RESEARCH ARTICLE SUMMARY

AVIAN GENOMICS

Complex evolutionary trajectories of sex chromosomes across bird taxa

Qi Zhou,* Jilin Zhang, Doris Bachtrog, Na An, Quanfei Huang, Erich D. Jarvis, M. Thomas P. Gilbert, Guojie Zhang*

INTRODUCTION: Sex chromosomes originate from ordinary autosomes. Ancient sex-specific W or Y chromosomes, like that of female chicken or those of male mammals, usually have lost most functional genes, owing to a loss of recombination with their former homologs, the Z or X chromosomes. Such a recombination restriction occurred in a stepwise manner along most of the sex chromosomes in parallel in birds and mammals (creating so-called "evolutionary strata"), apart from small pseudoautosomal regions (PARs) that maintain recombination. Sex chromosomes of some basal birds like ostrich and emu have exceptionally large recombining PARs, but little is known about the genomic composition of most bird species' sex chromosomes.

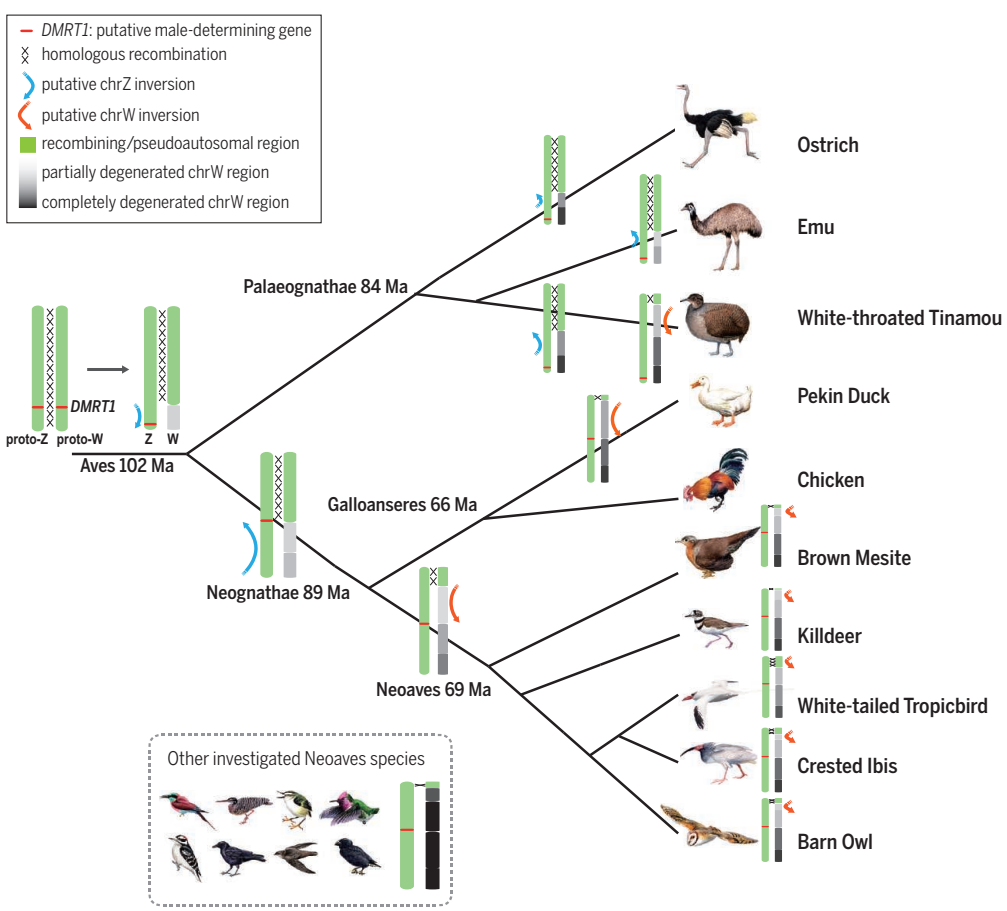
RATIONALE: Here we use the newly available genomes of 17 species spanning the entire avian phylogeny to decipher the genomic architecture and evolutionary history of bird sex chromosomes. We demarcate the PAR and the nonrecombining differentiated region between Z/W of each species by their different read depths relative to autosomes. We further assemble numerous W-linked genomic regions, whose abundance and sequence divergence level with Z chromosome reflect their ages of recombination loss.

RESULTS: Surprisingly, we find that more than half of the studied species have a W chromosome that is not completely degenerated. Besides ostrich and emu, some Neognathae species like tropicbird and killdeer also have long PARs. The nonrecombining regions between Z/W of many species exhibit a complex pattern of "evolutionary strata," resulting from the suppression of recombination in a stepwise and independent manner among some lineages. We conclude that the first evolutionary stratum that contains the putative male-determining gene *DMRT1* formed through a Z-linked chromosome inversion in an ancestor of all birds. This was followed by one stratum formed in the ancestor of Neognathae, one stratum in the ancestor of Neoaves, and independent emergence of more recent strata in most species. Many W-linked genes have disrupted protein function or reduced gene expression, and the rate of functional decay significantly slows down in older strata.

CONCLUSION: Our study uncovered an unexpected complexity of avian sex chromosomes, due to the lineage-specific recombination suppressions and different tempo of W degeneration. In contrast to mammals, some birds never experienced global recombination arrest, or differentiate at a very low rate between Z/W even after the recombination loss. This may relate to different intensities of sexual selection across bird species and explain their lack of a general chromosome-wide dosage compensation mechanism. ■

ON OUR WEB SITE

Read the full article at http://dx.doi.org/10.1126/science.1246338



Evolutionary strata of avian sex chromosomes. Bird sex chromosomes suppressed homologous recombination through shared or lineage-specific chromosomal inversions. Each defines an "evolutionary stratum" that exhibits a distinctive level of Z/W sequence divergence from neighboring strata. We demarcated and dated each stratum and show that bird sex chromosomes harbor great diversity regarding the lengths of recombining regions (green bars with crosses), or the degree of Z/W differentiation (scaled from gray to black).

The list of author affiliations is available in the full article online.
*Corresponding author. E-mail: zhouqi@berkeley.edu (Q.Z.); zhanggj@genomics.org.cn (G.Z.)
Cite this article as Q. Zhou *et al.*, *Science* **346**, 1246338 (2014). DOI: 10.1126/science.1246338

RESEARCH ARTICLE

Complex evolutionary trajectories of sex chromosomes across bird taxa

Qi Zhou,[1*†] Jilin Zhang,[2*] Doris Bachtrog,[1*] Na An,[2] Quanfei Huang,[2] Erich D. Jarvis,[3] M. Thomas P. Gilbert,[4,5] Guojie Zhang[2,6†]

Sex-specific chromosomes, like the W of most female birds and the Y of male mammals, usually have lost most genes owing to a lack of recombination. We analyze newly available genomes of 17 bird species representing the avian phylogenetic range, and find that more than half of them do not have as fully degenerated W chromosomes as that of chicken. We show that avian sex chromosomes harbor tremendous diversity among species in their composition of pseudoautosomal regions and degree of Z/W differentiation. Punctuated events of shared or lineage-specific recombination suppression have produced a gradient of "evolutionary strata" along the Z chromosome, which initiates from the putative avian sex-determining gene *DMRT1* and ends at the pseudoautosomal region. W-linked genes are subject to ongoing functional decay after recombination was suppressed, and the tempo of degeneration slows down in older strata. Overall, we unveil a complex history of avian sex chromosome evolution.

In many species with separate sexes, sex is determined by a pair of heteromorphic sex chromosomes that differ in their size, morphology, and gene content. Mammals have male heterogametic sex chromosomes (XX in female, XY in male), whereas birds have female heterogametic sex chromosomes (ZW in female and ZZ in male) (*1*). Both sex systems have originated independently from different ancestral autosomes after the two lineages diverged more than 300 million years ago (Ma) (*1*–*4*), yet they share many common features with respect to their genomic composition and evolutionary history. Most regions of mammalian Y or avian W chromosomes (chrW) lack recombination except for the pseudoautosomal regions (PARs) that are shared between X and Y, or Z and W, and which encompass only 2% of the human X (*5*). The nonrecombining region of the Y/W usually has lost almost all functional genes present on the ancestral chromosome except for a few loci, often with sex-specific functions (*6*), and Y/W chromosomes are often largely composed of repetitive and epigenetically silenced (heterochromatic) DNA (*7*). Recombination suppression is thought to follow the origination of a sex-determining locus and may initially encompass only a small chromosomal segment surrounding that locus, but then progressively spreads along Y or W chromosomes, possibly owing to the accumulation of sexually antagonistic alleles at PARs that benefit one sex but harm the other (*8*). Chicken and human sex chromosomes have in parallel evolved recombination suppression in a stepwise manner, manifested as "evolutionary strata" with a gradient of ages. Each stratum is characterized by a distinct range of levels of sequence homology between the remaining XY or ZW gene pairs, reflecting the punctuated evolutionary time points when recombination between the sex chromosomes ceased (*9*, *10*). To date, genomic sequences of 10 mammalian species' Y chromosomes have been determined (*11*, *12*), and the human Y was found to have five strata (*5*) probably resulting from Y-linked inversions (*13*), of which three are shared among all Eutherian mammals (*4*, *14*). No complete avian chrW has been sequenced, but partial annotation of W-linked genes in chicken (*Gallus gallus*) (*15*) implied that there are at least three strata (*10*, *16*), and some may have formed independently in different avian lineages (*17*–*20*).

A recent investigation of the karyotype of 200 avian species revealed great variation in the size and morphology of chrW, sometimes even between closely related species (*21*), suggesting a dynamic evolutionary history of bird sex chromosomes. Living birds consist of two sister lineages, the Palaeognathae (flightless ratites and Tinamous) and the Neognathae, which themselves consist of two major lineages, the Galloanserae (land and water fowls) and the Neoaves (all other neognaths). Neoaves are derived from an ancient species radiation and comprise more than 95% of all avian species (*22*, *23*). Karyotyping suggests that most ratite paleognaths, like ostrich (*Struthio camelus*) and emu (*Dromaius novaehollandiae*), have homomorphic sex chromosomes with extensive recombining and euchromatic PARs that resemble the ancestral state (*24*–*26*). In contrast, their sister group Tinamous shows interspecific variation in sizes and locations of PARs (*27*). Whether all Neognathae species have an almost completely degenerated chrW similar to that of chicken (*28*) is unknown. Here, we analyzed female genomes of 17 bird species newly generated by the Avian Genomics Consortium that span the avian taxonomic diversity (*23*, *29*), in order to unveil the broad evolutionary history of bird ZW chromosomes.

Results

Avian sex chromosomes have different lengths of PAR

We collected genome sequences of three Palaeognathae (ostrich, emu, white-throated tinamou), one Galloanserae (Pekin duck) (*30*), and 13 Neoaves species, including a representative of the sister lineage to all other Neoaves (brown mesite), the Caprimulgiformes (Anna's hummingbird and chimney swift), core waterbirds (crested ibis to white-tailed tropicbird), and core landbirds (rifleman to woodpecker) (*29*) (Fig. 1 and table S1). For three species—chicken, emu, and crested ibis—we analyzed sequence data of both sexes (*31*–*33*). We additionally improved the ostrich draft genome to a chromosome assembly with optical mapping (*34*) (N50: 17.7 Mb, table S2). Comparative cytogenetic mapping and genome sequence analysis show that there are very few interchromosomal rearrangements among bird species (*35*–*38*); thus, we anchored genome scaffold sequences of the remaining species using the ostrich Z chromosome (chrZ) for all Palaeognathae and chicken Z for all Neognathae (*3*).

To distinguish between the recombining PAR (that has nearly identical sequences between chrZ and chrW) and the region lacking recombination (that has become differentiated between chrZ and chrW) along the sex chromosome, we developed an approach that involves examination of read sequencing depth along chrZ (*34*). Although most sequences in the female-specific W-linked region may have lost almost all homology with chrZ, some ZW homolog pairs (often called gametologs) may remain. However, W-linked regions are probably too divergent to allow short sequencing reads derived from them [100 base pairs (bp) long in this study] to align with their Z homologs, especially if they are from more ancient strata. Consequently, PARs will show a level of read depth similar to that of autosomes, whereas differentiated regions will have only half the mapped read coverage depth in females (Fig. 1).

Based on this principle, we find evidence of extreme differences in the length of the PAR versus differentiated region on sex chromosomes across species (Fig. 1 and Table 1), in contrast to a lack of such variation in coverage from autosomes (fig. S1). The putative avian male sex-determining gene *DMRT1* (*39*) always resides in the differentiated region in all species, consistent with it being the initial seed for progressive recombination suppression (Fig. 1). However, the size of the PAR varies substantially among species (Table 1). In Palaeognathae, at least two-thirds of chrZ of ostrich and emu are pseudoautosomal and still

[1]Department of Integrative Biology, University of California, Berkeley, CA94720, USA. [2]China National Genebank, BGI-Shenzhen, Shenzhen, 518083. China. [3]Department of Neurobiology, Howard Hughes Medical Institute, Duke University Medical Center, Durham, NC 27710, USA. [4]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark. [5]Trace and Environmental DNA laboratory, Department of Environment and Agriculture, Curtin University, Perth, Western Australia 6102, Australia. [6]Centre for Social Evolution, Department of Biology, Universitetsparken 15, University of Copenhagen, DK-2100 Copenhagen, Denmark.
*These authors contributed equally to this work. †Corresponding author. E-mail: zhouqi@berkeley.edu (Q.Z.); zhanggj@genomics.org.cn (G.Z.)

recombine with chrW, in contrast to only 1% of chrZ of white-throated tinamou (*Tinamus guttatus*) being PAR (Fig. 1 and Table 1). We confirm that we have identified the true PAR in ostrich, as we find that eight genes that have been cytologically mapped to both chrZ and chrW (*26*) are appropriately distributed along our assembled ostrich chrZ (Fig. 1). Whereas most Neognathae have restricted recombination between their sex chromosomes into only a very short terminal PAR (Fig. 1 and fig. S2), unexpectedly two Neoavian species, killdeer (*Charadrius vociferus*) and white-tailed tropicbird (*Phaethon lepturus*), retain PARs of considerable length, encompassing between 5 and 9% of their chrZ (Table 1). Thus, unlike chicken (*40*), some Neognathae chrW can pair and recombine with chrZ through a sizable PAR. This also indicates that recombination suppression in the genomic segment adjacent to the PAR has evolved independently in at least some bird lineages (see below).

To further quantify Z/W divergence levels within the differentiated region, we searched for assembled W-linked sequences longer than 1 kb, which allow better alignments with chrZ than the raw reads. W gametologs may be embedded within repetitive DNA that cannot be readily assembled into large scaffolds, but unique segments of a species' chrW are expected to assemble into fragmented scaffold sequences separate from those of chrZ. If these W-segments are only differentiated to a moderate degree, they may have retained sufficient sequence homology to map to the differentiated region along chrZ. Using this approach, we identified candidate W-linked scaffolds up to 25% of the homologous chrZ length with low background mapping to autosomes, indicating effective discrimination from random fragments (*34*) (fig. S1). We identify abundant W-linked sequences from all three Palaeognathae species, one of the two Galloanserae species and in five of the Neoaves (Table 1 and table S3), and many fewer candidate W-derived fragments homologous to chrZs in chicken and the remaining Neoaves species (Fig. 1 and fig. S2).

Comparison of read depths between the two sexes in chicken, crested ibis, and emu verified that the candidate W-linked scaffolds are specific to females (figs. S3 and S4). Additionally, most of the candidate W-linked scaffolds of chicken exclusively align to W-linked or unmapped (chrUn) fragments of the reference chicken genome derived from Sanger sequencing (fig. S5). These results together confirmed that we have identified truly W-linked scaffolds. We further annotated genes in these W-derived scaffolds, ranging from 24 in the Pekin duck to 55 in ostrich and 198 in the white-throated tinamou (Table 1 and table S4), compared to 26 and 14 W-linked genes previously annotated through transcriptome sequencing in chicken and ostrich, respectively (figs. S6 and 7 and table S5) (*15*, *41*). Some W-linked genes may be missing or fragmented (table S5), if they are highly enriched for transposable elements or too similar to their Z-linked homologs in sequence. Nevertheless, we found the coding sequence lengths of the Z/W gene pairs identified in this work to generally correlate well with those of their chicken Z-linked orthologs (fig. S8). This indicates that, although our current chrW genome assemblies are interrupted by highly repetitive noncoding regions, they contain substantial amounts of coding regions of genes. Overall, we find that avian chrWs show great variation in their degree of degeneration, and more than one-third of the sampled Neognathae species harbor chrWs that are not as fully degenerated as that of chicken (fig. S2).

Evolutionary strata of Palaeognathae species

The identified W-linked fragments are stratified by differences in their occurrences and alignment identities along chrZ (figs. S9 and S10), exhibiting patterns of evolutionary strata. This provides evidence that, similar to those of mammals (*4*), avian sex chromosomes suppressed recombination

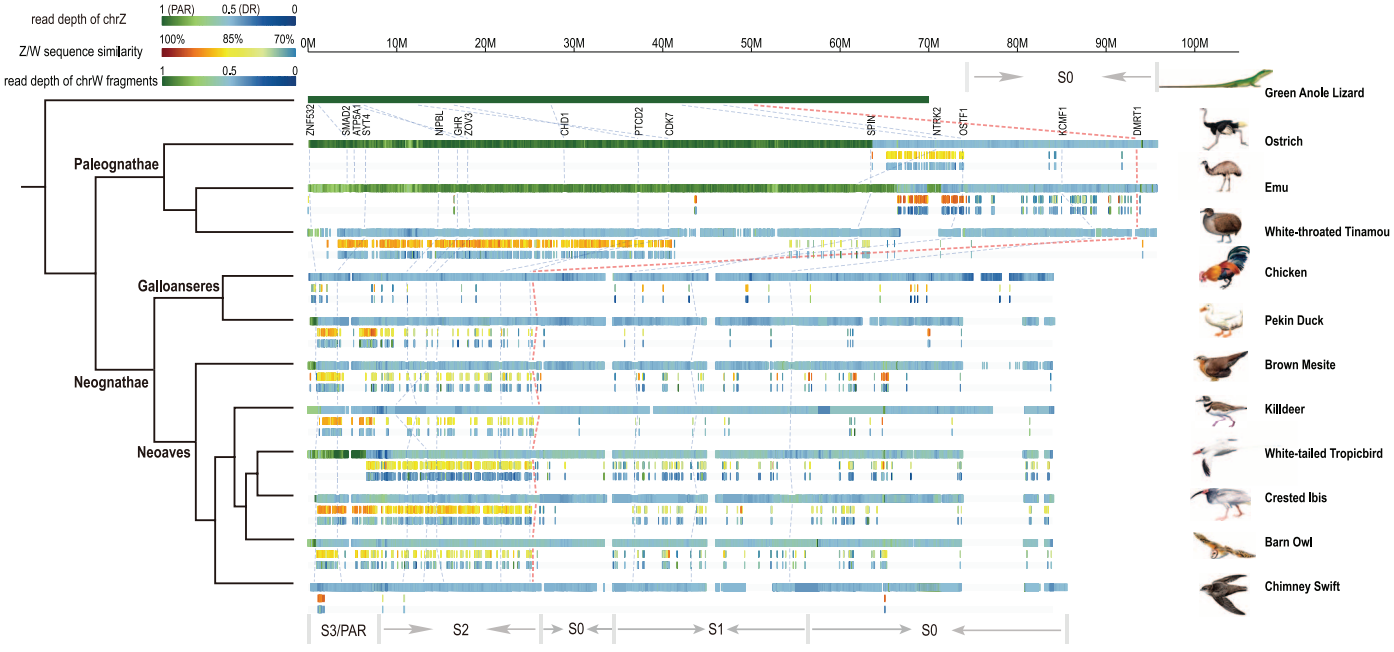


Fig. 1. Evolutionary strata of avian sex chromosomes. We present evolutionary strata patterns of investigated bird species with partially degenerated chrW, except for chicken and chimney swift as comparison, together with their revised phylogeny (*23*). A full pattern for all the investigated species is presented in fig. S2 and summarized in fig. S23. chrZs of tinamou and emu are ordered and scaled to the same length as that of ostrich, and all others to that of chicken. Each avian species includes three color-coded tracks from top to bottom: (i) female read depth of chrZ, where PAR is shown in green, differentiated region without recombination between Z/W in blue, and assembly gaps as blanks; (ii) pairwise alignment identities between Z/W plotted from 100% identical (red) to 70% identical (blue); and (iii) female read depth of W-linked sequence fragments, color-coded the same as (i). We labeled regions of each stratum at the bottom: "S0" corresponds to the most ancient evolutionary stratum 0 that is shared by all the species, while the younger strata are abbreviated as S1, S2, etc. Note that different species may not share the same S1, S2, and S3; and part of S0 (25 to 35 Mb in the figure) has been reshuffled because of chromosome inversions that occurred in the ancestor of Neognathae (Fig. 3 and fig. S15). The blank region around 75 to 85 Mb corresponds to a chicken-specific amplification of testis gene clusters (*3*) (table S7). We also labeled genes previously used for cytological mapping (*26*) and the reported S0 gene *KCMF1* (*4*) (traced through the phylogeny with blue dotted lines), as well as the putative male sex-determining gene *DMRT1* (traced with the dotted red line) (*39*).

through a series of punctuated events (*10*, *16*, *17*). The strata present a gradient of ages, reflected by higher levels of sequence homology between chrW and chrZ sequences as one moves toward the PAR (Fig. 1 and figs. S9 to S11). The putative male sex-determining gene of birds, *DMRT1* (*39*), is located within the first stratum (we denote this region as stratum S0) that evolved in the ancestor of all extant birds, around 100 (98 to 108) Ma (*23*). This is supported by a lack of abundant assembled W-linked fragments due to the ancient origin of S0, and also *DMRT1*'s close linkage in our new ostrich genome assembly to *KCMF1* (Fig. 1), which was also characterized as an S0 gene by an independent study through transcriptome sequencing (*4*). The boundaries of S0 can still be clearly traced by the nearby stratum in ostrich and emu (Fig. 1), both of which have maintained the most ancestral state of sex chromosome differentiation among the studied birds and share 90% of their Z-linked gene content of S0 (Fig. 2A).

Maximum-likelihood trees constructed with multiple S0 genes (including *KCMF1*) show that Z-linked gametologs of Palaeognathae and Neognathae cluster with each other by species rather than with their W-linked gametologs (fig. S12), confirming that recombination between chrZ and chrW at S0 was abolished in the common ancestor of all birds (*17*) (Fig. 2A and fig. S12). Consistent with their postulated ancient origin, almost no W-linked genes were identified within the orthologous region of S0 in all the studied species (Table 1), the exception being emu, whose W-linked S0 region has not completely degenerated (*23*, *31*). All residual W-linked fragments within emu S0 show similar levels of divergence from their homologous chrZ sequences (fig. S9), suggesting that they stopped recombining at the same time.

Adjacent to the S0 stratum, we identified one additional stratum (S1) in ostrich and emu, and at least two strata (S1 and S2) in the white-throated tinamou. In contrast to S0, S1 spans different regions in each of the three paleognaths: Each species shares less than 50% of its S1 gene content with the other two (Fig. 2B), and S2 is specific to tinamou, whose orthologous region in the ostrich and emu is PAR (Figs. 1 and 2C). This indicates that these younger strata formed independently in these three anciently diverged species. A small segment of the ostrich PAR is embedded within the inferred emu S1, which suggests a chromosomal rearrangement between these two birds (Fig. 1). Molecular clock dating with Z/W divergence levels at silent sites suggests that stratum S1 of ostrich formed about 75 (34 to 193) Ma and that of emu formed about 23 (11 to 59) Ma, and their sex chromosomes have maintained their current PARs ever since (table S6). White-throated tinamou has a much smaller PAR due to two separate strata that formed about 50 (21 to 95) Ma and 29 (7 to 82) Ma (table S6), after its divergence with ostrich about 58 to 95 Ma (*23*). Thus, common and lineage-specific events characterize sex chromosome evolution in Palaeognathae birds.

Evolutionary strata of Neognathae species

In comparison, Neognathae species exhibit either three or four strata along their sex chromosomes (Fig. 1). We infer that the oldest Neognathae-specific stratum S1 formed in an ancestor of all Neognathae adjacent to the S0 region described above (Fig. 1). We also recovered W-linked sequence fragments within this stratum by reordering them against the ostrich chrZ as a proxy for the proto-Z (fig. S13). Consistent with its ancestral emergence and previous results in chicken (*16*), Neognathae S1 harbors almost no identifiable W-linked fragments or genes in Galloanserae, and we find only few W-linked fragments with low sequence identities versus chrZ in Neoaves (figs. S9 and S10). The low density of W fragments within Neognathae S1 hampers us from determining the exact genomic boundaries of this stratum, but phylogenetic trees constructed from Z/W gametologs of S1 confirm that this stratum is shared among all Neognathae (Fig. 2D and fig. S14). Three Z/W gene pairs, *SPIN*, *HINT*, and *CHD1*, which were previously inferred to have stopped recombining before the Neoaves–Galloanserae split (*17*), are consistently included within the S1 region of Neognathae, despite the shuffling of gene order due to chromosomal rearrangements (fig. S15). By scaling the Z/W divergence level in noncoding regions to that of S2 (see below), we estimate that Neognathae S1 was formed about 71 to 119 Ma, which is consistent with its formation before the inferred split between Neoaves–Galloanserae about 83 to 94 Ma (*23*) (table S6).

Adjacent to S1, we identify a younger Neognathae stratum S2 that spans a similar region in all Neognathae (Fig. 1), but Z/W gametologs within this region cluster by lineage rather than by chromosomes (Fig. 2E and fig. S16), suggesting the Galliformes (chicken), Anseriformes (Pekin duck),

Table 1. Age and gene content of avian evolutionary strata.

Stratum	Species	Ostrich	Emu	White-throated tinamou	Pekin duck	Brown mesite	Killdeer	White-tailed tropicbird	Crested ibis	Barn owl
S0	Age (My)*					102				
	Z-linked genes	247	224	244	154	137	96	157	182	151
	W-linked genes†	0	6/8	0	0	1/1	0/0	2/3	5/4	1/2
	W/Z length (Mb)	0/21.4	1.2/19.8	0/18.3	0.01/32.5	0.2/18.4	0.1/31.9	0.8/20.9	0.8/28.5	0.3/19.4
S1‡	Age (My)	75	23	50			71–119			
	Z-linked genes	108	21	201	333	236	400	246	267	211
	W-linked genes	36/19	0/0	21/9	0/4	7/3	1/1	10/10	8/5	5/4
	W/Z length (Mb)	5.1/10.6	0.3/3.6	1.2/17.7	0.09/16.9	0.08/25.0	0.06/20.3	0.2/26.7	0.5/23.8	0.09/26.8
S2	Age (My)			29	92			69		
	Z-linked genes	PAR		461	237	256	279	224	229	203
	W-linked genes			94/74	14/9	23/8	27/11	37/19	33/32	12/10
	W/Z length (Mb)			16.6/35.1	0.5/17.7	1.6/19.8	2.2/21.2	5.0/19.1	6.3/20.0	1.7/17.8
S3	Age (My)				60	47	45		53	50
	Z-linked genes	PAR		PAR	96	21	24	PAR	25	36
	W-linked genes				13/5	7/3	3/3		2/1	5/2
	W/Z length (Mb)				1.3/8.3	1.5/3.7	2.4/4.7		0.6/3.2	1.1/6.2
PAR	Z-linked genes	718	619	12	18	19	37	55	8	13
	Z lengths (Mb)	63.6	65.5	0.7	1.05	0.7	4.2	6.6	0.4	0.9
Total	**W/Z genes**	55/1073	14/864	198/918	45/848	53/669	46/836	81/682	90/711	41/614
	W/Z lengths	5.1/95.6	1.6/88.9	17.9/71.9	1.9/76.5	3.4/67.7	4.7/82.4	6.0/73.2	8.3/75.9	3.3/71.1

*Ages (My, millions of years) were estimated based on Z/W noncoding sequence divergence level with the assumption of a molecular clock except for stratum S0 and Neoaves S2. Confidence intervals for the ages are shown in table S6. Note that strata with the same number are not necessarily shared among different species. † Number of genes with intact ORFs or disrupted ORFs on the W chromosome. ‡ Boundaries of stratum S0/S1 may not be accurate owing to the low number of identifiable W-linked sequences.

and Neoaves may have each independently formed stratum S2, after the shared origination of S1. On the basis of either the proposed divergence times between these lineages (*23*) or from Z/W sequence divergence level in noncoding regions, we estimate that the duck stratum S2 arose 68 (18 to 127) Ma and that of Neoaves arose 69 (66 to 72) Ma (table S6), whereas the chicken S2 was previously estimated to have arisen 34 to 54 Ma (*16*).

One Neoaves species, the white-tailed tropicbird, seems to have experienced no further recombination suppressions after stratum S2, and its current PAR appears to be comparable to the Neoaves ancestor PAR (Fig. 1). By contrast, orthologs of tropicbird PAR genes are located in differentiated regions of all the other Neoaves, including brown mesite (*Mesitornis unicolor*), suggesting that the youngest stratum S3 evolved independently in different lineages. Indeed, only five genes overlap between the different S3 regions (Fig. 2F), and Z-linked genes from S3 cluster with W-linked gametologs from the same species and not with Z-linked orthologs of other species (Fig. 2F and fig. S17). Two Z/W gene pairs within the S3 region, *ATP5A1* and *UBAP2*, were previously shown to have suppressed recombination independently among different bird species (*17*). Together, these results indicate that recombination in S3 ceased in some lineages after the Neoaves species radiation, and we estimate that S3 emerged at different time points in different lineages (between 45 and 53 Ma; Table 1 and table S6), by scaling its Z/W divergence level at noncoding regions to that of Neoaves S2.

Putative Z- and W-linked inversions causing recombination suppression

Recombination suppression at evolutionary strata may have been mediated by chromosomal inversions, as has been suggested for mammals (*9*, *13*). Genomic comparison of four avian species with de novo assembled genomes against the reptile outgroups green anole lizard (*Anolis carolinensis*) (*42*) and boa snake (*Boa constrictor*) (*43*) (Fig. 3 and fig. S18) allow us to infer the presence of inversions on chrZ and further inform on the

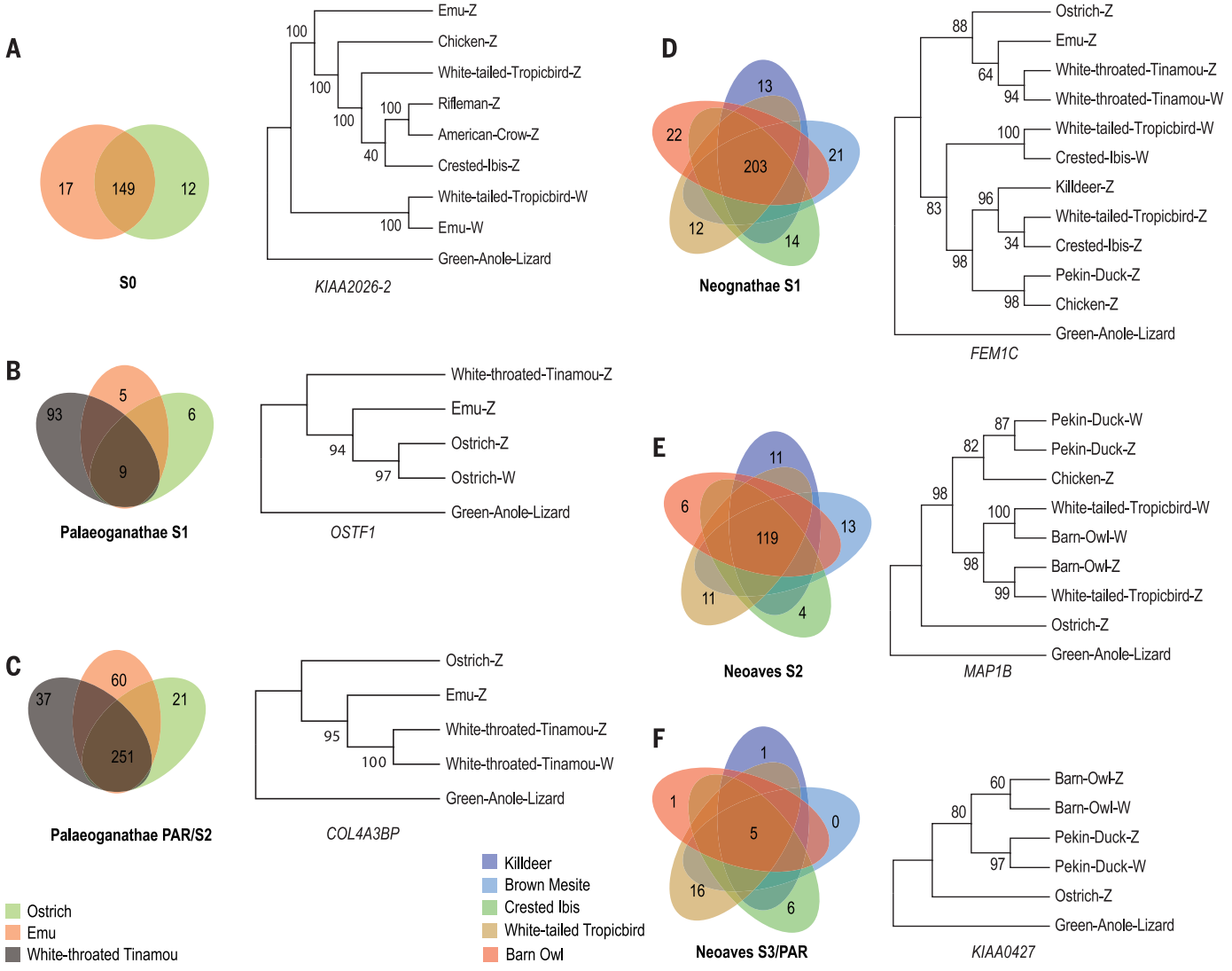


Fig. 2. Emergence of evolutionary strata in different avian lineages. Shown are Venn diagrams of overlapping gene content between orthologous strata of different Palaeognathae (**A** to **C**) and Neognathae (**D** to **F**) species. To the right of each diagram is a maximum-likelihood tree for a specific example gene constructed from Z- and W-linked gametologs of that stratum. Additional trees supporting the inference of a certain stratum's age are shown in the figs. S12, S14, S16, and S17. The tree is constructed with ostrich or green anole lizard as outgroup, with 100 bootstraps. A pattern in which Z gametologs are grouped together by species indicates that Z/W recombination was suppressed before their speciation, whereas clustering of Z and W gametologs of the same species indicates that Z/W recombination was suppressed after the speciation event. "Palaeognathae PAR" shows the number of genes that overlapped between ostrich PAR, emu PAR, and tinamou stratum S2. "Neoaves S3" shows the overlapped gene content between tropicbird PAR versus S3 of four other Neoaves species: killdeer, brown mesite, crested ibis, and barn owl.

formation paths of evolutionary strata in birds. Both cytogenetic mapping (*44*) and our genomic comparison have shown that ostrich has maintained the most ancestral gene synteny relative to other avian species with only a few intrachromosomal rearrangements (Fig. 3). We specifically compared gene synteny spanning the ostrich S0 with the assembled genomes of green anole lizard and boa snake and to cytogenetic data from three other reptiles (rat snake, gecko lizard, Chinese soft-shelled turtle) (*45*), and identified a chromosomal inversion on ostrich chrZ relative to reptiles, which are all collinear within their orthologous regions on the corresponding autosome (fig. S18). This suggests that an inversion on the proto-Z abolished recombination between chrZ and chrW at S0 in an ancestor of all birds. Gene content and synteny within the ostrich S1 region is largely conserved between the green anole lizard and ostrich, suggesting that a putative inversion creating the ostrich-specific S1 stratum must have occurred on chrW (Fig. 3).

Multiple genomic rearrangements within Neognathae S0 and S1 regions among Neognathae species preclude us from inferring the origin of Neognathae S1 (figs. S13 and S15). However, both cytogenetic (*35*) and genomic mapping (Fig. 3) place *DMRT1* close to the telomere of Palaeognathae chrZ but in the middle of chrZ for most other Neognathae species. This suggests that at least one Z-linked inversion, which may have been associated with the formation of Neognathae S1, has relocated *DMRT1* in the ancestor of all Neognathae. This putative inversion has shuffled part of the oldest stratum S0 containing *DMRT1* in between the younger strata Neoaves S2 and S1 in chicken (Fig. 3 and fig. S15), which explains why *DMRT1* was placed in a more recent stratum in a previous study using chicken Z chromosome coordinates as reference (fig. S15) (*16*). Gene synteny is to a large extent conserved across species within both S2 and S3 (Fig. 3 and figs. S18 and S19). This suggests that if inversions are responsible for recombination suppression between chrZ and chrW at S2 and S3, they originated on chrW.

Tempo of W degeneration in birds

Once recombination is suppressed between chrW and chrZ, the functions of W-linked genes are expected to decay (*6*). Y chromosome genome and transcriptome sequencing of 15 mammals has shown a loss of more than 95% of all genes hypothesized to have been present on the ancestral Y of each lineage (*4*, *12*). In contrast, we find extensive heterogeneity across birds in levels of W degeneration, as measured by the proportion of residual W-linked genes or sequences relative to their Z-linked homologs (Table 1). Heterogeneity in W degeneration is found not only between species groups whose strata evolved independently (e.g., Paleognathae versus Neognathae), but also within homologous strata dating back to a common ancestor, and which therefore have carried the same ancestral gene content and underwent degeneration for the same length of time (Fig. 1 and Table 1). This is exemplified by the W-linked region of Neoaves stratum S2 having become almost completely differentiated and lost nearly all genes in 8 out of 13 Neoaves species (fig. S2), whereas tropicbird and crested ibis have retained at least 20% of their ancestral W-linked genes relative to the homologous Z-linked regions (Table 1). Within each stratum, we estimate that chrWs of different species have lost 52 to 95% of their ancestral gene content (Fig. 4, A to G), which may be an overestimate owing to difficulties in assembling W gametologs within highly repetitive regions (figs. S6 and S7). Functional decay is also apparent at assembled W gametologs that are present in the nonrecombining region of chrW (*46*): We find that about 26 to 50% of W-linked genes within each stratum have likely become pseudogenes, owing to the accumulation of premature stop codons and/or frame-shift mutations (Fig. 4, A to G). In addition, W-linked genes in all species consistently show significantly higher rates of protein evolution than their Z-linked gametologs (comparing pairwise rates versus chicken; Wilcoxon test: $P < 0.05$, fig. S20), indicating that they experience less efficient purifying selection on the nonrecombining chrW than the Z gametologs (*46*, *47*). Disabling mutations may have also accumulated in regulatory regions: W-linked gametologs within the differentiated region of ostrich, emu, and duck are expressed at a significantly lower level than both their Z-linked gametologs and lizard orthologs (Wilcoxon test, $P < 0.05$, Fig. 4H and fig. S21) in different tissues, and W gametologs with disrupted open reading frames are down-regulated even more (fig. S22).

Older strata generally have lost more W-linked genes (Fig. 4, A to G), yet we find a significant negative correlation between the rate of gene loss (measured as percent of gene loss per million years) versus the age of stratum in birds ($R^2 = 0.87$, Pearson's test P-value $= 2.87 \times 10^{-12}$, Fig. 4I), similar to those seen in mammalian and chicken strata (*12*, *16*, *49*). Thus, older strata appear to lose genes at a lower rate in both clades,

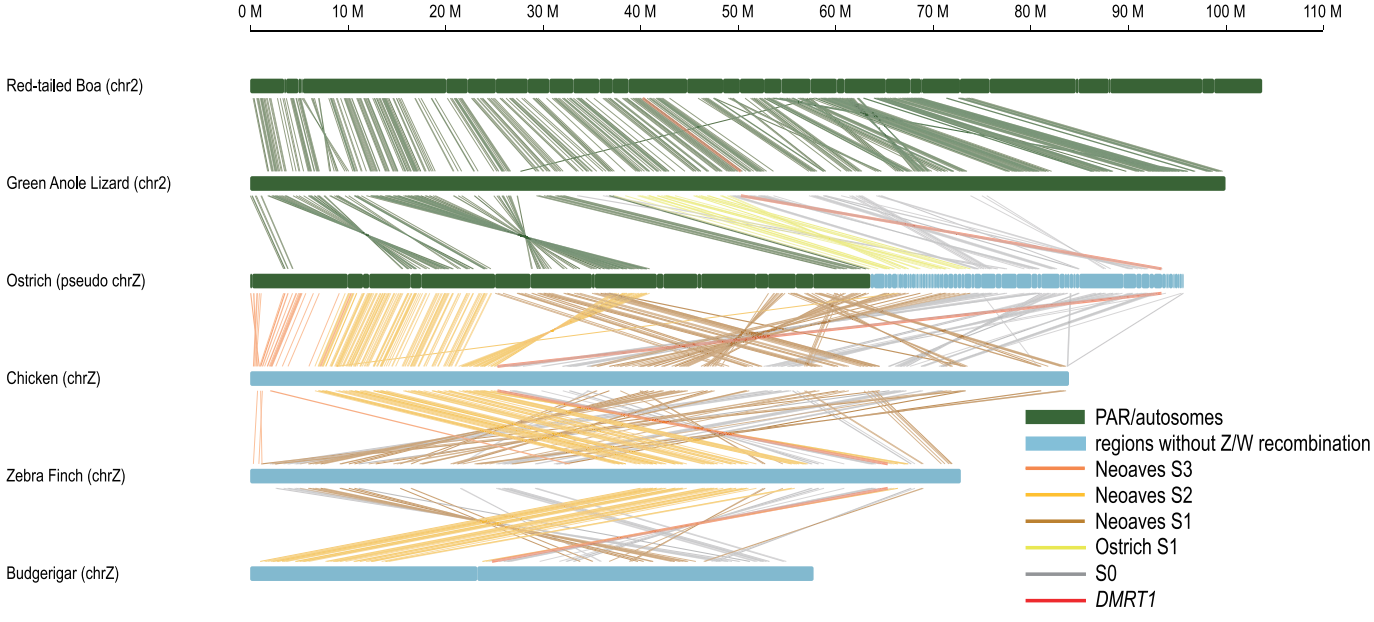


Fig. 3. Gene synteny within avian evolutionary strata. Mapped is the evolutionary strata gene synteny between sequenced reptile (boa and lizard) and avian species. Each line connects a pair of orthologous genes between species, and different colors of lines represent different strata of genes: S0 (gray), ostrich S1 (yellow), Neognathae S1 (brown), Neoaves S2 (orange), *DMRT1* (red). Genomic regions encompassing S0 and S1 have experienced multiple inversions across different species, relocating *DMRT1* to the middle of the Z chromosome after the divergence between Neognathae and Palaeognathae. In contrast, genomic regions within Neoaves S2 and S3 show high level of syntenic conservation across species.

regardless of their different types of sex chromosome inheritance. This matches theoretical predictions that interference effects between linked loci (Hill-Robertson effects), which are ultimately driving Y/W degradation, diminish over evolutionary time owing to the decrease in the number of functional genes that are targets for selection in older strata (*49*).

Discussion

Our results (Fig. 1, summarized in fig. S23) allow us to reconstruct a likely evolutionary formation trajectory of avian sex chromosome strata

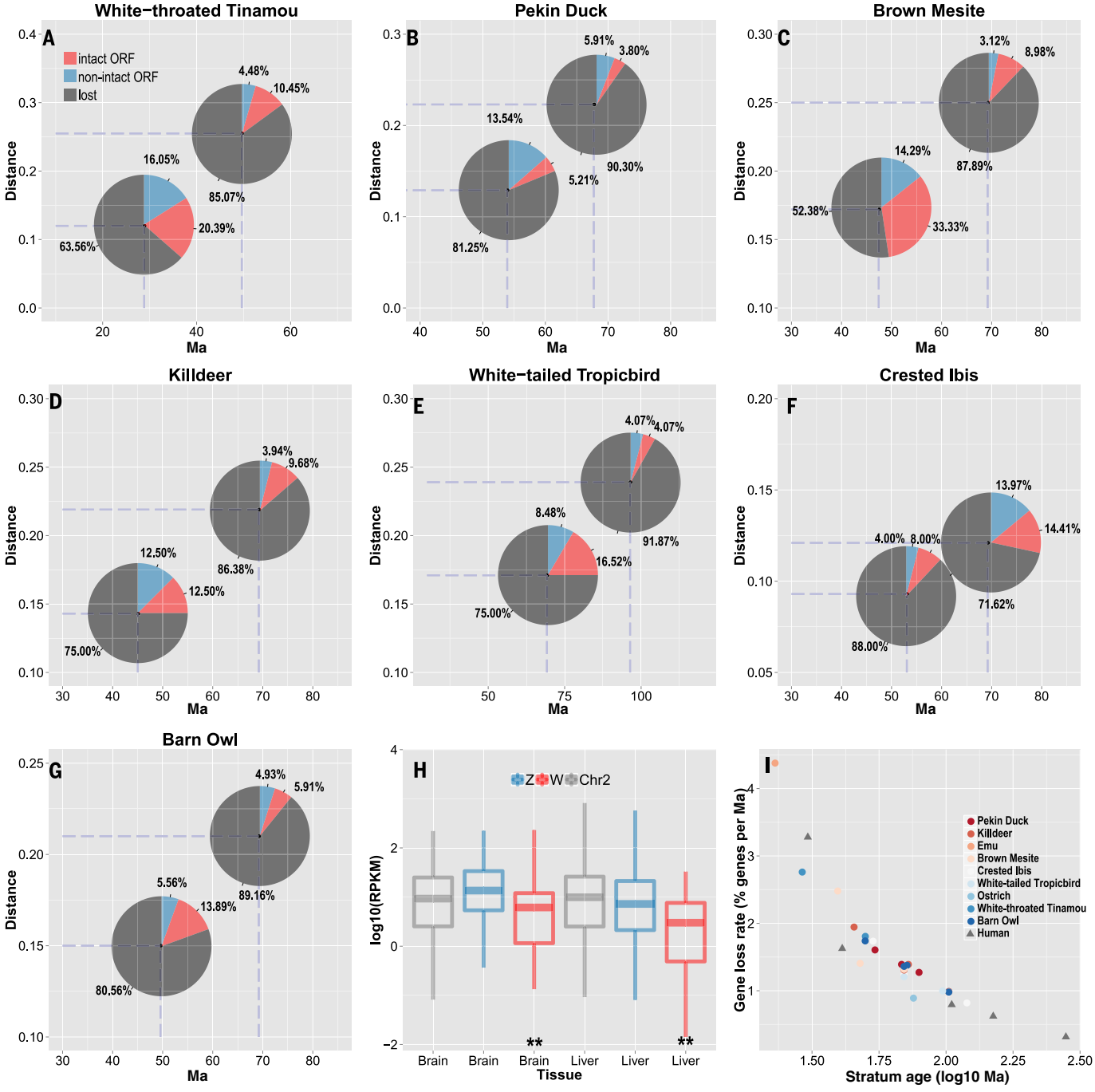


Fig. 4. Evolution of avian W-linked genes. (**A** to **G**) For each stratum of the example species, we show a pie chart of gene composition on the W with regards to whether their open reading frames (ORFs) are predicted to be intact or not. The *x* axis is the estimated age, and *y* axis is the Z/W divergence level of each stratum. Within each pie chart, we show W-linked genes that cannot be found within the assembled sequences (gray), genes with disrupted ORFs that contain premature stop codons or frameshift mutations (blue), and genes with intact ORFs (red). Note that the number of missing genes may be overestimated, owing to the difficulty of assembling W-linked genes surrounded by repetitive sequences. (**H**) Comparison of gene expression between Z and W linked gametologs in ostrich brain and liver samples, relative to their autosomal lizard orthologs. We show significance level of Wilcoxon test comparing the gametolog expression versus that of lizard (**$P < 0.01$). W gametologs show a significant down-regulation of gene expression. (**I**) Correlation between gene loss rate and age of stratum. Each data point represents a stratum of a certain bird species (dot) in Fig. 1 or human (triangle). The data for the human Y chromosome are derived from (*49*, *51*).

(detailed in Fig. 5). We hypothesize that genetic sex determination involving *DMRT1* originated in an ancestor of all birds, and a Z-linked inversion suppressing recombination between chrZ and chrW may have been selected by the accumulation of sexually antagonistic alleles nearby *DMRT1*, forming S0, the only stratum shared among all bird species. This event was followed by numerous independent inversions along the proto–sex chromosome pair accompanying various speciation events. Each event resulted in spreading recombination suppression across genomic regions adjacent to the non-recombining region and decreasing the size of the PAR. Separate evolutionary strata emerged among the different Palaeognathae species. We propose that Neognathae stratum S1 emerged in the ancestor of Neognathae, possibly involving the translocation of the Z-linked *DMRT1* gene toward the center of the Z chromosome. This was followed by the independent formation of stratum S2 in Galliformes, Anseriformes, and the ancestor of Neoaves. Finally some, but not all, Neognathae lineages formed independent S3 strata.

Several general patterns also emerge from comparing the evolutionary history of avian and mammalian sex chromosomes. Unlike in mammals, no sex-autosome fusions contributed to the formation of ancestral avian evolutionary strata, but recombination became abolished in a similar stepwise manner in both birds and mammals. The observation that recombination suppression gradually spreads outwards from the region containing the putative sex-determining gene *DMRT1* (Fig. 1 and fig. S9) in birds mimics patterns of strata formation in mammals, where the oldest stratum contains the sex-determining *SRY* gene. We infer that if inversions caused recombination suppression in birds, they must have occurred on both the Z and W chromosomes (Fig. 5). In contrast, all the inversions suppressing recombination in mammals probably occurred on the Y (*13*).

This different pattern is consistent with a model in which sexual antagonism drives the evolution of suppressed recombination between sex chromosomes (*8*, *50*). Assuming that most sexual antagonism is caused by sexual selection in males, alleles that are beneficial to males but detrimental to females may be more common than female-beneficial male-detrimental mutations (*8*, *50*). In XY systems, Y-linked inversions that trap such male-beneficial female-detrimental alleles into the male-limited region along the Y will be immediately selected for. In ZW systems, however, neither Z- nor W-linked inversions can restrict such alleles to males, but either can be favored by preventing male-beneficial sexual antagonistic alleles from recombining onto the female-specific W chromosome. In the chicken, male function genes are accumulating on the Z chromosome, including a large array of testis-expressed genes termed "Z amplicons" (*3*), providing empirical support for the accumulation of sex-specific genes along the sex chromosomes (table S7).

In conclusion, our results provide a detailed view of the evolutionary history of avian sex chromosomes, involving both shared and lineage-specific recombination suppression events. These events differ by their emergence times, the genomic regions spanned, and the tempo of Z/W differentiation within the affected region, which together underlie the diversity of avian sex chromosome composition. Even within homologous strata of the same age and size, we find marked variation in the extent to which they have lost their W-linked genes in different lineages. It is not entirely clear why rates of avian sex chromosome evolution should differ so dramatically, and whether, or to what extent, this can be attributed to sexual selection, which is widely manifested

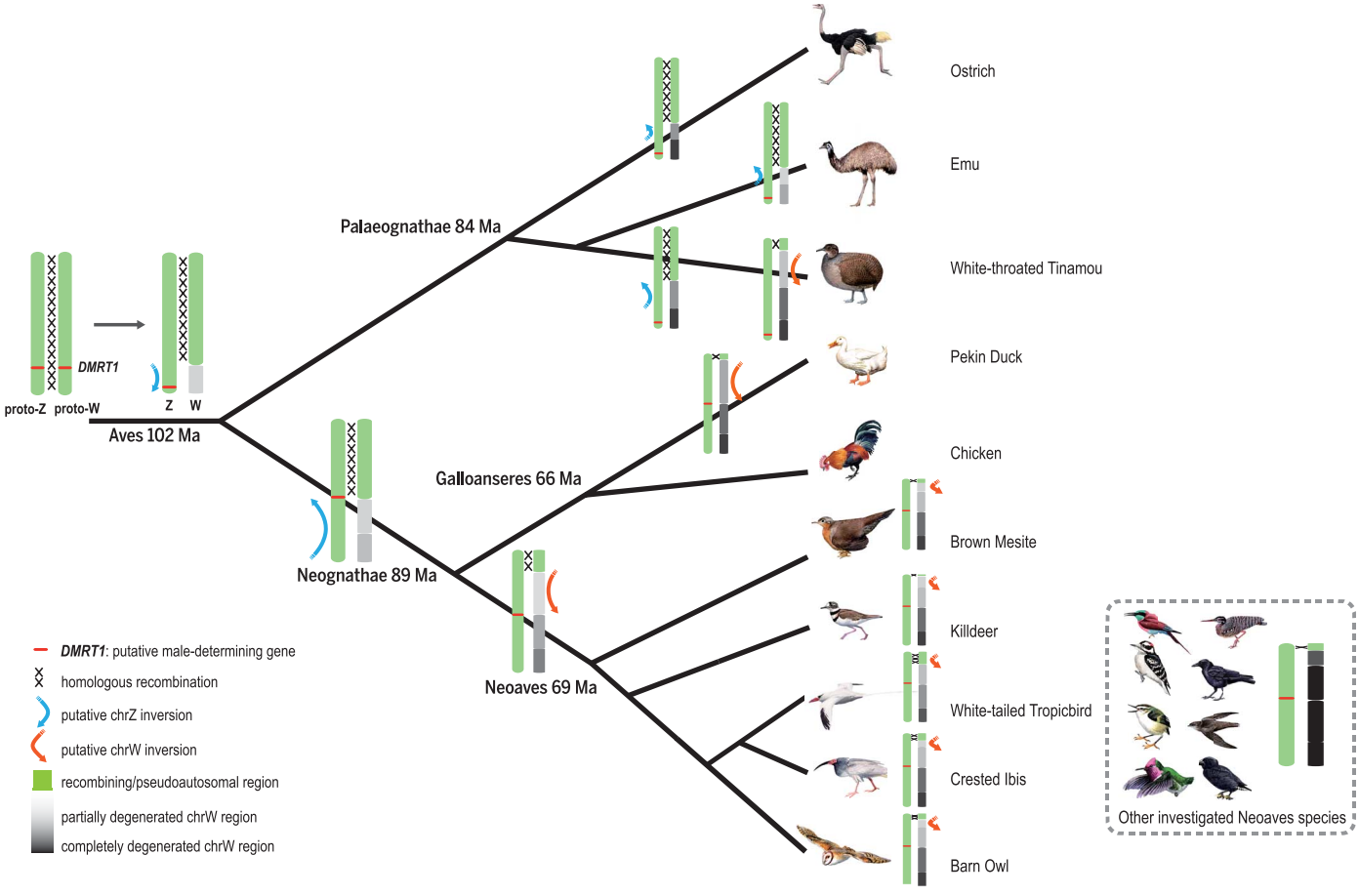


Fig. 5. Hypothesized formation scenarios of avian evolutionary strata. We designate *DMRT1* on the chromosome with the red bar, and homologous recombination between proto–sex chromosomes at PAR (in green) with intersecting lines. Gray or black regions on the W chromosome represent the female-specific region lacking recombination, with darker color representing higher level of differentiation between Z/W. We also show inferred Z- or W-linked inversions that have led to the formation of certain evolutionary strata on the tree.

in birds. Surprisingly, several Neognathae species have retained a large number of genes within the nonrecombining segment of their W chromosome (Table 1), contrary to the general assumption that ancient sex-linked regions invariably degenerate almost completely.

REFERENCE AND NOTES

1. J. J. Bull, *Evolution of Sex Determining Mechanisms* (Benjamin/Cummings, Menlo Park, CA, 1983).
2. A. K. Fridolfsson *et al*., Evolution of the avian sex chromosomes from an ancestral pair of autosomes. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8147–8152 (1998). doi: 10.1073/pnas.95.14.8147; pmid: 9653155
3. D. W. Bellott *et al*., Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature* **466**, 612–616 (2010). doi: 10.1038/nature09172; pmid: 20622855
4. D. Cortez *et al*., Origins and functional evolution of Y chromosomes across mammals. *Nature* **508**, 488–493 (2014). doi: 10.1038/nature13151; pmid: 24759410
5. M. T. Ross *et al*., The DNA sequence of the human X chromosome. *Nature* **434**, 325–337 (2005). doi: 10.1038/nature03440; pmid: 15772651
6. B. Charlesworth, D. Charlesworth, The degeneration of Y chromosomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **355**, 1563–1572 (2000). doi: 10.1098/rstb.2000.0717; pmid: 11127901
7. H. Skaletsky *et al*., The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003). doi: 10.1038/nature01722; pmid: 12815422
8. W. R. Rice, The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. *Evolution* **41**, 911–914 (1987). doi: 10.2307/2408899
9. B. T. Lahn, D. C. Page, Four evolutionary strata on the human X chromosome. *Science* **286**, 964–967 (1999). doi: 10.1126/science.286.5441.964; pmid: 10542153
10. K. Nam, H. Ellegren, The chicken (Gallus gallus) Z chromosome contains at least three nonlinear evolutionary strata. *Genetics* **180**, 1131–1136 (2008). doi: 10.1534/genetics.108.090324; pmid: 18791248
11. G. Li *et al*., Comparative analysis of mammalian Y chromosomes illuminates ancestral structure and lineage-specific evolution. *Genome Res.* **23**, 1486–1495 (2013). doi: 10.1101/gr.154286.112; pmid: 23788650
12. D. W. Bellott *et al*., Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* **508**, 494–499 (2014). doi: 10.1038/nature13206; pmid: 24759411
13. C. Lemaitre *et al*., Footprints of inversions at present and past pseudoautosomal boundaries in human sex chromosomes. *Genome Biol. Evol.* **1**, 56–66 (2009). doi: 10.1093/gbe/evp006; pmid: 20333177
14. S. A. Sandstedt, P. K. Tucker, Evolutionary strata on the mouse X chromosome correspond to strata on the human X chromosome. *Genome Res.* **14**, 267–272 (2004). doi: 10.1101/gr.1796204; pmid: 14762062
15. K. L. Ayers *et al*., RNA sequencing reveals sexually dimorphic gene expression before gonadal differentiation in chicken and allows comprehensive annotation of the W-chromosome. *Genome Biol.* **14**, R26 (2013). doi: 10.1186/gb-2013-14-3-r26; pmid: 23531366
16. A. E. Wright, H. K. Moghadam, J. E. Mank, Trade-off between selection for dosage compensation and masculinization on the avian Z chromosome. *Genetics* **192**, 1433–1445 (2012). doi: 10.1534/genetics.112.145102; pmid: 22997237
17. L. J. Handley, H. Ceplitis, H. Ellegren, Evolutionary strata on the chicken Z chromosome: Implications for sex chromosome evolution. *Genetics* **167**, 367–376 (2004). doi: 10.1534/genetics.167.1.367; pmid: 15166161
18. H. Ellegren, A. Carmichael, Multiple and independent cessation of recombination between avian sex chromosomes. *Genetics* **158**, 325–331 (2001). pmid: 11333240
19. R. S. de Kloet, S. R. de Kloet, Evolution of the spindlin gene in birds: Independent cessation of the recombination of sex chromosomes at the spindlin locus in neognathous birds and tinamous, a palaeognathous avian family. *Genetica* **119**, 333–342 (2003). doi: 10.1023/B:GENE.0000003842.72339.df; pmid: 14686612
20. J. García-Moreno, D. P. Mindell, Rooting a phylogeny with homologous genes on opposite sex chromosomes (gametologs): A case study using avian CHD. *Mol. Biol. Evol.* **17**, 1826–1832 (2000). doi: 10.1093/oxfordjournals.molbev.a026283; pmid: 11110898
21. J. Rutkowska, M. Lagisz, S. Nakagawa, The long and the short of avian W chromosomes: No evidence for gradual W shortening. *Biol. Lett.* **8**, 636–638 (2012). doi: 10.1098/rsbl.2012.0083; pmid: 22417794
22. S. J. Hackett *et al*., A phylogenomic study of birds reveals their evolutionary history. *Science* **320**, 1763–1768 (2008). doi: 10.1126/science.1157704; pmid: 18583609
23. E. D. Jarvis *et al*., Whole genome analyses resolve the early branches to the tree of life of modern birds. *Science* **346**, 1320–1331 (2014).
24. A. Ogawa, K. Murata, S. Mizuno, The location of Z- and W-linked marker genes and sequence on the homomorphic sex chromosomes of the ostrich and the emu. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4415–4418 (1998). doi: 10.1073/pnas.95.8.4415; pmid: 9539751
25. C. Nishida-Umehara *et al*., Differentiation of Z and W chromosomes revealed by replication banding and FISH mapping of sex-chromosome-linked DNA markers in the cassowary (Aves, Ratitae). *Chromosome Res.* **7**, 635–640 (1999). doi: 10.1023/A:1009236103013; pmid: 10628664
26. Y. Tsuda, C. Nishida-Umehara, J. Ishijima, K. Yamada, Y. Matsuda, Comparison of the Z and W sex chromosomal architectures in elegant crested tinamou (Eudromia elegans) and ostrich (Struthio camelus) and the process of sex chromosome differentiation in palaeognathous birds. *Chromosoma* **116**, 159–173 (2007). doi: 10.1007/s00412-006-0088-y; pmid: 17219176
27. M. I. Pigozzi, Diverse stages of sex-chromosome differentiation in tinamid birds: Evidence from crossover analysis in *Eudromia elegans* and *Crypturellus tataupa*. *Genetica* **139**, 771–777 (2011). doi: 10.1007/s10709-011-9581-1; pmid: 21567220
28. M. I. Pigozzi, Origin and evolution of the sex chromosomes in birds. *Biocell* **23**, 79–95 (1999). pmid: 10904535
29. G. Zhang *et al*., Comparative genomics across modern bird species reveal insights into avian genome evolution and adaptation. *Science* **346**, 1311–1320 (2014).
30. Y. Huang *et al*., The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. *Nat. Genet.* **45**, 776–783 (2013). doi: 10.1038/ng.2657; pmid: 23749191
31. B. Vicoso, V. B. Kaiser, D. Bachtrog, Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 6453–6458 (2013). doi: 10.1073/pnas.1217027110; pmid: 23547111
32. L. Ye *et al*., A vertebrate case study of the quality of assemblies derived from next-generation sequences. *Genome Biol.* **12**, R31 (2011). doi: 10.1186/gb-2011-12-3-r31; pmid: 21453517
33. W. L. Fan *et al*., Genome-wide patterns of genetic variation in two domestic chickens. *Genome Biol. Evol.* **5**, 1376–1392 (2013). doi: 10.1093/gbe/evt097; pmid: 23814129
34. Materials and methods are available as supplementary material on *Science* Online.
35. I. Nanda, K. Schlegelmilch, T. Haaf, M. Schartl, M. Schmid, Synteny conservation of the Z chromosome in 14 avian species (11 families) supports a role for Z dosage in avian sex determination. *Cytogenet. Genome Res.* **122**, 150–156 (2008). doi: 10.1159/000163092; pmid: 19096210
36. R. A. Dalloul *et al*., Multi-platform next-generation sequencing of the domestic turkey (Meleagris gallopavo): Genome assembly and analysis. *PLOS Biol.* **8**, e1000475 (2010). doi: 10.1371/journal.pbio.1000475; pmid: 20838655
37. W. C. Warren *et al*., The genome of a songbird. *Nature* **464**, 757–762 (2010). doi: 10.1038/nature08819; pmid: 20360741
38. H. Ellegren *et al*., The genomic landscape of species divergence in Ficedula flycatchers. *Nature* **491**, 756–760 (2012). pmid: 23103876
39. C. A. Smith *et al*., The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. *Nature* **461**, 267–271 (2009). doi: 10.1038/nature08298; pmid: 19710650
40. S. Guioli, R. Lovell-Badge, J. M. Turner, Error-prone ZW pairing and no evidence for meiotic sex chromosome inactivation in the chicken germ line. *PLOS Genet.* **8**, e1002560 (2012). doi: 10.1371/journal.pgen.1002560; pmid: 22412389
41. H. P. Yazdi, H. Ellegren, Old but not (so) degenerated—slow evolution of largely homomorphic sex chromosomes in ratites. *Mol. Biol. Evol.* **31**, 1444–1453 (2014). doi: 10.1093/molbev/msu101; pmid: 24618361
42. J. Alföldi *et al*., The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* **477**, 587–591 (2011). doi: 10.1038/nature10390; pmid: 21881562
43. K. R. Bradnam *et al*., Assemblathon 2: Evaluating de novo methods of genome assembly in three vertebrate species. *Gigascience* **2**, 10 (2013). doi: 10.1186/2047-217X-2-10; pmid: 23870653
44. A. Kawai *et al*., Different origins of bird and reptile sex chromosomes inferred from comparative mapping of chicken Z-linked genes. *Cytogenet. Genome Res.* **117**, 92–102 (2007). doi: 10.1159/000103169; pmid: 17675849
45. A. Kawai *et al*., The ZW sex chromosomes of Gekko hokouensis (Gekkonidae, Squamata) represent highly conserved homology with those of avian species. *Chromosoma* **118**, 43–51 (2009). doi: 10.1007/s00412-008-0176-2; pmid: 18685858
46. S. Berlin, H. Ellegren, Fast accumulation of nonsynonymous mutations on the female-specific W chromosome in birds. *J. Mol. Evol.* **62**, 66–72 (2006). doi: 10.1007/s00239-005-0067-6; pmid: 16320115
47. Y. Itoh, K. Kampf, A. P. Arnold, Disruption of FEM1C-W gene in zebra finch: Evolutionary insights on avian ZW genes. *Chromosoma* **118**, 323–334 (2009). doi: 10.1007/s00412-008-0199-8; pmid: 19139913
48. J. F. Hughes *et al*., Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature* **483**, 82–86 (2012). doi: 10.1038/nature10843; pmid: 22367542
49. D. Bachtrog, The temporal dynamics of processes underlying Y chromosome degeneration. *Genetics* **179**, 1513–1525 (2008). doi: 10.1534/genetics.107.084012; pmid: 18562655
50. D. Charlesworth, B. Charlesworth, Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. *Genet. Res.* **35**, 205–214 (1980). doi: 10.1017/S0016672300014051; pmid: 6930353
51. M. A. Wilson Sayres, K. D. Makova, Gene survival and death on the human Y chromosome. *Mol. Biol. Evol.* **30**, 781–787 (2013). doi: 10.1093/molbev/mss267; pmid: 23223713

ACKNOWLEDGMENTS

We thank B. Vicoso, R. Bowie, M. Sayres, I. Nanda, and D. Charlesworth for insightful comments and discussions; A. Thais for sharing the green anole lizard photo; O. Ryder and W. Wang for providing the ostrich sample; and the teams of N. Xiao and Z. Bin at OpGen Inc. for performing the optical mapping and technical assistance. We also thank five anonymous reviewers for constructive comments and suggestion. This work was supported by the National Key Basic Research Program of China (973 Program: 2013CB945200) to G.Z. Q.Z. is supported by grants from the National Institutes of Health (GM076007 and GM093182) and a Packard Fellowship to D.B. Raw sequence reads and assembled genomes and gene annotation are deposited at http://phybirds.genomics.org.cn/genome.jsp

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1246338/suppl/DC1
Materials and Methods
Figs. S1 to S23
Tables S1 to S7
References

23 September 2013; accepted 8 August 2014
10.1126/science.1246338

Complex evolutionary trajectories of sex chromosomes across bird taxa

Qi Zhou *et al.*

Science **346**, (2014);
DOI: 10.1126/science.1246338

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

www.sciencemag.org/content/346/6215/1246338/suppl/DC1

Supplementary Materials for

Complex evolutionary trajectories of sex chromosomes across bird taxa

Qi Zhou,* Jilin Zhang, Doris Bachtrog, Na An, Quanfei Huang, Erich D. Jarvis, M. Thomas P. Gilbert, Guojie Zhang*

*Corresponding author. E-mail: zhouqi@berkeley.edu (Q.Z.); zhanggj@genomics.org.cn (G.Z.)

Published 12 December 2014, *Science* **346**, 1246338 (2014)
DOI: 10.1126/science.1246338

This PDF file includes:

Other Supporting Online Material for this manuscript includes the following:
(available at www.sciencemag.org/content/346/6215/1246338/suppl/DC1)

Materials and Methods

Optical mapping of Ostrich genome

High molecular weight genomic DNA was extracted (Qiagen) from a blood sample collected from a male Ostrich in Kunming Zoo of China, and then passed to OpGen Inc. for collection of a single molecule restriction map (SMRM) on the ARGUS(R) Whole Genome Mapping System. DNA was captured by a microfluidic device, MapCard™ and subjected to restriction enzyme digestion with *Kpnl* enzyme. After digestion, the molecules were labeled with a fluorescent DNA intercalating dye, *JoJo-1* (Invitrogen) and then captured by an automated image-processing module. A total of 1.12 Tb SMRM data with 3925195 DNA molecules longer than 150 kb were collected. At the same time, the draft genome of a female Ostrich was converted into an *in silico* restriction map to provide seeds for selecting SMRM for alignment based on the restriction sites. SMRMs were preprocessed into the elongated consensus map (extended scaffolds) before the alignment, and low coverage regions toward ends were trimmed off to achieve high extension quality. We iterated the extension-alignment process four times to generate sufficient extensions. Then the draft genome scaffolds were ordered and oriented into super-scaffolds based on the alignment with SMRM, where conflicts among alignments were resolved by selecting those with the highest alignment scores. Scaffold N50 length has been greatly improved after optical mapping (**Table S2**). Z chromosome sequences were built by orienting and connecting super-scaffolds based on the fluorescence *in situ* hybridization (FISH) results from previous work (*1*), and filling in 600Ns in between.

Linkage group construction and identification of PAR

The detailed assembly and annotation of all the species are described in our comparative genomics study of all these species (*2*). About 70 fold of Illumina reads derived from a female chicken (#SRP005856) (*3*) and 17 fold of Illumina reads from a male chicken (#SRP022583) (*4*) were downloaded from NCBI SRA database, and Emu reads are from (*5*). In brief, the raw reads of all species were assembled into scaffold sequences by SOAPdenovo (*6*), and filled in the gaps within the scaffolds by GapCloser (http://sourceforge.net/projects/soapdenovo2/files/GapCloser/). We performed assemblies with different fold of chicken female reads, to evaluate the effect of read coverage on the assembly quality (**Fig. S6**), which seems to be minor. We used the chicken genome (galGal13, http://www.genome.ucsc.edu/) and its complete Z chromosome sequence from (*7*) as a reference to build the Neognathae pseudo-chromosomal sequences, and Ostrich genome from this study as a reference for other paleognaths. *DMRT1* was manually mapped to the chicken Z chromosome by its neighboring genes *DMRT2* and *DMRT3,* which together show conserved syntenic relationship within other vertebrate species (*8, 9*). In Ostrich, we found *DMRT1* located on the same super-scaffold with *ACO1*, and thus

we orientated the scaffold according to the previous FISH result of *ACO1* (*1*). Both reference genomes have been masked for their repetitive elements using a consensus avian repeat library by RepeatMasker (http://repeatmasker.org), prior to further alignments. We performed whole genome alignments of scaffolds against the reference sequence by LASTZ, with parameter set '--step=19 --hspthresh=2200 --inner=2000 --ydrop=3400 --gappedthresh=10000 --format=axt' (*10*) and a score matrix set for distant species comparison. Alignments were converted into a series of syntenic 'chains', 'net' and 'maf' results with different levels of alignment scores using UCSC Genome Browser's utilities (http://genomewiki.ucsc.edu/index.php/).

Based on the whole genome alignment, we first identified the best aligned scaffolds within the overlapping regions on the reference genome, according to their alignment scores with a cutoff of at least 50% of the whole scaffold aligned in the LASTZ net results. We further estimated the overall identity and coverage distributions with a 10kb non-overlapped sliding window for each scaffold along the reference sequence, and obtained the distributions of identity and coverage. Scaffolds within the lower 5% region of each distribution were removed to avoid spurious alignments. Finally, scaffolds were ordered and oriented into pseudo-chromosome sequences according to their unique positions on the reference.

We aligned the raw reads of each species to their pseudo-chromosome sequences by BWA (*11*) with the parameter set '-o 1 -e 50 -m 100000 -t 4 -i 15 -q 10 -I -k 0', adjusting for insert size of the Illumina sequencing library. We calculated read depth using SAMtools (*12*) within each 100kb non-overlapping window and normalized it against the median value of depths per single base pair throughout the entire genome, to allow comparison among species. The normalized depth was then converted by colorRampPalette in R and plotted against a 256-value color gradient array ranging from 0 to 1 (blue to green) along the Z chromosome to display the PAR and differentiated region. Boundaries of PAR were determined by a significant shift of depth values between neighboring windows.

Identification and validation of W-linked scaffolds

After excluding the best-aligned scaffolds used for building the Z chromosome, we performed a second round of LASTZ alignment against the Z chromosome sequence to identify candidate W-linked scaffolds, with a minimum length cutoff of 1kb, an alignment cutoff of at least 50% of the entire scaffold aligned and 70% identity. We also ran the same pipeline on an autosome (chr1) for comparison (**Fig. S1**). We mapped the re-sequencing data and calculated the read depths of a male chicken (17 fold coverage) (*4*), a male Crested Ibis (40 fold sequencing coverage) and a male Emu (*13*) (10 fold coverage) onto their candidate W-linked scaffolds, chrZ and chr1 sequences with the same parameter sets of BWA and SAMtools mentioned above. The histogram distribution

of normalized read depth of scaffolds was then compared between sexes in both species (**Fig. S3**) and along individual genes (**Fig. S4**), in order to confirm the expected female specific pattern of W-linked scaffolds, and a lack of mis-assembly with Z-linked sequences. A total of 98.9kb candidate chicken W-linked scaffolds were also aligned to the reference genome by Discontinuous MEGABLAST (http://www.ncbi.nlm.nih.gov/blast/html/megablast.html), only scaffolds with more than 50% of the sequence uniquely aligned at a minimum identity of 95% were assigned to certain chromosomes (**Fig. S5**).

To allow a more direct comparison, we scaled the Z-linked sequences of each species to the same length as the reference sequence of Ostrich or chicken, by filling the alignment gap with the same length of 'N's as the reference sequence based on the 'net' result of LASTZ. Then we calculated pairwise alignment identity between the Z/W based on the 'maf' result and oriented the candidate W-linked scaffolds along the Z chromosome sequences. We also color-coded the read depth and alignment identity of each scaffold for visualization. We scanned the Z/W pairwise alignment identities along the Z chromosome with a non-overlapping 100kb window, and determined the boundaries between the neighboring strata when there was a significant difference of the identities or occurrences of W-linked scaffolds (**Figs. S9-10**).

Inferring the age of evolutionary strata

For Palaeognathae and Galloanserae (Pekin Duck) species, we used a method similar to that of (*14*, *15*) assuming the molecular clock. We first scaled the species-specific mutation rate to that of chicken (2.5×10^{-9}/site/year) (*16*) by their ratio of median values of branch-specific synonymous substitution rates (dS) of genes linked to macro-chromosomes calculated by PAML (free-ratio model: *runmode = 0, mode = 1*) (*17*). We also calculated the degree of male-biased mutation (α) using the equation α=3(Z/A -2)/(4- 3Z/A) (*18*), where 'Z' and 'A' represent median substitution rates of intronic regions linked to autosomes and chrZ. The results of α are consistent with previous studies (*19, 20*), ranging from 2 to 6. W chromosome mutation rate was then estimated from the equation α=2A/W-1, where 'W' refers to the mutation rate of the W chromosome. The sum of Z and W chromosome mutation rates results in the divergence rate between the sex chromosomes. For the Z/W divergence level, we first removed the coding region from the pairwise Z/W genome alignments generated from LASTZ by customized perl scripts, and then estimated the divergence level within non-coding regions by baseml in the PAML package after 1000 bootstraps. The age of a stratum was finally estimated by dividing the divergence level by divergence rate.

For Neoaves species, we estimated the age by scaling the Z/W divergence level at non-coding regions of a certain stratum to that of Neoaves stratum S2, which evolved at least 69.2 MYA (*21*).

Annotation of W-linked genes

We used the protein sequences of Z-linked genes as a query and aligned them to the W scaffolds with BLAT (*22*). The best aligned (cutoff: identity>=70%, coverage>=50%) region with extended flanking sequences of 500 bp at both ends was then provided to GeneWise (*23*) (-tfor -genesf -gff –sum) to determine if it contains an intact open reading frame (ORF). We annotated the ORF as disrupted when GeneWise reported at least one premature stop codon or frame-shift mutation. CDS sequences of single-copy genes' Z/W gametologs were aligned by MUSCLE (*24*) and the resulting alignments were cleaned by Gblocks (-b4=5, -t=c, -e=-gb) (*25*). All alignments generated for W and Z gametologs are included in the Supplementary Data. Only alignments longer than 300bp were used for constructing maximum likelihood trees by RAxML (*26*) to infer whether their residing evolutionary stratum is shared among species or specific to lineages. Transcript sequences were also predicted by GeneWise with a minimum length cutoff of 150bp. We aligned the RNA-seq reads of ostrich females' liver and brain tissues (*27*) to the transcript sequences using TopHat (*28*) and calculated their expression levels measured by reads per kilo base per million (RPKM) on the basis of unique alignments.

We also produced pairwise alignments of Z/W linked CDS vs. their chicken orthologs by MUSCLE (*24*) and filtered the alignments by Gblocks (*25*). To compare the evolutionary rates of Z linked and W linked genes, we calculated their ratio of nonsynonymous substitution rate vs. synonymous substitution rate by PAML.

Rearrangement analysis

We produced 1:1 reciprocal best orthologs amongst between the Green Anole Lizard (29) Boa Snake (Boa constrictor constrictor) (*http://bioshare.bioinformatics.ucdavis.edu/Data/hcbxz0i7kg/Snake/*) (30), and all the bird species in this study by BLASTP. We found 1255 orthologous gene pairs between lizard and boa, 314 gene pairs between Ostrich and lizard, and 268 gene pairs shared amongst the four avian species plotted in **Fig. 3**. We connected them by dashed lines color-coded with the strata information where they reside.

Supplementary Figure Legend

figure S1. Patterns of read depth and alignment identities on autosomes (chromosome 1) for all the investigated species

We run the same pipeline for identifying the differentiated scaffolds as done in Figure 1 on an autosome (chromosome 1) as a control.

figure S2. Bird sex chromosomes have differentiated to different degrees

We show all the species studied here, whose phylogenetic tree is derived from (*21*).

Similar to Figure 1, each species has three tracks showing the read depth of chrZ, Z/W alignment identities and read depth of the identified W-linked fragments.

figure S3. Read depth distribution of scaffolds linked to different chromosomes of chicken, Crested Ibis and Emu
For each species, male reads are plotted in blue, and female reads in red. The identified chrW scaffolds in this work all show female-specific read depth patterns. Emu Z chromosome shows a bimodal pattern, corresponding to PAR and differentiated regions.

figure S4. Male and female read depths of Crested Ibis W-linked genes
Circos plot showing male (in blue) and female (in red) coverage along the individual W-linked gene of Crested Ibis in a 2kb window size and 400bp overlapping stepsize. Green line shows the median level of coverage of autosomal linked genes for comparison. This plot confirms that there is no chimeric assembly of Z-linked sequences into W-linked genes from our pipeline.

figure S5. Comparing candidate chicken W-linked scaffolds identified in this work to the chicken reference genome
This pie chart shows the proportions of candidate chicken W-linked sequences aligned to separate chromosomes of chicken reference genome.

figure S6. Comparing chicken W-linked genes derived from genomic assembly vs. those from transcriptomic assembly
We aligned the W-linked genes derived from our genomic assembly pipeline of this study to those derived from transcriptomic study (*15, 31*). We show for each gene of different evolutionary strata, their percentages of aligned transcript length under different sequencing coverages, in order to evaluate the effect of sequencing coverage on the W gene assembly. It seems for species like chicken with a highly degenerated W chromosome, we are only able to assemble short fragments of W-linked genes, due to the extremely high repeat content. And different coverages don't show a significant difference, although 20 fold is slightly better than the other two.

figure S7. Comparing Ostrich W-linked genes derived from genomic assembly vs. those from transcriptomic assembly
We assembled a total of 55 W-linked genes or their fragments of ostrich in this study (**Table S5**). Then we compared them to those derived from trascriptomic study (*32*), which have 9 W-linked genes' sequences available. 5 genes aligned more than 80% of their transcript lengths with an identity higher than 95%.

figure S8. The length distribution of annotated transcripts for Z and W linked genes in different species.

figure S9. The distribution of Z/W pairwise alignment identities along the Z chromosome
We plotted Z/W pairwise alignment identities along the Z chromosome for each species in a 100kb non-overlapping window.

figure S10. Pairwise Z/W alignment identities are significantly different amongst strata
Each box shown in the figure is the distribution of Z/W pairwise sequence identities within certain stratum. The color of the box is shown as the stratum color that is the same as that in Figure 1, and we show the statistical significance measuring the difference between the two neighboring strata as asterisks at the bottom (Wilcoxon test, *: $P<0.05$, **: $P<0.01$).

figure S11. Pairwise d*S* between Z/W gametologs are significantly different amongst strata
Each box shown in the figure is the distribution of pairwise synonymous substitution rates between Z and W gametologs of certain stratum. Older strata generally tend to have a higher d*S* value, and we show the statistical significance measuring the difference between the two neighboring strata as asterisks at the bottom (Wilcoxon test, *: $P<0.05$, **: $P<0.01$).

figure S12. Gene trees for Z and W linked gametologs in S0 region
Shown are maximum likelihood (ML) trees using coding regions of different Z and W linked genes, with gene name under each tree and bootstrapping values at each node. We collapsed branches to a polytomy when their most recent ancestral node has a bootstrapping value lower than 50. Trees that show separate clustering of Palaeognathae and Neognathae Z- or W- linked gametologs provide strong evidence that these genes suppressed recombination before the divergence of the two clades, i.e., at the common ancestor of all birds.

figure S13. Pattern of evolutionary strata in Neoaves when mapping their sequence scaffolds using the ostrich genome as a reference
We aligned Z or W linked scaffolds of Neoaves species against the genome of ostrich, and scaled them to the same length as ostrich chrZ. As the ostrich chrZ can be considered a proxy for the proto-Z chromosome with fewer species-specific rearrangements than chicken (**Figure 3**), we are able to identify the approximate regions where Neognathae stratum S1 spans. Note however, that a genomic rearrangement breaks apart stratum S2.

figure S14. Gene trees for Z and W linked gametologs in S1 region
Shown are maximum likelihood trees using coding regions of different Z and W linked genes in Neognathae and/or Palaeognathae S1 regions. We collapsed branches to a polytomy when their most recent ancestral node has a bootstrapping value lower than 50. Different Palaeognathae species and the ancestor of Neognathae formed their S1 stratum independently, as shown by the clustering of Z- and W- gametologs of the same species in Palaeognathae, but clustering of Z- or W- gametologs of different species in Neognathae.

figure S15. Comparison of strata definition between this study and previous study of chicken
We color coded lines connecting orthologous gene pairs between ostrich and chicken based on the strata definition from Figure 1: S0 is gray, Neognathae S1 is brown, S2 is orange and S3 is green. We labeled the location of gene *KCMF1,* which is close to putative avian male-determining gene *DMRT1*, and has been recently inferred to have lost Z/W recombination at the ancestor of all birds (*33*). We also labeled strata defined from the previous study (*15*) on the chicken Z chromosome.

figure S16. Gene trees for Z and W linked gametologs in Neognathae S2 region
Shown are maximum likelihood (ML) trees using coding regions of different Z and W linked genes, with gene name under each tree. We collapsed branches to a polytomy when their most recent ancestral node has a bootstrapping value lower than 50. We use orthologous sequence of lizard genes as outgroup to construct the gene trees, if we are able to determine the orthologous sequences that are not too divergent between birds and lizard. S2 emerged independently in different Galloanserae species and the ancestor of Neoaves, as shown by the clustering of Z- and W- gametologs of duck, but clustering of Z- or W- gametologs of different species in Neoaves. Some genes (e.g., *ZNF131*) may also show clustering of Z- and W- gametologs of the same species in Neoaves, probably due to gene conversion.

figure S17. Gene trees for Z and W linked gametologs in Neognathae S3 region
Shown are maximum likelihood (ML) trees using coding regions of different Z and W linked genes, with gene name under each tree. We collapsed branches to a polytomy when their most recent ancestral node has a bootstrapping value lower than 50. All the studied bird species formed their most recent stratum S3 independently, as shown by the clustering of Z- and W- gametologs of the same species.

figure S18. Comparison of gene syntenies of ostrich vs. other reptile species on the ostrich Z chromosome

(A) Gene synteny map between ostrich and boa snake (*Boa constrictor*) genomes at stratum S0 of ostrich. Each line presented in the plot corresponds to one orthologous gene pair between the two species, and both genomes are *de novo* assembled into chromosomes by optical mapping. Note that we did not incorporate the green lizard genome because one of the marker genes (*RPS6*) has not been mapped to chromosomes in lizard. We identify a chromosomal inversion between the boa and ostrich genome that encompasses the entire S0 region. (B) The inversion at S0 is absent in any other reptile species examined previously by cytogenetic mapping. This figure is adopted from Fig. 4 of (*34*), which determined the location of *DMRT1* in different species.

figure S19. Comparison of gene synteny of birds vs. lizard at Neoaves stratum S3
Each line shows one orthologous gene pair between species. And we only show those that are assigned as S3 in other Neoaves species, and color code them in green.

figure S20. Elevated rates of evolution of W-linked genes relative to their Z-linked homologs.
We calculated ratios of nonsynonymous substitution rate vs. synonymous substitution rate between Neoaves Z/W linked genes vs. their chicken orthologs, and compare them across different strata.

figure S21. Comparing expression level between Z and W gametologs in emu and duck
We compare the normalized expression levels (RPKM) between Z and W gametologs vs. their lizard orthologs. We scaled the expression level of lizard by either the difference between PAR genes of emu vs. their orthologs of lizard, or chr1 genes of duck vs. their orthologs of lizard; and divided all the lizard gene expression level by 2 to compare to the gametolog expression of birds.

figure S22. Comparing expression level between Z and W gametologs in ostrich
We divided W gametologs into those with or without intact ORFs and compared their expression vs. their Z-linked homologs and lizard orthologs. In both liver and brain tissues, there is a higher proportion of W gametologs that are silenced with their Z linked homologs expressed, if the W-linked ORFs are disrupted.

figure S23. Map of avian evolutionary strata
We summarize the results of Figure 1, and show the spanned region of evolutionary strata of each species. Both the strata and genes within are color-coded, but strata in different species with the same color don't necessarily mean they share the same origin. All species share stratum S0, and S1 is shared by Neognathae, S2 shared by Neoaves. Due to the genomic rearrangements, the strata region along the chicken Z chromosomes have

been reshuffled and become discontinuous (**Figure 3**). Since Neoaves species' chromosomes are organized using chicken as reference in this study, their strata regions shown here is only an approximation.

Supplementary Table Legend

Table S1. Genome sequencing and assembly statistics of the investigated species

Table S2. Comparison of Ostrich genome assemblies before and after optical mapping (OM)

Table S3. Statistics of W-linked scaffolds assembled for species shown in Figure 1

Table S4. Statistics of W-linked genes assembled in this study

Table S5. Distribution of 26 chicken W-linked genes in species of this study
We use '+' and '-' to show presence and absence of assembled W-linked orthologs of 26 chicken W-linked genes in the species of this study.

Table S6. Ages and their confidence intervals of avian evolutionary strata

Table S7. Numbers of 'Z-amplicon' genes in three Neognathae genomes
We used 'Z-amplicon' genes identified in the chicken genome (*7*) and searched for their orthologs in turkey and mallard genomes

figure S1. Patterns of read depth and alignment identities on autosomes (chromosome 1) for all the investigated species

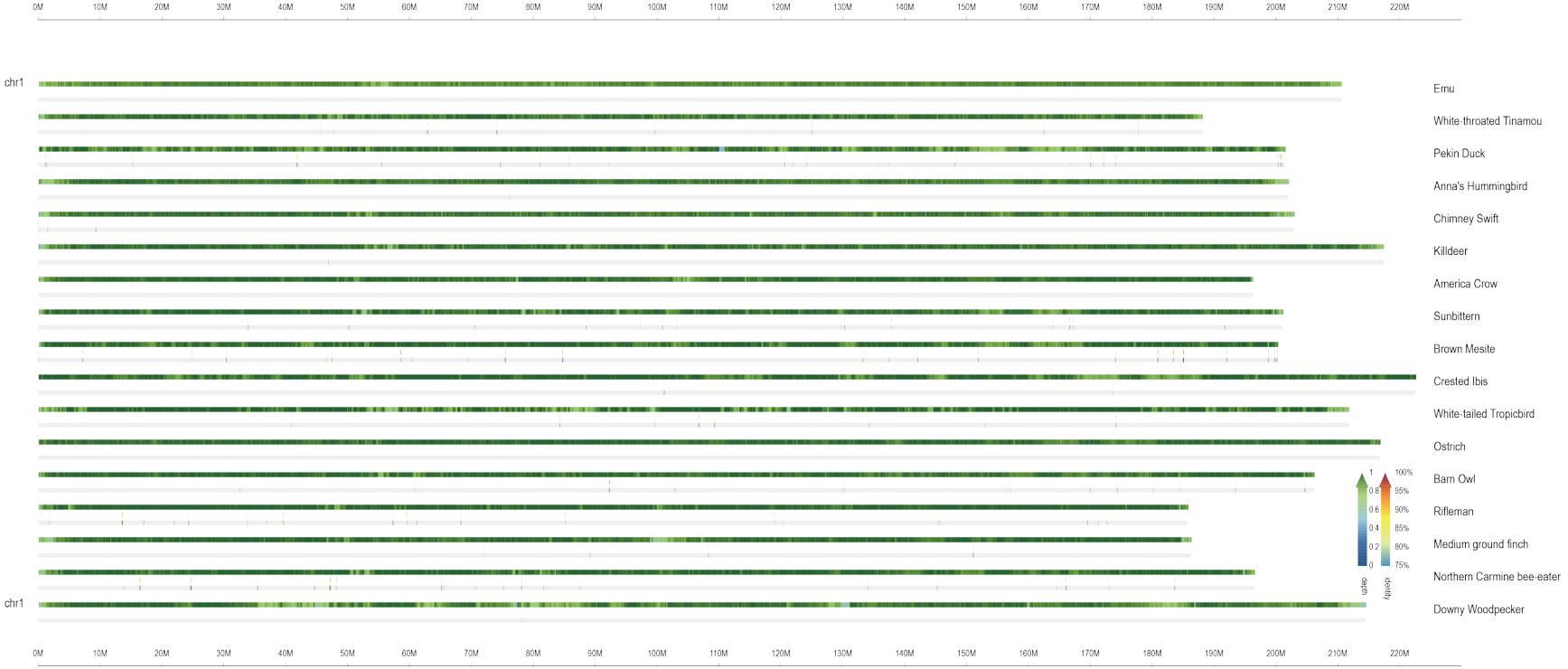


figure S2. Bird sex chromosomes have differentiated to different degrees

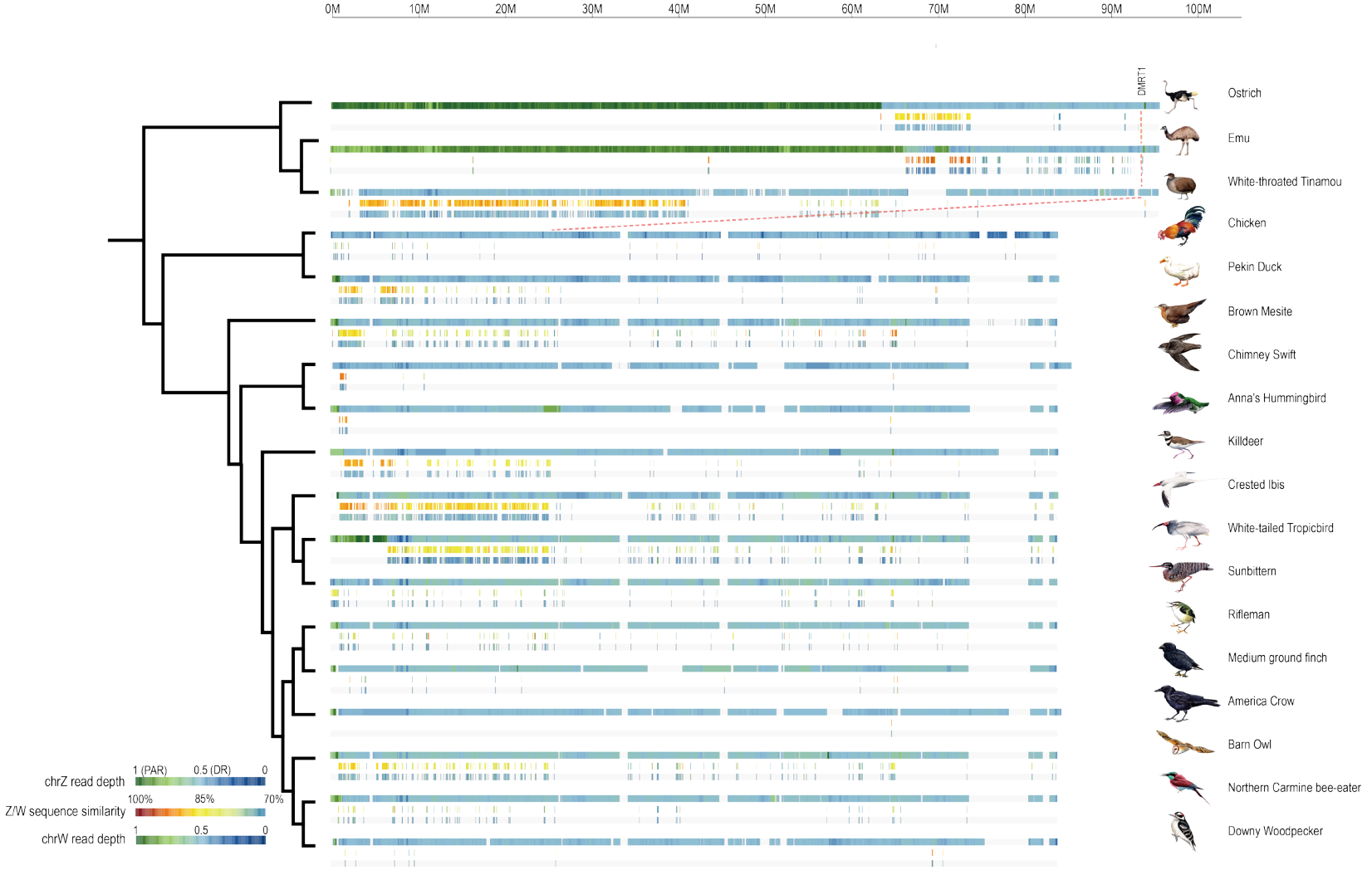


figure S3. Read depth distribution of scaffolds linked to different chromosomes of chicken, Crested Ibis and Emu

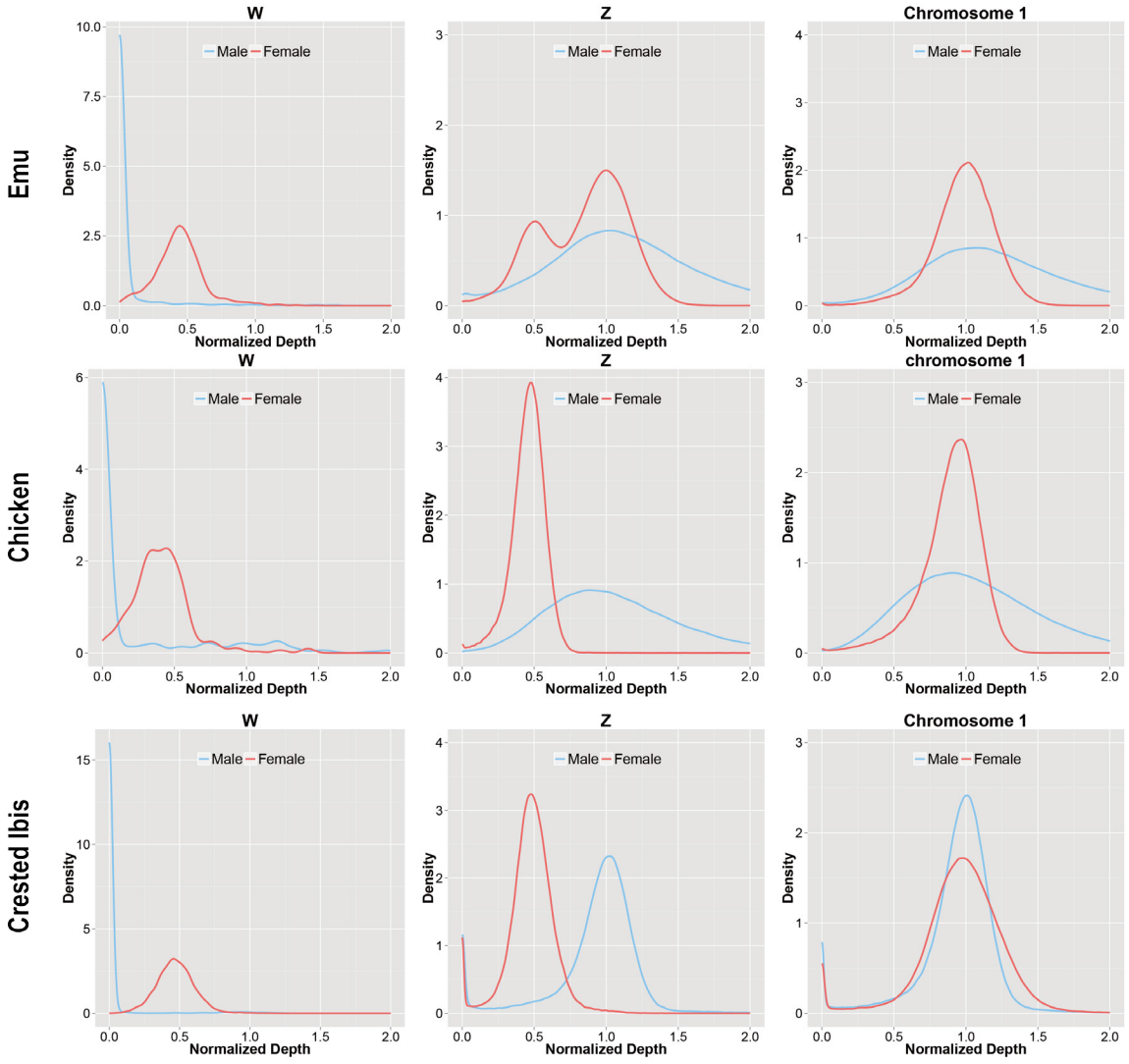


figure S4. Male and female read depths of Crested Ibis W-linked genes

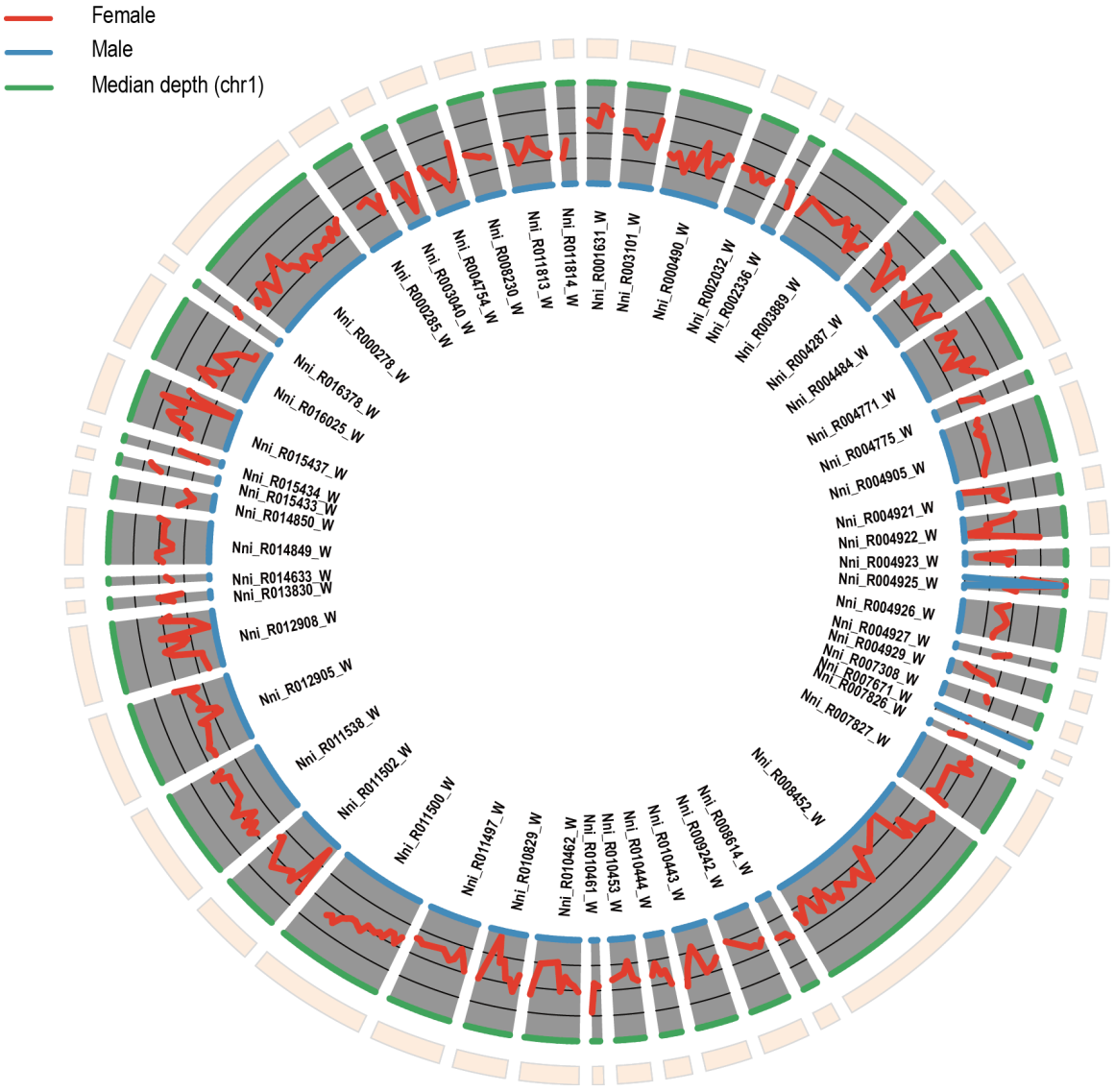


figure S5. Comparing candidate chicken W-linked scaffolds identified in this work to the chicken reference genome

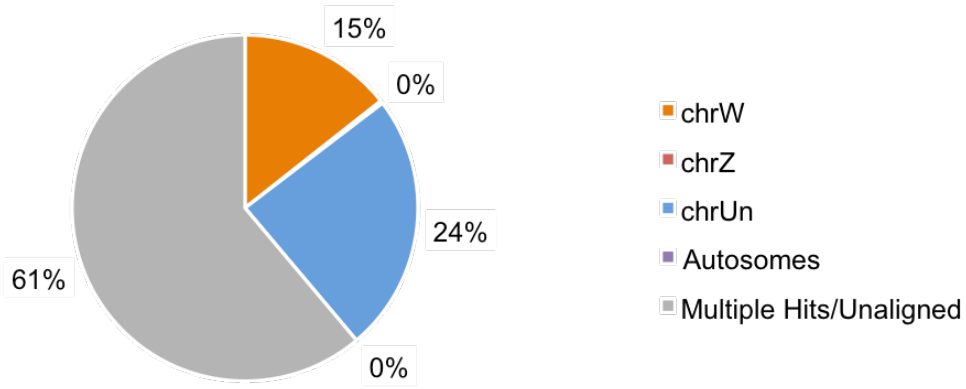


figure S6. Comparing chicken W-linked genes derived from genomic assembly vs. those from transcriptomic assembly

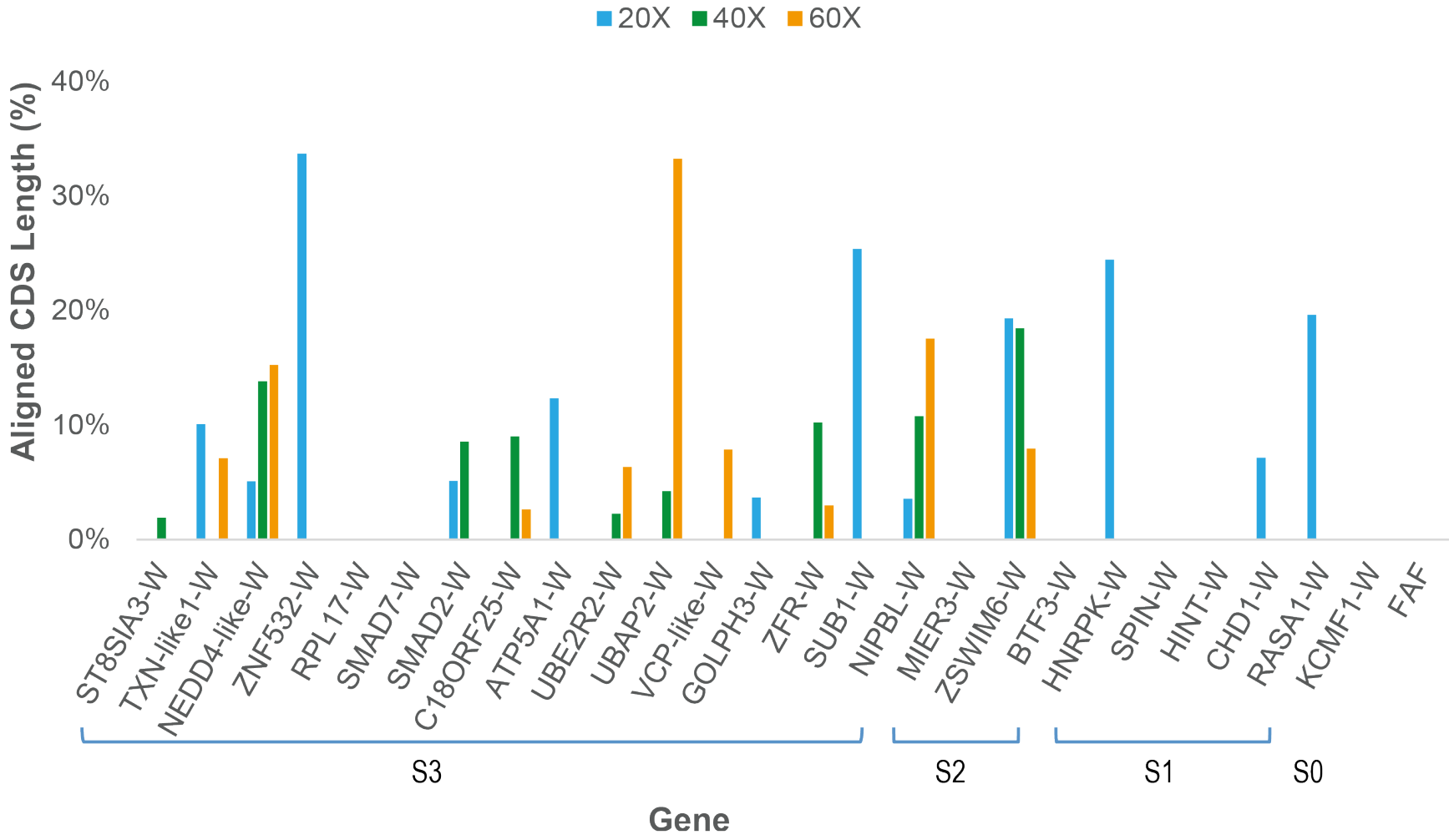


figure S7. Comparing Ostrich W-linked genes derived from genomic assembly vs. those from transcriptomic assembly

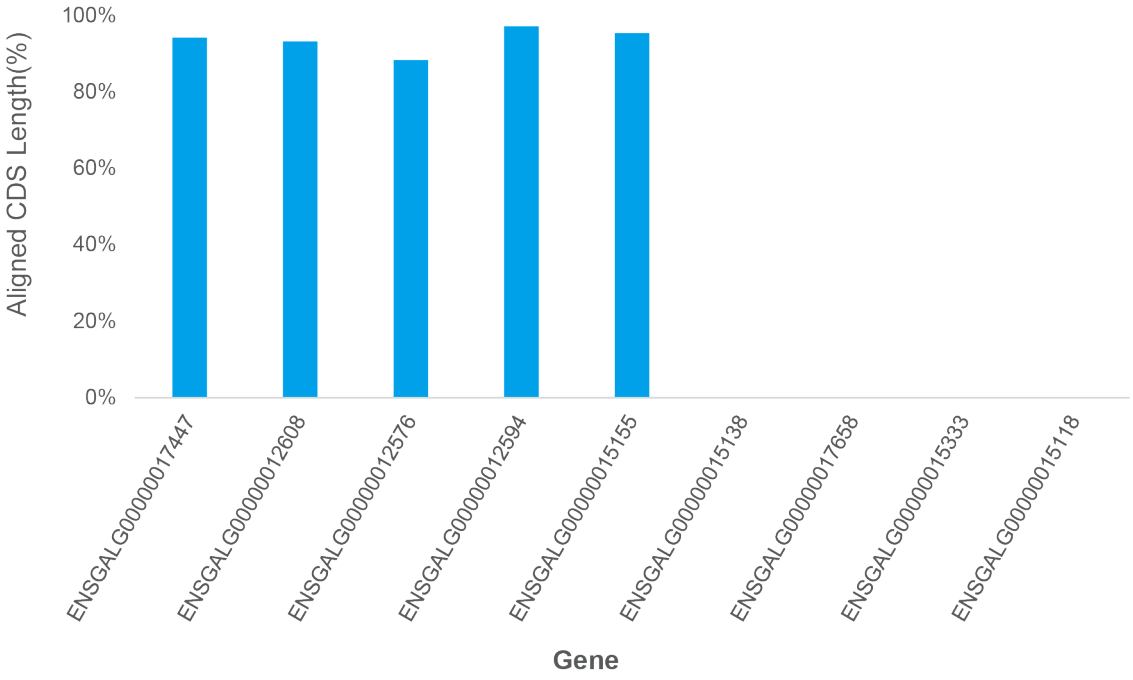


figure S8. The length distribution of annotated transcripts for Z and W linked genes in different species.

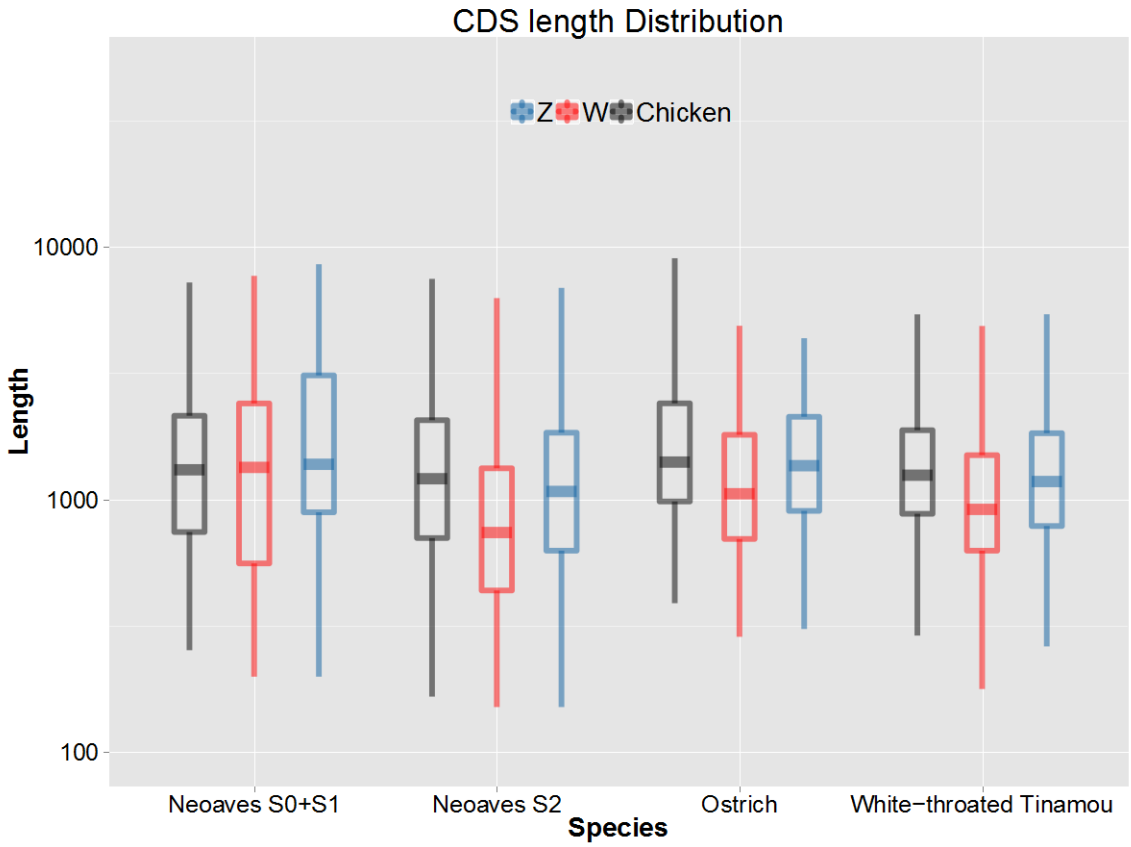


figure S9. The distribution of Z/W pairwise alignment identities along the Z chromosome

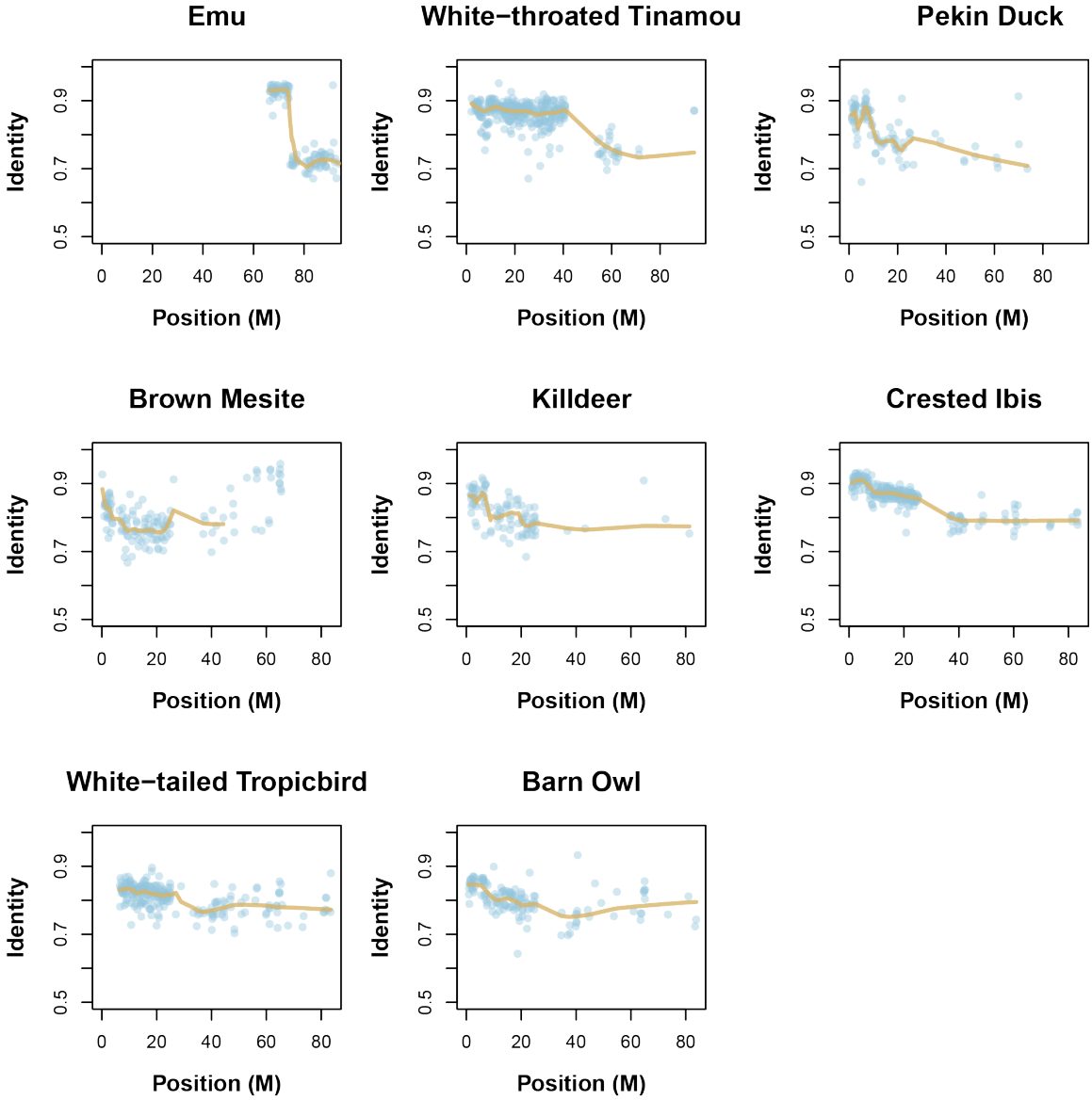


figure S10. Pairwise Z/W alignment identities are significantly different amongst strata

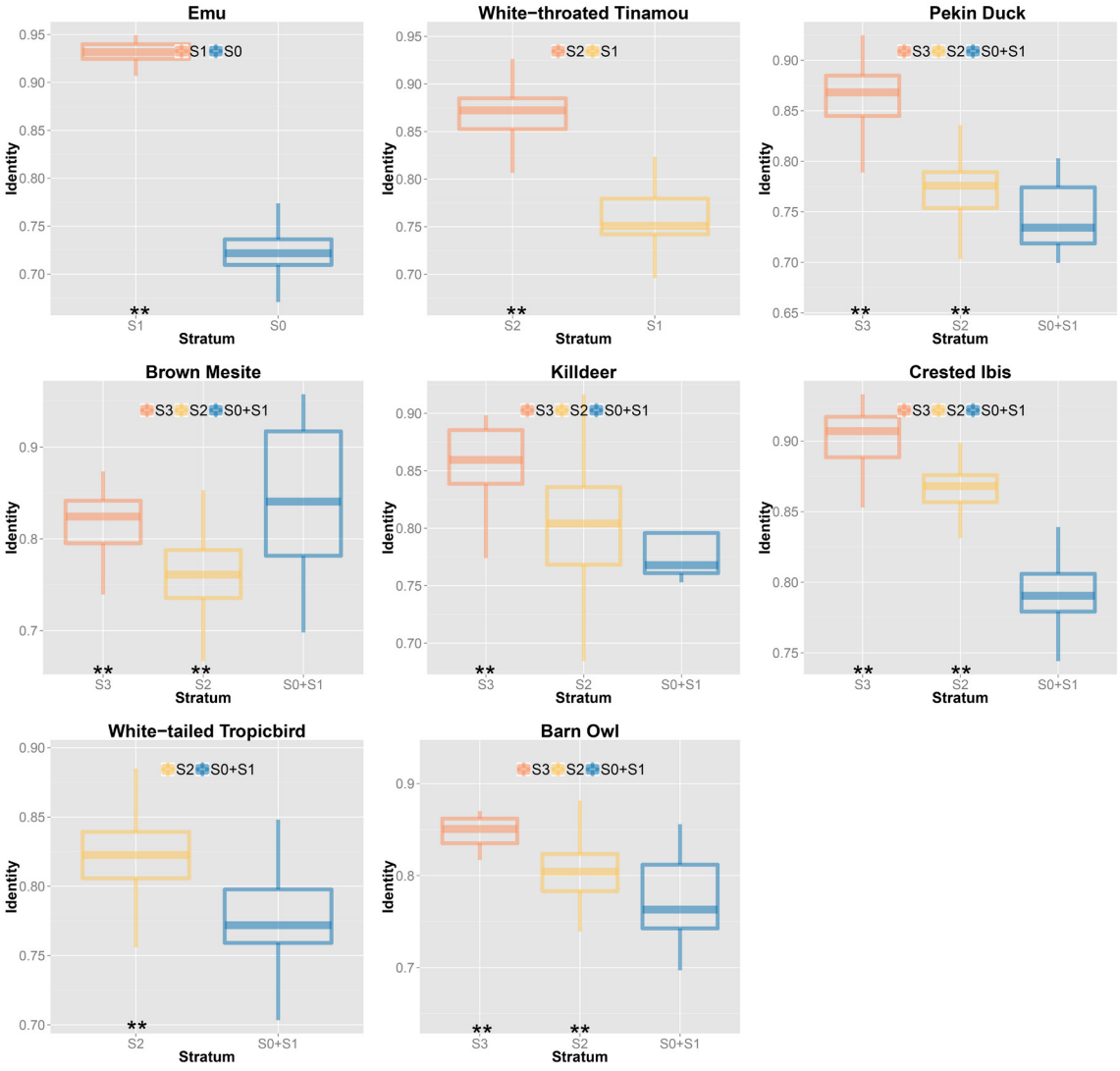


figure S11. Pairwise d*S* between Z/W gametologs are significantly different amongst strata

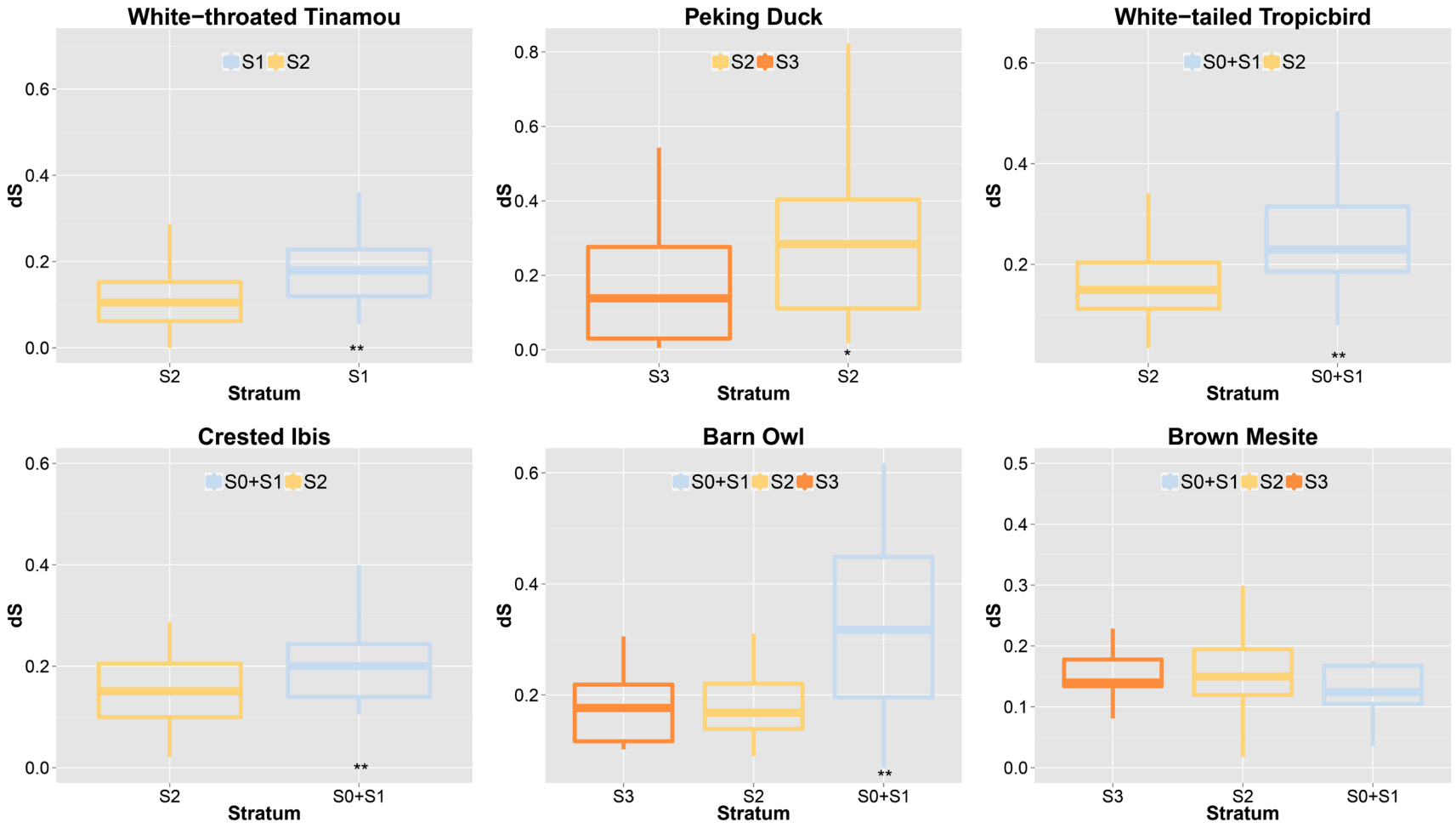


figure S12. Gene trees for Z and W linked gametologs in S0 region.

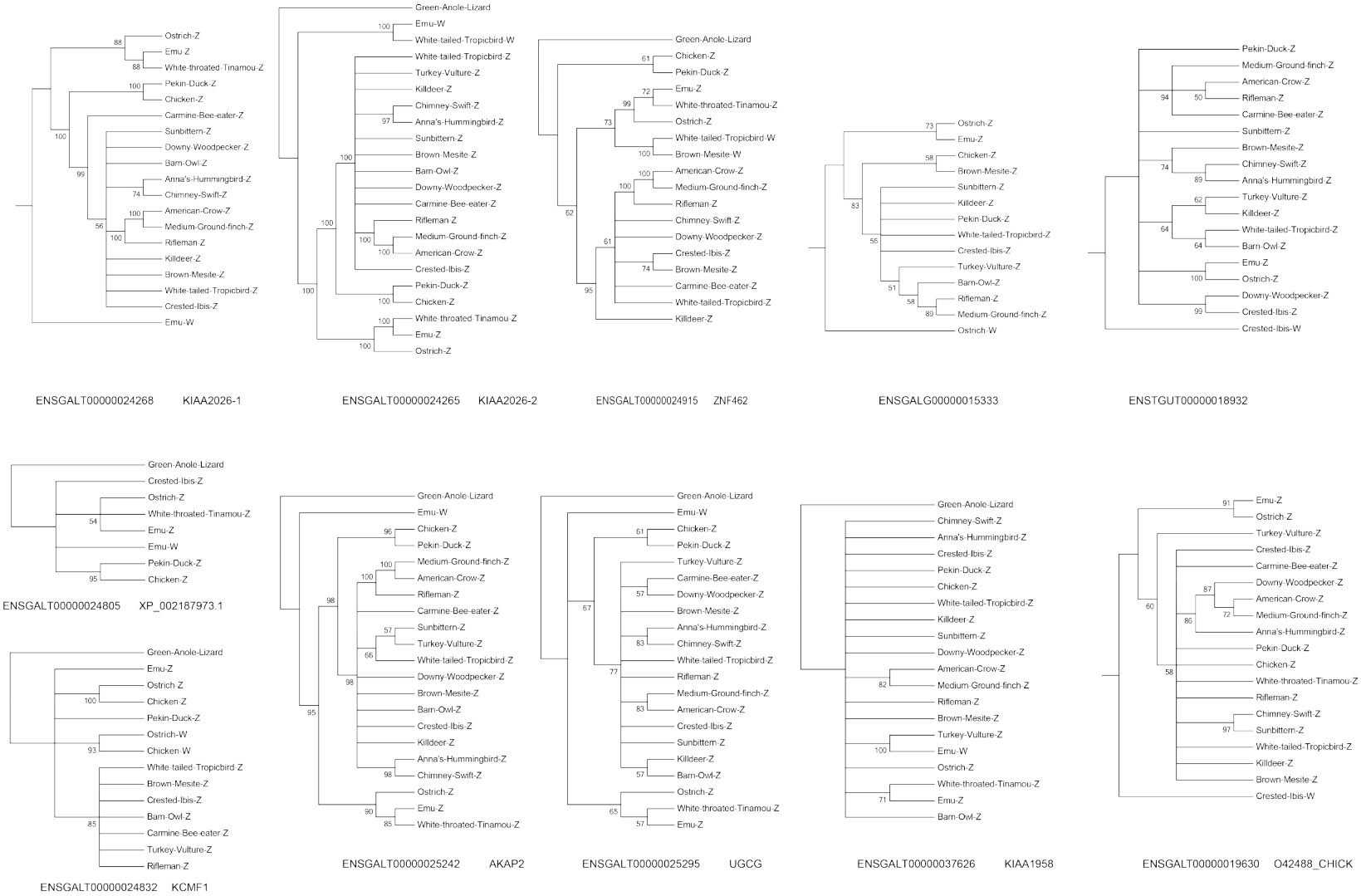


figure S13. Pattern of evolutionary strata in Neoaves when mapping their sequence scaffolds using the ostrich genome as a reference

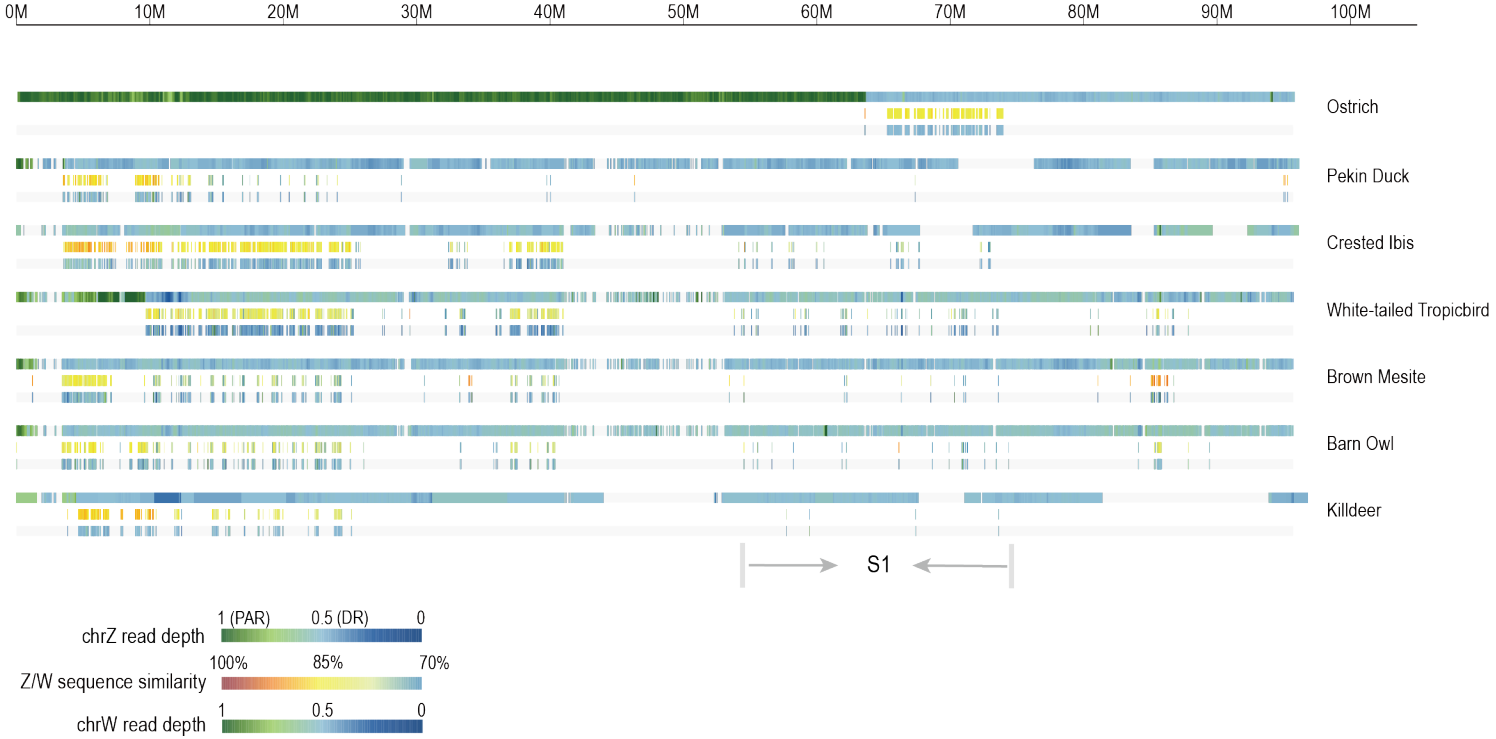


figure S14. Gene trees for Z and W linked gametologs in S1 region

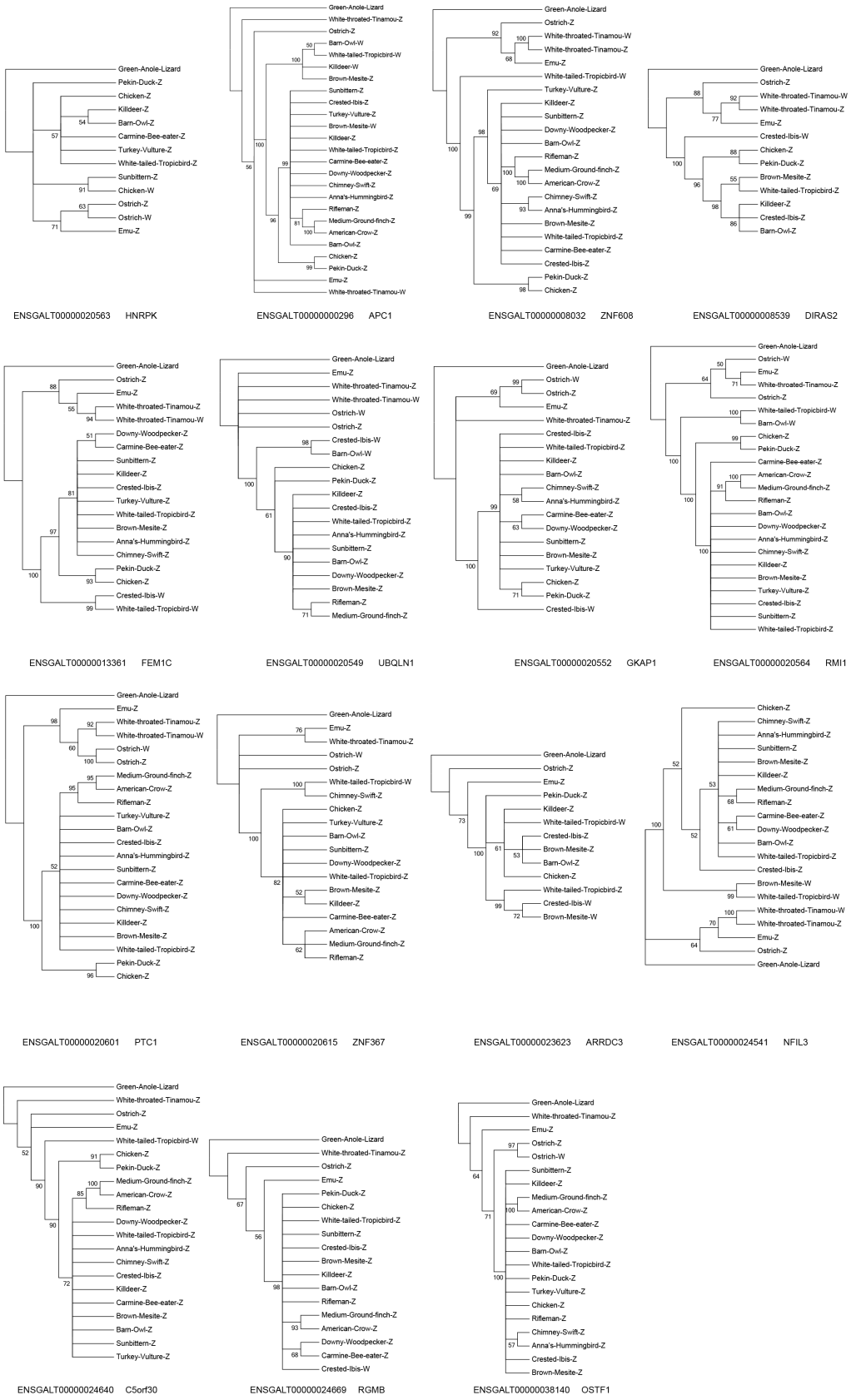


figure S15. Comparison of strata definition between this study and previous study of chicken

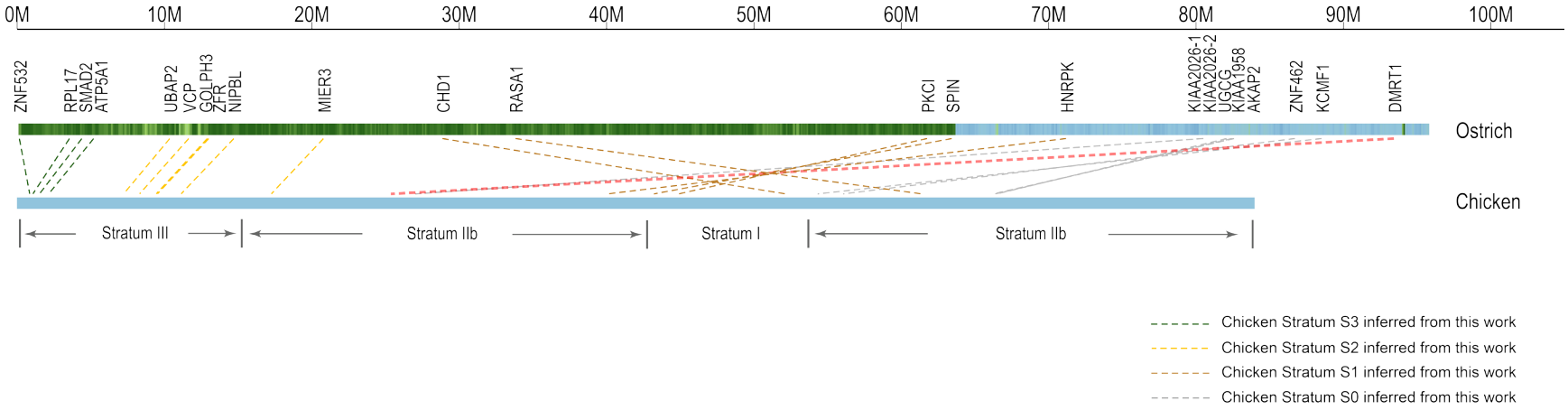
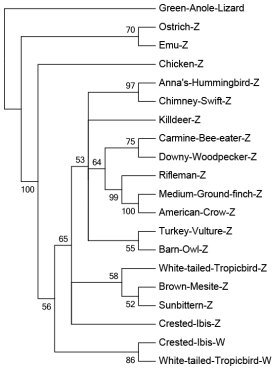
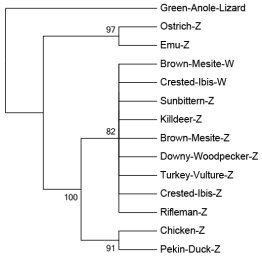


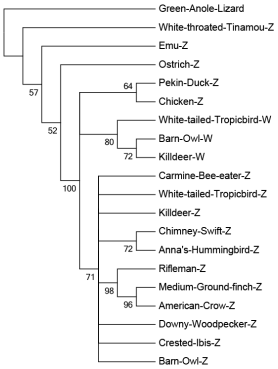
figure S16. Gene trees for Z and W linked gametologs in Neoaves S2 region



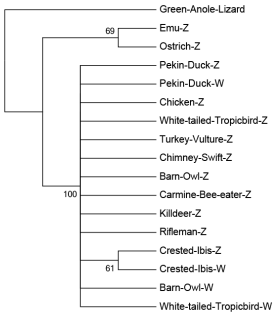
ENSGALT00000003737 RUSC2



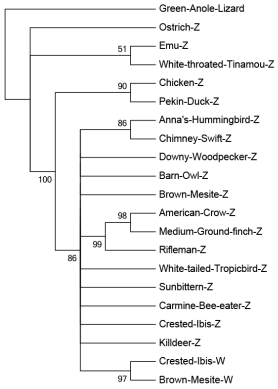
ENSGALT00000023766 KIF2A



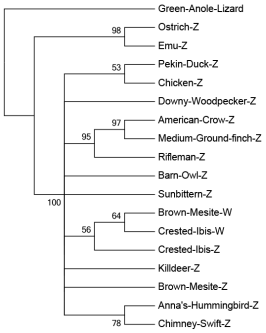
ENSGALT00000023926 JMY



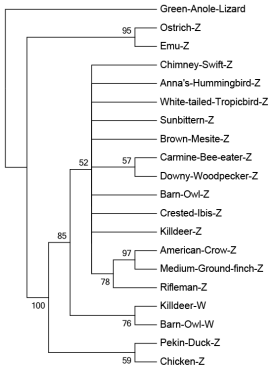
ENSGALT00000023979 ZNF131



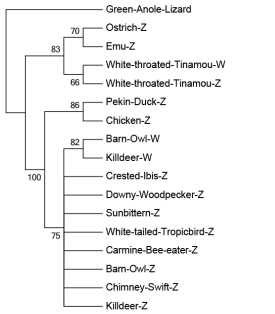
ENSGALT00000023981 LOC768482



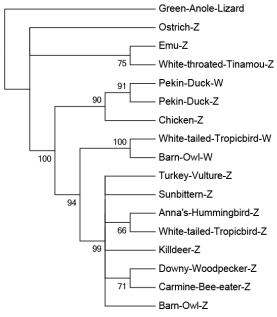
ENSGALT00000023990 NNT



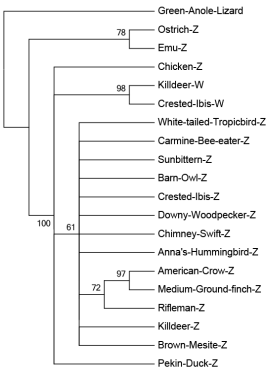
ENSGALT00000024114 HMGR



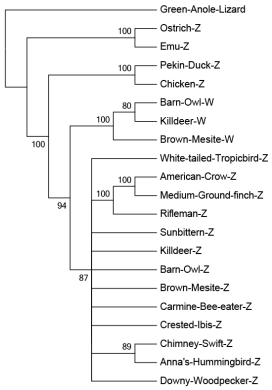
ENSGALT00000024118 COL4A3BP



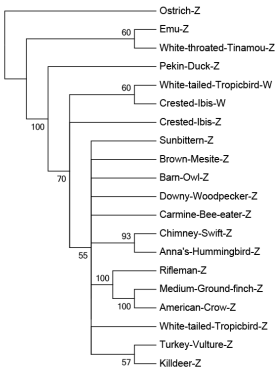
ENSGALT00000024188 MAP1B



ENSGALT00000038520 NSA2



ENSGALT00000038678 ERBB2IP



ENSTGUT00000002136 XP_002192423.1

figure S17. Gene trees for Z and W linked gametologs in Neognathae S3 region

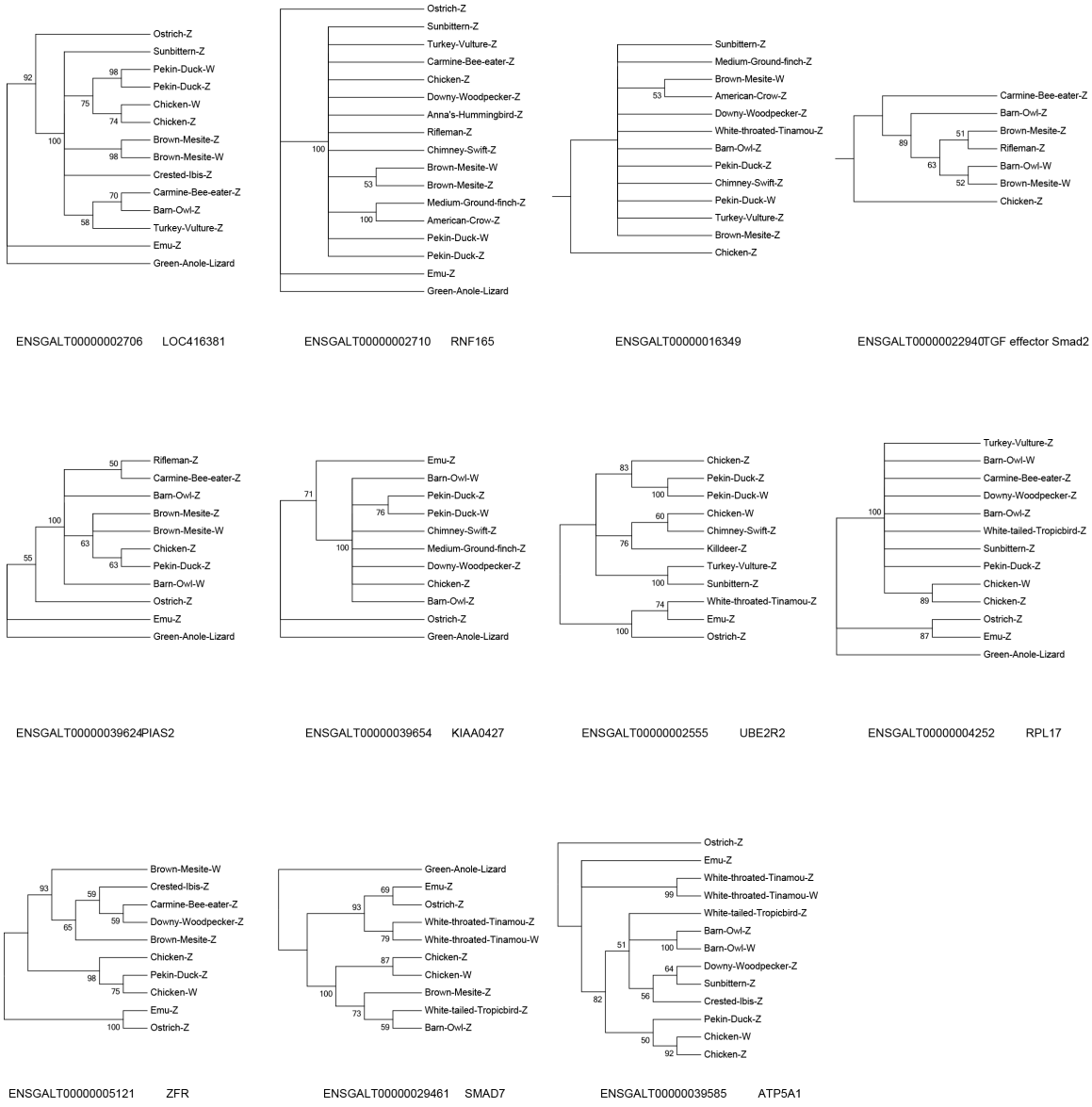


figure S18. Comparison of gene syntenies of ostrich vs. other reptile species on the ostrich Z chromosome

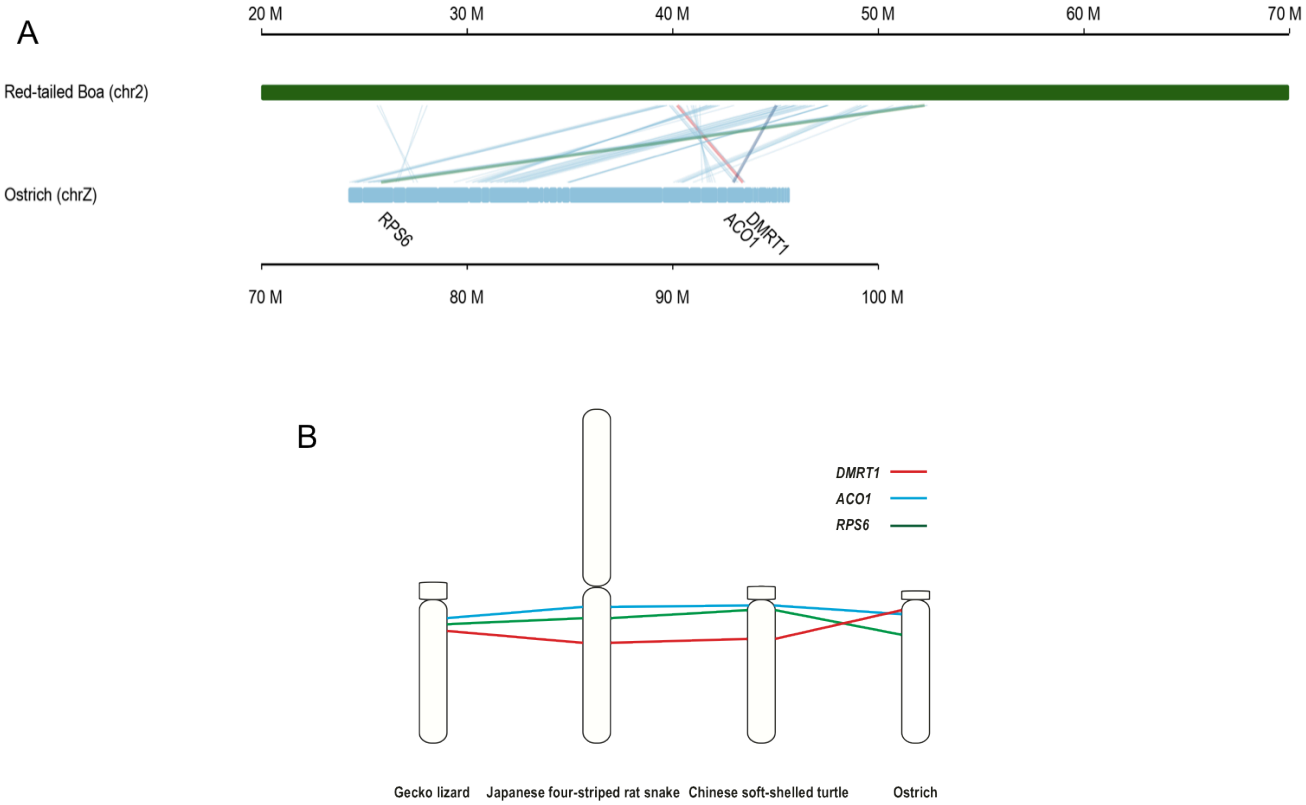


figure S19. Comparison of gene synteny of birds vs. lizard at Neoaves stratum S3

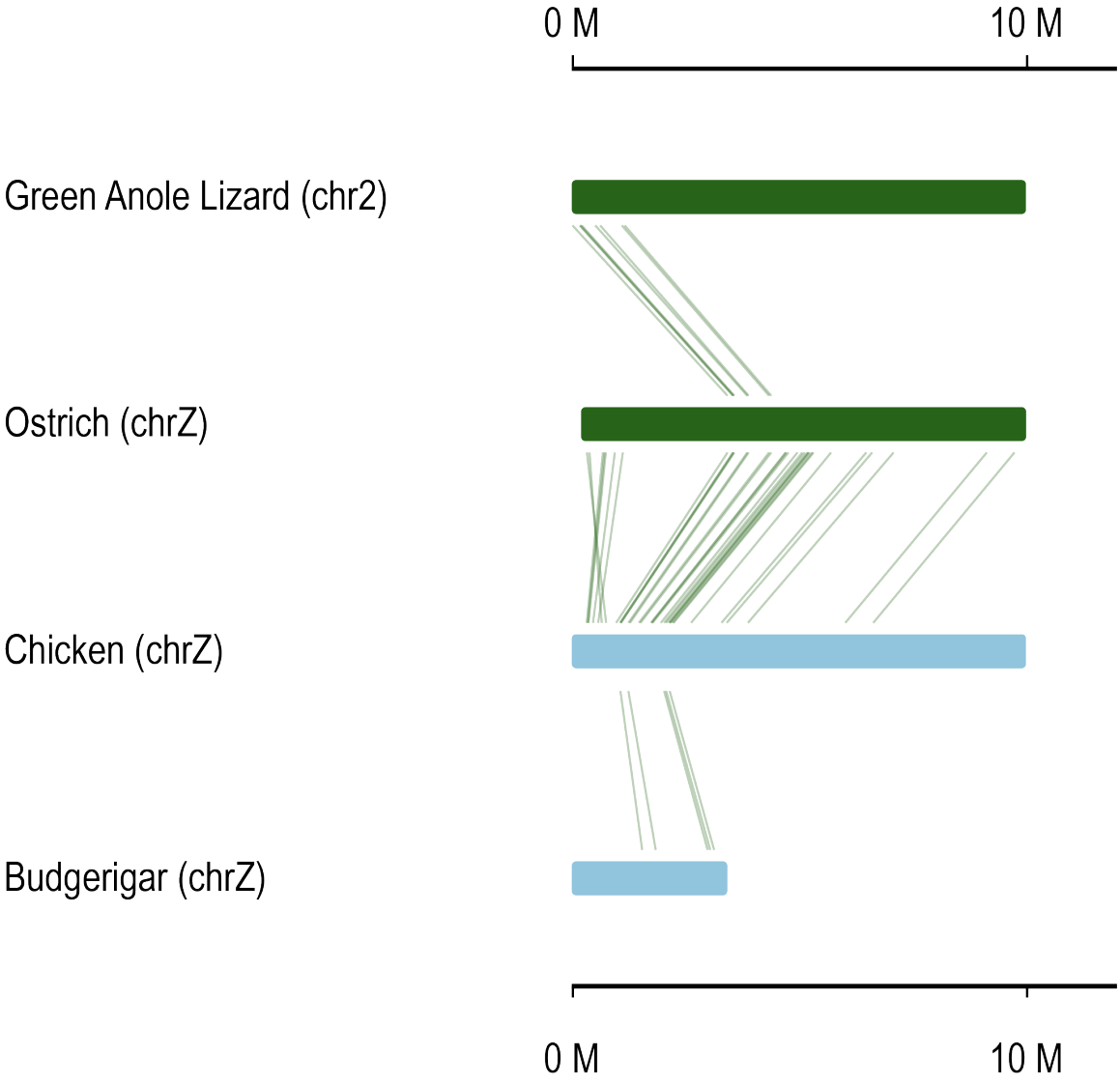


figure S20. Elevated rates of evolution of W-linked genes relative to their Z-linked homologs.

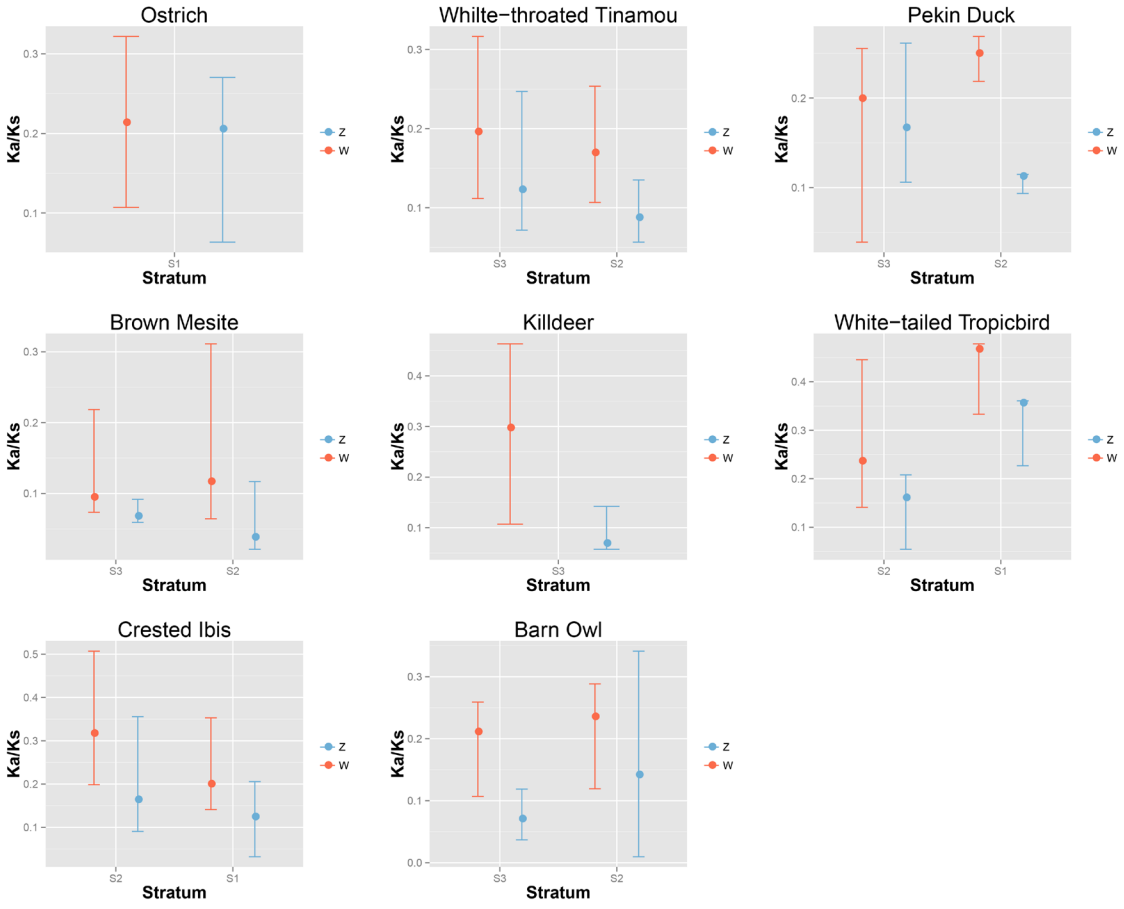


figure S21 Comparing expression level between Z and W gametologs in emu and duck

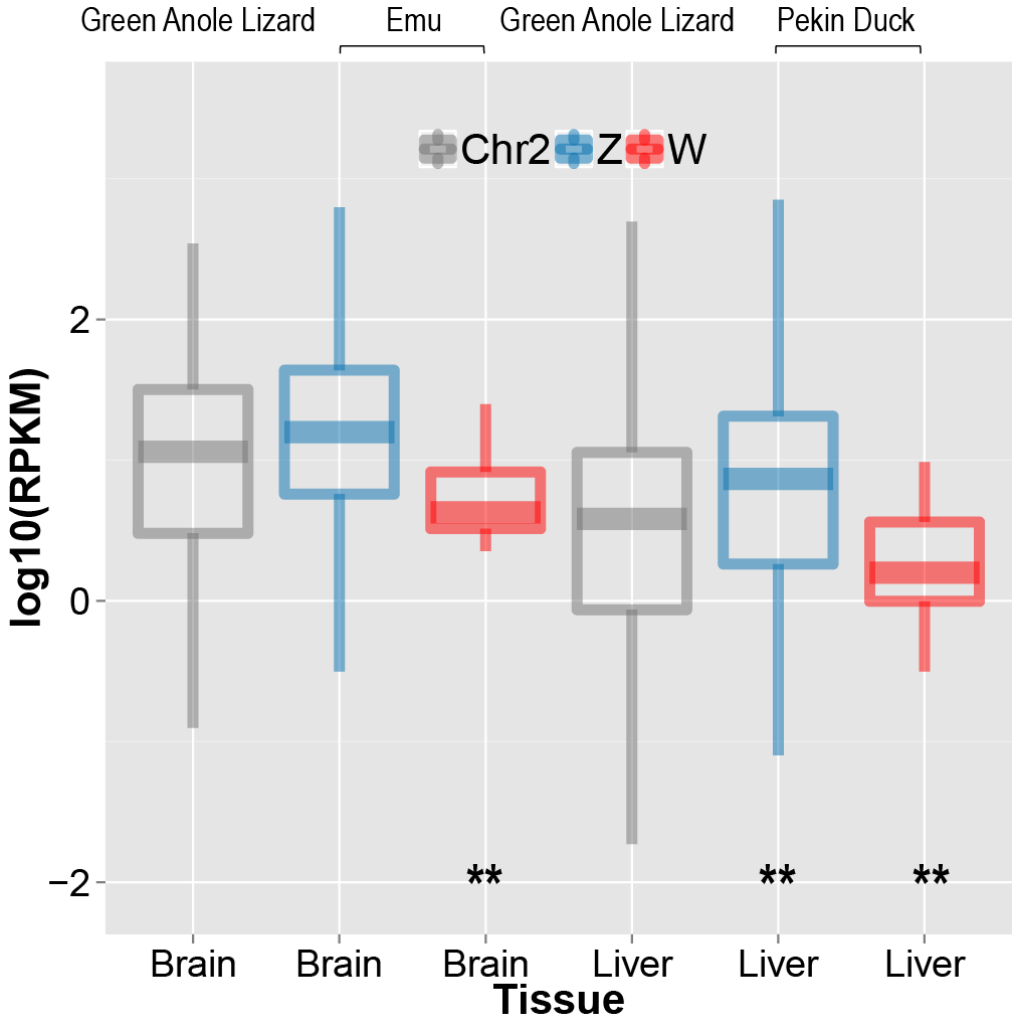


figure S22. Comparing expression level between Z and W gametologs in ostrich

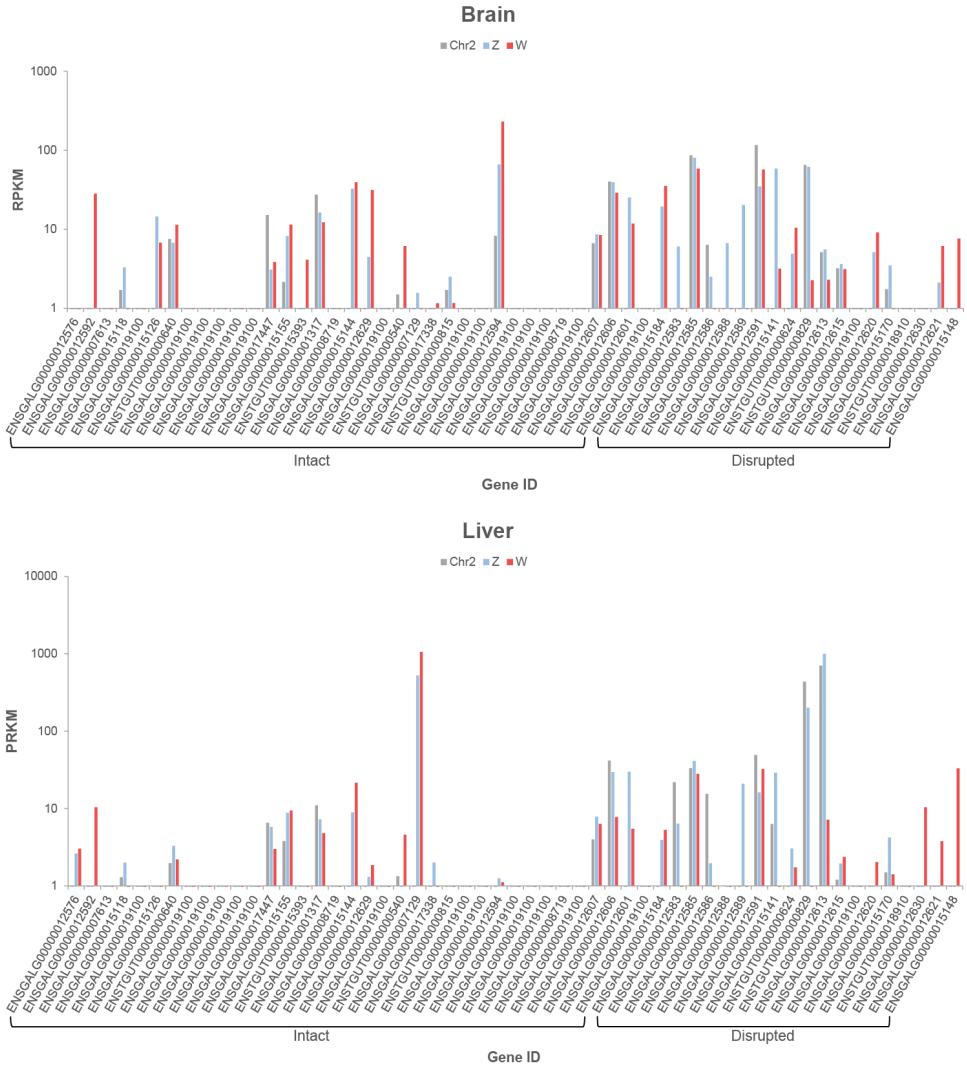
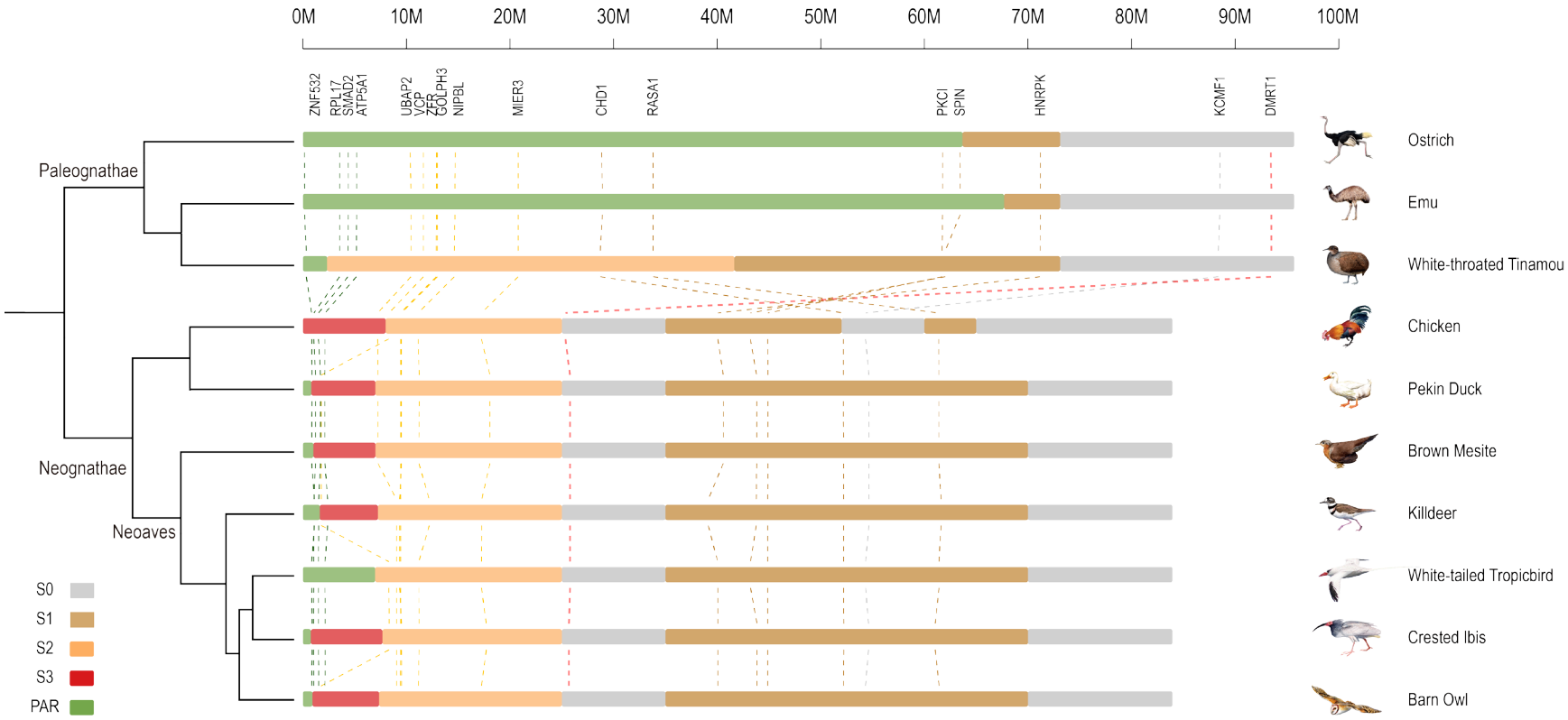


figure S23. Map of avian evolutionary strata



Supplementary References

1. Y. Tsuda, C. Nishida-Umehara, J. Ishijima, K. Yamada, Y. Matsuda, Comparison of the Z and W sex chromosomal architectures in elegant crested tinamou (Eudromia elegans) and ostrich (Struthio camelus) and the process of sex chromosome differentiation in palaeognathous birds. *Chromosoma* **116**, 159-173 (2007).
2. G. Zhang *et al.*, Comparative genomics across modern bird species reveal insights into avian genome evolution and adaptation. *Science* **346**, 1251385 (2014).
3. L. Ye *et al.*, A vertebrate case study of the quality of assemblies derived from next-generation sequences. *Genome Biol* **12**, R31 (2011).
4. W. L. Fan *et al.*, Genome-wide patterns of genetic variation in two domestic chickens. *Genome Biol Evol* **5**, 1376-1392 (2013).
5. B. Vicoso, V. B. Kaiser, D. Bachtrog, Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 6453-6458 (2013).
6. R. Li *et al.*, De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* **20**, 265-272 (2010).
7. D. W. Bellott *et al.*, Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature* **466**, 612-616 (2010).
8. B. Brunner *et al.*, Genomic organization and expression of the doublesex-related gene cluster in vertebrates and detection of putative regulatory regions for *DMRT1*. *Genomics* **77**, 8-17 (2001).
9. C. Ottolenghi, M. Fellous, M. Barbieri, K. McElreavey, Novel paralogy relations among human chromosomes support a link between the phylogeny of *doublesex*-related genes and the evolution of sex determination. *Genomics* **79**, 333-343 (2002).
10. R. S. Harris, *Improved pairwise alignment of genomic DNA*. (ProQuest, 2007).
11. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
12. H. Li *et al.*, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
13. B. Vicoso, V. B. Kaiser, D. Bachtrog, Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. *Proc. Natl. Acad. Sci. U. S. A.*, (2013).
14. K. Nam, H. Ellegren, The chicken (Gallus gallus) Z chromosome contains at least three nonlinear evolutionary strata. *Genetics* **180**, 1131-1136 (2008).
15. A. E. Wright, H. K. Moghadam, J. E. Mank, Trade-off between selection for dosage compensation and masculinization on the avian Z chromosome. *Genetics* **192**, 1433-1445 (2012).

16. D. E. Dimcheff, S. V. Drovetski, D. P. Mindell, Phylogeny of Tetraoninae and other galliform birds using mitochondrial 12S and ND2 genes. *Mol. Phylogenet. Evol.* **24**, 203-215 (2002).
17. Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586-1591 (2007).
18. T. Miyata, H. Hayashida, K. Kuma, K. Mitsuyasu, T. Yasunaga, Male-driven molecular evolution: a model and nucleotide sequence analysis. *Cold Spring Harb Symp Quant Biol* **52**, 863-867 (1987).
19. E. Axelsson, N. G. Smith, H. Sundstrom, S. Berlin, H. Ellegren, Male-biased mutation rate and divergence in autosomal, z-linked and w-linked introns of chicken and Turkey. *Mol. Biol. Evol.* **21**, 1538-1547 (2004).
20. A. Bartosch-Harlid, S. Berlin, N. G. Smith, A. P. Moller, H. Ellegren, Life history and the male mutation bias. *Evolution* **57**, 2398-2406 (2003).
21. E. D. Jarvis *et al.*, Avian Genome Consortium: Whole genome analyses resolve the early branches to the tree of life of modern birds. *submitted*, (2014).
22. W. J. Kent, BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656-664 (2002).
23. E. Birney, M. Clamp, R. Durbin, GeneWise and Genomewise. *Genome Res.* **14**, 988-995 (2004).
24. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792-1797 (2004).
25. G. Talavera, J. Castresana, Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564-577 (2007).
26. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
27. S. Adolfsson, H. Ellegren, Lack of dosage compensation accompanies the arrested stage of sex chromosome evolution in ostriches. *Mol. Biol. Evol.* **30**, 806-810 (2013).
28. C. Trapnell, L. Pachter, S. L. Salzberg, TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111 (2009).
29. J. Alfoldi *et al.*, The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* **477**, 587-591 (2011).
30. K. R. Bradnam *et al.*, Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *Gigascience* **2**, 10 (2013).
31. K. L. Ayers *et al.*, RNA sequencing reveals sexually dimorphic gene expression before gonadal differentiation in chicken and allows comprehensive annotation of the W-chromosome. *Genome Biol* **14**, R26 (2013).
32. H. P. Yazdi, H. Ellegren, Old but not (so) degenerated-slow evolution of largely

homomorphic sex chromosomes in ratites. *Mol. Biol. Evol.* **31**, 1444-1453 (2014).

33. D. Cortez *et al.*, Origins and functional evolution of Y chromosomes across mammals. *Nature* **508**, 488-493 (2014).
34. A. Kawai *et al.*, The ZW sex chromosomes of Gekko hokouensis (Gekkonidae, Squamata) represent highly conserved homology with those of avian species. *Chromosoma* **118**, 43-51 (2009).