

1

2 *Research article*

3

4 **Elasmobranch genome sequencing reveals evolutionary trends of**  
5 **vertebrate karyotype organization**

6

7 Kazuaki Yamaguchi<sup>1</sup>, Yoshinobu Uno<sup>1\*</sup>, Mitsutaka Kadota<sup>1</sup>, Osamu Nishimura<sup>1</sup>, Ryo  
8 Nozu<sup>2\*\*</sup>, Kiyomi Murakumo<sup>3</sup>, Rui Matsumoto<sup>3</sup>, Keiichi Sato<sup>2,3</sup>, Shigehiro Kuraku<sup>1,4,5#</sup>

9

10 <sup>1</sup>Laboratory for Phyloinformatics, RIKEN Center for Biosystems Dynamics Research  
11 (BDR), Kobe, Japan

12 <sup>2</sup>Okinawa Churashima Research Center, Okinawa Churashima Foundation,  
13 Okinawa, Japan.

14 <sup>3</sup>Okinawa Churaumi Aquarium, Okinawa, Japan.

15 <sup>4</sup>Molecular Life History Laboratory, Department of Genomics and Evolutionary  
16 Biology, National Institute of Genetics, Mishima, Japan.

17 <sup>5</sup>Department of Genetics, Sokendai (Graduate University for Advanced Studies),  
18 Mishima, Japan.

19

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

23 \*\*Present address: Graduate School of Integrated Sciences for Life, Hiroshima  
24 University, Hiroshima 739-0046, Japan

25 #To whom correspondence may be addressed. Shigehiro Kuraku

26

27 Running title: Evolution of shark chromosomes

28

29   **Abstract**

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

51

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

54 **Introduction**

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

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

79 (blood or embryos) (e.g., at aquariums) enabled karyotyping using culture cells for  
80 four orectolobiform shark species in Elasmobranchii (Uno et al. 2020). This has  
81 paved the way for rigid evaluation of whole genome sequences. Biological studies on  
82 cartilaginous fishes have been hindered by low accessibility to fresh material.  
83 Moreover, especially in studying elasmobranchs, genome analysis can encounter  
84 inherent difficulty incurred by their large genome sizes (reviewed in Kuraku 2021).  
85 These factors have prevented previous efforts on cartilaginous fishes from obtaining  
86 a suite of genome sequences supported by karyotype and genome size estimate as  
87 well as transcriptome sequencing (Rhie et al. 2021; Read et al. 2017; Tan et al.  
88 2021; Weber et al. 2020; Marra et al. 2019; Zhang et al. 2020).

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

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

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

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

127 In this study, we focused on the zebra shark *Stegostoma tigrinum* (or leopard  
128 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).

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

161 While we recognize a significant gap between the estimated and retrieved  
162 total sequence lengths, genome assemblies for the zebra shark and whale shark  
163 exhibited high completeness of protein-coding gene space of more than 90% (Figure  
164 1C). Prediction of protein-coding genes on the zebra shark and whale shark  
165 genomes, performed by incorporating homolog sequences and transcriptome data,  
166 resulted in 33,222 and 35,334 genes, respectively (Supplementary Tables 1 and 2),  
167 which allowed downstream molecular biological analysis.

168

### 169 **Karyotypic trends in sharks**

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

178 spotted bamboo shark (Zhang et al. 2020), and thorny skate (Rhie et al. 2021), and  
179 is assumed to be typical karyotypic organization of elasmobranchs (Supplementary  
180 Figure 2). For simplicity, we tentatively designated (1) zebra shark Chromosome 1 to  
181 14, longer than 70 Mb; (2) Chromosome 15 to 33, between 30 and 70 Mb; and (3)  
182 Chromosome 34 to 50, shorter than 30 Mb as elasmobranch macro-, middle-sized-,  
183 and micro-chromosomes (abbreviated into eMAC, eMID, and eMIC, respectively)  
184 that are differentially colored in Figures 1A and 2B. This categorization was also  
185 applied to the whale shark chromosomes (Figure 1A; Supplementary Figure 2).

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

195

196

### 197 **Vertebrate-wide comparisons including elasmobranchs**

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

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

212 Previous studies sought to reconstruct the process of karyotypic evolution of  
213 vertebrates but often lacked elasmobranchs in the dataset (Nakatani et al. 2021;  
214 Sacerdot et al. 2018). In the present study, we performed a gnathostome-wide  
215 comparison of the syntenic location of one-to-one orthologs, including the zebra  
216 shark (Figure 3B; Supplementary Figure 3). A considerable proportion of the one-to-  
217 one orthologs are shared between eMAC and large chromosomes of chicken and  
218 spotted gar (Figure 3B). The majority of the genomic regions in the zebra shark  
219 eMIC were shown not to share one-to-one orthologs with the so-called  
220 microchromosomes of the chicken or spotted gar (Figure 3B). Moreover, the smallest  
221 spotted gar chromosomes were frequently shown to be homologous to zebra shark  
222 eMID (chromosomes 15–33), and not to its eMIC (Chromosomes 34–50) (Figure 3B).  
223 It was also shown that zebra shark Chromosomes 10, 25, 26, 27, and 28 have  
224 chicken homologs of similar size but their *C. milii* homologs have been fused into  
225 larger chromosomes, while zebra shark Chromosomes 2, 3, 4, 5, 6, 8, 11, 18, 20, 23,  
226 33, 34, and 36 are likely to be products of fissions that occurred in the elasmobranch  
227 lineage (Figure 3C; Supplementary Figure 3). These results cast doubt on the

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

232

233

234 **Length-dependent properties of chromosomes**

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

253 exceptionally low GC-content, such as Chromosome 26 of the *C. milii*, Chromosome  
254 29 and 33 of the chicken, and Linkage Group 29 of the spotted gar (Figure 2;  
255 Supplementary Figures 1 and 2). These exceptions evoke a caution for the  
256 generalization of common characteristics of short chromosomes. It needs to be  
257 carefully examined whether the relatively short sequences with exceptionally low  
258 GC-content are fragments of large chromosomes that failed to be assembled to a  
259 chromosome scale. Such chromosomal scaffold sequences, with small length and  
260 relatively low GC-content, may be the cause of the insufficient support for the  
261 chromosome length-dependent pattern in some species (Supplementary Table 3).

262

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

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

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

283

284

285

286

### 287 **Intra-genomic repetitive element distribution**

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

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

312

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

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

328 of the putative X Chromosomes of the zebra shark to a part of human Chromosome  
329 12 and chicken Chromosome 34 (Figure 5C, magenta). Neither the human X, human  
330 Y, chicken Z, nor chicken W chromosomes exhibited pronounced homology to the  
331 putative zebra shark Chromosome X. The putative shark X Chromosome identified in  
332 this study do harbor orthologs of a number of well-studied regulatory factors but do  
333 not harbor the orthologs of the master sex determination genes identified in other  
334 vertebrates including teleost fishes, that is, *Dmrt1*- or *Sox3*-related transcription  
335 factors as well as components of TGF- $\beta$  signaling pathway, such as *Amh*, *Amhr2*,  
336 *Bmpr1b*, *Gsdf*, and *Gdf6* (Bertho et al. 2021).

337 In the approximately 17 Mb-long sequence of the putative zebra shark  
338 Chromosome X, one 1.5 Mb-long end exhibited a male–female sequencing depth  
339 ratio of nearly 1.0 (Figure 5D). This region, with a sequence depth comparable to  
340 that of the autosomes, is deduced to be a pseudoautosomal region (PAR) that is  
341 likely shared between the heterogametic sex chromosomes X and Y (Smeds et al.  
342 2014; Palmer et al. 2019). The identification of PAR was supported by a comparable  
343 level of genomic qPCR amplification for multiple genes in this region to that for  
344 autosomal regions (Figure 5B). We characterized possible unique patterns of  
345 molecular evolution typical of sex chromosomes (Figure 5D; Supplementary Table 5).  
346 In the X Chromosome, protein-coding genes exhibited a significant decrease of  
347 synonymous substitution rate ( $K_s$ ) supported by small effect and an increased  
348 median of non-synonymous substitution rate ( $K_a$ ), resulting in an increased  $K_a/K_s$   
349 ratio. Our comparison also revealed higher frequency of low-complexity repeats as  
350 well as higher GC-content in the PAR (Figure 5D and E), which is a hallmark of PAR  
351 observed in other species resulted from accelerated recombination (Galtier et al.  
352 2001; Galtier 2004; Smeds et al. 2014). The Hi-C contact map, derived from zebra

353 shark blood, shows subproximal contacts suggesting intermittent chromatin  
354 compartmentalization within the PAR and the rest of Chromosome X (Figure 5E).  
355 Our comparison of the ortholog location between the putative X  
356 Chromosomes of the two species supported a high cross-species conservation of  
357 the chromosome structure (Figure 5E). In the current whale shark assembly,  
358 Chromosome 41, which corresponds to the putative chromosome X, may not cover  
359 the whole chromosome, possibly excluding one end of the PAR (Figure 5E).

360

361

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

363 In the newly obtained sequence of the putative zebra shark X Chromosome, we  
364 identified an array of orthologs of the non-shark genes encoding homeobox proteins  
365 Hox C (Figure 5E). In elasmobranchs, Hox C genes were long thought missing from  
366 the genome (King et al. 2011) but later identified in several shark species, as rogue  
367 open reading frames (ORFs) flanked by massively repetitive sequences (Hara et al.  
368 2018; reviewed in Kuraku 2021). Our genome-wide gene prediction for the zebra  
369 shark detected the ORF of *Hoxc8*, -c11, and -c12, while the partial ORF of the  
370 putative *Hoxc6* ortholog was also identified by a manual search of the raw genomic  
371 sequence (Figure 6A). These Hox C genes were located in a 180 kb-long genomic  
372 segment in the PAR of the putative Chromosome X (Figure 6A), identified as a single  
373 cluster in an elasmobranch fish for the first time. Their orthologies were confirmed  
374 with molecular phylogenetic trees, which also indicated elevated molecular  
375 evolutionary rates with long branches for elasmobranch Hox C genes (Figure 6B).  
376 Our RNA-seq data showed the transcription of these Hox C genes (except for  
377 *Hoxc12*) in embryos and juvenile tissues (Figure 6C). The identified zebra shark Hox

378 C cluster is massively invaded by repetitive elements unlike the other Hox gene  
379 clusters (A, B, and D clusters) of this and many other vertebrate species (Figure 6A),  
380 as in the Hox C-containing genomic segments of other shark species (Hara et al.  
381 2018). Our search for zebra shark orthologs of the protein-coding genes located near  
382 the human Hox C cluster (e.g., *ATF7*, *CBX5*) revealed poor conservation of the gene  
383 compositions. Some of the zebra shark orthologs were not identified in its entire  
384 genome sequence, suggesting a divergent nature of the genomic regions flanking  
385 the Hox C cluster.

386

387

### 388 **Discussion**

389 In this study, we chose two orectolobiform shark species (zebra shark and whale  
390 shark) in Elasmobranchii and characterized their genomic organization with  
391 chromosome-scale DNA sequences. This study was achieved in support of  
392 epigenome and transcriptome data prepared using fresh tissue samples and  
393 previously obtained karyotype information. Our results suggest that their karyotypes  
394 are organized by chromosomes of gradual sizes marked with size-dependent  
395 sequence properties (Figure 2B). The length gradualism is a remarkable feature of  
396 elasmobranch karyotypes, although we tentatively grouped those chromosomes into  
397 three length-dependent categories as proposed recently for other elasmobranch  
398 species (Marlétaz et al. 2023; Stanhope et al. 2023). The pattern in the shark  
399 chromosomal organization is unique in its diversity among vertebrates (Uno et al.  
400 2020), which is characterized by abundant chromosomes (up to 106 for diploids) and  
401 variable chromosome sizes. The abundance and highly variable sizes of  
402 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 mili*,  
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

428 uniform definition (see Introduction). Our investigation, focusing on various aspects  
429 of chromosomal DNA sequences, provided a novel view of vertebrate karyotypes  
430 that cannot be understood with a simple binary classification, namely with or without  
431 microchromosomes, or between macro- and microchromosomes. This view is  
432 supported by the common pattern of high intrachromosomal heterogeneity within  
433 individual macrochromosomes (Figure 4; Supplementary Figure 4) as well as  
434 interchromosomal heterogeneity among different microchromosomes (Figure 2;  
435 Supplementary Figures 1 and 2). In particular, the heterogeneity within  
436 macrochromosomes, marked with high GC-content, high gene density, and small  
437 gene length of their ends, may be shared widely among diverse vertebrates (Figure  
438 4A, B). Intragenomic heterogeneity of GC-content was previously suggested to be  
439 caused by GC-biased gene conversion (Mugal et al. 2015). In addition, the uniform  
440 numbers of recombination per chromosome (Dumont and Payseur 2008) have been  
441 thought to explain the chromosome length-dependent GC-content variation between  
442 different chromosomes. Our observations did not show pronounced length-  
443 dependent variation of GC-content for chromosomes that were longer than 100 Mb  
444 (Figure 2B; Supplementary Figure 2). Taken together, we speculate that the peculiar  
445 nature of chromosome ends mainly account for the variation of GC-content among  
446 different chromosomes (Figure 4B). Importantly, ‘chromosome ends’ in this context  
447 not only harbor telomeric or other simple repeats but also hold complex sequences  
448 including protein-coding genes in sequence stretches that are longer than 1 Mb  
449 (Figure 4B; Supplementary Figure 4). In the wester clawed frog genome, such  
450 sequence stretches marked with elevated GC-content span much longer ranges than  
451 in other genomes (Supplementary Figure 1A and 2). The observed features of  
452 smaller chromosomes with larger proportions of such ‘ends’ in length are more

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

463 Our genome-wide sequencing depth investigation covering both sexes  
464 revealed the XY system for the two studied shark species and enabled the first  
465 sequence-based identification of shark sex chromosomes (namely, Chromosome X;  
466 Figure 5A). The genes on the shark Chromosome X tend to exhibit larger non-  
467 synonymous substitution rates ( $K_a$ ) and smaller synonymous substitution rates ( $K_s$ )  
468 than those on autosomes, resulting in a higher  $K_a/K_s$  ratio (Figure 5D). This  
469 resembles the pattern known in other species including birds and insects, which is  
470 known as the ‘Faster X (or Faster Z) hypothesis’ (Meisel and Connallon 2013;  
471 Charlesworth et al. 2018; Xu et al. 2019; Mank et al. 2007). The smaller  $K_s$  value is  
472 also indicative of the support for the male-driven evolution hypothesis (Li 2002;  
473 Miyata et al. 1987), which is to be examined by higher  $K_s$  values for genes on the  
474 presumptive Y chromosome that remains unidentified. Our comparison of protein-  
475 coding gene compositions showed that the zebra shark and the whale shark largely  
476 share Chromosome X that is homologous to each other (Figure 5E). The  
477 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

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

511

## 512 **Methods**

### 513 **Animals**

514 Fresh blood from a female adult zebra shark (total length, 2.2 m); Individual ID,  
515 sSteFas1 (also called F1) and a male adult whale shark (total length, 8.8 m;  
516 Individual ID, sRhiTyp1) were sampled at Okinawa Churaumi Aquarium and used for  
517 the preparation of whole-genome shotgun DNA libraries, Hi-C libraries, and RNA-seq  
518 libraries as well as for measuring nuclear DNA content by flow cytometry. Likewise,  
519 fresh blood of a male zebra shark (total length, 2.1 m; Individual ID, sSteFas2 [also  
520 called M1]) and a female whale shark (total length, 8.0 m; Individual ID, sRhiTyp2)  
521 were sampled and used for quantifying the male/female ratios of individual  
522 chromosomal regions. Extraction of ultra-high molecular weight DNA was performed  
523 by collecting blood cells by centrifugation, and the collected cells were embedded in  
524 agarose plugs ( $4.0 \times 10^5$  cells/plug). The agarose gel plugs were prepared and  
525 processed with the CHEF Mammalian Genomic DNA Plug Kit (BioRad, Cat. No.  
526 #1703591). Total RNAs used to construct RNA-seq libraries were extracted from  
527 various tissues of a female juvenile zebra shark (total length, 30 cm) born at

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

539

#### 540 **Genome sequencing and scaffolding**

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

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

566

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

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

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

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

590

### 591 **Repeat identification**

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

599

### 600 **Gene model construction**

601 The program Braker v2.1.6 was employed for gene prediction by inputting the results  
602 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 (Brúna 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 trimAIL 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**

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

640

#### 641 **Conserved synteny detection**

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

648

#### 649 **Statistical analysis**

650 Relationships of chromosome lengths with chromosome sequence features were  
651 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

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

680

681 **Identification of shark X Chromosome**

682 To identify a chromosome-scale scaffold with a distinct male–female sequencing  
683 depth ratio in the zebra shark, the same number of trimmed genomic shotgun reads  
684 (293,686,584) was prepared for both sexes. The reads were mapped with BWA-  
685 MEM2 (v2.2.1) onto a genome assembly (Li and Durbin 2009). Mapped reads were  
686 counted for male and female using the bamtobed program in the package BEDTools  
687 v2.29.2, for each scaffold in 10 kb non-overlapping windows, and male–female ratios  
688 were calculated (Quinlan and Hall 2010). Sequencing depth of male and female  
689 reads on the identified Chromosome X was also calculated for 10 kb non-  
690 overlapping windows. Windows with the proportion of ambiguous bases of more than  
691 50% were excluded from computation. This procedure was also applied to the whale  
692 shark, using 220,785,436 trimmed reads.

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

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

705

706 **Data access**

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

711

712 **Competing interests**

713 The authors declare no conflict of interests.

714

715 **Acknowledgments**

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

725

726 **References**

- 727 Bertho S, Herpin A, Schartl M, Guiguen Y. 2021. Lessons from an unusual  
728 vertebrate sex-determining gene. *Philos Trans R Soc B: Biol Sci* **376**:  
729 20200092.
- 730 Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, Amores  
731 A, Desvignes T, Batzel P, Catchen J, et al. 2016. The spotted gar genome  
732 illuminates vertebrate evolution and facilitates human-teleost comparisons.  
733 *Nat Genet* **48**: 427–437.
- 734 Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2:  
735 automatic eukaryotic genome annotation with GeneMark-EP+ and  
736 AUGUSTUS supported by a protein database. *NAR Genom Bioinform* **3**: 1–  
737 11.
- 738 Burt DW. 2002. Origin and evolution of avian microchromosomes. *Cytogenet  
739 Genome Res* **96**: 97–112.
- 740 Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an  
741 interactive, efficient and simple way. *PeerJ* **6**: e4958.
- 742 Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAI: a tool for  
743 automated alignment trimming in large-scale phylogenetic analyses.  
744 *Bioinformatics* **25**: 1972–1973.
- 745 Charlesworth B, Campos JL, Jackson BC. 2018. Faster-X evolution: theory and  
746 evidence from *Drosophila*. *Mol Ecol* **27**: 3753–3771.
- 747 Cosentino S, Iwasaki W. 2019. SonicParanoid: fast, accurate and easy  
748 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. *R/ideogram*: 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.

- 811 Koochekian N, Ascanio A, Farleigh K, Card DC, Schield DR, Castoe TA,
- 812 Jezkova T. 2022. A chromosome-level genome assembly and annotation of
- 813 the desert horned lizard, *Phrynosoma platyrhinos*, provides insight into
- 814 chromosomal rearrangements among reptiles. *Gigascience* **11**: giab098.
- 815 Kuraku S. 2021. Shark and ray genomics for disentangling their morphological
- 816 diversity and vertebrate evolution. *Dev Biol* **477**: 262–272.
- 817 Kuraku S, Ishiijima J, Nishida-Umehara C, Agata K, Kuratani S, Matsuda Y.
- 818 2006. cDNA-based gene mapping and GC3 profiling in the soft-shelled
- 819 turtle suggest a chromosomal size-dependent GC bias shared by
- 820 sauropsids. *Chromosome Res* **14**: 187–202.
- 821 Kuraku S, Zmasek CM, Nishimura O, Katoh K. 2013. aLeaves facilitates on-
- 822 demand exploration of metazoan gene family trees on MAFFT sequence
- 823 alignment server with enhanced interactivity. *Nucleic Acids Res* **41**: 22–28.
- 824 Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a bayesian software
- 825 package for phylogenetic reconstruction and molecular dating.
- 826 *Bioinformatics* **25**: 2286–2288.
- 827 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-
- 828 Wheeler transform. *Bioinformatics* **25**: 1754–1760.
- 829 Li W-H. 2002. Male-driven evolution. *Curr Opin Genet Dev* **12**: 650–656.

- 830 Long DJ. 1992. Sharks from the La Meseta Formation (Eocene), Seymour  
831 Island, Antarctic Peninsula. *J Vertebr Paleontol* **12**: 11-32.
- 832 Ma W, Xu L, Hua H, Chen M, Guo M, He K, Zhao J, Li F. 2021. Chromosomal-  
833 level genomes of three rice planthoppers provide new insights into sex  
834 chromosome evolution. *Mol Ecol Resour* **21**: 226–237.
- 835 Mank JE, Axelsson E, Ellegren H. 2007. Fast-X on the Z: rapid evolution of sex-  
836 linked genes in birds. *Genome Res* **17**: 618–624.
- 837 Marlétaz F, de la Calle-Mustienes E, Acemel RD, Paliou C, Naranjo S, Martínez-  
838 García PM, Cases I, Sleight VA, Hirschberger C, Marcet-Houben M, et al.  
839 2023. The little skate genome and the evolutionary emergence of wing-like  
840 fins. *Nature* **616**: 495–503.
- 841 Marra NJ, Stanhope MJ, Jue NK, Wang M, Sun Q, Bitar PP, Richards VP,  
842 Komissarov A, Rayko M, Kliver S, et al. 2019. White shark genome reveals  
843 ancient elasmobranch adaptations associated with wound healing and the  
844 maintenance of genome stability. *Proc Natl Acad Sci USA* **116**: 4446–4455.
- 845 Matsubara K, Kuraku S, Tarui H, Nishimura O, Nishida C, Agata K, Kumazawa  
846 Y, Matsuda Y. 2012. Intra-genomic GC heterogeneity in sauropsids:  
847 evolutionary insights from cDNA mapping and GC3 profiling in snake. *BMC  
848 Genomics* **13**: 1-14.
- 849 Meisel RP, Connallon T. 2013. The faster-X effect: integrating theory and data.  
850 *Trends Genet* **29**: 537–544.

- 851 Meyer A, Schloissnig S, Franchini P, Du K, Woltering JM, Irisarri I, Wong WY,
- 852 Nowoshilow S, Kneitz S, Kawaguchi A, Fabrizius A, Xiong P, Dechaud C,
- 853 Spaink HP, Wolff JN, Simakov O, Burmester T, Tanaka EM, Schartl M. 2021.
- 854 Giant lungfish genome elucidates the conquest of land by vertebrates.
- 855 *Nature* **590**: 284–289.
- 856 Miyata T, Hayashida H, Kuma K, Mitsuyasu K, Yasunaga T. 1987. Male-driven
- 857 molecular evolution: a model and nucleotide sequence analysis. *Cold*
- 858 *Spring Harb Symp Quant Biol* **52**: 863–867.
- 859 Mugal CF, Weber CC, Ellegren H. 2015. GC-biased gene conversion links the
- 860 recombination landscape and demography to genomic base composition.
- 861 *BioEssays* **37**: 1317–1326.
- 862 Mulley JF, Zhong Y-F, Holland PW. 2009. Comparative genomics of
- 863 chondrichthyan Hoxa clusters. *BMC Evol Biol* **9**: 218.
- 864 Nakatani Y, Shingate P, Ravi V, Pillai NE, Prasad A, McLysaght A, Venkatesh B.
- 865 2021. Reconstruction of proto-vertebrate, protocyclostome and proto-
- 866 gnathostome genomes provides new insights into early vertebrate evolution.
- 867 *Nat Commun* **12**: 4489.
- 868 Nakatani Y, Takeda H, Kohara Y, Morishita S. 2007. Reconstruction of the
- 869 vertebrate ancestral genome reveals dynamic genome reorganization in
- 870 early vertebrates. *Genome Res* **17**: 1254–1265.

- 871 Naylor G, Caira K, Jensen K, White W, Last P. 2012. A DNA sequence-based  
872 approach to the identification of shark and ray species and its implications  
873 for global elasmobranch diversity and parasitology *Bull Am Mus Nat Hist*  
874 **367:** 1-263.
- 875 Ohno S. 1970. Evolution by gene duplication. Springer Science & Business  
876 Media, Berlin.
- 877 Ohno S, Muramoto J, Stenius C, Christian L, Kittrell WA, Atkin NB. 1969.  
878 Microchromosomes in holocephalian, chondrostean and holostean fishes.  
879 *Chromosoma* **26:** 35–40.
- 880 Palmer DH, Rogers TF, Dean R, Wright AE. 2019. How to identify sex  
881 chromosomes and their turnover. *Mol Ecol* **28:** 4709–4724.
- 882 Pennell MW, Mank JE, Peichel CL. 2018. Transitions in sex determination and  
883 sex chromosomes across vertebrate species. *Mol Ecol* **27:** 3950–3963.
- 884 Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level  
885 expression analysis of RNA-seq experiments with HISAT, StringTie and  
886 Ballgown. *Nat Protoc* **11:** 1650–1667.
- 887 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing  
888 genomic features. *Bioinformatics* **26:** 841–842.
- 889 Read TD, Petit RA, Joseph SJ, Alam T, Weil MR, Ahmad M, Bhimani R, Vuong  
890 JS, Haase CP, Webb DH, et al. 2017. Draft sequencing and assembly of

- 891 the genome of the world's largest fish, the whale shark: *Rhincodon typus*
- 892 smith 1828. *BMC Genomics* **18**: 755.
- 893 Rhei A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva
- 894 M, Chow W, Fungtammasan A, Kim J. 2021. Towards complete and error-
- 895 free genome assemblies of all vertebrate species. *Nature* **592**: 737–746.
- 896 Sacerdot C, Louis A, Bon C, Berthelot C, Roest Crolli H. 2018. Chromosome
- 897 evolution at the origin of the ancestral vertebrate genome. *Genome Biol* **19**:
- 898 166.
- 899 Schwartz FJ, Maddock MB. 1986. Comparisons of karyotypes and cellular DNA
- 900 contents within and between major lines of elasmobranchs. In *Indo-Pacific*
- 901 *fish biology*, pp. 148–157. Ichthyological Soc. of Japan.
- 902 Schwartz FJ, Maddock MB. 2002. Cytogenetics of the elasmobranchs: genome
- 903 evolution and phylogenetic implications. *Mar Freshw Res* **53**: 491–502.
- 904 Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly
- 905 and annotation completeness. In *Gene Prediction Methods in Molecular*
- 906 *Biology* (ed. Kollmar M), pp. 227–245. Humana Press, New York.
- 907 Sheffer MM, Cordellier M, Forman M, Grewoldt M, Hoffmann K, Jensen C, Kotz
- 908 M, Král J, Kuss AW, Líznarová E, et al. 2022. Identification of sex
- 909 chromosomes using genomic and cytogenetic methods in a range-
- 910 expanding spider, *Argiope bruennichi* (Araneae: Araneidae). *Biological*
- 911 *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.org>.
- 921        Smit AFA, Hubley R, Green P. 2013. RepeatMasker Open-4.0.  
922                  <http://www.repeatmasker.org>.
- 923        Soderlund C, Bomhoff M, Nelson WM. 2011. SyMAP v3.4: A turnkey synteny  
924                  system with application to plant genomes. *Nucleic Acids Res* **39**: e68–e68.
- 925        Srikulnath K, Ahmad SF, Singchat W, Panthum T. 2021. Why do some  
926                  vertebrates have microchromosomes? *Cells* **10**: 1–33.
- 927        Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and  
928                  post-analysis of large phylogenies. *Bioinformatics* **30**: 1312-1313.
- 929        Stanhope MJ, Ceres KM, Sun Q, Wang M, Zehr JD, Marra NJ, Wilder AP, Zou  
930                  C, Bernard AM, Pavinski-Bitar P, et al. 2023. Genomes of endangered  
931                  great hammerhead and shortfin mako sharks reveal historic population  
932                  declines and high levels of inbreeding in great hammerhead. *iScience* **26**:  
933                  105815.

- 935 Suryamohan K, Krishnankutty SP, Guillory J, Jevit M, Schröder MS, Wu M,
- 936 Kuriakose B, Mathew OK, Perumal RC, Koludarov I, et al. 2020. The Indian
- 937 cobra reference genome and transcriptome enables comprehensive
- 938 identification of venom toxins. *Nat Genet* **52**: 106-117.
- 939 Tan M, Redmond AK, Dooley H, Nozu R, Sato K, Kuraku S, Koren S, Phillippy
- 940 AM, Dove ADM, Read TD. 2021. The whale shark genome reveals patterns
- 941 of vertebrate gene family evolution. *Elife* **10**: e65394.
- 942 Tymowska J. 1991. Polyploidy and cytogenetic variation in frogs of the genus
- 943 *Xenopus*. In *Amphibian cytogenetics and evolution* (eds. Green DM,
- 944 Sessions SK), pp. 259–297, Academic Press, Cambridge.
- 945 Uno Y, Nishida C, Tarui H, Ishishita S, Takagi C, Nishimura O, Ishijima J, Ota H,
- 946 Kosaka A, Matsubara K, et al. 2012. Inference of the protokaryotypes of
- 947 amniotes and tetrapods and the evolutionary processes of
- 948 microchromosomes from comparative gene mapping. *PLoS One* **7**: e53027.
- 949 Uno Y, Nozu R, Kiyatake I, Higashiguchi N, Sodeyama S, Murakumo K, Sato K,
- 950 Kuraku S. 2020. Cell culture-based karyotyping of orectolobiform sharks for
- 951 chromosome-scale genome analysis. *Commun Biol* **3**: 652.
- 952 Venkatesh B, Lee AP, Ravi V, Maurya AK, Lian MM, Swann JB, Ohta Y, Flajnik
- 953 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

1022 repeats' mode. The numbers given to the individual panels (1) to (8), correspond to  
1023 those at the nodes in the phylogenetic tree and are colored differentially to indicate  
1024 comparable divergence times. Diagonal lines are colored according to the level of  
1025 sequence divergence (dark green, 75-100%; light green, 50-75%; orange, 25-50%;  
1026 yellow, 0-25%). (B) Chromosomal homology suggested by synteny conservation of  
1027 one-to-one ortholog pairs. See Methods for details. The color of the ribbons  
1028 connecting the synteny represents each of the three categories of the zebra shark  
1029 chromosomes: eMAC (gray); eMID (magenta); and eMIC (cyan). (C) Chromosomal  
1030 homology between the zebra shark and other vertebrates. Conserved synteny is  
1031 visualized with the inter-specific correspondence of one-to-one orthologs (see  
1032 Methods).

1033

1034 **Figure 4.** Intrachromosomal heterogeneity of sequence characteristics. (A)  
1035 Comparison of global GC-content in 10 kb-long non-overlapping windows between  
1036 the 1-Mb-long ends and the remainders of relatively large chromosomes for diverse  
1037 vertebrates. (B) Comparison of GC-content of protein-coding regions, gene length,  
1038 gene density, and synonymous substitution rate ( $K_s$ ) among the 1 Mb-long  
1039 chromosome ends, their remainders, and relatively small chromosomes (zebra shark  
1040 eMIC and chicken microchromosomes (MIC)), for the zebra shark and chicken,  
1041 respectively. (C) Two-dimensional plots of chromosome lengths and the coverage of  
1042 interspersed repeats and simple tandem repeats. The coloring of the dots follows  
1043 that in Figure 2B. (D) Differential distribution of simple tandem repeats and  
1044 interspersed repeats. Proportions of the sequences identified as simple tandem  
1045 repeats and interspersed repeats in 10 kb-long windows were compared between  
1046 the 1 Mb-long ends of relatively large chromosomes (zebra shark eMAC and chicken

1047 macrochromosomes (MAC)), their remainders, and relatively small chromosomes  
1048 (zebra shark eMIC and chicken MIC). In B and D, significance of difference is  
1049 indicated as follows; \*,  $p$ -value < 0.05/number of tests; \*\*,  $p$ -value < 0.01/number of  
1050 tests; n.s., not significant. The effect sizes in statistical tests are indicated with a  
1051 hyphen for no effect, one dagger symbol ‘†’ for small effect, two dagger symbols for  
1052 medium effect ‘††’, and three dagger symbols ‘†††’ for large effect. Numbers of the  
1053 genomic regions sampled are included in parentheses. See Methods for more details  
1054 about statistical tests.

1055

1056 **Figure 5.** Genomic identification of the zebra shark Chromosome X. (A) Male–  
1057 female ratio of short-read sequencing depth in the shark chromosomes. (B) Male–  
1058 female copy-number difference of zebra shark Chromosome X. Amplification levels  
1059 of the genes on the PAR and the remainders of Chromosome X (scaffold 41) was  
1060 quantified using real-time PCR controlled with amplification of autosomal genomic  
1061 regions, by normalizing the PCR product abundance with that for the individual F1  
1062 (sSteFas1) (see Methods). (C) Cross-species synteny of sex chromosome-linked  
1063 genes based on 1-to-1 orthologs. (D) Characteristic comparisons between the  
1064 different chromosome categories in the zebra shark.  $K_s$  (synonymous substitution  
1065 rate) and  $K_a$  (nonsynonymous substitution rate) values were calculated for 1-to-1  
1066 ortholog shared with the whale shark. Only for GC-content, the PAR was shown  
1067 separately, because it contained few genes. Significance of difference is indicated as  
1068 follows; \*,  $p$ -value < 0.05/number of tests; \*\*,  $p$ -value < 0.01/number of tests; n.s.,  
1069 not significant. The effect sizes are indicated with a hyphen for no effect, one dagger  
1070 symbol ‘†’ for small effect, two dagger symbols ‘††’ for medium effect, and three  
1071 dagger symbols ‘†††’ for large effect in statistical tests. Numbers of the genomic

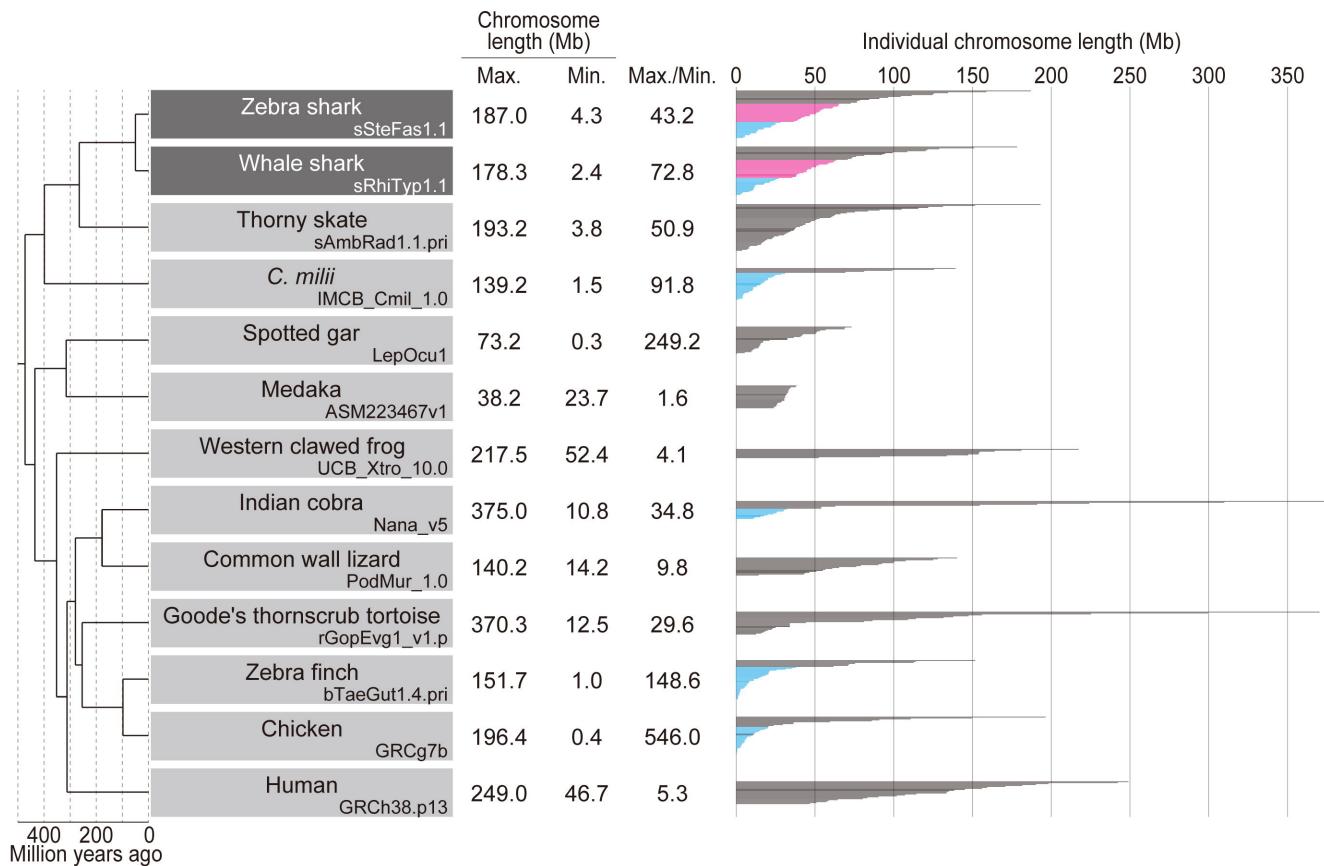
1072 regions sampled was shown in parentheses. See Methods for more details about  
1073 statistical tests. (*E*) Structural comparison between the zebra shark and whale shark  
1074 X Chromosome sequences. We focused more on the zebra shark because we could  
1075 not retrieve a part of the whale shark chromosome that is homologous to the putative  
1076 zebra shark Chromosome X.

1077

1078 **Figure 6.** Zebra shark Hox C genes. (*A*) Genomic structure of the zebra shark Hox  
1079 clusters and their neighboring regions. The exons of the Hox genes are shown in red  
1080 boxes. (*B*) Molecular phylogenetic tree of Hox12 group of genes. The tree was  
1081 inferred with the maximum-likelihood method as described in Methods. (*C*)  
1082 Expression profiles of the zebra shark Hox genes in embryos and various tissues of  
1083 a juvenile.

1084

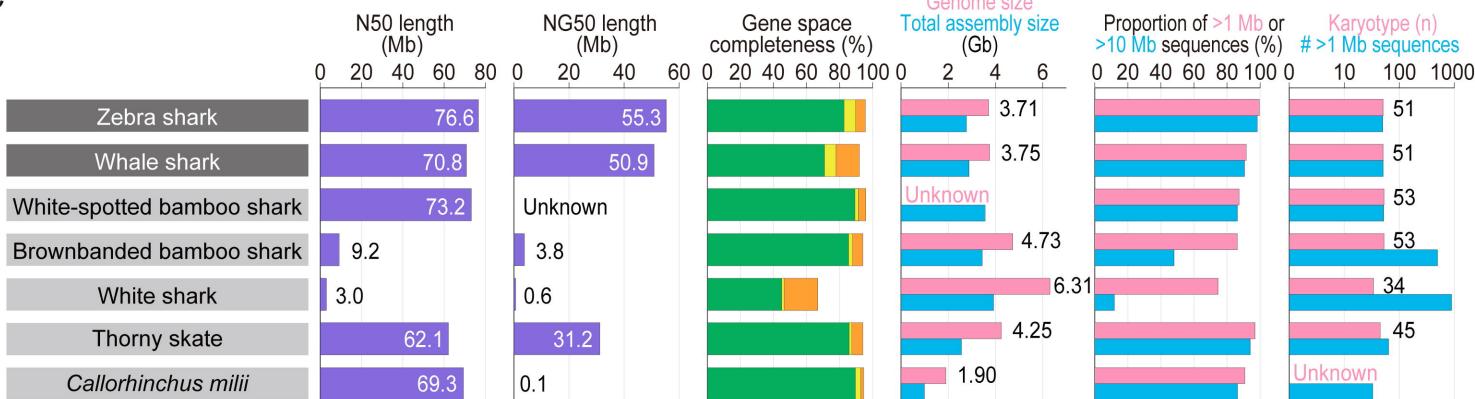
A



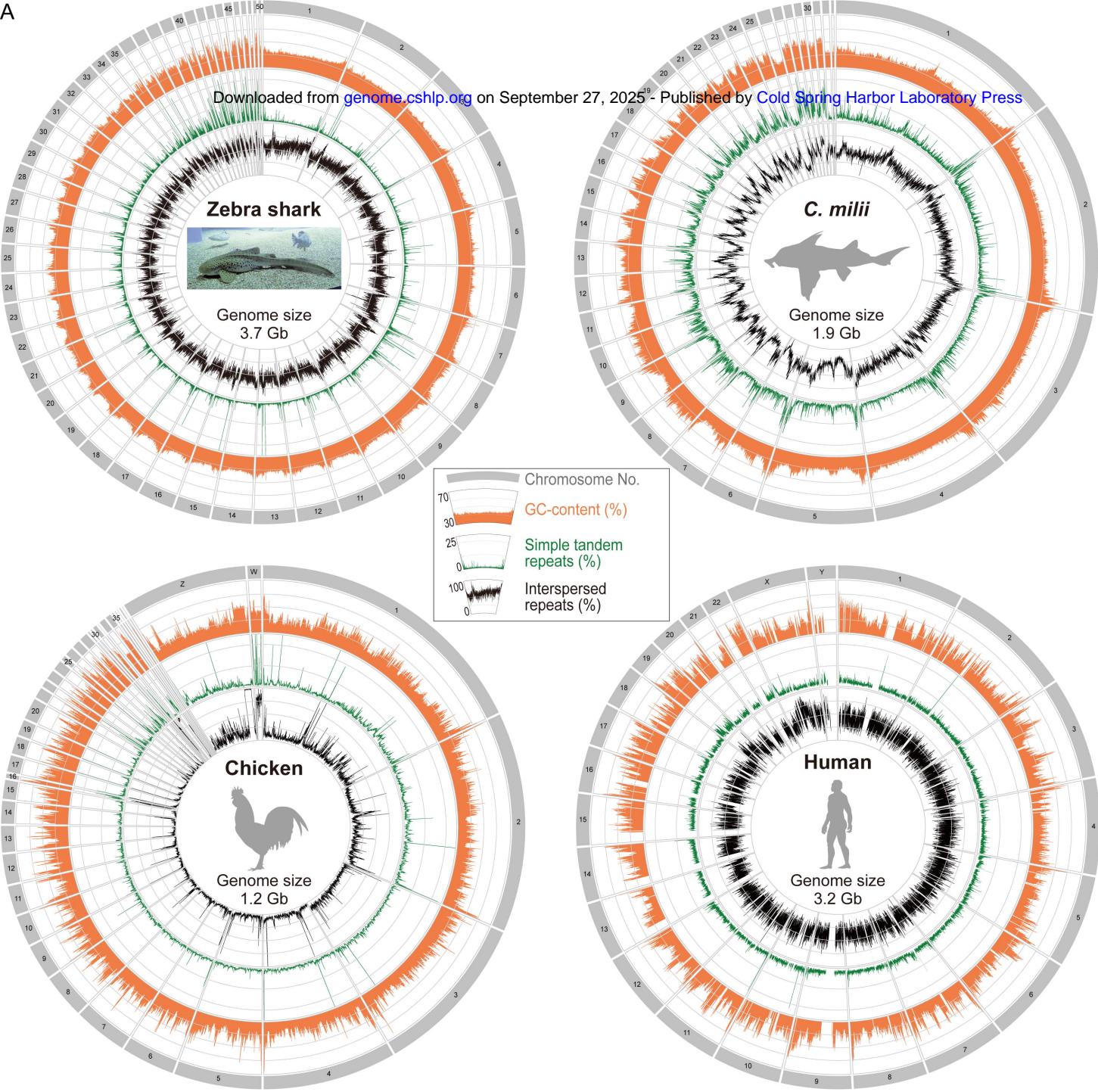
B



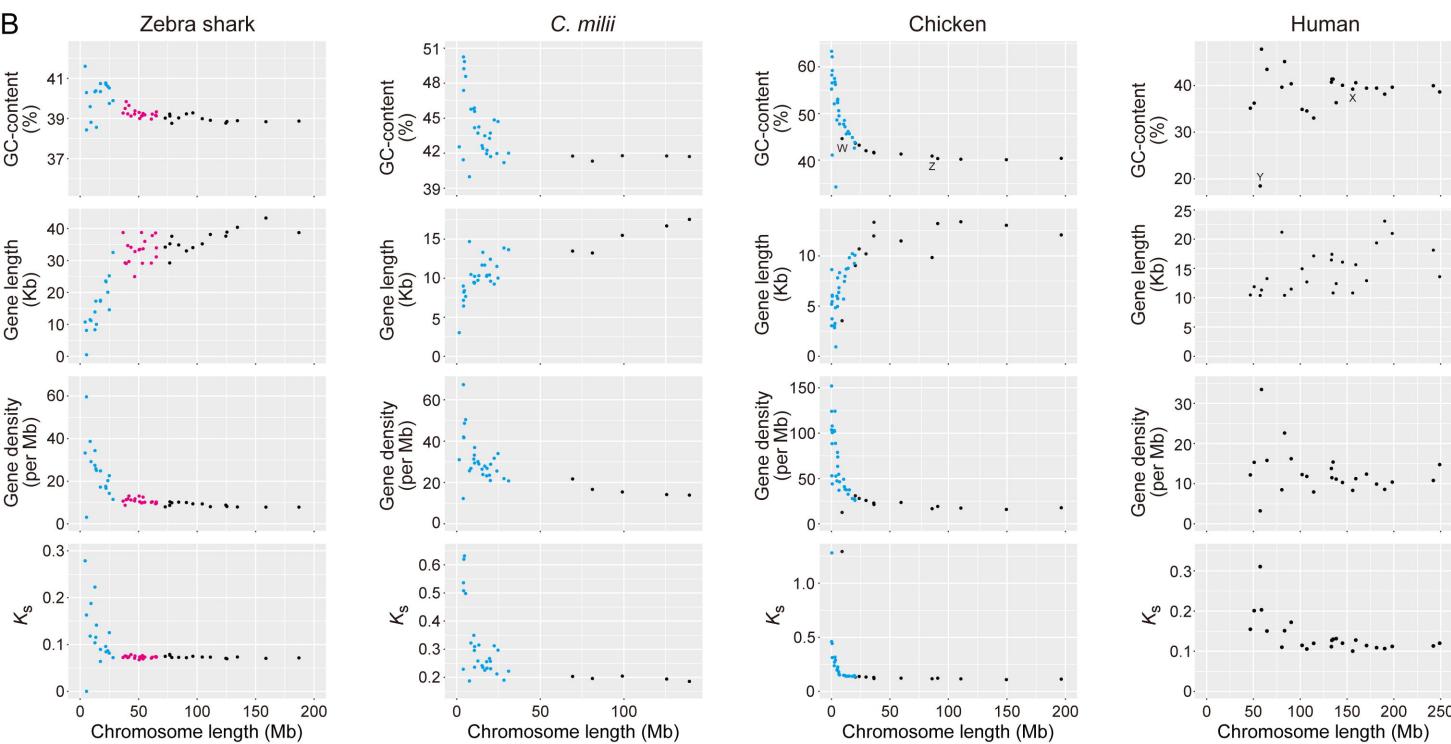
C

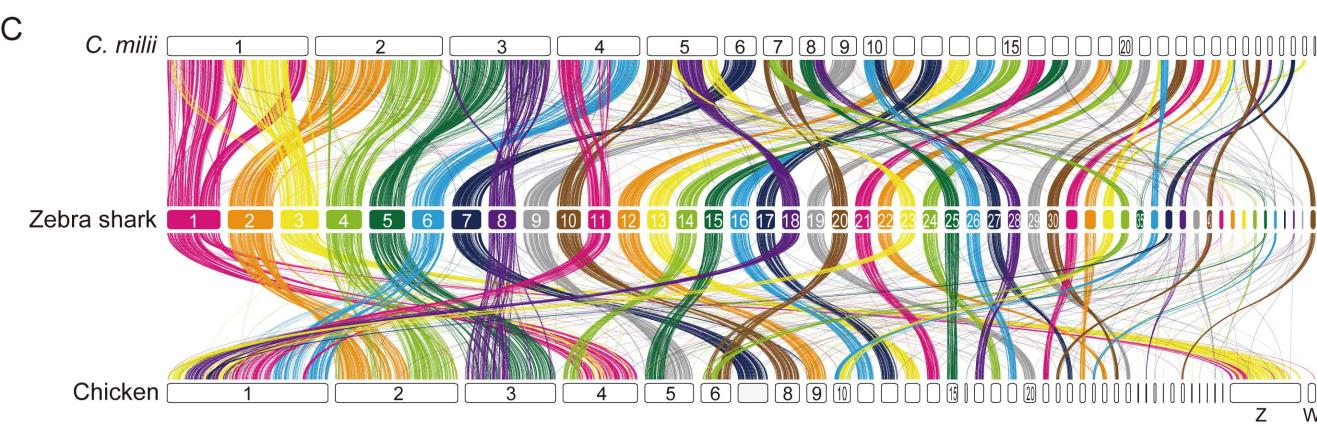
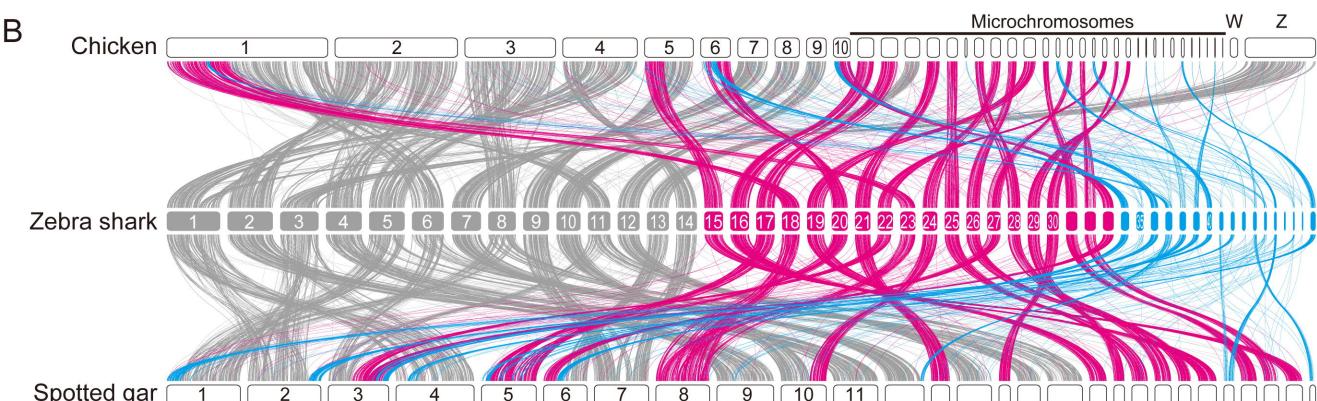
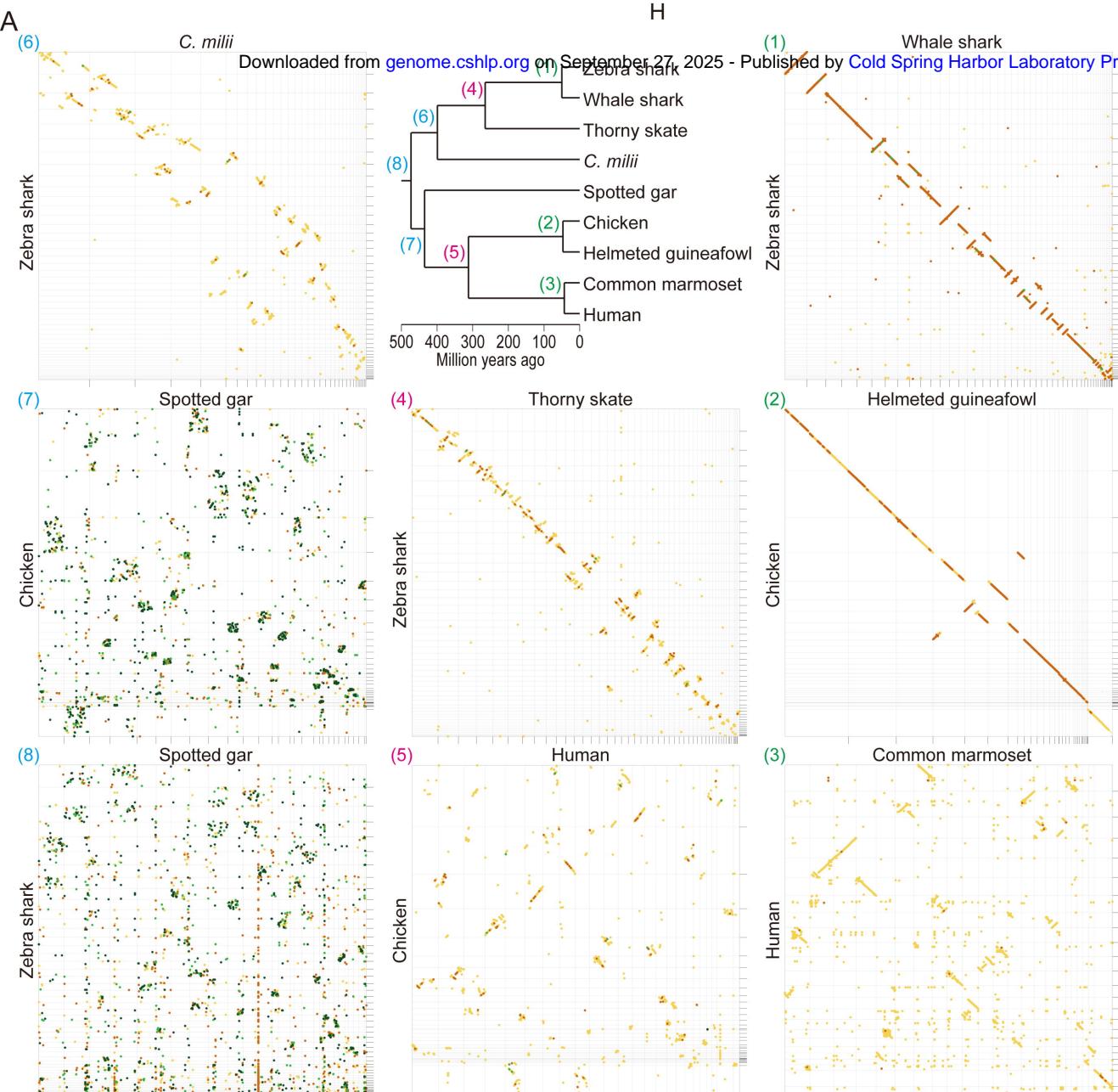


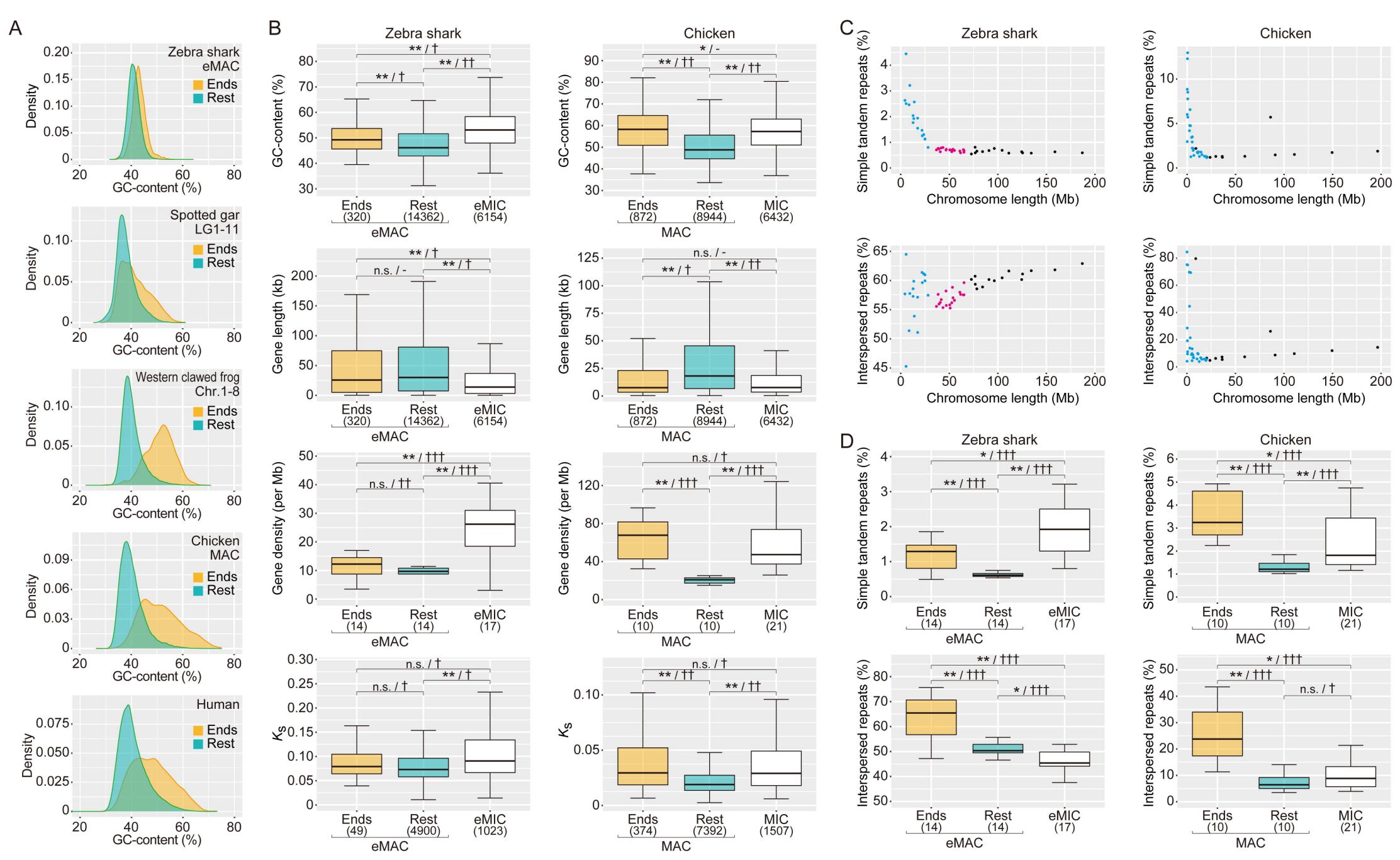
A

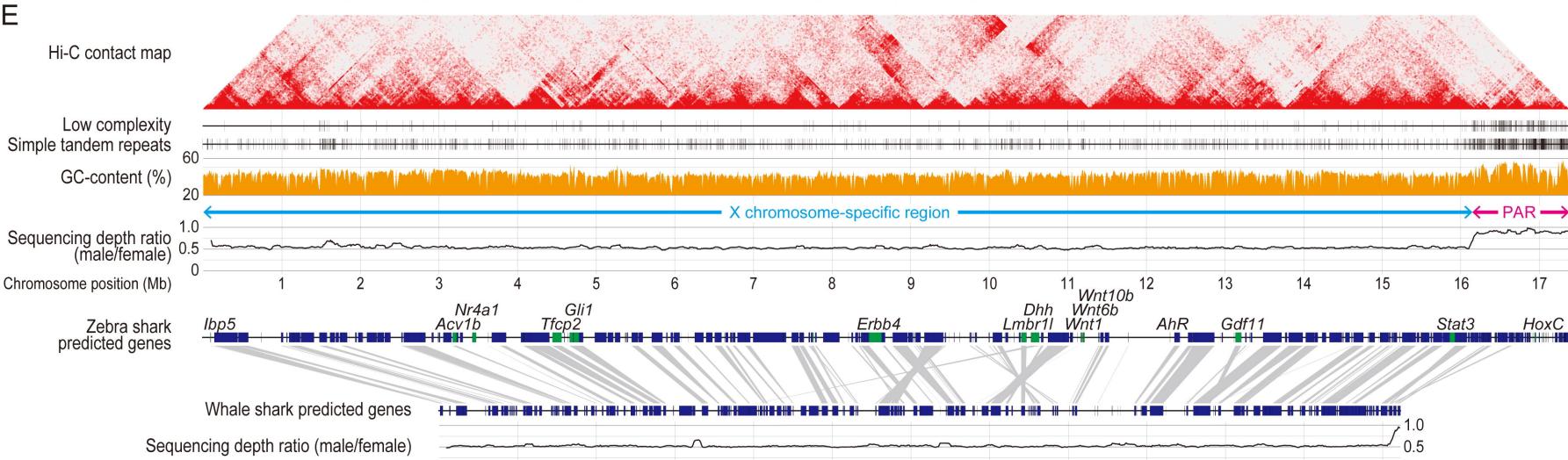
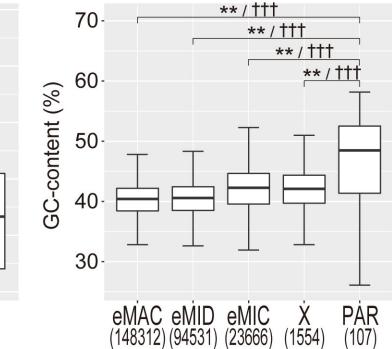
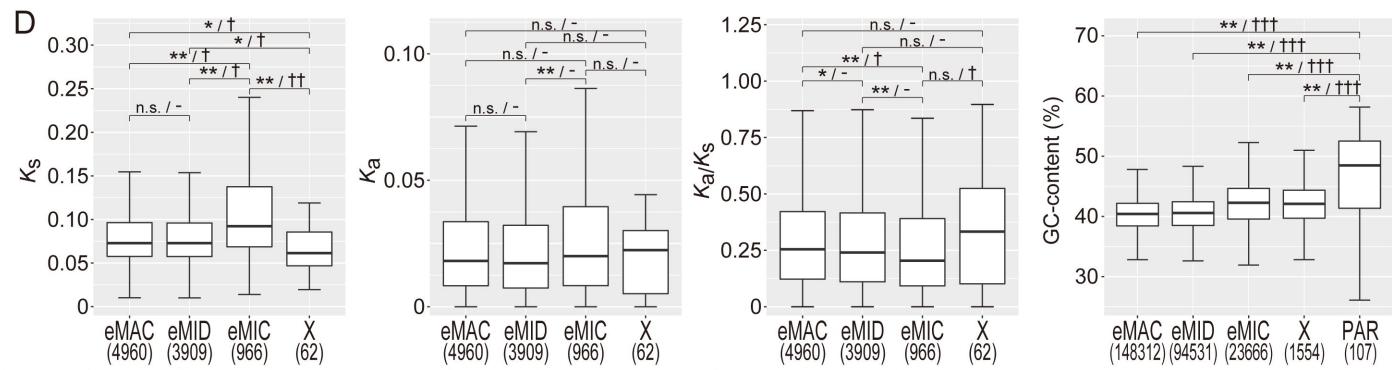
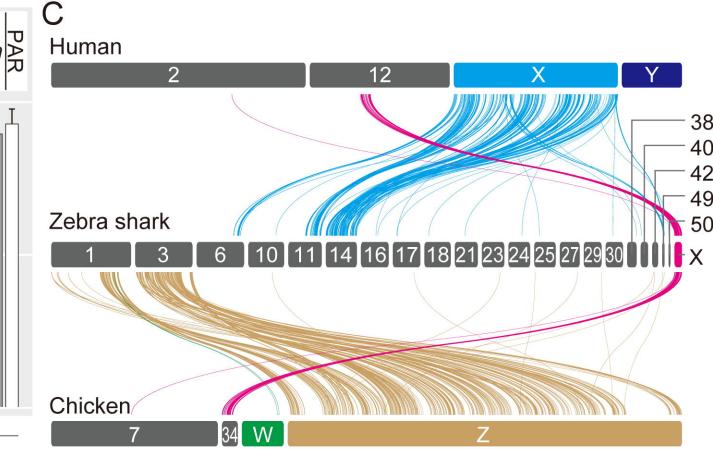
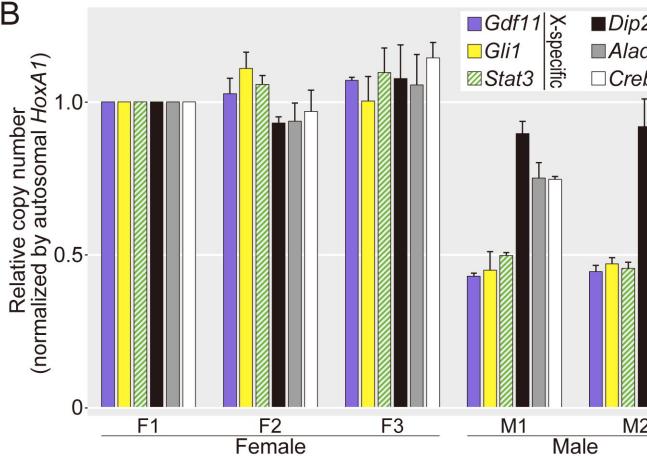
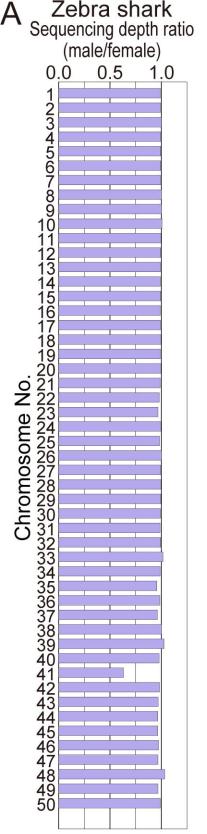


B

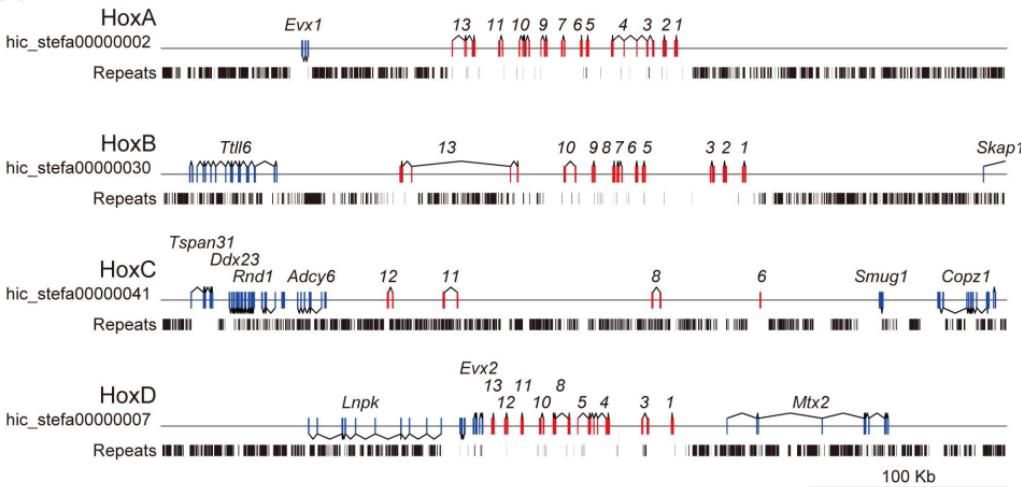




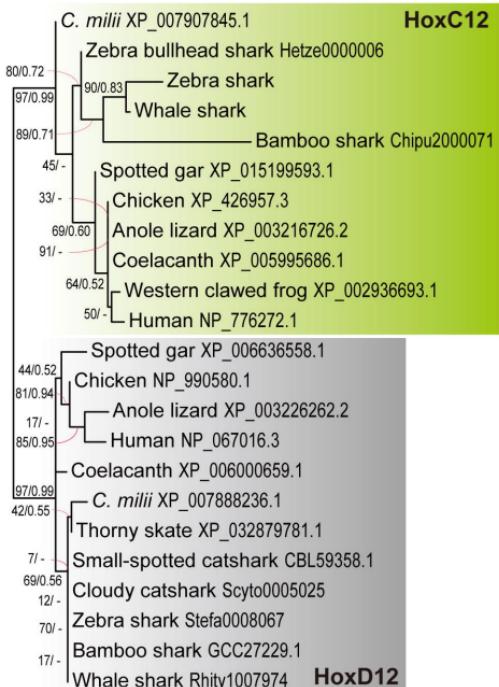




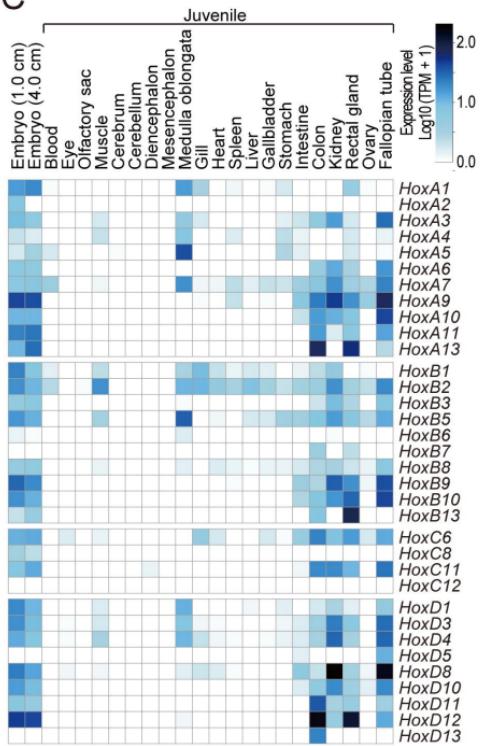
A



B



C





## 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](https://doi.org/10.1101/gr.276840.122)

---

**Supplemental Material** <http://genome.cshlp.org/content/suppl/2023/10/04/gr.276840.122.DC1>

**P<P** Published online August 17, 2023 in advance of the print journal.

**Accepted Manuscript** Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.

**Open Access** Freely available online through the *Genome Research* Open Access option.

**Creative Commons License** This manuscript is Open Access. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International license), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---



---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>