




# Genome Assembly of the Dyeing Poison Frog Provides Insights into the Dynamics of Transposable Element and Genome-Size Evolution

Carolin Dittrich<sup>1,2,\*†</sup>, Franz Hoelzl<sup>2,†</sup>, Steve Smith<sup>2</sup>, Chloe A. Fouilloux<sup>1</sup>, Darren J. Parker<sup>3</sup>, Lauren A. O'Connell <sup>4</sup>, Lucy S. Knowles<sup>5</sup>, Margaret Hughes<sup>6</sup>, Ade Fewings<sup>7</sup>, Rhys Morgan<sup>3</sup>, Bibiana Rojas <sup>1,2</sup>, and Aaron A. Comeault <sup>3,\*</sup>

<sup>1</sup>Department of Biology and Environmental Sciences, University of Jyväskylä, Jyväskylä, Finland

<sup>2</sup>Department of Interdisciplinary Life Sciences, Konrad Lorenz Institute of Ethology, University of Veterinary Medicine, Vienna, Austria

<sup>3</sup>School of Environmental and Natural Sciences, Molecular Ecology & Evolution Group, Bangor University, Bangor, UK

<sup>4</sup>Department of Biology, Stanford University, Stanford, USA

<sup>5</sup>NERC Environmental Omics Facility, University of Sheffield, Sheffield, UK

<sup>6</sup>Centre for Genomic Research, University of Liverpool, Liverpool, UK

<sup>7</sup>Supercomputing Wales, Digital Services, Bangor University, Bangor, UK

<sup>†</sup>These authors contributed equally to this work.

\*Corresponding authors: E-mails: a.comeault@bangor.ac.uk; carolin.dittrich@vetmeduni.ac.at.

Accepted: May 12, 2024

## Abstract

Genome size varies greatly across the tree of life and transposable elements are an important contributor to this variation. Among vertebrates, amphibians display the greatest variation in genome size, making them ideal models to explore the causes and consequences of genome size variation. However, high-quality genome assemblies for amphibians have, until recently, been rare. Here, we generate a high-quality genome assembly for the dyeing poison frog, *Dendrobates tinctorius*. We compare this assembly to publicly available frog genomes and find evidence for both large-scale conserved synteny and widespread rearrangements between frog lineages. Comparing conserved orthologs annotated in these genomes revealed a strong correlation between genome size and gene size. To explore the cause of gene-size variation, we quantified the location of transposable elements relative to gene features and find that the accumulation of transposable elements in introns has played an important role in the evolution of gene size in *D. tinctorius*, while estimates of insertion times suggest that many insertion events are recent and species-specific. Finally, we carry out population-scale mobile-element sequencing and show that the diversity and abundance of transposable elements in poison frog genomes can complicate genotyping from repetitive element sequence anchors. Our results show that transposable elements have clearly played an important role in the evolution of large genome size in *D. tinctorius*. Future studies are needed to fully understand the dynamics of transposable element evolution and to optimize primer or bait design for cost-effective population-level genotyping in species with large, repetitive genomes.

**Key words:** amphibians, genome evolution, transposable elements, long-reads, MobiSeq.

© The Author(s) 2024. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

## Significance

Amphibians display more variation in genome size than any other vertebrate lineage. Complexities associated with large genomes frequently hamper genome assembly and population genetic studies. Here we use long-read HiFi sequences to generate a high-quality 6.3 Gb genome assembly of the poison frog *Dendrobates tinctorius*. We use this genome and leverage comparative genomics and de novo annotations to quantify aspects of genome evolution driven by repetitive transposable genetic elements. Our results provide support for the dynamic role that transposable elements play in driving the evolution of “genomic gigantism” in amphibians. We also show how transposable elements can be leveraged for cost-efficient population genetic studies using limited input material.

## Introduction

Interspecific variation in genome size is a fundamental feature of biodiversity, and transposable elements play an important role in contributing to variation in genome size and structure (Kidwell 2002; Hawkins et al. 2006; Lee and Kim 2014). While historically referred to as “junk” DNA, it has been long known that the evolution of transposable elements can have profound effects on an organism’s phenotype. For example, the ability of transposable elements to drive gene expression and mosaic coloration in maize accompanied their discovery in the late 1940s by McClintock (1950). With the advancement of computational capabilities and the availability of cost-effective genetic methods, evidence of the phenotypic effects of transposable elements has increased. It is now widely appreciated that transposable elements can have a profound effect on evolution by influencing exon structure, telomeres, gene expression, and ultimately, adaptation or speciation (Feschotte 2008; Casacuberta and González 2013; Serrato-Capuchina and Matute 2018; Almojil et al. 2021). It has also been suggested that the mobilization and insertion of transposable elements can drive adaptation to global change when organisms are subject to environmental stress (Rey et al. 2016; Pimpinelli and Piacentini 2020). For a comprehensive classification and in-depth review of the impact transposable elements can have on the genome, we recommend the recent works of Kidwell and Lisch (1997), Wicker et al. (2007), Sotero-Caio et al. (2017), Bourque et al. (2018), and Almojil et al. (2021).

Characterizing the genomic landscape of transposable elements—such as the abundance of different types of elements and where they are located in the genome relative to genes, exons, and introns—is one approach that can be used to illuminate aspects of their evolution. With the increasing availability of whole-genome sequence data collected from many species, studies are increasingly quantifying the genomic landscape and “ecology” of transposable elements (Lamichhaney et al. 2021; Stitzer et al. 2021; Gozashti et al. 2023). A general pattern emerging from these studies is that different types of transposable elements contribute to genome evolution in different species. For example, long terminal repeat (LTR) retrotransposons

play an important role in genome size variation among plants (Lee and Kim 2014); while long and short interspersed nuclear elements (LINEs and SINEs, respectively) are more abundant than LTRs in mammals (Chalopin et al. 2015; Platt et al. 2018), LINEs are more abundant than SINEs and LTRs in bird (Zhang et al. 2014) and squamate genomes (Pasquasi et al. 2018), and SINEs are nearly absent in amphibian genomes (Zuo et al. 2023). In addition to the types of repetitive elements present within genomes, estimates of the timing of when different repetitive elements insert themselves vary both across species (Sun et al. 2015) and among the types of repetitive elements present within single genomes (Sun et al. 2015; Stitzer et al. 2021). Finally, different types of repetitive elements can display different insertion-site preferences or biases, with some inserting themselves into nonrandom locations in the genome, such as in intergenic regions, promoters, or introns (Bourque et al. 2018; Stitzer et al. 2021). Gaining a better understanding of the insertion-site preferences or biases displayed by different repetitive elements across different species is important because it will help facilitate a predictive understanding of how repetitive elements contribute to the evolution of genomic and genetic variation.

Among vertebrates, amphibians exhibit remarkable variation in genome size, surpassing that of any other group (Liedtke et al. 2018). The variation in amphibian genome size is influenced by the activity and accumulation of transposable elements (Sotero-Caio et al. 2017). However, the limited availability of high-quality amphibian genome assemblies—in large part due to the challenges of assembling large repetitive genomes with short-read sequencing technologies—makes the evolutionary dynamics and genomic ecology of transposable elements (TEs), alongside their impact on phenotypes and adaptation, difficult to test (however, see Zuo et al. 2023).

As complete genomes are often lacking for amphibians, the research community uses reduced representation library sequencing for studies of species delimitation and population genetics (Dufresnes et al. 2018; Funk et al. 2018; Nunziata and Weisrock 2018; Homola et al. 2019). Given their transposable element-rich genomes, it is surprising that methods that leverage repetitive elements, such as MobiSeq (Rey-Iglesia et al. 2019), have not yet

been applied to amphibians. MobiSeq is a method for constructing a reduced representation library by targeting the flanking regions of transposable elements to identify single nucleotide polymorphisms (SNPs) and genotypes (Rey-Iglesia et al. 2019). This approach requires minimal DNA input and does not necessarily require a reference genome, making it useful for investigating a wide array of research questions with regard to population genetics, evolutionary dynamics, and ecological interactions.

Poison frogs (family Dendrobatidae) are a group of Central and South American forest-dwelling amphibians with complex social behavior and elaborate parental care (Stynoski et al. 2015). A thriving research community has focused on this model clade for studying the effects of natural selection on phenotype, particularly the bright coloration coupled with chemical defenses that protects some species from predation (Noonan and Comeault 2009; Chouteau et al. 2011; Maan and Cummings 2012; Lawrence et al. 2019). A solid natural history background stemming from field observations done in the late 1900s (e.g. Sexton 1960; Silverstone 1973; Wells 1980; Myers and Daly 1983; Donnelly 1989; Summers 1989) had laid a firm foundation to ask both ultimate and proximate research questions. For example, research on the ultimate factors shaping their communication and territorial behavior, parental care, and space use, as well as the behavior of their larvae, has been steadily gaining traction (e.g. Pröhl 2005; Summers et al. 2006; Amézquita et al. 2011; Ringler et al. 2013; Schulte and Lötters 2013; Rojas 2014; Stynoski et al. 2014; Tumulty et al. 2014; Carvajal-Castro et al. 2021; Fouilloux et al. 2021). More recently, studies on the proximate mechanisms of such behaviors, e.g. neurobiology of egg provisioning, tadpole transport, and tadpole aggression (Fischer et al. 2019, 2020; Fischer and O'Connell 2020), and hormonal correlates of care, territoriality, and space use (Fischer and O'Connell 2020; Pašukonis et al. 2022; Rodríguez et al. 2022), now provide a more holistic understanding of what makes these frogs unique. However, the lack of genomic resources has precluded some of these topics from being addressed in depth. The one genome available for dendrobatid frogs (*Oophaga pumilio*) is highly fragmented, but still shows that transposable elements comprise a large portion of it (Rogers et al. 2018). Having access to reference genomes from more poison frogs would allow for more comprehensive approaches to questions related to demography, conservation, behavioral ecology, disease dynamics, and adaptation (Brandies et al. 2019).

In this study, we generate a high-quality reference genome for the dyeing poison frog, *Dendrobates tinctorius*. This species is aposematic—with drastic differences in coloration and toxicity across populations (Lawrence et al. 2019)—and displays complex social behaviors typical of many poison frog species including male parental care

and territoriality (Rojas 2014, 2015; Rojas and Pašukonis 2019; Fouilloux et al. 2021). We first leverage our genome assembly and three publicly available chromosome-scale assemblies of species belonging to Hyloidea to provide evidence for both large regions of synteny alongside significant structural evolution. This analysis also revealed that the evolution of genome size across these species is correlated with gene size, with *D. tinctorius* having both the largest genome and the longest genes. We then annotate transposable elements in the *D. tinctorius* genome to explore the genomic landscape of their evolution and find that transposable elements are more abundant in introns than in exons, likely contributing to the evolution of large genes in *D. tinctorius*. Finally, we explored the usefulness of the *D. tinctorius* reference genome in leveraging population genetic information from cost-effective mobile-element sequencing (MobiSeq, Rey-Iglesia et al. 2019). Applying this method to tadpoles collected from a wild population revealed that the highly repetitive nature of the *D. tinctorius* genome can generate challenges to MobiSeq genotyping, with a high degree of missing data across sites and individuals. We hope that the ability to use the *D. tinctorius* genome as a resource will refine the development of markers—such as additional mobile-element tags or alternate approaches such as baited capture—for genotyping individuals, thereby broadening the scope of behavioral ecology and population genomic research in amphibians.

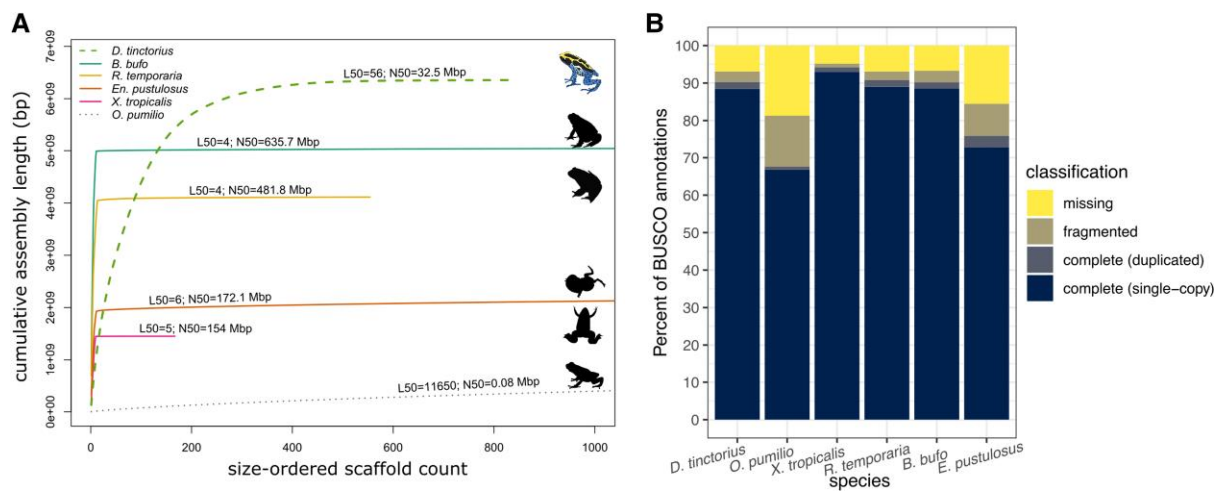
## Results

### Genome Assembly, Quality Control

Our final *D. tinctorius* assembly consists of 6.356 Gb assembled into 830 contigs. This assembly has a contig N50 and L50 of 32.539 Mb and 56 contigs, respectively, with a maximum contig size of 131.4 Mb. Our assembly is also highly accurate and complete, achieving an error rate of less than 0.0001—as indicated by a Quality Value score (Chen et al. 2021) of 41—and containing 4,796 of 5,310 curated tetrapod BUSCO genes (BUSCO summary: C: 90.3% [S: 88.4%, D: 1.9%], F: 2.8%, M: 6.9%, n: 5,310). While our assembly does not contain complete chromosomes, the contiguity and BUSCO scores were comparable to chromosome-level assemblies of other anurans available on NCBI (Fig. 1).

### Synteny Between *D. tinctorius* and Other Hyloidea Genomes

To explore patterns of conserved synteny between our assembly and closely related Anuran species, we compared the *D. tinctorius* assembly to chromosome-scale assemblies of *Bufo gargarizans* (Asiatic toad; Bufonidae), *Eleutherodactylus coqui* (common coqui; Eleutherodactylidae), and *Engystomops pustulosus* (Túngara frog; Leptodactylidae). *Bufo gargarizans*, *E. coqui*, and *E. pustulosus* shared a common ancestor with



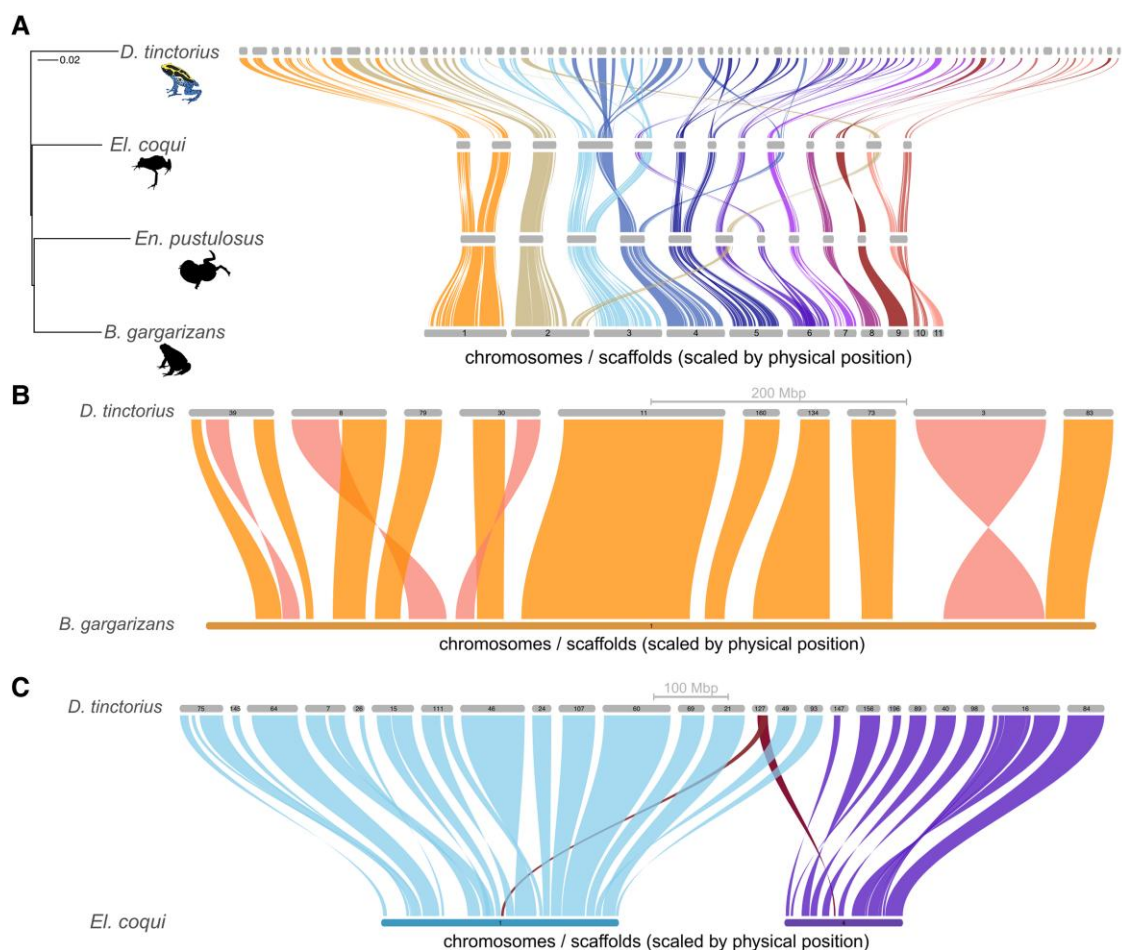
**Fig. 1.**—*Dendrobates tinctorius* assembly completeness. A) Cumulative assembly size in relation to size-ordered scaffolds for *D. tinctorius* and other published anuran genomes. B) BUSCO “tetrapoda” gene set completeness for the same genomes summarized in A). Pictograms used in panel A are from phylopic.org and are published under CC0 1.0 Universal Public Domain Dedication licence.

*D. tinctorius* ~65 Mya (Feng et al. 2017), and, while these four species have significantly diverged from one another, the former three were used for comparisons because they represent the three most closely related genera to *Dendrobates* that currently have publicly available chromosome-scale genome assemblies. Orthology-guided synteny analyses based on the tetrapoda BUSCO geneset (Simão et al. 2015; v5.2.2) in GENESPACE (Lovell et al. 2022; v1.3.1) provided evidence of both broad-scale synteny and genome evolution between *D. tinctorius* and the three other species we analyzed (Fig. 2). The number of synteny blocks identified between *D. tinctorius* and the chromosome-scale assemblies ranged from 116 (*E. pustulosus*) to 164 (*B. gargarizans*), with an average size of each block of 27.36 Mb (*E. pustulosus*) to 17.86 Mb (*B. gargarizans*). We focused this analysis on scaffolds of the *D. tinctorius* assembly that were greater than 20 Mb in length (see Materials and Methods), and these *D. tinctorius* scaffolds have median and mean lengths of 32.64 Mb and 41.81 Mb, respectively. Synteny between *D. tinctorius* scaffolds and the chromosome-scale assemblies therefore tends to span roughly half of entire *D. tinctorius* scaffolds, on average, while in many cases entire scaffolds showed collinearity with chromosomal regions of the chromosome-scale assemblies (Fig. 2B and C). In addition to collinear regions, we observed numerous rearrangements between *D. tinctorius* scaffolds and chromosomes of the other Anuran genomes we analyzed: for example, Fig. 2B highlights rearrangements between *D. tinctorius* scaffolds and *B. gargarizans*'s chromosome 1 (inversions highlighted in salmon pink color). Finally, we observed a single instance where a *D. tinctorius* scaffold contained synteny blocks that mapped to different chromosomes of a chromosome-scale assembly (*E. coqui* chromosomes 1 and 4; Fig. 2C). These results suggest that—at least at the course scale of ~65 My of evolution—

rearrangements within chromosomes are more numerous than changes in the overall evolution of chromosome number via fission and/or fusion events.

Our *D. tinctorius* assembly is larger than any of the genomes we used in the analyses presented above (*O. pumilio*, *B. gargarizans*, *E. coqui*, *E. pustulosus*, *Rana temporaria*, and *Xenopus tropicalis*), with genome size ranging from 1.451 Gb (*X. tropicalis*) to 6.356 Gb (*D. tinctorius*). We therefore explored how genome size evolution affects the size of gene regions by comparing the size of BUSCO annotations across five chromosome-scale genomes and the *D. tinctorius* assembly. For this analysis, we excluded the *O. pumilio* assembly due to high levels of fragmentation and included an assembly of *R. temporaria* to increase the phylogenetic diversity of genomes being analyzed—*R. temporaria* is outside of Hyloidea yet more closely related to Hyloidea than *X. tropicalis* (Fig. 3A). Across 1,949 single-copy orthologs that were present in all genomes, *D. tinctorius* orthologs ranged from 39.8% to 207.8% longer, on average, than the same orthologs in the *R. temporaria* and *X. tropicalis* assemblies, respectively. This percent difference in gene size was negatively correlated with the percent difference in genome size (Spearman's rho = 0.9;  $P = 0.083$ ), with larger genomes (i.e. *B. gargarizans* and *R. temporaria* vs. *D. tinctorius*) showing less difference in genes size than when genomes differed in size (i.e. *X. tropicalis* vs. *D. tinctorius*). In general, we found a positive correlation between genome size and the average size of BUSCO gene regions (rho = 0.94;  $P = 0.017$ , Fig. 3B), which was independent of phylogeny (Spearman's correlation on phylogenetically independent contrasts: rho = 0.9;  $P = 0.083$ ; Fig. 3C), indicating that the evolution of larger genomes in *D. tinctorius*, *B. gargarizans*, and *R. temporaria* has resulted in the evolution of larger gene regions. Because we focused on conserved orthologous





**FIG. 2.**—A) Riparian synteny plots generated using BUSCO gene regions as anchors showing broad patterns of synteny and rearrangements between *D. tinctorius* scaffolds and three publicly available chromosome-scale frog assemblies. B) Example *B. gargarizans* chromosome (chromosome 1) showing both synteny and structural evolution (e.g. inversions) with *D. tinctorius* scaffolds. C) Insertion of regions on *D. tinctorius* scaffold 127 into chromosomes 1 and 4 of the *E. coqui* assembly. Phylogeny shown in A) is pruned from the RAxML “L1-L379.concatenated”? tree reported in Hime et al. (2021).

gene regions that include introns, this pattern is likely driven by larger introns in larger genomes, and genome size is not correlated with predicted BUSCO amino acid lengths ( $\rho = -0.14$ ;  $P = 0.80$ ).

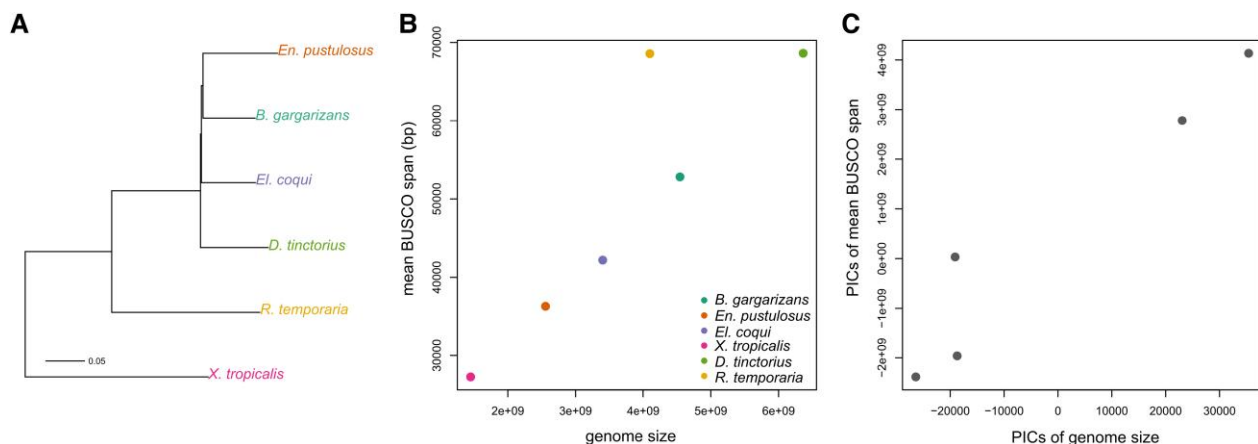
### Genome Annotation and Transposable Element Diversity

The majority of the *D. tinctorius* genome consists of repetitive elements: annotation with RepeatModeler (Flynn et al. 2020; v2.0.3) + RepeatMasker (Smith et al. 2021; v4.1.2-p1) identified 74.72% of our assembly as repetitive. Many of the repeats classified by RepeatMasker could not be assigned to specific types of repetitive elements (39.15%). Among assigned elements, 23.13% were retroelements and 12.45% were DNA transposons. The most common retroelements classified by RepeatMasker were LTR retrotransposons (15.69% of the assembly) and LINES (7.4%). LTR retrotransposons Ty3/DIRS1 were the most

abundant superfamily (13.79% of the assembly), while DNA transposons made up 12.45% of the assembly. Of DNA transposons, *Tc1-IS630-Pogo* (*Tc1*) and *hobo-Activator* (*hAT*) elements were the most abundant (6.57% and 3.91% of the assembly, respectively).

Independent analyses with LtrDetector (Valencia and Girgis 2019) annotated 262,486 LTRs spanning 2,343,246,106 bp, or 36.9% of our assembly. While this represents more LTRs in our assembly than when annotated with RepeatMasker, consistent with RepeatMasker, the most abundant LTR elements were found to belong to the Ty3/DIRS1 superfamily (99,884 elements, 819.4 Mb total), followed by *BEL/Pao* elements (1,762 elements, 13.9 Mb total). Also consistent with results from RepeatMasker, 66,046 of the LTR elements annotated by LtrDetector (spanning 588 Mb) had BLAST matches to the “Unknown”? category of elements identified by RepeatModeler.

To better understand the role that repetitive elements have played in the evolution of *D. tinctorius* genome



**Fig. 3.**—The relationship between genome size and the average size of annotated BUSCO gene regions in *D. tinctorius* compared with five publicly available chromosome-scale frog assemblies. A) Phylogenetic relationships between the species being compared (pruned RAxML “L1-L379.concatenated” tree from Hime et al. 2021). B) The relationship between genome size (in bp) and the mean size of BUSCO gene regions. C) The relationship between genome size and mean size of BUSCO gene regions, corrected for phylogeny using phylogenetically independent contrasts (PICs) with the phylogeny depicted in A).

**Table 1**

Counts of repeat elements annotated using RepeatModeler + RepeatMasker that overlap with coding DNA sequence (CDS) and introns annotated in the *D. tinctorius* assembly

Region of overlap	All elements	LTRs	LINEs	<i>Tc1</i> transposons	<i>hAT</i> transposons
CDS	115,330	3,500	1,826	10,227	507
Introns	528,441	42,844	27,152	45,806	2,907
Total N elements	11,331,718	1,048,437	702,556	897,264	265,361

structure, we compared the location of repetitive elements relative to de novo gene annotations we generated using the BRAKER2 pipeline (Brůna et al. 2021; v2.1.6). Of the 11,331,718 repeat annotations generated by RepeatMasker, 115,330 (1.02%) overlapped with coding DNA sequence (CDS) and 4,872 of these had a reciprocal overlap of at least 75% (Tables 1 and 2, respectively). More repetitive elements overlapped with introns compared to CDS (528,441 elements;  $\chi^2 = 270,118$ ;  $P < 2.2 \times 10^{-16}$ ); however, proportionally fewer TEs that overlapped introns showed reciprocal overlaps of greater than 75% compared to those overlapping CDS (1.05% vs. 4.22%, respectively;  $\chi^2 = 5,956.1$ ;  $P < 2.2 \times 10^{-16}$ ). These patterns are likely due to the fact that introns span nearly four times the number of bases and have a median size nearly twice that of CDS features (266 Mb vs. 69 Mb and 113 vs. 66 bp, respectively).

We next considered positions of the four most abundant classes/families of repetitive elements annotated by RepeatMasker in the *D. tinctorius* genome—LTRs, LINEs, *Tc1* DNA transposons, and *hAT* DNA transposons—in relation to CDS and introns. Of these four classes/families of repeats, LTRs and *Tc1* transposons showed the highest overlap with CDS and introns, with 0.33% (3,500) of LTR elements and 1.14% (10,227) of *Tc1* transposons showing

at least partial overlap with CDS, and 4.09% (42,844) and 5.11% (45,806) overlapping with introns, respectively. Only *Tc1* transposons were enriched in the proportion overlapping with CDS and introns compared to all non-*Tc1* elements (CDS: 1.14% vs. 1.01%, respectively; introns: 5.27% vs. 4.65%, respectively).

Given that a large portion of the *D. tinctorius* assembly is comprised of LTR elements—15.69% to 36.9% of the assembly annotated by RepeatModeler or LtrDetector, respectively—we next explored length distributions of these elements and estimated the timing of their insertion by estimating divergence between the left and right LTR of each element, assuming a substitution rate of  $2.5 \times 10^{-9}$  substitutions per site per year (Lau et al. 2020). Average estimates of insertion times for different types of LTR retroelements ranged from 10 to 28 Mya (Fig. 4). Retroelements with BLAST matches to the *DIRS* order of elements had the oldest average estimated insertion time (mean = 23.35 Mya, 95% empirical range = 1 to 60.4 Mya) while *ERV1* elements had the youngest insertion times (mean = 10.4 Mya, 95% empirical range = 0 to 45.1 Mya). The broad range of insertion times we estimated indicate that some LTR retroelements are old and may be shared with other species of poison frog (the MRCA of Dendrobatidae is estimated to have occurred ~36 Mya; Hime et al. 2021), while others

**Table 2**

Counts of repetitive elements that overlap with CDS and introns (see Table 1) that show greater than 75% reciprocal overlap with those features

Region of overlap	All elements	LTRs	LINEs	Tc1 transposons	hAT transposons
CDS	4,872	13	0	11	2
Introns	5,530	914	931	1,193	23

are young, potentially active, and species-specific (e.g. *Dendrobates* split from its sister genus *Oophaga* ~16 Mya; Guillory et al. 2019).

### A Test of the Utility of MobiSeq for Population Genetics

We next used the *D. tinctorius* reference genome to compare the genotyping success of three genotyping approaches using a MobiSeq dataset generated from 87 tadpoles collected from 17 phytotelmata at the Nouragues Research Station, French Guiana. Specifically, we compared the number of usable SNPs generated by either (i) de novo assembly of MobiSeq reads, (ii) mapping the reads to the other poison frog genome currently available on NCBI (*O. pumilio*), or (iii) mapping the reads to the *D. tinctorius* genome we generated as part of this study. We also called SNPs using two approaches: either the stacks pipeline (Catchen et al. 2011, 2013; v2.64) designed to assemble RADseq data de novo or using a reference genome; or the original MobiSeq pipeline using the program “analysis in next generation sequencing data”? (ANGSD; Korneliusson et al. 2014; v0.940) mapping sequence reads to the *D. tinctorius* assembly (Rey-Iglesia et al. 2019) (details in Materials and Methods).

The number of SNPs differed considerably between the two primers and different mapping approaches (de novo vs. reference genome) and showed high degrees of missing genotypes when genotyped using the stacks pipeline (LINE109:  $93.21 \pm 3.06\%$ , TE644:  $90.46 \pm 1.46\%$ ; mean percent missing data across sites  $\pm$  SD) and low coverage (supplementary table S1, Supplementary Material online). Regions amplified with the LINE109 primer resulted in considerably fewer SNPs than the TE644 primer. When we mapped sequences to the *O. pumilio* genome, gstacks only incorporated 10% to 12% of the reads and called considerably fewer SNPs than either the de novo approach or mapping to the *D. tinctorius* genome. The de novo approach with 50 bp fragments called the highest numbers of SNPs ( $n$  SNPs TE644 = 92,433); in contrast, the 100 bp fragments called the lowest numbers ( $n$  SNPs TE644 = 15,987). These differences were expected due to the large loss of SNPs due to truncation. Mapping to the *D. tinctorius* reference genome for both primers called the second most SNPs ( $n$  SNPs; LINE109 = 2,476, TE644 = 82,863), but missingness was still high and coverage low (supplementary

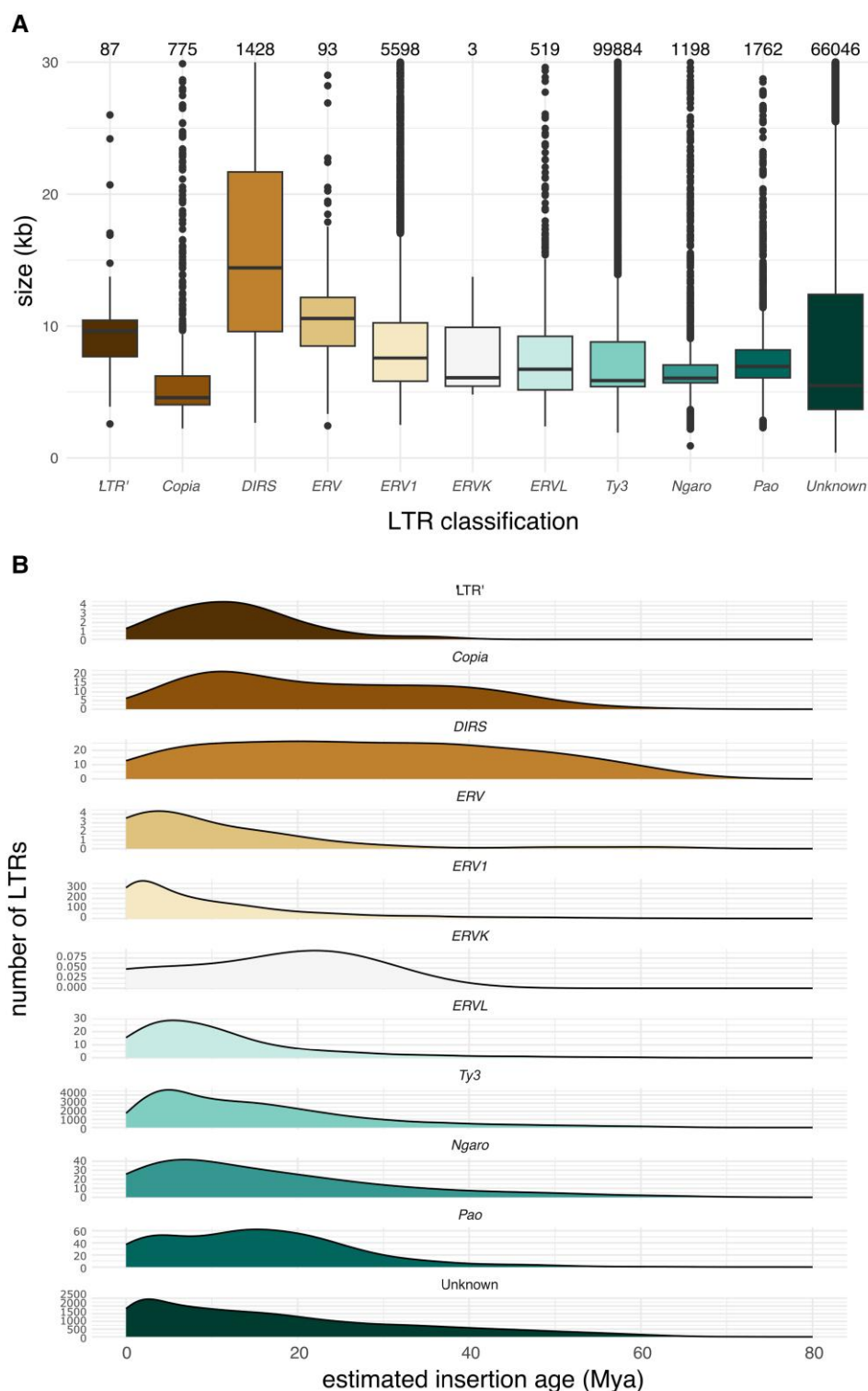
tables S1 and S2, Supplementary Material online). After applying a strong filter, where SNP had to be present in at least 50% of individuals, the numbers of SNP dropped dramatically (supplementary table S2, Supplementary Material online and Fig. 5B).

The SNP calling with the ANGSD pipeline gave  $17.67 \pm 0.17\%$  missing genotypes and 345 SNPs were called when the output was restricted to be present in at least 70 individuals. If SNPs were restricted to be present in at least 80 individuals, 291 SNPs could be called with  $11.07 \pm 0.08\%$  genotype missingness (Fig. 5A and B).

We used our called SNP dataset to test their usefulness in parentage analysis. Specifically, we ran a relatedness analysis with COLONY (Jones and Wang 2010; v2.0) on our dataset of 87 tadpoles. We included five sets of duplicate samples to test whether these samples will be reliably grouped together in our analysis. In general, we found significant differences in COLONY results depending on which SNP dataset we used for the analysis. Due to the low number of SNPs called with the LINE109 primers, we focused our COLONY analyses to data obtained using the TE644 primers. We considered the successful resolution of our duplicate samples, grouped together with a probability of 0.9 or greater, as evidence that a SNP dataset was reliable for parentage analysis. Generally, reliability was low when using SNPs that were genotyped using Stacks: neither the de novo nor the reference genome mapping approaches reliably identified the five duplicates in the dataset with high probability (Fig. 5C). When genotyped using the ANGSD pipeline, COLONY identified 4/5 duplicates when SNPs from at least 80 individuals were retained, and 5/5 duplicates when 70 individuals were the minimum threshold for calling SNPs, suggesting that the more complete datasets generated with ANGSD—with only 11% to 17% missing genotypes—provided sufficient information for identifying full-sibs (Fig. 5A and C). This method also gave more realistic estimated population sizes than the stacks method (supplementary table S2, Supplementary Material online).

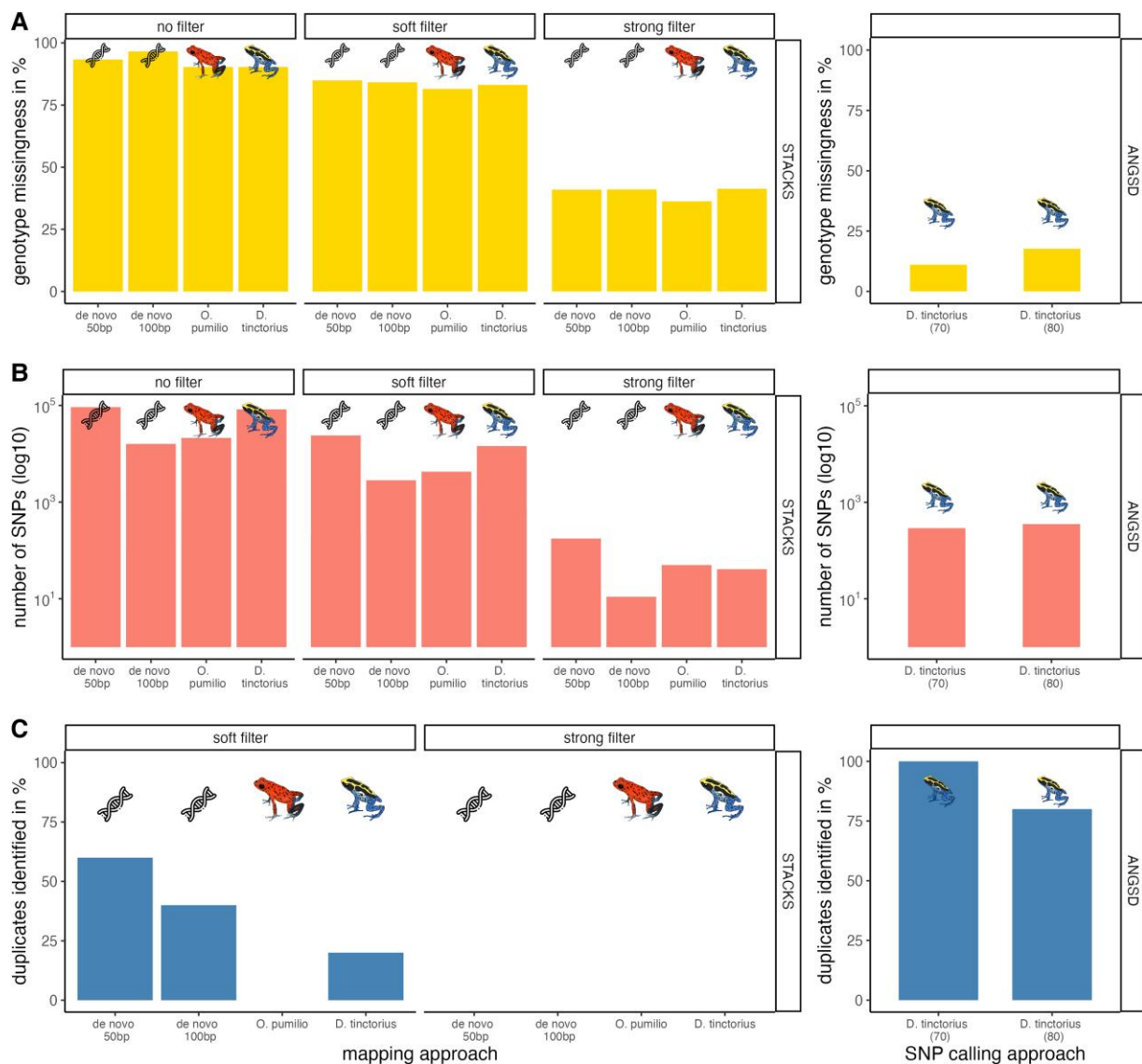
## Discussion

Access to genomic resources and tools holds the potential to transform our understanding of the ecology, evolution, life-history, and conservation of amphibians (e.g. Liedtke et al. 2018; Womack et al. 2019; Schloissnig et al. 2021; Kosch et al. 2022); yet amphibians have lagged behind other groups of animals in available genomic resources (Hotaling et al. 2021). The lack of genomic resources for frogs and toads (Anurans), in particular, is at least in part due to some species possessing large and complex genomes (Sun et al. 2015; Nowoshilow et al. 2018; Rogers et al. 2018): among frogs and toads genome size is highly variable ranging from 0.99 Gb in the plains spadefoot toad, *Spea*



**Fig. 4.**—Long terminal repeat (LTR) retroelement lengths A) and insertion times B). LTRs were annotated with LtrDetector and insertion times were estimated from divergence between the left and right LTR regions of each element independently assuming a substitution rate of  $2.5 \times 10^{-9}$  substitutions per site per year (Lau et al. 2020). In A), the total number of elements classified to each type is given above each boxplot.





**FIG. 5.**—Summary of A) genotype missingness in %, B) number of SNPs (log scale to account for the large differences), and C) percentage of duplicates found in the dataset using sequences derived from MobiSeq with the TE644 primer. Shown are the two pipeline approaches, stacks on the left (de novo assembly and mapping to reference genomes of *Oophaga pumilio* or *Dendrobates tinctorius*) with no filter SNPs, soft filter (filter\_rad) and strong filter (vcftools); and ANGSD on the right (calling of SNPs present in 70 or 80 individuals). For details, see Materials and Methods.

*bombifrons* (GenBank accession GCA\_027358695.2) to 10.2 Gb in the mountain yellow-legged frog, *Rana muscosa* (GenBank accession GCA\_029206835.1). Here we report a 6.8 Gb assembly of the dyeing poison frog, *D. tinctorius*, with a contig N50 of 32.5 Mb and 50% of the assembly being represented by only 56 contigs (Fig. 1). The lack of fully assembled chromosomes in our assembly limits our ability to test for chromosomal evolution in *D. tinctorius* compared to other species; however, we found only a single contig that mapped to multiple contigs of other available Hyloidea genomes (Fig. 2). Our analysis of synteny between these genomes suggests that the evolution of intrachromosomal

rearrangements occurs much more rapidly than the evolution of chromosome structure, despite 2-fold differences in genome size. Recent work comparing eight distantly related anurans supports the idea that the evolution of genome size greatly outpaces chromosomal changes (Bredeson et al. 2024).

We used multiple annotation approaches to show that over three quarters of the *D. tinctorius* genome consists of repetitive elements. These repetitive elements are more abundant in introns than exons, contributing to the evolution of gene sizes. Below, we discuss how our results advance the understanding of genome structure evolution

in frogs. Additionally, we highlight an application of the *D. tinctorius* genome to facilitate marker development for cost-effective, population-scale multilocus genotyping using MobiSeq (Rey-Iglesia et al. 2019).

### Transposable Elements and Genome Evolution

Amphibians are particularly useful models to provide insights into relationships between environmental or ecological factors and the dynamics of genome size evolution (Liedtke et al. 2018). Transposable elements are an interesting component of genome architecture, as their abundance and diversity has the potential to contribute to genetic diversity and subsequent adaptations and divergence within and between species (Ding et al. 2016; Schrader and Schmitz 2019). We find that over three quarters of the *D. tinctorius* genome consists of repetitive elements and that different elements can be found within—or overlapping—both exons and introns (Tables 1 and 2). This finding is consistent with the few studies that have generated chromosome-level genome assemblies for other amphibians. Notably, analyses of the Mexican axolotl (*Ambystoma mexicanum*; 32 Gb genome) and Tibetan frog (*Nanorana parkeri*; 2 Gb genome) genomes have reported LTR elements as the most abundant class of repetitive elements (Sun et al. 2015; Nowoshilow et al. 2018). By contrast, in the relatively small amphibian genome of the Western clawed frog (*X. tropicalis*; 1.5 Gb), LTR elements are less abundant, while DNA transposons are more abundant (Hellsten et al. 2010). The fact that LTR retroelements are abundant in large amphibian genomes has led to the hypothesis that these elements play a particularly important role in the evolution of “genomic gigantism”? (Sun et al. 2012). However, the mechanism underlying the expansion of LTR elements in large amphibian genomes, or the phenotypic consequences of their proliferation, remains to be tested. Our analyses of LTR abundances, size, and insertion times corroborate past findings (Sun et al. 2015; Nowoshilow et al. 2018) and indicate that LTRs have played a significant role in the genome size evolution of *D. tinctorius* (Fig. 4).

In addition to amphibians as a whole, transposable elements have likely played an important role in the evolution of genome size among poison frogs (family Dendrobatidae). For example, analyses of a fragmented assembly of *O. pumilio* estimated that this genome consists of over 4.5 Gb of repetitive sequence (Rogers et al. 2018). Consistent with our findings in *D. tinctorius*, Rogers et al. (2018) reported *Ty3/DIRS1* LTRs (referred to as “Gypsy-like”?) and *Tc1* retroelements as among the most abundant (1.0 Gb and 250 Mb of sequence, respectively). Rogers et al. (2018) also found that *Ty3* and *Tc1* elements are actively expressed in the oocytes of *O. pumilio*. Taken with the evidence of recent insertion events we estimate here (Fig. 4), it is likely that many elements are actively evolving and contributing to ongoing genome

size evolution in dendrobatid frogs. Research into the mechanisms governing TE insertion and regulation in anurans will be important to understand the mechanisms underlying genome size evolution in this group.

SINEs are a highly abundant class of transposable element in mammals and many other vertebrates, although they are almost entirely absent in amphibian genomes (Chalopin et al. 2015; Zuo et al. 2023), including *D. tinctorius*. This difference illustrates how the abundance of different transposable elements varies greatly across the tree of life. Interestingly, the majority of transposable elements we annotated in the *D. tinctorius* genomes could not be classified as known elements using our approach (39.15%). This is likely because reference libraries of described transposable elements used in annotation pipelines lack amphibian-specific elements, and this pattern of abundant “unclassified”? elements is common in studies of amphibian genomes (Sotero-Caio et al. 2017; Zuo et al. 2023). Future work that leverages ever-increasing amphibian genomic resources to describe the diversity and structure of the amphibian “dark matter”? elements would open doors for comparative analyses of transposable element evolution across taxa and greatly increase our understanding of the evolutionary history of parasitic DNA across the tree of life.

An ultimate consequence of TE evolution is their effect on genome size. For example, genome size increases with the abundance of repetitive elements in anurans (Zuo et al. 2023). Our results show that effects on genome size are not restricted to intergenic regions, with TEs being found within both exons and introns (Tables 1 and 2). Analyses of the Mexican axolotl, *A. mexicanum*, have shown a similar pattern, with introns being ten times larger in this species of salamander compared to other vertebrates (Smith et al. 2009). We also found that TEs affect the size of gene regions within the *D. tinctorius* genome (Fig. 3). These findings are contrary to a lack of relationship between gene size and genome size reported in a recent comparison of 14 anuran genomes (Zuo et al. 2023). A possible explanation for this discrepancy is that the largest poison frog genome, in Zuo et al. (2023)’s analysis, was found to contain significant contamination and has since been retracted (Stuckert et al. 2021, retracted).

While it is now clear that TEs comprise a large proportion of many amphibian genomes, and they are not restricted to intergenic regions (Sotero-Caio et al. 2017), there is still a knowledge gap in our understanding of how TEs influence the evolution and adaptability in this group. The fact that many repetitive elements are within or overlap gene features (introns and exons; Tables 1 and 2) suggests that their evolution may have important phenotypic consequences. For example, studies in other non-amphibian species have shown that TEs can have diverse effects on gene expression (Lanciano and Cristofari 2020; Rech et al. 2022); however, more work is needed to understand the phenotypic

consequences of TEs across ecologically and behaviorally diverse amphibian species. TEs have also been shown to contribute to the evolution of color phenotypes in a wide range of animals (Galbraith and Hayward 2023), yet links between TEs and coloration in amphibians are largely lacking (Galbraith and Hayward 2023). A possible mechanism linking aspects of genome size evolution—such as the accumulation of TEs in introns—and phenotypic evolution is the effect that intron size can have on gene expression (Castillo-Davis et al. 2002; Taft et al. 2007). Future work that leverages an increasing number of amphibian genomes with transcriptomic analysis—for example analyzing expression of different isoforms across species and developmental stages—could provide a fruitful avenue toward addressing the “consequences of genome-size evolution”? knowledge gap.

### Multilocus Genotyping Using MobiSeq

We used our genome assembly to show that leveraging repetitive elements to design and anchor primers may be a useful way to generate multilocus genotypic data at the population scale. These types of datasets would open exciting possibilities for understanding the adaptive processes and evolutionary dynamics of amphibian populations. The low amount of DNA that is needed for sequencing over larger parts of the genome makes MobiSeq a powerful tool for noninvasive sampling of target species. The target primers can be developed from the reference genome (if available), or from closely related species, as we have shown in our study. However, we recommend using a reference genome whenever possible to design species-specific primers and to map reads and genotype SNPs.

Our test of the MobiSeq approach showed that primers can differ considerably in their amplification and sequencing success. The element LINE109 could have been not abundant enough throughout the genome to give enough sequences: the LINE109 element was found 109 times in the *D. tinctorius* assembly, while the unknown transposable element TE644 was found 644 times. Therefore, we suggest that future marker development should design primers for markers which have at least 400 to 600 copies in the genome.

When genotyping and calling SNPs, de novo assembly with the program stacks gives better results than mapping to the *O. pumilio* genome where the primers were developed from (the *D. tinctorius* assembly was not available when we started the MobiSeq test). Although this is a closely related species, high variability in TE composition could contribute to poorer mapping and genotyping when relying on the *O. pumilio* genome. The de novo approach also gave better results than the mapping to the new *D. tinctorius* reference genome. A limitation of the stacks method we used for genotyping is that this approach resulted in high genotype missingness. Additionally, the

stacks program was developed for restriction site digested (RAD) approaches; therefore, we highly recommend using the original ANGSD pipeline, provided by the team that developed MobiSeq when genotyping MobiSeq data (Rey-Iglesia et al. 2019). This method gave the best results considering genotyping and further analysis with COLONY. Specifically, filtering SNPs to those called within a minimum of 70 (out of 92) individuals gave the most reliable relationship estimates in terms of identifying 5/5 of the duplicates in our dataset, as well as giving realistic population-size estimates, given our knowledge of the sample population and the biology of *D. tinctorius*.

MobiSeq was developed using mammalian genomes, which have lower levels of transposable elements, and focused on SINEs and LINEs, which are less common in amphibians (Sotero-Caio et al. 2017). Therefore, there are challenges in applying this method to amphibians with large genomes and high transposable element content, such as primer selection and mapping to an existing reference genome.

### Conclusions

Transposable elements are a major component of amphibian genomes and they play an important role in genome size evolution. By generating and analyzing a long-read assembly of the poison frog, *D. tinctorius*, we have shown that TE evolution impacts genome size, not only through their insertion in intergenic regions, but also within exons and introns. This “ecology”? of TEs provides a possible mechanism that links genome size evolution to phenotypic evolution. We also provide an example use of MobiSeq to generate low-input cost-effective population genetic data. This method could be used to study the evolutionary dynamics of amphibian populations, alongside aiding their conservation. Overall, our study adds to the growing body of knowledge on the evolution of amphibian genomes. We hope that the data and analyses we report here will be a valuable resource for future studies of amphibian genetics, evolution, behavior, and conservation.

### Materials and Methods

#### Sample Collection, DNA Extraction, and Sequencing for Whole-Genome Assembly

We generated a reference genome for *D. tinctorius* from a single male of the “azureus”? morph/population. This individual was captive bred by hobbyists in the United Kingdom, kept under license of the Home Office at Bangor University, and sacrificed by overdose of tricaine methanesulfonate followed by immediate pithing and decapitation. DNA was extracted in four parallel extractions performed at the NERC Environmental Omics Facility (NEOF) at the University of Sheffield using the

Macherey-Nagel NucleoBond High Molecular Weight DNA kit (see [supplementary SI, Supplementary Material](#) online for details). DNA was then cleaned and sheared to an average size of 19 kb before generating four HiFi libraries using SMRTbell template express kit 2.0 (PacBio). Final libraries were size-selected in the size range of 7 to 50 kb and sequenced across 18 SMRT cells (see [supplementary SI, Supplementary Material](#) online for details).

### Genome Assembly, Quality Control, and Synteny

We assembled HiFi reads greater than 10 kb in length using HiFiASM (Cheng et al. 2021, 2022; v0.16.1-r375) run with a bloom filter of 39 bits (option -f f39) and “aggressive” purging of haplotigs (option -l 2). To identify and correct assembly errors, we ran our primary assembly through the Inspector pipeline using `inspector.py` and `inspector-correct.py` scripts, respectively (Chen et al. 2021). Inspector also provides an estimate of assembly accuracy, reported as a Quality Value ( $QV = -10\log_{10}(\text{base-level errors}/\text{total assembly length})$ ). After error-correction, we identified and removed bacterial or viral contaminants using Kraken2 (Wood et al. 2019; v2.1.2) run using the standard Refseq database containing archaea, bacteria, viral, plasmid, human, and UniVec\_Core indexes (downloaded 27-5-2021). We estimated assembly completeness using BUSCO (Simão et al. 2015; v5.2.2) run in genome mode using the tetrapoda\_odb10 dataset, which consists of 5,310 single-copy orthologs derived from 38 genomes (created 2021-02-19). All other BUSCO options were left as default. We also compared assembly statistics of the *D. tinctorius* genome to the only other Dendrobatid genome currently available on NCBI (*O. pumilio*) and four chromosome-scale assemblies of other Anurans (*Bufo bufo* [NCBI RefSeq ID: GCF\_905171765.1], *R. temporaria* [RefSeq ID: GCF\_905171775.1], *E. pustulosus* [GenBank ID: GCA\_019512145.1], and *X. tropicalis* [RefSeq ID: GCF\_000004195.4]).

### Synteny Between *D. tinctorius* and Other Hyloidea Genomes

We compared the *D. tinctorius* assembly to chromosome-scale assemblies of *B. gargarizans* (Asiatic toad; Bufonidae; GenBank accession: [GCA\_014858855.1]), *E. coqui* (common coquí; Eleutherodactylidae; [GCA\_019857665.1]), and *E. pustulosus* (Túngara frog; Leptodactylidae; [GCA\_019512145.1]) using orthology-guided synteny map construction in GENESPACE (Lovell et al. 2022; v1.3.1) and BUSCO (Simão et al. 2015; v5.2.2) annotations of the 5,310 gene tetrapoda\_odb10 dataset (created 2021-02-19) as input. BUSCO annotations for *X. tropicalis* (GCA\_000004195.4) were used as outgroup sequences in the initial GENESPACE analysis. Because we were interested in identifying large regions of synteny, we constrained our analysis to scaffolds that have been assigned to

chromosomes in the chromosome-scale assemblies, and contigs at least 20 Mb long in the *D. tinctorius* assembly. This approach resulted in the number of annotated single-copy BUSCOs per genome being: 4,482, 3,579, 3,337, 3,440, and 4,939 for *B. gargarizans*, *E. pustulosus*, *E. coqui*, *D. tinctorius*, and *X. tropicalis*, respectively. We present summaries of synteny between *D. tinctorius* and the three chromosome-scale assemblies as the proportion of the *D. tinctorius* assembly contained within a synteny block and the average size of synteny blocks between genomes. Syntenic alignments were visualized using GENESPACE’s `plot_riparian` function (Lovell et al. 2022; v1.3.1).

We also tested whether the size of annotated BUSCO genes was positively correlated with genome size using a Spearman’s rank correlation test. We included the five genomes listed above and the chromosome-scale assembly of *R. temporaria* (GCA\_905171775.1). We focused this analysis on BUSCO annotations rather than de novo annotations generated for each genome because we annotated BUSCO genes using the same pipeline in all genomes, and BUSCO genes are chosen based on orthology across different vertebrate genomes, thereby reducing the likelihood that comparisons between paralogous genes affected our results. We note that restricting this analysis to conserved BUSCO genes could result in biases in gene size and evolutionary rate compared to all genes present in these genomes. To test whether the size of BUSCO gene regions correlated with genome size, we carried out Spearman’s rank correlations on the raw data as well as on phylogenetically independent contrasts (PICs) of both genome and BUSCO gene region sizes. PICs were calculated using the `pic()` function from the *ape* package in R (Paradis and Schliep 2019) and a pruned version of the phylogeny provided by Hime et al. (2021). Finally, we compared the *B. gargarizans* assembly—as it was the closest in size to that of *D. tinctorius*—to *D. tinctorius* using a non-ortholog based nucleotide alignment with `minimap2` (Li 2018, 2021; v2.24-r1122; options: -x asm20 -l10G -B3 -O4,24 -N10).

### Genome Annotation and Transposable Element Diversity

We annotated repetitive elements in our *D. tinctorius* assembly by first using RepeatModeler (Flynn et al. 2020; v2.0.3) with `rmblast` (v2.11.0+) to identify and generate a species-specific repeat library, followed by RepeatMasker (Smith et al. 2021; v 4.1.2-p1) run using the custom repeat library generated by RepeatModeler. RepeatModeler was run with default settings and RepeatMasker was run with `-xsmall -a -gff` options. Rather than conducting LTR structural analysis with RepeatMasker (`-LTRstruct` option), we independently annotated LTR retrotransposons using LtrDetector (Valencia and Girgis 2019). We used BLAST and the custom repeat library generated by RepeatModeler



to classify LTR retroelements annotated by LtrDetector. To estimate LTR insertion times, we extracted left and right LTR regions from the *D. tinctorius* genome in fasta format using bedtools getfasta function (Quinlan and Hall 2010) and the annotated coordinates generated by LtrDetector. We retained annotations where BLAST was able to match greater than 50% of the LTR retroelement with an element contained within the library generated by RepeatModeler (Flynn et al. 2020; v2.0.3). We then aligned left and right elements for each LTR retroelement using MAFFT (Kato and Standley 2013; v7.490; options: --globalpair --maxiterate 1000) and estimated insertion times from divergence between the left and right LTR, which we estimated from each MAFFT alignment using the dist.dna function in R run assuming Kimura's two-parameter model (Kimura 1980; option: model = "K80"?). To convert divergence estimates to time, we assumed a substitution rate of  $2.5 \times 10^{-9}$  substitutions per site per year (Lau et al. 2020).

In addition to repeat elements, we annotated protein-coding genes de novo using the BRAKER2 pipeline (Brůna et al. 2021; v2.1.6). Before annotation, we masked the genome using RepeatModeler and RepeatMasker with default settings. We ran BRAKER2 using evidence from 22 RNAseq libraries from brain, eggs, gut, liver, and skin. RNAseq reads were trimmed using Trimmomatic (Bolger et al. 2014; v. 0.39) to remove adaptor contamination and low-quality bases (options: LEADING:9 TRAILING:9 SLIDINGWINDOW:4:15 MINLEN:80). Trimmed reads were then mapped to the genome using STAR (Dobin et al. 2013; v. 2.7.8a) specifying the twopass Mode Basic option. Mapped RNAseq reads were then used as evidence in the BRAKER2 pipeline. UTR predictions were added to these predictions using the --addUTR=on option. Finally, UTR annotations were trimmed following a gap of over 1,000 bp. We used these annotations—alongside those generated by RepeatModeler + RepeatMasker—to test whether repetitive elements played a role in the evolution of intron/exon structure in the *D. tinctorius* genome using bedtools intersect (Quinlan and Hall 2010) with options -wa for the full comparison or -f 0.75 -r when testing for 75% reciprocal overlap between annotated gene features and repetitive elements.

### Tadpole Sampling and DNA Extraction for Relatedness Analysis

The aim of this part of the study was to test the MobiSeq protocol as a method for generating a reduced representation library in a non-model organism, the dyeing poison frog (*D. tinctorius*). We wanted to use SNP data to resolve the relatedness between tadpoles in different, small, confined environments such as phytotelmata (small bodies of water in tree holes or palm bracts, for natural history, see Rojas and Pašukonis 2019).

We sampled 87 tadpoles of *D. tinctorius* from 17 phytotelmata in the Nouragues Nature Reserve, French Guiana in 2020, by clipping a small part of the tail tip. Tissue samples were stored in 70% ethanol (EtOH) at  $-20^{\circ}\text{C}$  until further processing. To increase reliability and to test the appropriateness of the relatedness analysis, we included five samples as duplicates, giving a total sample size of 92. A modified salting-out method was used to extract DNA from small parts of the tail clips (supplementary methods, Supplementary Material online, section 1). The extracted DNA was eluted in 100  $\mu\text{l}$  TE buffer, DNA concentration measured with a NanoDrop spectrophotometer (Thermo Fisher) and stored at  $-20^{\circ}\text{C}$  until further processing.

### MobiSeq Primer Design

We designed specific primers for repeated elements as MobiSeq uses TEs in the DNA for target enrichment PCRs. No published reference genome for *D. tinctorius* was available when we started to work on the MobiSeq approach. To find highly repetitive TEs, we used the genome of a closely related species, the strawberry poison frog (*O. pumilio*), which was published in 2018 (Rogers et al.; NCBI GenBank: [GCA\_009801035.1]).

We used the free software RepeatMasker (Smith et al. 2021; v 4.1.2) to mask repeated elements and extracted those regions with Samtools (Danecek et al. 2021) to get the 22 last bp of transposable elements as a reversed complement and the number of occurrences. The list was cross checked and annotated with existing libraries of transposable element families provided by Dfam (Storer et al. 2021; v3.2). Based on this list, we choose elements with an average abundance of 400 to 600 times in the genome, as elements with higher abundances might be clustered and not evenly spread over the genome. LINEs and SINEs were generally less abundant so that we included only one primer for a lower abundance LINE. We initially chose six possible reverse transposable element primers and tested amplification and multiplexing with *D. tinctorius* DNA. We decided on the following two primers after running a target enrichment PCR, based on amplification success and minimization of primer dimers (AdapterTesequence):

- D\_tinct\_Line\_109  
(5'GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTATGTTACTATGTTACTATGT'3)
- D\_tinct\_TE\_644  
(5'GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTACTTTGGCCACCACTGTA'3).

In the first step of the MobiSeq protocol, the *D. tinctorius* DNA was digested using fragmentase. To ensure an even digestion across samples, the DNA was diluted to 10 ng/ $\mu\text{l}$ . We tested the correct incubation time for the fragmentase enzyme beforehand and decided on 20 min at  $37^{\circ}\text{C}$  for

each sample based on gel images (supplementary methods, Supplementary Material online, section 4.1). After fragmentation treatment, a Sera-Mag Speed beads clean-up was used to remove the enzyme mix and buffers (supplementary methods, Supplementary Material online, section 4.2). Clean-up was followed by end-repair of DNA fragments to create blunt ended fragments using T4 polymerase. After incubation, samples were cooled down to 10 °C and immediately used for the next step (supplementary methods, Supplementary Material online, section 4.3). For adapter ligation, double stranded modified P5 adapters (Meyer and Kircher 2010) were added to the end-repaired DNA fragments via a T4 DNA ligase. The modification of the adapter results in single-stranded adapters, which allow the use of universal adapter primers in the next target enrichment PCR step. The adapter ligation master mix (4 µl per sample) is added to the end-repaired samples. The samples were cooled down to 10 °C after the incubation (supplementary methods, Supplementary Material online, section 4.4). The resulting product was cleaned again, using the same Sera-Mag Speed beads as after DNA fragmentation (supplementary methods, Supplementary Material online, section 4.2) and eluted in 25 µl AE buffer and stored at 4 °C. We used the primers TE644 and Line109 as described above to enrich the fragments containing transposable elements in a multiplex approach (supplementary methods, Supplementary Material online, section 4.5). The remaining 15 µl of PCR was cleaned using the Sera-Mag Speed beads and eluted in 20 µl AE buffer. In the final step of the sequencing library preparation, a second PCR using Illumina indexed forward and reverse primers was used followed by agarose gel electrophoresis (supplementary methods, Supplementary Material online, section 4.6). The uniquely tagged samples were mixed, based on the brightness of the smear on the gel. The whole pool of samples was loaded on another gel to cut out fragments of the lengths between 200 and 500 bp. The innuPREP DOUBLEpure kit (Analytik Jena) was used to extract the DNA from the gel fragment, following the kit's protocol. The final library pool was sent to Novogene UK for sequencing in 150 bp paired end mode on a Novaseq instrument (Illumina).

### Bioinformatics for Variant Calling

After downloading the demultiplexed fastq files from Novogene, we visually checked the reads for quality with qiime2 (Bolyen et al. 2019; v2021.4) using the demux plugin (function summarize). As each fastq file contains Line109 and TE644 reads, we used cutadapt (Martin 2011; v4.1) to separate each file into two, based on the primer sequence. After separating the files, we used cutadapt to remove all versions of adapters and primers from the forward and reverse reads. Reverse-complement versions were used to remove primer fragments that can be present on the 5' end of

reads from fragments that are shorter than 150 bp. After trimming, the sequences were filtered for optical duplicates and remains of bacteriophage PhiX contamination using the shell scripts clumpify.sh and bbduke.sh from BBTools (Bushnell 2014).

For reduced representation sequencing approaches such as MobiSeq to be broadly applicable to non-model species that lack genomic resources, it is important to know how reliable markers are on the availability of genomic resources of closely related taxa. As such, we took three approaches to calling SNPs in our dataset: (i) de novo assembly of MobiSeq fragments, (ii) mapping reads to the *O. pumilio* reference, and (iii) mapping to the *D. tinctorius* genome we generated as part of this study. The de novo assembly was conducted twice using denovo\_map.pl from the stacks pipeline (Catchen et al. 2011, 2013; v2.64), using the default parameters. Therefore, all sequences were truncated to 50 and 100 bp, respectively, for both primers using process\_radtags. The reason for the double truncation is that the truncation process discards sequences shorter than the chosen length (50 or 100 bp). The 50 bp approach therefore retains more but shorter sequences than the 100 bp approach. We wanted to compare different truncation lengths to validate the robustness of de novo assembly. The reference genome mapping was done with ref\_map.pl from the stacks pipeline (Catchen et al. 2011, 2013; v2.64), using the default parameters. Before mapping, the reference genomes were indexed with the BWA-mem2 index function (Vasimuddin et al. 2019; v2.2.1) and subsequently sequences aligned with the mem function. The output was saved as a vcf file, with the following parameters: populations: -p 1 -r 0 --write-random-snp --max-obs-het 0.5 --ordered-export --vcf.

The soft filtering of SNP data was done in R (R Core Team 2022; v4.2.2) with Bioconductor (Morgan 2023; v3.16; BiocManager 1.30.19) and the filter\_rad function from the radiator package (Gosselin 2023; v1.2.8). To implement the radiator package from github, we used the package devtools (Wickham et al. 2022). All thresholds that we used for filtering can be found in supplementary methods, Supplementary Material online, section 5. We expected a high degree of missing data due to the MobiSeq approach, therefore the filter for maximum missingness was quite high (0.9).

The strong filtering of SNP data was done with vcftools (Danecek et al. 2011; v0.1.16) using the following parameters: --max-missing 0.5 --mac 2 --minDP 3.

Additionally, we used the ANGSD pipeline from Rey-Iglesia et al. (2019) (<https://github.com/shyamsg/MobiSeq/blob/master/code/pipeline.sh>), accounting for the fact that stacks was developed for restriction enzyme based RAD sequencing and might bias our output. We used the BWA-mem2 indexed reference genome of *D. tinctorius* to map our sequences, bedtools (Quinlan and Hall 2010; v2.26.0) to merge reads and kept only sites that are present in 90% of the cases. We called variants using ANGSD (Korneliussen

et al. 2014; v0.940) with minimum quality of 30, min mapping quality of 30, filter for SNPs with a  $P$ -value of  $1e^{-6}$ , major and minor allele were inferred directly from likelihoods, minor allele frequency was estimated (fixed major unknown minor), and genotypes and SNPs called. SNPs needed to be present in either a minimum of 80 or 70 individuals. We randomly called one SNP per contig to avoid linkage and created five vcf files for the 80 and 70 individuals dataset, respectively.

### Relatedness Analysis

The relatedness analysis was conducted with COLONY (Jones and Wang 2010; v2.0). We used the filtered SNPs of the TE644 primer only, derived from our four approaches (de novo stacks, *O. pumilio* reference genome or our new *D. tinctorius* genome) and two methods of calling SNPs (stacks vs. ANGSD). The write\_colony function in the radiator package was used to write a colony input function. The input parameters were as follows; the mating system of males and females was set to polygamous, no inbreeding and no update of allele frequencies, and the length of the run was set to 2 with full likelihood analysis and high precision. We ran an analysis with different random seeds to increase reliability of the clustering (1234 and 1789).

### Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

### Acknowledgments

We thank Lia Schlippe for providing drawings of *Oophaga pumilio* and *Dendrobates tinctorius*. Silhouettes of different frogs presented in Figs. 1, 2, and 3 were taken from the PhyloPic website ([phylopic.org](http://phylopic.org)). The DNA icon was taken from [flaticon.com](http://flaticon.com). This work was supported by funding from the NERC Environmental Omics Facility (project number 1382 to A.A.C.); Research Council of Finland (# 318404; 319949; 345974 to B.R.); and the National Science Foundation (IOS-1822025 and IOS-1827333 to L.A.O.).

### Author Contributions

A.A.C., S.S., and B.R. conceived and designed the project. R.M., C.A.F., and A.A.C. collected samples. C.D., F.H., L.S.K., and M.H. extracted DNA and prepared the libraries. S.S., L.A.O., A.F., and A.A.C. provided resources (data, computational). D.J.P., F.H., C.D., and A.A.C. performed bioinformatic analysis of the data. C.D., A.A.C., F.H., B.R., and S.S. discussed and interpreted the results, A.A.C., C.D., F.H., and L.A.O. wrote the first draft. A.A.C. and B.R. provided funding. All authors read, reviewed, and approved the final manuscript.

### Data Availability

The *D. tinctorius* genome assembly and BRAKER2 annotations are available on the Open Science Framework project page: [https://osf.io/tg24j/?view\\_only=13be51ec97184d888a12d4ba26dfc90f](https://osf.io/tg24j/?view_only=13be51ec97184d888a12d4ba26dfc90f) and the genome and HiFi reads are available on NCBI: BioProject: PRJNA977854; BioSample: SAMN381173 64. Illumina reads generated from MobiSeq libraries are publicly available on NCBI: BioProject: PRJNA1035148.

### Literature Cited

- Almojil D, Bourgeois Y, Falis M, Hariyani I, Wilcox J, Boissinot S. The structural, functional and evolutionary impact of transposable elements in eukaryotes. *Genes* (Basel). 2021;12(6):918. <https://doi.org/10.3390/genes12060918>.
- Amézquita A, Flechas SV, Lima AP, Gasser H, Hödl W. Acoustic interference and recognition space within a complex assemblage of dendrobatid frogs. *Proc Natl Acad Sci USA*. 2011;108(41):17058–17063. <https://doi.org/10.1073/pnas.1104773108>.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*. 2019;37(8):852–857. <https://doi.org/10.1038/s41587-019-0209-9>.
- Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, et al. Ten things you should know about transposable elements. *Genome Biol*. 2018;19(1):1–2. <https://doi.org/10.1186/s13059-018-1577-z>.
- Brandies P, Peel E, Hogg CJ, Belov K. The value of reference genomes in the conservation of threatened species. *Genes* (Basel). 2019;10(11):846. <https://doi.org/10.3390/genes10110846>.
- Bredeson JV, Mudd AB, Medina-Ruiz S, Mitros T, Smith OK, Miller KE, Lyons JB, Batra SS, Park J, Berkoff KC, et al. Conserved chromatin and repetitive patterns reveal slow genome evolution in frogs. *Nat Commun*. 2024;15(1):579. <https://doi.org/10.1038/s41467-023-43012-9>.
- Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform*. 2021;3(1):lqaa108. <https://doi.org/10.1093/nargab/lqaa108>.
- Bushnell B. BBMap short read aligner, and other bioinformatic tools. 2014 [accessed 2023 Jul 26]. <https://sourceforge.net/projects/bbmap/accessed>.
- Carvajal-Castro JD, Vargas-Salinas F, Casas-Cardona S, Rojas B, Santos JC. Aposematism facilitates the diversification of parental care strategies in poison frogs. *Sci Rep*. 2021;11(1):19047. <https://doi.org/10.1038/s41598-021-97206-6>.
- Casacuberta E, González J. The impact of transposable elements in environmental adaptation. *Mol Ecol*. 2013;22(6):1503–1517. <https://doi.org/10.1111/mec.12170>.
- Castillo-Davis CI, Mekhedov SL, Hartl DL, Koonin EV, Kondrashov FA. Selection for short introns in highly expressed genes. *Nat Genet*. 2002;31(4):415–418. <https://doi.org/10.1038/ng940>.
- Catchen J, Amores A, Hohenlohe P, Cresko W, Postlethwait J. Stacks: building and genotyping loci de novo from short-read sequences. *G3* (Bethesda). 2011;1(3):171–182. <https://doi.org/10.1534/g3.111.000240>.

- Catchen J, Hohenlohe P, Bassham S, Amores A, Cresko W. Stacks: an analysis tool set for population genomics. *Mol Ecol*. 2013;22(11):3124–3140. <https://doi.org/10.1111/mec.12354>.
- Chalopin D, Naville M, Plard F, Galiana D, Volff JN. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biol Evol*. 2015;7(2):567–580. <https://doi.org/10.1093/gbe/ew005>.
- Chen Y, Zhang Y, Wang AY, Gao M, Chong Z. Accurate long-read de novo assembly evaluation with Inspector. *Genome Biol*. 2021;22(1):312. <https://doi.org/10.1186/s13059-021-02527-4>.
- Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods*. 2021;18(2):170–175. <https://doi.org/10.1038/s41592-020-01056-5>.
- Cheng H, Jarvis ED, Fedrigo O, Koepfli K-P, Urban L, Gemmill NJ, Li H. Haplotype-resolved assembly of diploid genomes without parental data. *Nature Biotechnol*. 2022;40(9):1332–1335. <https://doi.org/10.1038/s41587-022-01261-x>.
- Chouteau M, Summers K, Morales V, Angers B. Advergence in Müllerian mimicry: the case of the poison dart frogs of Northern Peru revisited. *Biol Lett*. 2011;7(5):796–800. <https://doi.org/10.1098/rsbl.2011.0039>.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>.
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. Twelve years of SAMtools and BCFtools. *GigaScience*. 2021;10(2):giab008 [33590861]. <https://doi.org/10.1093/gigascience/giab008>.
- Ding Y, Berrocal A, Morita T, Longden KD, Stern DL. Natural courtship song variation caused by an intronic retroelement in an ion channel gene. *Nature*. 2016;536(7616):329–332. <https://doi.org/10.1038/nature19093>.
- Robin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
- Donnelly MA. Effects of reproductive resource supplementation on space-use patterns in *Dendrobates pumilio*. *Oecologia*. 1989;81(2):212–218. <https://doi.org/10.1007/BF00379808>.
- Dufresnes C, Mazepa G, Rodrigues N, Brelsford A, Litvinchuk SN, Sermier R, Lavanchy G, Betto-Colliard C, Blaser O, Borzée A, et al. Genomic evidence for cryptic speciation in tree frogs from the Apennine Peninsula, with description of *Hyla perrini* sp. nov. *Front Ecol Evol*. 2018;2018:144. <https://doi.org/10.3389/fevo.2018.00144>.
- Feng YJ, Blackburn DC, Liang D, Hillis DM, Wake DB, Cannatella DC, Zhang P. Phylogenomics reveals rapid, simultaneous diversification of three major clades of Gondwanan frogs at the Cretaceous–Paleogene boundary. *Proc Natl Acad Sci USA*. 2017;114(29):E5864–E5870. <https://doi.org/10.1073/pnas.1704632114>.
- Feschotte C. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet*. 2008;9(5):397–405. <https://doi.org/10.1038/nrg2337>.
- Fischer EK, Alvarez H, Lagerstrom KM, McKinney JE, Petrillo R, Ellis G, O'Connell LA. Neural correlates of winning and losing fights in poison frog tadpoles. *Physiol Behav*. 2020;223:112973. <https://doi.org/10.1016/j.physbeh.2020.112973>.
- Fischer EK, O'Connell LA. Hormonal and neural correlates of care in active versus observing poison frog parents. *Horm Behav*. 2020;120:104696. <https://doi.org/10.1016/j.yhbeh.2020.104696>.
- Fischer EK, Roland AB, Moskowitz NA, Tapia EE, Summers K, Coloma LA, O'Connell LA. The neural basis of tadpole transport in poison frogs. *Proc Soc B*. 2019;286(1907):20191084. <https://doi.org/10.1098/rspb.2019.1084>.
- Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci USA*. 2020;117(17):9451–9457. <https://doi.org/10.1073/pnas.1921046117>.
- Fouilloux C, Serrano Rojas SJ, Carvajal-Castro JD, Valkonen JK, Gaucher P, Fischer M-T, Pašukonis A, Rojas B. Pool choice in a vertical landscape: tadpole rearing site flexibility in phytotelm-breeding frogs. *Ecol Evol*. 2021;11(13):9021–9038. <https://doi.org/10.1002/ece3.7741>.
- Funk WC, Zamudio KR, Crawford AJ. Advancing understanding of amphibian evolution, ecology, behavior, and conservation with massively parallel sequencing. In: Hohenlohe PA, Rajora OP, editors. *Population genomics: wildlife*. Cham: Springer; 2018. p. 211–254.
- Galbraith JD, Hayward A. The influence of transposable elements on animal colouration. *Trends Genet*. 2023;39(8):624–638. <https://doi.org/10.1016/j.tig.2023.04.005>.
- Gosselin T. radiator: RADseq Data Exploration, Manipulation and Visualization using R. R package version 1.2.8; 2023 [accessed 2023 Jul 26]. <https://thierrygosselin.github.io/radiator/>.
- Gozashti L, Feschotte C, Hoekstra HE. Transposable element interactions shape the ecology of the deer mouse genome. *Mol Biol Evol*. 2023;40(4):msad069. <https://doi.org/10.1093/molbev/msad069>.
- Guillory WX, Muell MR, Summers K, Brown JL. Phylogenomic reconstruction of the Neotropical poison frogs (Dendrobatidae) and their conservation. *Diversity (Basel)*. 2019;11(8):126. <https://doi.org/10.3390/d11080126>.
- Hawkins JS, Kim H, Nason JD, Wing RA, Wendel JF. Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. *Genome Res*. 2006;16(10):1252–1261. <https://doi.org/10.1101/gr.5282906>.
- Hellsten U, Harland RM, Gilchrist MJ, Hendrix D, Jurka J, Kapitonov V, Ovcharenko I, Putnam NH, Shu S, Taher L, et al. The genome of the Western clawed frog *Xenopus tropicalis*. *Science*. 2010;328(5978):633–636. <https://doi.org/10.1126/science.1183670>.
- Hime PM, Lemmon AR, Lemmon ECM, Prendini E, Brown JM, Thomson RC, Kratovil JD, Noonan BP, Pyron RA, Peloso PLV, et al. Phylogenomics reveals ancient gene tree discordance in the amphibian tree of life. *Syst Biol*. 2021;70(1):49–66. <https://doi.org/10.1093/sysbio/syaa034>.
- Homola JJ, Loftin CS, Cammen KM, Helbing CC, Birol I, Schultz TF, Kinnison MT. Replicated landscape genomics identifies evidence of local adaptation to urbanization in wood frogs. *J Heredity*. 2019;110(6):707–719. <https://doi.org/10.1093/jhered/esz041>.
- Hotaling S, Kelley JL, Frandsen PB. Toward a genome sequence for every animal: where are we now?. *Proc Natl Acad Sci USA*. 2021;118(52):e2109019118. <https://doi.org/10.1073/pnas.2109019118>.
- Jones O, Wang J. COLONY: a program for parentage and sibship inference from multilocus genotype data. *Mol Ecol Res*. 2010;10(3):551–555. <https://doi.org/10.1111/j.1755-0998.2009.02787.x>.
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–780. <https://doi.org/10.1093/molbev/mst010>.
- Kidwell MG. Transposable elements and the evolution of genome size in eukaryotes. *Genetica*. 2002;115(1):49–63. <https://doi.org/10.1023/A:1016072014259>.
- Kidwell MG, Lisch D. Transposable elements as sources of variation in animals and plants. *Proc Natl Acad Sci USA*. 1997;94(15):7704–7711. <https://doi.org/10.1073/pnas.94.15.7704>.
- Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol*. 1980;16(2):111–120. <https://doi.org/10.1007/BF01731581>.



- Korneliusson TS, Albrechtsen A, Nielsen R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics*. 2014;15(1):356. <https://doi.org/10.1186/s12859-014-0356-4>.
- Kosch TA, Waddle AW, Cooper CA, Zenger KR, Garrick DJ, Berger L, Skerratt LF. Genetic approaches for increasing fitness in endangered species. *Trends Ecol Evol*. 2022;37(4):332–345. <https://doi.org/10.1016/j.tree.2021.12.003>.
- Lamichhaney S, Catullo R, Keogh JS, Clulow S, Edwards SV, Ezaz T. A bird-like genome from a frog: mechanisms of genome size reduction in the ornate burrowing frog, *Platyplectrum ornatum*. *Proc Natl Acad Sci USA*. 2021;118(11):e2011649118. <https://doi.org/10.1073/pnas.2011649118>.
- Lanciano S, Cristofari G. Measuring and interpreting transposable element expression. *Nat Rev Genet*. 2020;21(12):721–736. <https://doi.org/10.1038/s41576-020-0251-y>.
- Lau Q, Igawa T, Ogino H, Katsura Y, Ikemura T, Satta Y. Heterogeneity of synonymous substitution rates in the *Xenopus* frog genome. *PLoS One*. 2020;15(8):e0236515. <https://doi.org/10.1371/journal.pone.0236515>.
- Lawrence JP, Rojas B, Fouquet A, Mappes J, Blanchette A, Saporito RA, Bosque RJ, Courtois EA, Noonan BP. Weak warning signals can persist in the absence of gene flow. *Proc Natl Acad Sci USA*. 2019;116(38):19037–19045. <https://doi.org/10.1073/pnas.1901872116>.
- Lee SI, Kim NS. Transposable elements and genome size variations in plants. *Genomics Inform*. 2014;12(3):87. <https://doi.org/10.5808/GI.2014.12.3.87>.
- Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
- Li H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics*. 2021;37(23):4572–4574. <https://doi.org/10.1093/bioinformatics/btab705>.
- Liedtke HC, Gower DJ, Wilkinson M, Gomez-Mestre I. Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate. *Nature Ecol Evol*. 2018;2(11):1792–1799. <https://doi.org/10.1038/s41559-018-0674-4>.
- Lovell JT, Sreedasyam A, Schranz ME, Wilson M, Carlson JW, Harkess A, Emms D, Goodstein DM, Schmutz J. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *eLife*. 2022;11:e78526. <https://doi.org/10.7554/eLife.78526>.
- Maan ME, Cummings ME. Poison frog colors are honest signals of toxicity, particularly for bird predators. *Am Nat*. 2012;179(1):E1–E4. <https://doi.org/10.1086/663197>.
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 2011;17(1):10–12. <https://doi.org/10.14806/ej.17.1.200>
- McClintock B. The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci USA*. 1950;36(6):344–355. <https://doi.org/10.1073/pnas.36.6.344>.
- Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc*. 2010;2010(6):pdb-rot5448. <https://doi.org/10.1101/pdb.prot5448>.
- Morgan M. BiocManager: Access the Bioconductor Project Package Repository. R package version 1.30.20; 2023 [accessed 2023 Jul 27]. <https://CRAN.R-project.org/package=BiocManager>.
- Myers CW, Daly JW. Dart-poison frogs. *Sci Am*. 1983;248(2):96–105. <https://doi.org/10.1038/scientificamerican0283-120>.
- Noonan BP, Comeault AA. The role of predator selection on polymorphic aposematic poison frogs. *Biol Lett*. 2009;5(1):51–54. <https://doi.org/10.1098/rsbl.2008.0586>.
- Nowoshilow S, Schloissnig S, Fei J-F, Dahl A, Pang AWC, Pippel M, Winkler S, Hastie AR, Young G, Roscito JG, et al. The axolotl genome and the evolution of key tissue formation regulators. *Nature*. 2018;554(7690):50–55. <https://doi.org/10.1038/nature25458>.
- Nunziata SO, Weisrock DW. Estimation of contemporary effective population size and population declines using RAD sequence data. *Heredity* (Edinb). 2018;120(3):196–207. <https://doi.org/10.1038/s41437-017-0037-y>.
- Paradis E, Schliep K. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*. 2019;35(3):526–528. <https://doi.org/10.1093/bioinformatics/bty633>.
- Pasquesi GIM, Adams RH, Card DC, Schield DR, Corbin AB, Perry BW, Reyes-Velasco J, Ruggiero RP, Vandeweghe MW, Shortt JA, et al. Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. *Nat Commun*. 2018;9(1):2774. <https://doi.org/10.1038/s41467-018-05279-1>.
- Pašukonis A, Serrano-Rojas SJ, Fischer M-T, Loretto M-C, Shaykevich DA, Rojas B, Ringler M, Roland AB, Marcillo-Lara A, Ringler E, et al. Contrasting parental roles shape sex differences in poison frog space use but not navigational performance. *eLife*. 2022;11:e80483. <https://doi.org/10.7554/eLife.80483>.
- Pimpinelli S, Piacentini L. Environmental change and the evolution of genomes: transposable elements as translators of phenotypic plasticity into genotypic variability. *Funct Ecol*. 2020;34(2):428–441. <https://doi.org/10.1111/1365-2435.13497>.
- Platt RN, Vandeweghe MW, Ray DA. Mammalian transposable elements and their impacts on genome evolution. *Chromosome Res*. 2018;26(1-2):25–43. <https://doi.org/10.1007/s10577-017-9570-z>.
- Pröhl H. Clutch loss affects the operational sex ratio in the strawberry poison frog *Dendrobates pumilio*. *Behav Ecol Sociobiol*. 2005;58(3):310–315. <https://doi.org/10.1007/s00265-005-0915-9>.
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2022 [accessed 2022 Jul 15]. <https://www.R-project.org/>. Version 4.2.1.
- Rech GE, Radó S, Guirao-Rico S, Aguilera L, Horvath V, Green L, Lindstadt H, Jamilloux V, Quesneville H, González J. Population-scale long-read sequencing uncovers transposable elements associated with gene expression variation and adaptive signatures in *Drosophila*. *Nat Commun*. 2022;13(1):1948. <https://doi.org/10.1038/s41467-022-29518-8>.
- Rey-Iglesia A, Gopalakrishnan S, Carøe C, Alquezar-Planas DE, Ahlmann Nielsen A, Röder T, Bruhn Pedersen L, Næsberg-Nielsen C, Sinding M-HS, Fredensborg Rath M, et al. MobiSeq: de novo SNP discovery in model and non-model species through sequencing the flanking region of transposable elements. *Mol Ecol Res*. 2019;19(2):512–525. <https://doi.org/10.1111/1755-0998.12984>.
- Rey O, Danchin E, Mirouze M, Loot C, Blanchet S. Adaptation to global change: a transposable element–epigenetics perspective. *Trends Ecol Evol*. 2016;31(7):514–526. <https://doi.org/10.1016/j.tree.2016.03.013>.
- Ringler E, Pašukonis A, Hödl W, Ringler M. Tadpole transport logistics in a Neotropical poison frog: indications for strategic planning and adaptive plasticity in anuran parental care. *Front Zool*. 2013;10(1):67. <https://doi.org/10.1186/1742-9994-10-67>.
- Rodríguez C, Fusani L, Raboisson G, Hödl W, Ringler E, Canoine V. Androgen responsiveness to simulated territorial intrusions in *Allobates femoralis* males: evidence supporting the challenge hypothesis in a territorial frog. *Gen Comp Endocrinol*. 2022;326:114046. <https://doi.org/10.1016/j.ygcen.2022.114046>.
- Rogers RL, Zhou L, Chu C, Márquez R, Corl A, Linderoth T, Freeborn L, MacManes MD, Xiong Z, Zheng J, et al. Genomic takeover by transposable elements in the strawberry poison frog. *Mol Biol*

- Evol. 2018;35(12):2913–2927. <https://doi.org/10.1093/molbev/msy185>.
- Rojas B. Strange parental decisions: fathers of the dyeing poison frog deposit their tadpoles in pools occupied by large cannibals. *Behav Ecol Sociobiol.* 2014;68(4):551–559. <https://doi.org/10.1007/s00265-013-1670-y>.
- Rojas B. Mind the gap: treefalls as drivers of parental tradeoffs. *Ecol Evol.* 2015;5(18):4028–4036. <https://doi.org/10.1002/ece3.1648>.
- Rojas B, Pašukonis A. From habitat use to social behavior: natural history of a voiceless poison frog, *Dendrobates tinctorius*. *PeerJ.* 2019;7:e7648. <https://doi.org/10.7717/peerj.7648>.
- Schloissnig S, Kawaguchi A, Nowoshilow S, Falcon F, Otsuki L, Tardivo P, Timoshevskaya N, Keinath MC, Smith JJ, Voss SR, et al. The giant axolotl genome uncovers the evolution, scaling, and transcriptional control of complex gene loci. *Proc Natl Acad Sci USA.* 2021;118(15):e2017176118. <https://doi.org/10.1073/pnas.2017176118>.
- Schrader L, Schmitz J. The impact of transposable elements in adaptive evolution. *Mol Ecol.* 2019;28(6):1537–1549. <https://doi.org/10.1111/mec.14794>.
- Schulte LM, Lötters S. The power of the seasons: rainfall triggers parental care in poison frogs. *Evol Ecol.* 2013;27(4):711–723. <https://doi.org/10.1007/s10682-013-9637-z>.
- Serrato-Capuchina A, Matute DR. The role of transposable elements in speciation. *Genes (Basel).* 2018;9(5):254. <https://doi.org/10.3390/genes9050254>.
- Sexton OJ. Some aspects of the behavior and of the territory of a dendrobatid frog, *Prostherapis trinitatis*. *Ecology.* 1960;41(1):107–115. <https://doi.org/10.2307/1931944>.
- Silverstone PA. Observations on the behavior and ecology of a Colombian poison-arrow frog, *Kôkoé-pá (Dendrobates histrionicus Berthold)*. *Herpetologica.* 1973;29:295–301. <http://www.jstor.org/stable/3891565>.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31(19):3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
- Smith A, Hubley R, Green P. RepeatMasker version 4.1.2-p1; 2021 [accessed 2022 Mar 14]. <http://repeatmasker.org>.
- Smith JJ, Putta S, Zhu W, Pao GM, Verma IM, Hunter T, Bryant SV, Gardiner DM, Harkins TT, Voss SR. Genic regions of a large salamander genome contain long introns and novel genes. *BMC Genomics.* 2009;10(1):19. <https://doi.org/10.1186/1471-2164-10-19>.
- Sotero-Caio CG, Platt RN, Suh A, Ray DA. Evolution and diversity of transposable elements in vertebrate genomes. *Genome Biol Evol.* 2017;9(1):161–177. <https://doi.org/10.1093/gbe/evw264>.
- Stitzer MC, Anderson SN, Springer NM, Ross-Ibarra J. The genomic ecosystem of transposable elements in maize. *PLoS Genet.* 2021;17(10):e1009768. <https://doi.org/10.1371/journal.pgen.1009768>.
- Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mob DNA.* 2021;12(1):2. <https://doi.org/10.1186/s13100-020-00230-y>.
- Stuckert AMM, Chouteau M, McClure M, LaPolice TM, Linderoth T, Nielsen R, Summers K, MacManes MD. The genomics of mimicry: gene expression throughout development provides insights into convergent and divergent phenotypes in a Müllerian mimicry system. *Mol Ecol.* 2021;30(16):4039–4061. Retraction in: *Mol Ecol.* 2024. 33(8):e17345. <https://doi.org/10.1111/mec.16024>.
- Stynoski JL, Schulte LM, Rojas B. Poison frogs. Quick guide. *Curr Biol.* 2015; 25(21):R1026–R1028. <https://doi.org/10.1016/j.cub.2015.06.044>.
- Stynoski JL, Torres-Mendoza Y, Sasa-Marin M, Saporito RA. Evidence of maternal provisioning of alkaloid-based chemical defenses in the strawberry poison frog *Oophaga pumilio*. *Ecology.* 2014;95(3):587–593. <https://doi.org/10.1890/13-0927.1>.
- Summers K. Sexual selection and intra-female competition in the green poison-dart frog, *Dendrobates auratus*. *Anim Behav.* 1989;37: 797–805. [https://doi.org/10.1016/0003-3472\(89\)90064-X](https://doi.org/10.1016/0003-3472(89)90064-X).
- Summers K, Sea McKeon C, Heying H. The evolution of parental care and egg size: a comparative analysis in frogs. *Proc R Soc B.* 2006;273(1587):687–692. <https://doi.org/10.1098/rspb.2005.3368>.
- Sun C, Shepard DB, Chong RA, López Arriaza J, Hall K, Castoe TA, Feschotte C, Pollock DD, Mueller RL. LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. *Genome Biol Evol.* 2012;4(2):168–183. <https://doi.org/10.1093/gbe/evr139>.
- Sun YB, Xiong Z-J, Xiang X-Y, Liu S-P, Zhou W-W, Tu X-L, Zhong L, Wang L, Wu D-D, Zhang BL, et al. Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. *Proc Natl Acad Sci USA.* 2015;112(11): E1257–E1262. <https://doi.org/10.1073/pnas.1501764112>.
- Taft RJ, Pheasant M, Mattick JS. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays.* 2007;29(3):288–299. <https://doi.org/10.1002/bies.20544>.
- Tumulty J, Morales V, Summers K. The biparental care hypothesis for the evolution of monogamy: experimental evidence in an amphibian. *Behav Ecol.* 2014;25(2):262–270. <https://doi.org/10.1093/beheco/art116>.
- Valencia JD, Girgis HZ. LtrDetector: a tool-suite for detecting long terminal repeat retrotransposons de-novo. *BMC Genomics.* 2019;20(1):450. <https://doi.org/10.1186/s12864-019-5796-9>.
- Vasimuddin M, Sanchit M, Heng L, Srinivas A. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. *IEEE Parallel and Distributed Processing Symposium (IPDPS)*, Rio de Janeiro, Brazil. *IEEE (Institute of Electrical and Electronics Engineers)*; 2019. <https://doi.org/10.1109/IPDPS.2019.00041>.
- Wells KD. Behavioral ecology and social organization of a dendrobatid frog (*Colostethus inguinalis*). *Behav Ecol Sociobiol.* 1980;6(3): 199–209. <https://doi.org/10.1007/BF00569201>.
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. A unified classification system for eukaryotic transposable elements. *Nature Rev Gen.* 2007;8(12):973–982. <https://doi.org/10.1038/nrg2165>.
- Wickham H, Hester J, Chang W, Bryan J. devtools: Tools to Make Developing R Packages Easier. R package version 2.4.5; 2022 [accessed 2023 Jul 30]. <https://CRAN.R-project.org/package=devtools>.
- Womack MC, Metz MJ, Hoke KL. Larger genomes linked to slower development and loss of late-developing traits. *Am Nat.* 2019;194(6): 854–864. <https://doi.org/10.1086/705897>.
- Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 2019;20(1):1–3. <https://doi.org/10.1186/s13059-019-1891-0>.
- Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwald MJ, Meredith R, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science.* 2014;346(6215): 1311–1320. <https://doi.org/10.1126/science.1251385>.
- Zuo B, Nneji LM, Sun YB. Comparative genomics reveals insights into anuran genome size evolution. *BMC Genomics.* 2023;24(1):379. <https://doi.org/10.1186/s12864-023-09499-8>.

Associate editor: Sarah Schaack