Avian W and mammalian Y chromosomes convergently retained dosage-sensitive regulators

Daniel W Bellott[1,2], Helen Skaletsky[1,3], Ting-Jan Cho[1], Laura Brown[1,3], Devin Locke[4], Nancy Chen[5–7], Svetlana Galkina[8], Tatyana Pyntikova[1], Natalia Koutseva[1], Tina Graves[4], Colin Kremitzki[4], Wesley C Warren[4], Andrew G Clark[5,9], Elena Gaginskaya[8], Richard K Wilson[4] & David C Page[1–3]

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

After birds diverged from mammals, different ancestral autosomes evolved into sex chromosomes in each lineage. In birds, females are ZW and males are ZZ, but in mammals females are XX and males are XY. We sequenced the chicken W chromosome, compared its gene content with our reconstruction of the ancestral autosomes, and followed the evolutionary trajectory of ancestral W-linked genes across birds. Avian W chromosomes evolved in parallel with mammalian Y chromosomes, preserving ancestral genes through selection to maintain the dosage of broadly expressed regulators of key cellular processes. We propose that, like the human Y chromosome, the chicken W chromosome is essential for embryonic viability of the heterogametic sex. Unlike other sequenced sex chromosomes, the chicken W chromosome did not acquire and amplify genes specifically expressed in reproductive tissues. We speculate that the pressures that drive the acquisition of reproduction-related genes on sex chromosomes may be specific to the male germ line.

In birds and mammals, one pair of chromosomes differs between males and females. In birds, females are ZW and males are ZZ; in mammals, females are XX and males are XY. The sex chromosomes of birds and mammals are not orthologous: genes that are sex linked in birds are autosomal in mammals, and vice versa[1–3]. The orthologs of chicken sex-linked genes are found on human autosomes 5, 9, and 18, while the orthologs of human sex-linked genes are found on chicken autosomes 1 and 4 (refs. 1–4). The orthologs of sex-linked genes from birds and mammals are found on separate autosomes in outgroup species, like fish, indicating that the sex chromosomes of birds and mammals evolved independently, from what were once ordinary autosomes in the common ancestor[4].

Although the avian ZW sex chromosomes are the mirror image of the mammalian XY pair with respect to sex, these two chromosome pairs followed parallel evolutionary trajectories. In each lineage, a series of events, most likely inversions on the sex-specific (W or Y) chromosome, suppressed crossing-over between the sex chromosomes, leading to the formation of evolutionary strata[5–7]. In the absence of crossing-over, the sex-specific W and Y chromosomes diverged from their counterparts, the Z and X chromosomes. The Z and X chromosomes retained 98% of the genes that existed on the ancestral autosomes[4,8]. In contrast, the sex-specific W and Y chromosomes became subject to genetic decay[9]. Few ancestral genes remain on mammalian Y chromosomes; the opossum Y chromosome was among the most conservative, retaining 4–5% of ancestral genes, while decay was more severe on the mouse Y chromosome, where only 1% of ancestral genes remain[10,11]. The extent of divergence between the Z and W chromosomes varies widely across birds, from emu and ostrich—where two-thirds of the Z chromosome still crosses over with the W chromosome in lengthy pseudoautosomal regions—to the chicken, where the Z and W chromosomes are almost completely differentiated[7].

The current understanding of the biology and evolution of sex-specific chromosomes is largely based upon the reference sequences of several male-specific Y chromosomes[10–15]. Vertebrate sex chromosomes commonly contain ampliconic sequences, long stretches of duplicated sequences that have high nucleotide identity[4,8,10–15]. Resolving these sequences requires a methodology with an extraordinary level of accuracy and precision—specifically, the sequencing of large-insert clones derived from a single haplotype. High-quality, clone-based Y-chromosome reference sequences have identified two major phenomena in the evolution of male-specific Y chromosomes: the acquisition and amplification of testis-expressed gene families that preserve or enhance male fertility[11,13,15] and the preservation of widely expressed, dosage-sensitive ancestral genes that may have crucial roles in Turner syndrome and in sexual dimorphism in health and disease[10].

Analogous evolutionary pressures are expected to act on female-specific W chromosomes. Genes on the chicken W chromosome

[1]Whitehead Institute, Cambridge, Massachusetts, USA. [2]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [3]Howard Hughes Medical Institute, Whitehead Institute, Cambridge, Massachusetts, USA. [4]McDonnell Genome Institute, Washington University School of Medicine, St. Louis, Missouri, USA. [5]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York, USA. [6]Cornell Laboratory of Ornithology, Ithaca, New York, USA. [7]Department of Evolution and Ecology, University of California, Davis, Davis, California, USA. [8]Saint Petersburg State University, Saint Petersburg, Russian Federation. [9]Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York, USA. Correspondence should be addressed to D.C.P. (dcpage@wi.mit.edu).

Received 1 November 2016; accepted 29 December 2016; published online 30 January 2017; doi:10.1038/ng.3778

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

respond to female-specific selection[16] and are expected to accumulate genes expressed solely in female-specific tissues[17]. W-linked genes in chicken, turkey, and duck are evolving with significant contributions from purifying selection[18]. In chicken[19] and flycatcher[20], the combined expression of Z–W gene pairs in females is comparable to the expression from both Z genes in males, leading some investigators to hypothesize that the surviving ancestral W-linked genes in birds also should be enriched for broadly expressed, dosage-sensitive regulators[20].

Here we reconstruct the evolutionary trajectory of the genes ancestral to the avian sex chromosomes and examine whether evolutionary pressures analogous to those faced by the mammalian Y chromosomes generated biases in the gene content of female-specific avian W chromosomes. To enable a systematic and comprehensive analysis of gene acquisition and preservation, we produced the first high-quality, clone-based reference sequence from the female-specific chicken W chromosome, supported by physical, linkage, and cytological maps. These sequences, made immediately available in GenBank, have already enabled the design and interpretation of recent studies of avian sex chromosomes[18–20]. We took advantage of our previous reconstruction of the ancestral gene content of the avian sex chromosomes[4], as well as the candidate W-linked genes reported in draft genome assemblies from several avian lineages[7,18,20,21], to extend our parallel analysis across the surviving ancestral genes on the W chromosomes of 14 species of birds.

Genetic decay devastated the chicken W chromosome: only 28 of 685 ancestral genes remain. All of the genes on the chicken W chromosome derive from the ancestral autosomes and are expressed across a broad array of adult and embryonic tissues. Relative to other ancestral genes, surviving Z–W gene pairs on the W chromosomes of chicken and other birds are enriched for dosage-sensitive, broadly expressed genes, under strong purifying selection. We conclude that selection to maintain the ancestral dosage of homologous sex-chromosome gene pairs was the driving force behind the survival of ancestral W-chromosome genes in the chicken and across the avian lineage. Further, we speculate that differences in selective pressures operating on chromosomes in male and female germ lines may explain why no W-linked genes are expressed exclusively in female-specific tissues in the chicken.

RESULTS

Sequencing and analysis of the chicken W chromosome

We sequenced the euchromatic portion of the chicken W chromosome (**Fig. 1**), using the super-resolution methodology that we previously employed on mammalian Y, human X, and chicken Z chromosomes (Online Methods)[4,8,10–13,15]. We obtained a tiling path of 7 Mb in 13 contigs (**Fig. 1a**–**c**, **Supplementary Table 1**, and **Supplementary Data 1**), containing 28 genes (**Fig. 1b**). The resulting sequence is accurate to about 1 nt per 36 kb (Online Methods and **Supplementary Table 1**). We employed two methods to order and orient these contigs (Online Methods). First, we assigned each sequence contig to one of three distinct linkage groups on the radiation hybrid map (**Fig. 1a**,**d** and **Supplementary Data 2**), and, second, we ordered the three radiation hybrid linkage groups along the W chromosome using lampbrush FISH (**Fig. 1e**,**f** and **Supplementary Fig. 1**).

We validated each putative chicken W-chromosome gene by verifying transcriptional activity (**Supplementary Data 3**) and comparing its ORF to its human ortholog (**Supplementary Data 4**). All 28 genes on the W chromosome are broadly expressed across adult tissues (**Fig. 2**). Of the 28 genes, 27 are each present in a single copy on the chicken W chromosome; only *HINTW* has been amplified into a multicopy family (**Supplementary Fig. 2**) (refs. 18,22). Ampliconic

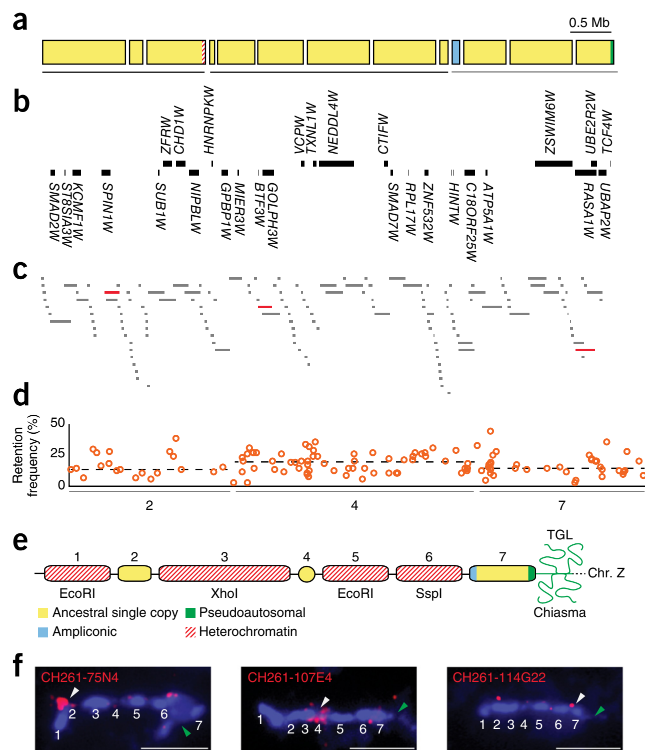

Figure 1 Structure of the chicken W chromosome. (**a**) Sequence map of the W chromosome, covering 7 Mb in 13 contigs. (**b**) Twenty-eight protein-coding genes. See also **Supplementary Data 2** and **3**. (**c**) Clone map; highlighted clones (red) were used as probes in lampbrush FISH. See also **Supplementary Table 1** and **Supplementary Data 1**. (**d**) Radiation hybrid retention frequencies for single-copy markers (orange circles) and the average for each chromomere (dashed lines). Chromomere 4, located near the centromere, displays the highest average retention frequency. See also **Supplementary Data 4**. (**e**) Schematic of the W chromosome at diplotene of female meiosis. The pseudoautosomal region (green) contains chiasma between terminal giant lumpy (TGL) loops at the W and Z termini. Chromomeres are numbered in ascending order from the free end of the W chromosome to the chiasma region. Heterochromatic repeat families (red hashes) occupy chromomeres 1, 3, 5, and 6. Chromomeres 2, 4, and 7 correspond to three distinct radiation hybrid linkage groups; most of their sequence is ancestral single-copy sequence (yellow). A small ampliconic region (blue) contains *HINTW*. See also **Supplementary Figure 2**. (**f**) Lampbrush FISH localizes BAC probes from each radiation hybrid linkage group to a different chromomere. The TGL site is marked with a green arrowhead; each chromomere is numbered in white. Scale bars, 5 μm. BAC probes contain interspersed repeats and give weak secondary signals at multiple sites on the W and other chromosomes; primary signal is marked with a white arrowhead. CH261-75N4 localizes to chromomere 2, CH261-107E4 localizes to chromomere 4, and CH261-114G22 localizes to chromomere 7 (all red). See also **Supplementary Figure 1**.

sequences, which are long stretches of duplicated sequence that have high nucleotide identity, are a common feature of mammalian Y chromosomes[11–15]. The *HINTW* array is the only ampliconic sequence on the chicken W chromosome, with approximately 40 copies of a 5-kb repeat unit, ranging from 95–99.9% nucleotide identity (**Fig. 1a**,**e** and **Supplementary Fig. 2**).

Reconstructing the ancestral autosomes

Our previous comparisons of the chicken Z chromosome with the orthologous human autosomes identified a set of 671 ancestral genes that were present on the ancestral amniote autosomes that became the chicken Z and W sex chromosomes (**Table 1** and **Supplementary**

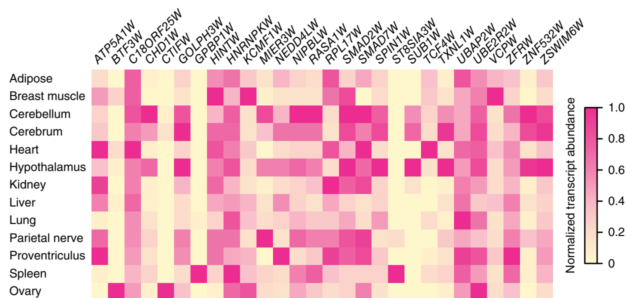

Figure 2 Chicken W-chromosome genes are broadly expressed across adult somatic tissues. The heat map shows the relative expression levels of W-chromosome genes in adult female tissues from the Chickspress RNA–seq data set (PRJNA204941). Expression for each gene is normalized to expression in the tissue with the highest expression level.

Table 1 Reconstruction of the gene content of the autosomal ancestors of the chicken sex chromosomes

	Lost	Ancestral	Gained
Bellott *et al.*[4]	49	671	493
Not unique	0	0	–427
	49	671	66
Updated annotations, new	+5	+5	+1
Updated annotations, withdrawn	–7	–49	–18
	47	627	49
Gained before birds diverged	0	+39	–39
Lost after birds diverged	–19	+19	0
Total	28	685	10

See also **Supplementary Table 2**.

Table 2) (ref. 4). We revisited our reconstruction of the gene content of the ancestral autosomes in light of recent improvements to the annotations of the chicken and human genomes, as well as newly published genome sequences from anole lizard[23], American alligator[24,25], and ostrich[26] that could allow us to determine whether gene gains or losses occurred before the common ancestor of extant birds (**Fig. 3a** and Online Methods). With this revised reconstruction, we identified 685 genes as present on the ancestral autosomes that became the avian Z and W sex chromosomes (**Table 1** and **Supplementary Table 2**).

Ancestral Z–W gene pairs in other avian species

We also searched for surviving ancestral Z–W gene pairs among the published W-linked genes from the 13 other avian species with published female genomes but without clone-based assemblies of the W chromosome (**Fig. 3a**)[7,18,21]. We stratified these candidate W-linked genes into two groups, on the basis of the amount of information used to identify W-linked genes for each species.

In three species—emu[27], crested ibis[28], and collared flycatcher[20]—candidate W-linked genes had been ascertained by comparing male and female genome assemblies to identify female-specific sequences[7,20], but W linkage was not confirmed by PCR or other additional mapping information. Fifty additional candidate W-linked genes were members of our set of 685 ancestral genes in one or more of these three species (**Fig. 3b** and **Supplementary Table 3**). Combining these 50 genes with the 28 genes from the chicken W chromosome gives a total of 78 genes of intermediate or high confidence in one or more of these four species (**Fig. 3b** and **Supplementary Table 3**). Together, these four species represent each of the three major lineages of birds—the paleognathae (emu), neoaves (crested ibis and collared flycatcher), and galloanserae (chicken)—allowing broad conclusions about W-chromosome evolution across all birds (**Fig. 3a**).

We regarded candidate W-chromosome genes in the ten remaining species[7] as lower-confidence predictions. In these species, candidate W-linked genes had been predicted directly from a female genome assembly. Without the control of a male genome assembly, two factors could potentially confound these gene predictions and diminish our ability to detect enrichments among surviving Z–W gene pairs. First, sequencing biases cause local variations in genome coverage that make it difficult to accurately identify the twofold changes in read depth that distinguish autosomal sequences from sex-linked sequences. Second, autosomal paralogs of Z-linked genes may appear similar to genuine Z–W gene pairs. Including predictions from these 10 species yielded another 79 ancestral genes, for a total of 157 putative ancestral genes in one or more of all 14 species (**Fig. 3b** and **Supplementary Table 3**).

Strategies for gene survival on sex-specific chromosomes

On the male-specific Y chromosomes of mammals, two evolutionary strategies contributed to gene survival despite widespread genetic decay: the retention and amplification of testis-expressed gene families and the conservation of ancestral X–Y gene pairs to maintain comparable expression between males and females[10,29]. Although analogous strategies should act on W chromosomes, and W chromosomes are expected to accumulate genes expressed solely in female-specific

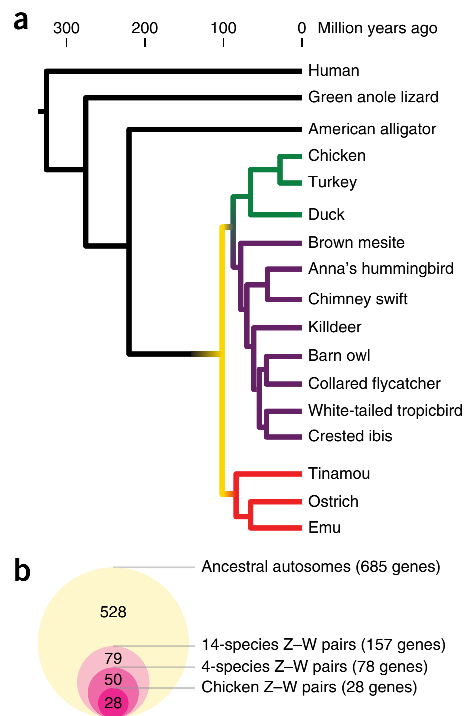

Figure 3 Ancestral Z–W gene pairs from 14 avian species. (**a**) Phylogenetic tree of the species in this study, with branches colored to highlight relationships among species. Humans diverged from birds 325 million years ago. Green anole lizard and American alligator diverged from birds 275 and 219 million years ago, respectively, and were used to resolve gene gains and losses between birds and mammals. Birds diverged from each other starting around 120 million years ago (yellow). The branches of major avian linages are colored: green, galloanserae; purple, neoaves; red, paleognaths. (**b**) Euler diagram showing overlapping sets of ancestral Z–W gene pairs identified in chicken (dark pink); 4 species (chicken, collared flycatcher, crested ibis, and emu) (medium pink); and all 14 published female avian genomes (light pink), as subsets of all 685 ancestral Z genes (light yellow). See also **Supplementary Table 3**.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

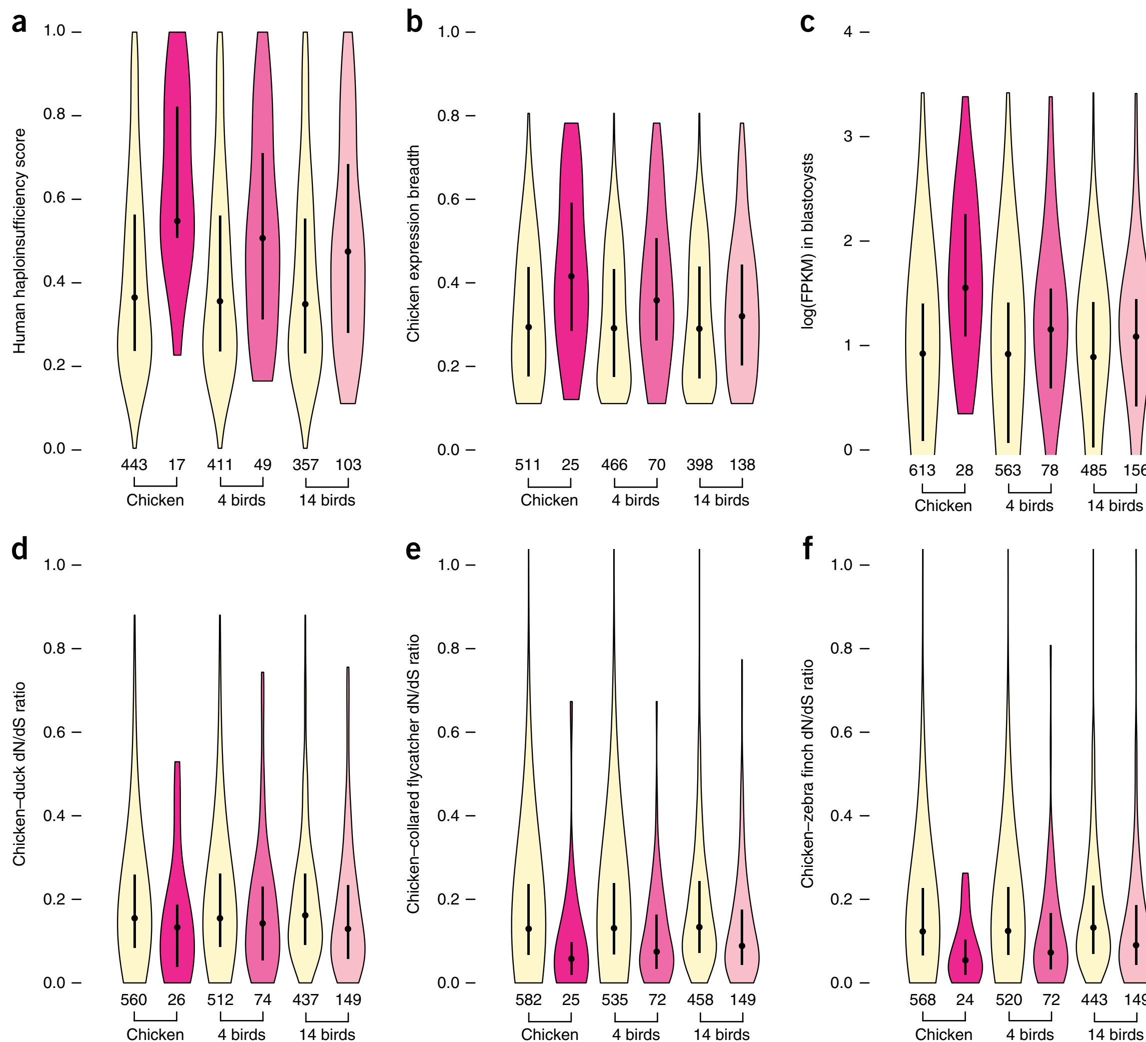

Figure 4 Factors in the survival of Z–W gene pairs. Violin plots, with the median (black circle) and interquartile range (black bar) indicated, compare annotations of ancestral Z–W gene pairs identified in chicken (dark pink); 4 species (chicken, collared flycatcher, crested ibis, and emu) (medium pink); and all 14 published female avian genomes (light pink) to annotations for the remainder of ancestral Z genes (light yellow). P values were obtained using one-tailed Mann–Whitney U tests. See the Online Methods and **Supplementary Figure 3**. (**a**) Human orthologs of ancestral Z–W gene pairs have higher probability of haploinsufficiency than other ancestral Z genes. Chicken, $P < 5.8 \times 10^{-5}$; 4 species, $P < 1.6 \times 10^{-3}$; 14 species, $P < 8.34 \times 10^{-4}$. (**b**) Chicken Z orthologs of ancestral Z–W gene pairs are more broadly expressed in adult chicken tissues than other ancestral Z genes. Chicken, $P < 2.1 \times 10^{-3}$; 4 species, $P < 3.8 \times 10^{-3}$; 14 species, $P < 0.059$. (**c**) Chicken Z orthologs of ancestral Z–W gene pairs are more highly expressed in chicken blastocysts. Chicken, $P < 7.7 \times 10^{-7}$; 4 species, $P < 1.1 \times 10^{-3}$; 14 species, $P < 2.8 \times 10^{-3}$. (**d**) Chicken Z orthologs of ancestral Z–W gene pairs have a reduced ratio of nonsynonymous substitutions per nonsynonymous site to synonymous substitutions per synonymous site (dN/dS) in alignments with orthologs in duck. Chicken, $P < 0.022$; 4 species, $P < 0.052$; 14 species, $P < 3.6 \times 10^{-3}$. (**e**) Chicken Z orthologs of ancestral Z–W gene pairs have a reduced dN/dS ratio in alignments with orthologs in collared flycatcher. Chicken, $P < 8.6 \times 10^{-5}$; 4 species, $P < 7.7 \times 10^{-5}$; 14 species, $P < 2.9 \times 10^{-5}$. (**f**) Chicken Z orthologs of ancestral Z–W gene pairs have a reduced dN/dS ratio in alignments with orthologs in zebra finch. Chicken, $P < 9.5 \times 10^{-5}$; 4 species, $P < 1.3 \times 10^{-4}$; 14 species, $P < 1.6 \times 10^{-4}$.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

tissues[16,17], we found that the female-specific chicken W chromosome has no genes that are exclusively expressed in sex-specific tissues (**Fig. 2**). In contrast to ampliconic genes on mammalian X and Y chromosomes[6,8,11–13,15], and even the chicken Z chromosome[4], the sole ampliconic gene on the W chromosome, *HINTW*, is broadly expressed (**Fig. 2**). Therefore, the first strategy cannot explain the survival of ancestral Z–W gene pairs in the chicken.

Despite widespread genetic decay on the sex-specific chromosome, dosage-sensitive genes functioning across many tissues and cell types may survive because their loss would have too great an impact on reproductive fitness and even viability. We looked for evidence that selection to maintain the correct dosage of ancestral genes might spare W-linked genes from genetic decay[10,29–31]. We compared each of these three lists of surviving ancestral Z–W gene pairs (from chicken alone, or 4 species, or all 14 species) to the other ancestral genes, reanalyzing published data sets for evidence that Z–W pair genes systematically differ from the ancestral genes on the Z chromosome that lack W homologs with regard to dosage sensitivity[32], breadth of expression[19,33,34], and intensity of purifying selection[35] (**Fig. 4**).

Z–W gene pairs are more dosage sensitive

First, we examined whether surviving Z–W gene pairs show signs of dosage sensitivity. We used published gene-by-gene estimates of the

probability of haploinsufficiency for the human genome[32], mapped on to their orthologs in the chicken genome. We found that the human orthologs of surviving Z–W gene pairs were more likely to be haploinsufficient than the human orthologs of ancestral Z-chromosome genes that lack W-chromosome homologs (one-tailed Mann–Whitney *U* test: chicken, $P < 5.8 \times 10^{-5}$; 4 species, $P < 1.6 \times 10^{-3}$; 14 species, $P < 8.34 \times 10^{-4}$) (**Fig. 4a** and **Supplementary Table 2**).

Additional evidence for the dosage sensitivity of specific Z–W gene pairs comes from human congenital disorders. Of the 28 chicken Z–W pairs, 3 have human orthologs implicated in congenital disorders caused by heterozygous loss-of-function mutations. Haploinsufficiency for *TCF4* is responsible for Pitt–Hopkins syndrome[36]. Parkes–Weber syndrome is caused by heterozygous inactivating mutations in *RASA1*. Cornelia de Lange syndrome results from haploinsufficiency for *NIPBL*[37]. To assess the statistical likelihood of finding 3 demonstrably haploinsufficient human orthologs among these 28 Z–W pairs, we examined all 4,562 human phenotypes with a known molecular basis catalogued in Online Mendelian Inheritance in Man (OMIM). Specifically, we searched for entries containing the word “haploinsufficient” or “haploinsufficiency” and found 238 disorders attributed to haploinsufficiency for a human gene. Three of 11 phenotypes mapping to human orthologs of Z–W pair genes were due to haploinsufficiency, whereas only 235 of 4,551 phenotypes in the rest of the human genome were due to haploinsufficiency (one-tailed Fisher’s exact test, $P < 0.017$). Taken together, the elevated haploinsufficiency probabilities and enrichment for human disorders caused by haploinsufficiency suggest that Z–W pairs are enriched for haploinsufficient genes.

Z–W gene pairs are more broadly expressed

Z–W gene pairs functioning across many tissues and cell types face additional selective constraints, which could prevent the loss of the W-linked copy, leading to an enrichment for broadly expressed genes among surviving Z–W pairs. Across adult chicken tissues, we observed that the Z homologs of Z–W gene pairs are more broadly expressed than ancestral Z genes that lack W homologs in chicken and in four species (one-tailed Mann–Whitney *U* test: chicken, $P < 2.1 \times 10^{-3}$; 4 species, $P < 3.8 \times 10^{-3}$; 14 species, $P < 0.059$) (**Fig. 4b** and **Supplementary Table 2**). This increased breadth of expression extends to the human orthologs of Z–W gene pairs; the human orthologs of Z–W gene pairs are more broadly expressed than the human orthologs of ancestral Z genes that lack W homologs in chicken and in four species (one-tailed Mann–Whitney *U* test: chicken, $P < 1.6 \times 10^{-3}$; 4 species, $P < 0.047$; 14 species, $P < 0.13$) (**Supplementary Fig. 3** and **Supplementary Table 2**). We conclude that the autosomal precursors of Z–W gene pairs were broadly expressed across adult tissues in the common ancestor of birds and mammals.

This breadth of expression also extends to the earliest stages of development. Ancestral Z–W pairs are more highly expressed in chicken blastocysts than are ancestral Z genes that lack W orthologs (one-tailed Mann–Whitney *U* test: chicken, $P < 7.7 \times 10^{-7}$; 4 species, $P < 1.1 \times 10^{-3}$; 14 species, $P < 2.8 \times 10^{-3}$) (**Fig. 4c** and **Supplementary Table 2**). We also examined the human orthologs of ancestral Z genes in published human embryonic transcriptome data[34]. We found that the human orthologs of ancestral Z–W pairs are more highly expressed in human blastocysts than are human orthologs of ancestral Z genes that lack W homologs in chicken and in 14 species (one-tailed Mann–Whitney *U* test: chicken, $P < 5.4 \times 10^{-5}$; 4 species, $P < 0.087$; 14 species, $P < 0.011$). We conclude that the autosomal precursors of the Z–W pairs were more broadly expressed across developmental time as well as across tissues in the amniote ancestor.

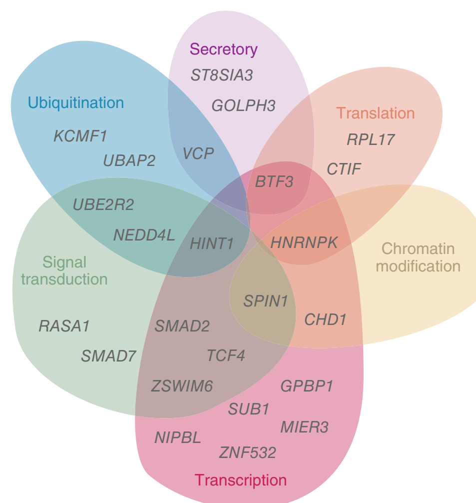

Figure 5 Regulatory annotations of chicken ancestral Z–W gene pairs. The Euler diagram depicts regulatory functions predicted for genes from selected Z–W gene pairs on the basis of UniProt annotations of human orthologs. See also **Supplementary Table 4**.

Z–W gene pairs are subject to stronger purifying selection

Previous comparisons among Z–W pairs in chicken, turkey, and duck showed that purifying selection has contributed significantly to the evolution of W-linked genes[18]. We reasoned that if Z–W gene pairs are haploinsufficient, alleles that impair the function of Z-linked homologs should be detrimental in both males and females, so that the Z homologs of Z–W gene pairs should also show signs of strong purifying selection. We examined Ensembl chicken ortholog alignment data for evidence that the Z-linked homologs of Z–W gene pairs were subject to stronger purifying selection than other ancestral Z-linked genes. In comparison to ancestral genes on the Z chromosome that lack W homologs, the Z-linked homologs of Z–W gene pairs have a reduced ratio of nonsynonymous to synonymous substitution (dN/dS) rates when chicken genes are compared to orthologs in duck for chicken and 14 species (one-tailed Mann–Whitney *U* test: chicken, $P < 0.022$; 4 species, $P < 0.052$; 14 species, $P < 3.6 \times 10^{-3}$) (**Fig. 4d** and **Supplementary Table 2**), and for all three groups both in collared flycatcher (one-tailed Mann–Whitney *U* test: chicken, $P < 8.6 \times 10^{-5}$; 4 species, $P < 7.7 \times 10^{-5}$; 14 species, $P < 2.9 \times 10^{-5}$) (**Fig. 4e** and **Supplementary Table 2**), and zebra finch (one-tailed Mann-Whitney *U* test: chicken, $P < 9.5 \times 10^{-5}$; 4 species, $P < 1.3 \times 10^{-4}$; 14 species, $P < 1.6 \times 10^{-4}$) (**Fig. 4f** and **Supplementary Table 2**). We conclude that, on avian W chromosomes, strong purifying selection has preserved a subset of ancestral genes that are more widely expressed and more dosage sensitive, just as it has on mammalian Y chromosomes.

Functional coherence of Z–W gene pairs

We recently characterized human X–Y gene pairs as performing an array of functions in gene expression and regulation, suggesting that X–Y pair genes could govern the expression of targets throughout the genome[10]. We asked whether our high-confidence set of ancestral genes that survived on the chicken W chromosome could be characterized as carrying out regulatory functions similar to the survivors on mammalian Y chromosomes. In comparison to ancestral genes on the Z chromosome that lack W homologs, Z–W pair genes are enriched for Gene Ontology (GO) annotations such as nucleic acid binding, nucleus, and transcription (**Supplementary Table 4**) that

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

suggest regulatory functions. We therefore looked more closely at the molecular functions of the 28 chicken Z–W pairs.

We observe that, in addition to the functions in transcription, translation, and protein stability attributed to mammalian X–Y pairs, chicken Z–W pairs are also predicted to act in protein secretion and signal transduction pathways (**Fig. 5**) (refs. 38,39). Specifically, several Z–W pairs share annotations that suggest roles in transducing TGF-β signaling (*SMAD2*, *SMAD7*, and *NEDD4L*) and modulating Wnt signaling (*UBE2R2*, *HINT1*, and *SPIN1*). Interactions between the TGF-β and Wnt pathways are critical for axis and pattern formation in early development[40], and, as morphogens, each can induce different cellular responses as a function of concentration, or dosage.

DISCUSSION

The preservation of broadly expressed, dosage-sensitive genes by purifying selection on avian W chromosomes offers a striking example of convergent evolution of the ZW and XY sex-chromosome systems. This survival strategy has been documented across diverse XY sex-chromosome systems; in *Drosophila miranda*, mammals, and threespine stickleback, purifying selection preserved a non-random set of ancestral Y-linked genes[10,30,31]. In *D. miranda*, surviving gene pairs on the Neo-X and Neo-Y chromosomes are expressed at higher levels and across more tissues than those genes that were lost to decay[30]. Likewise, in mammals, the surviving ancestral X–Y gene pairs are more broadly expressed across developmental time and adult tissues[10]. This strongly suggests that genes whose expression is required across a broad array of tissues are subject to greater constraints on gene dosage, making even the loss of a single copy costly. Like the Z–W gene pairs in birds we report here, the surviving X–Y gene pairs in mammals had higher predicted probabilities of haploinsufficiency, as well as ties to human syndromes caused by changes in gene dosage[10]. Similarly, the surviving X–Y gene pairs of both *D. miranda* and threespine stickleback were enriched for genes encoding proteins with many partners in protein–protein interaction networks[30,31]. Macromolecular complexes are sensitive to imbalances in the stoichiometry of their components, and an abundance of interactions is correlated with dosage sensitivity[41]. The repeated finding, across both female-specific (W) and male-specific (Y) chromosomes, that surviving ancestral genes are enriched for dosage-sensitive genes functioning across many tissues and cell types, contradicts the dire predictions of the imminent demise of sex-specific chromosomes due to genetic decay[42]. Purifying selection has been effective at preserving the ancestral dosage of critical genes on the sex-specific chromosome, even in the absence of crossing-over, through hundreds of millions of years of evolution.

Despite the evolutionary similarities, we note that the chicken W chromosome is remarkably divergent from all sequenced Y chromosomes, in that it lacks any genes expressed specifically in sex-specific organs or tissues. The gene content of mammalian Y chromosomes is frequently dominated by massively amplified testis-specific gene families that did not originate on the ancestral autosomes, even though they may have X-linked homologs[11,13–15]. In mammalian Y chromosomes, ancestral genes that narrowed their expression to male-specific tissues and became amplified into multicopy gene families were preserved across a greater number of species[10].

The relative simplicity of the W chromosome, with only broadly expressed ancestral genes and only one multicopy gene family, may be because its transmission is restricted to the female germ line. X, Y, and Z chromosomes pass through the male germ line, and all have acquired and amplified testis-expressed gene families[4,8,11–13,15,28]. This marked absence of acquired genes that are specifically expressed in the ovary or other female-specific tissues, even on a female-specific chromosome, suggests that, in amniotes, there is greater pressure to preserve or enhance male reproductive functions. Meiotic drive, which pits each chromosome against its homolog in a competition for transmission to the next generation, is one source of pressure on reproduction in males and females. However, there are more opportunities for meiotic drive to exert pressure during spermatogenesis than during oogenesis. Developing sperm are connected by cytoplasmic bridges, forming a syncytium that provides a venue for competition both during and after meiosis. For example, the ampliconic gene families on the X and Y chromosomes of the mouse are implicated in meiotic drive[11,43–45], even though they are expressed predominantly in post-meiotic germ cells[46]. During oogenesis, the arena for competition is narrower; any competition between homologous chromosomes must be resolved by the first meiotic division, when homologs separate and one is ejected into a polar body. Thus, W chromosomes, which only pass through the female germ line, may be subject to less disruptive selective pressures than those experienced by Y, X, and Z chromosomes in the male germ line. The complete sequences of W chromosomes from other birds, or the independently evolved Z and W chromosomes of snakes, could show whether the absence of acquired gene families that we observe in the chicken is a general feature of female-specific chromosomes.

We previously proposed that the dozen broadly expressed, dosage-sensitive genes on the human Y chromosome, along with their X-linked homologs that escape X-chromosome inactivation, are essential for the viability of 46,XY fetuses[10]. Two key observations support this hypothesis: first, that X–Y gene pairs are enriched for genes expressed in early development[10] and, second, that 99% of human 45,X conceptuses are not viable, while the remainder are often mosaic for all or part of a second sex chromosome[47–49].

Parallel lines of evidence in the chicken lead us to propose that the single-copy chicken Z–W pairs function to ensure female survival by providing the correct dosage of genes, especially those functioning in critical signaling pathways during early embryonic development. All 27 single-copy Z–W pairs are expressed in the developing chicken blastoderm, and the combined expression of Z–W gene pairs in females is comparable to the expression of the two Z homologs in males[19]. Both 2A:ZZW and 2A:Z0 aneuploid embryos have been observed in chicken at the blastocyst stage[50], but these embryos do not survive past 4 to 5 d of incubation[50,51], and sex chromosome aneuploidy is widely regarded as embryonic lethal in the chicken. Considering the severity of the three congenital developmental disorders linked to human orthologs of Z–W gene pairs, we conclude that hemizygosity for all Z–W gene pairs would likely result in early lethality.

In addition to their critical roles in maintaining embryonic viability, chicken Z–W and human X–Y gene pairs may have broader roles in sex determination and sexual dimorphism. Evidence of cell-autonomous sex determination in chickens has emerged from the study of lateral gynandromorphs[52], along with sexually dimorphic gene expression that precedes gonadal differentiation[19,52]. This leads us to speculate that one or more of the broadly expressed regulators found on the chicken W chromosome may have evolved to direct aspects of female fate in cell types across the body. In mammals, we are just beginning to understand the consequences of a fundamental sexual dimorphism, at the cellular level, arising from genetic differences between developmental regulators encoded by the X and Y chromosomes. In humans, for example, somatic mutations in the X-linked members of X–Y gene pairs were recently linked to the increased incidence of cancer in human males[53]. It will surely be of interest to compare and contrast

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

birds and mammals, taking advantage of the parallel evolutionary trajectories of avian ZW and mammalian XY chromosomes, to uncover new paradigms for understanding the regulation and development of sexual dimorphism in both health and disease.

URLs. Online Mendelian Inheritance in Man (OMIM) database (accessed 24 June 2015), ftp.omim.org/; RepeatMasker, http://www.repeatmasker.org/; custom Perl code for nucleotide dot plots, http://pagelab.wi.mit.edu/material-request.html; supplementary information mirror, http://jura.wi.mit.edu/page/papers/Bellott_et_al_2017/index.html.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS
We thank M. Delany (University of California, Davis) for UCD001 DNA, A. Vignal and M. Morrison (INRA Toulouse) for ChickRH6 radiation hybrid panel DNA, M. Lovett (Washington University, St. Louis) for chicken embryonic fibroblasts, C. Friedman and B. Trask for flow-sorted chicken W chromosomes, F. McCarthy for permission to use the Chickspress RNA–seq data set (PRJNA204941), and the "Chromas" Saint-Petersburg University Resource Center and L. Rapoport for technical assistance. This work was supported by the National Institutes of Health and the Howard Hughes Medical Institute. S.G. and E.G. were supported by the Russian Foundation of Basic Research (grant 16-04-01823a).

AUTHOR CONTRIBUTIONS
D.W.B., H.S., W.C.W., A.G.C., E.G., R.K.W., and D.C.P. planned the project. D.W.B., H.S., T.-J.C., D.L., and N.C. developed female-specific sequence-tagged sites. D.W.B., H.S., T.-J.C., and L.B. performed clone mapping. D.W.B., T.-J.C., N.K., T.G., and C.K. performed clone sequencing. S.G. and T.P. performed FISH analyses. D.W.B. and T.-J.C. performed RH mapping. D.W.B. and H.S. performed sequence analyses. D.W.B. and D.C.P. wrote the manuscript.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

1. Nanda, I. *et al.* 300 million years of conserved synteny between chicken Z and human chromosome 9. *Nat. Genet.* **21**, 258–259 (1999).
2. Ross, M.T. *et al.* The DNA sequence of the human X chromosome. *Nature* **434**, 325–337 (2005).
3. Fridolfsson, A.K. *et al.* Evolution of the avian sex chromosomes from an ancestral pair of autosomes. *Proc. Natl. Acad. Sci. USA* **95**, 8147–8152 (1998).
4. Bellott, D.W. *et al.* Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature* **466**, 612–616 (2010).
5. Handley, L.J., Ceplitis, H. & Ellegren, H. Evolutionary strata on the chicken Z chromosome: implications for sex chromosome evolution. *Genetics* **167**, 367–376 (2004).
6. Lahn, B.T. & Page, D.C. Four evolutionary strata on the human X chromosome. *Science* **286**, 964–967 (1999).
7. Zhou, Q. *et al.* Complex evolutionary trajectories of sex chromosomes across bird taxa. *Science* **346**, 1246338 (2014).
8. Mueller, J.L. *et al.* Independent specialization of the human and mouse X chromosomes for the male germ line. *Nat. Genet.* **45**, 1083–1087 (2013).
9. Charlesworth, B. & Charlesworth, D. The degeneration of Y chromosomes. *Phil. Trans. R. Soc. Lond. B* **355**, 1563–1572 (2000).
10. Bellott, D.W. *et al.* Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* **508**, 494–499 (2014).
11. Soh, Y.Q. *et al.* Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell* **159**, 800–813 (2014).
12. Hughes, J.F. *et al.* Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature* **483**, 82–86 (2012).
13. Hughes, J.F. *et al.* Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature* **463**, 536–539 (2010).
14. Li, G. *et al.* Comparative analysis of mammalian Y chromosomes illuminates ancestral structure and lineage-specific evolution. *Genome Res.* **23**, 1486–1495 (2013).
15. Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).
16. Moghadam, H.K., Pointer, M.A., Wright, A.E., Berlin, S. & Mank, J.E. W chromosome expression responds to female-specific selection. *Proc. Natl. Acad. Sci. USA* **109**, 8207–8211 (2012).
17. Mank, J.E., Hosken, D.J. & Wedell, N. Conflict on the sex chromosomes: cause, effect, and complexity. *Cold Spring Harb. Perspect. Biol.* **6**, a017715 (2014).
18. Wright, A.E., Harrison, P.W., Montgomery, S.H., Pointer, M.A. & Mank, J.E. Independent stratum formation on the avian sex chromosomes reveals inter-chromosomal gene conversion and predominance of purifying selection on the W chromosome. *Evolution* **68**, 3281–3295 (2014).
19. Ayers, K.L. *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).
20. Smeds, L. *et al.* Evolutionary analysis of the female-specific avian W chromosome. *Nat. Commun.* **6**, 7330 (2015).
21. Yazdi, H.P. & Ellegren, H. Old but not (so) degenerated—slow evolution of largely homomorphic sex chromosomes in ratites. *Mol. Biol. Evol.* **31**, 1444–1453 (2014).
22. Hori, T., Asakawa, S., Itoh, Y., Shimizu, N. & Mizuno, S. *Wpkci*, encoding an altered form of PKCI, is conserved widely on the avian W chromosome and expressed in early female embryos: implication of its role in female sex determination. *Mol. Biol. Cell* **11**, 3645–3660 (2000).
23. Alföldi, J. *et al.* The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* **477**, 587–591 (2011).
24. St John, J.A. *et al.* Sequencing three crocodilian genomes to illuminate the evolution of archosaurs and amniotes. *Genome Biol.* **13**, 415 (2012).
25. Green, R.E. *et al.* Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. *Science* **346**, 1254449 (2014).
26. Zhang, J., Li, C., Zhou, Q. & Zhang, G. Improving the ostrich genome assembly using optical mapping data. *Gigascience* **4**, 24 (2015).
27. Vicoso, B., Kaiser, V.B. & Bachtrog, D. Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. *Proc. Natl. Acad. Sci. USA* **110**, 6453–6458 (2013).
28. Li, S. *et al.* Genomic signatures of near-extinction and rebirth of the crested ibis and other endangered bird species. *Genome Biol.* **15**, 557 (2014).
29. Lahn, B.T. & Page, D.C. Functional coherence of the human Y chromosome. *Science* **278**, 675–680 (1997).
30. Kaiser, V.B., Zhou, Q. & Bachtrog, D. Nonrandom gene loss from the *Drosophila miranda* neo-Y chromosome. *Genome Biol. Evol.* **3**, 1329–1337 (2011).
31. White, M.A., Kitano, J. & Peichel, C.L. Purifying selection maintains dosage-sensitive genes during degeneration of the threespine stickleback Y chromosome. *Mol. Biol. Evol.* **32**, 1981–1995 (2015).
32. Huang, N., Lee, I., Marcotte, E.M. & Hurles, M.E. Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet.* **6**, e1001154 (2010).
33. Merkin, J., Russell, C., Chen, P. & Burge, C.B. Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science* **338**, 1593–1599 (2012).
34. Yan, L. *et al.* Single-cell RNA–Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1131–1139 (2013).
35. Cunningham, F. *et al.* Ensembl 2015. *Nucleic Acids Res.* **43**, D662–D669 (2015).
36. Zweier, C. *et al.* Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt–Hopkins syndrome). *Am. J. Hum. Genet.* **80**, 994–1001 (2007).
37. Tonkin, E.T., Wang, T.J., Lisgo, S., Bamshad, M.J. & Strachan, T. *NIPBL*, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. *Nat. Genet.* **36**, 636–641 (2004).
38. Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F. & Hamosh, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* **43**, D789–D798 (2015).
39. UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204–D212 (2015).
40. Skromne, I. & Stern, C.D. Interactions between Wnt and Vg1 signalling pathways initiate primitive streak formation in the chick embryo. *Development* **128**, 2915–2927 (2001).
41. Papp, B., Pál, C. & Hurst, L.D. Dosage sensitivity and the evolution of gene families in yeast. *Nature* **424**, 194–197 (2003).
42. Aitken, R.J. & Marshall Graves, J.A. The future of sex. *Nature* **415**, 963 (2002).
43. Cocquet, J. *et al.* A genetic basis for a postmeiotic X versus Y chromosome intragenomic conflict in the mouse. *PLoS Genet.* **8**, e1002900 (2012).
44. Cocquet, J. *et al.* The multicopy gene *Sly* represses the sex chromosomes in the male mouse germline after meiosis. *PLoS Biol.* **7**, e1000244 (2009).
45. Conway, S.J. *et al.* Y353/B: a candidate multiple-copy spermiogenesis gene on the mouse Y chromosome. *Mamm. Genome* **5**, 203–210 (1994).
46. Mueller, J.L. *et al.* The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. *Nat. Genet.* **40**, 794–799 (2008).

47. Cockwell, A., MacKenzie, M., Youings, S. & Jacobs, P. A cytogenetic and molecular study of a series of 45,X fetuses and their parents. *J. Med. Genet.* **28**, 151–155 (1991).
48. Hassold, T., Benham, F. & Leppert, M. Cytogenetic and molecular analysis of sex-chromosome monosomy. *Am. J. Hum. Genet.* **42**, 534–541 (1988).
49. Hook, E.B. & Warburton, D. The distribution of chromosomal genotypes associated with Turner's syndrome: livebirth prevalence rates and evidence for diminished fetal mortality and severity in genotypes associated with structural X abnormalities or mosaicism. *Hum. Genet.* **64**, 24–27 (1983).
50. Fechheimer, N.S. Origins of heteroploidy in chicken embryos. *Poult. Sci.* **60**, 1365–1371 (1981).
51. Bloom, S.E. Chromosome abnormalities in chicken (*Gallus domesticus*) embryos: types, frequencies and phenotypic effects. *Chromosoma* **37**, 309–326 (1972).
52. Zhao, D. *et al.* Somatic sex identity is cell autonomous in the chicken. *Nature* **464**, 237–242 (2010).
53. Dunford, A. *et al.* Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex bias. *Nat. Genet.* **49**, 10–16 (2017).

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

ONLINE METHODS

Single-haplotype iterative mapping and sequencing. We employed the single-haplotype iterative mapping and sequencing (SHIMS) strategy to assemble the chicken W chromosome sequence from 41 BAC and 123 fosmid clones (**Supplementary Table 1**). These clones were obtained from four BAC libraries (CH261, TAM31, TAM32, and TAM33) (refs. 54,55) and two fosmid libraries (J_AD and J_AE)[56], which provide combined ~16-fold coverage of the W chromosome of the single female red jungle fowl of the UCD001 line (RJF 256) (ref. 56). Thirty-seven BACs and 5 fosmids were sequenced on ABI 3730 machines at the McDonnell Genome Institute, and 4 BACs and 118 fosmids were sequenced on an Illumina MiSeq instrument at the Whitehead Institute; see individual GenBank records for assembly details. We previously used the SHIMS strategy to produce finished sequence from mammalian Y, human X and chicken Z chromosomes[4,8,10–13,15]. The major steps in the SHIMS strategy are outlined below.

Marker generation. We identified female-specific sequence contigs in the draft assembly of the chicken genome using existing genetic linkage data[56,57], direct sequencing of flow-sorted W-chromosome DNA, electronic searches for close homologs of Z-linked gene sequences[4], and electronic subtraction using short-read genomic data from a male White Leghorn[58]. We used these sequences to develop sequence-tagged sites (STS) and verified that they were female specific by PCR on DNA from a male and female red jungle fowl.

Initial BAC selection and sequencing. We identified large-insert BAC and fosmid clones and organized them into contigs of overlapping clones on the basis of (i) high-density filter hybridization using pools of overgo probes, (ii) electronic mapping of clone end sequences to female genomic sequences, and (iii) BAC fingerprint contig analysis. We confirmed the resulting contigs by PCR using female-specific STS markers and selected tiling paths for sequencing.

Distinguishing repeat copies and finding true tiling paths. We scrutinized overlaps between clones within repetitive regions for sequence differences or sequence family variants (SFVs). The presence of SFVs indicates that the clones belong to distinct copies of the same repeat family, and we used SFV patterns to identify clones that truly overlap. This produced new tiling paths. We reiterated this process until all overlaps were consistent.

Extension and joining of large-insert clone contigs. We identified clones that extend outward from or link existing contigs using high-density filter hybridization and electronic mapping of clone end sequences.

Gap closure. Regions composed of repeats with units less than 10 kb and greater than 99% identity frustrate the assembly of individual clones and are not well represented in our assemblies. These regions include both gene-poor regions, like centromeres, telomeres, and heterochromatin, and gene-rich regions, such as the *HINTW* array. No current technology is able to access these regions. Wherever possible, we attempted to find the boundaries of these arrays and obtain a representative repeat unit.

Calculation of sequence accuracy. The initial error rate estimated for clone sequencing and assembly was 1 in 28 kb. However, as 23% of our sequence was covered redundantly by two BACs, we were able to identify and resolve all discrepancies in redundantly covered regions so that the error rate for these regions was zero. Therefore, the final error rate was estimated to be $0.77 \times 1/28{,}000 + 0.23 \times 0 \approx 1/36{,}000$, or 1 in 36 kb.

Ordering and orienting sequence contigs. The structure of the chicken W chromosome presents a unique challenge to traditional techniques for chromosome mapping and assembly. Isolated islands of euchromatin are separated by massive heterochromatic tandem arrays, each composed of one of three families of genome-typical interspersed repeats[59–61]. We employed two independent and complementary methods to order and orient the 13 contigs of our sequence map along the W chromosome.

Radiation hybrid mapping. We tested 119 STS markers (**Fig. 1d** and **Supplementary Data 2**) on the ChickRH6 panel[62], a 6,000-rad panel consisting of 90 hybrid clones, and constructed an RH map using RHMAPPER[63]. We thereby assigned each of the 13 sequence contigs to one of three distinct linkage groups on the RH map (**Fig. 1a**,**d**).

Lampbrush FISH. We ordered the three RH linkage groups along the W chromosome using lampbrush FISH (**Fig. 1e–h** and **Supplementary Fig. 1**). The lampbrush W chromosome features a series of seven condensed heterochromatic chromomeres along its axis; these are numbered in ascending order from the tip of the long arm (**Fig. 1e**). The three major repetitive sequence families of the W chromosome were previously mapped to chromomeres 1, 3, 5, and 6 by lampbrush FISH[59,60] (**Fig. 1e**). We found that the three remaining chromomeres—2, 4, and 7—correspond to the three radiation hybrid linkage groups (**Fig. 1a**,**c**,**e**,**f** and **Supplementary Fig. 1**).

The first linkage group contains 3 contigs, spans 2 Mb, contains 8 genes, and corresponds to chromomere 2, on the long arm of the W chromosome (**Fig. 1e**,**f** and **Supplementary Fig. 1**). This linkage group terminated in sequences from the XhoI repeat family (**Fig. 1e** and **Supplementary Table 1**). The XhoI repeat family was previously mapped to the adjacent chromomere 3 (ref. 60), suggesting that we captured the border of this heterochromatic array (**Fig. 1e**).

The second linkage group contains 6 contigs, spans 3 Mb, contains 12 genes, and corresponds to chromomere 4 (**Fig. 1e**,**g**, **Supplementary Fig. 1**, and **Supplementary Table 1**), near the centromere of the W chromosome[64]. Consistent with proximity to the centromere, single-copy markers assigned to this linkage group were retained at higher frequency in the radiation hybrid panel than were markers in the other two linkage groups (**Fig. 1d**).

The third linkage group consists of 4 contigs, spans 2 Mb, contains 8 genes, and corresponds to chromomere 7, on the short arm of the W chromosome near the pseudoautosomal region where the Z and W chromosomes cross over during female meiosis (**Fig. 1e**,**h**, **Supplementary Fig. 1**, and **Supplementary Table 1**). Consistent with the results of lampbrush FISH, the proximal end of this linkage group contained the only ampliconic sequence on the W chromosome, a tandem array of *HINTW* genes (**Fig. 1c** and **Supplementary Fig. 1**). The *HINTW* array was previously mapped to the short arm of the W chromosome by metaphase FISH[22]. The third linkage group terminated in the pseudoautosomal region, which contained no genes but instead consisted entirely of telomeric and subtelomeric repeats shared with the Z chromosome (**Fig. 1e** and **Supplementary Fig. 1**).

Chromosomal FISH analyses. We performed FISH assays on *Gallus gallus domesticus* lampbrush chromosomes as previously described[64]. Briefly, lampbrush chromosomes were manually isolated from oocyte nuclei, dehydrated in 96% ethanol, air dried, and treated with RNase A. BAC probes were labeled with digoxygenin and denatured together with unlabeled competitor DNA and the lampbrush chromosomes before hybridization. Probes were detected with antibodies against digoxygenin conjugated to Cy3 (JacksonImmunoResearch Laboratories, 200-162-156) at a 1:400 dilution. Chromosomes were stained with DAPI and imaged on a fluorescence microscope. Experimental procedures involving chicken oocytes were approved by the Saint Petersburg State University Ethics Committee (statement 131-03-2).

Interspersed repeats. Interspersed repeats were electronically identified with RepeatMasker.

Identification of genes and transcription units. We identified genes and transcripts as previously described[4,10]. Briefly, we used TWINSCAN[65,66] with human as the informant genome and EST sequences from the BBSRC ChickEST database[67], supplemented by our own ESTs from adult ovary (SRP000097). We specifically searched for homologs of all genes found in the finished sequence of the chicken Z chromosome to detect ancestral genes as well as any genes co-acquired by the Z and W chromosomes. We validated transcription of predicted genes by RT–PCR and capillary sequencing, as well as 454 sequencing of ovary cDNA and Illumina-based RNA–seq (PRJNA204941).

Dot plots. Triangular dot plots (representing intrachromosomal sequence similarity) and square dot plots (representing interchromosomal sequence similarity) were generated with a custom Perl script (see URLs).

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

Reconstructing ancestral autosomes. Our previous comparisons of the chicken Z chromosome with orthologous regions of human autosomes identified 720 genes that were present on the ancestral amniote chromosomes that became the chicken Z and W sex chromosomes (**Table 1** and **Supplementary Table 2**) (ref. 4). Of these 720 genes, 671 had syntenic orthologs in both human and chicken. The other 49 genes had syntenic orthologs in human and an outgroup species (amphibians or fish), but not in chicken, indicating that these genes were lost along the lineage leading to chicken (**Table 1** and **Supplementary Table 2**). We also identified 66 distinct families of genes (493 genes in all) that had been added to the chicken Z chromosome but were not present on the ancestral amniote autosomes (**Table 1** and **Supplementary Table 2**) (ref. 4).

Previously, we relied on assignments of chicken and human orthologs in the Ensembl database (version 52). We reexamined all 786 distinct genes or families (720 ancestral plus 66 acquired) in light of recent improvements to annotations of the chicken and human genomes (Ensembl version 80) (ref. 35). This allowed us to eliminate genes that represented errors or redundancies in previous annotations of the chicken and human genomes and to add genes that had been overlooked by previous annotation efforts (**Table 1** and **Supplementary Table 2**). This reduced the number of ancestral genes maintained on the chicken Z chromosome from 671 to 627 (**Table 1** and **Supplementary Table 2**). This also reduced the number of ancestral genes evidently lost from the chicken Z chromosome from 49 to 47 and the number of distinct gene families added to the chicken Z chromosome from 66 to 49 (**Table 1** and **Supplementary Table 2**).

Our previous analyses were limited by the absence of genome sequences from species more closely related to chicken than humans; the 47 losses and 49 gains could have occurred at any time after birds diverged from mammals, about 325 million years ago[68] (**Fig. 3a**). To determine which of these gains and losses had occurred in the avian ancestor and which were specific to chicken, we looked for syntenic orthologs of genes in three species more closely related to chicken than human: anole lizard[23], American alligator[24,25], and ostrich[26] (**Fig. 3a**). Birds diverged from lizards around 275 million years ago and from crocodilians around 219 million years ago[68] (**Fig. 3a**). All birds have orthologous Z and W sex chromosomes, and therefore the ancestral autosomes must have begun to diverge before the earliest split in the avian tree—between the Palaeognathae (like the ostrich) and the Neognathae (like the chicken)—around 120 million years ago[68] (**Fig. 3a**).

Of the 47 human genes we had identified as lost in the lineage leading to chicken, 19 maintained a syntenic ortholog in the ostrich, indicating that these genes were actually present on the ancestral autosomes before the radiation of birds (**Table 1** and **Supplementary Table 2**). Similarly, of the 49 distinct genes we identified as gained in the lineage leading to chicken, 39 genes had a syntenic ortholog in lizard, alligator, or ostrich, indicating that these gains took place on the ancestral autosomes, before they evolved into the Z and W sex chromosomes (**Table 1** and **Supplementary Table 2**). Combining the 627 ancestral genes maintained on the chicken Z chromosome with the 19 genes lost from the chicken Z chromosome after the radiation of birds, plus the 39 genes gained by the ancestral autosomes before birds diverged, yields a total of 685 genes present on the ancestral autosomes that became the avian Z and W sex chromosomes (**Table 1** and **Supplementary Table 2**).

W-linked gene expression. We quantified the abundances of chicken transcripts from the Chickspress RNA–seq data set (PRJNA204941) using kallisto version 0.42.3 (ref. 69) and edgeR[70]. We normalized the transcript abundances for each gene to the abundance in the highest expressing tissue for that gene.

OMIM. We downloaded the full text of OMIM[38] and searched entries for "haploinsufficient" or "haploinsufficiency," limiting our search to phenotypes with a known molecular basis. We examined each of the resulting entries to verify that there was evidence that the phenotype was caused by haploinsufficiency.

Functional annotation. We mapped published functional annotation data onto our set of 685 ancestral genes and their human orthologs. For expression breadth, we normalized the expression of each gene to the highest RPKM in any tissue and took the average expression across all tissues. We used UniProt annotations to identify chicken Z–W pair genes involved in regulatory processes.

Statistics. We tested whether the human orthologs of Z–W gene pairs were enriched for phenotypes caused by haploinsufficiency, relative to the rest of the human genome, using a one-tailed Fisher's exact test, because of the small number of Z–W gene pairs whose orthologs are annotated in OMIM. We tested for enrichments in the annotations of ancestral Z–W gene pairs identified in chicken, 4 species (chicken, collared flycatcher, crested ibis, and emu), and all 14 published female avian genomes versus the remainder of ancestral Z genes using one-tailed Mann–Whitney U tests. We report all of our comparisons, and, in every case, all three classes of Z–W pairs differed from other ancestral genes in the expected direction, making correction for multiple comparisons unnecessary. We attribute the reduced significance for comparisons involving the sets of 4 and 14 species to noise from low-confidence gene predictions in these species.

The exact numbers used to calculate the P values for **Figure 4**, along with the associated test statistic, U, are as follows. The human orthologs of ancestral Z–W pairs had a higher probability of haploinsufficiency than other ancestral Z genes. Chicken Z–W pairs $n = 17$, other ancestral Z genes $n = 443$, $P < 5.8 \times 10^{-5}$, $U = 5{,}840.5$; 4-species Z–W pairs $n = 49$, other ancestral Z genes $n = 411$, $P < 1.6 \times 10^{-3}$, $U = 12{,}666$; 14-species Z–W pairs $n = 103$, other ancestral Z genes $n = 357$, $P < 8.34 \times 10^{-4}$, $U = 22{,}122.5$. The chicken Z orthologs of ancestral Z–W pairs were more broadly expressed in adult chicken tissues than other ancestral Z genes. Chicken Z–W pairs $n = 25$, other ancestral Z genes $n = 511$, $P < 2.1 \times 10^{-3}$, $U = 8{,}561$; 4-species Z–W pairs $n = 70$, other ancestral Z genes $n = 466$, $P < 3.8 \times 10^{-3}$, $U = 19{,}546$; 14-species Z–W pairs $n = 138$, other ancestral Z genes $n = 398$, $P < 0.059$, $U = 29{,}919$. The chicken Z orthologs of ancestral Z–W pairs were more highly expressed in chicken blastocysts than other ancestral Z genes. Chicken Z–W pairs $n = 28$, other ancestral Z genes $n = 613$, $P < 7.7 \times 10^{-7}$, $U = 13{,}188$; 4-species Z–W pairs $n = 78$, other ancestral Z genes $n = 563$, $P < 1.1 \times 10^{-3}$, $U = 26{,}684$; 14-species Z–W pairs $n = 156$, other ancestral Z genes $n = 485$, $P < 2.8 \times 10^{-3}$, $U = 43{,}410.5$. The chicken Z orthologs of ancestral Z–W pairs had a reduced dN/dS ratio in alignments with their orthologs in duck. Chicken Z–W pairs $n = 26$, other ancestral Z genes $n = 560$, $P < 0.022$, $U = 5{,}580.5$; 4-species Z–W pairs $n = 74$, other ancestral Z genes $n = 512$, $P < 0.052$, $U = 16{,}728.5$; 14-species Z–W pairs $n = 149$, other ancestral Z genes $n = 437$, $P < 3.6 \times 10^{-3}$, $U = 27{,}753$. The chicken Z orthologs of ancestral Z–W pairs had a reduced dN/dS ratio in alignments with their orthologs in collared flycatcher. Chicken Z–W pairs $n = 25$, other ancestral Z genes $n = 582$, $P < 8.6 \times 10^{-5}$, $U = 4{,}048$; 4-species Z–W pairs $n = 72$, other ancestral Z genes $n = 535$, $P < 7.7 \times 10^{-5}$, $U = 13{,}971$; 14-species Z–W pairs $n = 149$, other ancestral Z genes $n = 458$, $P < 2.9 \times 10^{-5}$, $U = 26{,}636$. The chicken Z orthologs of ancestral Z–W pairs had a reduced dN/dS ratio in alignments with their orthologs in zebra finch. Chicken Z–W pairs $n = 24$, other ancestral Z genes $n = 568$, $P < 9.5 \times 10^{-5}$, $U = 3{,}750.5$; 4-species Z–W pairs $n = 72$, other ancestral Z genes $n = 520$, $P < 1.3 \times 10^{-4}$, $U = 13{,}741.5$; 14-species Z–W pairs $n = 149$, other ancestral Z genes $n = 443$, $P < 1.6 \times 10^{-4}$, $U = 26{,}476.5$.

Data availability. GenBank accession numbers for BAC and fosmid sequences are listed in **Supplementary Table 1**. Z and W transcript sequences assembled from PRJNA204941 have been deposited at DDBJ, EMBL, and GenBank under accession GENL00000000. The version described in this paper is the first version, GENL01000000.

54. Lee, M.K. *et al.* Construction and characterization of three BAC libraries for analysis of the chicken genome. *Anim. Genet.* **34**, 151–152 (2003).
55. Wallis, J.W. *et al.* A physical map of the chicken genome. *Nature* **432**, 761–764 (2004).
56. International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004).
57. Zody, M.C. *Investigation of Mechanics of Mutation and Selection by Comparative Sequencing.* PhD thesis, Uppsala Univ. (2009).
58. Chen, N., Bellott, D.W., Page, D.C. & Clark, A.G. Identification of avian W-linked contigs by short-read sequencing. *BMC Genomics* **13**, 183 (2012).
59. Itoh, Y. & Mizuno, S. Molecular and cytological characterization of SspI-family repetitive sequence on the chicken W chromosome. *Chromosome Res.* **10**, 499–511 (2002).
60. Solovei, I., Ogawa, A., Naito, M., Mizuno, S. & Macgregor, H. Specific chromomeres on the chicken W lampbrush chromosome contain specific repetitive DNA sequence families. *Chromosome Res.* **6**, 323–327 (1998).

61. Saitoh, Y. & Mizuno, S. Distribution of XhoI and EcoRI family repetitive DNA sequences into separate domains in the chicken W chromosome. *Chromosoma* **101**, 474–477 (1992).
62. Morisson, M. *et al.* ChickRH6: a chicken whole-genome radiation hybrid panel. *Genet. Sel. Evol.* **34**, 521–533 (2002).
63. Slonim, D., Kruglyak, L., Stein, L. & Lander, E. Building human genome maps with radiation hybrids. *J. Comput. Biol.* **4**, 487–504 (1997).
64. Krasikova, A. *et al.* On the positions of centromeres in chicken lampbrush chromosomes. *Chromosome Res.* **14**, 777–789 (2006).
65. Flicek, P., Keibler, E., Hu, P., Korf, I. & Brent, M.R. Leveraging the mouse genome for gene prediction in human: from whole-genome shotgun reads to a global synteny map. *Genome Res.* **13**, 46–54 (2003).
66. Korf, I., Flicek, P., Duan, D. & Brent, M.R. Integrating genomic homology into gene structure prediction. *Bioinformatics* **17** (Suppl. 1), S140–S148 (2001).
67. Boardman, P.E. *et al.* A comprehensive collection of chicken cDNAs. *Curr. Biol.* **12**, 1965–1969 (2002).
68. Hedges, S.B., Dudley, J. & Kumar, S. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22**, 2971–2972 (2006).
69. Bray, N.L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA–seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
70. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.