





















# A chromosome-level genome assembly of the European green toad (*Bufo viridis*)

Patrik Rödin-Mörch <sup>1</sup>, Ignas Bunikis <sup>2</sup>, Eunkyong Choi <sup>3</sup>, Claudio Ciofi <sup>4</sup>, Genevieve Diedericks <sup>5</sup>, Maria Angela Diroma <sup>4</sup>, Elisabet Einarsdóttir <sup>3</sup>, Kristofer Försäter <sup>6</sup>, Julia Heintz <sup>2</sup>, Linnea Jonsäll <sup>2</sup>, Henrik Lantz <sup>7,8</sup>, Anssi Laurila <sup>1</sup>, Henrique G. Leitão <sup>5</sup>, Mai-Britt Mosbech <sup>2</sup>, Chiara Natali <sup>4</sup>, Remi-André Olsen <sup>9</sup>, Olga Vinnere Pettersson <sup>2</sup>, Lucile Soler <sup>7,8</sup>, Hannes Svoldal <sup>5,10</sup>, Estelle Proux-Wéra <sup>9</sup>, Jacob Höglund <sup>1,\*</sup>

<sup>1</sup>Department of Ecology and Genetics/Animal Ecology, Uppsala University, Norbyvägen 18D, Uppsala 752 36, Sweden

<sup>2</sup>Science for Life Laboratory, Department of Immunology, Genetics and Pathology, Uppsala University, National Genomics Infrastructure hosted by SciLifeLab, Box 518, Uppsala 751 08, Sweden

<sup>3</sup>Science for Life Laboratory, Department of Gene Technology, KTH-Royal Institute of Technology, Solna SE-171 21, Sweden

<sup>4</sup>Department of Biology, University of Florence, Sesto Fiorentino (FI) 50019, Italy

<sup>5</sup>Evolutionary Ecology Group, Department of Biology, University of Antwerp, Antwerp 2020, Belgium

<sup>6</sup>Foundation Nordens Ark, Åby Säteri, Hunnebostrand SE-456 93, Sweden

<sup>7</sup>Department of Medical Biochemistry and Microbiology, Uppsala University, NBIS—National Bioinformatics Infrastructure Sweden, Box 582, Uppsala 751 23, Sweden

<sup>8</sup>Department of Cell and Molecular Biology, Uppsala University; NBIS—National Bioinformatics Infrastructure Sweden, Box 596, Uppsala 751 24, Sweden

<sup>9</sup>Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Solna 171 65, Sweden

<sup>10</sup>Naturalis Biodiversity Center, Leiden 2333, Netherlands

\*Corresponding author: Jacob Höglund, Department of Ecology and Genetics/Animal Ecology, Uppsala University, Norbyvägen 18D, Uppsala 752 36, Sweden. Email: jacob.hoglund@ebc.uu.se

The European green toad (*Bufo viridis*) is geographically widely distributed. While the species global conservation status is labeled as of least concern by the IUCN, it is declining in many parts of its range where populations are fragmented and isolated. A high-quality reference genome is an important resource for conservation genomic researchers who are trying to understand and interpret the genomic signals of population decline, inbreeding, and the accumulation of deleterious mutations. Here, we assembled and annotated a chromosome-level reference genome for *B. viridis* as part of the European Reference Genome Atlas pilot project. The genome assembly, with a size of ~3.89 Gb consists of 11 chromosomes and an additional 2,096 unplaced scaffolds. The final assembly had a scaffold N50 value of 478.39 Mb and covered 90.4% single copy tetrapod orthologs, and 46.7% repetitive elements. Finally, a total of 23,830 protein-coding genes matching a known gene, together with 56,974 mRNAs were predicted. This high-quality reference genome will benefit amphibian evolutionary genomics research and enable conservation genetic studies to inform practical conservation work on this species.

**Keywords:** amphibians; *Bufo viridis*; chromosome-level assembly; genome annotation

## Introduction

Amphibians are considered to be the most threatened group of vertebrates, with more than 40% of the assessed amphibian species being threatened with extinction as a result of infectious disease, habitat loss, and climate change (Luedtke et al. 2023). Conservation genomics provides an approach for understanding basic biological properties relevant for the conservation of threatened taxa such as genetic diversity, population structure, amount of inbreeding, and genetic load (Formenti et al. 2022; Theissinger et al. 2023). As such, a high-quality reference genome is a valuable asset facilitating such work.

Among vertebrates, amphibians have the largest variance in genome size, ranging from 140 pg to as small as 0.95 pg (Liedtke et al. 2018). In anurans, the order consisting of frogs and toads, evidence suggests that the wide range in genome sizes is largely due

to the recent accumulation of transposable elements, as well as the failure to remove ancient transposable element insertions (Zuo et al. 2023). Due to their genome size and complexity, amphibians have until fairly recently been largely overlooked by large scale initiatives to produce reference genomes, with only a small fraction of the ~8,600 described amphibian species having a chromosome-level genome assembly available (Kosch et al. 2024).

The green toad genus *Bufo* is geographically wide-spread and contains up to 15 recognized lineages. These include lineages that are diploid or allopolyploids (the latter having formed through hybrid speciation and that occurs to the eastern part of the *Bufo* range) (Dufresnes et al. 2019). The *Bufo viridis* subgroup occurs from Western Europe and the Northern part of the Mediterranean region to the Ural Mountains in the east, and is globally listed as of least concern but decreasing throughout its range (IUCN 2023). For example, in Sweden *B. viridis* has historically been

declining and is considered to be the most vulnerable amphibian species in the country (Höglund et al. 2022), with its entire range confined to fragmented populations along the southern coast or on rocky islands off the coast, which are all geographically isolated from the European mainland populations.

Here, we present a chromosome-level, annotated reference genome for the pilot species *B. viridis* in the European Reference Genome Atlas (ERGA) pilot project (Mc Cartney et al. 2024). This reference genome will add to the growing number of amphibian reference genomes that will serve as useful tools for comparative and eco-evolutionary genomics studies, and to be used directly or indirectly to aid conservation practitioners working to safeguard amphibian diversity.

## Methods

### DNA and RNA extractions, sequencing library preparations, and sequencing

One male *B. viridis* from the ex situ breeding program at Nordens Ark (<https://nordensark.se/bevarande/bevarande-i-sverige/gron-flackig-padda/>) was humanely euthanized and dissected by a veterinarian on the 2021 May 11 (Supplementary Fig. S1) with the approval from the Ethical committee for animal experiments in Göteborg (Dnr 5.8.18-06182/2021). HMW-DNA for PacBio Sequel II sequencing, frozen ground tissue for Omni-C and RNA for Illumina RNA-seq and PacBio Iso-Seq using 7 tissues (skin, spleen, breast muscle, lung, heart, liver, and kidney) were sequenced (See Supplementary Material online for details).

### Genome assembly

The PacBio HiFi reads were investigated for potential contamination using Mash Screen v2.3 (Ondov et al. 2019) against the NCBI RefSeq database (<https://gembox.cbc.umd.edu/mash/refseq.genomes.k21s1000.msh>), and no contamination was detected. Further evaluation of coverage, genome size, and heterozygosity based on the HiFi reads was done using GeneScopeFK (<https://github.com/thegenemyers/GENESCOPE.FK>) using a k-mer size of 31, which is a modified version of GenomeScope v2.0 (Ranallo-Benavidez et al. 2020).

The PacBio HiFi <https://www.doi.org/10.5281/zenodo.3552717> reads were assembled using IPA v.1.5.0 (<https://github.com/PacificBiosciences/pbipa>). The initial assembly, the purged assembly and the scaffolded assembly were evaluated with BUSCO v5.4.6 (Manni et al. 2021) and K-mer frequency histogram and completeness statistics produced using MERQUERY.FK (<https://github.com/thegenemyers/MERQUERY.FK>), as well as general assembly summary statistics using QUAST v.5.0.2 (Gurevich et al. 2013) (Table 1). The HiFi assembly was purged of duplicates using the IPA internal purge\_dups v.1.2.5 (Guan et al. 2020). Omni-C reads were mapped to the contig assembly using BWA v. 0.7.17-r1188 (Li and Durbin 2009), with parameters  $-SSP$  and  $-T0$ . The alignments were then filtered and processed using pairtools v.0.3.0 (Open2C; Abdennur et al. 2023). The contig assembly was subsequently scaffolded using YaHS v.1.2.a.1.patch (Zhou et al. 2022). Manual curation of the scaffolded assembly was performed in Juicebox v.2.16.00 (Durand et al. 2016), and finalized using the juicer post command in YaHS.

### Genome annotation

Using the Nextflow pipeline AnnotationPreprocessing (<https://github.com/NBISweden/pipelines-nextflow>), an initial annotation preprocessing step was performed by assessing assembly statistics, presence of missing and ambiguous nucleotides, gene

**Table 1.** Summary statistics for the final assembly, gene, and repeat annotation.

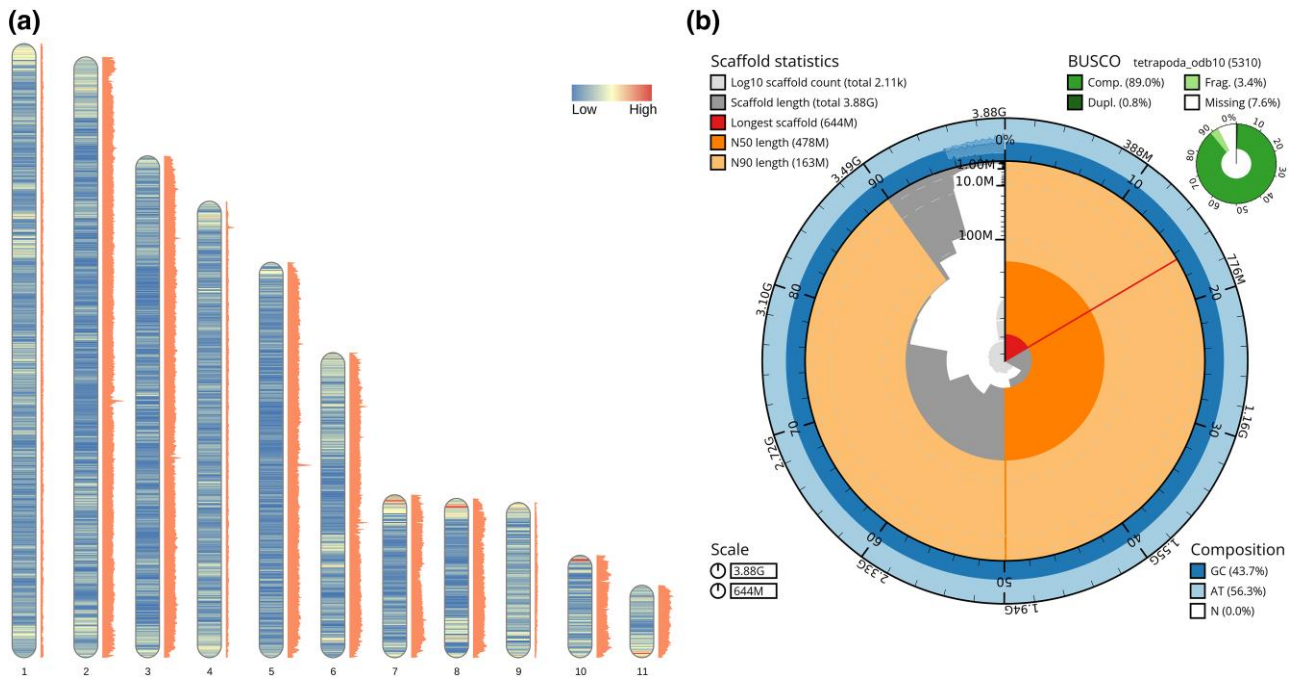
| Assembly statistics              |           |
|----------------------------------|-----------|
| Contig L50                       | 168       |
| Contig N50 (Mb)                  | 6.73      |
| Longest contig (Mb)              | 37        |
| Scaffold L50                     | 3         |
| Scaffold N50 (Mb)                | 478.39    |
| Longest scaffold (Mb)            | 643.6     |
| Number of scaffolds              | 2,107     |
| Annotation statistics            |           |
| Number of protein-coding genes   | 31,608    |
| Number of mRNA                   | 69,426    |
| Number of single exon genes      | 7,746     |
| Average number of exons per mRNA | 7.9       |
| Average exon length              | 310       |
| Average intron length            | 4,984     |
| Repeat statistics                |           |
| DNA_transposons                  | 366,392   |
| LINE                             | 178,322   |
| LTR                              | 110,821   |
| LOW_complexity                   | 183,696   |
| RC                               | 1,348     |
| Retroposon                       | 41        |
| SINE                             | 2,398     |
| Satellite                        | 3,213     |
| Simple_repeat                    | 1,016,805 |
| Unknown                          | 2,005,319 |
| rRNA                             | 1,845     |
| scRNA                            | 1         |
| snRNA                            | 18,749    |
| srpRNA                           | 1         |
| tRNA                             | 949       |

completeness, presence of organelles, and distance to other annotated species.

Curated protein sequences were downloaded from the UniProt database (Magrane and UniProt Consortium 2011, downloaded on 2022-12; 568,363 proteins), including the amphibian UniProt database (38,940 proteins, uniprot\_amphibia\_PE12.fasta). A species specific repeat library was created using RepeatModeler v.2.0.2a (Smit, Hubley and Green 2010; Flynn et al. 2020) and RepeatRunner (Yandell 2006). The identified repeats (excluding transposons), were then matched against the collected protein set in order to exclude nucleotide motifs originating from low-complexity coding sequences. Repetitive sequences were then masked in the genome using the identified repeats with RepeatMasker v.4.1.2\_p1 (Smit and Hubley 2010). Illumina RNA-seq data were aligned and assembled using an in-house pipeline from National Bioinformatics Infrastructure Sweden (NBIS) ([https://github.com/NBISweden/pipelines-nextflow/tree/master/subworkflows/transcript\\_assembly](https://github.com/NBISweden/pipelines-nextflow/tree/master/subworkflows/transcript_assembly)) that includes fastq preprocessing in fastp v. 0.23.2 (Chen et al. 2018), alignment of reads to the <https://www.doi.org/10.5281/zenodo.3552717> reference genome using HISAT v. 2.1.0 (Kim et al. 2015) and transcript assembly using StringTie v. 2.2.1 (Pertea et al. 2015).

The final assembly was annotated using the MAKER pipeline v.3.01.02 (Holt and Yandell 2011). First, evidence-based-annotations were constructed using both Illumina RNA-seq and Iso-seq assembled transcripts and Uni-Prot reference proteins. Second, using the initial MAKER evidence file, ab initio training was performed with Augustus v.3.3.3 (Stanke et al. 2006) using an NBIS in-house pipeline ([https://github.com/NBISweden/pipelines-nextflow/tree/master/subworkflows/abinitio\\_training](https://github.com/NBISweden/pipelines-nextflow/tree/master/subworkflows/abinitio_training)).

After running MAKER on a combination of evidence annotation together with curated ab initio profiles, a statistical evaluation of the final annotation was performed using the script



**Fig. 1.** a) The 11 putative chromosomes order by size with gene density represented as an overlaid heatmap and density of repeats in orange on the right side of each chromosome RIdiagram v.0.2.2 (Hao et al. 2020). The largest chromosomes have a total length of 644 Mb. b) snail plot produced in BlobToolKit (Challis et al. 2020) showing assembly statistics and BUSCO scores for the final scaffolded assembly.

agat\_sp\_statistics.pl from the AGAT package v.0.8.0 (Dainat 2021) followed by complementing the final evidence-based annotation with the ab initio genes missing in the evidence-based build with agat\_sp\_complement\_annotations.pl from the same toolkit.

Using BLAST v. 2.9.0 (Altschul et al. 1990), searching against the Uniprot/Swissprot databases, with a maximum e-value cutoff of  $1e-6$ , and Interproscan v. 5.59–91.0 (Hunter et al. 2012) functional annotation was performed on the translated CDS features of each coding transcript using the NBIS in house pipeline functional\_annotation ([https://github.com/NBISweden/pipelines-nextflow/tree/master/subworkflows/functional\\_annotation](https://github.com/NBISweden/pipelines-nextflow/tree/master/subworkflows/functional_annotation)). Furthermore, tRNAs were predicted using tRNAscan-SE v.1.3.1 (Lowe and Eddy 1997) and other ncRNAs were predicted using curated Rfam covariance models (Nawrocki et al. 2014) with the infernal package v.1.1.2 (Nawrocki and Eddy 2013).

## Results and discussion

From an adult male *B. viridis* (Supplementary Fig. S1), we generated 149.53 Gb PacBio HiFi, 468.78 Gb Omni-C and 12.65 Gb Iso-Seq sequencing reads together with 32.56 Gb Illumina RNA-seq data from 7 tissues (Supplementary Table S1). Using Genomescope2 with a k-mer length of 31, the predicted genome size was ~3.9 Gb with 0.27% heterozygosity (Supplementary Fig. S2). After purging of the initial HiFi assembly (Supplementary Fig. S3 and Table S2), we obtained a completeness score of 99.97 and QV score of 52.6 from MERQURY.FK (Supplementary Table S3). The final scaffolded assembly had a total size of ~3.89 Gb and consists of 2,107 scaffolds (Supplementary Table S4). The longest scaffold is ~644 Mb and the assembly has a scaffold N50 of ~478 Mb (Table 1), with 43% GC content (Supplementary Table S2). Eleven of the 2,107 scaffolds were considerably larger than the others (Supplementary Fig. S4) suggesting that they correspond to the 11 chromosomes. This is in accordance with other published true toad genome assemblies (*Bufo gargarizans*: Lu et al. 2021, *Bufo bufo*: Streicher et al. 2021).

BUSCO scores using the tetrapoda\_odb10 dataset containing 5,310 genes, for the final scaffolded assembly results in 91.3% complete tetrapod BUSCOs, out of which 90.4% are complete single copy and 0.9% complete duplicated BUSCOs (Supplementary Table S5).

A total of 46.7% (~1.81 Gb) of the genome was repeat masked with the combined repeat annotations from RepeatModeler v.2.0.2a (Smit and Hubley 2010; Flynn et al. 2020) and RepeatRunner (Yandell 2006) (Supplementary Table S7). The majority of the repeats were classified as unknown the rest consisted mostly of simple repeats, DNA transposons, LINEs, and LTRs (Table 1). This distribution of repeat families agrees with the expected distribution in Anuran genomes (Zuo et al. 2023).

The initial run of the MAKER software failed for scaffolds (chromosomes) 1, 4, and 11 using the species-specific repeat library for the toad. For those chromosomes, the built-in repeat library in MAKER was instead used in separate runs, and the 2 runs with the species-specific and default repeat library were then merged. The relative repeat densities across chromosomes 1, 4, and 11 suggest that the reason behind this might be that the species-specific repeat identification and masking for those chromosomes might not have been entirely optimal together with the large size of the chromosomes (Fig. 1a).

The structure of a total of 31,608 protein-coding genes was annotated, with an average exon and intron length of 310 and 4,984 bp, respectively (Table 1). The annotated gene set covers ~87% of complete tetrapod BUSCOs (Supplementary Table S6). Due to ab initio predictions showing elevated numbers of duplicated genes and fewer mRNAs, this final annotation was built based on evidence facilitated by Illumina RNA-seq and 219,633 full length cDNAs from Iso-Seq and complemented with ab initio predictions for loci that was missing in the initial evidence-based annotation. Furthermore, an elevated number of single exon genes without annotation was found and subsequently removed in the final build. In total 31,356 genes and 68,100 mRNAs could

be linked to functional and protein domains (Supplementary Table S8), out of which 23,830 genes and 56,974 mRNAs could be assigned a name (e-value < 1e-6) by BLAST searches against the Uniprot/Swissprot database. Furthermore, a total of 5,922 tRNAs and 6,112 exons were predicted by tRNAscan-SE v.1.3.1 (Lowe and Eddy 1997) with a total exon length of 434 Kb. However, only 23 tRNAs remained after discarding annotations with an annotation edit distance of 1. The ~66 Kb candidate sex determination locus for male-heterogametic XY found in Kuhl et al. 2024, consisting of the gene *bod11* (biorientation of chromosomes in cell division 1 was annotated and located in the region chr1: 579239124-579305844).

Finally, comparing the *B. viridis* assembly to the assemblies of the common toad (*Bufo viridis*, NCBI: GCF\_905171765.1) and the Asiatic toad (*B. gargarizans*: GCF\_014858855.1) reveal that while the genome size for *B. viridis* is smaller than the other 2 species, the contiguity (scaffold N50, Supplementary Table S9) is lower compared to both other assemblies but comparable to the Asiatic toad assembly. This is a result of the final assembly containing a relatively large number of unplaced scaffolds, the majority of which (67%) are under 100 Kb. It should also be noted that the 2 *Bufo* assemblies utilizes additional data for scaffolding (10X Genomics Chromium and BioNano data).

## Data availability

The final assembly and raw data have been deposited at the European Nucleotide Archive under the project accession number: PRJEB71764. Files are also available via the ERGA portal ([https://portal.erga-biodiversity.eu/data\\_portal/Bufotes%20viridis](https://portal.erga-biodiversity.eu/data_portal/Bufotes%20viridis)). Supplemental material available at G3 online.

## Acknowledgments

This study was part of the ERGA-pilot project. Agilent, Dovetail, PacBio, and Illumina kits were wholly or partially sponsored by the manufacturers. The authors would also like to acknowledge support from Antwerp University Hospital Center of Medical Genetics and Jarl Bastianen for access to sequencing RNA-seq library quality control equipment. We are thankful to SciLifeLab, the Knut, and Alice Wallenberg Foundation and the Council for Research Infrastructures and the Swedish Research Council for providing infrastructure for massive parallel sequencing and SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) for providing assistance computational infrastructure. Finally, the authors acknowledge the efforts and support of the ERGA-Pilot Committee: Ann Mc Cartney, Giulio Formenti, and Alice Mouton.

## Funding

This project was a part of a Pilot project run by the European Reference Genome Atlas (ERGA) consortium and included in the Biodiversity Genomics Europe project (Grant no. 101059492) funded by Horizon Europe under the Biodiversity, Circular Economy and Environment call (REA.B.3); co-funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract numbers 22.00173 and 24.00054; and by the UK Research and Innovation (UKRI) under the Department for Business, Energy and Industrial Strategy's Horizon Europe Guarantee Scheme. The Swedish partnership in the Pilot was made available by a grant to SciLifeLab from Uppsala University.

The use of Agilent, Dovetail, and Illumina kits were kindly sponsored by the manufacturers.

## Conflicts of interest

The authors declare no conflict of interest.

## Literature cited

- Open2C; Abdennur N, Fudenberg G, Flyamer IM, Galitsyna AA, Goloborodko A, Imakaev M, Venev SV. 2023. Pairtools: From sequencing data to chromosome contacts [Preprint]. bioRxiv 2023.02.13.528389. <https://doi.org/10.1101/2023.02.13.528389>.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215(3):403–410. doi:10.1016/S0022-2836(05)80360-2.
- Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. 2020. BlobToolKit—interactive quality assessment of genome assemblies. *G3 (Bethesda).* 10(4):1361–1374. doi:10.1534/g3.119.400908.
- Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 34(17):i884–i890. doi:10.1093/bioinformatics/bty560.
- Dufresnes C, Mazepa G, Jablonski D, Caliri Oliveira R, Wenseleers T, Shabanov DA, Auer M, Ernst R, Koch C, Ramírez-Chaves HE, et al. 2019. Fifteen shades of green: The evolution of Bufotes toads revisited. *Mol Phylogenet Evol.* 141:106615. doi:10.1016/j.ympev.2019.106615.
- Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Lieberman Aiden E. 2016. Juicebox provides a visualization system for hi-C contact maps with unlimited zoom. *Cell Syst.* 3(1):99–101. doi:10.1016/j.cels.2015.07.012.
- 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 U S A.* 117(17):9451–9457. doi:10.1073/pnas.1921046117.
- Formenti G, Theissinger K, Fernandes C, Bista I, Bombarely A, Bleidorn C, Ciofi C, Crottini A, Godoy JA, Höglund J, et al. 2022. The era of reference genomes in conservation genomics. *Trends Ecol Evol.* 37(3):197–202. doi:10.1016/j.tree.2021.11.008.
- Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics.* 36(9):2896–2898. doi:10.1093/bioinformatics/btaa025.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics.* 29(8):1072–1075. doi:10.1093/bioinformatics/btt086.
- Hao Z, Lv D, Ge Y, Shi J, Weijers D, Yu G, Chen J. 2020. Rideogram: Drawing SVG graphics to visualize and map genome-wide data on the ideograms. *PeerJ Comput Sci.* 6:e251. <http://doi.org/10.7717/peerj-cs.251>.
- Höglund J, Bolender L, Cortazar-Chinarro M, Meurling S, Laurila A, Hermaniuk A, Dufresnes C. 2022. Low neutral and immunogenetic diversity in northern fringe populations of the green toad *Bufo viridis*: implications for conservation. *Conserv Genet.* 23(1):139–149. doi:10.1007/s10592-021-01407-5.
- Holt C, Yandell M. 2011. MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics.* 12:491. doi:10.1186/1471-2105-12-491.
- Dainat J. AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format (Version v0.8.0). Zenodo. <https://www.doi.org/10.5281/zenodo.3552717>
- Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Bernard T, Binns D, Bork P, Burge S, et al. 2012. InterPro in 2011:



- New developments in the family and domain prediction database. *Nucleic Acids Res.* 40:D306–D312. doi:[10.1093/nar/gkr948](https://doi.org/10.1093/nar/gkr948).
- IUCN. The IUCN Red List of Threatened Species. Version 2023-1; 2023 <https://www.iucnredlist.org>. Accessed on 04-06-2024.
- Kim D, Langmead B, Salzberg SL. 2015. HISAT: A fast spliced aligner with low memory requirements. *Nat Methods.* 12(4):357–360. doi:[10.1038/nmeth.3317](https://doi.org/10.1038/nmeth.3317).
- Kosch TA, Crawford AJ, Lockridge Mueller R, Wollenberg Valero KC, Power ML, Rodríguez A, O'Connell LA, You ND, Skerratt LF. 2024. Comparative analysis of amphibian genomes: An emerging resource for basic and applied research. *Molec Ecol Resour.* 25(1):e14025. <https://doi.org/10.1111/1755-0998.14025>.
- Kuhl H, Tan Hui W, Klopp C, Kleiner W, Koyun B, Ciropac M, Feron R, Knytl M, Kloas W, Scharl M, et al. 2024. A candidate sex determination locus in amphibians which evolved by structural variation between X- and Y-chromosomes. *Nat Commun.* 15(1):4781. doi:[10.1038/s41467-024-49025-2](https://doi.org/10.1038/s41467-024-49025-2).
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 25(14):1754–1760. doi:[10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324).
- Liedtke HC, Gower DJ, Wilkinson M, Gomez-Mestre I. 2018. Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate. *Nat Ecol Evol.* 2(11):17921799. doi:[10.1038/s41559-018-0674-4](https://doi.org/10.1038/s41559-018-0674-4).
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25(5):955–964. doi:[10.1093/nar/25.5.955](https://doi.org/10.1093/nar/25.5.955).
- Lu B, Jiang J, Wu H, Chen X, Song X, Liao W, Fu J. 2021. A large genome with chromosome-scale assembly sheds light on the evolutionary success of a true toad (*Bufo gargarizans*). *Mol Ecol Resour.* 21(4):1256–1273. doi:[10.1111/1755-0998.13319](https://doi.org/10.1111/1755-0998.13319).
- Luedtke JA, Chanson J, Neam K, Hobin L, Maciel AO, Catenazzi A, Borzée A, Hamidy A, Aowphol A, Jean A, et al. 2023. Ongoing declines for the world's amphibians in the face of emerging threats. *Nature.* 622(7982):308–314. doi:[10.1038/s41586-023-06578-4](https://doi.org/10.1038/s41586-023-06578-4).
- Magrane M; UniProt Consortium. 2011. UniProt knowledgebase: A hub of integrated protein data. *Database (Oxford).* 2011:bar009. doi:[10.1093/database/bar009](https://doi.org/10.1093/database/bar009).
- Manni M, Berkeley MR, Seppely M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol.* 38(10):4647–4654. doi:[10.1093/molbev/msab199](https://doi.org/10.1093/molbev/msab199).
- Mc Cartney AM, Formenti G, Mouton A, De Panis D, Marins LS, Leitão HG, Diedericks G, Kirangwa J, Morselli M, Salces-Ortiz J, et al. 2024. The European reference genome atlas: piloting a decentralised approach to equitable biodiversity genomics. *NPJ Biodivers.* 3(1):28. doi:[10.1038/s44185-024-00054-6](https://doi.org/10.1038/s44185-024-00054-6).
- Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, Floden EW, Gardner PP, Jones TA, Tate J, et al. 2014. Rfam 12.0: Updates to the RNA families database. *Nucleic Acids Res.* 43(Database issue):D130–D137. doi:[10.1093/nar/gku1063](https://doi.org/10.1093/nar/gku1063).
- Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics.* 29(22):2933–2935. doi:[10.1093/bioinformatics/btt509](https://doi.org/10.1093/bioinformatics/btt509).
- Ondov B, Starrett G, Sappington A, Kostic A, Koren S, Buck CB, Phillippy AM. 2019. Mash screen: High-throughput sequence containment estimation for genome discovery. *Genome Biol.* 20(1):232. doi:[10.1186/s13059-019-1841-x](https://doi.org/10.1186/s13059-019-1841-x).
- Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-Seq reads. *Nat Biotechnol.* 33(3):290–295. doi:[10.1038/nbt.3122](https://doi.org/10.1038/nbt.3122).
- Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun.* 11(1):1432. doi:[10.1038/s41467-020-14998-3](https://doi.org/10.1038/s41467-020-14998-3).
- Smit A, Hubley R. 2010. RepeatModeler open-1.0.
- Smit A, Hubley R, Green P. 2010. RepeatMasker open-3.0.
- Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* 34(Web Server issue):W435–W439. doi:[10.1093/nar/gkl200](https://doi.org/10.1093/nar/gkl200).
- Streicher JW; Wellcome Sanger Institute Tree of Life programme; Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective; Tree of Life Core Informatics collective; Darwin Tree of Life Consortium. 2021. The genome sequence of the common toad, *Bufo bufo* (Linnaeus, 1758) [version 1; peer review: 2 approved]. *Wellcome Open Res.* 6:281. doi:[10.12688/wellcomeopenres.17298.1](https://doi.org/10.12688/wellcomeopenres.17298.1).
- Theissinger K, Fernandes C, Formenti G, Bista I, Berg PR, Bleidorn C, Bombarely A, Crottini A, Gallo GR, Godoy JA, et al. 2023. How genomics can help biodiversity conservation. *Trends Genet.* 39(7):545–559. doi:[10.1016/j.tig.2023.01.005](https://doi.org/10.1016/j.tig.2023.01.005).
- Yandell M. Comparative genomics library—RepeatRunner. [http://www.yandell-lab.org/repeat\\_runner/index.html](http://www.yandell-lab.org/repeat_runner/index.html). 2006.
- Zhou C, McCarthy SA, Durbin R. 2022. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics.* 39(1):btac808. doi:[10.1093/bioinformatics/btac808](https://doi.org/10.1093/bioinformatics/btac808).
- Zuo B, Nneji LM, Sun YB. 2023. Comparative genomics reveals insights into anuran genome size evolution. *BMC Genomics.* 24(1):379. <https://doi.org/10.1186/s12864-023-09499-8>.

Editor: R. Mallarino