LETTERS

The genome of a songbird

Wesley C. Warren[1], David F. Clayton[2], Hans Ellegren[3], Arthur P. Arnold[4], LaDeana W. Hillier[1], Axel Künstner[3], Steve Searle[5], Simon White[5], Albert J. Vilella[6], Susan Fairley[5], Andreas Heger[7], Lesheng Kong[7], Chris P. Ponting[7], Erich D. Jarvis[8], Claudio V. Mello[9], Pat Minx[1], Peter Lovell[9], Tarciso A. F. Velho[9], Margaret Ferris[2], Christopher N. Balakrishnan[2], Saurabh Sinha[2], Charles Blatti[2], Sarah E. London[2], Yun Li[2], Ya-Chi Lin[2], Julia George[2], Jonathan Sweedler[2], Bruce Southey[2], Preethi Gunaratne[10], Michael Watson[11], Kiwoong Nam[3], Niclas Backström[3], Linnea Smeds[3], Benoit Nabholz[3], Yuichiro Itoh[4], Osceola Whitney[8], Andreas R. Pfenning[8], Jason Howard[8], Martin Völker[11], Bejamin M. Skinner[12], Darren K. Griffin[12], Liang Ye[1], William M. McLaren[6], Paul Flicek[6], Victor Quesada[13], Gloria Velasco[13], Carlos Lopez-Otin[13], Xose S. Puente[13], Tsviya Olender[14], Doron Lancet[14], Arian F. A. Smit[15], Robert Hubley[15], Miriam K. Konkel[16], Jerilyn A. Walker[16], Mark A. Batzer[16], Wanjun Gu[17], David D. Pollock[17], Lin Chen[18], Ze Cheng[18], Evan E. Eichler[18], Jessica Stapley[18], Jon Slate[19], Robert Ekblom[19], Tim Birkhead[19], Terry Burke[19], David Burt[20], Constance Scharff[21], Iris Adam[21], Hugues Richard[22], Marc Sultan[22], Alexey Soldatov[22], Hans Lehrach[22], Scott V. Edwards[23], Shiaw-Pyng Yang[24], XiaoChing Li[25], Tina Graves[1], Lucinda Fulton[1], Joanne Nelson[1], Asif Chinwalla[1], Shunfeng Hou[1], Elaine R. Mardis[1] & Richard K. Wilson[1]

The zebra finch is an important model organism in several fields[1,2] with unique relevance to human neuroscience[3,4]. Like other songbirds, the zebra finch communicates through learned vocalizations, an ability otherwise documented only in humans and a few other animals and lacking in the chicken[5]—the only bird with a sequenced genome until now[6]. Here we present a structural, functional and comparative analysis of the genome sequence of the zebra finch (*Taeniopygia guttata*), which is a songbird belonging to the large avian order Passeriformes[7]. We find that the overall structures of the genomes are similar in zebra finch and chicken, but they differ in many intrachromosomal rearrangements, lineage-specific gene family expansions, the number of long-terminal-repeat-based retrotransposons, and mechanisms of sex chromosome dosage compensation. We show that song behaviour engages gene regulatory networks in the zebra finch brain, altering the expression of long non-coding RNAs, microRNAs, transcription factors and their targets. We also show evidence for rapid molecular evolution in the songbird lineage of genes that are regulated during song experience. These results indicate an active involvement of the genome in neural processes underlying vocal communication and identify potential genetic substrates for the evolution and regulation of this behaviour.

As in all songbirds, singing in the zebra finch is under the control of a discrete neural circuit that includes several dedicated centres in the forebrain termed the 'song control nuclei' (for an extensive series of reviews see ref. 8). Neurophysiological studies in these nuclei during singing have yielded some of the most illuminating examples of how vocalizations are encoded in the motor system of a vertebrate brain[9,10]. In the zebra finch, these nuclei develop more fully in the male than in the female (who does not sing), and they change markedly in size and organization during the juvenile period when the male learns to sing[11]. Analysis of the underlying cellular mechanisms of plasticity led to the unexpected discovery of neurogenesis in adult songbirds and life-long replacement of neurons[12]. Sex steroid hormones also contribute to songbird neural plasticity, in part by influencing the survival of new neurons[13]. Some of these effects are probably caused by oestrogen and/or testosterone synthesized within the brain itself rather than just in the gonads[14].

Song perception and memory also involve auditory centres that are present in both sexes, and the mere experience of hearing a song activates gene expression in these auditory centres[15]. The gene response itself changes as a song becomes familiar over the course of a day[16] or as the context of the experience changes[17]. The act of singing induces gene expression in the male song control nuclei, and these patterns of gene activation also vary with the context of the experience[18]. The function of this changing genomic activity is not yet understood, but it may support or suppress learning and help integrate information over periods of hours to days[19].

The chicken genome is the only other bird genome analysed to date[6]. The chicken and zebra finch lineages diverged about 100 million years ago near the base of the avian radiation[7]. By comparing their genomes we can now discern features that are shared (and thus

[1]The Genome Center, Washington University School of Medicine, Campus Box 8501, 4444 Forest Park Avenue, St Louis, Missouri 63108, USA. [2]University of Illinois, Urbana-Champaign, Illinois 61801 USA. [3]Uppsala University, Institute for Evolution and Genetics Systems, Norbyvägen 18D 752 36 Uppsala, Sweden. [4]University of California- Los Angeles, Los Angeles, California 90056, USA. [5]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. [6]EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. [7]MRC Functional Genomics Unit, University of Oxford, Department of Physiology, Anatomy and Genetics, South Parks Road, Oxford OX1 3QX, UK. [8]Howard Hughes Medical Institute, Department of Neurobiology, Box 3209, Duke University Medical Center, Durham, North Carolina 27710, USA. [9]Department of Behavioral Neuroscience, Oregon Health & Science University, Portland, Oregon 97239, USA. [10]Department of Biology & Biochemistry, University of Houston, Houston, Texas 77204, USA. [11]Department of Bioinformatics, Institute for Animal Health, Compton Berks RG20 7NN, UK. [12]Department of Biosciences, University of Kent, Canterbury, Kent CT2 7NJ, UK. [13]Instituto Universitario de Oncologia, Departamento de Bioquimica y Biologia Molecular, Universidad de Oviedo, 33006-Oviedo, Spain. [14]Crown Human Genome Center, Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel. [15]Institute for Systems Biology, 1441 North 34th Street, Seattle, Washington 98103-8904, USA. [16]Department of Biological Sciences, Louisiana State University, 202 Life Sciences Building, Baton Rouge, Louisiana 70803, USA. [17]Department of Biochemistry & Molecular Genetics, University of Colorado Health Sciences Center, Mail Stop 8101, Aurora, Colorado 80045, USA. [18]University of Washington, Genome Sciences, Seattle, Washington 98195, USA. [19]Department of Animal & Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK. [20]The Roslin Institute and Royal (Dick) School of Veterinary Studies, Edinburgh University, EH25 9OS, UK. [21]Freie Universitaet Berlin, Institut Biology, Takustr.6, 14195 Berlin, Germany. [22]Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, IhnestraBe 73 14195 Berlin, Germany. [23]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [24]Monsanto Company, 800 North Lindbergh Boulevard, St Louis, Missouri 63167, USA. [25]Neuroscience Center, Louisiana State University Health Sciences Center, New Orleans, Louisiana 70112, USA.

generally characteristic of birds), and features that are most conspicuously different between the two lineages—some of which will be related to the distinctive neural and behavioural traits of songbirds.

We sequenced and assembled a male zebra finch genome using methods described previously[6,20]. A male (the homogametic sex in birds) was chosen to maximize coverage of the Z chromosome. Of the 1.2 gigabase (Gb) draft assembly, 1.0 Gb has been assigned to 33 chromosomes and three linkage groups, by using zebra finch genetic linkage[21] and bacterial artificial chromosome (BAC) fingerprint maps. The genome assembly is of sufficient quality for the analysis presented here (see Supplementary Note 1 and Supplementary Table 1). A total of 17,475 protein-coding genes were predicted from the zebra finch genome assembly using the Ensembl pipeline supplemented by Gpipe gene models (Supplementary Note 1). To extend further the characterization of genes relevant to brain and behaviour, we also sequenced complementary DNAs from the forebrain of zebra finches at 50 (juvenile, during the critical song learning period) and 850 (adult) days post-hatch, mapping these reads (Illumina GA2) to the protein-coding models (Supplementary Note 1). Of the 17,475 protein-coding gene models we find 9,872 (56%) and 10,106 (57%) genes expressed in the forebrain at these two ages (90.7% overlap), respectively. In addition to evidence for developmental regulation, these reads show further splice forms, new exons and untranslated sequences (Supplementary Figs 1 and 2).

To address issues of large-scale genome structure and evolution, we compared the chromosomes of zebra finch and chicken using both sequence alignment and fluorescent *in situ* hybridization. These analyses showed overall conservation of synteny and karyotype in the two species, although the rate of intrachromosomal rearrangement was high (Supplementary Note 2). We were also surprised to see genes of the major histocompatibility complex (MHC) dispersed across several chromosomes in the zebra finch, in contrast to the syntenic organization of both chicken and human MHCs (Supplementary Note 2).

We assessed specific gene losses and expansions in the zebra finch lineage by constructing phylogenies of genes present in the last common ancestor of birds and mammals (Supplementary Note 2 and Supplementary Fig. 3). Both the zebra finch and the chicken genome assemblies lack genes encoding vomeronasal receptors, casein milk proteins, salivary-associated proteins and enamel proteins—not surprisingly, as birds lack vomeronasal organs, mammary glands and teeth. Unexpectedly, both species lack the gene for the neuronal protein synapsin 1 (*SYN1*); comparative analyses suggest that the loss of *SYN1* and flanking genes probably occurred in an ancestor to modern birds, possibly within the dinosaur lineage (Supplementary Note 2, Supplementary Table 2 and Supplementary Fig. 4). Both zebra finch and chicken have extensive repertoires of olfactory receptor-like sequences (Supplementary Note 2 and Supplementary Fig. 5), proteases (Supplementary Table 3), and a rich repertoire of neuropeptide and pro-hormone genes.

Compared to mammals, zebra finch has duplications of genes encoding several proteins with known neural functions, including growth hormone, (Supplementary Fig. 3), caspase-3 and β-secretase (Supplementary Table 3). Two large expansions of gene families expressed in the brain seem to have occurred in the zebra finch lineage after the split from mammals. One involves a family related to the *PAK3* (p21-activated kinase) gene. Thirty-one uninterrupted *PAK3*-like sequences have been identified in the zebra finch genome, of which 29 are expressed in testis and/or brain (Supplementary Note 2). The second involves the *PHF7* gene, which encodes a zinc-finger-containing transcriptional control protein. Humans only have a single *PHF7* gene, but remarkably the gene has been duplicated independently, many times in both the zebra finch and chicken lineages to form species-specific clades of 17 and 18 genes, respectively (Supplementary Fig. 6). In the zebra finch these genes are expressed in the brain (Supplementary Note 2).

An intriguing puzzle in avian genomics has been the evident lack of a chromosome-wide dosage compensation mechanism to balance the expression of genes on the Z sex chromosome, which is present in two copies in males but only one in females[22,23]. The chicken has been suspected of exerting dosage compensation on a more local level, by the non-coding RNA MHM (male hypermethylated)[24,25], to cause a characteristic variation of gene expression along the Z chromosome. The zebra finch genome assembly, however, lacks an MHM sequence, and genes adjacent to the comparable MHM chromosomal position show no special cluster of dosage compensation (Fig. 1 and Supplementary Note 2). Thus, the putative MHM-mediated mechanism of restricted Z-chromosome dosage compensation is not common to all birds. Chromosomal sex differences in the brain could have a direct role in the sex differences so evident in zebra finch neuroanatomy and singing behaviour.

In mammals, as much as half of their genomes represent interspersed repeats derived from mobile elements, whereas the interspersed repeat content of the chicken genome is only 8.5%. We find that the zebra finch genome also has a low overall interspersed repeat content (7.7%), containing a little over 200,000 mobile elements (Supplementary Tables 4 and 5). The zebra finch, however, has about three times as many retrovirus-derived long terminal repeat (LTR) element copies as the chicken, and a low copy number of short interspersed elements (SINEs), which the chicken lacks altogether. Expressed sequence tag (EST) analysis shows that mobile elements are present in about 4% of the transcripts expressed in the zebra finch brain, and some of these transcripts are regulated by song exposure (next section, Table 1). Figure 2 shows an example of an RNA that was identified in a microarray screening for genes specifically enriched in song control nuclei[26] and now seems to represent a long non-coding RNA (ncRNA) containing a CR1-like mobile element. These results indicate that further experiments investigating a possible role of mobile-element-derived repeated sequences in vocal communication are warranted.

A large portion of the genome is directly engaged by vocal communication. A recent study[27] defined distinct sets of RNAs in the

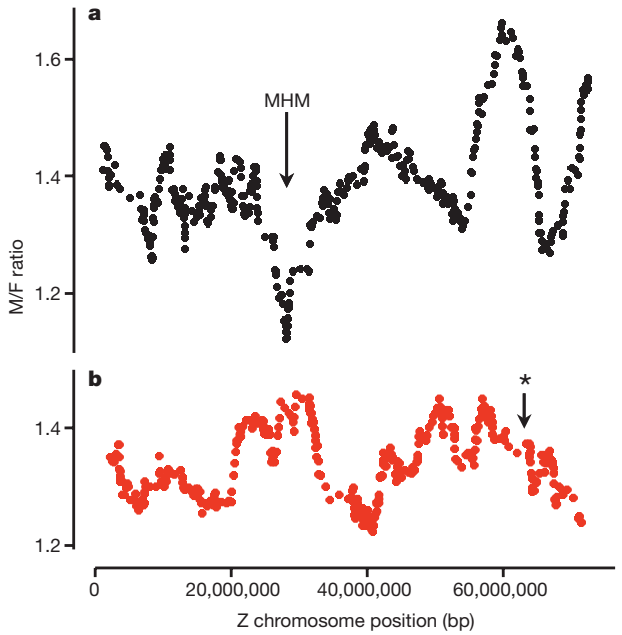

Figure 1 | Divergent patterns of dosage compensation in birds. a, b, The male to female (M/F) ratio of gene expression, measured by species-specific microarrays, is plotted along the Z chromosome of chicken (**a**) and zebra finch (**b**). Each point represents the average M/F ratio of a sliding window of 30 genes plotted at the median gene position and stepping one gene at a time along the chromosome. Note region of lower M/F ratios in chicken surrounding the locus of the MHM (male hypermethylated) ncRNA. In zebra finch, genes adjacent to the comparable MHM position (asterisk) show no special cluster of dosage compensation (low M/F ratios), and no MHM sequence appears in the genome assembly. bp, base pairs.

Table 1 | Structural features of the song responsive genome

	All genes analysed	Novel up	Novel down	Habituate up	Habituate down
All ESTs	17,877	145	461	1,531	1,774
Mapped loci	15,009	125	435	1,217	1,112
Ensembl genes	8,438	136	301	1,138	1,136
Mobile element content*					
Number with mobile elements	688	2	40	32	38
Percentage mobile elements	4	1	9	2	2
P-value	–	0.18	1.4×10^{-5}	0.005	0.004
Coding and non-coding content†					
mRNA transcripts (% (*P*-value))	59	86 (0.05)	32 (1×10^{-10})	65 (0.05)	71 (0.001)
EST loci mapped to introns (% (*P*-value))	6	1 (0.05)	21 (1×10^{-10})	3 (0.001)	6
Intergenic loci (% (*P*-value))	33	12 (0.001)	45 (0.05)	31	21 (0.001)
Protein-coding gene territories‡					
Mean gene length (kb)	30.4	21.7	78.8	34.8	31.2
Intergenic length (kb)	57.4	42.3	108.0	64.9	55.3
Territory size (kb)	87.8	64.1	186.8	99.7	86.4
P-value	–	3.9×10^{-3}	1.7×10^{-28}	9.3×10^{-10}	1.4×10^{-4}

A microarray made from non-redundant brain-derived ESTs[34] was used to define four subgroups of RNAs that show different responses in auditory forebrain to song exposures (novel up and down, habituated up and down)[27]. These ESTs were mapped to genome positions as described (Supplementary Note 3).
* All ESTs were analysed for mobile element content using RepeatMasker (Supplementary Note 2). *P*-value is for the comparison to all genes (Fisher's exact test).
† All ESTs that could be mapped uniquely to the genome assembly were assessed for overlap with Ensembl annotations of mRNA transcripts (protein coding and UTRs), intronic regions, or intergenic regions. *P*-value is for comparison to all mapped loci (Fisher's exact test). Results are the percentage with *P* values in parentheses where shown.
‡ The size of each unique protein-coding gene territory was determined by combining the length of the Ensembl gene model with its intergenic spacing. The *P*-value is for the comparison to all genes, using a two-tailed Wilcoxon rank sum test.

auditory forebrain that respond in different ways to song playbacks during the process of song-specific habituation, a form of learning[16]. We now map each of these song-responsive RNAs to the genome assembly (Table 1 and Supplementary Note 3). Notably, we find evidence that ~40% of transcripts in the unstimulated auditory forebrain are non-coding and derive from intronic or intergenic loci (Table 1). Among the RNAs that are rapidly suppressed in response to new vocal signals ('novel down'), two-thirds are ncRNAs.

The robust involvement of ncRNAs in the response to song led us to ask whether song exposure alters the expression of microRNAs—small ncRNAs that regulate gene expression by binding to target messenger RNAs. Indeed we find that miR-124, a conserved microRNA implicated in neurological function in other species[28], is rapidly suppressed in response to song playbacks (Fig. 3). We independently measured this effect by direct Illumina sequencing of small RNAs in the auditory forebrain, and also identified other known and new microRNAs, several of which also change in expression after song stimulation (Supplementary Note 2).

A potential site of action for microRNAs was shown by genomic mapping of transcripts that increase rapidly after new song exposure (Table 1, 'novel up'). Two of the cDNA clones that measured the most robust increases[27] align to an unusually long (3 kilobases (kb)) 3′ untranslated region (UTR) in the human gene that encodes the NR4A3 transcription factor protein (Fig. 4a). The entire UTR is similar in humans and zebra finches, with several long segments of >80% identity (Fig. 4b). Within these segments we find conserved predicted binding sites for 11 different microRNAs, including five new microRNAs found by direct sequencing of small RNAs from the zebra finch forebrain (Fig. 4b). These findings indicate that this *NR4A3* transcript element may function in both humans and songbirds to integrate many conserved microRNA regulatory pathways.

The act of singing also alters gene expression in song control nuclei[29], and we used the genome assembly to analyse the transcriptional control structure of this response. Using oligonucleotide microarrays, we identified 807 genes in which expression significantly changed as a result of singing. These were grouped by *k*-means clustering into 20 distinct expression profile clusters (Fig. 5a and Supplementary Note 3). Gene regulatory sequences (transcription-factor-binding sites) were predicted across the genome using a new motif-scanning approach (Supplementary Note 1), and we observed significant correlation between changes in expression of transcription factor genes and their predicted targets (Fig. 5b and Supplementary Table 6). Thus, the experience of singing and hearing song engages complex gene regulatory networks in the forebrain, altering the expression of microRNAs, transcription factor genes, and their targets, as well as of non-coding RNA elements that may integrate transcriptional and post-transcriptional control systems.

Learned vocal communication is crucial to the reproductive success of a songbird, and this behaviour evolved after divergence

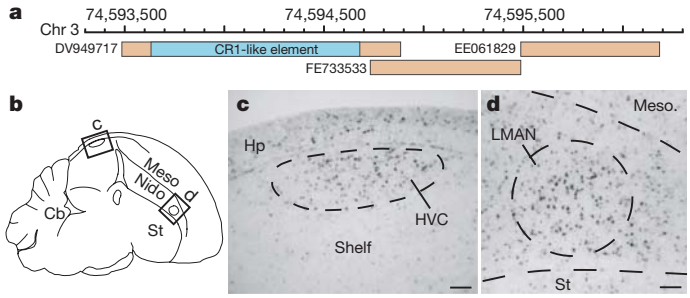

Figure 2 | Enriched expression of a CR1-like element in the zebra finch song system. **a**, Genomic alignment of an RNA containing a CR1-like retrotransposon element (in blue) and adjacent ESTs, with respective GenBank accession numbers. **b**–**d**, DV949717 is expressed in the brain of adult males with enrichment in song nuclei HVC (letter-based name) and LMAN (lateral magnocellular nucleus of the anterior nidopallium), as revealed by *in situ* hybridization. The diagram in **b** indicates areas shown in photomicrographs in **c** and **d**. Cb, cerebellum; Hp, hippocampus; Meso, mesopallium; Nido, nidopallium; Shelf, nidopallial shelf region; St, striatum. Scale bars, 0.1 mm.

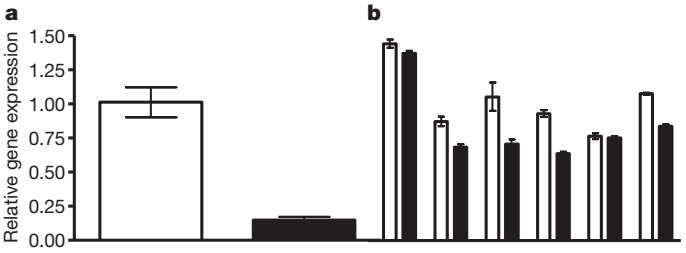

Figure 3 | miR-124 in the auditory forebrain is suppressed by exposure to new song. TaqMan assays comparing samples from the auditory lobule of adult male zebra finches in silence (open bars) or 30 min after onset of new song playback (filled bars). **a**, Comparison of two sample pools, each containing auditory forebrains of 20 birds. **b**, Comparisons of paired individual subjects, $n = 6$ pairs ($P = 0.03$, Wilcoxon paired test). Error bars denote s.e.m. of triplicate TaqMan assays. Parallel TaqMan analyses of the small RNA *RNU6B* were performed with all samples and showed no significant effect of treatment for this control RNA.

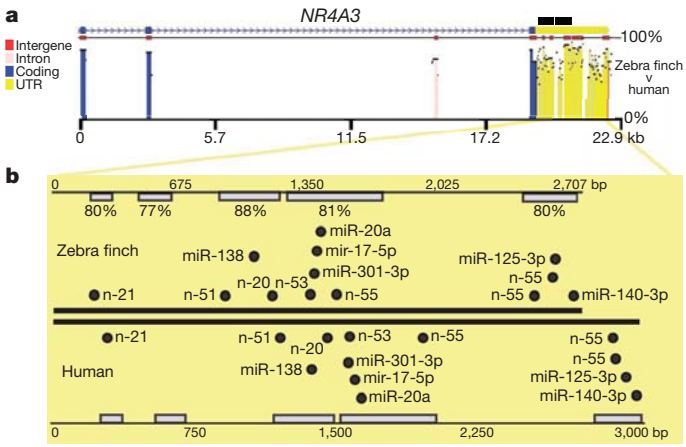

Figure 4 | Conserved *NR4A3* 3′UTR is a potential region for microRNA integration. **a**, zPicture alignment of 3′ portion of zebra finch to human gene[35] showing UTR region of high similarity beyond the coding exons. Dark red bars, regions with the highest sequence conservation; black rectangles, position of song-regulated ESTs[27] within the conserved UTR but outside the Ensembl gene model (ENSTGUG00000008853). **b**, Alignment of zebra finch and human 3′ UTR sequences showing the per cent sequence identity for each evolutionarily conserved region. Dots indicate positions of conserved new ('n-') or established ('miR-') microRNA-binding sites in both species within these regions.

of the songbird lineage[5]. Thus, it seems likely that genes involved in the neurobiology of vocal communication have been influenced by positive selection in songbirds. With this in mind, we examined the intersection of two sets of genes: (1) those that respond to song exposure in the auditory forebrain as discussed in the previous section; and (2) those that contain residues that seem to have been positively selected in the zebra finch lineage, as determined using phylogenetic analysis by maximum likelihood (PAML) (Supplementary Note 4). There are 214 genes that are common to both lists. Of these, 49 are suppressed by song exposure (Supplementary Table 7), and 6 of these 49 are explicitly annotated for ion channel activity (Table 2). This yields a highly significant statistical enrichment for the term 'ion channel activity' ($P = 0.0016$, false discovery rate (FDR) adjusted Fisher's exact test) and other related terms in this subset of genes (Supplementary Tables 8 and 9). Independent evidence has also demonstrated differential anatomical expression of ion channel genes in song control nuclei[26,30]. Ion channel genes have important roles in many aspects of behaviour, neurological function and disease[31]. This class of genes is highly likely to be linked to song behaviour and should be a major target for future functional studies.

Passerines represent one of the most successful and complex radiations of terrestrial animals[7]. Here we present the first, to our knowledge, analysis of the genome of a passerine bird. The zebra finch was chosen because of its well-developed status as a model organism for a number of fields in biology, including neurobiology, ethology, ecology, biogeography and evolution. In the zebra finch as in the chicken, we see a smaller, tighter genome compared to mammals, with a marked reduction of interspersed repeats. The zebra finch presents a picture of greater genomic plasticity than might have been expected from the chicken and other precedents, with a high degree of intrachromosomal rearrangements between the two avian species, gene copy number variations and transcribed mobile elements. Yet we also see an overall similarity to mammals in protein-coding gene content and core transcriptional control systems.

Our analysis suggests several channels through which evolution may have acted to produce the unique neurobiological properties of songbirds compared to the chicken and other animals. These include the management of sex chromosome gene expression, accelerated evolution of neuronal ion transport genes, gene duplications to produce new variants of *PHF7*, *PAK3* and other neurobiologically

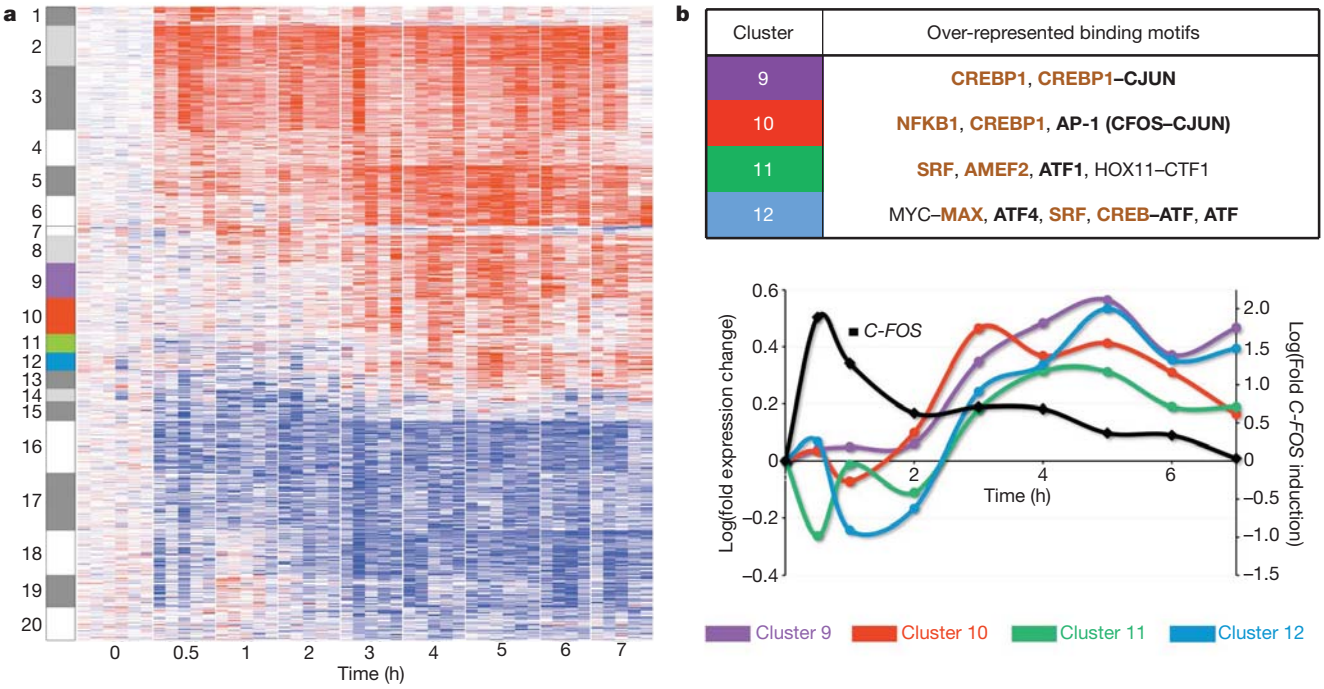

Figure 5 | Transcriptional control network in area X engaged by singing. **a**, Clustered (1–20) temporal expression profiles of 807 genes (rows) that change with time and amount of singing; red, increases; blue, decreases; white, no change relative to average 0-h control. Grey/coloured bars on left, clusters with enrichment of specific promoter motifs ($P < 0.01$). **b**, Enriched transcription-factor-binding motifs (abbreviations) found in the promoters of late response genes, clusters 9–12 (coloured as in **a**); bold, binding sites for known activity-dependent transcription factors (for example, CREBP1) or transcription factor complexes (for example, CREBP1–CJUN); black, sites for post-translationally activated transcription factors; brown, sites for transcriptionally activated transcription factors including by singing (for example, in cluster 1). Graph shows time course of average expression of all genes in the late response clusters, normalized to average 0 h for that cluster. Also plotted is the average expression of the *C-FOS* transcription factor mRNA, which binds to the AP-1 site over-represented in the promoters of cluster 10 genes.

Table 2 | Song-suppressed ion channel genes under positive selection

Gene	Description	Branch $\Delta\omega$	Sites PS/total
CACNA1B	Voltage-dependent N-type calcium channel subunit α-1B	0.016	9/2,484
CACNA1G	Voltage-dependent T-type calcium channel subunit α-1G	0.044*	2/2,468
GRIA2	Glutamate receptor 2 precursor (GluR-2, AMPA 2)	0.231*	17/948
GRIA3	Glutamate receptor 3 precursor (GluR-3, AMPA 3)	−0.010	4/894
KCNC2	Potassium voltage-gated channel subfamily C member 2 (Kv3.2)	0.315*	32/654
TRPV1	Transient receptor potential cation channel subfamily V member 1	−0.067	3/876

These six genes are suppressed by song exposure (FDR = 0.05)[27] and they show evidence of positive selection in the zebra finch relative to chicken ($P < 10^{-3}$, Supplementary Note 3). Branch $\Delta\omega$ denotes the difference in the non-synonymous to synonymous substitution ratio (*dN*/*dS*) between zebra finch and other birds (chicken and the ancestral branch leading to chicken and zebra finch). Positive values indicate that the gene is rapidly evolving, whereas negative values indicate genes evolving more slowly. Sites PS/total denotes the number of individual sites with empirical Bayes posterior probability greater than 0.95 of $\omega > 1$ (positive selection) in the finch versus the total number of residues in the protein, from branch-site model analysis implemented in PAML. Note that genes can show overall slower evolution in the branch model yet show evidence of significant positive selection at specific sites.
* Gene-wide differences that were significant ($P < 0.05$) by a likelihood ratio test.

important genes, and a new arrangement of MHC genes. Most notably, our analyses suggest a large recruitment of the genome during vocal communication, including the extensive involvement of ncRNAs. It has been proposed that ncRNAs have a contributing role in enabling or driving the evolution of greater complexity in humans and other complex eukaryotes[32]. Seeing that learned vocal communication itself is a phenomenon that has emerged only in some of the most complex organisms, perhaps ncRNAs are a nexus of this phenomenon.

Much work will be needed to establish the actual functional significance of many of these observations and to determine when they arose in avian evolution. This work can now be expedited with the recent development of a method for transgenesis in the zebra finch[33]. An important general lesson, however, is that dynamic and serendipitous aspects of the genome may have unexpected roles in the elaborate vocal communicative capabilities of songbirds.

METHODS SUMMARY

Sequence assembly. Sequenced reads were assembled and attempts were made to assign the largest contiguous blocks of sequence to chromosomes using a genetic linkage map[21], fingerprint map and synteny with the chicken genome assembly Gallus_gallus-2.1, a revised version of the original draft[6] (Supplementary Note 1).
Genes. Gene orthology assignment was performed using the EnsemblCompara GeneTrees pipeline and the OPTIC pipeline (Supplementary Note 1). Orthology rate estimation was performed with PAML (pairwise model = 0, Nssites = 0). In all cases, codon frequencies were estimated from the nucleotide composition at each codon position (F3X4 model).
Gene expression and evolution. Methods for Illumina read counting, *in situ* hybridization, TaqMan RT–PCR, microarrays, regulatory motif and evolutionary rate analyses are given in Supplementary Notes 1–4.

Received 30 September 2009; accepted 6 January 2010.

1. Zann, R. A. *The Zebra Finch: A Synthesis of Field and Laboratory Studies* (Oxford Univ, Press, 1996).
2. Clayton, D. F., Balakrishnan, C. N. & London, S. E. Integrating genomes, brain and behavior in the study of songbirds. *Curr. Biol.* **19,** R865–R873 (2009).
3. Nottebohm, F. in *Hope For a New Neurology* (ed. Nottebohm, F.) (New York Academy of Science, 1985).
4. Doupe, A. J. & Kuhl, P. K. Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* **22,** 567–631 (1999).
5. Jarvis, E. D. Learned birdsong and the neurobiology of human language. *Ann. NY Acad. Sci.* **1016,** 749–777 (2004).
6. Hillier, L. W. *et al.* Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432,** 695–716 (2004).
7. Hackett, S. J. *et al.* A phylogenomic study of birds reveals their evolutionary history. *Science* **320,** 1763–1768 (2008).
8. Zeigler, H. P. & Marler, P. *Behavioral Neurobiology of Bird Song* Vol. 1016 (New York Academy of Sciences, 2004).
9. Hahnloser, R. H., Kozhevnikov, A. A. & Fee, M. S. An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* **419,** 65–70 (2002).
10. Mooney, R. Neural mechanisms for learned birdsong. *Learn. Mem.* **16,** 655–669 (2009).
11. Konishi, M. & Akutagawa, E. Neuronal growth, atrophy and death in a sexually dimorphic song nucleus in the zebra finch brain. *Nature* **315,** 145–147 (1985).
12. Goldman, S. A. & Nottebohm, F. Neuronal production, migration, and differentiation in a vocal control nucleus of the adult female canary brain. *Proc. Natl Acad. Sci. USA* **80,** 2390–2394 (1983).
13. Nottebohm, F. The road we travelled: discovery, choreography, and significance of brain replaceable neurons. *Ann. NY Acad. Sci.* **1016,** 628–658 (2004).
14. London, S. E., Remage-Healey, L. & Schlinger, B. A. Neurosteroid production in the songbird brain: A re-evaluation of core principles. *Front. Neuroendocrinol.* **30,** 302–314 (2009).
15. Mello, C. V., Vicario, D. S. & Clayton, D. F. Song presentation induces gene expression in the songbird forebrain. *Proc. Natl Acad. Sci. USA* **89,** 6818–6822 (1992).
16. Dong, S. & Clayton, D. F. Habituation in songbirds. *Neurobiol. Learn. Mem.* **92,** 183–188 (2009).
17. Woolley, S. C. & Doupe, A. J. Social context-induced song variation affects female behavior and gene expression. *PLoS Biol.* **6,** e62 (2008).
18. Jarvis, E. D., Scharff, C., Grossman, M. R., Ramos, J. A. & Nottebohm, F. For whom the bird sings: context-dependent gene expression. *Neuron* **21,** 775–788 (1998).
19. Clayton, D. F. The genomic action potential. *Neurobiol. Learn. Mem.* **74,** 185–216 (2000).
20. Warren, W. C. *et al.* Genome analysis of the platypus reveals unique signatures of evolution. *Nature* **453,** 175–183 (2008).
21. Stapley, J., Birkhead, T. R., Burke, T. & Slate, J. A linkage map of the zebra finch *Taeniopygia guttata* provides new insights into avian genome evolution. *Genetics* **179,** 651–667 (2008).
22. Itoh, Y. *et al.* Dosage compensation is less effective in birds than in mammals. *J. Biol.* **6,** 2 (2007).
23. Ellegren, H. *et al.* Faced with inequality: chicken do not have a general dosage compensation of sex-linked genes. *BMC Biol.* **5,** 40 (2007).
24. Teranishi, M. *et al.* Transcripts of the MHM region on the chicken Z chromosome accumulate as non-coding RNA in the nucleus of female cells adjacent to the *DMRT1* locus. *Chromosome Res.* **9,** 147–165 (2001).
25. Arnold, A. P., Itoh, Y. & Melamed, E. A bird's-eye view of sex chromosome dosage compensation. *Annu. Rev. Genomics Hum. Genet.* **9,** 109–127 (2008).
26. Lovell, P. V., Clayton, D. F., Replogle, K. L. & Mello, C. V. Birdsong "transcriptomics": neurochemical specializations of the oscine song system. *PLoS One* **3,** e3440 (2008).
27. Dong, S. *et al.* Discrete molecular states in the brain accompany changing responses to a vocal signal. *Proc. Natl Acad. Sci. USA* **106,** 11364–11369 (2009).
28. Makeyev, E. V. & Maniatis, T. Multilevel regulation of gene expression by microRNAs. *Science* **319,** 1789–1790 (2008).
29. Wada, K. *et al.* A molecular neuroethological approach for identifying and characterizing a cascade of behaviorally regulated genes. *Proc. Natl Acad. Sci. USA* **103,** 15212–15217 (2006).
30. Wada, K., Sakaguchi, H., Jarvis, E. D. & Hagiwara, M. Differential expression of glutamate receptors in avian neural pathways for learned vocalization. *J. Comp. Neurol.* **476,** 44–64 (2004).
31. Cooper, E. C. & Jan, L. Y. Ion channel genes and human neurological disease: recent progress, prospects, and challenges. *Proc. Natl Acad. Sci. USA* **96,** 4759–4766 (1999).
32. Mattick, J. S. RNA regulation: a new genetics? *Nature Rev. Genet.* **5,** 316–323 (2004).
33. Agate, R. J., Scott, B. B., Haripal, B., Lois, C. & Nottebohm, F. Transgenic songbirds offer an opportunity to develop a genetic model for vocal learning. *Proc. Natl Acad. Sci. USA* **106,** 17963–17967 (2009).
34. Replogle, K. *et al.* The Songbird Neurogenomics (SoNG) Initiative: community-based tools and strategies for study of brain gene function and evolution. *BMC Genomics* **9,** 131 (2008).
35. Ovcharenko, I., Loots, G. G., Hardison, R. C., Miller, W. & Stubbs, L. zPicture: dynamic alignment and visualization tool for analyzing conservation profiles. *Genome Res.* **14,** 472–477 (2004).

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

Acknowledgements The sequencing of zebra finch was funded by the National Human Genome Research Institute (NHGRI). Further research support included grants to D.F.C. (NIH RO1 NS045264 and RO1 NS051820), H.E. (Swedish Research Council and Knut and Alice Wallenberg Foundation), E.D.J. (HHMI, NIH Directors Pioneer Award and R01 DC007218), M.A.B. (NIH RO1 GM59290) and J.S. (Biotechnology and Biological Sciences Research Council grant number BBE0175091). Resources for exploring the sequence and annotation data are

available on browser displays available at UCSC (http://genome.ucsc.edu), Ensembl (http://www.ensembl.org), the NCBI (http://www.ncbi.nlm.nih.gov) and http://aviangenomes.org. We thank K. Lindblad-Toh for permission to use the green anole lizard genome assembly, the Production Sequencing Group of The Genome Center at Washington University School of Medicine for generating all the sequence reads used for genome assembly, and the Clemson University Genome Institute for the construction of the BAC library. We would like to recognize all the important published work that we were unable to cite owing to space limitations.

Author Contributions W.C.W., D.F.C., H.E. and A.P.A. comprise the organizing committee of the zebra finch genome sequencing project. Project planning, management and data analysis: W.C.W., D.F.C., H.E. and A.P.A. Assembly annotation and analysis: L.W.H., P.M., S.-P.Y., L.Y., J.N., A.C., S.H., J.Sl., J.St., D.B. and S.-P.Y. Protein coding and non-coding gene prediction: S.S., C.B., P.F., S.W., A.H., C.P.P. and L.K. SNP analysis: P.F. and W.M.M. Orthology prediction and analysis: A.J.V., A.H., C.P.P., S.F. and L.K. Repeat element analysis: M.A.B., A.F.A.S., R.H., M.K.K., J.A.W., W.G. and D.D.P. Segmental duplication and gene duplication analysis: L.C., Z.C., E.E.E., L.K., C.P.P., M.F., C.N.B., R.E., J.G. and S.E.L. Protease annotation and analysis: X.S.P., V.Q., G.V. and C.L.-O. Neuropeptide hormone annotation: J.Sw. and B.S. Small non-coding RNA analysis: Y-C.L., Y.L., P.G., M.W. and X.L. Comparative mapping: D.K.G., M.V. and B.M.S. Singing induced gene network analysis: E.D.J., A.R.P., O.W. and J.H. Z-chromosome analysis: Y.I. and A.P.A. Gene expression and *in situ* analysis and synapsin synteny/loss analysis: C.V.M., P.L. and T.A.F.V. Adaptive evolution analysis: A.K., K.N., N.B., L.S., B.N. and C.N.B. Gene expression in the brain analysis: C.S., I.A., A.S., H.L., H.R. and M.S. MHC analysis: S.E., C.N.B. and R.E. Olfactory receptor analysis: T.O., D.L. and L.K. Sequencing management: R.K.W., E.R.M. and L.F. Physical map construction: T.G. Zebra finch tissue resources: T.Bu. and T.Bi. Zebra finch cDNA resources: D.F.C., E.D.J. and X.L.

Author Information The *Taeniopygia guttata* whole-genome shotgun project has been deposited in DDBJ/EMBL/GenBank under the project accession ABQF00000000. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to W.C.W. (wwarren@watson.wustl.edu), D.F.C (dclayton@illinois.edu), H.E. (hans.ellegren@ebc.uu.se) or A.P.A. (arnold@ucla.edu).