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

Tiffany A. Kosch[1], Andrew J. Crawford[2], Rachel L. Mueller[3], Katharina C. Wollenberg Valero[4], Ariel Rodríguez [5], Lauren A. O'Connell[6], Neil D. Young[1], and Lee F. Skerratt[1]

[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, 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

Running title: Comparative analysis of amphibian genomes

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, the sequencing and assembly of amphibian genomes has lagged far behind other taxa due to their comparatively large genome size. 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 repeat content of any vertebrate. We also provide evidence of conservation of genome synteny despite the long divergence times of this group but show that chromosome naming and orientation has been inconsistent across genome assemblies. Additionally, we discuss sequencing gaps in the phylogeny and suggest key targets for future sequencing endeavours. Lastly, we propose increased investment in amphibian genomics research to promote their conservation.

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 that 40% of species threatened by factors such as habitat change, disease, and over-exploitation (Scheele et al. 2019; IUCN 2022). Notably, many of these threats have been shown to be hard to reverse, suggesting that novel approaches that utilize genomic resources may lead to improved management decisions for some of the most endangered taxa (Scheele et al. 2014; Kosch et al. 2022).

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; Schulte et al. 2020; Nunes-de-Almeida et al. 2021). 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; Liu et al. 2020; De Angelis et al. 2021). Amphibians occur across habitat types from rainforests to deserts, freshwater streams to salt marshes, and tropical to arctic climates (Duellman 1999). 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 (Robert 2020; Paiola et al. 2023).

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 and four unpublished anuran assemblies. We evaluate assembly quality, sequencing technology, gene completeness and repeat content, taxonomic representation, and synteny.

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).

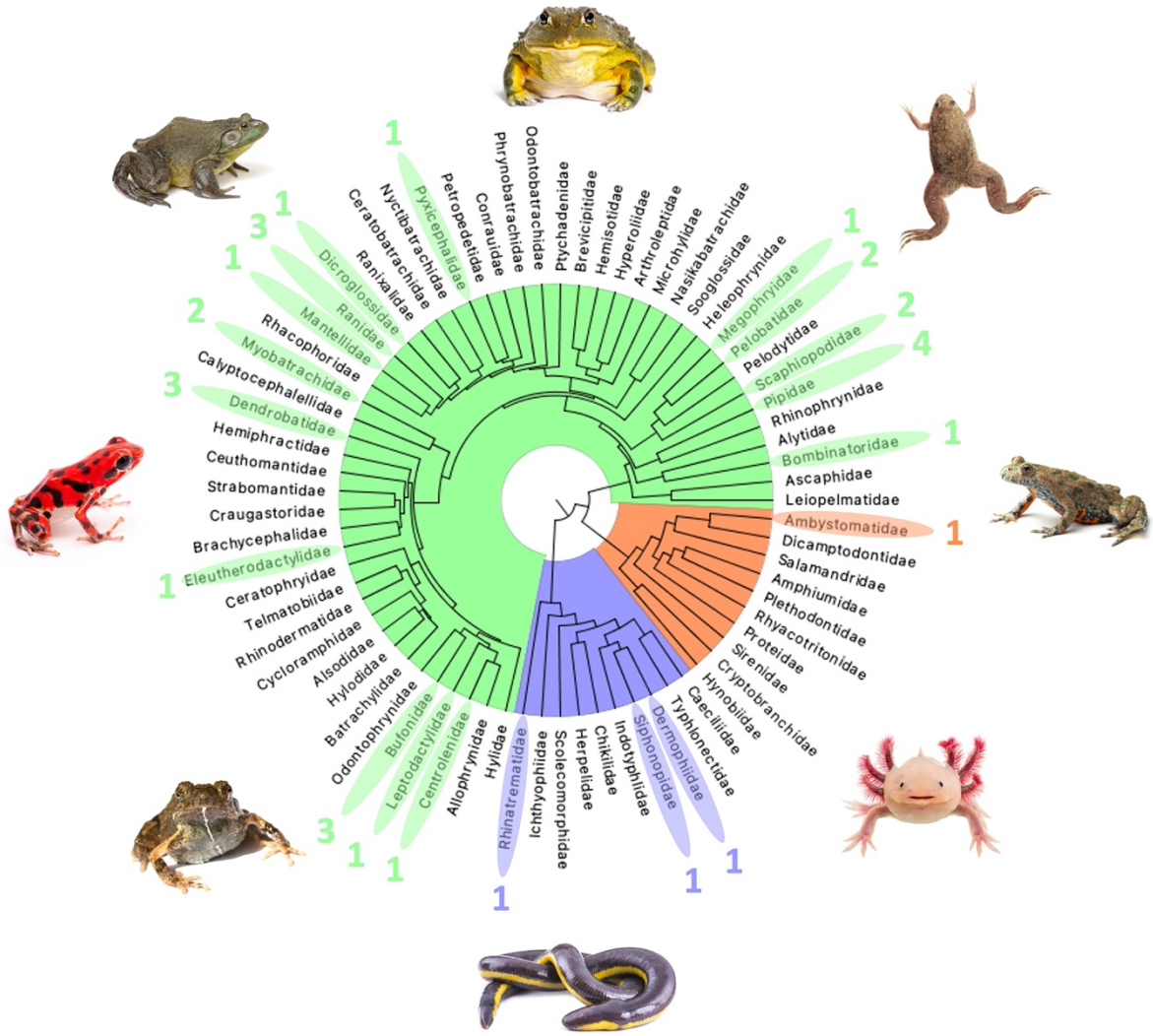

Figure 1. Phylogenetic tree of amphibian families. Amphibian families with representative genomes are highlighted and numbers indicate genome counts per family. (Green) anurans, (blue) caecilians, and (orange) salamanders. *E. pustulosus* (Family) image was taken by B. Gratwicke, other amphibian images were licenced to T. Kosch by Adobe Stock and Shutterstock.

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

Table 1. Genome quality measures.

Species	Family	N chrs	Assembly ID	N scaffolds	Genome size (Gb)	Gap %	Scaffold N50	Contig N50	% Chrs	% Repeats	BUSCO % complete
Allobates femoralis	Aromobatidae	0	in progress	6658	5.32	0.041	625512048	879094	0.00	74.02	81.37
Ambystoma mexicanum	Ambystomatidae	14	GCA002915635.3	27157	28.21	0.013	1205706733	217996	94.77	66.26	66.06
Bombina variegata	Bombinatoridae	0	GCA905336975.1	4065339	4.68	0.564	2005	1857	0.00	56.99	13.03
Bufo bufo	Bufonidae	11	GCA905171765.1	1307	5.04	0.196	635713434	3959750	98.92	77.47	90.11
Bufo gargarizans	Bufonidae	11	GCA014858855.1	747	4.55	1.792	539822115	1738317	96.21	74.54	86.74
Eleutherodactylus coqui	Eleutherodactylidae	13	GCA019857665.1	105233	2.79	15.115	109467076	10856	63.88	53.83	79.11
Engystomops pustulosus	Leptodactylidae	11	GCA019512145.1	108304	2.56	0.38	172109237	305357	75.27	52.32	75.48
Geotrypetes seraphini	Dermophiidae	19	GCA902459505.2	164	3.78	0.146	272612222	20656571	99.85	72.51	91.39
Glandirana rugosa	Ranidae	0	GCA018402905.1	561726	7.63	10.864	747629	11002	0.00	62.38	79.62
Hyalinobatrachium fleischmanni	Centrolenidae	0	in progress	227850	4.04	11.337	295403634	10459	0.00	61.2	66.27
Hymenochirus boettgeri	Pipidae	9	GCA019447015.1	25835	3.21	0.128	293320900	783846	82.53	52.72	87.38
Leptobrachium ailaonicum	Megophryidae	13	GCA018994145.1	5370	3.54	0.149	412424790	821125	98.22	67.92	81.79
Leptobrachium leishanense	Megophryidae	13	GCA009667805.1	5302	3.55	0.009	394693044	1946319	93.34	70.63	89.55
Limnodynastes dumerilii	Limnodynastidae	0	GCA011038615.1	520896	2.38	10.162	286041	10550	0.00	46.25	79.10
Lithobates catesbeianus	Ranidae	0	GCA002284835.2	1544635	6.25	11.584	39363	5994	0.00	62.17	44.05
Mantella aurantiaca	Mantellidae	0	in progress	135214	3.41	14.489	324858836	14079	0.00	52.56	58.57
Microcaecilia unicolor	Siphonopidae	14	GCA901765095.2	1081	4.69	1.012	376147139	3661507	97.06	69.21	90.89
Nanorana parkeri	Dicroglossidae	0	GCA000935625.1	25187	2.05	3.704	1069101	32798	0.00	51.39	90.40
Oophaga pumilio	Dendrobatidae	0	GCA009801035.1	125383	3.49	36.043	77984	5836	0.00	40.77	67.57
Oophaga sylvatica	Dendrobatidae	0	in progress	15287	5.19	0.109	571702597	97802	0.00	81.8	53.47
Pipa parva	Pipidae	0	GCA019650415.1	283592	1.37	0.596	7189	6215	0.00	43.01	26.72
Platyplectrum ornatum	Limnodynastidae	0	GCA016617825.1	148035	1.07	25.246	27267	4703	0.00	23.4	43.45
Pyxicephalus adspersus	Pyxicephalidae	14	GCA004786255.1	5411	1.56	10.517	157522388	30445	98.54	32.46	87.93
Rana temporaria	Ranidae	13	GCA905171775.1	555	4.11	0.637	481763206	6264810	98.26	69.11	90.72
Rhinatrema bivittatum	Rhinatrematidae	19	GCA901001135.2	1330	5.32	0.638	486875112	3216284	97.36	67.86	91.58
Rhinella marina	Bufonidae	0	GCA900303285.1	31391	2.55	0	167498	167498	0.00	67.71	85.89
Scaphiopus couchii	Scaphiopodidae	0	GCA009364435.1	1378905	0.48	0	362	362	0.00	26.94	0.70
Scaphiopus holbrookii	Scaphiopodidae	0	GCA009364455.1	1509266	0.71	0	514	514	0.00	31.42	2.60
Spea bombifrons	Pelobatidae	0	GCA009364475.1	1629896	0.77	0	522	522	0.00	33.73	3.71
Spea multiplicata	Pelobatidae	0	GCA009364415.1	49736	1.08	0.235	71513	30692	0.00	33.55	76.59
Xenopus laevis	Pipidae	18	GCA017654675.1	54	2.74	0.061	155250555	22451507	99.97	45.21	95.95
Xenopus tropicalis	Pipidae	10	GCA000004195.4	166	1.45	0.196	153961319	14634335	99.86	49.09	94.44

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 < 0.0000001; 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.

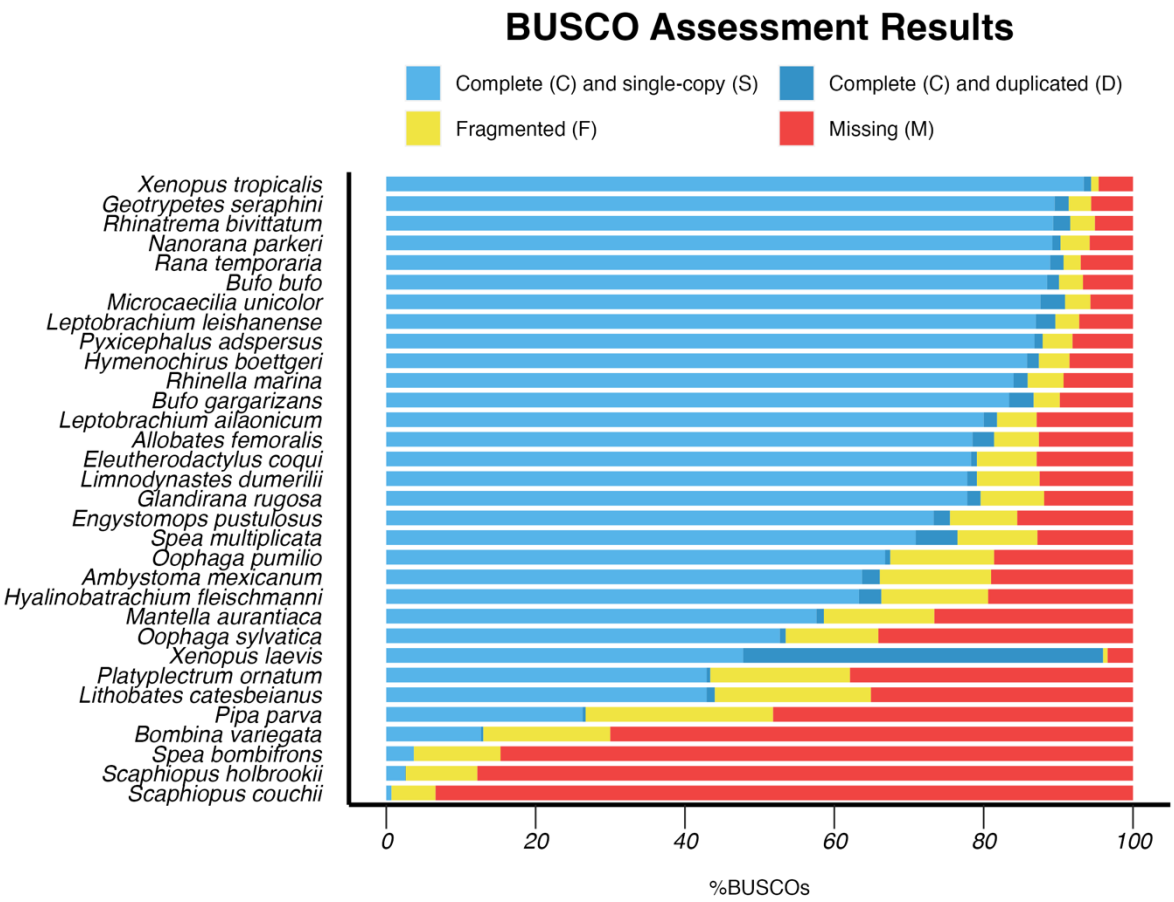

Figure 2. BUSCO assessment results for amphibian genomes.

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 (F_{29} = 12.8, p = 0.001), scaffold count (F_{29} = 7.80, p = 0.009), and percent BUSCO complete (F_{29} = 4.60, p = 0.041). The ability to classify repeats was also positively associated with read length, with longer reads resulting in better classification ($t_{26.241}$ = 3.57, p = 0.001).

These disparities in repeat percentage and content likely reflect differing evolutionary histories among species, as indicated by the 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).

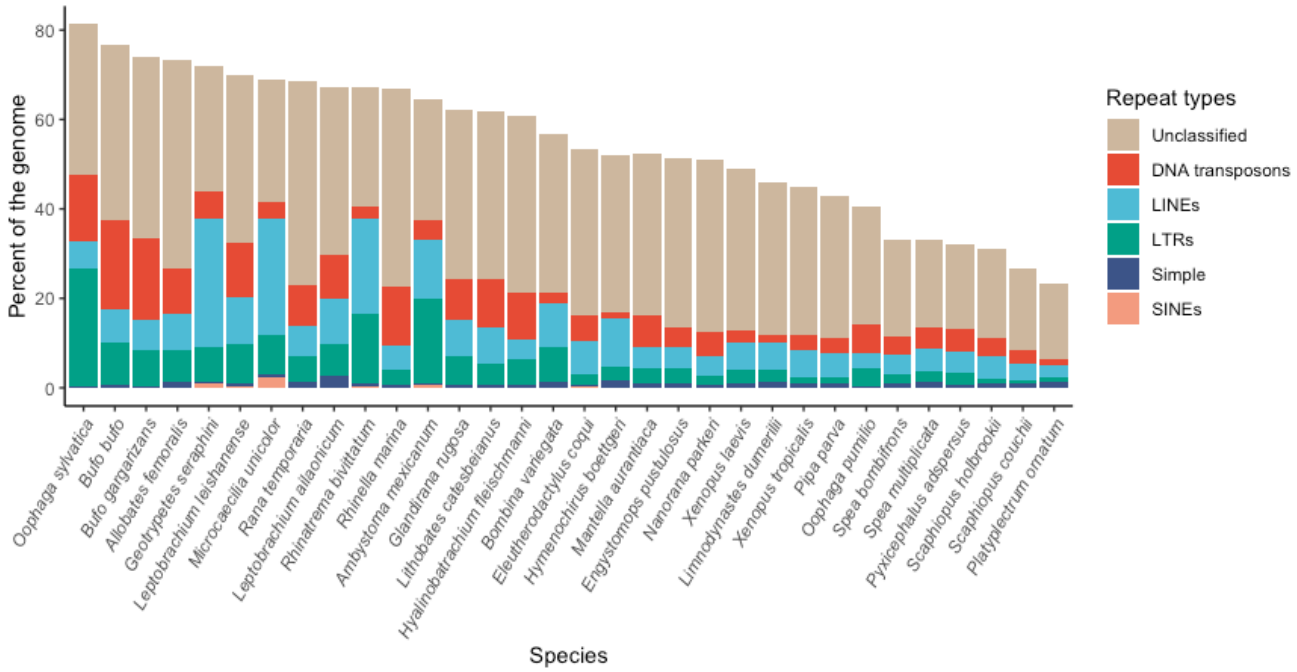

Figure 3. Repeat content across the amphibian genomes.

Genome synteny

Genome synteny of BUSCO genes was highly conserved within anurans (Fig. 4) and within caecilians (Fig. S6) but was less conserved across the amphibian orders (Fig. S5). However, chromosome naming was inconsistent across all taxa (Fig. 4, S5-6). For example, *Xenopus* sp. chr1 was chr12 in *Leptobrachium ailaonicum* (but not *L. leishanense*) and chr2 in *Bufo bufo* (but not *B. gargarizans*) (Fig. 4). Orientation of chromosomes was also inconsistent, including between species of the same genus (e.g., *Bufo*, *Leptobrachium*) (Fig. 4) and among the three caecilians (Fig. S6). Multiple inversions were evident including between chr5 of pipids *Xenopus laevis* (chr3 of *X. 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 4 and S6). 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. 4). 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.

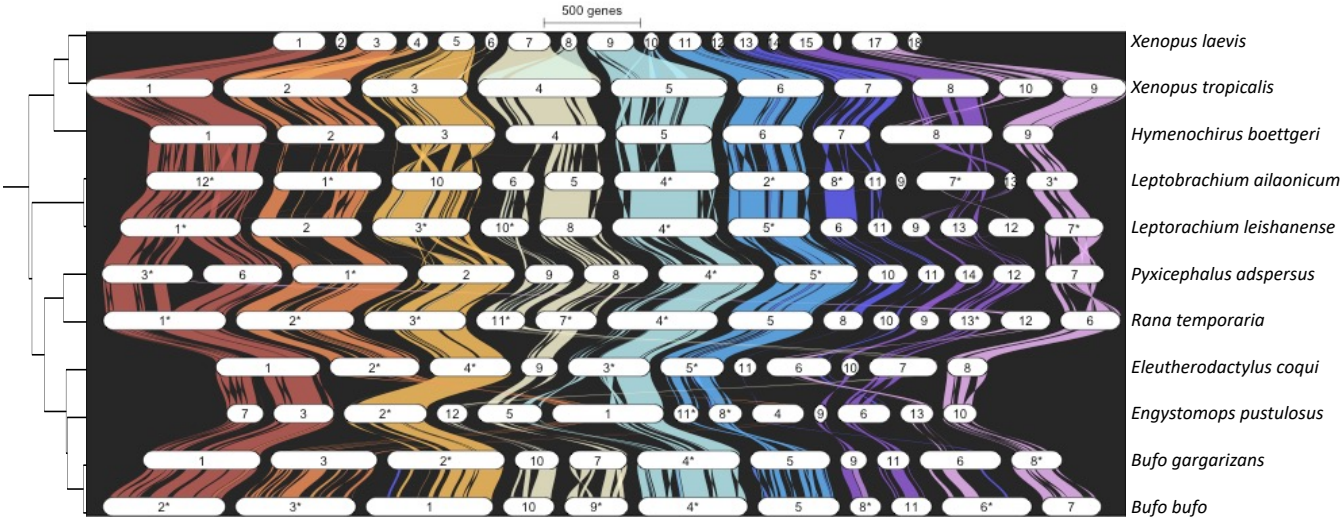

Figure 4. Synteny plot of BUSCOs for chromosome-level anuran genomes. Chromosomes are scaled by the number of BUSCOs. Phylogenetic tree created with Timetree.org. Reference genome is *Xenopus laevis*. *Indicate inverted chromosomes.

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). This is considerably less genomes 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; Rhie et al. 2021; Lawniczak et al. 2022).

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) (Sun et al. 2012; Frahry et al. 2015), 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) (Lynch and Walsh 2007; Liedtke et al. 2018) (but see Mohlhenrich and 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 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); and given 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 as well.

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).

Our study is one of the first to examine chromosomal synteny across amphibians. We show that overall synteny of amphibian genomes is relatively conserved, yet chromosome number and content vary across species, which seems to have been driven by multiple occurrences of chromosomal fusions and fissions (e.g., Fig. 4). Chromosomal rearrangements have occurred throughout vertebrate evolution, including the hypothesized

fusion of microchromosomes in the ancestor to 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 and Veltsos 2021), and unique features (Nowoshilow et al. 2018; Fischer et al. 2019; 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 identifucation 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 functional genetics for example for the purpose of immunological research to understand disease resistance, or for conservation purposes to enhance fitness.

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), 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 S6). 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 treeannotator (Drummond and 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://amphibiansoftheworld.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.

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 and 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 scripts. 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 https://doi.org/10.5281/zenodo.7679280.

Data and code availability

All original code has been deposited at GitHub and is publicly available at https://doi.org/10.5281/zenodo.7679280.

Competing interest statement

The authors declare no competing interests.

Supplemental Information

Supplemental information can be found online at https://doi.org/...

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.

Author contributions: conceptualization, T.A.K., A.J.C., L.A.O., A.R., and K.C.W.V; methodology, T.A.K, N.D.Y, R.L.M, and A.R.; formal analysis, T.A.K.; investigation, T.A.K., N.D.Y, and R.L.M.; resources, L.A.0. and A.R.; writing – original draft, T.A.K.; writing – review & editing, all authors; project administration, T.A.K; funding acquisition, T.A.K. and L.F.S.

References

Aganezov S, Yan SM, Soto DC, Kirsche M, Zarate S, Avdeyev P, Taylor DJ, Shafin K, Shumate A, Xiao C et al. 2022. A complete reference genome improves analysis of human genetic variation. *Science* **376**: eabl3533.

Biscotti MA, Carducci F, Olmo E, Canapa A. 2019. Vertebrate Genome Size and the Impact of Transposable Elements in Genome Evolution. In *Evolution, Origin of Life, Concepts and Methods*, doi:10.1007/978-3-030-30363-1_12 (ed. P Pontarotti), pp. 233-251. Springer International Publishing, Cham.

Blaxter M, Archibald JM, Childers AK, Coddington JA, Crandall KA, Di Palma F, Durbin R, Edwards SV, Graves JAM, Hackett KJ et al. 2022. Why sequence all eukaryotes? *Proceedings of the National Academy of Sciences* **119**: e2115636118.

Breton B-AA, Beaty L, Bennett AM, Kyle CJ, Lesbarrères D, Vilaça ST, Wikston MJH, Wilson CC, Murray DL. 2022. Testing the effectiveness of environmental DNA (eDNA) to quantify larval amphibian abundance. *Environmental DNA* **4**: 1229-1240.

Brown JL, Morales V, Summers K. 2010. A key ecological trait drove the evolution of biparental care and monogamy in an amphibian. *The american naturalist* **175**: 436-446.

Chestnut T, Anderson C, Popa R, Blaustein AR, Voytek M, Olson DH, Kirshtein J. 2014. Heterogeneous Occupancy and Density Estimates of the Pathogenic Fungus *Batrachochytrium dendrobatidis* in Waters of North America. *PLOS ONE* **9**: e106790.

Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Austine-Orimoloye O, Azov AG, Barnes I, Bennett R et al. 2022. Ensembl 2022. *Nucleic Acids Res* **50**: D988-d995.

Daly JW, Garraffo HM, Spande TF, Decker MW, Sullivan JP, Williams M. 2000. Alkaloids from frog skin: the discovery of epibatidine and the potential for developing novel non-opioid analgesics. *Natural product reports* **17**: 131-135.

De Angelis M, Casciaro B, Genovese A, Brancaccio D, Marcocci M, Novellino E, Carotenuto A, Palamara A, Mangoni M, Nencioni L. 2021. Temporin G, an amphibian antimicrobial peptide against influenza and parainfluenza respiratory viruses: insights into biological activity and mechanism of action. *The FASEB Journal* **35**: e21358.

Decena-Segarra LP, Bizjak-Mali L, Kladnik A, Sessions SK, Rovito SM. 2020. Miniaturization, genome size, and biological size in a diverse clade of salamanders. *The American Naturalist* **196**: 634-648.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**: 214.

Duellman WE. 1999. *Patterns of distribution of amphibians: a global perspective*. JHU Press.

Elewa A, Wang H, Talavera-López C, Joven A, Brito G, Kumar A, Hameed LS, Penrad-Mobayed M, Yao Z, Zamani N. 2017. Reading and editing the *Pleurodeles waltl* genome reveals novel features of tetrapod regeneration. *Nat Commun* **8**: 1-9.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**: 1-14.

Fischer EK, Roland AB, Moskowitz NA, Tapia EE, Summers K, Coloma LA, O'Connell LA. 2019. The neural basis of tadpole transport in poison frogs. *Proceedings of the Royal Society B: Biological Sciences* **286**: 20191084.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences* **117**: 9451-9457.

Frahry MB, Sun C, Chong RA, Mueller RL. 2015. Low levels of LTR retrotransposon deletion by ectopic recombination in the gigantic genomes of salamanders. *J Mol Evol* **80**: 120-129.

Gregory TR. 2002. Genome size and developmental complexity. *Genetica* **115**: 131-146.

Hime PM, Briggler JT, Reece JS, Weisrock DW. 2019. Genomic Data Reveal Conserved Female Heterogamety in Giant Salamanders with Gigantic Nuclear Genomes. *G3: Genes|Genomes|Genetics* doi:10.1534/g3.119.400556: g3.400556.402019.

Hogg CJ, Ottewell K, Latch P, Rossetto M, Biggs J, Gilbert A, Richmond S, Belov K. 2022. Threatened Species Initiative: Empowering conservation action using genomic resources. *Proceedings of the National Academy of Sciences* **119**: e2115643118.

Hotaling S, Kelley JL, Frandsen PB. 2021a. Toward a genome sequence for every animal: Where are we now? *Proceedings of the National Academy of Sciences* **118**: e2109019118.

Hotaling S, Sproul JS, Heckenhauer J, Powell A, Larracuente AM, Pauls SU, Kelley JL, Frandsen PB. 2021b. Long-reads are revolutionizing 20 years of insect genome sequencing. *Genome Biol Evol* doi:10.1093/gbe/evab138.

IUCN. 2022. The IUCN Red List of Threatened Species. Version 2022-2. Accessed on [6 January 2023].

Jetz W, Pyron RA. 2018. The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. *Nature Ecology & Evolution* **2**: 850-858.

Johnson BB, Searle JB, Sparks JP. 2021. Genome size influences adaptive plasticity of water loss, but not metabolic rates, in lungless salamanders. *J Exp Biol* **224**: jeb242196.

Kapusta A, Suh A, Feschotte C. 2017. Dynamics of genome size evolution in birds and mammals. *Proceedings of the National Academy of Sciences* **114**: E1460-E1469.

Keinath MC, Timoshevskiy VA, Timoshevskaya NY, Tsonis PA, Voss SR, Smith JJ. 2015. Initial characterization of the large genome of the salamander *Ambystoma mexicanum* using shotgun and laser capture chromosome sequencing. *Sci Rep* **5**: 16413.

Kosch TA, Waddle AW, Cooper CA, Zenger KR, Garrick DJ, Berger L, Skerratt LF. 2022. Genetic approaches for increasing fitness in endangered species. *Trends Ecol Evol* **37**: 332-345.

Kriventseva EV, Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, Zdobnov EM. 2018. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res* **47**: D807-D811.

Lamichhaney S, Catullo R, Keogh JS, Clulow S, Edwards SV, Ezaz T. 2021. A bird-like genome from a frog: Mechanisms of genome size reduction in the ornate burrowing frog, *Platyplectrum ornatum*. *Proceedings of the National Academy of Sciences* **118**: e2011649118.

Lawniczak MKN, Durbin R, Flicek P, Lindblad-Toh K, Wei X, Archibald JM, Baker WJ, Belov K, Blaxter ML, Marques Bonet T et al. 2022. Standards recommendations for the Earth BioGenome Project. *Proceedings of the National Academy of Sciences* **119**: e2115639118.

Lee BT, Barber GP, Benet-Pagès A, Casper J, Clawson H, Diekhans M, Fischer C, Gonzalez JN, Hinrichs AS, Lee CM et al. 2022. The UCSC Genome Browser database: 2022 update. *Nucleic Acids Res* **50**: D1115-d1122.

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. *Nature Ecology & Evolution* **2**: 1792-1799.

Liu Y, Shi D, Wang J, Chen X, Zhou M, Xi X, Cheng J, Ma C, Chen T, Shaw C. 2020. A novel amphibian antimicrobial peptide, phylloseptin-PV1, exhibits effective anti-

staphylococcal activity without inducing either hepatic or renal toxicity in mice. *Frontiers in Microbiology* **11**: 565158.
Lovell JT, Sreedasyam A, Schranz ME, Wilson M, Carlson JW, Harkess A, Emms D, Goodstein DM, Schmutz J. 2022. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *eLife* **11**: e78526.
Lynch M, Walsh B. 2007. *The origins of genome architecture*. Sinauer Associates, Sunderland.
Ma W-J, Veltsos P. 2021. The Diversity and Evolution of Sex Chromosomes in Frogs. *Genes* **12**: 483.
Manni M, Berkeley MR, Seppey 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**: 4647-4654.
Meyer A, Schloissnig S, Franchini P, Du K, Woltering JM, Irisarri I, Wong WY, Nowoshilow S, Kneitz S, Kawaguchi A et al. 2021. Giant lungfish genome elucidates the conquest of land by vertebrates. *Nature* **590**: 284-289.
Mohlhenrich ER, Mueller RL. 2016. Genetic drift and mutational hazard in the evolution of salamander genomic gigantism. *Evolution* **70**: 2865-2878.
Novák P, Guignard MS, Neumann P, Kelly LJ, Mlinarec J, Koblížková A, Dodsworth S, Kovařík A, Pellicer J, Wang W et al. 2020. Repeat-sequence turnover shifts fundamentally in species with large genomes. *Nature Plants* **6**: 1325-1329.
Nowoshilow S, Schloissnig S, Fei J-F, Dahl A, Pang AWC, Pippel M, Winkler S, Hastie AR, Young G, Roscito JG et al. 2018. The axolotl genome and the evolution of key tissue formation regulators. *Nature* doi:10.1038/nature25458 https://www.nature.com/articles/nature25458#supplementary-information.
Nunes-de-Almeida CHL, Haddad CFB, Toledo LF. 2021. A revised classification of the amphibian reproductive modes. *Salamandra* **57**: 413-427.
O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**: D733-745.
Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T et al. 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol* **20**: 275.
Paiola M, Dimitrakopoulou D, Pavelka MS, Robert J. 2023. Amphibians as a model to study the role of immune cell heterogeneity in host and mycobacterial interactions. *Dev Comp Immunol* **139**: 104594.
Platt RN, 2nd, Vandewege MW, Ray DA. 2018. Mammalian transposable elements and their impacts on genome evolution. *Chromosome Res* **26**: 25-43.
Rhie A McCarthy SA Fedrigo O Damas J Formenti G Koren S Uliano-Silva M Chow W Fungtammasan A Kim J et al. 2021. Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592**: 737-746.
Robert J. 2020. Experimental platform using the amphibian *Xenopus laevis* for research in fundamental and medical immunology. *Cold Spring Harbor Protocols* **2020**: pdb. top106625.
Saeed M, Rais M, Akram A, Williams MR, Kellner KF, Hashsham SA, Davis DR. 2022. Development and validation of an eDNA protocol for monitoring endemic Asian spiny frogs in the Himalayan region of Pakistan. *Sci Rep* **12**: 5624.

Scheele BC, Hunter DA, Grogan LF, Berger L, Kolby JE, McFadden MS, Marantelli G, Skerratt LF, Driscoll DA. 2014. Interventions for Reducing Extinction Risk in Chytridiomycosis-Threatened Amphibians. *Conserv Biol* **28**: 1195-1205.

Scheele BC, Pasmans F, Skerratt LF, Berger L, Martel A, Beukema W, Acevedo AA, Burrowes PA, Carvalho T, Catenazzi A et al. 2019. Amphibian fungal panzootic causes catastrophic and ongoing loss of biodiversity. *Science* **363**: 1459-1463.

Schloissnig S, Kawaguchi A, Nowoshilow S, Falcon F, Otsuki L, Tardivo P, Timoshevskaya N, Keinath MC, Smith JJ, Voss SR. 2021. The giant axolotl genome uncovers the evolution, scaling, and transcriptional control of complex gene loci. *Proceedings of the National Academy of Sciences* **118**.

Schulte LM, Ringler E, Rojas B, Stynoski JL. 2020. Developments in Amphibian Parental Care Research: History, Present Advances, and Future Perspectives. *Herpetol Monogr* **34**: 71-97.

Seidl F, Levis NA, Schell R, Pfennig DW, Pfennig KS, Ehrenreich IM. 2019. Genome of *Spea multiplicata*, a Rapidly Developing, Phenotypically Plastic, and Desert-Adapted Spadefoot Toad. *G3: Genes|Genomes|Genetics* **9**: 3909-3919.

Smit A, Hubley R, Green P. 2013. RepeatMasker Open-4.0. 2013-2015.

Sotero-Caio C, Challis R, Kumar S, Blaxter M. 2021. Genomes on a Tree (GoaT): A centralized resource for eukaryotic genome sequencing initiatives. *Biodiversity Information Science and Standards*.

Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. 2021. The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mobile DNA* **12**: 1-14.

Stuckert AM, Chouteau M, McClure M, LaPolice TM, Linderoth T, Nielsen R, Summers K, MacManes MD. 2021. The genomics of mimicry: gene expression throughout development provides insights into convergent and divergent phenotypes in a Müllerian mimicry system. *Mol Ecol*.

Suda K, Hayashi SR, Tamura K, Takamatsu N, Ito M. 2022. Activation of DNA Transposons and Evolution of piRNA Genes Through Interspecific Hybridization in *Xenopus* Frogs. *Front Genet* **13**.

Sun C, Shepard DB, Chong RA, López Arriaza J, Hall K, Castoe TA, Feschotte C, Pollock DD, Mueller RL. 2012. LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. *Genome Biol Evol* **4**: 168-183.

Sun Y-B, Zhang Y, Wang K. 2020. Perspectives on studying molecular adaptations of amphibians in the genomic era. *Zool Res* **41**: 351.

Team R. 2022. RStudio: integrated development for R. Boston, MA: RStudio. Inc.

Team RC. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. *http://www R-project org/*.

Wang J, Itgen MW, Wang H, Gong Y, Jiang J, Li J, Sun C, Sessions SK, Mueller RL. 2021. Gigantic Genomes Provide Empirical Tests of Transposable Element Dynamics Models. *Genomics, Proteomics & Bioinformatics* doi:https://doi.org/10.1016/j.gpb.2020.11.005.

Waters PD, Patel HR, Ruiz-Herrera A, Álvarez-González L, Lister NC, Simakov O, Ezaz T, Kaur P, Frere C, Grützner F et al. 2021. Microchromosomes are building blocks of bird, reptile, and mammal chromosomes. *Proceedings of the National Academy of Sciences* **118**: e2112494118.

Wu W, Gao YD, Jiang DC, Lei J, Ren JL, Liao WB, Deng C, Wang Z, Hillis DM, Zhang YP et al. 2022. Genomic adaptations for arboreal locomotion in Asian flying treefrogs. *Proc Natl Acad Sci U S A* **119**: e2116342119.