Title
Comparative genomics and chromosome evolution

Author
Mudd, Austin Beck

Publication Date
2019

Peer reviewed|Thesis/dissertation

Comparative genomics and chromosome evolution

By

Austin Beck Mudd

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Daniel Rokhsar, Chair
Professor Michael Eisen
Professor Richard Harland
Professor Rasmus Nielsen

Fall 2019

Abstract

Comparative genomics and chromosome evolution

by

Austin Beck Mudd

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Daniel Rokhsar, Chair

Viewed through the lens of comparative genomics, how have chromosomes and karyotypes evolved, particularly under extreme scenarios of rapid or torpid evolution? To examine this question, I produced and analyzed new high-quality, chromosome-scale genome assemblies, representing cases of extreme chromosome evolution. The interplay between convergent and divergent genomic architectures can further our understanding into the maintenance of chromosome structure and organization as well as the underlying biological mechanisms.

Despite their recent divergence, muntjac deer show striking karyotype differences. In Chapter 2, I produced new chromosome-scale genome assemblies for the Chinese and Indian muntjacs, *Muntiacus reevesi* (2n=46) and *Muntiacus muntjak* (2n=6/7), and analyzed their evolution and architecture. I identified six fusion events shared by both species relative to the cervid ancestor and, therefore, present in the muntjac common ancestor, six fusion events unique to the *M. reevesi* lineage, and 26 fusion events unique to the *M. muntjak* lineage. One of these *M. muntjak* fusions reversed an earlier fission in the cervid lineage. Although comparative Hi-C analysis uncovered differences in long-range genome contacts and A/B compartment structures, I discovered widespread conservation of local chromatin contacts between the muntjacs, even near the fusion sites. Analysis of the muntjac genomes revealed new insights into this unique case of rapid karyotype evolution and the resulting biological variation.

In contrast to muntjacs, frogs are a more phylogenetically ancient order, with two prominent species, *Xenopus laevis* and *Xenopus tropicalis*, being utilized as vertebrate cell and developmental model systems. In Chapter 3, I reported new chromosome-scale genome assemblies for three distantly related frogs: the common coquí (*Eleutherodactylus coqui*), the túngara frog (*Engystomops pustulosus*), and the Zaire dwarf clawed frog (*Hymenochirus boettgeri*). Through comparative sequence analysis of these and other frog genomes, I identified long-range and even chromosome-scale synteny, denoting the stability and pervasiveness of this type of genomic conservation across the frog phylogeny. These chromosome-scale comparisons also revealed the 13 ancestral chromosomes, with limited Robertsonian translocations and end-to-end fusions explaining the observed chromosome variations. These assemblies, encompassing over a billion years of evolutionary divergence, unmasked insights into torpid karyotype evolution in frogs.

To my mother and sister

In loving memory of my father

Table of contents

1 Introduction

History. The field of genomics emerged with the pivotal papers from James Watson and Francis Crick in 1953 about the structure and function of DNA [1,2] along with the first sequencing publications — the amino acid sequence of insulin in the 1950s [3], the nucleic acid sequence of yeast alanine transfer RNA in 1965 [4], and the genome of bacteriophage φX174 in 1977 [5]. However, the field's foundations stretch back to seminal discoveries in the 19th century including the laws of Mendelian inheritance [6] and Darwin's theory of evolution by natural selection [7], which were released 154 and 160 years ago, respectively. Similarly, chromosomes, or "chromatic nuclear figures", were first described in 1873, with the term chromosome coined in 1888 [8]. The fields of comparative genomics and chromosome evolution originated in the early and mid-1900s, with the findings by Alfred Sturtevant, Theodosius Dobzhansky, H. J. Muller, and other Drosophila geneticists about conserved genic content between chromosomes of different *Drosophila* spp. and the standardization of chromosome arm nomenclature (see Figure 1 from [9] and Figure 3 from [10]) [11,12].

In the following decades, technological developments improved the ability to explore chromosome evolution, from advancements in chromosome preparation in the 1950s to the development of chromosome banding in the 1960s to the introduction of fluorescent *in situ* hybridization in the 1970s. As a result, researchers were able to evaluate chromosome abnormalities in leukemia and Down's syndrome, accurately identify the 46 human chromosomes, and explore genomic changes across evolutionary history [12–14]. The newest iteration of technological development is the proliferation of genome sequencing and the ability to undertake chromosome-scale scaffolding of genome assemblies using technologies ranging from optical mapping [15] to chromosome conformation capture [16], an advancement that, in fact, has enabled this dissertation work. These improvements have greatly increased our understanding of chromosome biology, from broadscale chromosome territories to A/B compartments to sub-megabase topologically associating domains (see Figure 1 from [17]) [17,18], with hypotheses like the correlation between chromatin organization and banding patterns tying together the history of the field [19].

Comparative genomics. With the aforementioned rapid development of novel sequencing technologies [20], there has also been an expansion in the use of comparative genomics to better understand genome content, evolution, and variation. In particular, advances in sequencing and informatics have enabled the prediction and identification of various genomic features — coding elements like protein-coding genes; non-coding elements including non-coding RNAs, repeats, and regulatory elements; and structural variations such as copy number variations and translocations — through comparison with related or otherwise well-annotated reference genome assemblies and databases [21–23]. With these annotations, studies have examined genome variation through the lens of ecological and evolutionary forces, gene duplication, and polyploidization events [24–26]. For example, an influx of mammalian genome assemblies in 2011 enabled investigation into mammalian processes including hibernation, domestication, marine adaptation, dietary specialization, and disease [27–34]. Taxa-specific analyses have since expanded into other clades, such as birds [35], reptiles [36], and fish [37], with more assemblies and analyses underway [38,39]. New reference genome assemblies can enable further

investigation into genomic regions under evolutionary selection, ancestral genome content and thus ancestral biology, regulatory elements, and various other facets [40].

Chromosome evolution. Even with these technological innovations and new sources of data, however, the core question at the intersection of comparative genomics and chromosome evolution has remained the same: based on analysis of different species, populations, and individuals, what chromosome changes have occurred across evolutionary history? To this end, the observed variations fall into three categories: microstructural rearrangements with insertions, deletions, and duplications of segments or whole chromosomes; macrostructural rearrangements within and between chromosomes resulting in changes to chromosome morphology or number; and polyploidization events involving the entire genome [13,41]. As a few examples of these types of variations, studies have documented a variety of duplication, insertion, and deletion events leading to genetic disorders in humans and explored the potential detrimental effects caused by these types of chromosome abnormalities [42]. Case analyses in humans have also identified a variety of chromosome structural variations including Robertsonian translocations, reciprocal translocations, and chromosome inversions, enumerating any resulting phenotypic effects [43]. On a larger evolutionary scale (see Figure 5 from [44]), studies have described the end-to-end fusion of chromosome 2 in the human ancestor [45]; a whole genome duplication in the teleost ancestor [46]; whole genome duplication, fission, and fusion events in the vertebrate lineage [47,48]; and deep units of synteny across animals [49,50].

Rearrangements of chromosomes, however, must be evaluated in light of mitotic, meiotic, and related chromosome constraints [51] as well as the health and fecundity of individuals and populations. Chromosome rearrangement events can be deleterious or unable to be propagated, can result in a balanced polymorphism that is maintained in a population, or under select situations, can become fixed and potentially enable speciation [52]. Advantageous rearrangements are likely to become fixed in a population, although small population size, meiotic drive, epistatic interactions, and other factors can enable fixation of neutral or underdominant rearrangements [41,53]. Regarding constraints, chromosomes resulting from rearrangements must, in most eukaryotes, each retain a centromere and telomeres, fit within the nucleus and cell, and allow for proper mitotic recombination and segregation as well as other required cellular functions [52]. These mechanisms and constraints limit the means of viable chromosome evolution, with particular types of rearrangements therefore more probable than others [54].

Evolutionary rate. Over the past 160 years since Darwin's theory of evolution by natural selection [7], there has been continued disagreement about the rate of evolutionary change, with some arguing that evolution is fast, others slow, and still others both fast and slow [55]. An analysis of 10,000 generations of *Escherichia coli* found rapid evolution during the initial 2,000 generations, slower evolution during the following 3,000 generations, and limited evolution during the remaining 5,000 generations [56], demonstrating that evolutionary rate can be dynamic. Given that any finite partially ordered set has a maximal element and a minimal element [57], evolutionary rates must similarly have a maximum and a minimum, with two models describing this phenomenon. The first model, proposed by George Simpson, suggests different rates for particular taxonomic groups and classifies the organisms into one of three categories depending on rate: (1) bradytely for slow evolution, (2) horotely for normal evolution,

or (3) tachytely for fast evolution [58]. The second model, put forward by Thomas Schopf, proposes a singular set of rates that allows for variation, similar to a Gaussian distribution [59]. In either scenario, extreme rates of evolution exist, whether classified as bradytely and tachytely or present at the tails of the evolutionary rate distribution. When applied to karyotype evolution, terms such as karyotypic conservatism and karyotypic megaevolution [60] or chromosomal bradytely and chromosomal tachytely [61] have been coined.

Outline. Throughout this dissertation, I analyzed the evolution and architecture of inter-chromosome rearrangements between new and improved chromosome-scale genome assemblies under extreme scenarios of rapid or torpid evolution, with the results of each examination formatted as a manuscript. The second chapter is an analysis of rapid karyotype evolution in the Indian and Chinese muntjacs, *Muntiacus muntjak* and *Muntiacus reevesi*, characterizing shared and lineage-specific fusion events as well as documenting underlying changes in the chromatin architecture. The third chapter is an analysis of torpid karyotype evolution between seven frog assemblies, characterizing the ancestral chromosome state, the mechanisms of karyotype change, and the vast synteny across the frog phylogeny. These chapters are bounded by an introduction and a conclusion, providing the necessary context for the work as well as future directions.

2 Rapid chromosome evolution in muntjacs

This chapter is an update of the *bioRxiv* preprint [62], of which I am the first author. This work is included with permission from the authors below.

Authors' contributions. Austin B. Mudd assembled and annotated the genomes, completed the bioinformatic analyses, and wrote the manuscript. Jessen V. Bredeson assisted with the bioinformatic analyses and script development. Rachel Baum prepared the Hi-C libraries. Dirk Hockemeyer coordinated the cell line acquisitions, extracted the DNA, and prepared the Hi-C libraries. Dirk Hockemeyer and Daniel S. Rokhsar provided scientific leadership of the project and wrote the manuscript.

2.1 Background

Rapid karyotype evolution, also known as karyotypic megaevolution [60] or chromosomal tachytely [61], has been found in various species, such as rodents [63], bears [64], and gibbons [65], and as a byproduct of chromosome instability in cancer [66]. Perhaps the most spectacular example of rapid karyotype evolution is found in muntjacs, a genus of small deer with karyotypes ranging from 2n=46 for *Muntiacus reevesi* to 2n=6/7 for female/male *Muntiacus muntjak*, respectively, with *M. muntjak* having the smallest chromosome number of any mammal [67]. Cytogenetic analysis showed that muntjac karyotype diversity arose primarily through centromere-telomere (head-tail) tandem fusions and, to a lesser extent, centromere-centromere (head-head) tandem fusions [68,69], i.e., Robertsonian translocations [70]. Importantly, independent fusions occurred in each lineage after divergence from their common ancestor, such that the 2n=46 *M. reevesi* karyotype does not represent an intermediate stage between the ancestral 2n=70 cervid karyotype and the highly reduced *M. muntjak* karyotype [71,72].

Understanding the variation of genomic architectures in muntjacs has the potential to reveal new insights into chromosome evolution [48]. We therefore set out to explore karyotype changes in muntjacs by determining the number, distribution, and timing of shared and lineage-specific fusion events. To this end, we described the first chromosome-scale assemblies of *M. muntjak* and *M. reevesi*, with contiguity metrics that surpass those of earlier draft assemblies [73,74]. To infer the series of karyotype changes in muntjac, we leveraged existing assemblies of *Bos taurus* (cow) [75], *Cervus elaphus* (red deer) [76], and *Rangifer tarandus* (reindeer) [77]. In total, we characterized 38 muntjac fusion events, six of which are shared by *M. muntjak* and *M. reevesi*. The rate of 26 unique fusion events in the *M. muntjak* lineage over 4.9 million years represents more than an order of magnitude increase relative to the mammalian average [78]. Although the molecular mechanism driving these karyotype changes is unknown, we found that one fusion event in the *M. muntjak* lineage reversed a chromosome fission that occurred earlier in the cervid lineage. In another case, we found that a pair of ancestral cervid chromosomes likely fused independently in the *M. muntjac* and *M. reevesi* lineages. As such, some chromosomes may be more prone to karyotype changes than others, and care should be taken when applying the parsimony principle due to the possibility of convergent change.

We also took advantage of the extensive collinearity of the muntjac genomes to study changes in three-dimensional genome architecture that accompany chromosome fusions. While karyotype changes disrupted long-range three-dimensional genome structure, including A/B compartments, we found few changes at the local level. These analyses explored features of chromosome structure derived from the unique evolutionary history of these two karyotypically divergent species.

2.2 Results and discussion

Assembly and annotation. To investigate the tempo and mode of muntjac chromosome evolution, we generated high-quality, chromosome-scale genome assemblies for *M. muntjak* and *M. reevesi* (Table 2.1) using a combination of linked reads (10X Genomics Chromium Genome) and chromatin conformation capture (Dovetail Genomics Hi-C; Table 2.2, Methods). The resulting assemblies each contained 2.5 gigabases (Gb) of contig sequence with contig N50 lengths over 200 kilobases (kb). In both assemblies, over 92% of contig sequence was anchored to chromosomes. Compared with the publicly available assemblies [73,74], the assemblies described here represented a hundredfold improvement in scaffold N50 length and severalfold improvement in contig N50 length. As typical for short-read assemblies, our muntjac assemblies were largely complete with respect to genic sequences (see below) but were likely to underrepresent repetitive sequences such as pericentromeric heterochromatin and repetitive subtelomeric regions.

The assembled chromosome numbers recapitulated the karyotypes reported in the literature (2n=6 for female *M. muntjak* [79] and 2n=46 for *M. reevesi* [80]). *M. reevesi* chromosomes were validated against previously published chromosome painting data [81]. For *M. muntjak*, we aligned 377 previously sequenced bacterial artificial chromosomes (BACs) [82–84] and, based on corresponding fluorescent *in situ* hybridization (FISH) location data, found that 360 (95%) of BACs aligned to the expected chromosomes. Of the 17 BACs that aligned to a different chromosome than expected by FISH, 16 were well aligned to our assembly in regions of conserved collinearity among cow, red deer, and muntjac chromosomes, which suggested that the FISH-based chromosome assignments of these BACs were likely incorrect. Only one of these 17 BACs aligned to two of our assembled *M. muntjak* chromosomes, indicating a possible local misassembly or BAC construction error.

For each muntjac genome, we annotated ~26,000 protein-coding genes based on homology with *B. taurus* [75], *Ovis aries* (sheep) [85], and *Homo sapiens* (human) [86]. Over 98% of the resulting genes were functionally annotated by InterProScan (v5.34-73.0) [87]. We identified 19,649 one-to-one gene orthologs between the two muntjac species as well as 7,953 one-to-one gene orthologs present in the two muntjacs, *B. taurus* [75], *C. elaphus* [76], and *R. tarandus* [77]. These ortholog sets were used in the evolutionary and phylogenomic analyses below (Figure 2.1A and 2.1C, Table 2.3, Methods). Gene set comparisons (Figure 2.2) showed that the muntjac annotations included several thousand more conserved cervid genes than were found in the *C. elaphus* and *R. tarandus* annotations and demonstrated comparable completeness to *B. taurus*, supporting the accuracy of the muntjac assemblies in genic regions.

Comparative analysis. In order to study sequence and karyotype evolution, we aligned the two muntjac assemblies to each other and to *B. taurus* [75] as well as *B. taurus* to *C. elaphus* [76] and *R. tarandus* [77]. The pairwise alignment of the muntjac genomes contained 2.45 Gb of contig sequence, or over 97% of the assembled contig sequence lengths, with an average identity of 98.5% excluding indels. This reflected the degree of sequence conservation between the two species and their recent divergence. In comparison, alignments of red deer, reindeer, and muntjacs to *B. taurus* contained 1.80 to 2.21 Gb of contig sequences with 92.7% to 93.2% average identity. Sequence alignments were merged into runs of collinearity, and analysis of these runs revealed the timing of fission and fusion events in each lineage (Figures 2.1A–B and 2.3).

Chromosome evolution. We assessed chromosome evolution in muntjacs using *B. taurus* (BTA) and *C. elaphus* (CEL) as outgroups. For convenience, we referred to chromosome regions primarily by their BTA chromosome identifiers. We corroborated prior reports in literature [81] that:

1. In the last common ancestor of cow and deer, segments corresponding to the two cow chromosomes BTA26 and BTA28 were present as a single chromosome in the last common ancestor of cervids and *B. taurus*. This ancestral state, corresponding to BTA26_28, was retained in *C. elaphus* and the muntjacs.
2. Twelve chromosomes of the cervid ancestor arose by fission of chromosomes represented by six cow chromosomes (BTA1 => CEL19 and CEL31; BTA2 => CEL8 and CEL33; BTA5 => CEL3 and CEL22; BTA6 => CEL6 and CEL17; BTA8 => CEL16 and CEL29; and BTA9 => CEL26 and CEL28).
3. Although chromosomes homologous to BTA17 and BTA19 were fused in the *C. elaphus* lineage as CEL5, this fusion was unique to the *C. elaphus* lineage, and these cow chromosomes corresponded to distinct ancestral cervid chromosomes.

In the muntjacs, we found six fusions shared by *M. muntjak* and *M. reevesi* (BTA7/BTA3, BTA5prox/BTA22, BTA2dist/BTA11, BTA18/BTA25/BTA26_28 (the fusion of three ancestral chromosomes were counted as two fusion events), and BTA27/BTA8dist; Figure 2.4). All six of these fusions shared by *M. muntjak* and *M. reevesi* were also corroborated in previous BAC-FISH analyses of *Muntiacus crinifrons*, *Muntiacus feae*, and *Muntiacus gongshanensis* [88,89]. After the divergence of *M. muntjak* and *M. reevesi*, each lineage experienced additional fusions. In the *M. reevesi* lineage, there were six fusions (BTA7_3/BTA5dist, BTA18_25_26_28/BTA13, BTA2prox/BTA9dist/BTA2dist_11, BTA5prox_22/BTA24, and BTA29/BTA16). In the *M. muntjak* lineage, the three chromosomes arose via 26 lineage-specific fusions:

- *M. muntjak* chromosome 1: BTA7_3/BTA5prox_22/BTA17/BTA2prox/BTA1dist/ BTA29/BTA8prox/BTA9dist/BTA19/ BTA24/BTA23/BTA14/BTA2dist_11,
- *M. muntjak* chromosome 2*:* BTA15/BTA13/BTA18_25_26_28/BTA9prox/BTA20/ BTA21/BTA27_8dist/BTA5dist, and
- *M. muntjak* chromosome 3: BTAX/BTA1prox/BTA4/BTA16/BTA12/BTA6prox/ BTA6dist/BTA10.

While both *M. muntjak* and *M. reevesi* karyotypes included chromosomes that arose by fusion of BTA13 and BTA18_25_26_28, these events likely occurred independently. Consistent with our analysis, published BAC FISH mapping of *M. reevesi* against *M. crinifrons*, *M. feae*, and *M.*

gongshanensis found different locations of *B. taurus* chromosomes 13 and 18_25_26_28 in the muntjac species [88,89], which supported our finding that these were independent, lineage-specific fusion events.

In total, we found 38 fusion events and no fissions in the muntjac lineage (Figure 2.1A). All 12 of the *M. reevesi* fusions identified by our comparative analysis were substantiated by BAC-FISH [81], and 17 of the *M. muntjak* fusions were corroborated [90]. Our results were also consistent with the BAC-FISH findings of Chi et al. [69]. The rates of karyotype changes based on fission and fusion events in muntjacs were higher than the mammalian average of 0.4 changes per million years [78]. The *M. muntjak* lineage, with six fission events and 32 fusion events over the past 22.8 million years since the cervid ancestor, averaged 1.7 events per million years. In the 4.9 million years since the divergence from *M. reevesi*, this rate increased to 5.3 fusion events per million years, an order of magnitude greater than the mammalian average. The *M. reevesi* lineage, on the other hand, averaged 0.8 events per million years over the past 22.8 million years, with an accelerated rate of 1.2 events per million years over the past 4.9 million years. Although the calculated nucleotide divergence and time between the two muntjac species (Figure 2.1A and 2.1C, Table 2.3) mirrored the evolutionary distance between humans and chimpanzees [91,92], this number of fusion events since the muntjac last common ancestor far exceeded the rate in the chimpanzee and human lineages since their respective last common ancestor, i.e., a single fusion on the human lineage [45].

Reversal of a cervid-specific fission in *M. muntjak*. While analyzing the fission and fusion events, we discovered a fusion in *M. muntjak* that reversed, to the resolution of our assembly, the cervid-specific fission of the ancestral chromosome corresponding to BTA6 (Figure 2.5). To estimate the probability of such a reversion occurring by chance given the high rate of fusion in *M. muntjak*, we simulated a simplified model for karyotype change with four rules: (1) only one fission was allowed per chromosome; (2) all fissions occurred first, followed by all fusions; (3) for each fission, a chromosome was chosen at random; and (4) for each fusion, chromosomes and orientations were chosen at random. From a starting karyotype of n=29, representing the last common ancestor of cervids and *B. taurus* [81], we simulated the model of fissions and fusions to 1 million iterations per fission-fusion combination (Figure 2.6). The *M. muntjak* lineage, with six fissions and 32 fusions, had a 4.1% probability of at least one fusion reversing a prior fission. In comparison, the *C. elaphus* lineage, with six fissions and one fusion, had only a 0.13% probability of reversal by chance, and the *M. reevesi* lineage, with six fissions and 12 fusions, had a 1.5% chance of reversal. Even with the large number of fusions in muntjacs, the probability of a chance reversal of a previous fission was small. However, the reversal could have been aided by unmodeled effects of differential chromosome fusion probability arising, for example, by chromosome proximity in the nucleus. This analysis pointed to the importance of having multiple outgroups, here both *B. taurus* and *C. elaphus*, in phylogenetic analyses of karyotypes.

Changes in three-dimensional genome structure after karyotype change. Despite the extensive fusions documented above for *M. muntjak* and *M. reevesi*, the genomes were locally very similar with 98.5% identity in aligned regions and a nucleotide divergence based on fourfold degeneracy of 0.0130 substitutions per site. Our Hi-C chromatin conformation capture data allowed us to examine the impact of these chromosome rearrangements on megabase (Mb)

and longer length scales. Focusing first on the *M. muntjak* and *M. reevesi* lineage-specific fusion sites (Tables 2.4–2.7), we noted the maintenance of distinct Hi-C boundaries in several examples, such as the junction between *M. muntjak* chromosome segments X and 3 at 133 Mb on chromosome 3_X. Other fusion sites, however, showed no notable difference compared with the rest of the genome in *M. muntjak*. As expected, *M. reevesi* showed a clear distinction between intra- and inter-chromosome contacts, including across fusion sites in *M. muntjak* (Figure 2.7). To quantify the chromatin changes at these fusion sites, we divided the genomes into one Mb bins and compared normalized inter-bin Hi-C contacts between bins five Mb apart in the two species, using the *M. muntjak* assembly as the backbone for comparison (Figure 2.8). Supporting the initial visual analysis, we found that most bins containing a fusion site had fewer long-range chromatin contacts in *M. reevesi*, averaging 0.16 ± 0.09 normalized contacts per bin, compared with *M. muntjak*, averaging 0.62 ± 0.35 normalized contacts per bin, although we identified bins with few contacts in both species (Figure 2.8).

In order to test whether differences were present at a more local level, we next compared normalized one Mb intra-bin Hi-C contacts between the two species, again using the *M. muntjak* assembly as the backbone for comparison. We found that most of the chromatin contacts were consistent between the two muntjacs, including all but three of the bins containing fusion sites (Figures 2.9A and 2.10). Several regions, however, showed distinctive variation in chromatin contacts between the two species: the X chromosome and two regions on *M. muntjak* chromosome 1 (186–355 and 615–630 Mb). Since our sequenced *M. reevesi* sample was male [71] while the sequenced *M. muntjak* sample was female [93], we expected a difference in chromatin contacts on the X chromosome, a finding that was further supported by analysis of copy number across the genome using the 10X Genomics linked reads (Figure 2.9B). From this copy number analysis, we also hypothesized that the two regions on *M. muntjak* chromosome 1 (186–355 and 615–630 Mb) were a haplotype-specific duplication and a haplotype-specific deletion, respectively, explaining the difference in chromatin signal between the two muntjacs (Figure 2.9C–D). Although the inter-bin analysis identified long-range chromatin changes between sites five Mb apart, our quantitative comparison of one Mb intra-bin chromatin contacts found substantial chromatin conservation between the genome assemblies, including nearly all of the fusion sites. This conclusion was further supported by intra-bin analysis with 100 kb bins (Figure 2.11).

On a multi-megabase length scale, mammalian chromosomes can be subdivided into alternating A/B compartments based on intra-chromosome contacts. These compartments correspond to open and closed chromatin, respectively, and differ in gene density and GC content [94]. To test whether these compartments were conserved or disrupted by fusions, we computed the A/B chromatin compartment structures for *M. muntjak* and *M. reevesi* from the Hi-C data, again using the *M. muntjak* assembly as the backbone for comparison. We found that, in general, compartment boundaries were not well conserved between the muntjacs (Figure 2.12). Specifically, for A/B compartments larger than three Mb, only 17 compartments were completely conserved between the two species, out of 221 A/B compartments analyzed in *M. muntjak* and 161 in *M. reevesi*. We found that many of the compartments in *M. reevesi* were subdivided into multiple compartments in *M. muntjak*. Combining our analysis of A/B compartments and chromatin contacts, we found that the extensive set of fusions in the *M. muntjac* lineage altered the three-dimensional genome structure at the multi-megabase scale

while still maintaining conservation at the local level. These large-scale chromatin changes must have only limited effects on the underlying gene expression, since the two muntjac species can produce sterile hybrid offspring [95]. Similar uncoupling between genome topology and gene expression has been observed in *Drosophila melanogaster* [96].

Conclusions. We presented here new chromosome-scale assemblies of two muntjac deer that differ dramatically in karyotype, despite only limited sequence change, after ~4.9 million years of divergence. Analysis of these new assemblies revealed multiple changes in the underlying chromosome structure, including variation in the A/B compartments, despite maintenance of sub-megabase, three-dimensional genome contacts. One of the chromosome fusions reversed an earlier chromosome fission to the resolution of our assemblies, with the two events being separated by more than eight million years. Future studies will use these assemblies to resolve the nature of the fusion sites and to better understand the biological mechanisms related to chromosome fissions and fusions in muntjac.

2.3 Figures and tables

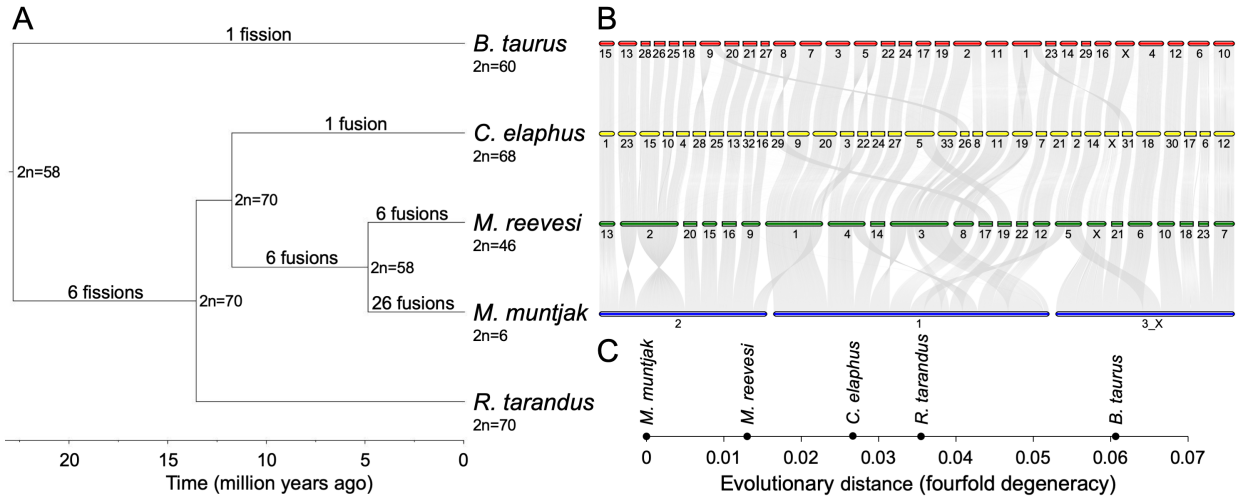


Figure 2.1. [A] The phylogenetic tree of the five analyzed species, calculated from fourfold degenerate sites and divergence time confidence intervals, was visualized with FigTree (commit 901211e; https://github.com/rambaut/figtree). The ancestral karyotype at each node and the six branches with fission and fusion events relative to the ancestral karyotype were labeled on the tree. The lack of fissions or fusions on the *R. tarandus*-specific branch as well as the timings of the cervid-specific and *B. taurus*-specific fissions were derived from literature [81]. [B] The alignment plot was generated with jcvi.graphics.karyotype (v0.8.12; https://github.com/tanghaibao/jcvi) using runs of collinearity containing at least 25 kb of aligned sequence between *B. taurus*, *C. elaphus*, *M. reevesi*, and *M. muntjak*, extracted from the ROAST-merged MAF file. *R. tarandus* was excluded, as it was not a chromosome-scale assembly. [C] Pairwise distances in substitutions per fourfold degenerate site extracted from the RAxML (v8.2.11) [97] phylogenetic tree using Newick utilities (v1.6) [98] were shown relative to the reference genome *M. muntjak*.

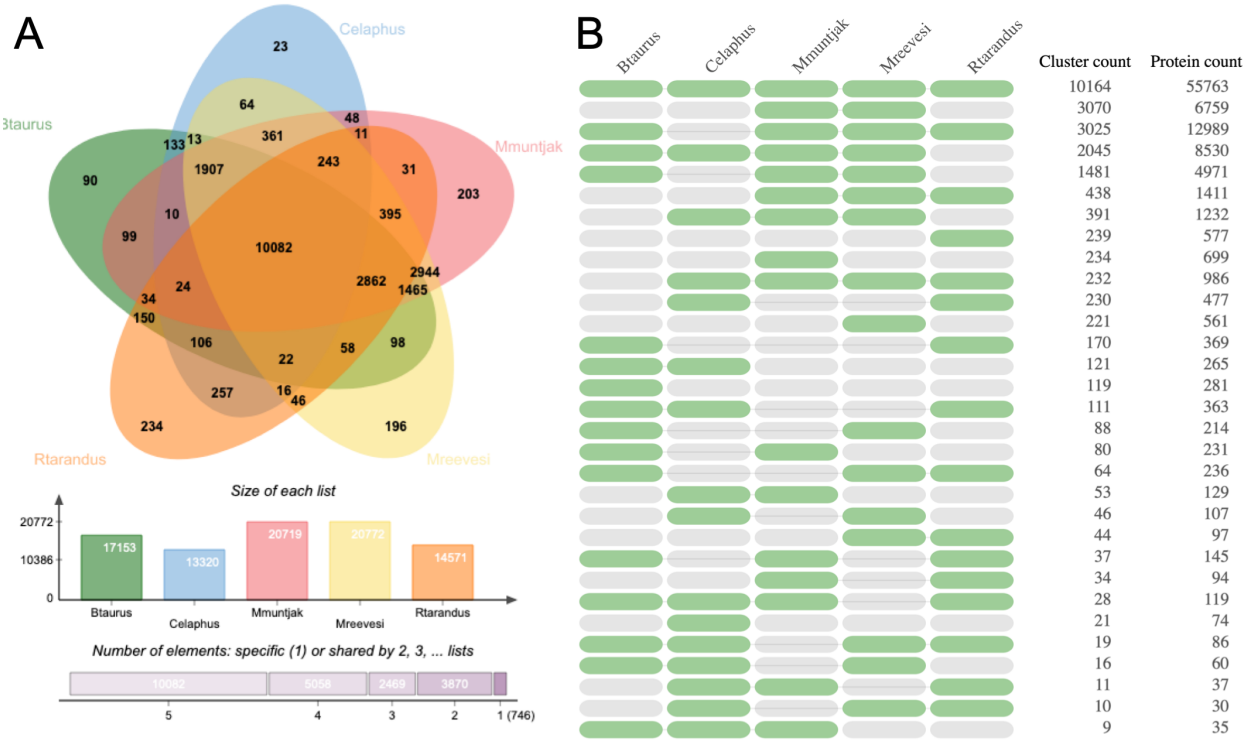


Figure 2.2. [A] Venn diagram of gene homology between the two muntjac annotations, *B. taurus* (Ensembl release 94 September 2011 genebuild of GCA_000003055.3) [75,99], *C. elaphus* (publication genebuild of GCA_002197005.1) [76], and *R. tarandus* [77,100] annotations analyzed with OrthoVenn [101] and [B] the occurrence table of gene homology clusters between these species reanalyzed with OrthoVenn2 [102] for visualization purposes. In the occurrence table, the green and grey ovals represent the presence or absence, respectively, of that species in the OrthoVenn2 clustering. The number of clusters and proteins were provided for all species combinations.

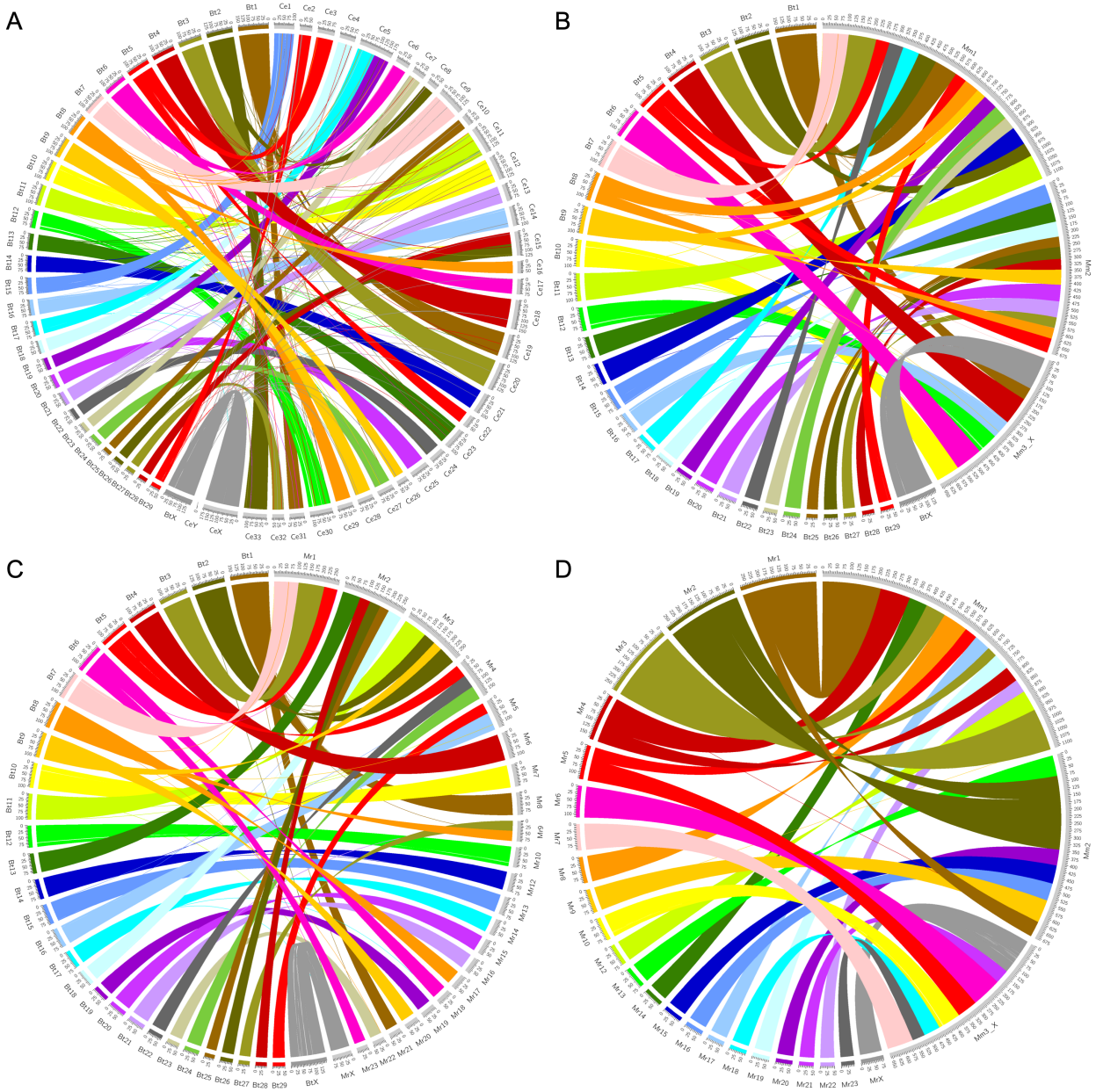


Figure 2.3. Circos (v0.69-6) [103] plots with runs of collinearity containing at least 25 kb of aligned sequence between [A] *B. taurus* (left, Bt) and *C. elaphus* (right, Ce), [B] *B. taurus* (left, Bt) and *M. muntjak* (right, Mm), [C] *B. taurus* (left, Bt) and *M. reevesi* (right, Mr), and [D] *M. reevesi* (left, Mr) and *M. muntjak* (right, Mm).

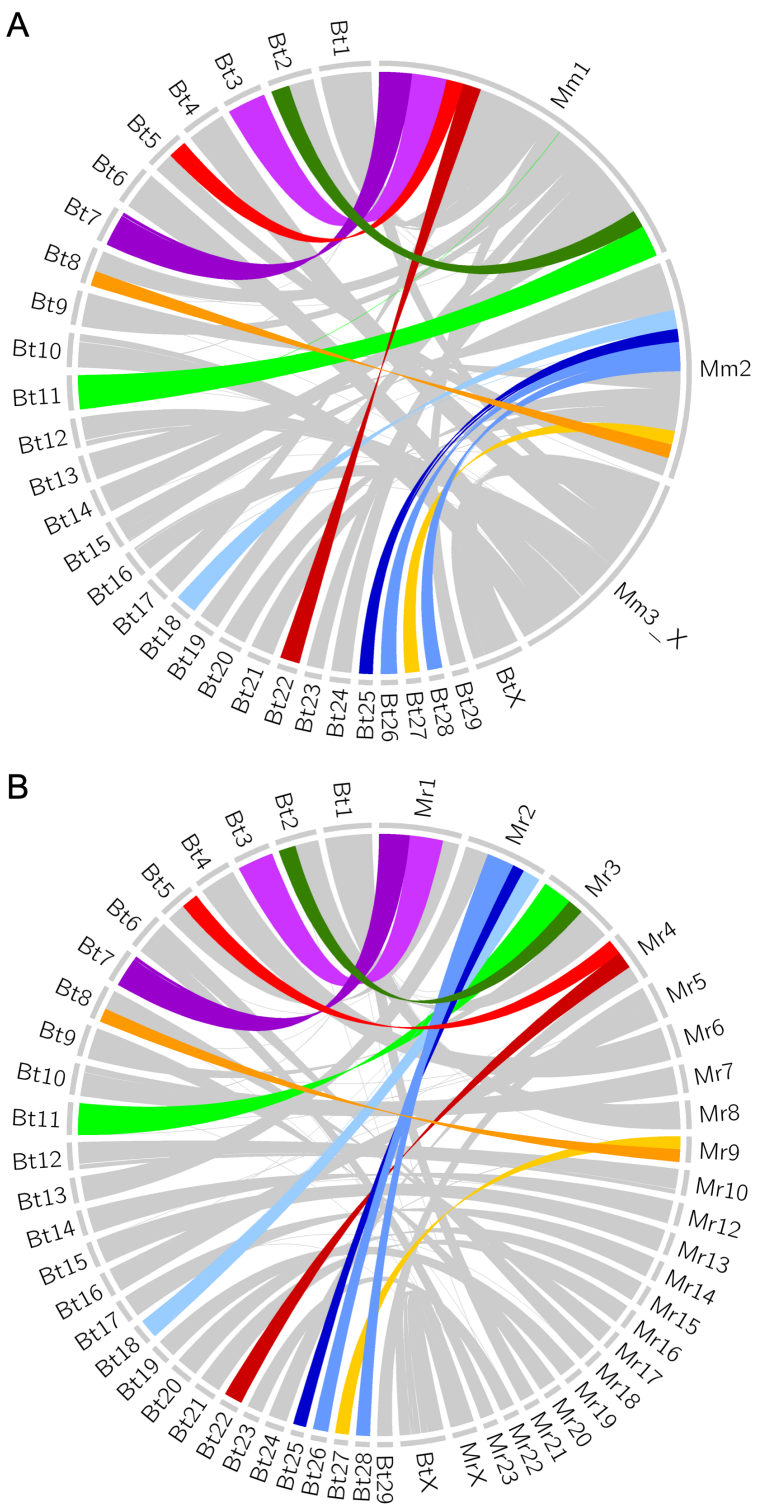


Figure 2.4. Circos (v0.69-6) [103] plots using runs of collinearity containing at least 25 kb of aligned sequence between [A] *B. taurus* (left, Bt) and *M. muntjak* (right, Mm) and [B] *B. taurus* (left, Bt) and *M. reevesi* (right, Mr) specifying the six shared muntjac fusions: BTA7/BTA3 (purple), BTA5prox/BTA22 (red), BTA2dist/BTA11 (green), BTA18/BTA25/BTA26_28 (blue), and BTA27/BTA8dist (orange).

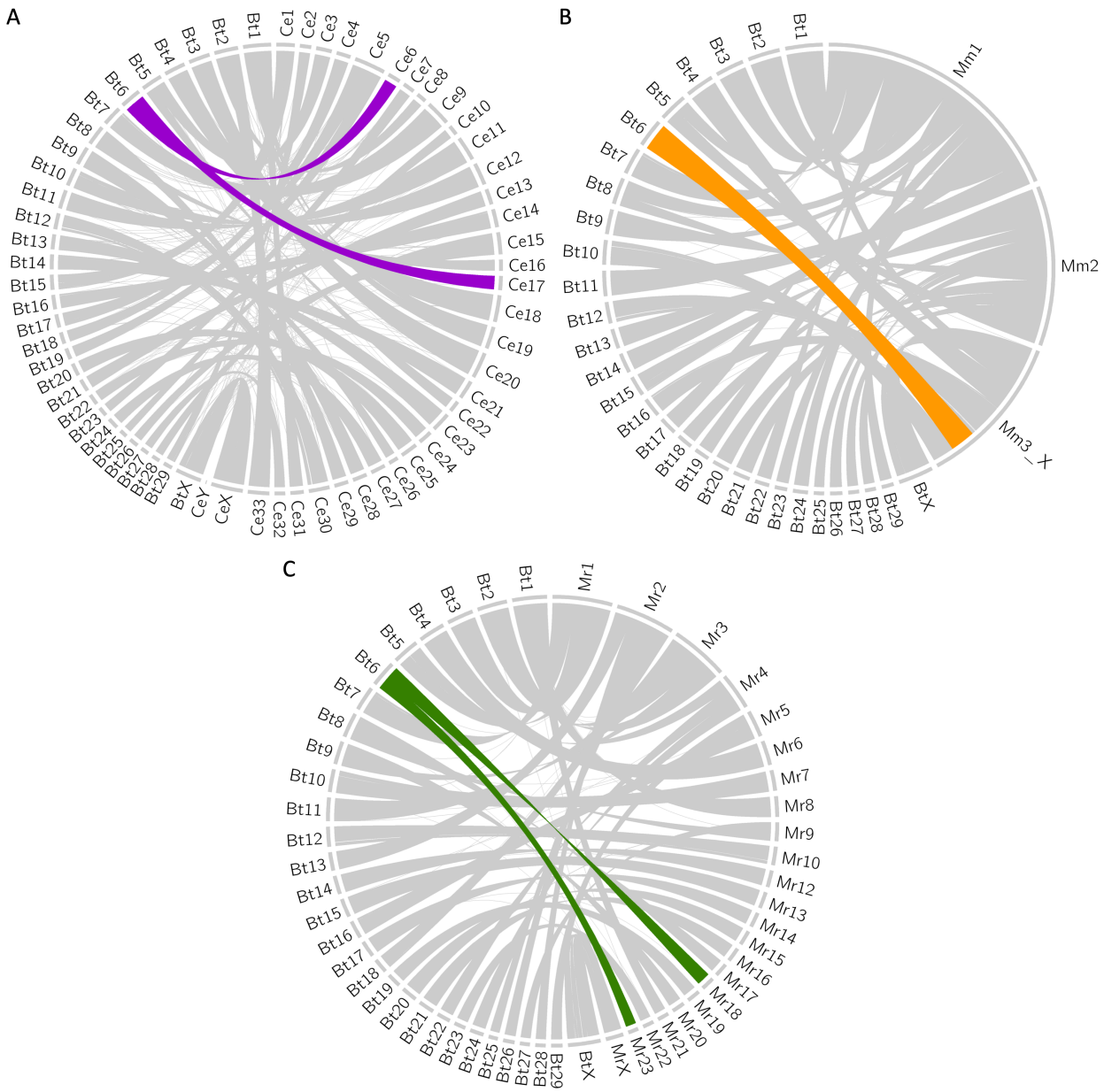


Figure 2.5. Circos (v0.69-6) [103] plots using runs of collinearity containing at least 25 kb of aligned sequence between [A] *B. taurus* (left, Bt) and *C. elaphus* (right, Ce) with the fission of BTA6 in purple; [B] *B. taurus* (left, Bt) and *M. muntjak* (right, Mm) with the fission-fusion reversal of BTA6 in orange; and [C] *B. taurus* (left, Bt) and *M. reevesi* (right, Mr) with the fission of BTA6 in green.

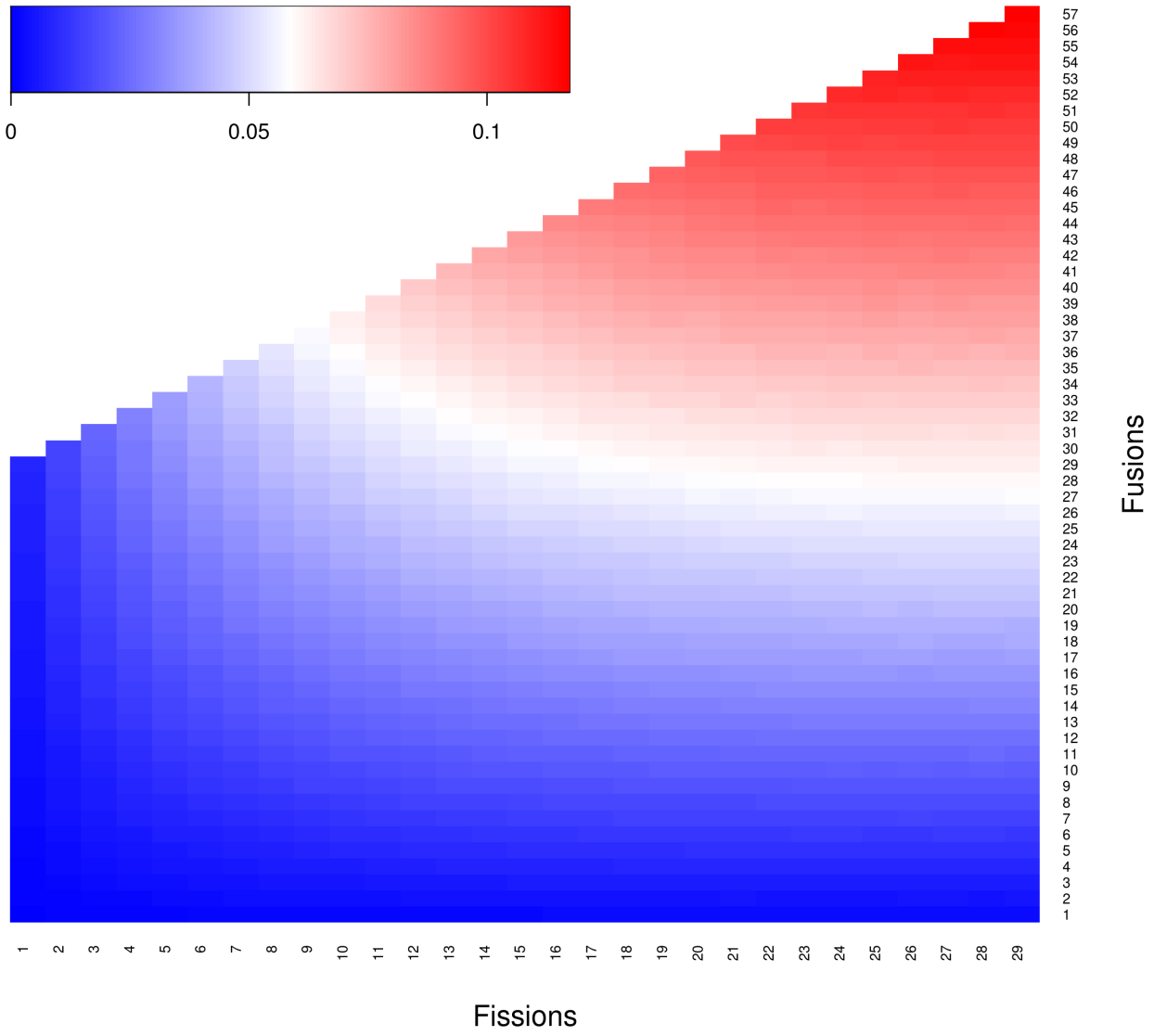


Figure 2.6. Heatmap of probabilities where at least one fusion reversed a prior fission modeled to 1 million iterations for each possible scenario from a starting karyotype of n=29, using custom script run_fission_fusion.sh (v1.0).

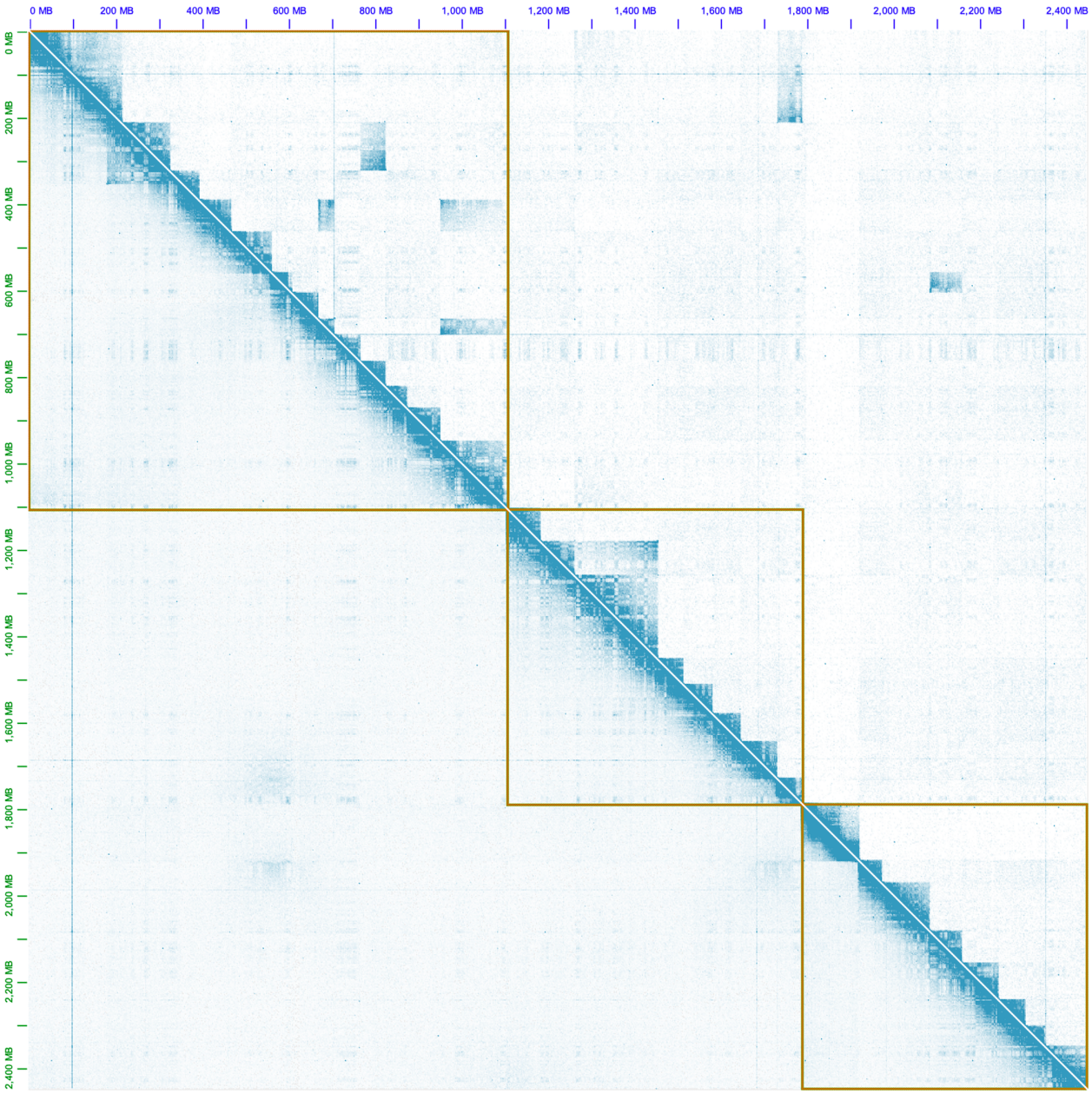


Figure 2.7. Visualization of the *M. muntjak* chromosomes' Hi-C contact map (bottom left) and the *M. reevesi* chromosomes' Hi-C contact map (top right) using the *M. muntjak* assembly as the reference in Juicebox (v1.11.08) [104]. The golden lines demarcate the boundaries of the three *M. muntjak* chromosomes.

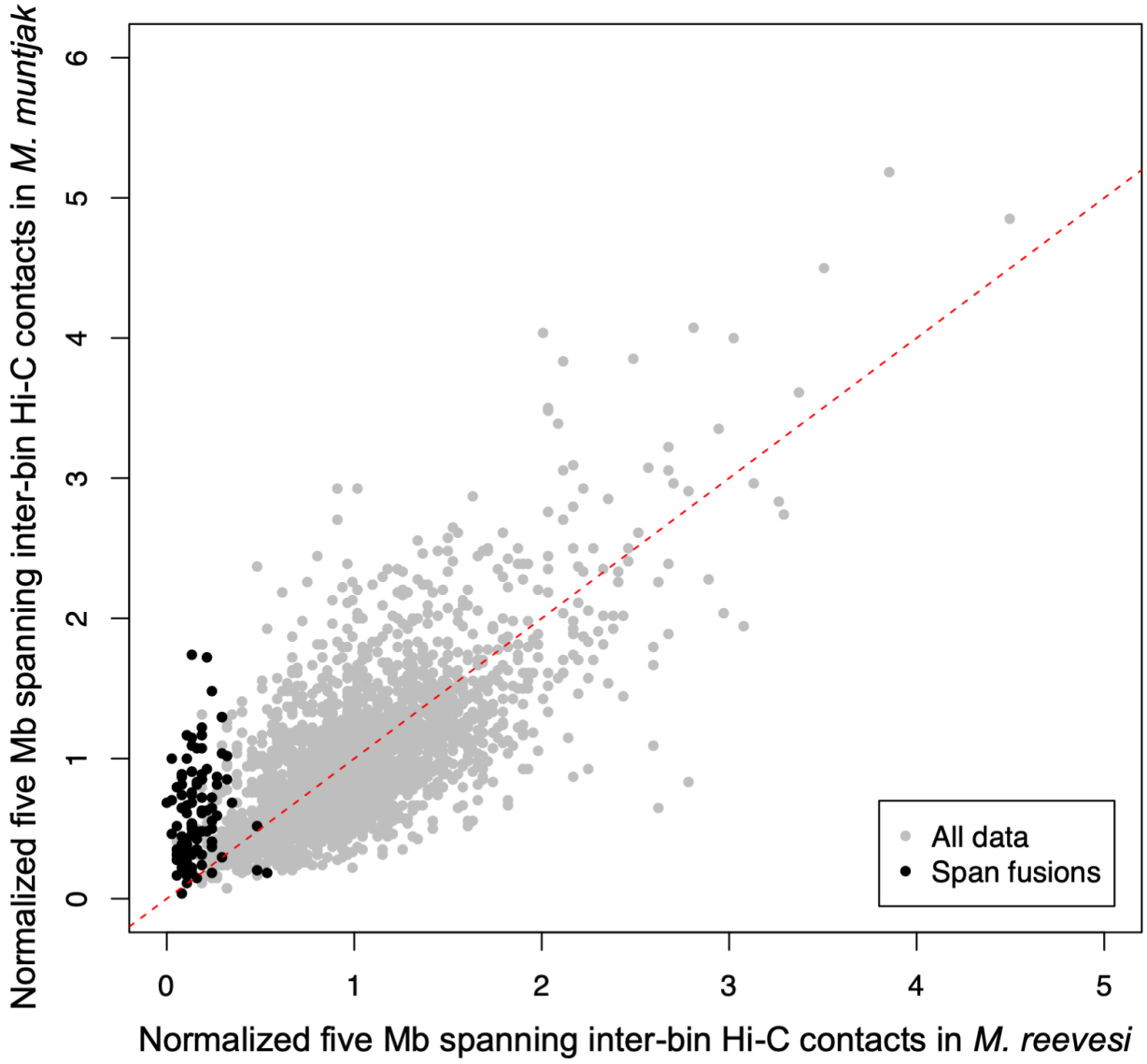


Figure 2.8. Normalized one Mb inter-bin Hi-C contacts between bins five Mb apart for *M. muntjak* (y axis) vs. *M. reevesi* (x axis) with the inter-bin contacts that span across but do not include the *M. muntjak* lineage-specific fusion sites (Table S6) colored black. The expected result of conserved Hi-C contacts was represented with a dashed red line.

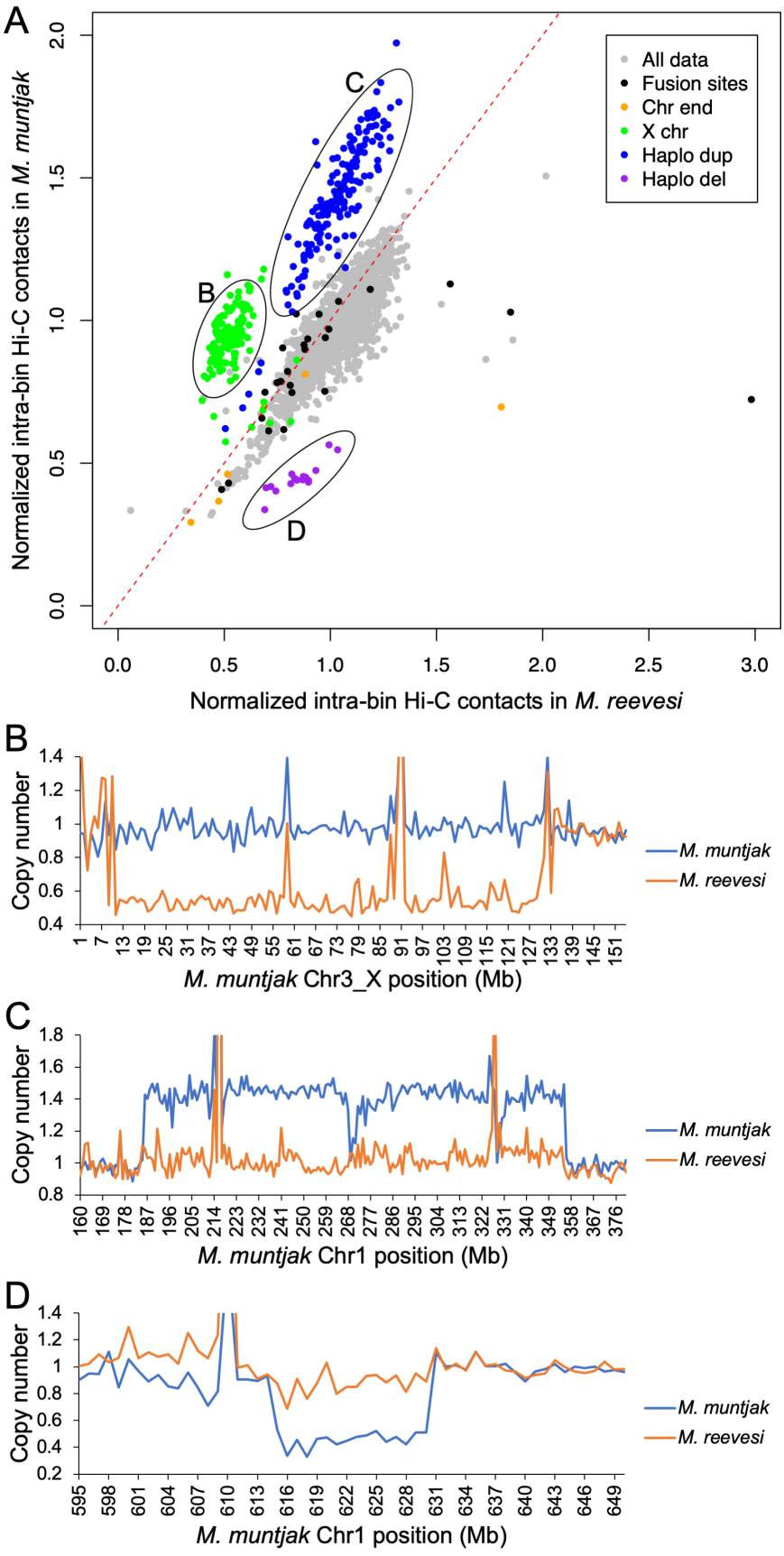

A
Normalized intra-bin Hi-C contacts in M. muntjak
Normalized intra-bin Hi-C contacts in M. reevesi
All data
Fusion sites
Chr end
X chr
Haplo dup
Haplo del
B
C
D
B
Copy number
M. muntjak Chr3_X position (Mb)
M. muntjak
M. reevesi
C
Copy number
M. muntjak Chr1 position (Mb)
M. muntjak
M. reevesi
D
Copy number
M. muntjak Chr1 position (Mb)
M. muntjak
M. reevesi

Figure 2.9. [A] Normalized one Mb intra-bin Hi-C contacts for *M. muntjak* (y axis) vs. *M. reevesi* (x axis) with the bins containing the *M. muntjak* lineage-specific fusion sites (Table S6), chromosome ends, the X chromosome, the potential *M. muntjak* haplotype-specific duplication, and the potential *M. muntjak* haplotype-specific deletion colored. The expected result of conserved Hi-C contacts was represented with a dashed red line. For fusion site ranges spanning two bins, the bin containing the majority of the fusion site range was deemed to be the fusion site bin. [B–D] Copy number was calculated from normalized coverage of adapter-trimmed 10X Genomics linked reads for three regions with variation in the chromatin contacts: [B] the X. chromosome, [C] the potential *M. muntjak* haplotype-specific duplication, and [D] the potential *M. muntjak* haplotype-specific deletion, with the copy number of *M. muntjak* in blue and *M. reevesi* in orange.

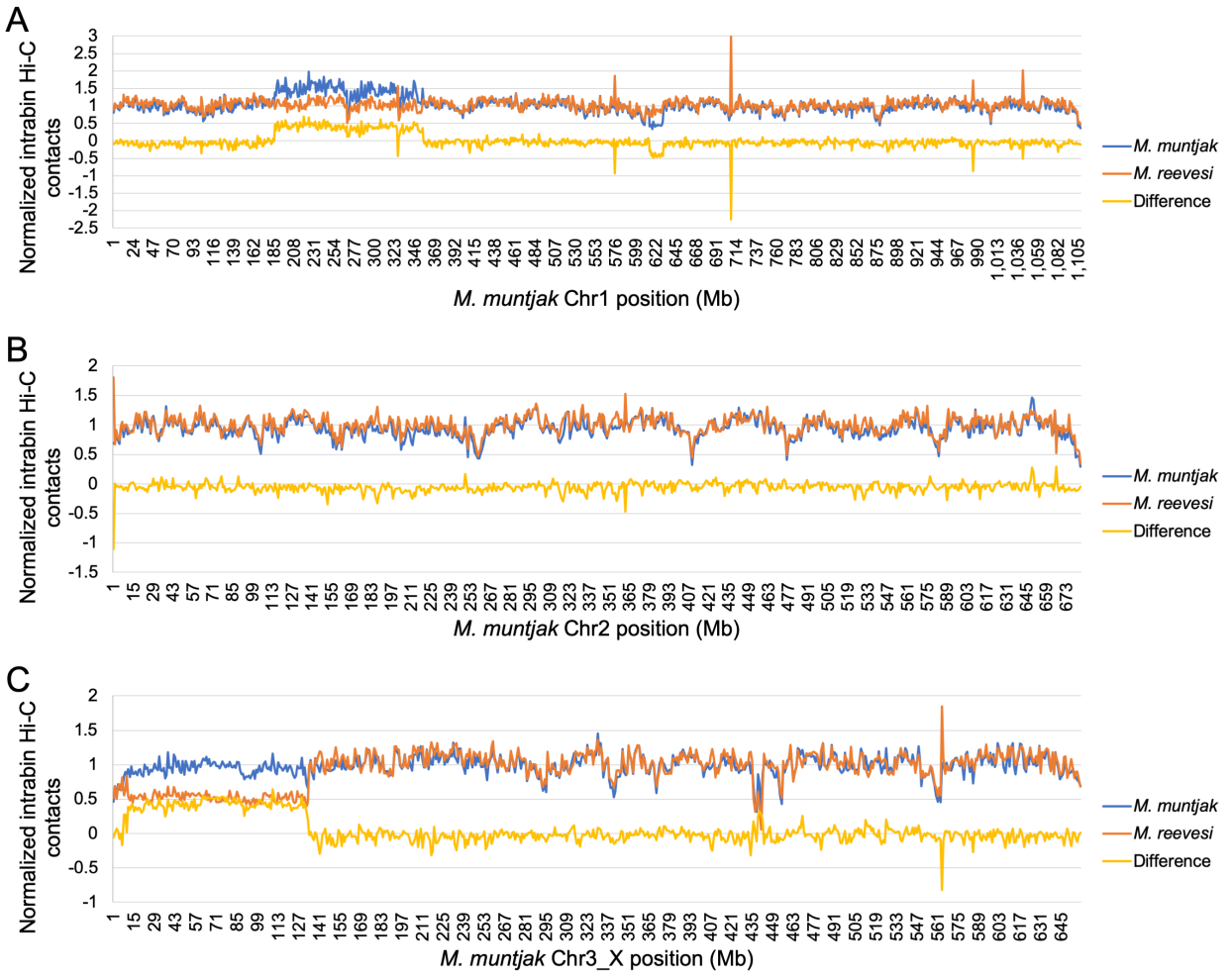


Figure 2.10. Using a bin size of one Mb and the *M. muntjak* assembly as the reference, normalized intra-bin Hi-C contacts for *M. muntjak* (blue) and *M. reevesi* (orange) at each position on [A] chromosome 1, [B] chromosome 2, and [C] chromosome 3_X. The difference of *M. muntjak* contacts minus *M. reevesi* contacts was displayed in yellow.

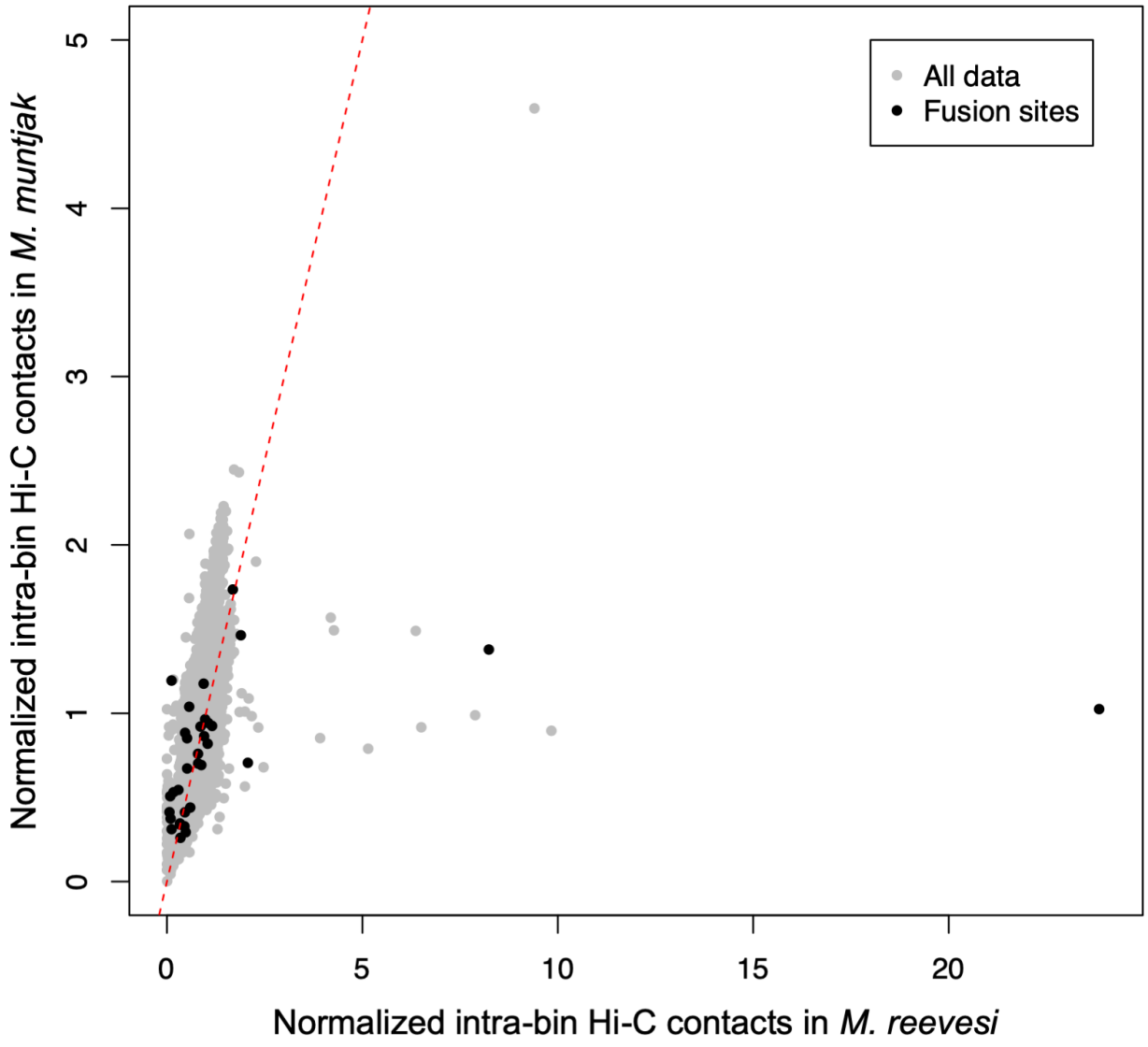


Figure 2.11. Normalized 100 kb intra-bin Hi-C contacts for *M. muntjak* (y axis) vs. *M. reevesi* (x axis) with the bins containing the *M. muntjak* lineage-specific fusion sites (Table S6) colored black. The expected result of conserved Hi-C contacts was represented with a dashed red line. For fusion site ranges spanning two bins, the bin containing the majority of the fusion site range was deemed to be the fusion site bin. For fusion site ranges spanning three or more bins, the middle 100 kb bin(s) was deemed to be the fusion site bin(s).

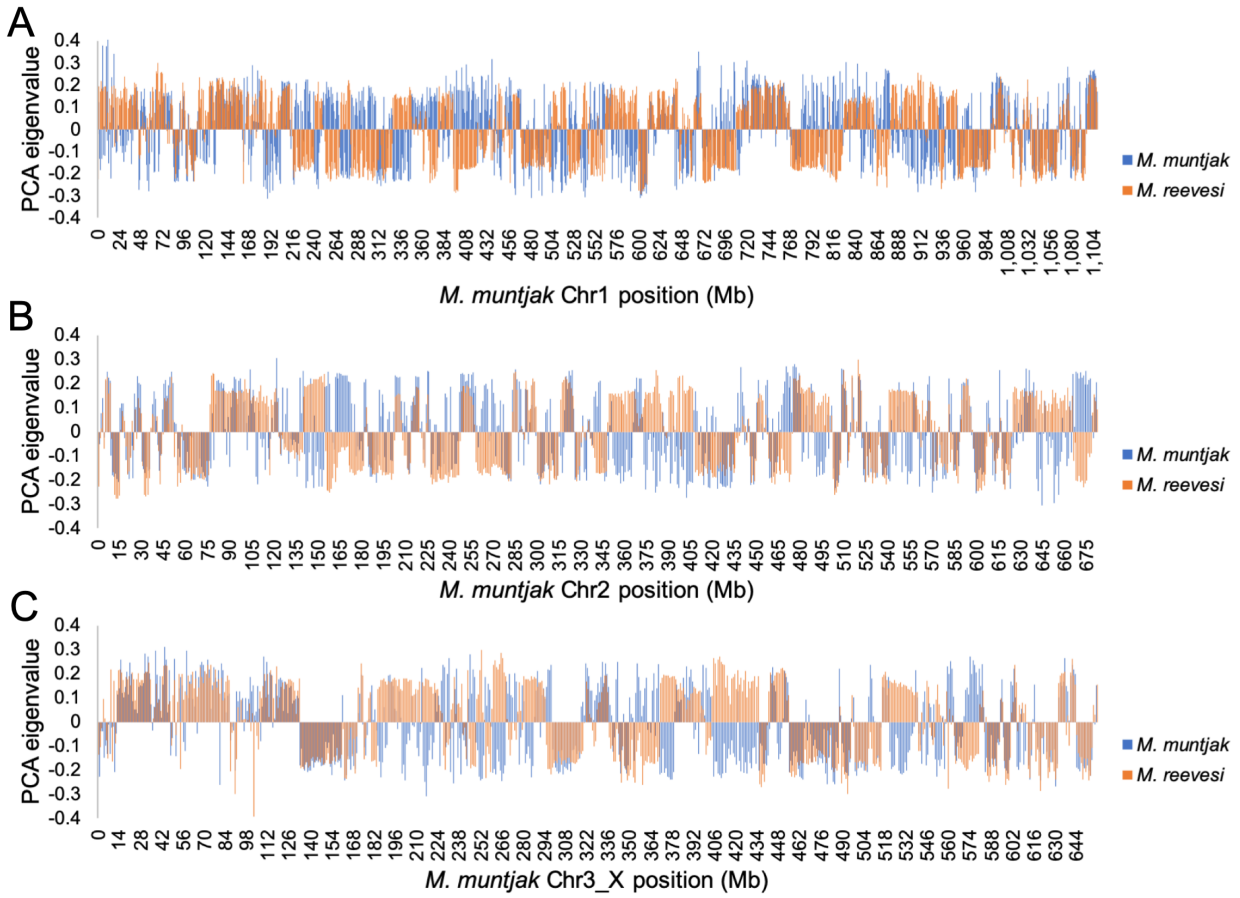


Figure 2.12. Using the *M. muntjak* assembly as the reference, identification of A/B compartment boundaries for *M. muntjak* (blue) and *M. reevesi* (orange) based on principle component analysis (PCA) eigenvalues from call-compartments.R (https://bitbucket.org/bredeson/artisanal).

Table 2.1. Summary of genome assemblies. Statistics were calculated using assemblathon_stats.pl (commit d1f044b) [105] and GenomeTools (v1.5.9) [106].

Genomic feature	*M. muntjak*	*M. reevesi*
Total scaffold length, bp	2,573,529,099	2,579,575,442
Number of scaffolds	25,651	29,705
Scaffold N50 length, bp	682,452,208	94,101,870
Total contig length, bp	2,518,738,577	2,514,747,046
Number of contigs	49,270	53,090
Contig N50 length, bp	215,534	225,142
Contigs sequence in chromosomes, %	95.06	92.93
Contig GC content, %	41.59	41.59
Masked contig repeat sequence, %	40.33	40.06
Number of genes	25,753	26,054
Genes with functional annotation, %	98.11	98.15
Average number of exons per gene	7.83	7.77

Median gene size, aa	328	326
Median exon size, bp	124	124
Median intron size, bp	921	911

Table 2.2. DNA sequencing generated for the genome assemblies, excluding index reads.

Species	Library type	Total number of reads	Total bases sequenced
M. muntjak (NCBI BioProject PRJNA542135)	10X Genomics Chromium Genome	768,921,264	115,338,189,600
	Dovetail Genomics Hi-C	521,749,568	78,784,184,768
M. reevesi (NCBI BioProject PRJNA542137)	10X Genomics Chromium Genome	696,864,964	104,529,744,600
	Dovetail Genomics Hi-C	530,002,086	80,030,314,986

Table 2.3. Pairwise nucleotide divergence in substitutions per site based on fourfold degeneracy between the examined species extracted from the RAxML (v8.2.11) [97] phylogenetic tree using Newick utilities (v1.6) [98].

Species	*B. taurus*	*C. elaphus*	*M. muntjak*	*M. reevesi*
C. elaphus	0.0550	–	–	–
M. muntjak	0.0606	0.0267	–	–
M. reevesi	0.0599	0.0259	0.0130	–
R. tarandus	0.0592	0.0298	0.0355	0.0347

Table 2.4. Locations of six cervid-specific fissions relative to the *B. taurus* genome assembly.

B. taurus chromosome	Using runs of collinearity from *C. elaphus*	Using runs of collinearity from *M. muntjak*	Using runs of collinearity from *M. reevesi*
BTA1	58,941,477 – 58,978,602	57,645,593 – 57,778,258	57,645,547 – 57,746,127
BTA2	93,282,776 – 93,424,724	79,668,309 – 79,719,766	79,668,935 – 79,719,765
BTA5	70,623,938 – 70,699,763	57,880,818 – 58,822,584	57,880,818 – 60,196,482
BTA6	63,301,740 – 63,370,450	Fusion reversal	68,435,554 – 68,455,313
BTA8	64,071,291 – 64,114,095	67,266,180 – 67,499,444	67,369,805 – 67,497,566
BTA9	63,670,677 – 64,013,115	64,824,832 – 65,087,945	64,824,832 – 65,087,945

Table 2.5. Locations of shared fusion events in the *M. muntjak* and *M. reevesi* genome assemblies using runs of collinearity from *B. taurus*.

B. taurus fused chromosomes	*M. muntjak*	*M. reevesi*
BTA7/BTA3	Chr1: 103,142,901 – 103,201,151	Chr1: 103,650,999 – 103,852,521
BTA5prox/BTA22	Chr1: 267,859,762 – 267,926,350	Chr4: 52,627,734 – 52,781,577
BTA2dist/BTA11	Chr1: 1,006,153,244 – 1,006,638,886	Chr3: 100,711,306 – 101,302,993
BTA18/BTA25	Chr2: 216,351,804 – 216,390,051	Chr2: 210,159,178 – 210,200,937
BTA25/BTA26	Chr2: 256,956,082 – 257,281,016	Chr2: 169,793,458 – 170,063,149
BTA26_28 (*B. taurus* fission)	Chr2: 305,072,202 – 305,072,202	Chr2: 121,918,267 – 122,082,397
BTA27/BTA8dist	Chr2: 580,274,743 – 582,942,905	Chr9: 40,769,993 – 43,560,918

Table 2.6. Locations of 26 unique fusion events in the *M. muntjak* genome assembly derived from one-to-one orthologs between *M. muntjak* and *M. reevesi* and then refined using runs of collinearity from *B. taurus* and *M. reevesi* against *M. muntjak*.

M. muntjak chromosome	*M. muntjak* start	*M. muntjak* end
Chromosome 1	215,667,096	215,740,550
Chromosome 1	326,596,489	326,664,606
Chromosome 1	394,376,597	394,423,120
Chromosome 1	468,279,958	468,421,169
Chromosome 1	562,054,424	562,154,407
Chromosome 1	609,147,303	609,442,186
Chromosome 1	669,798,392	669,917,570
Chromosome 1	707,332,274	707,411,696
Chromosome 1	767,481,614	767,858,594
Chromosome 1	825,563,679	825,664,460
Chromosome 1	875,352,976	875,473,556
Chromosome 1	952,277,739	952,439,995
Chromosome 2	76,554,689	76,587,068
Chromosome 2	155,096,035	155,468,627
Chromosome 2	348,208,540	348,522,144
Chromosome 2	407,305,863	407,476,405
Chromosome 2	474,891,789	475,146,554
Chromosome 2	540,052,390	540,055,842
Chromosome 2	624,503,504	624,522,918
Chromosome 3_X	133,000,163	133,001,250
Chromosome 3_X	184,069,851	184,122,591
Chromosome 3_X	295,103,485	295,251,858
Chromosome 3_X	370,301,578	370,307,164
Chromosome 3_X	454,989,747	454,992,643
Chromosome 3_X	516,012,504	516,138,154
Chromosome 3_X	562,995,659	563,046,092

Table 2.7. Locations of six unique fusion events in the *M. reevesi* genome assembly derived from one-to-one orthologs between *M. muntjak* and *M. reevesi* and then refined using runs of collinearity from *B. taurus* and *M. muntjak* against *M. reevesi*.

M. reevesi chromosome	*M. reevesi* start	*M. reevesi* end
Chromosome 1	216,558,007	216,594,231
Chromosome 2	78,441,940	78,562,328
Chromosome 3	154,290,838	154,298,872
Chromosome 3	191,988,183	192,099,419
Chromosome 4	111,704,918	111,709,617
Chromosome 5	47,104,391	47,224,935

2.4 Methods

DNA extraction and sequencing. High molecular weight DNA was extracted, as previously described [107], from fibroblast cell lines obtained from the University of Texas Southwestern Medical Center for *M. muntjak* (female) [93] and the University of Cambridge for *M. reevesi* (male) [71]. A 10X Genomics Chromium Genome library [108] was prepared for each species by the DNA Technologies and Expression Analysis Cores at the University of California Davis Genome Center and sequenced on the Illumina HiSeq X by Novogene Corporation. A Hi-C chromatin conformation capture library was also prepared for each species using the Dovetail Genomics Hi-C library preparation kit and sequenced on the Illumina HiSeq 4000 by the Vincent J. Coates Genomics Sequencing Laboratory at the University of California Berkeley.

Shotgun assembly. 10X Genomics linked reads were assembled with Supernova (v2.0.1) [108]. Putative archaeal, bacterial, viral, and vector contamination was identified and removed by querying the assemblies using BLAST+ (v2.6.0) [109] against the respective RefSeq and UniVec databases and removing sequences with at least 95% identity, E-value less than 1E-10, and hits aligning to more than half the scaffold size or 200 bases, using custom script general_decon.sh (v1.0). Putative mitochondrial sequence was also identified and removed by querying the assemblies using BLAST+ (v2.6.0) [109] against their respective mitochondrial assemblies (NCBI NC_004563.1 [110] and NC_004069.1 [111]) and removing sequences with at least 99% identity and E-value less than 1E-10, using custom script mt_decon.sh (v1.0). The decontamination removed 71 scaffolds totaling 836 kb from the *M. muntjak* assembly and 36 scaffolds totaling nine kb from the *M. reevesi* assembly.

Chromosome assembly. Hi-C reads were aligned to each assembly with Juicer (commit d3ee11b) [112]. A preliminary round of Hi-C-based scaffolding was performed with 3D-DNA (commit 745779b) [113], and residual redundancy due to split haplotypes was manually filtered through visualization of the Hi-C contact map in Juicebox (v1.9.0) [104], removing the smaller of any pair of duplicate scaffolds. This process removed 1.04 Gb of sequence from the *M. muntjak* assembly and 25 Mb of sequence from the *M. reevesi* assembly. The remaining scaffolds were organized into chromosomes by realigning the Hi-C reads to the deduplicated assembly with Juicer (commit d3ee11b) [112], ordering and orienting scaffolds into chromosomes with 3D-DNA (commit 745779b) [113], and then manually correcting using Juicebox (v1.9.0) [104]. After correction, gaps in the assembly were filled with adapter-trimmed 10X Genomics data using custom script trim_10X.py (v1.0) and Platanus (v1.2.1) [114].

Final assembly release and validation. Scaffolds smaller than one kb in the gap-filled assembly were removed with seqtk seq (v1.3-r106; https://github.com/lh3/seqtk), and chromosomes and scaffolds were numbered in order of size using SeqKit (v0.7.2-dev) [115]. X chromosomes were later renamed based on alignment with *B. taurus* [75]. Chromosomes in both species were oriented arbitrarily. For *M. reevesi*, the chromosome numbering in the assembly may differ from prior BAC-based studies. As *B. taurus* chromosome numbering is universally recognized, the extensive genomic collinearity of cervids, including both muntjacs, with cow provides a standard method for referencing homologous segments.

To validate the *M. muntjak* assembly, sequenced BACs [82–84] were aligned with BWA (v0.7.17-r1188) [116], and primary alignments were checked against the corresponding FISH locations, excluding unaligned BACs or those aligned to unplaced scaffolds.

Annotation and homology analysis. Repetitive elements were identified and classified with RepeatModeler (v1.0.11) [117] and combined for each species with ancestral Cetartiodactyla repeats from RepBase (downloaded November 8, 2018) [118]. The assemblies were then soft masked with RepeatMasker (v4.0.7) [119]. The assemblies were annotated using Gene Model Mapper (v1.5.3) [120] and BLAST+ (v2.6.0) [109] with the following assemblies and annotations from Ensembl release 94 [99] as input evidence: *B. taurus* (September 2011 genebuild of GCA_000003055.3) [75], *H. sapiens* (July 2018 genebuild of GCA_000001405.27) [86], and *O. aries* (May 2015 genebuild of GCA_000298735.1) [85]. Coding nucleotide and peptide sequences were extracted using gff3ToGenePred and genePredToProt from the UCSC Genomics Institute (binaries downloaded March 5, 2019) [121] with custom script postGeMoMa.py (v1.0), and functional annotation was run with InterProScan (v5.34-73.0) [87].

Pairwise gene homology of the two muntjac annotations as well as total gene homology of the two muntjac, *B. taurus* (Ensembl release 94 September 2011 genebuild of GCA_000003055.3) [75,99], *C. elaphus* (publication genebuild of GCA_002197005.1) [76], and *R. tarandus* [77,100] annotations were analyzed with OrthoVenn [101] using an E-value of 1E-5 and an inflation value of 1.5.

Comparative analysis. The two muntjac assemblies were aligned to each other with cactus (commit e4d0859) [122]. After removing any ambiguous sequence with seqtk randbase (v1.3-r106; https://github.com/lh3/seqtk), the muntjac assemblies, *C. elaphus* (GCA_002197005.1) [76], and *R. tarandus* [77,100] were each aligned pairwise against *B. taurus* (GCA_000003055.3) [75] with cactus (commit e4d0859) [122]. Using custom script cactus_filter.py (v1.0), all pairwise output HAL alignment files were converted into PSL format with halLiftover (commit f7287c8) [123]. Using tools from the UCSC Genomics Institute (binaries downloaded March 5, 2019) [121] unless noted otherwise, the PSL files were filtered and converted with pslMap, axtChain, chainPreNet, chainCleaner (commit aacca59) [124], chainNet, netSyntenic, netToAxt, axtSort, and axtToMaf. Runs of collinearity were extracted from each pairwise MAF file by linking together local alignment blocks where the locations of species one and species two, correspondingly, were in the same orientation and were neighboring in their respective genomes without intervening aligned sequence from elsewhere in the genomes. The pairwise MAF files from the alignments against *B. taurus* were also merged with ROAST/MULTIZ (v012109) [125], using the phylogenetic topology extracted with Newick utilities (v1.6) [98] from a consensus tree of the species from 10kTrees [126], and sorted with last (v912) [127].

Phylogeny. From the one-to-one orthologous genes of all five species identified by OrthoVenn, codons with potential fourfold degeneracy were extracted from the *B. taurus* Ensembl release 94 September 2011 genebuild, excluding codons spanning introns, using custom script 4Dextract.py (v1.0). Using the ROAST-merged MAF file with *B. taurus* as reference, the corresponding codons were identified in the other four species, checking for amino acid conservation and excluding any codons that span two alignment blocks in the MAF file. The output FASTA file containing fourfold degenerate bases was converted into PHYLIP format with BeforePhylo

(commit 0885849; https://github.com/qiyunzhu/BeforePhylo) and then analyzed with RAxML (v8.2.11) [97] using the GTR+Gamma model of substitution with outgroup *B. taurus*. As previously described [128], the divergence time confidence intervals from TimeTree [129] for all nodes except the outgroup *B. taurus* node were input into MEGA7 (v7.0.26) [130], using the Reltime method [131] and the GTR+Gamma model to create a time tree.

Chromatin conformation analysis. Hi-C reads from both species were aligned to the *M. muntjak* assembly with Juicer (commit d3ee11b) [112], and KR normalized intra-chromosome Hi-C contact matrices were extracted with Juicer Tools (commit d3ee11b) [112] at one Mb resolution. A sliding window-based localized PCA was used to call A/B compartment structure with custom script call-compartments.R (https://bitbucket.org/bredeson/artisanal). Localization of PCA 1 along the diagonal of the Pearson correlation matrix (40 windows of one Mb each with a step size of 20) amplifies compartment signal and mitigates confounding signal from large-scale, intra-chromosome and inter-arm contacts.

Hi-C contacts from the Juicer (commit d3ee11b) [112] merged_nodups.txt output file were split into one Mb and 100 kb bins using custom scripts HiCbins_1Mb.py and HiCbins_100kb.py, respectively. Intra-bin and inter-bin Hi-C contacts were extracted and normalized based on the average number of contacts per bin for each species.

Copy number analysis. To explore the three regions with variation in chromatin contacts, adapter-trimmed 10X Genomics data for each species was aligned to the *M. muntjak* assembly with BWA (v0.7.17-r1188) [116]. Alignment depth was extracted with SAMtools (v1.6) [132], and copy number was calculated from the average alignment depth per one Mb bin for each species.

2.5 Notes

Availability of data and materials. The assemblies, annotations, and raw data for *M. muntjak* and *M. reevesi* were deposited at NCBI under BioProjects PRJNA542135 and PRJNA542137, respectively. Supporting files for the repeat and gene annotations are available at https://doi.org/10.6078/D1KT16. Unless stated otherwise, custom code used in this study is available at https://github.com/abmudd/Assembly.

Acknowledgements. We thank Jerry Shay and Woody Wright for providing the *M. muntjak* cell line; Malcolm Ferguson-Smith and Fengtang Yang for providing the *M. reevesi* cell line; Karen Lundy and the Functional Genomics Laboratory at the University of California Berkeley for running quality control on the extracted DNA; Dovetail Genomics for providing the Hi-C library preparation kit and running quality control on the Hi-C libraries; Shana McDevitt and the Vincent J. Coates Genomics Sequencing Laboratory at the University of California Berkeley for sequencing the Hi-C libraries; Jessica Lyons for coordinating the preparation and sequencing of the *M. muntjak* 10X Genomics library; Diana Burkart-Waco and the DNA Technologies and Expression Analysis Cores at the University of California Davis Genome Center for preparing the 10X Genomics libraries; Novogene Corporation for sequencing the 10X Genomics libraries; and Gary Karpen for providing comments on the manuscript.

3 Torpid chromosome evolution in frogs

This chapter is an update of an unpublished manuscript, of which I am a co-first author. This work is included with permission from the authors below.

Authors' contributions. Austin B. Mudd extracted DNA and RNA, prepared RNA libraries for *E. pustulosus*, assembled and annotated the genomes, completed the bioinformatic analyses, and wrote the manuscript. Jessen V. Bredeson provided the *X. tropicalis* genome assembly, identified the pericentromeric regions based on Hi-C, and supported the bioinformatic analyses. Sofia Medina Ruiz provided the *X. tropicalis* genome annotation, curated the repeat library, and completed the repeat analysis. Kelly E. Miller completed the metaphase chromosome spreads in *H. boettgeri* and extracted *H. boettgeri* RNA for sequencing. Dirk Hockemeyer prepared Hi-C libraries. Richard M. Harland and Daniel S. Rokhsar provided scientific leadership of the project and wrote the manuscript.

3.1 Background

For the past century, Drosophila geneticists have explored the content and organization of fruit fly genomes [133]. The identification of Muller elements, or means of classifying chromosome arms in *Drosophila melanogaster*, gave rise to the study of synteny [11]. The categorization of Muller elements and the corresponding concept of syntenic block conservation has been analyzed in other fruit flies [10,134,135] and extended across the order Diptera and beyond [136,137]. In fact, Sved et al. [135] hypothesized that this conserved syntenic relationship and karyotype stability found in Diptera was unique and may be derived from facets of chromosome organization and structure, with particular emphasis on *D. melanogaster*'s lack of telomerase and inclusion of telomeric retrotransposons [138]. Although this hypothesis [135] was driven by comparison against taxa with large chromosome variation, such as mammals [139,140], estrildid finches [141], and cichlid fish [142], some vertebrate taxa with torpid karyotype evolution, also known as karyotypic conservatism [60] or chromosomal bradytely [61], have been described, such as cetaceans [143] and frogs [144]. This, therefore, questions whether the presence of long-range or even chromosome-scale conserved syntenic blocks, similar to Muller elements, can be found in species with decreased inter-chromosome variation.

Given the evolutionary significance of finding conserved synteny across disparate branches of the tree of life, we set out to explore karyotype relationships within the frog lineage, classify inter-chromosome variations, and discern syntenic blocks as well as the ancestral chromosome state. To this end, we leveraged new chromosome-scale assemblies of *Eleutherodactylus coqui*, *Engystomops pustulosus*, and *Hymenochirus boettgeri*; improved assemblies of *Leptobrachium (Vibrissaphora) ailaonicum* [145], *Pyxicephalus adspersus* [146], and *Xenopus tropicalis*; and published assemblies of *Ambystoma mexicanum* [147,148] and *Xenopus laevis* [149]. In total, we identified 17 fission, fusion, inter-chromosome translocation, and duplication events, which, based on the total analyzed branch length of 1.05 billion years, resulted in a rate of a single karyotype change every 62 million years. Given the limited chromosome variation, we also

characterized 13 ancestral chromosomes and found 95% of chromosome one-to-one gene orthologs maintained in the same syntenic block.

3.2 Results and discussion

Assembly. To address the nature of karyotype evolution in frogs, we generated high-quality, chromosome-scale genome assemblies for three new frogs: the Zaire dwarf clawed frog *H. boettgeri*, a member of the family Pipidae along with *Xenopus* spp., as well as the common coquí *E. coqui* and the túngara frog *E. pustulosus*, two neobatrachians from the families Eleutherodactylidae and Leptodactylidae, respectively (Table 3.1). These chromosome-scale assemblies were produced with chromatin conformation capture (Dovetail Genomics Hi-C) along with combinations of short reads (short-insert paired-ends and mate pairs), linked reads (10X Genomics Chromium Genome), and long reads (Pacific Biosciences SMRT; Figure 3.1, Table 3.2, Methods).

The assembled chromosome numbers recapitulated the karyotypes described in the literature (2n=26 for *E. coqui* [150] and 2n=22 for *E. pustulosus* [151]). Although the literature for *H. boettgeri* reported a karyotype of 2n=20–24 [152,153], our assembly contained fewer chromosomes, resulting in a karyotype of 2n=18. We confirmed the accuracy of our assembly's karyotype by performing chromosome spreads (n=75) from *H. boettgeri* samples, which exhibited both a mode and an average of 2n=18 (Figure 3.2). This karyotype discrepancy with the published literature may have resulted from cryptic sub-populations within *H. boettgeri*.

To augment our newly sequenced frog genomes and expand the phylogenetic scope of our study, we also included revised assemblies of two recently published frogs genomes in our analyses: the African bullfrog *P. adspersus*, a neobatrachian from the family Pyxicephalidae [146], and the Yunnan moustache toad *L. ailaonicum* from the family Megophryidae [145]. The assemblies were revised based on manual curation of Hi-C data using Juicebox (v1.9.0) [104], resulting in local and global rearrangements of the published assemblies (Methods). Notably, these curated revisions were carried out independently for each genome, without presuming conserved synteny with the other frogs.

Annotation and homology. For each species except *L. ailaonicum*, we annotated protein-coding genes based on RNA sequencing (Tables 3.2 and 3.3) and peptide homology with *X. tropicalis* (Methods), with the majority of resulting genes annotated by homology and/or domain content (Table 3.4). Gene set comparisons (Figure 3.3A–B) showed that the vast majority of protein-coding genes were conserved among all five frogs, supporting the completeness and accuracy of these assemblies in genic regions, although individual species differed in their completeness with regard to this metric. As expected from the inclusion of deep long-read sequencing and more manual review, *X. tropicalis* had the fewest missing gene clusters of the five species. Only 72 OrthoVenn2 [102] clusters were present in all of the other annotations but missing in *X. tropicalis*, supporting the completeness of the *X. tropicalis* assembly and annotation.

Comparing these gene sets, we identified 9,624 one-to-one gene orthologs (Figure 3.3C) between *X. tropicalis*, our three new frog assemblies, and a reannotation of *P. adspersus* (Methods), with

the additional requirement that an ortholog be either present in a single copy or absent in *L. ailaonicum* [145,154] and the L and S subgenomes of allotetraploid *X. laevis* (v9) [149]. The special treatment of these three (sub)genomes was due to the lower quality of the *L. ailaonicum* annotation and the known propensity for evolutionary gene loss in paleotetraploids. For analyses of synteny, we further restricted our attention to the 7,292 one-to-one gene orthologs that were only present on chromosomes, as opposed to unlinked scaffolds, in the core genomes *E. coqui*, *E. pustulosus*, *H. boettgeri*, *P. adspersus*, and *X. tropicalis* (Figure 3.4).

Synteny and ancestral state. Despite divergence times of up to 200 million years, the six frog genomes were readily aligned to *X. tropicalis*. We also aligned the greatly expanded and more deeply diverged axolotl *A. mexicanum* assembly (GCA_002915635.2) [147,148] to *X. tropicalis*. These sequence alignments were merged into runs of collinearity (Figures 3.5 and 3.6).

Analysis of these alignments revealed the timing of fission and fusion events in each lineage as well as the ancestral karyotype states (Figure 3.4B). From this, we identified the presence of 13 syntenic blocks, which we classified as elements A to M, in the examined frog species. Over 95% (6,952 of 7,292) of chromosome one-to-one gene orthologs were maintained in the same element among the five main frog species, similar to the 95% ortholog localization to Muller elements in *Drosophila* spp. (Figure 3.4B) [10], despite a total nucleotide divergence based on fourfold degeneracy between all examined frog species of 2.58 substitutions per site (Table 3.5). We noted no significant size variation between these elements or deviation in gene or repeat count across these five species (Figure 3.7; Table 3.6). We found that element A had the largest gene count in all examined species and that element H had the smallest gene count (Figure 3.7B). Although the number and boundaries of these elements may change as more chromosome-scale frog assemblies are analyzed in the future [155], this finding suggested a principle of long-range, and even chromosome-scale, synteny across multiple branches of the tree of life and could enable future chromosome naming in frogs based on these elements, similar to Muller elements in *Drosophila* spp. [10].

We also found that these 13 elements reflected the 13 chromosomes in the pipanuran ancestor, defined in the sense of Cannatella and Ford [156], which agreed with the prior assertion of 26 basal chromosomes in frogs [157]. This ancestral karyotype, however, may not apply to species phylogenetically basal to the analyzed species, such as those in superfamilies Leiopelmatoidea and Discoglossoidea. This delineation of ancestral karyotype may have phylogenetic relevance, given the historical use of karyotype to differentiate suborders Neobatrachia and Archaeobatrachia [157]. In order to identify the ancestral karyotype of all frogs, more data, particularly from superfamilies Leiopelmatoidea and Discoglossoidea, is needed.

Chromosome evolution. Using the runs of collinearity (Figures 3.5 and 3.6), chromosome one-to-one gene orthologs (Figure 3.4B), and Hi-C-based estimates of the pericentromeric locations (Methods), we assessed chromosome evolution and conservation between the aforementioned frog assemblies. For convenience, we referred to chromosome regions by their ancestral chromosome identifiers. Overall, we found broad pericentromeric conservation between the identified species (Figures 3.4B and 3.8), with some movement of the pericentromeric region across the frog lineage. Using the repeat landscape as a proxy for the state of the historic sequence, we identified several examples of centromere inactivation resulting from likely

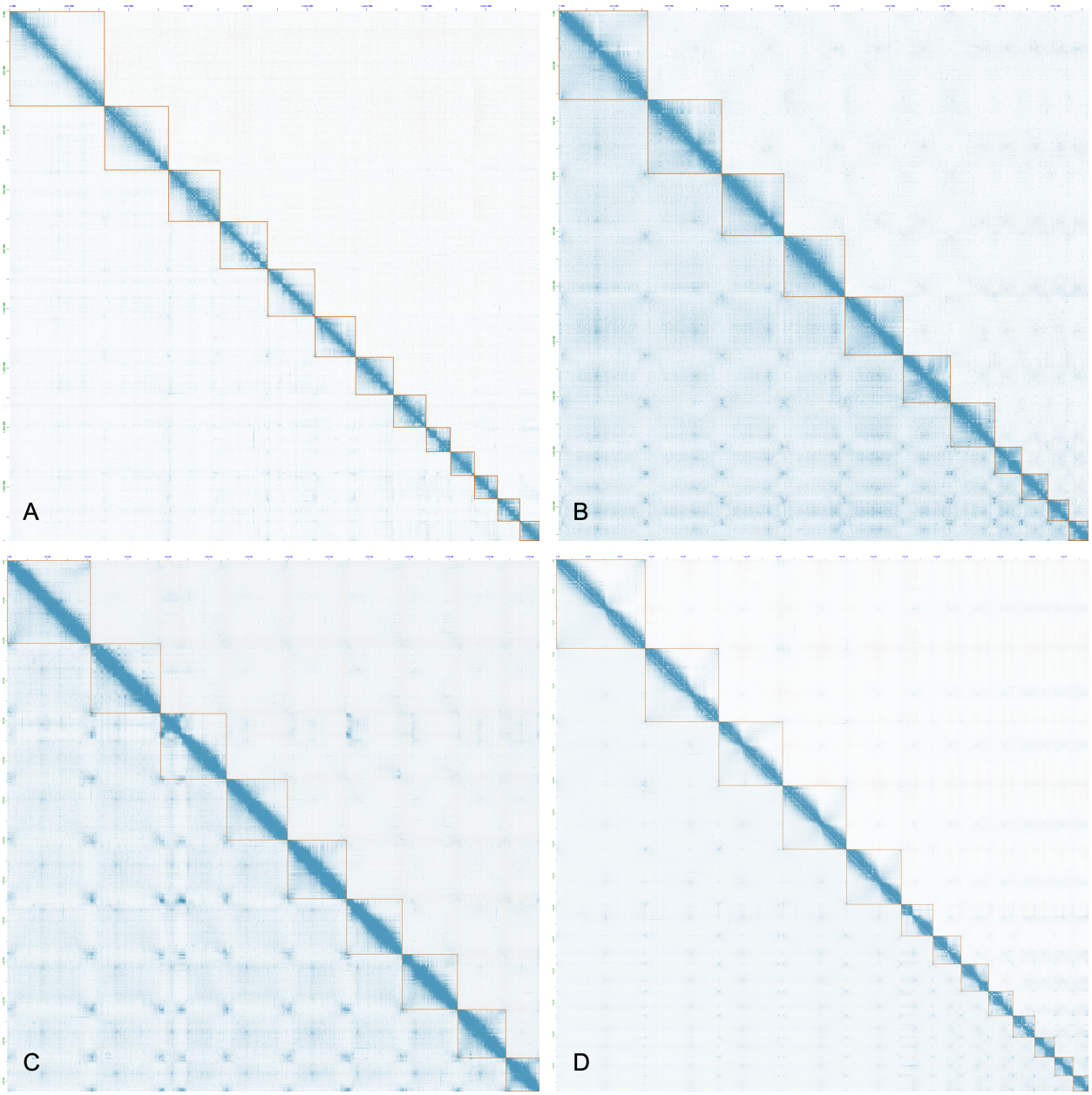
ancestral end-to-end fusions (Figures 3.4B and 3.8B). Further exploring these fission and fusion events, in addition to the previously reported end-to-end fusion of L/M and the whole-genome duplication in *X. laevis* [149], we identified a potential reciprocal chromosome arm translocation of A/M in *P. adspersus*, a potential Robertsonian fusion of I/M and an end-to-end fusion of D/K in *E. pustulosus*, and an end-to-end fusion of J_K/M in *H. boettgeri*. In *E. coqui*, we found two Robertsonian fissions of A and G, a potentially 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.

In the pipid lineage, we also found three ancestral fusions: two end-to-end fusions of D/E and J/K as well as a potentially Robertsonian fusion of H/I. We corroborated the separation of ancestral chromosomes D/E, H/I, and J/K in *Ascaphus truei* (unpublished). We additionally analyzed the state of these three ancestral chromosomes in *A. mexicanum* (Figure 3.5A), which supported the separation of J/K but did not support the separation of D/E and H/I. Our hypothesis that the three fusions occurred in the pipid common ancestor instead of fissions in the acosmanuran ancestor will be further tested with future chromosome-scale genome assemblies of frogs phylogenetically basal to the analyzed species [155]. Nonetheless, maintenance of the described collinearity must have biological significance, such as the role of recombination, chromosome territories, and aspects of three-dimensional chromosome structures like Rabl. Divergence from collinearity in frogs, much like in *Drosophila* spp. [11], may be nonrandom.

The long-range and, in most cases, chromosome-scale collinearity among the examined frog species, despite a total branch length of 1.05 billion years (Table 3.7), paralleled the synteny observed in birds [158] and reptiles [159] and differed from the significant chromosome variation found in mammals [139,140]. Assuming the rearrangement in *E. coqui* resulted from two Robertsonian fissions of B and F, followed by three Robertsonian fusions, we calculated a total of 17 fission, fusion, translocation, and duplication events, excluding intra-chromosome variations, resulting in a karyotype change every 62 million years (Figure 3.4B). This rate was slightly faster than the chromosome number change every 70 to 90 million years reported in literature [144,160] but still slower than karyotype change rates for mammals [78] and many reptiles [161]. Of course, the present rate calculation was limited, based on only seven species, and the rate may vary depending on the analyzed species. Some frog taxa will have a higher rate of karyotype change, such as *Eleutherodactylus* spp. (2n=16–32) or *Pristimantis* spp. (2n=22–38) [150]. On the other hand, some species may have had no significant inter-chromosome changes over the past 205 million years, such as *Rana temporaria* [162] or *L. ailaonicum* (Figure 3.4B). Nonetheless, this analysis of chromosome variation across the frog lineage suggested a slow rate of karyotype evolution.

Conclusions. Analysis of these new chromosome-scale assemblies identified the presence of syntenic blocks in frogs, denoting the existence of chromosome stability in multiple branches of the tree of life and the broader pervasiveness of this mechanism of conservation. Furthermore, this study examined inter-chromosome variation between the assemblies, calculating a rate of 0.016 karyotype changes per million years across the assessed frog lineage, despite a total nucleotide divergence based on fourfold degeneracy of 2.58 substitutions per site. Future studies will explore the presence of syntenic elements in frogs and other species with torpid karyotype evolution and use these assemblies in targeted, species-specific analyses.

3.3 Figures and tables



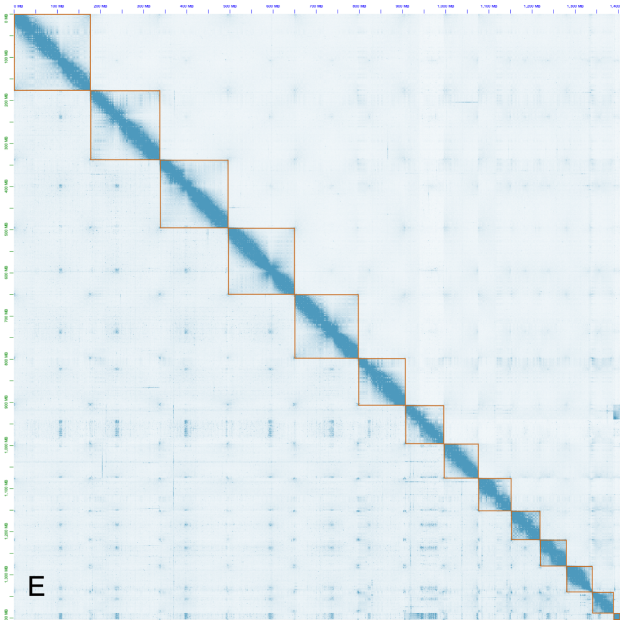


Figure 3.1. Visualization of chromosomes' Hi-C contact maps for [A] *E. coqui*, [B] *E. pustulosus*, [C] *H. boettgeri*, [D] *L. ailaonicum*, and [E] *P. adspersus* in Juicebox (v1.11.08) [104] with a minimum MAPping Quality (MAPQ) of zero or greater below the diagonal and MAPQ ≥ 30 above the diagonal. The golden lines demarcate the boundaries of the chromosomes.

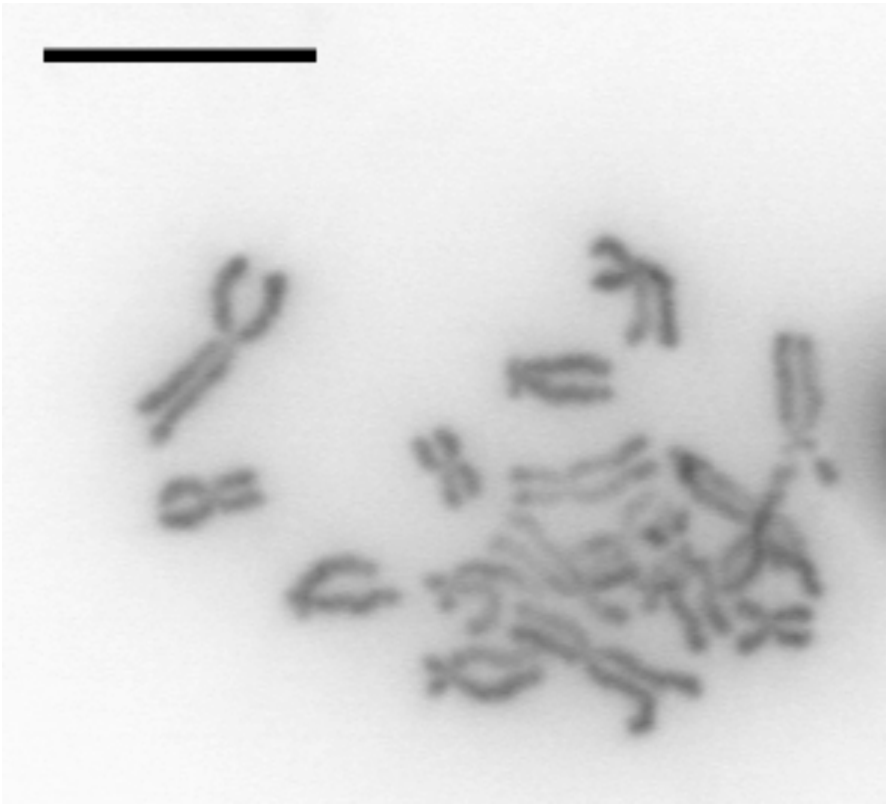


Figure 3.2. Example *H. boettgeri* metaphase chromosome spread with 18 chromosomes (scale bar = 10 μm).

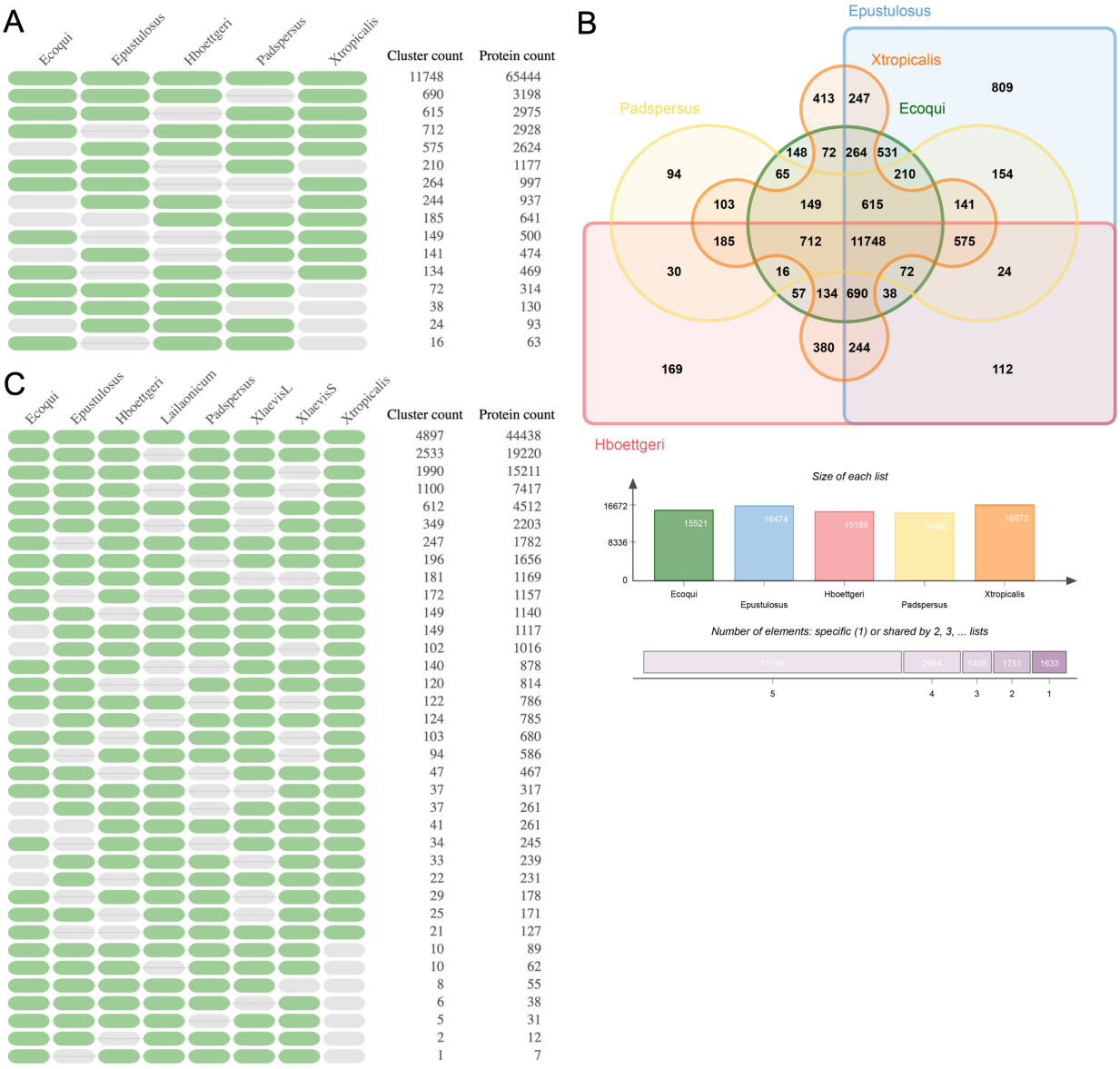


Figure 3.3. Analyzed with OrthoVenn2 [102], [A] the occurrence table of gene homology clusters containing three or more species between the five frog annotations completed in this analysis, [B] Venn diagram of gene homology between these five species and [C] the occurrence table of gene homology clusters containing six or more species between all seven chromosome-scale frog species. In the occurrence tables, the green and grey ovals represent the presence or absence, respectively, of that species in the OrthoVenn2 clustering.

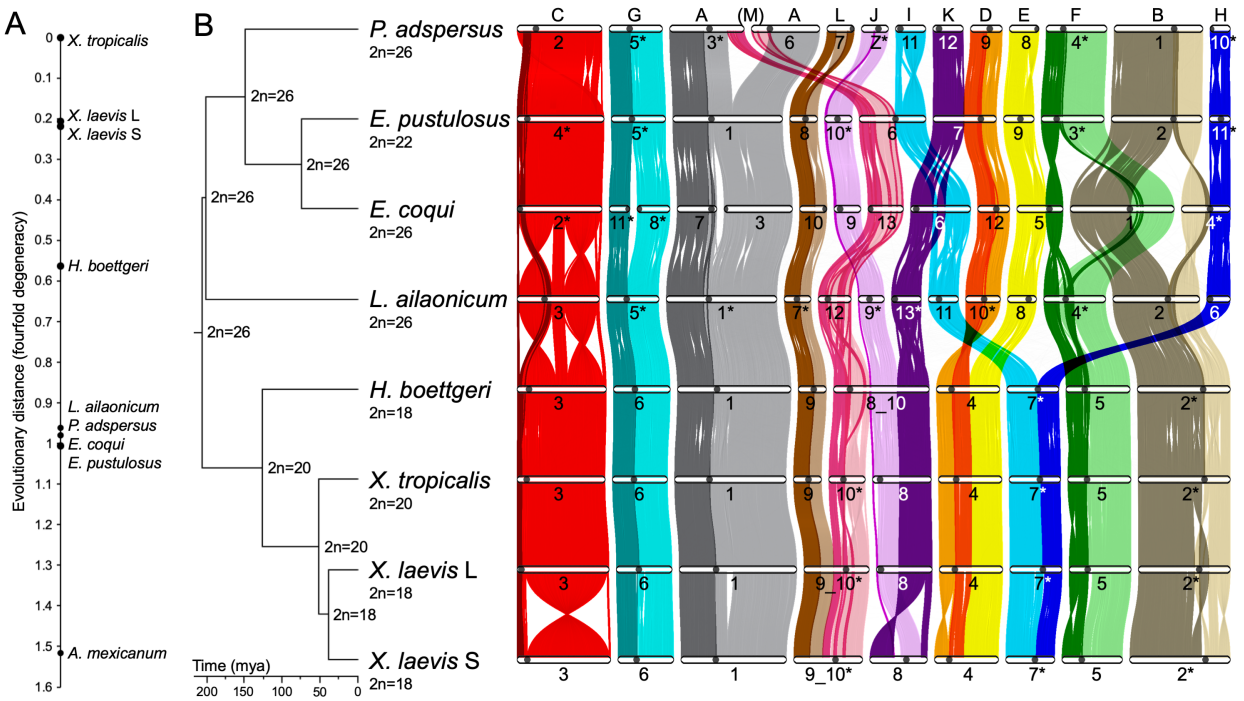
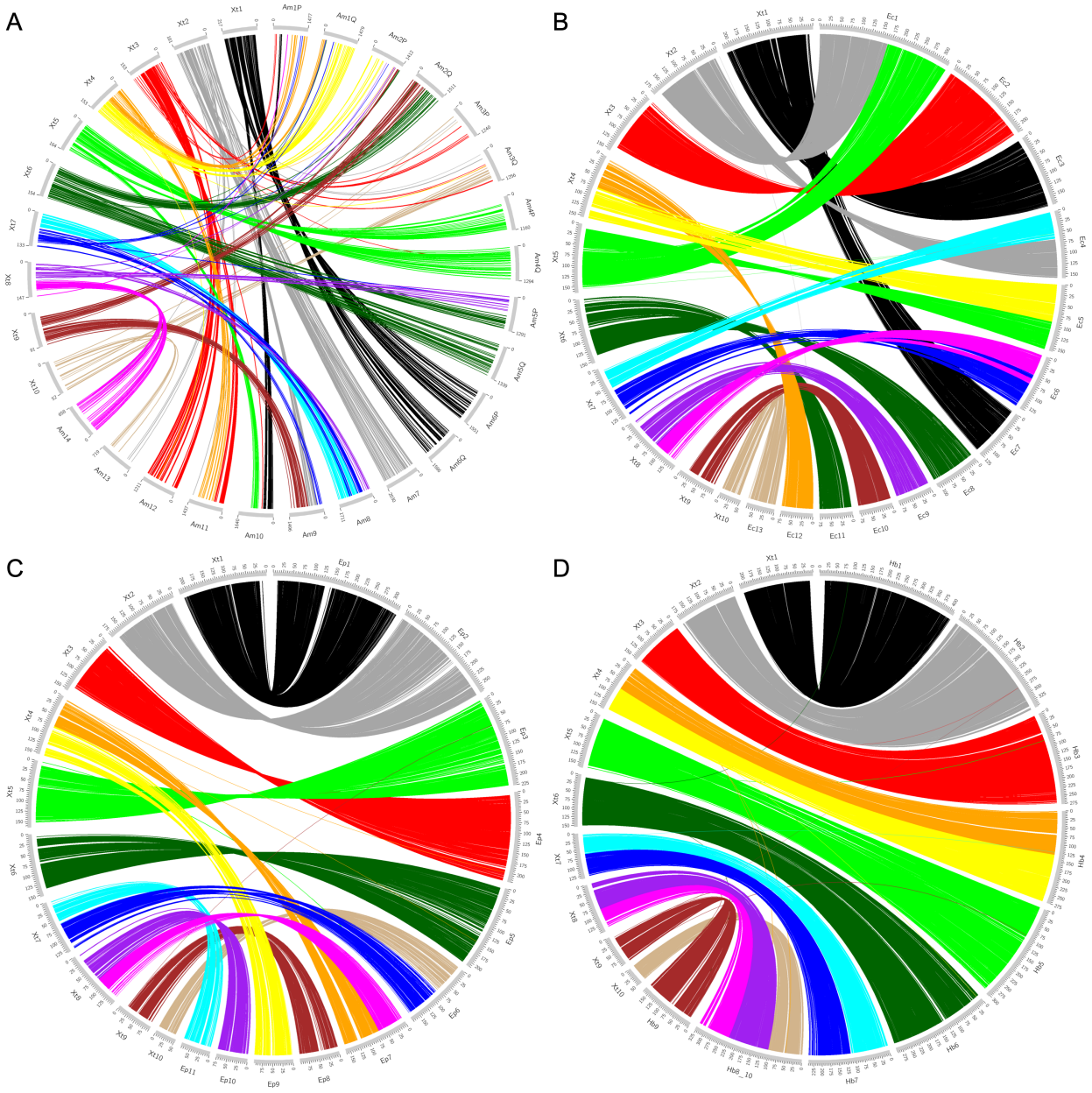


Figure 3.4. [A] Pairwise distances in substitutions per fourfold degenerate site extracted from the RAxML (v8.2.11) [97] phylogenetic tree using Newick utilities (v1.6) [98] were shown relative to the reference genome *X. tropicalis*. [B] Phylogenetic tree and gene ortholog alignment. The phylogenetic 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 (v0.8.12; https://github.com/tanghaibao/jcvi) using the 7,292 described chromosome one-to-one gene orthologs from OrthoVenn2 [102], followed by visual filtering of single stray orthologs. The pericentromeric region based on Hi-C 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 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.


A
B
C
D

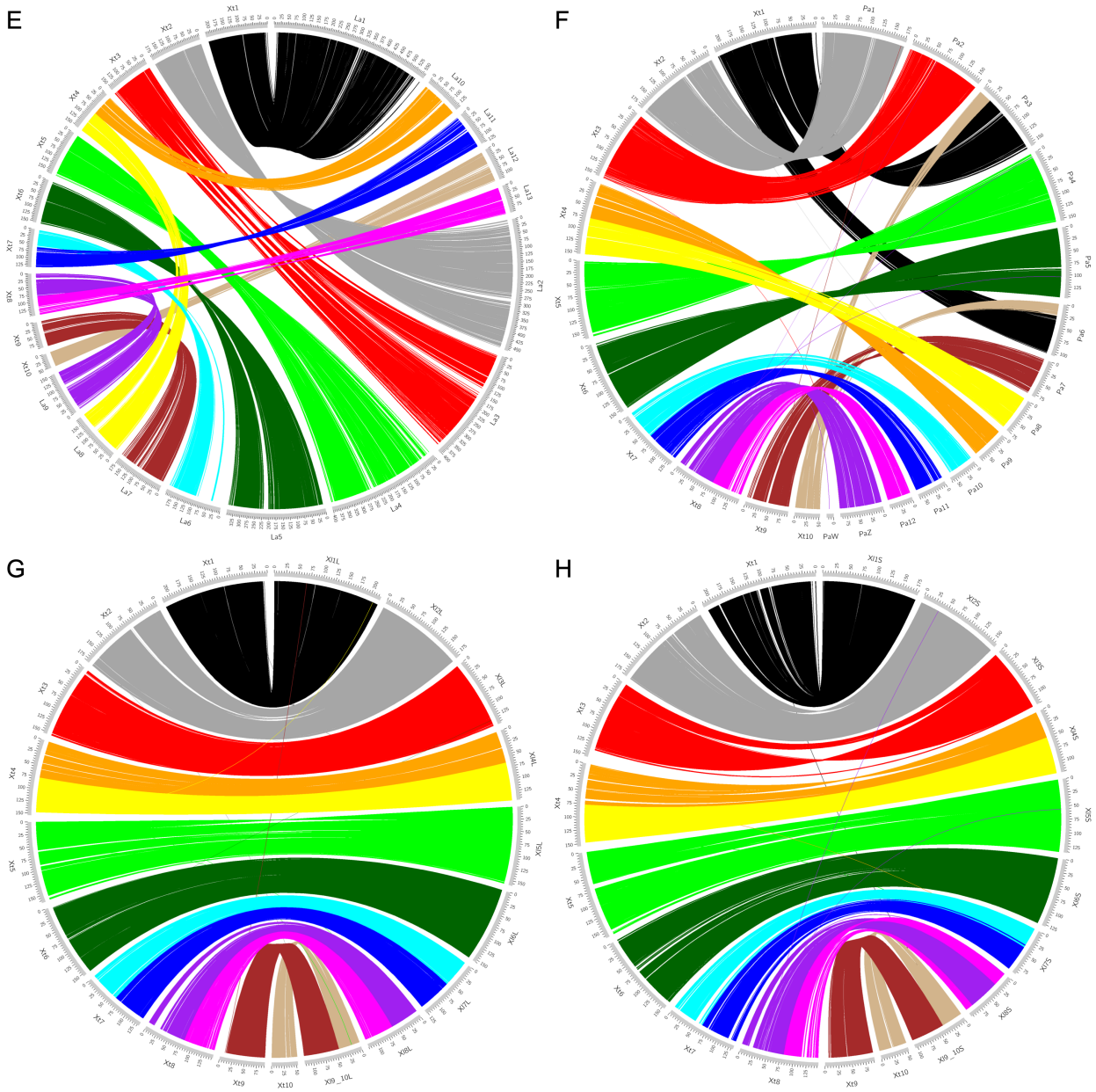


Figure 3.5. Circos (v0.69-6) [103] plots with runs of collinearity containing at least five kilobases (kb) of aligned sequence between [A] *X. tropicalis* (left, Xt) and *A. mexicanum* (right, Am), [B] *X. tropicalis* (left, Xt) and *E. coqui* (right, Ec), [C] *X. tropicalis* (left, Xt) and *E. pustulosus* (right, Ep), [D] *X. tropicalis* (left, Xt) and *H. boettgeri* (right, Hb), [E] *X. tropicalis* (left, Xt) and *L. ailaonicum* (right, La), [F] *X. tropicalis* (left, Xt) and *P. adspersus* (right, Pa), [G] *X. tropicalis* (left, Xt) and *X. laevis* L subgenome (right, Xl), and [H] *X. tropicalis* (left, Xt) and *X. laevis* S subgenome (right, Xl). For the plot of *X. tropicalis* and *A. mexicanum* [A], all chromosomes were scaled evenly. Runs of collinearity were colored with respect to the 13 ancestral chromosomes (A to M).

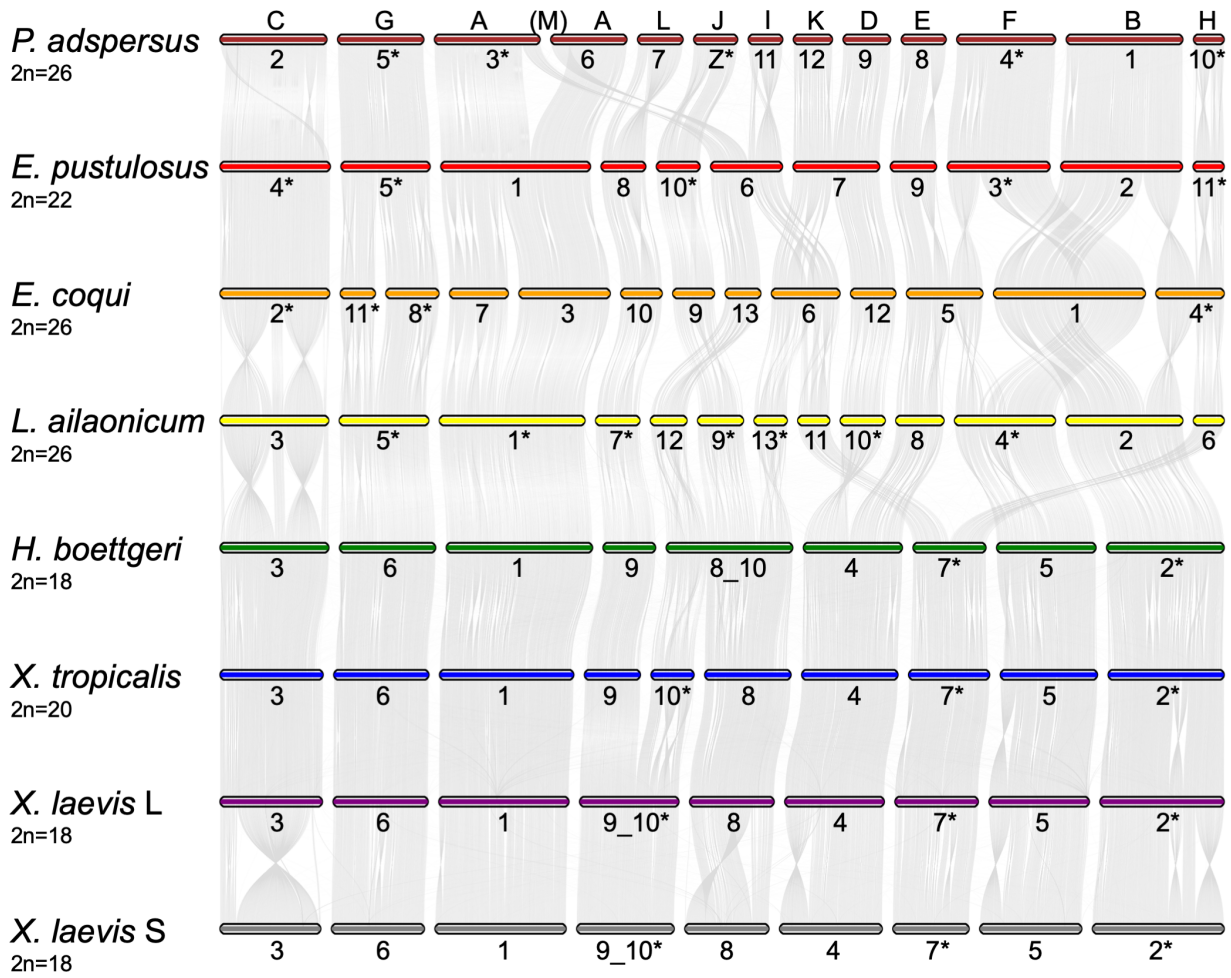


Figure 3.6. The alignment plot was generated with jcvi.graphics.karyotype (v0.8.12; https://github.com/tanghaibao/jcvi) using runs of collinearity containing at least one kb of aligned sequence between the species, extracted from the ROAST-merged MAF file. The ancestral chromosomes (A to M) were labeled at the top of the alignment based on the corresponding region in *P. adspersus*. Chromosomes labeled with an asterisk were reverse complemented in this image relative to the orientation in the assembly.

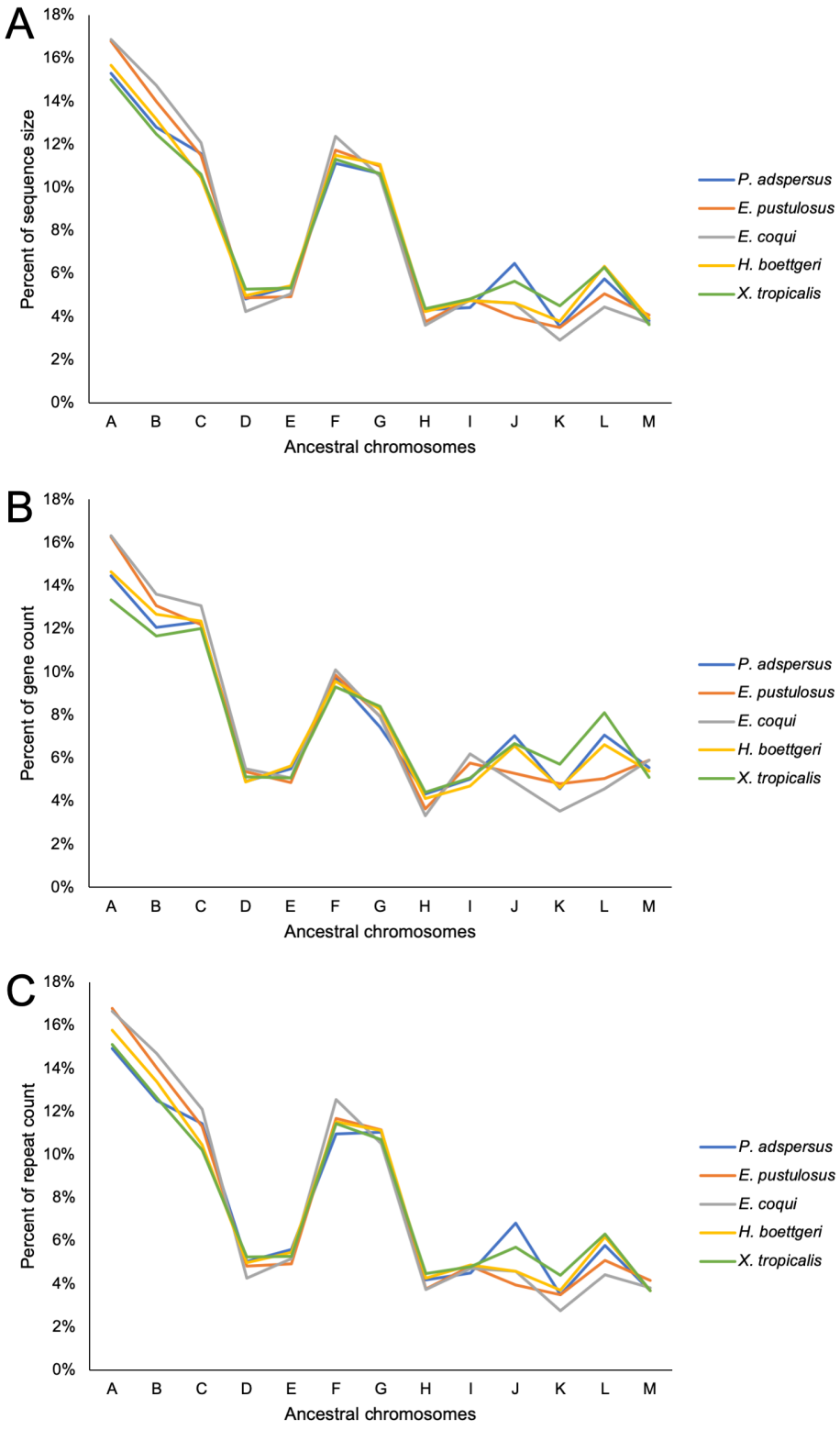


Figure 3.7. [A] Percent of sequence size for each ancestral chromosome, [B] percent of gene count for each ancestral chromosome, and [C] percent of repeat count for each ancestral chromosome. Boundaries of the ancestral chromosomes in each examined species were extracted from runs of collinearity containing at least one kb of aligned sequence against *L. ailaonicum*.

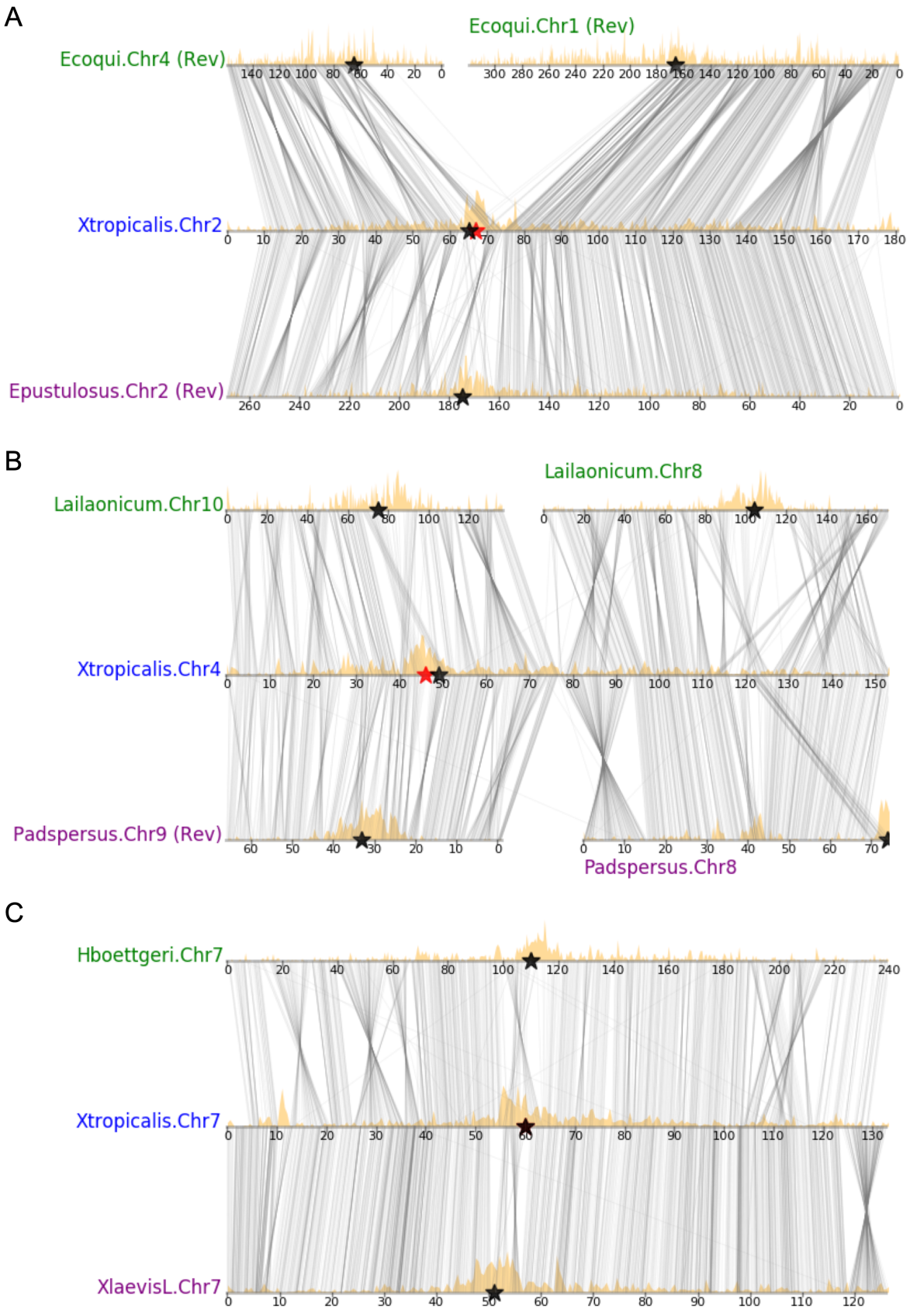

A
Ecoqui.Chr4 (Rev)
Ecoqui.Chr1 (Rev)
Xtropicalis.Chr2
Epustulosus.Chr2 (Rev)
B
Lailaonicum.Chr10
Lailaonicum.Chr8
Xtropicalis.Chr4
Padspersus.Chr9 (Rev)
Padspersus.Chr8
C
Hboettgeri.Chr7
Xtropicalis.Chr7
XlaevisL.Chr7

Figure 3.8. Examples of [A] a Robertsonian translocation, [B] an end-to-end fusion, and [C] pericentromeric conservation in the analyzed species visualized with custom script alignment_plots.py (v1.0). For each plot, the pericentromeric regions based on Hi-C inference were depicted with black stars, the centromeric satellite repeat in *X. tropicalis* 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.

Table 3.1. Summary of genome assemblies. Statistics were calculated using assembly-stats (commit 506a640; https://github.com/sanger-pathogens/assembly-stats).

Genomic feature	*E. coqui*	*E. pustulosus*	*H. boettgeri*
Total scaffold length, bp	2,789,403,129	2,592,984,374	3,214,299,233
Number of scaffolds	105,233	120,236	26,522
Scaffold N50 length, bp	109,468,876	172,109,237	293,320,900
Total contig length, bp	2,367,745,368	2,583,262,543	3,210,201,336
Number of contigs	480,045	137,711	42,879
Contig N50 length, bp	10,801	295,193	783,846
Contigs sequence in chromosomes, %	65.52	74.22	82.47
Contig GC content, %	43.46	42.53	39.51
Masked contig repeat sequence, %	56.1	47.92	48.6

Table 3.2. DNA and RNA sequencing generated, excluding index reads.

Species	Library type	Total number of reads	Total bases sequenced
E. coqui (NCBI BioProject PRJNA578591)	Illumina Short-Insert	1,705,673,540	283,472,133,530
	Illumina Mate Pair	597,450,992	149,960,198,992
	10X Genomics Chromium Genome	854,980,766	129,102,095,666
	Dovetail Genomics Hi-C	381,578,566	57,355,218,936
E. pustulosus (NCBI BioProject PRJNA578590)	10X Genomics Chromium Genome	852,054,914	128,660,292,014
	Pacific Biosciences SMRT	2,954,172	24,381,987,728
	Dovetail Genomics Hi-C	903,180,410	136,096,277,522
	Illumina mRNA	764,813,040	115,486,769,040
H. boettgeri (NCBI BioProject PRJNA578589)	Illumina Short-Insert	248,883,258	62,469,697,758
	10X Genomics Chromium Genome	924,540,580	139,605,627,580
	Pacific Biosciences SMRT	5,939,116	70,318,923,913
	Dovetail Genomics Hi-C	536,656,294	81,035,100,394
	Illumina mRNA	120,922,616	18,259,315,016

Table 3.3. Individuals and tissues used in RNA sequencing of *E. pustulosus*.

Individual	Sex	Tissue
237g6f1	Female	Brain
237g6f1	Female	Eggs
237g6f1	Female	Heart
237g6f1	Female	Larynx
237g6f1	Female	Skin, dorsal
237g6f1	Female	Skin, ventral
237g6f2	Female	Brain
237g6f2	Female	Eye
237g6f2	Female	Intestine
237g6f2	Female	Liver
237g6f2	Female	Lung
237g6f2	Female	Skin, dorsal
A10F2	Female	Brain
A10F2	Female	Eggs
A10F2	Female	Eye
A10F2	Female	Heart
A10F2	Female	Intestine
A10F2	Female	Larynx
A10F2	Female	Liver
A10F2	Female	Lung
A10F2	Female	Skin, ventral
A10M1	Male	Larynx
Brood 726 tadpole (stage ~56)	Unknown	Whole body minus gut
Brood 747 tadpole (stage ~45)	Unknown	Whole body minus gut

Table 3.4. Summary of genome annotations. Statistics were calculated by the Integrated Gene Call (IGC) pipeline [163].

Annotation feature	*E. coqui*	*E. pustulosus*	*H. boettgeri*	*P. adspersus*
Number of genes / primary transcripts	23,346	30,613	20,684	18,673
Number of alternate transcripts	11,220	34,629	7,374	2,574
Total number of transcripts	34,566	65,242	28,058	21,247
Average number of exons per gene	7.3	6.5	8.5	8.8
Median exon length	137	141	133	133
Median intron length	1,752	1,002	1,483	1,089
Number of complete genes	19,899	25,989	18,244	16,861
Number of incomplete genes with start codon	1,149	1,417	872	637

Number of incomplete genes with stop codon	1,662	2,125	1,164	941
Number of genes with Pfam annotation	17,101	20,393	16,677	15,711
Number of genes with Panther annotation	19,243	23,598	18,610	17,098
Number of genes with KOG annotation	9,633	10,365	10,162	10,347
Number of genes with KEGG Orthology annotation	9,799	10,528	9,973	9,779
Number of genes with E.C. number annotation	4,795	6,199	4,684	4,211

Table 3.5. Pairwise nucleotide divergence in substitutions per site based on fourfold degeneracy between the examined species extracted from the RAxML (v8.2.11) [97] phylogenetic tree using Newick utilities (v1.6) [98].

Species	*A. mexicanum*	*P. adspersus*	*E. pustulosus*	*E. coqui*	*L. ailaonicum*	*H. boettgeri*	*X. tropicalis*	*X. laevis* L
P. adspersus	1.1438	–	–	–	–	–	–	–
E. pustulosus	1.1697	0.7525	–	–	–	–	–	–
E. coqui	1.7050	0.7276	0.3560	–	–	–	–	–
L. ailaonicum	1.6863	1.0391	1.0650	1.0401	–	–	–	–
H. boettgeri	1.6215	1.0846	1.1106	1.0856	1.0669	–	–	–
X. tropicalis	1.5165	0.9795	1.0055	0.9806	0.9618	0.5621	–	–
X. laevis L	1.5367	0.9998	1.0258	1.0008	0.9821	0.5824	0.2056	–
X. laevis S	1.5496	1.0127	1.0386	1.0137	0.9949	0.5952	0.2185	0.1656

Table 3.6. Locations of ancestral chromosome fusions in the examined species using runs of collinearity from *L. ailaonicum*.

Species and chromosome	Fusion region	Fused ancestral chromosomes
E. coqui Chr1	167,326,341–167,391,435	B+F
E. coqui Chr4	64,368,134–64,506,892	H+B
E. coqui Chr5	90,030,948–90,111,922	E+F
E. coqui Chr6	51,920,009–52,085,381	K+I
E. pustulosus Chr6	78,294,430–79,251,447	M+I
E. pustulosus Chr7	67,080,716–67,193,535	K+D
H. boettgeri Chr4	132,010,981–132,334,876	D+E
H. boettgeri Chr7	112,455,236–114,827,918	H+I
H. boettgeri Chr8_10	103,984,075–104,182,343	M+J
H. boettgeri Chr8_10	227,189,815–227,531,594	J+K
P. adspersus Chr3	27,789,267–27,840,023	M+A
P. adspersus Chr6	25,252,335–25,589,174	M+A
X. tropicalis Chr4	76,607,359–76,630,690	D+E
X. tropicalis Chr7	63,487,360–63,537,129	H+I
X. tropicalis Chr8	81,812,475–81,878,653	J+K

Table 3.7. Divergence time input intervals from TimeTree [129] and estimated output times from MEGA7 (v7.0.26) [130] in million years ago.

Node	TimeTree minimum	TimeTree maximum	MEGA7 output
E. coqui – E. pustulosus	71	108	74
E. coqui – P. adspersus	147	162	149
E. coqui – L. ailaonicum	167	205	202
X. laevis L – *X. laevis* S	–	–	38
X. laevis L – *X. tropicalis*	34	79	52
X. laevis L – *H. boettgeri*	104	158	126
X. laevis L – *E. coqui*	187	220	205

3.4 Methods

***H. boettgeri* metaphase chromosome spread.** Stage 26 tadpoles (n=10) were incubated in 0.01% colchicine and 1X MMR for four to six hours at room temperature. 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 minutes. The dorsal portions were transferred to 0.2 mL of 60% acetic acid in deionized water and allowed to stand for five minutes. 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 five minutes. The slide and coverslip were then placed on dry ice for five minutes. The coverslip was removed from the frozen slide, and the slide was stained with 0.1 mg/mL Hoechst Stain solution for five minutes. A fresh coverslip was then mounted on the slide using VectaShield, and the edges were sealed with nail polish. Chromosomes in metaphase spreads were imaged on an Olympus BX51 Fluorescence Microscope run with Metamorph software using a 60x oil objective. Chromosome number was counted in 75 separate metaphase spreads.

Sequencing and assembly. The sequencing and assembly of *E. coqui, E. pustulosus*, and *H. boettgeri*, detailed below, followed a hierarchical assembly strategy, starting with short-read data and progressing to longer reads and linkages.

DNA extraction and sequencing of *E. pustulosus*. High molecular weight DNA was extracted, as previously described [149], from whole blood from two sisters [237g6f4 and 237g6f5] maintained at the University of the Pacific. Using DNA from one sister [237g6f4], a 10X Genomics Chromium Genome library [108] was prepared and sequenced on the Illumina HiSeq X by the HudsonAlpha Institute for Biotechnology. Using DNA from the other sister [237g6f5], Pacific Biosciences SMRT libraries were prepared and sequenced on the Pacific Biosciences Sequel by the DNA Technologies and Expression Analysis Cores at the University of California Davis Genome Center. Using liver dissected from a niece of the sisters [291g2f_3603], also maintained at the University of the Pacific, a Hi-C library was prepared using the Dovetail Genomics Hi-C library preparation kit and sequenced on the Illumina HiSeq 4000 by the Vincent J. Coates Genomics Sequencing Laboratory at the University of California, Berkeley. Two additional Hi-C libraries were prepared from the dissected liver and sequenced on the Illumina NextSeq by Dovetail Genomics.

Shotgun assembly of *E. pustulosus*. 10X Genomics linked reads were assembled with Supernova (v2.0.1) [108]. As previously described [62], putative archaeal, bacterial, viral, and vector contaminants were identified and removed by querying the assembly using BLAST+ (v2.6.0) [109] against the respective RefSeq and UniVec databases, using custom script general_decon.sh (v1.0). Putative mitochondrial sequence was also identified and removed by querying the assembly using BLAST+ (v2.6.0) [109] against the closest available mitochondrial assembly (NCBI JX564888.1) [164], using custom script mt_decon.sh (v1.0). Finally, putative nonvertebrate contamination was identified and removed through two rounds of filtering, using custom script nt_decon.sh (v1.0): (1) the assembly was queried using BLAST+ (v2.6.0) [109] against the NCBI NT database, flagging sequences with an E-value less than 1E-10 best hit to a nonvertebrate sequence, as identified by the corresponding taxonomic information; (2) flagged sequences were queried using BLAST+ (v2.6.0) [109] against previously published frog genomes (*Hyla arborea* v1 [165], *Nanorana parkeri* v2 [166], *P. adspersus* v29Jun2017 [146], *Rana catesbeiana* v3-20170621 [167], *R. temporaria* v3 [162], *X. laevis* v9 [149], and *X. tropicalis* v9 [168]) as well as frog sequences from NCBI EST, GSS, and Nucleotide databases, removing sequences without any hits based on a cutoff of 75% identity and an E-value less than 1E-10. The decontamination removed 8,581 scaffolds totaling 2.11 megabases (Mb) from the Supernova assembly.

Initial PacBio scaffolding of *E. pustulosus*. To improve the contiguity of the *E. pustulosus* assembly, decontaminated Supernova contigs were scaffolded with PacBio long-read data. This was achieved by performing a hybrid assembly of the filtered Supernova contigs and PacBio long reads using DBG2OLC (commit 1f7e752) [169]. PacBio reads were then mapped to the DBG2OLC output assembly with BLASR (commit 4323a52) [170] and polished with PBDAGCon (commit 1a2f1e7) [171] two times using the map4cns pipeline (commit dd89f52; https://bitbucket.org/rokhsar-lab/map4cns). Decontaminated Supernova contigs were mapped back to the polished DBG2OLC assembly using MUMmer (v3.23) [172] with a cutoff of 90% identity and then ordered and oriented into scaffolds using tsvtk get and maptk join from the GBS analysis pipeline (commit 80613d5) [173].

Initial chromosome assembly of *E. pustulosus*. After PacBio-based long-read scaffolding, the assembly was organized into chromosomes with Hi-C data using the Dovetail Genomics HiRise [174] pipeline. Hi-C reads were then aligned to the assembly with Juicer (commit d3ee11b) [112], and the assembly was manually corrected in Juicebox (v1.9.0) [104]. PacBio reads were aligned to the assembly with BWA (v0.7.17-r1188) [116], and gaps were resized using scripts pbGapLen and expand-gaps.py (https://bitbucket.org/bredeson/artisanal). Gaps in the assembly were then filled with PacBio data using PBJelly (PBSuite v15.8.24) [175].

Revised PacBio and 10X Genomics scaffolding of *E. pustulosus*. Given the limited resolution of the *E. pustulosus* Hi-C data for determining the proper order and orientation of scaffolds as well as the large number of gaps listed as overfilled by PBJelly, suggesting incorrect scaffolding, rearrangements, or other assembly errors, the gap-filled assembly was broken into contigs and then scaffolded with PacBio and 10X Genomics data. First, the contigs were scaffolded against the error corrected DBG2OLC assembly using MUMmer (v3.23) [172] as well as tsvtk and maptk in the GBS analysis pipeline (commit 80613d5) [173]. Next, the assembly was scaffolded with the 10X Genomics linked reads using Scaff10X (v2.1;

https://sourceforge.net/projects/phusion2/files/scaff10x/). Gaps were resized with PacBio data, as previously described, and filled with PBJelly (PBSuite v15.8.24) [175].

Revised chromosome assembly of *E. pustulosus*. The resulting assembly was organized back into chromosomes based on alignment against the initial chromosome assembly using MUMmer (v3.23) [172] as well as tsvtk and maptk in the GBS analysis pipeline (commit 80613d5) [173]. Gaps were again resized with PacBio data using pbGapLen and expand-gaps.py (https://bitbucket.org/bredeson/artisanal), and gaps were filled with PBJelly (PBSuite v15.8.24) [175].

Final assembly correction of *E. pustulosus*. The assembly was polished with two rounds of Illumina error correction. In this, 10X Genomics data was adapter trimmed using custom script trim_10X.py (v1.0) and aligned to the assembly with BWA (v0.7.17-r1188) [116]. Variants called by FreeBayes (commit 49413aa) [176] with a read depth within two standard deviations of the Gaussian fit (mean of 26.4 and standard deviation of 10.4) were corrected using the script ILEC in the map4cns pipeline (commit dd89f52; https://bitbucket.org/rokhsar-lab/map4cns).

After error correction, the Hi-C data was realigned to the assembly with Juicer (commit d3ee11b) [112]. Misjoins were identified and broken with the Hi-C data using Juicebox (v1.9.0) [104]. Remaining gaps were resized with the PacBio data, and closure was attempted with the adapter-trimmed 10X Genomics data using Platanus (v1.2.1) [114].

Final assembly release of *E. pustulosus*. Scaffolds smaller than one kb were removed from the final assembly with seqtk seq (v1.3-r106; https://github.com/lh3/seqtk), and chromosomes and scaffolds were numbered in order of size using SeqKit (v0.7.2-dev) [115]. Chromosomes were oriented arbitrarily.

RNA extraction and sequencing of *E. pustulosus*. In addition to the two whole tadpoles excluding gut at approximated stages 45 and 56, the following tissues were dissected from adult 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). All 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. Illumina TruSeq stranded mRNA libraries were prepared and sequenced on the Illumina HiSeq 4000 by the Vincent J. Coates Genomics Sequencing Laboratory at the University of California Berkeley.

DNA extraction, sequencing, and assembly of *H. boettgeri*. High molecular weight DNA was extracted, as previously described [149], from whole blood from one female [F2] purchased at the Albany Aquarium. A 10X Genomics Chromium Genome library [108] was prepared and sequenced on the Illumina HiSeq X by the HudsonAlpha Institute for Biotechnology. Pacific Biosciences SMRT libraries were prepared and sequenced on the Pacific Biosciences Sequel by the HudsonAlpha Institute for Biotechnology. Using liver dissected from a second, unrelated female [F3] purchased at the Albany Aquarium, a Hi-C library was prepared using the Dovetail Genomics Hi-C library preparation kit and sequenced on the Illumina HiSeq 4000 by the Vincent J. Coates Genomics Sequencing Laboratory at the University of California Berkeley. The

assembly process followed the same procedure as outlined for *E. pustulosus* with four differences: the closest available mitochondrial assembly used in decontamination was NCBI NC_015615.1 [177]; the decontamination removed 108 scaffolds totaling 39.9 kb from the Supernova assembly; variants with a read depth within only one standard deviation of the Gaussian fit (mean of 18.4 and standard deviation of 12.3) were corrected; and chromosomes were numbered and oriented based on alignment with MUMmer (v3.23) [172] to the *X. tropicalis* chromosomes.

RNA extraction and sequencing of *H. boettgeri*. Eggs were homogenized in TRIzol Reagent and processed according to manufacturer's instructions. RNA was then isolated using the RNeasy Kit (cat 74104). An mRNA library was prepared using the PrepX RNA-Seq for Illumina Library Kit (cat 640096/640097) by the Functional Genomics Laboratory at the University of California Berkeley and sequenced on the Illumina HiSeq 4000 by the Vincent J. Coates Genomics Sequencing Laboratory at the University of California Berkeley.

DNA extraction and sequencing of *E. coqui*. Kidney and liver tissue were dissected from one male collected in Hawaii [HN-11 male], and DNA was extracted from these tissues using the Zymo Research Quick gDNA MiniPrep Kit (cat D3007). Two Illumina short-insert libraries were prepared using kidney DNA by the Functional Genomics Laboratory at the University of California Berkeley and sequenced on the Illumina HiSeq 2500 and 4000 by the Vincent J. Coates Genomics Sequencing Laboratory at the University of California Berkeley. Two Illumina mate pair libraries were prepared using liver DNA and sequenced on the Illumina HiSeq 2500 by the HudsonAlpha Institute for Biotechnology. Using the liver tissue sample, a Hi-C library was prepared and sequenced on the Illumina NextSeq by Dovetail Genomics. High molecular weight DNA was extracted, as previously described [149], from whole blood from a second, unrelated male maintained at Harvard University [C4M]. Using DNA from this second male, a 10X Genomics Chromium Genome library [108] was prepared and sequenced on the Illumina HiSeq X by the HudsonAlpha Institute for Biotechnology.

Shotgun assembly of *E. coqui*. The short-insert libraries were adapter trimmed with ea-utils fastq-mcf (commit bd148d4) [178]. The mate pair libraries were adapter trimmed and split with NxTrim (commit 53c2193) [179]. Using custom script nxtrim_pipeline.sh (v1.0), the output from NxTrim was divided into two files: (1) reads flagged as mate pair or unknown were merged into a final mate pair file; (2) reads flagged as short-insert paired-end or single-end were merged into a final short-insert library, with the single-end reads given a corresponding blank second end. All trimmed data was then assembled with meraculous (v2.2.4) [180]. Mitochondrial sequence was assembled from adapter-trimmed short-insert data using custom script organelle_pipeline.py (v1.0) and NOVOPlasty (v2.6.3) [181], with other Hyloidea mitochondrial assemblies available on NCBI as input seeds.

Mirroring the *E. pustulosus* shotgun assembly process and as previously described [62], putative archaeal, bacterial, viral, and vector contamination was identified and removed by querying the assembly with BLAST+ (v2.3.0) [109] against the respective RefSeq and UniVec databases, using custom script general_decon.sh (v1.0). Putative mitochondrial sequence was also identified and removed by querying the assembly with BLAST+ (v2.3.0) [109] against the assembled mitochondrial sequence, using custom script mt_decon.sh (v1.0). In addition, the assembly was

queried with BLAST+ (v2.3.0) [109] against the NT database, previously published frog genomes (*H. arborea* v1 [165], *N. parkeri* v2 [166], *P. adspersus* v29Jun2017 [146], *R. catesbeiana* v3-20170621 [167], *R. temporaria* v3 [162], *X. laevis* v9 [149], and *X. tropicalis* v9 [168]), and frog sequences from NCBI EST, GSS, and Nucleotide databases to identify and remove non-vertebrate sequences, using script custom nt_decon.sh (v1.0). The decontamination removed 7,272 scaffolds totaling 2.06 Mb from the meraculous assembly.

Residual redundancy due to split haplotypes was identified and removed using custom script align_pipeline.sh (v1.0). To summarize, the adapter-trimmed libraries were aligned to the assembly with BWA (v0.7.15-r1140) [116]. Read depth was extracted from the alignments and used as a cutoff to separate half-depth and full-depth scaffolds. Half-depth scaffolds were then queried against each other with BLAST+ (v2.3.0) [109], and the smaller of each best-hit scaffold pair was extracted. Half-depth scaffolds were also kmer counted with Jellyfish (v2.1.4) [182], and the smaller of each scaffold pair with a unique, shared 31mer was extracted. Scaffolds identified in both BLAST+ and Jellyfish analyses were removed from the assembly. The redundancy pipeline removed 192,996 scaffolds totaling 31.1 Mb from the decontaminated assembly. The assembly was then scaffolded with SSPACE (v3.0) [183].

Chromosome assembly of *E. coqui*. The assembly was next scaffolded with 10X Genomics linked reads using Scaff10X (v2.1; https://sourceforge.net/projects/phusion2/files/scaff10x/). Attempts to further scaffold the assembly into chromosomes with the Hi-C data using Dovetail Genomics HiRise [174] and 3D-DNA (commit 745779b) [113] were unsuccessful. Therefore, the Scaff10X output was mapped to the *E. pustulosus* assembly using MUMmer (v3.23) [172] and then scaffolded based on synteny using tsvtk get and maptk join from the GBS analysis pipeline (commit 80613d5) [173]. Hi-C reads were aligned to the assembly with Juicer (commit d3ee11b) [112], and the synteny-based scaffolding was manually corrected with the Hi-C data using Juicebox (v1.9.0) [104]. Closure of the remaining gaps was attempted with the adapter-trimmed short-insert data using Platanus (v1.2.1) [114].

Final assembly release of *E. coqui*. Scaffolds smaller than one kb were removed from the final assembly with seqtk seq (v1.3-r106; https://github.com/lh3/seqtk), and chromosomes and scaffolds were numbered in order of size using SeqKit (v0.7.2-dev) [115]. Chromosomes were oriented arbitrarily.

Reassembly of *L. ailaonicum*. As previously described for *E. pustulosus*, putative archaeal, bacterial, viral, and vector contamination was checked by querying the published assembly [145,154] using BLAST+ (v2.9.0) [109] against the respective RefSeq and UniVec databases. Putative mitochondrial sequence was also checked by querying the assembly using BLAST+ (v2.9.0) [109] against the closest available mitochondrial assembly (NCBI NC_024427.1) [184]. No contaminant scaffolds were identified or removed from the assembly.

Hi-C reads (NCBI BioProject PRJNA523649) [145] were aligned to the assembly with Juicer (commit d3ee11b) [112], and existing scaffolding was manually error corrected using Juicebox (v1.11.08) [104]. All gaps were then resized to 100 bases. Chromosomes and scaffolds were numbered in order of size using SeqKit (v0.7.2-dev) [115]. Chromosomes were oriented arbitrarily.

Reassembly of *P. adspersus*. As previously described for *E. pustulosus*, putative archaeal, bacterial, viral, and vector contamination was identified and removed by querying a precursor (v29Jun2017) of the released assembly (GCA_004786255.1) [146] using BLAST+ (v2.6.0) [109] against the respective RefSeq and UniVec databases. Putative mitochondrial sequence was also identified and removed by querying the assembly using BLAST+ (v2.6.0) [109] against the closest available mitochondrial assembly (NCBI JX564898.1) [164]. The decontamination removed one scaffold totaling 1.45 kb from the assembly.

Chicago and Hi-C reads (NCBI BioProject PRJNA439445) [146] were aligned to the assembly with Juicer (commit d3ee11b) [112], and existing scaffolding was manually error corrected using Juicebox (v1.9.0) [104]. PacBio reads (NCBI BioProject PRJNA439445) [146] were aligned to the assembly with BWA (v0.7.17-r1188) [116], and gaps were resized using scripts pbGapLen and expand-gaps.py (https://bitbucket.org/bredeson/artisanal). Closure of the remaining gaps was attempted with ea-utils fastq-mcf (commit bd148d4) [178] adapter-trimmed TruSeq data (NCBI BioProject PRJNA439445) [146] using Platanus (v1.2.1) [114].

Scaffolds smaller than one kb were removed from the final assembly with seqtk seq (v1.3-r106; https://github.com/lh3/seqtk), and chromosomes and scaffolds were numbered in order of size using SeqKit (v0.7.2-dev) [115]. Chromosomes were later renamed and reoriented based on alignment with MUMmer (v3.23) [172] to the released assembly (GCA_004786255.1) [146].

Repeat library. RepeatModeler (v1.0.11) [117] was run for each chromosome-scale assembly, except *L. ailaonicum*, and the output was manually curated. Repeats from the following frog assemblies were combined for a pan-frog repeat library: *A. truei* [n=1,769], *E. coqui* [n=1,441], *E. pustulosus* [n=1,146], *H. boettgeri* [n=1,160], *P. adspersus* [n=908], *Xenopus borealis* [n=1,026], *X. laevis* v10 [n=913], and *X. tropicalis* v10 [n=973]. Ancestral frog repeats [n=934] from RepBase (v23.12) [118] were added to the curated repeats to create a final repeat library [n=10,270].

Annotation and homology analysis. Using the final repeat library, the chromosome-scale assemblies of *E. coqui, E. pustulosus*, *H. boettgeri, L. ailaonicum*, and *P. adspersus* were soft masked with RepeatMasker (v4.0.7 and v4.0.9) [119]. In addition to the aforementioned RNA sequencing, additional RNA data for *H. boettgeri* (NCBI BioProject PRJNA306175) [149] and *P. adspersus* (NCBI BioProject PRJNA439445) [146] was downloaded from NCBI SRA. Unpublished *E. coqui* RNA sequencing of stages 7, 10, and 13 hindlimb and stage 9-10 tail fin skin was obtained from Harvard University and the French National Center for Scientific Research, respectively. All RNA sequencing data was adapter trimmed with ea-utils fastq-mcf (commit bd148d4) [178] and aligned to the respective assemblies with STAR (v2.5.3a and v2.7.0f) [185], using the custom script STARalign.sh (v1.0).

Genome-guided transcriptomes were assembled with Trinity (v2.5.1) [186] for each individual RNA library: *E. coqui* [n=7], *E. pustulosus* [n=24], *H. boettgeri* [n=9], and *P. adspersus* [n=2]. The assembled transcriptomes for these four species were aligned to the respective assemblies with STARlong (v2.7.1a) [185] and then split into single-exon and multi-exon transcripts based on their alignments, using custom script filter_trinity.py (v1.0). Multi-exon transcripts were discarded if the first exon and/or last exon was less than 60 bp in length, if an intron was less

than 60 bp or greater than 300,000 bp in length, or if the total transcript alignment length was less than 250 bp. Single-exon transcripts larger than 80 amino acids and containing start and stop codons were extracted with TransDecoder (v3.0.1) [187].

Filtered single-exon and multi-exon transcripts were merged and used as mRNA evidence in the IGC pipeline [163] for genome annotation of the four species. Peptides from the v10 *X. tropicalis* annotation as well as SwissProt eukaryotes (downloaded November 2018) [188] were used as protein homology evidence in this pipeline. From the resulting annotations, gene homology between *E. coqui*, *E. pustulosus*, *H. boettgeri*, *L. ailaonicum* [145,154], *P. adspersus*, *X. laevis* (v9) [149], and *X. tropicalis* (v10) was analyzed with OrthoVenn2 [102] using an E-value of 1E-5 and an inflation value of 1.5. One-to-one gene orthologs between *E. coqui*, *E. pustulosus*, *H. boettgeri*, *P. adspersus*, and *X. tropicalis* were extracted from the OrthoVenn2 output, after requiring the ortholog sets to be either present in a single copy or absent in *L. ailaonicum* [145,154] and the L and S subgenomes of *X. laevis* (v9) [149]. Regarding the exceptions for *L. ailaonicum* and *X. laevis*, an initial analysis of the OrthoVenn2 output found that the *L. ailaonicum* annotation [145,154] was missing 37% (3,670) of the one-to-one gene orthologs found in the five main frog species, whereas the *X. laevis* annotation (v9) [149] has confounding factors associated with the allopolyploidization and resulting gene evolution. In constructing this set, we also excluded genes on the *P. adspersus* W chromosome.

Pericentromeric inference from Hi-C. Centurion (commit 985439c) [189] was run using MAPQ of zero or greater contact maps at one Mb matrix resolution with a coef parameter of 10. An initial set of pericentromeric positions determined by visualization in Juicebox (v1.11.08) [104] were supplied to the Centurion run. The Centurion results were reverted back to the initial visual estimates if the two values differed by more than 10%.

Comparative genomics. The assemblies for *A. mexicanum* (GCA_002915635.2) [147,148], *E. coqui*, *E. pustulosus*, *H. boettgeri*, *L. ailaonicum*, and *P. adspersus* were each aligned pairwise against *X. tropicalis* (v10) with cactus (commit e4d0859) [122]. *X. laevis* (v9) [149] was broken into subgenomes, and the chromosomes of each subgenome were aligned against *X. tropicalis* with cactus (commit e4d0859) [122]. As previously described [62], all pairwise output HAL alignment files were filtered and converted into MAF, using custom script cactus_filter.py (v1.0), and runs of collinearity were extracted from each pairwise MAF file. The pairwise MAF files were also merged with ROAST/MULTIZ (v012109) [125], using the phylogenetic topology from TimeTree [129], and sorted with last (v979) [127].

Phylogeny. Using the 9,624 identified one-to-one orthologous genes and the ROAST-merged MAF file, as previously described [62], fourfold degenerate bases were extracted with script 4Dextract.py (v1.0) and converted into PHYLIP format with BeforePhylo (commit 0885849; https://github.com/qiyunzhu/BeforePhylo). The maximum likelihood tree was estimated with RAxML (v8.2.11) [97] using the GTR+Gamma model of substitution with outgroup *A. mexicanum*. As previously described [128], the divergence time confidence intervals from TimeTree [129] for all nodes except the outgroup *A. mexicanum* node and the divergence between the L and S subgenomes in *X. laevis* were input into MEGA7 (v7.0.26) [130], using the Reltime method [131] and the GTR+Gamma model to create a time tree. The time calculated for the split between L and S subgenomes in *X. laevis* was substantiated in the literature [149,190].

Additional sequencing of *H. boettgeri*. DNA was extracted, as previously described [149], from whole blood from a female [F1] purchased at the Albany Aquarium. An Illumina short-insert library was prepared by the Functional Genomics Laboratory at the University of California Berkeley and sequenced on the Illumina HiSeq 2500 by the Vincent J. Coates Genomics Sequencing Laboratory at the University of California Berkeley.

Additional sequencing of *E. coqui*. Liver tissue was dissected from one female collected in Hawaii [HN-13 male], and DNA was extracted from the tissue using the Zymo Research Quick gDNA MiniPrep kit (cat D3007). Two Illumina short-insert libraries were prepared by the Functional Genomics Laboratory at the University of California Berkeley and sequenced on the Illumina HiSeq 2500 and 4000 by the Vincent J. Coates Genomics Sequencing Laboratory at the University of California Berkeley.

3.5 Notes

Availability of data and materials. The assemblies, annotations, and raw data were deposited at NCBI for *E. coqui* (BioProject PRJNA578591), *E. pustulosus* (BioProject PRJNA578590), *H. boettgeri* (BioProject PRJNA578589), *L. ailaonicum* (BioProject PRJNA578588), and *P. adspersus* (BioProject PRJNA578592). Unless stated otherwise, custom code used in this study is available at https://github.com/abmudd/Assembly.

Acknowledgements. We thank Mara Laslo and James Hanken for providing unpublished RNA sequencing data of *E. coqui* and *E. coqui* frogs for DNA extractions; Daniel Buchholz for providing *E. coqui* kidney and liver DNA and tissue; Nicolas Buisine and Laurent Sachs for providing unpublished RNA sequencing data of *E. coqui*; Marcos Gridi-Papp for providing *E. pustulosus* frogs for DNA and RNA extraction; Taejoon Kwon and Michael Ryan for providing funding for *E. pustulosus* sequencing; Rebecca Heald for providing *H. boettgeri* frogs for DNA extraction and funding for *H. boettgeri* work; 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; Jane Grimwood, Jeremy Schmutz, and the HudsonAlpha Institute for Biotechnology for preparing and sequencing 10X Genomics, PacBio, and Illumina mate pair 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 Hi-C library preparation kit, running quality control on Hi-C libraries, and preparing and sequencing Hi-C libraries; Shana McDevitt and the Vincent J. Coates Genomics Sequencing Laboratory at the University of California Berkeley for sequencing Hi-C and Illumina short-insert libraries; and Jessica Lyons for coordinating DNA extraction and sequencing.

4 Conclusion

Summary. Evolutionary conservation, or lack thereof, is a theme throughout this work, as I explored cases where nature strongly favored consistency and cases where variation occurred rapidly. Extreme variation at the chromosome scale was explored in Chapter 2 by comparing the genome assemblies of *Muntiacus muntjak* and *Muntiacus reevesi*, which diverged ~4.9 million years ago. With a recent rate of 5.3 fusion events per million years, *M. muntjak* has undergone rapid chromosome evolution. However, the underlying sequence of the two muntjacs has changed very little, and I found that the local chromatin contacts were conserved. I also explored the interesting cases of the fission-fusion reversal of chromosome BTA6 in *M. muntjak* and the uniquely independent fusions of chromosomes BTA13 and BTA18_25_26_28 in the two muntjac species. On the other hand, in Chapter 3, I identified an extreme case of torpid chromosome evolution by evaluating seven frog genomes (*Eleutherodactylus coqui*, *Engystomops pustulosus*, *Hymenochirus boettgeri*, *Leptobrachium ailaonicum*, *Pyxicephalus adspersus*, *Xenopus laevis*, and *Xenopus tropicalis*), finding that the frog chromosomes were highly conserved, despite diverging ~205 million years ago. In contrast to this limited karyotype variation, I calculated a total nucleotide divergence based on fourfold degeneracy between the examined frog species of 2.58 substitutions per site, revealing that the underlying sequence comparatively changed despite the vast chromosome conservation. Given these curious evolutionary incongruities, this work provides fodder for future inquiry into topics such as the evolutionary role and physiological constraint for karyotype change as well as the biological mechanisms related to chromosome conservation and variation.

Ongoing and future work. The superclass Tetrapoda encompasses an enormous range of species and ecological niches, from the 7.7 mm frog *Paedophryne amauensis* in the forest leaf litter [191] to the 32.6 m blue whale *Balaenoptera musculus* roaming the oceans [192]. Similarly, chromosome number has widely varied across tetrapods, ranging in count from six (*M. muntjak* [67]), 10 (*Macropus bicolor* [140]), and 14 (*Arthroleptis* spp. [193]) to 102 (*Tympanoctomys barrerae*), 108 (*Xenopus longipes*), and 138 (*Alcedo atthis*) [194], in addition to chromosome diversity resulting from supernumerary chromosomes [195] and microchromosomes [196]. Cytogenetic and comparative genomic analyses, such as the comparison of a basal relative of tetrapods, *Latimeria chalumnae*, against the frog *Ascaphus truei* [197], hypothesized that the tetrapod ancestor possessed a mixture of macrochromosomes and microchromosomes [198,199]. From examination of extant tetrapods, although microchromosomes are not present in mammals [200], they are found in reptiles [201] and amphibians [202], such as the aforementioned frog *A. truei* [197].

To evaluate this hypothesis of mixed macrochromosomes and microchromosomes throughout much of tetrapod and vertebrate evolutionary history [198,199] as well as more general karyotype conservation between tetrapods, I have been exploring the syntenic relationships between two pairs of karyotypically diverse tetrapods — the coastal tailed frog (*A. truei*; 2n=44) versus the western clawed frog (*X. tropicalis*; 2n=20 [168]) and the red jungle fowl (*Gallus gallus*; 2n=78 [203]) versus the American alligator (*Alligator mississippiensis*; 2n=30). With spotted gar (*Lepisosteus oculatus*; 2n=58 [199]) as an outgroup, I identified 44 tetrapod-specific synteny units (Table 4.1). Comparison of these synteny units against the Florida lancelet

(*Branchiostoma floridae*) enabled analysis of their origins relative to the ancestral chordate linkage groups [48,49]. From these synteny units, it has become possible to determine the mechanisms of chromosome rearrangements, both in ancestral and extant species, as well as other features including pericentromere conservation and gene evolution. This line of inquiry will further our understanding of the evolutionary and functional significance of chromosome synteny across the history of life.

Table 4.1. Chromosome locations and origin of the 44 tetrapod-specific synteny units. Non-primary chromosome locations were noted in parentheses.

Synteny unit	*Ascaphus truei*	*Xenopus tropicalis*	*Gallus gallus*	*Alligator mississippiensis*	*Lepisosteus oculatus*	Chordate linkage group notation
GNA1	Chr4	Chr5	Chr3	Chr1	LG1 (16)	AJK-2-alpha
GNA2	Chr3	Chr2	Chr1	Chr1	LG17 (3/14)	FKN-1-alpha
GNA3	Chr1	Chr1	Chr4	Chr2	LG4 (2)	FIQ-2-alpha
GNA4	Chr5	Chr3	Chr1	Chr4 (1)	LG8	EO-2-alpha
GNA5	Chr2	Chr6	Chr2	Chr3	LG9 (11)	DJ-1-alpha
GNA6	Chr1	Chr1	ChrZ	Chr3 (1/2)	LG2 (4)	CL-2-alpha
GNA7	Chr2	Chr6	Chr2	Chr5	LG11 (9)	B-1-alpha
GNA8	Chr10	Chr4	Chr8	Chr5 (7)	LG10	LM-1-alpha
GNA9	Chr7	Chr7	Chr6	Chr6 (2)	LG5	IQ-1-alpha
GNA10	Chr9	Chr8	Chr5	Chr2	LG7	A-1-alpha
GNA11	Chr8	Chr9	Chr7	Chr4 (8)	LG12	B-2-alpha
GNA12	Chr12	Chr9	Chr14	Chr13	LG13	H-1-alpha
GNA13	Chr14	Chr4	Chr11	Chr10 (1)	LG23	D-2-alpha
GNA14	Chr16	Chr4	Chr12	Chr12	LG5	E-1-alpha
GNA15	Chr13	Chr4	Chr5	Chr2	LG27	O-1-alpha
GNA16	Chr18	Chr1	Chr15	Chr10	LG20	G-1-alpha
GNA17	Chr15	Chr5	Chr9	Chr9	LG14	NP-2-alpha
GNA18	Chr5	Chr3	Chr13	Chr11	LG6	FIQ-2-beta
GNA19	Chr6	Chr7	Chr21	Chr13	LG25	P-1-alpha
GNA20	Chr3	Chr2	Chr19	Chr14	LG22	G-2-alpha
GNA21	Chr19	Chr8	Chr17	Chr12	LG21	M-2-alpha
GNA22	Chr11	Chr8	Chr4	Chr7	LG7 (1)	FKN-1-beta
GNA23	Chr7	Chr1	Chr28	Chr15	LG19	CL-2-beta
GNA24	Chr17	Chr3	Chr10	Chr8	LG3	C-1-alpha
GNA25	Chr6	Chr2	Chr23	Chr6	LG6	JK-2-beta
GNA26	Chr9	Chr2	Chr26	Chr14	LG23	EO-2-beta
GNA27	Chr21	Chr10	Chr20	Chr11	LG18	DJ-1-beta
GNA28	Chr22	Chr10	Chr18	Chr7	LG10	H-2-alpha
GNA29	Unknown	Chr10	Chr27	Chr4	LG15	B-2-beta
GNA30	Chr6	Chr7	Chr31 (22)	Chr5	LG8	H-1-beta
GNA31	Chr6	Chr7	Chr24	Chr15	LG26	A-2-beta
GNA32	Chr3	Chr2	Chr33	Chr8	LG4	B-1-beta

GNA33	Chr8	Chr8	Chr25	Chr8	LG24 (5)	C-1-beta
GNA34	Chr5	Chr3	Chr22	Chr9	LG1	IQ-1-beta
GNA35	Chr16	Chr4	Chr1	Chr4	LG12	H-2-beta
GNA36	Unknown	Chr3	Chr30	Chr7	LG6	LM-1-beta
GNA37	Unknown	Chr7	Chr5	Chr2	LG9	G-2-beta
GNA38	Chr19	Chr8	Unknown	Chr7	LG1	E-1-beta
GNA39	Chr7	Chr7	Chr1	Chr4 (1)	LG26	P-1-beta
GNA40	Chr8	Chr8	Chr32	Chr8	LG2	A-1-beta
GNA41	Chr8	Chr9	Chr9	Chr9	LG12	Unknown
GNA42	Unknown	Chr8	Chr16	Chr8	Unknown	M-2-beta
GNA43	Chr6	Chr7	Unknown	Chr8	LG24	Unknown
GNA44	Unknown	Chr3	Unknown	Chr8	LG2	NP-2-beta

References

1. Watson JD, Crick FHC. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. Nature. 1953;171: 737–738. doi:10.1038/171737a0
2. Watson JD, Crick FHC. Genetical implications of the structure of deoxyribonucleic acid. Nature. 1953;171: 964–967. doi:10.1038/171964b0
3. Sanger F. Chemistry of insulin: determination of the structure of insulin opens the way to greater understanding of life processes. Science. 1959;129: 1340–1344. doi:10.1126/science.129.3359.1340
4. Holley RW, Apgar J, Everett GA, Madison JT, Marquisee M, Merrill SH, et al. Structure of a ribonucleic acid. Science. 1965;147: 1462–1465. doi:10.1126/science.147.3664.1462
5. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes JC, et al. Nucleotide sequence of bacteriophage φX174 DNA. Nature. 1977;265: 687–695. doi:10.1038/265687a0
6. Abbott S, Fairbanks DJ. Experiments on plant hybrids by Gregor Mendel. Genetics. 2016;204: 407–422. doi:10.1534/genetics.116.195198
7. Darwin C. On The Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life. London: John Murray; 1859.
8. Zacharias H. Historical feature: key word: chromosome. Chromosom Res. 2001;9: 345–355. doi:10.1023/A:1016764113970
9. Schaeffer SW, Bhutkar A, McAllister BF, Matsuda M, Matzkin LM, O'Grady PM, et al. Polytene chromosomal maps of 11 Drosophila species: the order of genomic scaffolds inferred from genetic and physical maps. Genetics. 2008;179: 1601–1655. doi:10.1534/genetics.107.086074
10. Bhutkar A, Schaeffer SW, Russo SM, Xu M, Smith TF, Gelbart WM. Chromosomal rearrangement inferred from comparisons of 12 Drosophila genomes. Genetics. 2008;179: 1657–1680. doi:10.1534/genetics.107.086108
11. Schaeffer SW. Muller "elements" in Drosophila: how the search for the genetic basis for speciation led to the birth of comparative genomics. Genetics. 2018;210: 3–13. doi:10.1534/genetics.118.301084
12. Coghlan A, Eichler EE, Oliver SG, Paterson AH, Stein L. Chromosome evolution in eukaryotes: a multi-kingdom perspective. Trends Genet. 2005;21: 673–682. doi:10.1016/j.tig.2005.09.009
13. Fredga K. Chromosomal changes in vertebrate evolution. Proc R Soc London Ser B Biol Sci. 1977;199: 377–397. doi:10.1098/rspb.1977.0148
14. Balajee AS, Hande MP. History and evolution of cytogenetic techniques: current and future applications in basic and clinical research. Mutat Res - Genet Toxicol Environ Mutagen. 2018;836: 3–12. doi:10.1016/j.mrgentox.2018.08.008
15. Schwartz DC, Li X, Hernandez LI, Ramnarain SP, Huff EJ, Wang Y-K. Ordered restriction maps of Saccharomyces cerevisiae chromosomes constructed by optical mapping. Science. 1993;262: 110–114. doi:10.1126/science.8211116
16. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science. 2002;295: 1306–1311. doi:10.1126/science.1067799
17. Szabo Q, Bantignies F, Cavalli G. Principles of genome folding into topologically associating domains. Sci Adv. 2019;5: eaaw1668. doi:10.1126/sciadv.aaw1668

18. Cremer T, Cremer M. Chromosome territories. Cold Spring Harb Perspect Biol. 2010;2: a003889. doi:10.1101/cshperspect.a003889
19. Francke U. 2012 William Allan Award: adventures in cytogenetics. Am J Hum Genet. 2013;92: 325–337. doi:10.1016/j.ajhg.2013.01.010
20. Levy SE, Myers RM. Advancements in next-generation sequencing. Annu Rev Genomics Hum Genet. 2016;17: 95–115. doi:10.1146/annurev-genom-083115-022413
21. Jones SJM. Prediction of genomic functional elements. Annu Rev Genomics Hum Genet. 2006;7: 315–338. doi:10.1146/annurev.genom.7.080505.115745
22. Alexander RP, Fang G, Rozowsky J, Snyder M, Gerstein MB. Annotating non-coding regions of the genome. Nat Rev Genet. 2010;11: 559–571. doi:10.1038/nrg2814
23. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet. 2009;7: 85–97. doi:10.1038/nrg1767
24. Barrick JE, Lenski RE. Genome dynamics during experimental evolution. Nat Rev Genet. Nature Publishing Group; 2013;14: 827–839. doi:10.1038/nrg3564
25. Hermansen RA, Hvidsten TR, Sandve SR, Liberles DA. Extracting functional trends from whole genome duplication events using comparative genomics. Biol Proced Online. Biological Procedures Online; 2016;18: 11. doi:10.1186/s12575-016-0041-2
26. Levasseur A, Pontarotti P. The role of duplications in the evolution of genomes highlights the need for evolutionary-based approaches in comparative genomics. Biol Direct. 2011;6: 11. doi:10.1186/1745-6150-6-11
27. Demuth JP, De Bie T, Stajich JE, Cristianini N, Hahn MW. The evolution of mammalian gene families. PLoS One. 2006;1: e85. doi:10.1371/journal.pone.0000085
28. Necsulea A, Kaessmann H. Evolutionary dynamics of coding and non-coding transcriptomes. Nat Rev Genet. 2014;15: 734–748. doi:10.1038/nrg3802
29. Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, et al. A high-resolution map of human evolutionary constraint using 29 mammals. Nature. Nature Publishing Group; 2011;478: 476–482. doi:10.1038/nature10530
30. Villanueva-Cañas JL, Faherty SL, Yoder AD, Albà MM. Comparative genomics of mammalian hibernators using gene networks. Integr Comp Biol. 2014;54: 452–462. doi:10.1093/icb/icu048
31. Wang G-D, Xie H-B, Peng M-S, Irwin D, Zhang Y-P. Domestication genomics: evidence from animals. Annu Rev Anim Biosci. 2014;2: 65–84. doi:10.1146/annurev-animal-022513-114129
32. Foote AD, Liu Y, Thomas GWC, Vinař T, Alföldi J, Deng J, et al. Convergent evolution of the genomes of marine mammals. Nat Genet. Nature Publishing Group; 2015;47: 272–275. doi:10.1038/ng.3198
33. Kim S, Cho YS, Kim H-M, Chung O, Kim H, Jho S, et al. Comparison of carnivore, omnivore, and herbivore mammalian genomes with a new leopard assembly. Genome Biol. Genome Biology; 2016;17: 211. doi:10.1186/s13059-016-1071-4
34. Rogers J, Gibbs R a. Comparative primate genomics: emerging patterns of genome content and dynamics. Nat Rev Genet. Nature Publishing Group; 2014;15: 347–59. doi:10.1038/nrg3707
35. Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. Science. 2014;346: 1311–1320. doi:10.1126/science.1251385
36. Tollis M, Hutchins ED, Kusumi K. Reptile genomes open the frontier for comparative

analysis of amniote development and regeneration. Int J Dev Biol. 2014;58: 863–871. doi:10.1387/ijdb.140316kk
37. Yuan Z, Liu S, Zhou T, Tian C, Bao L, Dunham R, et al. Comparative genome analysis of 52 fish species suggests differential associations of repetitive elements with their living aquatic environments. BMC Genomics. BMC Genomics; 2018;19: 141. doi:10.1186/s12864-018-4516-1
38. Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, Crandall KA, et al. Earth BioGenome Project: sequencing life for the future of life. Proc Natl Acad Sci USA. 2018;115: 4325–4333. doi:10.1073/pnas.1720115115
39. Eisenstein M. Building an annotated ark. Lab Anim (NY). 2019;48: 105–108. doi:10.1038/s41684-019-0270-5
40. Worley KC, Richards S, Rogers J. The value of new genome references. Exp Cell Res. 2017;358: 433–438. doi:10.1016/j.yexcr.2016.12.014
41. Martinez PA, Jacobina UP, Fernandes R V., Brito C, Penone C, Amado TF, et al. A comparative study on karyotypic diversification rate in mammals. Heredity (Edinb). Nature Publishing Group; 2017;118: 366–373. doi:10.1038/hdy.2016.110
42. Theisen A, Shaffer LG. Disorders caused by chromosome abnormalities. Appl Clin Genet. 2010;3: 159–174. doi:10.2147/tacg.s8884
43. O'Neill ID. Homozygosity for constitutional chromosomal rearrangements: a systematic review with reference to origin, ascertainment and phenotype. J Hum Genet. 2010;55: 559–564. doi:10.1038/jhg.2010.80
44. Sacerdot C, Louis A, Bon C, Berthelot C, Roest Crollius H. Chromosome evolution at the origin of the ancestral vertebrate genome. Genome Biol. 2018;19: 116. doi:10.1186/s13059-018-1559-1
45. IJdo JW, Baldini A, Ward DC, Reeders ST, Wells RA. Origin of human chromosome 2: an ancestral telomere-telomere fusion. Proc Natl Acad Sci USA. 1991;88: 9051–9055. doi:10.1073/pnas.88.20.9051
46. Glasauer SMK, Neuhauss SCF. Whole-genome duplication in teleost fishes and its evolutionary consequences. Mol Genet Genomics. 2014;289: 1045–1060. doi:10.1007/s00438-014-0889-2
47. Dehal P, Boore JL. Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol. 2005;3: e314. doi:10.1371/journal.pbio.0030314
48. Nakatani Y, Takeda H, Kohara Y, Morishita S. Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. Genome Res. 2007;17: 1254–1265. doi:10.1101/gr.6316407
49. Putnam NH, Butts T, Ferrier DEK, Furlong RF, Hellsten U, Kawashima T, et al. The amphioxus genome and the evolution of the chordate karyotype. Nature. 2008;453: 1064–1071. doi:10.1038/nature06967
50. Simakov O, Kawashima T. Independent evolution of genomic characters during major metazoan transitions. Dev Biol. Elsevier Inc.; 2017;427: 179–192. doi:10.1016/j.ydbio.2016.11.012
51. Schubert I, Lysak MA. Interpretation of karyotype evolution should consider chromosome structural constraints. Trends Genet. 2011;27: 207–216. doi:10.1016/j.tig.2011.03.004
52. White MJD. Chromosomal rearrangements and speciation in animals. Annu Rev Genet. 1969;3: 75–98. doi:10.1146/annurev.ge.03.120169.000451
53. Potter S, Bragg JG, Blom MPK, Deakin JE, Kirkpatrick M, Eldridge MDB, et al.

Chromosomal speciation in the genomics era: disentangling phylogenetic evolution of rock-wallabies. Front Genet. 2017;8: 10. doi:10.3389/fgene.2017.00010
54. Lande R. Effective deme sizes during long-term evolution estimated from rates of chromosomal rearrangement. Evolution (N Y). 1979;33: 234–251. doi:10.2307/2407380
55. Gingerich PD. Rates of evolution. Annu Rev Ecol Evol Syst. 2009;40: 657–675. doi:10.1146/annurev.ecolsys.39.110707.173457
56. Lenski RE, Travisano M. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. Proc Natl Acad Sci USA. 1994;91: 6808–6814. doi:10.1073/pnas.91.15.6808
57. Kaye RW. The Mathematics of Logic: A Guide to Completeness Theorems and their Applications. Cambridge: Cambridge University Press; 2007.
58. Simpson GG. Horotely, Bradytely, and Tachytely. The Major Features of Evolution. New York, NY: Columbia University Press; 1953. pp. 313–337.
59. Schopf TJM. Rates of evolution and the notion of "living fossils." Annu Rev Earth Planet Sci. 1984;12: 245–292. doi:10.1146/annurev.ea.12.050184.001333
60. Baker RJ, Bickham JW. Karyotypic evolution in bats: evidence of extensive and conservative chromosomal evolution in closely related taxa. Syst Biol. 1980;29: 239–253. doi:10.1093/sysbio/29.3.239
61. Marks J. Rates of karyotype evolution. Syst Zool. 1983;32: 207–209. doi:10.2307/2413282
62. Mudd AB, Bredeson JV, Baum R, Hockemeyer D, Rokhsar DS. Muntjac chromosome evolution and architecture. bioRxiv. 2019; 772343. doi:10.1101/772343
63. Gladkikh OL, Romanenko SA, Lemskaya NA, Serdyukova NA, O'Brien PCM, Kovalskaya JM, et al. Rapid karyotype evolution in Lasiopodomys involved at least two autosome - sex chromosome translocations. PLoS One. 2016;11: e0167653. doi:10.1371/journal.pone.0167653
64. Nash WG, Wienberg J, Ferguson-Smith MA, Menninger JC, O'Brien SJ. Comparative genomics: tracking chromosome evolution in the family Ursidae using reciprocal chromosome painting. Cytogenet Genome Res. 1998;83: 182–192. doi:10.1159/000015176
65. Carbone L, Harris RA, Gnerre S, Veeramah KR, Lorente-Galdos B, Huddleston J, et al. Gibbon genome and the fast karyotype evolution of small apes. Nature. 2014;513: 195–201. doi:10.1038/nature13679
66. Gordon DJ, Resio B, Pellman D. Causes and consequences of aneuploidy in cancer. Nat Rev Genet. 2012;13: 189–203. doi:10.1038/nrg3123
67. Lin CC, Sasi R, Fan Y-S, Chen Z-Q. New evidence for tandem chromosome fusions in the karyotypic evolution of Asian muntjacs. Chromosoma. 1991;101: 19–24. doi:10.1007/BF00360682
68. Hartmann N, Scherthan H. Characterization of ancestral chromosome fusion points in the Indian muntjac deer. Chromosoma. 2004;112: 213–220. doi:10.1007/s00412-003-0262-4
69. Chi JX, Huang L, Nie W, Wang J, Su B, Yang F. Defining the orientation of the tandem fusions that occurred during the evolution of Indian muntjac chromosomes by BAC mapping. Chromosoma. 2005;114: 167–172. doi:10.1007/s00412-005-0004-x
70. Robertson WRB. Chromosome studies. I. Taxonomic relationships shown in the chromosomes of Tettigidae and Acrididae: V-shaped chromosomes and their significance in Acrididae, Locustidae, and Gryllidae: chromosomes and variation. J Morphol. 1916;27:

179–331. doi:10.1002/jmor.1050270202
71. Yang F, O'Brien PCM, Wienberg J, Ferguson-Smith MA. A reappraisal of the tandem fusion theory of karyotype evolution in the Indian muntjac using chromosome painting. Chromosom Res. 1997;5: 109–117. doi:10.1023/A:1018466107822
72. Wang W, Lan H. Rapid and parallel chromosomal number reductions in muntjac deer inferred from mitochondrial DNA phylogeny. Mol Biol Evol. 2000;17: 1326–1333. doi:10.1093/oxfordjournals.molbev.a026416
73. Farré M, Kim J, Proskuryakova AA, Zhang Y, Kulemzina AI, Li Q, et al. Evolution of gene regulation in ruminants differs between evolutionary breakpoint regions and homologous synteny blocks. Genome Res. 2019;29: 576–589. doi:10.1101/gr.239863.118
74. Chen L, Qiu Q, Jiang Y, Wang K, Lin Z, Li Z, et al. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. Science. 2019;364: eaav6202. doi:10.1126/science.aav6202
75. Zimin A V., Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, et al. A whole-genome assembly of the domestic cow, Bos taurus. Genome Biol. 2009;10: R42. doi:10.1186/gb-2009-10-4-r42
76. Bana NÁ, Nyiri A, Nagy J, Frank K, Nagy T, Stéger V, et al. The red deer Cervus elaphus genome CerEla1.0: sequencing, annotating, genes, and chromosomes. Mol Genet Genomics. 2018;293: 665–684. doi:10.1007/s00438-017-1412-3
77. Li Z, Lin Z, Ba H, Chen L, Yang Y, Wang K, et al. Draft genome of the reindeer (Rangifer tarandus). Gigascience. 2017;6: 1–5. doi:10.1093/gigascience/gix102
78. Maruyama T, Imai HT. Evolutionary rate of the mammalian karyotype. J Theor Biol. 1981;90: 111–121. doi:10.1016/0022-5193(81)90125-9
79. Wurster DH, Benirschke K. Indian muntjac, Muntiacus muntjak: a deer with a low diploid chromosome number. Science. 1970;168: 1364–1366. doi:10.1126/science.168.3937.1364
80. Wurster DH, Benirschke K. Chromosome studies in some deer, the springbok, and the pronghorn, with notes on placentation in deer. Cytologia (Tokyo). 1967;32: 273–285. doi:10.1508/cytologia.32.273
81. Frohlich J, Kubickova S, Musilova P, Cernohorska H, Muskova H, Vodicka R, et al. Karyotype relationships among selected deer species and cattle revealed by bovine FISH probes. PLoS One. 2017;12: e0187559. doi:10.1371/journal.pone.0187559
82. Zhou Q, Huang L, Zhang J, Zhao X, Zhang Q, Song F, et al. Comparative genomic analysis links karyotypic evolution with genomic evolution in the Indian muntjac (Muntiacus muntjak vaginalis). Chromosoma. 2006;115: 427–436. doi:10.1007/s00412-006-0066-4
83. Lin C-C, Hsu P-C, Li T-S, Liao S-J, Cheng Y-M, Hsieh L-J, et al. Construction of an Indian muntjac BAC library and production of the most highly dense FISH map of the species. Zool Stud. 2008;47: 282–292.
84. Tsipouri V, Schueler MG, Hu S, NISC Comparative Sequencing Program, Dutra A, Pak E, et al. Comparative sequence analyses reveal sites of ancestral chromosomal fusions in the Indian muntjac genome. Genome Biol. 2008;9: R155. doi:10.1186/gb-2008-9-10-r155
85. Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, Faraut T, et al. The sheep genome illuminates biology of the rumen and lipid metabolism. Science. 2014;344: 1168–1173. doi:10.1126/science.1252806
86. Schneider VA, Graves-Lindsay T, Howe K, Bouk N, Chen H-C, Kitts PA, et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the

enduring quality of the reference assembly. Genome Res. 2017;27: 849–864. doi:10.1101/gr.213611.116
87. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30: 1236–1240. doi:10.1093/bioinformatics/btu031
88. Huang L, Chi J, Wang J, Nie W, Su W, Yang F. High-density comparative BAC mapping in the black muntjac (Muntiacus crinifrons): molecular cytogenetic dissection of the origin of MCR 1p+4 in the X1X2Y1Y2Y3 sex chromosome system. Genomics. 2006;87: 608–615. doi:10.1016/j.ygeno.2005.12.008
89. Huang L, Wang J, Nie W, Su W, Yang F. Tandem chromosome fusions in karyotypic evolution of Muntiacus: evidence from M. feae and M. gongshanensis. Chromosom Res. 2006;14: 637–647. doi:10.1007/s10577-006-1073-2
90. Murmann AE, Mincheva A, Scheuermann MO, Gautier M, Yang F, Buitkamp J, et al. Comparative gene mapping in cattle, Indian muntjac, and Chinese muntjac by fluorescence in situ hybridization. Genetica. 2008;134: 345–351. doi:10.1007/s10709-008-9242-1
91. The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature. 2005;437: 69–87. doi:10.1038/nature04072
92. Locke DP, Hillier LW, Warren WC, Worley KC, Nazareth L V., Muzny DM, et al. Comparative and demographic analysis of orang-utan genomes. Nature. 2011;469: 529–533. doi:10.1038/nature09687
93. Zou Y, Yi X, Wright WE, Shay JW. Human telomerase can immortalize Indian muntjac cells. Exp Cell Res. 2002;281: 63–76. doi:10.1006/excr.2002.5645
94. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009;326: 289–293. doi:10.1126/science.1181369
95. Liming S, Pathak S. Gametogenesis in a male Indian muntjac x Chinese muntjac hybrid. Cytogenet Genome Res. 1981;30: 152–156. doi:10.1159/000131603
96. Ghavi-Helm Y, Jankowski A, Meiers S, Viales RR, Korbel JO, Furlong EEM. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. Nat Genet. 2019;51: 1272–1282. doi:10.1038/s41588-019-0462-3
97. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30: 1312–1313. doi:10.1093/bioinformatics/btu033
98. Junier T, Zdobnov EM. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. Bioinformatics. 2010;26: 1669–1670. doi:10.1093/bioinformatics/btq243
99. Cunningham F, Achuthan P, Akanni W, Allen J, Amode MR, Armean IM, et al. Ensembl 2019. Nucleic Acids Res. 2019;47: D745–D751. doi:10.1093/nar/gky1113
100. Li Z, Lin Z, Ba H, Chen L, Yang Y, Wang K, et al. Draft genomic data of the reindeer (Rangifer tarandus). GigaScience Database. 2017; doi:10.5524/100370
101. Wang Y, Coleman-Derr D, Chen G, Gu YQ. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res. 2015;43: W78–W84. doi:10.1093/nar/gkv487
102. Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H, et al. OrthoVenn2: a web server for whole-

genome comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res. 2019;47: W52–W58. doi:10.1093/nar/gkz333
103. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19: 1639–1645. doi:10.1101/gr.092759.109
104. Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 2016;3: 99–101. doi:10.1016/j.cels.2015.07.012
105. Bradnam KR, Fass JN, Alexandrov A, Baranay P, Bechner M, Birol I, et al. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. Gigascience. 2013;2: 10. doi:10.1186/2047-217X-2-10
106. Gremme G, Steinbiss S, Kurtz S. GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. IEEE/ACM Trans Comput Biol Bioinforma. 2013;10: 645–656. doi:10.1109/TCBB.2013.68
107. Hockemeyer D, Sfeir AJ, Shay JW, Wright WE, de Lange T. POT1 protects telomeres from a transient DNA damage response and determines how human chromosomes end. EMBO J. 2005;24: 2667–2678. doi:10.1038/sj.emboj.7600733
108. Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB. Direct determination of diploid genome sequences. Genome Res. 2017;27: 757–767. doi:10.1101/gr.214874.116
109. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10: 421. doi:10.1186/1471-2105-10-421
110. Shi YF, Shan XN, Li J, Zhang XM, Zhang HJ. Sequence and organization of the complete mitochondrial genome of the Indian muntjac (Muntiacus muntjak). Acta Zool Sin. 2003;49: 629–636. doi:10.3969/j.issn.1674-5507.2003.05.011
111. Zhang XM, Zhang HJ, Mou Y, Li J, Yi GC, Shan XN. Muntiacus reevesi mitochondrion, complete genome. NCBI Reference Sequence NC_004069.1; 2002.
112. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 2016;3: 95–98. doi:10.1016/j.cels.2016.07.002
113. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science. 2017;356: 92–95. doi:10.1126/science.aal3327
114. Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, et al. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res. 2014;24: 1384–1395. doi:10.1101/gr.170720.113
115. Shen W, Le S, Li Y, Hu F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS One. 2016;11: e0163962. doi:10.1371/journal.pone.0163962
116. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25: 1754–1760. doi:10.1093/bioinformatics/btp324
117. Smit AFA, Hubley R. RepeatModeler Open-1.0. 2015. http://www.repeatmasker.org
118. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015;6: 11. doi:10.1186/s13100-015-0041-9
119. Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2015. http://www.repeatmasker.org
120. Keilwagen J, Wenk M, Erickson JL, Schattat MH, Grau J, Hartung F. Using intron

position conservation for homology-based gene prediction. Nucleic Acids Res. 2016;44: e89. doi:10.1093/nar/gkw092
121. Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. Proc Natl Acad Sci USA. 2003;100: 11484–11489. doi:10.1073/pnas.1932072100
122. Paten B, Earl D, Nguyen N, Diekhans M, Zerbino D, Haussler D. Cactus: algorithms for genome multiple sequence alignment. Genome Res. 2011;21: 1512–1528. doi:10.1101/gr.123356.111
123. Hickey G, Paten B, Earl D, Zerbino D, Haussler D. HAL: a hierarchical format for storing and analyzing multiple genome alignments. Bioinformatics. 2013;29: 1341–1342. doi:10.1093/bioinformatics/btt128
124. Suarez HG, Langer BE, Ladde P, Hiller M. chainCleaner improves genome alignment specificity and sensitivity. Bioinformatics. 2017;33: 1596–1603. doi:10.1093/bioinformatics/btx024
125. Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AFA, Roskin KM, et al. Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res. 2004;14: 708–715. doi:10.1101/gr.1933104
126. Arnold C, Matthews LJ, Nunn CL. The 10kTrees website: a new online resource for primate phylogeny. Evol Anthropol. 2010;19: 114–118. doi:10.1002/evan.20251
127. Kielbasa SM, Wan R, Sato K, Horton P, Frith MC. Adaptive seeds tame genomic sequence comparison. Genome Res. 2011;21: 487–493. doi:10.1101/gr.113985.110
128. Mello B. Estimating timetrees with MEGA and the TimeTree resource. Mol Biol Evol. 2018;35: 2334–2342. doi:10.1093/molbev/msy133
129. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. Mol Biol Evol. 2017;34: 1812–1819. doi:10.1093/molbev/msx116
130. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33: 1870–1874. doi:10.1093/molbev/msw054
131. Tamura K, Battistuzzi FU, Billing-Ross P, Murillo O, Filipski A, Kumar S. Estimating divergence times in large molecular phylogenies. Proc Natl Acad Sci USA. 2012;109: 19333–19338. doi:10.1073/pnas.1213199109
132. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25: 2078–2079. doi:10.1093/bioinformatics/btp352
133. Hales KG, Korey CA, Larracuente AM, Roberts DM. Genetics on the fly: a primer on the Drosophila model system. Genetics. 2015;201: 815–842. doi:10.1534/genetics.115.183392
134. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow T a, et al. Evolution of genes and genomes on the Drosophila phylogeny. Nature. 2007;450: 203–218. doi:10.1038/nature06341
135. Sved JA, Chen Y, Shearman D, Frommer M, Gilchrist AS, Sherwin WB. Extraordinary conservation of entire chromosomes in insects over long evolutionary periods. Evolution (N Y). 2016;70: 229–234. doi:10.1111/evo.12831
136. Mandrioli M, Zambonini G, Manicardi GC. Comparative gene mapping as a tool to understand the evolution of pest crop insect chromosomes. Int J Mol Sci. 2017;18: 1919. doi:10.3390/ijms18091919

137. Artemov GN, Bondarenko SM, Naumenko AN, Stegniy VN, Sharakhova M V., Sharakhov I V. Partial-arm translocations in evolution of malaria mosquitoes revealed by high-coverage physical mapping of the Anopheles atroparvus genome. BMC Genomics. 2018;19: 278. doi:10.1186/s12864-018-4663-4
138. Melnikova L, Georgiev P. Drosophila telomeres: the non-telomerase alternative. Chromosom Res. 2005;13: 431–441. doi:10.1007/s10577-005-0992-7
139. Ferguson-Smith MA, Trifonov V. Mammalian karyotype evolution. Nat Rev Genet. 2007;8: 950–962. doi:10.1038/nrg2199
140. Deakin JE, Graves JAM, Rens W. The evolution of marsupial and monotreme chromosomes. Cytogenet Genome Res. 2012;137: 113–129. doi:10.1159/000339433
141. Hooper DM, Price TD. Rates of karyotypic evolution in estrildid finches differ between island and continental clades. Evolution (N Y). 2015;69: 890–903. doi:10.1111/evo.12633
142. Poletto AB, Ferreira IA, Cabral-de-Mello DC, Nakajima RT, Mazzuchelli J, Ribeiro HB, et al. Chromosome differentiation patterns during cichlid fish evolution. BMC Genet. 2010;11: 50. doi:10.1186/1471-2156-11-50
143. Árnason Ú. Comparative chromosome studies in Cetacea. Hereditas. 1974;77: 1–36. doi:10.1111/j.1601-5223.1974.tb01351.x
144. Wilson AC, Sarich VM, Maxson LR. The importance of gene rearrangement in evolution: evidence from studies on rates of chromosomal, protein, and anatomical evolution. Proc Natl Acad Sci USA. 1974;71: 3028–3030. doi:10.1073/pnas.71.8.3028
145. Li Y, Ren Y, Zhang D, Jiang H, Wang Z, Li X, et al. Chromosome-level assembly of the mustache toad genome using third-generation DNA sequencing and Hi-C analysis. Gigascience. Oxford University Press; 2019;8: 1–10. doi:10.1093/gigascience/giz114
146. Denton RD, Kudra RS, Malcom JW, Du Preez L, Malone JH. The African bullfrog (Pyxicephalus adspersus) genome unites the two ancestral ingredients for making vertebrate sex chromosomes. bioRxiv. 2018; 329847. doi:10.1101/329847
147. Nowoshilow S, Schloissnig S, Fei J-F, Dahl A, Pang AWC, Pippel M, et al. The axolotl genome and the evolution of key tissue formation regulators. Nature. 2018;554: 50–55. doi:10.1038/nature25458
148. Smith JJ, Timoshevskaya N, Timoshevskiy VA, Keinath MC, Hardy D, Voss SR. A chromosome-scale assembly of the axolotl genome. Genome Res. 2019;29: 317–324. doi:10.1101/gr.241901.118
149. Session AM, Uno Y, Kwon T, Chapman JA, Toyoda A, Takahashi S, et al. Genome evolution in the allotetraploid frog Xenopus laevis. Nature. 2016;538: 336–343. doi:10.1038/nature19840
150. Schmid M, Steinlein C, Bogart JP, Feichtinger W, León P, La Marca E, et al. The chromosomes of terraranan frogs. Schmid M, Bogart JP, Hedges SB, editors. Cytogenet Genome Res. 2010;130–131: 1–568. doi:10.1159/isbn.978-3-8055-9608-4
151. Nazareth Rabello M. Chromosomal studies in Brazilian anurans. Caryologia. 1970;23: 45–59. doi:10.1080/00087114.1970.10796362
152. Scheel JJ. The chromosomes of some African anuran species. In: Schröder JH, editor. Genetics and Mutagenesis of Fish. Berlin: Springer-Verlag; 1973. pp. 113–116. doi:10.1007/978-3-642-65700-9_11
153. Mezzasalma M, Glaw F, Odierna G, Petraccioli A, Guarino FM. Karyological analyses of Pseudhymenochirus merlini and Hymenochirus boettgeri provide new insights into the chromosome evolution in the anuran family Pipidae. Zool Anz. 2015;258: 47–53.

doi:10.1016/j.jcz.2015.07.001
154. Li Y, Ren Y, Zhang D, Jiang H, Wang Z, Li X, et al. Supporting data for "Chromosomal-level assembly of the mustache toad genome using third-generation DNA sequencing and Hi-C analysis." GigaScience Database. 2019; doi:10.5524/100624
155. Funk WC, Zamudio KR, Crawford AJ. Advancing understanding of amphibian evolution, ecology, behavior, and conservation with massively parallel sequencing. In: Hohenlohe P, Rajora OP, editors. Population Genomics. Cham: Springer; 2018. doi:10.1007/13836_2018_61
156. Ford LS, Cannatella DC. The major clades of frogs. Herpetol Monogr. 1993;7: 94–117. doi:10.2307/1466954
157. Morescalchi A. Evolution and karyology of the amphibians. Bolletino di Zool. 1980;47: 113–126. doi:10.1080/11250008009438709
158. Griffin DK, Robertson LBW, Tempest HG, Skinner BM. The evolution of the avian genome as revealed by comparative molecular cytogenetics. Cytogenet Genome Res. 2007;117: 64–77. doi:10.1159/000103166
159. Deakin JE, Ezaz T. Understanding the evolution of reptile chromosomes through applications of combined cytogenetics and genomics approaches. Cytogenet Genome Res. 2019;157: 7–20. doi:10.1159/000495974
160. Bush GL, Case SM, Wilson AC, Patton JL. Rapid speciation and chromosomal evolution in mammals. Proc Natl Acad Sci USA. 1977;74: 3942–3946. doi:10.1073/pnas.74.9.3942
161. Olmo E. Rate of chromosome changes and speciation in reptiles. Genetica. 2005;125: 185–203. doi:10.1007/s10709-005-8008-2
162. Palomar G, Ahmad F, Vasemägi A, Matsuba C, Nicieza AG, Cano JM. Comparative high-density linkage mapping reveals conserved genome structure but variation in levels of heterochiasmy and location of recombination cold spots in the common frog. G3 (Bethesda). 2017;7: 637–645. doi:10.1534/g3.116.036459
163. Shu S, Goodstein DM, Hayes D, Mitros T, Rokhsar DS. JGI plant genomics gene annotation pipeline. Plant Biology 2014. Portland, Oregon: American Society of Plant Biologists; 2014.
164. Zhang P, Liang D, Mao R-L, Hillis DM, Wake DB, Cannatella DC. Efficient sequencing of anuran mtDNAs and a mitogenomic exploration of the phylogeny and evolution of frogs. Mol Biol Evol. 2013;30: 1899–1915. doi:10.1093/molbev/mst091
165. Brelsford A, Dufresnes C, Perrin N. High-density sex-specific linkage maps of a European tree frog (Hyla arborea) identify the sex chromosome without information on offspring sex. Heredity (Edinb). 2016;116: 177–181. doi:10.1038/hdy.2015.83
166. Sun Y-B, Xiong Z-J, Xiang X-Y, Liu S-P, Zhou W-W, Tu X-L, et al. Whole-genome sequence of the Tibetan frog Nanorana parkeri and the comparative evolution of tetrapod genomes. Proc Natl Acad Sci USA. 2015;112: E1257–E1262. doi:10.1073/pnas.1501764112
167. Hammond SA, Warren RL, Vandervalk BP, Kucuk E, Khan H, Gibb EA, et al. The North American bullfrog draft genome provides insight into hormonal regulation of long noncoding RNA. Nat Commun. 2017;8: 1433. doi:10.1038/s41467-017-01316-7
168. Mitros T, Lyons JB, Session AM, Jenkins J, Shu S, Kwon T, et al. A chromosome-scale genome assembly and dense genetic map for Xenopus tropicalis. Dev Biol. 2019;452: 8–20. doi:10.1016/j.ydbio.2019.03.015
169. Ye C, Hill CM, Wu S, Ruan J, Ma Z. DBG2OLC: efficient assembly of large genomes

using long erroneous reads of the third generation sequencing technologies. Sci Rep. 2016;6: 31900. doi:10.1038/srep31900
170. Chaisson MJ, Tesler G. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. BMC Bioinformatics. 2012;13: 238. doi:10.1186/1471-2105-13-238
171. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods. 2013;10: 563–569. doi:10.1038/nmeth.2474
172. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004;5: R12. doi:10.1186/gb-2004-5-2-r12
173. ICGMC (International Cassava Genetic Map Consortium). High-resolution linkage map and chromosome-scale genome assembly for cassava (Manihot esculenta Crantz) from 10 populations. G3 (Bethesda). 2015;5: 133–144. doi:10.1534/g3.114.015008
174. Putnam NH, Connell BO, Stites JC, Rice BJ, Hartley PD, Sugnet CW, et al. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res. 2016;26: 342–350. doi:10.1101/gr.193474.115
175. English AC, Richards S, Han Y, Wang M, Vee V, Qu J, et al. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PLoS One. 2012;7: e47768. doi:10.1371/journal.pone.0047768
176. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv. 2012; 1207.3907.
177. Irisarri I, Vences M, San Mauro D, Glaw F, Zardoya R. Reversal to air-driven sound production revealed by a molecular phylogeny of tongueless frogs, family Pipidae. BMC Evol Biol. 2011;11: 114. doi:10.1186/1471-2148-11-114
178. Aronesty E. Comparison of sequencing utility programs. Open Bioinforma J. 2013;7: 1–8. doi:10.2174/1875036201307010001
179. O'Connell J, Schulz-Trieglaff O, Carlson E, Hims MM, Gormley NA, Cox AJ. NxTrim: optimized trimming of Illumina mate pair reads. Bioinformatics. 2015;31: 2035–2037. doi:10.1093/bioinformatics/btv057
180. Chapman JA, Ho I, Sunkara S, Luo S, Schroth GP, Rokhsar DS. Meraculous: de novo genome assembly with short paired-end reads. PLoS One. 2011;6: e23501. doi:10.1371/journal.pone.0023501
181. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2017;45: e18. doi:10.1093/nar/gkw955
182. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27: 764–770. doi:10.1093/bioinformatics/btr011
183. Boetzer M, Henkel C V., Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 2011;27: 578–579. doi:10.1093/bioinformatics/btq683
184. Xu Q, Liu S, Wan R, Yue B, Zhang X. The complete mitochondrial genome of the Vibrissaphora boringii (Anura: Megophryidae). Mitochondrial DNA A DNA Mapp Seq Anal. 2016;27: 758–759. doi:10.3109/19401736.2014.915527
185. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast

universal RNA-seq aligner. Bioinformatics. 2013;29: 15–21. doi:10.1093/bioinformatics/bts635
186. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat Biotechnol. 2011;29: 644–652. doi:10.1038/nbt.1883
187. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013;8: 1494–1512. doi:10.1038/nprot.2013.084
188. The UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2017;45: D158–D169. doi:10.1093/nar/gkw1099
189. Varoquaux N, Liachko I, Ay F, Burton JN, Shendure J, Dunham MJ, et al. Accurate identification of centromere locations in yeast genomes using Hi-C. Nucleic Acids Res. 2015;43: 5331–5339. doi:10.1093/nar/gkv424
190. Evans BJ, Kelley DB, Tinsley RC, Melnick DJ, Cannatella DC. A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implications for polyploid evolution. Mol Phylogenet Evol. 2004;33: 197–213. doi:10.1016/j.ympev.2004.04.018
191. Rittmeyer EN, Allison A, Gründler MC, Thompson DK, Austin CC. Ecological guild evolution and the discovery of the world's smallest vertebrate. PLoS One. 2012;7: e29797. doi:10.1371/journal.pone.0029797
192. Sears R, Perrin WF. Blue whale: Balaenoptera musculus. In: Perrin WF, Würsig B, Thewissen JGM, editors. Encyclopedia of Marine Mammals. San Diego, CA: Academic Press; 2009. pp. 120–124. doi:10.1016/B978-0-12-373553-9.00033-X
193. Bogart JP, Tandy M. Chromosome lineages in African ranoid frogs. Monit Zool Ital. 1981;15: 55–91. doi:10.1080/03749444.1981.10736629
194. Havelka M, Bytyutskyy D, Symonová R, Ráb P, Flajšhans M. The second highest chromosome count among vertebrates is observed in cultured sturgeon and is associated with genome plasticity. Genet Sel Evol. 2016;48: 12. doi:10.1186/s12711-016-0194-0
195. Ahmad SF, Martins C. The modern view of B chromosomes under the impact of high scale omics analyses. Cells. 2019;8: 156. doi:10.3390/cells8020156
196. Deakin JE, Ezaz T. Tracing the evolution of amniote chromosomes. Chromosoma. 2014;123: 201–216. doi:10.1007/s00412-014-0456-y
197. Bogart JP, Balon EK, Bruton MN. The chromosomes of the living coelacanth and their remarkable similarity to those of one of the most ancient frogs. J Hered. 1994;85: 322–325. doi:10.1093/oxfordjournals.jhered.a111470
198. Ohno S, Muramoto J, Stenius C, Christian L, Kittrell WA, Atkin NB. Microchromosomes in holocephalian, chondrostean and holostean fishes. Chromosoma. 1969;26: 35–40. doi:10.1007/BF00319498
199. Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, et al. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. Nat Genet. 2016;48: 427–437. doi:10.1038/ng.3526
200. O'Connor RE, Kiazim L, Skinner B, Fonseka G, Joseph S, Jennings R, et al. Patterns of microchromosome organization remain highly conserved throughout avian evolution. Chromosoma. 2019;128: 21–29. doi:10.1007/s00412-018-0685-6
201. Uno Y, Nishida C, Tarui H, Ishishita S, Takagi C, Nishimura O, et al. Inference of the protokaryotypes of amniotes and tetrapods and the evolutionary processes of microchromosomes from comparative gene mapping. PLoS One. 2012;7: e53027.

doi:10.1371/journal.pone.0053027
202. Wake MH, Hafner JC, Hafner MS, Klosterman LL, Patton JL. The karyotype of Typhlonectes compressicauda (Amphibia: Gymnophiona) with comments on chromosome evolution in caecilians. Experientia. 1980;36: 171–172.
203. Warren WC, Hillier LW, Tomlinson C, Minx P, Kremitzki M, Graves T, et al. A new chicken genome assembly provides insight into avian genome structure. G3 (Bethesda). 2017;7: 109–117. doi:10.1534/g3.116.035923