Ali F, Rossen ND, Kobiela ME, Craig RA, Garrison CE, Chen Z, et al. Evidence that toxin resistance in poison birds and frogs is not rooted in sodium channel mutations and may rely on "toxin sponge" proteins. Journal of General Physiology. 2021;153. doi:10.1085/jgp.202112872.

31. Saporito RA, Spande TF, Garraffo MH, Donnelly MA. Arthropod alkaloids in poison frogs: A review of the 'dietary hypothesis'. Heterocycles. 2009;79(1):277. doi:10.3987/REV-08-SR(D)11.

32. Mebs D, Alvarez JV, Pogoda W, Toennes SW, Köhler G. Poor alkaloid sequestration by arrow poison frogs of the genus *Phyllobates* from Costa Rica. Toxicon. 2014;80:73–77. doi:10.1016/j.toxicon.2014.01.006.

33. Protti-Sánchez F, Quirós-Guerrero L, Vásquez V, Willink B, Pacheco M, León E, et al. Toxicity and Alkaloid Profiling of the Skin of the Golfo Dulcean Poison Frog *Phyllobates vittatus* (Dendrobatidae). Journal of Chemical Ecology. 2019;doi:10.1007/s10886-019-01116-x.

34. Márquez R. The Evolutionary, Biogeographic, and Genetic Origin of Color Pattern Diversity in *Phyllobates* Poison-Dart Frogs [PhD Thesis]. University of Chicago; 2020. Available from: https://doi.org/10.6082/uchicago.2328.

35. Jacob Machado D, Janies D, Brouwer C, Grant T. A new strategy to infer circularity applied to four new complete frog mitogenomes. Ecology and Evolution. 2018;8(8):4011–4018. doi:10.1002/ece3.3918.

36. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–2120. doi:10.1093/bioinformatics/btu170.

37. Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, Marth GT. Bamtools: A C++ API and toolkit for analyzing and managing BAM files. Bioinformatics. 2011;27(12):1691–1692. doi:10.1093/bioinformatics/btr174.

38. Wang JR, Holt J, McMillan L, Jones CD. FMLRC: Hybrid long read error correction using an FM-index. BMC Bioinformatics. 2018;19(1):1–11. doi:10.1186/s12859-018-2051-3.

39. Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. Nature Methods. 2020;17(2):155–158. doi:10.1038/s41592-019-0669-3.

40. Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. The MaSuRCA genome assembler. Bioinformatics. 2013;29(21):2669–2677. doi:10.1093/bioinformatics/btt476.

41. Zimin AV, Puiu D, Luo MC, Zhu T, Koren S, Marçais G, et al. Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. Genome Research. 2017;27(5):787–792. doi:10.1101/gr.213405.116.

42. Zhu BH, Xiao J, Xue W, Xu GC, Sun MY, Li JT. P_RNA_scaffolder: A fast and accurate genome scaffolder using paired-end RNA-sequencing reads. BMC Genomics. 2018;19(1):1–13. doi:10.1186/s12864-018-4567-3.

43. 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. doi:10.1093/bioinformatics/btz891.

44. Ghurye J, Pop M, Koren S, Bickhart D, Chin CS. Scaffolding of long read assemblies using long range contact information. BMC Genomics. 2017;18(1):1–11. doi:10.1186/s12864-017-3879-z.

45. Ghurye J, Rhie A, Walenz BP, Schmitt A, Selvaraj S, Pop M, et al. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. PLoS Computational Biology. 2019;15(8):1–19. doi:10.1371/journal.pcbi.1007273.

46. Xu GC, Xu TJ, Zhu R, Zhang Y, Li SQ, Wang HW, et al. LR-Gapcloser: A tiling path-based gap closer that uses long reads to complete genome assembly. GigaScience. 2018;8(1):1–14. doi:10.1093/gigascience/giy157.

47. Warren RL. RAILS and Cobbler: Scaffolding and automated finishing of draft genomes using long DNA sequences. The Journal of Open Source Software. 2016;1(7):116. doi:10.21105/joss.00116.

48. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 2013; p. 1303.3997.

49. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–2079. doi:10.1093/bioinformatics/btp352.

50. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics. 2018;34(13):i142–i150. doi:10.1093/bioinformatics/bty266.

51. Zhang Y, Lu HW, Ruan J. GAEP: a comprehensive genome assembly evaluating pipeline. Journal of Genetics and Genomics. 2023;50(10):747–754. doi:https://doi.org/10.1016/j.jgg.2023.05.009.

52. 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. doi:10.1093/bioinformatics/btv351.

53. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO Update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Molecular Biology and Evolution. 2021;38(10):4647–4654. doi:10.1093/molbev/msab199.

54. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021;10(2). doi:10.1093/gigascience/giab008.

55. Chen SH, Rossetto M, van der Merwe M, Lu-Irving P, Yap JYS, Sauquet H, et al. Chromosome-level de novo genome assembly of Telopea speciosissima (New South Wales waratah) using long-reads, linked-reads and Hi-C. Molecular Ecology Resources. 2022;22(5):1836–1854. doi:https://doi.org/10.1111/1755-0998.13574.

56. Venables WN, Ripley BD. Modern Applied Statistics with S. 4th ed. New York: Springer; 2002. Available from: https://www.stats.ox.ac.uk/pub/MASS4/.

57. Chu C, Nielsen R, Wu Y. REPdenovo: Inferring De Novo repeat motifs from short sequence reads. PLoS ONE. 2016;11(3):1–17. doi:10.1371/journal.pone.0150719.

58. Chu C, Pei J, Wu Y. An improved approach for reconstructing consensus repeats from short sequence reads. BMC Genomics. 2018;19(Suppl 6). doi:10.1186/s12864-018-4920-6.

59. Smit A, Hubley R. RepeatModeler Open - 1.0; 2017. Available at https://www.repeatmasker.org/RepeatModeler/.

60. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA. 2015;6(1):4–9. doi:10.1186/s13100-015-0041-9.

61. Smit A, Hubley R, Green P. RepeatMasker - 4.0; 2018. Available at https://www.repeatmasker.org/RepeatMasker/.

62. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10(1):421. doi:10.1186/1471-2105-10-421.

63. Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genomics and Bioinformatics. 2021;3(1):1–11. doi:10.1093/nargab/lqaa108.

64. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21. doi:10.1093/bioinformatics/bts635.

65. Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nature Methods. 2021;18(4):366–368. doi:10.1038/s41592-021-01101-x.

66. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30(9):1236–1240. doi:10.1093/bioinformatics/btu031.

67. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Research. 2008;36(10):3420–3435. doi:10.1093/nar/gkn176.

68. Tarvin RD, Santos JC, O'Connell LA, Zakon HH, Cannatella DC. Convergent substitutions in a sodium channel suggest multiple origins of toxin resistance in poison frogs. Molecular Biology and Evolution. 2016;33(4):1068–1081. doi:10.1093/molbev/msv350.

69. Bricelj VM, Connell L, Konoki K, MacQuarrie SP, Scheuer T, Catterall WA, et al. Sodium channel mutation leading to saxitoxin resistance in clams increases risk of PSP. Nature. 2005;434(7034):763–767. doi:10.1038/nature03415.

70. Jost MC, Hillis DM, Lu Y, Kyle JW, Fozzard HA, Zakon HH. Toxin-resistant sodium channels: Parallel adaptive evolution across a complete gene family. Molecular Biology and Evolution. 2008;25(6):1016–1024. doi:10.1093/molbev/msn025.

71. Gendreau KL, Hornsby AD, Hague MTJ, McGlothlin JW. Gene Conversion Facilitates the Adaptive Evolution of Self-Resistance in Highly Toxic Newts. Molecular Biology and Evolution. 2021;38(10):4077–4094. doi:10.1093/molbev/msab182.

72. Montana KO, Ramírez-Castañeda V, Tarvin RD. Are Pacific Chorus Frogs (Pseudacris regilla) resistant to Tetrodotoxin (TTX)? Characterizing Potential TTX Exposure and Resistance in an Ecological Associate of Pacific Newts (*Taricha*). Journal of Herpetology. 2023;57(2):220 – 228. doi:10.1670/22-002.

73. Gazave E, Lapébie P, Richards GS, Brunet F, Ereskovsky AV, Degnan BM, et al. Origin and evolution of the Notch signalling pathway: an overview from eukaryotic genomes. BMC Evolutionary Biology. 2009;9(1):249. doi:10.1186/1471-2148-9-249.

74. Marlow H, Roettinger E, Boekhout M, Martindale MQ. Functional roles of Notch signaling in the cnidarian Nematostella vectensis. Developmental Biology. 2012;362(2):295–308. doi:10.1016/j.ydbio.2011.11.012.

75. Favarolo MB, López SL. Notch signaling in the division of germ layers in bilaterian embryos. Mechanisms of Development. 2018;154:122–144. doi:https://doi.org/10.1016/j.mod.2018.06.005.

76. Layden MJ, Martindale MQ. Non-canonical Notch signaling represents an ancestral mechanism to regulate neural differentiation. EvoDevo. 2014;5(1):30. doi:10.1186/2041-9139-5-30.

77. Walton KD, Croce JC, Glenn TD, Wu SY, McClay DR. Genomics and expression profiles of the Hedgehog and Notch signaling pathways in sea urchin development. Developmental Biology. 2006;300(1):153–164. doi:https://doi.org/10.1016/j.ydbio.2006.08.064.

78. Erkenbrack EM. Notch-mediated lateral inhibition is an evolutionarily conserved mechanism patterning the ectoderm in echinoids. Development Genes and Evolution. 2018;228(1):1–11. doi:10.1007/s00427-017-0599-y.

79. Hinman VF, Burke RD. Embryonic neurogenesis in echinoderms. WIREs Developmental Biology. 2018;7(4):e316. doi:10.1002/wdev.316.

80. Lloyd-Lewis B, Mourikis P, Fre S. Notch signalling: sensor and instructor of the microenvironment to coordinate cell fate and organ morphogenesis. Current Opinion in Cell Biology. 2019;61:16–23.

81. MacGrogan D, Münch J, de la Pompa JL. Notch and interacting signalling pathways in cardiac development, disease, and regeneration. Nature Reviews Cardiology. 2018;15(11):685–704. doi:10.1038/s41569-018-0100-2.

82. Hayward P, Brennan K, Sanders P, Balayo T, DasGupta R, Perrimon N, et al. Notch modulates Wnt signalling by associating with Armadillo β-catenin and regulating its transcriptional activity. Development. 2005;132(8):1819–1830. doi:10.1242/dev.01724.

83. Zhang R, Engler A, Taylor V. Notch: an interactive player in neurogenesis and disease. Cell and Tissue Research. 2018;371(1):73–89. doi:10.1007/s00441-017-2641-9.

84. Wang J, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu S, et al. The conserved domain database in 2023. Nucleic Acids Research. 2022;51(D1):D384–D388. doi:10.1093/nar/gkac1096.

85. Slater GSC, Birney E. Automated generation of heuristics for biological sequence comparison. BMC bioinformatics. 2005;6:31. doi:10.1186/1471-2105-6-31.

86. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28(23):3150–3152. doi:10.1093/bioinformatics/bts565.

87. Liedtke HC, Gower DJ, Wilkinson M, Gomez-Mestre I. Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate. Nature Ecology and Evolution. 2018;2(11):1792–1799. doi:10.1038/s41559-018-0674-4.

88. Bredeson JV, Mudd AB, Medina-Ruiz S, Mitros T, Smith OK, Miller KE, et al. Conserved chromatin and repetitive patterns reveal slow genome evolution in frogs. Nature Communications. 2024;15(1):579. doi:10.1038/s41467-023-43012-9.

89. Márquez R, Machado J Denis, Nouri R, Gendreau L Kerry, Janies D, Saporito A Ralph, et al.. Supporting data for "A draft genome assembly for the dart-poison frog Phyllobates terribilis"; 2025. Available from: http://gigadb.org/dataset/102722.

Supplementary material for: A draft genome assembly for the dart-poison frog *Phyllobates terribilis*

Roberto Márquez[1,*,†], Denis Jacob Machado[2,3,4,*,†], Reyhaneh Nouri[2,3], Kerry L. Gendreau[1], Daniel Janies[2,3], Ralph A. Saporito[5], Marcus R. Kronforst[6], and Taran Grant[4,*]

[1]Department of Biological Sciences, Virginia Tech, Blacksburg, VA. USA
[2]Center for Computational Intelligence to Predict Health and Environmental Risks, University of North Carolina at Charlotte, Charlotte, NC. USA
[3]Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC. USA
[4]Laboratório de Anfíbios, Departamento de Zoologia, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil
[5]Department of Biology, John Carroll University, University Heights, OH. USA
[6]Department of Ecology and Evolution, University of Chicago. Chicago, IL. USA

*__Correspondence:__ rmarquezp@vt.edu, dmachado@charlotte.edu, taran.grant@ib.usp.br
†Co-first authors

1 Supplementary Figures

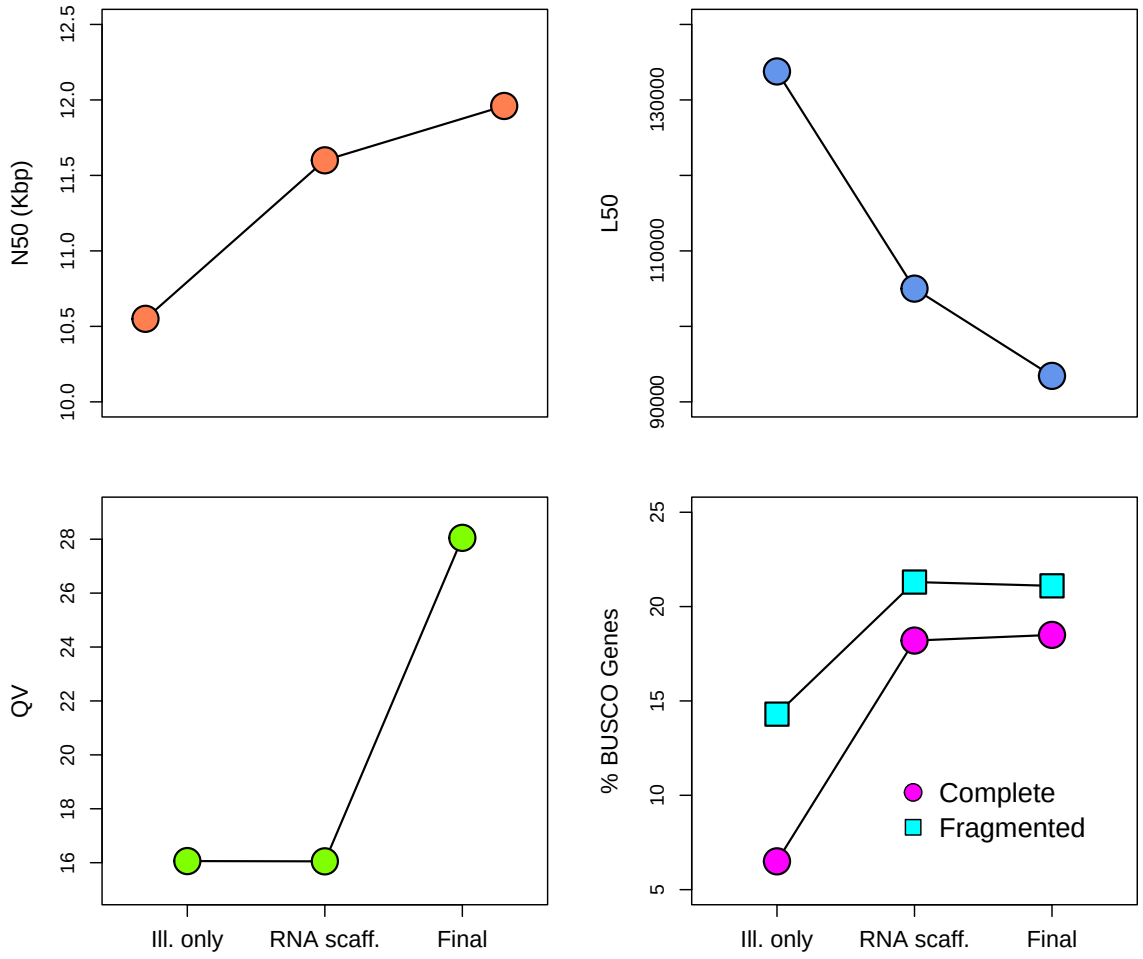

Figure S1: Contiguity and completeness statistics at three key steps of the assembly pipeline: The initial Illumina-only assembly generated with MaSuRCA, the assembly resulting from RNAseq-based scaffolding, and the final assembly incorporating PacBio and HiC scaffolding and gap filling. See Fig. 1 in the main text for further details on the assembly pipeline, and the Methods section for details on each statistic.

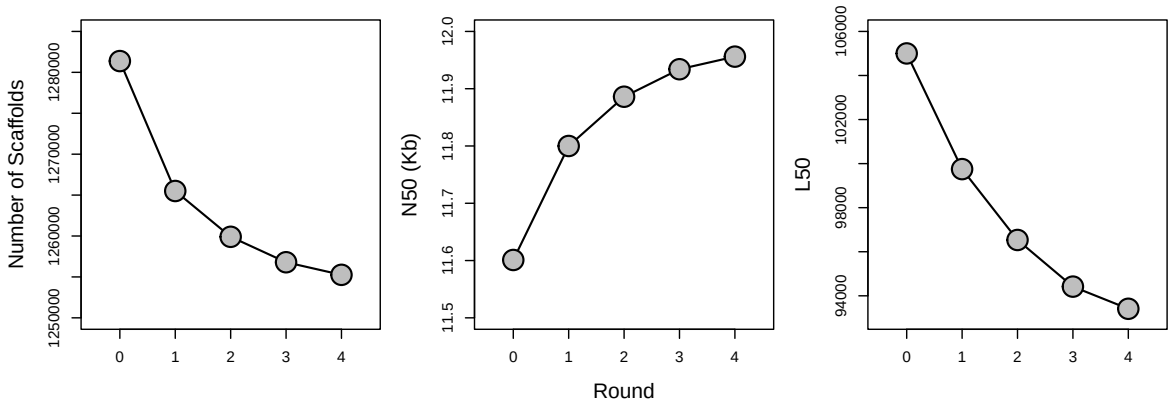

Figure S2: Contiguity statistics at each round of the iterative scaffolding and gap-filling step of assembly. Round 0 corresponds to the RANseq-scaffolded assembly. Further details on the pipeline can be found in Fig. 1 of the main text.