



Cite this article: Balakrishnan CN, Chapus C, Brewer MS, Clayton DF. 2013 Brain transcriptome of the violet-eared waxbill *Uraeginthus granatina* and recent evolution in the songbird genome. *Open Biol* 3: 130063. <http://dx.doi.org/10.1098/rsob.130063>

Received: 9 April 2013
Accepted: 30 July 2013

Subject Area:

genomics/neuroscience

Keywords:

zebra finch, genome, positive selection, sex chromosome, aggression, social behaviour

Author for correspondence:

Christopher N. Balakrishnan
email: balakrishnanc@ecu.edu

[†]Present address: Biological and Experimental Psychology, School of Biological and Chemical Sciences, Queen Mary University of London, London E1 4NS, UK.

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsob.130063>.

Brain transcriptome of the violet-eared waxbill *Uraeginthus granatina* and recent evolution in the songbird genome

Christopher N. Balakrishnan^{1,2}, Charles Chapus³,
Michael S. Brewer^{1,4} and David F. Clayton^{2,†}

¹Department of Biology, East Carolina University, Greenville, NC 27858, USA

²Institute for Genomic Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801, USA

³UMR-MD3, Institut de Recherche Biomédicale des Armées, Antenne Marseille, Marseille, 13007, France

⁴Department of Environmental Science, Policy and Management, University of California, Berkeley, CA 94720, USA

1. Summary

Songbirds are important models for the study of social behaviour and communication. To complement the recent genome sequencing of the domesticated zebra finch, we sequenced the brain transcriptome of a closely related songbird species, the violet-eared waxbill (*Uraeginthus granatina*). Both the zebra finch and violet-eared waxbill are members of the family Estrildidae, but differ markedly in their social behaviour. Using Roche 454 RNA sequencing, we generated an assembly and annotation of 11 084 waxbill orthologues of 17 475 zebra finch genes (64%), with an average transcript length of 1555 bp. We also identified 5985 single nucleotide polymorphisms (SNPs) of potential utility for future population genomic studies. Comparing the two species, we found evidence for rapid protein evolution (ω) and low polymorphism of the avian Z sex chromosome, consistent with prior studies of more divergent avian species. An intriguing outlier was putative chromosome 4A, which showed a high density of SNPs and low evolutionary rate relative to other chromosomes. Genome-wide ω was identical in zebra finch and violet-eared waxbill lineages, suggesting a similar demographic history with efficient purifying natural selection. Further comparisons of these and other estrildid finches may provide insights into the evolutionary neurogenomics of social behaviour.

2. Introduction

To date, nearly 50 mammalian genomes have been completely sequenced. These diverse genome sequences capture many of the major lineages in the mammalian tree of life, and frame the study of evolution on a broad scale. Detailed sampling of closely related species, however, provides a complementary perspective. Analyses of closely related primate species have begun to reveal the molecular [1,2] and regulatory [3] changes underlying species differences. Similarly, among insects the effort to sequence 12 *Drosophila*

species has yielded insights into patterns of nucleotide and gene family evolution [4–6].

Of over 10 000 species in the avian radiation, seven genomes have now been sequenced to near completion [7–12]. Two of these, the chicken (*Gallus gallus*) and the turkey (*Melagris gallopavo*), represent a single avian family (Phasianidae). The Phasianidae include a number of species of importance in the food industry, and the chicken in particular has been important in studies of developmental biology and immunology. Falcon (Falconidae), pigeon (Columbidae) and mallard duck (Anatidae) genomes have also recently been sequenced [10–12]. Passerines are the most diverse avian order, including over half of all bird species. Within this group, the only published genome is that of the zebra finch *Taeniopygia guttata* [8]. The zebra finch was selected for sequencing as a representative of the passerines, and specifically due to its prominent role as an experimental system for the study of neurobiology and behaviour [13–15].

Owing to the limited genomic data currently available for birds, studies of avian molecular evolution have largely been restricted to comparisons between the zebra finch and chicken [8,16,17] (but see [9,12]). Zebra finch and chicken lineages diverged between 75 and 150 Ma, thus bracketing all of avian diversity with the exception of the palaeognathes [18–20]. Comparisons of zebra finch and chicken therefore necessarily average evolutionary signatures across much of avian diversity. In doing so, the signal of recent adaptation in songbirds has been obscured [17]. Next-generation sequencing studies are driving a rapid increase in the genome-scale data available for birds [21,22], but these datasets too have targeted divergent taxa. Transcriptome datasets, however, offer the potential to rapidly fill in the avian tree of life [21,23–27], allowing comparisons of closely related taxa towards the understanding of molecular evolution on more recent time scales.

Birds in general display striking variation in social behaviour [28], and the estrildid finches (family Estrildidae, 143 species) in particular have been cited as important model systems for studying the behavioural, endocrinological and neural substrates of sociality [29–31]. Within the estrildids, the zebra finch in some ways represents one extreme along a continuum of variation in social behaviour. Zebra finches are highly colonial, living in large, nomadic flocks in Australia and southeast Asia. The other extreme is represented by the violet-eared waxbill (*Uraeginthus granatina*), native to Africa. In contrast to the zebra finch and many other estrildids, the waxbill is highly territorial, especially during the breeding season. These two species diverged around 10–15 Ma at the base of the estrildid finch radiation [32,33]. Notable progress has been made in defining neurobiological elements underlying vocal and social communication (especially in the zebra finch [34–40]) and territoriality (especially in the violet-eared waxbill [31]).

Thus, the violet-eared waxbill is both an emerging model organism in its own right, and an intriguing phylogenetic contrast to the zebra finch for studies of avian evolution and sociality. With this in mind, we sequenced the brain transcriptome of the violet-eared waxbill. We present here a detailed picture of transcription in the waxbill brain and molecular evolution in recently diverged songbird lineages, taking additional advantage of the recent publication of the transcriptome of a non-estrildid songbird species, the great tit (*Parus major*) [23].

3. Material and methods

3.1. 454 Library preparation sequencing and assembly

RNA was extracted from whole brain tissue (provided by Jim Goodson, Indiana University) of a male violet-eared waxbill that had been snap-frozen on dry ice. The frozen whole brain was manually fragmented and homogenized in Tri-reagent (Ambion). RNA was then extracted from the homogenate following the manufacturer's protocol. Purified RNA was checked for quality using an Agilent Biol analyzer. Fifty micrograms of total RNA were further purified using a Qiagen RNeasy spin column to further remove any possible DNA contamination. Polyadenylated RNA was selected from the total RNA with an Oligotex mini kit (Qiagen). Poly-A-selected RNA was then reverse transcribed using random primers. cDNA was fragmented using a nebulizer to generate fragments ranging from approximately 400 to 800 bp in length. We then normalized the cDNA library using the Trimmer Direct kit (Evrogen) and manufacturer's protocols. End repair, 3' addition of A bases and ligation of adaptors were done following Illumina library preparation guidelines but using Roche 454 adaptors. We then ran the library in agarose and gel purified the 400–800 bp band.

Sequences were assembled using GSASSEMBLER software (v. 2.3) from Roche. Reads that contained homopolymers (60% over the entire length of the read represented by one nucleotide) and reads that were shorter than 100 bp were filtered. The parameters used for the assembly were overlapMinMatchLength of 40 and overlapMinMatchIdentity of 90%. The resulting isotigs and singlets that were more than 100 bp were annotated using BLAST against non-redundant protein database from NCBI (<http://www.ncbi.nlm.nih.gov>), chicken proteins, and zebra finch transcripts, proteins and genome from Ensembl (<http://uswest.ensembl.org/>).

3.2. Single nucleotide polymorphism detection and analysis

We used DIAL (De novo Identification of Alleles [41]) to identify single nucleotide polymorphisms (SNPs) in the waxbill transcriptome. DIAL is specifically tailored to identify SNPs in species for which a reference genome is lacking. DIAL incorporates platform-specific pipelines (in this case, we specified Roche 454 sequencing) and a specific algorithm for cDNA sequencing (-transcript flag). Because we sequenced a single individual, our discovered waxbill SNPs are due to heterozygosity and are biased towards the identification of high-frequency polymorphisms. The approach in DIAL uses Newbler, the Roche assembly algorithm, to construct clusters within which polymorphisms are surveyed. SNPs discovered by DIAL were then mapped to the zebra finch genome. We assumed conservation of genome structure between waxbill and zebra finch to assess chromosomal location of waxbill SNPs and their location relative to genes (intergenic, intronic and genic). This assumption is reasonable given the broad conservation of synteny across birds [8,42]. Cytogenetic analyses (c- and g-banding) of estrildid finch chromosomes have also been conducted, revealing structural polymorphisms within the group [43]. These previous analyses, however, emphasized intrachromosomal polymorphisms, which are common, but would not impact the conclusions drawn here.

3.3. Sequence alignment and molecular rate analyses

To align sequences from the new violet-eared waxbill transcriptome, we compared assembled transcripts with annotated zebra finch cDNAs and with another recently published songbird transcriptome—that of the great tit [23]—using BLAST+2.2.25. The whole set of waxbill and great tit transcripts were compared with the zebra finch database using the following parameters: expected e -value = 1×10^{-5} and minimum length of the BLAST hit = 300 bp. For each waxbill and great tit transcript, we kept the zebra finch hit satisfying these criteria and with the best e -value.

In order to conduct downstream molecular rate analyses, we refined alignments to match homologous codons across species and remove any non-coding regions. To refine alignments, we used MUSCLE [44,45] with the `-diags` parameter invoked to improve alignment speed given the similarity of the species in question. We mapped the MUSCLE alignments onto the Ensembl zebra finch cDNA using the starting coordinate of the previously derived BLAST alignments. Where there were extra nucleotides in the transcriptome sequences (waxbill or great tit) that disrupted open reading frames, we treated them as sequencing errors and discarded the extra bases. Ensembl-annotated zebra finch cDNAs by definition all maintain an open reading frame, and we sought to maintain these annotated gene structures in our alignments. For waxbill and great tit, where multiple isoforms may have been assembled, this approach preferentially selects the most similar isoform to that represented in the zebra finch genome assembly. It is possible that different isoforms are present in the transcriptome(s) than in the Ensembl annotation, but in this case, exons not represented in the zebra finch gene models will simply not be represented in the analysis.

In many cases, there was more than one waxbill and/or great tit assembled transcripts mapped to each zebra finch cDNA. This was because individual isotigs often did not span the entire length of the zebra finch gene. We therefore determined a consensus sequence of each gene for each species. We trimmed the alignments by using the zebra finch CDS positions as a guide, and removed terminal stop codons for each sequence in the alignments. We used the same general approach to generate pairwise alignments (zebra finch–waxbill, zebra finch–great tit, great tit–waxbill). As these alignments required only two species to overlap, they resulted in a larger number of alignments and longer alignments. The quality and method of sequence alignment have important impacts on inferences regarding rates of evolution [46]. During the course of this study, we tested multiple alignment pipelines, including the use of amino acid sequence-based approaches [47]. Visual inspection of alignments supported the use of our MUSCLE-based pipeline.

We analysed molecular evolutionary patterns using PAML (phylogenetic analysis using maximum likelihood) [48,49]. For pairwise alignments, we used the pairwise ω (dN/dS) estimators implemented in codeml. We also used the likelihood approach in codeml and tested ‘branch’ models in which waxbill or zebra finch was allowed to have an independent rate relative to a null model where all three species had a fixed rate. We then used likelihood ratio tests to identify genes whose rate was significantly accelerated in either the waxbill or the zebra finch lineage. p -values were determined using a χ^2 -distribution with one degree of freedom and were adjusted

for multiple testing ($q < 0.05$) using Q -value [50]. To examine molecular rates at different time scales, we also compared rates from zebra finch/waxbill codeml comparisons with those from zebra finch/chicken comparisons. Estimates of ω from zebra finch relative to chicken were made using Ensembl Biomart (ensembl.org).

To estimate the overall rate of protein evolution (ω) across the genome, we generated bootstrap datasets based on our previous alignment following Heger & Ponting [5]. For each species pair, we generated 1000 bootstrap datasets by concatenating 150 randomly chosen alignments. These longer alignments generate more robust and reliable estimates of genome-wide ω [5]. Across each of the 1000 alignments, we compared average ω between pairwise analyses of zebra finch–great tit and waxbill–great tit to test for rate variation between these two species, as might be expected if the two species had different demographic histories.

Gene lists (genes represented in the transcriptome assembly, genes showing rate variation) were functionally described using gene ontology (GO) analyses using CORNA [51] as implemented in a public web server (www.ark-genomics.org/tools/GOfinch). Fisher’s exact tests and hypergeometric tests were conducted to test for statistical over- and under-representation of GO terms. All statistics were adjusted for multiple hypotheses testing using the method of Benjamini & Hochberg [52], and p -values given below are adjusted p -values unless otherwise noted.

4. Results

4.1. Assembly and annotation

Two plates of 454 sequencing yielded 975 606 and 1 055 860, reads with average read lengths of 323.5 bp and 380.5 bp, respectively. Raw sequence data have been deposited at the NCBI short read archive under accession no. SRX337999. The total waxbill brain transcriptome dataset therefore consisted of 2 031 466 reads, and 2 013 275 after filtering. Transcriptome assembly yielded 32 938 isogroups (genes) and 43 137 isotigs (transcripts) with an average size of 1555 bp and an N50 value of 2486 bp. This compares favourably with recent de novo transcriptome assemblies of the zebra finch (mean contig length = 150 bp [53]), great tit (mean contig length = 871 bp [23]) kiwi (mean contig length = 162 bp [26]) and other bird species [21]. The transcript set described here closely matches the transcript length profile for the full Ensembl transcript set for zebra finch (figure 1). The longest assembled isotig in our assembly spanned 17 589 bp, covering the primary transcript of the mitochondrial genome. Four hundred and seventy nine contigs could not be placed in isotigs, but were included in subsequent analyses, yielding 43 616 putative transcripts. A total of 233 903 singleton reads were not included in the assembly. BLAST analysis of assembled transcripts found significant matches to 7817 Ensembl genes out of 17 475 total zebra finch Ensembl genes (44.7%). Including singletons in our count of detected transcripts expands our transcriptome coverage to 11 084 Ensembl genes (63.4%).

4.2. Gene functional representation

We detected a total of 7027 GO categories in our dataset of which a number were significantly over- and under-represented

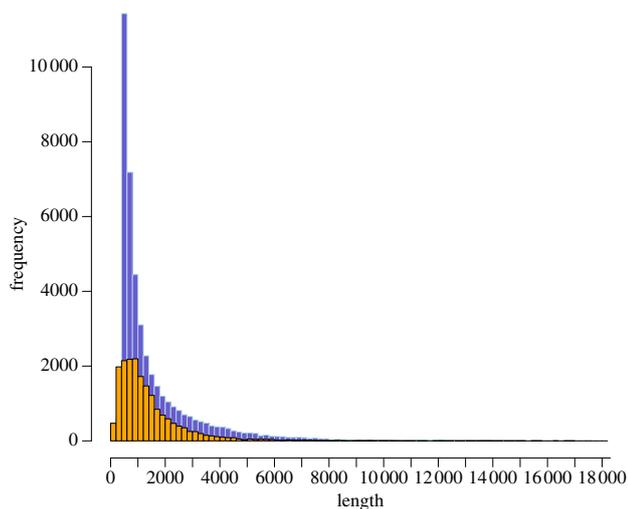


Figure 1. Distribution of 454 isotig size of violet-eared waxbill (purple, $n = 43\,616$) versus Ensembl cDNA sequences (orange, $n = 18\,597$).

(Fisher's exact test) relative to the full zebra finch Ensembl gene set (figures 2 and 3; electronic supplementary material, table S1). Terms describing cellular components including endoplasmic reticulum ($p = 1.3 \times 10^{-6}$), cytoplasm ($p = 8.3 \times 10^{-36}$) and cytosol ($p = 8.4 \times 10^{-8}$) were significantly enriched. As might be expected of a brain-derived library, categories of neurobiological function were also enriched (axon, $p = 0.0001$; dendrite, $p = 0.0048$; neuronal cell body, $p = 0.019$). A broad group of categories related to RNA processing and translation were also enriched. Under-represented categories were particularly intriguing and included categories related to transcriptional regulation (e.g. regulation of transcription, DNA-dependent, $p = 3.8 \times 10^{-6}$). Also lacking were immune-related transcripts (e.g. immune response, $p = 3.7 \times 10^{-9}$) and olfactory receptors (ORs; $p = 2.5 \times 10^{-79}$). Only two OR genes were detected relative to an expectation of 116 and total of 168 annotated ORs in the zebra finch genome (see also the electronic supplementary material, table S1).

4.3. Distribution of genetic variation in the transcriptome

Heterozygosity of the sequenced waxbill individual allowed us to identify SNPs in the transcriptome read data. Using DIAL, we identified 5985 SNPs, of which 5641 (94.3%) could be mapped to the zebra finch genome. Our transcriptome includes reads that map to known genes and reads that map outside of them, including both putatively intergenic and intronic compartments. Reads mapping outside of Ensembl-annotated genes may represent novel, previously undescribed transcripts. Reads mapping to annotated introns may represent alternative or incomplete splicing (precursor mRNA). A total of 22.9% of SNPs mapped to known coding sequences, and another 12.9% mapped to within 1 kb of the 3' flanking end of gene annotations. A total of 38.4% of the SNPs mapped intergenically (not including those in the 'flanking' category above) and 22.0% mapped to annotated introns. The remaining reads mapped annotated UTRs, telomeres and the 5' flanking region (within 1 kb) of known genes.

SNPs mapped to 28 zebra finch chromosomes and the number of SNPs detected per chromosome scaled linearly with chromosome size (figure 4). One striking exception to this pattern was the sex chromosome Z, which showed a marked reduction in SNP density. By contrast, chromosome 4A showed a much higher density of SNPs than any of the other chromosomes to which we mapped SNPs. Across all chromosomes, the SNPs were distributed at a density of 0.02 SNPs/kb, whereas chromosome 4A had 375 SNPs on an assembled chromosome of only 258 280 bp (0.54 SNPs/kb).

4.4. Patterns of gene and genome evolution

4.4.1. Pairwise comparisons between zebra finch and violet-eared waxbill

We generated alignments for 5310 putative orthologues for zebra finch and violet-eared waxbill. This gene set was restricted to alignments of 300 bp or greater, and averaged 1367 bp in length after removal of gapped and ambiguous positions. Of these alignments, 63 (0.01%) had $\omega > 1$, the traditional benchmark for adaptive evolution [54]. This list was significantly enriched for 13 GO categories ($p = 0.05$; see electronic supplementary material, table S1), but these tended to be small GO categories where the null expectation was 0 and the observation was one gene. Four categories pertaining to transcriptional regulation were enriched but fell short of statistical significance after correction for multiple testing (sequence-specific DNA binding, $p = 0.083$; regulation of transcription, DNA-dependent, adjusted $p = 0.11$; transcription factor activity, $p = 0.14$). These categories were represented by three, four and three genes, respectively.

To more broadly describe rate variation among genes, we also examined gene lists of $\omega > 0.8$ (112 genes) and the overall top 10% most rapidly evolving genes (530 genes; $\omega > 0.4075$). At $\omega > 0.8$, three GO categories (spindle astral microtubule organization, interkinetic nuclear migration and regulation of microtubule-based process) were significant ($p < 0.05$). GO categories neurogenesis, NADH dehydrogenase activity and cerebral cortex development bordered on statistical significance ($p = 0.075$). Among the top 10%, no GO categories were significant after correction for multiple comparisons. Four genes annotated as having cytokine receptor activity (out of six total such genes in the dataset) were in the top 10% of fastest-evolving genes, and this category also bordered on statistical significance ($p = 0.12$). Several functional categories were also moderately under-represented, suggesting stabilizing selection on the genes comprising these categories (protein serine/threonine kinase activity, protein tyrosine kinase activity, protein amino acid phosphorylation and protein kinase activity, $0.05 < p < 0.15$).

Pairwise ω estimates also revealed variation of molecular evolutionary rate by chromosome (ANOVA, $p = 0.0002$; figure 5). Genes of the Z sex chromosome are evolving faster than the other chromosomes (figure 4), and significantly faster than chromosome 4 (one-tailed t -test, $p = 0.01$), the chromosome closest to Z in the number of aligned genes (Chr Z = 326 genes and Chr 4 = 364 genes). Chromosome 4A is also evolving slowly relative to chromosome 4, from which it is derived (one-tailed t -test, $p = 0.005$), and chromosome 12, the chromosome most similar in gene number (one-tailed t -test, $p = 0.009$).

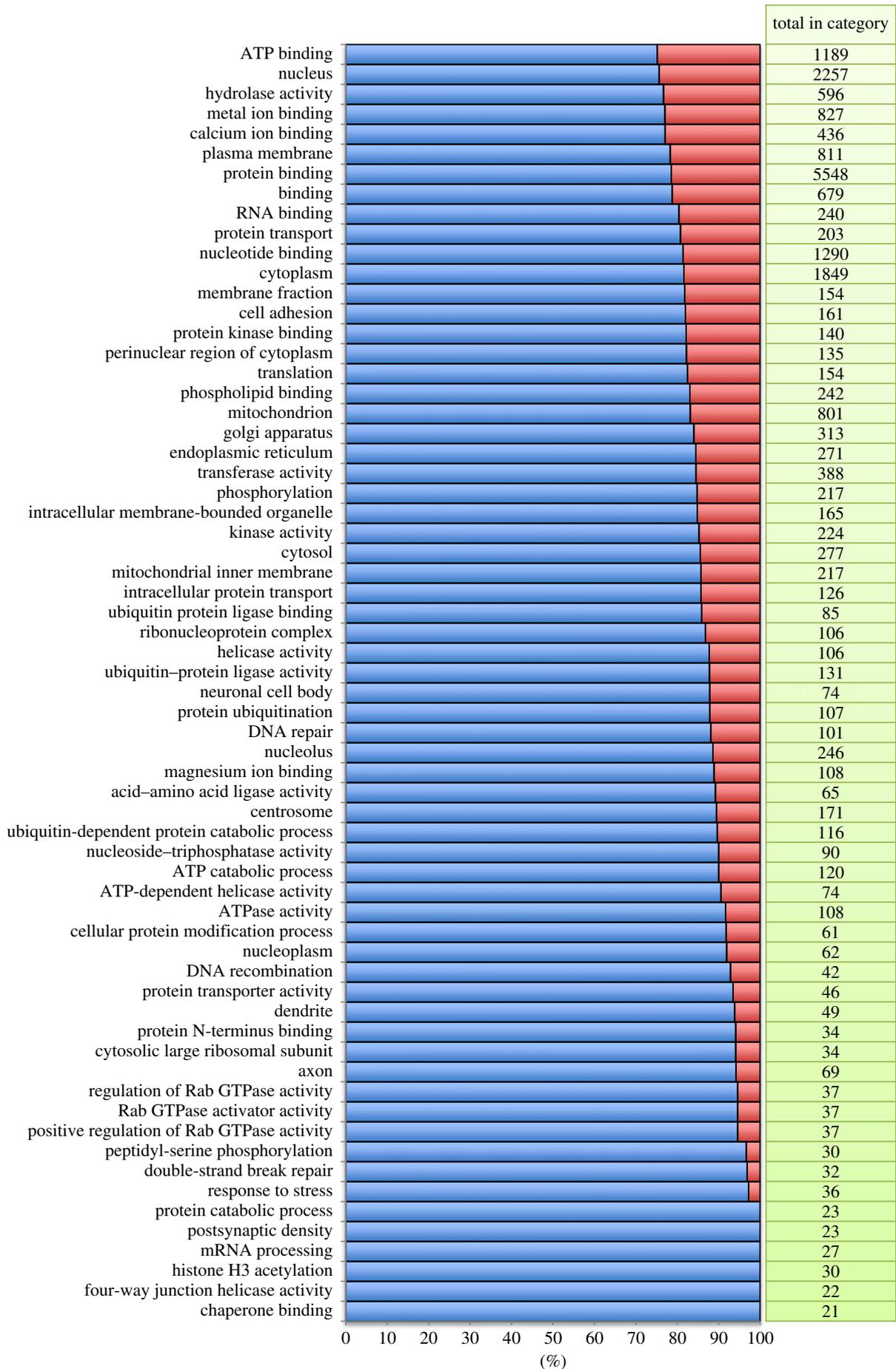


Figure 2. Significantly over-represented GO terms (Fisher's exact test $FDR < 0.05$) in the violet-eared waxbill brain transcriptome. The blue portion of each column represents the proportion of the total number genes within each GO category based on zebra finch that was detected in the waxbill assembly. The red portion of each bar represents the proportion that was not detected.

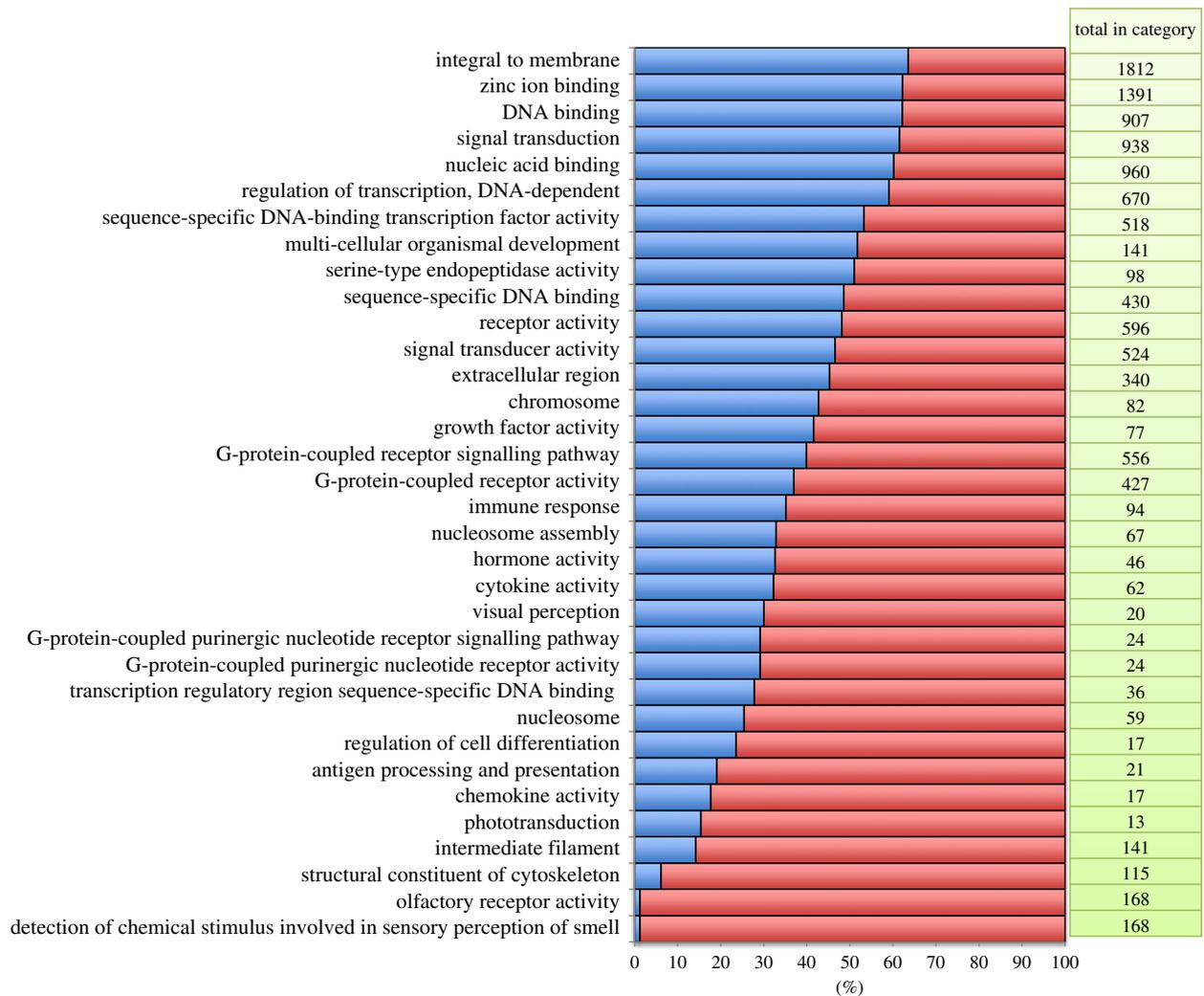


Figure 3. Significantly under-represented GO terms (Fisher's exact test $FDR < 0.05$) in the violet-eared waxbill brain transcriptome. The blue portion of each column represents the proportion of the total number genes within each GO category based on zebra finch that was detected in the waxbill assembly. The red portion of each bar represents the proportion that was not detected.

To estimate rates of evolution across the genome as a whole, we generated bootstrap datasets (see Material and methods) from pairwise alignments of each of our ingroup taxa (zebra finch and violet-eared waxbill) with an outgroup, the great tit. In each case, we estimated the average genomic ω between to be 0.13. Therefore, we found no difference in overall molecular rate between the two estrildid finch lineages. We also found a significant correlation between ω estimated from zebra finch/chicken orthologues in Ensembl and zebra finch/waxbill orthologues estimated here ($R = 0.35$, $p < 0.001$; figure 6), supporting broadly similar patterns of molecular evolution at these different time scales.

4.4.2. Phylogenetic rate analyses

Three-species alignments incorporating the great tit allowed us to examine positive and negative selection in a likelihood framework, testing models of branch-specific rate variation versus a null model of equal rates across the three branches of the tree. Adding a third species left us with 4721 alignments with a minimum length of 300 bp and an average length 1124.7 bases. We tested a null model of equal rates across all three lineages with alternatives in which either violet-eared waxbill or zebra finch was allowed an independent rate. By allowing rate to vary on the zebra finch lineage, we detected significant variation in rate in 347

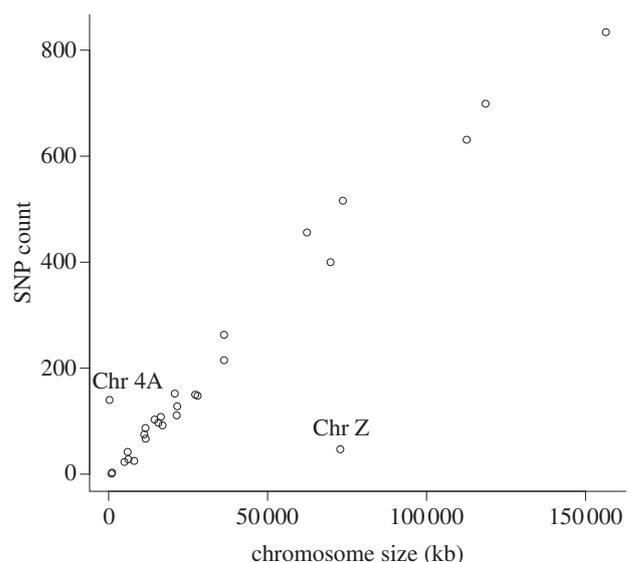


Figure 4. Distribution of SNPs among zebra finch chromosomes. In general, the number of discovered SNPs scales with chromosome size. Two notable exceptions, however, are the Z sex chromosome and chromosome 4A.

genes, 112 of which showed deceleration and 235 of which showed acceleration. However, only five of these genes (KIAA1712, DPM1, FOXK2, MXI1 and ATP6V0A) were

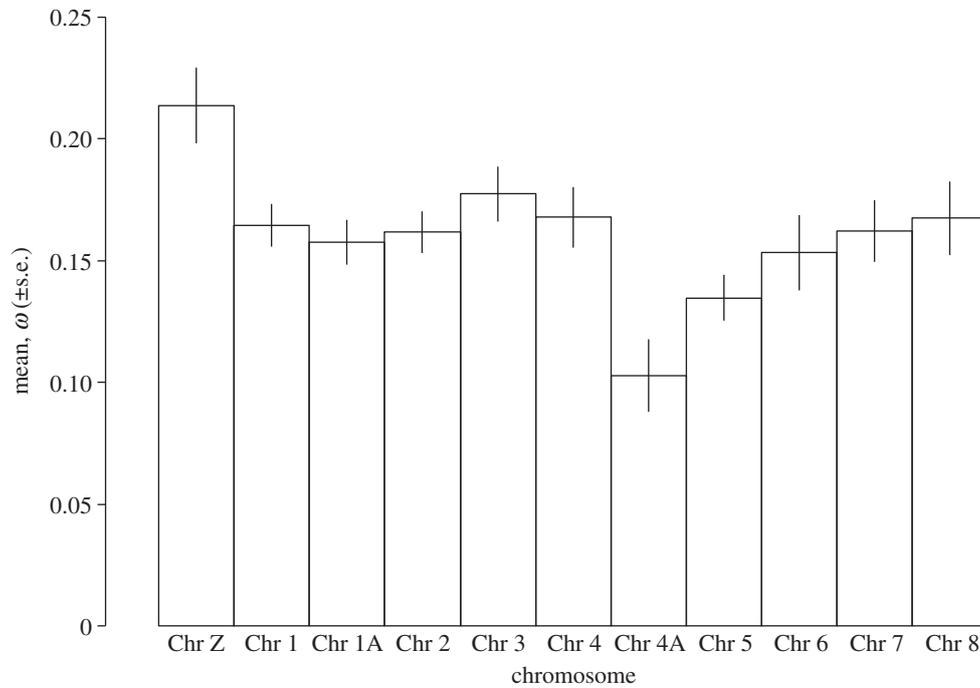


Figure 5. Pairwise ω across chromosomes 1–8 and Z (the macrochromosomes). There was significant variation among chromosomes in ω , with chromosomes Z and 4A showing a significant increase and decrease, respectively, from the other chromosomes.

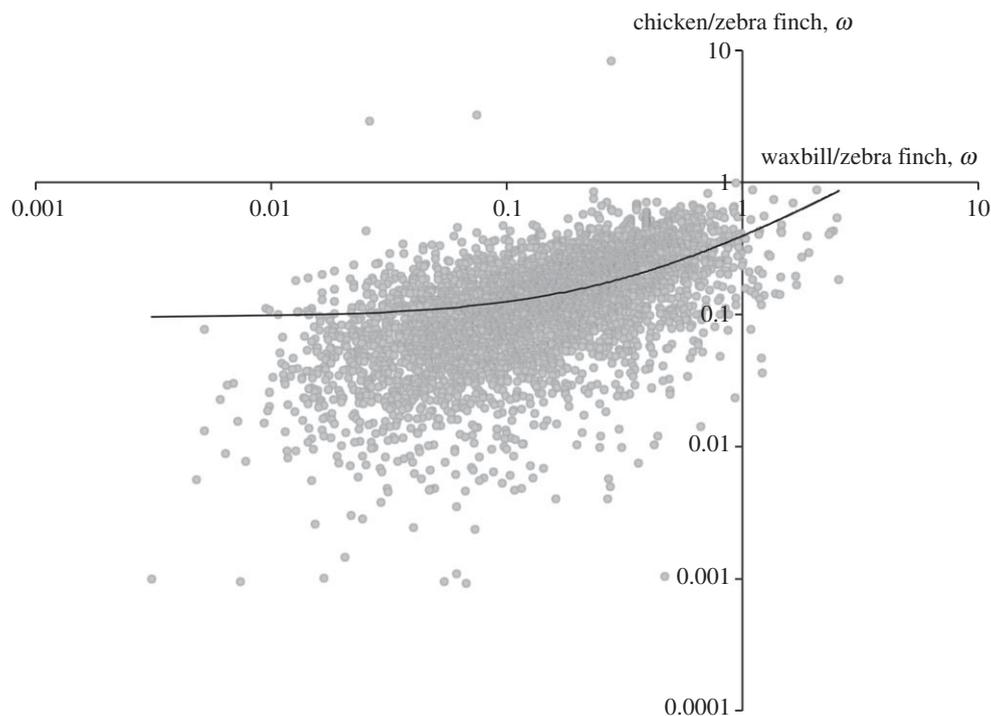


Figure 6. Significant correlation ($r = R = 0.35$, $p < 0.001$) in ω derived from chicken/zebra finch orthologues (Ensembl Biomart) versus those from zebra finch/waxbill orthologues (this study).

significantly accelerated after FDR correction and none were significantly decelerated. DPM1 is annotated with a number of intriguing GO functions related to brain and behaviour, such as axon target recognition, axonogenesis, eating behaviour, adult walking behaviour and response to stress. These GO categories were significantly enriched ($p < 0.05$); but in all cases, the observation was one gene relative to an expectation of 0 based on a sample of five accelerated genes. GO analysis of the full list of 235 accelerated genes (without correction for multiple comparisons) revealed enrichment of a number of other functional categories including cAMP-

dependent protein kinase complex, regulation of protein amino acid phosphorylation and forebrain development, among others (table 1). These enrichments, however, were non-significant after multiple testing corrections.

In the violet-eared waxbill lineage, we identified significant rate variation in 282 genes, of which 107 were decelerated and 175 accelerated. Only one gene, NCKAP1, was significant after FDR correction ($q < 0.05$). NCKAP1 is associated with five GO terms: apoptotic process, central nervous system development, integral to membrane, lamellipodium membrane and protein binding. The 175 accelerated genes are

Table 1. GO analysis of functional over-representation of genes showing significant rate acceleration ($p < 0.05$) in the zebra finch lineage relative to violet-eared waxbill and great tit.

GO	GO description	total	expected	observation	p -value	adjusted p -value
GO:0005952	cAMP-dependent protein kinase complex	4	0	3	0.00046	0.24
GO:0008603	cAMP-dependent protein kinase regulator activity	5	0	3	0.0011	0.24
GO:0005088	Ras guanyl-nucleotide exchange factor activity	6	0	3	0.0021	0.24
GO:0015386	potassium : hydrogen antiporter activity	2	0	2	0.0024	0.24
GO:0090181	regulation of cholesterol metabolic process	2	0	2	0.0024	0.24
GO:0019933	cAMP-mediated signalling	3	0	2	0.0071	0.24
GO:0017016	Ras GTPase binding	3	0	2	0.0071	0.24
GO:0015385	sodium : hydrogen antiporter activity	3	0	2	0.0071	0.24
GO:0051259	protein oligomerization	9	0	3	0.008	0.24
GO:0001932	regulation of protein phosphorylation	9	0	3	0.008	0.24
GO:0005083	small GTPase regulator activity	11	1	3	0.015	0.24
GO:0031594	neuromuscular junction	5	0	2	0.022	0.24
GO:0017157	regulation of exocytosis	5	0	2	0.022	0.24
GO:0016442	RNA-induced silencing complex	5	0	2	0.022	0.24
GO:0015299	solute : hydrogen antiporter activity	5	0	2	0.022	0.24
GO:0006885	regulation of pH	5	0	2	0.022	0.24
GO:0003729	mRNA binding	13	1	3	0.024	0.24
GO:0051056	regulation of small GTPase-mediated signal transduction	25	1	4	0.033	0.24
GO:0006814	sodium ion transport	15	1	3	0.035	0.24
GO:0016607	nuclear speck	16	1	3	0.042	0.24
GO:0000082	G1/S transition of mitotic cell cycle	7	0	2	0.043	0.24
GO:0000932	cytoplasmic mRNA processing body	7	0	2	0.043	0.24
GO:0045931	positive regulation of mitotic cell cycle	7	0	2	0.043	0.24
GO:0006417	regulation of translation	7	0	2	0.043	0.24
GO:0007265	Ras protein signal transduction	7	0	2	0.043	0.24
GO:0005488	binding	323	16	24	0.044	0.24

mildly (but not significantly) enriched for GO terms including neuropeptide receptor activity, mitochondrial respiratory chain complex 1, extracellular ligand-gated ion channel activity, neurotransmitter receptor activity (table 2).

5. Discussion

We have generated a brain transcriptome from a developing model species for neurobiology and behaviour, the violet-eared waxbill. The brain transcriptome encompasses partial or complete coverage of orthologues to 11 064 zebra finch genes, or 64% of currently annotated zebra finch genes. Our deep sequencing of two full Roche 454 plates reconstructed longer transcripts than any previously produced bird transcriptome, as measured by an N50 score and average transcript length. The combination of normalized libraries, long read lengths and deep sequencing probably explain this success. Despite this, our focus on a single tissue sample (whole brain), and

the challenges of detecting and assembling rare transcripts, prevented us from attaining 'complete' transcriptome coverage.

As might be predicted in a metabolically costly tissue like the brain, we found significant over-representation of GO terms associated with mitochondrial function and cellular energetics. We also observed a striking lack of expression of OR genes. Although birds were long considered not to have an important sense of smell, the discovery of numerous OR genes [55,56], and evidence for the use of smell [57–59], suggests that olfaction in birds has been underappreciated. In the waxbill brain transcriptome, however, we detected only expression of two ORs out of a large pool of such genes in the genome. While this pattern is striking, ORs are predominantly expressed in the olfactory epithelium [60,61]. Despite normalizing our library to minimize the impact of highly expressed genes, it is possible that OR expression is simply too low in the brain to be detected. While OR repertoires have begun to be characterized across bird lineages, there remains very little information on when, where and at what level these receptors are expressed [62–64]. A prior analysis in the zebra finch also

Table 2. GO analysis of genes showing significant rate acceleration in the violet-eared waxbill lineage relative to zebra finch and great tit ($p < 0.05$). A number of neurobiologically interesting categories are slightly enriched (Fishers test, $p < 0.05$), but fall short of statistical significance after correction for multiple testing (adjusted p -value).

GO	GO description	total	expected	observation	p -value	adjusted p -value
GO:0005230	extracellular ligand-gated ion channel activity	14	1	4	0.0013	0.13
GO:0006836	neurotransmitter transport	7	0	3	0.0015	0.13
GO:0016620	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor	10	0	3	0.0048	0.19
GO:0004222	metalloendopeptidase activity	21	1	4	0.0063	0.19
GO:0004890	GABA-A receptor activity	11	0	3	0.0064	0.19
GO:0007218	neuropeptide signalling pathway	13	0	3	0.01	0.19
GO:0045211	post-synaptic membrane	26	1	4	0.014	0.19
GO:0046854	phosphatidylinositol phosphorylation	15	1	3	0.016	0.19
GO:0005694	chromosome	16	1	3	0.019	0.19
GO:0031625	ubiquitin protein ligase binding	33	1	4	0.031	0.19
GO:0006351	transcription, DNA-dependent	49	2	5	0.032	0.19
GO:0045893	positive regulation of transcription, DNA-dependent	85	3	7	0.035	0.19
GO:0007166	cell surface receptor signalling pathway	20	1	3	0.035	0.19
GO:0016021	integral to membrane	476	17	26	0.037	0.19
GO:0045202	synapse	37	1	4	0.045	0.21
GO:0030054	cell junction	38	1	4	0.048	0.21

found scarce evidence for olfactory gene expression in the brain [65]. Detailed analyses of ORs in the zebra finch genome also suggest that current Ensembl annotations do not fully describe the OR repertoire [66]. Deeper and tissue-specific RNA sequencing in zebra finches and other passerines, as well as improvements to the genome assembly itself, will improve our understanding of OR repertoires and expression patterns.

We also found poor representation of the immune genome in the waxbill brain. This is consistent with the notion of the brain being 'immune privileged' [67,68]. Zebra finch brain ESTs have, however, revealed expression of MHC class I in the brain [8,53,69], and our waxbill transcriptome also includes an MHC class I gene. The finding of MHC class I genes in songbird brains is consistent with neurobiological roles for these genes in mammalian systems [70]. As a whole, however, immune genes are poorly represented in the brain transcriptome.

We have also described patterns of molecular evolution over the last 15 Myr in the estrildid finches. Although previous estimates of avian nucleotide substitution rates were derived from deep evolutionary divergences, we found that our estimate of genome-wide ω , 0.13, closely matches those derived in a previous study comparing chicken and zebra finch [17]. This ratio is similar to estimates from *Drosophila* [5] and rodents [71], and is lower than estimates from primates [71]. This supports the hypothesis that birds have had relatively large effective population sizes over their history, resulting in relatively efficient purifying natural selection. The fact that we focused on brain-expressed transcripts,

however, also probably biases this estimate downwards. Inclusion of RNAs from gonadal and immune-active tissues might incorporate a disproportionate number of fast-evolving genes, bumping the overall estimate up slightly. Based on our findings, we conclude that the estrildid finch brain transcriptome, at least, has evolved under efficient purifying selection. This pattern differs markedly from the recent observation of high rates of protein evolution among two closely related falcons [12].

We did not find any difference in ω between the two focal taxa here, zebra finch and violet-eared waxbill. This was somewhat contrary to our expectation. Zebra finches are colonial, abundant throughout Australia and appear to have had extremely large effective population sizes in their history [62]. Using a set of 30 sequence loci, Balakrishnan & Edwards [72] estimated effective population size (N_e) for Australian zebra finches to be around seven million. Given this large effective population size, we suspected that zebra finches would show relatively efficient purifying selection when compared with violet-eared waxbills. Violet-eared waxbills, however, also have a broad range in southern Africa [73], and population sizes for this species therefore must also be large enough to effectively purge slightly deleterious mutations that would otherwise elevate genomic estimates of ω [74]. Large effective population sizes and efficient natural selection may be the norm for passerine bird species. A lingering question is the extent to which domestication of the zebra finch over the last 100 years or more has shaped its genome sequence. The zebra finch genome assembly is based on a captive bird,

potentially influencing our estimates of nucleotide substitution relative to that derived from a wild bird.

Our analyses confirmed an increased rate of evolution on the Z sex chromosome. Previous studies of avian genomes have shown this and attributed fast evolution to the lower effective population size of the Z chromosome [16,75]. Despite this fast evolution of the Z chromosome, we did not find strong evidence of reproductive genes (expressed in the brain) being a specific target of positive selection in the genome. We also found a low rate of polymorphism on the Z chromosome, here measured as SNP density. This finding is also consistent with the hypothesis of purifying natural selection acting on the Z sex chromosome [76]. An unexpected outlier in our rate and SNP analyses was chromosome 4A. Chromosome 4A showed the opposite pattern from the Z chromosome in that ω was relatively low and SNP density was high. We speculate that this pattern reflects a relaxation of selection on the genes of chromosome 4A, perhaps in association with the fission of chromosome 4, which gave rise to 4A. Chromosome 4A is also of special interest as it has been described as a neo-sex chromosome in warblers (Sylvioidea), with linkage to the Z chromosome [77]. It is not clear whether the unusual patterns of variation and divergence we discovered might be related to the interesting biology of the chromosome.

Strong purifying selection was also evident in our gene-specific analyses of evolutionary rates. In pairwise comparisons, we found few genes with $\omega > 1$ and phylogenetic tests for rate acceleration similarly uncovered only a handful of cases of strongly accelerated evolution. In many analyses of positive selection, functional categories of immune response and reproduction are over-represented among positively selected genes [78]. We did not find such a signature in our analysis save for a signature of rapid evolution of cytokine genes. We attribute this primarily to the fact that such genes, those involved

in immunity and reproduction, are not well represented in our brain transcriptome. Immune genes were in fact significantly under-represented in our dataset as a whole (figure 2). Analyses of complete avian genomes, however, also failed to detect this common signature [17]. Nam *et al.* [17] attributed this to the fact that in their study, evolutionary comparisons of divergent species diluted the signature of adaptive bouts of evolution. Immune genes, however, have been shown to be evolving rapidly on relatively recent time scales in comparisons of turkey and chicken genomes [9]. Enhancing the waxbill brain transcriptome with genes better represented in other tissue (e.g. spleen and gonads) might help us to test for positive selection in immune- and reproduction-related genes.

Our study used 'branch' models to identify genes that show a signature of accelerated evolution across the entire gene. This approach is conservative because often only a few residues within a protein undergo adaptive evolution [69]. We refrained from using more powerful 'branch site' models given that our analysis involved only three passerine species for which high-quality transcriptomes were available, giving us limited power to investigate site-specific patterns. As new transcriptomes and whole genomes are sequenced for birds, we will have improved power to detect positive selection and uncover genes underlying traits of interest. The estrildid finches, in particular, represent a promising focal point for studying how genomic evolution is linked to neurobiological and behavioural change.

Acknowledgements. Jim Goodson provided the waxbill brain tissue sample used in this study. Aakrosh Ratan provided guidance in the use of DIAL software. Sequencing, assembly and preliminary annotation was performed at the Roy J. Carver Biotechnology Center at the University of Illinois by Alvaro Hernandez, Jyothi Thimapurram and their staff.

Funding statement. This work was funded by NIH NIGMS 1RC1GM091556 and East Carolina University.

References

- Dorus S, Vallender EJ, Evans PD, Anderson JR, Gilbert SL, Mahowald M, Wyckoff GJ, Malcom CM, Lahn BT. 2004 Accelerated evolution of nervous system genes in the origin of *Homo sapiens*. *Cell* **119**, 1027–1040. (doi:10.1016/j.cell.2004.11.040)
- Enard D, Depaulis F, Crollius HR. 2010 Human and non-human primate genomes share hotspots of positive selection. *PLoS Genet.* **6**, e1000840. (doi:10.1371/journal.pgen.1000840)
- Nowick K, Gernat T, Almaas E, Stubbs L. 2009 Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain. *Proc. Natl Acad. Sci. USA* **106**, 22 358–22 363. (doi:10.1073/pnas.0911376106)
- Hahn MW, Han MV, Han SG. 2007 Gene family evolution across 12 *Drosophila* genomes. *PLoS Genet.* **3**, e197. (doi:10.1371/journal.pgen.0030197)
- Heger A, Ponting CP. 2007 Evolutionary rate analyses of orthologs and paralogs from 12 *Drosophila* genomes. *Genome Res.* **17**, 1837–1849. (doi:10.1101/gr.6249707)
- Clark AG *et al.* 2007 Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203–218. (doi:10.1038/nature06341)
- Hillier LW *et al.* 2004 Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716. (doi:10.1038/nature03154)
- Warren WC *et al.* 2010 The genome of a songbird. *Nature* **464**, 757–762. (doi:10.1038/nature08819)
- Dalloul RA *et al.* 2010 Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. *PLoS Biol.* **8**, e1000475. (doi:10.1371/journal.pbio.1000475)
- Shapiro MD *et al.* 2013 Genomic diversity and evolution of the head crest in the rock pigeon. *Science* **339**, 1063–1067. (doi:10.1126/science.1230422)
- Huang Y *et al.* 2013 The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. *Nat. Genet.* **45**, 776–783. (doi:10.1038/ng.2657)
- Zhan X *et al.* 2013 Peregrine and saker falcon genome sequences provide insights into evolution of a predatory lifestyle. *Nat. Genet.* **45**, 563–566. (doi:10.1038/ng.2588)
- Clayton DF, Balakrishnan CN, London SE. 2009 Integrating genomes, brain and behavior in the study of songbirds. *Curr. Biol.* **19**, R865–R873. (doi:10.1016/j.cub.2009.07.006)
- Griffith SC, Buchanan KL. 2010 The zebra finch: the ultimate Australian supermodel. *Emu* **110**, v–xii. (doi:10.1071/MUv110n3_ED)
- Zann RA. 1996 *The zebra finch: a synthesis of field and laboratory studies*. Oxford, UK: Oxford University Press.
- Mank JE, Axelsson E, Ellegren H. 2007 Fast-X on the Z: rapid evolution of sex-linked genes in birds. *Genome Res.* **17**, 618–624. (doi:10.1101/gr.6031907)
- Nam K *et al.* 2010 Molecular evolution of genes in avian genomes. *Genome Biol.* **11**, R68. (doi:10.1186/gb-2010-11-6-r68)
- Brown JW, Rest JS, Garcia-Moreno J, Sorenson MD, Mindell DP. 2008 Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. *BMC Biol.* **6**, 6. (doi:10.1186/1741-7007-6-6)

19. Hackett SJ *et al.* 2008 A phylogenomic study of birds reveals their evolutionary history. *Science* **320**, 1763–1768. (doi:10.1126/science.1157704)
20. Sibley CG, Ahlquist JE. 1990 *Phylogeny and classification of birds: a study in molecular evolution*. New Haven, CT: Yale University Press.
21. Kunstner A *et al.* 2010 Comparative genomics based on massive parallel transcriptome sequencing reveals patterns of substitution and selection across 10 bird species. *Mol. Ecol.* **19**, 266–276. (doi:10.1111/j.1365-294X.2009.04487.x)
22. Haussler D *et al.* 2009 Genome 10K: a proposal to obtain whole-genome sequence for 10 000 vertebrate species. *J. Hered.* **100**, 659–674. (doi:10.1093/jhered/esp086)
23. Santure AW, Gratten J, Mossman JA, Sheldon BC, Slate J. 2011 Characterisation of the transcriptome of a wild great tit *Parus major* population by next generation sequencing. *BMC Genomics* **12**, 283. (doi:10.1186/1471-2164-12-283)
24. Srivastava A, Winker K, Shaw TI, Jones KL, Glenn TC. 2012 Transcriptome analysis of a North American songbird, *Melospiza melodia*. *DNA Res.* **19**, 325–333. (doi:10.1093/dnares/dss015)
25. Peterson MP *et al.* 2012 De novo transcriptome sequencing in a songbird, the dark-eyed junco (*Junco hyemalis*): genomic tools for an ecological model system. *BMC Genomics* **13**, 305. (doi:10.1186/1471-2164-13-305)
26. Subramanian S, Huynen L, Millar CD, Lambert DM. 2010 Next generation sequencing and analysis of a conserved transcriptome of New Zealand's kiwi. *BMC Evol. Biol.* **10**, 387. (doi:10.1186/1471-2148-10-387)
27. Wang B, Ekblom R, Castoe TA, Jones EP, Kozma R, Bongcam-Rudloff E, Pollock DD, Hoglund J. 2012 Transcriptome sequencing of black grouse (*Tetrao tetrix*) for immune gene discovery and microsatellite development. *Open Biol.* **2**, 120054. (doi:10.1098/rsob.120054)
28. Tinbergen N. 1953 *Social behaviour in animals: with special reference to vertebrates*. London, UK: Methuen.
29. Goodson JL. 1998 Vasotocin and vasoactive intestinal polypeptide modulate aggression in a territorial songbird, the violet-eared waxbill (Estrildidae: *Uraeginthus granatina*). *Gen. Comp. Endocrinol.* **111**, 233–244. (doi:10.1006/gcen.1998.7112)
30. Goodson JL, Kabelik D, Schrock SE. 2009 Dynamic neuromodulation of aggression by vasotocin: influence of social context and social phenotype in territorial songbirds. *Biol. Lett.* **5**, 554–556. (doi:10.1098/rsbl.2009.0316)
31. Goodson JL, Wang YW. 2006 Valence-sensitive neurons exhibit divergent functional profiles in gregarious and asocial species. *Proc. Natl Acad. Sci. USA* **103**, 17 013–17 017. (doi:10.1073/pnas.0606278103)
32. Sorenson MD, Balakrishnan CN, Payne RB. 2004 Clade-limited colonization in brood parasitic finches (*Vidua* spp.). *Syst. Biol.* **53**, 140–153. (doi:10.1080/10635150490265021)
33. Sorenson MD, Sefc KM, Payne RB. 2003 Speciation by host switch in brood parasitic indigobirds. *Nature* **424**, 928–931. (doi:10.1038/nature01863)
34. Mello CV, Clayton DF. 1994 Song-induced zenk gene-expression in auditory pathways of songbird brain and its relation to the song control-system. *J. Neurosci.* **14**, 6652–6666.
35. Mello CV, Nottebohm F, Clayton DF. 1995 Repeated exposure to one song leads to a rapid and persistent decline in an immediate early gene's response to that song in zebra finch telencephalon. *J. Neurosci.* **15**, 6919–6925.
36. Mello CV, Vicario DS, Clayton DF. 1992 Song presentation induces gene-expression in the songbird forebrain. *Proc. Natl Acad. Sci. USA* **89**, 6818–6822. (doi:10.1073/pnas.89.15.6818)
37. Dong S, Replogle KL, Hasadsri L, Imai BS, Yau PM, Rodriguez-Zas S, Southey BR, Sweedler JV, Clayton DF. 2009 Discrete molecular states in the brain accompany changing responses to a vocal signal. *Proc. Natl Acad. Sci. USA* **106**, 11 364–11 369. (doi:10.1073/pnas.0812998106)
38. Avey M, Phillmore L, Macdougall-Schackleton S. 2005 Immediate early gene expression following exposure to acoustic and visual components of courtship in zebra finches. *Behav. Brain Res.* **165**, 247–253. (doi:10.1016/j.bbr.2005.07.002)
39. Chew SJ, Vicario SD, Nottebohm F. 1996 A large-capacity memory system that recognizes the calls and songs of individual birds. *Proc. Natl Acad. Sci. USA* **93**, 1950–1955. (doi:10.1073/pnas.93.5.1950)
40. Bailey D, Wade J. 2003 Differential expression of the immediate early genes FOS and ZENK following auditory stimulation in the juvenile male and female zebra finch. *Mol. Brain Res.* **116**, 147–154. (doi:10.1016/S0169-328X(03)00288-2)
41. Ratan A, Yu Z, Hayes VM, Schuster SC, Miller W. 2010 Calling SNPs without a reference sequence. *BMC Bioinform.* **11**, 130. (doi:10.1186/1471-2105-11-130)
42. Volker M, Backstrom N, Skinner BM, Langley EJ, Bunzey SK, Ellegren H, Griffin DK. 2010 Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution. *Genome Res.* **20**, 503–511. (doi:10.1101/gr.103663.109)
43. Christidis L. 1987 Chromosomal evolution within the family Estrildidae (Aves) 0.3. The Estrildae (waxbill finches). *Genetica* **72**, 93–100. (doi:10.1007/bf00123167)
44. Edgar RC. 2004 MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* **5**, 113. (doi:10.1186/1471-2105-5-113)
45. Edgar RC. 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797. (doi:10.1093/nar/gkh340)
46. Markova-Raina P, Petrov D. 2011 High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 *Drosophila* genomes. *Genome Res.* **21**, 863–874. (doi:10.1101/gr.115949.110)
47. Suyama M, Torrents D, Bork P. 2006 PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–W612. (doi:10.1093/nar/gkl315)
48. Yang ZH. 2007 PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591. (doi:10.1093/molbev/msm088)
49. Yang ZH. 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. *Comp. Appl. Biosci.* **13**, 555–556.
50. Storey JD. 2002 A direct approach to false discovery rates. *J. R. Stat. Soc. B* **64**, 479–498. (doi:10.1111/1467-9868.00346)
51. Wu X, Watson M. 2009 CORNA: testing gene lists for regulation by microRNAs. *Bioinformatics* **25**, 832–833. (doi:10.1093/bioinformatics/btp059)
52. Benjamini H, Hochberg Y. 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300.
53. Ekblom R, Balakrishnan CN, Burke T, Slate J. 2010 Digital gene expression analysis of the zebra finch genome. *BMC Genomics* **11**, 219. (doi:10.1186/1471-2164-11-219)
54. Yang ZH, Bielawski JP. 2000 Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**, 496–503. (doi:10.1016/S0169-5347(00)01994-7)
55. Steiger SS, Fidler AE, Kempnaers B. 2009 Evidence for increased olfactory receptor gene repertoire size in two nocturnal bird species with well-developed olfactory ability. *BMC Evol. Biol.* **9**, 117. (doi:10.1186/1471-2148-9-117)
56. Steiger SS, Fidler AE, Valcu M, Kempnaers B. 2008 Avian olfactory receptor gene repertoires: evidence for a well-developed sense of smell in birds? *Proc. R. Soc. B* **275**, 2309–2317. (doi:10.1098/rspb.2008.0607)
57. De Groof G, Gwinner H, Steiger S, Kempnaers B, Van der Linden A. 2010 Neural correlates of behavioural olfactory sensitivity changes seasonally in European starlings. *PLoS ONE* **5**, e14337. (doi:10.1371/journal.pone.0014337)
58. Krause ET, Caspers BA. 2012 Are olfactory cues involved in nest recognition in two social species of estrildid finches? *PLoS ONE* **7**, e36615. (doi:10.1371/journal.pone.0036615)
59. Krause ET, Krueger O, Kohlmeier P, Caspers BA. 2012 Olfactory kin recognition in a songbird. *Biol. Lett.* **8**, 327–329. (doi:10.1098/rsbl.2011.1093)
60. Zhang XM, Rogers M, Tian HK, Zhang XH, Zou DJ, Jian L, Ma MH, Shepherd GM, Firestein SJ. 2004 High-throughput microarray detection of olfactory receptor gene expression in the mouse. *Proc. Natl Acad. Sci. USA* **101**, 14 168–14 173. (doi:10.1073/pnas.0405350101)
61. Zhang X, De la Cruz O, Pinto JM, Nicolae D, Firestein S, Gilad Y. 2007 Characterizing the expression of the human olfactory receptor gene family using a novel DNA microarray. *Genome Biol.* **8**, R86. (doi:10.1186/gb-2007-8-5-r86)
62. Nef S, Allaman I, Fiumelli H, DeCastro E, Nef P. 1996 Olfaction in birds: Differential embryonic expression

- of nine putative odorant receptor genes in the avian olfactory system. *Mech. Dev.* **55**, 65–77. (doi:10.1016/0925-4773(95)00491-2)
63. Nef S, Nef P. 1997 Olfaction: transient expression of a putative odorant receptor in the avian notochord. *Proc. Natl Acad. Sci. USA* **94**, 4766–4771. (doi:10.1073/pnas.94.9.4766)
 64. Steiger SS, Fidler AE, Kempnaers B. 2008 Detection of olfactory receptor transcripts in bird testes. *J. Hered.* **99**, 624–628. (doi:10.1093/jhered/esn051)
 65. Replogle K *et al.* 2008 The Songbird Neurogenomics (SoNG) Initiative: community-based tools and strategies for study of brain gene function and evolution. *BMC Genom.* **9**, 131. (doi:10.1186/1471-2164-9-131)
 66. Steiger SS, Kuryshev VY, Stensmyr MC, Kempnaers B, Mueller JC. 2009 A comparison of reptilian and avian olfactory receptor gene repertoires: species-specific expansion of group gamma genes in birds. *BMC Genom.* **10**, 446. (doi:10.1186/1471-2164-10-446)
 67. Galea I, Bechmann I, Perry VH. 2007 What is immune privilege (not)? *Trends Immunol.* **28**, 12–18. (doi:10.1016/j.it.2006.11.004).
 68. Shatz CJ. 2009 MHC class I: an unexpected role in neuronal plasticity. *Neuron* **64**, 40–45. (doi:10.1016/j.neuron.2009.09.044)
 69. Balakrishnan CN *et al.* 2010 Gene duplication and fragmentation in the zebra finch major histocompatibility complex. *BMC Biol.* **8**, 29. (doi:10.1186/1741-7007-8-29)
 70. Boulanger LM, Shatz CJ. 2004 Immune signalling in neural development, synaptic plasticity and disease. *Nat. Rev. Neurosci.* **5**, 521–531. (doi:10.1038/nrn1428)
 71. Kosiol C, Vinar T, da Fonseca RR, Hubisz MJ, Bustamante CD, Nielsen R, Siepel A. 2008 Patterns of positive selection in six mammalian genomes. *PLoS Genet.* **4**, e1000144. (doi:10.1371/journal.pgen.1000144)
 72. Balakrishnan CN, Edwards SV. 2009 Nucleotide variation, linkage disequilibrium and founder-facilitated speciation in wild populations of the zebra finch (*Taeniopygia guttata*). *Genetics* **181**, 645–660. (doi:10.1534/genetics.108.094250)
 73. Goodwin D. 1982 *Estrildid finches of the world*. Ithaca, NY: Cornell University Press.
 74. Ohta T. 2002 Near-neutrality in evolution of genes and gene regulation. *Proc. Natl Acad. Sci. USA* **99**, 16 134–16 137. (doi:10.1073/pnas.252626899).
 75. Mank JE, Nam K, Ellegren H. 2010 Faster-Z evolution is predominantly due to genetic drift. *Mol. Biol. Evol.* **27**, 661–670. (doi:10.1093/molbev/msp282).
 76. Sundstrom H, Webster MT, Ellegren H. 2004 Reduced variation on the chicken Z chromosome. *Genetics* **167**, 377–385. (doi:10.1534/genetics.167.1.377).
 77. Pala I, Naurin S, Stervander M, Hasselquist D, Bensch S, Hansson B. 2012 Evidence of a neo-sex chromosome in birds. *Heredity* **108**, 264–272. (doi:10.1038/hdy.2011.70).
 78. Nielsen R *et al.* 2005 A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **3**, 976–985. (doi:10.1371/journal.pbio.0030170)