Comparative analysis of amphibian genomes: an emerging resource for basic and applied research

Running title: Comparative analysis of amphibian genomes

Tiffany A. Kosch¹, Andrew J. Crawford², Rachel Lockridge Mueller³, Katharina C. Wollenberg Valero⁴, Megan L. Power⁴, Ariel Rodríguez⁵, Lauren A. O’Connell⁶, Neil D. Young¹, and Lee F. Skerratt¹

¹ Faculty of Science, University of Melbourne, Melbourne, Australia
² Departamento de Ciencias Biológicas, Universidad de los Andes, Bogotá, Colombia
³ Department of Biology, Colorado State University, Colorado, USA
⁴ School of Biology and Environmental Science, University College Dublin, Dublin, Ireland
⁵ Institute of Zoology, University of Veterinary Medicine of Hannover, Hannover, Germany
⁶ Department of Biology, Stanford University, California, USA

Corresponding author: tiffany.kosch@unimelb.edu.au
ABSTRACT

Amphibians are the most threatened group of vertebrates and are in dire need of conservation intervention to ensure their continued survival. They have many unique features including a high diversity of reproductive strategies, permeable and specialized skin capable of producing toxins and antimicrobial compounds, multiple genetic mechanisms of sex determination, and in some lineages even the ability to regenerate limbs and internal organs. Although genomics approaches would shed light on these unique phenotypic traits and aid in conservation management, sequencing and assembly of amphibian genomes has lagged far behind other taxa due to their comparatively large genome sizes. Fortunately, the development of long-read sequencing technologies and initiatives has led to a recent burst of new amphibian genome assemblies. Although growing, the field of amphibian genomics suffers from the lack of annotation resources, tools for working with challenging genomes, and lack of high-quality assemblies in multiple clades of amphibians. Here we analyze 32 publicly available amphibian genomes to evaluate their usefulness for functional genomics analysis. We report considerable variation in assembly quality and completeness, and report some of the highest transposable element and repeat contents of any vertebrate, which is associated with climate. We also provide evidence of conserved genome synteny despite the long divergence times of this group but show that chromosome naming and orientation have been inconsistent across genome assemblies. Additionally, we discuss sequencing gaps in the phylogeny and suggest key targets for future sequencing endeavors. Lastly, we propose increased investment in amphibian genomics research to promote their conservation.

KEYWORDS

Amphibian genomes; comparative genomics; transposable elements; repeat expansion; genome synteny
INTRODUCTION

Amphibians are an ancient lineage of vertebrates that predate amniotes by more than 100 million years. Despite the considerable age of this lineage, amphibians are now the most threatened group of vertebrates with more than 40% of species threatened by factors such as habitat change, disease, and over-exploitation (IUCN, 2022; Scheele et al., 2019). Notably, many of these threats are hard to reverse, suggesting that novel approaches that utilize genomic resources may lead to improved management decisions for some of the most endangered taxa (Kosch et al., 2022; Scheele et al., 2014).

We are only just beginning to understand the genetic basis of many of the unique features of amphibians. Amphibians exhibit a high diversity of reproductive strategies including biphasic and direct development, uniparental and biparental care, mouth and gastric brooding, and foam nesting (Brown et al., 2010; Nunes-de-Almeida et al., 2021; Schulte et al., 2020). They also have specialized skin capable of producing complex compounds of interest for drug discovery for the development of antimicrobial drugs and analgesics (Daly et al., 2000; De Angelis et al., 2021; Liu et al., 2020).

Amphibians occur across habitat types from rainforests to deserts, freshwater streams to salt marshes, and tropical to arctic climates (Duellman, 1999), but it is unclear how this ecological diversity is reflected in genome composition. Salamanders are an important resource for transplant and regeneration research due to their ability to regenerate limbs and internal organs (Elewa et al., 2017; Nowoshilow et al., 2018). Amphibians also have many of the same immune components of mammals making them an important model resource for immunology (Paiola et al., 2023; Robert, 2020).

Despite the obvious value of amphibian genomes for research on ecology, evolution, medicine, and improving their conservation, until recently, the generation of amphibian reference genomes has
been markedly slower than other vertebrates (Hotaling et al., 2021a). This lag can be attributed to high costs and the computational challenges of assembling their often large and complex genomes (Sun et al., 2020). Recent advances in sequencing technologies such as long read sequencing and assembly algorithms that incorporate hybrid approaches have circumvented many of these challenges leading to a surge of high quality, chromosome-level reference genomes. The next challenge will be developing the tools for annotation and comparative analyses of these large genomes.

In this study, we provide a state-of-the-field synthesis of amphibian reference genome assemblies by analyzing 32 publicly available amphibian genomes. We evaluate assembly quality, sequencing technology, gene completeness, transposable element and repeat content and its ecological correlates, taxonomic representation, and synteny.

**MATERIALS AND METHODS**

**Genomes**

A search conducted on the NCBI genome website using search term “amphibians” conducted on November 1, 2021 revealed there were 39 amphibian genomes from 30 species. All genome files in fasta format were downloaded for assessment. One genome was selected for each species. If there was more than one draft of a genome, the most recent draft and/or the primary haplotype was selected. In cases where there were multiple versions sequenced by different groups, the best genome was selected by lowest scaffold number. We excluded the *Ranitomeya imitator* genome as the genome was later retracted from NCBI. Entire genomes (including uncharacterized contigs but excluding mitochondrial genomes) were used for assessment unless indicated otherwise.
Genome databases NCBI Genomes, NCBI RefSeq (O'Leary et al., 2016), Ensembl (Cunningham et al., 2022), UCSC Genome Browser (Lee et al., 2022), and Genomes on a Tree (GoaT) (Sotero-Caio et al., 2021) were searched for information on the 34 amphibian species with reference genomes including chromosome number, annotation data, proteome availability, C-value, and sequencing technology. Read length was classified as “short” for Illumina sequencing and “long” for PacBio and Nanopore technologies.

A search for amphibian proteome datasets on NCBI RefSeq (O'Leary et al., 2016), Ensembl (Cunningham et al., 2022), and UCSC Genome Browser (Lee et al., 2022) databases on June 24, 2022 revealed 11 proteomics datasets.

A search of the NCBI Organelle database on 15, February 2023 using search term “amphibian” resulted in 353 mitochondrial genomes belonging to 345 species (Table S11). Seventeen mitochondrial genomes overlapped with the amphibian nuclear genomes analyzed in this study.

Reference genome availability summary

The GoaT online database (Sotero-Caio et al., 2021) was searched on January 6, 2023 to summarize genomes in progress or publicly available using the search terms “tax_tree(Amphibia) AND tax_rank(species) AND sequencing_status=in_progress” or “tax_tree(Amphibia) AND tax_rank(species) AND sequencing_status=insdc_open”. The same search terms were used to summarize publicly available genomes for mammals, birds, and non-avian reptiles with the “tax_tree” search term replaced by appropriate Class.
**Genome quality analyses**

Genome quality assessment was performed with BBMap (v. 38.76). Benchmarking Universal Single-Copy Orthologs (BUSCO) were summarized with the BUSCO tool (v. 5.1.2) (Manni et al., 2021) using the OrthoDB Tetrapoda ortholog library (v. odb10) (Kriventseva et al., 2018) (N=5310 orthologs). Percentage of the genome assembled to chromosomes was calculated with a custom bash script.

**Phylogenetic tree**

A species to family correspondence table was obtained from Jetz and Pyron (2018) (https://vertlife.org/files_20170703/) and was filtered to include only the species with the longest nucleotide sequence per family. This taxa subset was used to obtain a subset of 100 phylogenetic trees from the posterior distribution of the Jetz and Pyron (2018) dataset, as available from http://vertlife.org/phylosubsets. A consensus tree from these 100 trees was then obtained using treannotator (Drummond & Rambaut, 2007). The species names of the tree tips were then substituted with the corresponding family names using the “sub.taxa.label” function in the phylotools package (https://github.com/helixcn/phylotools) in R with the aid of the species to family correspondence table, which was updated with the most recent classification available in AmphibiaWeb https://amphibiaweb.org and the Amphibian Species of the World database https://amphibiaworld.amnh.org/.

**Repeat modelling and annotation**

Repeats were de novo modelled with RepeatModeler (v. 2.0.3) (Flynn et al., 2020). Genomes were then annotated using RepeatMasker (v. 4.1.2-p1) (Smit et al., 2013) with a concatenated library of genome-specific repeats generated from RepeatModeler and the Dfam amphibian repeat library (v. Dfam.h5) (Storer et al., 2021). Before annotation, any previous soft masking of the genomes was reversed. The results were summarized using a custom bash script.
**Ecological correlates of transposable element content**

On the basis of Global Biodiversity Information Facility (GBIF) data (https://www.gbif.org/; accessed May 2023), MaxEnt habitat suitability-weighted means of a range of bioclimatic variables were obtained for the 32 amphibian species. As previous studies have explored the relationship between amphibian genome size and environmental variables (Liedtke et al., 2022), here we focused on the relationship between bioclimatic variables and amphibian transposable elements. Influence of these bioclimatic variables on transposable element content was modelled using Bayesian mixed effect models (Hadfield, 2010), corrected for body size and phylogenetic signal (see Supplementary Methods for further information).

**Synteny analysis**

Synteny of BUSCO genes for chromosome level assemblies was analyzed with R Package GENESPACE (v. 0.9.4) (Lovell et al., 2022), which uses OrthoFinder (v. 2.5.4) (Emms & Kelly, 2019) to infer orthology. Synteny was analyzed using BUSCO “full_table.tsv” results files that were reformatted for GENESPACE input using a custom bash script. Synteny plots were generated for all chromosome level assemblies, all anuran chromosome level assemblies, and for the three caecilian genomes using the GENESPACE plotting tool “plot_riparianHits”. Chromosomes with reversed orientation compared to the reference genome were inverted to improve visualisation.

**Quantification and statistical analysis**

Regression analyses and Student’s t-tests for comparing genome quality measurements were conducted with the R statistics package (v. 4.1.2) (Team, 2013) in R Studio (v. 2022.02.3) (Team, 2022). Genome quality measures, contigN50, and scaffold count, were log transformed prior to
analysis. R-scripts for statistical analysis and plotting are available on GitHub at


RESULTS

Genome quality

A total of 32 nuclear amphibian genome assemblies were available for our study and were generated with a variety of sequencing technologies, including Illumina (NextSeq, HiSeq), PacBio (RS11, Sequel), and Oxford Nanopore. Sequenced genomes represented 20 of 73 amphibian families with reference genomes distributed unevenly across the phylogeny (Fig. 1). For example, there is only one salamander genome representing the 798 extant species, no genomes representing the microhylids or hyperoliids, yet there are five pipid and four pelobatid genomes (Fig. 1).

Genome assembly length ranged from 0.48 Gb in Scaphiopus couchii to 28.21 Gb in Ambystoma mexicanum and was strongly positively associated with c-value estimates of genome size ($F_{30} = 270.2$, $p < 1 \times 10^{-11}$) (Table 1, Fig. S2). Fifteen of these genomes were assembled to the chromosome level of which the percentage of the genome assigned to chromosomes ranged from 63.88 to 99.96% (Table 1). Percentage of the genome assigned to chromosomes was positively associated with contig N50 ($F_{13} = 10.6, p = 0.006$), number of scaffolds ($F_{13} = 21.35, p < 0.001$) and read length ($t_{28.4} = 3.65, p = 0.001$). There are additionally mitochondrial genome assemblies for 345 species of which 17 had nuclear reference genomes. Eleven of the species with genomes had proteomics data (Table S1).

The quality of the amphibian genomes varied considerably (Table 1). Genomes generated with short-read technologies were of lower quality than long-read or hybrid genome assemblies as indicated by significantly lower contig N50s ($t_{29} = 6.05, p < 0.00001$), percentage of complete Benchmarking
Universal Single-Copy Ortholog (BUSCO) genes ($t_{13,4} = 2.85, p = 0.013$), and higher scaffold numbers ($t_{29} = -4.86, p < 0.0001$).

Contig N50 ranged from 362 bp in *S. couchii* to 22.45 Mb in *Xenopus laevis* with a median of 32.80 Kb. Scaffold count varied considerably from 54 in *X. laevis* to more than four million in *Bombina variegata* with a median of 27.16 Kb (Table 1). Benchmarking Universal Single-Copy Orthologs (BUSCO) scores ranged from 0.7 to 96% completeness (Tables 1, S1; Fig. 2) and were positively associated with contig N50 ($F_{29} = 55.09, p < 0.000001$; Fig. S3) and scaffold count ($F_{29} = 42.98, p < 0.000001$). All genomes had low percentages of duplicate BUSCO genes (< 6%), suggesting they may be diploid except for the known tetraploid species, *X. laevis* (Fig. 2).

The variation we report here in genome quality, contiguity, and completeness may impact the value of the genomes for functional genomics research. However, the improvements in all these measures seen with the utilization of long read technologies or hybrid assemblies suggests that genome quality will continue to improve as these approaches are used more frequently.

**Repeat content**

Overall identified repeat percentage of the genomes ranged from 23% in *Platyplectrum ornatum* to 82% in *Oophaga sylvatica* and was positively associated with genome size ($F_{30} = 5.87, p = 0.022$) (Tables 1; S3). Repeat content varied across genomes with the anuran *O. sylvatica* and the salamander *Ambystoma mexicanum* dominated by Long Terminal Repeats (LTRs), the three caecilians dominated by Long Interspersed Nuclear Elements (LINEs), and the two bufonid anurans dominated by DNA transposons (Fig. 3; Tables S2-S4). The *Ambystoma mexicanum* genome had fewer repeats than might be predicted by its large size (Fig. 3).
The proportion of repeats that could be classified by RepeatMasker ranged from 6.6% in *P. ornatum* to 48.2% in *O. sylvatica* (Table S4) and was positively associated with genome quality measures contigN50 (F29 = 12.8, p = 0.001), scaffold count (F29 = 7.80, p = 0.009), and percent BUSCO complete (F29 = 4.60, p = 0.041). The ability to classify repeats was also positively associated with read length, with longer reads resulting in better classification (t26.241 = 3.57, p = 0.001).

These disparities in repeat percentage and content likely reflect differing evolutionary histories among species, as indicated by three of the four congeneric species pairs in our dataset having similar values (i.e., *Bufo*, *Leptobrachium*, and *Xenopus*; but not *Oophaga*). The differences we observed in *Oophaga pumilio* and *O. sylvatica* are likely due to assembly quality rather than genome content given that these two genomes were sequenced with different technologies and have dramatically different genome qualities (e.g., contig N50s of 5.8 vs. 97.8 Kbp respectively).

**Ecological correlates of transposable element content**

A Bayesian mixed effect modelling approach was employed to examine the relationships between proportion of transposable elements and environmental variables. Controlling for phylogenetic relationships (by estimating Pagel’s lambda, λ) and body size (Spearman correlation with TRANSPOSABLE ELEMENT content p = -0.772, p<0.001) and excluding the three globally invasive species (*Rhinella marina*, *X. laevis*, and *Lithobates catesbeianus*) our analysis revealed a significant influence (pMCMC = 0.004) of Bio18 (precipitation in the warmest quarter) on proportion of total transposable elements (Figs 4, S6; Table S7). Inclusion of these three invasives for a total of 32 species showed additional significant relationships with Bio15 (precipitation seasonality) and elevation (Table S6). Further analysis indicated that the relationship with Bio18 was not specific to a particular class of
transposable elements, such as retroelements or DNA transposons (Tables S7 and S8).

Phylogenetic signal (Pagel’s lambda, \( \lambda \)) was moderate when considering total transposable elements and retroelements (0.440; Tables S6 and S7) and increased when we considered DNA transposons alone (0.603; Table S8).

**Genome synteny**

Genome synteny of BUSCO genes was highly conserved within anurans (Fig. 5) and within caecilians (Fig. S8) but was less conserved across the amphibian orders (Fig. S7). However, chromosome naming was inconsistent across all taxa (Fig. 5, S10-11). For example, *X. tropicalis* chr1 is chr12 in *Leptobrachium ailaonicum* (but not *L. leishanense*) and chr2 in *Bufo bufo* (but not *Bufo gargarizans*) (Fig. 5). Orientation of chromosomes was also inconsistent, including between species of the same genus (e.g., *Bufo, Leptobrachium*) (Fig. 5) and among the three caecilians (Fig. S8). Multiple inversions were evident including between chr3 of pipids (*Xenopus tropicalis* and *Hymenochirus boettgeri*) and other anurans (chromosomes 1, 2, 3, 4, or 10), caecilians (chr3 and chr4/5/6), and even within species of the same genus (chr7 *Bufo gargarizans*, chr 9 *B. bufo*; Figs 5, S10-11). There was also evidence of several chromosomal fissions including the separation of chr1 of *Leptobrachium leishanense* into chr3 and chr6 in *Pyxicephalus adspersus* and into chr3 and chr7 in *Engystomops pustulosus*; however, this chromosome remained mostly intact in the other anuran genomes (Fig. 5). Given that these genomes were generated using different technologies, it is unclear whether or not the differences we observed in synteny are due to true differences or assembly errors.

**DISCUSSION**

In this study we analysed 32 amphibian reference genomes from the public domain to evaluate their content and usefulness for functional genetics research (Fig. 1, Table 1). There are considerably fewer reference genomes for amphibians than exist for birds (N=754), mammals (N=406), and non-
avian reptiles (N=108). This scarcity of reference genomes results in many gaps in genome
representation across the amphibian tree of life including many entirely unrepresented groups and
with only one genome representing the entire order Caudata. The unrepresented families include
many of interest from a conservation perspective due to their high number of IUCN RedList Critically
Endangered species (e.g., Cryptobranchidae, Plethodontidae, Strabomantidae, and Craugastoridae)
(IUCN, 2022). However, our search of the Genomes on a Tree (GoaT) database (Sotero-Caio et al.,
2021) indicated that there are a further 24 amphibian genome assemblies in progress (20 anurans, 4
caudatans) indicating that this resource will be increasing by nearly 70% in the next few years.

The quality and completeness of the genomes in our dataset varied considerably (e.g., Fig. 2). Much
of this variation can be attributed to the sequencing technology used to generate them, with short-
read sequencing approaches resulting in lower completeness and continuity (Fig. S3). These impacts
are a recognized limitation of short-read sequencing and have been reported to impact genome
quality in taxa from insects (Hotaling et al., 2021b) to other vertebrates (Rhie et al., 2021), but have
likely had a disproportionate impact in amphibian genomes due to the difficulty of assembling
genomes with high repeat content (Sun et al., 2020). Fortunately, most ongoing sequencing efforts
now use long-read or hybrid sequencing approaches, which along with improved sequencing
algorithms, should result in higher quality amphibian genomes (Hotaling et al., 2021a; Lawniczak et
al., 2022; Rhie et al., 2021).

Genome quality (i.e., high continuity, contiguity, accuracy, completeness (Rhie et al., 2021)) is critical
for applications such as quantitative genetics where assembly errors can lead to incorrect inferences
in genetic association or genetic prediction. Quality also enhances the usefulness of genomes. For
example, highly contiguous chromosome-level assemblies decrease computational requirements for
downstream analyses such as mapping, variant calling, and alignment (Aganezov et al., 2022).
One of the most intriguing features of amphibian genomes is the huge range they exhibit in genome size (Biscotti et al., 2019). This was exemplified in our dataset where assembly length ranged from 0.48 Gb in *Scaphiopus couchii* to 28 Gb in *Ambystoma mexicanum*. Why gigantic genomes exist in some species, but not others, remains a key evolutionary question (Kapusta et al., 2017; Wang et al., 2021). Explanations include differences in genome-level processes (e.g., insertion and deletion rates) (Frahry et al., 2015; Sun et al., 2012), development (e.g., developmental rate and complexity) (Gregory, 2002; Liedtke et al., 2018), physiology (e.g., water loss) (Johnson et al., 2021), body size (e.g. miniaturization) (Decena-Segarra et al., 2020), and demography (e.g., effective population size) (Liedtke et al., 2018; Lynch & Walsh, 2007) (but see Mohlhenrich & Mueller, 2016). As more amphibian genomes become available, these hypotheses can be more rigorously evaluated.

We report some of the largest estimates of repeat content of any vertebrate (82% in *Oophaga sylvatica* and 77% in *Bufo bufo*), exceeded only by the Australian lungfish at 90% (Meyer et al., 2021). As expected, genome size was correlated with repeat content affirming that much of the variation in amphibian genome size is due to an excess of repeats and transposable elements (Biscotti et al., 2019; Lamichhaney et al., 2021).

In contrast to mammals, whose repeat landscape is mainly dominated by LTR retrotransposons (Platt et al., 2018), amphibian repeat content varied considerably with some species dominated by DNA transposons (as previously reported (Suda et al., 2022), and others by non-LTR retrotransposons including the three caecilian genomes which were dominated by LINEs. This somewhat agrees with genomic data and transcriptomic data from the caecilian *Ichthyophis bannanicus*, where LINEs were the second most abundant type of repeat (26% of the genome) behind DIRS (30%) (Wang et al.,...
2021); this is a higher percentage of LINES than we report in the three caecilian genomes in this study (7 to 14%) (Table S4).

A considerable proportion of the repeats could not be classified. This was likely due to incorrect classification (e.g., genes categorized as repeats) and the lack of good amphibian-specific repeat resources (Ou et al., 2019) for classification via nucleotide sequence homology. The majority of amphibian curated repeat libraries are generated in reference to *Xenopus* species (e.g., Dfam); the large divergence times of this genus from the other amphibian species suggests that it may be a contributing factor to the lack of classification. However, we also report many unclassified repeats in the two *Xenopus* genomes.

The largest genome in our dataset, *A. mexicanum*, had fewer repeats than predicted given its size (Fig S3) (Nowoshilow et al., 2018). This may be due, in part, to the Dfam library used for repeat annotation being anuran-based; however, we did not observe this trend in the three caecilian genomes in our dataset. Also, we performed *de novo* annotation of this genome, which should have captured repetitive elements missing from Dfam. More likely, this low number of repeats reflects low deletion rates and, thus, persistence of repeats in the genome for extremely long periods of time, leading to their mutational decay into unique sequences whose repetitive origin is obscured (Keinath et al., 2015; Novák et al., 2020).

We also show that amphibian species that inhabit wet and warm climates particularly during the summer months have a greater proportion of transposable elements. This observed trend does not appear to be driven by a specific group of transposons suggesting it may be caused by climatic factors. Recent studies indicate that transposable elements exhibit greater activity in hotter climates (Baduel et al., 2021) with an increasing number of studies suggesting increased transposable element
activity contributes to genetic diversification and facilitates species adaptation (Li et al., 2018; Schrader & Schmitz, 2019; Stapley et al., 2015). The pattern observed here likewise suggests the potential for heightened transposable element activity, and may help explain transposable element accumulation and potentially the higher evolutionary rates observed in the genomes of tropical amphibians.

Our study is the first to examine chromosomal synteny across all amphibian orders. We show that overall synteny of amphibian genomes is relatively conserved, particularly within orders (Figs 5 and S8). This aligns with previous results from anurans that reported conserved genome organization in this group (Bredeson et al., 2021; Wu et al., 2022). However, chromosome content and number varied across species, which seems to have been driven by multiple occurrences of chromosomal fusions and fissions (e.g., Fig. 5). Chromosomal rearrangements have occurred throughout vertebrate evolution, including the hypothesized fusion of microchromosomes in the ancestor of tetrapods to create the larger macrochromosomes seen in amphibians and mammals and their subsequent fission to create the microchromosomes of modern birds and non-avian reptiles (Waters et al., 2021).

Some of the structural rearrangements we detected may be due to assembly errors and should be evaluated in future assemblies using long-read scaffolding approaches (e.g., Oxford nanopore sequencing), chromosome conformation capture technologies (e.g., Hi-C), or chromosome mapping approaches (e.g., FISH). We also identified incongruities with chromosome naming and orientation caused by differences in assembly methods. These were apparent even within species of the same genus (e.g., *Bufo*). We suggest potential revisions of existing genome annotations to improve congruity and that future assemblies are curated consistently against high-quality reference genomes (e.g., *Xenopus laevis*).
Conclusions

New sequencing technologies and assembly algorithms have resulted in enough genomes for comparative analyses spanning the amphibian phylogeny. This has already begun to yield important insights on the evolution (Lamichhaney et al., 2021; Wu et al., 2022), development (Schloissnig et al., 2021; Stuckert et al., 2021), sex determination (Hime et al., 2019; Ma & Veltsos, 2021), and unique features (Fischer et al., 2019; Nowoshilow et al., 2018; Seidl et al., 2019) of this interesting group of animals.

The increased availability of amphibian genomes can also aid conservation efforts in this highly threatened group by facilitating research on genome-wide functional diversity, which can be used to inform management decisions such as genetic rescue or targeted genetic intervention for species threatened by habitat loss or chytridiomycosis (Chestnut et al., 2014; Kosch et al., 2022). Additionally, well-annotated genomes can be used to create eDNA assays for population monitoring (Breton et al., 2022; Saeed et al., 2022).

Future research efforts should focus on generating more reference genomes to fill the gaps in the amphibian phylogeny and the identification of advantageous genetic traits against threats. Efforts should also be made to increase the quality of genomes and expand transcriptome and annotation databases. We suggest that these efforts strive to follow the recommendations of initiatives such as the Earth BioGenome Project (Lawniczak et al., 2022), the Darwin Tree of Life Project (Blaxter et al., 2022), and the Threatened Species Initiative (Hogg et al., 2022) to sequence at least one representative from each family to ensure taxonomic coverage. Species selection should prioritize species of interest for understanding valuable functional genetics traits; for example, for the purpose of immunological research to understand disease resistance, or for conservation purposes to enhance fitness.
ACKNOWLEDGEMENTS

This research was supported by The University of Melbourne’s Research Computing Services and the Petascale Campus Initiative, and Australian Research Council Grants (FT190100462, LP200301370).

We are grateful to A. Stuckert, N. Brajuka and J. Sproul for their advice on repeat modelling, and J.T. Li for providing suggestions on synteny analyses. We thank M. Moore for exploratory repeat analyses.

REFERENCES


Hotaling, S., Kelley, J. L., & Frandsen, P. B. (2021a). Toward a genome sequence for every animal: Where are we now? *Proceedings of the National Academy of Sciences, 118*(52), e2109019118. doi:10.1073/pnas.2109019118


https://www.nature.com/articles/nature25458#supplementary-information


Paiola, M., Dimitrakopoulou, D., Pavelka, M. S., & Robert, J. (2023). Amphibians as a model to study the role of immune cell heterogeneity in host and mycobacterial...
interactions. Developmental & Comparative Immunology, 139, 104594.
doi:https://doi.org/10.1016/j.dci.2022.104594

doi:10.1007/s10577-017-9570-z

Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M.,
Chow, W., Fungtammasan, A., Kim, J., Lee, C., Ko, B. J., Chaissin, M., Gedman, G. L.,
Cantin, L. J., Thibaud-Nissen, F., Haggerty, L., Bista, I., Smith, M., Haase, B.,
Mountcastle, J., Winkler, S., Paez, S., Howard, J., Vernes, S. C., Lama, T. M., Grutzner,
F., Warren, W. C., Balakrishnan, C. N., Burt, D., George, J. M., Biegler, M. T., Iorns, D.,
Kauitt, A. F., Franchini, P., Detrich, H. W., Svardal, H., Wagner, M., Naylor, G. J. P.,
Pippel, M., Malinsky, M., Mooney, M., Simbirsky, M., Hannigan, B. T., Pesout, T.,
Houck, M., Misuraca, A., Kingan, S. B., Hall, R., Kronenberg, Z., Sovic, I., Dunn, C.,
Ning, Z., Hastie, A., Lee, J., Selvaraj, S., Green, R. E., Putnam, N. H., Gut, I., Ghurye, J.,
Osiopova, E., Al-Ajli, F. O., Secomandi, S., Kim, H., Theofanopoulou, C., Hiller, M., Zhou,
Y., Harris, R. S., Makova, K. D., Medvedev, P., Hoffman, J., Masterson, P., Clark, K.,
Martin, F., Howe, K., Flicek, P., Walenz, B. P., Kwak, W., Clawson, H., Diekhans, M.,
Nassar, L., Paten, B., Kraus, R. H. S., Crawford, A. J., Gilbert, M. T. P., Zhang, G.,
Venkatesh, B., Murphy, R. W., Koepfli, K.-P., Shapiro, B., Johnson, W. E., Di Palma, F.,
Marques-Bonet, T., Teeling, E. C., Warnow, T., Graves, J. M., Ryder, O. A., Haussler,
D., O’Brien, S. J., Korfach, J., Lewin, H. A., Howe, K., Myers, E. W., Durbin, R., Phillips,
A. M., & Jarvis, E. D. (2021). Towards complete and error-free genome assemblies of
all vertebrate species. Nature, 592(7856), 737-746. doi:10.1038/s41586-021-03451-0

Robert, J. (2020). Experimental platform using the amphibian Xenopus laevis for research in
top106625.

(2022). Development and validation of an eDNA protocol for monitoring endemic
Asian spiny frogs in the Himalayan region of Pakistan. Scientific Reports, 12(1), 5624.
doi:10.1038/s41598-022-09084-1

Scheele, B. C., Hunter, D. A., Grogan, L. F., Berger, L., Kolby, J. E., McFadden, M. S.,
Extinction Risk in Chytridiomycosis-Threatened Amphibians. Conservation Biology,
28(5), 1195-1205. doi:10.1111/cobi.12322

Scheele, B. C., Pasmans, F., Skerratt, L. F., Berger, L., Martel, A., Beukema, W., Acevedo, A.
A., Burrowes, P. A., Carvalho, T., Catenazzi, A., De la Riva, I., Fisher, M. C., Flechas, S.,
V., Foster, C. N., Frias-Alvarez, P., Garner, T. W. J., Gratwicke, B., Guayasamin, J. M.,
Hirschfeld, M., Kolby, J. E., Kosch, T. A., La Marca, E., Lindemayer, D. B., Lips, K. R.,
Longo, A. V., Maneyro, R., McDonald, C. A., Mendelson, J., Palacios-Rodriguez, P.,
Parra-Olea, G., Richards-Zawacki, C. L., Rödel, M.-O., Rovito, S. M., Soto-Azat, C.,


doi: https://doi.org/10.1016/j.gpb.2020.11.005


**DATA ACCESSIBILITY AND BENEFIT-SHARING**

**Data Accessibility Statement**

**Genetic data:**
All the genomes used in this study are available on the NCBI Genomes database


**Code:**
All original code has been deposited on GitHub and is publicly available at

[https://doi.org/10.5281/zenodo.7679280](https://doi.org/10.5281/zenodo.7679280)

**AUTHOR CONTRIBUTIONS**

### Table 1. Genome quality measures.

<table>
<thead>
<tr>
<th>Species</th>
<th>Family</th>
<th>N chrs</th>
<th>Assembly ID</th>
<th>N scaffolds</th>
<th>Genome size (Gb)</th>
<th>Gap %</th>
<th>Scaffold N50 (kb)</th>
<th>Contig N50 (kb)</th>
<th>% Chrs</th>
<th>% Repea ts</th>
<th>Busco % complete</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Ambystoma mexicanum</em></td>
<td>Ambystomatidae</td>
<td>14</td>
<td>GCA002915</td>
<td>27157</td>
<td>28.21</td>
<td>0.013</td>
<td>1205706.7</td>
<td>217.996</td>
<td>94.77</td>
<td>66.26</td>
<td>66.06</td>
</tr>
<tr>
<td><em>Allobates femoralis</em></td>
<td>Aromobatidae</td>
<td>0</td>
<td>unpublishe d</td>
<td>6658</td>
<td>5.32</td>
<td>0.041</td>
<td>625512.0</td>
<td>879.1</td>
<td>0.00</td>
<td>74.02</td>
<td>81.37</td>
</tr>
<tr>
<td><em>Bombina variegata</em></td>
<td>Bombinatoridae</td>
<td>0</td>
<td>GCA009336</td>
<td>4065339</td>
<td>4.68</td>
<td>0.564</td>
<td>2.0</td>
<td>1.9</td>
<td>0.00</td>
<td>56.99</td>
<td>13.03</td>
</tr>
<tr>
<td><em>Rhinella marina</em></td>
<td>Bufonidae</td>
<td>11</td>
<td>GCA014858</td>
<td>747</td>
<td>4.55</td>
<td>1.792</td>
<td>539822.1</td>
<td>1738.3</td>
<td>96.21</td>
<td>74.54</td>
<td>86.74</td>
</tr>
<tr>
<td><em>Bufo bombifrons</em></td>
<td>Bufonidae</td>
<td>11</td>
<td>GCA009336</td>
<td>31391</td>
<td>2.55</td>
<td>0.196</td>
<td>167.5</td>
<td>167.5</td>
<td>0.00</td>
<td>67.71</td>
<td>85.89</td>
</tr>
<tr>
<td><em>Eleutherodactylus coqui</em></td>
<td>Dendrobatidae</td>
<td>11</td>
<td>GCA014858</td>
<td>1307</td>
<td>5.04</td>
<td>0.196</td>
<td>635713.4</td>
<td>3959.8</td>
<td>98.92</td>
<td>77.47</td>
<td>90.11</td>
</tr>
<tr>
<td><em>Hyalinobatrachium fleischmanni</em></td>
<td>Centrolenidae</td>
<td>0</td>
<td>unpublishe d</td>
<td>227850</td>
<td>4.04</td>
<td>11.137</td>
<td>295403.6</td>
<td>10.5</td>
<td>0.00</td>
<td>61.2</td>
<td>66.27</td>
</tr>
<tr>
<td><em>Diploglossus pumilus</em></td>
<td>Dendrobatidae</td>
<td>0</td>
<td>GCA009801</td>
<td>125383</td>
<td>3.49</td>
<td>36.043</td>
<td>78.0</td>
<td>5.8</td>
<td>0.00</td>
<td>40.77</td>
<td>67.57</td>
</tr>
<tr>
<td><em>Diploglossus sylvaticus</em></td>
<td>Dendrobatidae</td>
<td>0</td>
<td>GCA009801</td>
<td>15287</td>
<td>5.19</td>
<td>0.109</td>
<td>571702.6</td>
<td>97.8</td>
<td>0.00</td>
<td>81.8</td>
<td>53.47</td>
</tr>
<tr>
<td><em>Geotrypetes seraphini</em></td>
<td>Dermophiidae</td>
<td>19</td>
<td>GCA002459</td>
<td>164</td>
<td>3.78</td>
<td>0.146</td>
<td>272612.2</td>
<td>20666.6</td>
<td>99.85</td>
<td>72.51</td>
<td>91.39</td>
</tr>
<tr>
<td><em>Nanorana parkeri</em></td>
<td>Dicroglossidae</td>
<td>0</td>
<td>GCA000935</td>
<td>25187</td>
<td>2.05</td>
<td>3.704</td>
<td>10691.1</td>
<td>32.8</td>
<td>0.00</td>
<td>51.39</td>
<td>90.40</td>
</tr>
<tr>
<td><em>Eleutherodactylus coqui</em></td>
<td>Dendrobatidae</td>
<td>13</td>
<td>GCA018957</td>
<td>105233</td>
<td>2.79</td>
<td>15.115</td>
<td>109467.1</td>
<td>10.9</td>
<td>63.88</td>
<td>53.83</td>
<td>79.11</td>
</tr>
<tr>
<td><em>Engystomops pullulans</em></td>
<td>Leptodactylidae</td>
<td>11</td>
<td>GCA019512</td>
<td>108304</td>
<td>2.56</td>
<td>0.38</td>
<td>172109.2</td>
<td>305.4</td>
<td>75.27</td>
<td>52.32</td>
<td>75.48</td>
</tr>
<tr>
<td><em>Platyphlebium ornatum</em></td>
<td>Limnodynastidae</td>
<td>0</td>
<td>GCA001617</td>
<td>148035</td>
<td>1.07</td>
<td>25.246</td>
<td>27.3</td>
<td>4.7</td>
<td>0.00</td>
<td>23.4</td>
<td>43.45</td>
</tr>
<tr>
<td><em>Limnodynastes dumerili</em></td>
<td>Limnodynastidae</td>
<td>0</td>
<td>GCA011038</td>
<td>520896</td>
<td>2.38</td>
<td>10.162</td>
<td>286.0</td>
<td>10.6</td>
<td>0.00</td>
<td>46.25</td>
<td>79.10</td>
</tr>
<tr>
<td><em>Mantella aurantiaca</em></td>
<td>Mantellidae</td>
<td>0</td>
<td>unpublishe d</td>
<td>135214</td>
<td>3.41</td>
<td>14.489</td>
<td>324858.8</td>
<td>14.1</td>
<td>0.00</td>
<td>52.56</td>
<td>58.57</td>
</tr>
<tr>
<td><em>Leptobrachium allomurum</em></td>
<td>Megophryidae</td>
<td>13</td>
<td>GCA018994</td>
<td>5370</td>
<td>3.54</td>
<td>0.149</td>
<td>412424.8</td>
<td>821.1</td>
<td>98.22</td>
<td>67.92</td>
<td>81.79</td>
</tr>
<tr>
<td><em>Leptobrachium irishenense</em></td>
<td>Megophryidae</td>
<td>13</td>
<td>GCA009667</td>
<td>5302</td>
<td>3.55</td>
<td>0.009</td>
<td>394693.0</td>
<td>1946.3</td>
<td>93.34</td>
<td>70.63</td>
<td>89.55</td>
</tr>
<tr>
<td><em>Spea bombifrons</em></td>
<td>Pelobatidae</td>
<td>0</td>
<td>GCA009364</td>
<td>1629896</td>
<td>0.77</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.00</td>
<td>33.73</td>
<td>3.71</td>
</tr>
<tr>
<td><em>Spea multiplicata</em></td>
<td>Pelobatidae</td>
<td>0</td>
<td>GCA009364</td>
<td>49736</td>
<td>1.08</td>
<td>0.235</td>
<td>71.5</td>
<td>30.7</td>
<td>0.00</td>
<td>33.55</td>
<td>76.59</td>
</tr>
<tr>
<td><em>Pipa parva</em></td>
<td>Pipidae</td>
<td>0</td>
<td>GCA019650</td>
<td>283592</td>
<td>1.37</td>
<td>0.596</td>
<td>7.2</td>
<td>6.2</td>
<td>0.00</td>
<td>43.01</td>
<td>26.72</td>
</tr>
<tr>
<td><em>Xenopus tropicalis</em></td>
<td>Pipidae</td>
<td>10</td>
<td>GCA000004</td>
<td>166</td>
<td>1.45</td>
<td>0.196</td>
<td>153961.3</td>
<td>14634.3</td>
<td>99.86</td>
<td>49.09</td>
<td>94.44</td>
</tr>
<tr>
<td><em>Xenopus laevis</em></td>
<td>Pipidae</td>
<td>18</td>
<td>GCA017654</td>
<td>54</td>
<td>2.74</td>
<td>0.061</td>
<td>155250.6</td>
<td>22451.5</td>
<td>99.97</td>
<td>45.21</td>
<td>95.95</td>
</tr>
</tbody>
</table>
Figure 1. Phylogenetic tree of amphibian families. Amphibian families with representative genomes are highlighted and numbers indicate genome counts per family. (Green) anurans, (blue) caecilians, and (orange) salamanders. *Engystomops pustulosus* (Family) image was taken by B. Gratwicke, other amphibian images were licenced to T. Kosch by Adobe Stock and Shutterstock.
**Figure 2.** BUSCO (Benchmarking Universal Single-Copy Orthologs) assessment results for amphibian genomes.
Figure 3. Repeat content across the amphibian genomes.
Figure 4: Relationship between proportion of total transposable elements per genome (body size corrected and log transformed) and Bio18 (precipitation of the warmest quarter, log transformed). Amphibian genomes have a higher proportion of transposable elements with increasing wet and warm conditions. Colours range from purple - low to medium- green to orange - high values.
Figure 5. Synteny plot of BUSCOs (Benchmarking Universal Single-Copy Orthologs) for chromosome-level anuran genomes. The phylogenetic tree was created with Timetree.org. The reference genome is *Xenopus tropicalis*. Tetraploid anuran *Xenopus laevis* has been excluded. *Indicate inverted chromosomes. Chromosomes without BUSCOs were excluded from the plot.