# A bird-like genome from a frog: Mechanisms of genome size reduction in the ornate burrowing frog, *Platyplectrum ornatum*

Sangeet Lamichhaney[a,b,1], Renee Catullo[c,d], J. Scott Keogh[c], Simon Clulow[e], Scott V. Edwards[a,b,2], and Tariq Ezaz[f,2]

[a]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138; [b]Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138; [c]Division of Ecology and Evolution, Research School of Biology, Australian National University, Acton, ACT, Australia 2601; [d]Australian National Insect Collection and Future Science Platform Environomics, Commonwealth Scientific and Industrial Research Organization, Acton, ACT, Australia 2601; [e]Department of Biological Sciences, Macquarie University, Sydney, NSW, Australia 2109; and [f]Institute for Applied Ecology, Faculty of Science and Technology, University of Canberra, Canberra, ACT, Australia 2617

Contributed by Scott V. Edwards, January 6, 2021 (sent for review June 23, 2020; reviewed by Axel Meyer, Rachel Lockridge Mueller, and David B. Wake)

**The diversity of genome sizes across the tree of life is of key interest in evolutionary biology. Various correlates of variation in genome size, such as accumulation of transposable elements (TEs) or rate of DNA gain and loss, are well known, but the underlying molecular mechanisms driving or constraining genome size are poorly understood. Here, we study one of the smallest genomes among frogs characterized thus far, that of the ornate burrowing frog (*Platyplectrum ornatum*) from Australia, and compare it to other published frog and vertebrate genomes to examine the forces driving reduction in genome size. At ∼1.06 gigabases (Gb), the *P. ornatum* genome is like that of birds, revealing four major mechanisms underlying TE dynamics: reduced abundance of all major classes of TEs; increased net deletion bias in TEs; drastic reduction in intron lengths; and expansion via gene duplication of the repertoire of TE-suppressing Piwi genes, accompanied by increased expression of Piwi-interacting RNA (piRNA)-based TE-silencing pathway genes in germline cells. Transcriptomes from multiple tissues in both sexes corroborate these results and provide insight into sex-differentiation pathways in *Platyplectrum*. Genome skimming of two closely related frog species (*Lechriodus fletcheri* and *Limnodynastes fletcheri*) confirms a reduction in TEs as a major driver of genome reduction in *Platyplectrum* and supports a macroevolutionary scenario of small genome size in frogs driven by convergence in life history, especially rapid tadpole development and tadpole diet. The *P. ornatum* genome offers a model for future comparative studies on mechanisms of genome size reduction in amphibians and vertebrates generally.**

Anura | frog | intron | Piwi genes | transposable elements

Eukaryotic genome sizes are now known to vary by 200,000-fold (1), but the lack of correlation between genome size and organismal complexity (the C-value paradox) has been recognized for more than half a century (2). Larger genomes are associated with larger nucleus or cell sizes in diverse taxa (2–5), which in turn have consequences for organismal traits, such as body and brain size (6). Increases in genome size can be driven by polyploidization events (7), but our knowledge of the causes and consequences of genome size variation in diploids can be improved by investigation of additional species with unusually large or small genome sizes. Lineages characterized by larger genomes, such as flowering plants (8), conifers (9), or salamanders (10), invariably exhibit extensive amplification in transposable elements (TEs); the reverse trend is prevalent among lineages with smaller genomes, such as many fishes (11, 12), some plants (13), and insects (14). Recently, the tight correlation between TE abundance and genome size has come into question, and an alternative "accordion" model, which incorporates covariation between gains in genomic DNA by repeat element expansion and loss through deletion, has been documented in squamates (15), birds, and mammals (16). In vertebrates, additional mechanisms such as intron shortening and deletion biases, have also been linked to reduced genome size (17, 18)

In anurans (frogs) that are diploid, genome size is highly variable, ranging from up to 12 gigabases (Gb) in *Gymnopis* and *Alytes* to ∼1 Gb or less in *Platyplectrum* (19), *Spea* (20), and other genera (21). As in other tetrapods, there is a significant relationship between genome size and cell size in frogs, perhaps due to increased cell cycle time associated with larger genomes (22). In salamanders and other organisms, genome size tends to be smaller in lineages under selection for short development times, enabling faster cell replication (23, 24). In line with this expectation, a recent study identified a direct relationship between aridity and development time, and in turn between development time and genome size in frogs (21), proposing aridity as an indirect ecological driver of reduced genome size via selection for reduced development time in regions with intermittent availability of water. One of the smallest anuran genomes, that of the Mexican spadefoot toad (*Spea multiplicata*), was recently sequenced (20), and its small genome size is consistent with the

## Significance

The mechanisms by which genomes become smaller in lineages with large genomes, such as many frogs, are poorly understood. Here, we present the sequence of a very small genome from a frog—that of the ornate burrowing frog, *Platyplectrum ornatum*, whose genome is smaller than that of many birds, a group well known for small genomes. Our data show that this frog evolved molecular mechanisms to keep in check a major process characterizing large genomes: namely, proliferation of transposable elements. We also present basic data on gene expression and sex determination in this emerging model for genome miniaturization.

Author contributions: S.V.E. and T.E. designed research; S.L., R.C., and T.E. performed research; S.L., R.C., J.S.K., S.C., and T.E. contributed new reagents/analytic tools; S.L., R.C., J.S.K., S.V.E., and T.E. analyzed data; and S.L., R.C., J.S.K., S.C., S.V.E., and T.E. wrote the paper.

Reviewers: A.M., University of Konstanz; R.L.M., Colorado State University; and D.B.W., University of California, Berkeley.

Competing interest statement: J.S.K. and reviewer R.L.M. are coauthors on a 2020 article with multiple additional coauthors.



[1]Present address: Department of Biological Sciences, Kent State University, Kent, OH 44242.

[2]To whom correspondence may be addressed. Email: sedwards@fas.harvard.edu or tariq.ezaz@canberra.edu.au.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2011649118/-/DCSupplemental.

Published March 8, 2021.

trend of reduced development time at the tadpole stage associated with small genome size.

Here we describe the genome of Australia's ornate burrowing frog, *Platyplectrum ornatum* (Limnodynastidae), which has one of the smallest reported genome sizes among amphibians whose genome size has been measured (19, 21), as well as genomes of two related taxa, *Lechriodus* (*Le.*) *fletcheri* and *Limnodynastes* (*Li.*) *fletcheri. P. ornatum* and *Le. fletcheri* are closely related (21) whereas *Li. fletcheri* is more distantly related but a member of the same family, Limnodynastidae. These species are reported to have similarly small (*Le. fletcheri*) and much larger (*Li.*) genomes than *Platyplectrum*. The genus *Platyplectrum* comprises multiple species that are widely distributed across the eastern mesic biome, tropical northern savannahs, and arid deserts of Australia (Fig. 1). Its sister genus, *Lechriodus,* occupies rain forests of New Guinea and eastern Australia, but *Le. fletcheri*, unusually among rain forest species in this region, exclusively utilizes ephemeral pools for breeding (25). All species within the clade comprising *Platyplectrum* and *Lechriodus* share traits evolved convergently with the genus *Spea*, including short and plastic development times, carnivory as tadpoles, and breeding in ephemeral water sources (*SI Appendix*, Table S1) (26, 27). Related to this report of the genome of *Platyplectrum*, we explore diverse aspects of its biology through it. We therefore also report on patterns of sex-biased gene expression and the sex differentiation pathway, which can help inform the fundamental biology of this species and links between ecology and genomics. Additionally, we generated transcriptome data from multiple tissues of adult males and females and investigated expression profiles of genes and transcripts involved in sex and reproduction. Finally, we prepared metaphase chromosomes and karyotypes to assess evidence of heteromorphic sex chromosomes in *P. ornatum*.

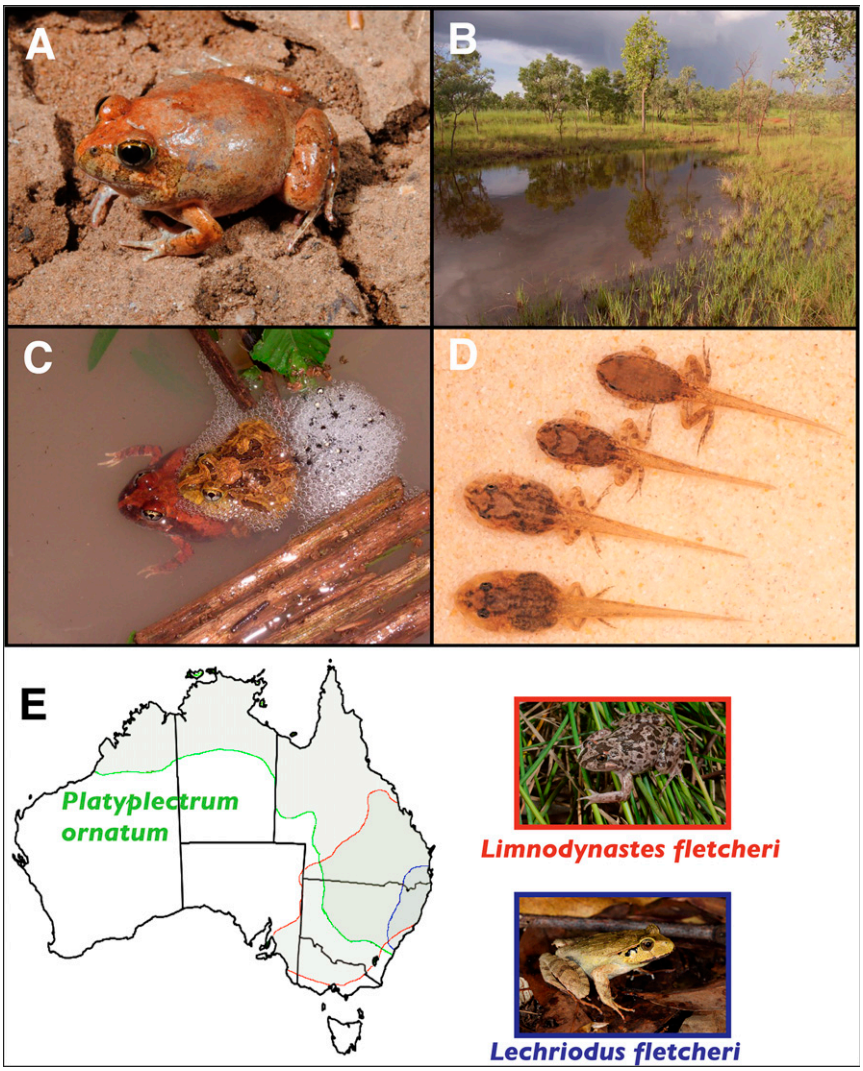


**Fig. 1.** Natural history and distribution of *P. ornatum*. (*A*) Adept at burrowing, these frogs are found widely throughout Australia in many different habitats on sandy soils where they are able to burrow to escape extended dry spells (*B*) When storms occur, they emerge to congregate and breed around ephemeral water bodies that form on the surface in clay pans, ditches, and swamps, such as here in the Kimberley region in the wet–dry tropics of Western Australia. (*C*) After amplexus occurs, the females deposit their eggs in small, floating clumps which they create by beating the jelly of the spawn into a frothy mass using a flange (paddle-like structure) on the second finger, a feature unique to the female for this purpose. (*D*) Upon hatching, the tadpoles progress rapidly through the developmental stages in order to metamorphose before pond desiccation occurs. Metamorphosis has been recorded in as little as 11 d postfertilization. (*E*) Distribution of *P. ornatum* and two closely related species studied here. Images credit: (*A*) Michael G. Swan (photographer), (*B* and *C*) Jennifer Francis (photographer), and (*D*) Marion Anstis (photographer).

## Results

**Genome Sequencing, Assembly, and Annotation.** The sequence of the genome of an ornate burrowing frog (*P. ornatum*) was generated from DNA isolated from the muscle of an adult female. Analysis of DNA quality indicated >85% of DNA fragments were larger than 60 kilobases (kb) (*SI Appendix*, Fig. S1). Based on the expected genome size of ∼0.95 Gb (19), the *P. ornatum* genome was sequenced to ∼140× coverage using multiple insert-size libraries (*SI Appendix*, Table S2) to generate an initial draft genome (*SI Appendix*), which, at a scaffold contiguity ($N_{50}$) of 24 kb, was quite fragmented. We attempted to improve the fragmented initial assembly using a chromatin conformation approach, Hi-C (28), and direct long read sequencing (29), but this effort only marginally improved the genome contiguity (*SI Appendix*, text and Table S3). The genome assembly presented here spanned a total of 1.05 Gb. We assessed genome contiguity and gene-space completeness using benchmarking universal single-copy orthologs (BUSCO) (30) (*SI Appendix*, Table S3). Among 3,950 gene orthologs highly conserved in tetrapods, 2,793 (70.7%) full-length genes were detected in the *P. ornatum* genome, a low number likely reflecting gene fragmentation in the assembly, but also likely a consequence of the lack of an anuran-specific BUSCO database. By contrast, 99.12% of the short reads aligned back to the draft genome assembly, indicating a high level of genome completeness. An oligonucleotide (kmer)-based statistical approach (31) estimated the genome size of *P. ornatum* at 1.06 Gb, indicating that our de novo genome assembly covered

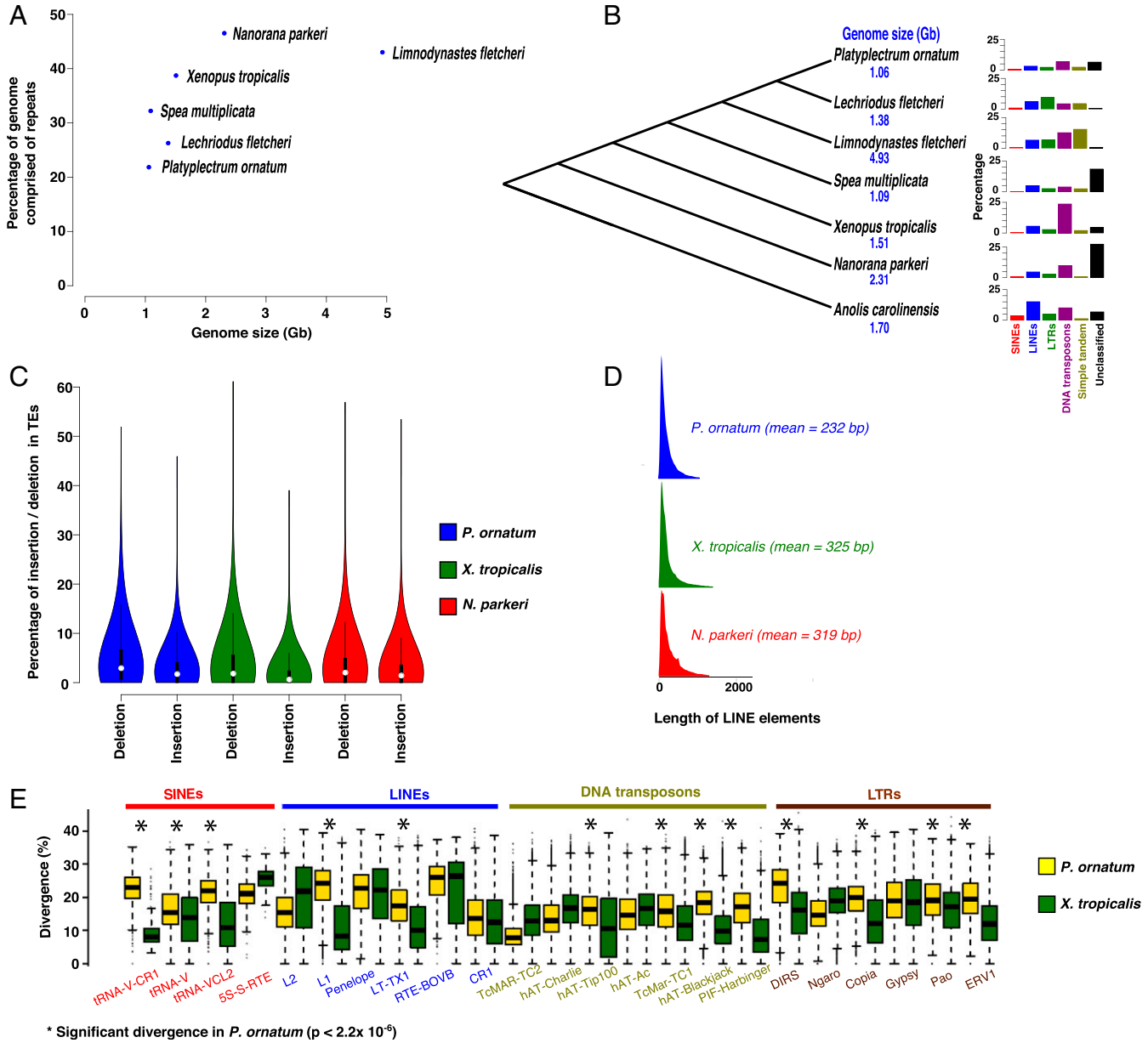


**Fig. 2.** Comparative analysis of TEs among anurans. (*A*) Correlation of TE content and genome size. (*B*) Comparison of various categories of TE across anuran genomes. (*C*) Distribution of insertion and deletion rates across TE in *P. ornatum*, *X. tropicalis*, and *N. parkeri*. (*D*) Distribution of the length of LINEs in *P. ornatum*, *X. tropicalis*, and *N. parkeri*. (*E*) Distribution of sequence divergence for different TE families in *P. ornatum* and *X. tropicalis*. Sequence divergence was calculated from alignments of TE families (*x*-axis) against consensus TE sequences from the Repbase database (50). Asterisks indicate that the divergence in *P. ornatum* for a given TE family is significantly different than that in *X. tropicalis* ($P < 2.2 \times 10^{-6}$).

about 98.3% (1.05 out of 1.06 Gb) of the expected genome size of *P. ornatum* (*SI Appendix*, Fig. S2). The overall assembly statistics indicated that >98% of expected genome of *P. ornatum* was assembled, despite the overall low genomic contiguity of the assembly (i.e., scaffold $N_{50}$ of 30 kb).

We also performed RNA sequencing from multiple tissues (brain, heart, muscle, and gonads) from three adult males and three adult females to aid in genome annotation (*SI Appendix*, Table S4). We identified 21,913 protein-coding genes (*SI Appendix*, Table S3) in the annotated *P. ornatum* genome. To better interpret genome size evolution in the Limnodynastidae and to obtain estimates of repeat content in related frogs, we skimmed the genomes of two species closely related to *P. ornatum*, *Le. fletcheri* (45×), and *Li. fletcheri* (21×), using a shallow sequencing approach (32) (*SI Appendix*, Table S5). The kmer-based approach (31) estimated the genome sizes of *Le. fletcheri* and *Li. fletcheri* to be 1.38 Gb and 4.93 Gb, respectively (*SI Appendix*, Table S6 and Fig. 2*A*), and, in the case of *Li. fletcheri*, our estimate was larger than previous estimates (19, 33). Additionally, kmer distributions were consistent with diploidy in this species (*SI Appendix*, Fig. S2).

**Reduction of Repeat Abundance in *P. ornatum*.** Amphibians have the largest reported variation in genome size among vertebrates (17, 21), and patterns of genome size among the sequenced frog genomes is consistent with that trend (*SI Appendix*, Table S6). The variation in genome size has been attributed mostly to stochastic genomic processes such as polyploidization (34), size and number of introns (35), complexity of regulatory regions (36), and, most importantly, proportion of repeats and TEs (16). We characterized the TEs in *P. ornatum* using homology-based (37) and de novo approaches (38). We also used similar methods to detect and annotate the repeat content in three additional publicly available frog genomes, *Xenopus tropicalis* (39), *Nanorana parkeri* (40), *S. multiplicata* (20), and the squamate reptile *Anolis carolinensis* (41) as an outgroup (Fig. 2*B*). The proportion of repeat content in the *P. ornatum* genome estimated directly from short sequence reads was 21.86%, and the estimate from the de novo genome assembly was 21.93%. The de novo approach provided improved repeat category annotations compared to homology-based estimation (*SI Appendix*, Fig. S3) so we used a similar method to characterize the repeat landscape in *Limnodynastes* and *Lechriodus*. The genome size and repeat content in *P. ornatum* is the lowest among published frog genomes (Fig. 2 *A* and *B*). Consistent with results from previous studies in plants (42) and animals (43), we observed a strong correlation between genome size and abundance of TEs in frogs (Fig. 2*A*). The *P. ornatum* genome exhibits consistent reduction in multiple classes of repetitive elements (Fig. 2*B*), suggesting that reduction across repetitive elements in general is an important process for genome size reduction in *P. ornatum.* These results also suggest that expansion of DNA transposons is a common feature in larger frog genomes (Fig. 2*B*) and is a driving mechanism in genome size evolution and variability in frogs.

Previous studies have demonstrated that evolutionary changes in genome size also result from differential rates of DNA gain or loss, particularly in sequences of TEs (44–46). We explored if similar processes also occur in frogs. We followed methods similar to ref. 47 and limited our analysis only to long interspersed nuclear elements (LINEs) (Table 1) because short interspersed nuclear elements (SINEs) only comprised a small portion of the total TE landscape; DNA transposons, due to their cut-and-paste mechanism, potentially bias estimates of rates of insertion and deletion; and long terminal repeats (LTRs), which are products of unequal intrastrand recombination (48), may bias estimates of divergence among repeats. We first aligned each LINE to its consensus sequence in three frog genomes (*P. ornatum, X. tropicalis*, and *N. parkeri*). We further estimated the rates of insertion and deletion by examining gaps in consensus and repeat sequence alignments. Rates of insertion and deletion within TEs were calculated by dividing the length sum of insertions or deletions by the length sum of TEs within the window in question. Rates of deletion and insertion were defined as the total number of base pairs deleted and inserted per repeat sequence, respectively. The deletion bias was calculated by dividing the number of deleted nucleotides by the number of inserted nucleotides within each window. This method will undercount insertions and deletions if they have occurred multiple times in a single region, but most of the events we observed in different TEs appeared dispersed and nonoverlapping, suggesting that multiple hits of indels are not common.

The rate of deletions outnumbered the rate of insertions in each species (Fig. 2*C*), a common pattern observed across all three domains of life (49). The deletion bias statistic indicated that deletion bias correlated with genome size among these three frog species and was the highest in *P. ornatum* (Table 1). Consistent with a higher deletion bias, the average length of TEs genome-wide was also shorter in *P. ornatum* compared to two other frog species with relatively larger genome size (Fig. 2*D*). These results indicate that a higher deletion bias and relatively shorter TEs have contributed to reduction in genome size in *P. ornatum.* Using the alignment of repeat elements to consensus sequences in the Repbase database (50), we also compared the sequence divergence of different categories of TEs in *P. ornatum* and *X. tropicalis*. The majority of TEs were more divergent in *P. ornatum* than those of *X. tropicalis* (Fig. 2*E*), indicating that they may have proliferated earlier, with subsequent reduction via deletion.

**Diversification and Increased Expression of Piwi Proteins and Piwi-Interacting RNAs.** Piwi proteins and Piwi-interacting RNAs (piRNAs) play key roles in suppressing TEs (51, 52). A recent study in fish with miniaturized genomes suggested copy number expansion of Piwi genes as a potential driver of reduced TE abundances via transposon silencing (12). We asked if similar mechanisms might exist in frogs. We identified six copies of Piwi genes in *P. ornatum* whereas only three copies are known in the *X. tropicalis*, whose genome is ~43% bigger (*SI Appendix*, Table S6). Through visual inspection, we confirmed that the presence of six full-length copies of Piwi genes was not a consequence of the fragmented assembly. We also confirmed the presence of six copies of Piwi genes in *P. ornatum* by examining their gene expression profiles in four different tissues from six individuals (*SI Appendix*, Table S7). To explore the evolution of Piwi gene families, we constructed a maximum likelihood phylogeny of all Piwi genes in *P. ornatum* and other major vertebrate lineages (*X. tropicalis, A. carolinensis, Gallus gallus, Danio rerio*, and *Homo sapiens*) (Fig. 3*A*). Five out of six copies of Piwi-like genes in *P. ornatum* clustered together with the Piwi-4 gene in human, indicating that copy-number expansion of the Piwi-4 gene has occurred on the lineage leading to *P. ornatum*. We examined evidence of episodic diversifying selection on the branches leading to Piwi-like genes in *P. ornatum* using the adaptive branch-site random effects likelihood (aBSREL) model in HyPhy (53). Significance was assessed using likelihood ratio tests ($P < 0.05$), which, after correcting for multiple testing, indicated that four out of six Piwi genes (all but Piwi-3 and -6) are under diversifying positive selection in *P. ornatum* (*SI Appendix*, Table S8).

Piwi genes have well-known and highly conserved roles in repressing TEs and are known to be expressed in germline cells (54). Using our transcriptome dataset, we examined expression levels of Piwi genes in gonads, brain, heart, and muscle of *P. ornatum*. We found a significantly higher expression of Piwi genes in gonads compared to other tissues (Fig. 3*B* and *SI Appendix*, Table S7). We also examined the expression of additional

Table 1. Mean rates of insertion, deletion, and deletion bias for LINEs in three frog genomes

| Species | Genome size, Gb | Insertion rate, % | Deletion rate, % | Deletion bias |
|---|---|---|---|---|
| *P. ornatum* | 1.06 | 4.90 | 1.94 | 2.52 |
| *N. parkeri* | 2.31 | 3.67 | 2.62 | 1.40 |
| *X. tropicalis* | 1.51 | 3.90 | 1.82 | 2.14 |

piRNA pathway genes in gonads (*SI Appendix*, Table S9). We used 19 piRNA pathway genes identified in *Drosophila melanogaster* (55) to identify orthologs in *P. ornatum* using a homology-based search on its genome and genome annotation. Expression analysis showed that the majority of piRNA pathway genes showed increased expression in gonads (*SI Appendix*, Table S9). These results indicated that gonad-enriched expression of Piwi and other piRNA pathway genes possibly contribute to similar mechanisms of TE repression in germline cells of *P. ornatum*. However, piRNA pathway genes can be more active in the germline than in soma in a genome of any size, and the ideal comparison would be to compare gene expression in germline among frog species with different genome sizes, which is outside the scope of this paper. Our measurement of Piwi and piRNA pathway gene expression in an amphibian germline will hopefully spur further research along these lines.

**Relative Contributions of Different Genomic Compartments to Genome Size Variation.** Previous studies of vertebrate genomes have shown that, in addition to lower repetitive content, reduction of genome size is also characterized by shorter gene lengths, mostly as a result of reduced intron size (12, 56). As expected, the *P. ornatum* and *X. tropicalis* genomes, the two genomes whose genomic compartments were readily calculated from annotations, both exhibit low percentages of coding regions, as calculated from 2,210 orthologous single copy genes, and higher percentages devoted to introns and intergenic regions (*SI Appendix*, Table S10). We identified a significant difference in intron sizes between the two species (Fig. 4*B*), and introns appear to contribute more to genome size reduction in *P. ornatum* relative to *X. tropicalis* than intergenic regions; whereas intergenic regions proportionally comprise a greater fraction of the genome in *P. ornatum* (81.5%) vs. *X. tropicalis* (62.9%), introns comprise proportionally less (16.6% vs. 35.5%, respectively). For example, the *KAZN* gene, involved in epidermal differentiation, exhibits similar intron–exon structure but drastic differences in intron length between the two species (Fig. 4*C*). TEs contribute only partially to the small intronic compartment in *P. ornatum*, comprising less than 5% of intronic DNA. These results suggest that reduction in TEs and introns comprise the major genome compartments contributing to the miniaturized genome of *P. ornatum*.

**Nucleotide Composition of the *P. ornatum* Genome.** The genomic landscape of TEs is shaped by local recombination rates (57). Because data on genome-wide recombination rates are not available for *P. ornatum,* we instead used guanine–cytosine (GC) content as a proxy for recombination rate as the local rate of recombination rates is positively correlated with GC content (58). Our expectation was that, if GC content predicts local recombination rate, in turn, it predicts higher local rates of deletion of TEs and intron shortening. However, we found no correlation of GC content with intron length (*SI Appendix*, Fig. S4) or length of individual TEs (*SI Appendix*, Fig. S5).

We further characterized the nucleotide composition (GC content) across the genome in *P. ornatum*, measuring it separately for coding and noncoding sequences. As expected, GC content was higher in coding regions (*SI Appendix*, Fig. S6), due in part to codon usage bias (59). In addition, we also examined the distribution of GC content, TE abundance, and number of heterozygous sites along 50-kb windows across the *P. ornatum* genome (*SI Appendix*, Fig. S7*A*). There was a positive correlation between TE abundance and GC content across the genome (*SI Appendix*, Fig. S7*B*), indicating higher occurrence of TE in GC-rich regions of the genome. However, our results indicated that neither TE abundance nor GC content is related to genome-wide heterozygosity (*SI Appendix*, Fig. S7*B*).

**Gene Family Evolution.** To examine the possible contribution of multigene families to genome reduction in *P. ornatum*, we compared distributions of gene family numbers among various vertebrates: human (*H. sapiens*), zebra fish (*D. rerio*), chicken (*G. gallus*), *Anolis* (*A. carolinensis*), and western clawed frog (*X. tropicalis*), together with the recently published anuran genome for the Mexican spadefoot toad (*S. multiplicata*) (20). We identified 15,083 orthologous gene families; 6,888 were shared among all seven species, and 9,007 were shared between the three frog species (*X. tropicalis, P. ornatum*, and *S. multiplicata*) (*SI Appendix*, Fig. S8 and Table S11). Of these 9,007 gene families, 117 were unique to the three frog species whereas 66 of these gene families were present only in *P. ornatum*.

We used a canonical phylogeny of vertebrates (60) (*SI Appendix*, Fig. S10) to analyze changes in gene family size across lineages leading to each of these six species. The high rate of gene duplication observed in *P. ornatum* (*SI Appendix*, Table S12) is not a remnant of a whole-genome duplication, given that kmer analysis strongly suggests a diploid genome (*SI Appendix*, Fig. S2), although it could be partly driven by the fragmented assembly; however, our assembly is of similar quality to the de novo assembly of *S. multipicata* (20). Using a Bayesian approach in which rates of birth and death are assumed to be similar across gene families but could vary across lineages (61), we analyzed a subset (9,900) of gene families to estimate global gene family dynamics and rates of birth and death across lineages (Fig. 3*C*). Surprisingly, we found little evidence for reductions in the size of gene families in *P. ornatum*. If anything, the probability of an orthologous gene family expanding in *P. ornatum* was higher than in the other lineages compared. We estimated that ∼1,170 gene families were contracting on the lineage leading to *Platyplectrum* whereas ∼5,269 and ∼3,460 gene families were unchanging or expanding in size, respectively, in *P. ornatum* (*SI Appendix*, Tables S13–S15). These numbers remained relatively similar between species and among processes (contraction, no change, expansion) when the birth–death rate was allowed to vary across lineages (*SI Appendix*, Table S14). Reductions in gene family size therefore do not appear to contribute to total genome size reduction in *P. ornatum*. Across the tree, the number of molecular functions that were enriched in the subset of gene families exhibiting higher-than-expected rates of birth and death (*SI Appendix*, Table S16) was greater than in the subset exhibiting lower-than-expected rates (*SI Appendix*, Table S17).

**Sex-Biased Gene Expression.** We further utilized *P. ornatum* genome annotation and RNA-seq data from gonads of three adult males and three adult females to identify sex-biased gene expression in *P. ornatum*. We utilized a reference-free approach (62) to quantify transcripts expressed in gonads of each

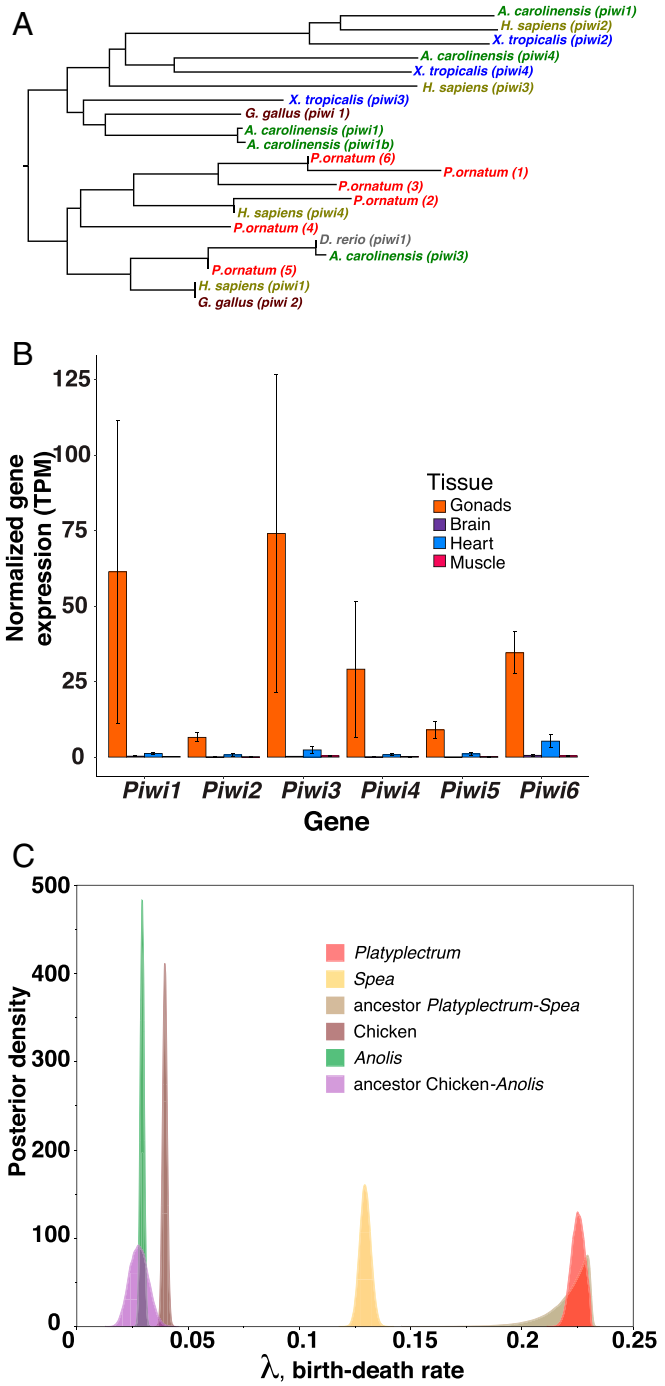


**Fig. 3.** Copy number expansion and expression of Piwi genes. (*A*) Maximum likelihood phylogenetic tree of the Piwi gene family in *P. ornatum* and other major vertebrate lineages. (*B*) Gene expression score (transcripts per million reads [TPM]) of the Piwi genes 1 to 6 in gonads and other tissues in *P. ornatum* females. Mean and SE are indicated ($n$ = 3 for all gene/tissue combinations). (*C*) Distribution of birth-date rate among 9,900 gene families estimated by a Bayesian method (61).

individual and identified differentially expressed transcripts between male (testis) and female gonads (ovaries), using a modeling approach that discriminates between biological variance and variance across the population from which the samples were derived (63). There were 90 differentially expressed ($P < 1e^{-10}$) transcripts (Fig. 5*A* and *SI Appendix*, Table S18) from 57 genes, 54 of which are highly expressed in males while only three are highly expressed in females (Table 2). We performed gene ontology analysis to identify significantly enriched gene ontology terms in this set of differentially expressed genes (false discovery rate [FDR] $P$ value < 0.05), revealing gene ontology terms related to reproduction (e.g., sexual reproduction, reproductive processes, gamete generation, folic acid transportation, triglyceride processes) as enriched (Fig. 5*B*). We further performed gene network analysis using these differentially expressed genes and their respective gene ontology terms. The results demonstrated how many key genes differentially expressed in males and females of *P. ornatum* were interconnected in metabolic pathways associated with reproduction (Fig. 5*C*). The list includes genes such as anti-Müllerian hormone, fibroblast growth factors, cytochrome P450 family 17 subfamily A member 1, and serpin family A member 5 (*SI Appendix*, Table S18), all of which are recognized as key genes associated with reproductive processes, particularly their role in testicular development in a wide range of vertebrates (64–67). Most of these genes are possible candidates associated with male gonad differentiation in *P. ornatum*.

We also performed a comparison of gene expression from adult somatic tissues (brain, heart, and muscle) from three males and three females to identify any evidence of intersexual differences in gene dosage that might offer clues about modes of sex determination in this species. Our analysis did not reveal evidence of gene dosage bias between male and female adult somatic tissues although a very small proportion of genes exhibited differential expression (Table 2 and *SI Appendix*, Fig. S9). These results are consistent with the sex chromosomes being homomorphic in *P. ornatum*, as in the majority of amphibians. To further corroborate this hypothesis, we prepared metaphase chromosomes and karyotypes from two mature adult males and two gravid females. Our karyotyping analysis revealed that both males and females have 11 pairs of chromosomes (2n = 22) without any obvious sex-specific heteromorphism in any pair of chromosomes (*SI Appendix*, Fig. S11). The karyotype results further strengthen the idea that *P. ornatum* has homomorphic sex chromosomes.

## Discussion

The genome of *P. ornatum* is remarkably like that of birds, with reduced repetitive DNA content, intron length, and intergenic distances. The wide variation in genome size in eukaryotes, and the causes and consequences of genome size differences, have fascinated evolutionary biologists (34, 43, 68). It is well appreciated that genome size correlates weakly, if at all, with the extent of coding sequence or organismal complexity but much more consistently with the extent of the noncoding, possibly regulatory sequence (69–71). The amount of repetitive DNA, now recognized as a critical evolutionary component of the genome (72, 73), is closely correlated with genome size in eukaryotes (74). Transposon amplification has been a common feature of genome expansion in larger genomes [e.g., flowering plants (8), conifers (9), or salamanders (21, 75, 76)]. Similarly, reduction in repeat content has been associated with smaller genomes [e.g., fish (12), plants (13), or insects (14)]. Genome size in birds and mammals shows little within-clade variation relative to other taxa, but even they demonstrate differential rates of transposon accumulation (16).

Despite technological advances providing improved sequencing capabilities, many published genome assemblies from frogs (77) are unusually fragmentary. The *P. ornatum* genome reported here is no exception, despite being highly reduced with a low incidence of TEs. In addition to building a draft genome using standard Illumina fragment and a mate pair libraries-based approach (78), we also attempted to improve the genome contiguity using chromatin confirmation and long read sequencing approaches, that have been widely used to generate chromosomal-scale genomes in other taxa (79). The resulting draft genome of *P. ornatum*, even after multiple genome-sequencing approaches,

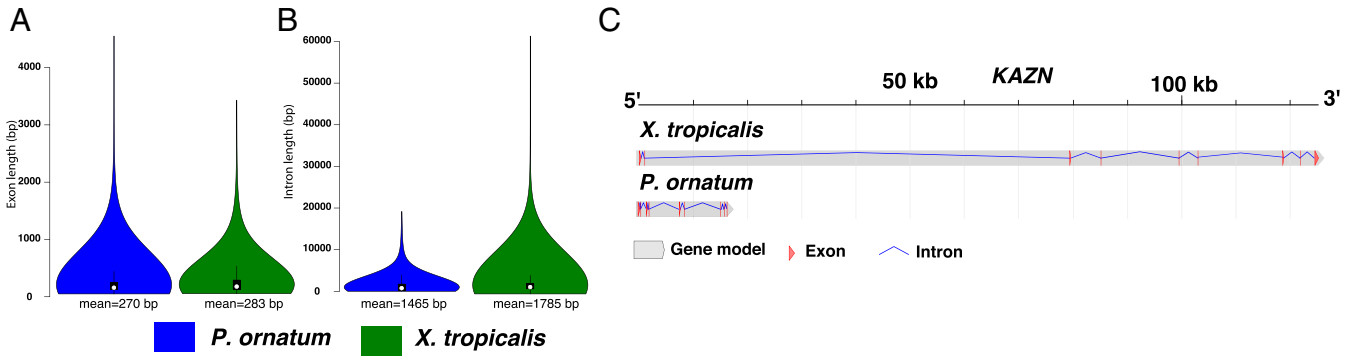


**Fig. 4.** Intron shortening in *P. ornatum*. (*A*) Distribution of exon and (*B*) intron lengths in *P. ornatum* and *X. tropicalis*. Only single copy genes were used for this analysis. (*C*) Example showing reduction of intron size in a specific gene (*KAZN*) in *P. ornatum*.

was not highly contiguous (*SI Appendix*, Table S3). Although there is considerable room for improvement of the current *P. ornatum* genome, this draft assembly nonetheless reveals much about mechanisms of genome size reduction in this burrowing frog.

***Platyplectrum* Genome Size Evolution in Context.** The fact that many of our comparisons involve few frog taxa calls for further accumulation of frog genomes for comparative analysis. A recent and comprehensive phylogenetic analysis of genome size evolution in amphibians helps place the bird-sized genome of *P. ornatum* in a larger phylogenetic context (21). We used their analysis to understand rates of genome size evolution on the lineage leading to *P. ornatum*. In their analysis of 464 amphibians, including 272 frogs, Liedtke et al. (21) documented a number of statistically significant jumps in genome size across the amphibian tree, with one of the most pronounced occurring on the branch leading to *P. ornatum*. We quantified rates of genome size evolution in the lineage leading to *Platyplectrum* by estimating from ref. 21 how much longer each branch would have to be so as to accommodate the observed genome size change, given the mean and variance parameters of the Brownian motion model that served as the null model in their study (*Materials and Methods* and *SI Appendix*, Fig. S12) (21). We found that the branch leading to *Platyplectrum* was among those departing most strongly from an underlying Brownian motion model (*SI Appendix*, Fig. S12*A*). We also found that the amount of apparent genome size change per million years on the branch leading to *Platyplectrum* was among the highest for those branches that were not unusually short and likely to produce spuriously high rates per unit time (*SI Appendix*, Fig. S12*B*). This analysis shows that not only is the genome of *Platyplectrum* one of the smallest among frogs, but its reduction was unusually fast compared to background rates among frogs and other amphibians.

**Mechanisms of Genome Size Reduction.** One key question of interest is to understand underlying mechanisms of differential transposon accumulation, which, in turn, strongly influences genome size. Biases in rates of insertions and deletions are a potential mechanism of genome size evolution (80), but the impact of this mechanism has been debated (18). Our results suggest that the small genome of *P. ornatum* is characterized by reduced content of several classes of TEs (Fig. 2 *A* and *B*). Comparative analysis of TEs in *P. ornatum* with those of two other published frog genomes [*X. tropicalis* (39) and *N. parkeri* (40)] indicated increased deletion bias in TE in *P. ornatum* relative to other frog species with larger genome size (Fig. 2*C*), which perhaps also led to the reduction of average TE length in *P. ornatum* (Fig. 2*D*). TEs in *P. ornatum* also exhibited higher sequence divergence compared to *X. tropicalis* (Fig. 2*E*), suggesting a more ancient proliferation.

Maintenance of genomic integrity over generations is a key selective process as rapid "cut/copy and paste" mechanisms of TEs constantly threaten genomic stability (81). To counter this, specific silencing pathways involving Piwi and piRNA genes are known to be widely employed across eukaryotic lineages (51, 52). In animals, these silencing pathways are particularly important in germline cells as a major line of defense to withstand genomic instability (54). We found evidence of copy number expansion in Piwi genes in *P. ornatum* (Fig. 3*A*) and their increased expression in gonads in relation to other tissues (Fig. 3*B* and *SI Appendix*, Table S7). We also found higher expression of piRNA pathway genes in gonads in relation to other tissues (*SI Appendix*, Table S9). These results show similar mechanisms of Piwi and piRNA-based TE suppression may have occurred in *P. ornatum*, facilitating its rapid genome miniaturization. However, this hypothesis needs detailed functional studies of germline cells in *P. ornatum* to further test this hypothesis. Previous studies of smaller genomes also identified the presence of shorter genes, mostly as a result of reduced intron size and number (12, 56, 82). The *P. ornatum* genome also showed similar features of reduced introns (Fig. 4*B*). However, we found no evidence of gene family size change in *P. ornatum* that could contribute substantially to changes in its genome size.

**Genomic Links with Life History Evolution in Frogs.** In vertebrates, genome size or correlates such as intron size have often been associated with life history complexity (3), metabolic rate (17, 83), flight (84, 85), body size (86), longevity (87), and many other factors (88). A recent study, however, found no link between the number of life stages and genome size in amphibians (21). Our data indicate the complexity of identifying the drivers of genome size evolution at macroevolutionary scales (89). Establishing the order of evolution of traits, such as reduced genome size, development time, and carnivory, is difficult without access to intermediate forms with some but not all of these traits. However, we found similarities in tadpole development and tadpole diet between three frog species: *Platyplectrum, Lechriodus*, and *Spea* (*SI Appendix*, Table S1). Although these comparisons are based on relatively few frog taxa, all three of these species have smaller genome sizes (Fig. 2). Additionally, there are many other frog species with larger genomes falling phylogenetically between these taxa, thus indicating a likely case of convergent genome size evolution, perhaps due to selection for rapid development at the sensitive tadpole stage under similar arid or ephemeral habitat conditions (Fig. 1). Our genome skimming in two related frogs (*Le. fletcheri* and *Li. fletcheri*) shows that genome size was indeed reduced in *Platyplectrum*, as opposed to being ancestrally retained deeper in its lineage, further strengthening our suggestion that the small genomes of *Platyplectrum*, *Spea*, and *Lechriodus* represent independent, convergent events. Genome size has been

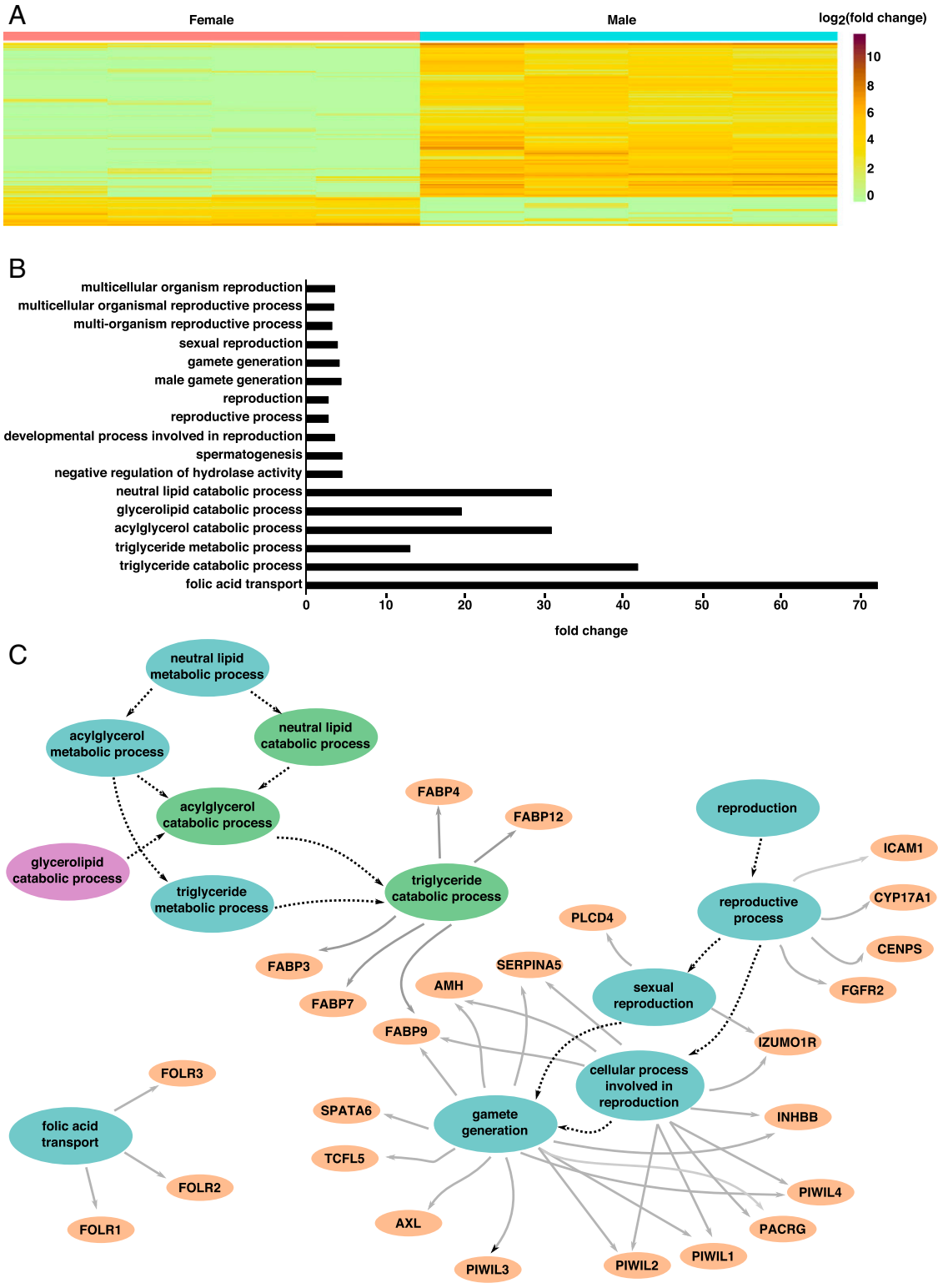


**Fig. 5.** Sex-biased gene expression in *P. ornatum*. (*A*) Heat map showing gene expression score (TPM) among differentially expressed genes ($P < 1e^{-10}$) among male and female gonads in *P. ornatum*. (*B*) Gene ontologies (molecular functions) of differentially expressed genes among male and female gonads in *P. ornatum*. (*C*) Network analysis showing functional interactions of differentially expressed genes among male and female gonads in *P. ornatum*.

found to be strongly linked to the length of the development period (21), which is consistent with findings in this study.

Our study also highlights the challenges faced in macroevolutionary studies attempting to ascertain the drivers of genome size evolution. Reduced genome size was possibly evolved in the ancestor to the *Platyplectrum/Lechriodus* clade (Fig. 2*B*) during the post-Miocene drying of Australia (90). Since that time, this clade has diversified into every major habitat in Australia,

**Table 2. Statistics on differential gene expression among multiple tissues on *P. ornatum***

| Tissues | Total transcripts | Differentially expressed transcripts | Differentially expressed genes | Higher expression in males | | Higher expression in females | |
|---|---|---|---|---|---|---|---|
| | | | | Transcripts | Genes | Transcripts | Genes |
| Gonads | 39,658 | 90 | 57 | 82 | 54 | 8 | 3 |
| Brain | 41,269 | 9 | 5 | 9 | 5 | 0 | 0 |
| Heart | 32,472 | 6 | 5 | 4 | 3 | 2 | 2 |
| Muscle | 25,801 | 5 | 2 | 5 | 2 | 0 | 0 |

including the arid zone, tropical savannahs, temperate forests, and subtropical rain forests. In each of these areas, tadpoles utilize ephemeral ponds for breeding, including, uniquely for a rain forest species, *Le. fletcheri*. This pattern suggests that the combination of reduced genome size, fast, flexible development time, and carnivorous tadpoles may facilitate ecological release of the clade into a wide variety of habitats. The association between genome size and aridity specifically may erode when a key innovation, such as rapid development time, facilitates dramatic ecological diversification into a wide array of nonarid habitats (21). Further studies investigating the link between genome size, aridity, and development are needed, but we currently are limited by the low number of genome size estimates in the Amphibia and detailed natural history information for many species.

**Sex Determination in *Platyplectrum*.** Analyses of sex-biased gene expression have received attention in the past decade due to the advancement of genome sequencing technologies and reduction of sequencing costs although the majority of those studies have focused on model organisms having heteromorphic sex chromosomes with well-assembled genomes. Those studies have revealed nonrandom distributions of sex-biased genes within the genome, rapid sequence divergence, and higher rates of protein evolution, compared to that of autosomes (e.g., faster Z or X effect). Sex-biased genes are under intense selection pressure (91–93), and sex-biased gene expression varies between tissues and stages of development, with a notably higher level of sex-biased gene expression differences occurring between gonadal tissues: e.g., testes and ovaries (91–93). Analysis of sex-biased gene expression in *P. ornatum* has provided us with insights on key candidate genes associated with gonadal development and differentiation.

Sex determination in frogs is primarily genetic, with the majority of the species displaying varying degrees of sexual dimorphisms. However, morphologically differentiated sex chromosomes are rare among frogs, irrespective of their state of heterogamety–whether they are either female (ZZ/WW) or male (XX/XY) heterogametic (94, 95). The occurrence of these nascent or evolutionarily young sex chromosomes in frogs has also facilitated rapid evolution and frequent turnover of sex chromosomes (94–97). In frogs, very few studies have addressed possible sex-biased gene expression, partly due to the scarcity of chromosome-anchored genome assemblies, as well as a lack of heteromorphic sex chromosomes. Recent studies of the European common frog *Rana temporaria*, a species with genotypic sex determination but lacking differentiated sex chromosomes and sex-specific gene dosage compensation, showed that expression of the sex-biased genes are associated with phenotypes rather than genotypes (98, 99). While the spatiotemporal patterns of sex-biased gene expression are evolutionarily conserved, the evolution of chromosome-specific enrichment of sex-biased genes may have followed a different evolutionary trajectory in species with homomorphic or cryptic sex chromosomes: for example, both autosomes and putative sex chromosomes displaying sex-biased expression in the European common frog. The evolutionary mechanisms of such a phenomenon are unknown. Therefore, investigating sex-biased gene expression in species with cryptic or evolutionarily young sex chromosomes, such as in frogs, will enable better understanding of the evolutionary dynamics and persistence of homomorphic sex chromosomes, sexual dimorphism, and sexual conflict, as well as the evolution of genes and genomes in general.

## Materials and Methods

**Genome Sequencing and De Novo Assembly.** Fieldwork and specimen acquisition was approved by the Macquarie University Animal Ethics Committee (protocol ARA 2019-010), and conducted under New South Wales Office of Environment and Heritage Scientific Licence SL101269 to S.C. Additional processing of specimens was approved by the University of Canberra Animal Ethics Committee (protocol number AEC 18-01 to T.E.). Genomic sequence data of an ornate burrowing frog (*P. ornatum*) was generated from DNA isolated from the muscle of an adult female. Based on the expected genome size of 0.96 Gb (19), the *P. ornatum* genome was sequenced to ~140× coverage (*SI Appendix*, Table S2), on an Illumina HiSeq 2500 platform using two different sequence libraries: 1) a fragment library with average insert size of 220 bp, and 2) a jumping library with average insert size of 6 kb, generating ~1.2 billion paired end reads with read length of 125 bp each. An initial draft genome was generated from these Illumina sequencing reads using ALLPATHS-LG (78). Furthermore, we also used additional sequencing approaches (chromatin confirmation and direct long reading sequencing) (see *SI Appendix* for additional details) for better assignment of contigs into contiguous scaffolds.

We assessed the genome contiguity using scaffold $N_{50}$ statistics (*SI Appendix*, Table S3). We also examined the gene-space completeness using BUSCO (100) that utilizes near-universal single-copy orthologs from the OrthoDB database to estimate the evolutionary-informed expectation of gene content in a de novo genome assembly. To assess the overall genome completeness, we also mapped short sequence reads generated from Illumina sequencing libraries back to the de novo genome assembly. In addition, we also used a kmer-based statistical approach (31) to get an independent estimation of genome size based on short sequence reads without relying on final de novo genome assembly.

**Transcriptome Sequencing and Genome Annotation.** We also collected brain, heart, muscle, and gonad tissues from one male and one adult female *P. ornatum* individual and generated ~540 million transcriptome reads to use the data for genome annotations (*SI Appendix*, Table S4). We first generated a female transcriptome assembly combining RNA-seq data from brain, gonad, heart, and muscle using Trinity (101). In addition, we mapped RNA-seq data from the male individual to the reference *P. ornatum* genome assembly using TopHat (102) and extracted exon/intron junction information for downstream usage for genome annotations. The annotation of gene content in the *P. ornatum* genome assembly was generated using the custom annotation pipeline using MAKER (103). We combined protein homology evidence (using publicly available protein sets of human, mouse, chicken, anole lizard, and western clawed frog, downloaded from the Ensemble v.93 database), RNA-based evidence (using female transcriptome assembly and exon/intron junction information from male RNA-seq data), and ab initio gene prediction methods for genome annotations.

**Genome Sequencing of Additional Two Closely Related Species.** We also collected liver samples from two additional species in the subfamily Limnodynastidae: the sister genus to *P. ornatum* (*Le. fletcheri*) and the more distantly related *Li. fletcheri* as a loan from Australian Museum collections to interpret genome size evolution within the subfamily Limnodynastidae (Australian ground frogs). We generated fragment libraries with an average insert size of 220 bp for both these species and sequenced in the Illumina HiSeq 2500 platform, generating ~1.2 billion paired end reads, each with a read length of 125 bp each. We utilized the kmer-based statistical approach (31) as above using these short sequence reads to estimate genome size in *Li. fletcheri and Le. fletcheri.*

**Characterizing Repeat Landscape in the Genomes.** To annotate TEs in *P. ornatum,* we combined homology-based and de novo approaches using Repeatmasker (37). We used the reference repeat library downloaded from the Repbase database (50) and an additional custom-built species-specific repeat library generated for *P. ornatum* using RepeatModeler for increased accuracy of detection and annotation of repeats. We also used a similar approach to detect and annotate TEs in three additional publicly available frog genomes—*X. tropicalis* (39), *N. parkeri* (40), and *S. multiplicata* (20)—and also in the squamate reptile *A. carolinensis* (41) as an outgroup.

The Repeatmasker-based approach discussed above is the most extensively used method to identify repeats in the genome (104). However, it requires a good quality reference genome and hence may not be appropriate for cases where such genomes are not available. We used dnaPipeTE (105), which performs de novo assembly of repeats directly from short sequence reads followed by characterization of repetitive landscape for comparative analysis and hence can be used for cases where reference genomes are not available. We first ran dnaPipeTE using *P. ornatum* short sequence data and compared the results to the Repeatmasker-based approach to evaluate the efficiency of dnaPipeTE. The proportion of repeat content in *P. ornatum* estimated directly from short sequence reads was 21.86%, and the one estimated from the de novo genome assembly was 21.93%. In addition, estimation directly from short sequence reads provided improved repeat category annotations, compared to reference genome-based estimation. This indicated that repeat content characterization based on short sequence reads provides similar results to that of "gold standard" methods based on reference genomes. Hence, we further used dnaPipeTE to characterize repeat landscapes in *Li. fletcheri and Le. fletcheri* where only data on short sequence reads were available.

We also estimated the amount of various repetitive elements (short- and long-interspersed repeats [SINEs and LINEs], LTRs, DNA transposons, and other repeats) and compared the repeat landscapes among six frog species: three species, *X. tropicalis*, *N. parkeri* and *S. multiplicata,* whose genomes are publicly available, and three Australian species (Limnodynastidae) from this study (*P. ornatum*, *Li. fletcheri*, and *Le. fletcheri*).

**Distribution of Intron and Exon Sizes across the Genomes.** We first generated an unbiased gene set for analyzing the relative lengths of genes, exons, and introns by selecting only 2,210 genes that had single copy across six species (human, chicken, zebrafish, anole lizard, western clawed frog, and ornate burrowing frog). Based on this gene set, we compared the distribution of exon and intron lengths among two species of frogs (*X. tropicalis* and *P. ornatum*) using the standard *t* test statistic.

**Gene Family Evolution.** To yield insights into the divergence of the *P. ornatum* genome in relation to major vertebrate lineages, we compared its genome with those of human (*H. sapiens*), zebra fish (*D. rerio*), chicken (*G. gallus*), Anole lizard (*A. carolinensis*), and western clawed frog (*X. tropicalis*). Clustering of orthologous genes among these six species was done using OrthoFinder (106), using the canonical tree for vertebrates (60) (see *SI Appendix* for details). We analyzed the changes in gene family size across six lineages by computing the Bayesian estimate of the birth and death parameter in multigene families across these lineages (61).

**Analysis of Rates of Genome Size Change in Frogs.** Liedtke et al. (21) estimated statistically significant jumps in genome size across amphibians using a Lévy process model (see *SI Appendix*, Fig. S12 and legend) and identified *Platyplectrum* as one of 10 branches on which deviations from a Brownian motion model occurred. We obtained estimates of the amount of genome size change per branch from their analysis. These estimates were then divided by the branch length in absolute time from their time tree, to yield approximate rates of genome size change per unit time. We plotted both the branch lengths in terms of trait change (*SI Appendix*, Fig. S12*A*) and in amount of change per unit time (*SI Appendix*, Fig. S12*B*) to place *Platyplectrum* genome size change in a larger context.

**Sex-Biased Gene Expression.** To enhance our understanding of molecular mechanisms associated with sex determination in *P. ornatum*, we characterized the transcriptomes of three adult males and three adult females (phenotypic sex in these individuals was determined by dissection, and all six individuals had mature gonads). A total of 87 Gb of RNA-seq data were generated from four tissues (brain, gonads, heart, and muscle) from these six individuals. A reference transcriptome assembly was first generated from a female individual using Trinity (101), and the abundance of each transcript was quantified using Kallisto (62), which does not require a reference genome for RNA-seq quantification.

**Mitotic Metaphase Chromosome Preparations and Karyotyping.** Mitotic metaphase chromosomes were prepared from bone marrow, kidney, and liver cells following the protocol described in ref. 107, with slight modifications. Briefly, tissue fragments were first dissociated in 10 mL of RPMI 1640 cell culture medium. Once dissociated, 400 μL (0.5%) of colchicine (0.0125% final concentration) was added to the cell suspension and incubated at room temperature for 30 min. After the incubation, cell suspension was centrifuged at 1,500 RPMI for 8 min. After centrifugation, the supernatant was discarded, and the cell pellet was resuspended in 5 mL of 0.075 M KCl and incubated for 30 min at 37 °C for hypotonic treatment. After the hypotonic treatment, cell suspension was centrifuged again as mentioned above, and the cell pellet was resuspended in 10 mL of freshly prepared Carnoy's solution (methanol:acetic acid 3:1). The resuspended cell resuspension was washed three times by centrifugation as above, and, after the third centrifugation, the cell pellet was fixed by resuspending in 500 μL of Carnoy's solution for further analysis and storage. For karyotyping analysis, 10 μL of fixed cell suspension was dropped onto a glass microscope slide and stained with 4′,6-diamidino-2-phenylindole (DAPI) mounted with an antifade medium (Vectasheld). Metaphase chromosome images were captured using a Zeiss Axioplan epifluorescence microscope equipped with a charge-coupled device (CCD) camera. ISIS scientific imaging software (Metasystems, Altlussheim, Germany) was used for image capture and karyotyping.

**Data Availability.** The draft genome of *P. ornatum* and all raw sequencing data in this study have been submitted to the National Center for Biotechnology Information (BioProject ID PRJNA637083; *P. ornatum* genome: JABWIB000000000; *P.ornatum* sequencing reads: SAMN15105358; *Lechriodus fletcheri* sequencing reads: SAMN17190749; *Limnodynastes fletcheri* sequencing reads: SAMN17191081). Transcriptome data for all samples are in GenBank (accession ID SAMN17191109–SAMN17191132). Additional data and files for analysis are available in the Dryad Repository (DOI: 10.5061/dryad.73n5tb2w9).

**ACKNOWLEDGMENTS.** We thank Simon Yung Wa Sin and Daren Card (Harvard University) for laboratory assistance in preparing sequencing libraries of *P. ornatum;* Zuzana Majtanova, Foyez Shams, and Rod Ubrihien for laboratory assistance in tissue collections, chromosome preparation, and karyotyping; Jodi Rowley for providing samples of *Li. fletcheri* and *Le. Fletcheri*; and the curators and curatorial associates of the Department of Herpetology of the Museum of Comparative Zoology at Harvard University for assistance with sample transportation and storage. We thank Yong Zhang for helpful comments on early drafts of this manuscript, the three reviewers for helpful reviews, and Jim Hanken for helpful discussion in the early stages of this project. We thank Christoph Liedtke and Ivan Gomez-Mestre for their help in placing *Platyplectrum* genome size change in context. Additional sequencing was performed at the Harvard University Bauer Core Facility and we thank the staff for assistance. Computations were performed on the Cannon computing cluster at Harvard University and supported by Harvard Research Computing. This work was supported in part by an internal research grant by the Institute for Applied Ecology, University of Canberra (to T.E.). S.L. was supported by a Wenner Gren Postdoctoral Fellowship. J.S.K. thanks the Australian Research Council for ongoing support. S.V.E. thanks the National Science Foundation and Harvard University for support. Page charges for this paper were covered in part by a grant from the Wetmore Colles Fund of the Museum of Comparative Zoology.

1. T. R. Gregory, Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biol. Rev. Camb. Philos. Soc.* **76**, 65–101 (2001).
2. A. E. Mirsky, H. Ris, The desoxyribonucleic acid content of animal cells and its evolutionary significance. *J. Gen. Physiol.* **34**, 451–462 (1951).
3. T. R. Gregory, Genome size and developmental complexity. *Genetica* **115**, 131–146 (2002).
4. R. L. Mueller, T. R. Gregory, S. M. Gregory, A. Hsieh, J. L. Boore, Genome size, cell size, and the evolution of enucleated erythrocytes in attenuate salamanders. *Zoology (Jena)* **111**, 218–230 (2008).
5. C. L. Organ, A. M. Shedlock, A. Meade, M. Pagel, S. V. Edwards, Origin of avian genome size and structure in non-avian dinosaurs. *Nature* **446**, 180–184 (2007).
6. G. Roth, J. Blanke, D. B. Wake, Cell size predicts morphological complexity in the brains of frogs and salamanders. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4796–4800 (1994).
7. A. C. Gerstein, S. P. Otto, Ploidy and the causes of genomic evolution. *J. Hered.* **100**, 571–581 (2009).
8. J. L. Bennetzen, Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica* **115**, 29–36 (2002).
9. B. Nystedt *et al.*, The Norway spruce genome sequence and conifer genome evolution. *Nature* **497**, 579–584 (2013).
10. S. Nowoshilow *et al.*, The axolotl genome and the evolution of key tissue formation regulators. *Nature* **554**, 50–55 (2018).

11. B. Venkatesh, P. Gilligan, S. Brenner, Fugu: A compact vertebrate reference genome. *FEBS Lett.* **476**, 3–7 (2000).
12. M. Malmstrøm *et al*., The most developmentally truncated fishes show extensive hox gene loss and miniaturized genomes. *Genome Biol. Evol.* **10**, 1088–1103 (2018).
13. E. Ibarra-Laclette *et al*., Architecture and evolution of a minute plant genome. *Nature* **498**, 94–98 (2013).
14. J. L. Kelley *et al*., Compact genome of the Antarctic midge is likely an adaptation to an extreme environment. *Nat. Commun.* **5**, 4611 (2014).
15. G. I. M. Pasquesi *et al*., Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. *Nat. Commun.* **9**, 2774 (2018).
16. A. Kapusta, A. Suh, C. Feschotte, Dynamics of genome size evolution in birds and mammals. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E1460–E1469 (2017).
17. E. Waltari, S. V. Edwards, Evolutionary dynamics of intron size, genome size, and physiological correlates in archosaurs. *Am. Nat.* **160**, 539–552 (2002).
18. T. R. Gregory, Insertion-deletion biases and the evolution of genome size. *Gene* **324**, 15–34 (2004).
19. E. Olmo, A. Morescalchi, Genome and cell size in frogs: A comparison with salamanders. *Experientia* **34**, 44–46 (1978).
20. F. Seidl *et al*., Genome of *Spea multiplicata*, a rapidly developing, phenotypically plastic, and desert-adapted spadefoot toad. *G3 (Bethesda)* **9**, 3909–3919 (2019).
21. H. C. Liedtke, D. J. Gower, M. Wilkinson, I. Gomez-Mestre, Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate. *Nat. Ecol. Evol.* **2**, 1792–1799 (2018).
22. H. A. Horner, H. C. Macgregor, C value and cell volume: Their significance in the evolution and development of amphibians. *J. Cell Sci.* **63**, 135–146 (1983).
23. G. A. Wyngaard, E. M. Rasch, N. M. Manning, K. Gasser, R. Domangue, The relationship between genome size, development rate, and body size in copepods. *Hydrobiologia* **532**, 123–137 (2005).
24. S. K. Sessions, A. Larson, Developmental correlates of genome size in plethodontid salamanders and their implications for genome evolution. *Evolution* **41**, 1239–1251 (1987).
25. S. Clulow, M. Swan, *A Complete Guide to Frogs of Australia* (Australian Geographic, 2018).
26. M. Anstis, *Tadpoles and Frogs of Australia* (New Holland Publishers, 2017).
27. P. Kern, R. L. Cramp, C. E. Franklin, Temperature and UV-B-insensitive performance in tadpoles of the ornate burrowing frog: An ephemeral pond specialist. *J. Exp. Biol.* **217**, 1246–1252 (2014).
28. J.-M. Belton *et al*., Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276 (2012).
29. S. L. Amarasinghe *et al*., Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* **21**, 30 (2020).
30. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
31. G. W. Vurture *et al*., GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
32. L. A. Ripma, M. G. Simpson, K. Hasenstab-Lehman, Geneious! Simplified genome skimming methods for phylogenetic systematic studies: A case study in Oreocarya (Boraginaceae). *Appl. Plant Sci.* **2**, 1400062 (2014).
33. E. Olmo, Quantitative variations in the nuclear DNA and phylogenesis of the Amphibia. *Caryologia* **26**, 43–68 (1973).
34. J. Pellicer, O. Hidalgo, S. Dodsworth, I. J. Leitch, Genome size diversity and its impact on the evolution of land plants. *Genes (Basel)* **9**, 88 (2018).
35. A. E. Vinogradov, Intron-genome size relationship on a large evolutionary scale. *J. Mol. Evol.* **49**, 376–384 (1999).
36. A. V. Markov, V. A. Anisimov, A. V. Korotayev, Relationship between genome size and organismal complexity in the lineage leading from prokaryotes to mammals. *Paleontol. J.* **44**, 363–373 (2010).
37. A. F. A. Smit, R. Hubley, P. Green, RepeatMasker Open-4.0. http://www.repeatmasker.org. Accessed 1 March 2019.
38. J. M. Flynn *et al*., RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9451–9457(2020).
39. U. Hellsten *et al*., The genome of the Western clawed frog Xenopus tropicalis. *Science* **328**, 633–636 (2010).
40. Y.-B. Sun *et al*., Whole-genome sequence of the Tibetan frog Nanorana parkeri and the comparative evolution of tetrapod genomes. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E1257–E1262 (2015).
41. J. Alföldi *et al*., The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* **477**, 587–591 (2011).
42. A. Kumar, J. L. Bennetzen, Plant retrotransposons. *Annu. Rev. Genet.* **33**, 479–532 (1999).
43. A. Canapa, M. Barucca, M. A. Biscotti, M. Forconi, E. Olmo, Transposons, genome size, and evolutionary insights in animals. *Cytogenet. Genome Res.* **147**, 217–239 (2015).
44. C. Sun, J. R. López Arriaza, R. L. Mueller, Slow DNA loss in the gigantic genomes of salamanders. *Genome Biol. Evol.* **4**, 1340–1348 (2012).
45. M. B. Frahry, C. Sun, R. A. Chong, R. L. Mueller, Low levels of LTR retrotransposon deletion by ectopic recombination in the gigantic genomes of salamanders. *J. Mol. Evol.* **80**, 120–129 (2015).
46. L. N. van de Lagemaat, L. Gagnier, P. Medstrand, D. L. Mager, Genomic deletions and precise removal of transposable elements mediated by short identical DNA segments in primates. *Genome Res.* **15**, 1243–1249 (2005).
47. K. Nam, H. Ellegren, Recombination drives vertebrate genome contraction. *PLoS Genet.* **8**, e1002680 (2012).
48. K. M. Devos, J. K. M. Brown, J. L. Bennetzen, Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis. *Genome Res.* **12**, 1075–1079 (2002).
49. C.-H. Kuo, H. Ochman, Deletional bias across the three domains of life. *Genome Biol. Evol.* **1**, 145–152 (2009).
50. W. Bao, K. K. Kojima, O. Kohany, Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
51. R. L. Mueller, piRNAs and evolutionary trajectories in genome size and content. *J. Mol. Evol.* **85**, 169–171 (2017).
52. C. Post, J. P. Clark, Y. A. Sytnikova, G.-W. Chirn, N. C. Lau, The capacity of target silencing by Drosophila PIWI and piRNAs. *RNA* **20**, 1977–1986 (2014).
53. M. D. Smith *et al*., Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* **32**, 1342–1353 (2015).
54. P. Rojas-Ríos, M. Simonelig, piRNAs and PIWI proteins: regulators of gene expression in development and stem cells. *Development* **145**, dev161786 (2018).
55. J. Song *et al*., Variation in piRNA and transposable element content in strains of Drosophila melanogaster. *Genome Biol. Evol.* **6**, 2786–2798 (2014).
56. D. E. Neafsey, S. R. Palumbi, Genome size evolution in pufferfish: A comparative analysis of diodontid and tetraodontid pufferfish genomes. *Genome Res.* **13**, 821–830 (2003).
57. Y. Ji, J. A. DeWoody, Genomic landscape of long terminal repeat retrotransposons (LTR-RTs) and solo LTRs as shaped by ectopic recombination in chicken and zebra finch. *J. Mol. Evol.* **82**, 251–263 (2016).
58. S. M. Fullerton, A. Bernardo Carvalho, A. G. Clark, Local rates of recombination are positively correlated with GC content in the human genome. *Mol. Biol. Evol.* **18**, 1139–1142 (2001).
59. J. Romiguier, C. Roux, Analytical biases associated with GC-content in molecular evolution. *Front. Genet.* **8**, 16 (2017).
60. I. Irisarri *et al*., Phylotranscriptomic consolidation of the jawed vertebrate timetree. *Nat. Ecol. Evol.* **1**, 1370–1378 (2017).
61. L. Liu, L. Yu, V. Kalavacharla, Z. Liu, A Bayesian model for gene family evolution. *BMC Bioinformatics* **12**, 426 (2011).
62. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
63. H. Pimentel, N. L. Bray, S. Puente, P. Melsted, L. Pachter, Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat. Methods* **14**, 687–690 (2017).
64. G. Zhai *et al*., Characterization of sexual trait development in cyp17a1-deficient zebrafish. *Endocrinology* **159**, 3549–3562 (2018).
65. B. Le Magueresse-Battistoni, Serine proteases and serine protease inhibitors in testicular physiology: The plasminogen activation system. *Reproduction* **134**, 721–729 (2007).
66. X. Jiang *et al*., The roles of fibroblast growth factors in the testicular development and tumor. *J. Diabetes Res.* **2013**, 489095 (2013).
67. R. P. Grinspon, S. Gottlieb, P. Bedecarrás, R. A. Rey, Anti-müllerian hormone and testicular function in prepubertal Boys with cryptorchidism. *Front. Endocrinol. (Lausanne)* **9**, 182 (2018).
68. K. Kraaijeveld, Genome size and species diversification. *Evol. Biol.* **37**, 227–233 (2010).
69. J. S. Mattick, RNA regulation: A new genetics? *Nat. Rev. Genet.* **5**, 316–323 (2004).
70. J. S. Mattick, M. E. Dinger, The extent of functionality in the human genome. *HUGO J.* **7**, 2 (2013).
71. J. S. Mattick, The genetic signatures of noncoding RNAs. *PLoS Genet.* **5**, e1000459 (2009).
72. E. B. Chuong, N. C. Elde, C. Feschotte, Regulatory activities of transposable elements: From conflicts to benefits. *Nat. Rev. Genet.* **18**, 71–86 (2017).
73. C. Biémont, A brief history of the status of transposable elements: From junk DNA to major players in evolution. *Genetics* **186**, 1085–1093 (2010).
74. M. G. Kidwell, Transposable elements and the evolution of genome size in eukaryotes. *Genetica* **115**, 49–63 (2002).
75. C. Sun *et al*., LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. *Genome Biol. Evol.* **4**, 168–183 (2012).
76. C. Sun, R. L. Mueller, Hellbender genome sequences shed light on genomic expansion at the base of crown salamanders. *Genome Biol. Evol.* **6**, 1818–1829 (2014).
77. NCBI Resource Coordinators, Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **44**, D7–D19 (2016).
78. S. Gnerre *et al*., High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1513–1518 (2011).
79. J. Dekker, K. Rippe, M. Dekker, N. Kleckner, Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
80. D. A. Petrov, Mutational equilibrium model of genome size evolution. *Theor. Popul. Biol.* **61**, 531–544 (2002).
81. H. H. J. Kazazian Jr, Mobile elements: Drivers of genome evolution. *Science* **303**, 1626–1632 (2004).
82. R. Greenhalgh *et al*., Genome streamlining in a minute herbivore that manipulates its host plant. *eLife* **9**, e56689 (2020).
83. N. A. Wright, T. R. Gregory, C. C. Witt, Metabolic "engines" of flight drive genome size reduction in birds. *Proc. Biol. Sci.* **281**, 20132780 (2014).
84. C. B. Andrews, S. A. Mackenzie, T. R. Gregory, Genome size and wing parameters in passerine birds. *Proc. Biol. Sci.* **276**, 55–61 (2009).
85. Q. Zhang, S. V. Edwards, The evolution of intron size in amniotes: A role for powered flight? *Genome Biol. Evol.* **4**, 1033–1043 (2012).
86. L. P. Decena-Segarra, L. Bizjak-Mali, A. Kladnik, S. K. Sessions, S. M. Rovito, Miniaturization, genome size, and biological size in a diverse clade of salamanders. *Am. Nat.* **196**, 634–648 (2020).

87. J. A. Weber *et al*., The whale shark genome reveals how genomic and physiological properties scale with body size. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 20662–20671 (2020).
88. T. R. Gregory, Ed., *The Evolution of the Genome*, (Elsevier Academic Press, Burlington, MA, 2005).
89. J. Blommaert, Genome size evolution: Towards new model systems for old questions. *Proc. R. Soc. B Biol. Sci.* **287**, 20201441 (2020).
90. M. Byrne *et al*., Decline of a biome: Evolution, contraction, fragmentation, extinction and invasion of the Australian mesic zone biota. *J. Biogeogr.* **38**, 1635–1656 (2011).
91. J. E. Mank, The transcriptional architecture of phenotypic dimorphism. *Nat. Ecol. Evol.* **1**, 6 (2017).
92. H. Ellegren, J. Parsch, The evolution of sex-biased genes and sex-biased gene expression. *Nat. Rev. Genet.* **8**, 689–698 (2007).
93. S. Grath, J. Parsch, Sex-biased gene expression. *Annu. Rev. Genet.* **50**, 29–44 (2016).
94. S. D. Sarre, T. Ezaz, A. Georges, Transitions between sex-determining systems in reptiles and amphibians. *Annu. Rev. Genomics Hum. Genet.* **12**, 391–406 (2011).
95. I. Miura, Sex determination and sex chromosomes in Amphibia. *Sex Dev.* **11**, 298–306 (2017).
96. B. L. S. Furman, B. J. Evans, Divergent evolutionary trajectories of two young, homomorphic, and closely related sex chromosome systems. *Genome Biol. Evol.* **10**, 742–755 (2018).
97. D. L. Jeffries *et al*., A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. *Nat. Commun.* **9**, 4088 (2018).
98. W.-J. Ma *et al*., Tissue specificity and dynamics of sex-biased gene expression in a common frog population with differentiated, yet homomorphic, sex chromosomes. *Genes (Basel)* **9**, E294 (2018).
99. W.-J. Ma, P. Veltsos, R. Sermier, D. J. Parker, N. Perrin, Evolutionary and developmental dynamics of sex-biased gene expression in common frogs with proto-Y chromosomes. *Genome Biol.* **19**, 156 (2018).
100. R. M. Waterhouse *et al*., BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **35**, 543–548 (2018).
101. M. G. Grabherr *et al*., Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
102. C. Trapnell, L. Pachter, S. L. Salzberg, TopHat: Discovering splice junctions with RNA-seq. *Bioinformatics* **25**, 1105–1111 (2009).
103. B. L. Cantarel *et al*., MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196 (2008).
104. N. Chen, Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **5**, 4.10.1–4.10.14 (2004).
105. C. Goubert *et al*., De novo assembly and annotation of the Asian tiger mosquito (Aedes albopictus) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (Aedes aegypti). *Genome Biol. Evol.* **7**, 1192–1205 (2015).
106. D. M. Emms, S. Kelly, OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
107. L. A. C. Bertollo, M. B. Cioffi, O. Moreira-Filho, "Direct chromosome preparation from freshwater teleost fishes" in *Fish Cytogenetic Techniques*, C. Ozouf-Costaz, E. Pisano, F. Foresti, L. F. de Almeida Toledo, Eds. (CRC Press, 2015), pp. 21–26.