

Additional file 3

Supplementary results

Data availability and supplementary data files

To allow this document to be read as a standalone companion to the main manuscript, key tables and figures that also appear in the main text are included here in their original position within the supplementary results. All supplementary tables, figures, and data files referenced in this document are archived in the Zenodo data package accompanying this study (DOI: 10.5281/zenodo.15298547). Several supplementary tables are provided as tab-separated or ODT files and supplementary figures as PNG or PDF files as referenced in the following. The genome annotation was deposited alongside the scaffolded genome assembly at Zenodo folder: Annotation_scaffolded_genome_assembly. All custom scripts used to generate the results reported here are available at GitHub (<https://github.com/koppk/Grenada-Frog-144>) and additionally archived in Zenodo.

Read quality analysis

Using NanoComp 1.24.2 (per-read metric) from the NanoPack suite [1], a total of 16,448,199 reads were obtained with real-time basecalling (FAST mode), with a mean read length of 3,268 bp (median 2,553 bp) and a mean read quality of Q11.4 (median Q11.9). High-accuracy (HAC) basecalling produced 17,071,244 reads with a similar mean read length of 3,257 bp (median 2,534 bp) but a higher mean quality of Q15.6 (median Q17.1). The read length N50 was ~4.8 kb for both modes (4,832 bp for FAST; 4,842 bp for HAC), and the length distributions had comparable spread (standard deviation ~3.57 kb for each). The total number of bases sequenced was 53.75 Gb for FAST and 55.60 Gb for HAC (Table SR1).

Table SR1. Raw read summary statistics. Calculated using NanoComp 1.24.2 from the NanoPack suite [1]. Basecalling was performed with MinKNOW v24.06.16. FAST mode: real-time basecalling during sequencing run. HAC mode: high-accuracy basecalling after the sequencing run.

General summary						
Basecalling mode	Real-time (FAST)			High-accuracy (HAC)		
Files	GrenadaFrog144_ONT_ALL.fastq.gz			GrenadaFrog144_ONT_HAC_all.fastq.gz		
Mean read length	3,268.1			3,256.7		
Mean read quality (Q)	11.4			15.6		
Median read length	2,553.0			2,534.0		
Median read quality (Q)	11.9			17.1		
Number of reads	16,448,199			17,071,244		
Read length N50	4,832			4,842		
STDEV read length	3,574.1			3,563.0		
Total bases	53,754,912,003			55,595,176,615		
Reads above quality cutoffs						
Quality scores (Q)	Number of reads	Percentage (%)	Number of bases (Mb)	Number of reads	Percentage (%)	Number of bases (Mb)
>Q10	13,838,221	84.1	46,739.2	16,675,929	97.7	54,703.3
>Q15	271,316	1.6	841.6	12,417,354	72.7	44,616.3
>Q20	5,583	0.0	22.1	2,074,827	12.2	9,190.8
>Q25	89	0.0	1.2	31,012	0.2	104.0
>Q30	10	0.0	0.0	1,265	0.0	0.3
Top 5 highest mean basecall quality scores						
Rank	Mean basecall quality score (Q)	Read length (bp)		Mean basecall quality score (Q)	Read length (bp)	
1	33.4	1,490		39.7	416	
2	33.4	873		39.0	115	
3	32.4	1,401		38.5	123	
4	31.7	766		38.2	155	
5	31.4	3,569		38.2	368	
Top 5 longest reads						

Rank	Read length (bp)	Mean basecall quality score	Read length (bp)	Mean basecall quality score
1	1,546,395	12.1	673,750	15.2
2	1,329,233	19.4	550,250	9.7
3	1,302,933	8.4	549,004	13.4
4	919,751	8.5	541,607	9.6
5	912,718	10.6	528,017	10.4

Calculated using NanoComp, quality score distributions indicated that 84.1% of FAST reads had Q>10 (13.8 million reads) compared to 97.7% of HAC reads (16.7 million). At the Q>15 threshold, 1.6% of FAST reads (0.27 million) remained versus 72.7% of HAC reads (12.4 million). Virtually no FAST reads surpassed Q>20 (0.03%, 5,583 reads), whereas 12.2% of HAC reads (2,074,827 reads) had Q>20. Small fractions of HAC reads even exceeded Q>25 (0.2%) and Q>30 (0.007%), while the FAST mode yielded only tens of reads at those high quality levels (Table SR1).

The highest individual read quality scores assessed using NanoComp were Q39.7 for HAC (read length 416 bp) and Q33.4 for FAST (1,490 bp). Correspondingly, the top five reads by quality in HAC had mean Q scores ranging from 39.7 to 38.2, with read lengths of 115-416 bp, whereas the top five FAST reads had Q scores from 33.4 to 31.4 and lengths of 766-3,569 bp. In terms of read length, the longest single read was 1,546,395 bp from the FAST run (quality Q12.1) compared to 673,750 bp from the HAC run (Q15.2). The five longest reads in the FAST dataset ranged from 912,718 bp to 1,546,395 bp (per-read Q scores 8.4-19.4), whereas in the HAC dataset they ranged from 528,017 bp to 673,750 bp (Q 9.6-15.2) (Table SR1).

In parallel, base-level quality metrics were computed using SeqKit v2.10.0 [2] for the combined HAC dataset. Based on this analysis, 83.1% of sequenced bases had Phred quality scores $\geq$ Q20 and 68.7% $\geq$ Q30 (Table SM1 in Additional file 2). In contrast to the NanoComp metrics reported above, which are based on per-read mean quality, the SeqKit values reflect the proportion of individual bases exceeding the respective thresholds, as SeqKit performs base-wise rather than read-wise quality assessment. The two quality assessments are complementary rather than directly comparable. The per-read arithmetic mean of Phred quality scores, averaged across all HAC reads, was Q32.45. The read-length-weighted mean, computed by summing all per-base Phred scores across all reads and dividing by total bases, was

Q33.10. One sequencing run (ERR14936785) showed a very high read-length N50 (155 kb) and a very small read count (1,946 reads).

Genome size and complexity estimation

Reference-free analysis of the HAC read dataset using a k-mer profile computed by Jellyfish v2.3.1 [3] followed by GenomeScope 2.0 [4] estimated the haploid genome size at 1.62 Gb (Table SR2; Figure SR1). Dividing the total number of high-accuracy basecalled bases (~55.6 Gb) by this estimate yielded an expected sequencing depth of approximately 34.3× ("Definition of coverage metrics": (i) expected depth). Heterozygosity was estimated at 0.62-0.68%, repeat content at ~30.3% (~490 Mb), and read error rate at ~0.84%. The GenomeScope 2.0 model fit ranged from 74.1% to 99.5%. The k-mer frequency distribution showed a distinct heterozygous peak at ~12× coverage and a homozygous peak at ~22× (Figure SR1, panel A). Read-based G+C content was 43.23%, computed from the total number of guanine and cytosine bases (24.04 Gb) relative to all sequenced bases (55.60 Gb). This read-derived estimate served as a reference-free baseline for comparison with the assembly-derived G+C values reported in the genome assembly section (Table SR3).

Table SR2. Genome size and complexity estimates for *Pristimantis euphronides*. K-mer frequency profiling (k = 21) was performed with Jellyfish v2.3.1 [3] on high-accuracy (HAC) basecalled Nanopore reads, followed by GenomeScope 2.0 [4]. A diploid model (ploidy = 2) was fitted.

Property	Min	Max
Homozygous (aa)	99.3166%	99.3801%
Heterozygous (ab)	0.619858%	0.683373%
Genome Haploid Length	1,617,259,408 bp	1,621,724,807 bp
Genome Repeat Length	490,426,395 bp	491,780,506 bp

Genome Unique Length	1,126,833,013 bp	1,129,944,300 bp
Model Fit	74.063%	99.5442%
Read Error Rate	0.838112%	0.838112%

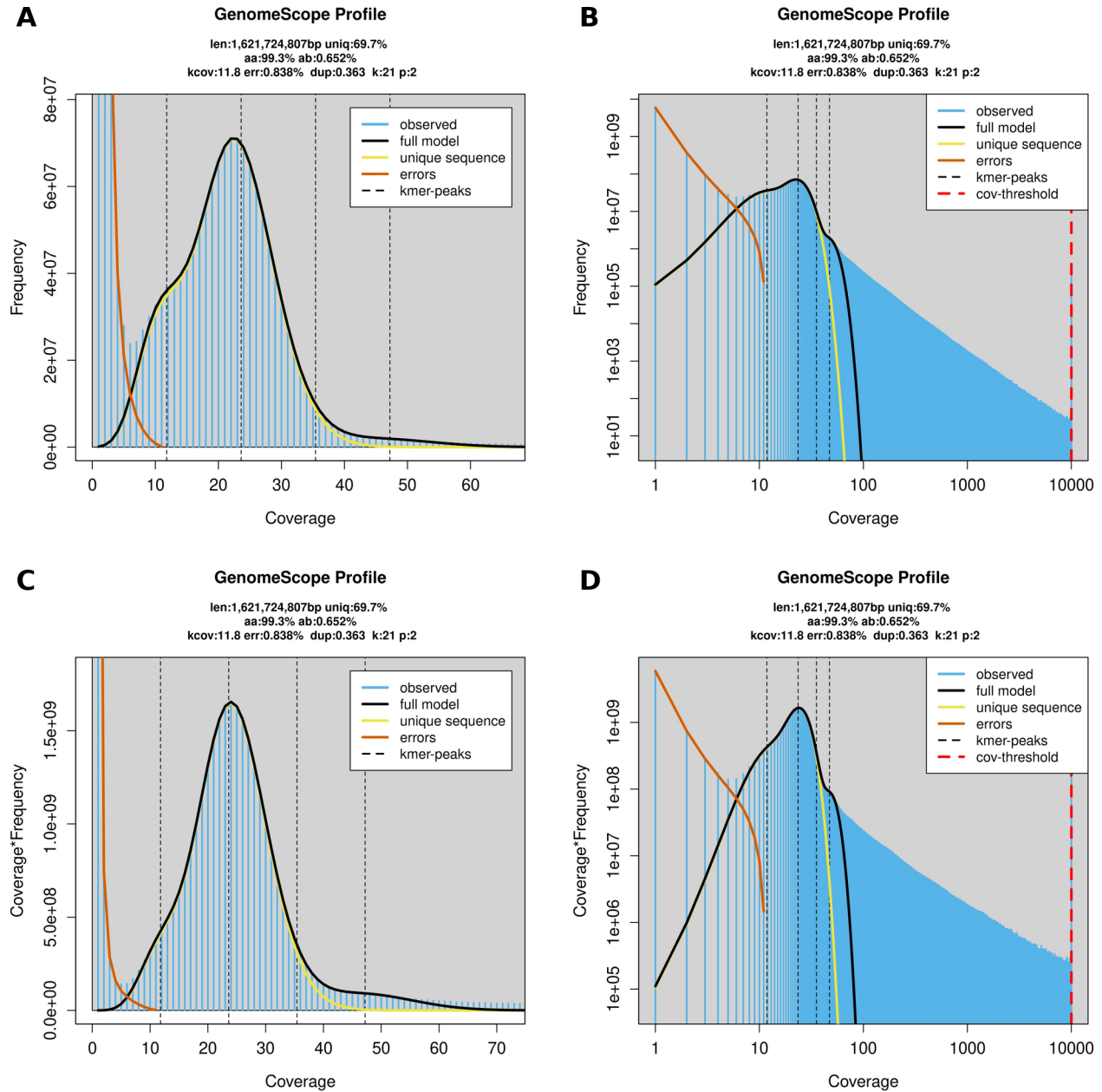


Figure SR1. Genome size and complexity estimates for *Pristimantis euphronides*. K-mer frequency profiling ($k = 21$) was performed with Jellyfish v2.3.1 [3] on high-accuracy (HAC) basecalled Oxford Nanopore reads, followed by GenomeScope 2.0 [4]. A diploid model (ploidy = 2) was fitted. (A) Linear scale. (B) Log scale. (C) Transformed linear scale (Coverage $\times$ Frequency). (D) Transformed log scale. Blue bars: observed k-mer frequencies. Black line: fitted diploid (2p) model. Yellow line: unique

sequence component. Orange line: sequencing error component. Dashed black lines: predicted k-mer coverage peaks (heterozygous at $\sim 12\times$, homozygous at $\sim 22\times$). Red dashed line: coverage threshold.

Taxonomic classification of reads

Kraken2 v2.1.2 [5] classified 823,241 of 16,448,199 FAST basecalled reads (5.01%) at the applied confidence threshold of 0.2. The remaining 15,624,958 reads (94.99%) were unclassified. Among the classified reads, 91.3% were assigned to the superfamily Hyloidea, totaling 751,664 reads (4.57% of all reads). Most of these were assigned at higher taxonomic ranks only. At the species level, 43,185 reads (0.26%) matched *Eleutherodactylus coqui* (Eleutherodactylidae), the most frequent hit. The second most represented species was *Ranitomeya imitator* (Dendrobatidae; 1,093 reads, 0.007%), followed by sparse hits across other Hyloidea taxa.

Non-eukaryotic taxa accounted for 0.02% of all reads. Bacterial sequences totaled 302 reads (0.0018%), primarily *Cutibacterium* (146 reads), *Micrococcus*, *Staphylococcus*, and *Streptococcus*. Fungal reads comprised 1,925 reads (0.0117%), mostly assigned to Dikarya, with minor species-level hits including *Fusarium musae*, *Drechmeria coniospora*, and *Malassezia restricta*. Protozoan reads totaled 804 (0.0049%), of which 795 were assigned to *Besnoitia besnoiti* and five to *Plasmodium* spp. Viral sequences comprised two reads, both assigned to the insect virus family Ascoviridae. No reads were classified as archaeal, plasmid, or UniVec contaminants.

Genome assembly statistics

Genome assembly statistics are summarized in Table SR3. The Flye v2.9.5 [6] *de novo* assembly consisted of 34,106 contigs with a sum length of 1.748 Gb before and 1.747 Gb after polishing using Medaka v2.0.1 [7] and an N50 of ~ 302 kb. The DNA composition of the Workflow 1-assembled contigs was 43.01% G + C.

Two haplotype-resolved assemblies were also produced to explore haplotypic diversity. However, only the primary collapsed (pseudo-haploid) genome assembly was selected for all downstream analyses and phylogenetic reconstructions. Methods and results for the haplotype-resolved assemblies are provided in Additional file 4 and Additional file 5, respectively.

The modified CSA2.6 [8] pipeline (Workflow 2) used Wtdbg2 v2.2 [9] for *de novo* assembly, resulting in 13,729 contigs and a contig N50 of ~333 kb. The total length of the contigs was approximately 1.6 Gb, with a G+C content of 42.53%.

Table SR3. Genome assembly and scaffolding statistics for *Pristimantis euphronides* across two independent workflows. Workflow 1: *de novo* assembly with Flye v2.9.5 [6], polished with Medaka v2.0.1 [7], scaffolded with RagTag v2.1.0 [10] against *Eleutherodactylus coqui* haplotype 1 (GCF_035609145.1). Workflow 2: *de novo* assembly with Wtdbg2 v2.2 [9], scaffolded with Ragout v2.3 [11], [12] within the modified CSA v2.6 [8] pipeline against the same reference. Estimated genome size from GenomeScope 2.0 [4]. bp: base pairs, Gb: gigabases. (Also presented, combined with Table SR4, as Table 1 in the main manuscript.)

Attribute	Workflow 1 (Flye/Medaka/RagTag)	Workflow 2 (modified CSA2.6)
Estimated genome size (Gb)	1.62	
After assembly (pre-scaffolding)	Flye and Medaka	WTDBG2
Number of contigs	34,106	13,729
Sum length of contigs (bp)	1,748,513,520 (Flye), 1,747,656,068 (Medaka-polished)	1,600,024,302
Contig N50 (bp)	302,327 (Flye), 302,355 (Medaka-polished)	333,065

Maximum contig length (bp)	3,444,756 (Flye), 3,443,368 (Medaka-polished)	2,647,371
Contig DNA G+C (%)	43.01	42.53
After scaffolding	RagTag	Ragout
Number of scaffolds and contigs	25,310	7,804
Sum length of scaffolds and contigs (bp)	1,748,533,034	1,601,209,302
Scaffold and contig N50 (bp)	117,773,236	115,239,735
Scaffold and contig L50 (N50_num)	6	6
Maximum scaffold and contig length (bp)	242,675,580	229,436,012
Number of gap sequences inserted during scaffolding	8,784 (of 100 bp length each)	5,925 (of 200 bp length each)
Sum length of gaps (bp)	878,400	1,185,000
Scaffold DNA G+C (%)	42.99	42.50

Assembly quality assessment

Inspector v1.2 [13] evaluated 25,680 contigs (≥ 3 kb; 1,734 Mb) and reported a consensus quality value (QV) of 36.2 for the Medaka-polished Flye assembly (Workflow 1). Small-scale errors totaled 293,398

(169.2/Mb): 149,478 base substitutions (50.9%), 73,703 small expansions (25.1%), and 70,217 small collapses (23.9%). Of 704 structural discordances, 653 (92.8%) were classified as haplotype switches, with 51 remaining (43 expansions, 6 collapses, 2 inversions) across 1,468 Mb. The read mapping rate was 98.2% at a mean depth of 32.0 $\times$, with a split-read rate of 12.4%. Merqury v1.3 [14] estimated QV 37.0 with k-mer completeness of 88.9%, identifying a haploid k-mer peak at 12 $\times$ and a diploid peak at 22 $\times$.

Of 17.1 million reads (55.6 Gb), 98.1% mapped to the pre-scaffolding assembly at a genome-wide bp-weighted mean coverage of 31.68 $\times$ and a mean read-to-assembly alignment identity of 95.6%. Of all contigs, 80.3% had a coverage coefficient of variation (CV) < 0.5 (assembly-wide mean CV 0.37) and 88.8% had a mean read base Phred quality $\geq$ 30 (bp-weighted mean Q 32.6). Genome-wide heterozygosity was 6.66 het/kb (het/total ratio 0.97).

Per-contig mapping quality was assessed across 34,040 contigs with aligned reads. The assembly-wide mean mapping quality (MAPQ) was 25.98. Of these, 4,934 contigs (14.5%) had a mean MAPQ $\geq$ 50, 13,853 (40.7%) had mean MAPQ 20-50, and 15,253 (44.8%) had mean MAPQ < 20. On average, 42.7% of alignments per contig had MAPQ = 0 and 35.8% had MAPQ $\geq$ 60.

Per-contig read base quality, computed from the Phred scores retained in the BAM, yielded a mean per-read quality of Q32.14 across 33,990 contigs (bp-weighted mean Q32.60). The majority of contigs (30,172; 88.8%) was covered by reads with mean base quality $\geq$ Q30, 3,711 (10.9%) contigs by reads with mean base quality Q20-30, and 107 (0.3%) contigs by reads with mean base quality below Q20.

Variant calling with BCFtools v1.19 [15] (mpileup -q 10 -Q 20 -d 200; call -mv) identified 12,020,184 variants, of which 11,634,865 (96.8%) were heterozygous and 385,319 (3.2%) homozygous-alternative. Genome-wide variant density was 6.88 per kb (6.66 het/kb). At the per-contig level, 2,824 contigs (8.3%) carried zero variants, 13,531 (39.7%) had < 5 variants/kb, 12,574 (36.9%) had 5-15 variants/kb, and 5,177 (15.2%) exceeded 15 variants/kb.

The mean per-contig soft-clip fraction was 0.2063, with 24,176 contigs (71.0%) in the 0.1-0.3 range, 3,896 (11.4%) below 0.1, and 5,968 (17.5%) above 0.3. The mean supplementary alignment rate was 0.1816. Cross-metric quality flagging identified 5,477 contigs (16.1%) meeting two or more of the following criteria: CV > 1, mean MAPQ < 10, soft-clip fraction > 0.3, or variant density > 20 per kb.

Gene-space completeness of both pre-scaffolding assemblies was assessed using the HANNO v0.4 pipeline [16] with identical evidence inputs. Benchmarking Universal Single-Copy Orthologs (BUSCO) v3.0.2 [17] evaluation against the tetrapoda_odb10 database (n = 5,310 orthologs) [18] on the BESTMODELS gene sets identified 4,718 complete orthologs (C: 88.9%) for Workflow 1 (Flye/Medaka), including 4,597 single-copy (S: 86.6%) and 121 duplicated (D: 2.3%), with 245 fragmented (F: 4.6%) and 347 missing (M: 6.5%) (Table SR4). Workflow 2 (Wtdbg2 v2.2) achieved 86.2% completeness (C), with 4,470 single-copy (S: 84.2%) and 109 duplicated (D: 2.1%), 294 fragmented (F: 5.5%), and 437 missing (M: 8.2%) (Table SR4).

Table SR4. Gene-space completeness across assembly and scaffolding workflows for *Pristimantis euphronides*. Benchmarking Universal Single-Copy Orthologs (BUSCO) [17] completeness was assessed using the HANNO v0.4 pipeline [16] against the tetrapoda_odb10 database (n = 5,310 orthologs) [18] with combined protein and mRNA evidence from seven Hyloidea RefSeq genome annotations: *Bufo gargarizans* (GCF_014858855.1), *Engystomops pustulosus* (GCF_040894005.1), *Bufo bufo* (GCF_905171765.1), *Ranitomeya imitator* (GCF_032444005.1), *Eleutherodactylus coqui* (GCF_035609145.1), *Dendropsophus ebraccatus* (GCF_027789765.1), and *Hyla sarda* (GCF_029499605.1). C: complete, S: single-copy, D: duplicated, F: fragmented, M: missing. All values in percent of the 5,310 tetrapoda_odb10 orthologs. (Also presented, combined with Table SR3, as Table 1 in the main manuscript.)

Assembly	Assembler	Scaffolder	C (%)	S (%)	D (%)	F (%)	M (%)
Workflow 1 - pre-scaffolding	Flye v2.9.5 + Medaka v2.0.1	-	88.85	86.57	2.28	4.61	6.53
Workflow 2 - pre-scaffolding	Wtdbg2 v2.2	-	86.23	84.18	2.05	5.54	8.23
Workflow 1 - scaffolded	Flye v2.9.5 + Medaka v2.0.1	RagTag v2.1.0	91.41	89.36	2.05	3.07	5.52
Workflow 2 - scaffolded	Wtdbg2 v2.2	Ragout v2.3	88.40	86.33	2.07	4.29	7.31

Whole-genome alignment using D-Genies v1.5.0 [19] of the Workflow 2 pre-scaffolding assembly (query) against the Workflow 1 pre-scaffolding assembly (target) showed genome-wide collinearity along the main diagonal (Figure SR2). Of 13,729 Workflow 2 contigs, 13,689 (99.7%) aligned to Workflow 1, representing 1,599.8 of 1,600.0 Mb (>99.9% of Workflow 2 sequence). In the target assembly, 18,238 of 34,106 contigs (1,649.3 Mb; 94.4%) received alignments; the 15,868 unmatched contigs totaled 98.3 Mb (mean 6.2 kb). When the alignment was repeated with query and target reversed, the number of unmatched Workflow 1 contigs decreased from 15,868 to 4,661.

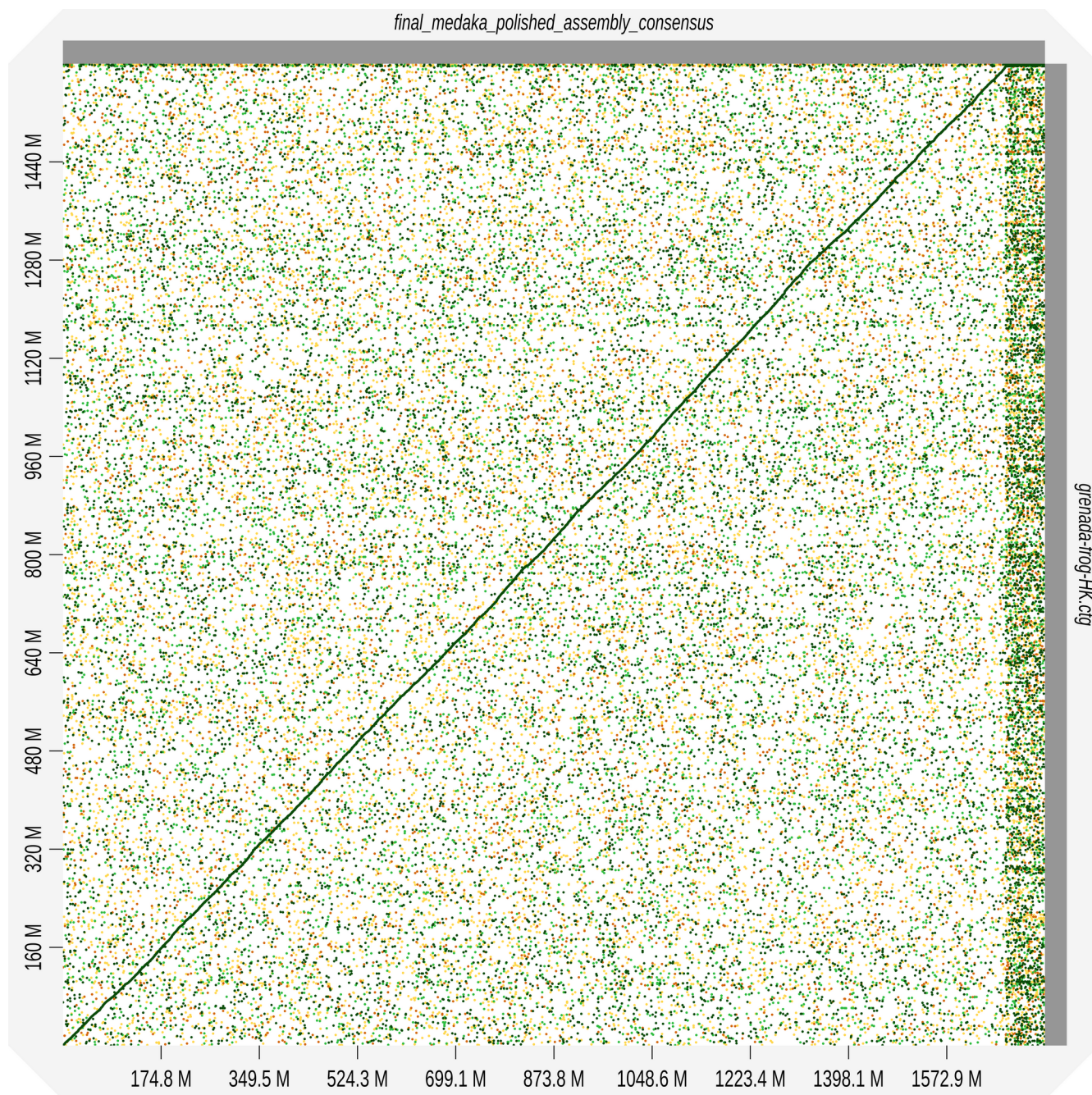


Figure SR2. Pre-scaffolding inter-workflow assembly concordance. D-Genies v1.5.0 [19] dotplot of the Workflow 2 pre-scaffolding assembly (Wtdbg2 v2.2 [9]; query, vertical axis) against the Workflow 1 pre-scaffolding assembly (Flye v2.9.5 [6]/Medaka v2.0.1 [7]; target, horizontal axis). Color reflects nucleotide identity in percentage (%) using the default D-Genies scale: white (0%), yellow (1-24%), orange (25-49%), light green (50-74%), and dark green (75-100%).

Reference genome selection via k-mer-based distance estimation

Pairwise genomic distances computed with Mash v2.3 [20] are visualized in the heatmap shown in Figure SR3 and detailed in Additional file 6 (Table). Among the 21 reference assemblies analyzed, *E. coqui* haplotype 2 (GenBank ID: GCA_035609135.1) exhibited the lowest sequence divergence to the Flye-assembled, Medaka-polished *Pristimantis euphronides* genome, with a value of 0.151946. This value represents an estimate of the Jaccard distance between the k-mer sets of two sequences, which approximates the fraction of sequence dissimilarity between them. While the NCBI RefSeq assembly, *E. coqui* haplotype 1 (RefSeq ID: GCF_035609145.1), showed a slightly higher distance (0.156619), it was chosen as the reference for RagTag scaffolding of the primary (collapsed) assembly because, as a RefSeq genome, it is curated and includes complete functional annotation for downstream analyses. In contrast, the more similar *E. coqui* haplotype 2 genome is neither part of RefSeq nor annotated. Among all reference assemblies, that of *Bufo viridis* (GenBank ID: GCA_037900795.1; Bufonidae) showed the highest sequence divergence (0.210897) to the new primary *P. euphronides* genome.

Because the haplotype-resolved assemblies were generated to facilitate direct structural and sequence-level comparison between the two *P. euphronides* haplotypes, scaffolding against the less divergent haplotype 2 reference was selected to minimize reference-induced differences. Results and methods for the two haplotype-resolved assemblies are provided in Additional file 5 and Additional file 4, respectively.

A broader examination of pairwise distances between assemblies of other Hyloidea species from different families revealed lower values than those observed for *P. euphronides*. The smallest such distance was 0.133864, found between *Anomaloglossus baeobatrachus* haplotype 1 (GCA_048569485.1; Aromobatidae) and *R. imitator* (GCF_032444005.1; Dendrobatidae), followed by 0.136948 between *A. baeobatrachus* haplotype 2 (GCA_048569465.1) and *R. imitator*. All other inter-family comparisons yielded Mash distances above 0.156619, ranging from 0.161822 between the two Leptodactylidae species *Engystomops pustulosus* (maternal haplotype; GCF_040894005.1) and *Leptodactylus fallax* (GCA_947044405.1) to 0.243761 between *A. baeobatrachus* haplotype 1 (GCA_048569485.1; Aromobatidae) and *Osteocephalus oophagus* (GCA_046270745.1; Hylidae).

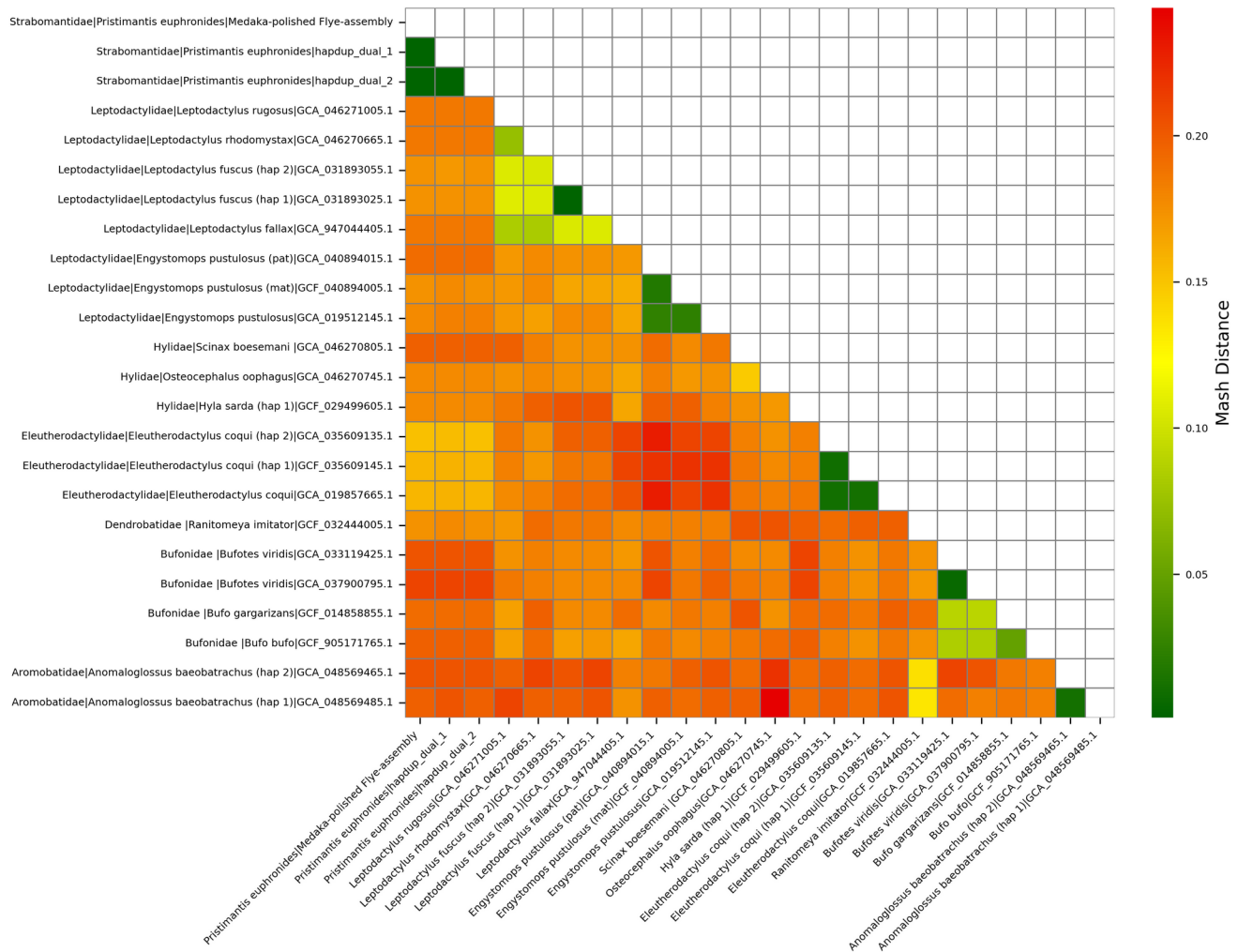


Figure SR3. Pairwise genomic distances between *Pristimantis euphronides* and 21 Hyloidea reference genome assemblies. Distances were computed with Mash v2.3 [20] using $k = 21$ and sketch size = 1,000. The lower triangle heatmap shows distances between three *P. euphronides* assemblies (primary Medaka-polished Flye assembly and two haplotype-resolved assemblies) and 21 publicly available Hyloidea genome assemblies. color scale: dark green (0, identical) to dark red (0.24, most divergent).

Reference genome selection via phylogenetic proximity

Of 46 anuran species with chromosome-level genome assemblies (24 families), 41 were present in the Portik et al. (2023) [21] phylogeny. Five species were absent: *Aquarana catesbeiana*, *Lithobates pipiens*, *Lithobates sylvaticus* (Ranidae), *Mixophyes fleayi* (Myobatrachidae), and *Ptychadena robeensis* (Ptychadenidae).

Among the 41 evaluated species, *E. coqui* exhibited the lowest divergence time to *P. euphronides*, with a time to most recent common ancestor (tMRCA) of 51.85 million years (Ma) and a topological distance of 28 edges (Table SR5). The next-closest species (*R. imitator*, *A. baeobatrachus*, *E. pustulosus*, and several Bufonidae species) were all approximately 10 Ma more distant, clustering at tMRCA values of 62.02 Ma (Table SR5).

The putative sister species *Pristimantis shrevei* (Saint Vincent and the Grenadines) had a tMRCA of 3.15 Ma and a topological distance of 2 edges but lacked a genome assembly. In total, 571 species in the phylogeny were closer to *P. euphronides* than *E. coqui*, none with a chromosome-level genome assembly, and 318 of these were congeneric *Pristimantis* species. Compared to *P. shrevei*, the closest species with an available chromosome-level reference genome (*E. coqui*) was found to be 16.5-fold ($51.85 \text{ Ma} / 3.15 \text{ Ma} = 16.5$) more distant from *P. euphronides*.

Table SR5. Ten closest chromosome-level anuran genomes ranked by phylogenetic proximity to *Pristimantis euphronides*. tMRCA: time to most recent common ancestor in million years (Ma). Topo: topological distance in number of edges. Both metrics derived from the Portik et al., 2023 [21] time-calibrated phylogeny. (Also presented as Table 2 in the main manuscript.)

Species	Family	tMRCA (Ma)	Topo (edges)
<i>Eleutherodactylus coqui</i>	Eleutherodactylidae	51.85	28
<i>Ranitomeya imitator</i>	Dendrobatidae	62.02	29
<i>Anomaloglossus baeobatrachus</i>	Aromobatidae	62.02	28
<i>Engystomops pustulosus</i>	Leptodactylidae	62.02	28
<i>Leptodactylus fuscus</i>	Leptodactylidae	62.02	30
<i>Bufo bufo</i>	Bufonidae	62.02	39
<i>Bufo gargarizans</i>	Bufonidae	62.02	40
<i>Bufotes viridis</i>	Bufonidae	62.02	40
<i>Ranitomeya variabilis</i>	Dendrobatidae	62.02	30

<i>Dendropsophus ebraccatus</i>	Hylidae	62.54	36
---------------------------------	---------	-------	----

Reference-guided scaffolding of genome assemblies

During Workflow 1 the Medaka-polished Flye-assembly was scaffolded using RagTag v2.1.0 [10] with the soft-masked *E. coqui* haplotype 1 genome (NCBI RefSeq ID: GCF_035609145.1) as reference. Dividing the total number of high-accuracy basecalled bases (~55.60 Gb) by the total length of the RagTag-scaffolded assembly (~1.75 Gb), including gap sequences, determined a genome coverage of 31.8× (Additional file 2: "Definition of coverage metrics": (ii) *bases per assembly length*, Table SM2).

Assembly statistics were collected before and after scaffolding to evaluate structural changes (Table SR3). RagTag reduced the number of sequences from 34,106 to 25,310. Total assembly length increased from ~1.747 Gb to ~1.749 Gb due to the insertion of gap sequences. The N50 increased from ~302 kb to ~118 Mb. The maximum scaffold length achieved was ~243 Mb. Of the 34,094 input contigs after the <200 bp filter, 8,815 (25.9%) were placed, accounting for 1,476,572,988 bp (84.5%) of the total scaffolded assembly length without gaps (1,747,654,634 bp); 8,785 of these were assigned to scaffolds 1-13 (1,474,662,941 bp; 84.4%) and 30 to scaffolds 14-31 (1,910,047 bp; 0.1%). The remaining 25,279 contigs (271,081,646 bp; 15.5%) were unplaced. RagTag inserted 8,784 gaps of default fixed length (100 bp). Gap sequences accounted for 0.05% of the total scaffolded genome length. The G+C content remained at ~43%.

The mean contig length within scaffolds 1-13 was 167,861 bp (median: 76,860 bp), while contigs in scaffolds 14-31 had a mean length of 63,668 bp (median: 46,287 bp) (Table SR6). The longest contig in scaffolds 1-13 was ~3.44 Mb compared to a maximum of ~351 kb in scaffolds 14-31. A comparative log-scale boxplot of these distributions is shown in Figure SR4.

Table SR6. Contig length statistics of the scaffolded *Pristimantis euphronides* genome assembly.

Contigs were classified by placement into the 13 longest RagTag v2.1.0 [10] scaffolds versus remaining scaffolds. Statistics were computed from the AGP placement file.

Metric (bp)	Contigs in scaffolds 1-13 (n = 8,785)	Contigs in scaffolds 14-31 (n = 30)
-------------	---------------------------------------	-------------------------------------

Mean length	167,861	63,668
Median (Q2)	76,860	46,287
Max length	3,443,368	350,850
25th percentile	27,422	22,186
75th percentile	196,379	80,410

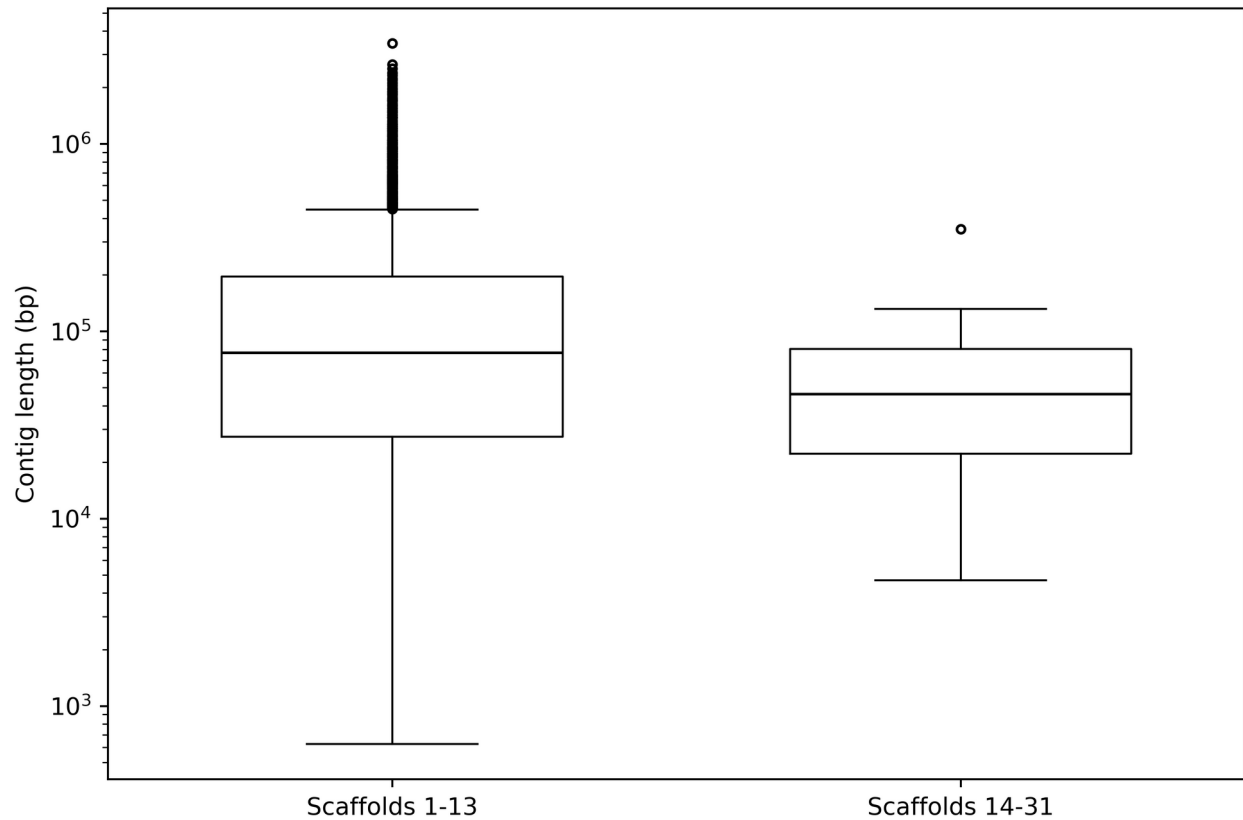


Figure SR4. Contig length distributions in the scaffolded *Pristimantis euphronides* genome assembly (Workflow 1). Comparison of contig lengths between contigs placed into the 13 longest scaffolds (1-13) and all remaining scaffolds (14-31). Boxplot on a logarithmic y-axis showing medians, interquartile ranges, whiskers, and outliers.

Workflow 2, employed as an alternative methodology to Workflow 1, utilized Ragout v2.3 [11], [12] for scaffolding the Wtdbg2 assembly of *P. euphronides* against the *E. coqui* haplotype 1 reference genome (NCBI RefSeq ID: GCF_035609145.1). This approach was implemented within a modified

CSA2.6 pipeline. The resulting genome comprised 7,804 scaffolds, with an N50 of ~115 Mb and a total length of ~1.6 Gb. The maximum scaffold length achieved was ~229 Mb. The GC content of these scaffolds was 42.50%.

Read mapping and genome coverage analysis

Minimap2 v2.28 [22] mapping followed by mosdepth v0.3.11 [23] analysis revealed that the HAC basecalled Nanopore reads mapped to the primary scaffolded *P. euphronides* genome assembly at a genome-wide mean depth of 30.9× (Additional file 2: "Definition of coverage metrics": (iii) mapped mean depth; 190,258 windows across 1.749 Gb). Across the full assembly, 81.4% of 10 kb windows were covered at $\geq 20\times$ and 5.9% fell below 10× coverage.

When restricted to the 13 longest scaffolds (147,560 windows, 1.476 Gb), 92.2% of 10 kb windows were covered at $\geq 20\times$ and 0.75% fell below 10×, with a size-weighted mean coverage of 29.3× (Figure SR5). Per-scaffold means ranged from 26.74× for scaffold_8 to 31.26× for scaffold_3. Read depth profiles across the 13 longest scaffolds are shown in Figure SR5 and individually in Additional file 7 (Figure panels A to M). Coverage was smoothed over short (2×10 kb) and long (1000×10 kb) window spans. Across most of these scaffolds, coverage was consistent with no complete coverage dropouts and only minor local variation.

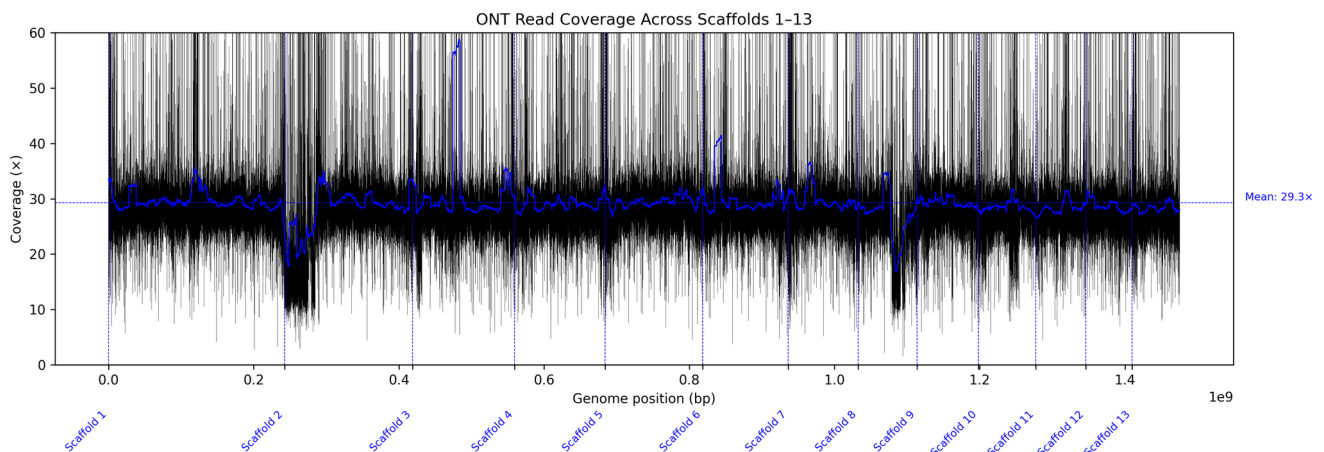


Figure SR5. Smoothed Nanopore read coverage across the 13 longest scaffolds. Blue dashed vertical lines: start position of individual scaffolds 1-13. Black line: read coverage variation over 2 x 10 kb windows. Blue bold line: long-range (10 Mb windows) smoothed trend. Blue dashed horizontal line: mean coverage (29.3×) across all 13 longest scaffolds. X-axis: genome position in gigabases (Gb, 1e9) scale. Y-axis: coverage as defined in Additional file 2: "Definition of coverage metrics": (iii) mapped mean depth. Read depth computed with mosdepth v0.3.11 [23] from Minimap2 [22] alignments. (Also presented as Figure 2 in the main manuscript.)

However, the following anomalies were detected from the smoothed 10 Mb coverage trend and confirmed against scaffold-specific thresholds. A total of two low-coverage regions were identified in scaffold_2, spanning 6.08 Mb from the beginning of the scaffold (17.2×) and 13.58 Mb from position 16.08-29.66 Mb (20.9×). A third dip of 410 kb was observed between 32.66-33.07 Mb with a mean smoothed depth of 23.1×. Scaffold_8 displayed a sustained low-coverage region of 13.03 Mb from 47.80-60.83 Mb (19.2×). These regions fell below the scaffold-specific smoothed coverage threshold and exceeded the minimum length criterion of 300 kb.

Multiple regions of elevated coverage were also detected. Scaffold_3 contained a 10.01 Mb high-coverage region between 55.16-65.17 Mb (57.2×), and scaffold_6 included a 10.00 Mb region from 15.95-25.95 Mb (40.4×). In scaffold_2, five discrete high-coverage regions were observed ranging from 1.38 Mb to 5.03 Mb in length, located between positions 46.85 Mb and 176.32 Mb, with mean smoothed coverage between 33.5× and 38.8×. Additional high-coverage regions were identified in scaffold_1 (790 kb and 3.77 Mb at the end of the scaffold), scaffold_7 (5.89 Mb, 36.3×), and scaffold_8 (9.99 Mb, 34.1×). These regions exceeded the respective scaffold-wide smoothed coverage means by more than five coverage units (×) and were sustained over lengths of at least 300 kb.

Evaluation of reference-guided scaffolding

RagTag placed 8,785 of 34,106 *P. euphronides* primary assembly contigs (25.8%) into the 13 largest scaffolds (scaffolds 1-13). Scaffold numbers follow *E. coqui* chromosome order (by *E. coqui*

chromosome size), therefore, scaffolds 4/5, 8/9, and 12/13 are pairwise reversed in *P. euphronides* size order. The remaining 25,321 contigs were assigned to smaller scaffolds (scaffolds 14-31) or remained unplaced.

Alignment of the pre-scaffolding *P. euphronides* primary assembly against the *E. coqui* reference genome using D-Genies produced 306,528 alignments, with 15,139 contigs (44.4%) lacking alignment to any *E. coqui* sequence and 823 *E. coqui* scaffolds receiving no aligned query segments. The whole-genome dotplot showed collinear clustering of contigs along all 13 *E. coqui* chromosomes (Figure SR6). The corresponding post-scaffolding dotplot displayed a continuous diagonal of alignments from scaffold_1 through scaffold_13, with most alignments on the forward strand and minimal off-diagonal signal (Figure SR7; Additional file 8: Figure plots A to M).

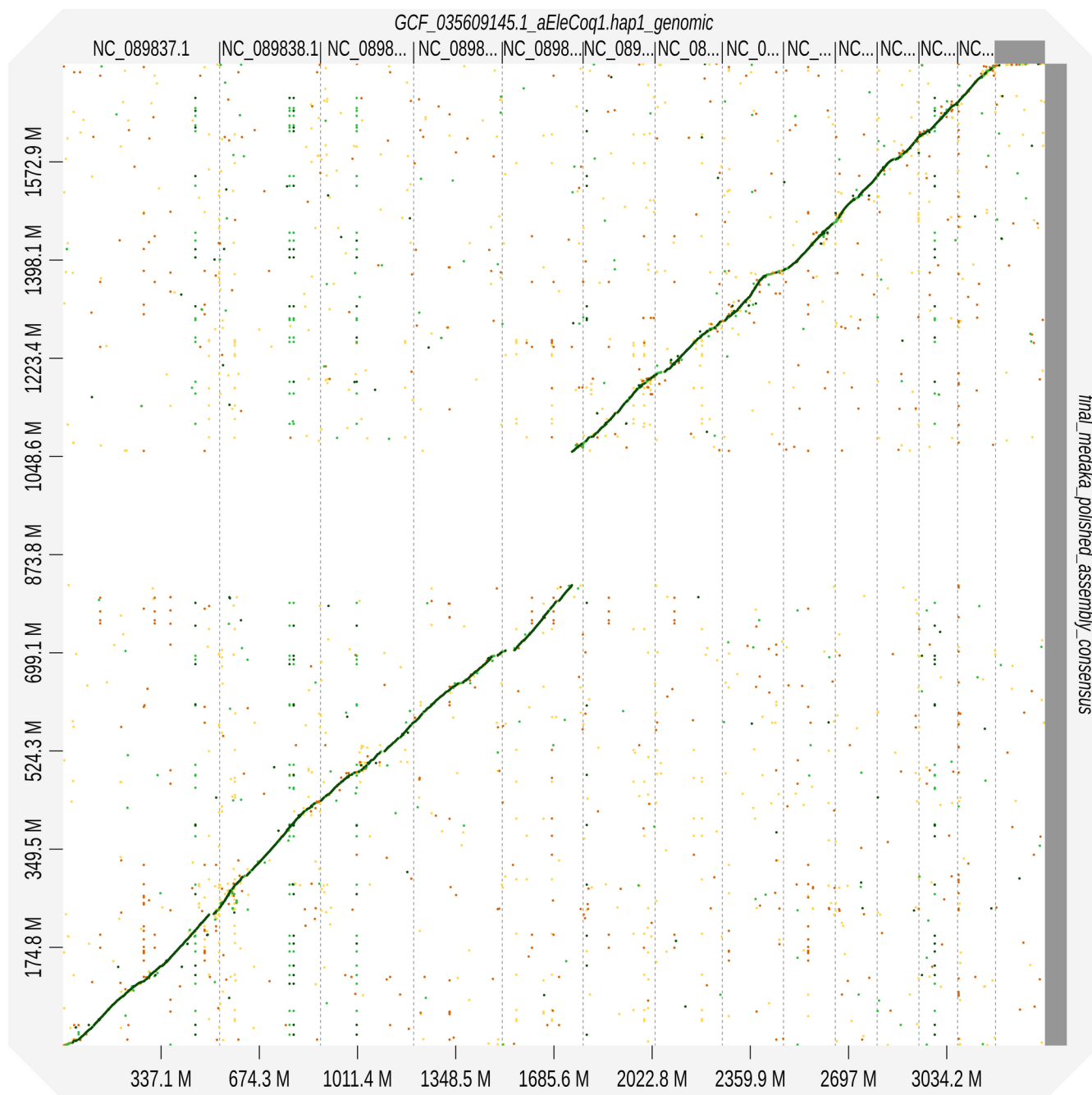


Figure SR6. Pre-scaffolding assembly aligned to the *Eleutherodactylus coqui* reference genome.

D-Genies v1.5.0 [19] dotplot of the Medaka-polished Flye assembly (34,106 contigs) against *E. coqui* haplotype 1 (RefSeq ID: GCF_035609145.1). Vertical axis: *Pristimantis euphronides* contigs.

Horizontal axis: *E. coqui* chromosomes. Color reflects nucleotide identity in percentage (%) using the

default D-Genies scale: white (0%), yellow (1-24%), orange (25-49%), light green (50-74%), and dark green (75-100%). (Also presented as Figure 3 in the main manuscript.)

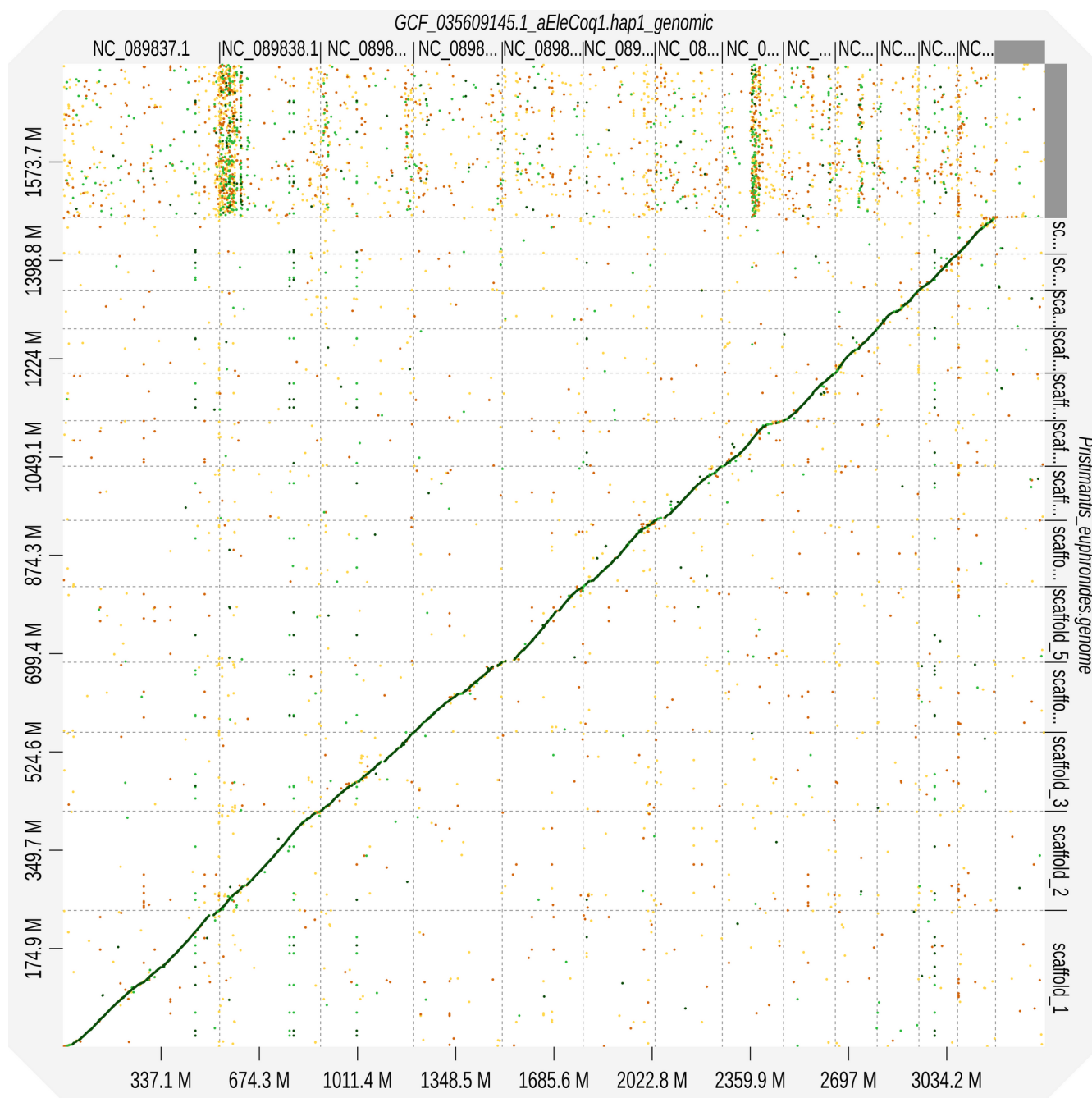


Figure SR7. Scaffolded assembly aligned to the *Eleutherodactylus coqui* reference genome. D-Genies v1.5.0 [19] dotplot of the RagTag-scaffolded *Pristimantis euphronides* genome against *E. coqui* haplotype 1 (RefSeq ID: GCF_035609145.1). Vertical axis: *P. euphronides* scaffolds. Horizontal axis: *E. coqui* chromosomes. Color reflects nucleotide identity in percentage (%) using the default D-Genies scale: white (0%), yellow (1-24%), orange (25-49%), light green (50-74%), and dark green (75-100%).

Pairwise alignment of the scaffolded *P. euphronides* assembly against the *Dendropsophus ebraccatus* genome (n = 15) revealed that four of the 13 scaffolds contained material aligning to two or three *D. ebraccatus* chromosomes (Table SR7, Figure SR8, panel A). Scaffold_1 (243 Mb) comprised material corresponding to *D. ebraccatus* chromosomes De6 (24 Mb aligned) and De5 (30 Mb aligned), with a predicted breakpoint at approximately 120 Mb (Figure SR9). Scaffold_3 (140 Mb) comprised material from three units: De8 (9 Mb aligned), De4 (16 Mb aligned), and De15 (15 Mb aligned), with predicted breakpoints at approximately 30 Mb and 95 Mb (Figure SR10). Scaffold_4 (125 Mb) comprised De8 (15 Mb aligned) and De11 (15 Mb aligned), with a predicted breakpoint at approximately 65 Mb (Figure SR11). Scaffold_6 (118 Mb) comprised De12 (15 Mb aligned) and De13 (14 Mb aligned), with a predicted breakpoint at approximately 70 Mb (Figure SR12). The remaining nine scaffolds each aligned to a single *D. ebraccatus* chromosome (Table SR8, Figure SR13). The *E. coqui* genome showed the same compound structure (Ec1 = De5 + De6, Ec3 = De4 + De15 + De8, Ec4 = De8 + De11, Ec6 = De12 + De13; Figure SR14).

Table SR7. *Dendropsophus ebraccatus* chromosome composition of compound scaffolds in the *Pristimantis euphronides* assembly. *D. ebraccatus* chromosome units listed in spatial order along the scaffold (left to right). Aligned Mb: total aligned sequence in the *P. euphronides* vs. *D. ebraccatus* pairwise comparison (minimum alignment block ≥ 1 Mb). Cross-validation: whether the same unit combination was also detected in the *P. euphronides* vs. *Engystomops pustulosus* comparison (n = 11). All alignments computed with D-Genies v1.5.0 [19] (Minimap2 v2.28 [22], default parameters; minimum alignment block ≥ 1 Mb).

Scaffold	Size (Mb)	<i>D. ebraccatus</i> units	Aligned (Mb)	<i>E. pustulosus</i> cross-validation
scaffold_1	243	De6 + De5	24 + 30	Ep2 + Ep3 (detected)
scaffold_3	140	De8 + De4 + De15	9 + 16 + 15	Ep10 + Ep3 (partial)
scaffold_4	125	De8 + De11	15 + 15	Ep2 + Ep11 (detected)
scaffold_6	118	De12 + De13	15 + 14	Ep6 + Ep7 (detected)

(dotted purple bar). Bars are colored by correspondence to *Dendropsophus ebraccatus* (n = 15) chromosome units. Four compound scaffolds (bold red labels: s1, s3, s4, s6) contain material from two or three *D. ebraccatus* units. Dashed red lines with scissors symbols mark breakpoints classified as predicted scaffolding artifacts. Solid green lines with filled circles mark breakpoints classified as predicted fusions. Black diagonal hatching: Z-candidate regions on s2 (0-42 Mb) and s8 (47-64 Mb). Scaffold sizes in Mb are shown in parentheses. Scaffold numbers follow *E. coqui* chromosome order, not *P. euphronides* size order. (B) Computationally inferred karyotype (n = 16), obtained by splitting at the three breakpoints classified as predicted scaffolding artifacts. Bold red labels: chromosomes derived from compound scaffolds. Green vertical lines within PrChr 3 and PrChr 6: retained predicted fusion boundaries. gray annotations (arrow s#): source scaffold. Z-candidate regions shown on PrChr 1 and PrChr 9. W-candidate contigs as in (A). colors as in (A). See section “Identification of candidate sex chromosome-linked regions” for Z- and W-candidate analysis. (Also presented as Figure 4 in the main manuscript.)

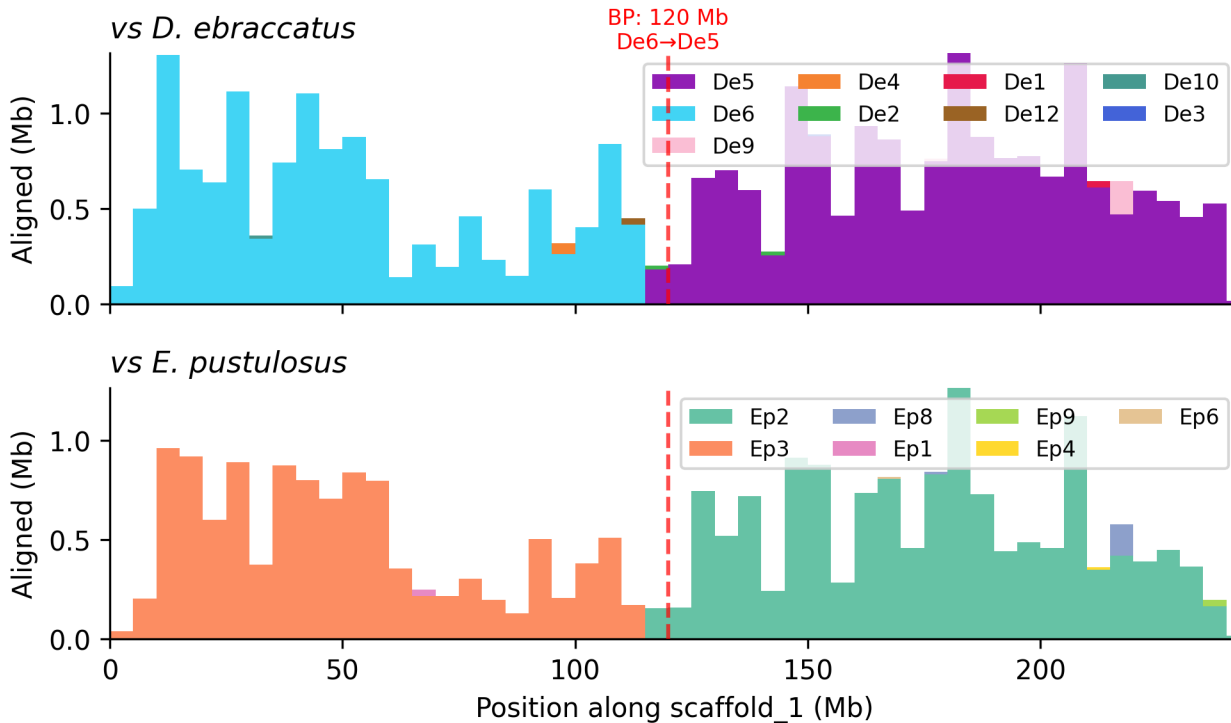


Figure SR9. Positional synteny along scaffold_1 (243 Mb) of *Pristimantis euphronides*. Aligned

sequence in 5 Mb bins along scaffold_1, colored by target chromosome. Upper panel: alignment to *Dendropsophus ebraccatus* (n = 15; GCF_027789765.1). Lower panel: alignment to *Engystomops pustulosus* (n = 11; GCF_040894005.1). Vertical axis: aligned bases per bin (Mb). Horizontal axis: position along scaffold_1 (Mb). colors represent individual target chromosomes (De1-De15 for *D. ebraccatus*, Ep1-Ep11 for *E. pustulosus*; see legends). Red dashed line: predicted breakpoint at ~120 Mb (De6 → De5). Alignments computed with D-Genies v1.5.0 [19] (Minimap2 v2.28 [22], default parameters); bins with <1 Mb total aligned sequence are included.

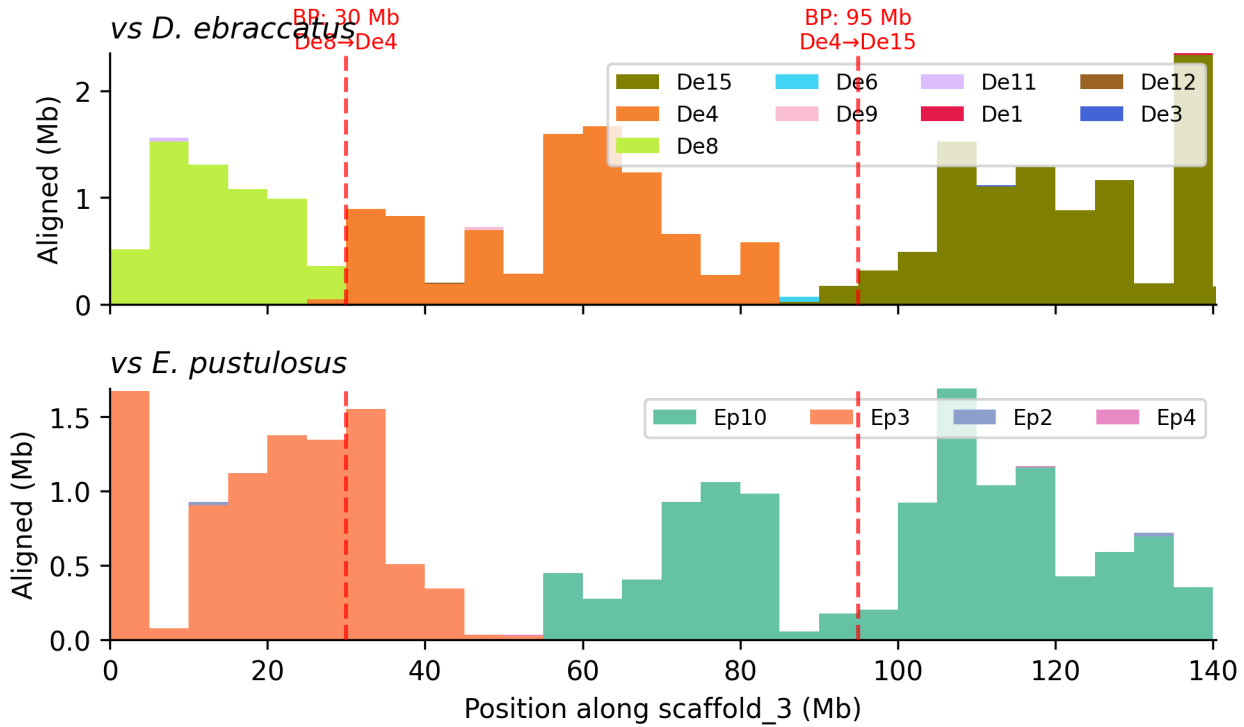


Figure SR10. Positional synteny along scaffold_3 (140 Mb) of *Pristimantis euphronides*. Aligned sequence in 5 Mb bins along scaffold_3, colored by target chromosome. Upper panel: alignment to *Dendropsophus ebraccatus* (n = 15; GCF_027789765.1). Lower panel: alignment to *Engystomops pustulosus* (n = 11; GCF_040894005.1). Vertical axis: aligned bases per bin (Mb). Horizontal axis: position along scaffold_3 (Mb). colors represent individual target chromosomes (De1-De15 for *D. ebraccatus*, Ep1-Ep11 for *E. pustulosus*; see legends). Red dashed lines: predicted breakpoints at ~30

Mb (De8 → De4) and ~95 Mb (De4 → De15). Alignments computed with D-Genies v1.5.0 [19] (Minimap2 v2.28 [22], default parameters).

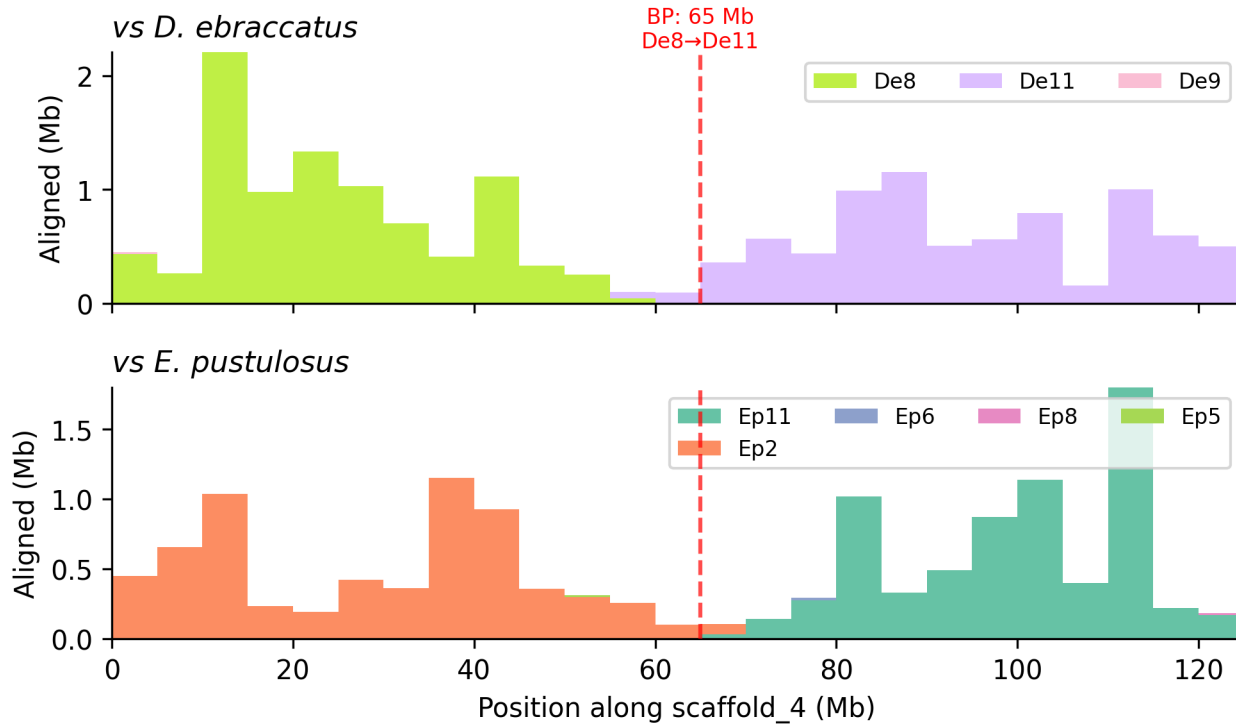


Figure SR11. Positional synteny along scaffold_4 (125 Mb) of *Pristimantis euphronides*. Aligned sequence in 5 Mb bins along scaffold_4, colored by target chromosome. Upper panel: alignment to *Dendropsophus ebraccatus* (n = 15; GCF_027789765.1). Lower panel: alignment to *Engystomops pustulosus* (n = 11; GCF_040894005.1). Vertical axis: aligned bases per bin (Mb). Horizontal axis: position along scaffold_4 (Mb). colors represent individual target chromosomes (De1-De15 for *D. ebraccatus*, Ep1-Ep11 for *E. pustulosus*; see legends). Red dashed line: predicted breakpoint at ~65 Mb (De8 → De11). Alignments computed with D-Genies v1.5.0 [19] (Minimap2 v2.28 [22], default parameters).

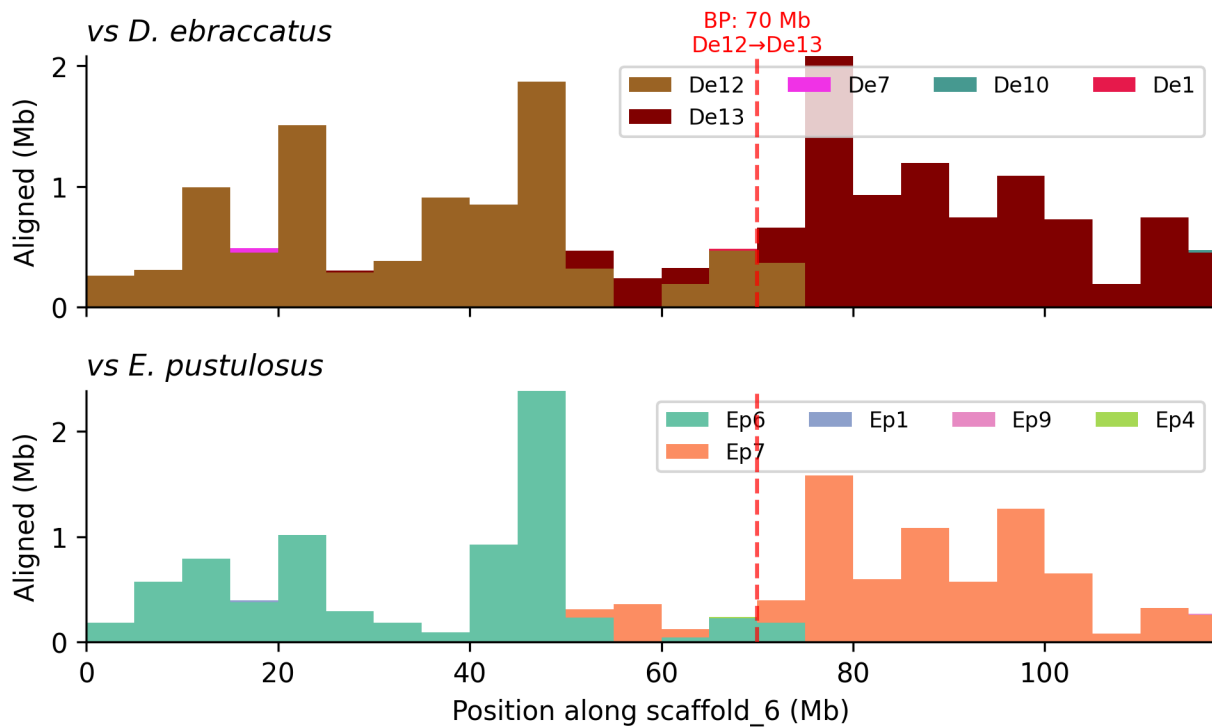


Figure SR12. Positional synteny along scaffold_6 (118 Mb) of *Pristimantis euphronides*. Aligned sequence in 5 Mb bins along scaffold_6, colored by target chromosome. Upper panel: alignment to *Dendropsophus ebraccatus* (n = 15; GCF_027789765.1). Lower panel: alignment to *Engystomops pustulosus* (n = 11; GCF_040894005.1). Vertical axis: aligned bases per bin (Mb). Horizontal axis: position along scaffold_6 (Mb). colors represent individual target chromosomes (De1-De15 for *D. ebraccatus*, Ep1-Ep11 for *E. pustulosus*; see legends). Red dashed line: predicted breakpoint at ~70 Mb (De12 → De13). Alignments computed with D-Genies v1.5.0 [19] (Minimap2 v2.28 [22], default parameters).

Table SR8. Karyotype reconstruction of *Pristimantis euphronides* across four hyloid species. Each row represents one *Dendropsophus ebraccatus* chromosome (De1-De15; n = 15; GCF_027789765.1). Columns: *D. ebraccatus* chromosome size (Mb); *P. euphronides* scaffolds with ≥5 Mb aligned sequence (aligned Mb in parentheses); number of scaffolds receiving alignments; status ("intact" = single scaffold, "split" = two or more); corresponding *Eleutherodactylus coqui* chromosome(s) (Ec1-Ec13; n = 13; GCF_035609145.1); corresponding *Engystomops pustulosus* chromosome(s) (Ep1-Ep11;

n = 11; GCF_040894005.1). Aligned Mb in parentheses throughout. All alignments computed with D-Genies v1.5.0 [19] (Minimap2 v2.28 [22], default parameters; minimum alignment block ≥ 1 Mb).

DenEbr unit	Accession	Mb	P. euphronides scaffolds	n	Status	E. coqui chr	E. pustulosus chr
De1	NC_091454.1	234	scaffold_2 (42 Mb)	1	intact	Ec2 (38 Mb)	Ep4 (37 Mb)
De2	NC_091455.1	220	scaffold_9 (23 Mb); scaffold_12 (15 Mb)	2	split	Ec9 (19 Mb); Ec12 (12 Mb); Ec1 (5 Mb)	Ep5 (37 Mb); Ep7 (6 Mb); Ep1 (6 Mb); Ep4 (6 Mb); Ep3 (5 Mb); Ep2 (5 Mb)
De3	NC_091456.1	202	scaffold_5 (35 Mb)	1	intact	Ec5 (29 Mb)	Ep1 (27 Mb)
De4	NC_091457.1	172	scaffold_11 (19 Mb); scaffold_3 (16 Mb)	2	split	Ec11 (16 Mb); Ec3 (12 Mb)	Ep7 (15 Mb); Ep10 (11 Mb)
De5	NC_091458.1	163	scaffold_1 (30 Mb)	1	intact	Ec1 (23 Mb)	Ep2 (24 Mb)
De6	NC_091459.1	153	scaffold_1 (24 Mb)	1	intact	Ec1 (19 Mb)	Ep3 (20 Mb)
De7	NC_091460.1	150	scaffold_7 (23 Mb)	1	intact	Ec7 (19 Mb)	Ep1 (22 Mb)
De8	NC_091461.1	127	scaffold_4 (15 Mb); scaffold_3 (9 Mb)	2	split	Ec4 (14 Mb); Ec3 (7 Mb)	Ep11 (12 Mb); Ep10 (7 Mb)
De9	NC_091462.1	122	scaffold_8 (23 Mb)	1	intact	Ec8 (20 Mb)	Ep8 (21 Mb)
De10	NC_091463.1	107	scaffold_10 (19 Mb)	1	intact	Ec10 (15 Mb)	Ep9 (15 Mb)
De11	NC_091464.1	102	scaffold_4 (15 Mb)	1	intact	Ec4 (13 Mb)	Ep2 (14 Mb)
De12	NC_091465.1	94	scaffold_6 (15 Mb)	1	intact	Ec6 (14 Mb)	Ep6 (13 Mb)
De13	NC_091466.1	84	scaffold_6 (14 Mb)	1	intact	Ec6 (11 Mb)	Ep7 (14 Mb)
De14	NC_091467.1	79	scaffold_13 (12 Mb)	1	intact	Ec13 (10 Mb)	Ep6 (11 Mb)
De15	NC_091468.1	76	scaffold_3 (15 Mb)	1	intact	Ec3 (12 Mb)	Ep3 (13 Mb)

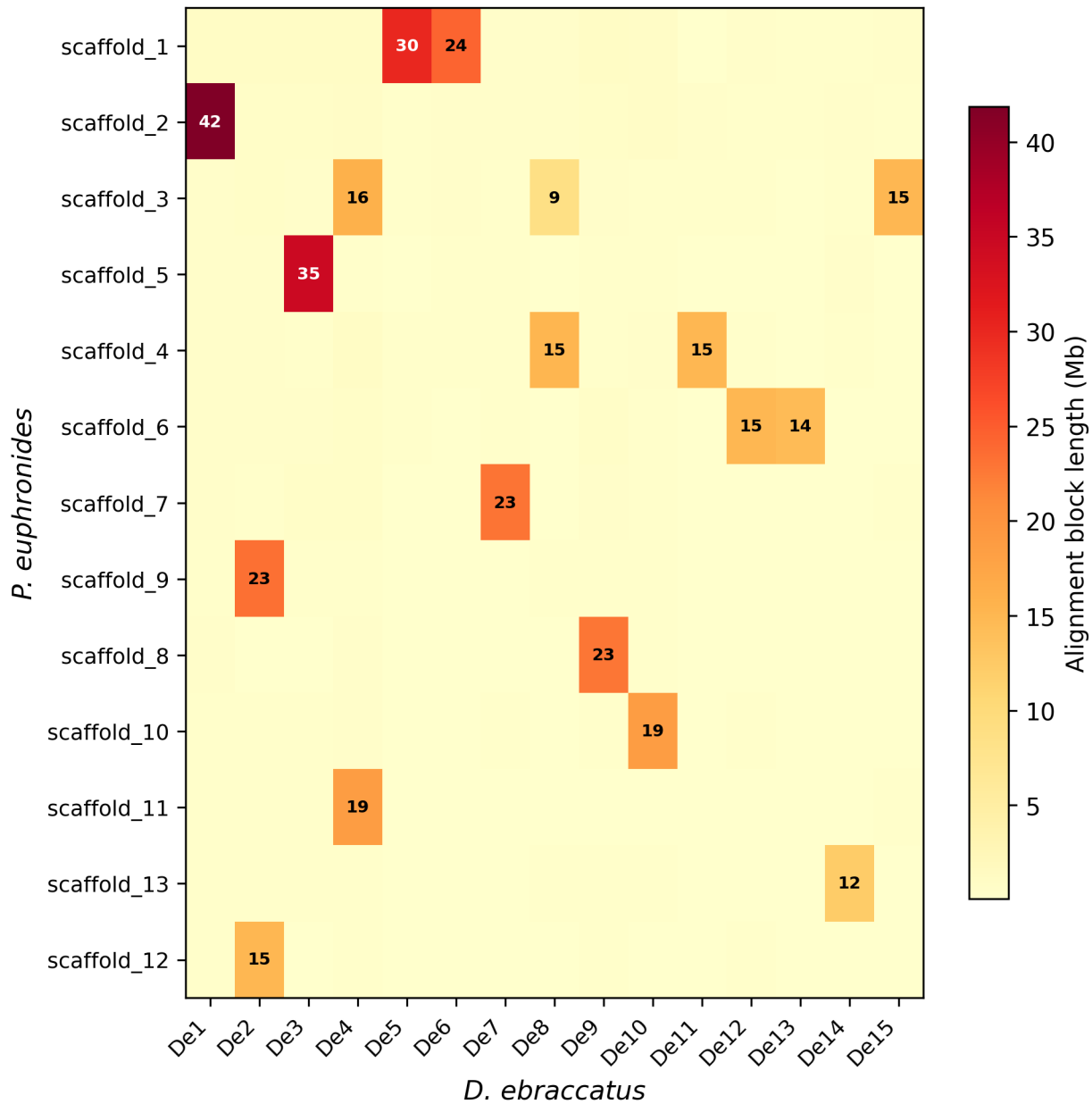


Figure SR13. Heatmap of alignment block length between *Pristimantis euphronides* scaffolds and *Dendropsophus ebraccatus* chromosomes. Total aligned sequence (Mb) between each *P. euphronides* scaffold (vertical axis; scaffolds 1-13) and each *D. ebraccatus* chromosome (horizontal axis; De1-De15; n = 15). Cell values: aligned Mb (minimum alignment block ≥ 1 Mb). color scale: pale (low) to dark red (high). Cells with <1 Mb total alignment not shown. Scaffolds with values in two or more columns are compound. Alignment computed with D-Genies v1.5.0 [19] (Minimap2 v2.28 [22], default

parameters). Genome assemblies: *P. euphronides* scaffolded assembly; *D. ebraccatus* paternal assembly (GCF_027789765.1).

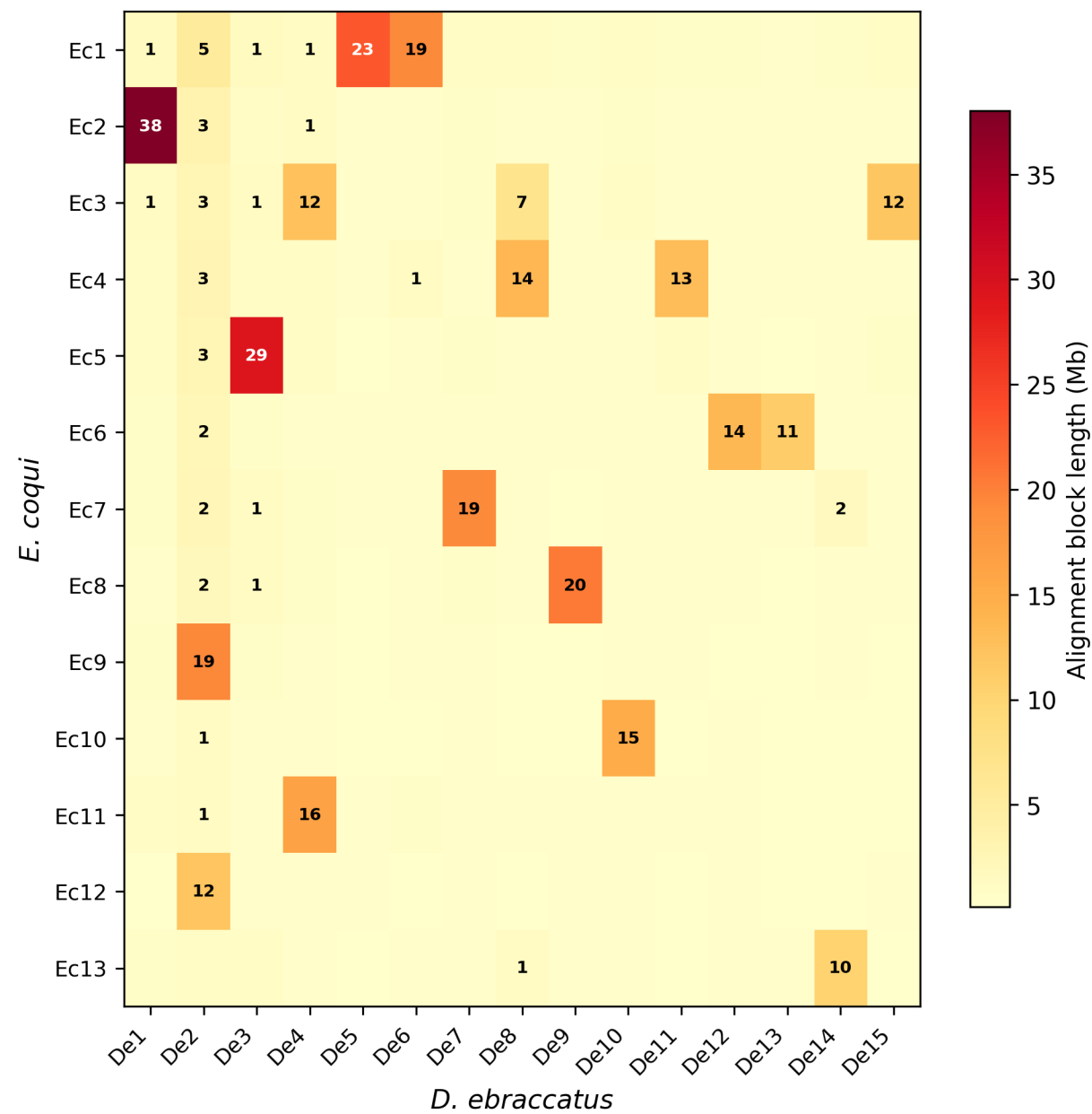


Figure SR14. Heatmap of alignment block length between *Eleutherodactylus coqui* and *Dendropsophus ebraccatus* chromosomes. Total aligned sequence (Mb) between each *E. coqui*

chromosome (vertical axis; Ec1-Ec13; n = 13) and each *D. ebraccatus* chromosome (horizontal axis; De1-De15; n = 15). Cell values: aligned Mb (minimum alignment block ≥ 1 Mb). color scale: pale (low) to dark red (high). Cells with <1 Mb total alignment not shown. *E. coqui* chromosomes with values in two or more columns are compound. Alignment computed with D-Genies v1.5.0 [19] (Minimap2 v2.28 [22], default parameters). Genome assemblies: *E. coqui* haplotype 1 (GCF_035609145.1); *D. ebraccatus* paternal assembly (GCF_027789765.1).

Alignment of *P. euphronides* against *E. pustulosus* (n = 11) detected three of the five breakpoints (scaffold_1, scaffold_4, scaffold_6), whereas the scaffold_3 breakpoints were not detected (Table SR8, Figure SR15). The *E. coqui* versus *E. pustulosus* comparison showed the corresponding pattern (Figure SR16). Alignment of *E. pustulosus* against *D. ebraccatus* completed the pairwise matrix (Figure SR17).

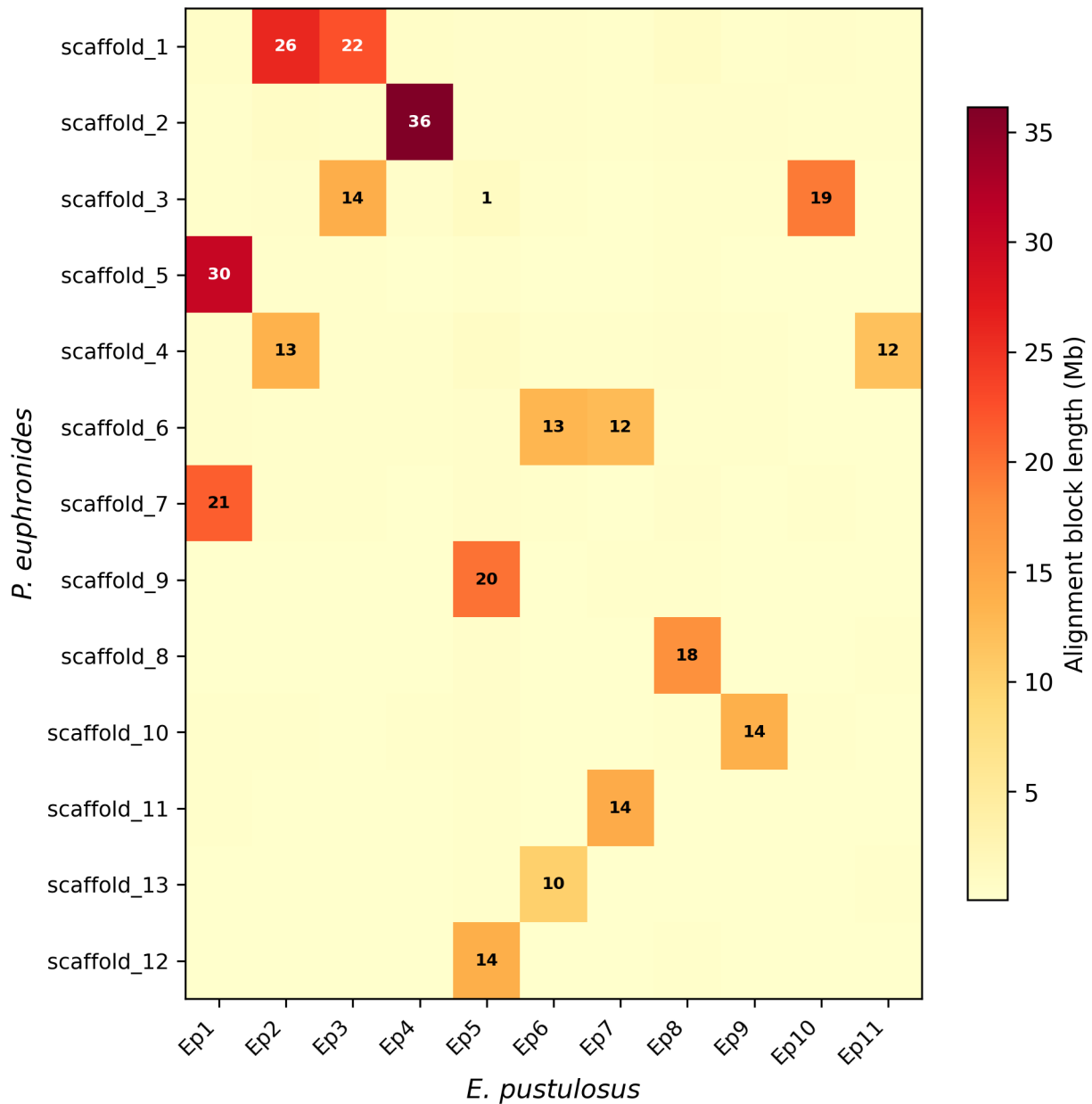


Figure SR15. Heatmap of alignment block length between *Pristimantis euphronides* scaffolds and *Engystomops pustulosus* chromosomes. Total aligned sequence (Mb) between each *P. euphronides* scaffold (vertical axis; scaffolds 1-13) and each *E. pustulosus* chromosome (horizontal axis; Ep1-Ep11; n = 11). Cell values: aligned Mb (minimum alignment block ≥ 1 Mb). color scale: pale (low) to dark red (high). Cells with < 1 Mb total alignment not shown. Alignment computed with D-Genies v1.5.0 [19]

(Minimap2 v2.28 [22], default parameters). Genome assemblies: *P. euphronides* scaffolded assembly; *E. pustulosus* maternal assembly (GCF_040894005.1).

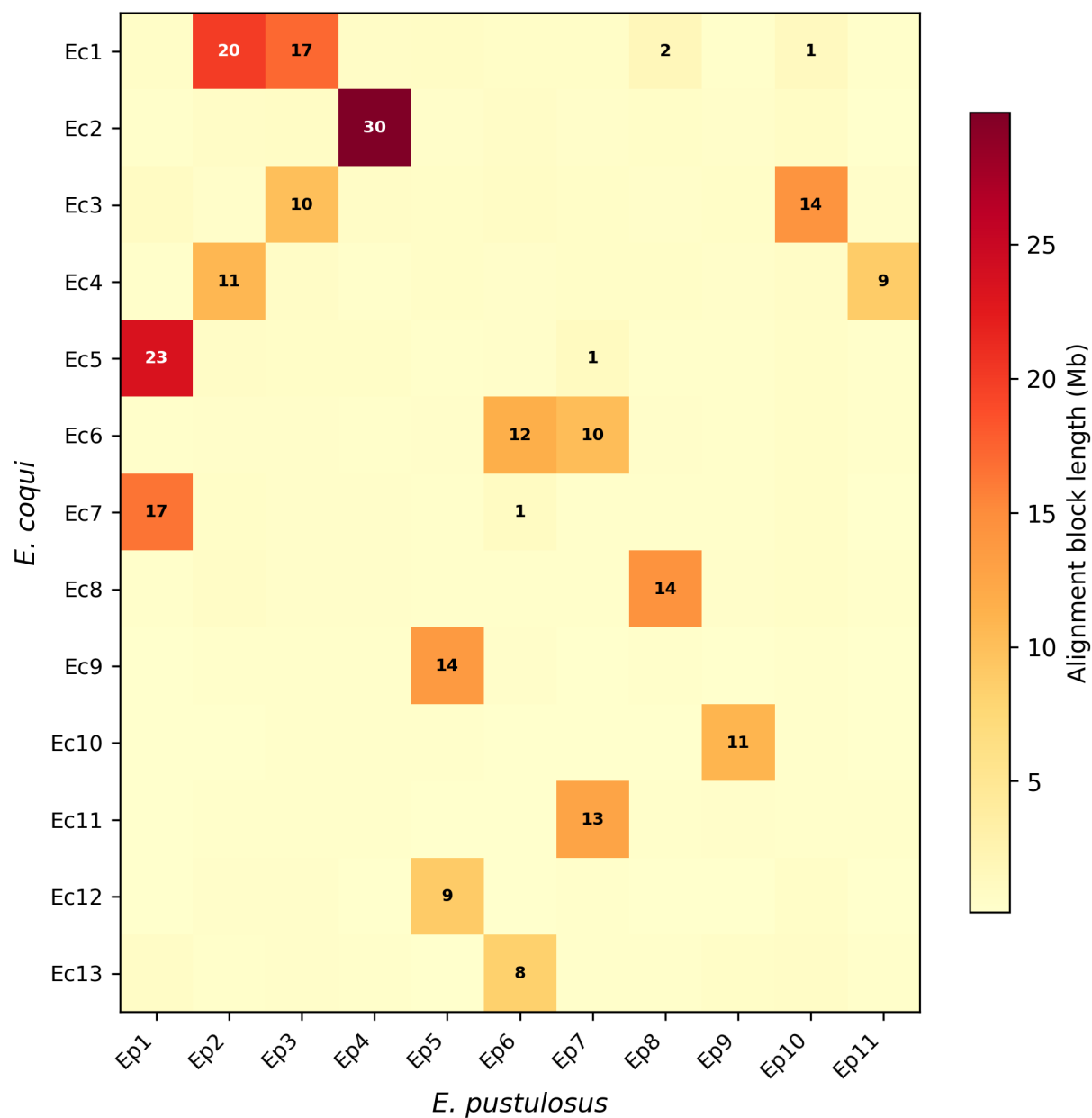


Figure SR16. Heatmap of alignment block length between *Eleutherodactylus coqui* and *Engystomops pustulosus* chromosomes. Total aligned sequence (Mb) between each *E. coqui*

chromosome (vertical axis; Ec1-Ec13; n = 13) and each *E. pustulosus* chromosome (horizontal axis; Ep1-Ep11; n = 11). Cell values: aligned Mb (minimum alignment block ≥ 1 Mb). color scale: pale (low) to dark red (high). Cells with <1 Mb total alignment not shown. Alignment computed with D-Genies v1.5.0 [19] (Minimap2 v2.28 [22], default parameters). Genome assemblies: *E. coqui* haplotype 1 (GCF_035609145.1); *E. pustulosus* maternal assembly (GCF_040894005.1).

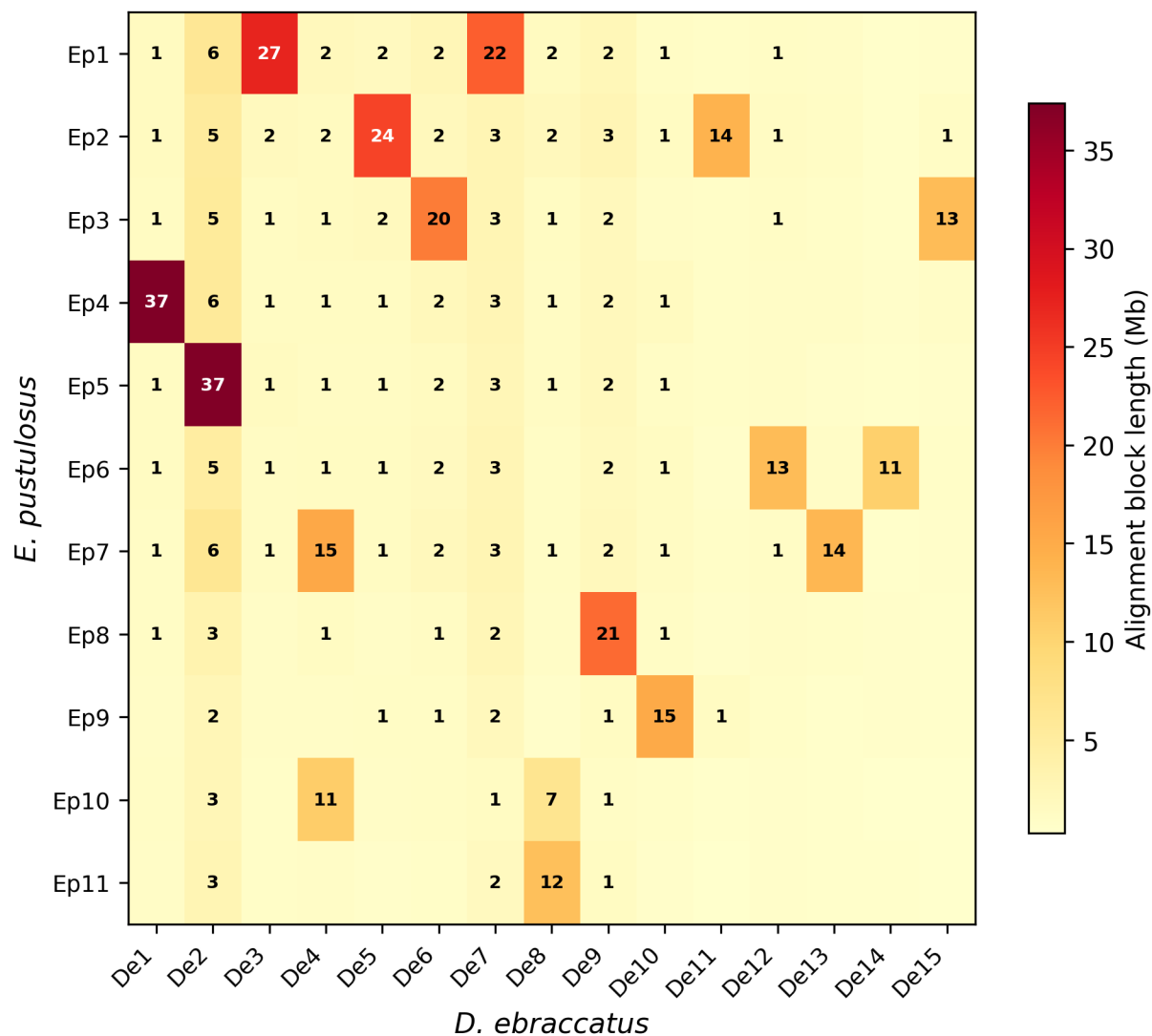


Figure SR17. Heatmap of alignment block length between *Engystomops pustulosus* and *Dendropsophus ebraccatus* chromosomes. Total aligned sequence (Mb) between each *E. pustulosus*

chromosome (vertical axis; Ep1-Ep11; n = 11) and each *D. ebraccatus* chromosome (horizontal axis; De1-De15; n = 15). Cell values: aligned Mb (minimum alignment block ≥ 1 Mb). color scale: pale (low) to dark red (high). Cells with < 1 Mb total alignment not shown. Alignment computed with D-Genies v1.5.0 [19] (Minimap2 v2.28 [22], default parameters). Genome assemblies: *E. pustulosus* maternal assembly (GCF_040894005.1); *D. ebraccatus* paternal assembly (GCF_027789765.1).

The contig allegiance analysis classified the five breakpoints within the four compound scaffolds as follows (Table SR9; Figure SR8, panel A; Figures SR18 and SR19). Three breakpoints were classified as predicted scaffolding artifacts: at scaffold_1 (~120 Mb; score = -3), the +/-5 Mb zone contained 21 contigs exclusively aligned to De5 and none to De6 (mixing = 0.00); at scaffold_3 (~30 Mb; score = -4), the zone was dominated by De15-aligned contigs with no De8 or De4 representation (mixing = 0.00); and at scaffold_3 (~95 Mb; score = -3), 16 contigs aligned exclusively to De4 (mixing = 0.00). At all three breakpoints classified as predicted scaffolding artifacts, the transition from one *D. ebraccatus* chromosome to another was abrupt, with no contigs carrying alignments to both *D. ebraccatus* chromosomes. Two breakpoints were classified as predicted fusions: at scaffold_4 (~65 Mb; score = +3), both De8- and De11-aligned contigs were present throughout the +/-5 Mb zone (mixing = 0.27) and the +/-15 Mb zone (mixing = 0.47); and at scaffold_6 (~70 Mb; score = +1), De12- and De13-aligned contigs co-occurred in the +/-5 Mb zone (mixing = 0.40), with the transition detected at 72.4 Mb.

Table SR9. Breakpoint classification within compound scaffolds. Classification based on contig allegiance analysis using pre-scaffolding (reference-naive) contigs aligned to *Dendropsophus ebraccatus*. Mixing: proportion of minority *D. ebraccatus* chromosome in the ± 5 Mb zone around the breakpoint (0.0 = pure; 0.5 = equal). Score: composite metric incorporating mixing in narrow and wide zones, transition sharpness, and split contig count (positive = predicted fusion; negative = predicted scaffolding artifact). Three breakpoints classified as predicted scaffolding artifacts + 13 scaffolds = 16 chromosomal units = n = 16 [24].

Scaffold	Position (Mb)	Boundary	Classification	Score	Mixing (±5 Mb)	Evidence
scaffold_1	~120	De6 De5	Predicted scaffolding artifact	−3	0.00	21 contigs exclusively De5; sharp transition (0.9 Mb)
scaffold_3	~30	De8 De4	Predicted scaffolding artifact	−4	0.00	Zone dominated by De15; no De8 or De4 in ±5 Mb
scaffold_3	~95	De4 De15	Predicted scaffolding artifact	−3	0.00	16 contigs exclusively De4; no transition detected
scaffold_4	~65	De8 De11	Predicted fusion	+3	0.27	De8 and De11 contigs intermixed; wide zone mixing = 0.47
scaffold_6	~70	De12 De13	Predicted fusion	+1	0.40	De12 and De13 co-occurring; transition at 72.4 Mb

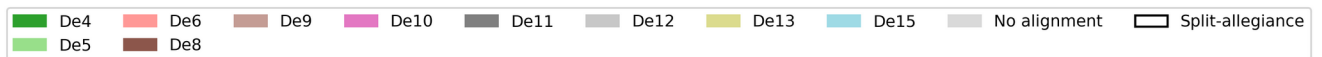
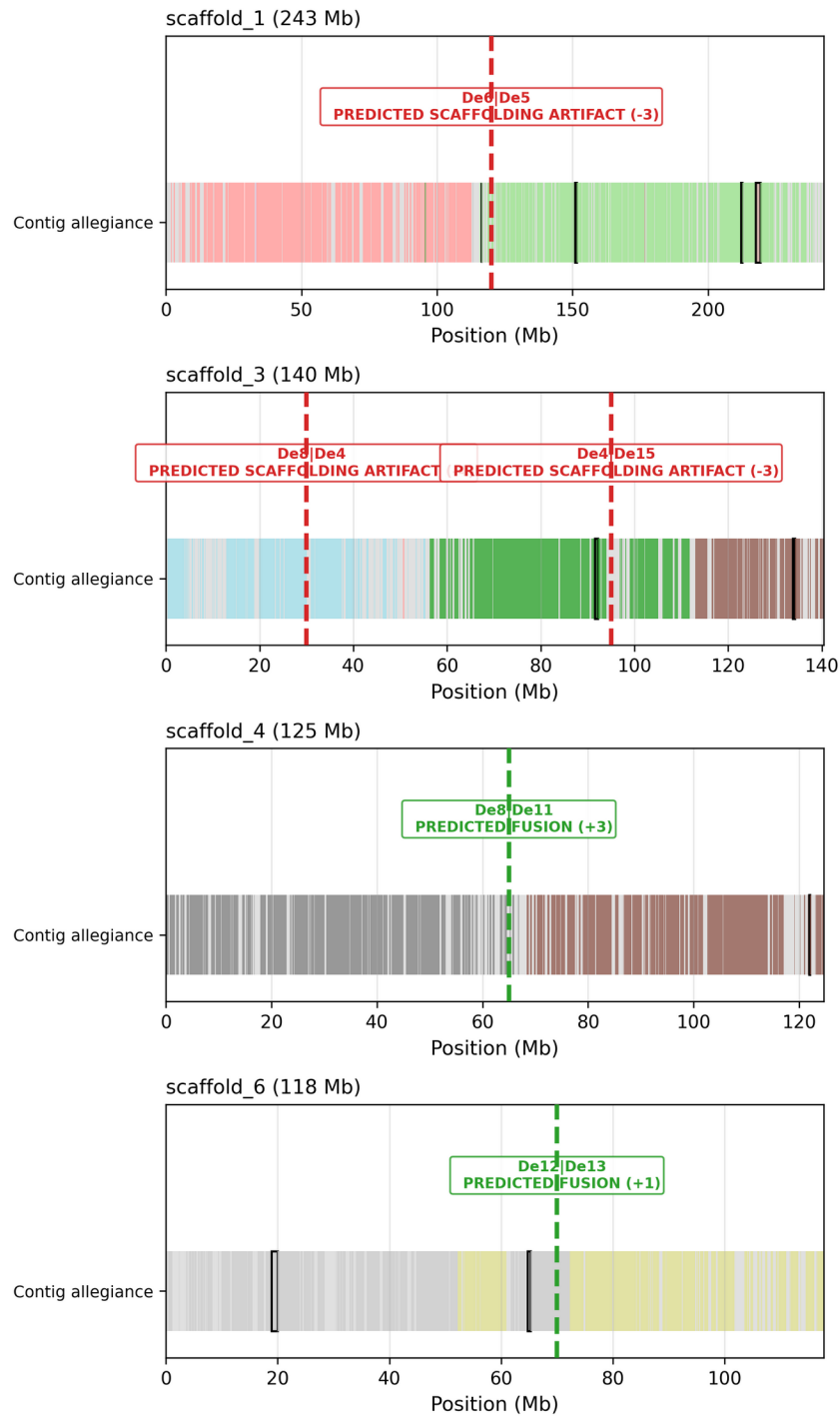


Figure SR18. Contig allegiance along the four compound scaffolds of *Pristimantis euphronides*.

Each panel shows one compound scaffold (scaffold_1, scaffold_3, scaffold_4, scaffold_6). Horizontal bars represent individual *P. euphronides* contigs positioned by their RagTag scaffold coordinates. Contigs are colored by their primary *Dendropsophus ebraccatus* chromosome alignment (De1-De15; see legend). gray: contigs with no *D. ebraccatus* alignment. Black outlines: contigs with split allegiance (alignments to two or more *D. ebraccatus* chromosomes). Vertical dashed red lines: breakpoints classified as predicted scaffolding artifacts. Vertical dashed green lines: breakpoints classified as predicted fusions. Horizontal axis: position along scaffold (Mb). Contigs were aligned to the *D. ebraccatus* paternal assembly (GCF_027789765.1) using Minimap2 v2.28 [22] (minimum alignment length ≥ 5 kb, minimum mapping quality ≥ 10). Scaffold coordinates derived from the RagTag [10] AGP file.

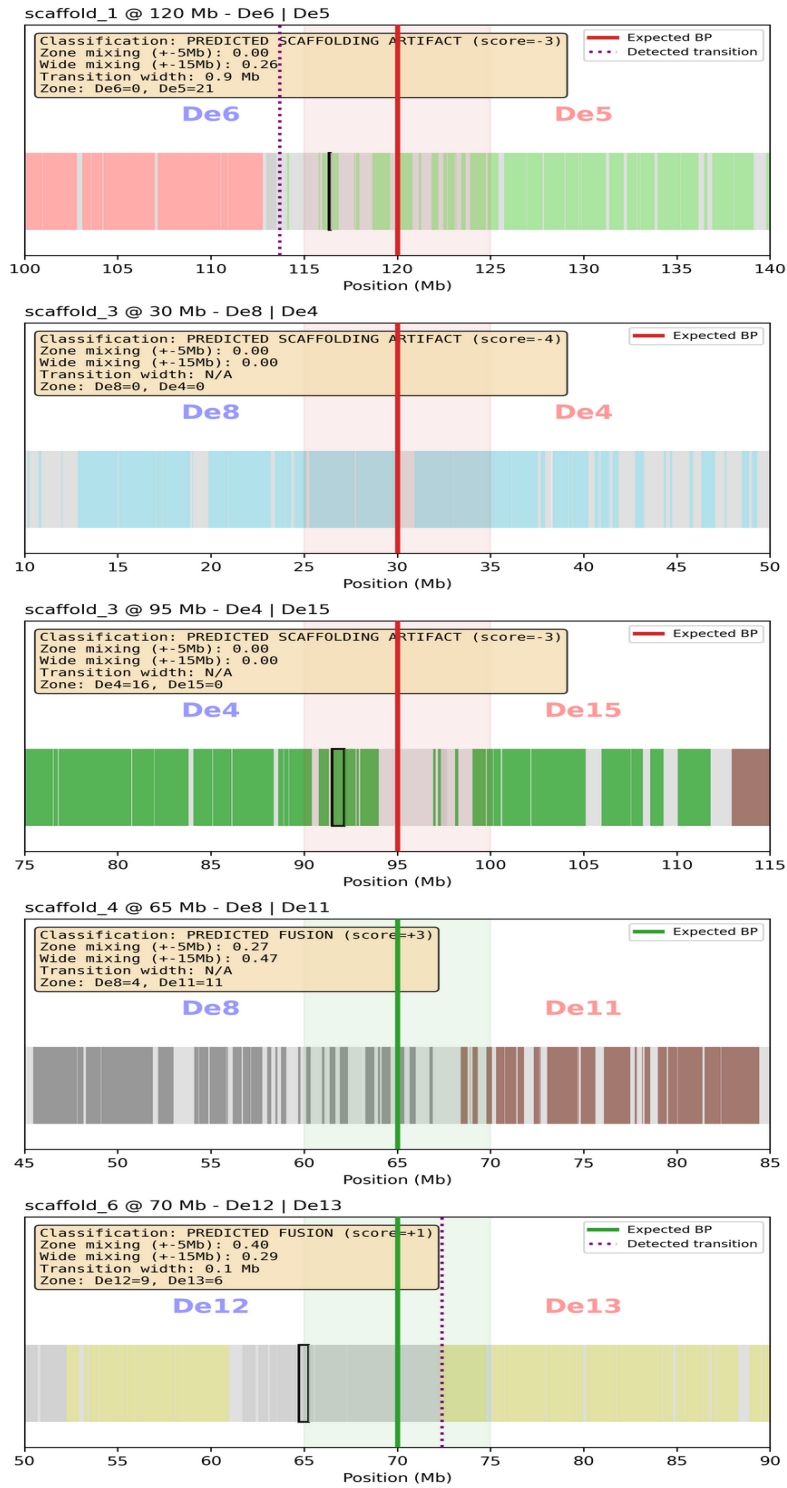


Figure SR19. Zoomed breakpoint views (± 20 Mb) for the five breakpoints within compound scaffolds of *Pristimantis euphronides*. Panels from top to bottom: scaffold_1 at ~ 120 Mb (De6|De5),

scaffold_3 at ~30 Mb (De8|De4), scaffold_3 at ~95 Mb (De4|De15), scaffold_4 at ~65 Mb (De8|De11), scaffold_6 at ~70 Mb (De12|De13). Horizontal bars: individual *P. euphronides* contigs colored by primary *Dendropsophus ebraccatus* chromosome alignment. Vertical solid line: expected breakpoint position. Vertical dotted purple line: detected transition position (where present). Shaded zone: ± 5 Mb around expected breakpoint. Inset box: classification, mixing score (± 5 Mb and ± 15 Mb zones), transition width, and contig counts per *D. ebraccatus* chromosome in the zone. Horizontal axis: position along scaffold (Mb). Contigs were aligned to the *D. ebraccatus* paternal assembly (GCF_027789765.1) using Minimap2 v2.28 [22] (minimum alignment length ≥ 5 kb, minimum mapping quality ≥ 10). Scaffold coordinates derived from the RagTag [10] AGP file.

Computational splitting of the 13 scaffolds at the three breakpoints classified as predicted scaffolding artifacts yields 16 predicted chromosomal units (13 original scaffolds + 1 additional unit from splitting scaffold_1 + 2 additional units from splitting scaffold_3 = 16), equal to the haploid chromosome count of $n = 16$ reported by Schmid et al. (2002) [24] on the basis of cytogenetic analysis of *P. euphronides* (then classified as *Eleutherodactylus euphronides*). The predicted 16-unit karyotype is presented schematically in Figure SR8, panel B. Of the 18 scaffolds outside the 13 largest (scaffolds 14-31; 7.5-351 kb), 17 produced chromosome-level alignments to one or more of the three Hyloidea references (Additional file 9: Table). Scaffold_23 (7,472 bp) produced no matches in the post-scaffolding D-Genies alignment. The karyotype reconstruction addresses autosomal scaffold composition. Sex chromosome candidates are presented in a following section (see “Identification of candidate sex chromosome-linked regions”).

Whole-genome alignment using D-Genies of the Workflow 2 post-scaffolding assembly (query) against the Workflow 1 post-scaffolding assembly (target) showed one-to-one correspondence between the 13 largest scaffolds produced by RagTag (Workflow 1) and Ragout (Workflow 2), with a continuous diagonal across the full genome length (Figure SR20). Of 7,804 Workflow 2 sequences, 7,765 (99.5%) aligned to Workflow 1, representing 1,600.9 Mb ($>99.9\%$ of Workflow 2 sequence). In the target assembly, 8,343 of 25,310 sequences (1,642.6 Mb; 93.9%) received alignments; the 16,967 unmatched sequences totaled 106.0 Mb (mean 6.2 kb). When the alignment was repeated with query and target reversed, the number of unmatched Workflow 1 sequences decreased from 16,967 to 4,613.

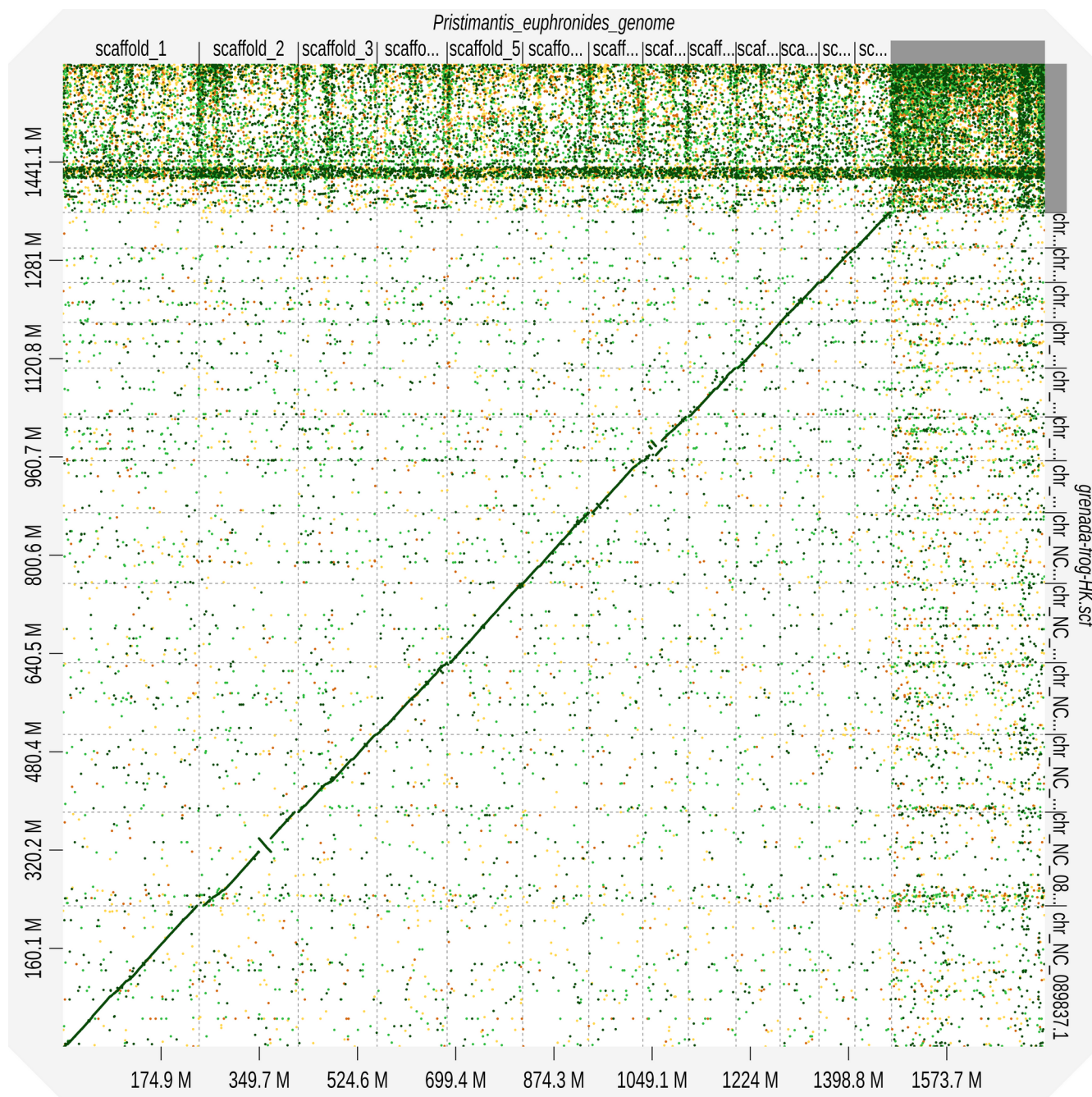


Figure SR20. Post-scaffolding inter-workflow scaffold concordance. D-Genies v1.5.0 [19] dotplot of the Workflow 2 scaffolded assembly (Ragout v2.3 [11]; query, vertical axis) against the Workflow 1

scaffolded assembly (RagTag v2.1.0 [10]; target, horizontal axis). Color reflects nucleotide identity in percentage (%) using the default D-Genies scale: white (0%), yellow (1–24%), orange (25–49%), light green (50–74%), and dark green (75–100%).

Alignment with previously published *Pristimantis euphronides* and *Eleutherodactylus johnstonei* sequences

To support species identification of the sequenced individual (“GrenadaFrog144”), all three publicly available *P. euphronides* short nucleotide reference sequences from a different individual (“voucher BWMC6918”) ranging from lengths of 493 bp to 2,570 bp, were aligned to the medaka-polished Workflow 1-assembly using Minimap2, and the resulting alignments were validated using BLASTn v2.14.1 [25]. Each of the reference loci mapped with high identity and full query coverage to distinct contigs in the assembly (Table SR10).

The mitochondrial fragment containing *tRNA-Phe*, *12S rRNA*, *tRNA-Val*, and *16S rRNA* (EF493527.1, 2570 bp) aligned to contig_57691, with 2570/2571 bp or 99.96% identity, containing a single gap but no mismatches. The *RAG-1* partial coding sequence (EF493427.1, 578 bp) aligned to contig_15875, with 577/578 bp matching bases or 99.83% identity, containing a single mismatch and no gaps. The *Tyr* exon 1 and partial CDS (EF493489.1, 493 bp) aligned to contig_19362, with 491/493 bp or 99.59% identity, containing two mismatches and no gaps.

The *RAG-1* partial coding sequence aligned on the same strand as its respective reference sequence, while the *Tyr* exon 1 and the mitochondrial locus were located on the reverse strand of their respective target contig. All alignments were contiguous, showed full-length coverage of the query sequences, and contained no insertions or deletions apart from a single gap in the alignment of the mitochondrial fragment. These BLASTn alignments confirmed that each contig showed the highest similarity, with percentages of identity ranging from 99.59% to 99.96% to the original *P. euphronides* short sequences published by [26], thereby providing molecular support for the species identification of the sequenced individual as *P. euphronides*, the Grenada frog.

When the same three assembly contigs were queried against the NCBI Nucleotide collection restricted to *Eleutherodactylus johnstonei* (taxid:350008), the top-scoring hits corresponded to the same three loci: the mitochondrial fragment containing *tRNA-Phe*, *12S rRNA*, *tRNA-Val*, and *16S rRNA* (OM914617.1) for contig_57691, the RAG1 partial coding sequence (JX298190.1) for contig_15875, and the Tyr partial coding sequence (OM928401.1) for contig_19362. Identity values were substantially lower at all three loci compared to the *P. euphronides* alignments (Table SR10): 81.16% (491/605 bp, 99 mismatches, 15 gaps) for the mitochondrial locus, 91.40% (1,201/1,314 bp, 113 mismatches, no gaps) for RAG-1, and 90.04% (443/492 bp, 49 mismatches, no gaps) for Tyr.

Table SR10. Alignment of Medaka-polished Flye assembly contigs against published *Pristimantis euphronides* and *Eleutherodactylus johnstonei* sequences. Three *P. euphronides* GenBank reference sequences (voucher BWMC6918; [26]) were aligned to assembly contigs identified by Minimap2 v2.28 [22]. BLASTn [25] pairwise alignment was performed via the NCBI web interface against the respective *P. euphronides* accessions (pairwise mode) and against the NCBI Nucleotide database restricted to *E. johnstonei* (taxid:350008). Identities: matching bases / alignment length (percent identity). Query region: coordinates on the assembly contig. Internal heading rows separate the two target species. For the *E. johnstonei* RAG-1 comparison, the full contig was used as query because the *E. johnstonei* RAG-1 reference (JX298190.1) covers a region extending beyond the *P. euphronides* amplicon. For mt and Tyr, the exact assembly subsequence aligning to the *P. euphronides* reference was used. *E. johnstonei* accessions: OM914617.1 (mt 16S rRNA), JX298190.1 (RAG-1), OM928401.1 (Tyr).

Locus	Assembly contig	Query region	GenBank ID	Identities	Number of mismatches	Number of gaps
Target species: <i>P. euphronides</i>						
mt (12S-16S rRNA)	57691	21168-23738	EF493527.1	2570 / 2571 (99.96%)	0	1
RAG-1	15875	29701-30278	EF493427.1	577 / 578 (99.83%)	1	0
Tyr	19362	95583-96075	EF493489.1	491 / 493 (99.59%)	2	0

Target species: <i>E. johnstonei</i>						
mt (12S-16S rRNA)	57691	21168-23738	OM914617.1	491 / 605 (81.16%)	99	15
RAG-1	15875	30607-31920	JX298190.1	1201 / 1314 (91.40%)	113	0
Tyr	19362	95583-96075	OM928401.1	443 / 492 (90.04%)	49	0

Annotation of scaffolded genome assembly

Annotation of the Medaka-polished, RagTag-scaffolded *P. euphronides* assembly (Workflow 1) using the HANNO pipeline with combined protein and mRNA evidence from seven Hyloidea RefSeq genomes yielded 28,071 predicted protein-coding gene models after removal of 119 coordinate-identical duplicates (Table SR11). Of these, 21,763 (77.5%) were supported by protein alignment evidence (miniprot v0.13-r248 [27]) and 6,308 (22.5%) by transcript alignment evidence (Minimap2). Gene structures comprised 226,443 exons (mean 8.07 per gene) with a total exon length of 57.2 Mb (3.27% of genome) and a total coding sequence length of 38.3 Mb (203,416 CDS features; 2.19% of the scaffolded genome size). Gene models spanned 1,011.4 Mb of genomic sequence (57.84%), including introns and UTRs. Mean CDS and mRNA lengths were 1,364 bp and 2,038 bp, respectively. Functional annotation through eggNOG-mapper v2.1.13 [28] assigned eggNOG orthologous groups to 20,213 genes (72.0%), functional descriptions to 19,993 (71.2%), COG categories to 19,836 (70.7%), Pfam domains to 19,399 (69.1%), gene symbols to 16,891 (60.2%), Gene Ontology terms to 15,185 (54.1%), KEGG orthologs to 13,618 (48.5%), and KEGG pathway mappings to 8,471 (30.2%). Of these 28,071 gene models, 23,485 (83.7%) were located on the 13 longest scaffolds, 46 (0.2%) on scaffolds 14-31, and 4,540 (16.2%) on unplaced contigs.

Completeness assessment using BUSCO against the tetrapoda_odb10 database (n = 5,310 orthologs) on the final BESTMODELS gene set identified 4,854 complete orthologs (C: 91.4%), including 4,745 single-copy (S: 89.4%) and 109 duplicated (D: 2.1%) (Table SR4). Additionally, 163 orthologs (F: 3.1%) were classified as fragmented and 293 (M: 5.5%) were missing (Table SR4). In comparison, annotation of the scaffolded *P. euphronides* assembly of Workflow 2 (Wtdbg2 v2.2, Ragout v2.3) using

the HANNO pipeline with identical evidence inputs achieved a lower BUSCO completeness of 88.4% (C), with 4,584 single-copy (S: 86.3%) and 110 duplicated (D: 2.1%) orthologs, 228 fragmented (F: 4.3%), and 388 missing (M: 7.3%) (Table SR4).

Table SR11. Structural and functional annotation statistics for the *Pristimantis euphronides* genome. Annotation was performed using the HANNO v0.4 pipeline [16] with combined protein and mRNA evidence from seven Hyloidea RefSeq genomes. Functional annotation was performed with eggNOG-mapper v2.1.13 [28]. Coordinate-identical duplicate gene models (same chromosome, start, end, and strand) were identified and excluded. Percentages in the "% of genome" column are calculated relative to the total scaffolded assembly size (1,748,533,034 bp including gaps). Percentages in the "% of genes" column refer to the total number of gene models (28,071). ¹ Gene span includes introns and UTRs (genomic distance from transcription start to end). ² CDS features = one per coding exon; CDS length = sum of coding bases only (excluding introns). ³ Functional description from eggNOG-mapper. (Also presented as Table 3 in the main manuscript.)

Category	Count	Total length (bp)	% of genome	% of genes
Structural features				
Protein-coding gene models	28,071	1,011,417,551 ¹	57.84	
mRNA	28,071	1,011,417,551 ¹	57.84	
Exons	226,443	57,210,245	3.27	
CDS features ²	203,416	38,293,302	2.19	
Mean exons per gene	8.07			
Mean CDS length per gene	1,364 bp			
Mean mRNA length per gene	2,038 bp			
Evidence source				
Protein-based models (miniprot)	21,763			77.5
Transcript-based models (minimap2)	6,308			22.5
Functional annotation				
Genes with gene symbol (Preferred_name)	16,891			60.2
Genes with functional description ³	19,993			71.2
Genes with eggNOG orthologous groups	20,213			72.0
Genes with Pfam domains	19,399			69.1

Genes with GO terms	15,185			54.1
Genes with COG category	19,836			70.7
Genes with KEGG orthologs (KO)	13,618			48.5
Genes with KEGG pathway mapping	8,471			30.2

The annotation is deposited in two formats. The primary set uses the scaffold and contig names used throughout this manuscript with HANNO internal gene model identifiers. A second set, reformatted with INSDC accessions and sequential locus tags (PRIEUP_00001-PRIEUP_28071), is directly comparable to the NCBI genome annotation package for GCA_965278355.2. A scaffold-to-accession mapping and the NCBI assembly report are included.

Identification of candidate sex chromosome-linked regions

The reconstructed ideogram (Figure SR21), based on the cytogenetic data of Schmid et al., 2002 [24] for *P. euphronides* and *P. shrevei* (reported as *E. euphronides* and *E. shrevei*), shows the Z chromosome as acrocentric ($1.4 \pm 0.1 \mu\text{m}$, 3.66% of haploid karyotype) and the W chromosome as submetacentric ($5.5 \pm 0.1 \mu\text{m}$, 13.54%), with 10-14 C-bands visible at underexposure. Fluorochrome staining revealed intense Hoechst fluorescence at the W centromere, four Q-bright bands along the W that were consistently M-negative and vice versa, weaker Hoechst in the distal q-arm, and DAPI positivity restricted to the centromeric region. The Z was uniformly dim except for M-positive telomeric heterochromatin.

Flow cytometry data were only available for *P. shrevei* [24] serving as proxy for *P. euphronides* genome parameter estimations. Converted from nuclear DNA content to megabases, the male haploid (1C) genome size was ~836 Mb, the W–Z chromosome size difference ~499 Mb, and the female haploid (1C) genome size ~1,335 Mb for this *Pristimantis* species (Table SR12).

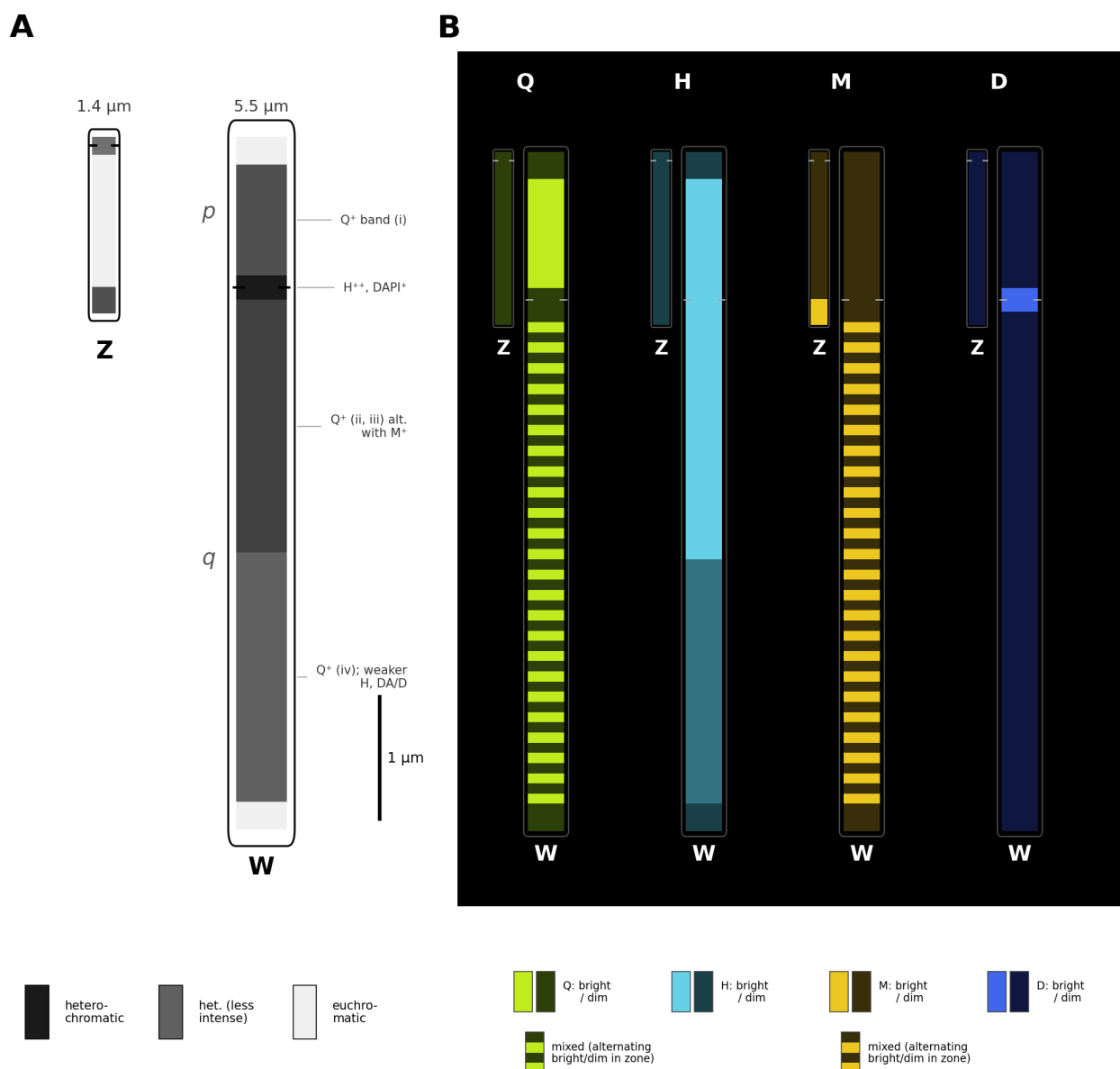


Figure SR21. ZW sex chromosome ideogram of *Pristimantis euphronides*. Reconstructed from the cytogenetic banding data of Schmid et al., 2002 [24]. Chromosomes are drawn to scale (Z: acrocentric, $\sim 1.4 \mu\text{m}$; W: submetacentric, $\sim 5.5 \mu\text{m}$). **(A)** C-banding ideogram in standard grayscale convention (dark: heterochromatic; light: euchromatic). Annotations indicate the four quinacrine-bright bands (i-iv) and the Hoechst-bright, DAPI-positive centromeric region. **(B)** Fluorochrome response per chromosomal region using representative emission colors: Q: quinacrine mustard (green); H: Hoechst 33258 (cyan); M: distamycin A/mithramycin (golden-yellow); D: DAPI (blue). Bright and dim colors represent strong versus quenched fluorescence; alternating stripes indicate zones where Q-positive and

M-positive bands alternate but individual band boundaries are not resolved. Zone boundaries are approximate, based on regional descriptions in the source publication ([24], pp. 85-86, 92; see Figures 3a-d, 5a, and 8a-b therein for original photomicrographs). Scale bar: 1 μ m.

Table SR12. Genome parameters for *Pristimantis shrevei*. Derived from flow cytometry data [24] using 1 pg DNA $\approx$ 978 Mb. W-Z: size difference between W and Z chromosome.

Parameter	Value
Male diploid (2C) nuclear DNA content (pg)	1.71
Female diploid (2C) nuclear DNA content (pg)	2.22
W-Z (Mb)	~499
Male haploid (1C) genome size (Mb)	~836
Female haploid (1C) genome size (Mb)	~1335

The primary assembly comprised 34,106 contigs. Partitioning by the RagTag AGP assigned 8,815 contigs (1,476.6 Mb, N50 386,935 bp, GC 42.48%) to the placed set and 25,291 contigs (271.1 Mb, N50 21,959 bp, GC 45.9%) to the unplaced set (Table SR13).

Table SR13. Summary statistics of placed and unplaced contig sets. Contigs were partitioned based on the RagTag [10] AGP file. Assembly statistics were computed with SeqKit v2.10.0 [2] (stats -a).

Set	Contigs	Total length (bp)	N50 (bp)	Min (bp)	Max (bp)	GC (%)
Placed	8,815	1,476,572,988	386,935	626	3,443,368	42.48
Unplaced	25,291	271,083,080	21,959	44	358,340	45.90
Total	34,106	1,747,656,068				

A multi-criterion analysis was applied to the 13 longest scaffolds (1,476 Mb total) of the RagTag-scaffolded primary assembly. The scaffold-level coverage screen identified two scaffolds with significant enrichment of half-coverage bins. The genome-wide reference coverage was 28.99× (median of per-scaffold means), yielding a half-coverage cutoff of 18.12×. Of 1,482 1 Mb bins across the 13 scaffolds, 28 (1.9%) fell below this threshold. Scaffold 2 contained 18 half-coverage bins out of 177 (10.2%; Bonferroni-corrected $p = 1.3 \times 10^{-7}$) and scaffold 8 contained 10 out of 82 (12.2%; Bonferroni-corrected $p = 4.7 \times 10^{-5}$). No other scaffold contained any half-coverage bins (Table SR14).

Table SR14. Scaffold-level screening for enrichment of half-coverage 1 Mb bins. The genome-wide reference coverage was 28.99× (median of per-scaffold means). Half-coverage bins were defined as those with mean depth $< 18.12 \times (0.625 \times \text{reference})$. The genome-wide background rate was $28/1,482 = 1.89\%$. Each scaffold was tested for enrichment by a one-sided binomial test with Bonferroni correction for 13 tests ($\alpha = 0.05$, corrected $\alpha = 0.0038$).

Scaffold	1 Mb bins	Observed	Expected	Fraction (%)	p (raw)	p (Bonferroni)	Flagged
scaffold_1	243	0	4.59	0.0	1.00	1.00	no
scaffold_2	177	18	3.34	10.2	1.00×10^{-8}	1.30×10^{-7}	YES
scaffold_3	141	0	2.66	0.0	1.00	1.00	no
scaffold_4	125	0	2.36	0.0	1.00	1.00	no
scaffold_5	135	0	2.55	0.0	1.00	1.00	no
scaffold_6	118	0	2.23	0.0	1.00	1.00	no
scaffold_7	97	0	1.83	0.0	1.00	1.00	no
scaffold_8	82	10	1.55	12.2	3.59×10^{-6}	4.66×10^{-5}	YES
scaffold_9	85	0	1.61	0.0	1.00	1.00	no
scaffold_10	79	0	1.49	0.0	1.00	1.00	no
scaffold_11	70	0	1.32	0.0	1.00	1.00	no
scaffold_12	64	0	1.21	0.0	1.00	1.00	no
scaffold_13	66	0	1.25	0.0	1.00	1.00	no

Boundary detection by optimal binary segmentation identified the half-coverage block on scaffold 2 at Mb 0-42 (43 Mb; mean coverage 21.24×, flanking mean 31.92×, ratio 0.67) and on scaffold 8 at Mb

47-64 (18 Mb; mean coverage 18.96×, flanking mean 27.94×, ratio 0.68). Both regions were best described by a three-segment model. The two standard deviation extension did not shift the core boundaries on either scaffold. Together, the two regions comprised 61 Mb of candidate Z-candidate-linked sequence (Table SR15).

Table SR15. Z-candidate region boundaries determined by optimal binary segmentation.

Boundaries were identified using three-segment models selected by Bayesian Information Criterion (BIC) on per-Mb coverage profiles with outlier capping at 2× per-scaffold median. Extension used a flanking mean minus 2 SD threshold. Coordinates are inclusive Mb.

Scaffold	Start (Mb)	End (Mb)	Length (Mb)	Mean cov. (×)	Flanking mean (×)	Ratio	Model
scaffold_2	0	42	43	21.24	31.92	0.665	3-segment
scaffold_8	47	64	18	18.96	27.94	0.679	3-segment
Total			61				

Inter-haplotype assembly concordance was reduced in the Z-candidate regions. The two Z-candidate regions showed a mean pairwise identity of 96.5% (scaffold 2: 96.51%, 596 alignments; scaffold 8: 96.51%, 261 alignments), compared to 97.9% across the 14 autosomal regions (range 97.0-98.2%). Both Z-candidate values ranked below all 14 autosomal values (exact Mann-Whitney U test, $U = 28/28$, $p = 0.008$). In the within-scaffold comparison, Z-candidate identity (96.5%) was lower than the three flanking autosomal regions on the same scaffolds (mean 97.7%), with maximum rank separation ($U = 6/6$); however, with $n = 2$ versus $n = 3$, the minimum achievable exact p-value is 0.10 ($1/C(5,2) = 1/10$), precluding significance at conventional thresholds despite complete separation of ranks (Table SR16).

Table SR16. Inter-haplotype assembly concordance for Z-candidate and autosomal regions. Mean pairwise identity was computed from minimap2 [22] asm5 alignments between the two independently RagTag v2.1.0 [10]-scaffolded haplotype assemblies. Significance was assessed by exact one-tailed

Mann-Whitney U tests. For the within-scaffold comparison, the minimum achievable p-value with $n = 2$ versus $n = 3$ is 0.10.

Region		Type	Identity (%)			Alignments			
Chr 2: 0-42 Mb		Z_candidate	96.51			596			
Chr 2: 42-185 Mb		Autosomal	98.19			816			
Chr 8: 47-64 Mb		Z_candidate	96.51			261			
Chr 8: 0-47 Mb		Autosomal	98.01			318			
Chr 8: 64-87 Mb		Autosomal	96.99			277			
Chr 1 (full)		Autosomal	98.21			1,611			
Chr 3 (full)		Autosomal	97.75			1,263			
Chr 4 (full)		Autosomal	98.06			1,010			
Chr 5 (full)		Autosomal	98.21			946			
Chr 6 (full)		Autosomal	98.15			804			
Chr 7 (full)		Autosomal	98.18			694			
Chr 9 (full)		Autosomal	98.24			719			
Chr 10 (full)		Autosomal	97.45			766			
Chr 11 (full)		Autosomal	97.54			571			
Chr 12 (full)		Autosomal	98.12			606			
Chr 13 (full)		Autosomal	97.81			577			
Comparison		Z mean (%)	Autosomal mean (%)		n_Z	n_Auto	U	U_max	p (exact)
Z vs all autosomal		96.51	97.92		2	14	28	28	0.008
Z vs same-scaffold flanks		96.51	97.73		2	3	6	6	0.10

Repeat density was significantly elevated in the Z-candidate regions. Across the 13 scaffolds, Z-candidate bins contained 47.8% RepeatMasker v4.2.3-masked [29] sequence versus 39.6% for autosomal bins (Mann-Whitney U, $p = 2.7 \times 10^{-12}$). This enrichment was consistent across repeat classes: LTR elements (4.3% versus 2.8%, $p = 2.1 \times 10^{-13}$), DNA transposons (13.4% versus 10.6%, $p = 3.1 \times 10^{-10}$), simple repeats including low complexity (2.7% versus 1.6%, $p = 3.5 \times 10^{-7}$), LINEs (6.3% versus 4.5%, $p = 1.3 \times 10^{-5}$), and satellites (1.3% versus 0.7%, $p = 5.1 \times 10^{-5}$). SINEs and unclassified repeats showed no significant difference ($p = 0.69$ and $p = 0.40$, respectively). WindowMasker v1.0.0 [30] confirmed the overall pattern (40.3% versus 35.1%, $p = 3.8 \times 10^{-11}$). The within-scaffold comparison yielded equivalent results: Z-candidate bins were significantly more repeat-

rich than flanking autosomal bins on the same scaffolds (RepeatMasker total 47.8% versus 38.8%, $p = 1.9 \times 10^{-11}$; WindowMasker 40.3% versus 34.5%, $p = 5.8 \times 10^{-11}$) (Table SR17).

Table SR17. Repeat density in Z-candidate versus autosomal 1 Mb bins. Repeat content was characterized using RepeatMasker v4.2.3 [29] with Dfam 3.9 [31] (taxon Anura) and WindowMasker [30] with shared k-mer counts from the full primary assembly. Significance was assessed by one-tailed Mann-Whitney U tests with normal approximation (H_1 : Z-candidate > autosomal).

Z-candidate (61 bins, 61.0 Mb) versus all autosomal (1,421 bins, 1,414.5 Mb)			
Repeat class	Z-candidate (%)	Autosomal (%)	p
RepeatMasker total	47.84	39.64	$2.73 \times 10^{-12}****$
WindowMasker total	40.32	35.15	$3.82 \times 10^{-11}****$
LTR	4.34	2.75	$2.11 \times 10^{-13}****$
DNA transposon	13.35	10.65	$3.08 \times 10^{-10}****$
Simple repeat	2.68	1.62	$3.47 \times 10^{-7}****$
LINE	6.32	4.49	$1.33 \times 10^{-5}****$
Satellite	1.29	0.71	$5.08 \times 10^{-5}****$
SINE	0.07	0.08	0.69
Other/Unclassified	19.79	19.35	0.40
Z-candidate (61 bins) versus same-scaffold flanking autosomal (198 bins)			
Repeat class	Z-candidate (%)	Flanking (%)	p
RepeatMasker total	47.84	38.82	$1.89 \times 10^{-11}****$
WindowMasker total	40.32	34.48	$5.82 \times 10^{-11}****$
LTR	4.34	2.72	$5.76 \times 10^{-12}****$
DNA transposon	13.35	10.49	$6.41 \times 10^{-10}****$
Simple repeat	2.68	1.60	$1.36 \times 10^{-6}****$
LINE	6.32	4.61	$6.27 \times 10^{-5}****$
Satellite	1.29	0.66	$1.44 \times 10^{-4}****$
SINE	0.07	0.13	0.57
Other/Unclassified	19.79	18.59	0.07

Gene density did not differ between Z-candidate and autosomal regions. Z-candidate bins contained 17.6 genes per Mb (1,071 genes across 61 bins), compared to 15.8 genes per Mb across all autosomal bins (22,494 genes, 1,420 bins; Mann-Whitney U, $p = 0.91$) and 17.8 genes per Mb in the flanking autosomal bins on the same scaffolds (3,497 genes, 197 bins; $p = 0.47$). The autosomal bin count (1,420) differs by one from the repeat analysis (1,421 bins) because one terminal bin contained sequence but no gene midpoint. Gene body length was not elevated in Z-candidate genes (median 13,039 bp versus 17,000 bp in autosomal genes and 16,972 bp in flanking genes; Mann-Whitney U, $p \geq 0.99$ for both comparisons) (Table SR18).

Table SR18. Gene density and gene body length in Z-candidate versus autosomal regions. Gene density (genes per 1 Mb bin) was computed from the HANNO v0.4 [16] BESTMODELS annotation on the scaffolded assembly. Gene body length is the genomic span (end – start) of each gene model. Significance was assessed by one-tailed Mann-Whitney U tests with normal approximation: H_1 for gene density was Z-candidate < autosomal; H_1 for gene body length was Z-candidate > autosomal.

Gene density per 1 Mb bin					
Comparison	Group	Bins	Genes	Mean genes/Mb	p
Z vs all autosomal	Z-candidate	61	1,071	17.56	0.91
	Autosomal	1,420	22,494	15.84	
Z vs same-scaffold flanks	Z-candidate	61	1,071	17.56	0.47
	Flanking autosomal	197	3,497	17.75	
Gene body length					
Comparison	Group	Genes	Mean (bp)	Median (bp)	p
Z vs all autosomal	Z-candidate	1,071	38,348	13,039	1.00
	Autosomal	22,494	41,237	17,000	
Z vs same-scaffold flanks	Z-candidate	1,071	38,348	13,039	1.00
	Flanking autosomal	3,497	40,303	16,972	

Per-gene coverage classification of 16,912 named genes on the scaffolded assembly at a reference coverage of 29.09× assigned 13,851 (81.9%) to Auto_1.0×, 963 (5.7%) to Hemi_0.5×, 815 (4.8%) to

High_Coverage/Repeat, and 1,283 (7.6%) to Low_Coverage/Other. Of these, 15,180 resided on placed contigs and 1,732 on unplaced contigs within the scaffolded assembly (Table SR19).

Table SR19. Per-gene coverage classification on the scaffolded assembly. Gene body coordinates from the HANNO v0.4 [16] BESTMODELS annotation were supplied to mosdepth v0.3.11 [23] (--fast-mode). Reference coverage (total_region): 29.09×. Coverage classes are defined relative to this reference.

Coverage class	Genes	% of total
Auto_1.0×	13,851	81.9
Hemi_0.5×	963	5.7
High_Coverage/Repeat	815	4.8
Low_Coverage/Other	1,283	7.6
Total	16,912	
Location	Genes	
Placed	15,180	
Unplaced	1,732	

The bp-weighted mean coverage of placed contigs on the pre-scaffolding assembly was 29.81×. Placed contigs had a per-contig median coverage of 27.22× (mean 29.33×, SD 48.31). Unplaced contigs had a per-contig median of 15.62× (mean 46.76×, SD 455.22, maximum 33,407×). Among placed contigs, 5,975 (1,343.3 Mb) were classified as Auto_1.0×, 1,203 (45.2 Mb) as Hemi_0.5×, 666 (62.8 Mb) as High_Coverage/Repeat, and 971 (25.2 Mb) as Low_Coverage/Other. Among unplaced contigs, 6,302 (66.5 Mb) were classified as Hemi_0.5×, 4,310 (87.2 Mb) as Auto_1.0×, 4,513 (55.2 Mb) as High_Coverage/Repeat, 10,049 (62.1 Mb) as Low_Coverage/Other, and 117 (0.2 Mb) as Zero (Table SR20).

Table SR20. Per-contig coverage classification on the pre-scaffolding assembly. Per-contig mean coverage was computed with mosdepth v0.3.11 [23] (--fast-mode, --no-per-base). The autosomal reference baseline was the bp-weighted mean coverage of placed contigs (29.81×). Coverage class

boundaries: Hemi_0.5× (0.375-0.625× reference = 11.18-18.63×), Auto_1.0× (0.75-1.25× reference = 22.36-37.26×), High_Coverage/Repeat (>37.26×), Low_Coverage/Other (all remaining), Zero (no mapped reads).

Set	Coverage class	Contigs	Total (Mb)
Placed	Auto_1.0×	5,975	1,343.3
	Hemi_0.5×	1,203	45.2
	High_Coverage/Repeat	666	62.8
	Low_Coverage/Other	971	25.2
Unplaced	Hemi_0.5×	6,302	66.5
	Auto_1.0×	4,310	87.2
	High_Coverage/Repeat	4,513	55.2
	Low_Coverage/Other	10,049	62.1
	Zero	117	0.2

At the contig level, repeat content was assessed separately for the placed and unplaced contig sets. WindowMasker masked 35.4% of placed and 69.0% of unplaced sequence (per-contig median 35.3% and 87.6%, respectively). RepeatMasker classified 40.0% of placed and 78.6% of unplaced sequence as repetitive (Table SR21). Among unplaced contigs stratified by coverage class, High_Coverage/Repeat contigs had 85.1% RepeatMasker-masked content, Hemi_0.5× contigs had 81.5%, Auto_1.0× contigs had 73.3%, and Low_Coverage/Other contigs had 77.2%. DNA transposons were the most abundant repeat class across all unplaced coverage categories (19.4-25.8%), followed by unknown repeats (22.1-26.1%) and LINEs (10.5-12.1%). Satellite content was 7.3% in High_Coverage/Repeat contigs, 3.9% in Hemi_0.5× contigs, and 3.1% in Auto_1.0× contigs. Among placed contigs, those within Z-candidate regions had 47.4% RepeatMasker-masked content compared to 39.7% in the autosomal remainder. DNA transposon content in Z-candidate placed contigs was 13.0% compared to 10.7% in autosomal placed contigs. LTR content was 4.3% in Z-candidate and 2.8% in autosomal placed contigs. LINE content was 6.1% in Z-candidate and 4.5% in autosomal placed contigs (Table SR21).

Table SR21. Repeat content of placed and unplaced contig sets. RepeatMasker v4.2.3 [29] with Dfam 3.9 [31] (taxon Anura, --uncurated, -pa 6). WindowMasker [30] with shared k-mer counts from the full pre-scaffolding primary assembly. RepeatMasker was run on the placed contig FASTA and the length-filtered (≥ 500 bp) unplaced contig FASTA (25,079 of 25,291 unplaced contigs). Values are percentage of total sequence per set. RM masked (%): total RepeatMasker-masked fraction. WM masked (%): total WindowMasker-masked fraction. SINE: short interspersed nuclear elements. LINE: long interspersed nuclear elements. LTR: long terminal repeat elements. DNA: DNA transposons. RC/Helitron: rolling-circle transposons. Satellite: satellite repeats. Simple/Low: simple repeats and low-complexity DNA. Unknown: unclassified interspersed repeats. Other: remaining RepeatMasker categories including small RNA and artifacts.

By contig set													
Set	Contigs	Total (bp)	RM masked (%)	WM masked (%)	SINE	LINE	LTR	DNA	RC/Helitron	Satellite	Simple/Low	Unknown	Other
Placed	8,815	1,476,572,988	40.00	35.4	0.08	4.57	2.83	10.76	0.22	0.72	1.66	18.47	0.69
Unplaced	25,079	271,000,222	78.63	69.0	0.25	11.23	8.68	21.97	1.22	4.35	6.03	23.44	1.46
Unplaced contigs by coverage class													
Coverage class	Contigs	Total (bp)	RM masked (%)		SINE	LINE	LTR	DNA	RC/Helitron	Satellite	Simple/Low	Unknown	Other
Hemi_0.5x	6,281	66,469,231	81.53		0.36	10.46	10.22	25.76	1.28	3.93	5.04	23.16	1.30
Auto_1.0x	4,300	87,164,797	73.30		0.17	11.09	8.32	19.97	1.33	3.12	4.62	22.91	1.76
High_Coverage/Repeat	4,496	55,151,868	85.08		0.16	12.06	7.83	19.44	0.97	7.30	9.72	26.14	1.48
Low_Coverage/Other	9,903	62,039,967	77.19		0.31	11.44	8.33	23.00	1.21	3.92	5.70	22.09	1.21
Zero	99	174,359	106.78		0.42	31.53	5.05	10.04	0.00	5.35	31.56	22.75	0.07
Placed contigs by scaffold zone													
Scaffold zone	Contigs	Total (bp)	RM masked (%)		SINE	LINE	LTR	DNA	RC/Helitron	Satellite	Simple/Low	Unknown	Other
Z-candidate	1,186	58,905,817	47.38		0.07	6.13	4.33	13.04	0.20	1.27	2.70	18.94	0.70
Autosomal	7,629	1,417,667,171	39.69		0.08	4.51	2.77	10.66	0.22	0.70	1.62	18.45	0.69

HANNO annotation of the pre-scaffolding primary assembly with seven Hyloidea proteomes yielded 30,256 gene models, of which 18,099 received a functional gene name via eggNOG-mapper. Post-hoc partitioning assigned 25,513 models (16,295 named) to the placed set and 4,743 models (1,804 named) to the unplaced set across 2,965 contigs. Of the placed genes, 22,631 were on Auto_1.0× contigs, 989 on Hemi_0.5× contigs, 1,262 on High_Coverage/Repeat contigs, and 631 on Low_Coverage/Other contigs. Of all placed genes, 910 mapped to the Z-candidate region on scaffold 2 and 341 to the Z-candidate region on scaffold 8. Of the unplaced genes, 1,879 were on Auto_1.0× contigs, 1,025 on Hemi_0.5× contigs, 911 on High_Coverage/Repeat contigs, 927 on Low_Coverage/Other contigs, and 1 on a zero-coverage contig. Miniprot identified 5,657 hits on 2,006 placed contigs and 1,758 hits on 688 unplaced contigs (Table SR22). The HANNO annotation of the scaffolded assembly (28,190 models, 16,912 named, of which 23,611 placed and 4,579 unplaced across 2,995 contigs) was used for the per-gene coverage classification reported above.

Table SR22. Gene content of placed and unplaced contig sets. HANNO v0.4 [16] annotation on the pre-scaffolding primary assembly with seven Hyloidea protein and mRNA evidence datasets. Functional gene names from eggNOG-mapper v2.1.13 [28] (Preferred_name field, converted to uppercase). Miniprot [27] with reviewed Anuran UniProt proteins (taxonomy ID 8292; accessed February 2026).

HANNO gene models by set and coverage class			
Set	Coverage class	Total models	Named models
Placed	Auto_1.0×	22,631	
	Hemi_0.5×	989	
	High_Coverage/Repeat	1,262	
	Low_Coverage/Other	631	
	Placed total	25,513	16,295
Unplaced	Auto_1.0×	1,879	
	Hemi_0.5×	1,025	
	High_Coverage/Repeat	911	
	Low_Coverage/Other	927	
	Zero	1	

	Unplaced total	4,743	1,804
All		30,256	18,099
Miniprot hits			
	Set	Hits	Contigs with hits
	Placed	5,657	2,006
	Unplaced	1,758	688

The tiered gametolog candidate classification assigned 48 genes to Tier 1, 9 to Tier 2, 18 to Tier 3, 11 to Tier 4, 46 to Tier 5, 530 to Tier 6, and 13,508 to Tier 7. The 48 Tier 1 genes resided on contig pairs sub-classified as Tier 1a (5 contig pairs carrying 13 Tier 1 genes), Tier 1b (29 contig pairs each carrying 1 Tier 1 gene), or Tier 1c (6 Tier 1 genes with ambiguous contig pairing) (Table SR23).

Table SR23. Gametolog candidate tier classification of named genes. Tiers 1-7 classify genes based on the distribution of hemizygous (Hemi_0.5×) hits across placed and unplaced contigs, using the HANNO v0.4 [16] annotation of the pre-scaffolding primary assembly. Sub-tiers 1a and 1b classify the contig pairs carrying Tier 1 genes. Sub-tier 1c classifies Tier 1 genes with ambiguous contig pairing. Z-candidate regions were defined by the half-coverage bin enrichment test and boundary segmentation. See Methods Additional file 2: “Identification of candidate sex chromosome-linked regions” for tier definitions.

Tier	Entity	Description	Count
1	Genes	1 Hemi Z-region placed + 1 Hemi unplaced, no paralogs	48
1a	Contig pairs	Contig pairs sharing ≥ 2 Tier 1 genes	5 (carrying 13 genes)
1b	Contig pairs	Contig pairs sharing 1 Tier 1 gene	29 (carrying 29 genes)
1c	Genes	Tier 1 genes with ambiguous contig pairing	6

2	Genes	1:1 pair, any scaffold	9
3	Genes	Z-region pair with paralogs	18
4	Genes	Non-Z pair with paralogs	11
5	Genes	Multiple hemizygous copies	46
6	Genes	Hemizygous on one side only	530
7	Genes	No hemizygous hits	13,508

The 5 Tier 1a contig pairs carried 13 Tier 1 genes. Four pairs mapped to scaffold 2, spanning scaffold coordinates 6.4 to 29.8 Mb: contig_54110 and contig_2556 (CDPF1, TTLL12, at 6.4 Mb), contig_4883 and contig_8781 (CYB5R3, REP15, at 28.7 Mb), contig_24880 and contig_15441 (FOXN1, NRIP2, RHNO1, at 29.5 Mb), and contig_37028 and contig_56404 (CCDC77, KDM5A, at 29.8 Mb). One pair mapped to scaffold 8 at 58.0 Mb: contig_16139 and contig_12320 (ACSM3, BRI3, ERI2, THUMP1) (Table SR24, Figure SR22, Figure SR23).

Table SR24. Tier 1a gametolog candidate contig pairs. Each pair consists of a placed (Z-candidate-linked) contig within a Z-candidate region and an unplaced (candidate W-linked) contig sharing ≥ 2 Tier 1 genes. Scaffold coordinates refer to the RagTag v2.1.0 [10]-scaffolded assembly. W coverage is the per-contig mean depth ($\times$) on the pre-scaffolding assembly. Gene names from eggNOG-mapper v2.1.13 [28] “Preferred_name field”. See Methods Additional file 2: “Identification of candidate sex chromosome-linked regions” for tier definitions.

Scaffold	Position (Mb)	Z contig	W contig	W cov. ($\times$)	Genes
scaffold_2	6.4	contig_54110	contig_2556	11.55	CDPF1, TTLL12

scaffold_2	28.7	contig_4883	contig_8781	15.10	CYB5R3, REP15
scaffold_2	29.5	contig_24880	contig_15441	13.91	FOX M1, NRIP2, RHNO1
scaffold_2	29.8	contig_37028	contig_56404	16.10	CCDC77, KDM5A
scaffold_8	58.0	contig_16139	contig_12320	17.53	ACSM3, BRI3, ERI2, THUMPD1

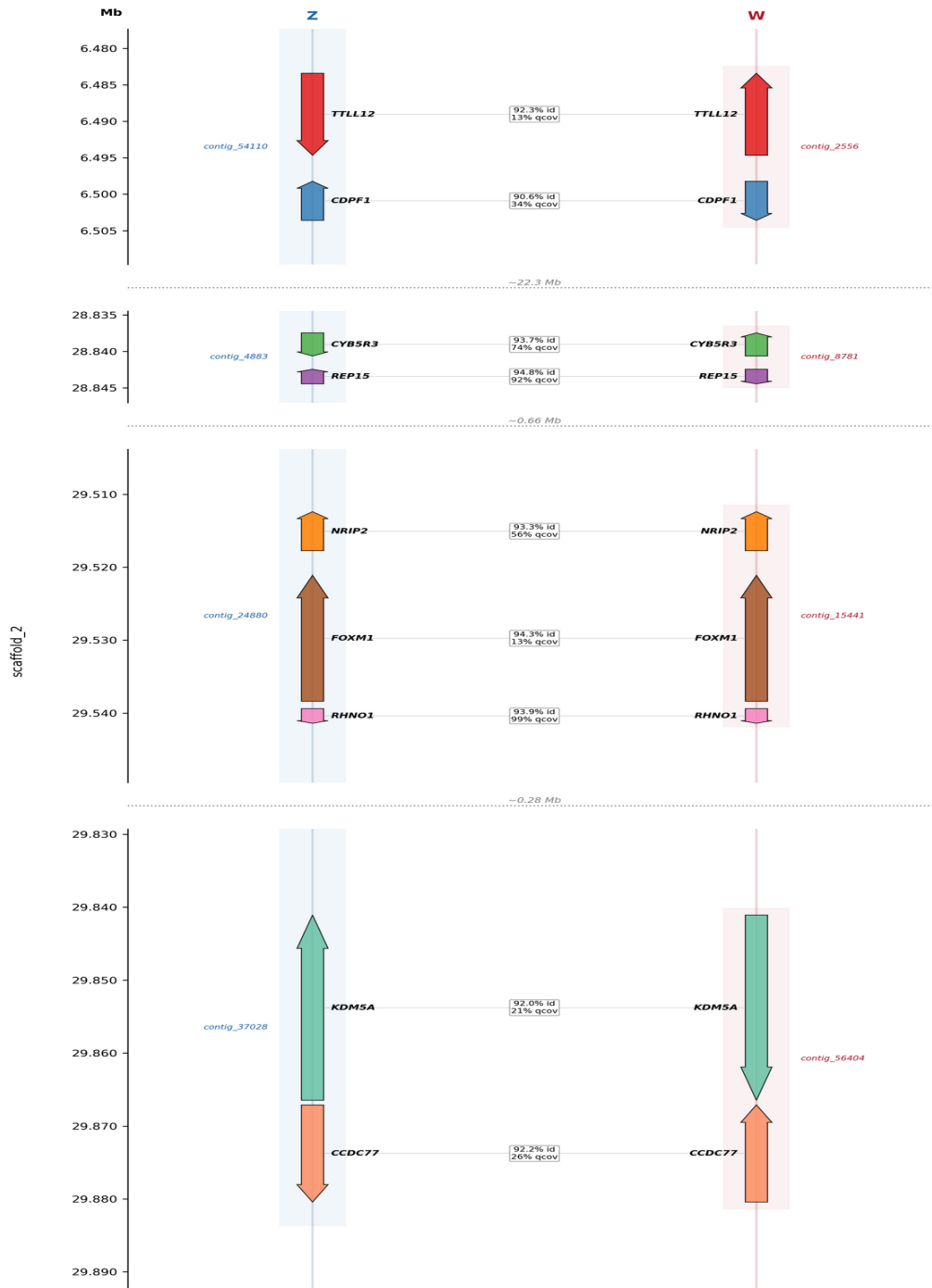


Figure SR22. Tier 1a Z-W gametolog candidate synteny on scaffold_2. Each panel shows one Z-W contig pair. Blue axis: Z-candidate-linked (placed) contig. Red axis: W-candidate-linked (unplaced) contig. colored arrows depict genes; arrow direction indicates transcriptional orientation. Gene names are italicized. gray lines connect Z and W copies; midpoint boxes show BLASTn [25] nucleotide

identity (% id) and query coverage (% qcov). Y-axis: scaffold_2 coordinates in megabases (Mb). Light blue and red shading marks contig extent. Contig identifiers are in italics beside each axis. Dashed lines between panels indicate gaps with approximate genomic distances.

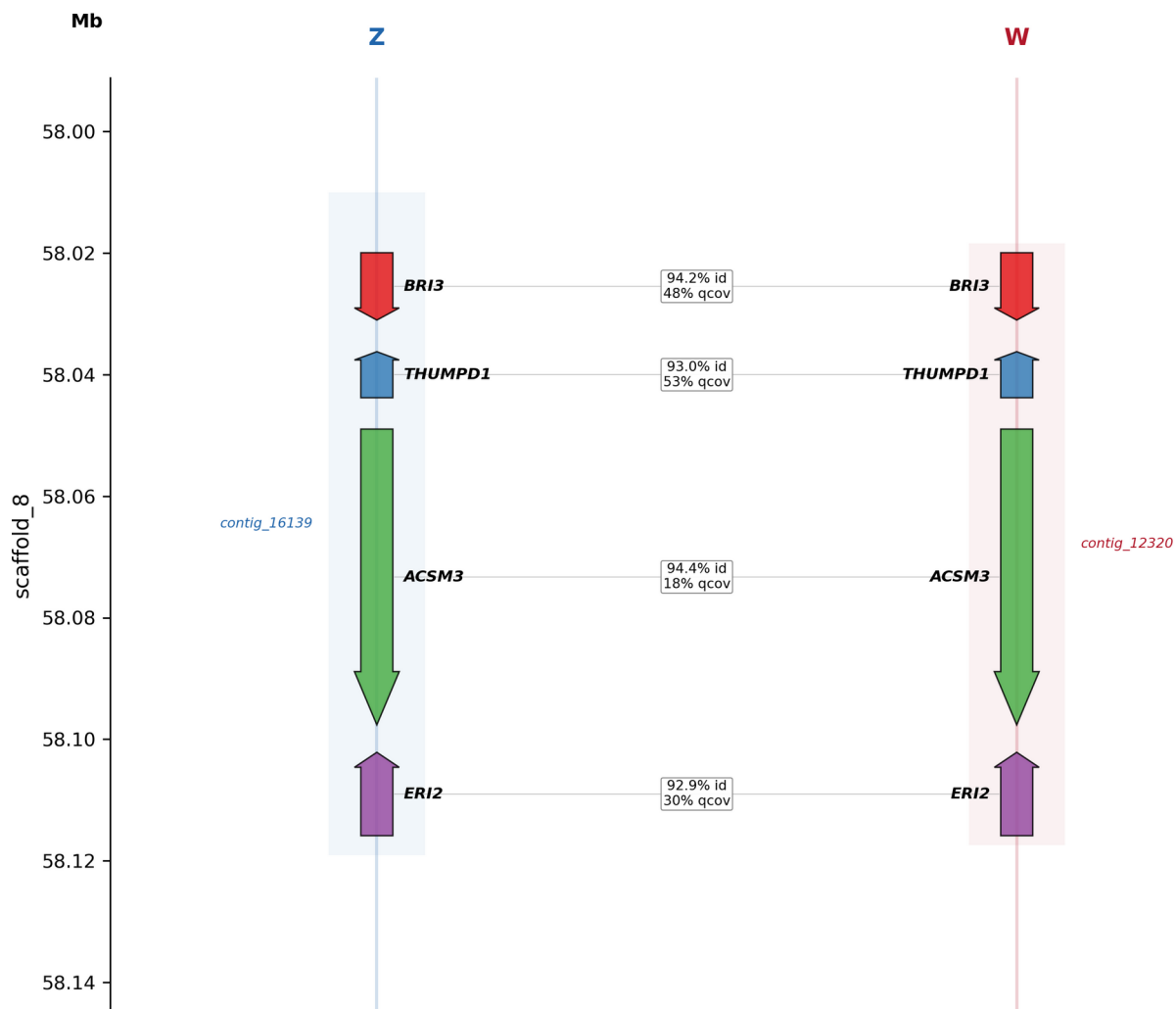


Figure SR23. Tier 1a Z-W gametolog candidate synteny on scaffold_8. Blue axis: Z-candidate-linked (placed) contig. Red axis: W-candidate-linked (unplaced) contig. colored arrows depict genes; arrow direction indicates transcriptional orientation. Gene names are italicized. gray lines connect Z and W copies; midpoint boxes show BLASTn [25] nucleotide identity (% id) and query coverage (%)

qcov). Y-axis: scaffold_8 coordinates in megabases (Mb). Light blue and red shading marks contig extent. Contig identifiers are in italics beside each axis.

BLASTn alignment of the 13 Tier 1 genes on Tier 1a contig pairs yielded alignments for all 13, with a mean nucleotide identity of 93.19% (median 93.27%, range 90.63-94.80%). Of the 29 Tier 1 genes on Tier 1b contig pairs, 27 produced alignments with a mean identity of 93.54% (median 93.21%, range 90.20-99.43%). IKBIP and TDRD1 produced no alignment. All 6 Tier 1c genes produced alignments with a mean identity of 93.93%. Across the 46 Tier 1 genes that produced alignments, the median nucleotide identity was 93.22% (Table SR25).

Table SR25. Pairwise alignment of gametolog candidates. BLASTn v2.14.1 [25] was used with default parameters. For each Tier 1 gene, the genomic region on the placed (Z) contig was used as query and the corresponding region on the unplaced (W) contig as subject. Genes producing no significant alignment are listed separately. See Methods Additional file 2: 'Identification of candidate sex chromosome-linked regions' for tier definitions.

Tier	Genes aligned	Mean identity (%)	Median identity (%)	Min (%)	Max (%)	No-hit genes
1a	13/13	93.19	93.27	90.63	94.80	
1b	27/29	93.54	93.21	90.20	99.43	IKBIP, TDRD1
1c	6/6	93.93	93.94	91.06	96.31	

Compositional banding of scaffold_2 revealed spatially structured repeat composition within the Z-candidate region (0-42 Mb; Figure SR24; Table SR26). Segment analysis identified an AT-rich repetitive zone spanning the first 6.9 Mb (48% AT-rich repetitive rendered pixels, 0% GC-rich repetitive), followed by a predominantly non-repetitive interval from 6.9 to 25.4 Mb (72-91% non-

repetitive), and a GC-rich repetitive block from 25.4 to 41.9 Mb (53-78% GC-rich repetitive). AT-rich repetitive sequence was exclusive to the Z-candidate region and absent from the remainder of scaffold_2. Beyond the Z-region boundary, the autosomal portion (47-170 Mb) was almost entirely non-repetitive (>93%), with GC-rich repetitive segments restricted to the scaffold termini (170-176 Mb, 81-94% GC-rich repetitive).



Figure SR24. Repeat and GC-content banding of scaffold_2. Vertical strip represents the full scaffold of 176 megabases (Mb). Banding is computed from GC content and WindowMasker [30] repeat density in 2 kb sliding windows (400 bp step). AT-rich repetitive (yellow-green), GC-rich repetitive (ochre), non-repetitive (light gray). Blue dashed line marks Z-candidate region boundary; blue bracket with "Z" label indicates the Z-candidate interval (0-42 Mb).

Table SR26. Compositional segments of scaffold_2 from pixel-based banding analysis. Segments identified on the rendered banding image (Figure SR24). GC content and WindowMasker [30] repeat density were computed in 2 kb sliding windows (400 bp step), averaged into 1,200 evenly spaced bins, and classified as AT-rich repetitive (GC < 50%, masked), GC-rich repetitive (GC ≥ 50%, masked), or non-repetitive (unmasked). Segments shorter than 1 Mb were merged into their longer neighbor. Pixel fractions denote the proportion of rendered bins within each segment assigned to each class.

Segment start (Mb)	Segment end (Mb)	Span (Mb)	Rendered class	Z/non- Z	AT-rep (%)	GC-rep (%)	Non-rep (%)
0.00	6.91	6.91	AT-rich repetitive	Z	47.8	0.0	52.2
6.91	9.55	2.64	Non-repetitive	Z	27.8	0.0	72.2
9.55	12.05	2.50	Non-repetitive	Z	11.8	0.0	88.2
12.05	14.99	2.94	Non-repetitive	Z	15.0	0.0	85.0
14.99	20.72	5.73	Non-repetitive	Z	5.1	30.8	64.1
20.72	25.42	4.70	Non-repetitive	Z	0.0	9.4	90.6
25.42	32.18	6.76	GC-rich repetitive	Z	0.0	78.3	21.7

32.18	41.88	9.70	GC-rich repetitive	Z	0.0	53.0	47.0
41.88	47.02	5.14	GC-rich repetitive	non-Z	0.0	74.3	25.7
47.02	170.01	122.99	Non-repetitive	non-Z	0.0	0.9-29.9	70.2-99.1
170.01	173.83	3.82	GC-rich repetitive	non-Z	0.0	80.8	19.2
173.83	176.32	2.50	GC-rich repetitive	non-Z	0.0	94.4	5.6

Scaffold_8 showed a different compositional pattern within its Z-candidate region (47-64 Mb; Figure SR25; Table SR27). The Z-region was predominantly non-repetitive (61-84%) with interspersed GC-rich repetitive windows (16-39%) but no AT-rich repetitive sequence. AT-rich repetitive sequence on scaffold_8 was confined to the first 2.6 Mb (0-2.6 Mb, non-Z-candidate-region), while a GC-rich repetitive block occupied the terminal 3.9 Mb (77-81 Mb, 69% GC-rich repetitive). The two Z-candidate regions thus differed in repeat composition: the scaffold_2 Z-region contained both AT-rich and GC-rich repetitive sequence with a clear transition from AT-rich repetitive (0-6.9 Mb) to GC-rich repetitive (25.4-41.9 Mb), whereas the scaffold_8 Z-region contained only GC-rich repetitive sequence.

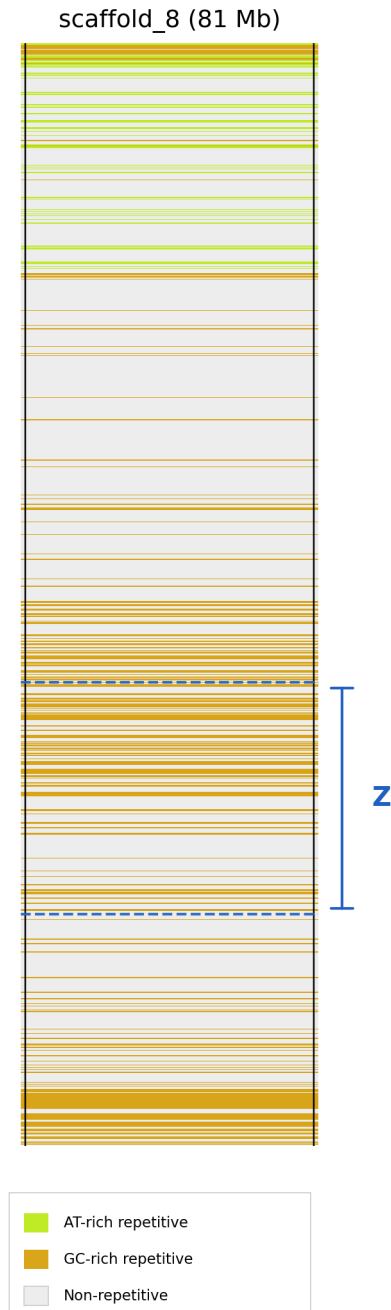


Figure SR25. Repeat and GC-content banding of scaffold_8. Vertical strip represents the full scaffold of 81 megabases (Mb). Banding is computed from GC content and WindowMasker [30] repeat density in 2 kb sliding windows (400 bp step). AT-rich repetitive (yellow-green), GC-rich repetitive (ochre), non-repetitive (light gray). Blue dashed lines mark Z-candidate region boundaries; blue bracket with "Z" label indicates the Z-candidate interval (47-64 Mb).

Table SR27. Compositional segments of scaffold_8 from pixel-based banding analysis. Segments identified on the rendered banding image (Figure SR25). GC content and WindowMasker [30] repeat density were computed in 2 kb sliding windows (400 bp step), averaged into 1,200 evenly spaced bins, and classified as AT-rich repetitive (GC < 50%, masked), GC-rich repetitive (GC ≥ 50%, masked), or non-repetitive (unmasked). Segments shorter than 1 Mb were merged into their longer neighbor. Pixel fractions denote the proportion of rendered bins within each segment assigned to each class.

Segment start (Mb)	Segment end (Mb)	Span (Mb)	Rendered class	Z/non- Z	AT-rep (%)	GC-rep (%)	Non-rep (%)
0.00	2.63	2.63	AT-rich repetitive	non-Z	34.2	31.6	34.2
2.63	7.77	5.13	Non-repetitive	non-Z	19.7	2.6	77.6
7.77	17.42	9.65	Non-repetitive	non-Z	9.7-14.9	0.0-6.6	78.7-85.7
17.42	39.97	22.55	Non-repetitive	non-Z	0.0	2.2-10.8	89.2-97.8
39.97	58.20	18.23	Non-repetitive	Z	0.0	38.9	61.1
58.20	64.48	6.28	Non-repetitive	Z	0.0	16.1	83.9
64.48	77.18	12.70	Non-repetitive	non-Z	0.0	3.6-22.7	77.3-96.4
77.18	81.03	3.85	GC-rich repetitive	non-Z	0.0	69.0	31.0

The hemizygous fraction varied across the compositional zones of both Z-candidate regions (Table SR28). On scaffold_2, the AT-rich repetitive zone (0-6.9 Mb) had the highest proportion of hemizygous genes (71.9%, 146/203) and the lowest proportion of autosomal-coverage genes (3.4%, 7/203). The non-repetitive zone (6.9-25.4 Mb) was intermediate (67.9% hemizygous, 9.7% autosomal). The GC-

rich repetitive zone (25.4-41.9 Mb) had the lowest hemizygous fraction (55.1%, 185/336) and the highest autosomal fraction (22.6%, 76/336). On scaffold_8, the 40-58 Mb Z-segment showed 66.4% hemizygous genes (156/235), whereas the 58-64 Mb segment contained only 26.4% hemizygous genes (28/106) and 61.3% autosomal-coverage genes (65/106).

Table SR28. Gene coverage classification by compositional zone within Z-candidate regions.

Genes from the HANNO v0.4 [16] BESTMODELS annotation of the scaffolded assembly were assigned to compositional zones defined by pixel-based segment detection (Tables SR26, SR27). Coverage classes were determined by mosdepth v0.3.11 [23] (--fast-mode) relative to the autosomal reference coverage of 29.12×. Adjacent Z-region segments of the same rendered class were combined into single zones for scaffold_2; the two non-repetitive Z-segments on scaffold_8 (40–58 Mb and 58–64 Mb) were retained separately due to different coverage profiles.

Scaffold	Zone (Mb)	Rendered class	Span (Mb)	Total genes	Hemi_0.5 × (%)	Auto_1.0 × (%)	High/Repeat (%)	Low/Other (%)
scaffold_2	0-6.9	AT-rich repetitive	6.9	203	146 (71.9)	7 (3.4)	10 (4.9)	40 (19.7)
scaffold_2	6.9-25.4	Non-repetitive	18.5	371	252 (67.9)	36 (9.7)	21 (5.7)	62 (16.7)
scaffold_2	25.4-41.9	GC-rich repetitive	16.5	336	185 (55.1)	76 (22.6)	22 (6.5)	53 (15.8)
scaffold_8	40-58	Non-repetitive	18.2	235	156 (66.4)	37 (15.7)	3 (1.3)	39 (16.6)
scaffold_8	58-64	Non-repetitive	6.3	106	28 (26.4)	65 (61.3)	1 (0.9)	12 (11.3)

Side-by-side banding of the five Tier 1a contig pairs showed higher repeat content in the W copies than in their Z counterparts (Table SR29). On scaffold_2, the four Z contigs ranged from 29.8% to 72.7% non-repetitive, while their W partners ranged from 48.1% to 61.4% non-repetitive (Figure SR26). The scaffold_8 pair (contig_16139 / contig_12320) showed 61.9% versus 55.7% non-repetitive for Z and W, respectively (Figure SR27).

Table SR29. Compositional banding summary of Tier 1a gametolog candidate contig pairs. Per-contig fractions of AT-rich repetitive, GC-rich repetitive, and non-repetitive windows were computed from GC content and WindowMasker [30] repeat density in 2 kb sliding windows (400 bp step). WindowMasker masking density threshold: 0.4. See Methods Additional file 2: “Identification of candidate sex chromosome-linked regions” for tier definitions.

Scaffold	Pair	Side	Contig	Length (bp)	GC (%)	WM density (%)	AT-rep (%)	GC-rep (%)	Non-rep (%)	Genes
scaffold_2	1	Z	contig_5 4110	75,473	45.7	40.9	29.3	13.6	57.1	TTLL12, CDPF1
scaffold_2	1	W	contig_2 556	23,370	47.2	45.2	31.5	20.4	48.1	TTLL12, CDPF1
scaffold_2	2	Z	contig_4 883	161,353	44.4	67.7	65.9	4.3	29.8	CYB5R3, REP15
scaffold_2	2	W	contig_8 781	94,665	44.9	49.2	43.1	6.5	50.4	CYB5R3, REP15
scaffold_2	3	Z	contig_2 4880	94,800	57.5	62.4	9.4	56.6	33.9	NRIP2, FOXMI, RHNO1

scaffold_2	3	W	contig_1 5441	42,181	49.5	40.5	38.6	0.0	61.4	NRIP2, FOXN1, RHNO1
scaffold_2	4	Z	contig_3 7028	75,012	50.3	31.8	15.3	12.0	72.7	KDM5A, CCDC77
scaffold_2	4	W	contig_5 6404	103,912	47.6	53.3	31.8	20.0	48.2	KDM5A, CCDC77
scaffold_8	1	Z	contig_1 6139	109,112	44.9	41.2	26.1	11.9	61.9	BRI3, THUMPD1, ACSM3, ERI2
scaffold_8	1	W	contig_1 2320	85,988	45.7	44.3	35.2	9.0	55.7	BRI3, THUMPD1, ACSM3, ERI2

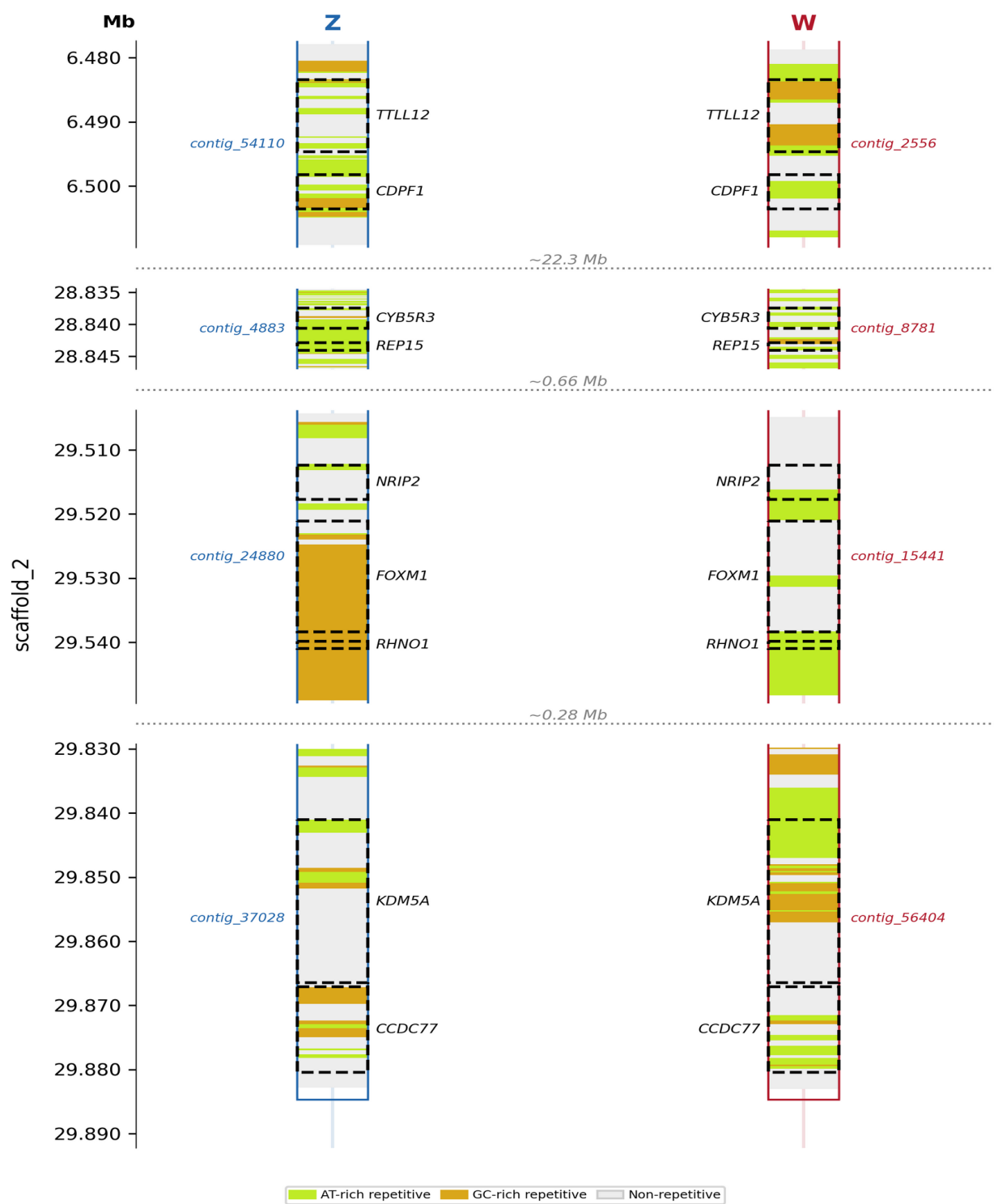


Figure SR26. Compositional banding of Tier 1a gametolog candidate contig pairs on scaffold_2. Vertical strips show the Z-candidate contig (left, blue outline) and its candidate W-genic counterpart

(right, red outline) for each of four Tier 1a contig pairs. Strip fill represents three categories of sequence composition computed from GC content and WindowMasker [30] repeat density in 2 kb sliding windows (400 bp step): yellow-green, AT-rich repetitive; amber/gold, GC-rich repetitive; light gray, non-repetitive. Dashed black rectangles indicate annotated gene positions spanning full gene length. Gene names are italicized. Y-axis: scaffold coordinate in megabases (Mb). W-side contigs are unplaced and drawn at the corresponding Z-side scaffold position. Contig names are shown to the left (Z) and right (W) of each strip. Dotted lines between panels indicate scaffold span between displayed contigs.

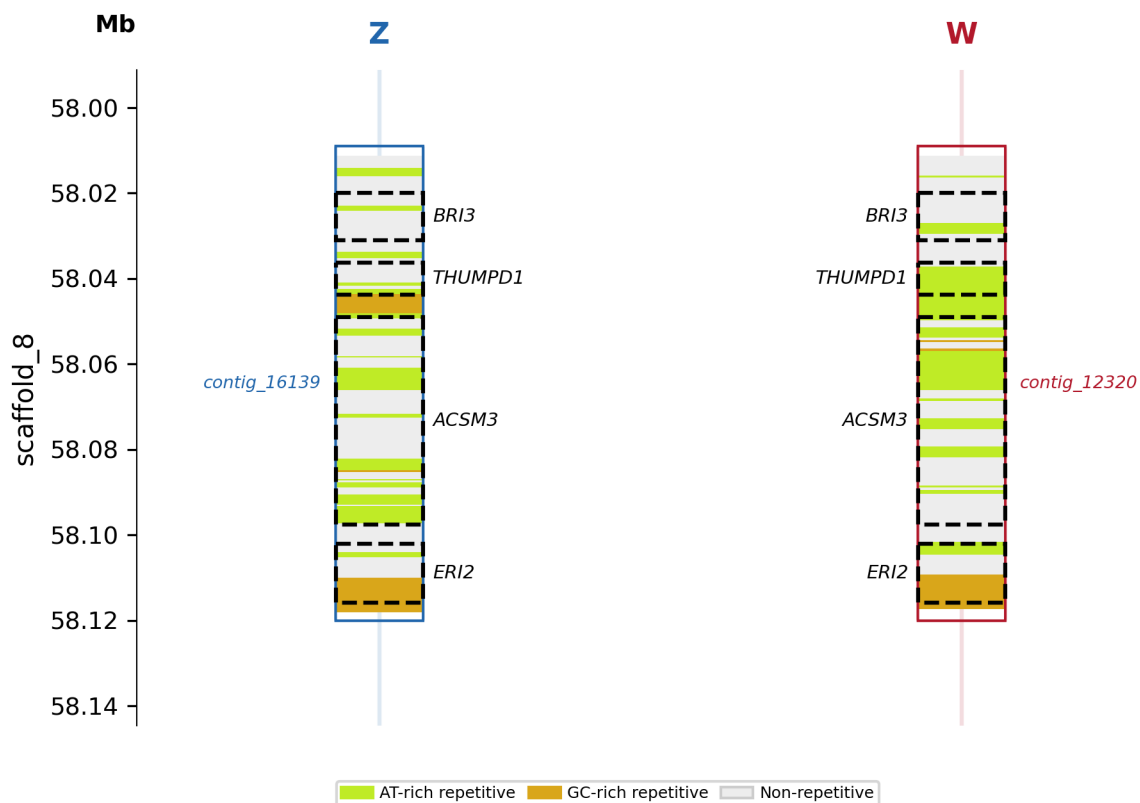


Figure SR27. Compositional banding of Tier 1a gametolog candidate contig pairs on scaffold_8.

Vertical strips show the Z-candidate contig (left, blue outline) and its candidate W-genic counterpart (right, red outline) for one Tier 1a contig pair. Strip fill represents three categories of sequence composition computed from GC content and WindowMasker [30] repeat density in 2 kb sliding

windows (400 bp step): yellow-green, AT-rich repetitive; amber/gold, GC-rich repetitive; light gray, non-repetitive. Dashed black rectangles indicate annotated gene positions spanning full gene length. Gene names are italicized. Y-axis: scaffold coordinate in megabases (Mb). The W-side contig is unplaced and drawn at the corresponding Z-side scaffold position. Contig names are shown to the left (Z) and right (W) of the strip.

Composite banding of W-genic contigs showed predominantly non-repetitive composition across all three tiers (Figure SR28; Table SR30). Tier 1a contigs (5 contigs, 350 kb) were 52.2% non-repetitive (36.5% AT-rich, 11.3% GC-rich repetitive). Tier 1b contigs (6 contigs, 446 kb) were 67.1% non-repetitive (26.7% AT-rich, 6.2% GC-rich). Tier 1c contigs (6 contigs, 152 kb) were 50.7% non-repetitive (30.2% AT-rich, 19.1% GC-rich).

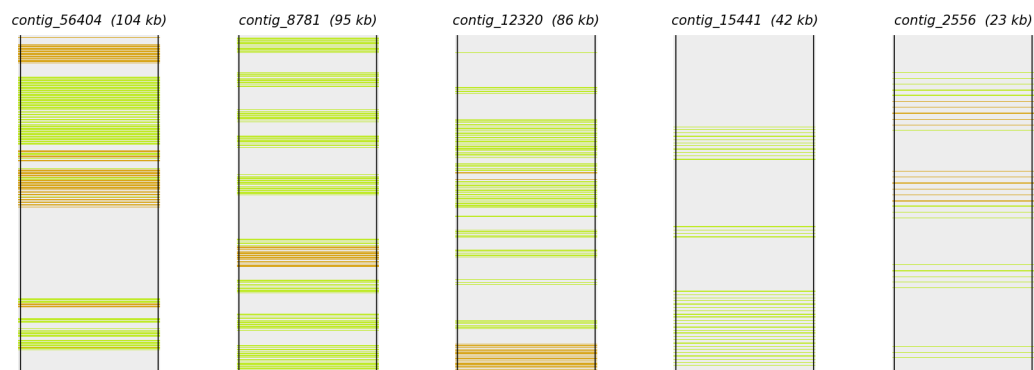
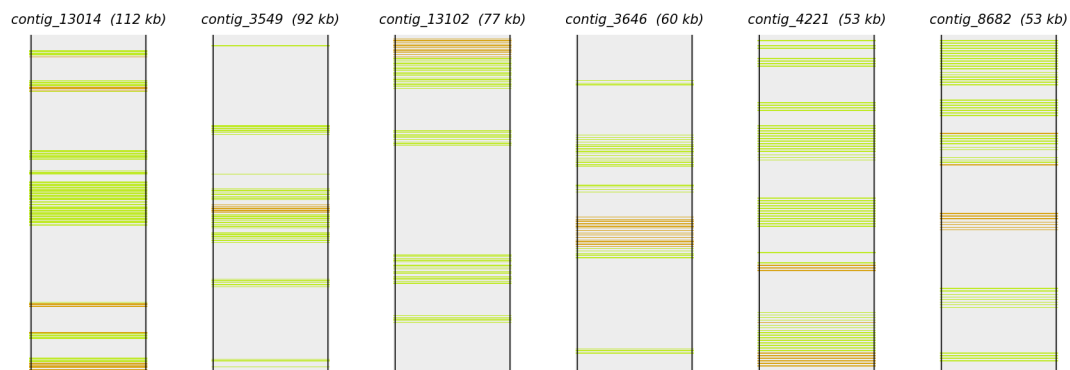
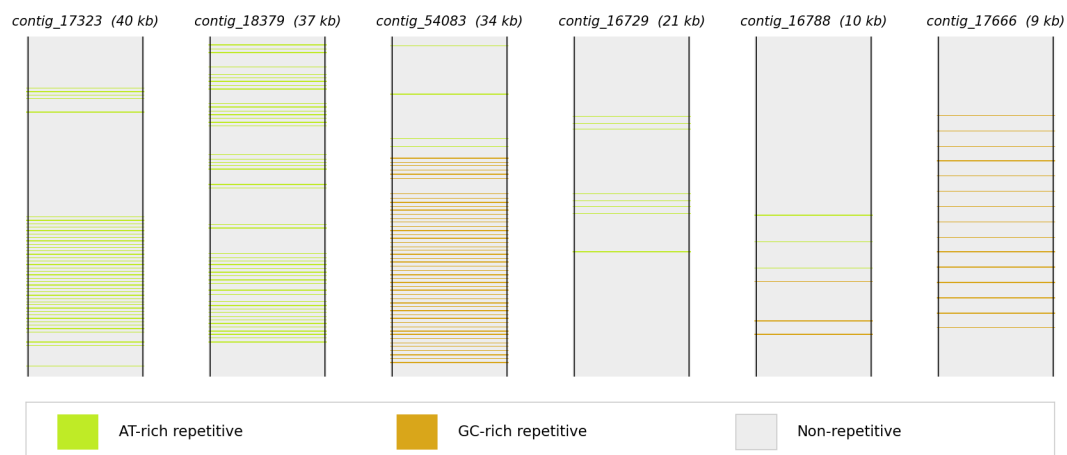
A**B****C**

Figure SR28. Composite banding of W-genic contigs. Three panels show unplaced (W-candidate-linked) contigs from Tier 1 gametolog candidate pairs. (A) Unplaced contigs from Tier 1a contig pairs. (B) Six largest unplaced contigs from Tier 1b contig pairs. (C) Contigs carrying Tier 1c genes. Each vertical strip represents one contig; contig name and length (kb) are given above. Banding is computed from GC content and WindowMasker [30] repeat density in 2 kb sliding windows (400 bp step). AT-rich repetitive (yellow-green), GC-rich repetitive (ochre), non-repetitive (light gray). Contigs are ordered by length within each panel.

Table SR30. Compositional banding summary of W-genic contigs by tier. Per-contig fractions of AT-rich repetitive, GC-rich repetitive, and non-repetitive windows were computed from GC content and WindowMasker [30] repeat density in 2 kb sliding windows (400 bp step). WindowMasker masking density threshold: 0.4. Tier 1b shows the six largest contigs only. See Methods Additional file 2: “Identification of candidate sex chromosome-linked regions” for tier definitions.

Tier	Contig	Length (bp)	GC (%)	WM density (%)	AT-rep (%)	GC-rep (%)	Non-rep (%)
tier1a	contig_56404	103,912	47.6	53.3	31.8	20.0	48.2
tier1a	contig_8781	94,665	44.9	49.2	43.1	6.5	50.4
tier1a	contig_12320	85,988	45.7	44.3	35.2	9.0	55.7
tier1a	contig_15441	42,181	49.5	40.5	38.6	0.0	61.4
tier1a	contig_2556	23,370	47.2	45.2	31.5	20.4	48.1
tier1a total	5 contigs	350,116	46.6	47.9	36.5	11.3	52.2
tier1b	contig_13014	112,076	46.5	38.8	25.4	6.2	68.5
tier1b	contig_3549	91,949	43.1	30.4	19.6	2.7	77.8

tier1b	contig_13102	76,819	44.6	35.1	25.0	5.9	69.2
tier1b	contig_3646	60,011	43.2	40.1	20.5	9.6	69.9
tier1b	contig_4221	52,757	45.5	46.7	42.5	7.9	49.6
tier1b	contig_8682	52,720	49.1	41.1	36.2	7.1	56.7
tier1b total	6 contigs	446,332	45.2	37.8	26.7	6.2	67.1
tier1c	contig_17323	40,255	43.3	48.9	44.8	0.0	55.2
tier1c	contig_18379	37,198	42.8	43.0	54.5	0.0	45.5
tier1c	contig_54083	33,988	53.4	68.6	5.0	61.3	33.8
tier1c	contig_16729	21,104	43.1	28.0	16.7	0.0	83.3
tier1c	contig_16788	10,274	43.8	36.0	14.3	14.3	71.4
tier1c	contig_17666	8,977	65.7	79.6	0.0	83.3	16.7
tier1c total	6 contigs	151,796	46.6	49.9	30.2	19.1	50.7

The 60 largest W-heterochromatin contigs (5.5 Mb total) were 93.1% repetitive (44.3% AT-rich, 48.8% GC-rich; mean WindowMasker density 91.0%; Figure SR29; Table SR31), with 11 contigs near-exclusively AT-rich repetitive (>80% AT-rich, 0% GC-rich), 13 near-exclusively GC-rich repetitive (>80% GC-rich, 0% AT-rich), and 36 showing mixed profiles. Only 8 of 60 contigs contained more than 20% non-repetitive sequence.

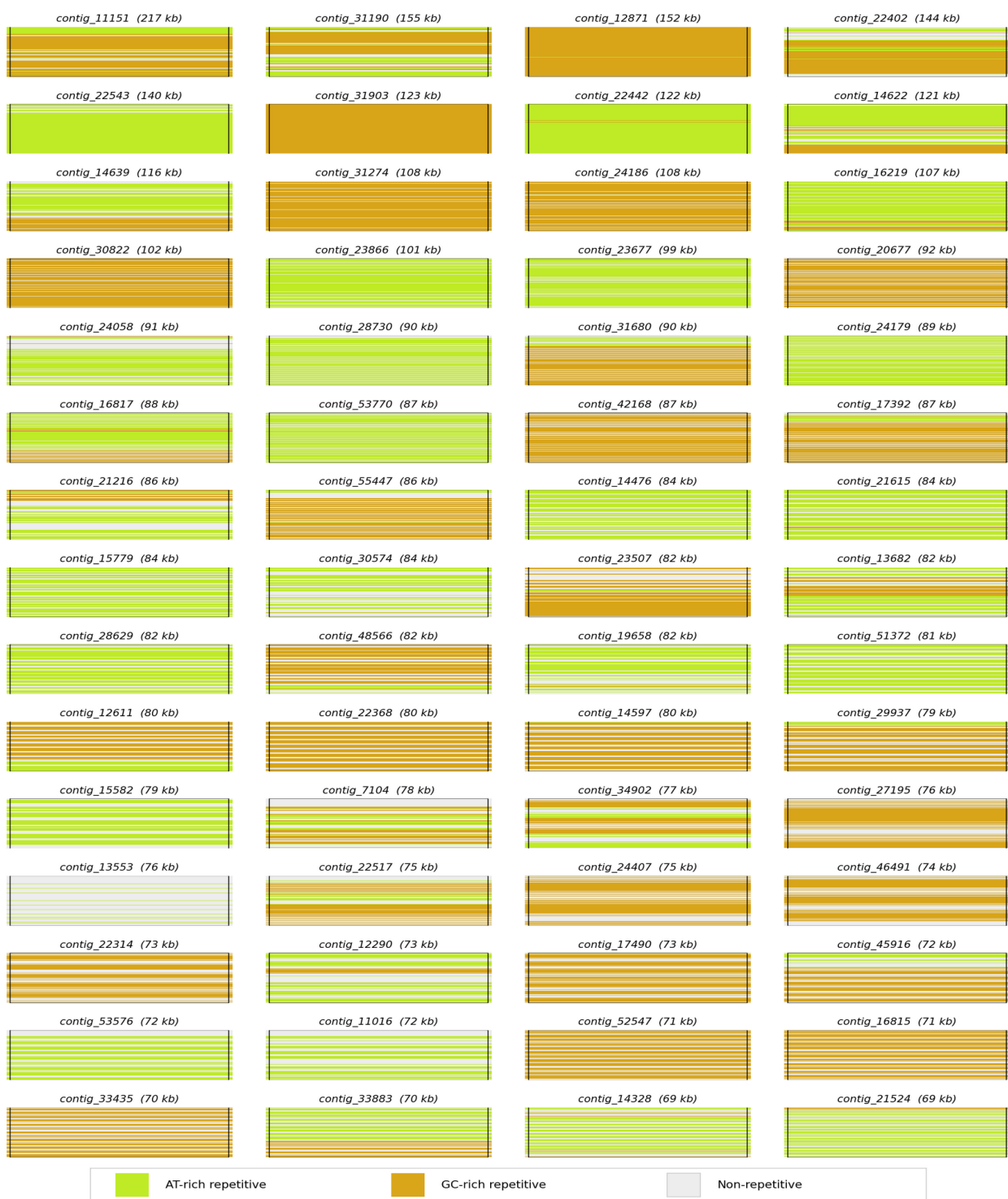


Figure SR29. Composite banding of the 60 largest W-heterochromatin contigs. Each vertical strip represents one contig classified as W-heterochromatin (repeat-strong); contig name and length (kb) are

given above. Banding is computed from GC content and WindowMasker [30] repeat density in 2 kb sliding windows (400 bp step). AT-rich repetitive (yellow-green), GC-rich repetitive (ochre), non-repetitive (light gray). Contigs are ordered by length, left to right, top to bottom.

Table SR31. Compositional banding summary of the 60 largest W-heterochromatin (repeat-strong) contigs. Per-contig fractions of AT-rich repetitive, GC-rich repetitive, and non-repetitive windows were computed from GC content and WindowMasker [30] repeat density in 2 kb sliding windows (400 bp step). WindowMasker masking density threshold: 0.4. Contigs ordered by length.

Contig	Length (bp)	GC (%)	WM density (%)	AT-rep (%)	GC-rep (%)	Non-rep (%)
contig_11151	217,085	51.7	85.1	21.8	67.5	10.8
contig_31190	154,953	52.6	69.9	28.5	48.8	22.7
contig_12871	151,731	51.8	99.5	1.9	98.1	0.0
contig_22402	143,607	48.2	80.0	18.3	61.4	20.3
contig_22543	140,150	39.4	88.5	91.3	0.0	8.7
contig_31903	123,032	56.2	100.0	0.0	100.0	0.0
contig_22442	121,869	42.3	99.5	97.3	2.7	0.0
contig_14622	121,040	46.5	78.5	61.1	21.8	17.1
contig_14639	116,006	46.7	70.2	47.2	29.7	23.1
contig_31274	107,755	56.4	100.0	0.0	100.0	0.0
contig_24186	107,682	55.6	99.1	3.0	97.0	0.0
contig_16219	107,426	43.7	97.1	87.5	12.5	0.0

contig_30822	102,423	55.1	100.0	0.0	100.0	0.0
contig_23866	101,207	46.3	98.6	98.8	0.0	1.2
contig_23677	98,505	38.1	99.9	100.0	0.0	0.0
contig_20677	91,761	56.3	99.8	0.0	100.0	0.0
contig_24058	91,067	43.5	72.4	65.9	8.5	25.6
contig_28730	90,487	46.1	96.2	96.4	0.0	3.6
contig_31680	89,805	52.6	90.3	14.5	79.5	5.9
contig_24179	89,041	35.2	99.9	100.0	0.0	0.0
contig_16817	88,340	53.9	98.0	70.4	29.6	0.0
contig_53770	87,198	39.6	99.7	100.0	0.0	0.0
contig_42168	87,075	54.8	99.8	0.0	100.0	0.0
contig_17392	86,985	50.6	90.6	12.2	82.6	5.2
contig_21216	85,648	42.8	60.6	42.9	21.0	36.2
contig_55447	85,512	52.7	85.4	8.1	81.3	10.5
contig_14476	84,419	48.4	100.0	100.0	0.0	0.0
contig_21615	84,387	39.6	99.7	96.1	3.9	0.0
contig_15779	83,911	43.2	99.3	100.0	0.0	0.0
contig_30574	83,693	40.1	64.0	68.8	1.5	29.8
contig_23507	82,083	57.0	81.0	4.5	81.6	13.9

contig_13682	82,020	46.9	93.4	63.2	32.8	4.0
contig_28629	81,721	46.5	99.9	100.0	0.0	0.0
contig_48566	81,605	52.5	88.9	12.0	81.5	6.5
contig_19658	81,572	44.5	86.7	83.9	3.5	12.6
contig_51372	80,768	30.6	99.8	100.0	0.0	0.0
contig_12611	80,210	53.4	97.7	23.0	76.0	1.0
contig_22368	79,925	57.1	99.2	0.0	100.0	0.0
contig_14597	79,683	51.1	98.6	0.5	99.5	0.0
contig_29937	79,219	54.3	98.7	6.2	93.8	0.0
contig_15582	78,535	39.5	81.5	89.6	0.5	9.9
contig_7104	78,450	49.7	73.6	35.9	41.7	22.4
contig_34902	77,074	47.0	91.1	40.4	53.7	5.9
contig_27195	76,201	56.5	100.0	0.0	100.0	0.0
contig_13553	75,574	47.8	97.4	96.7	1.1	2.2
contig_22517	75,218	54.0	81.4	25.0	62.5	12.5
contig_24407	74,662	56.9	100.0	0.0	100.0	0.0
contig_46491	74,349	51.2	100.0	0.0	100.0	0.0
contig_22314	73,455	55.4	99.9	0.0	100.0	0.0
contig_12290	72,855	43.6	85.8	70.8	16.9	12.4

contig_17490	72,812	53.3	98.9	1.1	98.9	0.0
contig_45916	71,982	50.2	83.3	17.7	70.3	12.0
contig_53576	71,706	46.5	91.6	90.3	0.0	9.7
contig_11016	71,616	41.3	60.8	62.3	2.3	35.4
contig_52547	70,723	55.5	99.9	0.0	100.0	0.0
contig_16815	70,627	52.6	99.6	0.0	100.0	0.0
contig_33435	70,419	62.9	99.6	0.0	100.0	0.0
contig_33883	70,053	53.0	96.9	67.8	32.2	0.0
contig_14328	69,218	46.4	87.9	82.2	10.1	7.7
contig_21524	68,955	44.0	89.4	94.0	4.8	1.2
TOTAL	5,497,090	48.8	91.0	44.3	48.8	6.9

The multi-evidence master contig table classified 815 unplaced contigs (18.4 Mb) as W-genic (hemizygous coverage with at least one annotated gene), 1,310 (30.1 Mb) as W-genic-weak (gene-bearing contigs outside both autosomal and hemizygous coverage ranges), 14,582 (105.3 Mb) as W-heterochromatin (no annotated genes and repeat content $\geq 60\%$), 4,310 (87.2 Mb) as autosomal-unscaffolded (autosomal-level coverage), 117 (0.2 Mb) as zero-coverage, and 4,157 (29.9 Mb) as unclassified. In total, 16,707 unplaced contigs (153.9 Mb) were flagged as W-candidate-linked (Table SR32).

Table SR32. W-candidate classification of unplaced contigs. Each unplaced contig was classified based on its coverage class, gene content (HANNO v0.4 [16] and miniprot [27]), and repeat content (WindowMasker [30] masked fraction ≥ 0.6 or RepeatMasker v4.2.3 [29] total masked $\geq 60\%$). See Methods Additional file 2: 'Identification of candidate sex chromosome-linked regions' for classification criteria.

Category	Contigs	Total (Mb)
W-genic	815	18.4
W-genic-weak	1,310	30.1
W-heterochromatin	14,582	105.3
W-flagged total	16,707	153.9
autosomal-unscaffolded	4,310	87.2
zero-coverage	117	0.2
unclassified	4,157	29.9
Unplaced total	25,291	271.1

All six Doublesex and Mab-3 Related Transcription factor (DMRT) family genes identified in the HANNO BESTMODELS annotation were located on autosomal scaffolds at diploid coverage (Table SR33). DMRT1, DMRT3, and DMRT2 formed a tandem cluster at 106.05-106.26 Mb on scaffold_5, with DMRTA1 at 113.87 Mb on the same scaffold. DMRTA2 and DMRTB1 were located on scaffold_3 at 119.86 Mb and 120.95 Mb, respectively. None were located within the Z-candidate regions on scaffolds 2 or 8. A tBLASTn [25] search using the *Xenopus laevis* Dm-W [32] DM domain (Universal Protein Resource (UniProt) [33] accession Q1L8R5, residues 20-86) and the *P. euphronides* DMRT1

protein as queries against the full pre-scaffolding assembly returned no hits on any unplaced contig at e-value thresholds of 0.01 and 10. All hits were confined to the DMRT loci on scaffolds 3 and 5.

Table SR33. Doublesex and Mab-3 Related Transcription factor (DMRT) family genes in the *Pristimantis euphronides* genome. Scaffold coordinates from the RagTag v2.1.0 [10] scaffolded assembly. Coverage class and mean depth from mosdepth v0.3.11 [23] relative to the autosomal reference (29.09×). tBLASTn [25] search of the DM domain (*Xenopus laevis* Dm-W [32], Universal Protein Resource (UniProt) [33] accession Q1L8R5, residues 20-86) against the full pre-scaffolding primary assembly (e-value 10) returned hits only at these loci and none on unplaced contigs. Cluster: DMRT1, DMRT3, and DMRT2 tandem cluster at 106.05-106.26 Mb on scaffold_5.

Gene	Scaffold	Position (Mb)	Mean depth (×)	Coverage class	Context
<i>DMRT1</i>	scaffold_5	106.13-106.18	26.98	Auto_1.0×	Cluster
<i>DMRT3</i>	scaffold_5	106.19-106.19	23.95	Auto_1.0×	Cluster
<i>DMRT2</i>	scaffold_5	106.25-106.26	30.32	Auto_1.0×	Cluster
<i>DMRTA1</i>	scaffold_5	113.87-113.87	35.86	Auto_1.0×	
<i>DMRTA2</i>	scaffold_3	119.86-119.87	34.05	Auto_1.0×	
<i>DMRTB1</i>	scaffold_3	120.95-120.96	28.81	Auto_1.0×	

To identify genes on candidate W-linked contigs with no counterpart on placed scaffolds, named genes on W-genic and W-genic-weak contigs were cross-referenced against all named genes on placed scaffolds (Table SR34). In the W-genic set, 234 of 315 named genes matched a placed gene by name. Of 478 genes sent to BLASTn [25] (e-value 1e-5) against a database of placed scaffold sequences, 375 produced hits. The remaining 103 genes (10 named, 93 unnamed) had no placed counterpart. In the W-genic-weak set, 254 of 451 named genes matched a placed gene by name. Of 1,034 genes sent to BLASTn, 814 produced hits. The remaining 220 genes (31 named, 189 unnamed) had no placed counterpart. Among the 41 named genes with no placed counterpart across both sets, none had a recognized function in sex determination or sex differentiation. The 282 unnamed genes with no placed counterpart were screened for protein domains associated with sex determination or differentiation using the Protein families database (Pfam) [34] annotations from the HANNO BESTMODELS output. Four carried Pfam domains (two DDE_Tnp_4 transposase domains, one 7tm_4 olfactory receptor

domain, one PAE domain). None carried DM, HMG-box, Forkhead, steroid hormone receptor, cytochrome P450, Tudor, Piwi, or Homeobox domains. BLASTn of all 282 sequences against the NCBI Nucleotide database [35] identified hits for 34 genes, all to housekeeping or structural genes in other anuran species. No sex-related gene was identified. The remaining 248 genes produced no significant alignment to any sequence in the database.

Table SR34. Identification of genes on candidate W-linked contigs with no counterpart on placed scaffolds. W-genic contigs: unplaced contigs classified as Hemi_0.5× (0.375-0.625× autosomal reference coverage) carrying at least one annotated gene. W-genic-weak contigs: unplaced contigs carrying at least one gene but not classified as either Hemi_0.5× or Auto_1.0×. Named genes on these contigs were cross-referenced by gene name against all named genes on placed scaffolds (HANNO v0.4 [16] annotation). Genes with a name match were classified as gametologs. Unmatched named genes and all unnamed genes were queried by BLASTn v2.14.1 [25] (e-value 1e-5) against a database of placed scaffold sequences only. Genes with no placed counterpart: no name match and no significant BLASTn alignment to any placed scaffold sequence. Pfam [34] screening used domain annotations from the HANNO BESTMODELS output. Sex-related domains screened: DM, HMG-box, Forkhead, steroid hormone receptor, cytochrome P450, Tudor, Piwi, and Homeobox. NCBI nt BLASTn: v2.14.1, e-value 1e-5, max_target_seqs 3, against a local copy of the NCBI Nucleotide database [35].

Step	W-genic	W-genic-weak
Name-based cross-reference		
Named genes	315	451
Name match on placed (gametologs)	234	254
No name match (named)	81	197
Unnamed genes	381	794
BLASTn vs placed scaffolds		
Sent to BLASTn	478	1,034
BLASTn hit (diverged gametologs)	375	814
No placed counterpart	103	220
named	10	31
unnamed	93	189
Combined 282 unnamed genes with no placed counterpart		

With Pfam domain (none sex-related)	4	
NCBI nt hit (none sex-related)	34	
No hit in NCBI nt	248	

Phylogenetic analysis using nuclear gene orthologs

Of the 19 nuclear genes targeted [21], 18 were identified and extracted in CDS format from the HANNO annotation of the scaffolded *P. euphronides* genome (Workflow 1, see “Annotation of scaffolded genome assembly”): CXCR4, RAG1, TYR, SLC8A1 (NCX1), SLC8A3, BDNF, BMP2, MYC (CMYC), CRYBA1 (CRYBA), MC1R, MYH6 (MYH), NTF3 (NT3), POMC, RAG2, RHO, SIAH1 (SIA), TNS3, and ZFX. The 19th gene, H3A (Histone H3A), was not present in the HANNO annotation.

Homologous sequences for all 18 genes were retrieved from NCBI GenBank across 20 Hyloidea families (Additional file 10: Table). GenBank sequence availability varied widely, ranging from 57 sequences across six families for MC1R to 5,424 sequences across 19 families for RAG1. All 18 genes were aligned and subjected to quality filtering. During terminal gap trimming, the *P. euphronides* sequences for MYC, RAG1, and SLC8A1 were reduced entirely to alignment gaps, indicating that the *P. euphronides* CDS did not overlap positionally with the majority of the shorter GenBank sequences for these genes; these three genes were excluded despite high sequence availability in GenBank (202, 5,424, and 485 sequences, respectively). During subsequent manual curation in Jalview v2.11.4.1 [36], three additional genes were excluded: CRYBA1 showed excessive internal fragmentation of the *P. euphronides* sequence (281 internal alignment gaps); MYH6 was represented by a single genus (*Eleutherodactylus*, three species) in GenBank, precluding multi-family phylogenetic inference; and ZFX showed poor positional overlap between the *P. euphronides* CDS (30.3% of alignment length) and the Hyloidea sequences. This left 12 genes with sufficiently complete and gap-free alignments across multiple Hyloidea families (Additional file 10: Table): BDNF, BMP2, CXCR4, MC1R, NTF3, POMC, RAG2, RHO, SIAH1, SLC8A3, TNS3, and TYR.

Of the 12 retained genes, TYR was the only nuclear gene for which a *P. euphronides* sequence was also available in GenBank (EF493489.1) prior to this study. The remaining 11 orthologs were obtained from the newly assembled scaffolded and annotated *P. euphronides* genome.

Based on sequence availability across families, two concatenated datasets were constructed. A 5-gene supermatrix (CXCR4, POMC, SIAH1, SLC8A3, TYR) included 16 species from nine Hyloidea families: *Ischnocnema guentheri* (Brachycephalidae), *Bufo bufo* and *Bufo gargarizans* (Bufonidae), *Ceratophrys ornata* and *Lepidobatrachus laevis* (Ceratophryidae), *Ceuthomantis smaragdinus* and *Microkayla wettsteini* (Craugastoridae), *R. imitator* (Dendrobatidae), *Adelophryne gutturosa* and *Diasporus diastema* (Eleutherodactylidae), *D. ebraccatus* and *Hyla sarda* (Hylidae), *E. pustulosus* (Leptodactylidae), *Niceforonia brunnea*, *Phrynopus bracki* and *P. euphronides* (Strabomantidae), plus *Rana temporaria* (Neobatrachia, Ranoidea, Ranidae) as outgroup. A 12-gene supermatrix comprising all retained loci included species for which all 12 genes were available, resulting in six species from five Hyloidea families: *B. bufo* (Bufonidae), *R. imitator* (Dendrobatidae), *D. ebraccatus* and *H. sarda* (Hylidae), *E. pustulosus* (Leptodactylidae), and *P. euphronides* (Strabomantidae), plus *R. temporaria* (Ranidae) as outgroup.

Maximum-likelihood trees were inferred using IQ-TREE v2.2.0.3 [37] and visualized using iTOL v6 [38] for both concatenated datasets and single-gene alignments. The 5-gene supermatrix analysis (Figure SR30) recovered the direct-developing frog families (Terrarana: Strabomantidae, Craugastoridae, Eleutherodactylidae, Brachycephalidae) as a clade (bootstrap support 68.4%) distinct from other hyloid families (Hylidae, Bufonidae, Dendrobatidae, Leptodactylidae, Ceratophryidae). Within the Terrarana group, *P. euphronides* was placed within a group (bootstrap support 89.6%) including two other Strabomantidae species (*P. bracki* and *N. brunnea*), the Craugastoridae species *M. wettsteini*, and the Brachycephalidae member *I. guentheri*.

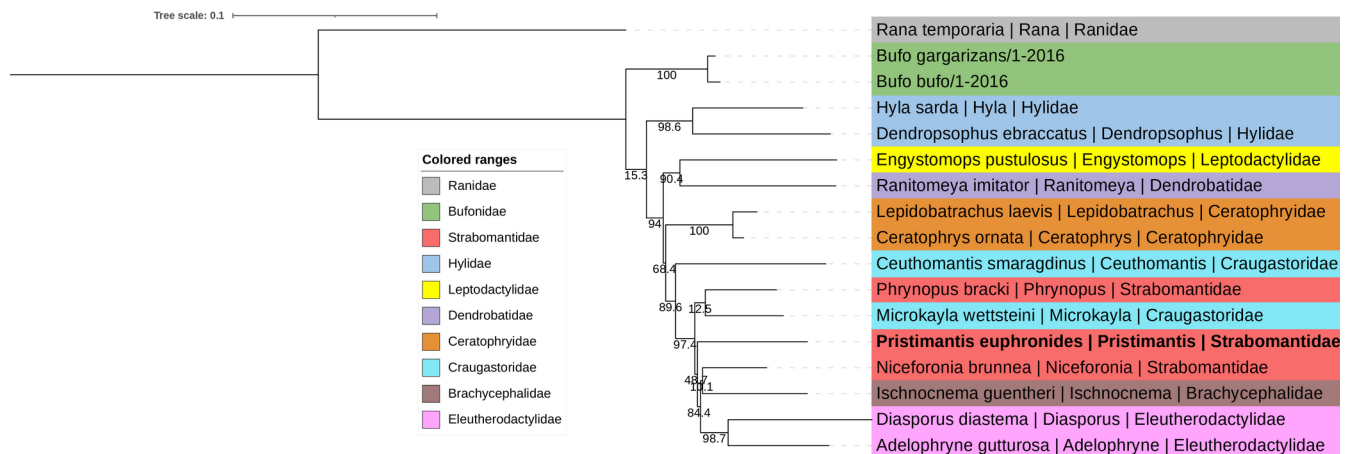


Figure SR30. Five-gene supermatrix phylogeny of *Pristimantis euphronides* and 15 other Hyloidea species. Maximum likelihood tree inferred with IQ-TREE v2.2.0.3 [37] from five concatenated nuclear genes (CXCR4, POMC, SIAH1, SLC8A3, TYR) representing nine Hyloidea families in total, rooted with *Rana temporaria* (Neobatrachia, Ranoidea, Ranidae) as outgroup. Tree visualized with iTOL v6 [38]. colors indicate taxonomic family (legend at right). *P. euphronides* (family Strabomantidae) is highlighted in red and bold. Branch support values are ultrafast bootstrap percentages (%). Tree scale: 0.1. (Also presented as Figure 5 in the main manuscript.)

The 12-gene supermatrix analysis (Figure SR31) also placed *P. euphronides* as a distinct lineage, with some topological differences relative to the 5-gene tree. In this dataset, *P. euphronides*, the only Terrarana representative included, clustered adjacent to the monophyletic Hylidae clade with low bootstrap support (41%). This grouping was nested within a larger clade that also contained representatives of Leptodactylidae (*E. pustulosus*) and Bufonidae (*B. bufo*), supported by a bootstrap value of 52.7%. The Dendrobatidae representative (*R. imitator*) was placed outside this clade as a separate lineage. The outgroup *R. temporaria* (Ranidae) was placed on a separate branch.

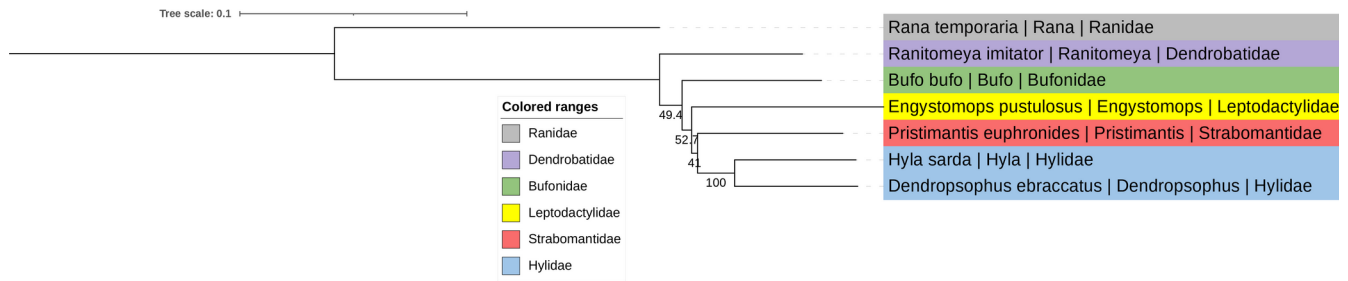


Figure SR31. Twelve-gene supermatrix phylogeny of *Pristimantis euphronides* and five other Hyloidea species. Maximum likelihood tree inferred with IQ-TREE v2.2.0.3 [37] from 12 concatenated nuclear genes (BDNF, BMP2, CXCR4, MC1R, NTF3, POMC, RAG2, RHO, SIAH1, SLC8A3, TNS3, TYR) representing five Hyloidea families in total, rooted with *Rana temporaria* (Neobatrachia, Ranoidea, Ranidae) as outgroup. Tree visualized with iTOL v6 [38]. colors indicate taxonomic family (legend at right). *P. euphronides* (family Strabomantidae) is highlighted in red and bold. Branch support values are ultrafast bootstrap percentages (%). Tree scale: 0.1.

Maximum-likelihood phylogenies were inferred for each of the 12 individual gene alignments (Additional file 11: Figures 1-12). In each single-gene tree, *P. euphronides* was placed within the clade Terrarana, a group of direct-developing Neotropical frogs within the superfamily Hyloidea that comprises the families Strabomantidae, Craugastoridae, Eleutherodactylidae, and Brachycephalidae.

In the phylogenetic tree for the gene BDNF (Additional file 11: Figure 1), the *P. euphronides* sequence grouped with that of *E. coqui* (Eleutherodactylidae) with 75.9% bootstrap support. As in the following, bootstrap values refer to ultrafast bootstrap (UFBoot) percentages based on 1,000 replicates, as calculated by IQ-TREE. No additional sequences from the family Strabomantidae were available for this gene.

The phylogenetic tree for the gene BMP2 (Additional file 11: Figure 2) showed a similar placement of *P. euphronides* next to *E. coqui*, with even higher bootstrap support (94.4%). No other Terrarana representative sequences were available for the BDNF and BMP2 gene.

In the phylogenetic tree for the gene CXCR4 (Additional file 11: Figure 3), *P. euphronides* grouped with *Pristimantis thymelensis* (Strabomantidae) with a bootstrap support value of 86.2%. Two other

genera of this family, *Strabomantis* and *Niceforonia* formed genus-specific clusters (bootstrap support > 99%). These were embedded within a broader group of all four Terrarana families, in which Eleutherodactylidae, including *E. coqui*, *E. planirostris*, *E. marnockii*, *D. diastema*, and *A. gutturosa*, formed a separate and well-supported (bootstrap support 95%) lineage.

In the phylogenetic tree for gene MC1R (Additional file 11: Figure 4), *P. euphronides* clustered with *E. coqui* (bootstrap support 75.5%), the only other terraranan species included in this dataset. *P. euphronides* also grouped with *E. coqui* in the phylogenetic tree for the gene NTF3 (Additional file 11: Figure 5), within a broader group that also included non-terraranan families such as Centrolenidae and Bufonidae. The dataset for this gene did not include representatives of other terraranan families, and the NTF3 sequence of *P. euphronides* grouped closest with *Ceratophrys cornuta* (Ceratophryidae), with bootstrap support of 70.3%.

In the phylogenetic tree for the gene POMC (Additional file 11: Figure 6), *P. euphronides* was placed within a distinct *Pristimantis* clade (bootstrap support 97%) with 12 other species of that Strabomantidae genus. Species of two other Strabomantidae genera, *Oreobates* and *Lynchi*, formed a clade within a broader group in which, apart from the Strabomantidae genus *Niceforonia*, four species of Craugastoridae were also included with 78.6% bootstrap support. Eleutherodactylidae sequences of that gene were grouped in their own clade with 95.9% bootstrap support. Additional terraranan sequences from Craugastoridae and Brachycephalidae were present and formed family-specific clusters.

In the phylogenetic tree for the gene RAG2 (Additional file 11: Figure 7), among Strabomantidae members, *P. euphronides* grouped closest with *P. thymelensis* (bootstrap support 100%) and *N. brunnea* closest with *Strabomantis sulcatus* (bootstrap support 57.7%). These sequences joined in a broader group (bootstrap support 97.6%) with clusters of Craugastoridae and Eleutherodactylidae RAG2 genes. The phylogenetic tree for the gene RHO (Additional file 11: Figure 8) showed *P. euphronides* in a group with *P. bracki* and *Niceforonia dolops* (Strabomantidae) that also included *E. coqui* and *A. gutturosa* (Eleutherodactylidae) as well as *Craugastor milesi* and *Craugastor xucanebi* (Craugastoridae) with a bootstrap support value of 85.3%.

In the phylogenetic tree for the gene SIAH1 (Additional file 11: Figure 9), *P. euphronides* clustered with *Pristimantis cruentus* (bootstrap support 76.5%). That was also the case for the phylogenetic tree for the gene SLC8A3 (Additional file 11: Figure 10), where they joined the other two Strabomantidae members (bootstrap support 79.2%) *Strabomantis biporcatus* and *N. brunnea* in a broader group that comprised not only four Eleutherodactylidae, two Brachycephalidae, and three Craugastoridae members but also three non-terraranan species: *Odontophrynus occidentalis* (Alsodidae) as well as *Pleurodema* sp. VUB 1030 and *E. pustulosus* (Leptodactylidae).

In the phylogenetic tree for the gene TNS3 (Additional file 11: Figure 11), *P. euphronides* and *Ischnocnema* species (Brachycephalidae) were the only Terrarana members present and clustered with bootstrap support of 97.2%. No Eleutherodactylidae or Craugastoridae gene sequences were available for this gene.

In the phylogenetic tree for the gene TYR (Additional file 11: Figure 12), which included the largest number of Strabomantidae sequences among the single-gene datasets (262 sequences representing 148 species across nine genera), *P. euphronides* clustered with *P. urichi* and *P. thymelensis* in a clade supported by 66% ultrafast bootstrap. This *Pristimantis* clade was embedded within a larger lineage of Strabomantidae that included several well-supported clusters of congeneric species. For example, *P. achatinus* sequences grouped with *P. condor*, *P. buccinator*, and *P. conspicillatus* (89% support), while *P. brevifrons* formed a clade with *P. acatallelus* (99%), and *P. scoloblepharus* was closely associated with *P. parectatus* (76.4%). Other *Pristimantis* lineages included *P. rozei*, *P. mondolfii*, *P. factiosus*, and multiple *P. zeuctotylus* specimens, which grouped in a clade with 99.7% support.

Outside the *Pristimantis* genus but within the Strabomantidae family, distinct clades were recovered for *Phrynopus*, *Oreobates*, *Lynchi*, *Tachiramantis*, *Niceforonia*, and *Urkuphryne*. *Phrynopus* formed a strongly supported monophyletic group including *P. peruanus*, *P. remotum*, and *P. mirosławae* (95.2% support), and *Oreobates* sequences formed a separate, well-supported clade (e.g., 92.5% for *O. saxatilis* and *O. machiguenga*). *N. brunnea* and *N. dolops* grouped with *Lynchi* species (99.6% support), forming a distinct lineage within Strabomantidae. *Tachiramantis* also formed a monophyletic clade (100% bootstrap support). These major Strabomantidae lineages were separated from Craugastoridae and Brachycephalidae, which each formed their own clades. For instance, *Craugastor*

species such as *C. mexicanus*, *C. pygmaeus*, and *C. longirostris* grouped in a larger clade supported by 93.7%, and *Ischnocnema* species (Brachycephalidae) formed a distinct lineage as well.

Divergence times and biogeographic context

Among the five anuran species listed for Grenada in AmphibiaWeb [39], the endemic *P. euphronides* is phylogenetically the most divergent from all other co-occurring species (Table SR35). The nearest relative on the island, *E. johnstonei*, shares a common ancestor with *P. euphronides* at approximately 51.9 Ma (25 edges). The two *Leptodactylus* species and *Rhinella marina* are more distant still, with tMRCAs of approximately 62.0 Ma (31-40 edges). AmphibiaWeb [39] lists *Leptodactylus wagneri* with a Grenada country code, whereas Amphibian Species of the World [40] does not record *L. wagneri* from the Lesser Antilles and recognizes only *Leptodactylus validus* on these islands.

Table SR35. Phylogenetic distance to *Pristimantis euphronides* of anuran species listed for Grenada. Divergence times and topological distances derived from the Portik et al., 2023 [21] time-calibrated phylogeny. tMRCA: time to most recent common ancestor in million years. Topo. edges: topological distance in number of edges. Geographic distributions from AmphibiaWeb [39].

Species	tMRCA (Ma)	Topo. edges	Endemic to Grenada	Distribution
<i>Pristimantis euphronides</i>	-	-	Yes	Grenada
<i>Eleutherodactylus johnstonei</i>	51.85	25	No	Widespread Caribbean
<i>Leptodactylus validus</i>	62.02	31	No	Southern Lesser Antilles, Trinidad and Tobago, South America
<i>Leptodactylus wagneri</i>	62.02	32	No	Southern Lesser Antilles, Trinidad and Tobago, South America
<i>Rhinella marina</i>	62.02	40	No	Pan-tropical

The closest relative of *P. euphronides* in the full phylogeny [21] is *P. shrevei* from Saint Vincent and the Grenadines (tMRCA $\approx$ 3.15 Ma, 2 edges). The next closest relatives are a clade of 12 mainland species at tMRCA $\approx$ 18.8 Ma, 5 to 10 edges from *P. euphronides*. Ten of these species occur in Colombia only: *P. anolirex*, *P. merostictus*, *P. savagei*, *P. lutitus*, *P. bowara*, *P. medemi*, *P. carrangerorum*, *P. affinis*, *P. lynchi*, and *P. nervicus*. The remaining two species are recorded from both Colombia and Venezuela in Amphibian Species of the World [40], although AmphibiaWeb [39] lists *P. mondolfii* as occurring only in Venezuela.

Geographic enrichment of the full phylogeny [21] using AmphibiaWeb [39] occurrence data revealed that five of the six Caribbean *Pristimantis* are single-island endemics. The three species *P. charlottevillensis*, *P. urichi*, and *P. turpinorum* are restricted to Trinidad and Tobago, *P. shrevei* to Saint Vincent and the Grenadines, and *P. euphronides* to Grenada. Only *P. euphronides* and *P. shrevei* occur on oceanic Lesser Antillean islands, while the genus *Pristimantis* is absent from the Greater Antilles. The three species from Trinidad and Tobago occur on continental-shelf islands closer to the South American mainland than the oceanic Lesser Antilles. The species *P. terraebolivaris* occurs on Trinidad and in Venezuela. The minimum great-circle distance from Grenada (country centroid) to the closest *Pristimantis* congener is $\sim$ 87 km to *P. shrevei* on Saint Vincent and the Grenadines.

The 12 mainland congeners at the next phylogenetic node lie at country-centroid distances of $\sim$ 849 km in Venezuela and $\sim$ 1,636 km in Colombia, with *P. nicefori* recorded from both. Per-species type-locality distances from Pointe Salines, the southwesternmost point of Grenada, range from 1,267 km for *P. mondolfii* in Táchira, Venezuela, to 1,666 km for *P. savagei* in the Sierra de la Macarena, Meta, Colombia, with a mean of 1,430 km (Table SR36). All 12 species occur in highland Andean habitats along the Cordillera Oriental of Colombia, the Sierra de la Macarena, or the contiguous Venezuelan Andes.

Table SR36. Type-locality coordinates and distances to Grenada of the 12 closest mainland *Pristimantis* congeners. Species at the 18.83 million years (Ma) node of the Portik et al., 2023 [21] time-calibrated phylogeny, ranked by topological distance (Topo. edges). Type-locality strings were retrieved from Amphibian Species of the World v6.2 [40]. The Source column indicates how each

coordinate was obtained. Coordinates labeled OriginalDescription_DMS or OriginalDescription_Decimal are explicit coordinates from the original taxonomic description, in degrees-minutes-seconds or decimal-degree form respectively. Coordinates labeled GoogleMaps were extracted from the Google Maps URL obtained by locating the most specific named place in the type-locality string. Distances are great-circle distances by haversine to Pointe Salines (12.00°N, 61.79°W), the southwesternmost point of Grenada.

Species	Topo. edges	Coordinates (°N, °W)	Source	Distance (km)
<i>P. mondolfii</i>	5	7.5680, 72.4480	GoogleMaps	1,267
<i>P. anolirex</i>	6	7.1371, 72.6681	GoogleMaps	1,309
<i>P. merostictus</i>	6	6.0093, 73.2233	GoogleMaps	1,421
<i>P. nicefori</i>	6	6.9833, 72.7385	GoogleMaps	1,324
<i>P. savagei</i>	6	2.6416, 73.6031	GoogleMaps	1,666
<i>P. lutitus</i>	7	6.2860, 73.1526	GoogleMaps	1,399
<i>P. bowara</i>	7	7.1687, 72.2655	OriginalDescription_Decimal	1,268
<i>P. medemi</i>	7	4.2693, 73.5996	GoogleMaps	1,558
<i>P. carrangerorum</i>	8	5.2933, 72.7053	GoogleMaps	1,412
<i>P. affinis</i>	9	4.5476, 73.7360	GoogleMaps	1,553
<i>P. lynchi</i>	10	5.4333, 72.7333	OriginalDescription_DMS	1,406
<i>P. nervicus</i>	10	4.2974, 73.7809	GoogleMaps	1,572

References

1. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics*. 2018;34(15):2666-2669.
2. Shen W, Sipos B, Zhao L. SeqKit2: A Swiss army knife for sequence and alignment processing. *Imeta*. 2024;3(3):e191.
3. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. 2011;27(6):764-770.
4. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 2020 Mar 18;11(1):1432.
5. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol*. 2019;20:257.

6. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol.* 2019 May;37(5):540-546.
7. Oxford Nanopore Technologies (ONT). Medaka: sequence correction provided by ONT Research. GitHub. Version 2.0.1. <https://github.com/nanoporetech/medaka/releases/tag/v2.0.1>. Accessed 1 Mar 2025.
8. Kuhl H, Li L, Wuertz S, Stöck M, Liang XF, Klopp C. CSA: A high-throughput chromosome-scale assembly pipeline for vertebrate genomes. *Gigascience.* 2020;9(5):giaa034.
9. Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. *Nat Methods.* 2020;17(2):155-158.
10. Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, et al. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* 2022;23(1):258.
11. Kolmogorov M, Raney B, Paten B, Pham S. Ragout - a reference-assisted assembly tool for bacterial genomes. *Bioinformatics.* 2014;30(12):i302-i309.
12. Kolmogorov M, Armstrong J, Raney BJ, Streeter I, Dunn M, Yang F, et al. Chromosome assembly of large and complex genomes using multiple references. *Genome Res.* 2018;28(11):1720-1732.
13. Chen Y, Zhang Y, Wang AY, Gao M, Chong Z. Accurate long-read de novo assembly evaluation with Inspector. *Genome Biol.* 2021;22:312.
14. Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* 2020;21:245.
15. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience.* 2021;10(2):giab008.
16. Kuhl H. 2024. HANNO: efficient High-throughput ANNOtation of protein coding genes in eukaryote genomes. doi:10.5281/zenodo.11532370. Accessed 10 Apr 2025.
17. 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. *Mol Biol Evol.* 2021;38(10):4647-4654.
18. Kriventseva EV, Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, et al. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* 2019;47(D1):D807–11.
19. Cabanettes F, Klopp C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ.* 2018 Jun 4;6:e4958.
20. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 2016;17(1):132.
21. Portik DM, Streicher JW, Wiens JJ. Frog phylogeny: A time-calibrated, species-level tree based on hundreds of loci and 5,242 species. *Mol Phylogenet Evol.* 2023 Nov;188:107907.

22. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018 Sep 15;34(18):3094-3100.
23. Pedersen BS, Quinlan AR. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics*. 2018;34(5):867-868.
24. Schmid M, Feichtinger W, Steinlein C, Rupprecht A, Haff T, Kaiser H. Chromosomal banding in Amphibia. XXIII. Giant W sex chromosomes and extremely small genomes in *Eleutherodactylus euphronides* and *Eleutherodactylus shrevei* (Anura, Leptodactylidae). *Cytogenet Genome Res*. 2002;97(1-2):81-94.
25. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10:421.
26. Heinicke MP, Duellman WE, Hedges SB. Major Caribbean and Central American frog faunas originated by ancient oceanic dispersal. *Proc Natl Acad Sci U S A*. 2007 Jun 12;104(24):10092-7.
27. Li H. Protein-to-genome alignment with minimap2. *Bioinformatics*. 2023;39(1):btad014.
28. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol*. 2021 Dec 9;38(12):5825-5829.
29. Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2013-2015. <http://www.repeatmasker.org>. Accessed 10 Dec 2025.
30. Morgulis A, Gertz EM, Schäffer AA, Agarwala R. WindowMasker: window-based masker for sequenced genomes. *Bioinformatics*. 2006;22(2):134-141.
31. Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mob DNA*. 2021 Jan 12;12(1):2.
32. Yoshimoto S, Okada E, Umemoto H, Tamura K, Uno Y, Nishida-Umehara C, et al. A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. *Proc Natl Acad Sci U S A*. 2008;105(7):2469–74.
33. UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Res*. 2025 Jan 6;53(D1):D609-D617.
34. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: the protein families database in 2021. *Nucleic Acids Res*. 2021;49(D1):D412–9.
35. NCBI Nucleotide. National Center for Biotechnology Information. <https://www.ncbi.nlm.nih.gov/nucleotide/>. Accessed 22 Mar 2025.
36. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009 May 1;25(9):1189-91.

37. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 2020;37(5):1530–4.
38. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* 2024 Jul 5;52(W1):W78-W82.
39. AmphibiaWeb. 2026. <https://amphibiaweb.org>. University of California, Berkeley, CA, USA. Accessed 16 Jan 2026.
40. Frost DR. Amphibian Species of the World: an Online Reference. Version 6.2. American Museum of Natural History, New York. <https://amphibiansoftheworld.amnh.org>. Accessed 26 April 2026.