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

Dittrich, Carolin; Hoelzl, Franz; Smith, Steve; Fouilloux, Chloe A; Parker, Darren J; O'Connell, Lauren A; Knowles, Lucy S; Hughes, Margaret; Fewings, Ade; Morgan, Rhys; Rojas, Bibiana; Comeault, Aaron A

Genome Biology and Evolution

Peer reviewed version

[Cyswllt i'r cyhoeddiad / Link to publication](#)

24. Jun. 2024

1 **Title**: Genome assembly of the dyeing poison frog provides insights into

2 the dynamics of transposable element and genome-size evolution

3

4 Carolin Dittrich[1,2,8]*, Franz Hoelzl[2,8], Steve Smith[2], Chloe A. Fouilloux[1], Darren J. Parker[3],

5 Lauren A. O'Connell[4], Lucy S. Knowles[5], Margaret Hughes[6], Ade Fewings[7], Rhys Morgan[3],

6 Bibiana Rojas[1,2], Aaron A. Comeault[3]*

7

8 [1]Department of Biology and Environmental Sciences, University of Jyväskylä, Jyväskylä,

9 Finland

10 [2]Department of Interdisciplinary Life Sciences, Konrad Lorenz Institute of Ethology, University of

11 Veterinary Medicine, Vienna, Austria

12 [3]School of Environmental and Natural Sciences, Molecular Ecology & Evolution Group, Bangor

13 University, Bangor, UK

14 [4]Department of Biology, Stanford University, Stanford, USA

15 [5]University of Sheffield, Sheffield, UK

16 [6]University of Liverpool, Liverpool, UK

17 [7]Supercomputing Wales, Digital Services, Bangor University, Bangor, UK

18

19 [8]these authors contributed equally to this work

20

21 *Authors for Correspondence: Aaron Comeault, School of Environmental and Natural Sciences,

22 Bangor University, Bangor, UK, +44 1248 383216, a.comeault@bangor.ac.uk & Carolin Dittrich,

23 [2]Department of Interdisciplinary Life Sciences, Konrad Lorenz Institute of Ethology, University of

24 Veterinary Medicine, Vienna, Austria, carolin.dittrich@vetmeduni.ac.at.

1

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 optimise primer or bait design for cost-effective population-level genotyping in species with large, repetitive genomes.

2

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 Barbara McClintock (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 (Almojil et al. 2021; Casacuberta and González 2013; Feschotte 2008; Serrato-Capuchina and Matute 2018). It has also been suggested that the mobilisation and insertion of transposable elements can drive adaptation to global change when organisms are subject to environmental stress (Pimpinelli and Piacentini 2020; Rey et al. 2016). For a comprehensive classification and in-depth review of the impact transposable elements can have on the genome, we recommend the recent works of (Almojil et al. 2021; Bourge et al. 2018; Kidwell and Lisch 1997; Sotero-Caio et al. 2017; Wicker et al. 2007).

Characterising 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 (Pasquesi 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 non-random locations in the genome, such as in intergenic regions, promotors, 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 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; Homola et al. 2019; Nunziata and Weisrock 2018). 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 regards 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 behaviour 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 colouration coupled with chemical defences that protects some species from predation (Noonan and Comeault 2009; Chouteau et al. 2011; Maan and Cummings 2012; Lawrence and Rojas 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 behaviour, parental care, and space use, as well as the behaviour 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 et al. 2013; Tumulty et al. 2014; Rojas 2014; Stynoski et al. 2014; Carvajal-Castro et al. 2021; Fouilloux et al. 2021). More recently, studies on the proximate mechanisms of such behaviours,

1   e.g., neurobiology of egg provisioning, tadpole transport and tadpole aggression (Fischer and

2   O'Connell 2020; Fischer et al. 2019, 2020), and hormonal correlates of care, territoriality and

3   space use (Fischer and O'Connell 2020; Pašukonis et al. 2022; Rodríguez et al. 2022), now

4   provide a more holistic understanding of what makes these frogs unique. However, the lack of

5   genomic resources has precluded some of these topics from being addressed in depth. The one

6   genome available for dendrobatid frogs (*Oophaga pumilio*) is highly fragmented, but still shows

7   that transposable elements comprise a large portion of it (Rogers et al. 2018). Having access to

8   reference genomes from more poison frogs would allow for more comprehensive approaches to

9   questions related to demography, conservation, behavioural ecology, disease dynamics and

10   adaptation (Brandies et al. 2019).

11   In this study, we generate a high-quality reference genome for the dyeing poison frog,

12   *Dendrobates tinctorius*. This species is aposematic—with drastic differences in coloration and

13   toxicity across populations (Lawrence and Rojas et al. 2019)—and displays complex social

14   behaviours typical of many poison frog species including male parental care and territoriality

15   (Rojas and Pašukonis 2019; Fouilloux et al. 2021; Rojas 2014; 2015). We first leverage our

16   genome assembly and three publicly available chromosome-scale assemblies of species

17   belonging to Hyloidea to provide evidence for both large regions of synteny alongside significant

18   structural evolution. This analysis also revealed that the evolution of genome size across these

19   species is correlated with gene size, with *D. tinctorius* having both the largest genome and the

20   longest genes. We then annotate transposable elements in the *D. tinctorius* genome to explore

21   the genomic landscape of their evolution and find that transposable elements are more

22   abundant in introns than in exons, likely contributing to the evolution of large genes in *D.*

23   *tinctorius*. Finally, we explored the usefulness of the *D. tinctorius* reference genome in

24   leveraging population genetic information from cost-effective mobile element sequencing

25   (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 behavioural ecology and population genomic research in amphibians.

Results

Genome assembly, quality control

Our final *Dendrobates 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:5310). 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 coquí; Eleutherodactylidae), and *Engystomops pustulosus* (Túngara frog; Leptodactylidae). *Bufo gargarizans*, *El. coqui*, and *En. pustulosus* shared a common ancestor with *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 analysed (fig. 2). The number of synteny blocks identified between *D. tinctorius* and the chromosome-scale assemblies ranged from 116 (*En. pustulosus*) to 164 (*B. gargarizans*), with an average size of each block of 27.36 Mb (*En. 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 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 analysed: for example, Figure 2B highlights rearrangements between *D. tinctorius* scaffolds and *B. gargarizans*'s chromosome 1 (inversions highlighted in salmon pink colour). Finally, we observed a single instance where a *D. tinctorius* scaffold contained synteny blocks that mapped to different chromosomes of a chromosome-scale assembly (*El. 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*, *El. coqui*, *En. pustulosus*, *R. temporaria* and *Xenopus tropicalis*), with genome size ranging from 1.451 Gb (*X. tropicalis*) to 6.356 Gb (*D.*

9

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 analysed — *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* versus *D. tinctorius*) showing less difference in genes size than when genomes differed in size (i.e., *X. tropicalis* versus *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 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

1   assigned elements, 23.13% were retroelements and 12.45% were DNA transposons. The most

2   common retroelements classified by RepeatMasker were LTR retrotransposons (15.69% of the

3   assembly) and LINEs (7.4%). LTR retrotransposons Ty3*DIRS1* were the most abundant

4   superfamily (13.79% of the assembly), while DNA transposons made up 12.45% of the

5   assembly. Of DNA transposons, *Tc1-IS630-Pogo* (*Tc1*) and *hobo-Activator* (*hAT*) elements

6   were the most abundant (6.57% and 3.91% of the assembly, respectively).

7   Independent analyses with LtrDetector (Valencia and Girgis 2018) annotated 262,486

8   LTRs spanning 2,343,246,106 bp, or 36.9% of our assembly. While this represents a higher

9   percentage of LTRs in our assembly than when annotated with RepeatMasker, consistent with

10   RepeatMasker, the most abundant LTR elements were found to belong to the Ty3/*DIRS1*

11   superfamily (99,884 elements, 819.4 Mb total), followed by *BEL/Pao* elements (1,762 elements,

12   13.9 Mb total). Also consistent with results from RepeatMasker, 66,046 of the LTR elements

13   annotated by LtrDetector (spanning 588 Mb) had BLAST matches to the "Unknown" category of

14   elements identified by RepeatModeler.

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

16   *tinctorius* genome structure, we compared the location of repetitive elements relative to *de novo*

17   gene annotations we generated using the BRAKER2 pipeline (Brůna et al. 2021; v2.1.6). Of the

18   11,331,718 repeat annotations generated by RepeatMasker, 115,330 (1.02%) overlapped with

19   coding DNA sequence (CDS) and 4,872 of these had a reciprocal overlap of at least 75% (table

20   1 and table 2, respectively). More repetitive elements overlapped with introns compared to CDS

21   (525,441 elements; χ^2 = 270,118; P < 2.2 x 10-16); however, proportionally fewer TEs that

22   overlapped introns showed reciprocal overlaps of greater than 75% compared to those

23   overlapping CDS (1.05 versus 4.22% respectively; χ^2 = 5,956.1; P < 2.2 x 10-16). These

24   patterns are likely due to the fact that introns span nearly 4 times the number of bases and have

11

a median size nearly twice that of CDS features (266 Mb versus 69 Mb and 113 versus 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% versus 1.01%, respectively; introns: 5.27% versus 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 2.5×10^{-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 - 60.4 Mya) while *ERV1* elements had the youngest insertion times (mean = 10.4 Mya, 95% empirical range = 0 - 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. 2018), while others are young, potentially active, and species-specific (e.g. *Dendrobates* split from its sister genus *Oophaga* ~16 Mya; Guillory et al. 2019).

1 *A test of the utility of MobiSeq for population genetics*

2   We next used the *D. tinctorius* reference genome to compare the genotyping success of

3 three genotyping approaches using a MobiSeq dataset generated from 87 tadpoles collected

4 from 17 phytotelmata at the Nouragues Research Station, French Guiana. Specifically, we

5 compared the number of usable SNPs generated by either (1) de novo assembly of MobiSeq

6 reads, (2) mapping the reads to the other poison frog genome currently available on NCBI (*O.*

7 *pumilio*), or (3) mapping the reads to the *D. tinctorius* genome we generated as part of this

8 study. We also called SNPs using two approaches: either the stacks pipeline (Catchen et al.

9 2011, 2013; v2.64) designed to assemble RADseq data de novo or using a reference genome;

10 or the original MobiSeq pipeline using the program "analysis in next generation sequencing

11 data" (ANGSD; Korneliussen et al. 2014; v0.940) mapping sequence reads to the *D. tinctorius*

12 assembly (Rey-Iglesia et al. 2019) (details in Methods).

13   The number of SNPs differed considerably between the two primers and different

14 mapping approaches (de novo vs. reference genome) and showed high degrees of missing

15 genotypes when genotyped using the stacks pipeline (LINE109: 93.21 ± 3.06%, TE644: 90.46 ±

16 1.46%; mean percent missing data across sites ± sd) and low coverage (supplementary table

17 S1, Supplementary Material online). Regions amplified with the LINE109 primer resulted in

18 considerably fewer SNPs than the TE644 primer. When we mapped sequences to the *O.*

19 *pumilio* genome, gstacks only incorporated 10-12% of the reads and called considerably fewer

20 SNPs than either the de novo approach or mapping to the *D. tinctorius* genome. The de novo

21 approach with 50bp fragments called the highest numbers of SNPs (n SNPs TE644 = 92,433);

22 in contrast, the 100bp fragments called the lowest numbers (n SNPs TE644 = 15,987). These

23 differences were expected due to the large loss of SNPs due to truncation. Mapping to the *D.*

24 *tinctorius* reference genome for both primers called the second most SNPs (n SNPs; LINE109 =

25 2,476, TE644 = 82,863), but missingness was still high and coverage low (supplementary table

1 S1 and S2, Supplementary Material online). After applying a strong filter, where SNP had to be

2 present in at least 50% of individuals, the numbers of SNP dropped dramatically (table S2 and

3 fig. 5B).

4   The SNP calling with the ANGSD pipeline gave 17.67 ± 0.17% missing genotypes and

5 345 SNPs were called when the output was restricted to be present in at least 70 individuals. If

6 SNPs were restricted to be present in at least 80 individuals, 291 SNPs could be called with

7 11.07 ± 0.08% genotype missingness (fig. 5A and B).

8   We used our called SNP dataset to test their usefulness in parentage analysis.

9 Specifically, we ran a relatedness analysis with COLONY (Jones and Wang 2010; v2.0) on our

10 dataset of 87 tadpoles. We included 5 sets of duplicate samples to test whether these samples

11 will be reliably grouped together in our analysis. In general, we found significant differences in

12 COLONY results depending on which SNP dataset we used for the analysis. Due to the low

13 number of SNPs called with the LINE109 primers, we focused our COLONY analyses to data

14 obtained using the TE644 primers. We considered the successful resolution of our duplicate

15 samples, grouped together with a probability of 0.9 or greater, as evidence that a SNP data set

16 was reliable for parentage analysis. Generally, reliability was low when using SNPs that were

17 genotyped using Stacks: neither the de novo nor the reference genome mapping approaches

18 reliably identified the 5 duplicates in the dataset with high probability (fig. 5C). When genotyped

19 using the ANGSD pipeline, COLONY identified 4/5 duplicates when SNPs from at least 80

20 individuals were retained, and 5/5 duplicates when 70 individuals were the minimum threshold

21 for calling SNPs, suggesting that the more complete datasets generated with ANGSD—with

22 only 11-17% missing genotypes—provided sufficient information for identifying full-sibs (fig. 5A

23 and C). This method also gave more realistic estimated population sizes than the stacks method

24 (supplementary table S2, Supplementary Material online).

1 **Discussion**

2   Access to genomic resources and tools holds the potential to transform our

3 understanding of the ecology, evolution, life-history, and conservation of amphibians (e.g.

4 Liedtke et al. 2018; Womack et al. 2019; Schloissnig et al. 2021; Kosch et al. 2022); yet

5 amphibians have lagged behind other groups of animals in available genomic resources

6 (Hotaling et al. 2021). The lack of genomic resources for frogs and toads (Anurans), in

7 particular, is at least in part due to some species possessing large and complex genomes

8 (Rogers et al. 2018; Nowoshilow et al. 2018; Sun et al. 2015): among frogs and toads genome

9 size is highly variable ranging from 0.99 Gb in the plains spadefoot toad, *Spea bombifrons*

10 (GenBank accession GCA_027358695.2) to 10.2 Gb in the mountain yellow-legged frog, *Rana*

11 *muscosa* (GenBank accession GCA_029206835.1). Here we report a 6.8 Gb assembly of the

12 dyeing poison frog, *D. tinctorius*, with a contig N50 of 32.5 Mb and 50% of the assembly being

13 represented by only 56 contigs (fig. 1). The lack of fully assembled chromosomes in our

14 assembly limits our ability to test for chromosomal evolution in *D. tinctorius* compared to other

15 species; however, we found only a single contig that mapped to multiple contigs of other

16 available Hyloidea genomes (fig. 2). Our analysis of synteny between these genomes suggests

17 that the evolution of intrachromosomal rearrangements occurs much more rapidly than the

18 evolution of chromosome structure, despite two-fold differences in genome size. Recent work

19 comparing 8 distantly related anurans supports the idea that the evolution of genome size

20 greatly outpaces chromosomal changes (Bredeson et al. 2024).

21   We used multiple annotation approaches to show that over three quarters of the *D.*

22 *tinctorius* genome consists of repetitive elements. These repetitive elements are more abundant

23 in introns than exons, contributing to the evolution of gene sizes. Below, we discuss how our

24 results advance the understanding of the 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 (Schrader and Schmitz 2018; Ding et al. 2016). 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 (Nowoshilow et al. 2018; Sun et a. 2015). By contrast, in the relatively small amphibian genome of the Western clawed frog (*Xenopus 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, remain to be tested. Our analyses of LTR abundances, size, and insertion times corroborate past findings (Nowoshilow et al. 2018, Sun et al. 2015) and indicated that LTRs have played a significant role in the genome size evolution of *D. tinctorius* (fig. 4).

16

1   In addition to amphibians as a whole, transposable elements have likely played an

2   important role in the evolution of genome size among poison frogs (family Dendrobatidae). For

3   example, analyses of a fragmented assembly of *O. pumilio* estimated that this genome consists

4   of over 4.5 Gb of repetitive sequence (Rogers et al. 2018). Consistent with our findings in *D.*

5   *tinctorius*, Rogers et al. (2018) reported *Ty3*/*DIRS1* LTRs (referred to as "Gypsy-like") and *Tc1*

6   retroelements as among the most abundant (1.0 Gb and 250 Mb of sequence, respectively).

7   Rogers et al. (2018) also found that *Ty3* and *Tc1* elements are actively expressed in the oocytes

8   of *O. pumilio*. Taken with the evidence of recent insertion events we estimate here (fig. 4), it is

9   likely that many elements are actively evolving and contributing to ongoing genome size

10   evolution in Dendrobatid frogs. Research into the mechanisms governing TE insertion and

11   regulation in anurans will be important to understand the mechanisms underlying genome size

12   evolution in this group.

13   SINEs are a highly abundant class of transposable element in mammals and many other

14   vertebrates, although they are almost entirely absent in amphibian genomes (Chalopin et al.

15   2015; Zuo et al. 2023), including *D. tinctorius*. This difference illustrates how the abundance of

16   different transposable elements varies greatly across the tree of life. Interestingly, the majority of

17   transposable elements we annotated in the *D. tinctorius* genomes could not be classified as

18   known elements using our approach (39.15%). This is likely because reference libraries of

19   described transposable elements used in annotation pipelines lack amphibian-specific

20   elements, and this pattern of abundant "unclassified" elements is common in studies of

21   amphibian genomes (Sotero-Caio et al. 2017; Zuo et al. 2023). Future work that leverages ever-

22   increasing amphibian genomic resources to describe the diversity and structure of the

23   amphibian 'dark matter' elements would open doors for comparative analyses of transposable

24   element evolution across taxa and greatly increase our understanding of the evolutionary history

25   of parasitic DNA across the tree of life.

17

1	An ultimate consequence of TE evolution is their effect on genome size. For example,

2	genome size increases with the abundance of repetitive elements in anurans (Zuo et al. 2023).

3	Our results show that effects on genome size are not restricted to intergenic regions, with TEs

4	being found within both exons and introns (tables 1 and 2). Analyses of the Mexican axolotl,

5	*Ambystoma mexicanum*, have shown a similar pattern, with introns being ten times larger in this

6	species of salamander compared to other vertebrates (Smith et al. 2009). We also found that

7	TEs affect the size of gene regions within the *D. tinctorius* genome (fig. 3). These findings are

8	contrary to a lack of relationship between gene size and genome size reported in a recent

9	comparison of 14 anuran genomes (Zuo et al. 2023). A possible explanation for this discrepancy

10	is that the largest poison frog genome, in Zuo et al. (2023)'s analysis was found to contain

11	significant contamination and has since been retracted (Stuckert et al. 2021, retracted).

12	While it is now clear that TEs comprise a large proportion of many amphibian genomes,

13	and they are not restricted to intergenic regions (Sotero-Caio et al. 2017), there is still a

14	knowledge gap in our understanding of how TEs influence the evolution and adaptability in this

15	group. The fact that many repetitive elements are within or overlap gene features (introns and

16	exons; tables 1 and 2) suggests that their evolution may have important phenotypic

17	consequences. For example, studies in other non-amphibian species have shown that TEs can

18	have diverse effects on gene expression (Lanciano and Cristofari 2020; Rech et al. 2022);

19	however, more work is needed to understand the phenotypic consequences of TEs across

20	ecologically and behaviourally diverse amphibian species. TEs have also been shown to

21	contribute to the evolution of colour phenotypes in a wide range of animals (Galbraith and

22	Hayward 2023), yet links between TEs and colouration in amphibians are largely lacking

23	(Galbraith and Hayward 2023). A possible mechanism linking aspects of genome size

24	evolution—such as the accumulation of TEs in introns—and phenotypic evolution, is the effect

25	that intron size can have on gene expression (Castillo-Davies et al. 2002; Taft et al. 2007).

1 Future work that leverages an increasing number of amphibian genomes with transcriptomic

2 analysis—for example analysing expression of different isoforms across species and

3 developmental stages—could provide a fruitful avenue towards addressing the 'consequences

4 of genome-size evolution' knowledge gap.

5 ***Multilocus genotyping using MobiSeq***

6 We used our genome assembly to show that leveraging repetitive elements to design

7 and anchor primers may be a useful way to generate multilocus genotypic data at the population

8 scale. These types of datasets would open exciting possibilities for understanding the adaptive

9 processes and evolutionary dynamics of amphibian populations. The low amount of DNA that is

10 needed for sequencing over larger parts of the genome makes MobiSeq a powerful tool for non-

11 invasive sampling of target species. The target primers can be developed from the reference

12 genome (if available), or from closely related species, as we have shown in our study. However,

13 we recommend using a reference genome whenever possible to design species-specific

14 primers and to map reads and genotype SNPs.

15 Our test of the MobiSeq approach showed that primers can differ considerably in their

16 amplification and sequencing success. The element LINE109 could have been not abundant

17 enough throughout the genome to give enough sequences: the LINE109 element was found

18 109 times in the *D. tinctorius* assembly, while the unknown transposable element TE644 was

19 found 644 times. Therefore, we suggest that future marker development should design primers

20 for markers which have at least 400 to 600 copies in the genome.

21 When genotyping and calling SNPs, de novo assembly with the program stacks gives

22 better results than mapping to the *O. pumilio* genome where the primers were developed from

23 (the *D. tinctorius* assembly was not available when we started the MobiSeq test). Although this

24 is a closely related species, high variability in TE composition could contribute to poorer

19

1 mapping and genotyping when relying on the *O. pumilio* genome. The de novo approach also

2 gave better results than the mapping to the new *D. tinctorius* reference genome. A limitation of

3 the stacks method we used for genotyping is that this approach resulted in high genotype

4 missingness. Additionally, the stacks program was developed for restriction site digested (RAD)

5 approaches; therefore, we highly recommend using the original ANGSD pipeline, provided by

6 the team that developed MobiSeq when genotyping MobiSeq data (Rey-Iglesias et al. 2019).

7 This method gave the best results considering genotyping and further analysis with COLONY.

8 Specifically, filtering SNPs to those called within a minimum of 70 (out of 92) individuals gave

9 the most reliable relationship estimates in terms of identifying 5/5 of the duplicates in our

10 dataset, as well as giving realistic population-size estimates, given our knowledge of the sample

11 population and the biology of *D. tinctorius*.

12 MobiSeq was developed using mammalian genomes, which have lower levels of

13 transposable elements, and focused on SINEs and LINEs, which are less common in

14 amphibians (Sotero-Caio et al. 2017). Therefore, there are challenges in applying this method to

15 amphibians with large genomes and high transposable element content, such as primer

16 selection and mapping to an existing reference genome.

17 **Conclusions**

18 Transposable elements are a major component of amphibian genomes and they play an

19 important role in genome size evolution. By generating and analysing a long-read assembly of

20 the poison frog, *Dendrobates tinctorius*, we have shown that TE evolution impacts genome size,

21 not only through their insertion in intergenic regions, but also within exons and introns. This

22 "ecology" of TEs provides a possible mechanism that links genome size evolution to phenotypic

23 evolution. We also provide an example use of MobiSeq to generate low-input cost-effective

24 population genetic data. This method could be used to study the evolutionary dynamics of

25 amphibian populations, alongside aiding their conservation. Overall, our study adds to the

1   growing body of knowledge on the evolution of amphibian genomes. We hope the data and

2   analyses we report here will be a valuable resource for future studies of amphibian genetics,

3   evolution, behaviour, and conservation.

4

5   **Materials and Methods**

6   *Sample collection, DNA extraction, and sequencing for whole genome assembly*

7   We generated a reference genome for *Dendrobates tinctorius* from a single male of the

8   "azureus" morph / population. This individual was captive bred by hobbyists in the United

9   Kingdom, kept under licence of the Home Office at Bangor University, and sacrificed by

10   overdose of Tricaine Methanesulfonate followed by immediate pithing and decapitation. DNA

11   was extracted in four parallel extractions performed at the NERC Environmental Omics Facility

12   (NEOF) at the University of Sheffield using the Macherey-Nagel NucleoBond High Molecular

13   Weight DNA kit (see SI for details). DNA was then cleaned and sheared to an average size of

14   19 kb before generating four HiFi libraries using SMRTbell template express kit 2.0 (PacBio).

15   Final libraries were size-selected in the size range of 7-50kb and sequenced across 18 SMRT

16   cells (see SI for details).

17

18   *Genome assembly, quality control, and synteny*

19   We assembled HiFi reads greater than 10kb in length using HiFiASM (Cheng et al.

20   2021, 2022; v0.16.1-r375) run with a bloom filter of 39 bits (option -f f39) and "aggressive"

21   purging of haplotigs (option -l 2). To identify and correct assembly errors, we ran our primary

22   assembly through the Inspector pipeline using inspector.py and inspector-correct.py scripts,

23   respectively (Chen et al. 2021). Inspector also provides an estimate of assembly accuracy,

24   reported as a Quality Value (QV = $-10\log_{10}$(base-level errors / total assembly length)). After

25   error-correction we identified and removed bacterial or viral contaminants using Kraken2 (Wood

26   et al. 2019; v2.1.2) run using the standard Refseq database containing archaea, bacteria, viral,

1 plasmid, human, and UniVec_Core indexes (downloaded 27-5-2021). We estimated assembly

2 completeness using BUSCO (Simão et al. 2015; v5.2.2) run in genome mode using the

3 tetrapoda_odb10 dataset, which consists of 5,310 single-copy orthologs derived from 38

4 genomes (created 2021-02-19). All other BUSCO options were left as default. We also

5 compared assembly statistics of the *D. tinctorius* genome to the only other Dendrobatid genome

6 currently available on NCBI (*Oophaga pumilio*) and four chromosome-scale assemblies of other

7 Anurans (*Bufo bufo* [NCBI RefSeq ID: GCF_905171765.1], *Rana temporaria* [RefSeq ID:

8 GCF_905171775.1], *Engystomops pustulosus* [GenBank ID: GCA_019512145.1], and *Xenopus*

9 *tropicalis* [RefSeq ID: GCF_000004195.4]).

10

11 *Synteny between D. tinctorius and other Hyloidea genomes*

12   We compared the *D. tinctorius* assembly to chromosome-scale assemblies of *Bufo*

13 *gargarizans* (Asiatic toad; Bufonidae; GenBank accession: [GCA_014858855.1]),

14 *Eleutherodactylus coqui* (Common coquí; Eleutherodactylidae; [GCA_019857665.1]), and

15 *Engystomops pustulosus* (Túngara frog; Leptodactylidae; [GCA_019512145.1]) using orthology-

16 guided synteny map construction in GENESPACE (Lovell et al. 2022; v1.3.1) and BUSCO

17 (Simão et al. 2015; v5.2.2) annotations of the 5,310 gene tetrapoda_odb10 dataset (created

18 2021-02-19) as input. BUSCO annotations for *Xenopus tropicalis* [GCA_000004195.4] were

19 used as outgroup sequences in the initial GENESPACE analysis. Because we were interested

20 in identifying large regions of synteny, we constrained our analysis to scaffolds that have been

21 assigned to chromosomes in the chromosome-scale assemblies, and contigs at least 20 Mb

22 long in the *D. tinctorius* assembly. This approach resulted in the number of annotated single-

23 copy BUSCOs per genome being: 4,482, 3,579, 3,337, 3,440, and 4,939 for *B. gargarizans*, *En.*

24 *pustulosus*, *El. coqui*, *D. tinctorius*, and *X. tropicalis*, respectively. We present summaries of

25 synteny between *D. tinctorius* and the three chromosome-scale assemblies as the proportion of

26 the *D. tinctorius* assembly contained within a synteny block and the average size of synteny

22

1   blocks between genomes. Syntenic alignments were visualised using GENESPACE's

2   plot_riparian function (Lovell et al. 2022; v1.3.1).

3   We also tested whether the size of annotated BUSCO genes was positively correlated

4   with genome size using a Spearman's rank correlation test. We included the five genomes listed

5   above and the chromosome-scale assembly of *Rana temporaria* [GCA_905171775.1]. We

6   focused this analysis on BUSCO annotations rather than *de novo* annotations generated for

7   each genome because we annotated BUSCO genes using the same pipeline in all genomes,

8   and BUSCO genes are chosen based on orthology across different vertebrate genomes,

9   thereby reducing the likelihood that comparisons between paralogous genes affected our

10   results. We note that restricting this analysis to conserved BUSCO genes could result in biases

11   in gene size and evolutionary rate compared to all genes present in these genomes. To test

12   whether the size of BUSCO gene regions correlated with genome size we carried out

13   Spearman's rank correlations on the raw data as well as on phylogenetically independent

14   contrasts (PICs) of both genome and BUSCO gene region sizes. PICs were calculated using

15   the *pic()* function from the *ape* package in R (Paradis and Schliep 2019) and a pruned version of

16   the phylogeny provided by Hime et al. (2021). Finally, we compared the *B. gargarizans*

17   assembly—as it was the closest in size to that of *D. tinctorius*—to *D. tinctorius* using a non-

18   ortholog based nucleotide alignment with minimap2 (Li 2018, 2021; v2.24-r1122; options:-x

19   asm20 -l10G -B3 -O4,24 -N10).

20

21   *Genome annotation and transposable element diversity*

22   We annotated repetitive elements in our *D. tinctorius* assembly by first using

23   RepeatModeler (Flynn et al. 2020; v2.0.3) with rmblast (v2.11.0+) to identify and generate a

24   species-specific repeat library, followed by RepeatMasker (Smith et al. 2021; v 4.1.2-p1) run

25   using the custom repeat library generated by RepeatModeler. RepeatModeler was run with

26   default settings and RepeatMasker was run with -xsmall -a -gff options. Rather than conducting

23

LTR structural analysis with RepeatMasker (-LTRStruct option), we independently annotated Long Terminal Repeat (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 (Katoh 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 2-parameter model (Kimura 1980; option: model = "K80"). To convert divergence estimates to time, we assumed a substitution rate of 2.5×10^{-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 1000bp. 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*

24

1 genome using bedtools intersect (Quinlan and Hall 2010) with options -wa for the full

2 comparison or -f 0.75 -r when testing for 75% reciprocal overlap between annotated gene

3 features and repetitive elements.

4

5 *Tadpole sampling and DNA extraction for relatedness analysis*

6 The aim of this part of the study was to test the MobiSeq protocol as a method for

7 generating a reduced representation library in a non-model organism, the dyeing poison frog (*D.*

8 *tinctorius*). We wanted to use single nucleotide polymorphism (SNP) data to resolve the

9 relatedness between tadpoles in different, small, confined environments such as phytotelmata

10 (small bodies of water in tree holes or palm bracts, for natural history see Rojas and Pašukonis

11 2019).

12 We sampled 87 tadpoles of *D. tinctorius* from 17 phytotelmata in the Nouragues Nature

13 Reserve, French Guiana in 2020, by clipping a small part of the tail tip. Tissue samples were

14 stored in 70% ethanol (EtOH) at -20°C until further processing. To increase reliability and to test

15 the appropriateness of the relatedness analysis, we included 5 samples as duplicates, giving a

16 total sample size of 92. A modified salting-out method was used to extract DNA from small parts

17 of the tail clips (supplementary methods, section 1, Supplementary Material online). The

18 extracted DNA was eluted in 100µl TE buffer, DNA concentration measured with a NanoDrop™

19 spectrophotometer (Thermo Fisher) and stored at -20°C until further processing.

20 *MobiSeq primer design*

21 We designed specific primers for repeated elements as MobiSeq uses transposable

22 elements (TEs) in the DNA for target enrichment PCRs. No published reference genome for *D.*

23 *tinctorius* was available when we started to work on the MobiSeq approach. To find highly

24 repetitive TEs, we used the genome of a closely related species, the strawberry poison frog

1 (*Oophaga pumilio*), which was published in 2018 (Rogers et al.; NCBI GenBank:

2 [GCA_009801035.1]).

3   We used the free software RepeatMasker (Smith et al. 2021; v 4.1.2) to mask repeated

4 elements and extracted those regions with Samtools (Danecek et al. 2021) to get the 22 last bp

5 of transposable elements as a reversed complement and the number of occurrences. The list

6 was cross checked and annotated with existing libraries of transposable element families

7 provided by Dfam (Storer et al. 2021; v3.2). Based on this list we choose elements with an

8 average abundance of 400-600 times in the genome, as elements with higher abundances

9 might be clustered and not evenly spread over the genome. LINES and SINEs were generally

10 less abundant so that we included only one primer for a lower abundance LINE. We initially

11 chose six possible reverse transposable element primers and tested amplification and

12 multiplexing with *D. tinctorius* DNA. We decided on the following two primers after running a

13 target enrichment PCR, based on amplification success and minimisation of primer dimers

14 (AdapterTEsequence):

15 D_tinct_Line_109

16 (5'GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTATGTTACTATGTTACTATGT'3)

17 D_tinct_TE_644

18 (5'GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTACTTTTGGCCACCACTGTA'3).

19

20   In the first step of the MobiSeq protocol, the *D. tinctorius* DNA was digested using

21 fragmentase. To ensure an even digestion across samples, the DNA was diluted to 10ng/µl. We

22 tested the correct incubation time for the fragmentase enzyme beforehand and decided on 20

23 min at 37°C for each sample based on gel images (supplementary methods, section 4.1,

24 Supplementary Material online). After fragmentase treatment, a Sera-Mag Speed beads clean-

26

1　up was used to remove the enzyme mix and buffers (supplementary methods, section 4.2,

2　Supplementary Material online). Clean-up was followed by end-repair of DNA fragments to

3　create blunt ended fragments using T4 polymerase. After incubation, samples were cooled

4　down to 10°C and immediately used for the next step (supplementary methods, section 4.3,

5　Supplementary Material online). For adapter ligation, double stranded modified P5 adapters

6　(Meyer and Kirchner 2010) were added to the end-repaired DNA fragments via a T4 DNA

7　ligase. The modification of the adapter results in single-stranded adapters, which allow the use

8　of universal adapter primers in the next target enrichment PCR step. The adapter ligation

9　master mix (4µl per sample) is added to the end-repaired samples. The samples were cooled

10　down to 10°C after the incubation (supplementary methods, section 4.4, Supplementary

11　Material online). The resulting product was cleaned again, using the same Sera-Mag Speed

12　beads as after DNA fragmentation (supplementary methods, section 4.2, Supplementary

13　Material online) and eluted in 25µl AE buffer and stored at 4°C. We used the primers TE644 and

14　Line109 as described above to enrich the fragments containing transposable elements in a

15　multiplex approach (supplementary methods, section 4.5, Supplementary Material online). The

16　remaining 15µl of PCR were cleaned using the Sera-Mag Speed beads and eluted in 20µl AE

17　buffer. In the final step of the sequencing library preparation, a second PCR using Illumina

18　indexed forward and reverse primers was used followed by agarose gel electrophoresis

19　(supplementary methods, section 4.6, Supplementary Material online). The uniquely tagged

20　samples were mixed, based on the brightness of the smear on the gel. The whole pool of

21　samples was loaded on another gel to cut out fragments of the lengths between 200 and 500bp.

22　The innuPREP DOUBLEpure kit (Analytik Jena) was used to extract the DNA from the gel

23　fragment, following the kit's protocol. The final library pool was sent to Novogene UK for

24　sequencing in 150bp paired end mode on a Novaseq instrument (Illumina).

25

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 150bp. 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 reliant markers are on the availability of genomic resources of closely related taxa. As such, we took three approaches to calling SNPs in our dataset: (1) de novo assembly of MobiSeq fragments, (2) mapping reads to the *O. pumilio* reference, and (3) 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 50bp and 100bp, 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 100bp). The 50bp approach therefore retains more but shorter sequences than the 100bp 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

28

1 sequences aligned with the mem function. The output was saved as a vcf file, with the following

2 parameters: populations: -p 1 -r 0 --write-random-snp --max-obs-het 0.5 --ordered-export --vcf.

3   The soft filtering of SNP data was done in R (R Core Team 2022; v4.2.2) with

4 Bioconducter (Morgan 2022; v3.16; BiocManager 1.30.19) and the filter_rad function from the

5 radiator package (Gosselin 2020; v1.2.8). To implement the radiator package from github we

6 used the package devtools (Wickham et al. 2022). All thresholds that we used for filtering can

7 be found in supplementary methods (section 5, Supplementary Material online). We expected a

8 high degree of missing data due to the Mobiseq approach, therefore the filter for maximum

9 missingness was quite high (0.9).

10   The strong filtering of SNP data was done with vcftools (Danecek et al. 2011; v0.1.16)

11 using the following parameters: --max-missing 0.5 --mac 2 --minDP 3.

12   Additionally, we used the ANGSD pipeline from Rey-Iglesia et al. 2019

13 (https://github.com/shyamsg/MobiSeq/blob/master/code/pipeline.sh), accounting for the fact that

14 stacks was developed for restriction enzyme based RAD sequencing and might bias our output.

15 We used the BWA-mem2 indexed reference genome of *D. tinctorius* to map our sequences,

16 bedtools (Quinlan and Hall 2010; v2.26.0) to merge reads and kept only sites that are present in

17 90% of the cases. We called variants using ANGSD (Korneliussen et al. 2014; v0.940) with

18 minimum quality of 30, min mapping quality of 30, filter for SNPs with a p-value of 1e-6, major

19 and minor allele were inferred directly from likelihoods, minor allele frequency was estimated

20 (fixed major unknown minor), genotypes and SNPs called. SNPs needed to be present in either

21 a minimum of 80 or 70 individuals. We randomly called one SNP per contig to avoid linkage and

22 created 5 vcf files for the 80 and 70 individuals dataset respectively.

23

24

29

1   *Relatedness analysis*

2   The relatedness analysis was conducted with COLONY (Jones and Wang 2010; v2.0).

3   We used the filtered SNPs of the TE644 primer only, derived from our 4 approaches (de novo

4   stacks, *O. pumilio* reference genome or our new *D. tinctorius* genome) and two methods of

5   calling SNPs (stacks vs. ANGSD). The write_colony function in the radiator package was used

6   to write a colony input function. The input parameters were as follows; the mating system of

7   males and females was set to polygamous, no inbreeding and no update of allele frequencies,

8   the length of the run was set to 2 with full likelihood analysis and high precision. We ran an

9   analysis with different random seeds to increase reliability of the clustering (1234 and 1789).

10

11 **Acknowledgements**

18

19 **Data availability**

20 The *D. tinctorius* genome assembly and BRAKER2 annotations are available on the Open

21 Science   Framework   project   page:

22 https://osf.io/tg24j/?view_only=13be51ec97184d888a12d4ba26dfc90f and the genome is being

23 processed to be made available on NCBI: BioProject: PRJNA977854; BioSample:

1 SAMN38117364. Illumina reads generated from MobiSeq libraries are publicly available on

2 NCBI: BioProject: PRJNA1035148.

3

4 **Authors' contributions**

5 AAC, SS and BR conceived and designed the project. RM, CAF and AAC collected samples.

6 CD, FH, LSK, MH extracted DNA and prepared the libraries. SS, LAO, AF, AAC provided

7 resources (data, computational). DP, FH, CD, AAC performed bioinformatic analysis of the data.

8 CD, AAC, FH, BR and SS discussed and interpreted the results, AAC, CD, FH and LAO wrote

9 the first draft. AAC and BR provided funding. All authors read, reviewed and approved the final

10 manuscript.

11

12 **References**

13 Almojil D et al. 2021. The structural, functional and evolutionary impact of transposable

14 elements in eukaryotes. Genes. 12(6):918.

15 Amézquita A, Flechas SV, Lima AP, Gasser H, Hödl W. 2011. Acoustic interference and

16 recognition space within a complex assemblage of dendrobatid frogs. Proc Natl Acad Sci USA.

17 108:17058–17063.

18 Bourque G et al. 2018. Ten things you should know about transposable elements. Genome Biol.

19 19:1–2.

20 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence

21 data. Bioinformatics. 30:2114–2120.

22 Bolyen E et al. 2019. Reproducible, interactive, scalable and extensible microbiome data

23 science using QIIME 2. Nature Biotechnol. 37: 852–857.

1   Brandies P et al. 2019. The value of reference genomes in the conservation of threatened

2   species. Genes. 10(11): 846.

3   Bredeson JV et al. 2024. Conserved chromatin and repetitive patterns reveal slow genome

4   evolution in frogs. Nat Commun. 15:579.

5   Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic

6   eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein

7   database. NAR Genom Bioinform. 3: lqaa108.

8   Bushnell B. 2014. BBMap short read aligner, and other bioinformatic tools.

9   https://sourceforge.net/projects/bbmap/.

10   Carvajal-Castro JD et al. 2021. Aposematism facilitates the diversification of parental care

11   strategies in poison frogs. Sci Rep. 11(1): 19047.

12   Casacuberta E, González J. 2013. The impact of transposable elements in environmental

13   adaptation. Mol Ecol. 22(6):1503–17.

14   Castillo-Davis CI, Mekhedov SL, Hartl DL, Koonin EV, Kondrashov FA. 2002. Selection for short

15   introns in highly expressed genes. Nat Genet. 31:415–418.

16   Catchen J, Amores A, Hohenlohe P, Cresko W, Postlethwait J. 2011. Stacks: building and

17   genotyping loci de novo from short-read sequences. G3: Genes, Genomes, Genetics. 1:171–

18   182.

19   Catchen J, Hohenlohe P, Bassham S, Amores A, Cresko W. 2013. Stacks: an analysis tool set

20   for population genomics. Mol Ecol. 22:3124–3140.

1 Chalopin D, Naville M, Plard F, Galiana D, Volff JN. 2015. Comparative analysis of transposable

2 elements highlights mobilome diversity and evolution in vertebrates. Genome Biol Evol.

3 7(2):567–80.

4 Chen Y, Zhang Y, Wang AY, Gao M, Chong Z. 2021. Accurate long-read de novo assembly

5 evaluation with Inspector. Genome Biol. 22:312.

6 Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly

7 using phased assembly graphs with hifiasm. Nature Methods. 18(2):170–5.

8 Cheng H et al. 2022. Haplotype-resolved assembly of diploid genomes without parental data.

9 Nature Biotechnol. 40(9):1332–1335.

10 Chouteau M, Summers K, Morales V, Angers B. 2011. Advergence in Müllerian mimicry: the

11 case of the poison dart frogs of Northern Peru revisited. Biol Lett. 7(5):796–800.

12 Danecek P et al. 2011. The variant call format and VCFtools. Bioinformatics. 27(15):2156–2158.

13 Danecek P et al. 2021. Twelve years of SAMtools and BCFtools. GigaScience. 10(2), giab008

14 [33590861].

15 Ding Y, Berrocal A, Morita T, Longden KD, Stern DL. 2016. Natural courtship song variation

16 caused by an intronic retroelement in an ion channel gene. Nature. 536:329–332.

17 Dobin A et al. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29:15–21.

18 Donnelly MA. 1989. Effects of reproductive resource supplementation on space-use patterns in

19 *Dendrobates pumilio*. Oecologia. 81:212–218.

20 Dufresnes C et al. 2018. Genomic evidence for cryptic speciation in tree frogs from the

21 Apennine Peninsula, with description of *Hyla perrini* sp. nov. Front Ecol Evol. 2018:144.

33

1 Feng YJ et al. 2017. Phylogenomics reveals rapid, simultaneous diversification of three major

2 clades of Gondwanan frogs at the Cretaceous–Paleogene boundary. Proc Natl Acad Sci USA.

3 114(29):E5864–70.

4 Feschotte, C. 2008. Transposable elements and the evolution of regulatory networks. Nat Rev

5 Genet. 9:397–405.

6 Fischer EK et al. 2019. The neural basis of tadpole transport in poison frogs. Proc Soc B.

7 286:20191084.

8 Fischer EK, O'Connell LA. 2020. Hormonal and neural correlates of care in active versus

9 observing poison frog parents. Horm Behav. 120:104696.

10 Fischer EK et al. 2020. Neural correlates of winning and losing fights in poison frog tadpoles.

11 Physiol Behav. 223:112973.

12 Flynn JM et al. (2020). RepeatModeler2 for automated genomic discovery of transposable

13 element families. Proc Natl Acad Sci USA. 117: 9451–9457.

14 Fouilloux C et al. 2021. Pool choice in a vertical landscape: tadpole rearing site flexibility in

15 phytotelm-breeding frogs. Ecol Evol. 11: 9021–9038.

16 Funk WC, Zamudio KR, Crawford AJ. 2018. Advancing understanding of amphibian evolution,

17 ecology, behavior, and conservation with massively parallel sequencing. In: Population

18 Genomics: Wildlife, (eds) Hohenlohe PA, Rajora OP, Springer, Cham, pp. 211-254.

19 Galbraith JD, Hayward A. 2023. The influence of transposable elements on animal colouration.

20 Trends in Genetics. 39(8):624–638.

21 Gozashti L, Feschotte C, Hoekstra HE. 2023. Transposable element interactions shape the

22 ecology of the deer mouse genome. Mol Biol Evol. 40:msad069.

1   Gosselin T. 2023. radiator: RADseq Data Exploration, Manipulation and Visualization using R. R

2   package version 1.2.8 https://thierrygosselin.github.io/radiator/

3   Guillory WX, Muell MR, Summers K, Brown JL. 2019. Phylogenomic reconstruction of the

4   Neotropical poison frogs (Dendrobatidae) and their conservation. Diversity. 2019; 11(8):126.

5   Hawkins JS, Kim H, Nason JD, Wing RA, Wendel JF. 2006. Differential lineage-specific

6   amplification of transposable elements is responsible for genome size variation in Gossypium.

7   Genome Res. 16(10):1252–61.

8   Hellsten U. et al. 2010. The genome of the Western clawed frog *Xenopus tropicalis*. Science.

9   328(5978):633–636.

10   Hime PM et al. 2021. Phylogenomics reveals ancient gene tree discordance in the amphibian

11   tree of life. Syst Biol.70(1):49–66.

12   Homola JJ et al. 2019. Replicated Landscape Genomics Identifies Evidence of Local Adaptation

13   to Urbanization in Wood Frogs. J Heredity. 110:707–719.

14   Hotaling S, Kelley JL, Frandsen PB. 2021. Toward a genome sequence for every animal: where

15   are we now?. Proc Natl Acad Sci USA. 118(52):e2109019118.

16   Jones O, Wang J. 2010. COLONY: a program for parentage and sibship inference from

17   multilocus genotype data. Mol Ecol Res. 10:551–555.

18   Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:

19   improvements in performance and usability. Mol Biol Evol. 30:772–780

20   Kidwell MG, Lisch D. 1997.Transposable elements as sources of variation in animals and

21   plants. Proc Natl Acad Sci USA. 94(15):7704–7711.

1   Kidwell MG. 2002. Transposable elements and the evolution of genome size in eukaryotes.

2   Genetica. 115:49–63.

3   Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions

4   through comparative studies of nucleotide sequences. J Mol Evol, 16:111–120.

5   Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of Next Generation

6   Sequencing Data. BMC Bioinformatics. 15:356.

7   Kosch TA et al. 2022. Genetic approaches for increasing fitness in endangered species. Trends

8   Ecol Evol. 37:332–345.

9   Lamichhaney S et al. 2021. A bird-like genome from a frog: Mechanisms of genome size

10 reduction in the ornate burrowing frog, *Platyplectrum ornatum*. Proc Natl Acad Sci USA.

11 118(11):e2011649118.

12 Lanciano S, Cristofari G. 2020. Measuring and interpreting transposable element expression.

13 Nat Rev Genet. 21:721–736.

14 Lau Q et al. 2020. Heterogeneity of synonymous substitution rates in the Xenopus frog genome.

15 PLoS One. 15(8):e0236515.

16 Lawrence JP, Rojas B et al. 2019. Weak warning signals can persist in the absence of gene

17 flow. Proc Natl Acad Sci USA. 116(38):19037–45.

18 Lee SI, Kim NS. 2014. Transposable elements and genome size variations in plants. Genomics

19 & Informatics. 12(3):87.

20 Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics.

21 34(18):3094–3100.

1 Li H. 2021. New strategies to improve minimap2 alignment accuracy. Bioinformatics. 37:4572–

2 4574.

3 Liedtke HC, Gower DJ, Wilkinson M, Gomez-Mestre I. 2018. Macroevolutionary shift in the size

4 of amphibian genomes and the role of life history and climate. Nature Ecol Evol. 2(11):1792–

5 1799.

6 Lovell JT et al. 2022. GENESPACE tracks regions of interest and gene copy number variation

7 across multiple genomes. eLife. 11:e78526.

8 Maan ME, Cummings ME. 2012. Poison frog colors are honest signals of toxicity, particularly for

9 bird predators. Am Nat. 179(1):E1–4.

10 Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads.

11 EMBnet.journal. 17(1):10–12.

12 McClintock B. 1950. The origin and behavior of mutable loci in maize. Proc Natl Acad Sci USA.

13 *36*(6):344–355.

14 Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target

15 capture and sequencing. Cold Spring Harbor Protocols. 2010(6):pdb-rot5448.

16 Morgan M. 2020. BiocManager: Access the Bioconductor Project Package Repository. R

17 package version 1.30.19, https://CRAN.R-project.org/package=BiocManager

18 Myers CW, Daly JW. 1983. Dart-poison frogs. Sci Am. 248:96–105.

19 Noonan BP, Comeault AA. 2009. The role of predator selection on polymorphic aposematic

20 poison frogs. Biol Lett. 5(1):51–4.

21 Nowoshilow S, et al. 2018. The axolotl genome and the evolution of key tissue formation

22 regulators. Nature. 554:50–55.

1   Nunziata SO, Weisrock DW. 2018. Estimation of contemporary effective population size and

2   population declines using RAD sequence data. Heredity. 120(3):196–207.

3   Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary

4   analyses in R. Bioinformatics. 35:526–528.

5   Pasquesi GIM et al. 2018. Squamate reptiles challenge paradigms of genomic repeat element

6   evolution set by birds and mammals. Nat Commun. 9:2774.

7   Pašukonis A et al. 2022. Contrasting parental roles shape sex differences in poison frog space

8   use but not navigational performance. eLife. 11:e80483.

9   Pimpinelli, S, Piacentini, L. 2020. Environmental change and the evolution of genomes:

10 Transposable elements as translators of phenotypic plasticity into genotypic variability. Funct

11 Ecol. 34: 428–441.

12 Platt RN, Vandewege MW, Ray DA. 2018. Mammalian transposable elements and their impacts

13 on genome evolution. Chromosome Res. 26:25–43.

14 Pröhl H. 2005. Clutch loss affects the operational sex ratio in the strawberry poison frog

15 *Dendrobates pumilio*. Behav Ecol Sociobiol. 58:310–315.

16 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic

17 features. Bioinformatics. 26:841–842.

18 R Core Team. R: A language and environment for statistical computing. R Foundation for

19 Statistical Computing, Vienna, Austria. 2020. URL https://www.R-project.org/. Version 4.2.2

20 Rech GE et al. 2022. Population-scale long-read sequencing uncovers transposable elements

21 associated with gene expression variation and adaptive signatures in Drosophila. Nature

22 Commun. 13(1):1948.

1   Rey O, Danchin E, Mirouze M, Loot C, Blanchet S. 2016. Adaptation to global change: a

2   transposable element–epigenetics perspective. Trends Ecol Evol. 31(7):514–526.

3   Rey-Iglesia A et al. 2019. MobiSeq: De novo SNP discovery in model and non-model species

4   through sequencing the flanking region of transposable elements. Mol Ecol Res. 19(2):512–25.

5   Ringler E, Pašukonis A, Hödl W, Ringler M. 2013. Tadpole transport logistics in a Neotropical

6   poison frog: indications for strategic planning and adaptive plasticity in anuran parental care.

7   Front Zool. 10:67.

8   Rodríguez C et al. 2022. Androgen responsiveness to simulated territorial intrusions in *Allobates*

9   *femoralis* males: evidence supporting the challenge hypothesis in a territorial frog. Gen Comp

10 Endocrinol. 326:114046.

11 Rojas B. 2014. Strange parental decisions: fathers of the dyeing poison frog deposit their

12 tadpoles in pools occupied by large cannibals. Behav Ecol Sociobiol. 68:551–559.

13 Rojas B. 2015. Mind the gap: treefalls as drivers of parental tradeoffs. Ecol Evol. 5:4028–4036.

14 Rojas B, Pašukonis A. 2019. From habitat use to social behavior: natural history of a voiceless

15 poison frog, *Dendrobates tinctorius*. PeerJ. 7:e7648.

16 Rogers RL et al. 2018. Genomic takeover by transposable elements in the strawberry poison

17 frog. Mol Biol Evol. 35:2913–2927.

18 Schloissnig S et al. 2021. The giant axolotl genome uncovers the evolution, scaling, and

19 transcriptional control of complex gene loci. Proc Natl Acad Sci USA. 118(15):e2017176118.

20 Schrader L, Schmitz J. 2019. The impact of transposable elements in adaptive evolution. Mol

21 Ecol. 28(6):1537–49.

1 Schulte LM, Lötters S. 2013. The power of the seasons: rainfall triggers parental care in poison

2 frogs. Evol Ecol. 27:711–723.

3 Serrato-Capuchina A, Matute DR. 2018. The Role of Transposable Elements in Speciation.

4 Genes. 9(5):254.

5 Sexton OJ. 1960. Some aspects of the behavior and of the territory of a dendrobatid frog,

6 *Prostherapis trinitatis*. Ecology. 41:107–115.

7 Silverstone PA. 1973. Observations on the behavior and ecology of a Colombian poison-arrow

8 frog, Kõkoé-pá (*Dendrobates histrionicus* Berthold). Herpetologica. 29:295–301.

9 Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO:

10 assessing genome assembly and annotation completeness with single-copy orthologs.

11 Bioinformatics. 31(19):3210–3212.

12 Smith JJ et al. 2009. Genic regions of a large salamander genome contain long introns and

13 novel genes. BMC Genomics. 10:19.

14 Smith A, Hubley R, Green P. RepeatMasker version 4.1.2-p1. 2021. http://repeatmasker.org

15 Sotero-Caio CG, Platt RN, Suh A, Ray DA. 2017. Evolution and diversity of transposable

16 elements in vertebrate genomes. Genome Biol Evol. 9(1):161–177.

17 Stitzer MC, Anderson SN, Springer NM, Ross-Ibarra J. 2021. The genomic ecosystem of

18 transposable elements in maize. PLoS Genet. 17(10):e1009768.

19 Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. 2021. The Dfam community resource of

20 transposable element families, sequence models, and genome annotations. Mobile DNA. 12:2.

21 Stuckert AMM, Chouteau M, McClure M, LaPolice TM, Linderoth T, Nielsen R, Summers K,

22 MacManes MD. 2021. The genomics of mimicry: Gene expression throughout development

1 provides insights into convergent and divergent phenotypes in a Müllerian mimicry system. Mol

2 Ecol. 30(16):4039-4061. Retraction in: Mol Ecol. 2024. 33(8):e17345.

3 Stynoski JL, Torres-Mendoza Y, Sasa-Marin M, Saporito RA. 2014. Evidence of maternal

4 provisioning of alkaloid-based chemical defenses in the strawberry poison frog *Oophaga*

5 *pumilio*. Ecology. 95(3):587–93.

6 Stynoski JL, Schulte LM, Rojas B. 2015. Poison frogs. Quick Guide. Curr Biol. 25:R1026–

7 R1028.

8 Summers K. 1989. Sexual selection and intra-female competition in the green poison-dart frog,

9 *Dendrobates auratus*. Anim Behav. 37:797–805.

10 Summers K, Sea McKeon C, Heying H. 2006. The evolution of parental care and egg size: a

11 comparative analysis in frogs. Proc R Soc B. 273:687–692.

12 Sun C et al. 2012. LTR retrotransposons contribute to genomic gigantism in plethodontid

13 salamanders. Genome Biol Evol. 4(2):168–183.

14 Sun YB et al. 2015. Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the

15 comparative evolution of tetrapod genomes. Proc Natl Acad Sci USA. 112(11):E1257–1262.

16 Taft RJ, Pheasant M, Mattick JS. 2007. The relationship between non-protein-coding DNA and

17 eukaryotic complexity. Bioessays. 29(3):288–99.

18 Tumulty J, Morales V, Summers K. 2014. The biparental care hypothesis for the evolution of

19 monogamy: experimental evidence in an amphibian. Behav Ecol. 25(2):262–70.

20 Valencia JD, Girgis HZ. 2019. LtrDetector: A tool-suite for detecting long terminal repeat

21 retrotransposons de-novo. BMC Genomics. 20:450.

1 Vasimuddin M, Sanchit M, Heng L, Srinivas A. 2019. Efficient Architecture-Aware Acceleration

2 of BWA-MEM for Multicore Systems. IEEE Parallel and Distributed Processing Symposium

3 (IPDPS). doi:10.1109/IPDPS.2019.00041.

4 Wells KD.1980. Behavioral ecology and social organization of a dendrobatid frog (*Colostethus*

5 *inguinalis*). Behav Ecol Sociobiol. 6:199–209.

6 Wicker T et al. 2007. A unified classification system for eukaryotic transposable elements.

7 Nature Rev Gen. 8(12):973–82.

8 Wickham H, Hester J, Chang W, Bryan J. 2022. devtools: Tools to Make Developing R

9 Packages Easier. R package version 2.4.5, https://CRAN.R-project.org/package=devtools.

10 Womack MC, Metz MJ, Hoke KL. 2019. Larger genomes linked to slower development and loss

11 of late-developing traits. Am Nat. 194(6):854–64.

12 Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. Genome

13 Biol. 20:1–3.

14 Zhang G et al. 2014. Comparative genomics reveals insights into avian genome evolution and

15 adaptation. Science. 346:1311–1320.

16 Zuo B, Nneji LM, Sun YB. 2023. Comparative genomics reveals insights into anuran genome

17 size evolution. BMC Genomics. 24(1):379.

18

1 **Figure and Table Legends**

2 **Figure 1.** *Dendrobates tinctorius* assembly completeness. (**A**) Cumulative assembly size in

3 relation to size-ordered scaffolds for *D. tinctorius* and other published anuran genomes. (**B**)

4 BUSCO "tetrapoda" gene set completeness for the same genomes summarised in (A).

5 **Figure 2. A**) Riparian synteny plots generated using BUSCO gene regions as anchors showing

6 broad patterns of synteny and rearrangements between *D. tinctorius* scaffolds and three publicly

7 available chromosome-scale frog assemblies. **B**) Example *B. gargarizans* chromosome showing

8 both synteny and structural evolution with *D. tinctorius* scaffolds. Inversions in panel B are

9 shown in salmon pink, collinear regions in orange. **C**) Insertion of regions on *D. tinctorius*

10 scaffold 127 (highlighted in red) into chromosomes 1 and 4 of the *E. coqui* assembly. Phylogeny

11 shown in panel A) is pruned from the RAxML "L1-L379.concatenated" tree reported in Hime et

12 al. (2021).

13 **Figure 3.** The relationship between genome size and the average size of annotated BUSCO

14 gene regions in *D. tinctorius* compared with five publicly available chromosome-scale frog

15 assemblies. A) Phylogenetic relationships between the species being compared (pruned

16 RAxML "L1-L379.concatenated" tree from Hime et al. 2021). B) The relationship between

17 genome size (in bp) and the mean size of BUSCO gene regions. C) The relationship between

18 genome size and mean size of BUSCO gene regions, corrected for phylogeny using

19 phylogenetically independent contrasts (PICs) with the phylogeny depicted in A.

20 **Figure 4.** Long terminal repeat (LTR) retroelement lengths **(A)** and insertion times **(B)**. LTRs

21 were annotated with LtrDetector and insertion times were estimated from divergence between

22 the left and right LTR regions of each element independently assuming a substitution rate of 2.5

23 x 10^{-9} substitutions per site per year (Lau et al. 2020). In (A) the total number of elements

24 classified to each type is given above each boxplot.

25

1 **Figure 5.** Summary of **(A)** genotype missingness in %, **(B)** number of SNPs (log scale to

2 account for the large differences) and **(C)** percentage of duplicates found in the dataset using

3 sequences derived from MobiSeq with the TE644 primer. Shown are the two pipeline

4 approaches, stacks on the left (de novo assembly and mapping to reference genomes of

5 *Oophaga pumilio* or *Dendrobates tinctorius*) with no filter SNPs, soft filter (filter_rad) and strong

6 filter (vcftools); and ANGSD on the right (calling of SNPs present in 70 or 80 individuals). For

7 details see Methods.

8

9 **Table 1.** Counts of repeat elements annotated using RepeatModeler + RepeatMasker that

10 overlap with coding DNA sequence (CDS) and introns annotated in the *D. tinctorius* assembly.

Region of overlap	All elements	LTRs	LINEs	*Tc1* transposons	*hAT* 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

11

12

13

14 **Table 2.** Counts of repetitive elements that overlap with CDS and introns (see Table 1) that

15 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

16

17

44
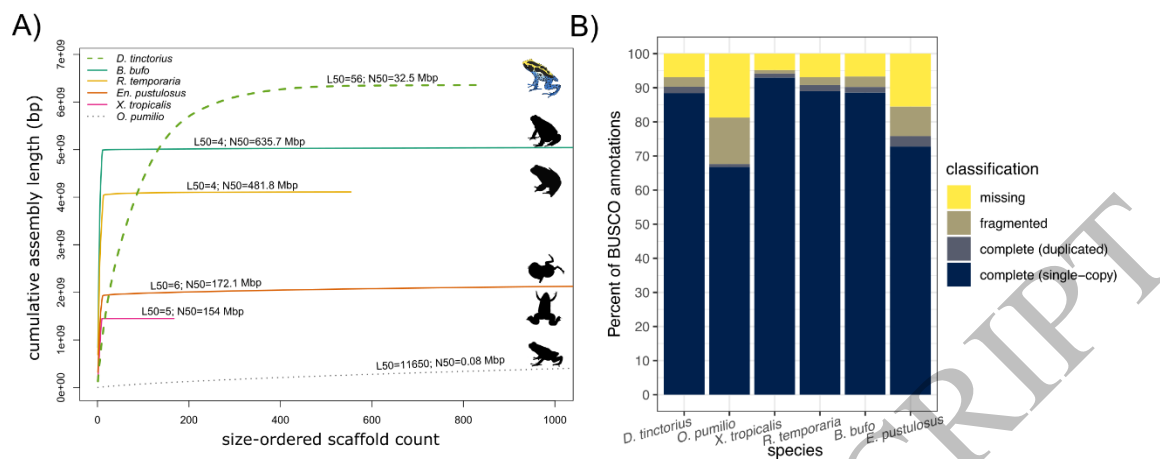
Figure 1

148x59 mm (x DPI)

1
2
3
4

Figure 2
165x134 mm (x DPI)

A)

En. pustulosus
B. gargarizans
El. coqui
D. tinctorius
R. temporaria
X. tropicalis

0.05

B)

mean BUSCO span (bp)

genome size

- B. gargarizans
- En. pustulosus
- El. coqui
- X. tropicalis
- D. tinctorius
- R. temporaria

C)

PICs of mean BUSCO span
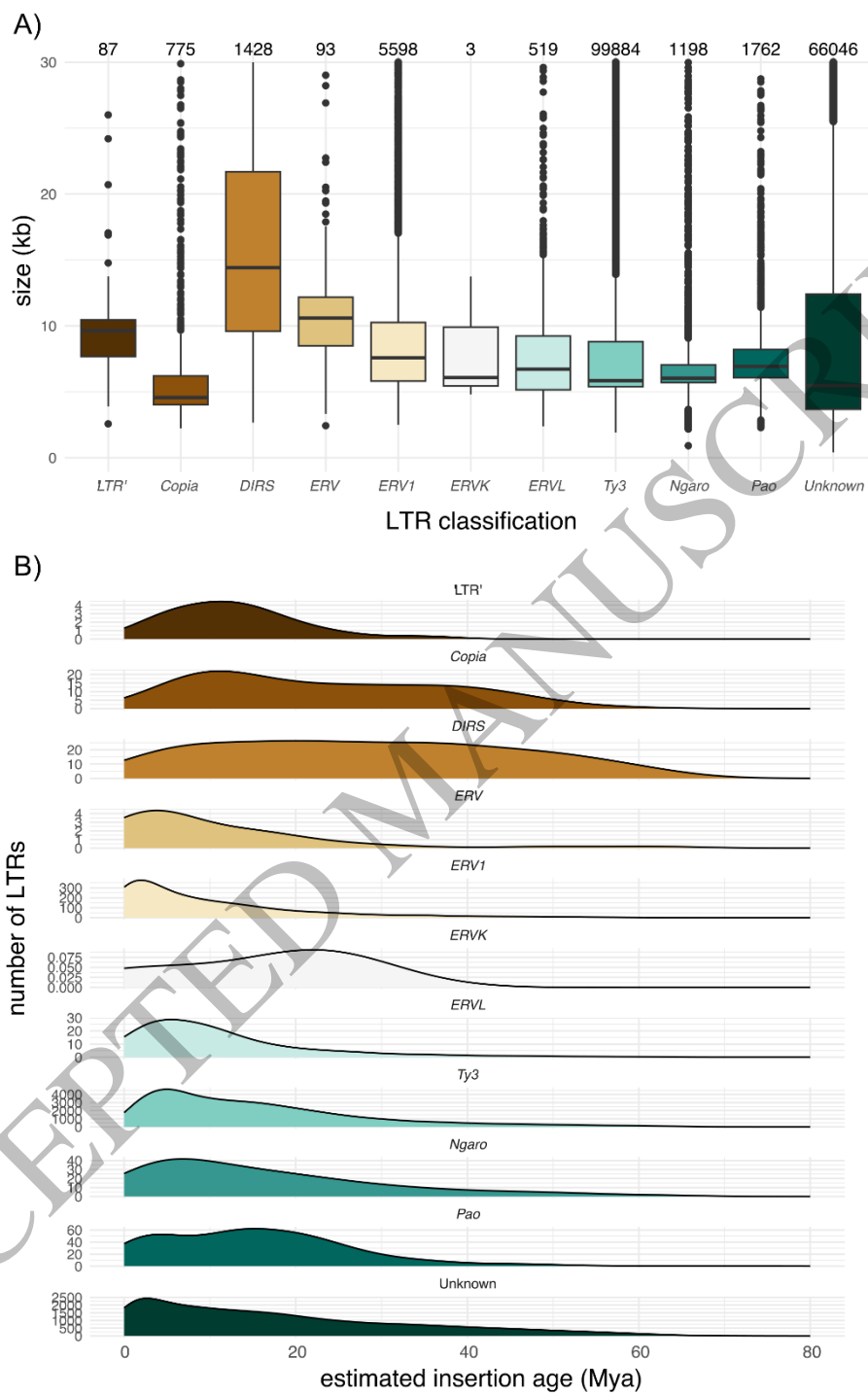
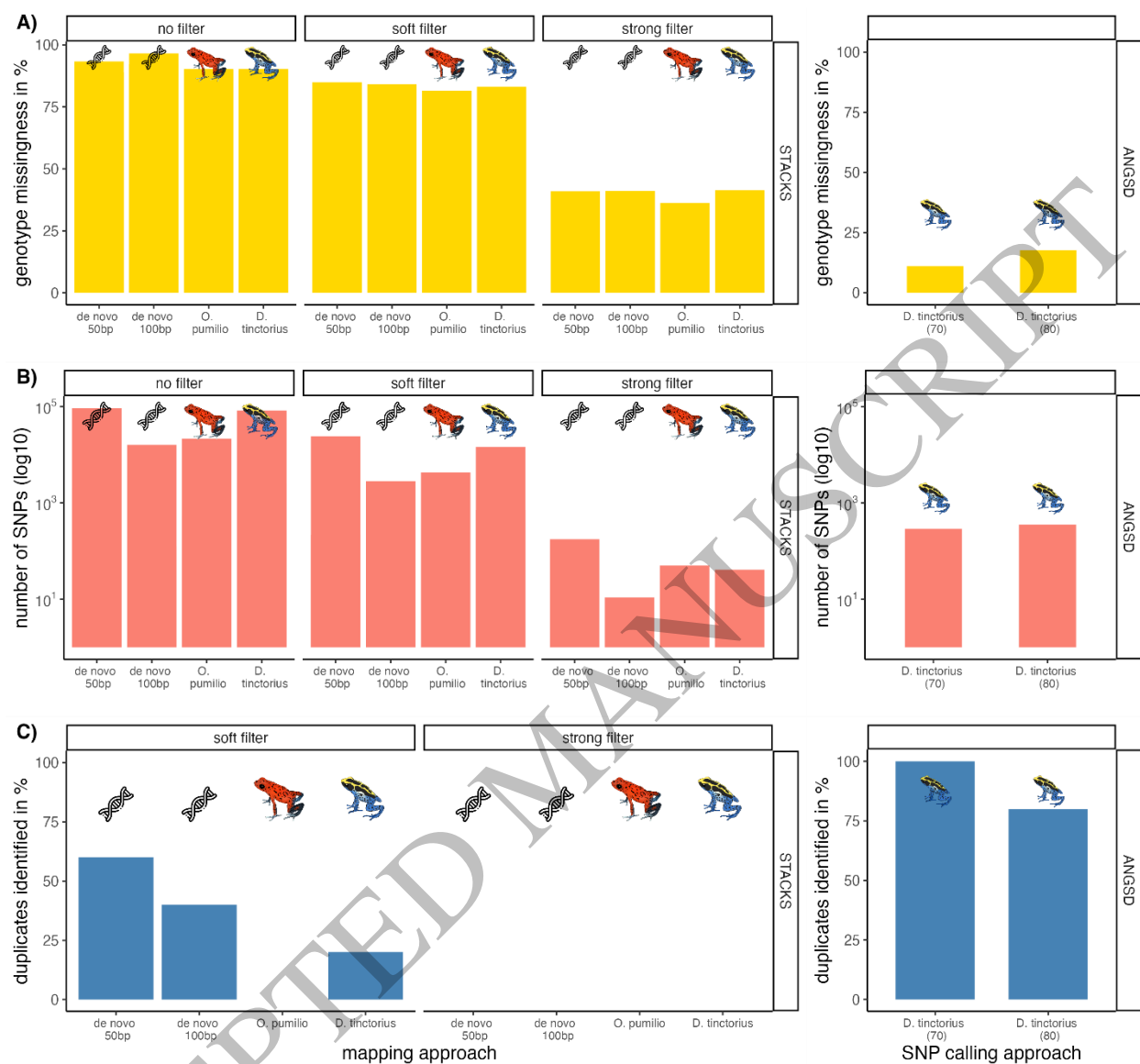PICs of genome size

1

2 *Figure 3*
3 *165x59 mm (x DPI)*

4

1

2

3
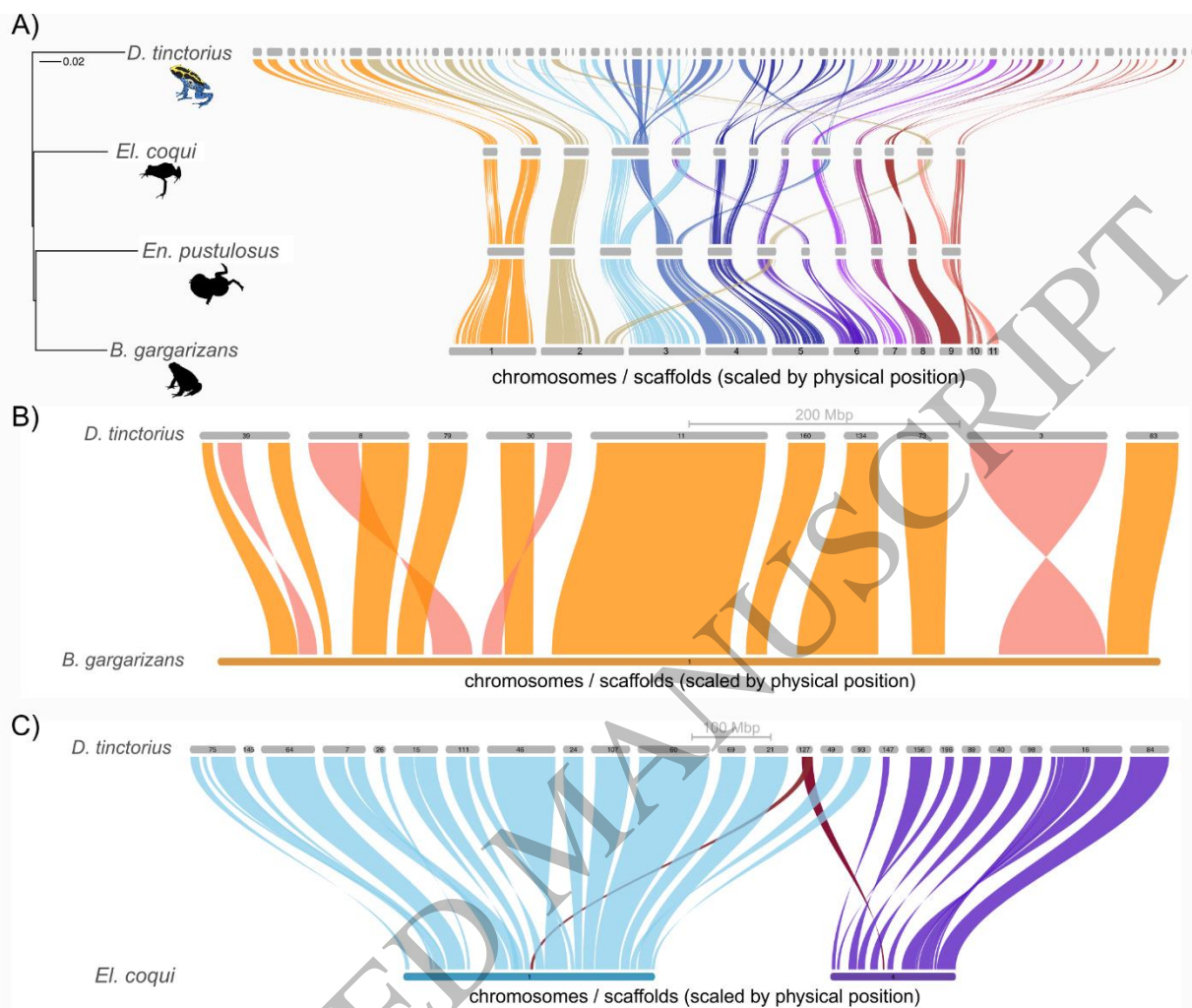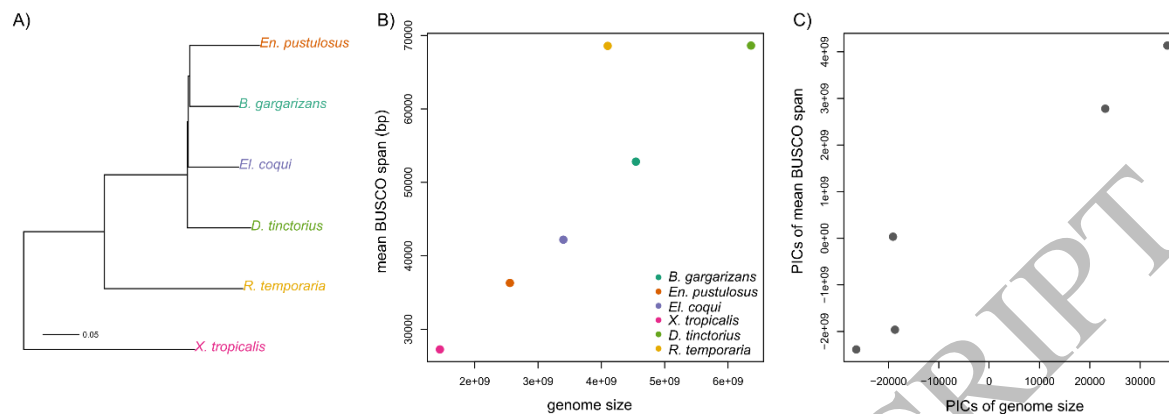
Figure 4
144x229 mm (x DPI)

4

Figure 5
165x153 mm (x DPI)