1 **Dynamic evolutionary history and gene content of sex chromosomes across diverse**

2 **songbirds**

3

4 Luo-hao Xu[1,2], Gabriel Auer[2], Valentina Peona[3], Alexander Suh[3], Yuan Deng[4], Shao-hong

5 Feng[4], Guo-jie Zhang[4,5,6], Mozes P.K. Blom[7], Les Christidis[8,9], Stefan Prost[10,11], Martin Irestedt[7,*],

6 Qi Zhou[1,2,*]

7

8   1. MOE Laboratory of Biosystems Homeostasis & Protection, Life Sciences Institute,

9   Zhejiang University, Hangzhou, China

10   2. Department of Molecular Evolution and Development, University of Vienna, Vienna,

11   Austria

12   3. Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University,

13   Uppsala, Sweden

14   4. China National Genebank, BGI-Shenzhen, Shenzhen, China

15   5. State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology,

16   Chinese Academy of Sciences, Kunming, China

17   6. Section for Ecology and Evolution, Department of Biology, University of Copenhagen,

18   Copenhagen, Denmark

19   7. Department of Bioinformatics and Genetics, Swedish Museum of Natural History,

20   Stockholm, Sweden

21   8. National Marine Science Centre, Southern Cross University, Coffs Harbour, NSW,

22   Australia

23   9. School of BioSciences, University of Melbourne, Parkville, VIC, Australia

24   10. Department of Integrative Biology, University of California, Berkeley, Berkeley, United

25   States

26   11. Department of Biology, Stanford University, Stanford, United States

27   * Correspondence should be addressed to M.I. (Martin.Irestedt@nrm.se) or Q.Z.

28   (zhouqi1982@zju.edu.cn)

Abstract

Songbirds have a species number almost equivalent to that of mammals, and are classic models for studying mechanisms of speciation and sexual selection. Sex chromosomes are hotspots of both processes, yet their evolutionary history in songbirds remains unclear. To elucidate that, we characterize female genomes of 11 songbird species having ZW sex chromosomes, with 5 genomes of bird-of-paradise species newly produced in this work. We conclude that songbird sex chromosomes have undergone at least four steps of recombination suppression before their species radiation, producing a gradient pattern of pairwise sequence divergence termed 'evolutionary strata'. Interestingly, the latest stratum probably emerged due to a songbird-specific burst of retrotransposon CR1-E1 elements at its boundary, or chromosome inversion on the W chromosome. The formation of evolutionary strata has reshaped the genomic architecture of both sex chromosomes. We find stepwise variations of Z-linked inversions, repeat and GC contents, as well as W-linked gene loss rate that are associated with the age of strata. Over 30 W-linked genes have been preserved for their essential functions, indicated by their higher and broader expression of orthologs in lizard than those of other sex-linked genes. We also find a different degree of accelerated evolution of Z-linked genes vs. autosomal genes among different species, potentially reflecting their diversified intensity of sexual selection. Our results uncover the dynamic evolutionary history of songbird sex chromosomes, and provide novel insights into the mechanisms of recombination suppression.

Introduction

49

50   Songbirds (Oscines, suborder Passeri) have over 5000 species and comprise the majority of

51   passerines and nearly half of the all extant bird species[1]. This is a result of the largest avian

52   species radiation occurred about 60 million years (MY) ago[2]. Facilitated by the development of

53   genomics, many species besides the zebra finch (*Taeniopygia guttata*) are now transforming

54   into important models for studying molecular patterns and mechanisms of speciation[3,4],

55   supergenes[5] and cognition[6], out of their long history of ecological or behavioral studies, out of

56   their long history of ecological or behavioral studies. One major reason that has been fueling

57   biologists' fascination with songbirds is their staggering and diversified sexual traits. Many

58   species possess striking plumage forms and colors, sophisticated songs and mating rituals, all

59   of which can undergo rapid turnovers even between sister species. Theories predict that sex

60   chromosomes play a disproportionately large role in speciation (the 'large X/Z' effect), sexual

61   selection and evolution of sexually dimorphic traits[7-9]. However, the evolutionary history of

62   songbird sex chromosome remains unclear, because there were few genomic studies

63   characterizing songbird sex chromosomes except for the Collared Flycatcher (*Ficedula*

64   *albicollis*)[10]. In contrast to the mammalian XY system, birds have independently evolved a pair

65   of female heterogametic sex chromosomes that are usually heteromorphic in females (ZW) and

66   homomorphic in males (ZZ). A recent cytological investigation of over 400 passerine species

67   found a higher fixation rate of chromosome inversions on the Z chromosome than autosomes

68   within species. Gene flow in the Z chromosome is thus more likely reduced in the face of

69   hybridization[11]. Indeed, a significantly lower level of introgression, and a higher level of *F*st in Z-

70   linked genes compared to autosomal genes has been reported from studying pairs of recently

71   diverged songbird species[12-15]. Such a large-Z pattern is probably caused by several factors

72   which act in an opposite manner to the XY sex system. First, Z chromosomes are more often

73   transmitted in males, thus are expected to have a higher mutation rate than the rest of the

74   genome, due to the 'male-driven evolution' effect[16]. Second, as sexual selection more frequently

75   targets males, the variation in male reproductive success will further reduce the effective

76   population size of Z chromosome from three quarters of that of autosomes[17]. The consequential

77   stronger effect of genetic drift is expected to fix excessive slightly deleterious mutations on the Z

78   chromosome, and lead to a faster evolutionary rate than on autosomes (the 'fast-Z' effect)[18].

79   This has been demonstrated in the Galloanserae (e.g., chicken and duck) species, those of

80   which undergo strong sperm competition, i.e., more intensive male sexual selection, exhibit a

81   larger difference between the Z chromosome and autosomes in their evolutionary rates[19].

3

82   In contrast to the avian Z chromosome, or more broadly the mammalian XY

83   chromosomes, the genomic studies of avian W chromosomes, especially those of songbirds

84   have not started only until recently[10,20,21]. This is because most genomic projects prefer to

85   choose the homogametic sex (e.g., male birds or female mammals) for sequencing, in order to

86   avoid the presumably gene-poor and highly repetitive Y or W chromosomes. The Y/W

87   chromosomes have undergone suppression of recombination to prevent the sex-determining

88   gene or sexually antagonistic genes (beneficial to one sex but detrimental to the other) from

89   being transmitted to the opposite sex[22]. As a result, interference between linked loci ('Hill-

90   Robertson' effect) reduces the efficacy of natural selection and drives the ultimate genetic decay

91   of non-recombining regions of Y/W chromosomes[23]. This process can be accelerated by

92   positive selection targeting, for example, male-related genes on the Y chromosome[24]; or by

93   background selection purging the deleterious mutations from highly dosage-sensitive genes[25].

94   Simulation showed that both forces play a different role at different stages of Y/W

95   degeneration[26]. Both have been implicated in analyses of mammalian[24,27] and *Drosophila*[28,29] Y-

96   linked genes. However, no evidence has been found for female-specific selection among the W-

97   linked genes (also called gametologs) of chicken[21] or flycatcher[30].

98   Intriguingly, in both birds[20] and mammals[31], as well as several plant species (e.g. *Silene*

99   *latifolia*[32]), recombination suppression has proceeded in a stepwise manner presumably through

100   chromosome inversions, leaving a stratified pattern of sequence divergence between sex

101   chromosomes termed 'evolutionary strata'[33]. Eutherian mammalian X and Y chromosomes have

102   been inferred to share at least three strata, with another two more recent ones shared only

103   among catarrhines (old world monkeys and great apes)[27]. It has been recently discovered that

104   the history and tempo of avian sex chromosome evolution is much more complicated than that

105   of mammals[20]. All bird sex chromosomes only share the first step of recombination suppression

106   (stratum 0, Aves S0) encompassing the avian male-determining gene *DMRT1*. This was

107   followed by the independent formation of S1 in the Palaeognathae (e.g., ratites and tinamous)

108   and in the ancestor of the Neognathae (all other extant avian radiations). Ratites have halted

109   any further recombination loss and maintained over two thirds of the entire sex chromosome

110   pair as the exceptionally long recombining pseudoautosomal regions (PAR). Therefore, their W

111   chromosomes are unusually homomorphic and gene-rich comparing to the Z chromosomes. In

112   contrast, all species of Neognathae examined have suppressed recombination throughout most

113   regions of the sex chromosomes with short and varying sizes of PAR[34]. Overall, avian W

114   chromosomes seem to have retained more genes and decayed at a slower rate than the

115   mammalian Y chromosomes. Furthermore, sexually monomorphic species (e.g., most ratites)

4

116   seem to differentiate even slower than sexually dimorphic species (chicken and most Neoaves)

117   in their sex chromosomes, consistent with the hypothesis that sexually antagonistic genes have

118   triggered the expansion of recombination suppression between sex chromosomes[35]. However,

119   due to the ratites' deep divergence from other birds, and also an expected much lower mutation

120   rate due to their larger body size and longer generation time, it is unclear what the actual

121   influence of sexual selection is on the rate of sex chromosome evolution. All Neoaves species

122   share one stratum S2, with the more recent evolutionary history of sex chromosomes of

123   songbirds unclear. So far, only one songbird, the collared flycatcher has been extensively

124   characterized for its W-linked genes[30], whose number is within the range of 46 to 90 W-linked

125   genes reported for other Neoaves[20]. To elucidate the evolutionary history of songbird sex

126   chromosomes, we produced high-quality female genomes of five birds-of-paradise (BOP).

127   Together with a re-analysis of 6 other published female genomes of songbird species[30,36-39], our

128   analyses cover the two major songbird lineages (Corvida and Passerida) that rather diverged in

129   the last 50 MY[2,40].

130

131   **Results**

132   **Characterization of songbird sex chromosome sequences**

133   We produced between 36-150 fold genomic coverage of sequencing data for each BOP species,

134   and performed *de novo* genome assembly followed by chromosome mapping using the great tit

135   (*Parus major*) genome as reference[6]. The high continuity and completeness of the draft

136   genomes are reflected by their scaffold N50 lengths (all >3 Mb except for the raggiana BOP,

137   *Paradisaea raggiana*) and BUSCO core gene completeness scores (92.9% to 94.0%)

138   (**Supplementary Table 1**). To reconstruct the evolutionary history of sampled songbird sex

139   chromosomes, we first identified the sex-linked sequences from the draft genome of each

140   species. We searched for the scaffolds from sexually differentiated regions (SDR) as those that

141   show half the female sequencing depth of autosomes (**Fig. 1A**, **Supplementary Fig. 1**).

142   Scaffolds from putative PARs were inferred by their homology to that of flycatcher[41], and also

143   sequencing depth similar to autosomes. It is noteworthy that our method cannot identify very

144   recent fusion/translocation of autosomal fragments to the sex chromosome pair ('neo-sex'

145   chromosome), as in the case of warblers[42]. All the studied songbirds have a short putative PAR

146   ranging from 564 kb to 781 kb. We then categorized SDR sequences as either Z- or W-linked

147   with the expectation that the W chromosome would diverge much faster than the homologous Z

148   chromosome of the same species, when being compared to the Z chromosome of an outgroup

149   species (**Materials and Methods**), as a result of rapid accumulation of deleterious mutations

150 and repetitive elements after recombination suppression. It is possible that Z-linked paralogs,

151 although not expected to be abundant in the compact avian genomes[36], would confound our

152 identified W-linked sequences by showing a similar sequence divergence pattern. Thus, among

153 the five species with male sequencing data, we further verified the sex-linkage by confirming all

154 the putative W-linked scaffolds with a clear female-specific pattern (**Fig. 1A, Supplementary**

155 **Fig. 2**). The assembled lengths of the largely euchromatic parts of W chromosomes range from

156 1.33 to 6.52 Mb, corresponding to only 1.9% to 8.5% of the Z chromosome length across

157 species (**Fig. 1B**, **Supplementary Table 2**), probably as a result of large deletions and massive

158 invasions of repetitive elements which might be difficult to assemble with existing technologies.

159 Indeed, the repeat content of the assembled W chromosomes is 2.5 to 4.9 fold higher than that

160 of Z chromosomes on the chromosome-wide average (**Supplementary Fig. 3 and Table 2**),

161 consistent with the patterns of W chromosomes of chicken and flycatcher[43,44]. We have not

162 found shared syntenic W-linked regions across species that may have been deleted in the

163 ancestor of songbirds compared to the chicken or ostrich genome, as suggested by previous

164 cytological studies[45].

165

166 **Age-dependent genomic impact of evolutionary strata**

167 If recombination was suppressed between sex chromosomes in a stepwise manner, we expect

168 to find a gradient of Z/W sequence divergence levels along the chromosome sequence of the Z

169 chromosome, like what has been reported along the human X chromosome[46]. Previous work

170 showd that the Z chromosomes of the Neognathae have undergone dramatic intrachromosomal

171 rearrangements which resulted in a misleading extant synteny for ordering different evolution

172 strata[20]. By contrast, the Palaeognathae (e.g. emu *Dromaius novaehollandiae* and ostrich

173 *Struthio camelus*) have maintained highly conserved sequence synteny even with reptile

174 species, with over two thirds of their sex-linked regions still recombining as an appropriate

175 approximate of proto-sex chromosomes of all bird species[20]. We first confirmed these patterns

176 by comparative mapping of the nearly chromosome-level assemblies of Z chromosomes of

177 collared flycatcher, hooded crow (*Corvus corone*)[47] vs. chicken and emu (**Fig. 2A**). This allows

178 us to reconstruct the spanned regions of S0 shared by all birds, and a part of S1 shared by all

179 Neognathae species in the studied songbird genomes by their homology to the emu genome.

180 They are mapped as two continuous regions on the emu Z chromosome, but have become

181 severely reshuffled into dispersed fragments in songbirds (**Fig. 2A, Supplementary Fig. 4-5**).

182 Two recently formed strata (Neoaves S2 and S3) are highly conserved for their synteny across

183 avian species, and each show significantly ($P < 0.05$, Wilcoxon test) different levels of Z/W

184   sequence divergence (**Fig. 2B, Supplementary Fig. 6**), GC3 (GC content at the third codon

185   positions, **Supplementary Fig. 7**) and Z-linked long terminal repeat (LTR) content (**Fig. 2C,**

186   **Supplementary Fig. 7**) from each other. The drastic change of Z/W sequence similarity allows

187   us to precisely map the boundaries between these two strata. In general, a series of

188   recombination suppression has reshaped the genomic architecture of Z chromosomes in a

189   chronological order. Regions of younger strata exhibit much fewer Z-linked intrachromosomal

190   rearrangements between species, suggesting the reduced selective constraints on gene

191   synteny after recombination was suppressed in the older strata[48]. In particular, GC3 content

192   decreases, while the repeat content increases by the age of stratum. This is probably because

193   of weaker effects of GC-biased gene conversion (gBGC)[49], and purifying selection against TE

194   insertions[50] under reduced recombination, both of which have been acting for a longer time

195   within Z-linked regions of older strata. Consistently, a similar pattern has also been found

196   contrasting PAR vs. the remaining Z-linked regions in the collared flycatcher[41].

197

198   **Lineage-specific burst of retrotransposon probably has induced recombination**

199   **suppression between sex chromosomes**

200   The distribution of long interspersed nuclear elements (LINEs), mainly the retrotransposon

201   chicken repeat 1 (CR1) elements shows an exceptional pattern compared to that of LTR

202   elements (**Fig. 2C**, **Supplementary Fig. 7**) associated with the age of strata. S3 presents a

203   similar proportion of CR1 with S0 that is much higher than the rest of the Z-linked regions. A

204   close examination shows that this is due to the specific accumulation of CR1 elements at the

205   boundary between PAR and S3. Such a burst of CR1, particularly only the CR1-E1 subfamily[51],

206   extends into about one third of the entire S3 region and is shared by all investigated songbirds

207   but absent in the deep-branching passerine rifleman (*Acanthisitta chloris*) and other Neoaves

208   (**Fig. 2D**, **Supplementary Fig. 8**). The exact boundary sequence is not assembled probably due

209   to such accumulation of CR1-E1, and also previously reported multiple deletions in passerines

210   that removes a gene *DCC* (Deleted in Colorectal Carcinoma) highly conserved across other

211   vertebrates[52]. This gene is responsible for axon guidance for brain midline crossing and has

212   independently been lost in some but not all passerines and Galliformes[53].

213   In addition, we find evidence that the burst of CR1-E1 elements coincides with the S3

214   emergence. Our phylogenetic reconstruction of Z- and W-linked gametolog sequences shows

215   that songbird-derived sequences are always grouped by chromosome instead of species,

216   compared to outgroup birds (**Fig. 2E**, **Supplementary Fig. 9-12**). This indicates that all

217   songbirds share four evolutionary strata, with the latest stratum S3 formed after the divergence

7

218   between songbirds and the rifleman. The highly conserved synteny between songbird species,

219   and between songbirds and chicken of S3 on the Z chromosome (**Fig. 2A, Supplementary Fig.**

220   **5**), suggests that there was no chromosomal inversion on the Z chromosome at S3, and the

221   recent burst of the CR1-E1 subfamily probably contributed to the formation of S3, although we

222   cannot exclude the contribution of possible W-linked chromosomal inversions. Interestingly,

223   other CR1 subfamilies (CR1-E4, CR1-E5, CR1-E6) have an independent burst at the PAR/S3

224   boundary in rifleman (**Fig. 2D**, **Supplementary Fig. 8** and **Table 3**). Given that this boundary

225   region has been shown to have frequent but different degrees of multiple gene loss in different

226   lineages of birds[52,53], it is very likely a hotspot for structural changes (including LINE

227   accumulation) that have recurrently contributed to the independent formation of S3 in many bird

228   species.

229

230   **Fast-Z pattern of songbirds implies dynamic evolution of sexual selection**

231   The formation of evolutionary strata has subjected the Z chromosome to male-biased

232   transmission and a reduced effective population size, which are expected to produce faster

233   mutation and evolutionary rates of Z-linked genes[17]. Indeed, we found a larger branch-specific

234   synonymous substitution rate (dS) of Z-linked genes (statistically not significant), but a

235   significantly smaller dS of W-linked genes, compared to that of autosomal genes ($P < 0.01$,

236   Wilcoxon rank sum test, **Supplementary Fig. 13**), as a result of male-driven evolution. The

237   branch-specific evolutionary rates (ω) measured by the ratios of nonsynonymous substitution

238   rate (dN) over dS have significantly ($P < 0.01$, Wilcoxon rank sum test, **Supplementary Fig. 14**)

239   increased for both Z- and W-linked gametologs relative to autosomal genes, indicating a 'fast-Z'

240   effect and degeneration of W-linked genes (see below). Previous simulation work and

241   experimental evidence in Galloanserae have suggested that different degrees of sexual

242   selection targeting males will influence the male-mating success, hence the genetic drift effect

243   on the Z chromosome to a different degree[17,19]. Songbirds, especially BOPs, have been

244   frequently used as textbook demonstrations for sexual selection, though the long-term

245   evolutionary history of sexual selection remained unclear. To reconstruct this, we approximate

246   the intensity of sexual selection targeting males by measuring the fast-Z effect (Z/A value, the

247   ratio of branch-specific ω values of Z-linked genes vs. autosomal genes) in a phylogenetic

248   context (**Fig. 3, Supplementary Table 4**). The varying Z/A values at different lineages suggest

249   a dynamic change of intensity of sexual selection, even among the five BOP species that

250   diverged within the last 15 MY[54]. While the significant (permutation test, $P < 0.05$) fast-Z pattern

251   of the sexually monochromatic American crow (*Corvus brachyrhynchos*) may reflect the sexual

252   selection acting on the ancestral lineage leading to the Corvidae (crows, jays and allies), a lack

253   of such a pattern in the raggiana BOP and magnificent BOP (*Cicinnurus magnificus*) is

254   somewhat unexpected. These species are known for their lekking behaviors[55], with which very

255   few males dominate almost all females for copulation through outcompeting other males with

256   displays. This produces a strongly biased male-mating success, and direct challenge for

257   maintaining genetic variation in the population ('the lekking paradox')[56]. Very few field

258   quantitative studies have been performed on BOP species, and it will be interesting to

259   investigate whether raggiana and magnificent BOPs female individuals may solve the 'lekking

260   paradox' by changing their mating preference, and in fact mate with more males than is

261   presumed.

262

263   **Conserved gene content of the songbird W chromosomes**

264   In contrast to the dynamic evolution of Z-linked genes and sequences, W chromosomes of all

265   the studied songbirds have undergone dramatic gene loss but exhibit an unexpected

266   conservation of the retained gene repertoire across species. The numbers of assembled W-

267   linked genes range from 31 in house sparrow (*Passer domesticus*) to 63 in the king BOP

268   (*Cicinnurus regius*), compared to about 600 to 800 Z-linked genes in each species (**Fig. 4A**,

269   **Supplementary Tables 2 and 7**). These numbers are likely an underestimate because genes

270   embedded in highly repetitive regions maybe missing from the current W chromosome

271   assemblies. In general, Corvida species retain more W-linked genes than Passerida species

272   (**Supplementary Table 5**), likely due to their longer generation time thus a lower mutation rate.

273   Most W-linked genes are single-copy without lineage-specific expansion, except for HINT1W

274   (**Supplementary Fig. 15**). Despite the relative rarity of gene duplication in birds compared to

275   mammals[57], we find one W-linked gene that is a duplicated copy from an autosomal gene *NARF*

276   in American crow (**Supplementary Fig. 16**). This duplicated gene is also present in another

277   Corvida species *Lycocorax pyrrhopterus* (Alexander Suh, personal communication), suggesting

278   that the duplication event likely occurred in the ancestor of Corvida. It will be interesting to see

279   whether this gene shows signatures of female-specific selection, e.g, novel patterns of ovary-

280   specific expression, which drives its fixation on the W. Fifty-seven genes are shared by at least

281   one Corvida and another Passerida species, and 23 genes are shared between at least one

282   songbird species and the chicken[21]. This suggests they were still preserved on the W

283   chromosome before the divergence of passerine or Neognathae species. Interestingly, despite

284   the independent origin of S2 in the chicken and Neoaves[20], all the chicken W-linked genes but

285   one are also found in passerines, indicating similar underlying evolutionary forces governing

9

286 their convergent retention since Galloanserae and Neoaves diverged from each other 89 MY

287 ago.

288   To examine such forces, we perform gene ontology (GO) analyses on the 79 genes that

289 are present on the W chromosome of at least one songbird species. They are enriched ($P <$

290 0.01, Fisher's exact test) for two GO terms of 'DNA binding' and 'transcription factor activity,

291 sequence-specific DNA binding' (**Supplementary Table 6**). This suggests that similar to the

292 mammalian Y-linked genes[27], some W-linked genes are retained for their important functions of

293 regulating gene activities elsewhere in the genome. The Z-linked homologs of lost genes evolve

294 significantly ($P <$ 0.01, Wilcoxon rank sum test) faster with their ω ratios higher than those of the

295 retained genes on the W chromosome (**Fig. 4B**, **Supplementary Fig. 17**). This shows a

296 different selective pressure acting on these two sets of genes on the proto-sex chromosomes.

297 As this pattern maybe confounded by the 'faster-Z' effect of hemizygous Z-linked genes, we

298 further study the autosomal orthologs of these genes in the green anole lizard (*Anolis*

299 *carolinensis*). We find that the orthologs of retained genes have significantly ($P <$ 1.497e-05,

300 Wilcoxon rank sum test, **Fig. 4C**) higher expression levels in all lizard tissues of both males and

301 females, and also a broader tissue expression pattern than those of the lost genes across all the

302 tissues (**Fig. 4D**). The patterns are also consistent among the four songbird evolutionary strata.

303 These results are robust if we use emu to infer the ancestral expression pattern

304 (**Supplementary Fig. 18**), whose sex chromosomes are largely a PAR. Consistent with the

305 result of collared flycatcher[10], we find no evidence that female-specific selection may prevent

306 the gene loss, or drive certain genes to undergo positive selection on the songbird W

307 chromosomes. We found no excess of ovary-biased lizard orthologs among those of the

308 retained W-linked genes: only 6 out of 72 (8.3%) are ovary-biased while the genome-wide

309 proportion is about 20%.

310

311 **Comparing gene loss between avian W chromosomes and mammalian Y chromosomes**

312 Overall, 4.6% to 9.2% of single-copy W-linked genes per songbird species (**Fig. 4A**), compared

313 to 1.6% to 3.0% single-copy Y-linked genes per mammalian species[27] have been retained for

314 their essential or sex-specific functions. A seemingly high retention ratio of W-linked genes in

315 birds can be partially attributed to the generally much lower mutation rate of W chromosome

316 relative to Y chromosome by male-driven evolution effect (**Supplementary Fig. 13**), assuming a

317 similar generation time between mammals and birds. In addition, a much more frequent and

318 stronger sex-specific selection acting on the Y chromosome than on the W chromosome,

319 sometimes driving the massive expansion of Y-linked gene copies with male-related function[24],

10

320 probably also has contributed to a faster rate of Y chromosome gene loss by the hitchhiking

321 effect[24,26]. To scrutinize the tempo of gene loss throughout the evolution of songbird sex

322 chromosomes, we conservatively reconstructed the numbers of retained W-linked gametologs

323 at each phylogenetic node of the avian tree (**Fig. 5A, Supplementary Table 8**), by identifying

324 the genes present on any of the studied avian W chromosomes. We find that within each

325 stratum, the percentage of gene loss is always much larger at an earlier evolutionary time point

326 than in the recent ones, and this is consistent between birds and mammals (**Fig. 5B**). Thus the

327 majority of gene loss has probably occurred during the early stage of W or Y chromosome

328 evolution, and the rate of gene loss dramatically decreases by the less retained genes. Although

329 convergent gene loss may cause an overestimation of lost genes at more ancestral time points

330 (e.g., in S0 region), this probably has little influence on our estimate in the most recent songbird-

331 specific stratum S3 which has already lost 69.8% of the W-gametologs within 50 MY. We also

332 find that the retained genes of songbird W chromosomes are often close to each other

333 (**Supplementary Fig.19**), suggesting large sequence deletions have had an important

334 contribution to drastic gene loss.

335 The decrease of gene loss rate on Y/W chromosomes through evolutionary time can be

336 explained by a weaker Hill-Robertson effect that the less retained genes can induce, which has

337 been previously shown in a simulation study[26]. Thus, the ancestral gene number of older

338 evolutionary strata which would have undergone more serious gene loss must have a larger

339 influence on the extant number of retained genes. A lower rate of retained mammalian Y-linked

340 genes relative to avian W-linked genes, besides the cause of male-driven evolution, can be also

341 attributed to the fact that the first two or three mammalian evolutionary strata emerged before

342 the divergence of eutherians which together account for over 93.2% of the entire gene content

343 of ancestral Y chromosome, while those of birds only account for 53.3% of the entire ancestral

344 W-linked gene content (**Fig. 5C**).

345

346 **Discussion**

347 The evolution of sex chromosomes is usually, but not always (e.g., in frogs[58], ratite birds[20] and

348 pythons[59]), marked with episodes of recombination suppression that eventually restricts the

349 recombining region to one or two small PARs at the end of the sex chromosomes. The resulting

350 patterns of evolutionary strata have been widely reported in many animal and plant species,

351 with the responsible formation mechanism presumed to be chromosomal inversions[60]. Indeed,

352 footprints of inversions in the latest two strata between mammalian X and Y chromosomes have

353 been found by examining the synteny order between X/Y, and particularly the X-linked boundary

11

354   sequence that has been disrupted into two dispersed sequences on the Y chromosome[61]. In

355   songbirds, we show here that four evolutionary strata were formed before their species radiation.

356   Importantly, we provide evidence suggesting an alternative scenario of recombination

357   suppression. We find in the latest stratum S3, a genome-wide burst of a certain transposable

358   element (TE) led to its specific accumulation at the mutation/insertion hotspot near the PAR

359   boundary of the sex chromosomes (**Fig. 2, Supplementary Table 3**), which probably further

360   mediated recombination suppression. It has been reported with some exceptions, in many

361   species that local recombination rates and abundance of TE elements generally have a

362   negative association, with their causal relationship difficult to be disentangled[50]. In the case of

363   the songbird S3 region, several patterns suggest that TE accumulation might be the cause

364   rather than the result of recombination suppression. First, in several species (e.g., Lawes's

365   parotia (*Parotia lawesi*) and the king BOP, **Fig. 2D**, **Supplementary Fig. 8**), the responsible

366   CR1-E family is also enriched close to the PAR boundary, where normal levels of recombination

367   can be expected. Second, the abundance of CR1-E gradually decreases further away from the

368   boundary, which also has undergone genomic deletions or rearrangements independently in

369   other bird species[52,53], thus is likely to be a mutation hotspot. Third, only CR1-E repeats, but not

370   any other type of CR1 or repeat families, are enriched at the PAR/S3 boundary. These results

371   are not likely to be expected if S3 stratum formation occurred before the CR1-E accumulation,

372   which predicts a uniformly distributed accumulation of various kinds of repeat elements (e.g.,

373   LTR elements, **Supplementary Fig. 20**) by Hill-Robertson interference effects that would not

374   extend into the PAR. Indeed, our comparative analyses between species suggest no Z-linked

375   inversion in S3 (**Fig. 2A, Supplementary Fig. 5**), though we cannot exclude the possibility of a

376   W-linked inversion that may have instead contributed to the formation of S3. The verification of

377   the latter requires future improvements of genome assembly using for example,

378   PacBio/Nanopore sequencing technology to assemble the highly repetitive W-linked sequence.

379   Under our proposed scenario, TEs probably reduced the recombination rate in PAR

380   through, for example, changing the chromatin structure or disrupting recombination hotspots[50].

381   The CR1-E retrotransposon accumulation occurred at the mutation hotspot located at PAR/S3

382   boundary despite the selection against its deleterious effects of disrupting gene functions. This

383   has resulted in partial or complete deletion of several genes in songbirds[52,53]. However, the

384   consequential reduction of recombination rate can provide the selective advantage of

385   accelerating the fixation of pre-existing sexually antagonistic (SA) alleles in PAR through sex-

386   biased transmission; or subjecting the PAR for the 'fast-Z' evolution by male-driven evolution

387   effect (**Fig. 3**) and increasing its exposure for male-biased selection, so that novel SA alleles

388   may more frequently emerge and become fixed. The latter has been implicated by the recent

389   findings in songbirds that male-specific trait genes, for example those related to sperm

390   morphology[62] or plumage colors[63], which have recently become diverged within or between

391   populations, are enriched on the Z chromosome. In addition, TE accumulation is likely to

392   increase the chance of chromosome inversions through ectopic recombination, or by reducing

393   the selective constraints on gene synteny[48,50]. The latter is supported by our result that older

394   evolutionary strata have undergone more Z-linked genomic rearrangements between songbird

395   species than the younger ones (**Fig. 2A**), which creates a positive feedback once the

396   recombination suppression was initiated. This provides a mechanistic explanation for the more

397   frequent fixation of Z-linked inversions found among passerines.

398   While the Z chromosome is predicted to accumulate dominant male-beneficial mutations,

399   the W chromosome is expected to accumulate female-beneficial mutations responding to

400   female-specific transmission[64]. Here we find that more genes to have been retained on the W

401   chromosomes of songbirds than that of chicken (**Fig. 4**). However, both previous works on the

402   chicken and collared flycatcher[21,30], as well as our present study, have not found evidence for

403   such 'feminization' of W chromosome. This is in contrast to multiple reported cases of

404   'masculinization' of the Y chromosome in the ancient mammalian Y chromosome systems[24] or

405   the recently-evolved Y chromosome of *Drosophila miranda*[28]. Genes specifically expressed in

406   the male germline have either massively amplified their copy numbers, or upregulated their

407   expression levels on these evolving Y chromosomes. Such a difference can be explained by the

408   fact that regardless of the sex chromosome type, sexual selection is often targeting males in

409   most species, thus the Z/Y chromosomes are more frequently influenced than the W/X

410   chromosome due to their male-biased transmission, although X chromosomes are nevertheless

411   expected to accumulate recessive male-beneficial alleles[9]. The convergently evolved pattern

412   shared between the mammalian Y and avian W chromosomes is largely attributed to the

413   essential genes that have important regulatory functions and are preferentially retained over the

414   long period of recombination suppression (**Fig. 4**)[21,27]. However, previous transcriptome

415   comparison of chicken breeds selected for egg-laying vs. fighting, i.e. female-specific vs. male-

416   specific traits, has found that most W-linked genes are upregulated in the former[64]. Few high-

417   quality avian W chromosome sequences are available except for that of chicken[21,65]. Songbirds

418   provide a rich resource with many species having a reversed direction of sexual selection and

419   ornamented females[66]. Application of long-read sequencing technology in future will help to

420   elucidate the role of the W chromosome in sexual selection and speciation of birds[67].

421

Materials and Methods

Genome assembly and annotation

Genomic DNA were extracted from fresh female tissue samples of from BOP species *Cicinnurus regius* (ANWC B24969), *Cicinnurus magnificus* (ANWC B27061), *Paradisaea rubra* (YPM84686), *Parotia lawesii* (ANWC B26535), using Thermo Scientific™ KingFisher™ Duo Prime purification system. Paired-end and mate pair libraries for these samples were prepared by SciLifeLab in Stockholm, Sweden. All libraries were sequenced on Illumina HiSeq 2500 or Hiseq X v4 at SciLifeLab in Stockholm, Sweden. For *Paradisaea raggiana* (USNM638608), genomic DNA were extracted using EZNA SQ Tissue DNA Kit. One paired-end and one mate-pair libraries were sequenced on Illumina HiSeq 2000. We also used published female genomes of *Corvus brachyrhynchos*, *Serinus canaria, Passer domesticus, Geospiza fortis, Ficedula albicollis, Pseudopodoces humilis* for analysis in this work (**Supplementary Table 9**)[30,36-39]. The BOP genomes were assembled using ALLPATHS-LG[68] with 'HAPLOIDIFY=True'. For *P. raggiana*, due to the lack of overlapping paired-end reads, SOAPdenovo2[69] was used instead. Gene models were annotated using the MAKER pipeline in two rounds[70]. The reference protein sequences of zebra finch, great tit, hooded crow, American crow, collared flycatcher and chicken were downloaded from NCBI RefSeq (**Supplementary Table 9**). Using the reference protein sequences and chicken HMM (Hidden Markov Models), an initial set of gene models was obtained by using MAKER, and those models were taken for SNAP model training[71]. In addition, 3000 gene models were selected for Augustus training[72]. The trained gene models and the protein sequences were taken as input for MAKER in the second run. To annotate repeats, first we used RepeatModeler[73] to identify and classify repeat elements for each species, including the published genomes[74]. Then we combined each individual library with an avian repeat library and further manually curated BOP repeat consensus sequences, followed by annotation repeats using RepeatMasker (v1.7)[73].

Identification of sex-linked sequences

We used the sequence of the great tit Z chromosome as reference to search for homologous Z-linked sequences in the studied species, as it has the best assembly quality (contig N50: 133kb, scaffold N50: 7.7Mb) among the published genomes of songbirds[6]. We used nucmer from MUMMer package (v4.0)[75] for genome-wide pairwise sequence alignment. For any scaffold larger than 10 kb that has more than 60% of the sequence aligned to great tit chrZ, it was identified as candidate Z-linked sequences. They were further inspected by comparing the female sequencing coverage to that of autosomes. To do so, the raw female reads were

14

456 mapped to the genome references by using bwa[76] then the mean sequencing depth of every 50-

457 kb window was calculated. Sequencing depth for single sites was counted using 'samtools

458 depth' before calculating window-based coverage. Any site with low mapping quality (less than

459 60) and very high coverage (3 times larger than average) was excluded.

460   To identify the W-linked scaffolds, first we identified the scaffolds that show half-

461 coverage relative to that of autosomes. We plotted the distribution of coverage of all scaffolds to

462 decide the cutoff of 'half-coverage' (**Supplementary Fig.1**). Those half-coverage scaffolds were

463 expected to be either Z-linked or W-linked. To distinguish between the Z and W, we aligned the

464 half-coverage scaffolds to the Z chromosome from the hooded crow, a closely related species to

465 BOP, whose genome is derived from a male without contamination of W-linked sequences. We

466 used nucmer for sequence alignment and only kept 1-to-1 best alignments. We then calculated

467 the proportion of sequences of each scaffold that was aligned to the hooded crow chrZ, and

468 decided a cutoff to separate the Z and W based on the distribution of the proportion of Z-linked

469 alignment. We further excluded the candidate W-linked scaffolds that over 10% of the

470 sequences were aligned to hooded crow autosomes, or a larger portion of sequences aligned to

471 the Z chromosome than the W. Finally, only the scaffolds that were larger than 50 kb were kept.

472 We also retrieved additional Z-linked scaffolds that were absent in the results from the

473 homology-based approach, likely due to the missing Z-linked sequences in the great tit Z

474 chromosome assembly. For ground tit, medium ground finch, house sparrow and collared

475 flycatcher in which both male and female sequencing reads are available, the W-linked

476 sequences were further verified by mapping the male reads. Specifically, for every scaffold, the

477 number of nucleotide sites that were covered by male and female sequencing data were

478 counted respectively as Nm and Nf, and the ratios of Nm to Nf were calculated. W-linked

479 scaffolds were expected to have Nm/Nf ratios close to zero and one for autosomal or Z-linked

480 scaffolds. For PARs, we used the known sequences of the zebra finch[77] and collared

481 flycatcher[41] to search for the homologous sequences in other species using nucmer. The female

482 sequencing depths of those candidate PARs were compared to autosomes and required to be

483 similar.

484

485 **Demarcation of evolutionary strata**

486 We ordered and oriented the identified Z-linked scaffolds into one pseudo-chromosomal

487 sequence (pseudo-chrZ) based on their alignments against the chromosomal assembly of great

488 tit. Hooded crow has only 15 Z-linked scaffolds and 10 out of them are larger than 1 Mb[47], thus

489 was used as a representative Corvida species for comparison on the Z chromosome. We

15

490   determined the relative order and orientation of the scaffolds according to their alignment on the

491   great tit Z chromosome. Similarly, for BOP species, we created pseudo-chromosome Z using

492   great tit as guiding reference. The pseudo-Z chromosome of emu was built using ostrich Z

493   chromosome[78] [79]as reference. We used nucmer for pairwise alignment of the Z or pseudo-Z

494   chromosomes. Alignments short than 2kb were excluded.

495   The W-linked scaffolds were then aligned to the pseudo-chrZ using lastz[80], after masking

496   repetitive sequences. Sequence similarity of the alignments between the chrZ and chrW was

497   calculated by the script pslScore from UCSC Genome Browser (https://genome.ucsc.edu/).

498   Individual alignments that had sequence similarity lower than 60 or higher than 96, or alignment

499   length shorter than 65 were removed. After that, we concatenated alignments within non-

500   overlapping sliding windows of 100 kb, and calculated sequence similarity for the concatenated

501   alignments. When the length of concatenated alignments was shorter than 2 kb within a 100-kb

502   window, the window was excluded from further analyses. The window-based sequence

503   similarity was then plotted along the pseudo-chrZ. The shift of sequence similarity was used to

504   demarcate the boundaries of S3/S2 and S2/S1. Since very few W-linked sequences have been

505   assembled for the most ancient stratum S0, we mapped its reshuffled fragments in songbirds

506   based on their homology with the emu S0. Our previous study showed emu has a recent

507   species-specific stratum (S1) while the first stratum (S0) is ancient and shared by all birds[20].

508   This allows for the demarcation of S1 and S0 by detecting their differential degree of ZW

509   differentiation. Specifically, by using relatively relaxed mapping criteria (bwa mem) to map

510   female sequencing reads, only S0 showed reduced coverage relative to autosomes or PAR

511   (**Supplementary Fig. 4**), while S1 showed reduced coverage when stringent mapping was

512   applied (bwa sampe –a 900 –n 1 –N 0 –o 10000).

513   To scrutinize the accumulated LINE (mostly CR1) elements at the PAR/S3 boundary, we

514   first divided them into each subfamily (approximated by each repeat consensus sequence)

515   according to the RepeatMasker annotation. Among all subtypes, CR1-E1 is usually ranked with

516   the highest or second highest number at the S3 region across all songbird species. Other high-

517   ranking subtypes included CR1-E3, CR1-E5, CR1-E4, CR1-E6, CR1-J2 and CR1-Y2. Then we

518   plotted each subtype's abundance with a 100 kb non-overlapping window along the Z

519   chromosome, in all the studied songbirds, as well as outgroup species rifleman and falcon[81], to

520   identify the burst of CR1-E1.

521

522   **Sex-linked gene analyses**

16

523   After removing LTR-derived genes, we used BLAT[82] to align the annotated coding sequences of

524   W-linked genes to the Z chromosome to search for their gametologous pairs. Then we produced

525   pairwise gametolog alignments using MUSCLE,[83] and then manually inspected the alignments

526   to remove genes with short or ambiguous alignments. For species other than the BOPs, gene

527   models of the W chromosomes were directly retrieved from the RefSeq genome annotation,

528   with some of them subjected to manual inspections. To determine the orthologous relationship

529   among the studied species, we first extracted the sequence of the longest protein of each gene.

530   Those protein sequences were subjected to all-vs-all BLAST search that was implemented

531   through the program proteinortho[84]. BLAST hits with identity lower than 50% or alignment

532   coverage lower than 50% were removed. We also took gene synteny information into account

533   when grouping orthologous genes. Besides the twelve female genomes for which we studied

534   the sex chromosomes, we also included high-quality genomes of great tit, hooded crow and

535   ostrich (**Supplementary Table 9**). We retained those orthologous groups if they contained

536   sequences of at least ten species.

537   To estimate the substitution rates of coding sequences, first we performed multiple

538   sequence alignments for orthologous genes. We used the guidance2 pipeline

539   (http://guidance.tau.ac.il/ver2/source.php) which employs PRANK to align sequences of codons.

540   To filter low-quality sites in the alignments, we ran trimal (http://trimal.cgenomics.org/) to further

541   filter the ambiguous alignments with '-gt 0.8'. The phylogeny of the birds was extracted from

542   Jetz et al. (2012)[85]. We used codeml from the PAML package[86] to estimate the synonymous

543   substitution rates (dS) and non-synonymous substitution rates (dN). To estimate chromosome-

544   wide dN and dS, the sums of synonymous or nonsynonymous substitutions were divided by

545   those of the number of synonymous or nonsynonymous sites, as applied in Wright et al.[19].

546   Individual genes with abnormal dN (higher than 0.1, in total 179 genes) or dS (higher than 0.8,

547   in total 135 genes) out of 111,748 orthologous gene groups were removed. Confidence intervals

548   were calculated by 100 bootstraps. The GC content of the third position codons (GC3) was

549   calculated using codonW[87] for the longest isoform of each gene. Chromosome-wide dN/dS (ω)

550   was calculated using the ratios of chromosome-wide dN to chromosome-wise dS. The fast-Z

551   effect is measured by Z/A value, the ratio of ω values of Z-linked genes to autosomal genes,

552   and we calculated the Z/A value for every terminal branch and internal branch. To determine

553   whether the difference of ω between Z-linked and autosomal genes wasis significant, we

554   performed permutation test by resampling 1000 times. The genes of chromosome 4 and

555   chromosome 5 were used to represent autosomal genes as the sizes of those two

556   chromosomes are similar to the Z chromosome.

17

557   For each gametologous pair, we grouped together Z-linked genes and assembled W-

558   linked genes and performed multiple sequence alignment. The same guidance2 pipeline was

559   used as in sequence divergence analysis. For S3 genes, we also included rifleman[36]. We used

560   IQTREE[88] to construct maximum likelihood phylogenetic trees. The best substitution model was

561   automatically selected in by IQ-TREE. We ran 100 bootstraps to evaluate the confidence levels

562   of phylogenies. Ostrich was used as the outgroup to root the tree. The gene ontology (GO) term

563   annotations for both gametolog-pairs genes (list) and entire Z-linked genes (background) of

564   chicken were analyzed using DAVID 6.8[89]. GO term enrichment was analyzed by comparing the

565   number of appearance of GO terms of 'list' gene versus 'background' gene.

566

567   **Gene loss analysis**

568   We identified a total of 673 Z-linked orthologous genes that are shared between chicken and

569   emu as the putative ancestral genes on the proto-sex chromosomes of birds. For the gene

570   cluster that was lost in chicken at the DCC locus of S3, an ancestral gene content was inferred

571   based on Fig. 3 of Patthey et al.[53]. They were then grouped into four evolutionary strata

572   according to the strata annotation of songbird Z chromosomes. At each node of the avian

573   phylogenetic tree, we calculated the ratio of the number of lost genes to the number of ancestral

574   genes at that node. For the nodes leading to Passerida and Corvida, if there were at least one

575   species retaining a W-linked gene, we inferred that this gene was present in their ancestor.

576   Similarly, we defined the presence of ancestral genes in Passeriformes, Neoaves and

577   Neognathae using other published avian W-linked gene information[20,21].

578

579   **Gene expression analysis**

580   We downloaded the raw RNA-seq reads of green anole (brain, gonad, liver, heart and kidney)

581   and emu (brain, gonad and spleen) from SRA (**Supplementary Table 9**). In addition, we

582   collected transcriptomes of adult emu kidneys of both sexes. We used the RSEM pipeline[90] to

583   quantify the gene expression levels. We used STAR[91] to map raw reads to the transcriptomes

584   which was constructed based on gene annotations. The expression levels were estimated at the

585   gene level, in the form of TPM (Transcripts Per Million). The mean TPM value of biological

586   replicates was calculate for each gene. Tissue specificity of gene expression was estimated by

587   calculating tau[92].

588

589   **Data availability**

590 Genome sequencing and RNA-seq data, and genome assemblies generated in this study have

591 been deposited in the NCBI SRA under PRJNA491255.

592

607

608 **Figure Legend**

609 **Figure 1. The Z and W chromosomes of different songbirds.**

610 **A)** We use medium ground finch as an example to demonstrate our identification and

611 verification of sex-linked sequences. For each scaffold shown as a circle with scaled size to its

612 length, the ratio of nucleotide sites that were mapped by male vs. female genomic reads is

613 plotted against the sequencing depth of this scaffold. Scaffold sequences are clustered

614 separately by their derived chromosomes with W-linked (red circles) and Z-linked (blue circles)

615 sequences showing the expected half the autosome (green) sequencing depth, and W-linked

616 sequences showing almost no mappable sites from male reads. **B)** The lengths of Z and W

617 chromosomes across the studied songbird species. The shorter length of raggiana BOP W

618 chromosome is probably caused by the low sequencing coverage. Species name marked in

619 blue are those that have male reads available for verifying the female-specificity of W-linked

620 sequences.

621

622 **Figure 2. Evolution strata of songbirds.**

623   **A)** Genomic synteny of the avian Z chromosomes. Each color represents one evolutionary

624   stratum of songbirds which does not apply to chicken or emu, as they have independent origins

625   of evolutionary strata except for S0. The location of *DMRT1*, the avian male-determining gene is

626   marked by the red dashed line. Generally, the synteny is more conserved in younger strata

627   between species. **B)** We use Lawes's parotia as an example to demonstrate the pairwise

628   sequence similarity pattern of evolutionary strata. The size of circles is scaled to the length of

629   sequence alignments between Z/W chromosomes. **C)** Transposable elements (LINEs and LTRs)

630   are more strongly enriched in older strata (S0 is the first stratum) except for LINEs at S3. Levels

631   of significance comparing neighboring strata are tested by Wilcoxon test and shown with

632   asterisks: '***': $P<0.001$, '**': $P<0.01$, '*': $P<0.05$. **D)** Lineage-specific burst of CR1-E1 (a

633   subfamily of CR-1 LINEs, red line) at the boundary of the PAR and S3 in songbirds, since their

634   divergence with other passerine species. Other subfamilies of CR1 elements are also plotted

635   with the green line for comparison. **E)** Phylogenetic tree using Z- and W-linked gametolog

636   sequences of the gene *C18orf25* located at S3. Lineages are clustered by chromosomes (red or

637   blue), not by species, suggesting S3 independently formed in rifleman, chicken and the ancestor

638   of songbirds.

639

640   **Figure 3. Fast-Z evolution of songbirds.**

641   We show the difference of evolutionary rates between Z-linked genes vs. autosomal genes (Z/A

642   value), as a measurement of fast-Z effect throughout the lineages of studied songbird species.

643   The tree length and color is scaled to the Z/A value, with lineages that show a significant

644   (permutation test, $P < 0.05$) fast-Z pattern labelled with asterisks. '***': $P<0.001$, '**': $P<0.01$, '*':

645    $P<0.05$. We also labelled their information mating systems ('monogamy' vs. 'polygamy'), and

646   male display type[54] ('lekking', 'exploded lekking', 'solitary display').

647

648   **Figure 4. W-linked genes are preserved by purifying selection.**

649   **A)** We show the retained W-linked genes of each studied songbird species, as well as those of

650   chicken, with homologous genes aligned vertically. The order of genes follows that of their emu

651   homologs along the Z chromosome. The colors represent the evolutionary strata among

652   songbirds. **B)** The Z-linked genes without W-linked homologs (green, 'Lost') evolve faster than

653   those with W-linked homologs retained (red, 'Retained'), as indicated by their branch lengths

654   scaled to dN/dS ratios. **C)** The Z-linked genes whose W-linked homologs have become lost

655   (upper panel) tend to have a higher expression level (measured by TPM) in their lizard orthologs

656   than those with W-linked homologs retained (lower panel). The genes are divided further by the

20

657   stratum they reside on, and the expression level is shown by log-transformed medium

658   expression values of each category as color-coded heatmap. **D)** Gene expression tissue

659   specificity in green anole lizard for the homologous avian Z-linked genes.

660

661   **Figure 5. Comparison of gene loss between W chromosomes of songbirds and Y**

662   **chromosomes of primates.**

663   **A)** We show the percentage of gene loss, and ancestral gene number for each evolutionary

664   stratum at each phylogenetic node. **B)** Similar analyses for the Y-linked gametologs of primates

665   based on the data of [21], with S1 as the first stratum of eutherian mammals. **C)** We show the

666   length of songbird W or primate Y chromosomes scaled to the ancestral gene number of each

667   evolutionary stratum, with the color scaled to the overall percentage of gene loss. The ages of

668   evolutionary strata are indicated by the number (in millions of years) at the nodes below the

669   bars. As eutherian mammals have much larger ancestral evolutionary strata than those of birds,

670   they probably suffer a more severe gene loss on the Y chromosome.

671

672   **Reference**

673   1.   IOC World Bird List 8.2.
674   2.   Barker, F.K., Cibois, A., Schikler, P., Feinstein, J. & Cracraft, J. Phylogeny and
675   diversification of the largest avian radiation. *Proc Natl Acad Sci U S A* **101**, 11040-5
676   (2004).
677   3.   Ellegren, H. *et al.* The genomic landscape of species divergence in Ficedula flycatchers.
678   *Nature* **491**, 756-760 (2012).
679   4.   Poelstra, J.W. *et al.* The genomic landscape underlying phenotypic integrity in the face
680   of gene flow in crows. *Science* **344**, 1410-1414 (2014).
681   5.   Tuttle, E.M. *et al.* Divergence and functional degradation of a sex chromosome-like
682   supergene. *Curr Biol* **26**, 344-50 (2016).
683   6.   Laine, V.N. *et al.* Evolutionary signals of selection on cognition from the great tit genome
684   and methylome. *Nat Commun* **7**, 10474 (2016).
685   7.   Mank, J.E. Sex chromosomes and the evolution of sexual dimorphism: lessons from the
686   genome. *Am Nat* **173**, 141-50 (2009).
687   8.   Coyne, J.A. Genetics and speciation. *Nature* **355**, 511 (1992).
688   9.   Charlesworth, B., Coyne, J. & Barton, N. The relative rates of evolution of sex
689   chromosomes and autosomes. *Am Nat* **130**, 113-146 (1987).
690   10.   Smeds, L. *et al.* Evolutionary analysis of the female-specific avian W chromosome. *Nat.*
691   *Commun.* **6**, 7330 (2015).
692   11.   Hooper, D.M. & Price, T.D. Chromosomal inversion differences correlate with range
693   overlap in passerine birds. *Nat Ecol Evol* **1**, 1526-1534 (2017).
694   12.   Backström, N. *et al.* A high-density scan of the Z chromosome in Ficedula flycatchers
695   reveals candidate loci for diversifying selection. *Evolution* **64**, 3461-3475 (2010).
696   13.   Elgvin, T.O. *et al.* Hybrid speciation in sparrows II: a role for sex chromosomes? *Mol*
697   *Ecol* **20**, 3823-37 (2011).

698 14. Storchová, R., Reif, J. & Nachman, M.W. Female heterogamety and speciation: reduced
699   introgression of the Z chromosome between two species of nightingales. *Evolution* **64**,
700   456-471 (2010).
701 15. Ellegren, H. *et al.* The genomic landscape of species divergence in Ficedula flycatchers.
702   *Nature* **491**, 756-60 (2012).
703 16. Li, W.H., Yi, S. & Makova, K. Male-driven evolution. *Curr Opin Genet Dev* **12**, 650-6
704   (2002).
705 17. Vicoso, B. & Charlesworth, B. Effective population size and the faster-X effect: an
706   extended model. *Evolution* **63**, 2413-26 (2009).
707 18. Mank, J.E., Nam, K. & Ellegren, H. Faster-Z evolution is predominantly due to genetic
708   drift. *Mol Biol Evol* **27**, 661-670 (2010).
709 19. Wright, A.E. *et al.* Variation in promiscuity and sexual selection drives avian rate of
710   Faster-Z evolution. *Mol Ecol* **24**, 1218-35 (2015).
711 20. Zhou, Q. *et al.* Complex evolutionary trajectories of sex chromosomes across bird taxa.
712   *Science* **346**, 1246338 (2014).
713 21. Bellott, D.W. *et al.* Avian W and mammalian Y chromosomes convergently retained
714   dosage-sensitive regulators. *Nat Genet* **49**, 387-394 (2017).
715 22. Charlesworth, D., Charlesworth, B. & Marais, G. Steps in the evolution of heteromorphic
716   sex chromosomes. *Heredity* **95**, 118-28 (2005).
717 23. Bachtrog, D. Y-chromosome evolution: emerging insights into processes of Y-
718   chromosome degeneration. *Nat Rev Genet* **14**, 113-24 (2013).
719 24. Soh, Y.Q. *et al.* Sequencing the mouse Y chromosome reveals convergent gene
720   acquisition and amplification on both sex chromosomes. *Cell* **159**, 800-13 (2014).
721 25. Charlesworth, B. & Charlesworth, D. The degeneration of Y chromosomes. *Philos Trans
722   R Soc Lond B Biol Sci* **355**, 1563-72 (2000).
723 26. Bachtrog, D. The temporal dynamics of processes underlying Y chromosome
724   degeneration. *Genetics* **179**, 1513-25 (2008).
725 27. Bellott, D.W. *et al.* Mammalian Y chromosomes retain widely expressed dosage-
726   sensitive regulators. *Nature* **508**, 494-499 (2014).
727 28. Zhou, Q. & Bachtrog, D. Sex-specific adaptation drives early sex chromosome evolution
728   in Drosophila. *Science* **337**, 341-345 (2012).
729 29. Koerich, L.B., Wang, X., Clark, A.G. & Carvalho, A.B. Low conservation of gene content
730   in the Drosophila Y chromosome. *Nature* **456**, 949-51 (2008).
731 30. Smeds, L. *et al.* Evolutionary analysis of the female-specific avian W chromosome. *Nat
732   Commun* **6**, 7330 (2015).
733 31. Cortez, D. *et al.* Origins and functional evolution of Y chromosomes across mammals.
734   *Nature* **508**, 488-93 (2014).
735 32. Bergero, R., Forrest, A., Kamau, E. & Charlesworth, D. Evolutionary strata on the X
736   chromosomes of the dioecious plant Silene latifolia: evidence from new sex-linked genes.
737   *Genetics* **175**, 1945-1954 (2007).
738 33. Lahn, B.T. & Page, D.C. Four evolutionary strata on the human X chromosome. *Science*
739   **286**, 964-7 (1999).
740 34. Gorelick, R. *et al.* Abrupt shortening of bird W chromosomes in ancestral Neognathae.
741   *Biol. J. Linn. Soc. Lond.* **119**, 488-496 (2016).
742 35. Rice, W.R. The accumulation of sexually antagonistic genes as a selective agent
743   promoting the evolution of reduced recombination between primitive sex chromosomes.
744   *Evolution* **41**, 911-914 (1987).
745 36. Zhang, G. *et al.* Comparative genomics reveals insights into avian genome evolution and
746   adaptation. *Science* **346**, 1311-1320 (2014).
747 37. Frankl-Vilches, C. *et al.* Using the canary genome to decipher the evolution of hormone-
748   sensitive gene regulation in seasonal singing birds. *Genome Biol* **16**, 19 (2015).

749 38. Qu, Y. *et al.* Ground tit genome reveals avian adaptation to living at high altitudes in the
750   Tibetan plateau. *Nat Commun* **4**, 2071 (2013).
751 39. Elgvin, T.O. *et al.* The genomic mosaicism of hybrid speciation. *Sci Adv* **3**, e1602996
752   (2017).
753 40. Claramunt, S. & Cracraft, J. A new time tree reveals Earth historys imprint on the
754   evolution of modern birds. *Science Advances* **1**, e1501005-e1501005 (2015).
755 41. Smeds, L. *et al.* Genomic identification and characterization of the pseudoautosomal
756   region in highly differentiated avian sex chromosomes. *Nat Commun* **5**, 5448 (2014).
757 42. Pala, I. *et al.* Evidence of a neo-sex chromosome in birds. *Heredity* **108**, 264-72 (2012).
758 43. Suh, A., Smeds, L. & Ellegren, H. Abundant recent activity of retrovirus-like
759   retrotransposons within and among flycatcher species implies a rich source of structural
760   variation in songbird genomes. *Mol Ecol* **27**, 99-111 (2018).
761 44. Kapusta, A. & Suh, A. Evolution of bird genomes-a transposon's-eye view. *Ann N Y*
762   *Acad Sci* **1389**, 164-185 (2017).
763 45. Gorelick, R. *et al.* Abrupt shortening of bird W chromosomes in ancestral Neognathae.
764   *Biological Journal of the Linnean Society* **119**, 488-496 (2016).
765 46. Ross, M.T. *et al.* The DNA sequence of the human X chromosome. *Nature* **434**, 325-37
766   (2005).
767 47. Poelstra, J.W. *et al.* The genomic landscape underlying phenotypic integrity in the face
768   of gene flow in crows. *Science* **344**, 1410-4 (2014).
769 48. Wright, A.E., Dean, R., Zimmer, F. & Mank, J.E. How to make a sex chromosome. *Nat*
770   *Commun* **7**, 12087 (2016).
771 49. Duret, L. & Galtier, N. Biased gene conversion and the evolution of mammalian genomic
772   landscapes. *Annu Rev Genomics Hum Genet* **10**, 285-311 (2009).
773 50. Kent, T.V., Uzunović, J. & Wright, S.I. Coevolution between transposable elements and
774   recombination. *Phil Trans R Soc B* **372**, 20160458 (2017).
775 51. Suh, A. *et al.* Mesozoic retroposons reveal parrots as the closest living relatives of
776   passerine birds. *Nat Commun* **2**, 443 (2011).
777 52. Friocourt, F. *et al.* Recurrent DCC gene losses during bird evolution. *Sci Rep* **7**, 37569
778   (2017).
779 53. Patthey, C., Tong, Y.G., Tait, C.M. & Wilson, S.I. Evolution of the functionally conserved
780   DCC gene in birds. *Sci Rep* **7**, 42029 (2017).
781 54. Irestedt, M., Jonsson, K.A., Fjeldsa, J., Christidis, L. & Ericson, P.G. An unexpectedly
782   long history of sexual selection in birds-of-paradise. *BMC Evol Biol* **9**, 235 (2009).
783 55. Diamond, J. Biology of birds of paradise and bowerbirds. *Ann Rev Ecol Syst* **17**, 17-37
784   (1986).
785 56. Kirkpatrick, M. & Ryan, M.J. The evolution of mating preferences and the paradox of the
786   lek. *Nature* **350**, 33-38 (1991).
787 57. Suh, A. The Specific Requirements for CR1 Retrotransposition Explain the Scarcity of
788   Retrogenes in Birds. *J. Mol. Evol.* **81**, 18-20 (2015).
789 58. Stock, M. *et al.* Ever-young sex chromosomes in European tree frogs. *PLoS Biol* **9**,
790   e1001062 (2011).
791 59. Vicoso, B., Emerson, J.J., Zektser, Y., Mahajan, S. & Bachtrog, D. Comparative sex
792   chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global
793   dosage compensation. *PLoS Biol* **11**, e1001643 (2013).
794 60. Charlesworth, D. Evolution of recombination rates between sex chromosomes. *Philos*
795   *Trans R Soc Lond B Biol Sci* **372**(2017).
796 61. Lemaitre, C. *et al.* Footprints of inversions at present and past pseudoautosomal
797   boundaries in human sex chromosomes. *Genome Biol Evol* **1**, 56-66 (2009).

798 62.   Knief, U. *et al.* Fitness consequences of polymorphic inversions in the zebra finch
799   genome. *Genome Biol* **17**, 199 (2016).
800 63.   Campagna, L. *et al.* Repeated divergent selection on pigmentation genes in a rapid finch
801   radiation. *Sci Adv* **3**, e1602404 (2017).
802 64.   Moghadam, H.K., Pointer, M.A., Wright, A.E., Berlin, S. & Mank, J.E. W chromosome
803   expression responds to female-specific selection. *Proc Natl Acad Sci U S A* **109**, 8207-
804   11 (2012).
805 65.   Peona, V., Weissensteiner, M.H. & Suh, A. How complete are "complete" genome
806   assemblies?—An avian perspective. *Mol Ecol Resour.* **0**.
807 66.   Amundsen, T. Why are female birds ornamented? *Trends Ecol Evol* **15**, 149-155 (2000).
808 67.   Irwin, D.E. Sex chromosomes and speciation in birds and other ZW systems. *Mol Ecol*
809   **27**, 3831-3851 (2018).
810 68.   Gnerre, S. *et al.* High-quality draft assemblies of mammalian genomes from massively
811   parallel sequence data. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 1513-1518 (2011).
812 69.   Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de
813   novo assembler. *Gigascience* **1**, 18 (2012).
814 70.   Cantarel, B.L. *et al.* MAKER: an easy-to-use annotation pipeline designed for emerging
815   model organism genomes. *Genome Res* **18**, 188-196 (2008).
816 71.   Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
817 72.   Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron
818   submodel. *Bioinformatics* **19**, ii215-ii225 (2003).
819 73.   http://www.repeatmasker.org.
820 74.   Prost, S. *et al.* Genome analyses show strong selection on coloration, morphological and
821   behavioral phenotypes in birds-of-paradise. *bioRxiv*, 287086 (2018).
822 75.   Marçais, G. *et al.* MUMmer4: A fast and versatile genome alignment system. *PLoS*
823   *Comput Biol* **14**, e1005944 (2018).
824 76.   Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
825   (2013).
826 77.   Singhal, S. *et al.* Stable recombination hotspots in birds. *Science* **350**, 928-32 (2015).
827 78.   Yazdi, H.P. & Ellegren, H. A genetic map of ostrich Z chromosome and the role of
828   inversions in avian sex chromosome evolution. *Genome Biol Evol* **10**, 2049-2060 (2018).
829 79.   Zhang, J., Li, C., Zhou, Q. & Zhang, G. Improving the ostrich genome assembly using
830   optical mapping data. *GigaScience* **4**, 24 (2015).
831 80.   Harris, R.S. The Pennsylvania State University (2017).
832 81.   Zhan, X. *et al.* Peregrine and saker falcon genome sequences provide insights into
833   evolution of a predatory lifestyle. *Nat Genet* **45**, 563-6 (2013).
834 82.   Hancock, J.M. BLAT (BLAST-like Alignment Tool). in *Dictionary of Bioinformatics and*
835   *Computational Biology* (2004).
836 83.   Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high
837   throughput. *Nucleic Acids Res* **32**, 1792-1797 (2004).
838 84.   Lechner, M. *et al.* Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC*
839   *Bioinformatics* **12**, 124 (2011).
840 85.   Jetz, W., Thomas, G.H., Joy, J.B., Hartmann, K. & Mooers, A.O. The global diversity of
841   birds in space and time. *Nature* **491**, 444-448 (2012).
842 86.   Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-
843   1591 (2007).
844 87.   Peden, J. http://codonw.sourceforge.net//culong.html.
845 88.   Nguyen, L.-T., Schmidt, H.A., von Haeseler, A. & Minh, B.Q. IQ-TREE: a fast and
846   effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol*
847   *Evol* **32**, 268-274 (2015).

848   89.   Jiao, X. *et al.* DAVID-WS: a stateful web service to facilitate gene/protein list analysis.
849   *Bioinformatics* **28**, 1805-1806 (2012).
850   90.   Li, B. & Dewey, C.N. RSEM: accurate transcript quantification from RNA-Seq data with
851   or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
852   91.   Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21
853   (2013).
854   92.   Yanai, I. *et al.* Genome-wide midrange transcription profiles reveal expression level
855   relationships in human tissue specification. *Bioinformatics* **21**, 650-9 (2005).
856

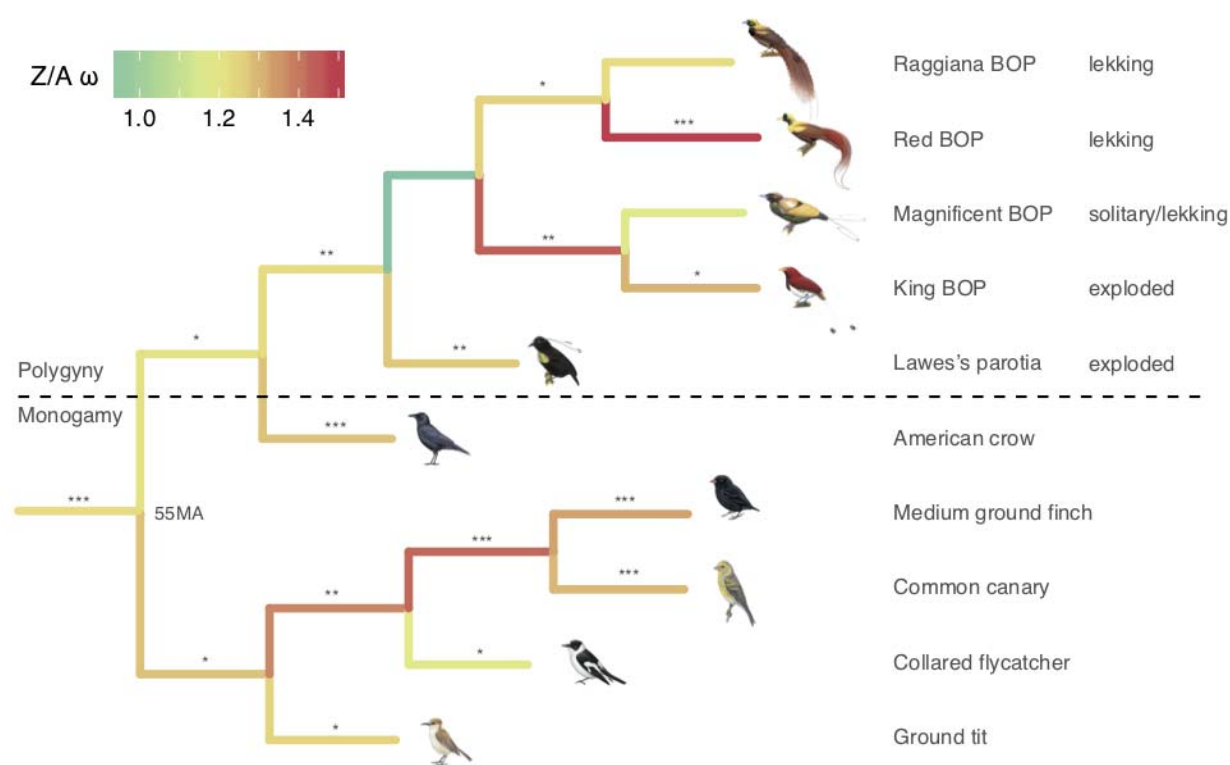857   **Figure 1. The Z and W chromosomes of different songbirds.**
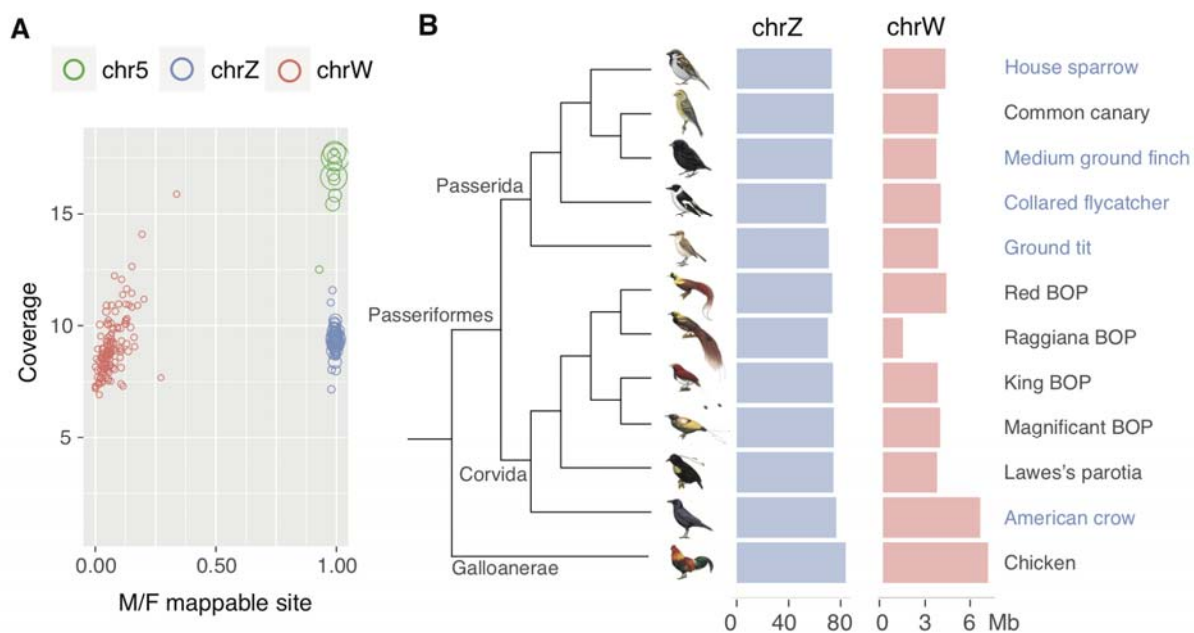
858

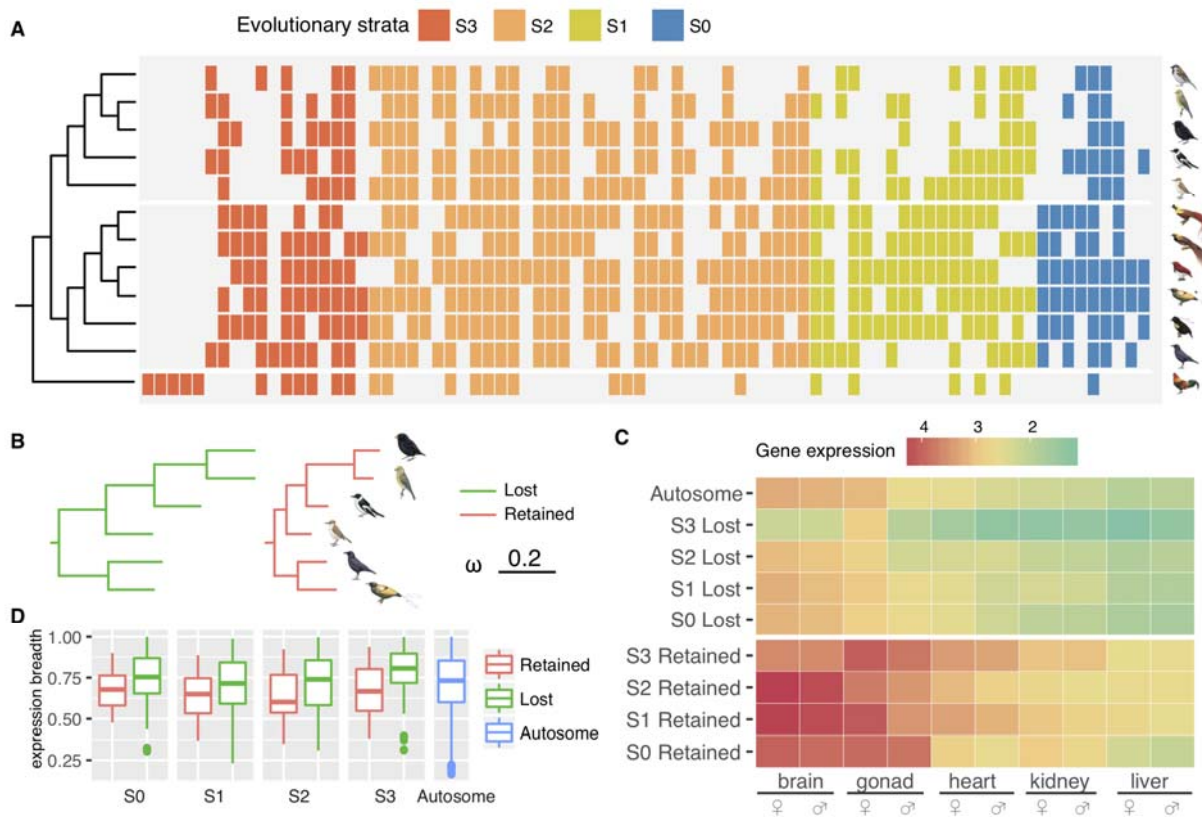859 **Figure 2. Evolutionary strata of songbirds.**
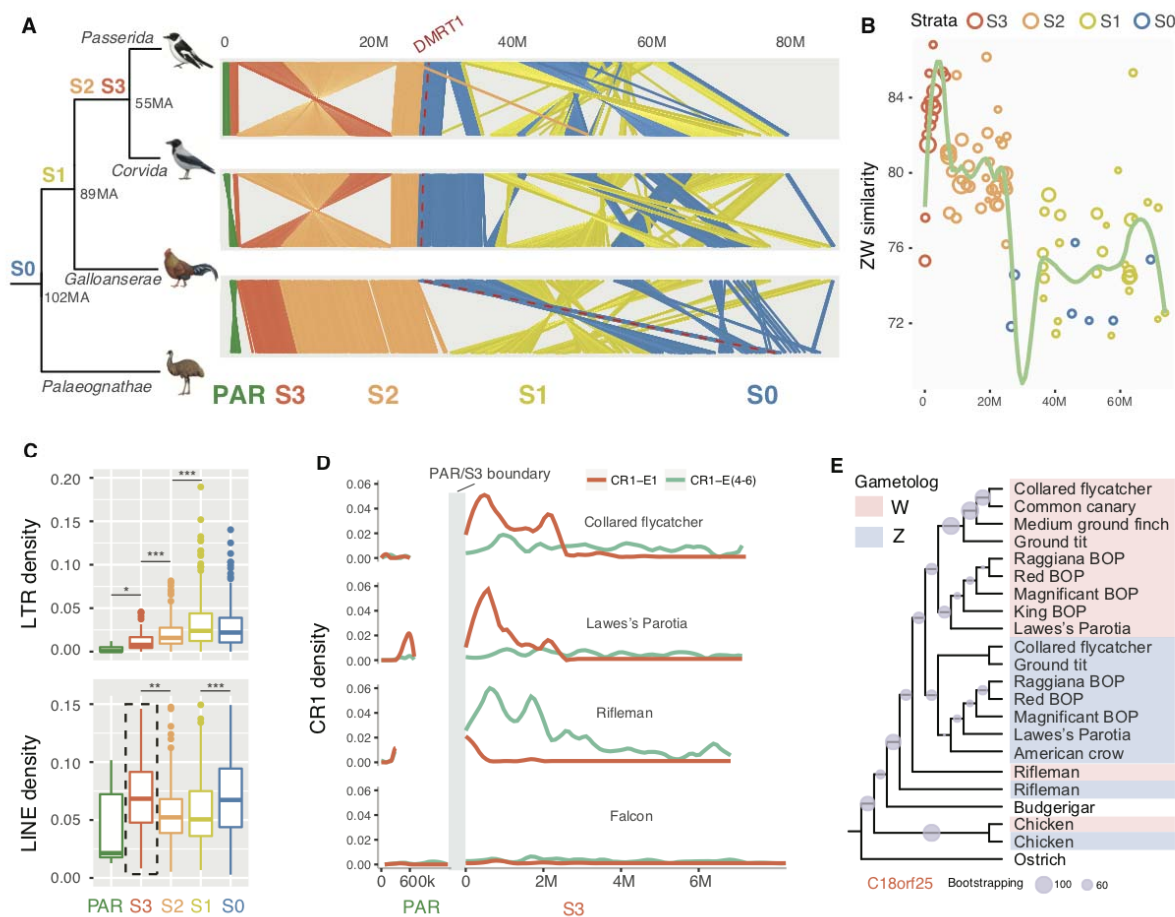
860

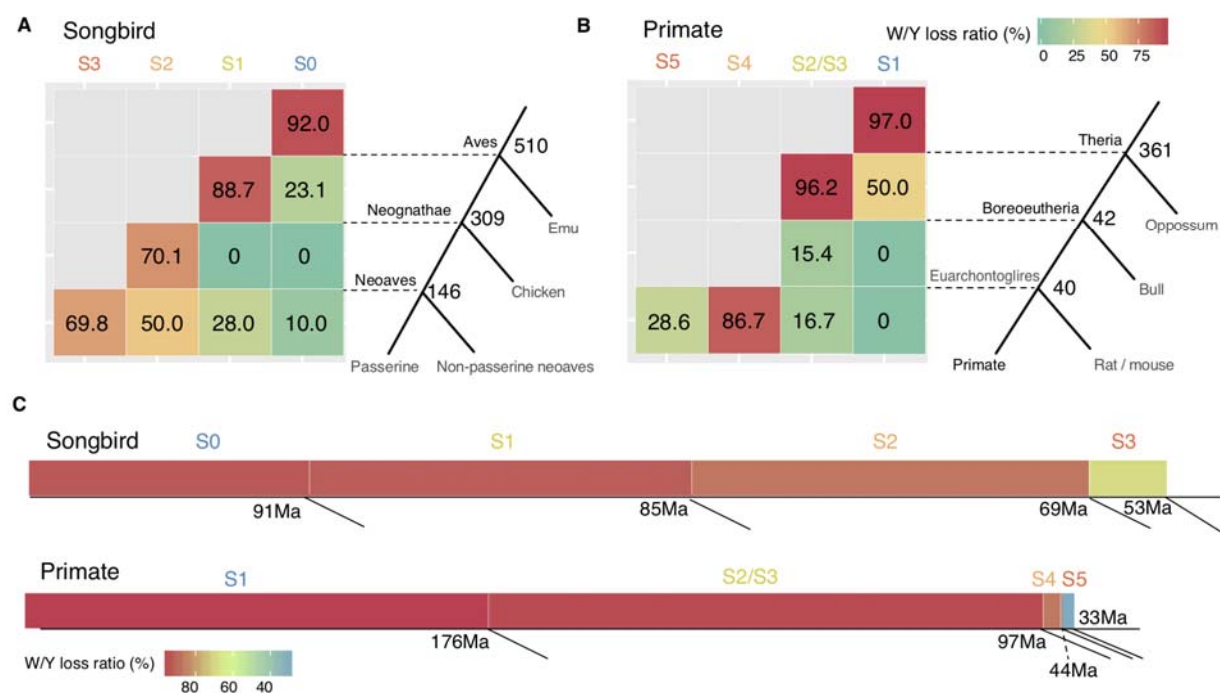861   **Figure 3. Fast-Z evolution of songbirds**

862

863 **Figure 4. W-linked genes are are preserved by purifying selection.**

864

Figure 5. Comparison of gene loss between W chromosomes of songbirds and Y

chromosomes of primates.