Conserved chromatin and repetitive patterns reveal slow genome evolution in frogs

Jessen V. Bredeson[1†], Austin B. Mudd[1†], Sofia Medina-Ruiz[1†], Therese Mitros[1], Owen K. Smith[2], Kelly E. Miller[1], Jessica B. Lyons[1,3], Sanjit S. Batra[4], Joseph Park[1], Kodiak C. Berkoff[1], Christopher Plott[5], Jane Grimwood[5], Jeremy Schmutz[5], Guadalupe Aguirre-Figueroa[2], Mustafa K. Khokha[6], Maura Lane[6], Isabelle Philipp[1], Mara Laslo[7], James Hanken[7], Gwenneg Kerdivel[8], Nicolas Buisine[8], Laurent M. Sachs[8], Daniel R. Buchholz[9], Taejoon Kwon[10], Heidi Smith-Parker[11], Marcos Gridi-Papp[12], Michael J. Ryan[11], Robert. D. Denton[13], John H. Malone[13], John B. Wallingford[14], Aaron F. Straight[2], Rebecca Heald[1], Dirk Hockemeyer[1,3,16], Richard M. Harland[1], Daniel S. Rokhsar[1,3,15,16,*]

[1]Department of Molecular and Cell Biology, Weill Hall, University of California, Berkeley, CA, 94720, USA.
[2]Department of Biochemistry, Stanford University School of Medicine, 279 Campus Drive, Beckman Center 409, Stanford, CA, 94305-5307, USA.
[3]Innovative Genomics Institute, University of California, Berkeley, CA, 94720, USA.
[4]Computer Science Division, University of California Berkeley, 2626 Hearst Avenue, Berkeley, CA, 94720, USA.
[5]Genome Sequencing Center, HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA.
[6]Pediatric Genomics Discovery Program, Departments of Pediatrics and Genetics, Yale University School of Medicine, 333 Cedar Street, New Haven, CT, 06510, USA.
[7]Department of Organismic and Evolutionary Biology, and Museum of Comparative Zoology, Harvard University, Cambridge, MA, 02138, USA.
[8]Département Adaptation du Vivant, UMR 7221 CNRS, Muséum National d'histoire Naturelle, Paris, France.
[9]Department of Biological Sciences, University of Cincinnati, Cincinnati, OH, USA.
[10]Department of Biomedical Engineering, Ulsan National Institute of Science and Technology, Ulsan, 44919, Republic of Korea.
[11]Department of Integrative Biology, Patterson Labs, 2401 Speedway, University of Texas, Austin, TX, 78712, USA.
[12]Department of Biological Sciences, University of the Pacific, 3601 Pacific Avenue, Stockton, CA, 95211, USA.
[13]Department of Molecular and Cell Biology and Institute of Systems Genomics, University of Connecticut, 181 Auditorium Road, Unit 3197, Storrs, CT, 06269, USA.
[14]Department of Molecular Biosciences, Patterson Labs, 2401 Speedway, The University of Texas at Austin, Austin, TX, 78712, USA.
[15]Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa, 9040495, Japan.
[16]Chan-Zuckerberg BioHub, 499 Illinois Street, San Francisco, CA, 94158, USA.
[†]These authors contributed equally
[*]Corresponding authors

Abstract

Frogs are an ecologically diverse and phylogenetically ancient group of living amphibians that include important vertebrate cell and developmental model systems, notably the genus *Xenopus*. Here we report a high-quality reference genome sequence for the western clawed frog, *Xenopus tropicalis*, along with draft chromosome-scale sequences of three distantly related emerging model frog species, *Eleutherodactylus coqui*, *Engystomops pustulosus* and *Hymenochirus boettgeri*. Frog chromosomes have remained remarkably stable since the Mesozoic Era, with limited Robertsonian (i.e., centric) translocations and end-to-end fusions found among the smaller chromosomes. Conservation of synteny includes conservation of centromere locations, marked by centromeric tandem repeats associated with Cenp-a binding, surrounded by pericentromeric LINE/L1 elements. We explored chromosome structure across frogs, using a dense meiotic linkage map for *X. tropicalis* and chromatin conformation capture (HiC) data for all species. Abundant satellite repeats occupy the unusually long (~20 megabase) terminal regions of each chromosome that coincide with high rates of recombination. Both embryonic and differentiated cells show reproducible association of centromeric chromatin, and of telomeres, reflecting a Rabl configuration similar to the "bouquet" structure of meiotic cells. Our comparative analyses reveal 13 conserved ancestral anuran chromosomes from which contemporary frog genomes were constructed.

Main Text

Introduction

Amphibians are widely used models in developmental and cell biology[1–5], and their importance extends to the fields of infectious disease, ecology, pharmacology, environmental health, and biological diversity[6–10]. While the principal model systems belong to the genus *Xenopus* (notably the diploid western clawed frog *X. tropicalis* and the paleo-allotetraploid African clawed frog *X. laevis*, other amphibian models have increasingly been introduced due to their diverse developmental, cell biological, physiological, and behavioral adaptations[11–14].

While genome evolution has been extensively studied in mammals[15] and birds[16,17], the relative lack of phylogenetically diverse chromosome-scale frog genomes has limited the study of genome evolution in anuran amphibians. Here, we report a high-quality assembly for *X. tropicalis* and three new chromosome-scale genome assemblies for the direct-developing Puerto Rican coquí (*Eleutherodactylus coqui*), the túngara frog (*Engystomops pustulosus*), which is a model for vocalization, and the Zaire dwarf clawed frog (*Hymenochirus boettgeri*), which has an unusually small embryo and is a model for regulation of cell and body sizes. Genome assemblies are essential resources for further work to exploit the experimental possibilities of these diverse animals. The new high quality *X. tropicalis* genome upgrades previous draft assemblies[18,19] and our new genomes complement draft chromosome-scale sequences for the African clawed frog[20] (*Xenopus laevis*), the African bullfrog[21] (*Pyxicephalus adspersus*), the Leishan moustache toad[22] (*Leptobrachium leishanense*), the Ailao moustache toad[23] (*Leptobrachium* [*Vibrissaphora*] *ailaonicum*), and Asiatic toad[24] (*Bufo gargarizans*), as well as scaffold- and contig-scale assemblies for other species[25]. The rapidly increasing number

of chromosome-scale genome assemblies makes anurans ripe for comparative genomic and evolutionary analysis.

Chromosome number variation among frogs is limited[26]. Based on cytological[27,28] and sequence comparisons[18,20,26,29,30] most frogs have *n* ~ 10–12 pairs of chromosomes. The constancy of the frog karyotype is in contrast with the more dramatic variation seen across mammals[15,31], which as a group is considerably younger than frogs. The constancy of the frog karyotype parallels the static karyotypes of birds[16], although birds typically have nearly three times more chromosomes than frogs, including numerous microchromosomes (among frogs, only the basal *Ascaphus*[32] has microchromosomes). Despite the stable frog chromosome number, however, fusions, fissions, and other inter-chromosomal rearrangements do occur, and we can use comparisons among chromosome-scale genome sequences to (1) infer the ancestral chromosomal elements, (2) determine the rearrangements that have occurred along frog phylogeny, and (3) characterize the patterns of chromosomal change among frogs. These findings of conserved synteny among frogs are consistent with prior demonstrations of conservation between *Xenopus tropicalis* with other tetrapods, including human and chicken[18,33].

Since frog karyotypes are so highly conserved, *X. tropicalis* can be used as a model for studying chromosome structure, chromatin interaction, and recombination for the entire clade. Features that can be illuminated at the sequence level include the structure and organization of centromeres and nature of the unusually long subtelomeres relative to mammals (frog subtelomeres are ~20 megabases, compared with the mammalian subtelomeres that are typically shorter than a megabase). The extended subtelomeres of frogs form interacting chromatin structures in interphase nuclei that reflect three-dimensional intra-chromosome and inter-chromosome subtelomeric contacts, which are consistent with a Rabl configuration. As in other animals, subtelomeres of frogs have an elevated GC content and recombination rate.

Here we show that the unusually high enrichment of recombination in the subtelomeres likely reflects similar structural and functional properties in other vertebrates, though the quality of the assembly reveals that the length of subtelomeres, enrichment of transposon subsequences by unequal crossing over, and high recombination rates are considerably greater than in mammals. We use Cenp-a binding at satellites to confirm centromere identity and extend the predictive power of the repeat structures to centromeres of other frogs. We address the unusually high recombination rate in subtelomeric regions, correlating with the landscape of base composition and transposons. Over the 200 million years of evolution that we address here, centromeres have generally been stable, but the few karyotypic changes reveal the predominant Robertsonian translocations at centromeric regions; we also document the slow degeneration that occurs to inactivated centromeres and fused telomeres, changes that are obscured in animals with rapidly evolving karyotypes.

Results

New frog genome sequences

High-quality chromosome-scale genome assembly for *X. tropicalis*

To establish a high-quality chromosomal reference genome sequence to study the structure and organization of *Xenopus tropicalis* chromosomes and for comparisons with other frog genomes, we integrated multiple sequencing technologies, including Single-Molecule Real-Time long reads (SMRT sequencing; Pacific Biosciences), linked read sets (10x Genomics), short-read shotgun sequencing, *in vivo* chromatin conformation capture, and meiotic mapping, combined with previously generated dideoxy shotgun sequence (**Supplementary Data 1**, **Supplementary Figs. 1A–D** and **2**, and **Supplementary Notes 1** and **2**). New sequences were generated from 17th generation individuals from the same inbred Nigerian line that was used in the original

Sanger shotgun sequencing[33]. The completeness, protein coding capacity, repeat structure and sequence variation are discussed in supplementary information (**Supplementary Figs. 1–6** and **Supplementary Tables 1–4**) providing the basis for comparisons.

The new v10 reference assembly spans 1,448.4 Mb and is substantially more complete than the previous v9 sequence[18], assigning 219.2 Mb more sequence to chromosomes (**Supplementary Table 1**). The v10 assembly is also far more contiguous, with half of the sequence contained in 32 contigs longer than 14.6 Mb (vs. 71.0 kb in v9). The assembly captures 99.6% of known coding sequence (**Supplementary Table 2**, **Supplementary Note 2**). We found that the fragmented quality of earlier assemblies was due in part to the fact that 68.3 Mb (4.71%) of the genome was not sampled by the 8× redundant Sanger dideoxy whole-genome shotgun dataset[33] (**Supplementary Fig. 3**, **Supplementary Note 2**). These missing sequences, apparently due to non-uniformities in shotgun cloning and/or sequences (**Supplementary Fig. 1E**), are distributed across an estimated 140.5k blocks of mean size 485.7 bp (longest 50.0 kb) on the new reference assembly and capture an additional 6,774 protein-coding exons (**Supplementary Fig. 1F**–**G**). The enhanced contiguity of v10 is accounted for by the relatively uniform coverage of PacBio long-read sequences on the genome, as expected from other studies[34–37]. Most remaining gaps are in highly repetitive and satellite-rich centromeres and subtelomeric regions (see below) (**Supplementary Figs. 1H** and **3**).

Additional chromosome-scale frog genomes

To assess the evolution of chromosome structure across a diverse set of frogs, we generated chromosome-scale genome assemblies for three new emerging model species, including the Zaire dwarf clawed frog *Hymenochirus boettgeri* (a member of the family Pipidae along with *Xenopus* spp.) and two neobratrachians: the Puerto Rican coquí *Eleutherodactylus coqui* (family

Eleutherodactylidae) and the túngara frog *Engystomops pustulosus* (family Leptodactylidae). These chromosome-scale draft genomes were primarily assembled from short-read datasets and chromatin conformation capture (HiC) data (**Supplementary Data 1**, **Supplementary Table 5**, **Supplementary Note 3**). To further expand the scope of our comparisons, we also updated the assemblies of two recently published frog genomes: the African bullfrog *Pyxicephalus adspersus*[21] from the neobatrachian family Pyxicephalidae, and the Ailao moustache toad *Leptobrachium* (*Vibrissaphora*) *ailaonicum*[22] from the family Megophryidae (**Supplementary Fig. 7**, **Supplementary Note 3**). These species span the pipanuran clade, which comprises all extant frogs except for a small number of phylogenetically basal taxa, such as *Ascaphus*[38].

The chromosome numbers of the new assemblies agree with previously described karyotypes for *E. coqui*[39] (2*n* = 26) and *E. pustulosus*[40] (2*n* = 22). The literature for *H. boettgeri,* however, is more equivocal, with reports[41,42] of 2*n* = 20–24. The *n* = 9 chromosomes in our *H. boettgeri* assembly are consistent with our chromosome spreads (**Supplementary Fig. 7A**). The karyotype variability in the published literature and discrepancy with karyotypes of our *H. boettgeri* samples may be the result of cryptic sub-populations within this species, or segregating chromosome polymorphisms.

Protein-coding gene set for *X. tropicalis*

The improved *X. tropicalis* genome encodes an estimated 25,016 protein-coding genes (**Supplementary Table 3**), which we predicted by taking advantage of 8,580 full-length-insert *X. tropicalis* cDNAs from the “Mammalian” Gene Collection[43] (MGC), 1.27 million Sanger-sequenced expressed sequence tags[33] (ESTs), and 334.5 Gbp of RNA-seq data from an aggregate of 16 conditions and tissues[44,45] (**Supplementary Data 1**, **Supplementary Note 2**).

The predicted gene set is a notable improvement on previous annotations, both in completeness and in full-length gene-level accuracy, due in part to the more complete assembly (**Supplementary Fig. 1**, **Supplementary Table 2**, **Supplementary Note 2**). In particular, gaps in the earlier genome assemblies arising from cloning biases in the Sanger sequencing process and encompassing exons embedded in highly repetitive sequences have been filled by single molecule long reads (**Supplementary Figures 1** and **3**).

A measure of this completeness and the utility of the *X. tropicalis* genome is provided by comparing its gene set with those of vertebrate model systems with reference-quality genomes, including chicken[46], zebrafish[47], mouse[48] and human[49,50] (**Supplementary Fig. 4**). Notably, despite the closer phylogenetic relationship between birds and mammals, *X. tropicalis* shares more orthologous gene families (and mutual best hits) with human than does chicken, possibly because of the loss of genomic segments in the bird lineage[16,51] and/or residual incompleteness of the chicken reference sequence, due to the absence of several microchromosomes[46]. For example, of 13,009 vertebrate gene families with representation from at least four of the vertebrate reference species, only 341 are missing from *X. tropicalis* versus 1,110 from chicken (**Supplementary Fig. 4**). The current *X. tropicalis* genome assembly also resolves gene order and completeness of gene structures in the long subtelomeres that were missed in previous assemblies due to their highly repetitive nature (**Supplementary Fig. 1F–G**).

Protein-coding gene sets for other frogs

We annotated the new genomes of *E. coqui*, *E. pustulosus*, *H. boettgeri,* and *P. adspersus* using transcriptome data from these species (**Supplementary Data 1**) and peptide homology with *X. tropicalis* (**Supplementary Tables 6** and **7**). To include mustache toad in our cross-frog comparisons we adopted the published annotation of Li et al.[22] (**Supplementary Note 3**). We

found 14,412 orthologous groups across the five genera with OrthoVenn2[52], including genes found in at least four of the five frog genera represented (**Supplementary Fig. 7B**). As expected, due to its reference-quality genome and well-studied transcriptome, only 72 of these clusters were not represented in *X. tropicalis*; the other frog genomes each had between 575 and 712 of these genes missing (or mis-clustered), suggesting better than 95% completeness in the other species. For analyses of synteny, we further restricted our attention to 7,292 one-to-one gene orthologs that were present on chromosomes (as opposed to unlinked scaffolds) in the "core" genomes *X. tropicalis, H. boettgeri, E. coqui*, *E. pustulosus*, and *P. adspersus*. The total branch length in the pipanuran tree shown in **Fig. 1** (including both *X. laevis* subgenomes) is 2.58 substitutions per four-fold synonymous site.

Repetitive landscape

Centromeric and telomeric tandem repeats play a critical role in the stability of chromosome structure[53]. Nonetheless, other kinds of repeats also play a role in the preservation of these important chromosome landmarks[54,55]. The new *X. tropicalis* v10 assembly captures sequences from centromeres and distal sub-telomeres that were fragmented in the previous assemblies[18,33]. The percentage of the genome covered by transposable elements is slightly higher than previously reported[33] (36.82% vs. 34%) (**Supplementary Table 4**).

Insertional bias in the pericentromeric regions is observed for specific families of long interspersed elements (LINEs), including the relatively young Chicken Repeat 1 (CR1, ref.[56]) (3.14% of the genome) and the ancient L1 (1.06%) (**Fig. 2** and **Supplementary Fig. 5**). The *X. tropicalis* v10 assembly captures significantly more tandem repeats in the distal subtelomeric portions of the genome relative to earlier assemblies. An exhaustive search for tandem repeats using Tandem Repeat Finder[57] determined that 10.67% of the chromosomes is covered by

194 tandem arrays consisting of 5 or more monomeric units greater than 10 bp. Many tandem
195 repeat footprints are in gaps from previous assemblies[18,33] (**Supplementary Fig. 3**). Our new
196 hybrid genome assembly closed many gaps containing centromeric and subtelomeric tandem
197 repeats and captured numerous subtelomeric genes (**Supplementary Fig. 1**). The overall
198 repeat landscape derived from the *X. tropicalis* assembly is mirrored in the other frog
199 assemblies, with similar centromeric repeats, and lengthy subtelomeres, as discussed below
200 (**Supplementary Fig. 9**).
201

202 Genetic variation

203 Although the *X. tropicalis* reference genotype is highly inbred, it nevertheless retains 15 long
204 heterozygous blocks ranging in size from 1.34 to 74.6 Mb. This exceeds the expectations based
205 on 17 generations of brother-sister mating, suggesting residual heterozygosity could be
206 maintained by balancing selection. Within the heterozygous blocks we observe 3.0 single
207 nucleotide variants per kilobase. To begin to develop a catalog of segregating variation we
208 sequenced pools of frogs from the Nigerian and Ivory Coast B populations, which have been
209 previously analysed using SSLP markers[58]. From our light pool shotgun analysis we identified a
210 total of 6,546,379 SNPs. There were 2,482,703 variants in the Nigerian pool and 4,661,928 in
211 the Ivory Coast B pool, with 598,252 shared by both pools, pointing to substantial differentiation
212 between populations (**Supplementary Fig. 6**, **Supplementary Note 2**).
213

214 Evolutionary dynamics of frog chromosomes

215 Conserved synteny and ancestral chromosomes

216 Comparison of the chromosomal positions of orthologs across seven frog genomes reveals
217 extensive conservation of synteny and collinearity (**Fig. 1**, **Supplementary Fig. 8**). We identified

13 conserved pipanuran syntenic units that we denote A through M (**Methods**, **Supplementary Note 4**). Each unit likely represents an ancestral pipanuran chromosome, an observation consistent with the 2*n* = 26 ancestral karyotype inferred from cytogenetic comparisons across frogs[27,59]. Over 95% (6,952 of 7,292) of chromosomal one-to-one gene orthologs are maintained in the same unit across the five frog species, attesting to the stability of these chromosomal elements (**Fig. 1**). The conservation of gene content per element is comparable to the 95% ortholog maintenance in the Muller elements in *Drosophila* spp.[60]. Despite an over two-fold difference in total genome size across the sampled genomes, each ancestral pipanuran element accounts for a nearly constant proportion of the total genome size, gene count, and repeat count in each species, implying uniform expansions and contractions during the history of the clade (**Supplementary Fig. 7C**).

At least some of these pipanuran elements have a deeper ancestry within amphibians. Comparison with the genome of the axolotl, *Ambystoma mexicanum*—a member of the order Caudata (salamanders and newts), and ~292 million years divergent from pipanurans[61]—reveals conservation of multiple syntenic units (**Supplementary Fig. 8A**). For example, axolotl chromosomes 4, 6, 7, and 14 are in near 1:1 correspondence with pipanuran elements F, A, B and K, respectively, although small pieces of F and A can be found on axolotl 10, and parts of B can be found on axolotl 9 and 13. Other axolotl chromosomes are fusions of parts of two or more pipanuran elements. For example, axolotl chromosome 5 is a fusion of a portion of J with most of G; the remainder of G is fused with a portion of L on the q arm of axolotl 2. Further comparisons are needed to determine which of these rearrangements occurred on the axolotl vs. the stem pipanuran lineage. Genomes from the superfamilies Leiopelmatoidea and Alytoidea, which diverged prior to the radiation of pipanurans, will also be informative.

Chromosome evolution

Block rearrangements of the 13 ancestral elements dominate the evolutionary dynamics of pipanuran karyotypes (**Table 1**, **Fig. 1**). While element C has remained intact as a single chromosome across the group (except for internal inversions), all the other elements have experienced translocations during pipanuran evolution. During these translocations, the elements have remained intact with the exception of the breakage of elements A and M by reciprocal partial arm exchange observed in *P. adspersus* chromosomes 3 and 6.

To trace the evolutionary history of centromeres shown in **Fig. 1**, we inferred their positions using HiC contact map patterns as in *X. tropicalis* (where centromeres were also confirmed by analysis of Cenp-a binding as described below). In general, the pericentromeres of other pipanurans were characterized by the same repetitive element families found in *Xenopus*, further corroborating their identification. Overall, we found broad pericentromeric conservation among the species analyzed (**Figs. 1** and **3A**).

Robertsonian or centric translocations involving breaks and joins near centromeres account for several of the rare rearrangements (**Figs. 1** and **3B**). For example, element G clearly experienced a centric fission in the *E. coqui* lineage. Conversely, I and M underwent centric fusion in the *E. pustulosus* lineage. *E. coqui* has experienced the most intense rearrangement, including Robertsonian fissions of A and G, a Robertsonian fusion of I/K, and a significant series of Robertsonian rearrangements involving B, E, F, and H that resulted in Bprox/H, Bdist/Fdist, and E/Fprox (**Table 1**, **Supplementary Table 8**). (Mechanistically, these “fissions” and “fusions” likely occur by translocations; for a discussion see ref.[62].) Elements I and H form the two arms of a metacentric chromosome in pipids (**Fig. 3A**), and therefore the pipid ancestor, but are found as either independent acrocentric chromosomes (e.g., in *P. adspersus* and *L. ailaonicum*) or as arms of metacentrics formed by centric fusion with other elements (**Supplementary Table 8**).

We also observed end-to-end “fusions" of metacentric chromosomes, for example, the joining of D with K in *E. pustulosus*, and with element E in the common ancestor of pipids (*Hymenochirus* and *Xenopus*) (**Figs. 1** and **3C**). Since bicentric chromosomes are not stably propagated through mitosis, one of the two ancestral centromeres brought together by end-to-end fusion must be lost or inactivated, as shown in **Fig. 3C** for the ancient D-E fusion in pipids. We note that the D centromere persists in both end-to-end fusions involving D, suggesting that centromeres derived from different ancestral elements may be differentially susceptible to silencing.

Using the pericentromere and subtelomere repeat landscape as a proxy, we found several examples of end-to-end chromosome fusions in which residual subtelomeric signals are preserved near the presumptive junction (**Fig. 3**, **Supplementary Fig. 9**). These include the end-to-end fusion of *X. tropicalis*-like chromosomes 9 and 10 (elements L and M) to produce the Chr9_10 progenitor of *X. laevis* that is found in both the L and S subgenomes of this allotetraploid[20]. These *X. laevis* chromosomes display evidence of decaying subtelomeric signatures in the region surrounding the ancestral L-M fusion (**Fig. 1** and **Supplementary Fig. 9A**,**B**). Similarly, enrichment of subtelomerically associated repeats is observed in *H. boettgeri* chromosome 8_10 (**Supplementary Fig. 9C–E**) near the junction between the portions of the chromosome with M and J/K ancestry (the J/K fusion occurred near the base of pipids). In both cases, the centromere from element M (i.e., the centromere in *X. tropicalis* chromosome 9) is maintained after fusion. The inversion of the p-arm from Chr8S also has evidence of decaying sequence but the median is less than the median JC distance at the Chr9_10 fusion, suggesting that the fusion preceded the inversion.

Rate of karyotype change

The long-range and, in most cases, chromosome-scale collinearity (**Supplementary Fig. 8**, **Supplementary Table 9**) among the frog species we examined, despite a combined branch length of 1.05 billion years (**Supplementary Tables 10** and **11**), parallels the synteny observed in birds[63] and reptiles[64] but differs from the substantial chromosome variation found in mammals[15,31]. Maintenance of collinear blocks may reflect an intrinsically slow rate of rearrangement in frogs, perhaps a consequence of large regions devoid of recombination, or selection favoring retention of specific gene order and chromosome structure related to chromosomal functions. We inferred a total of 17 fission, fusion, translocation, and duplication events (excluding smaller intra-chromosome rearrangements) resulting in a karyotype change every 62 million years (**Fig. 1**). This rate is similar to the rate of one chromosome-number change every 70 to 90 million years as previously proposed for frogs and mammals[26,28] but still slower than karyotype change rates for most mammals[65] and many reptiles[66]. Of course, our rate calculation is based on only seven species, and the rate may vary depending on the species analyzed. Some frog taxa, such as *Eleutherodactylus* spp. (2*n* = 16–32) and *Pristimantis* spp.[39] (2*n* = 22–38), have had a higher rate of karyotype change. On the other hand, some species, such as *Leptobrachium ailaonicum*, *L. leishanense*[14], and *Rana temporaria*[102], may have had no significant inter-chromosome changes over the past 205 million years (**Fig. 1**). Nonetheless, this analysis of chromosome variation across the frog lineage suggests a remarkably slow rate of karyotype evolution.

Chromosome structure and conformation

The stasis of *Xenopus* chromosomes relative to other frogs (see below) allows us to examine the repetitive landscape of chromosomes that are not frequently rearranged by translocation, and may be approaching a structural equilibrium.

Centromeres, satellites, and pericentromeric repeats

Vertebrate centromeres are typically characterized by tandem families of centromeric satellites (e.g., the alpha satellites of humans) that bind to the centromeric histone H3 protein, Cenp-a, a centromere-specific variant of histone H3[53,67]. Cenp-a binding satellites have been described in *X. laevis*[68], and here we find distantly related *X. tropicalis* satellite sequences that also co-precipitate with Cenp-a. Thus, chromatin immunoprecipitation and sequencing (ChIP-seq) shows that Cenp-a binding coincides with the predictions of centromere positions derived from chromatin conformation analysis and repetitive content (**Supplementary Fig. 5A–C**, **Supplementary Tables 12** and **13**). Importantly, this concordance supports the prediction of centromere position for other species that we infer below. The Cenp-a bound-sequences are arrays of 205-bp monomers that share a mean sequence identity greater than 95% at the nucleotide level, with a specific segment of the repeating unit showing greatest variability (**Supplementary Fig. 10**). The *X. tropicalis* centromere sequence is different from centromeric-associated repeats found in *X. laevis*[68,69], suggesting the sequences evolve rapidly after speciation but are maintained within the species.

All metacentric pericentromeric regions of *X. tropicalis* chromosomes are enriched in retrotransposable repetitive elements (15 Mb regions shown in **Fig. 2**). In other vertebrate species and *Drosophila*, retrotransposable elements from the pericentromeric regions are involved in the recruitment of constitutive heterochromatin components[70,71]. Among the pericentromerically enriched repeats we identified specific families belonging to LTR retrotransposons (Ty3), non-LTR retrotransposons (CR1, Penelope, and L1), and DNA transposable elements (PIF-Harbinger and piggyBac families) (**Fig. 2**, **Supplementary Fig. 5**). CR1 (CR1-2_XT) is the most prevalent and it is among the youngest of all pericentromeric retrotransposons (mean Jukes-Cantor (JC) distance to consensus of 0.05). In contrast, L1 and Penelope types have a mean JC greater than 0.4 (**Supplementary Fig. 5**). The age of the

repeats, indirectly measured by the JC distance, suggests that pericentromeric retrotransposons have experienced different bursts of activity and tendency to insert near the centromere. Expression of active retrotransposons and random insertion can compromise chromosome stability, and because silencing of these is crucial, genomes develop mechanisms to rapidly silence them. Such insertions may be positively selected, and therefore amplified, to establish pericentromeric heterochromatin, but may be counter selected when they insert in gene rich chromosome arms.

Recombination and extended subtelomeres

Although meiotic recombination is distributed across chromosomes, it is enriched near the chromosome ends (**Supplementary Fig. 11A**). While in humans, meiotic recombination is suppressed close to centromeres and elevated near telomeres, recombination is still regularly distributed on chromosome arms[72]. Other groups, including birds and fish, experience most recombination events 5 Mb away from the telomeres and only modest recombination is observed outside those regions[73–76]. Binding events for the protein PRDM9, present in mouse, rat and human, mark recombination hotspots in the chromosome arms in these species[77]. Given that amphibians lack the *prdm9* gene[78], we analyzed the genomic features that colocalized in areas prone to recombination.

We studied the distribution of recombination along *X. tropicalis* chromosomes using a previously generated Nigerian-Ivory Coast F_2 cross[18] (**Supplementary Note 5**, **Supplementary Data 2**). Half of the observed recombination is concentrated in only 160 Mb (11.0% of the genome) and 90% of the observed recombination occurs in 540 Mb (37.3%). In contrast, the central regions of each chromosome are “cold”, with recombination rates below 0.5 cM/Mb (**Supplementary Fig. 11B**, **Supplementary Table 14**). Strikingly, we find that (sex-averaged) recombination is

370 concentrated within just 30 Mb of the ends of each chromosome and occurs only rarely
371 elsewhere (**Supplementary Fig. 11A**); the regions of the subtelomeres experiencing high
372 recombination are nearly 6-fold larger than in non-amphibian genomes[73,74]. These rates of
373 recombination were not previously determined, since the repeat-rich subtelomeres were absent
374 from the assemblies, and markers that happened to lie in those regions showed insufficient
375 linkage to be incorporated into the maps.
376
377 Due to the elevated recombination, and repeat structure discussed below, we defined the
378 extended sub-telomeres as the terminal 30 Mb of all metacentric chromosomes, and terminal 30
379 Mb excluding the 15 Mb surrounding the pericentromeric regions of acrocentric chromosomes
380 (Chr3, Chr8, and Chr10) (**Fig. 2**). The median recombination rate in the extended subtelomeres
381 (1.73 cM/Mb) is ten-fold higher than the median rate observed in the rest of the chromosome
382 arms (0.16 cM/Mb). The recombination rate in the 5-Mb region surrounding the centromeric
383 tandem repeats is even lower (0.04 cM/Mb). Since constitutive heterochromatin in
384 pericentromeric regions is known to repress recombination, this observation is expected
385 (reviewed in refs.[79,80]). However, the centromeres of acrocentric chromosomes lie within 30 Mb
386 of the telomere, which precludes the extended sub-telomeric associated repeats (**Fig. 2** and
387 **Supplementary Fig. 12A**).
388
389 We examined the relationship between rates of recombination against repetitive elements and
390 sequence motifs associated with recombination hotspots in other vertebrate species
391 (**Supplementary Fig. 13A**, **Supplementary Table 14**). Similar to chicken and zebra finch,
392 recombination is the highest in subtelomeres and positively correlates with GC content[73,76,81],
393 which is consistent with GC-biased gene conversion[82–84] in recombinogenic regions (median GC
394 = 42.5% in the 74 Mb in which half of the recombination occurs) vs. the non-recombinogenic
395 centers of chromosomes (median 38.8%). As in zebra finch (**Supplementary Fig. 14**),

396 recombination in *X. tropicalis* is strongly correlated with satellite repeats (Pearson correlation =
397 +0.68, R^2 = 0.457). The high density of satellite repeats (**Supplementary Table 15**) in highly
398 recombinogenic subtelomeric regions suggests that unequal crossing over during meiotic
399 recombination mediates tandem repeat expansions[85,86]. Notably in the extended subtelomeric
400 regions, tandem repeats are enriched in specific tetrameric sequences (TGGG, AGGG, and
401 ACAG) compared to non-tandem repeats (**Supplementary Fig. 13B**). In contrast, centromeric
402 tandem repeats are completely devoid of these short sequences.
403
404 Some of the tandem arrays enriched in the terminal 30 Mb from all chromosomes derive from
405 portions of transposable elements such as SINE/tRNA-V, LINE/CR1, DNA/Kolobok-2
406 (**Supplementary Fig. 12**, **Supplementary Table 16**). For example the minisatellite expansion
407 that arose from the family of SINE/tRNA-V present in the pipid lineage[87] amplified a 52-bp
408 portion of the 3'UTR-tail from the SINE/tRNA-V element in *Xenopus tropicalis* and other frog
409 species (**Supplementary Table 17**). Although intact SINE/tRNA-V elements are distributed
410 throughout the genome, the minisatellite fragment is only expanded in subtelomeric SINE/tRNA-
411 Vs, suggesting that recombination in subtelomeres has driven minisatellite expansion
412 (**Supplementary Figs. 12** and **15**). Interestingly, although the satellite expansions are similar in
413 *X. laevis* and *X. tropicalis*, they differ in other frogs, suggesting that different satellite expansions
414 can occur repeatedly during the maintenance of the long subtelomeric regions (see below).
415

416 ## Chromatin conformation correlates with cytogenetic features

417 To further refine our understanding of chromosome structure in *X. tropicalis*, we studied
418 chromatin conformation capture ("HiC") data from nucleated blood cells. These experiments link
419 short reads representing sequences in close three-dimensional proximity[88]. **Fig. 4** shows
420 mapped HiC read pairs for chromosomes 1 and 2, with different minimum mapping quality

thresholds above and below the diagonal (**Fig. 4** and **Supplementary Fig. 2**, **Supplementary Note 5**). We consistently observe a "wing" of intra-chromosome contacts transverse to the main diagonal, which (1) intersects the main diagonal near the cytogenetically defined, Cenp-a-binding centromere, and (2) indicates contacts between p and q arms (**Supplementary Figs. 2** and **16**). These observations imply that interphase chromosomes are "folded" at their centromeres, with contacts between distal arms. We also observe inter-chromosome contacts among centromeres of different chromosomes and between their telomeres (**Supplementary Fig. 10A**).

Taken together, these intra- and inter-chromosome contacts are consistent with a Rabl configuration of chromosomes[89,90] in *Xenopus* blood cells. This configuration is understood as a relict structure from the previous mitosis[91,92], in which the chromosomes have become elongated and telomeres clustered on the nuclear membrane. Associations between centromeres and between telomeres, first observed in salamander embryos[89], are also observed in other animals[93,94], fungi[95], and plants[96–98]. These findings suggest that remnants of this 'Rabl configuration'[89] may be a common feature of post-mitotic cells across a wide range of eukaryotes. Here, we quantified the degree to which chromosomes are compacted in the Rabl configuration using HiC data and find that, with the exception of blood cell nuclei (sum of squared distances (SSD) 1.465), chromosomes from early frog development (NF stages 8 to 23) appear more tightly constrained (mean SSD 1.384) in Rabl configuration than the more specialized (liver and brain) adult tissues and sperm (mean SSD 5.583; **Supplementary Fig. 16**, **Supplementary Table 18**, **Supplementary Note 5**). Although it is possible some amount of HiC signal may be due to residual incompleteness in the assembly and concomitant mismapping of reads to repeat sequences, these observations are robust to quality filtering, even when using single-copy sequences. Furthermore, such contacts are weakest in sperm[16], a

control that argues strongly against sequence mismapping artefacts (**Supplementary Fig. 10B**, **Supplementary Note 5**).

We also observed three-dimensional associations between pericentromeric regions of different chromosomes, based on enriched HiC contacts[90,99] (**Fig. 4**). As with the Rabl signal, these "contacts" are accentuated when HiC reads are allowed to map permissively (**Methods**), which suggests that they may be influenced by common repetitive pericentromeric sequences shared among chromosomes. The signal persists in weaker form with more stringent read mapping, however, and either represents *bona fide* signal or residual incompleteness of the pericentromeric assembly. Notably, the correlation between centromere position and the observed intra-chromosome folding and inter-chromosome contacts at centromeres allows us to use HiC analysis and principal component analysis (PCA) of intra- and inter-chromosome contacts[97] to infer the likely centromeric positions based purely on HiC data in frogs whose cytogenetics are less well-studied (see below).

Chromatin compartments

Chromatin contacts in human[88,100,101], mouse[101], chicken[102] and other phylogenetically diverse species[103–105] often show a characteristic checkerboard pattern that is superimposed on the predominant near-diagonal signal. This pattern implies an alternating 'A/B' compartment structure with enriched intra-compartment contacts within chromosomes (**Fig. 5A**), which has been linked with G-banding in humans[106]. *X. tropicalis* also exhibits an A/B compartment pattern, which emerges as alternating gene rich ('A') and gene poor ('B') regions (median 19.99 genes/Mb in A and 9.99 genes/Mb in B) (**Fig. 5B**). A/B compartments are also differentiated by repetitive content[101], with A-compartment domains showing slight enrichment (1.21–1.44 fold) in DNA transposons of the DNA/Kolobok-T2, DNA/hAT-Charlie, and Mariner-Tc1 families. B-

compartment domains had significantly higher enrichment for DNA transposons (DNA/hAT-Ac, Mar-Tigger) and retrotransposons (Ty3/metaviridae and CR1), among other repeats (1.12–2.11 fold) (**Fig. 5C**). The association between repeats overrepresented in A and B compartments is also captured in one of the principal components obtained from the repeat densities of all chromosomes (**Supplementary Note 5**); we detect a modest negative correlation (R = −0.44) between the HiC eigenvectors that classified A/B compartments and the third principal component eigenvectors obtained from the repeat density matrix (**Supplementary Fig. 5B**). The association between chromatin condensation and repeat type could be due to a preference for certain transposable elements to insert in specific chromatin contexts, or chromatin condensation to be controlled in part by transposable element content, or a combination of these factors. However, we were unable to find any correlation of AB compartments with the G-banding of condensed chromosomes in *X. tropicalis*[107,108].

Higher-order interactions

Chromatin conformation contacts also provide clues to the organization of chromosomes within the nucleus. We observe non-random (χ^2 (81, n = 24,987,749) = 3,049,787; $p < 2.2\times10^{-308}$) associations between chromosomes in blood cell nuclei (**Fig. 4B**, **Supplementary Tables 19** and **20**): (a) chromosome 1 is enriched for contacts with chromosomes 2–8 (mean 1.05× enrichment), and depleted of contacts with 9 and 10 (mean 0.89×); (b) among themselves, chromosomes 2–8 show differential contact enrichment or depletion; and (c) chromosomes 9 and 10 are enriched (1.17×) for contacts with one another but are depleted with respect to all other chromosomes. These observations suggest the presence of distinct chromosome territories[89,109–111], where chromosomes 2–8 localized more proximal to—and arrayed around—chromosome 1, with chromosomes 9 and 10 relatively sequestered from chromosome 1 (**Fig. 4C**). The contact enrichment between chromosomes 9 and 10 is particularly notable because

these short chromosomes (91.2 and 52.4 Mb, respectively) have become fused in the *X. laevis* lineage[112], which might have been enabled by their persistent nuclear proximity[113].

Between chromosomes, p-p and q-q arm interactions exhibit a small but significant enrichment (1.059× enrichment; χ^2 (1, n = 24,786,496) = 17,037; $p < 2.2\times10^{-308}$) over p-q arm contacts. This is a general feature of metacentric and sub-metacentric chromosomes also observed in other frog genomes (see below), except *E. coqui* (0.928× enrichment; χ^2 (1, n = 6,850,547) = 3,914; $p < 2.2\times10^{-308}$), the chromosomes of which are predominantly acrocentric. Finally, the p-arms of chromosomes 3, 4, 8, and 9 are enriched for contacts with chromosome 10, with the sub-metacentric chromosomes 3 and 8 showing the strongest enrichment (and a slight preference between p-arms). The q-arms of chromosomes 3 and 8, however, exhibit a slight enrichment for contacts with chromosomes 1, 2, 4, and 5. Taken together, these observations suggest colocalization of the p and q arms of chromosomes 3 and 8 in blood cell nuclei.

Conclusions

Anuran amphibians play a central role in biology, not simply as a globally distributed animal taxon, but also as key subjects for research in areas that range from ecology and evolution to cell and developmental biology. The genomic resources generated here will thus provide important tools for further studies. Given the crucial role of *X. tropicalis* for genomic analysis of development and regeneration[114,115], the improvements to our understanding of its genome reported here will provide a more finely grained view of biomedically important genetic and epigenetic mechanisms. This new genome is also important from the standpoint of evolutionary genomics, as comparisons between the genomes of *X. tropicalis* and *X. laevis* shed light on mechanisms of genome duplication[115]. The new genome described here for *H. boettgeri,* another pipid frog, is also significant in this regard, as it enables an interesting comparison of

Xenopus genomes to that of a closely related outgroup. Moreover, the genomes of *E. coqui* and *E. pustulosus* provide a foundation for future studies of the evolution of ontogenies and of their underlying developmental mechanisms, as *E. coqui* is a direct-developing frog with no tadpole stage[116] and *E. pustulosus,* a foam-nesting frog, is a model for studying mating calls and female mate choice[116]. In addition to their interesting life histories, both frogs display interesting patterns of gastrulation[117,118]. Finally, recent work has demonstrated the efficacy of genetic or genomic analysis for understanding the impact of chytrid fungus on various amphibian species[119]. A deeper and broader understanding of amphibian genomes will be useful in the context of the global decline of amphibian populations[120,121].

Online Methods

Genomic extraction and sequencing.

High molecular weight DNA was extracted from blood of an F_{17} *Xenopus tropicalis* Nigerian strain female (ref.[18]; **Supplementary Note 1**). Paired-end (PE) Illumina shotgun libraries were constructed by the QB3 Functional Genomics Laboratory (FGL) and sequenced on an Illumina HiSeq 2500 as 2×250 bp reads at the Vincent J. Coates Genomics Sequencing Lab (VCGSL) at the University of California, Berkeley. Single-molecule real-time (SMRT) continuous long-read (CLR) sequencing was performed at the HudsonAlpha Institute for Biotechnology on PacBio RSII machines with P6-C4 chemistry (**Supplementary Note 1**, **Supplementary Data 1**). 10x Genomics Chromium linked-read sequencing was carried out at HudsonAlpha on HiSeq XTen (**Supplementary Note 1**).

***Xenopus tropicalis* genome assembly and annotation.**

Chromium linked reads (10x Genomics) were assembled with Supernova[122] (v1.1.5). This assembly was used to seed the assembly of continuous long reads (PacBio) using DBG2OLC[123]

(commit 1f7e752). An independent PacBio-only assembly was constructed with Canu[124] (v1.6-132-gf9284f8). These two assemblies were combined, or metassembled, using MUMmer[125] (v3.23) and quickmerge[126] commit e4ea490 (**Supplementary Fig. 1A**, **Supplementary Note 2**). Residual haplotypic redundancy was identified and removed (**Supplementary Fig. 1B**, **Supplementary Note 2**). The non-redundant metassembly was scaffolded with Sanger paired-ends and BAC-ends[33] using SSPACE[127] v3 and HiC using 3D-DNA[94,128] commit 2796c3b, then manually curated in JuiceBox[129,130] v1.9.0 (**Supplementary Note 2**). The assembly was polished with Arrow[131], Pilon[132] v1.23, and then FreeBayes[133] (v1.1.0-54-g49413aa) with ILEC (map4cns; https://bitbucket.org/rokhsar-lab/map4cns). The genome was annotated with the DOE-JGI Integrated Gene Call pipeline[134] (IGC) using transcript assemblies (TAs) generated with Trinity[135,136] v2.5.1 from multiple developmental stages and tissues (**Supplementary Data 1**, **Supplementary Note 2**). RepeatModeler[137] v1.0.11 was run on all frog species. The frog and ancestral repeat libraries from RepBase[138] v23.12 were combined with the repeats consensuses identified by Repeat Modeler. The merged repeat library was used to annotate repeats of all frogs with RepeatMasker[139] v4.0.7 (**Supplementary Notes 2** and **3**).

***Hymenochirus boettgeri* metaphase chromosome spread.**

Stage 26 tadpoles (*n* = 10) were incubated at room temperature in 0.01% colchicine and 1× MMR for 4–6 hr. After removing the yolky ventral portion of the tadpoles, the remaining dorsal portions were pooled together in deionized water and allowed to stand for 20 min. The dorsal portions were transferred to 0.2 mL of 60% acetic acid in deionized water and allowed to stand for 5 min. The tissue was then pipetted onto a positively charged microscope slide, and excess acetic acid was blotted away. To flatten the tissue and promote chromosome spreading, the slide was covered with a coverslip and a lead brick was placed on top of it for 5 min. The slide and coverslip were then placed on dry ice for 5 min. The coverslip was removed from the frozen slide, and the slide was stained with 0.1 mg/mL Hoechst Stain solution for 5 min. A fresh

coverslip was then mounted on the slide using VectaShield, and the edges were sealed with nail polish. Chromosomes in metaphase spreads (**Supplementary Fig. 7A**) were imaged on an Olympus BX51 Fluorescence Microscope run with Metamorph software using a 60× oil objective. Chromosome number was counted in 75 separate metaphase spreads.

Genome and transcriptome sequencing of five pipanurans.

Illumina PE 10x Genomics Chromium linked-read whole-genome libraries for *E. pustulosus* (from liver), *E. coqui* (from blood) and *H. boettgeri* (from liver) were sequenced on an HiSeq X at the HudsonAlpha Institute for Biotechnology. PacBio SMRT Sequel I CLR data were generated at UC Davis DNA Technologies and Expression Analysis Core for each of *E. pustulosus* and *H. boettgeri* from liver samples. In addition, two TruSeq Illumina PE libraries (from kidney) and two Nextera mate-pair libs (from liver) for *E. coqui* were prepared. HiC chromatin conformation capture libraries were prepared for *H. boettgeri*, *E. pustulosus*, and *E. coqui* using the Dovetail™ HiC Kit for Illumina following the “Animal Tissue Samples” protocol. HiC libraries were sequenced on the Illumina HiSeq 4000 by the VCGSL.

Illumina TruSeq Stranded mRNA Library Prep Kit libraries were prepared from *E. pustulosus* stages 45 and 56 whole tadpoles (gut excluded) and various adult tissues dissected from frogs maintained at the University of the Pacific. Brain (n = 3), dorsal skin (n = 2), eggs (n = 2), eye (n = 2), heart (n = 2), intestine (n = 2), larynx (n = 3), liver (n = 2), lung (n = 2), and ventral skin (n = 2) samples were washed twice with PBS, homogenized in TRIzol Reagent, and centrifuged, followed by flash freezing of the supernatant. RNA was isolated following the *TRIzol Reagent User Guide* (Pub. No. MAN0001271 Rev. A.0) protocol. In addition, *H. boettgeri* eggs were homogenized in TRIzol Reagent and processed according to manufacturer's instructions. RNA was then isolated using the QIAGEN RNeasy Mini Kit (cat 74104). An Illumina mRNA library was prepared using the Takara PrepX RNA-Seq for Illumina Library Kit by the Functional

598 Genomics Laboratory at the University of California Berkeley. All libraries were sequenced at
599 the VCGSL on an HiSeq 4000 as 151 bp PE reads. See **Supplementary Note 3** for additional
600 details about DNA/RNA extractions and library preparations, and **Supplementary Data 1** for a
601 complete list of DNA/RNA sequencing data generated for *E. coqui*, *E. pustulosus*, and *H.*
602 *boettgeri*.
603
604 **Assembly and annotation of five pipanuran genomes.**
605 Contigs were assembled with Supernova[122] v2.0.1 (*E. pustulosus* and *H. boettgeri*) or
606 Meraculous[140,141] v2.2.4 (*E. coqui*). For *E. coqui*, residual haplotypic redundancy was removed
607 using custom scripts (https://github.com/abmudd/Assembly) prior to scaffolding with SSPACE[127]
608 (v3.0). *E. pustulosus* and *H. boettgeri* contigs were ordered and oriented using MUMmer[125]
609 (v3.23) alignments to PBEC-polished (map4cns commit dd89f52; https://bitbucket.org/rokhsar-
610 lab/map4cns) DBG2OLC[123] (commit 1f7e752) hybrid contigs (**Supplementary Note 3**). All three
611 assemblies were scaffolded further with linked-reads and Scaff10X (v2.1;
612 https://sourceforge.net/projects/phusion2/files/scaff10x).
613
614 *E. pustulosus* and *H. boettgeri* chromosome-scale scaffolds were constructed with Dovetail
615 Genomics HiC via the HiRise scaffolder[142], followed by manual curation in JuiceBox[128–130]
616 v1.9.0. Due to the fragmented nature of the *E. coqui* assembly, initial chromosome-scale
617 scaffolds were first constructed by synteny with *E. pustulosus*, then refined in JuiceBox[128–130]
618 v1.9.0. Gaps in the *E. pustulosus* and *H. boettgeri* assemblies bridged by PacBio reads were
619 resized using custom scripts (pbGapLen; https://bitbucket.org/bredeson/artisanal) and filled with
620 PBJelly[143] (PBSuite v15.8.24). These two assemblies were polished with FreeBayes and ILEC
621 (map4cns commit dd89f52; https://bitbucket.org/rokhsar-lab/map4cns). A final round of gap-
622 filling was then performed on the three assemblies using Platanus[144] (v1.2.1).
623

Previously published *L. ailaonicum*[23] (GCA_018994145.1) *and P. adspersus*[21] (GCA_004786255.1) assemblies were manually corrected in JuiceBox[128–130] (v1.11.08) using their respective HiC and Chicago data (**Supplementary Data 1**). Gaps in the corrected *P. adspersus* scaffolds were resized with PacBio reads (as described above) and filled using Platanus[144] (v1.2.1) with published TruSeq Illumina data obtained from NCBI (PRJNA439445). As described elsewhere[145], all assemblies were screened for contaminants prior to scaffolding, and only final scaffolds and contigs longer than 1 kb were retained for downstream analyses. More details on assembly procedures can be found in (**Supplementary Note 3**).

Genomic repeats in all five species were annotated with RepeatMasker[137,139] (v4.0.7 and v4.0.9) using the repeat library generated above. Protein-coding genes were annotated for *E. coqui*, *E. pustulosus*, *H. boettgeri*, and *P. adspersus* using the IGC[134] pipeline with homology and transcript evidence. For each respective species, newly generated RNA-seq data were combined with public *H. boettgeri*[20] (BioProject PRJNA306175) and *P. adspersus*[21] (BioProject PRJNA439445) data, and unpublished *E. coqui* data (stages 7, 10, and 13 hindlimb [Harvard University]; stage 9–10 tail fin skin [French National Center for Scientific Research]). Transcript assemblies used as input to IGC were assembled with Trinity[135,136] (v2.5.1) and filtered using the heuristics described in **Supplementary Note 3.**

Synteny and ancestral chromosome inference.

One-to-one gene ortholog set between frog proteomes was obtained from the output from OrthoVenn2[52] using an E-value of 1×10^{-5} and an inflation value of 1.5 (**Supplementary Note 4**). The assemblies of all frog species and axolotl were pairwise aligned against the *X. tropicalis* genome using cactus[146] (commit e4d0859) (**Supplementary Note 4**). Pairwise collinearity runs were merged into runs of collinearity with ROAST/MULTIZ[147] (v012109) using the phylogenetic topology from TimeTree[148] and sorted with last[149] (v979) (**Supplementary Note 4**).

Phylogeny and estimation of sequence divergence.

Fourfold degenerate bases of one-to-one orthologs were obtained and reformatted from the MAFFT alignment as described in Mudd et al.[145] (**Supplementary Note 4**). The maximum-likelihood phylogeny was obtained with RAxML[150] (v8.2.11) using the GTR+Gamma model of substitution with outgroup *Ambystoma mexicanum*. Divergence times were calculated with MEGA7[151] (v7.0.26) with the GTR+Gamma model of substitution using Reltime method[152].

Chromosome evolution.

Customized scripts[145] were used to extract pairwise alignments from the ROAST-merged MAF file and converted into runs of collinearity. The runs of collinearity were visualized with Circos[153] (v0.69-6) (**Supplementary Note 4**).

Centromeres, satellites, and pericentromeric repeats.

Tandem repeats were called using Tandem Repeat Finder[57] (trf genome.fa 2 5 7 80 10 50 2000 -l 6 -d -ngs*)*. To identify tandem repeats enriched in pericentromeric and subtelomeric regions, we extracted the monomer sequences of all tandem repeats overlapping the region of interest. A database of non-redundant monomers was created by making a dimer database. Dimers were clustered with BlastClust[154] v2.2.26 (-S 75 -p F -L 0.45 -b F -W 10). A non-redundant monomer database was created using the most common monomer size from each cluster. The non-redundant sequences were mapped to the genome with BLASTN[155] (-outfmt 6 -evalue 1e3). The enriched monomeric sequences in centromeres and subtelomeres were identified by selecting the highest normalized rations of tandem sequence footprints in the region of interest over the remaining portions of the genome. For more detail, see **Supplementary Note 5**.

675 **Genetic variation.**

676 Reads were aligned with BWA-MEM[156] (v0.7.17-r1188) and alignments were processed using
677 SAMtools[157] (v1.9-93-g0ca96a4), keeping only properly paired reads (samtools view -f3 -F3852)
678 for variant calling. Variants were called with FreeBayes[133] (v1.1.0-54-g49413aa; --standard-
679 filters --genotype-qualities --strict-vcf --report-monomorphic). Only bi-allelic SNPs with depth
680 within mode ± 1.78SDs were retained. An allele-balance filter [0.3–0.7] for heterozygous
681 genotypes was also applied. Segmental heterozygosity/homozygosity were estimated using
682 windows of 500 kb with 50-kb step using BEDtools[158] (v2.28.0) for pooled samples or snvrate[159]
683 (v2.0; https://bitbucket.org/rokhsar-lab/wgs-analysis). For more detail, see **Supplementary**
684 **Note 2**.
685
686 **GC-content, gene, and repeat landscape.**
687 GC-content percentages were calculated in 1-Mb bins sliding every 50 kb. Gene densities were
688 obtained using a window size of 250 kb sliding every 12.5 kb. The repeat density matrix for *X.*
689 *tropicalis* was obtained by counting base pairs per 1 Mb (sliding every 200 kb) covered by
690 repeat families and classes of repeats. The principal component analysis (PCA) was performed
691 on the density matrix composed of 7,253 1-Mb bins and 3,070 repeats (**Supplementary Note**
692 **5**). The first (PC1) and second (PC2) components were smoothed using a cubic spline method.
693
694 **Chromatin immunoprecipitation.**
695 *Xenopus tropicalis* XTN-6 cells (Gorbsky and Horb, unpublished) were grown in 70% calcium-
696 free L-15 (US Biologicals cat# L2101-02-50L), pH 7.2/10% Fetal Bovine Serum/Penicillin-
697 Streptomycin (Invitrogen cat# 15140-163) at RT. Native MNase ChIP-seq protocol performed as
698 described previously in Smith et al.[69]. Approximately 40 million cells were trypsinized and
699 collected; nuclei were isolated by dounce extraction and collected with a sucrose cushion.
700 Chromatin was digested to mononucleosomes by MNase. Nuclei were lysed and soluble

nucleosomes were extracted overnight at 4 °C. Extracted mononucleosomes were precleared with Protein A dynabeads (Invitrogen cat# 100-02D) for at least 4 h at 4 °C. A sample was taken for input after preclearing. Protein A dynabeads were bound to 10- μg antibody (either Rb-anti-Xl Cenp-a, cross-reactive with *X. tropicalis*, Rb-anti-H4 Abcam cat# 7311 or Rb-anti-H3 Abcam cat# 1791) and incubated overnight with precleared soluble mononucleosomes at 4 °C. Dynabeads bound to Rabbit IgG antibody (Jackson ImmunoResearch cat#011-000-003) were collected with a magnet and washed three times with TBST (0.1% Triton X-100) before elution with 0.1% SDS in TE and proteinase K incubation at 65 °C with shaking for at least 4 h. Isolated and input mononucleosomes were size-selected using Ampure beads (Beckman cat# A63880) and prepared for sequencing using the NEBNext ultra ii DNA library prep kit for Illumina (NEB cat# E7654). Three replicates were sequenced on an Illumina HiSeq 4000 lane 2×150 bp by the Stanford Functional Genomics Facility. PE reads were trimmed with Trimmomatic[160] v0.39 filtering for universal Illumina primers and for Nextera-PE indices. Processed PE reads were mapped with minimap2 (ref.[161]) v2.17-r941 against the unmasked genome reference. Samtools[157] v1.9 was used for sorting and indexing the alignment. Read counts (MapQ0) per 10-kb bin (non-overlapping) for all samples were calculated with multiBamSummary from deeptools[162] v3.3.0. Read counts were normalized by the total number of counts in the chromosomes per sample (**Supplementary Note 5**).

Recombination and extended subtelomeres.

The reads from the F_2 mapping population[18] were aligned to the v10 genome using BWA-MEM[156] (v0.7.17-r1188). Variants were called using FreeBayes[133] (v1.1.0-54-g49413aa; --standard-filters --genotype-qualities --strict-vcf). SNPs were filtered, and valid F_2 mapping sites were selected when the genotypes of the Nigerian F0 and the ICB F_0 were fixed and different and there was a depth of at least 10 for each F_0 SNP. Maps were calculated using JoinMap[163] v4.1 (**Supplementary Note 5**, **Supplementary Data 2**). The variation on the linkage map was

727 smoothed using the cubic spline function calculated every 500 kb. The Pearson correlation
728 coefficient was calculated between recombination rates and genomic features that include GC
729 content, repeat densities, and densities of reported CTCF and recombination hotspots[164,165].
730
731 **Chromatin conformations and higher-order interactions.**
732 HiC read pairs were mapped with Juicer[128] (v1.5.6) and observed counts were extracted at 1 Mb
733 resolution with JuicerTools. Centromeres were estimated manually in JuiceBox[129] and refined
734 with Centurion[99] v0.1.0-3-g985439c using ICE-balanced MapQ0 matrices
735 (https://bitbucket.org/rokhsar-lab/xentr10/src/master/hic). Rabl chromatin structure was visualized
736 with PCA from Knight-Ruiz[166]-balanced MapQ30 matrices, and significance estimated by
737 permutation testing using custom R scripts. Rabl constraint between p- and q-arms was
738 measured as the sum of square distances (SSD) in PC1-PC2 dimensions, calculated between
739 non-overlapping bins traveling sequentially away from the centromere. Inter-/intra-chromosomal
740 contact enrichment analyses were quantified from MapQ30 matrices using χ^2 tests in R[167] v3.5.0
741 (**Supplementary Note 5**; https://bitbucket.org/rokhsar-lab/xentr10/src/master/hic).
742
743 **A/B compartments.**
744 A/B compartments were called with custom R[167] (v3.5.0) scripts from Knight-Ruiz-balanced
745 (observed / expected normalized) MapQ30 HiC contact correlation matrices generated with
746 Juicer[128] (**Supplementary Note 5**). Pearson's correlation between eigenvectors of PC1 from the
747 HiC correlation matrix and gene density were used to designate A and B compartments per
748 chromosome.
749

Data Availability

The assemblies, annotations, and raw data are deposited in NCBI for v10 *X. tropicalis* (BioProjects PRJNA577946 and PRJNA726269), *E. coqui* (BioProject PRJNA578591), *E. pustulosus* (BioProject PRJNA578590), *H. boettgeri* (BioProject PRJNA578589), *L. ailaonicum* (BioProject PRJNA578588), and *P. adspersus* (BioProject PRJNA578592).

Code Availability

All custom scripts used in this work can be found at https://bitbucket.org/rokhsar-lab/xentr10 and https://github.com/abmudd/Assembly.

References

1. Cannatella, D. C. & de Sá, R. O. *Xenopus laevis* as a model organism. *Syst. Biol.* **42**, 476–507 (1993).
2. Beetschen, J. C. How did urodele embryos come into prominence as a model system? *Int. J. Dev. Biol.* **40**, 629–636 (1996).
3. Brown, D. D. A tribute to the *Xenopus laevis* oocyte and egg. *J. Biol. Chem.* **279**, 45291–45299 (2004).
4. Harland, R. M. & Grainger, R. M. *Xenopus* research: metamorphosed by genetics and genomics. *Trends Genet.* **27**, 507–515 (2011).
5. Gurdon, J. B. & Hopwood, N. The introduction of *Xenopus laevis* into developmental biology: of empire, pregnancy testing and ribosomal genes. *Int. J. Dev. Biol.* **44**, 43–50 (2000).
6. Blaustein, A. R. & Dobson, A. A message from the frogs. *Nature* **439**, 143–144 (2006).

7. Farrer, R. A. *et al.* Multiple emergences of genetically diverse amphibian-infecting chytrids include a globalized hypervirulent recombinant lineage. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 18732–18736 (2011).
8. Whiles, M. R. *et al.* Disease-driven amphibian declines alter ecosystem processes in a tropical stream. *Ecosystems* **16**, 146–157 (2013).
9. Gomes, A. *et al.* Bioactive molecules from amphibian skin: their biological activities with reference to therapeutic potentials for possible drug development. *Indian J. Exp. Biol.* **45**, 579–593 (2007).
10. McCallum, M. L. Amphibian decline or extinction? Current declines dwarf background extinction rate. *hpet* **41**, 483–491 (2007).
11. Ryan, M. J., Fox, J. H., Wilczynski, W. & Rand, A. S. Sexual selection for sensory exploitation in the frog *Physalaemus pustulosus*. *Nature* **343**, 66–67 (1990).
12. Romero-Carvajal, A. *et al.* Embryogenesis and laboratory maintenance of the foam-nesting túngara frogs, genus *Engystomops* (= *Physalaemus*). *Dev. Dyn.* **238**, 1444–1454 (2009).
13. Miller, K. E., Session, A. M. & Heald, R. Kif2a scales meiotic spindle size in *Hymenochirus boettgeri*. *Curr. Biol.* **29**, 3720–3727.e5 (2019).
14. Minsuk, S. B. & Keller, R. E. Surface mesoderm in *Xenopus*: a revision of the stage 10 fate map. *Dev. Genes Evol.* **207**, 389–401 (1997).
15. Ferguson-Smith, M. A. & Trifonov, V. Mammalian karyotype evolution. *Nat. Rev. Genet.* **8**, 950–962 (2007).
16. Zhang, G. *et al.* Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311–1320 (2014).
17. Kiazim, L. G. *et al.* Comparative mapping of the macrochromosomes of eight avian species provides further insight into their phylogenetic relationships and avian karyotype evolution. *Cells* **10**, (2021).
18. Mitros, T. *et al.* A chromosome-scale genome assembly and dense genetic map for

Xenopus tropicalis. *Dev. Biol.* **452**, 8–20 (2019).

19. Niu, L. *et al.* Three-dimensional folding dynamics of the *Xenopus tropicalis* genome. *Nat. Genet.* **53**, 1075–1087 (2021).

20. Session, A. M. *et al.* Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature* **538**, 336–343 (2016).

21. Denton, R. D., Kudra, R. S., Malcom, J. W., Du Preez, L. & Malone, J. H. The African Bullfrog (*Pyxicephalus adspersus*) genome unites the two ancestral ingredients for making vertebrate sex chromosomes. *Cold Spring Harbor Laboratory* 329847 (2018) doi:10.1101/329847.

22. Li, J. *et al.* Genomic and transcriptomic insights into molecular basis of sexually dimorphic nuptial spines in *Leptobrachium leishanense*. *Nat. Commun.* **10**, 5551 (2019).

23. Li, Y. *et al.* Chromosome-level assembly of the mustache toad genome using third-generation DNA sequencing and Hi-C analysis. *Gigascience* **8**, (2019).

24. Lu, B. *et al.* A large genome with chromosome-scale assembly sheds light on the evolutionary success of a true toad (*Bufo gargarizans*). *Mol. Ecol. Resour.* **21**, 1256–1273 (2021).

25. Sun, Y.-B., Zhang, Y. & Wang, K. Perspectives on studying molecular adaptations of amphibians in the genomic era. *Zool Res* **41**, 351–364 (2020).

26. Wilson, A. C., Sarich, V. M. & Maxson, L. R. The importance of gene rearrangement in evolution: evidence from studies on rates of chromosomal, protein, and anatomical evolution. *Proc. Natl. Acad. Sci. U. S. A.* **71**, 3028–3030 (1974).

27. Morescalchi, A. Evolution and karyology of the amphibians. *Boll. Zool.* **47**, 113–126 (1980).

28. Bush, G. L., Case, S. M., Wilson, A. C. & Patton, J. L. Rapid speciation and chromosomal evolution in mammals. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 3942–3946 (1977).

29. Nowoshilow, S. *et al.* The axolotl genome and the evolution of key tissue formation regulators. *Nature* **554**, 50–55 (2018).

30. Smith, J. J. *et al.* A chromosome-scale assembly of the axolotl genome. *Genome Res.* **29**, 317–324 (2019).

31. Deakin, J. E., Graves, J. A. M. & Rens, W. The evolution of marsupial and monotreme chromosomes. *Cytogenet. Genome Res.* **137**, 113–129 (2012).

32. Bogart, J. P., Balon, E. K. & Bruton, M. N. The chromosomes of the living coelacanth and their remarkable similarity to those of one of the most ancient frogs. *J. Hered.* **85**, 322–325 (1994).

33. Hellsten, U. *et al.* The genome of the Western clawed frog *Xenopus tropicalis*. *Science* **328**, 633–636 (2010).

34. Carneiro, M. O. *et al.* Pacific biosciences sequencing technology for genotyping and variation discovery in human data. *BMC Genomics* **13**, 375 (2012).

35. Koren, S. *et al.* Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. *Nat. Biotechnol.* **30**, 693–700 (2012).

36. Quail, M. A. *et al.* A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**, 341 (2012).

37. Loomis, E. W. *et al.* Sequencing the unsequenceable: Expanded CGG-repeat alleles of the fragile X gene. *Genome Research* vol. 23 121–128 (2013).

38. Feng, Y.-J. *et al.* Phylogenomics reveals rapid, simultaneous diversification of three major clades of Gondwanan frogs at the Cretaceous-Paleogene boundary. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E5864–E5870 (2017).

39. Schmid, M. *et al.* The chromosomes of Terraranan frogs. Insights into vertebrate cytogenetics. *Cytogenetic and Genome Research* vols 130-131 1–14 (2010).

40. Rabello, M. N. Chromosomal studies in Brazilian anurans. *Caryologia* **23**, 45–59 (1970).

41. Scheel, J. J. The chromosomes of some African anuran species. in *Genetics and Mutagenesis of Fish* 113–116 (Springer Berlin Heidelberg, 1973). doi:10.1007/978-3-642-

65700-9_11.

42. Mezzasalma, M., Glaw, F., Odierna, G., Petraccioli, A. & Guarino, F. M. Karyological analyses of *Pseudhymenochirus merlini* and *Hymenochirus boettgeri* provide new insights into the chromosome evolution in the anuran family Pipidae. *Zoologischer Anzeiger - A Journal of Comparative Zoology* **258**, 47–53 (2015).
43. Temple, G. *et al.* The completion of the mammalian gene collection (MGC). *Genome Res.* **19**, 2324–2333 (2009).
44. Marin, R. *et al.* Convergent origination of a *Drosophila*-like dosage compensation mechanism in a reptile lineage. *Genome Research* vol. 27 1974–1987 (2017).
45. Owens, N. D. L. *et al.* Measuring absolute RNA copy numbers at high temporal resolution reveals transcriptome kinetics in development. *Cell Reports* vol. 14 632–647 (2016).
46. Warren, W. C. *et al.* A new chicken genome assembly provides insight into avian genome structure. *G3: Genes|Genomes|Genetics* vol. 7 109–117 (2017).
47. Howe, K. *et al.* The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498–503 (2013).
48. Mouse Genome Sequencing Consortium *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
49. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
50. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
51. Lovell, P. V. *et al.* Conserved syntenic clusters of protein coding genes are missing in birds. *Genome Biol.* **15**, 565 (2014).
52. Xu, L. *et al.* OrthoVenn2: A web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Research* vol. 47 W52–W58 (2019).
53. Hartley, G. & O'Neill, R. Centromere repeats: Hidden gems of the genome. *Genes* vol. 10

223 (2019).

54. Chueh, A. C., Wong, L. H., Wong, N. & Choo, K. H. A. Variable and hierarchical size distribution of L1-retroelement-enriched CENP-A clusters within a functional human neocentromere. *Hum. Mol. Genet.* **14**, 85–93 (2005).

55. Kuznetsova, I. S. *et al.* LINE-related component of mouse heterochromatin and complex chromocenters' composition. *Chromosome Res.* **24**, 309–323 (2016).

56. Suh, A. The specific requirements for CR1 retrotransposition explain the scarcity of retrogenes in birds. *J. Mol. Evol.* **81**, 18–20 (2015).

57. Benson, G. Tandem Repeats Finder: A program to analyze DNA sequences. *Nucleic Acids Research* vol. 27 573–580 (1999).

58. Igawa, T. *et al.* Inbreeding ratio and genetic relationships among strains of the Western clawed frog, *Xenopus tropicalis*. *PLoS One* **10**, e0133963 (2015).

59. Ford, L. S. & Cannatella, D. C. The major clades of frogs. *Herpetological Monographs.* **7**, 94–117 (1993).

60. Bhutkar, A. *et al.* Chromosomal rearrangement inferred from comparisons of 12 *Drosophila* genomes. *Genetics* **179**, 1657–1680 (2008).

61. Pyron, R. A. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. *Syst. Biol.* **60**, 466–481 (2011).

62. Schubert, I. & Lysak, M. A. Interpretation of karyotype evolution should consider chromosome structural constraints. *Trends Genet.* **27**, 207–216 (2011).

63. Griffin, D. K., Robertson, L. B. W., Tempest, H. G. & Skinner, B. M. The evolution of the avian genome as revealed by comparative molecular cytogenetics. *Cytogenet. Genome Res.* **117**, 64–77 (2007).

64. Deakin, J. E. & Ezaz, T. Understanding the evolution of reptile chromosomes through applications of combined cytogenetics and genomics approaches. *Cytogenet. Genome Res.* **157**, 7–20 (2019).

903 65. Maruyama, T. & Imai, H. T. Evolutionary rate of the mammalian karyotype. *J. Theor. Biol.*
904 **90**, 111–121 (1981).
905 66. Olmo, E. Rate of chromosome changes and speciation in reptiles. *Genetica* **125**, 185–203
906 (2005).
907 67. Jagannathan, M., Cummings, R. & Yamashita, Y. M. A conserved function for
908 pericentromeric satellite DNA. *Elife* **7**, (2018).
909 68. Edwards, N. S. & Murray, A. W. Identification of *Xenopus* CENP-A and an associated
910 centromeric DNA repeat. *Molecular Biology of the Cell* vol. 16 1800–1810 (2005).
911 69. Smith, O. K. *et al.* Identification and characterization of centromeric sequences in *Xenopus*
912 *laevis*. *Cold Spring Harbor Laboratory* 2020.06.23.167643 (2020)
913 doi:10.1101/2020.06.23.167643.
914 70. Penke, T. J. R., McKay, D. J., Strahl, B. D., Matera, A. G. & Duronio, R. J. Direct
915 interrogation of the role of H3K9 in metazoan heterochromatin function. *Genes Dev.* **30**,
916 1866–1880 (2016).
917 71. Di Giacomo, M. *et al.* Multiple epigenetic mechanisms and the piRNA pathway enforce
918 LINE1 silencing during adult spermatogenesis. *Mol. Cell* **50**, 601–608 (2013).
919 72. Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nat. Genet.* **31**,
920 241–247 (2002).
921 73. Backstrom, N. *et al.* The recombination landscape of the zebra finch *Taeniopygia guttata*
922 genome. *Genome Research* vol. 20 485–495 (2010).
923 74. Dréau, A., Venu, V., Avdievich, E., Gaspar, L. & Jones, F. C. Genome-wide recombination
924 map construction from single individuals using linked-read sequencing. *Nat. Commun.* **10**,
925 4309 (2019).
926 75. Shanfelter, A. F., Archambeault, S. L. & White, M. A. Divergent fine-scale recombination
927 landscapes between a freshwater and marine population of threespine stickleback fish.
928 *Genome Biol. Evol.* **11**, 1573–1585 (2019).

929 76. Singhal, S. *et al.* Stable recombination hotspots in birds. *Science* **350**, 928–932 (2015).
930 77. Jensen-Seaman, M. I. *et al.* Comparative recombination rates in the rat, mouse, and
931 human genomes. *Genome Res.* **14**, 528–538 (2004).
932 78. Baker, Z. *et al.* Repeated losses of PRDM9-directed recombination despite the
933 conservation of PRDM9 across vertebrates. *Elife* **6**, (2017).
934 79. Kuhl, L.-M. & Vader, G. Kinetochores, cohesin, and DNA breaks: Controlling meiotic
935 recombination within pericentromeres. *Yeast* **36**, 121–127 (2019).
936 80. Termolino, P., Cremona, G., Consiglio, M. F. & Conicella, C. Insights into epigenetic
937 landscape of recombination-free regions. *Chromosoma* vol. 125 301–308 (2016).
938 81. Groenen, M. A. M. *et al.* A high-density SNP-based linkage map of the chicken genome
939 reveals sequence features correlated with recombination rate. *Genome Res.* **19**, 510–519
940 (2009).
941 82. Duret, L. & Galtier, N. Biased gene conversion and the evolution of mammalian genomic
942 landscapes. *Annu. Rev. Genomics Hum. Genet.* **10**, 285–311 (2009).
943 83. Galtier, N., Piganeau, G., Mouchiroud, D. & Duret, L. GC-content evolution in mammalian
944 genomes: the biased gene conversion hypothesis. *Genetics* **159**, 907–911 (2001).
945 84. Meunier, J. & Duret, L. Recombination drives the evolution of GC-content in the human
946 genome. *Mol. Biol. Evol.* **21**, 984–990 (2004).
947 85. Lam, B. S. & Carroll, D. Tandemly repeated DNA sequences from *Xenopus laevis*. I.
948 Studies on sequence organization and variation in satellite 1 DNA (741 base-pair repeat).
949 *J. Mol. Biol.* **165**, 567–585 (1983).
950 86. Cohen, S., Menut, S. & Méchali, M. Regulated formation of extrachromosomal circular
951 DNA molecules during development in *Xenopus laevis*. *Mol. Cell. Biol.* **19**, 6682–6689
952 (1999).
953 87. Ogiwara, I. V-SINEs: A new superfamily of vertebrate SINEs that are widespread in
954 vertebrate genomes and retain a strongly conserved segment within each repetitive unit.

Genome Research vol. 12 316–324 (2002).

88. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
89. Rabl, C. Über Zelltheilung. *Morphologisches Jahrbuch* 214–330 (1885).
90. Muller, H., Gil, J., Jr & Drinnenberg, I. A. The impact of centromeres on spatial genome architecture. *Trends Genet.* **35**, 565–578 (2019).
91. Therizols, P., Duong, T., Dujon, B., Zimmer, C. & Fabre, E. Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2025–2030 (2010).
92. Buttrick, G. J. *et al.* Nsk1 ensures accurate chromosome segregation by promoting association of kinetochores to spindle poles during anaphase B. *Mol. Biol. Cell* **22**, 4486–4502 (2011).
93. Stevens, T. J. *et al.* 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59–64 (2017).
94. Dudchenko, O. *et al. De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
95. Duan, Z. *et al.* A three-dimensional model of the yeast genome. *Nature* vol. 465 363–367 (2010).
96. Armstrong, S. J., Franklin, F. C. & Jones, G. H. Nucleolus-associated telomere clustering and pairing precede meiotic chromosome synapsis in *Arabidopsis thaliana*. *J. Cell Sci.* **114**, 4207–4217 (2001).
97. Mascher, M. *et al.* A chromosome conformation capture ordered sequence of the barley genome. *Nature* **544**, 427–433 (2017).
98. Cowan, C. R., Carlton, P. M. & Cande, W. Z. The polar arrangement of telomeres in interphase and meiosis. Rabl organization and the bouquet. *Plant Physiol.* **125**, 532–538 (2001).

981 99. Varoquaux, N. *et al.* Accurate identification of centromere locations in yeast genomes
982 using Hi-C. *Nucleic Acids Res.* **43**, 5331–5339 (2015).
983 100. Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat.*
984 *Rev. Genet.* **19**, 789–800 (2018).
985 101. Lu, J. Y. *et al.* Homotypic clustering of L1 and B1/Alu repeats compartmentalizes the 3D
986 genome. *Cell Res.* (2021) doi:10.1038/s41422-020-00466-6.
987 102. Fishman, V. *et al.* 3D organization of chicken genome demonstrates evolutionary
988 conservation of topologically associated domains and highlights unique architecture of
989 erythrocytes' chromatin. *Nucleic Acids Research* vol. 47 648–665 (2019).
990 103. Kaaij, L. J. T., van der Weide, R. H., Ketting, R. F. & de Wit, E. Systemic loss and gain of
991 chromatin architecture throughout zebrafish development. *Cell Rep.* **24**, 1–10.e4 (2018).
992 104. Eagen, K. P., Aiden, E. L. & Kornberg, R. D. Polycomb-mediated chromatin loops revealed
993 by a subkilobase-resolution chromatin interaction map. *Proc. Natl. Acad. Sci. U. S. A.* **114**,
994 8764–8769 (2017).
995 105. Dong, P. *et al.* 3D chromatin architecture of large plant genomes determined by local A/B
996 compartments. *Mol. Plant* **10**, 1497–1509 (2017).
997 106. Francke, U. 2012 William Allan Award: Adventures in cytogenetics. *Am. J. Hum. Genet.* **92**,
998 325–337 (2013).
999 107. Uno, Y. *et al.* Diversity in the origins of sex chromosomes in anurans inferred from
1000 comparative mapping of sexual differentiation genes for three species of the Raninae and
1001 Xenopodinae. *Chromosome Res.* **16**, 999–1011 (2008).
1002 108. Uno, Y. *et al.* Inference of the protokaryotypes of amniotes and tetrapods and the
1003 evolutionary processes of microchromosomes from comparative gene mapping. *PLoS One*
1004 **7**, e53027 (2012).
1005 109. Parada, L. A., McQueen, P. G., Munson, P. J. & Misteli, T. Conservation of relative
1006 chromosome positioning in normal and cancer cells. *Curr. Biol.* **12**, 1692–1697 (2002).

1007 110. Parada, L. A., McQueen, P. G. & Misteli, T. Tissue-specific spatial organization of
1008 genomes. *Genome Biol.* **5**, R44 (2004).
1009 111. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals
1010 folding principles of the human genome. *Science* **326**, 289–293 (2009).
1011 112. Uno, Y., Nishida, C., Takagi, C., Ueno, N. & Matsuda, Y. Homoeologous chromosomes of
1012 *Xenopus laevis* are highly conserved after whole-genome duplication. *Heredity* vol. 111
1013 430–436 (2013).
1014 113. Rosin, L. F. *et al.* Chromosome territory formation attenuates the translocation potential of
1015 cells. *Elife* **8**, (2019).
1016 114. Bright, A. R. *et al.* Combinatorial transcription factor activities on open chromatin induce
1017 embryonic heterogeneity in vertebrates. *EMBO J.* **40**, e104913 (2021).
1018 115. Kakebeen, A. D., Chitsazan, A. D., Williams, M. C., Saunders, L. M. & Wills, A. E.
1019 Chromatin accessibility dynamics and single cell RNA-Seq reveal new regulators of
1020 regeneration in neural progenitors. *Elife* **9**, (2020).
1021 116. Elinson, R. P. Metamorphosis in a frog that does not have a tadpole. *Curr. Top. Dev. Biol.*
1022 **103**, 259–276 (2013).
1023 117. del Pino, E. M. *et al.* A comparative analysis of frog early development. *Proc. Natl. Acad.*
1024 *Sci. U. S. A.* **104**, 11882–11888 (2007).
1025 118. Vargas, A. & Del Pino, E. M. Analysis of cell size in the gastrula of ten frog species reveals
1026 a correlation of egg with cell sizes, and a conserved pattern of small cells in the marginal
1027 zone. *J. Exp. Zool. B Mol. Dev. Evol.* **328**, 88–96 (2017).
1028 119. Oswald, P. *et al.* Locality, time and heterozygosity affect chytrid infection in yellow-bellied
1029 toads. *Dis. Aquat. Organ.* **142**, 225–237 (2020).
1030 120. Alford, R. A., Dixon, P. M. & Pechmann, J. H. Ecology. Global amphibian population
1031 declines. *Nature* vol. 412 499–500 (2001).
1032 121. Leung, B. *et al.* Clustered versus catastrophic global vertebrate declines. *Nature* **588**, 267–

1033 271 (2020).

1034 122. Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M. & Jaffe, D. B. Direct determination of
1035 diploid genome sequences. *Genome Res.* **27**, 757–767 (2017).

1036 123. Ye, C., Hill, C. M., Wu, S., Ruan, J. & Ma, Z. S. DBG2OLC: Efficient assembly of large
1037 genomes using long erroneous reads of the third generation sequencing technologies. *Sci.*
1038 *Rep.* **6**, 31900 (2016).

1039 124. Koren, S. *et al.* Canu: Scalable and accurate long-read assembly via adaptive k-mer
1040 weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).

1041 125. Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol.* **5**,
1042 R12 (2004).

1043 126. Chakraborty, M., Baldwin-Brown, J. G., Long, A. D. & Emerson, J. J. Contiguous and
1044 accurate *de novo* assembly of metazoan genomes with modest long read coverage.
1045 *Nucleic Acids Research* gkw654 (2016) doi:10.1093/nar/gkw654.

1046 127. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-
1047 assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579 (2011).

1048 128. Durand, N. C. *et al.* Juicer provides a one-click system for analyzing loop-resolution Hi-C
1049 experiments. *Cell Syst* **3**, 95–98 (2016).

1050 129. Durand, N. C. *et al.* Juicebox provides a visualization system for Hi-C contact maps with
1051 unlimited zoom. *Cell Systems* vol. 3 99–101 (2016).

1052 130. Dudchenko, O., Shamim, M. S., Batra, S. S. & Durand, N. C. The Juicebox Assembly
1053 Tools module facilitates *de novo* assembly of mammalian genomes with chromosome-
1054 length scaffolds for under $1000. *Biorxiv* (2018).

1055 131. Chin, C.-S. *et al.* Nonhybrid, finished microbial genome assemblies from long-read SMRT
1056 sequencing data. *Nat. Methods* **10**, 563–569 (2013).

1057 132. Walker, B. J. *et al.* Pilon: An integrated tool for comprehensive microbial variant detection
1058 and genome assembly improvement. *PLoS ONE* vol. 9 e112963 (2014).

133. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv [q-bio.GN]* (2012).

134. Shu, S., Rokhsar, D., Goodstein, D., Hayes, D. & Mitros, T. *JGI Plant Genomics Gene Annotation Pipeline*. https://www.osti.gov/biblio/1241222 (2014).

135. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).

136. Haas, B. J. *et al. De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).

137. Smit, A. F. A. & Hubley, R. *RepeatModeler Open-1.0*. (2008-2015).

138. Jurka, J. *et al.* Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005).

139. Smit, A. F. A., Hubley, R. & Green, P. *RepeatMasker Open-4.0*. (2013-2015).

140. Chapman, J. A. *et al.* Meraculous: *De novo* genome assembly with short paired-end reads. *PLoS One* **6**, e23501 (2011).

141. Goltsman, E., Ho, I. & Rokhsar, D. Meraculous-2D: Haplotype-sensitive assembly of highly heterozygous genomes. *arXiv [q-bio.GN]* (2017).

142. Putnam, N. H. *et al.* Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.* **26**, 342–350 (2016).

143. English, A. C. *et al.* Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One* **7**, e47768 (2012).

144. Kajitani, R. *et al.* Efficient *de novo* assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* **24**, 1384–1395 (2014).

145. Mudd, A. B., Bredeson, J. V., Baum, R., Hockemeyer, D. & Rokhsar, D. S. Analysis of muntjac deer genome and chromatin architecture reveals rapid karyotype evolution. *Communications Biology* vol. 3 (2020).

146. Paten, B. *et al.* Cactus: Algorithms for genome multiple sequence alignment. *Genome Res.*

21, 1512–1528 (2011).

147. Blanchette, M. *et al.* Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**, 708–715 (2004).

148. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: A resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).

149. Kiełbasa, S. M., Wan, R., Sato, K., Horton, P. & Frith, M. C. Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21**, 487–493 (2011).

150. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

151. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).

152. Tamura, K. *et al.* Estimating divergence times in large molecular phylogenies. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 19333–19338 (2012).

153. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).

154. Dondoshansky, I. & Wolf, Y. Blastclust (NCBI Software Development Toolkit). *ScienceOpen* https://www.scienceopen.com/document?vid=b654ab9a-231d-410a-832d-37c7c7bc7165 (2002).

155. Camacho, C. *et al.* BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

156. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* (2013).

157. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

158. Quinlan, A. R. BEDTools: The Swiss-army tool for genome feature analysis. *Curr. Protoc. Bioinformatics* **47**, 11.12.1–34 (2014).

159. Bredeson, J. V. *et al.* Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat. Biotechnol.* **34**, 562–570 (2016).

160. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* vol. 30 2114–2120 (2014).

161. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

162. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–5 (2016).

163. Van Ooijen, J. W. Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. *Genet. Res.* **93**, 343–349 (2011).

164. Myers, S., Bottolo, L., Freeman, C., McVean, G. & Donnelly, P. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**, 321–324 (2005).

165. Shifman, S. *et al.* A high-resolution single nucleotide polymorphism genetic map of the mouse genome. *PLoS Biol.* **4**, e395 (2006).

166. Knight, P. A. & Ruiz, D. A fast algorithm for matrix balancing. *IMA J. Numer. Anal.* **33**, 1029–1047 (2012).

167. R Core Team. R Core Team. R: A language and environment for statistical computing. *Foundation for Statistical Computing* (2013).

168. Tang, H. *et al.* Synteny and collinearity in plant genomes. *Science* **320**, 486–488 (2008).

Acknowledgements

We thank Karen Lundy and the Functional Genomics Laboratory at the University of California Berkeley for running quality control on extracted DNA and RNA and for preparing Illumina short-insert libraries; Oanh Nguyen and the DNA Technologies and Expression Analysis Cores at the University of California Davis Genome Center for preparing and sequencing PacBio libraries; Dovetail Genomics for providing the HiC library preparation kit, running quality control on HiC libraries, and preparing and sequencing HiC libraries; Shana McDevitt and the Vincent J. Coates Genomics Sequencing Laboratory at the University of California Berkeley for sequencing HiC and Illumina short-insert libraries; Shengqiang Shu for advice on the use of the IGC annotation pipeline. We thank Rick Elinson for providing *E. coqui* frogs and tissues. We thank Gary Gorbsky from the Oklahoma Medical Research Foundation and Marko Horb and the National *Xenopus* Resource at the MBL for providing the XTN-6 cell lines. Finally, we thank Chunhui Hou and colleagues for permission to access their HiC data prior to publication.

Funding

This study was supported by NIH grants R01HD080708 to D.S.R.; R01GM086321, R01HD065705 to D.S.R. and R.M.H.; R35GM127069 to R.M.H.; R35 GM118183 to R.B.H. A.B.M. was supported by NIH grants T32GM007127 and T32HG000047 and a David L. Boren Fellowship. D.S.R. is grateful for support from the Marthella Foskett Brown Chair in Biological Sciences; R.M.H., the C.H. Li Distinguished Chair in Molecular and Cell Biology; and R.H., the Flora Lamson Hewlett chair in biochemistry. A.F.S. and O.K.S. were supported by R01GM074728, O.K.S. by NIH T32 GM113854-02 and NSF GRFP; M.K.K. and M.L. by R01HD102186; J.H. by NSF grants DEB-1701591 and DBI-1702263; M.L., a Women in Science Fellowship; T.K. by the Basic Science Research Program, National Research Foundation of

Korea (NRF), Ministry of Education (2018R1A6A1A03025810), Future-leading Project Research Fund (1.200094.01) of UNIST and the Institute for Basic Science (IBS-R022-D1); J.B.W. and H.S.P. by R01GM104853, R01HD085901; M.J.R. by NSF IOS-0910112; Smithsonian Tropical Research Institute; Clark Hubbs Regents Professorship; L.M.S. by the “Centre National de la Recherche Scientifique” (PEPS ExoMod “Triton”) and the “Muséum National d’Histoire Naturelle” (Action Transversale du Muséum “Cycles biologiques: Evolution et adaptation”) and a Scientific council post-doctoral position to G.K.

This work used the Vincent J. Coates Genomics Sequencing Laboratory at the University of California Berkeley, supported by NIH grant S10OD018174, and the DNA Technologies and Expression Analysis Cores at the University of California Davis Genome Center, supported by NIH grant S10OD010786. This research used the National Energy Research Scientific Computing Center, a Department of Energy Office of Science User Facility supported by contract number DE-AC02-05CH11231. LMS acknowledges the “Ecole Normale Supérieure de PARIS” genomic platform for RNA-sequencing and the PCIA high performance computing platform at “Muséum National d’Histoire Naturelle”.

Author Information

Contributions

J.V.B., A.B.M., S.M-R., T.M., R.M.H., and D.S.R. wrote the manuscript with feedback from M.L., H.P.S., J.H., J.B.L., J.B.W., M.J.R., O.K.S., D.R.B., M.G-P., J.H., N.B., T.K., L.M.S., R.H., J.S., M.K.K., A.F.S., and D.H. Genomes were assembled by J.V.B., S.S.B. (*Xtr*); A.B.M., and K.C.B. (other frogs). S.M-R., A.B.M., and G.K. assembled transcripts and annotated genomes. S.M-R. and J.V.B. assessed gene completeness; S.M-R. analyzed repeat and recombination landscapes. S.M-R. and J.P. identified centromeric repeats. O.K.S., G.A-F. and A.F.S.

conducted ChIP-seq experiments and S.M-R. performed analysis. J.V.B. analyzed HiC features. T.M. constructed the linkage map. T.M. and J.V.B. analyzed heterozygosity. A.B.M. performed genome-wide comparisons. K.E.M. and R.H. examined *Hbo* metaphase spreads. M.K.K. and M.L. inbred *Xtr* frogs. R.M.H. (*Xtr*); M.G-P. (*Epu*); K.E.M. and R.H. (*Hbo*); M.L. and J.H. (*Eco*) collected frogs. R.M.H. (*Xtr*); M.G-P., H.S-P. (*Epu*); and D.R.B. (*Eco*) collected tissue samples. A.B.M., D.R.B. (*Eco*); J.B.L., and I.P. (*Xtr*) extracted DNA. A.B.M., S.M-R. (*Epu*); K.E.M., R.H. (*Hbo*); and L.M.S. (*Eco*) extracted RNA and libraries were prepared by A.B.M. (*Epu*). M.L., J.H. (*Eco*); K.E.M., and R.H. (*Hbo*) provided RNA-seq data. T.K., M.J.R., J.B.W. (*Epu*); and J.B.L. (*Xtr*) coordinated sequencing. C.P., J.G., and J.S. prepared and sequenced 10x Genomics, PacBio, and Illumina mate-pair libraries. D.H. prepared HiC libraries. R.D.D. and J.H.M. provided early access to the *Pad* assembly. N.B. (*Eco*) provided bioinformatic support. L.M.S. led the *Eco* efforts. R.M.H. and D.S.R. led the project.

Ethics Declarations

Competing Interests

D.S.R. is a member of the Scientific Advisory Board of, and a minor shareholder in, Dovetail Genomics LLC, which provides as a service the high-throughput chromatin conformation capture (HiC) technology used in this study.

M.K.K. is President and co-founder of Victory Genomics, Inc.

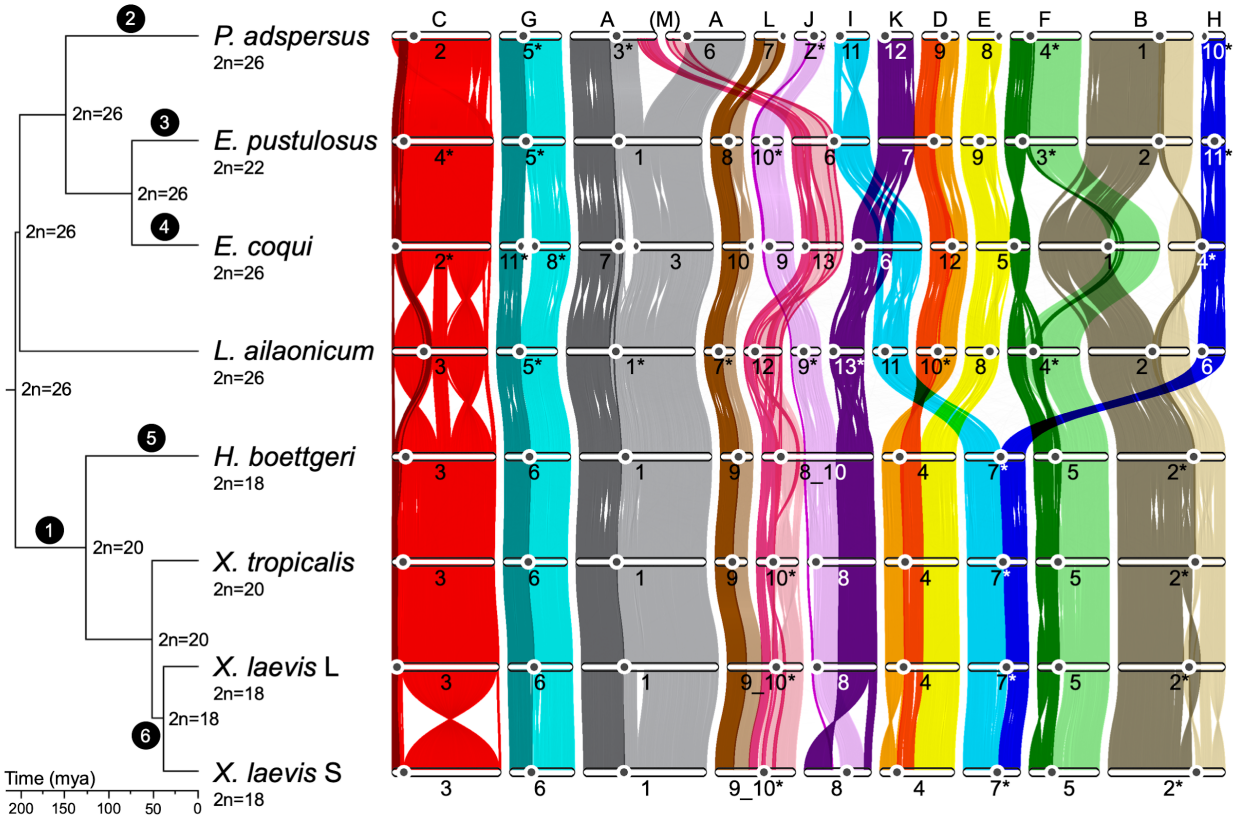

Fig. 1 Phylogenetic tree and gene ortholog alignment.

The phylogenetic tree of the seven analyzed species, calculated from fourfold degenerate sites and divergence time confidence intervals, was visualized with FigTree (commit 901211e; https://github.com/rambaut/figtree). The ancestral karyotype at each node was labeled on the tree. The alignment plot was generated with jcvi.graphics.karyotype[168] (v0.8.12; https://github.com/tanghaibao/jcvi) using the 7,292 described chromosome one-to-one gene orthologs from OrthoVenn2 (ref.[52]), followed by visual filtering of single stray orthologs. The pericentromeric region based on HiC inference was represented with a black circle on each chromosome. The ancestral chromosomes (A to M) were labeled at the top of the alignment based on the corresponding region in *P. adspersus*. The alignments for each ancestral chromosome were colored uniquely, with those upstream and downstream of the *X. tropicalis* centromeric satellite repeat from tandem repeat analysis shaded with a light versus dark shade of the ancestral chromosome color. Chromosomes labeled with an asterisk were reverse complemented in this image relative to the orientation in the assembly. Black circles with white text reference chromosome changes outlined in **Table 1**.

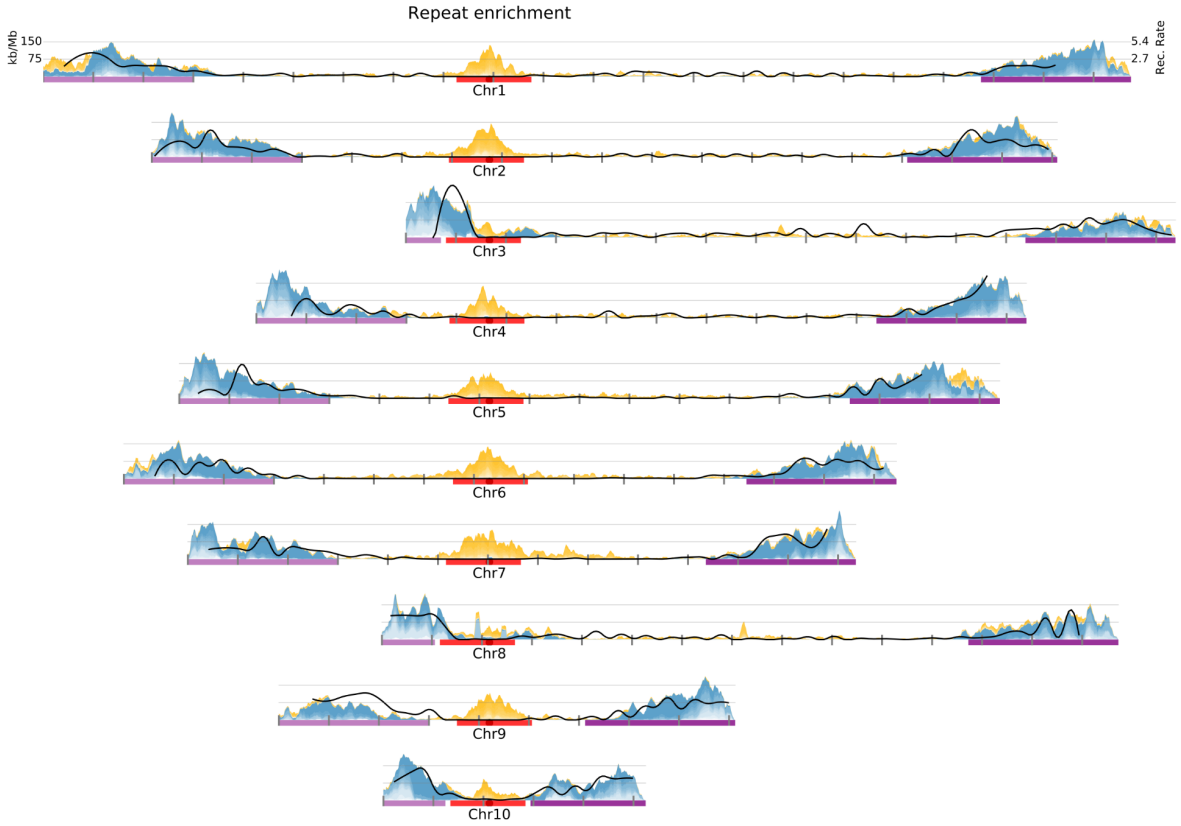

Fig. 2 Density of pericentromeric and subtelomeric repeats in *Xenopus tropicalis*.

Pericentromeric boundaries (red) and subtelomeric boundaries (purple) were used to obtain enriched repeats excluding chromosomes with short p-arms (chromosomes 3, 8, and 10). Pericentromeric repeats (yellow) correspond to selected subsets of non-LTR retrotransposons (CR1, L1, and Penelope), LTR retrotransposons (Ty3), and DNA transposons (PiggyBac and Harbinger). Subtelomeric enriched repeats (blue) correspond mainly to Satellite repeats and LTR retrotransposons (Ty3, Ngaro). Chromosomes are centered by the position of centromeric tandem repeats (black dot and dotted vertical line). The rate of recombination (cM/Mb) is shown as a solid black line. Tick marks indicate 10 Mb blocks (**Supplementary Fig. 16**).

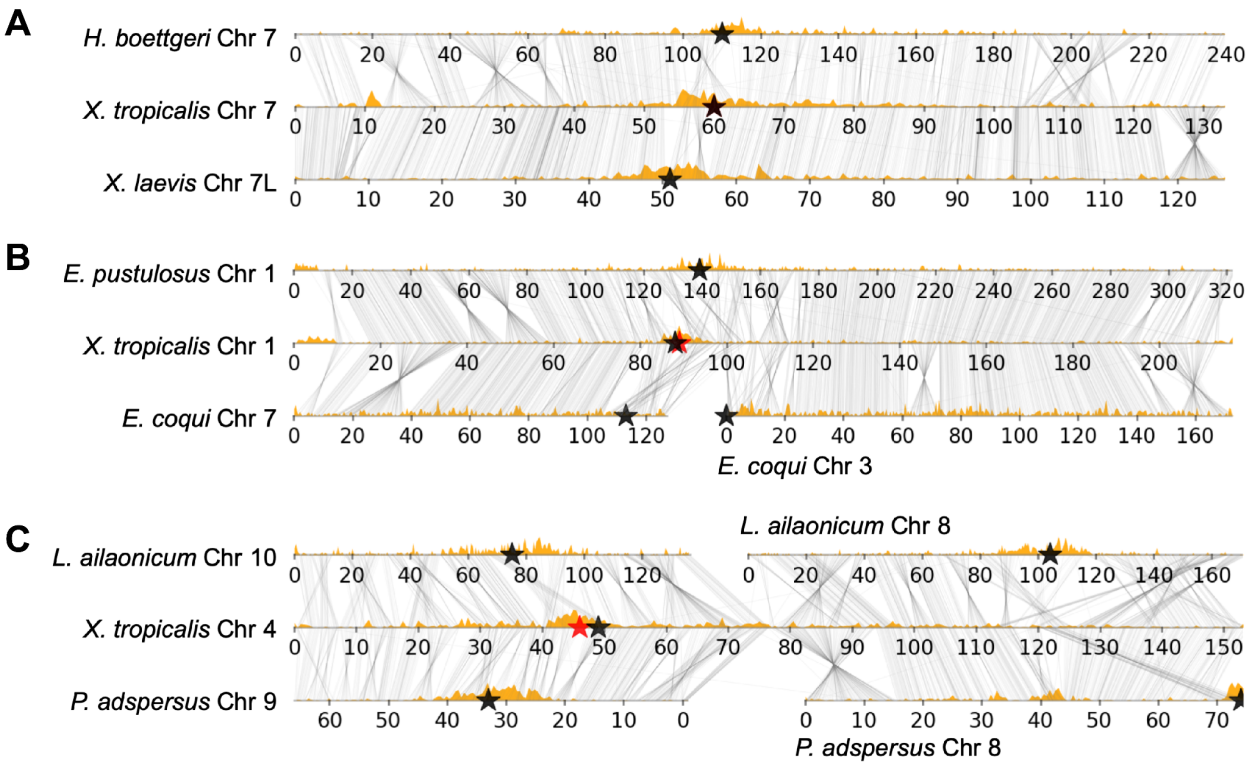

Fig. 3 Subtelomeric repeats highlight regions of chromosome fusion.

Examples of (**A**) conserved chromosome structure and pericentromere maintenance, (**B**) a Robertsonian translocation in the lineage leading to *E. coqui*, and (**C**) an end-to-end fusion that occurred in the lineage giving rise to *X. tropicalis* and subsequent pericentromere loss. The analyzed species were visualized with a custom script, alignment_plots.py (v1.0; https://github.com/abmudd/Assembly). For each plot, the HiC inference-based pericentromeric regions are depicted with black stars, the *X. tropicalis* centromeric satellite repeat from tandem repeat analysis with a red star, the density of L1 repeats per chromosome with light brown histograms, and the runs of collinearity containing at least one kb of aligned sequence between the species with connecting black lines.

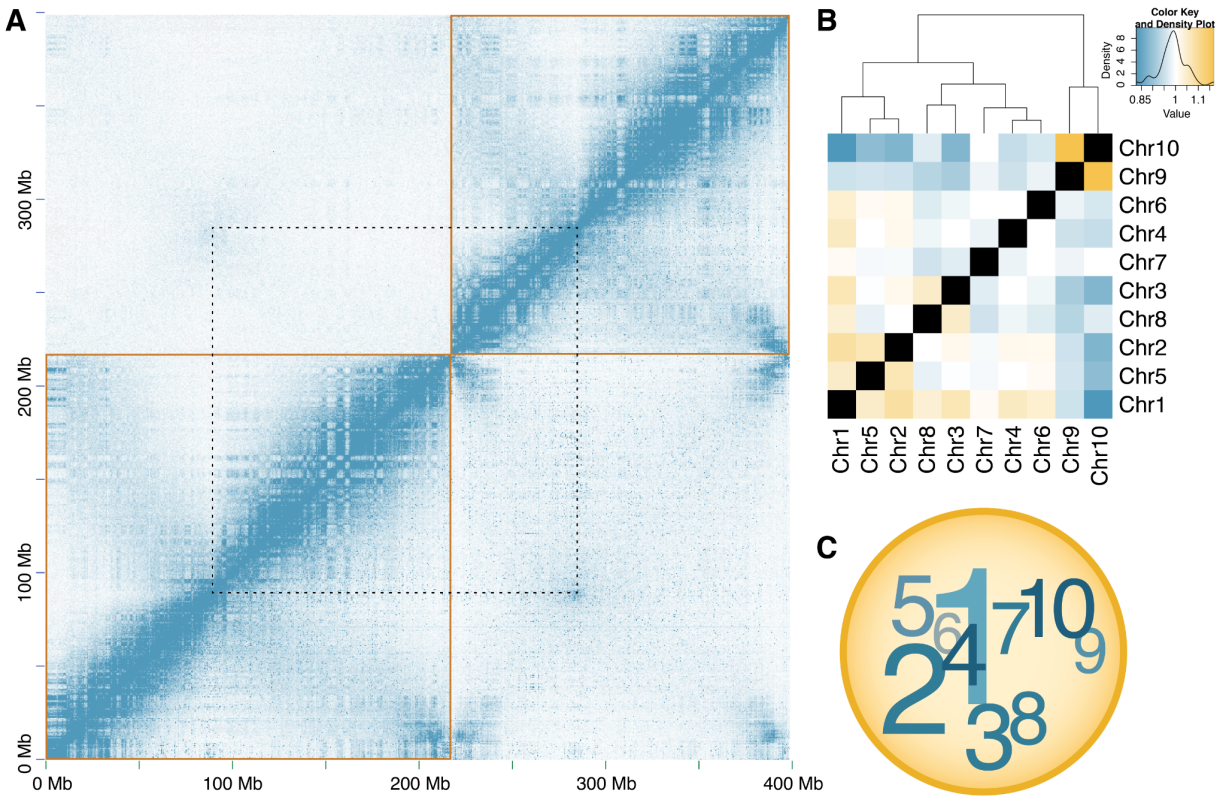

Fig. 4 Organization of *X. tropicalis* chromosomes into Rabl configuration and distinct nuclear territories.

(**A**) HiC contact matrices at 500 kb resolution for chromosomes 1 and 2 (lower-left and upper-right gold boxes, respectively) showing features of the three-dimensional chromatin architecture within *X. tropicalis* blood cell nuclei. Blue pixels represent chromatin contacts between X-Y pairs of genomic loci, and their intensity is proportional to their contact frequency. HiC read pairs are mapped stringently (MapQ ≥ 30) above the diagonal and permissively (MapQ ≥ 0) below the diagonal. The characteristic A/B compartment ("checkerboard") and Rabl-like inter-arm ("angel wing") contact patterns within each chromosome are evident. Above the diagonal, an increased frequency of inter-chromosomal chromatin contacts is observed between pericentromeres (connected by dotted lines) and between chromosome arms, suggesting a centromere-clustered organization of chromosomes in Rabl configuration. Below the diagonal, high-intensity pixels not present above the diagonal are present near the ends of chromosomes, suggesting a telomere-proximal spatial bias in the distributions of similar genomic repeats. See **Supplementary Fig. 1D** for a plot showing all chromosomes. (**B**) Chromosome territories within the

1251 nucleus. Yellow, white, and blue colors indicate the normalized relative enrichment, parity, and depletion
1252 of chromatin contacts between non-homologous chromosomes in the nucleus. For example,
1253 Chromosome 1 exhibits higher relative contact frequencies with all chromosomes except chromosomes
1254 Chr7, Chr9, and Chr10, which are generally depleted of contacts except among themselves. (MapQ ≥ 30;
1255 χ^2 (81, n = 24,987,749) = 3,049,787; $p < 2.2\times10^{-308}$; Relative range: 0.82774–1.16834). (**C**) Schematic
1256 representation of chromosome territories within the nucleus. Chromosome number size is proportional to
1257 the number of enriched interactions. Darker and lighter colors indicate chromosomes nearer or distant to
1258 the reader, respectively.

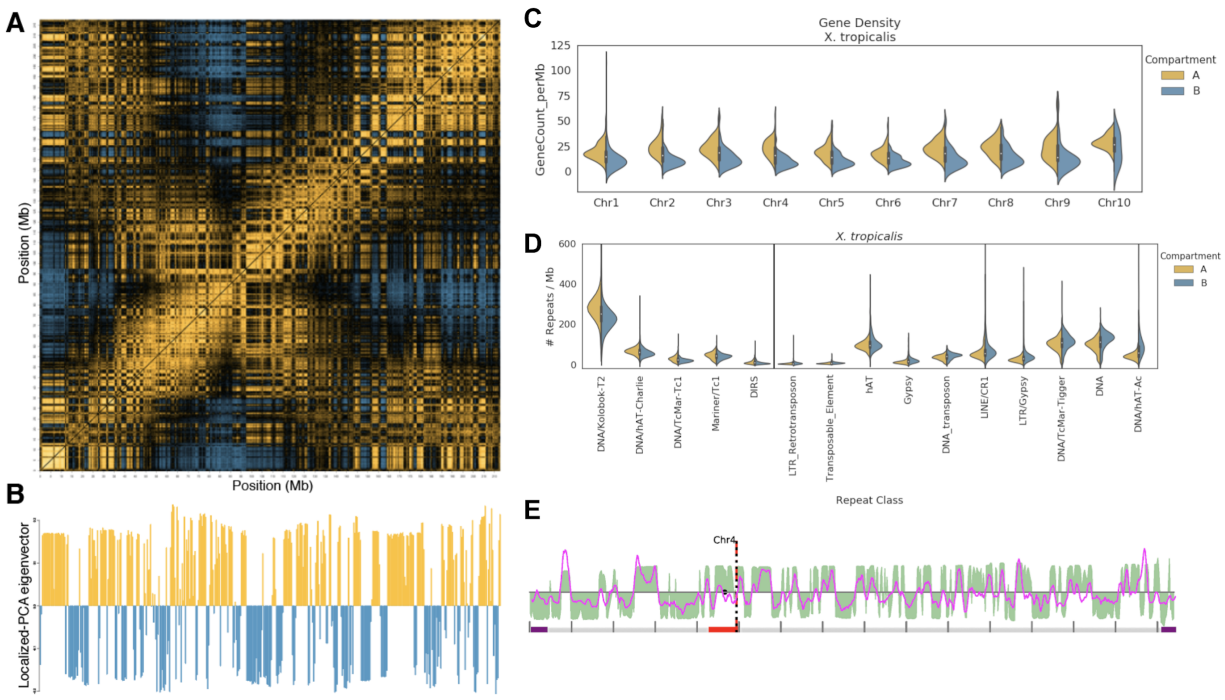

Fig. 5 A/B compartment structure and gene/repeat densities.

(**A**) Correlation matrix of intra-chromosomal HiC contact densities between all pairs of non-overlapping 250 kb loci on chromosome 1. Yellow and blue pixels indicate correlation and anti-correlation, respectively, and reveal which genomic loci occupy the same or different chromatin compartment. Black pixels indicate weak/no correlation. (**B**) The first principal component eigenvectors revealing compartment structure along chromosome 1, obtained by singular value decomposition of the correlation matrix in **A**. Yellow (positive) and blue (negative) eigenvectors indicate regions of chromosome 1 partitioned into the A and B compartments, respectively. (**C**) Gene density distributions in A vs. B compartments per chromosome. (**D**) Repeat classes significantly enriched in A vs. B compartments. (**E**) The eigenvectors obtained from the HiC correlation matrix (PC1, green) and the eigenvectors from the repeat density matrix (PC3, pink) coincide at the transitions of predicted A/B compartments (**Supplementary Fig. 16B**).

Table 1 Organization and conservation of the 13 ancestral chromosomes of pipanuran genomes

	Phylogenetic position	Structural event
(1)	Stem pipid lineage	J + K ⟹ JK D. + E. ⟹ D.E I• + •H ⟹ I•H (Rob. fusion)
(2)	*P. adspersus* lineage after divergence from *R. temporaria*	A + M ⟹ A1.m1 + m2.A2
(3)	*E. pustulosus* lineage after divergence from *E. coqui*	M + I ⟹ M.I (Rob) K + D ⟹ K.D (Possible end-end)
(4)	*E. coqui* lineage after divergence from *E. pustulosus*	G1•G2 ⟹ G1• + •G2 (Rob. fission) A1•A2 ⟹ A1• + •A2 (Rob. fission) I + K ⟹ I•K (Rob. fusion + inversion) E + F1•F2 +B1•B2 + H ⟹ E•F1 + F2•B2 + B1•H
(5)	*H. boettgeri* lineage after divergence from *Xenopus*	M + J•K ⟹ MJK
(6)	*X. laevis* progenitor lineage after divergence from *X. tropicalis*	L + M ⟹ LM

Rob, Robertsonian. Middle-dots (i.e., "•") represent centromeres. Periods (i.e., ".") represent translocation breakpoints.

Supplementary Information

(Provided in a separate document)

Supplementary Fig. 1 Genome assembly and recovery of missing genes.

Supplementary Fig. 2 *Xenopus tropicalis* genome-wide HiC contact map.

Supplementary Fig. 3 GC landscape and tandem repeats.

Supplementary Fig. 4 Comparison of gene content in assemblies of model vertebrates.

Supplementary Fig. 5 PCA eigenvectors projected on genomic coordinates.

Supplementary Fig. 6 *Xenopus tropicalis* Nigerian strain residual heterozygosity.

Supplementary Fig. 7 Assembly and annotation of other frog species.

Supplementary Fig. 8 Pairwise gene colinearity of frog genomes.

Supplementary Fig. 9 Chromosome fusions in *Xenopus laevis* and *Hymenochirus boettgeri*.

Supplementary Fig. 10 Estimating the positions of *Xenopus tropicalis* centromeres.

Supplementary Fig. 11 *Xenopus tropicalis* recombination landscape.

Supplementary Fig. 12 Distribution of Satellite repeats in *Xenopus tropicalis*.

Supplementary Fig. 13 Correlates of recombination rate.

Supplementary Fig. 14 Zebra finch subtelomeric tandem repeats.

Supplementary Fig. 15 Microsatellite origin SINE/tRNA evolved into a microsatellite sequence.

Supplementary Fig. 16 *Xenopus tropicalis* 3D chromatin structure and nuclear organization.

Supplementary Table 1 Sequence Completeness.

Supplementary Table 2 Transcript coverage of *X. tropicalis* assemblies v9 and v10.

Supplementary Table 3 *Xenopus tropicalis* protein coding loci annotation summary statistics.

Supplementary Table 4 *Xenopus tropicalis* repeat abundances.

Supplementary Table 5 Summary of other frog genome assemblies.

Supplementary Table 6 Summary of annotations for other frog genomes.

Supplementary Table 7 BUSCO genome scores of other frog genome assemblies.

Supplementary Data Files

Supplementary Data 1: Table of sequencing data.

An MS Excel file summarizing the sequencing data used to construct the six frog genome assemblies new or updated in this study, as well as the RNA-seq data used for annotating their protein-coding genes. The *X. tropicalis* ChIP-seq data are also included.

Supplementary Data 2: Genetic markers.

An MS Excel file containing the marker number, locus identifier, chromosome name, centiMorgan position, and chromosome coordinate for each genetic marker in the F_2 *X. tropicalis* genetic linkage map.