Supplemental Methods

TABLE OF CONTENTS

Sample collection

L. boringii samples used in this study were collected from the Badagongshan National Nature Reserve, Hunan Province, China. For genomic sequencing, we collected muscle and liver from one adult male and flash-froze them in liquid nitrogen. Genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA, USA). Muscle and other seven tissues from the same individual (including skin, brain, testis, liver, heart, kidney, and spleen) were collected for transcriptomic sequencing. Liver tissue DNA was extracted from another male individual for construction of Hi-C and HiFi libraries.

For whole-genome resequencing data, we sampled and sequenced 20 male and 20 female individuals (Supplemental Dataset 1) and selected two individuals to generate a full-sib family (including two parents and 147 offspring, Supplemental Dataset 2), each individual was euthanized in MS222 and dissected under a stereomicroscope. Muscle tissues were stored at −80°C for sequencing. DNA extraction was carried out using EasyPure® Genomic DNA Kit (TransGen Biotech, China).

All the samples from the wild population were subjected to 15~20× coverage and offspring with ~5× coverage. For transcriptomic sequencing, RNA was extracted from gonads in two sexes at four developmental stages. We further collected both gonad and somatic tissues (muscle, heart, kidney, liver, lung and brain) in adult frogs (Supplemental Dataset 6). Each sample included three biological replicates.

All experiments involving animals in this study were approved by the Animal Ethics Committee of the School of Life Sciences, Central China Normal University (CCNU-IACUC-2022–010). We have complied with all relevant ethical regulations for animal testing and research.

Genome sequencing, assembly and annotation

A combination of Illumina sequencing, PacBio sequencing, and Hi-C sequencing was used to generate the *L. boringii* genome assembly. Paired-end libraries with insert sizes of 300~350 bp were constructed and sequenced on the Illumina NovaSeq 6000 system (Illumina, San Diego, CA, USA). After the removal of sequencing adapters, contaminant reads (mitochondrial, bacterial, and viral sequences), and low-quality reads, we finally obtained 329.7 Gb (~97.5× coverage) of clean reads. The high-quality reads were used for genome size estimation by the *k*-mer method.

The DNA extracted from the muscle and liver was used for sequence library construction using the PacBio Sequel II Platform. For 20 kb template library preparation, ten micrograms (μg) of *L. boringii* genomic DNA was used, following the manufacturer's protocol with the BluePippin Size Selection system (Sage Science, Beverly, MA, USA). The PacBio single molecule real-time (SMRT) library was prepared using the SMRT bell express template prep Kit 2.0 (Pacific Biosciences, Menlo Park, CA, USA) and sequenced on the PacBio Sequel II platform. After filtering low-quality reads, joints, and short reads, we recovered 333.60 Gb of subreads (~98.6× coverage, Supplemental Table S1), with an average length of 13,213 kb. We then generated a draft assembly with Canu v2.0 (Koren et al. 2017). The corrected subreads were used for genome assembly using WTDBG v1.1.006 (https://github.com/ruanjue/wtdbg). The draft assembly was polished using raw PacBio sequencing data by arrow, and with Illumina paired-end reads with Pilon v1.18 (Utturkar et al. 2017). The preliminary assembled genome was de-redundant using purge_haplotigs, which identifies and removes redundant heterozygous contigs based on read depth distribution and sequence similarity. The statistics of the reads mapping rate were summarized with BWA v0.7.17 (Li and Durbin 2009) and SAMtools v1.9 (Li et al. 2009). Finally, we assessed genome completeness using the BUSCO v4.0.1 (Simão et al. 2015) with the metazoan_odb9 lineage (n = 978).

To further improve the continuity of the assembled genomes and anchor the assemblies into chromosomes, Hi-C sequencing was performed to order and orient the contigs, as well as to correct mis-joined sections and merge overlaps. According to the protocol, nuclear DNA from liver was cross-linked and enzymatically digested with Hind III restriction enzyme overnight, leaving pairs of distally located but physically interacting DNA molecules attached to each other. The sticky ends of the digested fragments were biotinylated and ligated to each other to form chimeric circles. Biotinylated circles, which are chimeras of physically associated DNA molecules from the original cross-linking, were enriched, sheared, and sequenced with the Illumina NovaSeq 6000 system. A total of ~1,126.92 million clean Hi-C reads pairs (173.6 GB, ~51.4× coverage) were obtained and then mapped to the draft assembly using BWA v0.7.17 (Li and Durbin 2009) and were filtered to obtain valid pairs. Then, the contigs were anchored into chromosomes by Hi-C sequencing reads through the Juicer v1.5 (Durand et al. 2016) and 3D-DNA v180922 (Dudchenko et al. 2017) software workflows. Based on the Hi-C correction, we assembled 1,570 contigs

to 13 pseudochromosomes and generated a 3.38 Gb of the *L. boringii* reference genome, which was used for subsequent analyses.

Repeat sequences were identified by two different methods. First, we identified known TEs using two programs (RepeatMasker and RepeatProteinMask, http://www.repeatmasker.org). Then, we used these two programs to identify TEs by aligning the genome sequence to a self-generated curated TE protein database separately. Second, we constructed a *de novo* repeat library using RepeatModeler (Flynn et al. 2020) and LTR-FINDER (Castelo et al. 2002), which yielded consensus sequences and classification information for each repeat family. The RepeatMasker program was then applied to annotate these genome sequences.

We integrated three approaches, namely, *de novo* prediction, homology search, and transcript-based assembly, to annotate protein-coding genes in a repeat-masked genome. Consensus gene structures were generated by integrating the homolog protein prediction and *de novo* prediction. *De novo* prediction of protein-coding genes using the GlimmerHMM v 3.0.4 (Majoros et al. 2004) and AUGUSTUS v 3.3.2 (Stanke et al. 2006). The genes predicted from above methods were integrated into a non-redundant and more complete gene set using MAKER2 v 2.31.10 (Holt and Yandell 2011), and the final reliable gene set was obtained using the HiCESAP pipeline (Gooalgene Co., Ltd., Wuhan, China, https://www.gooalgene.com/). To assign gene functions, the predicted gene sequences were searched against the NR, GO, KEGG, KOG, Pfam, SwissProt, and TrEMBL databases. Annotation integrity was estimated by comparison with reference genome annotations and BUSCO v4.0.1 (Simão et al. 2015).

Cytogenetic karyotype analysis of *L. boringii*

Metaphase chromosomes were prepared from the kidneys of female and male *L. boringii* tadpoles following the method described previously (Phimphan and Aiumsumang 2021), with slight adaptations as follows. We injected colchicine into the tadpole's abdominal cavity at a dose of 1-5 μg/g for 2-3 hours prior to tissue collection. Kidney tissues were first washed with Phosphate-buffer saline (PBS) and then minced, filtered and centrifuged to obtain precipitated kidney cells. We then resuspended cells with 0.34% KCl solution and dropped them onto a glass slide to let stand for 30 minutes. Steam treatment was performed on the mixture of fixed solution (ethanol: acetic acid: water=1:2:3) and anhydrous ethanol for 2 hours and 30 minutes,

respectively. We then rinsed the glass slides with another fixed solution (ethanol: acetic acid=1:2) 3-4 times and air-dried them. Conventional staining was done using 10% Giemsa's solution for 30 minutes.

Variants calling and primary data filtration

All Illumina raw data were quality-checked, demultiplexed, and filtered by FastQC v0.11.5 (Andrews 2010) sequencing reads were mapped to the reference genome with BWA-MEM (Li 2013; Li and Durbin 2009). Individual sam files were converted to bam files and sorted with SAMtools v1.9 (Li et al. 2009), followed by the removal of duplicate reads using Picard v2.1 (http://broadinstitute.github.io/picard/).

We first applied a hard filter to the raw data by sets using GATK v4.1.3.0 (McKenna et al. 2010) with the following criteria: QUAL < 30.0; QualByDepth (QD) < 2.0; FilterStrand (FS) > 60.0; RMS Mapping Quality (MQ) < 20.0; ReadPosRankSum <−8.0. Secondly, bi-allelic SNPs with minor allele frequency (MAF) ≥ 0.01, mean depth values (min-meanDP) ≥ 5, and proportion of missing data < 0.20 were kept using VCFtools v0.1.15 (Danecek et al. 2011).

Identification of sex chromosome and the SLR

A GWAS was performed by mixed-model association using EMMA eXpedited (EMMAX, Kang et al. 2010), using sex as a phenotype. Phased genotypes were processed with PLINK v1.90b6.10 (Purcell et al. 2007) to generate the input for EMMAX. The threshold for significance in the GWAS was set with the *p*-value of 2.699e-8 by dividing 0.05 by the number of total SNPs. Sex-linked regions were inferred based on the presence of SNPs significantly associated with sex by the GWAS analysis. We used 200 kb overlapping sliding windows with a step size of 50 kb to calculate the F_{ST} values between male and female populations using VCFtools v0.1.15 (Danecek et al. 2011). The top 1% was selected as the significance threshold of $F_{ST.}$

To identify the sex-specific SNPs, we filtered the SNPs and retained those that were present in at least 75% of all individuals (males and females combined) and a minor allele frequency (MAF) ≥ 0.05 and a heterozygosity threshold < 0.75 by VCFtools v0.1.15 (Danecek et al. 2011) was used to retain SNPs. Sex-specific SNPs were defined based on sex differences in allele frequencies (Brelsford et al. 2017). The screening criteria were defined as SNPs with a allele frequency ≥ 0.95 in females and an allele frequency differential (ΔAF) between females and males ≥ 0.4 to identify

male-specific supporting XY, and vice versa for female-specific SNPs supporting ZW. Thus, SNPs heterozygous in males and homozygous in females and are regarded as male-specific SNPs supporting XY sex chromosomes, and vice versa for female-specific SNPs supporting ZW chromosomes. As a result, we obtained 274,384 male-specific heterozygous SNPs and 240 female-specific heterozygous SNPs. Variants annotation of the sex-linked SNPs was performed using SnpEff (Cingolani et al. 2012). We identified 607 missense variants belonging to 395 genes.

Validation of sex-specific markers by conventional Sanger sequencing

To obtain more accurate sex markers, the sex-linked SNPs obtained from the previous step were further verified. The 300-bp upstream and downstream sequences of each male-specific SNPs were used to design primers. We used DNA samples from toes of 24 males and 24 females to further validate these SNPs by Sanger sequencing. After rounds of screening and validation, we finally generated four pairs of primers (Supplemental Tables S19, S20) that were heterozygous in all male individuals and homozygous in all female individuals. These strictly validated sex-linked markers could be used for accurately separating the genotype sex of *L. boringii* in the sex reversal identification and transcriptome analysis.

Haploid genome, *k*-mer analysis and coverage

We supplemented the original PacBio HiFi data using the PacBio Sequel II Platform and generated additional CCS reads (166.03Gb, ~49.1X coverage). We then conducted a *de novo* assembly of two haploid genomes. These long and highly accurate HiFi reads were assembled using Hifiasm (https://github.com/chhylp123/hifiasm) and HiCanu (Nurk et al. 2020). Each haploid genome was then used for the second round of improvement using the same procedure of Hi-C assembly described above. We generated two high-quality haploid genomes of a male *L. boringii*. The length of two haplotype chromosomes was 3.923 Gb and 3.740 Gb for HapA and HapB, respectively. We quantified genome completeness for each haplotype genome using the BUSCO v4.0.1 (Simão et al. 2015) with the metazoan_odb9 lineage (n = 978).

To identify the Y Chromosome in *L. boringii*, we followed the *k*-mer analysis method described previously (Morris et al. 2018). In brief, we utilized the HAWK pipeline (Rahman et al. 2018) to count *k*-mers from paired-end DNA-seq reads. Because of the extensive sequencing depth and large sample number, comparing all males to all

females was computationally prohibitive. Therefore, we divided the individuals into four groups (five males and five females), identified male and female unique *k*-mers in each group and filtered by more than 20× normalized coverage. Then, in all four groups, we filtered female-specific *k*-mers (female-mers) and male-specific *k*-mers (Y-mers) shared in at least two groups (Supplemental Fig. S6). All filtered sex-specific *k*-mers were further aligned to haploid genomes by using BWA v0.7.17 (Li and Durbin 2009) to infer the Y-linked haploid genome.

We also aligned male and female paired-end DNA-seq reads to the XY reference genome (Lbor.v1), HapA and HapB reference genome by BWA v0.7.17 (Li and Durbin 2009) and extracted uniquely mapping reads. We then used BEDtools (Quinlan and Hall 2010) to calculate the coverage (number of times each site was sequenced divided by the total number of sequenced sites) of each scaffold in each sample. For each scaffold, we calculated the male-to-female (M:F) FC coverage as $\log_2$(average male coverage) – $\log_2$(average female coverage).
Additionally, we analyzed synteny between HAChr1 and HBChr1 by pairwise mapping whole genomes using Minimap2 v2.24 (Li 2018), identified structural variants with SyRI v1.6 (Goel et al. 2019), and plotted syntenic blocks larger than 20 kb using plotsr v0.5.4 (Goel and Schneeberger 2022).

High-density genetic map

The male heterozygous and female homozygous genotypes were encoded as lm×ll, while the male homozygous and female heterozygous genotypes were encoded as nn×np.
According to the population type, the developed markers were filtered following four criteria to remove: (1) loci with missing data in parents and those loci where both parents were homozygous or heterozygous. (2) loci with a missing rate >25% or within a physical distance of 300kb. (3) loci with parental genotypes "lm×ll" or "nn×np" that did not conform to the 1:1 segregation ratio. (4) markers with biased segregation based on chi-square tests and $\alpha < 0.05$.We excluded markers that were heterozygous in both parents, for this class of marker, in heterozygous offspring, we would be unable to determine the parent of origin for each allele, rendering them uninformative for sex-specific linkage mapping.
Based on the selected SNP markers, the Lep-MAP3 (Rastas 2017) software was used to partition linkage groups, with an LOD threshold of 3.0. Subsequently, the maximum likelihood method was employed to order the linkage groups. Post-processing of the

genetic map for each LG was done with the online software MareyMap (Siberchicot et al. 2017). We built a genetic map by plotting SNP genetic distance against SNP physical distance for each LG and sex. The integrated genetic map of *L. boringii* was constructed, consisting of 13 linkage groups (Supplemental Figs. S15, S16). A final set of 10,884 curated informative SNPs was used to calculate sex-specific local recombination rates using a locally weighted regression model (LOESS) with a span parameter of 0.2 in MareyMap online. This method estimated the local recombination rates (cM/Mb) as the slope of the curve describing the relationship between the physical (Mb) and genetic (cM) positions.

RNA sequencing and gene expression analysis

Based on the results of histological data of gonads, we found four critical stages (G25, G28, G42, Adult) of development in sex differentiation. (1) Stage G25: the gonads are undifferentiated, but primordial germ cells begin to proliferate. (2) Stage G28: the sex of the gonads can be identified based on morphological characteristics in dissection. (3) Stage G42: tadpoles develop to the peak of metamorphosis, and the gonads are more mature. (4) Adult frogs: fully sexually mature.

To investigate biased gene expression, we collected gonad tissues from both sexes at these critical stages. To compare gene expression between gonad tissues and somatic tissues, we also collected tissues from muscle, heart, kidney, lung, and brain from adult frogs. These tissues were collected from each individual and immediately maintained in RNAlater reagent. All samples were further confirmed for genotypic sex by Sanger sequencing of sex-linked markers (Supplemental Tables S19, S20). Due to the inability to determine the sex of the samples during early gonadal development, this step was necessary to ensure the accuracy of downstream analyses excluded any sex-reversed individuals.

For each stage and sex, we prepared three biological replicates for RNA extractions by using an RNA extraction kit (Omega Bio-Tek) in combination with TRIzol reagent (Invitrogen). The integrity and concentration of RNA were tested with an Agilent 2100 Bioanalyzer instrument, and the qualified RNA was used for transcriptome library sequencing. The cDNA sequencing library was constructed separately for each individual and was sequenced using the Illumina NovaSeq 6000 system. The raw data obtained were subjected to data quality control and filtering to obtain valid data. RNA-seq reads were mapped to the reference genome using HISAT2 (Kim et al. 2015),

and the reads mapped to each gene were counted using featureCounts v1.6.2 (Liao et al. 2014). Read counts were normalized using the TPM method. TPM = (CDS read count × mean read length × 10^6)/ (CDS length × total transcript count). Differentially expressed gene analyses to compare tissue types, developmental stages and sexes were performed with the edge R package (Robinson et al. 2010). Differentially expressed genes in male and female individuals were identified using DEseq2 (Love et al. 2014), with differential $|FC| \geq 2$ and FDR ≤ 0.05.

Sex-biased genes were classified into four categories of |FC| 2–4 (low), 4–8 (mid), and > 8 (high), and expressed as a $\log_2$ ratio of female-to-male (which has negative values for male-biased genes and positive values for female-biased genes). As suggested by Montgomery and Mank (2016), only $|FC| \geq 2$ will be interpreted throughout, in order to minimize possible scaling issues due to whole-body sampling (ovaries are slightly larger than testes, which may potentially lead to bias in calling sex-biased gene expression). Thus, both conditions FDR < 0.05 and $|\log_2 FC| \geq 1$ will have to be met when calling the sex-biased gene.

Allele-specific expression (ASE) analysis.

To estimate ASE patterns from RNAseq data, we tailored previously published pipelines (Quinn et al. 2014). We called SNPs separately for males and females using SAMtools mpileup v1.9 (Li et al. 2009). We performed initial SNP filtering using VarScan (Koboldt et al. 2012) with the following parameter: --min-coverage 2 --min-avg-qual 20 --min-freq-for-homs 0.90 --p-value 1 --strand-filter 0 --min-var-freq 1e-10. We filtered SNPs to retain only those located in exonic regions. To enable comparative analysis between sex chromosomes and autosomes, we partitioned SNPs into autosomal and sex-chromosomal categories based on chromosomal positional information. To exclude potential sequencing errors from our SNP dataset, we applied coverage filtering thresholds by Zimmer et al. 2016. RNA-Seq data have an intrinsic bias for the estimation of ASE, because those reads that resemble the reference genome have a higher probability of aligning successfully. To avoid the potential bias in our ASE estimations from preferential assignment of reads to the reference allele (Stevenson et al. 2013), we removed clusters of more than 5 SNPs in 100 bp windows.

If genes have biallelic expression, meaning that alleles from both chromosomes are expressed at the same level, we expect a probability of around 0.5 of recovering reads

from either chromosome. For each SNP in the final filtered dataset, we tested for ASE by identifying significant deviations from the expected probability of 0.5 using a two-tailed binomial test ($p < 0.05$). We corrected for multiple testing when running binomial tests on autosomal SNPs. Additionally, we called SNPs ASE if a minimum of 70% of the reads stemmed from one of the chromosomes. We called genes ASE if they had at least one SNP with a consistent ASE pattern across all heterozygous samples. We tested for significant differences in ASE patterns between the sexes and between the autosomes and the sex chromosome using Wilcoxon rank sum tests.

Gene coexpression analysis.

To cluster genes with similar expression patterns across samples, we conducted a coexpression analysis based on 24 samples using WGCNA v1.63 (Langfelder and Horvath 2007). We constructed an unsupervised network for transcriptome data using the function blockwiseModules with default parameters. First, a matrix of Pearson's correlations between genes was generated based on TPM values across samples. Then an adjacency matrix representing the connection strength among genes was constructed by raising the correlation matrix to a soft threshold power to achieve a scale-free topology fit index of 0.80. Next, the adjacency matrix was used to calculate the topological overlap matrix (TOM). Genes with similar coexpression patterns across samples were grouped using hierarchical clustering of dissimilarity among the topological overlap measures (1 − TOM). Coexpressed modules were determined using a dynamic tree-cutting algorithm setting with a minimum module size of 30 and a cut height of 0.998. An eigengene value (the first principal component of the scaled module expression profiles) was calculated to characterize the overall expression trend for each module. The intramodular connectivity was measured as kME values that represent the Pearson's correlation between the expression level of that gene and the ME. Then the Pearson's correlations between ME values and sampling trait values were calculated to measure the strength and direction of association between modules and traits. Fisher's asymptotic *p* values were calculated for given correlations using the corPvalueFisher module. Significant module–trait associations were considered when $p < 0.05$.

We integrated the previously identified sex-related genes, genes with missense mutations, and sex-associated gene sets from WGCNA to jointly confirm key regulatory pathways involved in sex development. We performed KEGG enrichment analysis on these gene sets to investigate the biological processes. First, these genes were aligned to the KEGG database (http://www.genome.ad.jp/kegg/). Then, we

applied KEGG enrichment using the R package clusterProfiler (Yu et al. 2012), with the strict cutoff of p values <0.01 and FDR <0.05.

Reference

Andrews S. 2010. *FastQC: a quality control tool for high throughput sequence data*. Babraham Bioinformatics Web site.

Brelsford A, Lavanchy G, Sermier R, Rausch A, Perrin N. 2017. Identifying homomorphic sex chromosomes from wild-caught adults with limited genomic resources. *Mol Ecol Resour* **17**: 752–759.

Castelo AT, Martins W, Gao GR. 2002. TROLL—tandem repeat occurrence locator. *Bioinformatics* **18**: 634-636.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**: 80–92.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. 2017. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**: 92–95.

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**: 95–98.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci USA* **117**: 9451–9457.

Goel M, Schneeberger K. 2022. plotsr: visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics* **38**: 2922–2926.

Goel M, Sun H, Jiao W-B, Schneeberger K. 2019. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol* **20**: 277.

374 Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database
375 management tool for second-generation genome projects. *BMC Bioinformatics*
376 **12**: 491.
377 Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S-Y, Freimer NB, Sabatti C, Eskin
378 E. 2010. Variance component model to account for sample structure in
379 genome-wide association studies. *Nat Genet* **42**: 348–354.
380 Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low
381 memory requirements. *Nat Methods* **12**: 357–360.
382 Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis
383 ER, Ding L, Wilson RK. 2012. VarScan 2: Somatic mutation and copy
384 number alteration discovery in cancer by exome sequencing. *Genome Res* **22**:
385 568–576.
386 Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu:
387 scalable and accurate long-read assembly via adaptive *k* -mer weighting and
388 repeat separation. *Genome Res* **27**: 722–736.
389 Langfelder P, Horvath S. 2007. Eigengene networks for studying the relationships
390 between co-expression modules. *BMC Syst Biol* **1**: 54.
391 Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with
392 BWA-MEM. http://arxiv.org/abs/1303.3997.
393 Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*
394 **34**:3094−100
395 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler
396 transform. *Bioinformatics* **25**: 1754–1760.
397 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G,
398 Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The
399 Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:
400 2078–2079.
401 Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program
402 for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923–930.
403 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and
404 dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 1–21.
405 Majoros WH, Pertea M, Salzberg SL. 2004. TigrScan and GlimmerHMM: two open
406 source ab initio eukaryotic gene-finders. *Bioinformatics* **20**: 2878-2879.
407 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella
408 K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit:

409 a MapReduce framework for analyzing next-generation DNA sequencing data.
410 *Genome Res* **20**: 1297–1303.
411 Montgomery SH, Mank JE. 2016. Inferring regulatory change from gene expression:
412 the confounding effects of tissue scaling. *Mol Ecol* **25**: 5114–5128.
413 Morris J, Darolti I, Bloch N, Wright A, Mank J. 2018. Shared and Species-Specific
414 Patterns of Nascent Y Chromosome Evolution in Two Guppy Species. *Genes*
415 **9**: 238.
416 Nurk S, Walenz BP, Rhie A, Vollger MR, Logsdon GA, Grothe R, Miga KH, Eichler
417 EE, Phillippy AM, Koren S. 2020. HiCanu: accurate assembly of segmental
418 duplications, satellites, and allelic variants from high-fidelity long reads.
419 *Genome research* **30**: 1291–1305.
420 Phimphan S, Aiumsumang S. 2021. Chromosomal characteristics of Taolor's stream
421 frog (*Limnonectes taylori*)(Amphibia, Anura) from Thailand. *The Nucleus* **64**:
422 129–133.
423 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J,
424 Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for
425 whole-genome association and population-based linkage analyses. *Am J Hum*
426 *Genet* **81**: 559–575.
427 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing
428 genomic features. *Bioinformatics* **26**: 841-842.
429 Quinn A, Juneja P, Jiggins FM. 2014. Estimates of allele-specific expression in
430 *Drosophila* with a single genome sequence and RNA-seq data. *Bioinformatics*
431 **30**: 2603–2610.
432 Rahman A, Hallgrímsdóttir I, Eisen M, Pachter L. 2018. Association mapping from
433 sequencing reads using k-mers. *eLife* **7**: e32920.
434 Rastas P. 2017. Lep-MAP3: robust linkage mapping even for low-coverage whole
435 genome sequencing data. *Bioinformatics* **33**: 3726-3732.
436 Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for
437 differential expression analysis of digital gene expression data. *Bioinformatics*
438 **26**: 139–140.
439 Siberchicot A, Bessy A, Guéguen L, Marais GA. 2017. MareyMap Online: A
440 User-Friendly Web Application and Database Service for Estimating
441 Recombination Rates Using Physical and Genetic Maps. *Genome Biol Evol* **9**:
442 2506–2509.

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

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res* **34**: W435-439.

Stevenson KR, Coolon JD, Wittkopp PJ. 2013. Sources of bias in measures of allele-specific expression derived from RNA-seq data aligned to a single reference genome. *BMC genomics* **14**: 1–13.

Utturkar SM, Klingeman DM, Hurt RA, Brown SD. 2017. A Case Study into Microbial Genome Assembly Gap Sequences and Finishing Strategies. *Front Microbiol* **8**: 1272.

Yu G, Wang L-G, Han Y, He Q-Y. 2012. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS* **16**: 284-287.

Zimmer F, Harrison PW, Dessimoz C, Mank JE. 2016. Compensation of Dosage-Sensitive Genes on the Chicken Z Chromosome. *Genome Biol Evol* **8**: 1233-1242.