# Elasmobranch genome sequencing reveals evolutionary trends of vertebrate karyotypic organization

Kazuaki Yamaguchi, Yoshinobu Uno, Mitsutaka Kadota, et al.

*Genome Res.* published online August 17, 2023
Access the most recent version at doi:10.1101/gr.276840.122

*Research article*

# Elasmobranch genome sequencing reveals evolutionary trends of vertebrate karyotype organization

Kazuaki Yamaguchi[1], Yoshinobu Uno[1*], Mitsutaka Kadota[1], Osamu Nishimura[1], Ryo Nozu[2**], Kiyomi Murakumo[3], Rui Matsumoto[3], Keiichi Sato[2,3], Shigehiro Kuraku[1,4,5#]

[1]Laboratory for Phyloinformatics, RIKEN Center for Biosystems Dynamics Research (BDR), Kobe, Japan

[2]Okinawa Churashima Research Center, Okinawa Churashima Foundation, Okinawa, Japan.

[3]Okinawa Churaumi Aquarium, Okinawa, Japan.

[4]Molecular Life History Laboratory, Department of Genomics and Evolutionary Biology, National Institute of Genetics, Mishima, Japan.

[5]Department of Genetics, Sokendai (Graduate University for Advanced Studies), Mishima, Japan.

[*]Present address: Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo, 153-8902, Japan

[**]Present address: Graduate School of Integrated Sciences for Life, Hiroshima University, Hiroshima 739-0046, Japan

[#]To whom correspondence may be addressed. Shigehiro Kuraku

Running title: Evolution of shark chromosomes

**Abstract**

Genomic studies of vertebrate chromosome evolution have long been hindered by the scarcity of chromosome-scale DNA sequences of some key taxa. One of those limiting taxa has been the elasmobranchs (sharks and rays), which harbor species often with numerous chromosomes and enlarged genomes. Here, we report the chromosome-scale genome assembly for the zebra shark *Stegostoma tigrinum*, an endangered species that has a relatively small genome among sharks (3.71 Gb), as well as for the whale shark *Rhincodon typus*. Our analysis employing a male–female comparison identified an X Chromosome, the first genomically characterized shark sex chromosome. The X Chromosome harbors the Hox C cluster whose intact linkage has not been shown for an elasmobranch fish. The sequenced shark genomes exhibit a gradualism of chromosome length with remarkable length-dependent characteristics—shorter chromosomes tend to have higher GC-content, gene density, synonymous substitution rate, and simple tandem repeat content as well as smaller gene length and lower interspersed repeat content. We challenge the traditional binary classification of karyotypes as with and without so-called microchromosomes. Even without microchromosomes, the length-dependent characteristics persist widely in non-mammalian vertebrates. Our investigation of elasmobranch karyotypes underpins their unique characteristics and provides clues for understanding how vertebrate karyotypes accommodate intragenomic heterogeneity to realize a complex readout. It also paves the way to dissecting more genomes with variable sizes to be sequenced on high quality.

Key words: shark, chromosome-scale assembly, microchromosome, karyotype evolution, sex chromosome

## Introduction

Genomes accommodate coexisting regions with differential characteristics, and these characteristics are manifested not only in DNA sequences (e.g., GC-content) but also in intra-genomic heterogeneity of non-sequence features such as chromatin openness, replication timing, and recombination frequency. These features are thought to be associated with how karyotypes of individual species are organized. For example, in the chicken genome, early replicating regions tend to be found in small chromosomes, known as ‘microchromosomes’ (see below). It is still unknown how such intragenomic heterogeneity is accommodated by variable karyotypes as well as how it arose during evolution. To reconstruct the ancestor of all extant vertebrates and the evolutionary process thereafter, information from the evolutionary lineages that branched off in the early phase of vertebrate evolution is instrumental. Among those lineages, whole genome sequence information for cartilaginous fishes has been scarce. The importance of studying cartilaginous fishes is doubled when we consider their genomic trends. No additional whole genome duplication has been reported for cartilaginous fishes, while drastic lineage-specific genomic changes are being reported for the other extant non-osteichthyan taxon, cyclostomes (Nakatani et al. 2021).

Cartilaginous fishes (chondrichthyans) are divided into two groups, Holocephala (chimaera and ratfish) and Elasmobranchii (sharks and rays). Even long after whole genome sequences of a holocephalan species, *Callorhinuchus milii*, were made available (Venkatesh et al. 2014), those sequences have not been validated with any karyotyping report. This limitation stems mainly from the technical difficulty in reproducibly preparing chromosome spreads from a stable supply of metaphase cells. Only recently has our repeated sampling of fresh shark tissues

(blood or embryos) (e.g., at aquariums) enabled karyotyping using culture cells for four orectolobiform shark species in Elasmobranchii (Uno et al. 2020). This has paved the way for rigid evaluation of whole genome sequences. Biological studies on cartilaginous fishes have been hindered by low accessibility to fresh material. Moreover, especially in studying elasmobranchs, genome analysis can encounter inherent difficulty incurred by their large genome sizes (reviewed in Kuraku 2021). These factors have prevented previous efforts on cartilaginous fishes from obtaining a suite of genome sequences supported by karyotype and genome size estimate as well as transcriptome sequencing (Rhie et al. 2021; Read et al. 2017; Tan et al. 2021; Weber et al. 2020; Marra et al. 2019; Zhang et al. 2020).

Chromosome-level analysis is broadening our scope of comparative genomics (Rhie et al. 2021; Deakin et al. 2019). The difficulty of elasmobranch genome sequencing is manifested in the retrieval of the Hox C cluster, an array of homeobox-containing genes (reviewed in Kuraku 2021). While their Hox A, B, and D gene clusters have been reliably assembled with few gaps (Hara et al. 2018; Mulley et al. 2009), assembling the Hox C cluster, which was initially reported as missing from elasmobranch genomes (King et al. 2011), suffered from high GC-content and frequent repetitive elements, resulting in fragmentary sequences (Hara et al. 2018; reviewed in Kuraku 2021). This difficulty is expected to be overcome by the application of a prevailing approach to scaffolding the genomic fragments up to the chromosome scale using chromatin contact data (Yamaguchi et al. 2021; Dudchenko et al. 2017) as well as long-read sequencing.

Another expectation from chromosome-level genome analysis is the identification of sex chromosomes, although sequencing and assembling sex chromosomes often suffer from difficulties caused by high repetitiveness or uneven

sequence depth (Rhie et al. 2021; Ma et al. 2021). While sex determination mechanisms have been revealed by an increasing number of studies on osteichthyan vertebrates (Graves 2016; Pennell et al. 2018), no report is available for vertebrate species outside osteichthyans, namely cyclostomes and chondrichthyans. The quest for sex determination mechanisms can be initiated by the identification of sex chromosomes as already improvised in several vertebrate species (Franchini et al. 2018).

Some vertebrate karyotypes, including most bird karyotypes, consist of small-sized chromosomes, or microchromosomes (Figure 1A) that were initially recognized as shorter than 1 μm in cytogenetic observations (Ohno et al. 1969; Ohno 1971). Microchromosomes are known to have higher GC-content, higher gene density, and different chromatin states compared with the remaining macrochromosomes (International Chicken Genome Sequencing Consortium (ICGSC) 2004; Burt 2002; Waters et al. 2021). Some genome sequence-based studies regard chromosomes smaller than 20 Mb as microchromosomes and investigated their possible common origin (Nakatani et al. 2021; ICGSC 2004), but they have not been unambiguously defined on a cross-species basis. Accumulating information from synteny-based analysis suggests that the last common jawed vertebrate ancestor already possessed microchromosomes (Braasch et al. 2016; Meyer et al. 2021; Simakov et al. 2020; Waters et al. 2021; Nakatani et al. 2007, 2021). However, this hypothesis needs to be examined by incorporating more diverse vertebrate taxa into the comparison, on a solid basis of experimentally validated karyotypic configurations of individual species.

In this study, we focused on the zebra shark *Stegostoma tigrinum* (or leopard shark; Figure 1B) and report its whole-genome sequences for the first time. This

129 species has the smallest genome size (3.71 Gb) among the elasmobranch species
130 whose genomes have been sequenced to date (as of December 2022). Each of the
131 resultant chromosome-scale genome assemblies of the zebra shark, as well as the
132 whale shark (Figure 1B), have been constructed using samples from a single
133 individual (which was not achieved in earlier efforts: Tan et al. 2021; Read et al.
134 2017; Weber et al. 2020) and controlled by referring to their karyotypes. Using the
135 obtained sequences, we performed comparative investigations to characterize the
136 diversity of chromosomal organization.
137
138
139 **Results**
140 **The smallest shark genome sequenced to date**
141 We focused on the zebra shark (or leopard shark) *Stegostoma tigrinum* (formerly, *S.*
142 *fasciatum*) with the haploid nuclear DNA content of 3.79 pg (3.71 Gb; Kadota et al.
143 2023) and the karyotype of 2n = 102 (Uno et al. 2020). Using genomic DNA
144 extracted from blood cells of a female adult, the whole genome was sequenced and
145 assembled with short reads. The resultant sequences were further scaffolded using
146 Hi-C data obtained from blood cell nuclei of the same individual (see Methods). The
147 number of output scaffold sequences longer than 1 Mbp (thereafter tentatively
148 designated ‘chromosomes’) was 50, which closely approximates its chromosome
149 number revealed by the cytogenetic observation and karyotypic organization
150 previously characterized by us using primary cultured cells (Figure 1C). The retrieval
151 of the chromosomal scale was also guaranteed by the N50 scaffold length of 76.6
152 Mb (Figure 1C).

We also performed *de novo* whole genome sequencing of an adult male whale shark, *Rhincodon typus*, using Linked-Read data and Hi-C scaffolding (see Methods). The number of sequences greater than 1 Mbp matched the number of chromosomes in the karyotype (n = 51) (Figure 1C), and the N50 scaffold length of the resultant assembly reached 70.8 Mb, significantly exceeding that of the assemblies previously published for this species (Figure 1C). To date, our product is the only chromosome-scale genome assembly for this species that was built consistently from a single individual.

While we recognize a significant gap between the estimated and retrieved total sequence lengths, genome assemblies for the zebra shark and whale shark exhibited high completeness of protein-coding gene space of more than 90% (Figure 1C). Prediction of protein-coding genes on the zebra shark and whale shark genomes, performed by incorporating homolog sequences and transcriptome data, resulted in 33,222 and 35,334 genes, respectively (Supplementary Tables 1 and 2), which allowed downstream molecular biological analysis.

**Karyotypic trends in sharks**

The obtained zebra shark genome assembly consisted of chromosome sequences of highly variable length with a gradual slope, in accordance with our previous cytogenetic observation (Uno et al. 2020), spanning from 187.0 Mb down to 4.3 Mb (Figure 2A). This pattern does not resemble the length variation in the *Callorhinchus milii* or the chicken (Figure 2A). The chicken especially exhibits a steep slope, marked by a number of chromosomes shorter than 20 Mb (conventionally called microchromosomes) (Figures 1A and 2A). Gradualism in the chromosome length distribution is also observed in the whale shark (Supplementary Figure 1), white-

spotted bamboo shark (Zhang et al. 2020), and thorny skate (Rhie et al. 2021), and is assumed to be typical karyotypic organization of elasmobranchs (Supplementary Figure 2). For simplicity, we tentatively designated (1) zebra shark Chromosome 1 to 14, longer than 70 Mb; (2) Chromosome 15 to 33, between 30 and 70 Mb; and (3) Chromosome 34 to 50, shorter than 30 Mb as elasmobranch macro-, middle-sized-, and micro-chromosomes (abbreviated into eMAC, eMID, and eMIC, respectively) that are differentially colored in Figures 1A and 2B. This categorization was also applied to the whale shark chromosomes (Figure 1A; Supplementary Figure 2).

The zebra shark genome exhibits consistent intra-chromosomal GC-content compared with other species, and its chromosome ends in general, as well as the shorter chromosomes, tend to have relatively higher GC-content (Figure 2A). Zebra shark DNA sequences show a uniformly high frequency of interspersed repeats throughout the genome (Figure 2A). These characteristics are also observed in the whale shark genome (Supplementary Figure 1A), but the co-occurrence of these two features (small intra-chromosomal GC-content variation and uniformly high interspersed repeat frequency) was not explicitly observed in the other species compared (Figure 2A; Supplementary Figure 1A).

**Vertebrate-wide comparisons including elasmobranchs**

To analyze how the karyotypes of these shark species were derived, chromosomal nucleotide sequences were compared between species pairs with variable divergence times (Figure 3A). The comparison between the zebra shark and the whale shark exhibited a high similarity in chromosomal organization with few intrachromosomal breaks (panel 1 in Figure 3A). The high similarity of this

orectolobiform shark species pair suggests high conservation of genomic sequences from around or earlier than 50 million years ago (see Discussion), compared with osteichthyan species pairs in a similar divergence time range—the human-marmoset pair diverged about 43 million years ago (panel 3 in Figure 3A). The relatively high conservation of chondrichthyan chromosome organization is also supported by the comparisons between more distantly related species pairs. The similarity of the zebra shark genome sequences to the thorny skate (panel 4) and the *C. milii* (panel 6) exceeded those for the species pairs with about 300- and 400-million-year divergences, respectively (panel 5 and panels 7 in Figure 3A).

Previous studies sought to reconstruct the process of karyotypic evolution of vertebrates but often lacked elasmobranchs in the dataset (Nakatani et al. 2021; Sacerdot et al. 2018). In the present study, we performed a gnathostome-wide comparison of the syntenic location of one-to-one orthologs, including the zebra shark (Figure 3B; Supplementary Figure 3). A considerable proportion of the one-to-one orthologs are shared between eMAC and large chromosomes of chicken and spotted gar (Figure 3B). The majority of the genomic regions in the zebra shark eMIC were shown not to share one-to-one orthologs with the so-called microchromosomes of the chicken or spotted gar (Figure 3B). Moreover, the smallest spotted gar chromosomes were frequently shown to be homologous to zebra shark eMID (chromosomes 15–33), and not to its eMIC (Chromosomes 34–50) (Figure 3B). It was also shown that zebra shark Chromosomes 10, 25, 26, 27, and 28 have chicken homologs of similar size but their *C. milii* homologs have been fused into larger chromosomes, while zebra shark Chromosomes 2, 3, 4, 5, 6, 8, 11, 18, 20, 23, 33, 34, and 36 are likely to be products of fissions that occurred in the elasmobranch lineage (Figure 3C; Supplementary Figure 3). These results cast doubt on the

common origins of microchromosomes among chicken, spotted gar, and zebra shark and indicate a more drastic reorganization of karyotypes at the base of jawed vertebrates than previously inferred from the comparison involving only the *C. milii* as a cartilaginous fish (Nakatani et al. 2021).

**Length-dependent properties of chromosomes**

Previous studies showed intragenomic heterogeneity of sequence features depending on chromosome lengths in birds and reptiles (Kuraku et al. 2006; Matsubara et al. 2012; Burt 2002; Waters et al. 2021; Srikulnath et al. 2021) To characterize shark chromosomes in depth, base compositions, gene length and density, and molecular evolutionary rates quantified with genic synonymous substitutions ($K_s$) were investigated (Figure 2B). Our statistical tests supported a negative correlation for GC-content, gene density, and $K_s$ with chromosome length, as well as a positive correlation of gene length (Supplementary Table 3; see Methods). This chromosome length-dependent pattern was not supported in human, sea lamprey and teleost fishes, but was supported for multiple features (among GC-content, gene length and density, and $K_s$) in some other vertebrates with large chromosome length heterogeneity including other elasmobranch species (Figure 2; Supplementary Figure 2; Supplementary Table 3). The correlation of GC-content with chromosome length was also observed for the western clawed frog *Xenopus tropicalis* which conventionally is thought to have no microchromosomes recognized cytogenetically (Uno et al. 2012) (Supplementary Figures 2; Supplementary Table 3).

Although this chromosome length-dependent pattern is observed in phylogenetically diverse vertebrate species, some short chromosomes exhibit

exceptionally low GC-content, such as Chromosome 26 of the *C. milii*, Chromosome 29 and 33 of the chicken, and Linkage Group 29 of the spotted gar (Figure 2; Supplementary Figures 1 and 2). These exceptions evoke a caution for the generalization of common characteristics of short chromosomes. It needs to be carefully examined whether the relatively short sequences with exceptionally low GC-content are fragments of large chromosomes that failed to be assembled to a chromosome scale. Such chromosomal scaffold sequences, with small length and relatively low GC-content, may be the cause of the insufficient support for the chromosome length-dependent pattern in some species (Supplementary Table 3).

**Do ‘macrochromosome’ ends resemble microchromosomes?**

We analyzed regional variations within individual chromosomes of diverse vertebrates, including the zebra shark. To further characterize the distinct trend of chromosomal ends indicated in Figure 2, we separated the 1-Mb-long ends from relatively large chromosomes and analyzed the trends of genomic sequences in five vertebrate species (Figure 4A). Intact chromosome ends are known to be occupied by telomeric or subtelomeric simple repeats. To eliminate the effect of those repeats, we focused on the regions harboring protein-coding genes, which recapitulated the higher GC-content in the chromosome ends (Supplementary Figure 4). To examine other characteristics of chromosome ends, we focused on zebra shark and chicken with a large chromosome length variation. Our further comparison consistently revealed an increase of the medians of GC-content, gene density and synonymous substitution rate, as well as a decrease in gene length, in the ends of relatively large chromosomes, compared with their remainders (Figure 4B). In both the zebra shark and chicken, the medians of these features for large chromosome ends were closer

to those for small chromosomes (zebra shark eMIC and chicken MIC) than those for the reminders of large chromosomes (Figure 4B). This pattern is less pronounced in the zebra shark than in chicken, according to the variable support levels from statistical tests (indicated with the number of the symbol ' † ' in Figure 4B; Supplementary Table 4; also see Methods).

**Intra-genomic repetitive element distribution**

We also analyzed the distribution of repetitive elements on chromosomes of variable length. In fact, previous studies yielded equivocal observations. Some of those studies showed a higher abundance of repetitive elements on larger chromosomes (Koochekian et al. 2022), whereas others indicated localization biased toward smaller chromosomes (Hara et al. 2018). So far, no solid gnathostome-wide comparison has been made by taking the difference of repeat classes into account. In the newly obtained shark genome sequences, we separately quantified the sequence proportions identified as interspersed repeats (LINE, SINE, LTR, and DNA elements) and simple tandem repeats (simple repeats and low-complexity DNA sequences, including satellites). In the zebra shark, the interspersed repeat content is positively correlated with chromosome length, while the simple tandem repeat content exhibits a negative correlation (Figure 4C; Supplementary Figure 5; Supplementary Table 3). These patterns were also observed in the chicken and the *C. milii* but in a less pronounced manner, while they were not observed in human (Figure 4C; Supplementary Figure 5; Supplementary Table 3).

As examined above for other characteristics, we dissected the observed chromosome length-dependent trend of repeat distribution, again by isolating the 1-Mb-long ends versus the remainder of relatively large chromosomes (Figure 4D; Supplementary Table 4). In this comparison, we observed a higher content of interspersed repeats in the ends of relatively large chromosomes, namely zebra shark eMAC and chicken MAC (Figure 4D). The higher repeat content was commonly observed in relatively small chromosomes (zebra shark eMIC and chicken MIC), except that interspersed repeat content is reduced in zebra shark eMIC (Figure 4D).

**First genomic characterization of a shark sex chromosome**

So far, there has been no intensive DNA sequence-based characterization of sex chromosomes for chondrichthyan species. Expecting that a sex chromosome will exhibit a distinct male–female ratio of sequencing depth (Palmer et al. 2019), we performed short-read sequencing of the whole genomes for both sexes in zebra shark and whale shark. For these species, our previous cytogenetic analysis did not detect any heteromorphic sex chromosomes (Uno et al. 2020). Our comparison among different chromosomes detected a lower male-to-female sequencing depth ratio of close to 0.5 for Chromosome 41 of both these species (Figure 5A). This suggests male as a heterogametic sex and the XY system for these species, which was validated for zebra shark by genomic qPCR (Figure 5B). Next, we investigated the origin of the putative X Chromosome of these species—was it derived from the same ancestral chromosome that were differentiated later into the sex chromosomes of other vertebrate lineages, particularly of mammals and birds that have long-standing sex chromosomes? Our comparison revealed chromosome-level homology

of the putative X Chromosomes of the zebra shark to a part of human Chromosome 12 and chicken Chromosome 34 (Figure 5C, magenta). Neither the human X, human Y, chicken Z, nor chicken W chromosomes exhibited pronounced homology to the putative zebra shark Chromosome X. The putative shark X Chromosome identified in this study do harbor orthologs of a number of well-studied regulatory factors but do not harbor the orthologs of the master sex determination genes identified in other vertebrates including teleost fishes, that is, *Dmrt1*- or *Sox3*-related transcription factors as well as components of TGF-β signaling pathway, such as *Amh*, *Amhr2*, *Bmpr1b*, *Gsdf*, and *Gdf6* (Bertho et al. 2021).

In the approximately 17 Mb-long sequence of the putative zebra shark Chromosome X, one 1.5 Mb-long end exhibited a male–female sequencing depth ratio of nearly 1.0 (Figure 5D). This region, with a sequence depth comparable to that of the autosomes, is deduced to be a pseudoautosomal region (PAR) that is likely shared between the heterogametic sex chromosomes X and Y (Smeds et al. 2014; Palmer et al. 2019). The identification of PAR was supported by a comparable level of genomic qPCR amplification for multiple genes in this region to that for autosomal regions (Figure 5B). We characterized possible unique patterns of molecular evolution typical of sex chromosomes (Figure 5D; Supplementary Table 5). In the X Chromosome, protein-coding genes exhibited a significant decrease of synonymous substitution rate ($K_s$) supported by small effect and an increased median of non-synonymous substitution rate ($K_a$), resulting in an increased $K_a/K_s$ ratio. Our comparison also revealed higher frequency of low-complexity repeats as well as higher GC-content in the PAR (Figure 5D and E), which is a hallmark of PAR observed in other species resulted from accelerated recombination (Galtier et al. 2001; Galtier 2004; Smeds et al. 2014). The Hi-C contact map, derived from zebra

shark blood, shows subproximal contacts suggesting intermittent chromatin compartmentalization within the PAR and the rest of Chromosome X (Figure 5E).

Our comparison of the ortholog location between the putative X Chromosomes of the two species supported a high cross-species conservation of the chromosome structure (Figure 5E). In the current whale shark assembly, Chromosome 41, which corresponds to the putative chromosome X, may not cover the whole chromosome, possibly excluding one end of the PAR (Figure 5E).

**Identification of Hox C cluster on the putative X Chromosome**

In the newly obtained sequence of the putative zebra shark X Chromosome, we identified an array of orthologs of the non-shark genes encoding homeobox proteins Hox C (Figure 5E). In elasmobranchs, Hox C genes were long thought missing from the genome (King et al. 2011) but later identified in several shark species, as rogue open reading frames (ORFs) flanked by massively repetitive sequences (Hara et al. 2018; reviewed in Kuraku 2021). Our genome-wide gene prediction for the zebra shark detected the ORF of *Hoxc8*, *-c11*, and *-c12*, while the partial ORF of the putative *Hoxc6* ortholog was also identified by a manual search of the raw genomic sequence (Figure 6A). These Hox C genes were located in a 180 kb-long genomic segment in the PAR of the putative Chromosome X (Figure 6A), identified as a single cluster in an elasmobranch fish for the first time. Their orthologies were confirmed with molecular phylogenetic trees, which also indicated elevated molecular evolutionary rates with long branches for elasmobranch Hox C genes (Figure 6B). Our RNA-seq data showed the transcription of these Hox C genes (except for *Hoxc12*) in embryos and juvenile tissues (Figure 6C). The identified zebra shark Hox

C cluster is massively invaded by repetitive elements unlike the other Hox gene clusters (A, B, and D clusters) of this and many other vertebrate species (Figure 6A), as in the Hox C-containing genomic segments of other shark species (Hara et al. 2018). Our search for zebra shark orthologs of the protein-coding genes located near the human Hox C cluster (e.g., *ATF7*, *CBX5*) revealed poor conservation of the gene compositions. Some of the zebra shark orthologs were not identified in its entire genome sequence, suggesting a divergent nature of the genomic regions flanking the Hox C cluster.

## Discussion

In this study, we chose two orectolobiform shark species (zebra shark and whale shark) in Elasmobranchii and characterized their genomic organization with chromosome-scale DNA sequences. This study was achieved in support of epigenome and transcriptome data prepared using fresh tissue samples and previously obtained karyotype information. Our results suggest that their karyotypes are organized by chromosomes of gradual sizes marked with size-dependent sequence properties (Figure 2B). The length gradualism is a remarkable feature of elasmobranch karyotypes, although we tentatively grouped those chromosomes into three length-dependent categories as proposed recently for other elasmobranch species (Marlétaz et al. 2023; Stanhope et al. 2023). The pattern in the shark chromosomal organization is unique in its diversity among vertebrates (Uno et al. 2020), which is characterized by abundant chromosomes (up to 106 for diploids) and variable chromosome sizes. The abundance and highly variable sizes of chromosomes are known for some avian species, but the shark genome

403 organization is distinct from avian counterparts in that shark chromosomes generally
404 have higher repeat content than those of the chicken (Figure 2). In our comparison,
405 the shark karyotype is also characterized by the ratio of the largest and smallest
406 chromosome lengths of 40 to 100, compared with < 10 for most vertebrates, except
407 for species with microchromosomes (Figure 1A). In fact, the length of the shortest
408 chromosomal sequence in typical genome assemblies deposited currently in public
409 databases is often dependent on sequence length cutoff by researchers' decision.
410 Especially for species with no solid karyotypic reference such as *Callorhinchus milii*,
411 the range of sequences considered as chromosomes needs to be carefully
412 examined.

413 Our chromosome-scale genome sequencing and analysis were enabled by
414 access to fresh tissue samples. Because of low accessibility, no previous efforts
415 could provide a set of DNA sequences, karyotypic configuration, and reliable
416 measure of nuclear DNA content for a single chondrichthyan species. Of these, the
417 two latter elements serve as indispensable references to validate the output of
418 sequencing. These requirements are satisfied for both zebra shark and whale shark
419 in our study. Especially for the whale shark, no published studies employed
420 chromatin contact data for Hi-C scaffolding and transcriptome sequencing (Tan et al.
421 2021; Marra et al. 2019; Weber et al. 2020) . In our study, the access to embryos
422 and blood not only enabled chromosome-scale genome scaffolding but also provided
423 transcriptional evidence of most shark Hox C genes that have been shown to exist in
424 a cluster for the first time (Figure 6).

425 Peculiar fractions of vertebrate karyotypes with small sizes and higher GC-
426 content have traditionally been designated as microchromosomes, which usually
427 denote chromosomes shorter than 20 Mb (e.g., Nakatani et al. 2021) but have no

uniform definition (see Introduction). Our investigation, focusing on various aspects of chromosomal DNA sequences, provided a novel view of vertebrate karyotypes that cannot be understood with a simple binary classification, namely with or without microchromosomes, or between macro- and microchromosomes. This view is supported by the common pattern of high intrachromosomal heterogeneity within individual macrochromosomes (Figure 4; Supplementary Figure 4) as well as interchromosomal heterogeneity among different microchromosomes (Figure 2; Supplementary Figures 1 and 2). In particular, the heterogeneity within macrochromosomes, marked with high GC-content, high gene density, and small gene length of their ends, may be shared widely among diverse vertebrates (Figure 4A, B). Intragenomic heterogeneity of GC-content was previously suggested to be caused by GC-biased gene conversion (Mugal et al. 2015). In addition, the uniform numbers of recombination per chromosome (Dumont and Payseur 2008) have been thought to explain the chromosome length-dependent GC-content variation between different chromosomes. Our observations did not show pronounced length-dependent variation of GC-content for chromosomes that were longer than 100 Mb (Figure 2B; Supplementary Figure 2). Taken together, we speculate that the peculiar nature of chromosome ends mainly account for the variation of GC-content among different chromosomes (Figure 4B). Importantly, ‘chromosome ends’ in this context not only harbor telomeric or other simple repeats but also hold complex sequences including protein-coding genes in sequence stretches that are longer than 1 Mb (Figure 4B; Supplementary Figure 4). In the wester clawed frog genome, such sequence stretches marked with elevated GC-content span much longer ranges than in other genomes (Supplementary Figure 1A and 2). The observed features of smaller chromosomes with larger proportions of such ‘ends’ in length are more

affected than those of longer chromosomes, which likely explains the length-dependent nature of chromosomes. The nature of macrochromosome ends (e.g., with higher GC -content) is thought to be a remnant of the fusion of one or more microchromosome(s) to a macrochromosome (Waters et al. 2021). This hypothesis is not supported by our observation that even species possessing no explicit microchromosomes (e.g., western clawed frog) have chromosome ends with the peculiar nature of DNA sequences (Supplementary Figures 1A and 4). No close relatives of the western clawed frog have been shown to possess microchromosomes; Tymowska 1991), and thus microchromosome fusions cannot account for the characteristics of their chromosome ends with higher GC-content.

Our genome-wide sequencing depth investigation covering both sexes revealed the XY system for the two studied shark species and enabled the first sequence-based identification of shark sex chromosomes (namely, Chromosome X; Figure 5A). The genes on the shark Chromosome X tend to exhibit larger non-synonymous substitution rates ($K_a$) and smaller synonymous substitution rates ($K_s$) than those on autosomes, resulting in a higher $K_a/K_s$ ratio (Figure 5D). This resembles the pattern known in other species including birds and insects, which is known as the ‘Faster X (or Faster Z) hypothesis’ (Meisel and Connallon 2013; Charlesworth et al. 2018; Xu et al. 2019; Mank et al. 2007). The smaller $K_s$ value is also indicative of the support for the male-driven evolution hypothesis (Li 2002; Miyata et al. 1987), which is to be examined by higher $K_s$ values for genes on the presumptive Y chromosome that remains unidentified. Our comparison of protein-coding gene compositions showed that the zebra shark and the whale shark largely share Chromosome X that is homologous to each other (Figure 5E). The Chromosome X harbors the Hox C cluster that was previously shown to be highly

478 divergent and degenerative (Hara et al. 2018), but its localization in the PAR (Figure
479 5E) suggests a balanced dosage of the Hox C genes and their expressions between
480 males and females.

481 The whale shark is known as the largest extant ‘fish’, and one of its extant
482 closest relatives is the zebra shark (Naylor et al. 2012). In elasmobranch evolution,
483 the lineages leading to these two species diverged no later than 48.6 million years
484 ago (Long 1992). The high similarity of the chromosome-scale sequence
485 organization (panel 1 in Figure 3; Supplementary Figure 3) as well as the gene
486 compositions on the Chromosome X between these two species (Figure 5E)
487 indicates a lower rate of chromosomal rearrangement in these lineages compared
488 with those of species pairs of similar divergence times in other vertebrate lineages
489 (Figure 3A). Among vertebrates, mammals and birds have relatively long-standing
490 sex chromosomes (X/Y and Z/W, respectively) shared throughout these individual
491 taxa that arose more than 48.6 million years ago (Long 1992). Although still limited in
492 number, sex chromosomes have been identified in some elasmobranch species by
493 cytogenetic analyses, all of which have a male-heterogametic system (Uno et al.
494 2020). Our study with genome sequencing showed that it also holds for the zebra
495 shark and whale shark. Sharks, or a phylogenetically wider subset of cartilaginous
496 fishes, may possess even older sex chromosomes, depending on phylogenetic
497 prevalence of their homologs in more distantly related shark and even ray species.

498 Our synteny analysis yielded novel insights into genome evolution
499 encompassing the whole diversity of vertebrates. It showed homology of the shark
500 Chromosome X to human Chromosome 12 and chicken Chromosome 34 (Figure
501 5C), which suggests a difference in ancestral autosomes that were adopted as sex
502 chromosome between elasmobranchs and other vertebrates for which sex

chromosomes have been characterized. The synteny analysis involving sharks also provided clues to the origin of microchromosomes. It suggests that eMIC, the shark's small chromosomes, are homologous to several of the large chromosomes in both chicken and spotted gar and are not necessarily homologous to their microchromosomes (Figure 3B). Also, microchromosomes of the chicken and gar were not shown to be homologous to eMICs (Figure 3B). It is crucial for any effort for reconstructing the diversity of chromosome organization in vertebrates to incorporate diverse elasmobranchs into the comparisons.

## Methods

### Animals

Fresh blood from a female adult zebra shark (total length, 2.2 m); Individual ID, sSteFas1 (also called F1) and a male adult whale shark (total length, 8.8 m; Individual ID, sRhiTyp1) were sampled at Okinawa Churaumi Aquarium and used for the preparation of whole-genome shotgun DNA libraries, Hi-C libraries, and RNA-seq libraries as well as for measuring nuclear DNA content by flow cytometry. Likewise, fresh blood of a male zebra shark (total length, 2.1 m; Individual ID, sSteFas2 [also called M1]) and a female whale shark (total length, 8.0 m; Individual ID, sRhiTyp2) were sampled and used for quantifying the male/female ratios of individual chromosomal regions. Extraction of ultra-high molecular weight DNA was performed by collecting blood cells by centrifugation, and the collected cells were embedded in agarose plugs ($4.0 \times 10^5$ cells/plug). The agarose gel plugs were prepared and processed with the CHEF Mammalian Genomic DNA Plug Kit (BioRad, Cat. No. #1703591). Total RNAs used to construct RNA-seq libraries were extracted from various tissues of a female juvenile zebra shark (total length, 30 cm) born at

Okinawa Churaumi Aquarium and a female juvenile whale shark (total length, 7.7 m; Individual ID, sRhiTyp3) (Supplementary Table 6). These animals were introduced into the aquarium in accordance with local regulations before those species were assessed as endangered. Animal handling and sample collections at the aquarium were conducted by veterinary staff without restraining the individuals under the experiment ID AT19002 approved by Institutional Animal Care and Use Committee of Okinawa Churashima Foundation in accordance with the Husbandry Guidelines approved by the Ethics and Welfare Committee of Japanese Association of Zoos and Aquariums. All other experiments were conducted in accordance with the Guideline of the Institutional Animal Care and Use Committee (IACUC) of RIKEN Kobe Branch (Approval ID: H16-11).

**Genome sequencing and scaffolding**

For a female zebra shark, paired-end and mate-pair DNA libraries for *de novo* genome sequencing were prepared and sequenced as previously described (Hara et al. 2018; Yamaguchi et al. 2021). The amount of starting DNA and numbers of PCR cycles for the library preparation are included in Supplementary Table 6. The total sequencing coverage amounted to 95.8 times the genome size based on the reference measured previously by flow cytometry (3.71 Gb; Kadota et al., in preparation). Low-quality bases from paired-end reads were removed by TrimGalore v0.6.6 (https://github.com/FelixKrueger/TrimGalore, Accessed 4 Jan 2019) with the options '--stringency 2 --quality 20 --length 25 --paired --retain_unpaired'. As described previously (Hara et al. 2018), short-read assembly of the zebra shark, as well as scaffolding with mate-pair reads followed by gap closure, was performed using Platanus v1.2.4 (Kajitani et al. 2014).

Whole genome sequencing for a male whale shark employed the 10x Genomics Chromium to produce Linked-Read data. A DNA library was prepared using 12 ng of gDNA extracted from blood cells according to the user guide of the Chromium Genome Library Kit v2 Chemistry using the Chromium Genome Library Kit & Gel Bead Kit v2 (10x Genomics, Cat. No. #120258) and the Chromium Genome Chip Kit v2 (10x Genomics, Cat. No. #120257). The library was sequenced on a HiSeq X (Illumina) platform to obtain 151 nt-long paired-end reads. Sequence assembly using the Linked-Read data of 46.4 times the genome size was performed with the program Supernova v2.0 (Weisenfeld et al. 2017). The resultant sequences were subjected to scaffolding with the program P_RNA_scaffolder (commit 7941e0f in GitHub) (Zhu et al. 2018) using the result of the alignment of transcriptome sequence reads (obtained as described below) performed with the program HISAT2 v2.1.0 onto those genome sequences (Kim et al. 2019).

**Hi-C data production and chromosome-scale genome scaffolding**

Hi-C libraries of the zebra shark and whale shark were constructed using restriction enzymes DpnII and HindIII, respectively, as previously reported (Kadota et al. 2020). Blood cells collected as described above were fixed in 1% formaldehyde solution. A fixed tissue containing 10 μg of DNA was used for the preparation of Hi-C DNA via *in situ* restriction digestion and ligation. The Hi-C library was prepared using 2 μg of the ligated DNA with five cycles of PCR amplification. Quality controls of the ligated DNA and the Hi-C libraries were performed as described previously (Kadota et al. 2020).

Each of the zebra shark and the whale shark genome assemblies was used for Hi-C read mapping with Juicer v1.5 (Durand et al., 2016a) and chromosome-scale scaffolding with the program 3d-dna (v180922) (Dudchenko et al. 2017). In the

scaffolding, three different lengths were tested (5, 10, and 15 kb) for the option '-i' defining the input sequence length threshold. For each species, the three resulting scaffolding outputs, as well as the original assembly before Hi-C scaffolding, were assessed based on sequence length distribution and protein-coding gene completeness. Among all the scaffolding outputs compared, the output with the option '-i 10000' was judged to be optimal and was subjected to a "review" of the scaffolding results on Juicebox v1.11.08 (Durand et al. 2016)b) to minimize inconsistent signals of chromatin contacts (Supplementary Figure 6). The review was facilitated by referring to nucleotide sequence-level similarity between different scaffolding outputs visualized by SyMAP v5.0 (Soderlund et al. 2011). After the review, the sequences judged as contaminants from other organisms were removed, as previously reported (Hara et al. 2018).

**Repeat identification**

To obtain a species-specific repeat library, RepeatModeler v2.0.2a was run on the genome assembly of the individual species with default parameters (Smit and Hubley 2008). Detection of repeat elements in the genome was performed by RepeatMasker v4.1.2-p1 (Smit et al. 2013) with RMBlast v2.6.0+, using the species-specific repeat library obtained above. For quantification of the content of interspersed repeats and simple tandem repeats, RepeatMasker was run separately with the options '-nolow -norna' and '-noint -norna', respectively.

**Gene model construction**

The program Braker v2.1.6 was employed for gene prediction by inputting the results of RNA-seq read mapping to a genome assembly in which repetitive sequences are

603 soft-masked by RepeatMasker with the options '-nolow -xsmall', as well as amino
604 acid sequences of closely related species as homolog hints (Brůna et al. 2021; Smit
605 et al. 2013). To build the homolog hints based on the amino acid sequences, we
606 used the previously reported amino acid sequence sets of the brownbanded bamboo
607 shark and the whale shark.
608 Gene space completeness in Figure 1C was obtained by the BUSCO pipeline
609 ver. 5 (Seppey et al. 2019) using the BUSCO's Vertebrata ortholog set.
610

611 **Synonymous and non-synonymous substitution quantification**

612 In order to calculate the number of synonymous substitutions per synonymous site
613 ($K_s$) and the number of nonsynonymous substitutions per non-synonymous site ($K_a$),
614 the 1-to-1 orthologs shared by the four elasmobranch species (zebra shark, whale
615 shark, brownbanded bamboo shark, and cloudy catshark) were selected by
616 SonicParanoid v1.3.4 (Cosentino and Iwasaki 2019) as follows. First, peptide
617 sequences of the retrieved orthologs were aligned with MAFFT v7.475 with the
618 option '-linsi' (Katoh and Standley 2013). The individual alignments were trimmed
619 and back-translated into nucleotides with trimAl v1.4.rev15 with the options '-
620 automated1 -backtrans' followed by the removal of gapped sites using trimAlL with
621 the option '-nogaps' (Capella-Gutiérrez et al. 2009). Ortholog groups containing
622 fewer than 100 aligned codons or a stop codon were discarded. For the selected
623 ortholog groups, $K_s$ and $K_a$ were computed with yn00 in the PAML v4.9c88 (Yang
624 2007). Computed values larger than 0.01 and smaller than 99 were included in the
625 results (Figures 2B, 4B, and 5D).
626

627 **RNA-seq and transcriptome data processing**

Total RNAs were extracted with TRIzol reagent (Thermo Fisher Scientific). Quality control of the RNA treated with DNase I was performed with Bioanalyzer 2100 (Agilent Technologies). Libraries were prepared with TruSeq RNA Sample Prep Kit (Illumina) or TruSeq Stranded mRNA LT Sample Prep Kit (Illumina) as previously described (Hara et al. 2018). The amount of starting total RNA and numbers of PCR cycles are included in Supplementary Table 6. To remove adaptor sequences and low-quality bases, the obtained sequence reads were trimmed with TrimGalore v0.6.6 as outlined above, and *de novo* transcriptome assembly was performed with the program Trinity v2.11.0 with the option '--SS_lib_type RF' (Grabherr et al. 2011). The trimmed RNA-seq reads were aligned to the genome assembly using the program HISAT2 v2.1.1 (Kim et al. 2019), which was followed by gene expression quantification with StringTie v2.0.6 (Pertea et al. 2016).

**Conserved synteny detection**

Characterization of chromosomal homology among different species employed predicted protein sequence datasets available at the NCBI RefSeq database. After alternative splicing variants were removed, one-to-one orthologs were selected by SonicParanoid v1.3.4 with the option 'most-sensitive' (Cosentino and Iwasaki 2019). Conserved synteny between species was visualized based on single copy 1-to-1 orthologs using RIdeogram (Hao et al. 2020).

**Statistical analysis**

Relationships of chromosome lengths with chromosome sequence features were tested with the correlation coefficient through Pearson's correlation analysis (Figures

652 2B, 4C, Supplementary Figures 2, and 5), and the details are included in
653 Supplementary Table 3.
654 Statistical significance of between-group differences in the box plots in
655 Figures 4B, 4D, and 5D was tested with the nonparametric Kruskal-Wallis test and
656 the Mann-Whitney *U* test, and the detailed results are included in Supplementary
657 Tables 4 and 5. The Kruskal-Wallis test was used to evaluate significance of
658 difference among groups, the Mann-Whitney *U* test was used to evaluate the
659 significance of difference between two groups, and Rank-biserial correlation was
660 calculated as effect size. Bonferroni correction was performed for multiple
661 comparisons. A large number of samples results in a smaller standard error of the
662 mean, which is more likely to be significant. Therefore, the degree of difference
663 between two groups was evaluated not only by *p*-value, but also by effect size.
664
665 **Molecular phylogeny inference**
666 Amino acid sequences were retrieved from aLeaves (Kuraku et al. 2013). Multiple
667 sequence alignment was performed with MAFFT with the option '-linsi'. The aligned
668 sequence sets were processed using trimAl v1.4 rev15 with the option '-automated1'
669 (Capella-Gutiérrez et al. 2009). This was followed by another trimAl run with the
670 option '-nogaps'. Molecular phylogenetic trees were inferred by RAxML with the '-m
671 PROTCATWAG -f a -# 1000' options unless stated otherwise (Stamatakis 2014).
672 Tree inference in the Bayesian framework was performed with the program
673 PhyloBayes v4.1c with the options '-cat -dgam 4 -wag -nchain 2 1000 0.3 50' unless
674 stated otherwise. This was followed by an execution of bpcomp in the PhyloBayes
675 v4.1c package with the option '-x 100' (Lartillot et al. 2009). The support values at
676 the nodes of molecular phylogenetic trees included are, in order, bootstrap values

and Bayesian posterior probabilities. The latter was shown only when the relationship at the node in the visualized tree was supported by the Bayesian inference.

**Identification of shark X Chromosome**

To identify a chromosome-scale scaffold with a distinct male–female sequencing depth ratio in the zebra shark, the same number of trimmed genomic shotgun reads (293,686,584) was prepared for both sexes. The reads were mapped with BWA-MEM2 (v2.2.1) onto a genome assembly (Li and Durbin 2009). Mapped reads were counted for male and female using the bamtobed program in the package BEDTools v2.29.2, for each scaffold in 10 kb non-overlapping windows, and male–female ratios were calculated (Quinlan and Hall 2010). Sequencing depth of male and female reads on the identified Chromosome X was also calculated for 10 kb non-overlapping windows. Windows with the proportion of ambiguous bases of more than 50% were excluded from computation. This procedure was also applied to the whale shark, using 220,785,436 trimmed reads.

The sequencing-based identification of the putative zebra shark Chromosome X was validated with multiple male and female individuals by real-time quantitative PCR (qPCR). The PCR was performed with the Luna Universal qPCR Master Mix (New England Biolabs) according to the manufacturer's instruction using the DNA template of 68.4 ng and oligonucleotide primers designed to avoid intronic regions (Supplementary Table 7). The reaction was performed in triplication on a CFX96 Real-time PCR System (Bio-Rad) with pre-heating at 95 °C for 1 minute and two-step cycling (40 cycles) of denaturing at 95 °C for 15 seconds and annealing/extension at 60 °C for 30 seconds, followed by post-amplification step for

dissociation curve analysis. Male-female difference was quantified with the $2^{-\Delta\Delta Ct}$ method (Rao et al. 2013) as demonstrated conventionally in validating sex difference (e.g., Sheffer et al. 2022).

## Data access

All raw and processed sequencing data generated in this study have been submitted to the NCBI Genome (https://www.ncbi.nlm.nih.gov/genome/) under accession numbers JAHMAH000000000 and JAFIRC000000000 and the NCBI BioProject database under accession number PRJNA703743.

## Competing interests

The authors declare no conflict of interests.

## Acknowledgments

We thank Tomoyuki Furuyashiki and Masayuki Taniguchi for their cooperation in obtaining Linked-Read sequence data, Chiharu Tanegashima and Kaori Tatsumi for their support in sequence data acquisition, Hatsune Makino-Itou for qPCR experiments, animal caretakers including Yano Nagisa at Okinawa Churaumi Aquarium, Keisuke Yonehara, and Akane Kawaguchi for their assistance, and Yukiko Imai and Taiki Niwa for valuable discussion. Computations were partially performed on the NIG supercomputer at ROIS National Institute of Genetics. This study was funded by intramural budgets granted by RIKEN and the National Institute of Genetics, as well as JSPS KAKENHI Grant numbers. 20H03269 and 22K15088.

## References

Bertho S, Herpin A, Schartl M, Guiguen Y. 2021. Lessons from an unusual vertebrate sex-determining gene. *Philos Trans R Soc B: Biol Sci* **376**: 20200092.

Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, Amores A, Desvignes T, Batzel P, Catchen J, et al. 2016. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat Genet* **48**: 427–437.

Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform* **3**: 1–11.

Burt DW. 2002. Origin and evolution of avian microchromosomes. *Cytogenet Genome Res* **96**: 97–112.

Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **6**: e4958.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.

Charlesworth B, Campos JL, Jackson BC. 2018. Faster-X evolution: theory and evidence from Drosophila. *Mol Ecol* **27**: 3753–3771.

Cosentino S, Iwasaki W. 2019. SonicParanoid: fast, accurate and easy orthology inference. *Bioinformatics* **35**: 149–151.

749 Deakin JE, Potter S, O'Neill R, Ruiz-Herrera A, Cioffi MB, Eldridge MDB, Fukui
750 K, Marshall Graves JA, Griffin D, Grutzner F, et al. 2019. Chromosomics:
751 bridging the gap between genomes and chromosomes. *Genes* **10**: 627.
752 Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC,
753 Shamim MS, Machol I, Lander ES, Aiden AP, et al. 2017. *De novo*
754 assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-
755 length scaffolds. *Science* **356**: 92–95.
756 Dumont BL, Payseur BA. 2008. Evolution of the genomic rate of recombination
757 in mammals. *Evolution* **62**: 276–294.
758 Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL.
759 2016a. Juicer provides a one-click system for analyzing loop-resolution Hi-
760 C experiments. *Cell Syst* 3: 95-8.
761 Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden
762 EL. 2016b. Juicebox provides a visualization system for Hi-C contact maps
763 with unlimited zoom. *Cell Syst* **3**: 99–101.
764 Franchini P, Jones JC, Xiong P, Kneitz S, Gompert Z, Warren WC, Walter RB,
765 Meyer A, Schartl M. 2018. Long-term experimental hybridisation results in
766 the evolution of a new sex chromosome in swordtail fish. *Nat Commun* **9**:
767 5136.Galtier N. 2004. Recombination, GC-content and the human
768 pseudoautosomal boundary paradox. *Trends in Genetics* **20**: 347–349.

769 Galtier N, Piganeau G, Mouchiroud D, Duret L. 2001. GC-content evolution in
770 mammalian genomes: the biased gene conversion hypothesis. *Genetics*
771 **159**: 907–911.
772 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X,
773 Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome
774 assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*
775 **29**: 644–652.
776 Graves JAM. 2016. Evolution of vertebrate sex chromosomes and dosage
777 compensation. *Nat Rev Genet* **17**: 33–46.
778 Hao Z, Lv D, Ge Y, Shi J, Weijers D, Yu G, Chen J. 2020. *RIdeogram*: drawing
779 SVG graphics to visualize and map genome-wide data on the idiograms.
780 *PeerJ Comput Sci* **6**: e251.
781 Hara Y, Yamaguchi K, Onimaru K, Kadota M, Koyanagi M, Keeley SD, Tatsumi
782 K, Tanaka K, Motone F, Kageyama Y, et al. 2018. Shark genomes provide
783 insights into elasmobranch evolution and the origin of vertebrates. *Nat Ecol*
784 *Evol* **2**: 1761-1771.
785 Hardie DC, Hebert PDN. 2004. Genome-size evolution in fishes. *Can J Fish*
786 *Aquat Sci* **61**: 1636–1646.
787 International Chicken Genome Sequencing Consortium. 2004. Sequence and
788 comparative analysis of the chicken genome provide unique perspectives
789 on vertebrate evolution. *Nature* **432**: 695–716.

790 Kadota M, Nishimura O, Miura H, Tanaka K, Hiratani I, Kuraku S. 2020.
791 Multifaceted Hi-C benchmarking: what makes a difference in chromosome-
792 scale genome scaffolding? *Gigascience* **9**: giz158.
793 Kadota M, Tatsumi K, Yamaguchi K, Uno Y, Kuraku S. 2023. Shark and ray
794 genome size estimation: methodological optimization for inclusive and
795 controllable biodiversity genomics. *bioRxiv*. 2023.02.23.529029
796 Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M,
797 Harada M, Nagayasu E, Maruyama H, et al. 2014. Efficient de novo
798 assembly of highly heterozygous genomes from whole-genome shotgun
799 short reads. *Genome Res* **24**: 1384–1395.
800 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software
801 version 7: improvements in performance and usability. *Mol Biol Evol* **30**:
802 772–780.
803 Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome
804 alignment and genotyping with HISAT2 and HISAT-genotype. *Nat*
805 *Biotechnol* **37**: 907–915.
806 King BL, Gillis JA, Carlisle HR, Dahn RD. 2011. A natural deletion of the HoxC
807 cluster in elasmobranch fishes. *Science* **334**: 1517.
808 Knief U, Forstmeier W. 2016. Mapping centromeres of microchromosomes in
809 the zebra finch (*Taeniopygia guttata*) using half-tetrad analysis.
810 *Chromosoma* **125**: 757-768.

Koochekian N, Ascanio A, Farleigh K, Card DC, Schield DR, Castoe TA, Jezkova T. 2022. A chromosome-level genome assembly and annotation of the desert horned lizard, *Phrynosoma platyrhinos*, provides insight into chromosomal rearrangements among reptiles. *Gigascience* **11**: giab098.

Kuraku S. 2021. Shark and ray genomics for disentangling their morphological diversity and vertebrate evolution. *Dev Biol* **477**: 262–272.

Kuraku S, Ishijima J, Nishida-Umehara C, Agata K, Kuratani S, Matsuda Y. 2006. cDNA-based gene mapping and GC3 profiling in the soft-shelled turtle suggest a chromosomal size-dependent GC bias shared by sauropsids. *Chromosome Res* **14**: 187–202.

Kuraku S, Zmasek CM, Nishimura O, Katoh K. 2013. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Res* **41**: 22–28.

Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**: 2286–2288.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Li W-H. 2002. Male-driven evolution. *Curr Opin Genet Dev* **12**: 650–656.

Long DJ. 1992. Sharks from the La Meseta Formation (Eocene), Seymour Island, Antarctic Peninsula. *J Vertebr Paleontol* **12**: 11-32.

Ma W, Xu L, Hua H, Chen M, Guo M, He K, Zhao J, Li F. 2021. Chromosomal-level genomes of three rice planthoppers provide new insights into sex chromosome evolution. *Mol Ecol Resour* **21**: 226–237.

Mank JE, Axelsson E, Ellegren H. 2007. Fast-X on the Z: rapid evolution of sex-linked genes in birds. *Genome Res* **17**: 618–624.

Marlétaz F, de la Calle-Mustienes E, Acemel RD, Paliou C, Naranjo S, Martínez-García PM, Cases I, Sleight VA, Hirschberger C, Marcet-Houben M, et al. 2023. The little skate genome and the evolutionary emergence of wing-like fins. *Nature* **616**: 495–503.

Marra NJ, Stanhope MJ, Jue NK, Wang M, Sun Q, Bitar PP, Richards VP, Komissarov A, Rayko M, Kliver S, et al. 2019. White shark genome reveals ancient elasmobranch adaptations associated with wound healing and the maintenance of genome stability. *Proc Natl Acad Sci USA* **116**: 4446–4455.

Matsubara K, Kuraku S, Tarui H, Nishimura O, Nishida C, Agata K, Kumazawa Y, Matsuda Y. 2012. Intra-genomic GC heterogeneity in sauropsids: evolutionary insights from cDNA mapping and GC3 profiling in snake. *BMC Genomics* **13**: 1-14.

Meisel RP, Connallon T. 2013. The faster-X effect: integrating theory and data. *Trends Genet* **29**: 537–544.

Meyer A, Schloissnig S, Franchini P, Du K, Woltering JM, Irisarri I, Wong WY, Nowoshilow S, Kneitz S, Kawaguchi A, Fabrizius A, Xiong P, Dechaud C, Spaink HP, Volff JN, Simakov O, Burmester T, Tanaka EM, Schartl M. 2021. Giant lungfish genome elucidates the conquest of land by vertebrates. *Nature* **590**: 284–289.

Miyata T, Hayashida H, Kuma K, Mitsuyasu K, Yasunaga T. 1987. Male-driven molecular evolution: a model and nucleotide sequence analysis. *Cold Spring Harb Symp Quant Biol* **52**: 863–867.

Mugal CF, Weber CC, Ellegren H. 2015. GC-biased gene conversion links the recombination landscape and demography to genomic base composition. *BioEssays* **37**: 1317–1326.

Mulley JF, Zhong Y-F, Holland PW. 2009. Comparative genomics of chondrichthyan Hoxa clusters. *BMC Evol Biol* **9**: 218.

Nakatani Y, Shingate P, Ravi V, Pillai NE, Prasad A, McLysaght A, Venkatesh B. 2021. Reconstruction of proto-vertebrate, protocyclostome and proto-gnathostome genomes provides new insights into early vertebrate evolution. *Nat Commun* **12**: 4489.

Nakatani Y, Takeda H, Kohara Y, Morishita S. 2007. Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome Res* **17**: 1254–1265.

Naylor G, Caira K, Jensen K, White W, Last P. 2012. A DNA sequence–based approach to the identification of shark and ray species and its implications for global elasmobranch diversity and parasitology *Bull Am Mus Nat Hist* **367**: 1-263.

Ohno S. 1970. Evolution by gene duplication. Springer Science & Business Media, Berlin.

Ohno S, Muramoto J, Stenius C, Christian L, Kittrell WA, Atkin NB. 1969. Microchromosomes in holocephalian, chondrostean and holostean fishes. *Chromosoma* **26**: 35–40.

Palmer DH, Rogers TF, Dean R, Wright AE. 2019. How to identify sex chromosomes and their turnover. *Mol Ecol* **28**: 4709–4724.

Pennell MW, Mank JE, Peichel CL. 2018. Transitions in sex determination and sex chromosomes across vertebrate species. *Mol Ecol* **27**: 3950–3963.

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* **11**: 1650–1667.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

Read TD, Petit RA, Joseph SJ, Alam T, Weil MR, Ahmad M, Bhimani R, Vuong JS, Haase CP, Webb DH, et al. 2017. Draft sequencing and assembly of

the genome of the world's largest fish, the whale shark: *Rhincodon typus* smith 1828. *BMC Genomics* **18**: 755.

Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Kim J. 2021. Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592**: 737–746.

Sacerdot C, Louis A, Bon C, Berthelot C, Roest Crollius H. 2018. Chromosome evolution at the origin of the ancestral vertebrate genome. *Genome Biol* **19**: 166.

Schwartz FJ, Maddock MB. 1986. Comparisons of karyotypes and cellular DNA contents within and between major lines of elasmobranchs. In *Indo-Pacific fish biology*, pp. 148–157. Ichthyological Soc. of Japan.

Schwartz FJ, Maddock MB. 2002. Cytogenetics of the elasmobranchs: genome evolution and phylogenetic implications. *Mar Freshw Res* **53**: 491–502.

Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. In *Gene Prediction Methods in Molecular Biology* (ed. Kollmar M), pp. 227–245. Humana Press, New York.

Sheffer MM, Cordellier M, Forman M, Grewoldt M, Hoffmann K, Jensen C, Kotz M, Král J, Kuss AW, Líznarová E, et al. 2022. Identification of sex chromosomes using genomic and cytogenetic methods in a range-expanding spider, *Argiope bruennichi* (Araneae: Araneidae). *Biological Journal of the Linnean Society* **136**: 405–416.

912 Simakov O, Marlétaz F, Yue JX, O'Connell B, Jenkins J, Brandt A, Calef R,
913 Tung CH, Huang TK, Schmutz J, Satoh N, Yu JK, Putnam NH, Green RE,
914 Rokhsar DS. 2020. Deeply conserved synteny resolves early events in
915 vertebrate evolution. *Nat Ecol Evol* **4**: 820–830.
916 Smeds L, Kawakami T, Burri R, Bolivar P, Husby A, Qvarnström A, Uebbing S,
917 Ellegren H. 2014. Genomic identification and characterization of the
918 pseudoautosomal region in highly differentiated avian sex chromosomes.
919 *Nat Commun* **5**: 5448.
920 Smit AFA, Hubley R. 2008. RepeatModeler Open-1.0 http://www. repeatmasker.
921 org.
922 Smit AFA, Hubley R, Green P. 2013. RepeatMasker Open-4.0.
923 http://www.repeatmasker.org.
924 Soderlund C, Bomhoff M, Nelson WM. 2011. SyMAP v3.4: A turnkey synteny
925 system with application to plant genomes. *Nucleic Acids Res* **39**: e68–e68.
926 Srikulnath K, Ahmad SF, Singchat W, Panthum T. 2021. Why do some
927 vertebrates have microchromosomes? *Cells* **10**: 1–33.
928 Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and
929 post-analysis of large phylogenies. *Bioinformatics* **30**: 1312-1313.
930 Stanhope MJ, Ceres KM, Sun Q, Wang M, Zehr JD, Marra NJ, Wilder AP, Zou
931 C, Bernard AM, Pavinski-Bitar P, et al. 2023. Genomes of endangered
932 great hammerhead and shortfin mako sharks reveal historic population
933 declines and high levels of inbreeding in great hammerhead. *iScience* **26**:
934 105815.

Suryamohan K, Krishnankutty SP, Guillory J, Jevit M, Schröder MS, Wu M, Kuriakose B, Mathew OK, Perumal RC, Koludarov I, et al. 2020. The Indian cobra reference genome and transcriptome enables comprehensive identification of venom toxins. *Nat Genet* **52**: 106-117.

Tan M, Redmond AK, Dooley H, Nozu R, Sato K, Kuraku S, Koren S, Phillippy AM, Dove ADM, Read TD. 2021. The whale shark genome reveals patterns of vertebrate gene family evolution. *Elife* **10**: e65394.

Tymowska J. 1991. Polyploidy and cytogenetic variation in frogs of the genus *Xenopus*. In *Amphibian cytogenetics and evolution* (eds. Green DM, Sessions SK), pp. 259–297, Academic Press, Cambridge.

Uno Y, Nishida C, Tarui H, Ishishita S, Takagi C, Nishimura O, Ishijima J, Ota H, Kosaka A, Matsubara K, et al. 2012. Inference of the protokaryotypes of amniotes and tetrapods and the evolutionary processes of microchromosomes from comparative gene mapping. *PLoS One* **7**: e53027.

Uno Y, Nozu R, Kiyatake I, Higashiguchi N, Sodeyama S, Murakumo K, Sato K, Kuraku S. 2020. Cell culture-based karyotyping of orectolobiform sharks for chromosome-scale genome analysis. *Commun Biol* **3**: 652.

Venkatesh B, Lee AP, Ravi V, Maurya AK, Lian MM, Swann JB, Ohta Y, Flajnik MF, Sutoh Y, Kasahara M, et al. 2014. Corrigendum: elephant shark

954 genome provides unique insights into gnathostome evolution. *Nature* **513**:
955 574.
956 Waters PD, Patel HR, Ruiz-Herrera A, Alvarez-Gonzalez L, Lister NC, Simakov
957 O, Ezaz T, Kaur P, Frere C, Grutzner F, et al. 2021. Microchromosomes are
958 building blocks of bird, reptile, and mammal chromosomes. *Proc Natl Acad*
959 *Sci USA* **118**: e2112494118.
960 Weber JA, Park SG, Luria V, Jeon S, Kim HM, Jeon Y, Bhak Y, Jun JH, Kim SW,
961 Hong WH, et al. 2020. The whale shark genome reveals how genomic and
962 physiological properties scale with body size. *Proc Natl Acad Sci USA* **117**:
963 20662–20671.
964 Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB. 2017. Direct
965 determination of diploid genome sequences. *Genome Res* **27**: 757-767.
966 Xu L, Wa Sin SY, Grayson P, Edwards S v, Sackton TB. 2019. Evolutionary
967 dynamics of sex chromosomes of paleognathous birds. *Genome Biol Evol*
968 **11**: 2376–2390.
969 Yamaguchi K, Kadota M, Nishimura O, Ohishi Y, Naito Y, Kuraku S. 2021.
970 Technical considerations in Hi-C scaffolding and evaluation of
971 chromosome-scale genome assemblies. *Mol Ecol* **30**: 5923–5934.
972 Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol*
973 *Evol* **24**: 1586–1591.

974 Zhang Y, Gao H, Li H, Guo J, Ouyang B, Wang M, Xu Q, Wang J, Lv M, Guo X,
975 et al. 2020. The white-spotted bamboo shark genome reveals chromosome
976 rearrangements and fast-evolving immune genes of cartilaginous fish.
977 *iScience* **23**: 101754.
978 Zhu B-H, Xiao J, Xue W, Xu G-C, Sun M-Y, Li J-T. 2018. P_RNA_scaffolder: a
979 fast and accurate genome scaffolder using paired-end RNA-sequencing
980 reads. *BMC Genomics* **19**: 175.

981

982 **Figure 1**. Shark species studied and comparative statistics of their genome
983 assemblies. (*A*) The karyotypes of diverse vertebrate species are depicted with the
984 length of individual chromosomal DNA sequences. The maximum and minimum
985 chromosome lengths are based on the records in NCBI Genomes.
986 Microchromosomes for osteichthyans are shown in light blue according to individual
987 original reports (Nakatani et al. 2021; Suryamohan et al. 2020; Knief and Forstmeier
988 2016; ICGSC 2004), while the eMID and eMIC of the two shark species whose
989 genomes have been sequenced in the present study (see Results) are shown in
990 magenta and light blue, respectively. (*B*) Zebra shark (left) and whale shark (right).
991 (*C*) Statistics of the genome assemblies. The identifiers of the genome assemblies,
992 as well as statistical comparison of more metrics, are included in Supplementary
993 Table 1. Gene space completeness shows the proportions of selected one-to-one
994 protein-coding orthologs with ‘complete’ (green), ‘duplicated’ (yellow), and
995 ‘fragmented’ (orange) coverages retrieved in the sequences by the BUSCO pipeline
996 (see Methods). The details of the genome sizes and karyotypes included are based

997 on existing literature (Uno et al. 2020; Schwartz and Maddock 1986, 2002; Hardie
998 and Hebert 2004).
999
1000
1001 **Figure 2.** Chromosomal sequence compositions of the zebra shark and selected
1002 vertebrate species. (*A*) Sequence characterization of different chromosomal
1003 segments. The orange areas show GC-content (30–70%), while the green and black
1004 lines show content of simple tandem repeats (0–25%) and interspersed repeats (0–
1005 100%), respectively, in 100-kb-long non-overlapping windows. (*B*) Two-dimensional
1006 plots of GC-content and median values of gene length, gene density, and the median
1007 of synonymous substitutions per synonymous site ($K_s$) for protein-coding genes on
1008 individual chromosomes. See Methods for statistical tests for correlation of these
1009 features with chromosome length. Computation of $K_s$ was performed for each of the
1010 four species in order by involving a pair species, namely, whale shark *Rhincodon*
1011 *typus*; small-eyed rabbitfish *Hydrolagus affinis*; helmeted guineafowl *Numida*
1012 *meleagris*; and common marmoset *Callithrix jacchus*. Zebra shark chromosomes are
1013 shown as three groups, eMAC, eMID, and eMIC (see text for details), in black,
1014 magenta, and cyan, respectively, while microchromosomes of the chicken and *C.*
1015 *milii* are colored in cyan. At the moment, *C. milii* is the only holocephalan species
1016 with a chromosome-scale genome assembly.
1017
1018 **Figure 3**. Cross-species investigation of chromosomal homology. (*A*) Dot matrices
1019 showing genome sequence similarities for selected pairs of vertebrate species with
1020 variable divergence times. Sequences of high similarity are shown with diagonal
1021 lines by the program D-GENIES (Cabanettes and Klopp 2018) with the ‘Many

repeats' mode. The numbers given to the individual panels (1) to (8), correspond to those at the nodes in the phylogenetic tree and are colored differentially to indicate comparable divergence times. Diagonal lines are colored according to the level of sequence divergence (dark green, 75-100%; light green, 50-75%; orange, 25-50%; yellow, 0-25%). (*B*) Chromosomal homology suggested by synteny conservation of one-to-one ortholog pairs. See Methods for details. The color of the ribbons connecting the synteny represents each of the three categories of the zebra shark chromosomes: eMAC (gray); eMID (magenta); and eMIC (cyan). (*C*) Chromosomal homology between the zebra shark and other vertebrates. Conserved synteny is visualized with the inter-specific correspondence of one-to-one orthologs (see Methods).

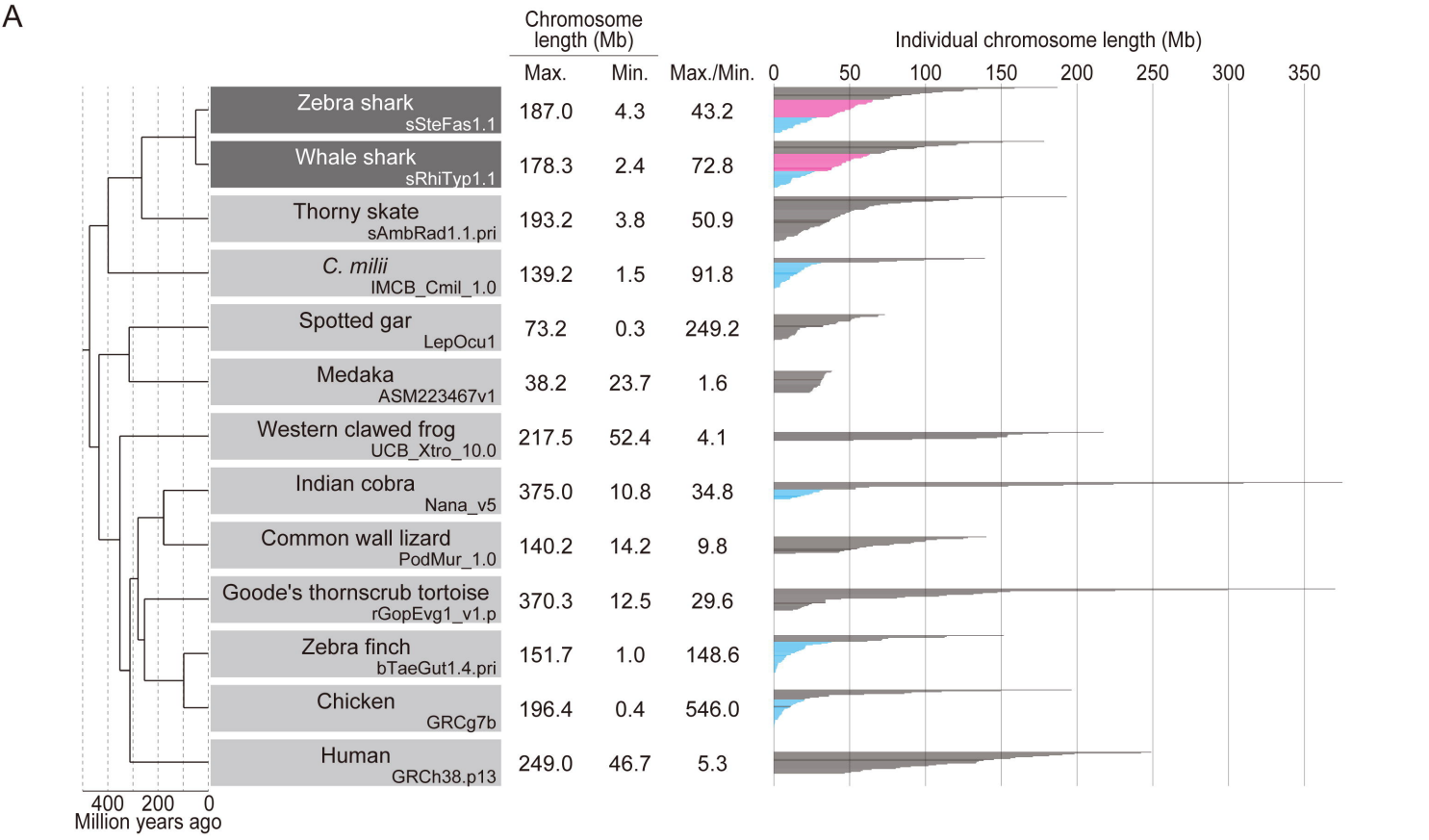
**Figure 4.** Intrachromosomal heterogeneity of sequence characteristics. (*A*) Comparison of global GC-content in 10 kb-long non-overlapping windows between the 1-Mb-long ends and the remainders of relatively large chromosomes for diverse vertebrates. (*B*) Comparison of GC-content of protein-coding regions, gene length, gene density, and synonymous substitution rate ($K_s$) among the 1 Mb-long chromosome ends, their remainders, and relatively small chromosomes (zebra shark eMIC and chicken microchromosomes (MIC)), for the zebra shark and chicken, respectively. (*C*) Two-dimensional plots of chromosome lengths and the coverage of interspersed repeats and simple tandem repeats. The coloring of the dots follows that in Figure 2B. (*D*) Differential distribution of simple tandem repeats and interspersed repeats. Proportions of the sequences identified as simple tandem repeats and interspersed repeats in 10 kb-long windows were compared between the 1 Mb-long ends of relatively large chromosomes (zebra shark eMAC and chicken

macrochromosomes (MAC)), their remainders, and relatively small chromosomes (zebra shark eMIC and chicken MIC). In B and D, significance of difference is indicated as follows; *, *p*-value < 0.05/number of tests; **, *p*-value < 0.01/number of tests; n.s., not significant. The effect sizes in statistical tests are indicated with a hyphen for no effect, one dagger symbol '†' for small effect, two dagger symbols for medium effect '††', and three dagger symbols '†††' for large effect. Numbers of the genomic regions sampled are included in parentheses. See Methods for more details about statistical tests.

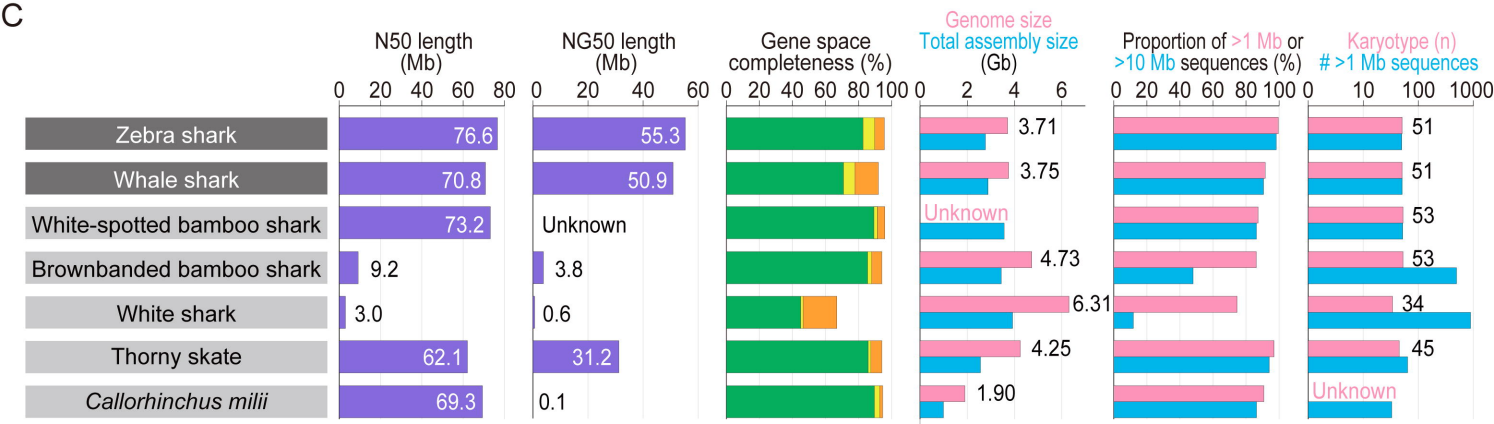
**Figure 5.** Genomic identification of the zebra shark Chromosome X. (*A*) Male–female ratio of short-read sequencing depth in the shark chromosomes. (*B*) Male-female copy-number difference of zebra shark Chromosome X. Amplification levels of the genes on the PAR and the remainders of Chromosome X (scaffold 41) was quantified using real-time PCR controlled with amplification of autosomal genomic regions, by normalizing the PCR product abundance with that for the individual F1 (sSteFas1) (see Methods). (*C*) Cross-species synteny of sex chromosome-linked genes based on 1-to-1 orthologs. (*D*) Characteristic comparisons between the different chromosome categories in the zebra shark. $K_s$ (synonymous substitution rate) and $K_a$ (nonsynonymous substitution rate) values were calculated for 1-to-1 ortholog shared with the whale shark. Only for GC-content, the PAR was shown separately, because it contained few genes. Significance of difference is indicated as follows; *, *p*-value < 0.05/number of tests; **, *p*-value < 0.01/number of tests; n.s., not significant. The effect sizes are indicated with a hyphen for no effect, one dagger symbol '†' for small effect, two dagger symbols '††' for medium effect, and three dagger symbols '†††' for large effect in statistical tests. Numbers of the genomic

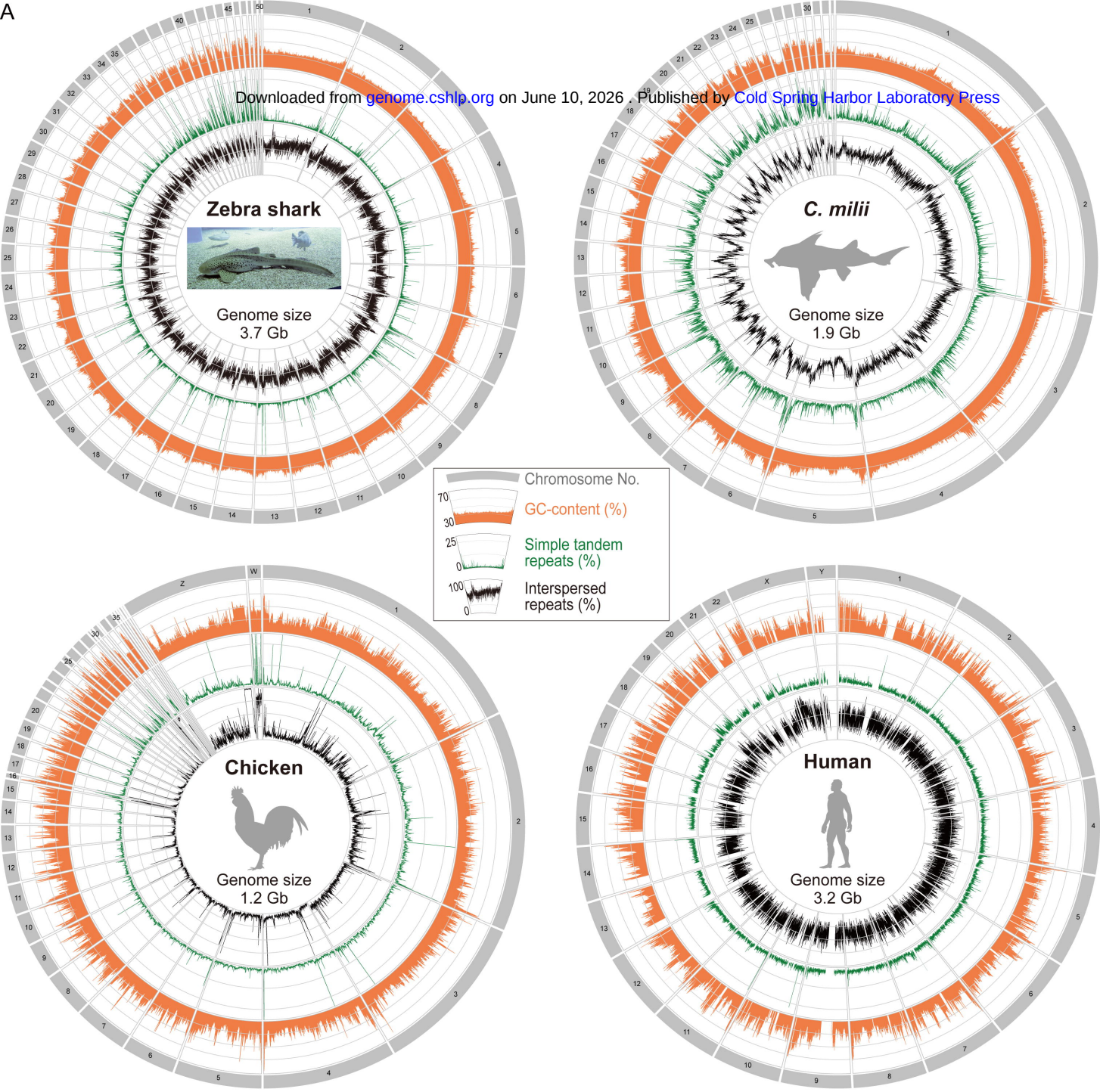
regions sampled was shown in parentheses. See Methods for more details about statistical tests. (*E*) Structural comparison between the zebra shark and whale shark X Chromosome sequences. We focused more on the zebra shark because we could not retrieve a part of the whale shark chromosome that is homologous to the putative zebra shark Chromosome X.

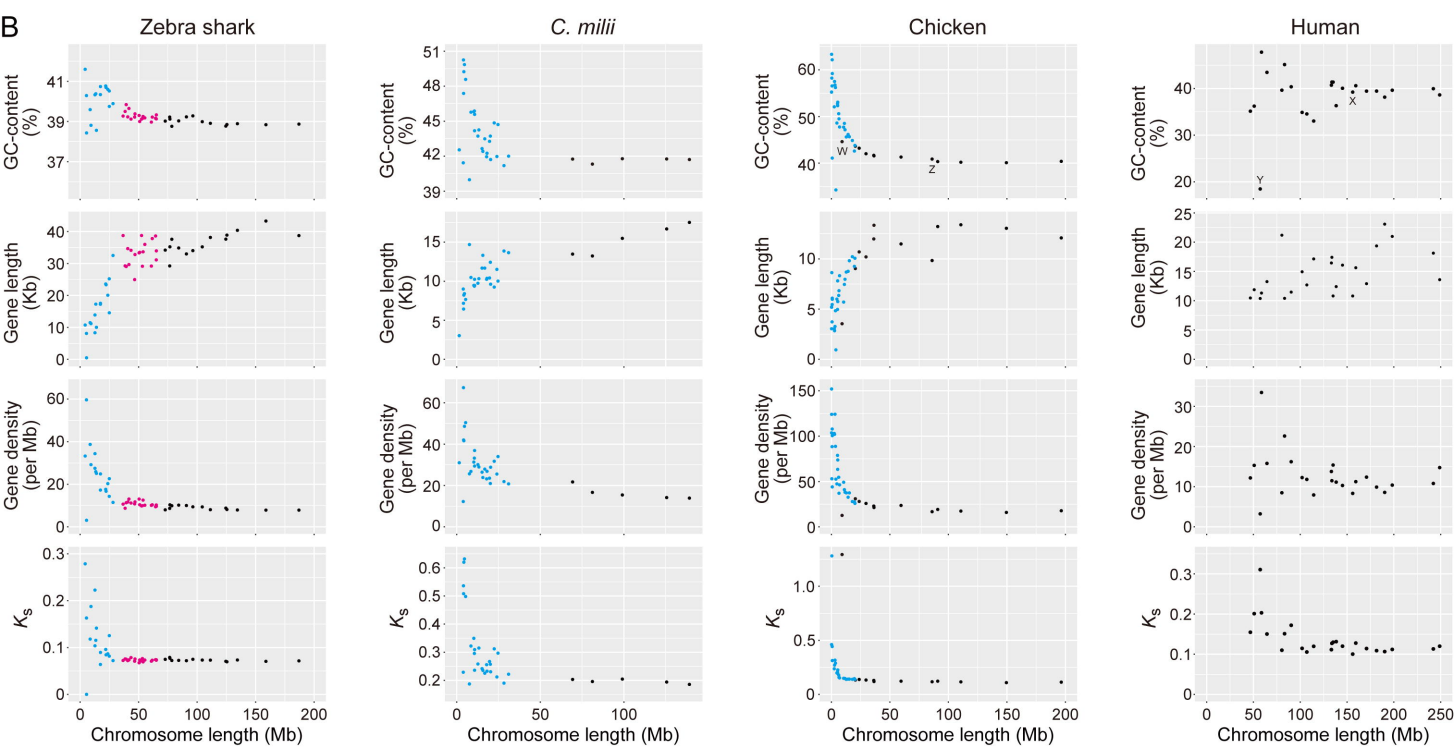
**Figure 6.** Zebra shark Hox C genes. (*A*) Genomic structure of the zebra shark Hox clusters and their neighboring regions. The exons of the Hox genes are shown in red boxes. (*B*) Molecular phylogenetic tree of Hox12 group of genes. The tree was inferred with the maximum-likelihood method as described in Methods. (*C*) Expression profiles of the zebra shark Hox genes in embryos and various tissues of a juvenile.

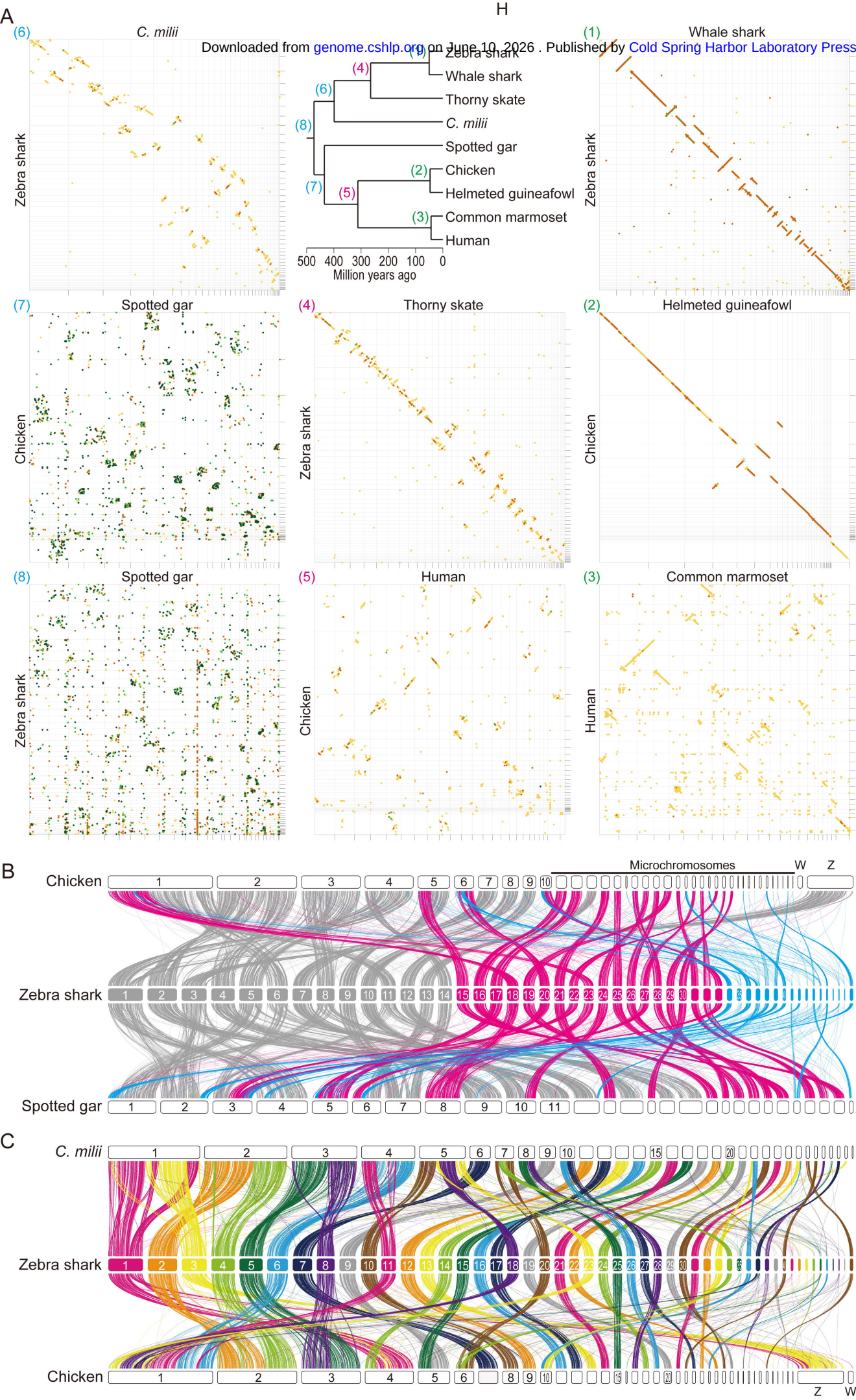

A
Chromosome length (Mb)
Max.
Min.
Max./Min.
Individual chromosome length (Mb)
0
50
100
150
200
250
300
350
Zebra shark sSteFas1.1 187.0 4.3 43.2
Whale shark sRhiTyp1.1 178.3 2.4 72.8
Thorny skate sAmbRad1.1.pri 193.2 3.8 50.9
C. milii IMCB_Cmil_1.0 139.2 1.5 91.8
Spotted gar LepOcu1 73.2 0.3 249.2
Medaka ASM223467v1 38.2 23.7 1.6
Western clawed frog UCB_Xtro_10.0 217.5 52.4 4.1
Indian cobra Nana_v5 375.0 10.8 34.8
Common wall lizard PodMur_1.0 140.2 14.2 9.8
Goode's thornscrub tortoise rGopEvg1_v1.p 370.3 12.5 29.6
Zebra finch bTaeGut1.4.pri 151.7 1.0 148.6
Chicken GRCg7b 196.4 0.4 546.0
Human GRCh38.p13 249.0 46.7 5.3
400
200
0
Million years ago

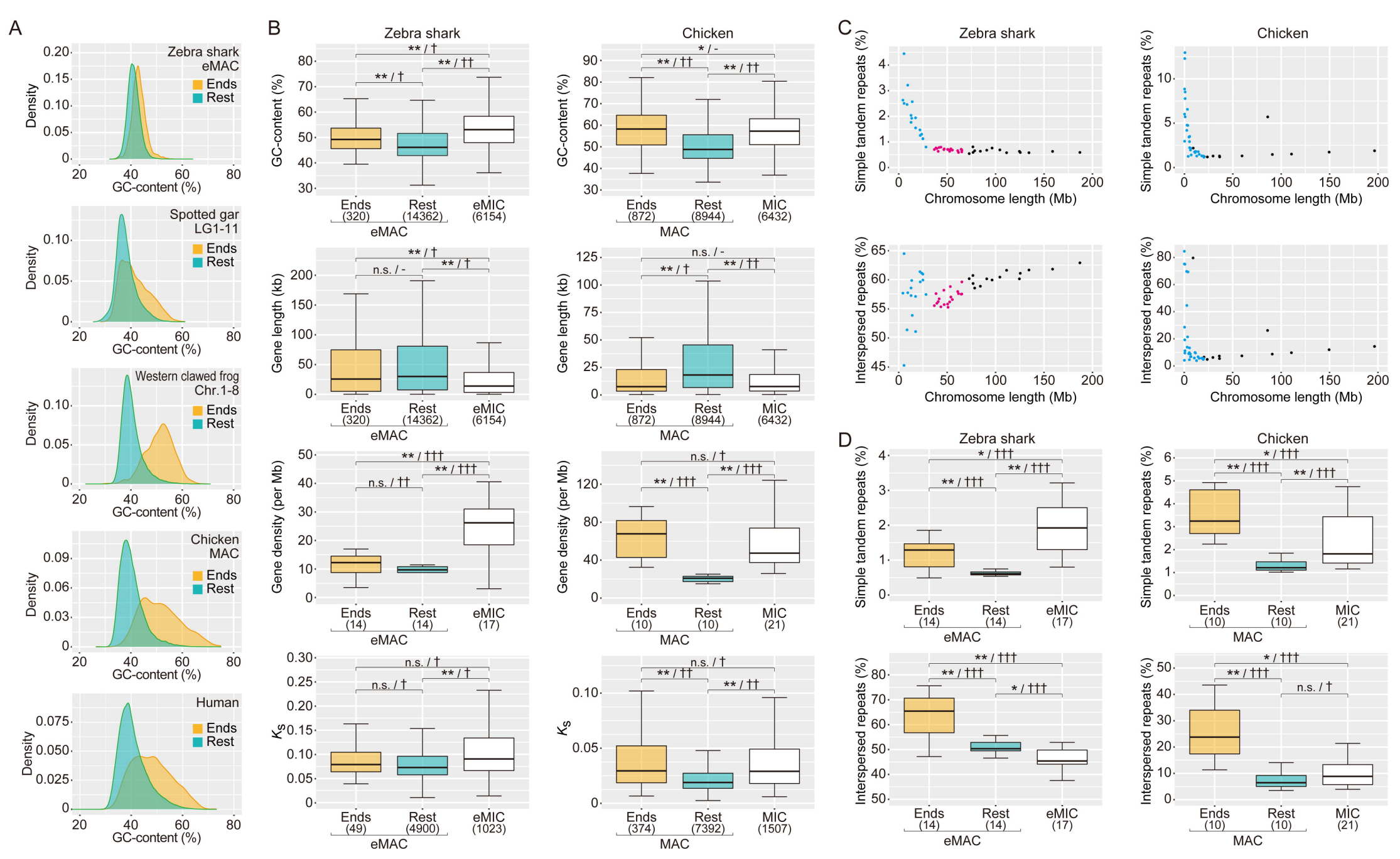

B

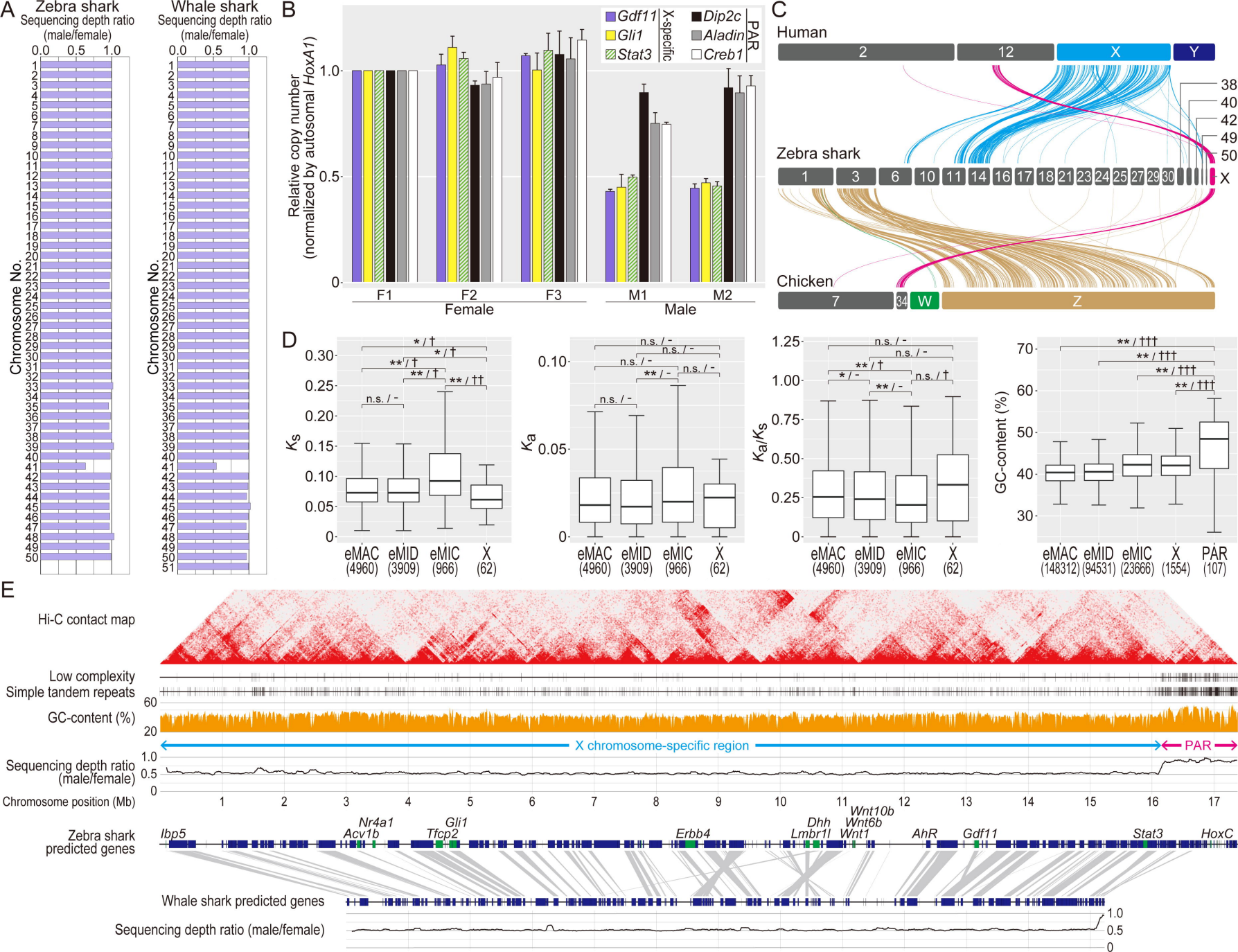

C
N50 length (Mb)
NG50 length (Mb)
Gene space completeness (%)
Genome size
Total assembly size (Gb)
Proportion of >1 Mb or >10 Mb sequences (%)
Karyotype (n)
# >1 Mb sequences
Zebra shark 76.6 55.3 3.71 51
Whale shark 70.8 50.9 3.75 51
White-spotted bamboo shark 73.2 Unknown Unknown 53
Brownbanded bamboo shark 9.2 3.8 4.73 53
White shark 3.0 0.6 6.31 34
Thorny skate 62.1 31.2 4.25 45
Callorhinchus milii 69.3 0.1 1.90 Unknown

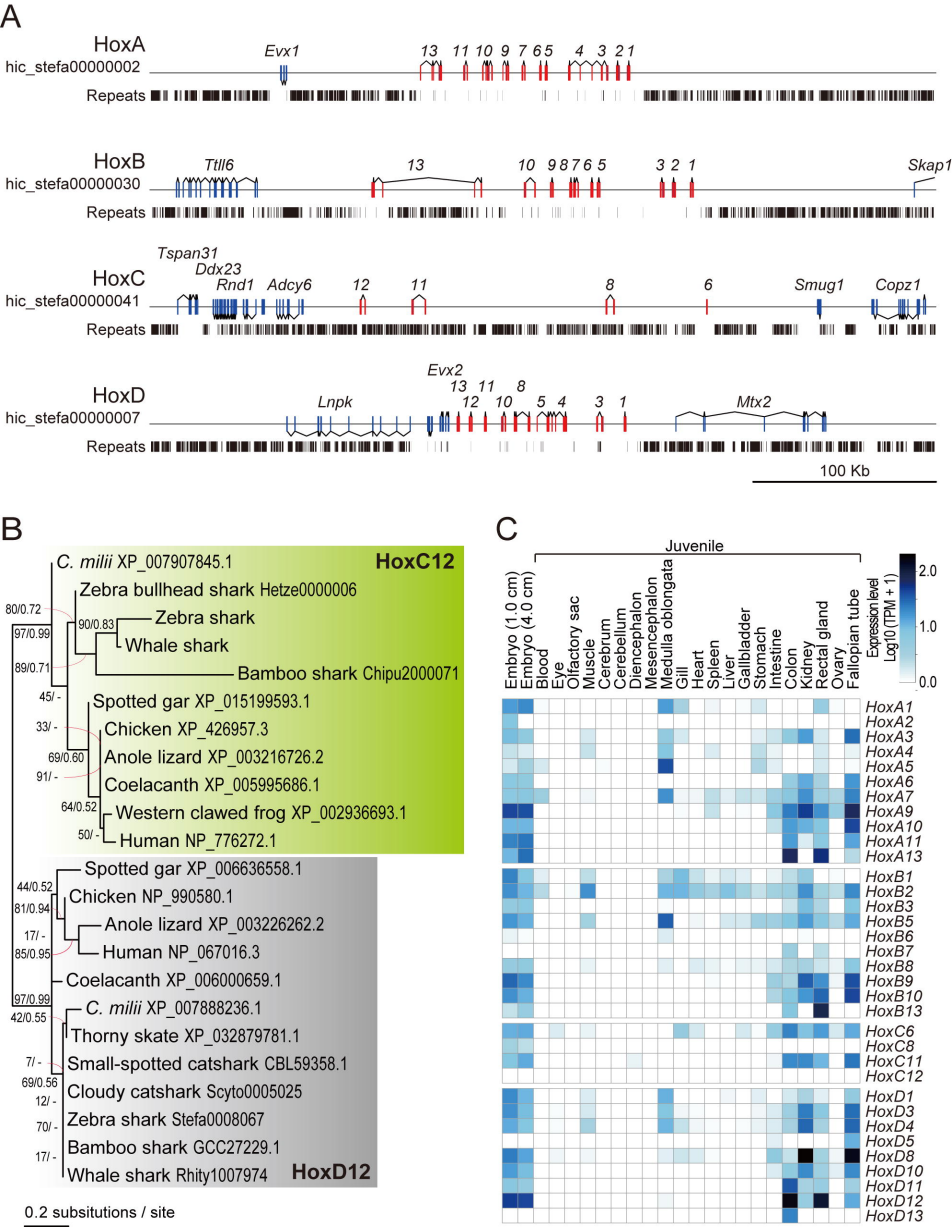





A
Zebra shark
eMAC
Ends
Rest
Spotted gar
LG1-11
Western clawed frog
Chr.1-8
Chicken
MAC
Human
Density
GC-content (%)
B
Zebra shark
Chicken
GC-content (%)
Gene length (kb)
Gene density (per Mb)
Ks
Ends
Rest
eMIC
MIC
eMAC
MAC
C
Simple tandem repeats (%)
Interspersed repeats (%)
Chromosome length (Mb)
D


A
Zebra shark
Sequencing depth ratio (male/female)
Whale shark
Sequencing depth ratio (male/female)
Chromosome No.
B
Relative copy number (normalized by autosomal HoxA1)
Gdf11
Gli1
Stat3
Dip2c
Aladin
Creb1
X-specific
PAR
F1
F2
F3
M1
M2
Female
Male
C
Human
Zebra shark
Chicken
D
Ks
Ka
Ka/Ks
GC-content (%)
eMAC (4960)
eMID (3909)
eMIC (966)
X (62)
eMAC (148312)
eMID (94531)
eMIC (23666)
X (1554)
PAR (107)
E
Hi-C contact map
Low complexity
Simple tandem repeats
GC-content (%)
X chromosome-specific region
PAR
Sequencing depth ratio (male/female)
Chromosome position (Mb)
Zebra shark predicted genes
Ibp5
Nr4a1
Acv1b
Gli1
Tfcp2
Erbb4
Dhh
Lmbr1l
Wnt10b
Wnt6b
Wnt1
AhR
Gdf11
Stat3
HoxC
Whale shark predicted genes
Sequencing depth ratio (male/female)


A
HoxA
hic_stefa00000002
Evx1
13 11 10 9 7 6 5 4 3 2 1
Repeats
HoxB
hic_stefa00000030
Ttll6
13
10 9 8 7 6 5
3 2 1
Skap1
Repeats
Tspan31
Ddx23
HoxC
Rnd1
Adcy6
12
11
8
6
Smug1
Copz1
hic_stefa00000041
Repeats
Evx2
13 11 8
HoxD
Lnpk
12 10 5 4 3 1
Mtx2
hic_stefa00000007
Repeats
100 Kb
B
C. milii XP_007907845.1
HoxC12
Zebra bullhead shark Hetze0000006
Zebra shark
Whale shark
Bamboo shark Chipu2000071
Spotted gar XP_015199593.1
Chicken XP_426957.3
Anole lizard XP_003216726.2
Coelacanth XP_005995686.1
Western clawed frog XP_002936693.1
Human NP_776272.1
Spotted gar XP_006636558.1
Chicken NP_990580.1
Anole lizard XP_003226262.2
Human NP_067016.3
Coelacanth XP_006000659.1
C. milii XP_007888236.1
Thorny skate XP_032879781.1
Small-spotted catshark CBL59358.1
Cloudy catshark Scyto0005025
Zebra shark Stefa0008067
Bamboo shark GCC27229.1
Whale shark Rhity1007974
HoxD12
80/0.72
97/0.99
90/0.83
89/0.71
45/ -
33/ -
69/0.60
91/ -
64/0.52
50/ -
44/0.52
81/0.94
17/ -
85/0.95
97/0.99
42/0.55
7/ -
69/0.56
12/ -
70/ -
17/ -
0.2 subsitutions / site
C
Juvenile
Embryo (1.0 cm)
Embryo (4.0 cm)
Blood
Eye
Olfactory sac
Muscle
Cerebrum
Cerebellum
Diencephalon
Mesencephalon
Medulla oblongata
Gill
Heart
Spleen
Liver
Gallbladder
Stomach
Intestine
Colon
Kidney
Rectal gland
Ovary
Fallopian tube
Expression level
Log10 (TPM + 1)
2.0
1.0
0.0
HoxA1
HoxA2
HoxA3
HoxA4
HoxA5
HoxA6
HoxA7
HoxA9
HoxA10
HoxA11
HoxA13
HoxB1
HoxB2
HoxB3
HoxB5
HoxB6
HoxB7
HoxB8
HoxB9
HoxB10
HoxB13
HoxC6
HoxC8
HoxC11
HoxC12
HoxD1
HoxD3
HoxD4
HoxD5
HoxD8
HoxD10
HoxD11
HoxD12
HoxD13