1

2   *Research article*

3

Elasmobranch genome sequencing reveals evolutionary trends of vertebrate karyotype organization

6

7   Kazuaki Yamaguchi[1], Yoshinobu Uno[1*], Mitsutaka Kadota[1], Osamu Nishimura[1], Ryo

8   Nozu[2**], Kiyomi Murakumo[3], Rui Matsumoto[3], Keiichi Sato[2,3], Shigehiro Kuraku[1,4,5#]

9

10   [1]Laboratory for Phyloinformatics, RIKEN Center for Biosystems Dynamics Research

11   (BDR), Kobe, Japan

12   [2]Okinawa Churashima Research Center, Okinawa Churashima Foundation,

13   Okinawa, Japan.

14   [3]Okinawa Churaumi Aquarium, Okinawa, Japan.

15   [4]Molecular Life History Laboratory, Department of Genomics and Evolutionary

16   Biology, National Institute of Genetics, Mishima, Japan.

17   [5]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 the smallest genome sequenced to date among sharks

36 (3.71 Gb), as well as for the whale shark *Rhincodon typus*. Our analysis employing a

37 male–female comparison identified an X chromosome, the first genomically

38 characterized shark sex chromosome. The X chromosome harbors a Hox C cluster

39 whose intact linkage has not been shown for an elasmobranch fish. The sequenced

40 shark 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, which resemble the edges of longer chromosomes. This

44 pattern of intragenomic heterogeneity, previously recognized as peculiar to species

45 with so-called microchromosomes, occurs in more vertebrates including

46 elasmobranchs. We challenge the traditional binary classification of karyotypes as

47 with and without microchromosomes, as even without microchromosomes, shorter

48 chromosomes tend to have higher contents of GC and simple tandem repeats and

49 harbor shorter and more rapid-evolving genes. Such characteristics also appear on

50 the edges of longer chromosomes. Our investigation of elasmobranch karyotypes

51 underpins their unique characteristics and provides clues for understanding how

52 vertebrate karyotypes accommodate intragenomic heterogeneity to realize a

53 complex readout.

1

54

55   Key words: shark, chromosome-scale assembly, microchromosome, karyotype

56   evolution, sex chromosome

57

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

3

83   (blood or embryos) (e.g., at aquariums) enabled karyotyping using culture cells for

84   four orectolobiform shark species in Elasmobranchii (Uno et al. 2020). This has

85   paved the way for rigid evaluation of whole genome sequences. Biological studies on

86   cartilaginous fishes have been hindered by low accessibility to fresh material.

87   Moreover, especially in studying elasmobranchs, genome analysis can encounter

88   inherent difficulty incurred by their large genome sizes (reviewed in Kuraku 2021).

89   These factors have prevented previous efforts on cartilaginous fishes from obtaining

90   a suite of genome sequences supported by karyotype and genome size estimate as

91   well as transcriptome sequencing (Rhie et al. 2021; Read et al. 2017; Tan et al.

92   2021; Weber et al. 2020; Marra et al. 2019; Zhang et al. 2020).

93   Chromosome-level analysis is broadening our scope of comparative

94   genomics (Rhie et al. 2021; Deakin et al. 2019). The difficulty of elasmobranch

95   genome sequencing is manifested in the retrieval of the Hox C cluster, an array of

96   homeobox-containing genes (reviewed in Kuraku 2021). While shark Hox A, B, and

97   D gene clusters have been reliably assembled with few gaps (Hara et al. 2018;

98   Mulley et al. 2009), assembling the Hox C cluster, which was initially reported as

99   missing from elasmobranch genomes (King et al. 2011), suffered from high GC-

100   content and frequent repetitive elements, resulting in fragmentary sequences (Hara

101   et al. 2018; reviewed in Kuraku 2021). This difficulty is expected to be overcome by

102   the application of a rising approach to scaffolding the genomic fragments up to the

103   chromosome scale using chromatin contact data (Yamaguchi et al. 2021;

104   Dudchenko et al. 2017) as well as long-read sequencing.

105   Another expectation from chromosome-level genome analysis is the

106   identification of sex chromosomes, although sequencing and assembling sex

107   chromosomes often suffer from difficulties caused by high repetitiveness or uneven

4

108   sequence depth (Rhie et al. 2021; Ma et al. 2021). While sex determination

109   mechanisms have been revealed by an increasing number of studies on

110   osteichthyan vertebrates (Graves 2016; Pennell et al. 2018), no report is available

111   for vertebrate species outside osteichthyans, namely cyclostomes and

112   chondrichthyans. The quest for sex determination mechanisms can be initiated by

113   the identification of sex chromosomes as already improvised in several vertebrate

114   species (Franchini et al. 2018).

115   Some vertebrate karyotypes, including most bird karyotypes, consist of small-

116   sized chromosomes, or microchromosomes (Figure 1A) that were initially recognized

117   as shorter than 1 μm in cytogenetic observations (Ohno et al. 1969; Ohno 1970).

118   Microchromosomes are known to have higher GC-content, higher gene density, and

119   different chromatin states compared with the remaining macrochromosomes

120   (International Chicken Genome Sequencing Consortium (ICGSC) 2004; Burt 2002;

121   Waters et al. 2021). Some genome sequence-based studies regard chromosomes

122   smaller than 20 Mb as microchromosomes and investigated their possible common

123   origin (Nakatani et al. 2021; ICGSC 2004), but they have not been unambiguously

124   defined on a cross-species basis. Accumulating information from synteny-based

125   analysis suggests that the last common jawed vertebrate ancestor already

126   possessed microchromosomes (Braasch et al. 2016; Waters et al. 2021; Nakatani et

127   al. 2007, 2021). However, this hypothesis needs to be examined by incorporating

128   more diverse vertebrate taxa into the comparison, on a solid basis of experimentally

129   validated karyotypic configurations of individual species.

130   In this study, we focused on the zebra shark *Stegostoma tigrinum* (or leopard

131   shark; Figure 1B) and report its whole-genome sequences for the first time. This

132   species has the smallest genome size (3.71 Gb) among the elasmobranch species

5

133 whose genomes have been sequenced to date. Each of the resultant chromosome-

134 scale genome assemblies of the zebra shark, as well as the whale shark (Figure 1B),

135 have been constructed using samples from a single individual (which was not

136 achieved in earlier efforts: Tan et al. 2021; Read et al. 2017; Weber et al. 2020) and

137 controlled by referring to their karyotypes. Using the obtained sequences, we

138 performed comparative investigations to characterize the diversity of chromosomal
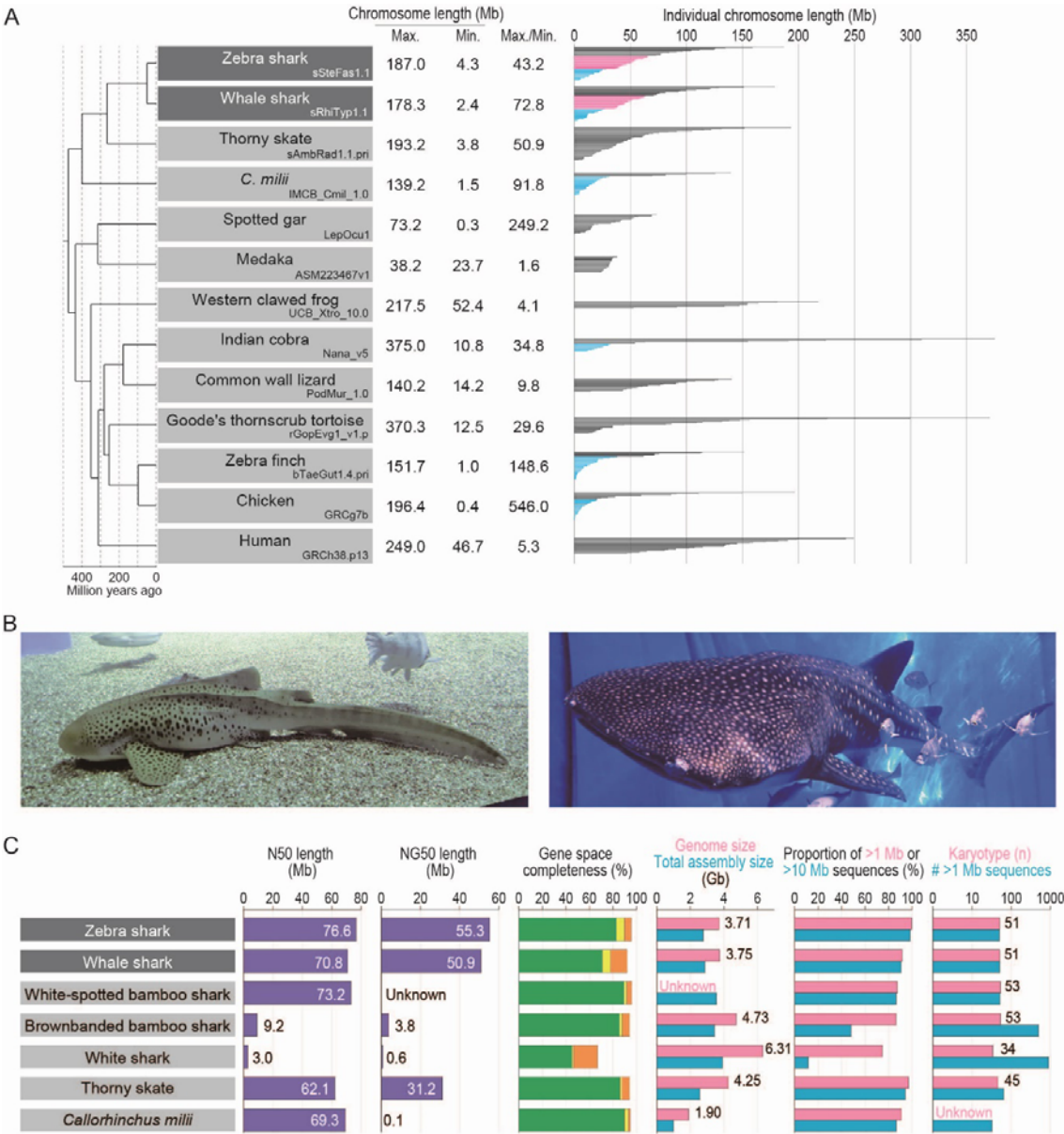
139 organization.

140

141   **Figure 1**. Shark species studied and comparative statistics of their genome

142   assemblies. (*A*) The karyotypes of diverse vertebrate species are depicted with the

143   length of individual chromosomal DNA sequences. The maximum and minimum

144   chromosome lengths are based on the records in NCBI Genomes.

145   Microchromosomes for osteichthyans are shown in light blue according to individual

146   original reports (Nakatani et al. 2021; Suryamohan et al. 2020; Knief and Forstmeier

147   2016; ICGSC 2004), while the eMID and eMIC of the two shark species whose

148   genomes have been sequenced in the present study (see Results) are shown in

149   orange and light blue, respectively. (*B*) Zebra shark (left) and whale shark (right). (*C*)

150   Statistics of the genome assemblies. The identifiers of the genome assemblies, as

151   well as statistical comparison of more metrics, are included in Supplementary Table

152   1. Gene space completeness shows the proportions of selected one-to-one protein-

153   coding orthologs with 'complete' (green), 'duplicated' (yellow), and 'fragmented'

154   (orange) coverages retrieved in the sequences by the BUSCO pipeline (see

155   Methods). The details of the genome sizes and karyotypes included are based on

156   existing literature (Uno et al. 2020; Schwartz and Maddock 1986, 2002; Hardie and

157   Hebert 2004).

158

159   **Results**

160   **The smallest shark genome sequenced to date**

161   We focused on the zebra shark (or leopard shark) *Stegostoma tigrinum* (formerly, *S.*

162   *fasciatum*) with the haploid nuclear DNA content of 3.79 pg (3.71 Gb; Kadota et al. in

163   prep) and the karyotype of 2n = 102 (Uno et al. 2020). Using genomic DNA extracted

164   from blood cells of a female adult, the whole genome was sequenced and

165   assembled with short reads. The resultant sequences were further scaffolded using

7

166 Hi-C data obtained from blood cell nuclei of the same individual (see Methods). The

167 number of output scaffold sequences longer than 1 Mbp (thereafter tentatively

168 designated 'chromosomes') was 50, which closely approximates its chromosome

169 number revealed by the cytogenetic observation and karyotypic organization

170 previously characterized by us using primary cultured cells (Figure 1C). The retrieval

171 of the chromosomal scale was also guaranteed by the N50 scaffold length of 76.6

172 Mb (Figure 1C).

173   We also performed *de novo* whole genome sequencing of an adult male

174 whale shark, *Rhincodon typus*, using Linked-Read data and Hi-C scaffolding (see

175 Methods). The number of sequences greater than 1 Mbp matched the number of

176 chromosomes in the karyotype (n = 51) (Figure 1C), and the N50 scaffold length of

177 the resultant assembly reached 70.8 Mb, significantly exceeding that of the

178 assemblies previously published for this species (Figure 1C). To date, our product is

179 the only chromosome-scale genome assembly for this species that was built

180 consistently from a single individual.

181   While we recognize a significant gap between the estimated and retrieved

182 total sequence lengths, genome assemblies for the zebra shark and whale shark

183 exhibited high completeness of protein-coding gene space of more than 90% (Figure

184 1C). Prediction of protein-coding genes on the zebra shark and whale shark

185 genomes, performed by incorporating homolog sequences and transcriptome data,

186 resulted in 33,222 and 35,334 genes, respectively (Supplementary Tables 1 and 2),

187 which allowed downstream molecular biological analysis.

188

189 **Karyotypic trends in sharks**

8

190   The obtained zebra shark genome assembly consisted of chromosome sequences

191   of highly variable length with a gradual slope, in accordance with our previous

192   cytogenetic observation (Uno et al. 2020), spanning from 187.0 Mb down to 4.3 Mb

193   (Figure 2A). This pattern does not resemble the length variation in the *Callorhinchus*

194   *milii* or the chicken (Figure 2A). The chicken especially exhibits a steep slope,

195   marked by a number of chromosomes shorter than 20 Mb (microchromosomes)

196   (Figures 1A and 2A). Gradualism in the chromosome length distribution is also

197   observed in the whale shark (Supplementary Figure 1), white-spotted bamboo shark

198   (Zhang et al. 2020), and thorny skate (Rhie et al. 2021), and is assumed to be typical

199   karyotypic organization of elasmobranchs (Supplementary Figure 2). For simplicity,

200   we tentatively designated (1) zebra shark chromosomes 1 to 14, longer than 70 Mb;

201   (2) chromosomes 15 to 33, between 30 and 70 Mb; and (3) chromosomes 34 to 50,

202   shorter than 30 Mb as elasmobranch macro-, middle-sized-, and micro-

203   chromosomes (abbreviated into eMAC, eMID, and eMIC, respectively) that are

204   differentially colored in Figures 1A and 2B. This categorization also applies directly to

205   the whale shark chromosomes (Figure 1A; Supplementary Figure 2).

206   The zebra shark genome exhibits consistent intra-chromosomal GC-content

207   compared with other species, and its chromosome ends in general, as well as the

208   shorter chromosomes, tend to have relatively higher GC-content (Figure 2A). Zebra

209   shark DNA sequences show a uniformly high frequency of interspersed repeats

210   throughout the genome, while its shorter chromosomes tend to have higher simple

211   tandem repeat content (Figure 2A). These characteristics are also observed in the

212   whale shark genome (Supplementary Figure 1), but the co-occurrence of these two

213   features was not explicitly observed in the other species compared (Figure 2A;

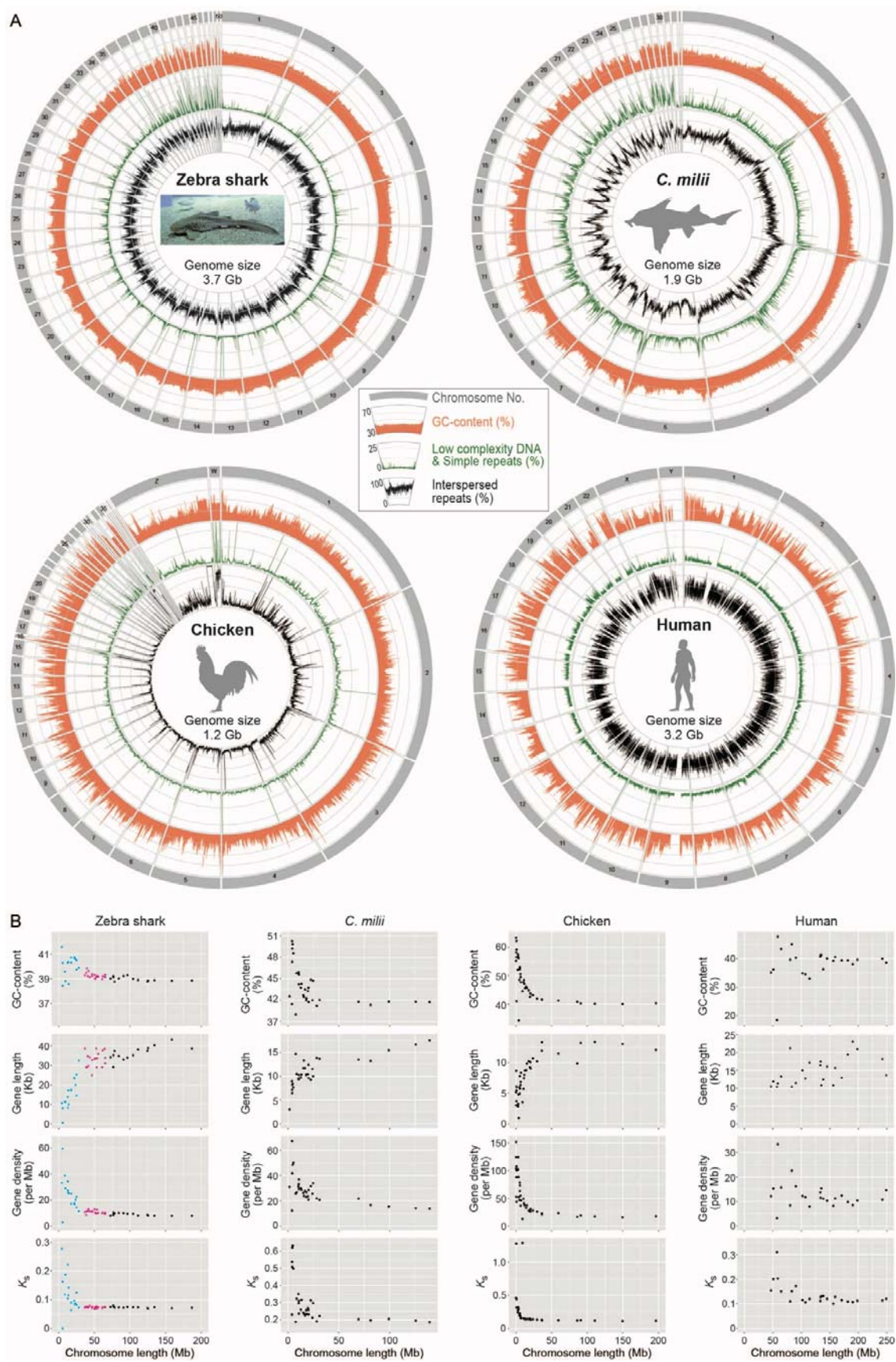214   Supplementary Figure 1).

215

10

216   **Figure 2.** Chromosomal sequence compositions of the zebra shark and selected

217   vertebrate species. (*A*) Sequence characterization of different chromosomal

218   segments. The orange areas show GC-content (30%–70%), while the green and

219   black lines show content of simple tandem repeats (0%–25%) and interspersed

220   repeats (0%–100%), respectively, in 100-kb-long non-overlapping windows. (*B*) Two-

221   dimensional plots of GC-content and median values of gene length, gene density,

222   and the median of synonymous substitutions per synonymous site (K_s) for protein-

223   coding genes on individual chromosomes. Computation of K_s was performed for

224   each of the four species in order by involving a pair species, namely, whale shark

225   *Rhincodon typus*; small-eyed rabbitfish *Hydrolagus affinis*; helmeted guineafowl

226   *Numida meleagris*; and common marmoset *Callithrix jacchus*. Zebra shark

227   chromosomes are shown as three groups, eMAC, eMID, and eMIC (see text for

228   details), in black, magenta, and cyan, respectively.

229

230   **Vertebrate-wide comparisons including elasmobranchs**

231   To analyze how the karyotypes of these shark species were derived, chromosomal

232   nucleotide sequences were compared between species pairs with variable

233   divergence times (Figure 3A). The comparison between the zebra shark and the

234   whale shark exhibited a high similarity in chromosomal organization with few

235   intrachromosomal breaks (panel 1 in Figure 3A). The high similarity of this

236   orectolobiform shark species pair suggests high conservation of genomic sequences

237   from around or earlier than 50 million years ago (see Discussion), compared with

238   osteichthyan species pairs in a similar divergence time range—the human-marmoset

239   pair diverged about 43 million years ago (panel 3 in Figure 3A). The relatively high

240   conservation of chondrichthyan chromosome organization is also supported by the

11

241   comparisons between more distantly related species pairs. The similarity of the

242   zebra shark genome sequences to the thorny skate (panel 4) and the *C. milii* (panel

243   6) exceeded those for the species pairs with about 300- and 400-million-year

244   divergences, respectively (panel 5 and panels 7 in Figure 3A).

245   Previous studies sought to reconstruct the process of karyotypic evolution of

246   vertebrates but often lacked elasmobranchs in the dataset (Nakatani et al. 2021;

247   Sacerdot et al. 2018). In the present study, we performed a gnathostome-wide

248   comparison of the syntenic location of one-to-one orthologs, including the zebra

249   shark (Figure 3B; Supplementary Figure 3). Remarkably, the majority of the genomic

250   regions in the zebra shark eMIC were shown not to share one-to-one orthologs with

251   the so-called microchromosomes of the chicken or spotted gar (Figure 3B).

252   Moreover, the smallest spotted gar chromosomes were frequently shown to be

253   homologous to zebra shark eMID (chromosomes 15–33), and not to its eMIC

254   (chromosomes 34–) (Figure 3B). It was also shown that zebra shark chromosomes

255   10, 25, 26, 27, and 28 have chicken homologs of similar size but their *C. milii*

256   homologs have been fused into larger chromosomes, while zebra shark

257   chromosomes 2, 3, 4, 5, 6, 8, 11, 18, 20, 23, 33, 34, and 36 are likely to be products

258   of fissions that occurred in the elasmobranch lineage (Figure 3C; Supplementary

259   Figure 3). These results refute the common origins of microchromosomes among

260   chicken, spotted gar, and zebra shark and indicate a more drastic reorganization of

261   karyotypes at the base of jawed vertebrates than previously inferred from the

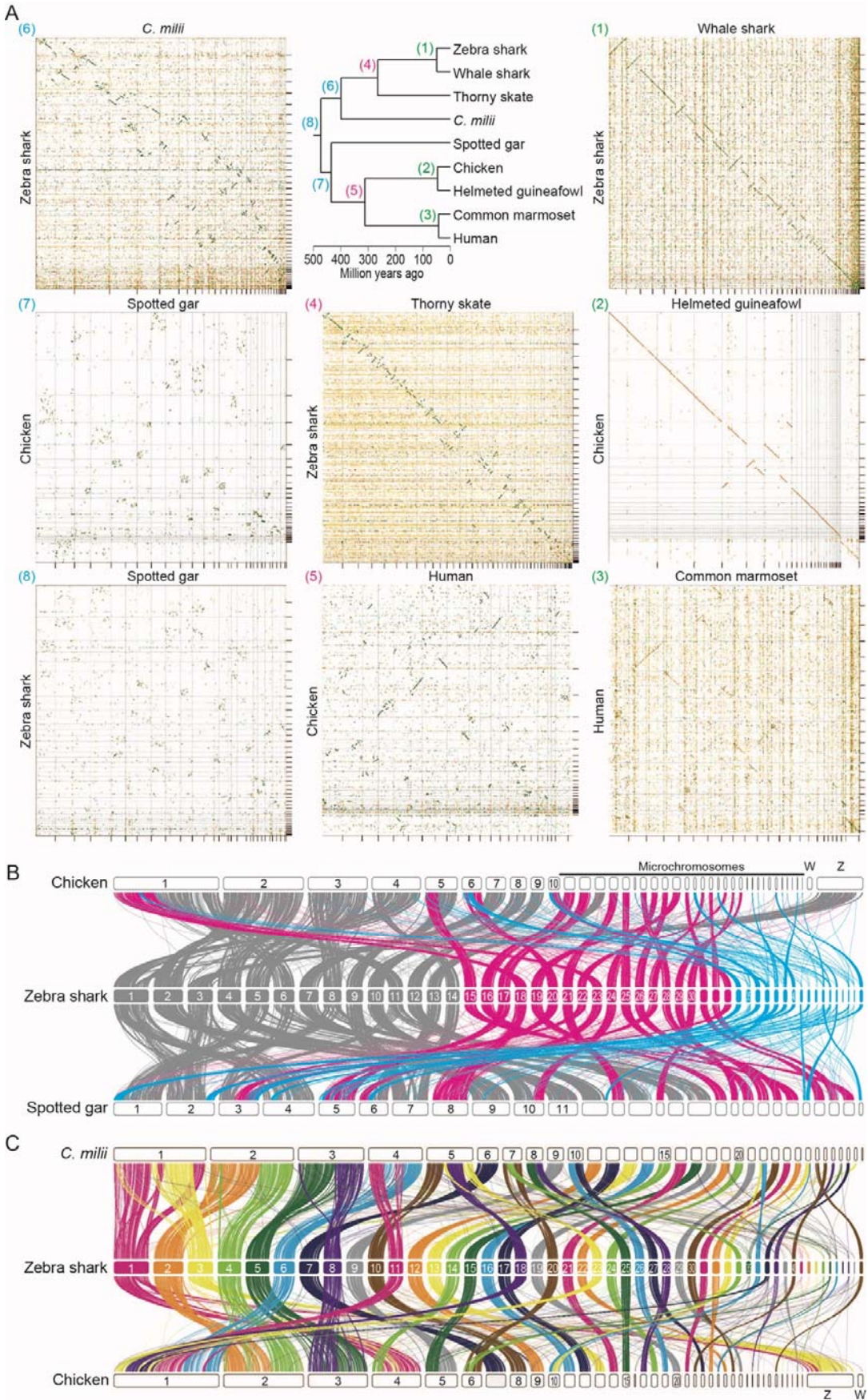262   comparison involving only the *C. milii* as a cartilaginous fish (Nakatani et al. 2021).

263

13

264　**Figure 3**. Cross-species investigation of chromosomal homology. (*A*) Dot matrices

265　showing genome sequence similarities for selected pairs of vertebrate species with

266　variable divergence times. Sequences of high similarity are shown with diagonal

267　lines by the program D-GENIES (Cabanettes and Klopp 2018). The numbers given

268　to the individual panels,(1) to (8), correspond to those at the nodes in the

269　phylogenetic tree and are colored differentially to indicate comparable divergence

270　times. Diagonal lines are colored according to the level of sequence similarity (dark

271　green, 75-100%; light green, 50-74%; orange, 25-49%; yellow, 0-24%), and frequent

272　repetitive sequences cause yellow dotted signals in the background. (*B*)

273　Chromosomal homology suggested by synteny conservation of one-to-one ortholog

274　pairs. See Methods for details. The color of the ribbons connecting the synteny

275　represents each of the three categories of the zebra shark chromosomes: eMAC

276　(gray); eMID (magenta); and eMIC (cyan). (*C*) Chromosomal homology between the

277　zebra shark and other vertebrates. Conserved synteny is visualized with the inter-

278　specific correspondence of one-to-one orthologs (see Methods).

279

280　**Length-dependent properties of chromosomes**

281　To characterize shark chromosomes in depth, base compositions, gene distributions,

282　and molecular evolutionary rates quantified with genic synonymous substitutions (K_s)

283　were investigated (Figure 2B). As previously observed in birds and reptiles (Kuraku

284　et al. 2006; Matsubara et al. 2012; Burt 2002; Waters et al. 2021; Srikulnath et al.

285　2021), relatively short zebra shark chromosomes tend to have higher GC-content,

286　higher gene density, and smaller gene length than its larger chromosomes (Figure

287　2B). This chromosome length-dependent pattern is most pronounced in the chicken

288　in which microchromosomes are canonically implicated (Figure 2B). Remarkably,

14

289   this chromosome length-dependent pattern is observed not only in the spotted gar

290   for which homology of microchromosomes to chicken was previously implicated

291   (Braasch et al. 2016) but also in the western clawed frog *Xenopus tropicalis* which

292   conventionally is thought to have no microchromosomes recognized cytogenetically

293   (Uno et al. 2012) (Supplementary Figures 2).

294   Although this chromosome length-dependent pattern seems prevalent among

295   phylogenetically diverse vertebrate species, some short chromosomes exhibit

296   exceptionally low GC-content, such as chromosome 26 of the *C. milii*, chromosome

297   29 and 33 of the chicken, and Linkage Group 29 of the spotted gar (Figure 2;

298   Supplementary Figures 1 and 2). These exceptions evoke a caution for the

299   generalization of common characteristics of short chromosomes. It needs to be

300   carefully examined whether the relatively short sequences with exceptionally low

301   GC-content are fragments of large chromosomes that failed to be assembled to a

302   chromosome scale.

303

304   **Do 'macrochromosome' ends resemble microchromosomes?**

305   We analyzed regional variations within individual chromosomes of diverse

306   vertebrates, including the zebra shark. To further characterize the distinct trend of

307   chromosomal ends indicated in Figure 2, we separated the 1-Mb-long ends from

308   relatively large chromosomes and analyzed the trends of genomic sequences as well

309   as those of the coding genes (Figure 4A). Intact chromosome ends are known to be

310   occupied by telomeric or subtelomeric simple repeats. Thus, we first examined

311   whether the end of the chromosomal sequences used in this study harbor sufficient

312   protein-coding genome segments that provide unambiguous trends. Our comparison

313   detected a remarkable difference in GC-content of protein-coding regions between

314   the 1-Mb-long chromosome ends and the remainder of the chromosome

315   (Supplementary Figure 4), which resembled the pattern of the genomic sequences

316   (Figure 4A). This observation suggests that the 1-Mb-long margins of the

317   chromosomal DNA sequences employed in this study are not occupied solely by

318   telomeric or subtelomeric simple repeats, but also engulf functional protein-coding

319   genes that are affected by the trends of flanking genomic environments. This

320   prompted us to analyze more genomic and genic properties of chromosome

321   sequence ends. Our further comparison consistently revealed an increase of gene

322   density and synonymous substitution rate, as well as a decrease in gene length, in

323   the ends of relatively large chromosomes, compared with their remainders (Figure

324   4B). In both the zebra shark and chicken, the characteristics of chromosome ends

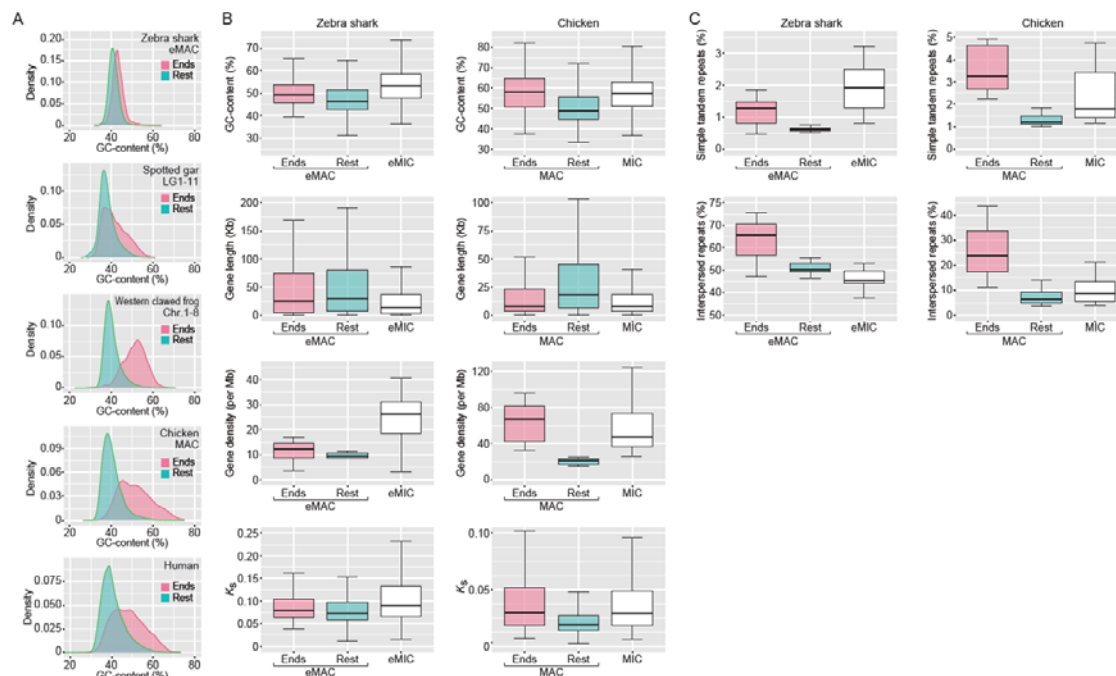325   resemble those of relatively small chromosomes (Figure 4B).

326

Figure 4. Intrachromosomal heterogeneity of sequence characteristics. (*A*)

Comparison of global GC-content in 10 Kb-long non-overlapping windows between

330 the 1-Mb-long ends and the remainders of relatively large chromosomes for diverse

331 vertebrates. (*B*) Comparison of global GC-content of protein-coding regions, gene

332 length, gene density, and synonymous substitution rate (K_s) in 10 Kb-long windows

333 among the 1 Mb-long chromosome ends, their remainders, and relatively small

334 chromosomes (zebra shark eMIC and chicken microchromosomes (MIC)), for the

335 zebra shark and chicken, respectively. (*C*) Differential distribution of simple tandem

336 repeats and interspersed repeats. Proportions of the sequences identified as simple

337 tandem repeats and interspersed repeats in 10 Kb-long windows were compared

338 between the 1 Mb-long ends of relatively large chromosomes (zebra shark eMAC

339 and chicken macrochromosomes (MAC)), their remainders, and relatively small

340 chromosomes (zebra shark eMIC and chicken MIC).

341

342 **Intra-genomic repetitive element distribution**

343 We also analyzed the distribution of repetitive elements on chromosomes of variable

344 length. In fact, previous studies yielded equivocal observations. Some of those

345 studies showed a higher abundance of repetitive elements on larger chromosomes

346 (Koochekian et al. 2022), whereas others indicated localization biased toward

347 smaller chromosomes (Hara et al. 2018). So far, no solid cross-species comparison

348 has been made by taking the difference of repeat classes into account. In the newly

349 obtained shark genome sequences, we separately quantified the sequence

350 proportions identified as interspersed repeats (LINE, SINE, LTR, and DNA elements)

351 and simple tandem repeats (simple repeats and low-complexity DNA sequences,

352 including satellites). This resulted in contrasting patterns between these two repeat

353 categories. Along with chromosome length increase, the interspersed repeat content

354 increases, whereas the simple tandem repeat content decreases (Figure 2A;

17

355 Supplementary Figure 1). Remarkably, the higher frequency of simple tandem

356 repeats on smaller chromosomes is commonly observed in other vertebrates,

357 including the chicken and the *C. milii* (Figure 2A).

358   As examined above for other characteristics, we dissected the observed

359 chromosome length-dependent trend of repeat distribution, again by isolating the 1-

360 Mb-long ends versus the remainder of relatively large chromosomes (Figure 4C). In

361 this comparison, we observed a higher content of interspersed repeats in the ends of

362 relatively large chromosomes, namely zebra shark eMAC and chicken MAC (Figure

363 4C). The higher repeat content was commonly observed in relatively small

364 chromosomes (zebra shark eMIC and chicken MIC), except that interspersed repeat

365 content is reduced in zebra shark eMIC (Figure 4C).

366

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

368 So far, there has been no intensive DNA sequence-based characterization of sex

369 chromosomes for chondrichthyan species. Expecting that a sex chromosome will

370 exhibit a distinct male–female ratio of sequencing depth (Palmer et al. 2019), we

371 performed short-read sequencing of the whole genomes for both sexes in zebra

372 shark and whale shark. For these species, our previous cytogenetic analysis did not

373 detect any heteromorphic sex chromosomes (Uno et al. 2020). Our comparison

374 among different chromosomes detected a remarkably lower male-to-female

375 sequencing depth ratio of close to 0.5 for chromosome 41 of both these species

376 (Figure 5A). This suggests male as a heterogametic sex and the XY system for

377 these species. Next, we investigated the origin of this putative X chromosome of

378 these species by locating human and chicken orthologs of the genes on shark

379 chromosomes (Figure 5B, magenta). This revealed chromosome-level homology of

18

380   the putative X chromosomes of the zebra shark to a part of human chromosome 12

381   and chicken chromosome 34 (Figure 5B, magenta). Neither the human X, human Y,

382   chicken Z, nor chicken W chromosomes exhibited pronounced homology to the

383   putative zebra shark X chromosome. The putative shark X chromosome identified in

384   this study do harbor orthologs of a number of well-studied regulatory factors but do

385   not harbor the orthologs of the sex determination genes identified in other

386   vertebrates, Dmrt1- or Sox3-related transcription factors as well as components of

387   TGF-β signaling pathway, such as Amh, Amhr2, Bmpr1b, Gsdf, and Gdf6 (Bertho et

388   al. 2021).

389   In the approximately 17 Mb-long sequence of the putative zebra shark X

390   chromosome, one 1.5 Mb-long end exhibited a male–female sequencing depth ratio

391   of nearly 1.0 (Figure 5D). This region, with a sequence depth comparable to that of

392   the autosomes, is deduced to be a pseudoautosomal region (PAR) that is likely

393   shared between the heterogametic sex chromosomes X and Y (Smeds et al. 2014;

394   Palmer et al. 2019). We characterized possible unique patterns of molecular

395   evolution typical of sex chromosomes (Figure 5C). In the X chromosome, protein-

396   coding genes exhibited a relatively low synonymous substitution rate (K_s) and high

397   non-synonymous substitution rate (K_a), resulting in an increased K_a/K_s ratio. Our

398   comparison also revealed higher frequency of low-complexity repeats as well as

399   higher GC-content in the PAR (Figure 5C and D).

400   Our comparison of the ortholog location between the putative X chromosomes

401   of the two species supported a high cross-species conservation of the chromosome

402   structure (Figure 5D). In the current whale shark assembly, chromosome 41, which

403   corresponds to the putative X chromosome, may not cover the whole chromosome,

404   possibly excluding one end of the PAR (Figure 5D).

19

405

406
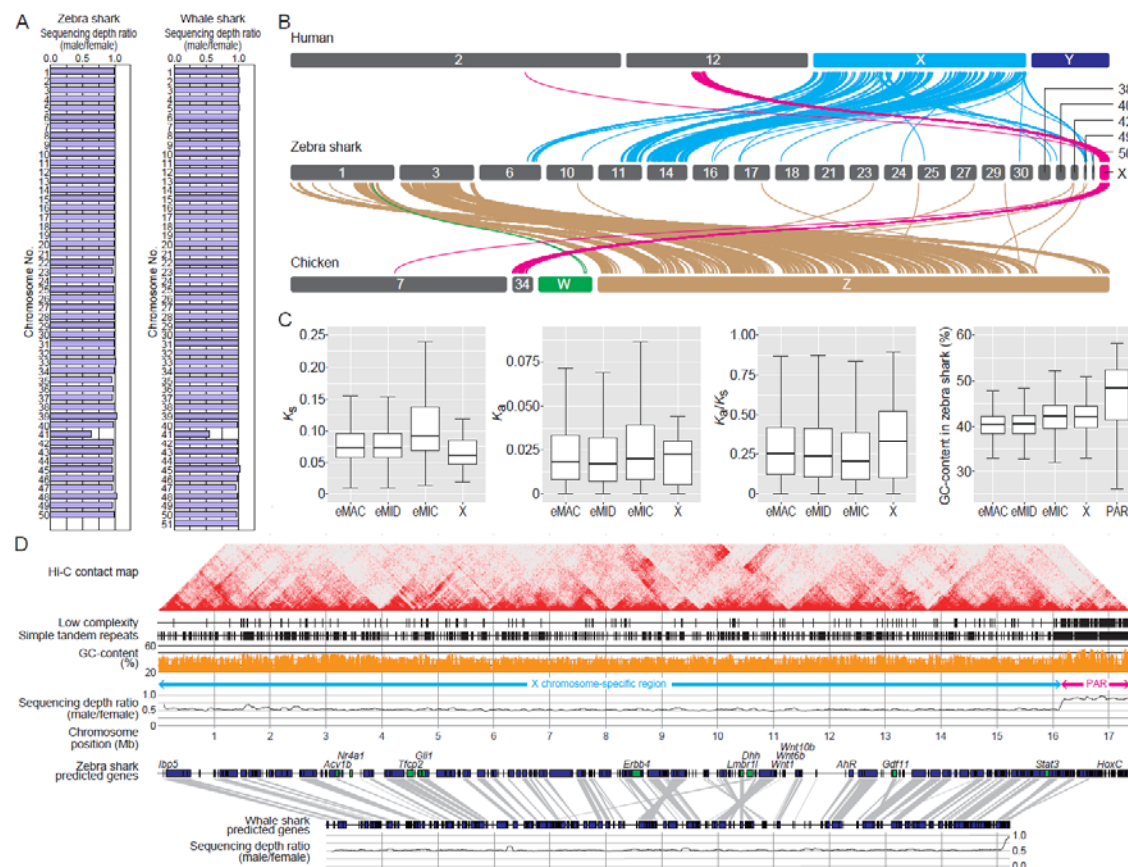
Figure 5. Genomic identification of the zebra shark X chromosome. (*A*) Male–female ratio of short-read sequencing depth in the shark chromosomes. (*B*) Cross-species synteny of sex chromosome-linked genes based on 1-to-1 orthologs. (*C*) Characteristic comparisons between the different chromosome categories 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. (*D*) Structural comparison between the zebra shark and whale shark X chromosome sequences.

Identification of Hox C cluster on the putative X chromosome

418   In the newly obtained sequence of the putative zebra shark X chromosome, we

419   identified an array of orthologs of the non-shark genes encoding homeobox proteins

420   Hox C (Figure 5D). In elasmobranchs, Hox C genes were long thought missing from

421   the genome (King et al. 2011) but later identified in several shark species, as rogue

422   open reading frames (ORFs) flanked by massively repetitive sequences (Hara et al.

423   2018; Kuraku 2021). Our genome-wide gene prediction for the zebra shark detected

424   the ORF of *Hoxc8*, *-c11*, and *-c12*, while the partial ORF of the putative *Hoxc6*

425   ortholog was also identified by a manual search of the raw genomic sequence

426   (Figure 6A). These Hox C genes were located in a 180 Kb-long genomic segment in

427   the PAR of the putative X chromosome (Figure 6A), identified as a single cluster in

428   an elasmobranch fish for the first time. Their orthologies were confirmed with

429   molecular phylogenetic trees, which also indicated elevated molecular evolutionary

430   rates with long branches for elasmobranch Hox C genes (Figure 6B). Our RNA-seq

431   data showed the transcription of these Hox C genes (except for *Hoxc12*) in embryos

432   and juvenile tissues (Figure 6C). The identified zebra shark Hox C cluster is

433   massively invaded by repetitive elements unlike the other Hox gene clusters (A, B,

434   and D clusters) of this and many other vertebrate species (Figure 6A), as in the Hox

435   C-containing genomic segments of other shark species (Hara et al. 2018). Our

436   search for zebra shark orthologs of the protein-coding genes located near the human

437   Hox C cluster (e.g., ATF7, CBX5) revealed poor conservation of the gene

438   compositions. Some of the zebra shark orthologs were not identified in its entire

439   genome sequence, suggesting a divergent nature of the genomic regions flanking
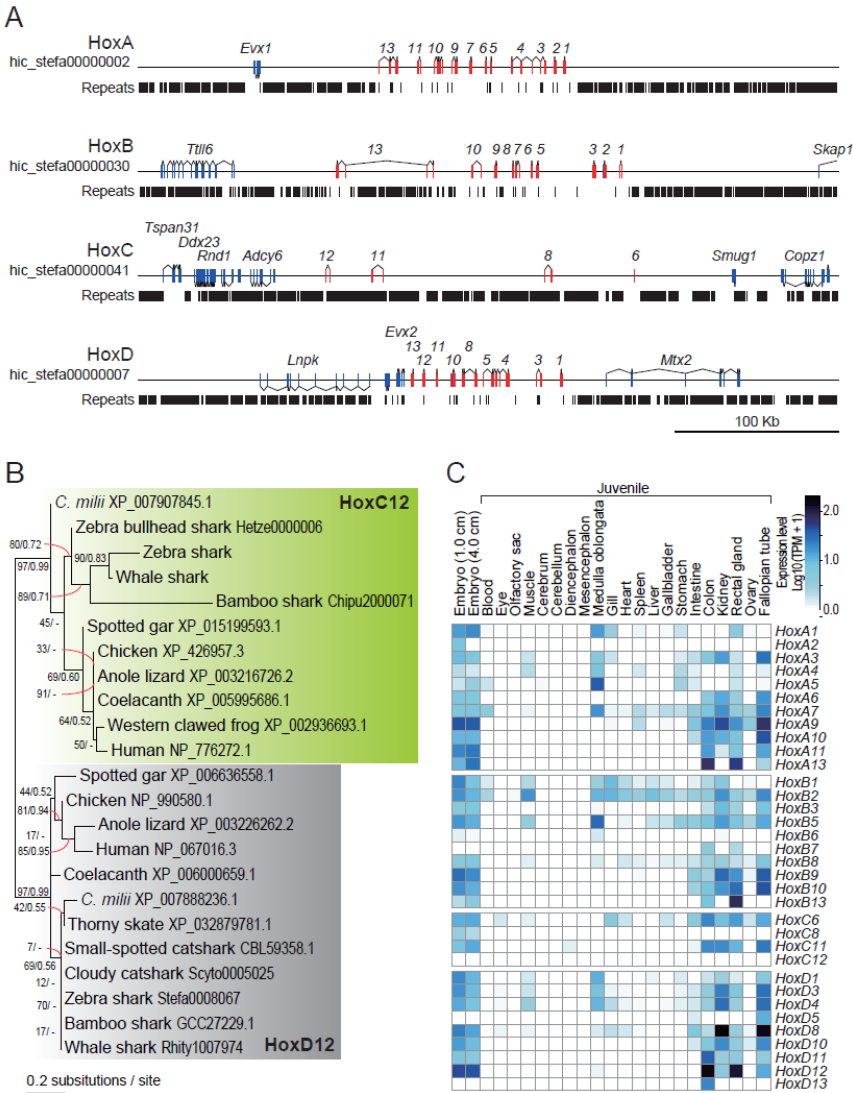
440   the Hox C cluster.

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.

Discussion

450 In this study, we chose two orectolobiform shark species (zebra shark and whale

451 shark) in Elasmobranchii and characterized their genomic organization with

452 chromosome-scale DNA sequences. This study was achieved in support of

453 epigenome and transcriptome data prepared using fresh tissue samples and

454 previously obtained karyotype information. Our results suggest that their karyotypes

455 are organized by chromosomes of gradual sizes marked with size-dependent

456 sequence properties (Figure 2). The pattern in the shark chromosomal organization

457 is unique in its diversity among vertebrates (Uno et al. 2020), which is characterized

458 by abundant chromosomes (up to 106 for diploids) and variable chromosome sizes.

459 The abundance and highly variable sizes of chromosomes are known for some avian

460 species, but the shark genome organization is distinct from avian counterparts in that

461 shark chromosomes generally have higher repeat content than those of the chicken

462 (Figure 2). In our comparison, the shark karyotype is also characterized by the ratio

463 of the largest and smallest chromosome lengths of 40 to 100, compared with < 10 for

464 most vertebrates, except for species with microchromosomes (Figure 1A). In fact,

465 the length of the shortest chromosomal sequence in typical genome assemblies

466 deposited currently in public databases is often dependent on sequence length cutoff

467 by researchers' decision. Especially for species with no solid karyotypic reference

468 such as *Callorhinchus milii*, the range of sequences considered as chromosomes

469 needs to be carefully examined.

470   Our chromosome-scale genome sequencing and analysis were enabled by

471 access to fresh tissue samples. Because of low accessibility, no previous efforts

472 could provide a set of DNA sequences, karyotypic configuration, and reliable

473 measure of nuclear DNA content for a single chondrichthyan species. Of these, the

474 two latter elements serve as indispensable references to validate the output of

23

475   sequencing. These requirements are satisfied for both zebra shark and whale shark

476   in our study. Especially for the whale shark, no published studies employed

477   chromatin contact data for Hi-C scaffolding and transcriptome sequencing (Tan et al.

478   2021; Marra et al. 2019; Weber et al. 2020) . In our study, the access to embryos

479   and blood not only enabled chromosome-scale genome scaffolding but also provided

480   transcriptional evidence of most shark Hox C genes that have been shown to exist in

481   a cluster for the first time (Figure 6).

482   Peculiar fractions of vertebrate karyotypes with small sizes and higher GC-

483   content have traditionally been designated as microchromosomes, which usually

484   denote chromosomes shorter than 20 Mb (e.g., Nakatani et al. 2021) but have no

485   uniform definition (see Introduction). Our investigation, focusing on various aspects

486   of chromosomal DNA sequences, provided a novel view of vertebrate karyotypes

487   that cannot be understood with a simple binary classification, namely with or without

488   microchromosomes, or between macro- and microchromosomes. This view is

489   supported by the common pattern of high intrachromosomal heterogeneity within

490   individual macrochromosomes (Figure 4; Supplementary Figure 4) as well as

491   interchromosomal heterogeneity among different microchromosomes (Figure 2;

492   Supplementary Figures 1 and 2). In particular, the heterogeneity within

493   macrochromosomes, marked with high GC-content, high gene density, and small

494   gene length of their ends, may be shared widely among diverse vertebrates (Figure

495   4A, B). Intragenomic heterogeneity of GC-content was previously suggested to be

496   caused by GC-biased gene conversion (Mugal et al. 2015). In addition, the uniform

497   numbers of recombination per chromosome (Dumont and Payseur 2008) have been

498   thought to explain the chromosome length-dependent GC-content variation between

499   different chromosomes. Our observations did not show pronounced length-

24

500   dependent variation of GC-content for chromosomes that were longer than 100 Mb

501   (Figure 2B; Supplementary Figure 2). Taken together, we speculate that the peculiar

502   nature of chromosome ends mainly account for the variation of GC-content among

503   different chromosomes (Figure 4B). Importantly, 'chromosome ends' in this context

504   not only harbor telomeric or other simple repeats but also hold complex sequences

505   including protein-coding genes in sequence stretches that are longer than 1 Mb

506   (Figure 4B; Supplementary Figure 4). In the *X. tropicalis* genome, such sequence

507   stretches marked with elevated GC-content span much longer ranges than in other

508   genomes (Supplementary Figure 1 and 2). The observed features of smaller

509   chromosomes with larger proportions of such 'ends' in length are more affected than

510   those of longer chromosomes, which likely explains the length-dependent nature of

511   chromosomes. The nature of macrochromosome ends (e.g., with higher GC -

512   content) is thought to be a remnant of the fusion of one or more

513   microchromosome(s) to a macrochromosome (Waters et al. 2021). This hypothesis

514   is not supported by our observation that even species possessing no explicit

515   microchromosomes (e.g., *Xenopus tropicalis*) have chromosome ends with the

516   peculiar nature of DNA sequences (SupplementaySupplementary Figures 1, 4).

517   Remarkably no close relatives of *X. tropicalis* have been shown to possess

518   microchromosomes),(Tymowska 1991), and thus microchromosome fusions cannot

519   account for the characteristics of their chromosome ends with higher GC-content.

520   Our genome-wide sequencing depth investigation covering both sexes

521   revealed the XY system for the two studied shark species and enabled the first

522   sequence-based identification of shark sex chromosomes (namely, X chromosomes;

523   Figure 5A). The genes on the shark X chromosome tend to exhibit larger non-

524   synonymous substitution rates (K_a) and smaller synonymous substitution rates (K_s)

525   than those on autosomes, resulting in a higher K_a/K_s ratio (Figure 5D). This

526   resembles the pattern known in other species including birds and insects, which is

527   known as the 'Faster X (Faster Z) hypothesis' (Meisel and Connallon 2013;

528   Charlesworth et al. 2018; Xu et al. 2019; Mank et al. 2007). The smaller K_s value is

529   also indicative of the support for the male-driven hypothesis (Li 2002; Miyata et al.

530   1987), which is also to be examined by higher K_s values for genes on the

531   presumptive Y chromosome even though it remains unidentified. Our comparison of

532   protein-coding gene compositions showed that the zebra shark and the whale shark

533   largely share the X chromosome that is homologous to each other (Figure 5D). The

534   X chromosome harbors the Hox C cluster that was previously shown to be highly

535   divergent and degenerative, but its localization in the PAR (Figure 5D) suggests a

536   balanced dosage between males and females.

537   The whale shark is known as the largest extant 'fish', and one of its extant

538   closest relatives is the zebra shark (Naylor et al. 2012). In elasmobranch evolution,

539   the lineages leading to these two species diverged no earlier than 48.6 million years

540   ago (Long 1992). The remarkably high similarity of the chromosome-scale sequence

541   organization (panel 1 in Figure 3; Supplementary Figure 3) as well as the gene

542   compositions on the X chromosome between these two species (Figure 5D)

543   indicates a lower rate of chromosomal rearrangement in these lineages compared

544   with those of species pairs of similar divergence times in other vertebrate lineages

545   (Figure 3A). Among vertebrates, mammals and birds have relatively long-standing

546   sex chromosomes (X/Y and Z/W, respectively) shared throughout these individual

547   taxa that arose more than 48.6 million years ago (Long 1992). Although still limited in

548   number, sex chromosomes have been identified in some elasmobranch species by

549   cytogenetic analyses, all of which have a male-heterogametic system (Uno et al.

550   2020), as shown for the zebra shark and whale shark with genome sequencing in

551   this study. Sharks, or a phylogenetically wider subset of cartilaginous fishes, may

552   possess even older sex chromosomes, depending on phylogenetic prevalence of

553   their homologs in more distantly related shark and even ray species.

554   Our synteny analysis yielded novel insights into genome evolution

555   encompassing the whole diversity of vertebrates. It showed homology of the shark X

556   chromosome to human chromosome 12 and chicken chromosome 34 (Figure 5B),

557   which suggests an independent origin of the sex chromosome in elasmobranchs.

558   The synteny analysis involving sharks also provided clues to the origin of

559   microchromosomes. It suggests that eMIC, the shark's small chromosomes, are

560   homologous to several of the large chromosomes in both chicken and spotted gar

561   and are not necessarily homologous to their microchromosomes (Figure 3B). Also,

562   microchromosomes of the chicken and gar were not shown to be homologous to

563   eMICs (Figure 3B). It is crucial for any effort to reconstruct the diversity of

564   chromosome organization in vertebrates to incorporate elasmobranchs into the

565   comparisons.

566

567   **Methods**

568   **Animals**

569   Fresh blood from a female adult zebra shark (total length, 2.2 m); Individual ID,

570   sSteFas1 and a male adult whale shark (total length, 8.8 m; Individual ID, sRhiTyp1)

571   were sampled at Okinawa Churaumi Aquarium and used for the preparation of

572   whole-genome shotgun DNA libraries, Hi-C libraries, and RNA-seq libraries as well

573   as for measuring nuclear DNA content by flow cytometry. Likewise, fresh blood of a

574   male zebra shark (total length, 2.1 m; Individual ID, sSteFas2) and a female whale

575   shark (total length, 8.0 m; Individual ID, sRhiTyp2) were sampled and used for

576   quantifying the male/female ratios of individual chromosomal regions. Extraction of

577   ultra-high molecular weight DNA was performed by collecting blood cells by

578   centrifugation, and the collected cells were embedded in agarose plugs (4.0×10^5

579   cells/plug). The agarose gel plugs were prepared and processed with the CHEF

580   Mammalian Genomic DNA Plug Kit (BioRad, Cat. No. #1703591). Total RNAs used

581   to construct RNA-seq libraries were extracted from various tissues of a female

582   juvenile zebra shark (total length, 30 cm) born at Okinawa Churaumi Aquarium and a

583   female juvenile whale shark (total length, 7.7 m; Individual ID, sRhiTyp3)

584   (Supplementary Table 3). These animals were introduced into the aquarium in

585   accordance with local regulations before those species were assessed as

586   endangered. Animal handling and sample collections at the aquarium were

587   conducted by veterinary staff without restraining the individuals under the experiment

588   ID AT19002 approved by Institutional Animal Care and Use Committee of Okinawa

589   Churashima Foundation in accordance with the Husbandry Guidelines approved by

590   the Ethics and Welfare Committee of Japanese Association of Zoos and Aquariums.

591   All other experiments were conducted in accordance with the Guideline of the

592   Institutional Animal Care and Use Committee (IACUC) of RIKEN Kobe Branch

593   (Approval ID: H16-11).

594

595   **Genome sequencing and scaffolding**

596   For a female zebra shark, paired-end and mate-pair DNA libraries for *de novo*

597   genome sequencing were prepared and sequenced as previously described (Hara et

598   al. 2018; Yamaguchi et al. 2021). The amount of starting DNA and numbers of PCR

599   cycles for the library preparation are included in Supplementary Table 3. The total

28

600 sequencing coverage amounted to 95.8 times the genome size based on the

601 reference measured previously by flow cytometry (3.71 Gb; Kadota et al., in

602 preparation). Low-quality bases from paired-end reads were removed by TrimGalore

603 v0.6.6 with the options '--stringency 2 --quality 20 --length 25 --paired --

604 retain_unpaired'. As described previously (Hara et al. 2018), short-read assembly of

605 the zebra shark, as well as scaffolding with mate-pair reads followed by gap closure,

606 was performed using Platanus v1.2.4 (Kajitani et al. 2014).

607   Whole genome sequencing for a male whale shark employed the 10x

608 Genomics Chromium to produce Linked-Read data. A DNA library was prepared

609 using 12 ng of gDNA extracted from blood cells according to the user guide of the

610 Chromium Genome Library Kit v2 Chemistry using the Chromium Genome Library

611 Kit & Gel Bead Kit v2 (10x Genomics, Cat. No. #120258) and the Chromium

612 Genome Chip Kit v2 (10x Genomics, Cat. No. #120257). The library was sequenced

613 on a HiSeq X (Illumina) platform to obtain 151 nt-long paired-end reads. Sequence

614 assembly using the Linked-Read data of 46.4 times the genome size was performed

615 with the program Supernova v2.0. The resultant sequences were subjected to

616 scaffolding with the program P_RNA_scaffolder (commit 7941e0f in GitHub) (Zhu et

617 al. 2018) using the result of the alignment of transcriptome sequence reads

618 (obtained as described below) performed with the program HISAT2 v2.1.0 onto

619 those genome sequences (Kim et al. 2019).

620

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

622 Hi-C libraries of the zebra shark and whale shark were constructed using restriction

623 enzymes DpnII and HindIII, respectively, as previously reported (Kadota et al. 2020).

624 Blood cells collected as described above were fixed in 1% formaldehyde solution. A

625 fixed tissue containing 10 µg of DNA was used for the preparation of Hi-C DNA via *in*

626 *situ* restriction digestion and ligation. The Hi-C library was prepared using 2 µg of the

627 ligated DNA with five cycles of PCR amplification. Quality controls of the ligated DNA

628 and the Hi-C libraries were performed as described previously (Kadota et al. 2020).

629   Each of the zebra shark and the whale shark genome assemblies was used

630 for Hi-C read mapping with Juicer v1.5 and chromosome-scale scaffolding with the

631 program 3d-dna (v180922) (Dudchenko et al. 2017). In the scaffolding, three

632 different lengths were tested (5, 10, and 15 kb) for the option '-i' defining the input

633 sequence length threshold. For each species, the three resulting scaffolding outputs,

634 as well as the original assembly before Hi-C scaffolding, were assessed based on

635 sequence length distribution and protein-coding gene completeness. Among all the

636 scaffolding outputs compared, the output with the option '-i 10000' was judged to be

637 optimal and was subjected to a "review" of the scaffolding results on Juicebox

638 v1.11.08 (Durand et al. 2016) to minimize inconsistent signals of chromatin contacts

639 (Supplementary Figure 5). The review was facilitated by referring to nucleotide

640 sequence-level similarity between different scaffolding outputs visualized by SyMAP

641 v5.0 (Soderlund et al. 2011). After the review, the sequences judged as

642 contaminants from other organisms were removed, as previously reported (Hara et

643 al. 2018). The resulting sequences were deposited as JAHMAH000000000 and

644 JAFIRC000000000 in DDBJ/NCBI under BioProject PRJNA703743.

645

646 **Repeat identification**

647 To obtain a species-specific repeat library, RepeatModeler v2.0.2a was run on the

648 genome assembly of the individual species with default parameters (Smit and

649 Hubley 2008). Detection of repeat elements in the genome was performed by

650 RepeatMasker v4.1.2-p1 (Smit et al. 2013) with RMBlast v2.6.0+, using the species-

651 specific repeat library obtained above. For quantification of the content of

652 interspersed repeats and simple tandem repeats, RepeatMasker was run separately

653 with the options '-nolow -norna' and '-noint -norna', respectively.

654

655 **Gene model construction**

656 The program Braker v2.1.6 was employed for gene prediction by inputting the results

657 of RNA-seq read mapping to a genome assembly in which repetitive sequences are

658 soft-masked by RepeatMasker with the options '-nolow -xsmall', as well as amino

659 acid sequences of closely related species as homolog hints (Brůna et al. 2021; Smit

660 et al. 2013). To build the homolog hints based on the amino acid sequences, we

661 used the previously reported amino acid sequence sets of the brownbanded bamboo

662 shark and the whale shark.

663 Gene space completeness in Figure 1C was obtained by the BUSCO pipeline

664 ver. 5 (Seppey et al. 2019) using the BUSCO's Vertebrata ortholog set.

665

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

667 In order to calculate the number of synonymous substitutions per synonymous site

668 (K_s) and the number of nonsynonymous substitutions per non-synonymous site (K_a),

669 the 1-to-1 orthologs shared by the four elasmobranch species (zebra shark, whale

670 shark, brownbanded bamboo shark, and cloudy catshark) were selected by

671 SonicParanoid v1.3.4 (Cosentino and Iwasaki 2019) as follows. First, peptide

672 sequences of the retrieved orthologs were aligned with MAFFT v7.475 with the

673 option '-linsi' (Katoh and Standley 2013). The individual alignments were trimmed

674 and back-translated into nucleotides with trimal v1.4.rev15 with the options '-

31

675   automated1 -backtrans' followed by the removal of gapped sites using trimAl with the

676   option '-nogaps' (Capella-Gutiérrez et al. 2009). Ortholog groups containing fewer

677   than 100 aligned codons or a stop codon were discarded. For the selected ortholog

678   groups, K_s and K_a were computed with yn00 in the PAML v4.9c88 (Yang 2007).

679   Computed values larger than 0.01 and smaller than 99 were included in the results

680   (Figures 2B, 4B, and 5C).

681

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

683   Total RNAs were extracted with Trizol reagent (Thermo Fisher Scientific). Quality

684   control of the RNA treated with DNase I was performed with Bioanalyzer 2100

685   (Agilent Technologies). Libraries were prepared with TruSeq RNA Sample Prep Kit

686   (Illumina) or TruSeq Stranded mRNA LT Sample Prep Kit (Illumina) as previously

687   described (Hara et al. 2018). The amount of starting total RNA and numbers of PCR

688   cycles are included in Supplementary Table 3. To remove adaptor sequences and

689   low-quality bases, the obtained sequence reads were trimmed with TrimGalore

690   v0.6.6 as outlined above, and *de novo* transcriptome assembly was performed with

691   the program Trinity v2.11.0 with the option '--SS_lib_type RF' (Grabherr et al. 2011).

692   The trimmed RNA-seq reads were aligned to the genome assembly using the

693   program HISAT2 v2.1.1 (Kim et al. 2019), which was followed by gene expression

694   quantification with StringTie v2.0.6 (Pertea et al. 2016).

695

696   **Conserved synteny detection**

697   Characterization of chromosomal homology among different species employed

698   predicted protein sequence datasets available at the NCBI RefSeq database. After

699   alternative splicing variants were removed, one-to-one orthologs were selected by

32

700   SonicParanoid v1.3.4 with the option 'most-sensitive' (Cosentino and Iwasaki 2019).

701   Conserved synteny between species was visualized based on single copy 1-to-1

702   orthologs using RIdeogram (Hao et al. 2020).

703

704   **Molecular phylogeny inference**

705   Amino acid sequences were retrieved from aLeaves (Kuraku et al. 2013). Multiple

706   sequence alignment was performed with MAFFT with the option '-linsi'. The aligned

707   sequence sets were processed using trimAl v1.4 rev15 with the option '-automated1'

708   (Capella-Gutiérrez et al. 2009). This was followed by another trimAl run with the

709   option '-nogaps'. Molecular phylogenetic trees were inferred by RAxML with the '-m

710   PROTCATWAG -f a -# 1000' options unless stated otherwise (Stamatakis 2014).

711   Tree inference in the Bayesian framework was performed with the program

712   PhyloBayes v4.1c with the options '-cat -dgam 4 -wag -nchain 2 1000 0.3 50' unless

713   stated otherwise. This was followed by an execution of bpcomp in the PhyloBayes

714   v4.1c package with the option '-x 100' (Lartillot et al. 2009). The support values at

715   the nodes of molecular phylogenetic trees included are, in order, bootstrap values

716   and Bayesian posterior probabilities. The latter was shown only when the

717   relationship at the node in the visualized tree was supported by the Bayesian

718   inference.

719

720   **Identification of the X chromosome**

721   To identify a chromosome-scale scaffold with a distinct male–female sequencing

722   depth ratio in the zebra shark, the same number of trimmed genomic shotgun reads

723   (293,686,584) was prepared for both sexes. The reads were mapped with BWA-

724   mem2 (v2.2.1) onto a genome assembly (Li and Durbin 2009). Mapped reads were

725   counted for male and female using the bamtobed program in the package bedtools

726   v2.29.2, for each scaffold in 10 Kb non-overlapping windows, and male–female ratios

727   were calculated (Quinlan and Hall 2010). Sequencing depth of male and female

728   reads on the identified X chromosome was also calculated for 10 Kb non-overlapping

729   windows. Windows with the proportion of ambiguous bases of more than 50% were

730   excluded from computation. This procedure was also applied to the whale shark,

731   using 220,785,436 trimmed reads.

732

733   **Data access**

734   The sequence data resulting from the present study were deposited as

735   JAHMAH000000000 and JAFIRC000000000 in DDBJ/NCBI under BioProject

736   PRJNA703743.

737

738   **Competing interests**

739   The authors declare no conflict of interests.

740

750

References

751 Bertho S, Herpin A, Schartl M, Guiguen Y. 2021. Lessons from an unusual vertebrate sex-

753   determining gene. *Philos Trans R Soc B: Biol Sci* **376**: 20200092.

754 Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, Amores A,

755   Desvignes T, Batzel P, Catchen J, et al. 2016. The spotted gar genome illuminates

756   vertebrate evolution and facilitates human-teleost comparisons. *Nat Genet* **48**: 427–437.

757 Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic

758   eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a

759   protein database. *NAR Genom Bioinform* **3**: 1–11.

760 Burt DW. 2002. Origin and evolution of avian microchromosomes. *Cytogenet Genome Res*

761   **96**: 97–112.

762 Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient

763   and simple way. *PeerJ* **6**: e4958.

764 Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated

765   alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972-1973.

766 Charlesworth B, Campos JL, Jackson BC. 2018. Faster-X evolution: theory and evidence

767   from *Drosophila*. *Mol Ecol* **27**: 3753–3771.

768 Cosentino S, Iwasaki W. 2019. SonicParanoid: fast, accurate and easy orthology inference.

769   *Bioinformatics* **35**: 149–151.

770 Deakin JE, Potter S, O'Neill R, Ruiz-Herrera A, Cioffi MB, Eldridge MDB, Fukui K, Marshall

771   Graves JA, Griffin D, Grutzner F, et al. 2019. Chromosomics: bridging the gap between

772   genomes and chromosomes. *Genes (Basel)* **10**: 627.

773   Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS,

774   Machol I, Lander ES, Aiden AP, et al. 2017. *De novo* assembly of the *Aedes aegypti*

775   genome using Hi-C yields chromosome-length scaffolds. *Science* **356**: 92–95.

776   Dumont BL, Payseur BA. 2008. Evolution of the genomic rate of recombination in mammals.

777   *Evolution* **62**: 276–294.

778   Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. 2016.

779   Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom.

780   *Cell Syst* **3**: 99–101.

781   Franchini P, Jones JC, Xiong P, Kneitz S, Gompert Z, Warren WC, Walter RB, Meyer A,

782   Schartl M. 2018. Long-term experimental hybridisation results in the evolution of a new

783   sex chromosome in swordtail fish. *Nat Commun* **9**: 5136.

784   Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,

785   Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-

786   Seq data without a reference genome. *Nat Biotechnol* **29**: 644–652.

787   Graves JAM. 2016. Evolution of vertebrate sex chromosomes and dosage compensation.

788   *Nat Rev Genet* **17**: 33–46.

789   Hao Z, Lv D, Ge Y, Shi J, Weijers D, Yu G, Chen J. 2020. *RIdeogram*: drawing SVG

790   graphics to visualize and map genome-wide data on the idiograms. *PeerJ Comput Sci*

791   **6**: e251.

792   Hara Y, Yamaguchi K, Onimaru K, Kadota M, Koyanagi M, Keeley SD, Tatsumi K, Tanaka K,

793   Motone F, Kageyama Y, et al. 2018. Shark genomes provide insights into

794   elasmobranch evolution and the origin of vertebrates. *Nat Ecol Evol* **2**: 1761-1771.

795   Hardie DC, Hebert PDN. 2004. Genome-size evolution in fishes. *Can J Fish Aquat Sci* **61**:

796   1636–1646.

797   International Chicken Genome Sequencing Consortium. 2004. Sequence and comparative

798   analysis of the chicken genome provide unique perspectives on vertebrate evolution.

799   *Nature* **432**: 695–716.

800   Kadota M, Nishimura O, Miura H, Tanaka K, Hiratani I, Kuraku S. 2020. Multifaceted Hi-C

801   benchmarking: what makes a difference in chromosome-scale genome scaffolding?

802   *Gigascience* **9**: giz158.

803   Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada M,

804   Nagayasu E, Maruyama H, et al. 2014. Efficient *de novo* assembly of highly

805   heterozygous genomes from whole-genome shotgun short reads. *Genome Res* **24**:

806   1384–1395.

807   Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:

808   improvements in performance and usability. *Mol Biol Evol* **30**: 772–780.

809   Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment

810   and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**: 907–915.

811   King BL, Gillis JA, Carlisle HR, Dahn RD. 2011. A natural deletion of the HoxC cluster in

812   elasmobranch fishes. *Science* **334**: 1517.

813   Knief U, Forstmeier W. 2016. Mapping centromeres of microchromosomes in the zebra finch

814   (*Taeniopygia guttata*) using half-tetrad analysis. *Chromosoma* **125**: 757-768.

815   Koochekian N, Ascanio A, Farleigh K, Card DC, Schield DR, Castoe TA, Jezkova T. 2022. A

816   chromosome-level genome assembly and annotation of the desert horned lizard,

817   *Phrynosoma platyrhinos*, provides insight into chromosomal rearrangements among

818   reptiles. *Gigascience* **11**: giab098.

819   Kuraku S. 2021. Shark and ray genomics for disentangling their morphological diversity and

820   vertebrate evolution. *Dev Biol* **477**: 262–272.

821   Kuraku S, Ishijima J, Nishida-Umehara C, Agata K, Kuratani S, Matsuda Y. 2006. cDNA-

822   based gene mapping and GC3 profiling in the soft-shelled turtle suggest a

823   chromosomal size-dependent GC bias shared by sauropsids. *Chromosome Res* **14**:

824   187–202.

825   Kuraku S, Zmasek CM, Nishimura O, Katoh K. 2013. aLeaves facilitates on-demand

826   exploration of metazoan gene family trees on MAFFT sequence alignment server with

827   enhanced interactivity. *Nucleic Acids Res* **41**: 22–28.

828   Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a bayesian software package for

829   phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**: 2286–2288.

830   Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler

831   transform. *Bioinformatics* **25**: 1754–1760.

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

833   Long DJ. 1992. Sharks from the La Meseta Formation (Eocene), Seymour Island, Antarctic

834   Peninsula. *J Vertebr Paleontol* **12**: 11-32.

835   Ma W, Xu L, Hua H, Chen M, Guo M, He K, Zhao J, Li F. 2021. Chromosomal-level

836   genomes of three rice planthoppers provide new insights into sex chromosome

837   evolution. *Mol Ecol Resour* **21**: 226–237.

838   Mank JE, Axelsson E, Ellegren H. 2007. Fast-X on the Z: rapid evolution of sex-linked genes

839   in birds. *Genome Res* **17**: 618–624.

840   Marra NJ, Stanhope MJ, Jue NK, Wang M, Sun Q, Bitar PP, Richards VP, Komissarov A,

841   Rayko M, Kliver S, et al. 2019. White shark genome reveals ancient elasmobranch

842   adaptations associated with wound healing and the maintenance of genome stability.

843   *Proc Natl Acad Sci USA* **116**: 4446–4455.

844   Matsubara K, Kuraku S, Tarui H, Nishimura O, Nishida C, Agata K, Kumazawa Y, Matsuda

845   Y. 2012. Intra-genomic GC heterogeneity in sauropsids: evolutionary insights from

846   cDNA mapping and GC3 profiling in snake. *BMC Genomics* **13**: 1-14.

847   Meisel RP, Connallon T. 2013. The faster-X effect: integrating theory and data. *Trends

848   Genet* **29**: 537–544.

849   Miyata T, Hayashida H, Kuma K, Mitsuyasu K, Yasunaga T. 1987. Male-driven molecular

850   evolution: a model and nucleotide sequence analysis. *Cold Spring Harb Symp Quant

851   Biol* **52**: 863–867.

852   Mugal CF, Weber CC, Ellegren H. 2015. GC-biased gene conversion links the recombination

853   landscape and demography to genomic base composition. *BioEssays* **37**: 1317–1326.

854   Mulley JF, Zhong Y-F, Holland PW. 2009. Comparative genomics of chondrichthyan Hoxa

855   clusters. *BMC Evol Biol* **9**: 218.

856   Nakatani Y, Shingate P, Ravi V, Pillai NE, Prasad A, McLysaght A, Venkatesh B. 2021.

857   Reconstruction of proto-vertebrate, protocyclostome and proto-gnathostome genomes

858   provides new insights into early vertebrate evolution. *Nat Commun* **12**: 4489.

859   Nakatani Y, Takeda H, Kohara Y, Morishita S. 2007. Reconstruction of the vertebrate

860   ancestral genome reveals dynamic genome reorganization in early vertebrates.

861   *Genome Res* **17**: 1254–1265.

862   Naylor G, Caira K, Jensen K, White W, Last P. 2012. A DNA sequence–based approach to

863   the identification of shark and ray species and its implications for global elasmobranch

864   diversity and parasitology *Bull Am Mus Nat Hist* **367**: 1-263.

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

866   Ohno S, Muramoto J, Stenius C, Christian L, Kittrell WA, Atkin NB. 1969.

867   Microchromosomes in holocephalian, chondrostean and holostean fishes.

868   *Chromosoma* **26**: 35–40.

869   Palmer DH, Rogers TF, Dean R, Wright AE. 2019. How to identify sex chromosomes and

870   their turnover. *Mol Ecol* **28**: 4709–4724.

871   Pennell MW, Mank JE, Peichel CL. 2018. Transitions in sex determination and sex

872   chromosomes across vertebrate species. *Mol Ecol* **27**: 3950–3963.

873   Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression

874   analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* **11**:

875   1650–1667.

876   Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic

877   features. *Bioinformatics* **26**: 841–842.

878   Read TD, Petit RA, Joseph SJ, Alam T, Weil MR, Ahmad M, Bhimani R, Vuong JS, Haase

879   CP, Webb DH, et al. 2017. Draft sequencing and assembly of the genome of the

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

881   755.

882   Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W,

883   Fungtammasan A, Kim J. 2021. Towards complete and error-free genome assemblies

884   of all vertebrate species. *Nature* **592**: 737–746.

885   Sacerdot C, Louis A, Bon C, Berthelot C, Roest Crollius H. 2018. Chromosome evolution at

886   the origin of the ancestral vertebrate genome. *Genome Biol* **19**: 166.

887   Schwartz FJ, Maddock MB. 1986. Comparisons of karyotypes and cellular DNA contents

888   within and between major lines of elasmobranchs. In *Indo-Pacific fish biology*, pp. 148–

889   157. Ichthyological Soc. of Japan.

890   Schwartz FJ, Maddock MB. 2002. Cytogenetics of the elasmobranchs: genome evolution

891   and phylogenetic implications. *Mar Freshw Res* **53**: 491–502.

892   Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and

893   annotation completeness. In *Gene Prediction Methods in Molecular Biology* (ed.

894   Kollmar M), pp. 227–245. Humana Press, New York.

895   Smeds L, Kawakami T, Burri R, Bolivar P, Husby A, Qvarnström A, Uebbing S, Ellegren H.

896   2014. Genomic identification and characterization of the pseudoautosomal region in

897   highly differentiated avian sex chromosomes. *Nat Commun* **5**: 5448.

898   Smit AFA, Hubley R. 2008. RepeatModeler Open-1.0 http://www.repeatmasker.org.

899   Smit AFA, Hubley R, Green P. 2013. RepeatMasker Open-4.0. http://www.repeatmasker.org.

900   Soderlund C, Bomhoff M, Nelson WM. 2011. SyMAP v3.4: A turnkey synteny system with

901   application to plant genomes. *Nucleic Acids Res* **39**: e68–e68.

41

902   Srikulnath K, Ahmad SF, Singchat W, Panthum T. 2021. Why do some vertebrates have

903   microchromosomes? *Cells* **10**: 1–33.

904   Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of

905   large phylogenies. *Bioinformatics* **30**: 1312-1313.

906   Suryamohan K, Krishnankutty SP, Guillory J, Jevit M, Schröder MS, Wu M, Kuriakose B,

907   Mathew OK, Perumal RC, Koludarov I, et al. 2020. The Indian cobra reference genome

908   and transcriptome enables comprehensive identification of venom toxins. *Nat Genet* **52**:

909   106-117.

910   Tan M, Redmond AK, Dooley H, Nozu R, Sato K, Kuraku S, Koren S, Phillippy AM, Dove

911   ADM, Read TD. 2021. The whale shark genome reveals patterns of vertebrate gene

912   family evolution. *Elife* **10**: e65394.

913   Tymowska J. 1991. Polyploidy and cytogenetic variation in frogs of the genus *Xenopus*. In

914   *Amphibian cytogenetics and evolution* (eds. Green DM, Sessions SK), pp. 259–297,

915   Academic Press, Cambridge.

916   Uno Y, Nishida C, Tarui H, Ishishita S, Takagi C, Nishimura O, Ishijima J, Ota H, Kosaka A,

917   Matsubara K, et al. 2012. Inference of the protokaryotypes of amniotes and tetrapods

918   and the evolutionary processes of microchromosomes from comparative gene mapping.

919   *PLoS One* **7**: e53027.

920   Uno Y, Nozu R, Kiyatake I, Higashiguchi N, Sodeyama S, Murakumo K, Sato K, Kuraku S.

921   2020. Cell culture-based karyotyping of orectolobiform sharks for chromosome-scale

922   genome analysis. *Commun Biol* **3**: 652.

923   Venkatesh B, Lee AP, Ravi V, Maurya AK, Lian MM, Swann JB, Ohta Y, Flajnik MF, Sutoh Y,

924   Kasahara M, et al. 2014. Corrigendum: elephant shark genome provides unique

925   insights into gnathostome evolution. *Nature* **513**: 574.

926   Waters PD, Patel HR, Ruiz-Herrera A, Alvarez-Gonzalez L, Lister NC, Simakov O, Ezaz T,

927   Kaur P, Frere C, Grutzner F, et al. 2021. Microchromosomes are building blocks of bird,

928   reptile, and mammal chromosomes. *Proc Natl Acad Sci USA* **118**: e2112494118.

929   Weber JA, Park SG, Luria V, Jeon S, Kim HM, Jeon Y, Bhak Y, Jun JH, Kim SW, Hong WH,

930   et al. 2020. The whale shark genome reveals how genomic and physiological properties

931   scale with body size. *Proc Natl Acad Sci USA* **117**: 20662–20671.

932   Xu L, Wa Sin SY, Grayson P, Edwards S v, Sackton TB. 2019. Evolutionary dynamics of sex

933   chromosomes of paleognathous birds. *Genome Biol Evol* **11**: 2376–2390.

934   Yamaguchi K, Kadota M, Nishimura O, Ohishi Y, Naito Y, Kuraku S. 2021. Technical

935   considerations in Hi-C scaffolding and evaluation of chromosome-scale genome

936   assemblies. *Mol Ecol* **30**: 5923–5934.

937   Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**:

938   1586–1591.

939   Zhang Y, Gao H, Li H, Guo J, Ouyang B, Wang M, Xu Q, Wang J, Lv M, Guo X, et al. 2020.

940   The white-spotted bamboo shark genome reveals chromosome rearrangements and

941   fast-evolving immune genes of cartilaginous fish. *iScience* **23**: 101754.

942   Zhu B-H, Xiao J, Xue W, Xu G-C, Sun M-Y, Li J-T. 2018. P_RNA_scaffolder: a fast and

943   accurate genome scaffolder using paired-end RNA-sequencing reads. *BMC Genomics*

944   **19**: 175.

945

946

947