LETTERS

Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition

Daniel W. Bellott[1], Helen Skaletsky[1], Tatyana Pyntikova[1], Elaine R. Mardis[2], Tina Graves[2], Colin Kremitzki[2], Laura G. Brown[1], Steve Rozen[1], Wesley C. Warren[2], Richard K. Wilson[2] & David C. Page[1]

In birds, as in mammals, one pair of chromosomes differs between the sexes. In birds, males are ZZ and females ZW. In mammals, males are XY and females XX. Like the mammalian XY pair, the avian ZW pair is believed to have evolved from autosomes, with most change occurring in the chromosomes found in only one sex—the W and Y chromosomes[1–5]. By contrast, the sex chromosomes found in both sexes—the Z and X chromosomes—are assumed to have diverged little from their autosomal progenitors[2]. Here we report findings that challenge this assumption for both the chicken Z chromosome and the human X chromosome. The chicken Z chromosome, which we sequenced essentially to completion, is less gene-dense than chicken autosomes but contains a massive tandem array containing hundreds of duplicated genes expressed in testes. A comprehensive comparison of the chicken Z chromosome with the finished sequence of the human X chromosome demonstrates that each evolved independently from different portions of the ancestral genome. Despite this independence, the chicken Z and human X chromosomes share features that distinguish them from autosomes: the acquisition and amplification of testis-expressed genes, and a low gene density resulting from an expansion of intergenic regions. These features were not present on the autosomes from which the Z and X chromosomes originated but were instead acquired during the evolution of Z and X as sex chromosomes. We conclude that the avian Z and mammalian X chromosomes followed convergent evolutionary trajectories, despite their evolving with opposite (female versus male) systems of heterogamety. More broadly, in birds and mammals, sex chromosome evolution involved not only gene loss in sex-specific chromosomes, but also marked expansion and gene acquisition in sex chromosomes common to males and females.

A century ago, Herman Muller proposed the first theory of sex chromosome evolution—that the X and Y chromosomes of *Drosophila* evolved from an ordinary pair of autosomes, and that genes on the Y chromosome had gradually deteriorated while their counterparts on the X had been preserved[1]. In the 1960s, Susumu Ohno applied Muller's theory to the sex chromosomes of vertebrates, arguing that while the sex-specific W and Y chromosomes of birds and mammals had degenerated, the content of the Z and X chromosomes remained intact[2]. Four decades on, comparisons of the human X and Y chromosomes have underscored the dramatic evolutionary changes on the Y chromosome[3–5], but the assumption that the X chromosome has been evolutionarily stable remains unexamined.

The evolutionary relationship between the mammalian X chromosome and the avian Z chromosome has been the subject of much speculation, but it also remains unresolved. Ohno conjectured that the X chromosomes of mammals were orthologous to the Z chromosomes of birds[2]. However, comparative mapping of 30 Z-linked genes indicated that the chicken Z chromosome was orthologous to human chromosomes 5, 8, 9 and 18, and not to the human X chromosome[6,7]. These findings were supported by the draft sequence of the chicken genome, but only about one-third of the sequence of the Z chromosome was present in the assembly, leaving open the possibility that regions of orthology between the avian Z and mammalian X chromosomes had yet to be detected[8]. The recent discovery that a subset of the five platypus X chromosomes contains orthologues of genes on the chicken Z chromosome renewed speculation that the avian Z and mammalian X chromosomes have a common origin[9–12]. To accommodate the results of comparative gene mapping experiments, it has been proposed that the chicken Z and human X chromosomes were derived from different portions of an ancestral proto-sex chromosome that broke apart, leaving Z orthologues autosomal in mammals and X orthologues autosomal in birds[11,12].

To reconstruct and compare the evolutionary trajectories of the avian Z and mammalian X chromosomes, we have produced the finished sequence of the chicken Z chromosome (Supplementary Figs 1–3). The resulting sequence spans roughly 80 megabases (Mb), is complete apart from four gaps and is accurate to about one nucleotide per megabase. The chicken Z chromosome contains ~1,000 genes (Supplementary Table 1). This makes the Z chromosome less gene-dense than any chicken autosome, with 11 genes per megabase, which is less than half of the chicken autosomal average of 25 genes per megabase[8] (Table 1). Conversely, the density of interspersed repeats is 60% higher in the Z chromosome than in chicken autosomes (Supplementary Fig. 1 and Table 1). Most of these repeats are long interspersed nuclear elements (LINEs), whose abundance in the Z chromosome is 70% higher than in autosomes (Supplementary Fig. 1 and Table 1). As a result, the Z chromosome is structurally distinct from the rest of the chicken genome.

The Z chromosome's most prominent feature is a previously unrecognized tandem array of testis-expressed genes, extending over 11 Mb at the distal end of the long arm (Fig. 1 and Supplementary Fig. 4). This array constitutes nearly 15% of the Z chromosome, one-fifth of all chicken segmental duplications and 1% of the entire chicken genome[8] (Fig. 1a, b). This sequence was initially reported as heterochromatin[13], but we find that three genes are present in each repeat

Table 1 | Comparison of structural features of chromosomes Z and X with autosomes

	Chicken Z chromosome	Chicken autosomes	Human X chromosome	Human autosomes
Genes per megabase	12	25	7	12
Interspersed repeats	15%	9.4%	56%	45%
LINEs	11%	6.4%	32%	21%
Average gene size	21 kb	27 kb	49 kb	57 kb

kb, kilobase.

[1]Howard Hughes Medical Institute, Whitehead Institute, and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. [2]The Genome Center, Washington University School of Medicine, 4444 Forest Park Boulevard, St Louis, Missouri 63108, USA.

unit and that a smaller flanking array contains a fourth (Fig. 1c, d, Supplementary Fig. 5 and Supplementary Table 2). Together, these four gene families total hundreds of copies and constitute almost one-third of the protein-coding genes on the Z chromosome (Fig. 1d and Supplementary Table 2). All four gene families are expressed predominantly in the testis (Fig. 1e). We have termed this massive array of testis-expressed genes the 'Z amplicon'.

With the finished sequence of the Z chromosome in hand, we set out to test Ohno's hypothesis that the avian Z and mammalian X chromosomes are orthologous. To reconstruct and visualize evolutionary relationships between chicken and human chromosomes, we systematically plotted the locations of orthologous gene pairs (Supplementary Figs 6 and 7). We find that none of the ~1,000 genes on the chicken Z chromosome has an orthologue on the human X chromosome (Fig. 2a, b and Supplementary Table 1). The Z chromosome is orthologous only to portions of human autosomes 5, 9 and 18 (Fig. 2a). Contrary to initial reports[7], the Z chromosome is not orthologous to human chromosome 8 (Supplementary Fig. 6). In reciprocal fashion, the human X chromosome is orthologous only to portions of chicken autosomes 1 and 4, and not to the Z chromosome[5] (Fig. 2b). On the basis of this comprehensive analysis, we conclude that genes that are sex-linked in chickens are autosomal in humans, and vice versa, in broad agreement with earlier comparative mapping experiments[6,7].

Although the Z and X chromosomes show no signs of orthology, it is possible that they were recruited from different portions of a proto-sex chromosome in the common ancestor of birds and mammals[11]. Some investigators have raised this possibility on the basis of comparative gene mapping in the platypus[11]. However, the platypus does not form an outgroup to birds and mammals, and cannot resolve which state is the ancestral one: a platypus-like linkage of Z-orthologous genes and X-orthologous genes, or the separation we observe in chicken and human. Other researchers have attempted to resolve this question by comparisons with an outgroup genome that is far from complete[12]. Instead, we compared the Z and X chromosomes with the genomes of the four closest outgroup species whose genomes have been sequenced and assembled. Each species represents a different order of teleost fish, which diverged from land vertebrates more than 450 million years ago[14]. After they diverged from birds and mammals, but before they diverged from each other, these fish species experienced a whole-genome duplication, complicating the identification of 1:1 orthologues[14]. Nevertheless, we observe that most orthologues of Z- and X-linked genes occupy separate portions of each fish genome (Supplementary Figs 8–11). For example, three-spine stickleback (*Gasterosteus aculeatus*) linkage groups 13 and 14 carry the bulk of Z-orthologous genes, whereas X-orthologous genes mostly reside on stickleback linkage groups 1, 4, 7 and 16 (Fig. 2c). Because we observe that Z-orthologous genes are separated from

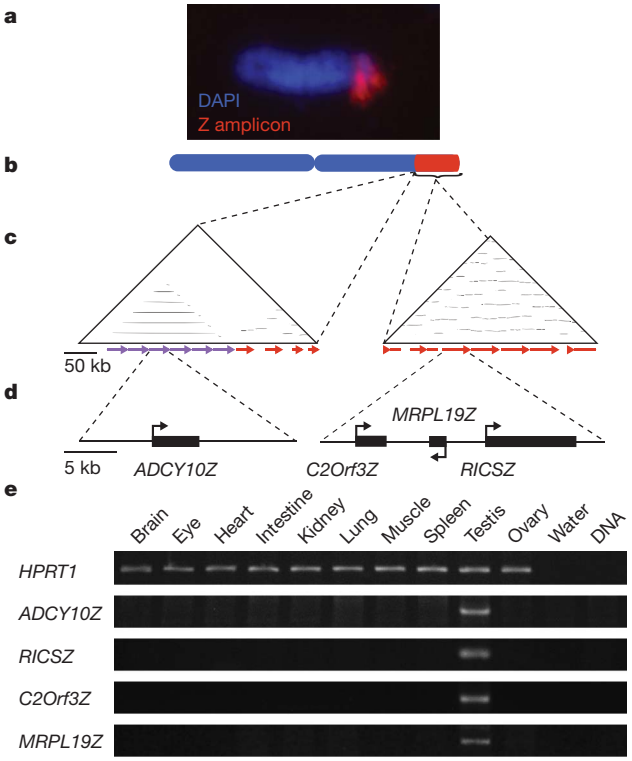

Figure 1 | The Z amplicon. **a**. Fluorescence *in situ* hybridization (FISH) of Z-amplicon bacterial artificial chromosome (BAC) CH261-77N6 (red) to the distal long arm of the Z chromosome (blue). DAPI, 4′,6-diamidino-2-phenylindole. **b**, The Z amplicon (red) constitutes the most distal 11 Mb of the Z chromosome. **c**, Triangular dot plots each comparing the sequence of a Z-chromosome BAC with itself. Within the plot, each dot represents a perfect match of 50 base pairs (bp). Direct repeats appear as horizontal lines. On the left, BAC CH261-73L15 contains six tandem repeats covering 120 kb immediately proximal to the Z amplicon. On the right, BAC CH261-137P21, a representative Z-amplicon clone. Each 25–30-kb repeat unit is ~95% similar to any other, though some units have been disrupted by insertions and deletions. **d**, Genes in repeat units of the Z amplicon. Each 20-kb repeat unit of small array in CH261-73L15 contains one copy of *ADCY10Z*. Each 25–30kb repeat unit of Z amplicon contains one copy each of *C2Orf3Z*, *MRPL19Z*, and *RICSZ*. **e**, Reverse transcriptase (RT)–PCR analysis of Z-amplicon gene expression in adult tissues. *HPRT1* is widely expressed in adult tissues and serves as positive control for reverse transcriptase reaction. All Z-amplicon genes are expressed in testis, but not other tissues.

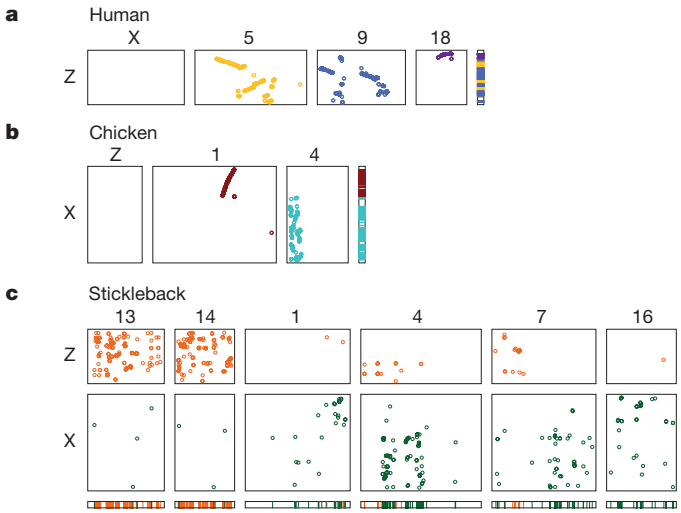

Figure 2 | Independent origin of chicken Z and human X chromosomes. Rectangular dot plots show chromosomal locations of Z-orthologous or X-orthologous genes in other species. **a**, Chicken Z chromosome versus selected human chromosomes. The chicken Z chromosome is not orthologous to the human X chromosome, but is orthologous to portions of human autosomes 5 (yellow), 9 (blue) and 18 (purple). At right: three-colour projection of dot plots onto a unified schematic of the chicken Z chromosome, showing that orthology to human chromosomes 5, 9 and 18 accounts for most of the Z chromosome, with the exception of the Z amplicon on the distal long arm. **b**, Human X chromosome versus selected chicken chromosomes. The human X chromosome is not orthologous to the chicken Z chromosome, but is orthologous to portions of chicken autosomes 1 (red) and 4 (cyan). At right: two-colour projection of dot plots onto a unified schematic of the human X chromosome, showing that orthology to chicken chromosomes 1 and 4 spans the X chromosome. **c**, Chicken Z chromosome (orange) and human X chromosome (green) versus selected stickleback chromosomes. Chicken Z and human X orthologues occupy separate and distinct locations within the stickleback genome. Chicken Z orthologues are present on stickleback chromosomes 13 and 14, whereas human X orthologues are present on stickleback chromosomes 1, 4, 7 and 16. At bottom: two-colour projection of dot plots onto unified schematics of stickleback chromosomes, showing the relative contribution of chicken Z and human X orthologues.

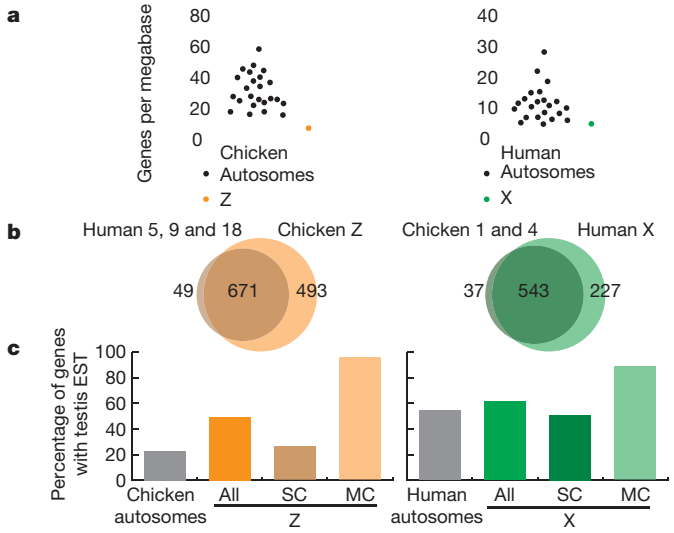

Figure 3 | Convergent gene gain on the chicken Z and human X chromosomes. **a**, Gene density of Z and X chromosomes compared with autosomes. Both are unusually gene poor, with about half the gene density of a typical autosome. **b**, Venn diagrams comparing gene content of chicken Z and human X chromosomes with orthologous autosomes. Most genes on orthologous autosomes remain on the sex chromosomes; few have been lost. Both the chicken Z chromosome and the human X chromosome gained hundreds of genes not present on orthologous autosomes. **c**, Percentage of protein coding genes with testis ESTs in Unigene. On left: in comparison with chicken autosomes, the Z chromosome is enriched for testis-expressed genes. Single-copy Z chromosome genes (SC) show no enrichment for testis ESTs relative to autosomal genes, but nearly all multicopy (MC) genes are expressed in testis. On right: similar results obtain on the human X chromosome.

X-orthologous genes in birds, mammals and each of these four fish, we conclude that the Z and the X chromosomes have evolved independently from distinct portions of the ancestral vertebrate genome.

Although we rejected the hypothesis that the avian Z and mammalian X chromosomes share a common origin, we discovered that the chicken Z and human X chromosomes share common features. Like the chicken Z chromosome, the human X chromosome has a low gene density; there are half as many genes per megabase on the X chromosome as on the average human autosome[5] (Fig. 3a and Table 1). Other investigators have observed that low gene density is often associated with increased interspersed repeat content, specifically LINEs[5,15]. We also observe this association on the Z and X chromosomes (Supplementary Fig. 1 and Table 1).

There are two possible explanations for these features of the chicken Z and human X chromosomes. Either the Z and X chromosomes arose from autosomes pre-adapted for the role of sex chromosomes, or they arose from ordinary autosomes that convergently evolved into specialized sex chromosomes. If the Z and X chromosomes arose from pre-adapted autosomes, then the structural features shared by the Z and X chromosomes should also be found on the orthologous autosomal regions. We tested this theory by comparing each sex chromosome to the orthologous autosomes in the other species (Fig. 2, Table 2, and Supplementary Tables 3 and 4). As a group, the autosomal regions that correspond to the Z and X chromosomes are typical of their respective genomes (Table 2). Although these regions show a slight deficit in gene density relative to the average within their respective genomes, the difference is too small to account for the extremely low gene density of the Z and X chromosomes. Because the orthologous autosomes in the other species do not share the structural features common to the Z and X chromosomes, we infer that these convergent features arose during the process of sex chromosome evolution, and not before.

To explain the paucity of genes on the Z and X chromosomes, we looked for evidence that both chromosomes lost genes during sex chromosome evolution. Instead, we observed that both the Z chromosome and the X chromosome gained protein-coding genes. We compared the gene content of the Z and the X chromosomes to the orthologous autosomes from the other species as a surrogate for the ancestral gene content of the Z and X chromosomes (Figs 2 and 3b, Table 2, and Supplementary Tables 3 and 4). We found that only a few dozen genes present on the orthologous autosomes are absent from the Z and X chromosomes (Fig. 3b). In contrast, hundreds of genes present on the Z and X chromosomes are absent from the orthologous autosomes (Fig. 3b). We conclude that both the Z chromosome and the X chromosome experienced substantial net gene gain.

The majority of genes gained by the Z and X chromosomes are members of multicopy families (Fig. 3b and Supplementary Tables 3 and 4). On the chicken Z chromosome, these are the genes of the Z amplicon. The human X chromosome has gained thirteen different cancer/testis antigen gene families[5]. All of the Z-amplicon genes are expressed predominantly in testis (Fig. 1e), as are the cancer/testis antigen genes of the human X chromosome[16]. The addition of these multicopy gene families has biased the Z and X chromosomes towards testis-expressed genes (Fig. 3c). Both the Z chromosome and the X chromosome have an elevated proportion of genes expressed in testis tissue in comparison with autosomes as measured by the number of genes with a testis expressed sequence tag (EST) in Unigene[17] data sets (Fig. 3c). However, when multicopy genes are removed, the remaining conserved single-copy genes show no bias (Fig. 3c). Others have observed a bias towards sex- and reproduction-related genes on the human X chromosome[18]. Our comparison suggests that the Z chromosome shares this bias. This bias was not a feature of the autosomes that gave rise to the sex chromosomes of birds and mammals; it arose by gene acquisition and amplification during sex chromosome evolution in each lineage.

In light of this convergent gene gain, we looked for factors other than gene loss that could account for the low gene density of the Z and X chromosomes. Low gene density could result from Z- and X-linked genes that are larger than those on autosomes, resulting in fewer genes in the same amount of sequence. However, we find that genes on both the Z chromosome and the X chromosomes are smaller, on average, than autosomal genes (Table 1). The only remaining explanation for the unusually low gene density of the Z and X chromosomes is a massive expansion of non-coding intergenic sequences that spread the genes farther apart. We estimate that intergenic regions were expanded by about 40 Mb in the case of the Z chromosome and 80 Mb in the case of the X chromosome—nearly half the present lengths of these chromosomes. No single class of non-coding sequence can account for this change, but the enrichment for LINEs on both the Z chromosome and the X chromosome (Table 1) suggests that the doubling of intergenic sequence may have been driven by recurrent insertion and divergence of transposable elements. In mammalian genomes, high LINE density is associated with reduced rates of crossing over[19], and suppression of crossing over is a key step in the evolution of differentiated sex chromosomes. However, the Z and X

Table 2 | Comparison of structural features of Z-orthologous and X-orthologous autosomal regions with all autosomes

	Z-orthologous regions of human chromosomes 5, 9 and 18	Human autosomes	X-orthologous regions of chicken chromosomes 1 and 4	Chicken autosomes
Genes per megabase	10	12	23	25
Interspersed repeats	48%	45%	8.9%	9.4%
LINEs	23%	21%	6.0%	6.4%

chromosomes are enriched for LINEs in comparison with autosomal regions with similarly low rates of crossing over (Supplementary Fig. 12 and Supplementary Note 1).

Our comparison of the finished sequences of the chicken Z and human X chromosomes reveals that each evolved independently from different portions of the ancestral genome, from separate pairs of ordinary autosomes. In each lineage, different portions of the ancestral genome were substantially remodelled to become specialized sex chromosomes. The Z and X chromosomes have converged on a set of structural features that distinguish them from autosomes: a high density of interspersed repeats, and long intergenic distances resulting in low gene density. Furthermore, the Z and X chromosomes have both gained multicopy gene families that are expressed in testis, biasing the gene content of both chromosomes towards male reproductive functions.

This convergent specialization of Z and X chromosomes for male reproduction is surprising given that the Z chromosome evolved with female heterogamety and the X chromosome evolved in opposite circumstances, with male heterogamety. One might have anticipated that the Z and X chromosome would have opposing rather than convergent biases in gene content. Although strong selective pressures drive the evolution of genes related to male reproduction[20,21], these selective pressures influence autosomes as well as sex chromosomes. However, unlike autosomes, sex chromosomes are uniquely susceptible to selection for traits that benefit one sex more than the other[22]. Our results suggest that, in amniotes, selective pressures to preserve or enhance male reproductive functions have trumped the differences between ZW and XY systems to produce the changes in gene content that we observe.

For nearly 100 years, it has been thought that sex chromosome evolution involved drastic modification of sex-specific chromosomes but only modest change in chromosomes shared by the sexes[1,2]. In the past decade, this understanding was reinforced by comprehensive molecular comparisons between the human X and Y chromosomes[3–5] and by more limited comparison of sex chromosome pairs in other plants and animals[23–25]. These X–Y or Z–W comparisons revealed extensive genetic decay in the sex-specific Y or W chromosome, while assuming that Z and X chromosomes faithfully represent their autosomal progenitors. By contrast, the Z–autosome and X–autosome comparisons in this study reveal that the chicken Z chromosome and the human X chromosome have undergone dramatic evolutionary changes that were not anticipated and that previous studies could not detect. In birds and mammals, sex chromosome evolution was not limited to gene loss from sex-specific chromosomes, but extended to expansion and gene acquisition on the chromosomes shared between the sexes.

METHODS SUMMARY

Mapping and sequencing. All Z-chromosome BAC and fosmid clones that we selected for sequencing (see Supplementary Table 5 for GenBank accession numbers) were from six libraries generated from the same female of the inbred line of red jungle fowl (UCD001) as was used for the whole-genome shotgun sequence of the chicken[8,26]. As a result, the sequence we obtained is that of a single Z-chromosome haplotype. We made use of available BAC fingerprint maps to select tiling paths across the Z chromosome. Contigs were ordered and oriented by radiation hybrid mapping[27] and confirmed by FISH[28] (Supplementary Figs 2 and 3).

Sequence analysis. We used REPEATMASKER (http://www.repeatmasker.org) to identify and mask interspersed repeats. We used BLAT[29] to detect intrachromosomal similarity and custom Perl scripts to construct triangular dot plots (http://jura.wi.mit.edu/page/Y/azfc/self_dot_plot.pl).

Comparative genomics. We detected orthology by using BLAT to align peptide sequences (Ensembl version 52)[30] and identifying the best reciprocal hit between species. We constructed the interspecies dot plots using the chromosomal coordinates extracted from Ensembl (or, in the case of the Z chromosome, the coordinates from this study). For counts of gene gain and loss, we used the list of orthologues of human and chicken genes compiled by Ensembl, which we then manually reviewed to ensure accuracy. In cases where genes on the sex chromosomes did not have a 1:1 orthologue on a corresponding autosome in the other species (or vice versa), we used outgroup species (fish and amphibians) to determine the lineage (chicken or human) on which a gene was gained or lost.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 February; accepted 13 May 2010.
Published online 11 July 2010.

1. Muller, H. J. A gene for the fourth chromosome of *Drosophila*. *J. Exp. Zool.* **17,** 325–336 (1914).
2. Ohno, S. *Sex Chromosomes and Sex-Linked Genes* (Springer, 1967).
3. Lahn, B. T. & Page, D. C. Four evolutionary strata on the human X chromosome. *Science* **286,** 964–967 (1999).
4. 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).
5. Ross, M. T. *et al.* The DNA sequence of the human X chromosome. *Nature* **434,** 325–337 (2005).
6. Nanda, I. *et al.* 300 million years of conserved synteny between chicken Z and human chromosome 9. *Nature Genet.* **21,** 258–259 (1999).
7. Nanda, I., Haaf, T., Schartl, M., Schmid, M. & Burt, D. W. Comparative mapping of Z-orthologous genes in vertebrates: implications for the evolution of avian sex chromosomes. *Cytogenet. Genome Res.* **99,** 178–184 (2002).
8. International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432,** 695–716 (2004).
9. Grützner, F. *et al.* In the platypus a meiotic chain of ten sex chromosomes shares genes with the bird Z and mammal X chromosomes. *Nature* **432,** 913–917 (2004).
10. Rens, W. *et al.* Resolution and evolution of the duck-billed platypus karyotype with an X1Y1X2Y2X3Y3X4Y4X5Y5 male sex chromosome constitution. *Proc. Natl Acad. Sci. USA* **101,** 16257–16261 (2004).
11. Ezaz, T., Stiglec, R., Veyrunes, F. & Marshall Graves, J. A. Relationships between vertebrate ZW and XY sex chromosome systems. *Curr. Biol.* **16,** R736–R743 (2006).
12. Smith, J. J. & Voss, S. R. Bird and mammal sex-chromosome orthologs map to the same autosomal region in a salamander (*Ambystoma*). *Genetics* **177,** 607–613 (2007).
13. Hori, T. *et al.* Characterization of DNA sequences constituting the terminal heterochromatin of the chicken Z chromosome. *Chromosome Res.* **4,** 411–426 (1996).
14. Kasahara, M. *et al.* The medaka draft genome and insights into vertebrate genome evolution. *Nature* **447,** 714–719 (2007).
15. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).
16. Scanlan, M. J., Simpson, A. J. & Old, L. J. The cancer/testis genes: review, standardization, and commentary. *Cancer Immun.* **4,** 1 (2004).
17. Wheeler, D. L. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **33,** D39–D45 (2005).
18. Saifl, G. M. & Chandra, H. S. An apparent excess of sex- and reproduction-related genes on the human X chromosome. *Proc. R. Soc. Lond. B* **266,** 203–209 (1999).
19. Wichman, H. A., Bussche, R. A., Hamilton, M. J. & Baker, R. J. Transposable elements and the evolution of genome organization in mammals. *Genetica* **86,** 287–293 (1992).
20. Wyckoff, G. J., Wang, W. & Wu, C. I. Rapid evolution of male reproductive genes in the descent of man. *Nature* **403,** 304–309 (2000).
21. Swanson, W. J. & Vacquier, V. D. The rapid evolution of reproductive proteins. *Nature Rev. Genet.* **3,** 137–144 (2002).
22. Rice, W. R. Sex chromosomes and the evolution of sexual dimorphism. *Evolution* **38,** 735–742 (1984).
23. Filatov, D. A. Evolutionary history of Silene latifolia sex chromosomes revealed by genetic mapping of four genes. *Genetics* **170,** 975 (2005).
24. Handley, L. J. L., Ceplitis, H. & Ellegren, H. Evolutionary strata on the chicken Z chromosome: implications for sex chromosome evolution. *Genetics* **167,** 367 (2004).
25. Nicolas, M. *et al.* A gradual process of recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. *PLoS Biol.* **3,** e4 (2005).
26. Wallis, J. W. *et al.* A physical map of the chicken genome. *Nature* **432,** 761–764 (2004).
27. Morisson, M. *et al.* ChickRH6: a chicken whole-genome radiation hybrid panel. *Genet. Sel. Evol.* **34,** 521–533 (2002).
28. Saxena, R. *et al.* Four DAZ genes in two clusters found in the AZFc region of the human Y chromosome. *Genomics* **67,** 256–267 (2000).
29. Kent, W. J. BLAT–the BLAST-like alignment tool. *Genome Res.* **12,** 656–664 (2002).
30. Hubbard, T. J. *et al.* Ensembl 2009. *Nucleic Acids Res.* **37,** D690–D697 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Rapoport for technical assistance, S. Repping and S. van Daalen for experimental advice, and E. Anderson, T. Endo, M. Gill, A. Hochwagen, C. Hongay, Y. Hu, J. Hughes, J. Marszalek, J. Mueller and Y. Soh for comments on the manuscript. We thank the Broad Institute Genome Sequencing Platform and Genome Sequencing and Analysis Program, F. Di Palma and K. Lindblad-Toh for making the unpublished data for *Gasterosteus aculeatus*

available. This work was supported by the National Institutes of Health and the Howard Hughes Medical Institute.

Author Contributions D.W.B., H.S., W.C.W., S.R., R.K.W. and D.C.P. planned the project. D.W.B. and L.G.B. performed BAC mapping. D.W.B. performed RT–PCR analysis. T.G. and C.K. were responsible for finished BAC sequencing. D.W.B. and H.S. performed comparative sequence analyses. T.P. performed FISH analysis. E.R.M. performed 454 sequencing. D.W.B. and D.C.P. wrote the paper.

Author Information Predicted Z-amplicon transcript sequences and the complete assembled sequence of the Z chromosome are available at http://jura.wi.mit.edu/page/papers/Bellott_et_al_2010/ (see Supplementary Table 5 for GenBank accession numbers). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.C.P. (dcpage@wi.mit.edu).

METHODS

Mapping and sequencing. All Z-chromosome BAC and fosmid clones selected for sequencing (Supplementary Table 5) were from six libraries (CH261, TAM31, TAM32, TAM33, J_AD and J_AE) generated from the same female of the inbred line of red jungle fowl (UCD001) as was used for the whole-genome shotgun sequence of the chicken[8,26]. As a result, the Z chromosome we obtained is that of a single haplotype. We made use of publicly available BAC fingerprint maps and BAC-end sequences as a source of mapping information and markers. Individual BAC fingerprint contigs were ordered and oriented by radiation hybrid mapping using CHICKRH6[27]. No cell line is available from any bird of the UCD001 line, so we used chicken embryonic fibroblasts derived from White Leghorn (available from Charles River Labs) for FISH experiments to provide independent confirmation of the order and orientation of the sequence (Supplementary Figs 9–11).

Chromosomal FISH. One- or two-colour FISH to chicken chromosomes was performed as previously described[28].

Z-chromosome sequence similarity. Analyses of intrachromosomal similarity were performed using BLAT[29] (version 34) to compare all 5-kb sequence segments, in 1-kb steps, to the entire remainder of the Z-chromosome sequence. For each segment, we recorded the highest percentage identity to a non-overlapping segment.

Genes and transcription units. We identified potential transcripts in three ways.

(1) We used human (Ensembl version 52, NCBI 36)[30] as the informant genome and chicken EST sequences as additional evidence to identify potential transcripts on the repeat-masked chicken Z chromosome, using TWINSCAN[31,32] (version 3.5). We compared the output with the Ensembl 52 annotations for chicken and human to identify previously unrecognized genes in our prediction. We considered previously unrecognized chicken genes valid if they were spliced in chicken and conserved to human.

(2) For the novel genes in the Z-amplicon region, we relied on BLAST[33] (NCBI version 2.2.19) matches to complementary DNA sequences to identify copies of *ADCY10Z*, *C2Orf3Z*, *MRPL19Z* and *RICSZ* that showed evidence of splicing. We then tested for transcription by RT–PCR across a panel of adult tissues.

(3) We used a combination of methods to locate non-coding transcripts on the chicken Z chromosome. We used TRNASCAN-SE[34] (version 1.23) for transfer RNA predictions. For other non-coding RNAs, we used BLAST to compare our sequence with those of known chicken non-coding RNAs in GenBank[35] and miRBase[36]

Interspersed repeats. We electronically identified interspersed repeats with REPEATMASKER[37] (version 3.2.7).

Triangular dot plots. We performed dot plot analysis using a custom Perl script[38].

Expressed sequence tags. We used EST sequences from the BBSRC ChickEST database[39], supplemented by our own 454 EST runs on ovary and testis (SRA# SRP000097).

Z-amplicon size. To estimate the amount of Z-ampliconic sequence missing from our assembly, we compared the average depth of chicken fosmid end sequences in the single-copy region of the Z chromosome with the depth in the ampliconic region, reasoning that the excess depth in the ampliconic region could be attributed to sequence we could not obtain because the similarity between individual repeat units precluded either cloning or the assembly of BACs. We used BLAT to map 23,977 fosmid end sequences to 72.2 Mb of single-copy Z sequence, giving an average of 331 ends per megabase. In the 5.6 Mb of the Z amplicon, we found 3,787 fosmid end hits, for 666 ends per megabase, roughly a twofold enrichment. Therefore, we concluded that the Z amplicon comprises roughly 11.4 Mb.

RT–PCR. We used chicken total RNA (Zyagen) and the RETROscript Kit (Ambion). We amplified 1 μl of the RT product through 30 cycles of PCR with an annealing temperature of 55 °C.

Primers are as follows. *HPRT1* (116-bp product): GGATTTGAAGTGCCAGA CAAA (forward); GCTTTGTACTTCTGCTTCCCC (reverse). *ADCY10Z* (145-bp product): GTTTGTCAGGTCTCTGTGGGA (forward); GTAGAGGTCCT CGAGCAAGGC (reverse). *RICSZ* (144-bp product): GACAGAGATCAGGGA CATGGA (forward); AAACAGGAACACCAACTGCAT (reverse). *C2Orf3Z* (131-bp product): TGTTCAAAATTCCAAGGCAGA (forward); AGGTAACGA TTCAGCAGCTTG (reverse). *MRPL19Z* (242-bp and 60-bp products): CAAGCAGAAGCAGAGAGAGGA (forward); TGACCATGGTTGAGGTTTCA (reverse).

Orthologous chromosomes. To identify orthologous chromosomes in inter-species comparisons, we relied on a gene-based approach. We conducted reciprocal BLAT searches using Ensembl 52 peptide sequence databases from *Gallus gallus*, *Homo sapiens*, *Danio rerio*, *Gasterosteus aculeatus*, *Oryzias latipes* and *Tetraodon nigroviridis*. Considering only the longest peptide sequences for each Ensembl gene, we flagged best reciprocal BLAT hits between two genomes as orthologous genes. We constructed the interspecies dot plots using the chromosomal coordinates extracted from Ensembl (or in the case of the Z chromosome, the coordinates from this study). Each individual dot represents a pair of orthologous genes.

Gene gain and loss. We relied on the assignments of chicken and human orthologues in Ensembl 52. However, we manually reviewed genes on chicken chromosomes 1, 4 and Z as well as human chromosomes 5, 9, 18 and X that did not have simple 1:1 orthologues in the Ensembl database, to find pairs of orthologous genes that were missing from the database or not properly identified.

To study gene gain and loss on the chicken Z chromosome, we divided genes into the following categories on the basis of their locations in chickens, humans and outgroup species: (A) Z-linked genes with orthologues on human autosomes 5, 9 and 18; (B i) Z-linked genes present only in birds, but not in outgroups or human; (B ii) Z-linked genes present in birds and outgroups but not in human; (B iii a) Z-linked genes with human orthologues not on autosomes 5, 9 or 18 that were not syntenic with neighbours in outgroups or human; (B iii b) Z-linked genes with human orthologues not on autosomes 5, 9 or 18 that were syntenic with neighbours in outgroups; (C i) genes on human autosomes 5, 9 and 18 with orthologues only in mammals, but not in outgroups or chicken; (C ii) genes on human autosomes 5, 9 and 18 with orthologues only in mammals and outgroups, but not in chicken; (C iii a) genes on human autosomes 5, 9 and 18 with chicken orthologues not on the Z chromosome that were not syntenic with neighbours in outgroups or chicken; (C iii b) genes on human autosomes 5, 9 and 18 with chicken orthologues not on the Z chromosome that were syntenic with neighbours in outgroups.

Category (A) was counted as shared, categories (B i) and (B iii a) were counted as gains to the Z chromosome, and categories (C ii) and (C iii b) were counted as losses from the Z chromosome. Categories (B ii) and (B iii b) (representing losses from the human autosomes) and categories (C i) and (C iii a) (representing gains to the human autosomes) were excluded. We carried out an analogous analysis on the human X chromosome and chicken autosomes 1 and 4.

Biased gene content. We searched for Unigene[17] EST clusters from normal chicken and human testis that corresponded with chicken and human genes to identify genes expressed in the testis. We determined the percentage of genes with at least one testis EST for each category (autosomes, Z- or X-chromosome total, Z- or X-chromosome single copy, Z- or X-chromosome multicopy). Multicopy genes on the Z chromosome include those of the Z amplicon, and multicopy genes on the X chromosome include those identified as cancer/testis antigen genes in the finished sequence of the X chromosome[5].

31. Korf, I., Flicek, P., Duan, D. & Brent, M. R. Integrating genomic homology into gene structure prediction. *Bioinformatics* **17** (suppl. 1), S140–S148 (2001).
32. 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).
33. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
34. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25,** 955–964 (1997).
35. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. GenBank. *Nucleic Acids Res.* **37,** D26–D31 (2009).
36. Griffiths-Jones, S., Saini, H. K., van Dongen, S. & Enright, A. J. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* **36,** D154–D158 (2008).
37. Smit, A. F. A., Hubley, R. & Green, P. RepeatMasker Open-3.0. ⟨http://www.repeatmasker.org⟩ (2007).
38. Kuroda-Kawaguchi, T. *et al.* The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nature Genet.* **29,** 279–286 (2001).
39. Boardman, P. E. *et al.* A comprehensive collection of chicken cDNAs. *Curr. Biol.* **12,** 1965–1969 (2002).