INDEPENDENT STRATUM FORMATION ON THE AVIAN SEX CHROMOSOMES REVEALS INTER-CHROMOSOMAL GENE CONVERSION AND PREDOMINANCE OF PURIFYING SELECTION ON THE W CHROMOSOME

Alison E. Wright,1,2,3 Peter W. Harrison,2 Stephen H. Montgomery,2 Marie A. Pointer,1 and Judith E. Mank2
1Department of Zoology, Edward Grey Institute, University of Oxford, Oxford OX1 3PS, United Kingdom
2Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, United Kingdom
3E-mail: alison.e.wright@ucl.ac.uk

Received May 4, 2014
Accepted July 15, 2014

We used a comparative approach spanning three species and 90 million years to study the evolutionary history of the avian sex chromosomes. Using whole transcriptomes, we assembled the largest cross-species dataset of W-linked coding content to date. Our results show that recombination suppression in large portions of the avian sex chromosomes has evolved independently, and that long-term sex chromosome divergence is consistent with repeated and independent inversions spreading progressively to restrict recombination. In contrast, over short-term periods we observe heterogeneous and locus-specific divergence. We also uncover four instances of gene conversion between both highly diverged and recently evolved gametologs, suggesting a complex mosaic of recombination suppression across the sex chromosomes. Lastly, evidence from 16 gametologs reveal that the W chromosome is evolving with a significant contribution of purifying selection, consistent with previous findings that W-linked genes play an important role in encoding sex-specific fitness.

KEY WORDS: Chromosome inversion, female heterogamety, recombination suppression.

Recombination suppression initiates the divergence of sex chromosomes from an ancestral autosomal pair, and promotes the degeneration of the W (in female heterogamety) and Y (in male heterogamety) chromosome coding content (Ohno 1967; Charlesworth 1991; Charlesworth and Charlesworth 2000; Charlesworth et al. 2005; Bachtrog et al. 2008). In spite of the pivotal role recombination suppression plays in sex chromosome evolution, our understanding of the underlying mechanisms and dynamics is largely limited to either highly divergent sex chromosomes (e.g., mammals and Drosophila), or incipient sex chromosome systems, such as those observed in plants and fish (Chibalina and Filatov 2011; Muyle et al. 2012; Bergero et al. 2013; Natri et al. 2013; Qiu et al. 2013).

Older sex chromosomes support the classic model of sex chromosome evolution, where intra-chromosomal rearrangements are proposed to initiate the suppression of recombination across large regions (Charlesworth et al. 2005). Theoretically, these inversions result in instantaneous recombination suppression between the sex chromosomes. This leaves a spatial signature of clusters of gametologs, orthologous genes on the X and Y (or Z and W) chromosomes, of similar divergence, often referred to as strata. Strata are thought to form independently in a...
stepwise process over millions of years. At least four chromosomal rearrangements occurring over 300 million years have been documented on the human X and Y chromosomes (Lahn and Page 1999; Ross et al. 2005; Lemaître et al. 2009; Wilson and Makova 2009; Pandey et al. 2013; Bellott et al. 2014), and there is evidence of four strata on the avian sex chromosomes spanning 130 million years (Wright et al. 2012).

Newly evolved sex chromosomes show less concordant support for the strata model. The threespine stickleback (Gasterosteus aculeatus) and white campion (Silene latifolia) X and Y chromosomes originated around 30 million years ago (Mya) and less than 10 Mya, respectively. Work on these systems indicates that recombination suppression evolves heterogeneously along the length of the chromosome (Chibalina and Filatov 2011; Muyle et al. 2012; Bergero et al. 2013; Natri et al. 2013; Qiu et al. 2013; Roesti et al. 2013). This suggests that although selection to suppress recombination between the X and Y chromosomes occurs over large regions, within those regions, the effects are variable and genetic exchange between the sex chromosomes persists in some places.

We need to fully integrate short- and long-term views of sex chromosome evolution to provide a complete temporal overview of the dynamics of recombination suppression. Additionally, much of our understanding of sex chromosome evolution is based on studies of neo-sex chromosomes in Drosophila (Bachtrog et al. 2008; Kaiser et al. 2011; Zhou and Bachtrog 2012). These have been highly informative, however, males are achiasmatic in Drosophila and therefore cannot be used to study progressive recombination suppression events. Instead, comparative analyses of sex chromosome divergence across a range of species with recombination in both sexes are required to provide an evolutionary framework to characterize the temporal dynamics of sex chromosome evolution.

The avian sex chromosomes are homologous across the entire clade, however recombination suppression between the Z and W has evolved repeatedly in different lineages across 130 million years (Tsuda et al. 2007; Suh et al. 2011). We use the repeated evolution of avian sex chromosome strata to create a cohesive view of short-, medium-, and long-term dynamics in Z-W divergence patterns. To do this, we assembled the largest cross-species dataset of W-linked coding content to date spanning the Galiformes (the landfowl) and the Anseriformes (the waterfowl). These two sister orders, the Galloanserae, last shared a common ancestor approximately 90 Mya (van Tuinen and Hedges 2001). Our analysis indicates that the majority of the Z and W chromosomes formed independently in each of these orders. Additionally, our results suggest ongoing recombination between gametologs in the most recent region of the Anseriform sex chromosomes, and gene conversion throughout the sex chromosomes.

Materials and Methods
IDENTIFICATION OF W-LINKED GENES

Previous efforts to identify W-linked sequences in birds (Chen et al. 2012) have resulted primarily in noncoding sequence, which makes comparisons to the Z chromosome difficult. To expand the known coding content of both the mallard duck (Anas platyrhynchos) and wild turkey (Meleagris gallopavo) W chromosomes, we used a combined approach based on RNA-seq data and sequence similarity to known red jungle fowl (Gallus gallus) W-linked genes.

RNA-seq data were obtained from captive populations of A. platyrhynchos and M. gallopavo, taken at the start of their first breeding season (year 1 for A. platyrhynchos, year 2 for M. gallopavo). All samples were collected in accordance with national guidelines and with permission from institutional ethical review committees. The left gonad was collected separately from five male and five female A. platyrhynchos, and seven male and five female M. gallopavo. The spleen was collected from five male and five female A. platyrhynchos and four male and two female M. gallopavo. The samples were homogenized and stored in RNAlater until preparation. We extracted RNA using the Animal Tissue RNA Kit (Qiagen) and The Wellcome Trust Centre for Human Genetics, University of Oxford, prepared and barcoded samples using standard methods. RNA was sequenced on an Illumina HiSeq 2000 resulting in on average 26 million 100 bp paired-end reads per sample.

The data were assessed for quality using FastQC version 0.10.1 (www.bioinformatics.babraham.ac.uk/projects/fastqc) and filtered using Trimmomatic version 0.22 (Lohse et al. 2012). Specifically, reads with residual adaptor sequences were removed and reads were trimmed if the leading or trailing bases had a Phred score <4 or the sliding window average Phred score over four bases was <15. Post filtering, reads were removed if either read pair was <25 bases in length. Filtered reads from each species were mapped to the respective reference genomes obtained from Ensembl version 74 (Fliece et al. 2013; M. gallopavo version 2.01/GCA_000146605.1; Dalloul et al. 2010; Dalloul et al. 2014) and A. platyrhynchos version 1.0/GCA_000355885.1; Huang et al. 2013) using TopHat2 version 2.0.9 (Trapnell et al. 2009; Kim et al. 2013) and Bowtie2 version 2.1.1 (Langmead et al. 2009; Langmead and Salzberg 2012). For recently evolved gametologs with high sequence similarity, it is possible that Z-linked reads will spuriously map to W-linked exons. To avoid this and allow accurate identification of female-limited genes and therefore putative W-linkage, both reads of a pair had to map concordantly to the reference sequence and no mismatches in the alignment were permitted. This strategy, although essential to accurately differentiate gametologs, could fail to identify W-linked genes.

Materials and Methods
IDENTIFICATION OF W-LINKED GENES

Previous efforts to identify W-linked sequences in birds (Chen et al. 2012) have resulted primarily in noncoding sequence, which makes comparisons to the Z chromosome difficult. To expand the known coding content of both the mallard duck (Anas platyrhynchos) and wild turkey (Meleagris gallopavo) W chromosomes, we used a combined approach based on RNA-seq data and sequence similarity to known red jungle fowl (Gallus gallus) W-linked genes.

RNA-seq data were obtained from captive populations of A. platyrhynchos and M. gallopavo, taken at the start of their first breeding season (year 1 for A. platyrhynchos, year 2 for M. gallopavo). All samples were collected in accordance with national guidelines and with permission from institutional ethical review committees. The left gonad was collected separately from five male and five female A. platyrhynchos, and seven male and five female M. gallopavo. The spleen was collected from five male and five female A. platyrhynchos and four male and two female M. gallopavo. The samples were homogenized and stored in RNAlater until preparation. We extracted RNA using the Animal Tissue RNA Kit (Qiagen) and The Wellcome Trust Centre for Human Genetics, University of Oxford, prepared and barcoded samples using standard methods. RNA was sequenced on an Illumina HiSeq 2000 resulting in on average 26 million 100 bp paired-end reads per sample.

The data were assessed for quality using FastQC version 0.10.1 (www.bioinformatics.babraham.ac.uk/projects/fastqc) and filtered using Trimmomatic version 0.22 (Lohse et al. 2012). Specifically, reads with residual adaptor sequences were removed and reads were trimmed if the leading or trailing bases had a Phred score <4 or the sliding window average Phred score over four bases was <15. Post filtering, reads were removed if either read pair was <25 bases in length. Filtered reads from each species were mapped to the respective reference genomes obtained from Ensembl version 74 (Fliece et al. 2013; M. gallopavo version 2.01/GCA_000146605.1; Dalloul et al. 2010; Dalloul et al. 2014) and A. platyrhynchos version 1.0/GCA_000355885.1; Huang et al. 2013) using TopHat2 version 2.0.9 (Trapnell et al. 2009; Kim et al. 2013) and Bowtie2 version 2.1.1 (Langmead et al. 2009; Langmead and Salzberg 2012). For recently evolved gametologs with high sequence similarity, it is possible that Z-linked reads will spuriously map to W-linked exons. To avoid this and allow accurate identification of female-limited genes and therefore putative W-linkage, both reads of a pair had to map concordantly to the reference sequence and no mismatches in the alignment were permitted. This strategy, although essential to accurately differentiate gametologs, could fail to identify W-linked genes.
when the entire locus is not encompassed on a single scaffold in the genome assembly.

To identify polymorphisms between Z and W coding sequences the mapping criteria were relaxed. For each sample, we used Cufflinks version 2.1.0 (Trapnell et al. 2010, 2013) to estimate transcript abundances for Ensembl annotated genes as fragments per kilobase per million mappable reads. We identified strongly female-biased and female-limited genes as putative W-linked genes using gene expression profiles in males and females (Fig. S1).

Putative W-linked genes were also identified by high sequence similarity to known G. gallus W-linked genes. *Gallus gallus* (Galgal4.0/GCA_000002315.2), *M. gallopavo*, and *A. platyrhynchos* cDNA sequences were obtained from Ensembl version 74 (Flicek et al. 2013) and for each species the longest transcript for each gene was identified. We determined putative orthology using BLAST version 2.2.26 (Altschul et al. 1990) with an e-value cutoff of $1 \times 10^{-10}$ and identified putative *A. platyrhynchos* and *M. gallopavo* W-linked genes that had the highest BLAST score to known *G. gallus* W-linked genes.

For both *A. platyrhynchos* and *M. gallopavo*, we validated putative W-linked genes by PCR using genomic DNA from three individuals of each sex.

**IDENTIFICATION OF GAMETOLOGS**

Previously, we used newly identified *G. gallus* gametologs to identify four strata on the Z chromosome (Wright et al. 2012). Subsequent to this work, the Ensembl annotation of the *G. gallus* W and Z chromosomes has been revised (Flicek et al. 2013) and additional W-linked genes have been identified (Ayers et al. 2013). To account for these changes, we updated gametologs and reanalyzed divergence estimates using coding sequences from Ensembl version 74 (Flicek et al. 2013) and PAML version 4.7 (Yang 2007). Revised divergence estimates were compared to previous estimates (Wright et al. 2012) with Spearman’s correlation in the rcorr function (Hmisc package [Schemper 2003; Tian et al. 2007] in R [R Core Team 2011]). Across gametologs, revised $d_S$ estimates were strongly correlated with previous estimates from Wright et al. 2012 (Spearman’s rank correlation P-value < 0.01, correlation coefficient 0.78). The $d_S$ of three newly identified gametologs were also consistent with the previously identified strata, with overlapping 95% confidence intervals (Fig. 1, Table S1), in line with the previous findings that the *G. gallus* Z chromosome was formed by four independent recombination suppression events.

*Anas platyrhynchos*, *M. gallopavo*, *G. gallus*, and *T. guttata* (taeGut3.2.4) coding sequences were obtained from Ensembl version 74 (Flicek et al. 2013), and additional *M. gallopavo* sequences were obtained from Genbank (Backstrom et al. 2005; Berlin and Ellegren 2006). For each species the longest transcript for each gene was identified. *Gallus gallus* sequences were BLASTed reciprocally against *A. platyrhynchos*, *M. gallopavo*, and *T. guttata* coding sequences using an e-value cutoff of $1 \times 10^{-10}$ and the highest BLAST score was used to identify the best BLAST hit.

For *A. platyrhynchos* and *M. gallopavo* W-linked genes, we identified orthologous Z-linked genes using the following steps. For each W-linked gene (1) we identified the orthologous *G. gallus* W-linked gene using BLAST and the *G. gallus* Z-linked ortholog from Wright et al. (2012); (2) *A. platyrhynchos* and *M. gallopavo* Z-linked orthologs were defined as the reciprocal best hit of the *G. gallus* Z-linked ortholog; (3) in a few instances there was no reciprocal best hit, often because of high sequence similarity between Z and W gametologs. For these genes, the Z-linked reciprocal ortholog was identified as the next best BLAST hit of the *G. gallus* Z-linked ortholog.

Avian genomes are particularly stable with few major genomic rearrangements (Stiglec et al. 2007) and little gene movement off and onto the avian Z chromosome, potentially due to a lack of active transposons (Toups et al. 2011). We therefore expect Z-linkage and synteny to be highly conserved across *A. platyrhynchos*, *M. gallopavo*, and *G. gallus*. We verified Z-linkage of reciprocal orthologs using positional information in the *M. gallopavo* genome assembly. The *A. platyrhynchos* genome assembly lacks annotated chromosomes and as recently diverged Z- and W-linked genes share a high degree of sequence similarity, to avoid misidentifying the Z-linked reciprocal ortholog as a W-linked paralog, we verified the reciprocal ortholog was not located on the W chromosome using PCR. However, this approach does not distinguish Z chromosome to autosomal gene duplication, but given the pronounced stability of avian genomes (Toups et al. 2011), this is unlikely to bias our results.

**DIVERGENCE OF GAMETOLOGS**

To identify the number and boundaries of evolutionary strata in *A. platyrhynchos* and *M. gallopavo*, we estimated rates of synonymous divergence ($d_S$) (Wright et al. 2012), as $d_S$ provides a relative measure of the relative length of time gametologs have differentiated from an ancestral autosomal gene. For each species, we obtained Z and W coding sequences from Biomart, and aligned the translated protein sequences with MUSCLE (Edgar 2004a,b) in MEGA5 (Tamura et al. 2011). Alignments were visually inspected and poorly aligned regions were removed. The CODEML package in PAML version 4.7 (Yang 2007) was used to calculate maximum-likelihood estimates of synonymous ($d_S$) and nonsynonymous ($d_N$) divergence in pairwise comparisons and generate standard errors. We verified $d_S \leq 1$ for all gametologs, thereby avoiding inaccurate divergence estimates due to mutational saturation and double hits (Axelsson et al. 2008).

Three *A. platyrhynchos* gametologs had $d_S < 0.02$. We verified that the annotated Z and W genes were distinct coding
sequences by identifying female-limited SNPs using RNA-seq data and W-linked SNPs from Ensembl annotated sequences.

Some gametologs had multiple W paralogs, and to identify the true W-linked ortholog we used the method outlined in Wright et al. (2012). Briefly, we used the gametologous pair with the lowest $d_S$ estimate to avoid potential problems of relaxed or diversifying selection acting on gene copies (Graur and Li 2000) likely due to gene duplications following the suppression of recombination (Busby et al. 2011). However, the A. platyrhynchos draft assembly contains short sequences that may represent partial fragments of longer W-linked genes. To avoid biasing divergence estimates on short stretches of sequence, we excluded W-linked genes if coverage of the G. gallus W-linked ortholog was <25% and the aligned length with the Z-linked ortholog was <150 bp. This resulted in the removal of three A. platyrhynchos W-linked genes from the analysis. All M. gallopavo W-linked genes met these criteria, and therefore none were excluded.

For each species, we mapped $d_S$ estimates to the Z chromosome, taking physical position from the orthologous G. gallus Z-linked ortholog. Divergence dates for gametologs were calculated using a molecular clock. Briefly, based on male-mutation bias estimates in birds, the molecular clock accounts for sex-specific mutation rates in Z- and W-linked genes. This results in a divergence time of $d_S/(3.8 \times 10^{-9})$ years (Dimcheff et al. 2002b; Hedges 2002; Harlid et al. 2003; Axellsson et al. 2004; Berlin et al. 2006; Webster et al. 2006; Nam and Ellegren 2008).

We used M. gallopavo gametologs to independently confirm the four strata previously identified in G. gallus (Wright et al. 2012) and to refine the dates at which recombination was suppressed. We assessed longer term dynamics of sex chromosome divergence using A. platyrhynchos gametologs. To do so, we built gene trees to determine whether recombination ceased independently between gametologs in each species, or before the last common ancestor of the Galloanserae. Specifically we (1) obtained coding sequences for gametologs in the G. gallus, A. platyrhynchos, M. gallopavo, and T. guttata from Biomart, and aligned the translated protein sequences with MUSCLE (Edgar 2004a, b) in MEGA5 (Tamura et al. 2011). We checked alignments by eye and removed poorly aligned regions. Taeniopygia guttata Z-linked orthologs, identified from a reciprocal BLAST with G. gallus, were used as outgroups; (2) we constructed maximum-likelihood gene trees with 1000 bootstrap replicates in MEGA5 (Tamura et al. 2011). In the absence of overlapping sequences, gene trees were constructed between G. gallus and A. platyrhynchos or M. gallopavo gametologs separately. In all cases, the T. guttata Z-linked ortholog was designated as the outgroup.

To measure the completeness of recombination suppression between gametologs, we tested for evidence of gene conversion using GENECONV (Sawyer 1999). GENECONV identifies potential gene conversion sites as long fragments of identical sequence bounded by variable sites, assessing significance by 10,000 random permutations of variable sites across the alignment. $P$-values for global fragments are corrected for multiple comparisons and sequence length. We tested for gene conversion using the multiple sequence alignments described above (used to build gene trees) and specified the –group option to look only for conversion between gametologs in G. gallus, A. platyrhynchos, and M. gallopavo. We used the –Seqtype = SILENT option to specify coding sequence and that gene conversion fragments should only be detected using silent site amino acid polymorphisms. High false-positive rates have been documented when this option is not implemented, particularly if genes are subject to contrasting selective regimes (Ezawa et al. 2006, 2010). We used the strict default -gscale value (gscale = 0) where no mismatches are permitted between aligned putative conversion fragments and repeated the analysis with a relaxed mismatch penalty (gscale = 2) to ensure the number of mismatches did not bias our analysis. GENECONV detects inner fragments, which arise from gene conversion events between the ancestors of two aligned sequences and outer fragments, which represent a conversion events originating outside of the alignment or events whose signature has been weakened by subsequent mutation. We identified significant inner gene conversion fragments between gametologs as evidence of ongoing recombination between the sex chromosomes.

On closer inspection, gene conversion events identified in this study span several exons that in turn are separated by extremely dissimilar intronic sequences. As GENECONV identifies potential gene conversion sites as long fragments of sequence bounded by variable sites, it estimates the maximum sequence length over which recombination acts. To avoid falsely identifying gene conversion sites, we excluded exons where GENECONV fragments span less than 50 bp. We used informative polymorphic sites and maximum-likelihood exon trees to establish the direction of gene conversion.

**SEQUENCE EVOLUTION OF Z- AND W-LINKED GENES**

To assess the selective regime acting on Z and W coding sequences, we used the branch models and branch-site test in the CODEML package in PAML version 4.7 (Yang 2007).

Branch models (model = 2, nssites = 0) allow $\omega (d_S/d_S)$ to vary across branches. By comparing a model where $\omega$ is estimated to one where it is fixed to equal 1 (the expected $d_S/d_S$ under neutral evolution) or 0 (strict purifying selection) it is possible to assess the contribution of neutral and purifying selection to coding sequence evolution. To do this, we used the multiple sequence alignments described above (used to build gene trees and test for gene conversion). If A. platyrhynchos and M. gallopavo gametologs ceased recombining before the last common ancestor of the Galloanserae the following gene tree was specified (T. guttata Z, (((G. gallus W, M. gallopavo W), A. platyrhynchos W),...
Results

CHARACTERISING THE CODING CONTENT OF THE AVIAN W CHROMOSOME ACROSS SPECIES

We used a combination of bioinformatics, molecular genetic, and phylogenetic methods to validate putative W-linked genes and identify previously unknown W-linked coding content in the wild turkey (M. gallopavo) and mallard duck (A. platyrhynchos). Our combined dataset comprised 14 W-linked genes in M. gallopavo and 21 in A. platyrhynchos, together with 27 previously identified W-linked genes in the red jungle fowl (G. gallus) (Table 1). Some W-linked genes are present in multiple copy number (Backstrom et al. 2005; Moghadam et al. 2012) and these W-linked genes correspond to seven gametologs in M. gallopavo, 11 gametologs in A. platyrhynchos, and 20 gametologs in G. gallus. To date, this is the largest cross-species dataset of W-linked coding sequences yet assembled.

FOUR Z CHROMOSOME STRATA ARE CONSERVED IN THE GALLIFORMES

Sex chromosome strata are identified from clusters of gametologs with similar synonymous (dS) divergence estimates (Lahn and Page 1999; Skaletsky et al. 2003), which provide a measure of the relative length of time gametologs have been isolated from each other. Previously, we identified four evolutionary strata on the G. gallus Z chromosome, where recombination was suppressed progressively over 130 million years (Wright et al. 2012).

We used seven gametologs in M. gallopavo to independently verify the G. gallus strata. Z-linked physical position was conserved between G. gallus and M. gallopavo orthologs, and each stratum is represented by at least one M. gallopavo gametolog. For all gametologs, dS estimates were consistent with orthologous G. gallus Z–W dS estimates and 95% confidence intervals overlapped (Fig. 1A, Table S2). Maximum-likelihood gene trees of G. gallus and M. gallopavo gametologs, with the zebra finch (T. guttata) Z-linked ortholog included as an outgroup, indicate that all four strata predate the G. gallus–M. gallopavo divergence estimated at roughly 30 MYA (Dimcheff et al. 2002a). For all seven gametologs, the gene trees clustered by sex chromosome rather than by species, and the phylogenetic grouping of G. gallus and M. gallopavo W-linked genes showed high bootstrap support (in all cases ≥95%, Figs. 1A, S2), indicating that gametolog divergence predates the common ancestor of G. gallus and M. gallopavo. This provides independent verification that the four recombination suppression events previously identified in G. gallus are conserved across the Galliformes.

Divergence dates for recombination suppression in the youngest strata should predate the divergence between G. gallus and M. gallopavo approximately 30 Mya (Dimcheff et al. 2002a). Using a molecular clock that accounts for sex-specific mutation rates in Z- and W-linked genes (Wright et al. 2012), we found that Stratum III, corresponding to 17–43 Mb (dS = 0.156–0.268), arose between 41 and 71 Mya and Stratum IV, corresponding to 0–11 Mb (dS = 0.137–0.257), between 36 and 68 Mya (Fig. 1A). These dates are consistent with the finding that four strata are conserved in the Galliformes.

RECOMBINATION SUPPRESSION EVOLVED INDEPENDENTLY IN THE ANSERIFORMES

Using 11 newly identified A. platyrhynchos gametologs, we calculated dS estimates and maximum-likelihood gene trees to assess the dynamics of recombination suppression and sex chromosome divergence. Avian genomes are relatively stable, with few major genomic rearrangements, potentially due to a lack of active transposons (Toups et al. 2011). As synteny of the Z chromosome is highly conserved among all extant birds (Vicoso et al. 2013b), including the Galloanserae A. platyrhynchos and G. gallus ( Skinner et al. 2009), we took physical position of the A. platyrhynchos gametologs from the location of the corresponding G. gallus Z-linked ortholog. To ensure W-linked paralogs were not misidentified as Z-linked orthologs, we used PCR to confirm they were not physically female-limited.

For the three A. platyrhynchos gametologs located in the two oldest Galliform strata, dS estimates were consistent with orthologous G. gallus Z–W dS estimates, where 95% confidence intervals overlapped (Fig. 1B, Table S3). Maximum-likelihood gene trees for all loci cluster by sex chromosome rather than by species, with 100% bootstrap support for the grouping of G. gallus and A. platyrhynchos W-linked orthologs distinct from Z-linked genes (Figs. 1B, S3). For RASA1Z and RASA1W, we have coding sequence for G. gallus, A. platyrhynchos, and M. gallopavo gametologs, and W-linked sequences cluster together with 100% bootstrap support. The three gametolog sets therefore convergently indicate that the two oldest strata are conserved across the Galloanserae. These strata correspond to 45–51 Mb (dS = 0.285–0.404, Conserved Stratum I), 54–61 Mb
Table 1. Known and newly identified *G. gallus*, *A. platyrhynchos*, and *M. gallopavo* W-linked coding genes.

| Name              | Z-linked gene ID | W-linked gene ID |
|-------------------|------------------|------------------|
| **G. gallus**     |                  |                  |
| HINT1             | 00428            | 22674, 22683, 22690, 22679. |
| CHD1              | 14642            | *CHDW*           |
| RASA1             | 17706            | 22614. |
| KCMF1             | 15391            | 14441            |
| MIER3             | 14721            | 00140, 14584. |
| SPIN              | 14916            | *SPINW*          |
| HNRNPK            | 12591            | 14366. |
| CZH18ORF25        | 01763            | 01585. |
| VCP               | 01986            | 00386. |
| SIAT8C            | 03049            | 26991            |
| ZFR               | 03235            | 14545. |
| NIPBL             | 03605            | 13312. |
| UBAP2             | 13809            | 05785. |
| ATP5A1            | 14644            | 01756. |
| MADH2             | 14697            | 14184. |
| BTF3              | 13512            | 00395. |
| UBE2R2            | 01668            | 09227. |
| ZSWIM6            | 14734            | 27170. |
| ZNF532            | 02852            | 14003. |
| RPL17L            | 02696            | 22174. |
| Novel             | -                | 29099. |
| Novel             | -                | 25736. |
| Novel             | -                | 25865. |
| **M. gallopavo**  |                  |                  |
| HINT1             | 06655            | 16898, 09917, *HINTW*(AY713488.1). |
| RASA1             | 08292            | *RASA1W*(AH015047.1). |
| SPIN              | 02500            | *SPINW*(AH015050.1), 06743 |
| VCP               | -                | 09037. |
| NIPBL             | 01989            | 13606, *IDN3W*(AH015048.1). |
| UBAP2             | 01837            | *UBAP2W*(AY188758.1), 05595 |
| ATP5A1            | 01414            | 09963. |
| MADH2             | *MADH2Z*         | *MADH2W*(AH015049.1). |
| Novel             | -                | 16394. |
| **A. platyrhynchos** |              |                  |
| CHD1              | 09965            | 05191, 02506. |
| RASA1             | 05627            | 05611, 11371, 10611 |
| KCMF1             | 13426            | 03026, 03106. |
| MIER3             | 06634            | 10850. |
| SPIN              | 13922            | 02923. |
| HNRNPK            | 10856            | 10986. |
| VCP               | 06634            | 05806. |
| ZFR               | 15627            | 15519. |
| NIPBL             | 08473            | 10290, 05315, 10560, 02953, 03022 |
| UBE2R2            | 03800            | 16000. |
| ZSWIM6            | 06992            | 13555, 14338. |
| UBAP2             | 06449            | 16155. |

1 Ensgalg00000000(G. gallus)/Ensmgag00000000(M. gallopavo)/Ensaplg00000000(A. platyrhynchos).
2 Moghadam et al. (2012).
3 Ayers et al. (2013).
4 Backstrom et al. (2005).
5 Berlin and Ellegren (2006).
6 Amalgamation of Z and W sequences in current genome assembly.
(\(d_S = 0.171-0.271\), Conserved Stratum II), and arose between 75 and 106 Mya, and 45 and 71 Mya, respectively.

In contrast, of the eight \(A.\) platyrhynchos gametologs located in the youngest Galliform strata, \(d_S\) estimates for five gametologs are significantly lower than the corresponding \(G.\) gallus gametolog \(d_S\) estimates, where 95% confidence intervals are nonoverlapping (Fig. 1B, Table S3). Of these eight \(A.\) platyrhynchos gametologs, four cluster together in maximum-likelihood gametologs, the signal of which was not detected for three gametologs, where the signal of which was not detected for three gametologs. Using a molecular clock, we calculated that recombination ceased independently in Anseriformes between 0 and 39 Mya (Fig. 1B), well after the divergence of Galliformes and Anseriformes approximately 90 Mya (van Tuinen and Hedges 2001).

**HETEROGENEOUS DIVERGENCE OF YOUNG ANSERIFORM GAMAGELOGS**

There is marked heterogeneity in \(d_S\) across the youngest region of the Anseriform sex chromosomes, ranging from 0 to 0.148. This variation is greater than that observed across any other avian strata, and is not a result of a single outlier. Three gametologs have \(d_S < 0.02\), indicating either very recent divergence or persistent recombination. We verified that \(Z\) - and \(W\)-linked coding sequences were distinct using female-limited SNPs identified from RNA-seq data and \(W\)-linked SNPs from Ensembl annotated sequences. Therefore, the heterogeneity in divergence is not consistent with instantaneous suppression of recombination resulting from a large-scale intra-chromosomal event such as an inversion. Instead, this may suggest more gradual divergence of gametologs with residual recombination across large tracts of the sex chromosomes.

To further assess the possibility of ongoing recombination between gametologs, we identified signatures of gene conversion between \(Z\)- and \(W\)-linked coding sequences in \(G.\) gallus, \(A.\) platyrhynchos, and \(M.\) gallopavo using GENECONV (Sawyer 1999). Three gametologs in \(A.\) platyrhynchos showed evidence of significant inner gene conversion following a permutation test with 10,000 iterations (Table 2). We repeated the analysis allowing mismatches between gene conversion fragments \((g_scale = 2)\) and found no significant difference between the results. These gametologs are uniformly distributed across the \(Z\) chromosome, located in Conserved Strata I, II, and Anseriform-specific Stratum III, indicating ongoing recombination along the whole length of the avian sex chromosomes. In total, we detected four \(Z\)- and four \(W\)-linked exons subject to gene conversion (Table 2). Of these, we could determine the direction of gene conversion for the exons in the Conserved Strata I and II. Informative variant sites indicate genetic material was transferred from \(KCMF1W\) to \(KCMF1Z\) and from \(CHD1Z\) to \(CHD1W\). Exon tree topologies support the proposed direction of gene conversion in \(KCMF1\) and cluster by species for \(CHD1\), revealing putative gene conversion across \(G.\) gallus \(CHD1\) gametologs, the signal of which was not detected with GENECONV.

Gene conversion between gametologs can lower \(d_S\) estimates and generate molecular gene trees where gametologs cluster by species. This may lead to false conclusions that recombination was suppressed independently in a given lineage. We detected no gene conversion between the four \(A.\) platyrhynchos gametologs where gene trees cluster by species, or the three gametologs with \(d_S < 0.02\). However, previous studies have shown that GENECONV’s power to detect gene conversion events between similar sequences is limited (McGrath et al. 2009; Ezawa et al. 2010).

We recalculated \(d_S\) for three \(A.\) platyrhynchos gametologs after removing exons subject to gene conversion and as expected, the estimated divergence time increased. However, the range of divergence dates over which recombination was suppressed in Conserved Stratum I and II remained constant. The recalculated gene tree for the gametolog in Anseriform-specific Stratum III revealed a complex mosaic of recombination across specific exons. Before excluding the exons subject to gene conversion, the phylogenetic topology was unresolved, however after removal of these exons, \(A.\) platyrhynchos and \(G.\) gallus \(HNRNPKW\) clustered together with high bootstrap support (\(\geq 95\%\)).

**COMPARATIVE ANALYSIS OF \(W\)-LINKED CODING CONTENT EVOLUTION**

Studies examining the evolutionary forces acting after recombination suppression are mainly limited to male heterogametic systems (Chibalina and Filatov 2011; Hughes et al. 2012; Bachtrog 2013). As a result, the factors driving the differentiation and degeneration of \(Z\) and \(W\) sex chromosomes are partially unclear.

Now equipped with the largest dataset of \(W\)-linked genes to date, we used the branch models and branch-site test in the CODEML package in PAML version 4.7 (Yang 2007) to assess the selective regime driving \(Z\) and \(W\) coding sequence change over 90 million years of avian evolution.

For \(W\)-linked branches, \(\omega (d_N/d_S)\) was significantly lower than 1 in 14 gametolog families, indicating these genes are evolving primarily due to purifying selection (Tables 3, S4). The exception was \(HINT1W\), where \(\omega\) was not significantly different to 1. However, \(HINT1W\) is a known ampliconic gene with documented gene conversion among \(W\)-linked copies (Backstrom et al. 2005). This likely violates the assumptions made in CODEML, where \(d_S\) is equivalent to the neutral mutation rate (Yang 2007). For \(Z\)-linked branches, \(\omega\) was not significantly different to 0 across four loci, indicating these are extremely conserved and evolving under strict purifying selection. For the other 10 loci, \(\omega\) was
Figure 1. Dynamics of recombination suppression on the avian sex chromosomes. The distribution of synonymous ($d_s$) divergence estimates between *G. gallus* (black), *M. gallopavo* (yellow, panel A) and *A. platyrhynchos* (green, panel B) gametologs, mapped against physical position of the *G. gallus* Z-linked ortholog. The 95% confidence intervals were calculated using standard errors. Bootstrap values were calculated using 1000 permutations. AZ, GZ, MZ, TZ corresponds to *A. platyrhynchos*, *G. gallus*, *M. gallopavo*, and *T. guttata* Z-linked genes. The physical Z-linked position of the loci for which gene trees are shown are indicated (*). *Meleagris gallopavo* $d_s$ estimates map closely onto the *G. gallus* strata boundaries, and maximum-likelihood gene trees for all gametologs cluster by sex chromosome, not species, indicating that four strata are conserved across the Galliformes (Fig. S2). *Anas platyrhynchos* $d_s$ estimates map closely onto the two oldest *G. gallus* strata boundaries, and maximum-likelihood gene trees for gametologs within these regions cluster by sex chromosome indicating that these strata are conserved across the Galloanserae. For the remaining gametologs, maximum-likelihood gene trees reveal that recombination was suppressed independently in the *A. platyrhynchos* lineage (Fig. S3). The $d_s$ estimates are highly heterogeneous and locus specific, indicating gradual divergence of gametologs with residual recombination across large tracts of the sex chromosomes.
Table 2. Gene conversion between gametologs in *A. platyrhynchos*.

| Gene name | Gene ID (Z/W) | Stratum         | P-value b/c   | Spanning Z-linked exon no. | Spanning W-linked exon no. | Length of gene conversion track (bp) |
|-----------|---------------|-----------------|---------------|---------------------------|---------------------------|-------------------------------------|
| CHD1      | 09965/05191   | Conserved I     | 0.0006/0.0027 | 27                        | 7                         | 174                                 |
| KCMF1     | 13426/03026   | Conserved II    | 0.0056/0.0273 | 4                         | 3                         | 142                                 |
| HNRNPK    | 10856/10986   | Anseriform-specific III | 0.0081/0.0294 | 10                        | 8                         | 255                                 |
| HNRNPK    | 10856/10986   | Anseriform-specific III | 0.0314/0.0775 | 18                        | 13                        | 159                                 |

*ENsaglp000000.*

*b Simulated P-values are based on 10,000 permutations.

c Karlin–Altshul P-values are Bonferroni-corrected KQ (BLAST-like).

d Exons where gene conversion fragments span <50 bp are not listed.

* Significant after correcting for multiple comparisons.

We tested for positive selection using the branch-site model. This model allows $\omega$ to vary across both branches and among sites, and tests for adaptive evolution acting on a subset of codons along a branch(es) defined a priori (Zhang et al. 2005). We found no evidence of positive selection on any Z-linked gene families (Table S5) but found significant evidence for positive selection acting on one W-linked loci, *ATP5A1W* (proportion of sites where $\omega > 1$ was 0.012 and $\omega = 17.70$) (Tables 3, S4).

Gene conversion violates the assumption in PAML that $d_S$ is constant across a gene and the GC bias in the mismatch repair process can inflate $\omega (d_S/d_S)$ (Berglund et al. 2009). Consequently, previous studies have found that gene conversion can lead to false signals of positive selection (Casola and Hahn 2009; Ratnakumar et al. 2010). Although we found no evidence of gene conversion between *ATP5A1W* and *ATP5A1Z* in any species, *ATP5A1W* is present in two copies in the *G. gallus* genome (Moghadam et al. 2012) and may be subject to intra-chromosomal gene conversion. In contrast, our polymorphism data indicate there is only a single copy in the *M. gallopavo* genome, as we failed to identify polymorphisms in this locus within any single female, which would indicate multiple copy number. To exclude the possibility that gene conversion between W-linked *G. gallus* paralogs is driving a false signal of positive selection, we removed *G. gallus ATP5A1W* and repeated the analysis. After excluding this gene, the branch-site test for positive selection remained significant (proportion of sites where $\omega > 1$ was 0.006 and $\omega = 41.87$).

**Discussion**

The process of recombination suppression is a crucial component of sex chromosome evolution, facilitating the divergence and differentiation of gametologs (Charlesworth and Charlesworth 1978). Studies from highly differentiated sex chromosomes suggest that recombination is suppressed instantaneously in a stepwise manner by intra-chromosomal rearrangements (Lahn and Page 1999; Pandey et al. 2013). However, our ability to determine whether inversions are a cause or consequence of this process is hampered by the limited numbers of sex-linked genes in these species. Conversely, in newly evolved sex chromosomes, recombination appears to be suppressed in a gradual and heterogeneous nature (Chibalina and Filatov 2011; Muyle et al. 2012; Bergero et al. 2013; Natri et al. 2013; Qiu et al. 2013).

We used a comparative approach to study the dynamics of sex chromosome recombination suppression on the largest collection of W-linked coding content yet assembled across 90 million years of avian evolutionary history. Our results reveal that recombination between the Z and W chromosomes ceased independently multiple times across different avian lineages. The pattern we uncover is consistent with both the inversion model (Charlesworth et al. 2005) where evolutionary strata form in a stepwise process, and heterogeneous and locus specific recombination suppression over short-term periods. Additionally, we show that recombination suppression is not necessarily complete, as highly diverged gametologs recombine across specific exons via gene conversion.

**DYNAMICS OF SEX CHROMOSOME DIVERGENCE**

Using both maximum-likelihood gene trees and synonymous divergence estimates, we confirm that four strata are conserved across the Galliformes (Fig. 2). These four strata were previously identified in *G. gallus* (Wright et al. 2012), and follow the
Table 3. Branch models and branch-site test for W-linked branches.

| Gene name         | Species included | \( \omega \) | Branch model test \((\omega=1)\) likelihood ratio | Branch model test \((\omega=0)\) likelihood ratio | Branch-site test \((\omega=1)\) likelihood ratio |
|-------------------|------------------|--------------|---------------------------------|---------------------------------|---------------------------------|
| HINT1W            | G, M, T          | 1.10         | 0.05                            | 3697.05**                       | 0.05                            |
| CHD1W             | G, A, T          | 0.14         | 94.45**                         | 7012.15**                       | 0.00                            |
| RASA1W            | G, A, M, T       | 0.09         | 31.33**                         | 1386.95**                       | 0.00                            |
| KCME1W            | G, A, T          | 0.22         | 35.49**                         | 5674.71**                       | 0.22                            |
| MIER3W            | G, A, T          | 0.24         | 17.97**                         | 167.25**                        | 0.00                            |
| SPINW             | G, A, M, T       | 0.09         | 28.45**                         | 1749.98**                       | 0.00                            |
| HNRNPKW           | G, A, T          | 0.03         | 108.75**                        | 84.21**                         | 0.03                            |
| VCPW              | G, A, T          | 0.03         | 25.58**                         | 19.13**                         | 0.16                            |
| ZFRW              | G, A, T          | 0.18         | 27.16**                         | 173.50**                        | 0.00                            |
| NIPBLW            | G, A, T          | 0.05         | 44.08**                         | 63.91**                         | 0.00                            |
| NIPBLW            | G, M, T          | 0.03         | 38.40**                         | 343.5**                         | 0.00                            |
| UBAF2W            | G, M, T          | 0.15         | 6.19*                           | 354.82**                        | 0.00                            |
| ATP5A1W           | G, M, T          | 0.09         | 16.54**                         | 650.83**                        | 3.58*                           |
| MADH2W            | G, M, T          | 0.06         | 22.37**                         | 333.50**                        | 0.01                            |
| UBE2R2W           | G, A, T          | 0.04         | 15.03**                         | 17.39**                         | 0.00                            |
| ZSWIM6W           | G, A, T          | 0.25         | 11.25**                         | 128.88**                        | 0.00                            |

\(^1G = G.\) gallus, \(M = M.\) gallopavo, \(A = A.\) platyrhyncos, \(T = T.\) guttata.

\(^{**}P\)-value < 0.01.

\(^{*}P\)-value < 0.05.

Likelihood values for each model and the gene IDs are listed in Table S4. Significant likelihood ratios are in bold.
SEX CHROMOSOME EVOLUTION

Figure 2. Independent expansion of the nonrecombining regions across the Galloanserae. Phylogeny illustrating the independent cessation of recombination on the sex chromosomes in the Galliformes and Anseriformes. Estimated date of recombination suppression (millions of years), calculated using a molecular clock that accounts for sex-specific mutation rates in Z- and W-linked genes, is mapped onto the phylogeny.

same stepwise pattern observed across the mammalian X chromosome (Lahn and Page 1999; Sandstedt and Tucker 2004; Pearks Wilkerson et al. 2008; Van Laere et al. 2008). Using newly identified gametologs, we were able to refine the dates at which recombination was suppressed across the Galliform strata. We find that the oldest strata originated approximately 90 Mya, and the youngest 46 Mya, both of which are consistent with the divergence date between *G. gallus* and *M. gallopavo* (Dimcheff et al. 2002a).

Similarly, maximum-likelihood gene trees reveal that the two oldest Galliform strata are conserved on the Anseriform Z chromosome. In contrast, recombination suppression is heterogeneous across the remainder of the *A. platyrhynchos* sex chromosomes, and the topology of gene trees revealed that gametologs diverged independently in this region (Fig. 2). Divergence estimates vary from 0 to 39 Mya, a range that is greater than that observed across any other avian strata, and not a result of a single outlier. We verified that Z- and W-linked sequences were in fact distinct using polymorphism data and Ensembl coding sequences. This heterogeneity in divergence is not consistent with instantaneous suppression of recombination resulting from a large-scale intra-chromosomal event and is instead indicative of Z–W divergence combined with ongoing recombination. However, due to the limited number of gametologs, we lack sufficient power to formally test for differences in the variance of $d_S$ across strata.

The lack of positional information for the *A. platyrhynchos* Z chromosome means we cannot rule out the possibility that Anseriform-specific Stratum III is actually made up of two distinct strata formed by a lineage specific inversion. Under this scenario, the three gametologs with $d_S$ estimates $<0.02$ would comprise the youngest stratum with ongoing recombination between the W and Z orthologs. Positional information may support this scenario, as two of the three gametologs are located between 0 and 10 Mb on the Z chromosome.

Some gametologs are subject to gene conversion, lending further support to the notion that recombination suppression is not complete. Three *A. platyrhynchos* gametologs showed evidence of significant gene conversion, and were distributed across the length of the Z chromosome. Therefore, even for highly differentiated gametologs, certain exons are prevented from diverging. This is consistent with previous studies documenting gene conversion between X- and Y-linked orthologs in mammals (Iwase et al. 2010; Trombetta et al. 2014). Gene conversion may therefore be an important mechanism halting the degeneration of sex-limited gametologs (Rosser et al. 2009) in addition to the intra-chromosomal gene conversion documented on the avian W and mammalian Y chromosomes (Backstrom et al. 2005; Marais et al. 2010; Hallast et al. 2013; Bellott et al. 2014). Interestingly, a previous study used whole chromosome painting to show that the *A. platyrhynchos* W chromosome is highly divergent in comparison to closely related avian W chromosomes (Stiglec et al. 2007). We find a high degree of conservation among coding genes, indicating exon-specific gene conversion may facilitate large-scale turnover of intronic and repetitive regions while still conserving gene function.

This heterogeneous pattern of sex chromosome divergence and ongoing recombination is consistent with *G. aculeatus* and *S. latifolia* sex chromosomes. The *S. latifolia* XY system originated 10 Mya, and a neo-sex chromosome formed 2 Mya in *G. aculeatus*. In both these system, recombination suppression evolves heterogeneously (Chibalina and Filatov 2011; Muyle et al. 2012; Bergero et al. 2013; Natri et al. 2013; Natri et al. 2013; Qiu et al. 2013; Roesti et al. 2013). The underlying mechanism of recombination suppression is unclear, but changes in heterochromatin and methylation, which have been shown to alter patterns of recombination (Mirouze et al. 2012), or local changes in the binding sequence that recruits recombination machinery to the chromosome (Baudat et al. 2010; Myers et al. 2010), may be responsible.
Our divergence estimates for recombination suppression are consistent with divergence dates for G. gallus, A. platyrhynchos, and M. gallopavo with the exception of Conserved Stratum II. Using a molecular clock that accounts for sex-specific mutation rates in Z- and W-linked genes, the divergence of Conserved Stratum II was estimated to have occurred between 45 and 71 Mya. Galliformes and Anseriformes diverged 90 Mya (van Tuinen and Hedges 2001). Although these estimates are not vastly distant, they are inconsistent and may indicate that the molecular clock underestimates the time of gametolog divergence due to the unique evolutionary forces acting on sex chromosomes (Bachtrog et al. 2011).

EVOLUTION OF HETEROGAMETIC SEX CHROMOSOMES AFTER RECOMBINATION SUPPRESSION

Studies examining the evolutionary forces acting after recombination has ceased have mainly been limited to male heterogametic systems, such as the Y chromosomes of mammals and Drosophila (Chibalina and Filatov 2011; Hughes et al. 2012; Bachtrog 2013). These studies have uncovered strong purifying selection acting on the Y chromosome (Kaiser et al. 2011; Hughes et al. 2012; Bellott et al. 2014; Wilson Sayres et al. 2014) together with signatures of positive selection (Gerrard and Filatov 2005; Li et al. 2013). However, the selective forces driving the evolution of the W chromosome are unclear. This is partly due to a previous paucity of known W-linked genes, which has restricted the potential of interspecific analyses (Berlin and Ellegren 2006).

Theoretical models predict key differences between W and Y chromosome evolution (Kirkpatrick and Hall 2004; Mank 2012; Wright and Mank 2013). When sexual selection is acting primarily on males, the Y chromosome has a lower effective population size than W-linked sequences (Wright and Mank 2013), and is subject to male-mutation bias (Kirkpatrick and Hall 2004). These factors have been hypothesized to result in rapid degeneration of Y-linked coding sequences, a pattern that has been observed on Drosophila and primate Y chromosomes (Hughes et al. 2010, 2012; Bachtrog 2013).

Using the largest catalog of W-linked genes to date, we find that avian W-linked genes are evolving with a significant contribution of purifying selection. This is consistent with a previous study showing that sex-specific selection drives evolutionary changes in the expression of W-linked genes (Moghadam et al. 2012), and indicates that although the W chromosome does not determine sex directly (Smith et al. 2009), it has an important role in encoding sex-specific fitness. The large number of genes conserved across 90 million years of avian evolution is also potentially indicative of lower degeneration rates on the W compared to the Y chromosome.

We found evidence of positive selection in one W-linked gene that encodes a mitochondrial ATP synthase subunit. As the W chromosome and mitochondria are both maternally inherited and are effectively linked (Berlin et al. 2007), we might expect strong coevolution between these portions of the genome, and this may be reflected by rapid divergence in sequence across different avian lineages.

Our findings indicate that adaptive evolution together with conservation of gene sequence is not restricted to male heterogametic systems and may be a fundamental feature of heterogametic sex chromosome evolution.

CONCLUDING REMARKS

Recombination suppression is a key force driving sex chromosome evolution, and sex chromosomes are therefore important for the study of genome evolution, particularly the interplay between selective forces and recombination. Here, we use a comparative approach to examine short- and long-term dynamics of sex chromosome evolution. Our results reveal that recombination has ceased multiple and independent times over 90 million years of avian evolution. We show that sex chromosome divergence over the long term is characterized by regions of similar divergence among gametologs, consistent with the inversion model of sex chromosome evolution (Lahn and Page 1999; Charlesworth et al. 2005; Vicoso et al. 2013a). In contrast, over short time periods we observe heterogeneous and locus specific divergence. Additionally, we uncover gene conversion between highly diverged gametologs across both long and short evolutionary trajectories. Moreover, if sex chromosomes do evolve via inversions, additional mechanisms may be required to completely suppress recombination. Our data also show that the W chromosome is evolving primarily due to purifying selection, although we also detect the signature of positive selection at one locus. Our findings indicate that the W chromosome plays an important role in encoding sex-specific fitness, and potentially exhibits a lower rate of degeneration compared to Y chromosome.

ACKNOWLEDGMENTS

We thank V. Roberts and J. Hall for collection assistance, G. Parker for assistance in securing samples, Z. Yang for helpful comments on the analyses conducted in PAML, and D. Burt for providing A. platyrhynchos coding sequences in advance of the release of the full genome annotation. We also thank J. Perry, F. Zimmer, V. Oostra, and three anonymous reviewers for helpful suggestions and comments on the manuscript. Sequencing was performed by the Wellcome Trust Centre for Human Genetics Sequencing Hub (funded by Wellcome Trust Grant reference 090532/Z/09/Z and MRC Hub Grant G0900747 91070). The authors acknowledge the use of the UCL Legion High Performance Computing Facility (Legion@UCL) and UCL Unity SMP Facility, and associated support services, in the completion of this work. The work described here was funded by the European Research Council under the Framework 7 agreement (grant agreement 260233). AEW is funded by a National Environment Research Council...
Ph.D. studentship. SHM is funded by a research fellowship from the Royal Commission for the Exhibition of 1851.

**LITERATURE CITED**

Achtschel, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.

Axelsson, E., N. G. C. Smith, H. Sundström, S. Berlin, and H. Ellegren. 2004. Male-biased mutation rate and divergence in autosomal, Z-linked and W-linked introns of chicken and turkey. Mol. Biol. Evol. 21:1538–1547.

Axelsson, E., L. Hultin-Rosenberg, M. Brandström, M. Zvahalen, D. F. Clayton, and H. Ellegren. 2008. Natural selection in avian protein-coding genes expressed in brain. Mol. Ecol. 17:3008–3017.

Ayers, K. L., N. M. Davidson, D. Demyah, K. N. Roeszler, F. Gruetzner, A. H. Sinclair, A. Oshlack, and C. A. Smith. 2013. RNA sequencing reveals sexually dimorphic gene expression before gonadal differentiation in chicken and allows comprehensive annotation of the W-chromosome. Genome Biol. 14:R24.

Bachtrog, D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat. Rev. Genet. 14:113–124.

Bachtrog, D., E. Horn, K. M. Wong, X. Maside, and P. de Jong. 2008. Genomic degradation of a young Y chromosome in Drosophila miranda. Genome Biol. 9:R30.

Bachtrog, D., M. Kirkpatrick, J. E. Mank, S. F. McDaniel, J. C. Pires, W. R. Rice, and N. Valenzuela. 2011. Are all sex chromosomes created equal? Trends Genet. 27:350–357.

Backstrom, N., H. Ceplitis, S. Berlin, and H. Ellegren. 2005. Gene conversion drives the evolution of the HINTW, an ampiclicon gene on the female-specific avian W chromosome. Mol. Biol. Evol. 22:1992–1999.

Baudat, F., J. Buard, C. Grey, A. Fledel-Alon, C. Ober, M. Przeworski, G. Coop, and B. de Massy. 2010. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science 327:836–840.

Bellott, D. W., J. F. Hughes, H. Skaletsky, L. G. Brown, T. Pyntikova, T. Graves, R. S. Fulton, S. Dugan, Y. Ding, C. J. Buhay, C. A. Smith. 2012. Identification of avian W-linked contigs by short-read sequencing. BMC Genomics 13:183. doi:10.1186/1471-2164-13-183.

Chibalina, M. V., and D. A. Filatov. 2011. Plant Y chromosome degeneration is retarded by haploid purification selection. Curr. Biol. 21:1475–1479.

Dalloul, R. A., J. A. Long, A. V. Zimin, L. Aslam, K. Beal, L. A. Blomberg, P. Bouffard, D. W. Burt, O. Crasta, R. P. M. A. Crooijmans, et al. 2010. Multi-platform next-generation sequencing of the domestic turkey (Meleagris gallopavo): genome assembly and analysis. PLoS Biol. 8:e1000475.

Dalloul, R. A., A. V. Zimin, R. E. Settlage, S. Kim, and K. M. Reed. 2014. Next-generation sequencing strategies for characterizing the turkey genome. Poult. Sci. 93:479–484.

Dimcheff, D. E., S. V. Drovetski, and D. P. Mindell. 2002a. Phylogeny of Tetraoninae and other galliform birds using mitochondrial 12S and ND2 genes. Mol. Phylogenet. Evol. 24:203–215.

———. 2002b. Phylogeny of Tetraoninae and other galliform birds using mitochondrial 12S and ND2 genes. Mol. Phylogenet. Evol. 24:203–215.

Edgar, R. C. 2004a. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:1–19.

———. 2004b. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Ezawa, K., S. Oota, and N. Saitou. 2006. Genome-wide search of gene conversions in duplicated genes of mouse and rat. Mol. Biol. Evol. 23:927–940.

Ezawa, K., K. Ikeo, T. Gojobori, and N. Saitou. 2010. Evolutionary pattern of gene homogenization between primate-specific paralogs after human and macaque speciation using the 4–2–4 method. Mol. Biol. Evol. 27:2152–2171.

Flicek, P., I. Ahmed, M. R. Amode, D. Barrett, K. Beal, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fairley, et al. 2013. Ensembl 2013. Nucleic Acids Res. 41:D48–D55.

Gerrard, D. T., and D. A. Filatov. 2005. Positive and negative selection on mammalian Y chromosomes. Mol. Biol. Evol. 22:1423–1432.

Graur, D., and W.-H. Li. 2000. Fundamentals of molecular evolution. Sinauer Associates, Sunderland, MA.

Hallast, P., P. Balaresque, G. R. Bowden, S. Ballereau, and M. A. Jobling. 2013. Recombination dynamics of a human Y-chromosomal palindrome: rapid GC-biased gene conversion, multi-kilobase conversion tracts, and rare inversions. PLoS Genet. 9:e1003666.

Harlid, A. B., S. Berlin, N. G. C. Smith, A. P. Mosllier, and H. Ellegren. 2003. Life history and the male mutation bias. Evolution 57:2398–2406.

Hedges, S. B. 2002. The origin and evolution of model organisms. Nature 415:771–776.

Huang, Y., Y. Li, D. W. Burt, H. Chen, Y. Zhang, W. Qian, H. Kim, S. Gan, Y. Zhao, J. Li, et al. 2013. The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. Nat. Genet. 45:776–783.

Hughes, J. F., H. Skalskaya, R. A. Zimin, L. G. Brown, T. Graves, R. S. Fulton, S. Dugan, Y. Ding, C. J. Buhay, C. Kremzitki, et al. 2012. Strict evolutionary conservatism followed rapid gene loss on human and rhesus Y chromosomes. Nature 483:82–86.
Iwase, M., Y. Satta, H. Hirai, Y. Hirai, and N. Takahata. 2010. Frequent gene conversion events between the X and Y homologous chromosomal regions in primates. BMC Evol. Biol. 10:225. doi:10.1186/1471-2148-10-225.

Kaiser, V. B., Q. Zhou, and D. Bachtrog. 2011. Nonrandom gene loss from the Drosophila miranda neo-Y chromosome. Genome Biol. Evol. 3:1329–1337.

Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S. L. Salzberg. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14:R36.

Kirkpatrick, M., and D. W. Hall. 2004. Male-biased mutation, sex linkage, and the rate of adaptive evolution. Evolution 58:437–440.

Lahn, B. T., and D. C. Page. 1999. Four evolutionary strata on the human X chromosome. Science 286:964–967.

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:357–359.

Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10:R25.

Lemaître, C., M. D. V. Braga, C. Gautier, M.-F. Sagot, E. Tannier, and G. A. B. Marais. 2009. Footprints of inversions at present and past pseudautosomal boundaries in human sex chromosomes. Genome Biol. Evol. 1:56–66.

Li, G., B. W. Davis, T. Raudsepp, A. J. P. Wilkerson, T. Graves, D. Albracht, W. Warren, B. P. Chowdhary, L. C. Skow, and W. J. Murphy. 2008. Gene discovery and comparative analysis of X-degenerate genes from the domestic cat Y chromosome. Genomics 92:329–338.

Qiu, S., R. Bergero, and D. Charlesworth. 2013. Testing for the footprint of sexually antagonistic polymorphisms in the pseudautosomal region of a plant sex chromosome pair. Genetics 194:663–672.

Ratnakumar, A., S. Mousset, S. Glémin, J. Berglund, N. Galtier, L. Duret, and M. T. Webster. 2010. Detecting positive selection within genomes: the problem of biased gene conversion. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 365:2571–2580.

Roesti, M., D. Moser, and D. Berner. 2013. Recombination in the threespine stickleback genome—patterns and consequences. Mol. Ecol. 22: 3014–3027.

Ross, M. T., D. V. Graham, A. J. Coffey, S. Scherer, K. McIay, D. Muzny, M. Platzer, G. R. Howell, C. Burrows, C. P. Bird, et al. 2005. The DNA sequence of the human X chromosome. Nature 434:325–337.

Rosser, Z. H., P. Balaresque, and M. A. Jobling. 2009. Gene conversion between the X chromosome and the male-specific region of the Y chromosome at a translocation hotspot. Am. J. Med. Genet. 85:130–134.

Sandstedt, S. A., and P. K. Tucker. 2004. Evolutionary strata on the mouse X chromosome correspond to strata on the human X chromosome. Genome Res. 14:267–272.

Sawyer, S. A. 1999. GENECONV: a computer package for the statistical detection of gene conversion. Distributed by the author www.math.wustl.edu/~sawyer/geneconv/.

Schemper, M. 2003. Predictive accuracy and explained variation. Stat. Med. 22:2299–2308.

Skaletsky, H., T. Kuroda-Kawaguchi, P. J. Minx, H. S. Cordum, L. Hillier, L. G. Brown, S. Repping, T. Pintikoboda, J. Ali, T. Bieri, et al. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825–837.

Skinner, B. M., L. B. W. Robertson, H. G. Tempest, E. J. Langley, D. Ianno, K. E. Fowler, R. P. A. Crooijmans, A. D. Hall, D. K. Griffin, and M. Voelker. 2009. Comparative genomics in chicken and Pekin duck using FISH mapping and microarray analysis. BMC Genomics 10:357. doi:10.1186/1471-2164-10-357.

Smith, C. A., K. N. Roeszler, T. Ohnesorg, D. M. Cummins, P. G. Farlie, T. J. Doran, and A. H. Sinclair. 2009. The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. Nature 461:267–271.

Stiglec, R., T. Ezaz, and J. A. M. Graves. 2007. A new look at the evolution of avian sex chromosomes. Cytogenet. Genome Res. 117:103–109.

Suh, A., J. O. Kriegs, J. Brosius, and J. Schmitz. 2011. Retroposition insertions and the chronology of avian sex chromosome evolution. Mol. Biol. Evol. 28:2993–2997.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28:2731–2739.
Tian, L., T. Cai, E. Goetghebeur, and L. J. Wei. 2007. Model evaluation based on the sampling distribution of estimated absolute prediction error. Biometrics 94:297–311.

Toups, M. A., J. B. Pease, and M. W. Hahn. 2011. No excess gene movement is detected off the avian or lepidopteran Z chromosome. Genome Biol. Evol. 3:1381–1139.

Trapnell, C., L. Pachter, and S. L. Salzberg. 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111.

Trapnell, C., B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. van Baren, S. L. Salzberg, B. J. Wold, and L. Pachter. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28:511–515.

Trapnell, C., D. G. Hendrickson, M. Sauvageau, L. Goff, J. L. Rinn, and L. Pachter. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. 31:46–53.

Trombetta, B., D. Sellitto, R. Scozzari, and F. Cruciani. 2014. Inter- and intra-species phylogenetic analyses reveal extensive X-Y gene conversion in the evolution of gametologous sequences of human sex chromosomes. Mol. Biol. Evol. 31:2108–2123.

Tsuda, Y., C. Nishida-Umehara, J. Ishijima, K. Yamada, and Y. Matsuda. 2007. 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.

van Tuinen, M., and S. B. Hedges. 2001. Calibration of avian molecular clocks. Mol. Biol. Evol. 18:1203–1216.

Wilson, M. A., and K. D. Makova. 2009. Evolution and survival on eutherian sex chromosomes. PLoS Genet. 5: e1000568.

Wilson Sayres, M. A., K. E. Lohmueller, and R. Nielsen. 2014. Natural selection reduced diversity on Human Y chromosomes. PLoS Genet. 10:e1004064.

Wright, A. E., and J. E. Mank. 2013. The scope and strength of sex-specific selection in genome evolution. J. Evol. Biol. 26:1841–1853.

Wright, A. E., H. K. Moghadam, and J. E. Mank. 2012. Trade-off between selection for dosage compensation and masculinization on the avian Z chromosome. Genetics 192:1433–1445.

Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591.

Zhang, J., R. Nielsen, and Z. Yang. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol. Biol. Evol. 22:2472–2479.

Zhou, Q., and D. Bachtrog. 2012. Chromosome-wide gene silencing initiates Y degeneration in Drosophila. Curr. Biol. 22:522–525.

Supporting Information
Additional Supporting Information may be found in the online version of this article at the publisher’s website:

- **Figure S1.** Identification of W-linked genes.
- **Figure S2.** Gene trees for M. gallopavo gametologs.
- **Figure S3.** Gene trees for A. platyrhynchos gametologs.
- **Table S1.** Sequence divergence between G. gallus gametologs.
- **Table S2.** Sequence divergence between M. gallopavo gametologs.
- **Table S3.** Sequence divergence between A. platyrhynchos gametologs.
- **Table S4.** Branch models and branch-site test across W-linked branches.
- **Table S5.** Branch models and branch-site test across Z-linked branches.