Highly Variable Recombinational Landscape Modulates Efficacy of Natural Selection in Birds

Toni I. Gossmann[1],*, Anna W. Santure[1,2], Ben C. Sheldon[3], Jon Slate[1], and Kai Zeng[1],*

[1]Department of Animal and Plant Sciences, University of Sheffield, United Kingdom

[2]School of Biological Sciences, University of Auckland, New Zealand

[3]Edward Grey Institute, Department of Zoology, University of Oxford, United Kingdom

*Corresponding author: E-mail: toni.gossmann@googlemail.com, k.zeng@sheffield.ac.uk.

Accepted: July 18, 2014

Abstract

Determining the rate of protein evolution and identifying the causes of its variation across the genome are powerful ways to understand forces that are important for genome evolution. By using a multitissue transcriptome data set from great tit (*Parus major*), we analyzed patterns of molecular evolution between two passerine birds, great tit and zebra finch (*Taeniopygia guttata*), using the chicken genome (*Gallus gallus*) as an outgroup. We investigated whether a special feature of avian genomes, the highly variable recombinational landscape, modulates the efficacy of natural selection through the effects of Hill–Robertson interference, which predicts that selection should be more effective in removing deleterious mutations and incorporating beneficial mutations in high-recombination regions than in low-recombination regions. In agreement with these predictions, genes located in low-recombination regions tend to have a high proportion of neutrally evolving sites and relaxed selective constraint on sites subject to purifying selection, whereas genes that show strong support for past episodes of positive selection appear disproportionally in high-recombination regions. There is also evidence that genes located in high-recombination regions tend to have higher gene expression specificity than those located in low-recombination regions. Furthermore, more compact genes (i.e., those with fewer/shorter introns or shorter proteins) evolve faster than less compact ones. In sum, our results demonstrate that transcriptome sequencing is a powerful method to answer fundamental questions about genome evolution in nonmodel organisms.

Key words: protein evolution, natural selection, Hill–Robertson interference (HRI), tissue specificity in gene expression, recombination, RNAseq.

Introduction

It is well known that the rate of protein evolution varies across the genome (Li 1997). Determining the causes of this variation is a powerful way to quantify the relative importance of natural selection and genetic drift and to identify factors that are important in shaping patterns of molecular evolution (Kimura 1983; Li 1997; Pál et al. 2006). When protein-coding DNA sequences are analyzed, the rate of protein evolution is often measured by the ratio $\omega = d_n/d_s$, where d_n and d_s are, respectively, the rates of nonsynonymous and synonymous substitutions (Li 1997; Nei and Kumar 2000; Yang 2006). By using the ratio of d_n to d_s, ω is expected to be less sensitive to variation in mutation (Nei and Gojobori 1986; Goldman and Yang 1994; Li 1997), which is known to exist in the genome (Lynch 2010; Hodgkinson and Eyre-Walker 2011). Therefore, variation in ω is considered to reflect selective pressures on the protein (Li 1997; Nei and Kumar 2000; Yang 2006). Specifically, under the assumption that synonymous changes are neutral, $\omega < 1$ is regarded as evidence of purifying selection acting on nonsynonymous mutations, $\omega = 1$ reflects neutral evolution, and $\omega > 1$ can be viewed as support for past episodes of positive selection driving nonsynonymous mutations to fixation (Li 1997; Nei and Kumar 2000; Yang 2006).

Estimates of ω have been obtained from a large array of different taxa. ω is generally less than 1 when considering the genome as a whole, reflecting the widely accepted theory that most nonsynonymous mutations have harmful effects on fitness and are therefore removed by purifying selection (Pál et al. 2006; Eyre-Walker and Keightley 2007). However, ω varies substantially across genes in the genome. Attempts to understand biological/selective causes of this variation have uncovered that ω is associated with factors such as protein dispensability, protein structure and stability, the number of protein–protein interactions, developmental timing, and patterns of gene expression (both in terms of expression level and tissue specificity); the extensive literature on these topics have been reviewed by Pál et al. (2006) and Choi and Hannenhalli (2013) (see also Marais et al. 2005; Parmley et al. 2007; Axelsson et al. 2008; Larracuente et al. 2008; Cai and Petrov 2010). Gene expression pattern appears to be a major correlate of protein evolutionary rate. For instance, in multicellular organisms, broadly expressed genes tend to have lower ω than genes with high tissue specificity in expression (Duret and Mouchiroud 2000; Axelsson et al. 2008; Larracuente et al. 2008; Slotte et al. 2011). Furthermore, genes involved in certain biological processes such as immunity and reproduction (e.g., spermatogenesis) tend to evolve faster than other genes in the genome and are often enriched for targets of positive selection, probably as a result of both inter- and intraspecies arms races (Nielsen et al. 2005; Haerty et al. 2007; Axelsson et al. 2008; Kosiol et al. 2008; Obbard et al. 2009). In contrast, genes with neural functions such as those expressed primarily in the brain exhibit lower evolutionary rates, which is likely to be a consequence of strong selective pressures to minimize the damaging effects induced by protein misfolding (Drummond and Wilke 2008). It should be noted that direction and intensity of correlations between ω and the factors mentioned above are sometimes inconsistent between species (reviewed by Pál et al. 2006 and Choi and Hannenhalli 2013), highlighting the importance of investigating these effects in distantly related species to verify their generality.

Variation in ω can also be induced by heterogeneity in recombination rate across the genome (Smukowski and Noor 2011). Physical linkage between loci on the same chromosome may affect the rate of protein evolution through Hill–Robertson interference (HRI), whereby any locus linked to other loci subject to directional selection experiences a reduction in local N_e, the effective population size (Hill and Robertson 1966; McVean and Charlesworth 2000; Comeron et al. 2008; Sella et al. 2009; Charlesworth 2012; Cutter and Payseur 2013). Because the efficacy of selection is determined by $N_e s$, where s is the selection coefficient (Kimura 1983; Charlesworth B and Charlesworth D 2010), tight linkage between selected sites hinders both the fixation of beneficial mutations and the elimination of deleterious mutations. Recombination reduces the interference by breaking up the association between variants at different loci, which in turn increases N_e, and hence the effectiveness of selection. The HRI theory therefore predicts that adaptive substitutions should appear more frequently in high-recombination regions, whereas low-recombination regions may accumulate more fixations of slightly deleterious mutations.

Despite having clear theoretical predictions, the importance of HRI in shaping protein evolution remains unclear (Webster and Hurst 2012; Cutter and Payseur 2013). In fact, as pointed out in a recent review, empirical studies have documented "extreme disparities among species" (Cutter and Payseur 2013). In *Drosophila melanogaster*, ω in regions that lack recombination (e.g., the fourth chromosome) is significantly higher than in regions where crossing-over occurs, consistent with relaxed purifying selection, but there is little difference in ω between regions with high, intermediate, and low crossover frequencies (Haddrill et al. 2007; Larracuente et al. 2008). In a recent analysis of the Drosophila Population Genomics Project data (Campos et al. 2014), it was found that the efficacy of natural selection, both positive and purifying, increases with local recombination rate, which may explain the lack of difference in ω within crossover regions, if the differential effects of positive and purifying selection on substitution rates at selected sites (the former elevates the rate and the latter depresses it) balance out (Campos et al. 2014). In contrast, in humans, recombination rate and ω were found to be uncorrelated, and there is little evidence that the efficiency of selection varies across regions with different recombination frequencies. This may be partly be explained by the small N_e in humans ($\approx 10^4$), which may render a general reduction in efficacy of selection, which in turn makes detecting the effects of HRI harder than in species with large N_e such as *Drosophila* ($\approx 10^6$; Bullaughey et al. 2008). As a third example, ω and recombination rate were found to be negatively correlated in yeast (e.g., Connallon and Knowles 2007; Cutter and Moses 2011). However, this pattern appears to be mediated by variation in gene expression, whereby slow-evolving, highly expressed genes tend to be located in high-recombination regions. After controlling for differences in expression, no evidence of substantial variation in selection efficacy across the yeast genome was found (Pál et al. 2001; Weber and Hurst 2009). These disparities between species call for analysis of data from species with different N_e and/or recombinational landscape, so that the HRI theory can be further tested and missing elements in the existing models identified (Webster and Hurst 2012; Cutter and Payseur 2013).

There are approximately 10,000 species in the class Aves (Jetz et al. 2012). Understanding how genome evolution occurs in this group of organisms has been an important topic in evolutionary genetics (reviewed by Ellegren 2013). In light of the discussion presented above, comparative genomic analysis of avian genomes will help to understand what factors, especially those characteristic of birds, correlate with ω, and whether these correlations are comparable to those observed in other species. Despite recent progress (reviewed by Ellegren 2013), important questions remain. For instance, it is unknown whether gene compactness (i.e., intron number, intron length, and protein length; reviewed by Choi and

Hannenhalli 2013) is correlated with rates of protein evolution. It is also unclear whether genes situated in different genomic locations (e.g., subtelemeric versus central regions of macrochromosomes) tend to have different average specificity in gene expression. Answers to these questions are important for the study of HRI.

Avian genomes have a rather similar karyotype, with the number of chromosomes being almost constant across species (Griffin et al. 2007; Ellegren 2010; Skinner and Griffin 2012). A typical avian genome contains 40 pairs of chromosomes, of which, depending on definition, around a dozen are large macrochromosomes, and the remainder are microchromosomes. The lengths of macro- and microchromosomes can differ by more than an order of magnitude, which is substantially more than in, for example, mammals (reviewed by Ellegren 2010). Because at least one crossing-over per chromosome is needed for proper segregation during meiosis, a consequence of the large difference in chromosome size is that microchromosomes have substantially higher average recombination rates ($\approx$10 cM/Mb) than macrochromosomes (0.5–2 cM/Mb), which has been confirmed by analyses of genetic maps in several birds (Stapley et al. 2008, 2010; Groenen et al. 2009; Backström et al. 2010; van Oers et al. 2014). This is more variable than average recombination rates observed in humans (1.07-2.10 cM/Mb; Jensen-Seaman et al. 2004). These genetic maps also reveal that, within macrochromosomes, the distribution of recombination frequency is nonuniform, with the majority of recombination events clustered in small regions close to telomeres. Although similar "telomere effects" have also been observed in other organisms such as humans (Jensen-Seaman et al. 2004), the clustering in birds appears to be stronger. For instance, the recombination rate drops very close to zero in regions more than 15 Mb away from the telomeres of zebra finch macrochromosomes (Backström et al. 2010; Stapley et al. 2010).

It has been suggested that HRI has probably played a role in driving the negative correlation between recombination rate (which is inversely related to chromosome size at a broad scale) and ω in birds (Axelsson et al. 2005; Nam et al. 2010; Künstner et al. 2010; Balakrishnan et al. 2013). However, the relative importance of positive and purifying selection to this observation is unknown. To provide better support for the HRI model, we intend to test 1) whether the elevation of ω in low-recombination regions is due to relaxed purifying selection, instead of enrichment of fast-evolving genes driven by positive selection and 2) whether positively selected genes are more likely to be found in high-recombination regions. In light of the highly variable recombinational landscape within macrochromosomes (Backström et al. 2010; Stapley et al. 2010; van Oers et al. 2014), it is essential to consider subtelomeric (i.e., ends) and central regions separately, which have high and low recombination frequencies, respectively.

Here, we focus on sequence divergence in protein-coding regions between two passerine birds, zebra finch and great tit (*Parus major*), with the latter being a model organism for addressing key topics in evolutionary ecology (Drent et al. 2003; Visser et al. 2003; Bouwhuis et al. 2010). By making use of a multitissue transcriptome data set in great tits (Santure et al. 2011) and the zebra finch genome (Warren et al. 2010), we seek to address the following questions: 1) How do variables such as tissue specificity in gene expression, intron number, intron length, and protein length correlate with ω? 2) do genes specifically expressed in different tissues evolve at different rates compared with other genes? and 3) is the efficacy of natural selection higher in regions with more frequent recombination, as predicted by the HRI theory?

Materials and Methods

Pairwise Sequence Alignments

The great tit transcriptome sequencing data were obtained from Santure et al. (2011). Briefly, in that article, normalized cDNA was sequenced from eight tissues; cDNA was pooled from ten different birds, all from Wytham Woods (Oxfordshire, UK). We focused on 95,979 assembled contigs with four or more reads. Because the contigs may contain noncoding sequences originating from pre-mRNA, UTRs, and other genomic parts, (e.g., due to leaky expression, Santure et al. 2011), we identified coding regions by mapping the contigs to cDNA of an outgroup species. We obtained outgroup information for *Gallus gallus* (chicken), *Taeniopygia guttata* (zebra finch), *Anas platyrhynchos* (mallard duck), *Ficedula albicollis* (collared flycatcher), *Meleagris gallopavo* (turkey), and *Melopsittacus undulatus* (budgerigar) from Ensembl (Flicek et al. 2012) and *Geospiza fortis* (medium ground finch) from Zhang et al. (2012). We used a nucleotide-based alignment strategy to map the great tit contigs to the corresponding regions of the outgroup genomes. First, we conducted a whole-genome BLAT (Kent 2002) search of the contigs against the cDNA of the outgroup species. For each pairwise BLAT hit, we obtained a pairwise alignment using bl2seq from BLASTALL (Altschul et al. 1990) and extracted from this alignment the longest ORF (minimum size 300 nucleotides) based on the outgroup sequence. The corresponding protein sequence of great tit was obtained by adjusting for frameshifts and stop codons, which were masked using PAL2NAL (Suyama et al. 2006). Input files for PAML (Yang 2007) were generated. We used rumode = −2 with the F3x4 codon model to obtain d_n/d_s ratios using the codeml program of the PAML suite. Because one contig can have hits in multiple outgroup loci, we used the hit with the lowest d_s value. We also discarded hits for which the overall substitution rate was too high (tree length > 1.2, which is likely the effect of incorrect alignments). If one outgroup locus had several great tit hits, we combined the longest nonoverlapping stretches to obtain d_n/d_s for this locus.

Multiple Sequence Alignments and PAML Analysis

To conduct site-specific analyses of substitution rates, we used sequence triplets (3-way alignments) of chicken, zebra finch, and the great tit contigs. First, we identified homologous genes between chicken and zebra finch using Inparanoid (Remm et al. 2001; Ostlund et al. 2010). We then excluded those ortholog pairs, which either did not map or inconsistently mapped to the great tit contigs based on the pairwise sequence alignments. The remaining sequence triplets were aligned using MUSCLE (Edgar 2004). Uncertain sequence positions were removed based on scores from ZORRO (Wu et al. 2012), and the final alignment was processed with PAL2NAL (Suyama et al. 2006). The resulting alignments were used as input for PAML (Yang 2007). Because alignment errors may affect the downstream substitution rate analysis, to check the robustness of our results, we also applied a different alignment strategy using PRANK (Löytynoja and Goldman 2005) and GUIDANCE (Landan and Graur 2008; Penn et al. 2010). Only a small proportion of alignments were different between the two alignment strategies, and there was no difference in alignment inconsistencies between recombination jungles and deserts (G test, high vs. low/very low recombination rate and inner macrochromosomes vs. outer macrochromosomes and microchromosomes, $P=0.59$ and $P=0.51$), suggesting that alignment quality is only a minor issue in our case.

PAML uses a maximum likelihood approach to obtain substitution rate estimates for the provided phylogeny based on certain model assumptions. To obtain ω ($= d_n/d_s$) for each sequence triplet, we assumed a constant ω across the tree (model M0, one-ratio model). We also conducted site tests to identify heterogeneity in ω within genes (but not heterogeneity between branches). For this, the likelihood of a more complex model was compared with a nested simpler model, and significance was assessed using a likelihood ratio test. To test for evidence of positive selection, we compared the site models M7 and M8 (Yang et al. 1998, 2000). M7 and M8 assume that ω among sites follows a $\beta(0, 1)$ distribution, but M8 additionally allows for sites with $\omega > 1$. To identify genes with evidence for positive selection, we used a likelihood ratio test comparing M7 and M8 assuming a χ^2 distribution with df $= 2$ and corrected for multiple testing using the method of Benjamini and Hochberg (1995) with false discovery rates (FDRs) ranging from 10% to 50%. Approximately 90% (depending on the applied FDR) of genes under positive selection based on the PRANK alignments are also detected when using MUSCLE alignments, suggesting that the majority of positively selected genes are consistent between the two alignment approaches. To test for the role of purifying selection for genes that did not show evidence of positive selection, we extracted parameter estimates from model M1a in PAML (Yang 2007), which allows a proportion of sites to be neutrally evolving (i.e., $\omega = 1$), and the remaining sites to be subject to purifying selection (i.e., $\omega < 1$).

Physical Position and Estimates of Recombination Rate

Linkage maps constructed in several birds have consistently shown that 1) microchromosomes tend to have much higher per-site recombination rate than macrochromosomes and that 2) for macrochromosomes, most of the recombination events take place in subtelomeric regions with a large sections of the inner part of these chromosomes with much lower recombination rates (the telomere effect; Groenen et al. 2009; Backström et al. 2010; Stapley et al. 2010). We inferred the physical position of each gene using the zebra finch genome and classified genes into three categories: Microchromosome (chromosomes 13–28), macrochromosome with telomeric location (chromosomes 1–12 and Z within three megabases from the chromosome tip), and macrochromosome within the inner 25% of the total chromosome length. Our definition of microchromosomes followed that of Backström et al. (2010), which was based on the observation that recombination rates of chromosomes less than 20 Mb in length appear to be high (i.e., comparable to subtelomeric regions of macrochromosomes) and uniform across the length of the chromosome.

Even though the karotype within birds is relatively stable (Griffin et al. 2007), there were two major chromosomal fission and fusion events along the chicken and passerine lineages. We therefore excluded genes located at the beginning of chromosomes 1 and 4 as well as genes located at the end of chromosome 1A, where beginning and end are defined according to the zebra finch genetic map (Stapley et al. 2008). We also excluded genes on chromosome 4A, which is a microchromosome in the passerines but part of chromosome 4 (which is large) in chicken. We also classified genes according to the local recombination rates inferred by comparing the physical map with the genetic map in zebra finch (Backström et al. 2010). We defined three categories of genes according to their estimates of recombination rate as follows: Very low recombination (regions with no detected recombination events), low recombination (lower 25% of genes with nonzero recombination rate estimates), and high recombination (upper 75% of nonzero recombination rate genes).

Extraction of Gene Features

We retrieved information on expression specificity for each gene from Santure et al. (2011) who followed the approach of Mank et al. (2008) to account for small levels of undetected expression. Expression specificity is measured by τ (Yanai et al. 2005), which ranges from 0 for genes with equal expression in all tissues to 1 for highly biased genes for whom most transcripts were found in only one tissue. Expression for each contig was standardized to the number of reads per million (TPM, see Santure et al. 2011 for details) and τ was calculated as follows;

$$\tau = \frac{\sum_{i=1}^{N}\left(1 - \frac{\ln(\mathrm{TPM}_i)}{\ln(\mathrm{TPM}_{\max})}\right)}{N-1}$$

where N is the number of tissues, TPM_i is the level of expression in tissue i, and $\mathrm{TPM}_{\max}$ is the highest level of expression of a given contig over all i tissues examined. The number and size of introns, the size of exons, gene density (proportion of coding sites per Mb), and the chromosome size were inferred from the zebra finch genome (Warren et al. 2010). The proportion of sites near intron–exon boundaries was calculated as the number of introns divided by protein length.

Statistical Analysis

The statistical package R was used to carry out statistical tests and generate box plots (using boxplot.stats function with default parameters). In a box plot, the box represents the range between upper and lower quartiles, the horizontal line within the box shows the median, and the whiskers show the most extreme data point, which is no more than 1.5 times the length of the box away from the box. To test for enrichment of genes with different gene ontology (GO) classifications, we used goatools (https://github.com/tanghaibao/goatools, last accessed July 28, 2014).

Results

Mapping the Great Tit Transcriptome to Other Avian Genomes

We investigated rates of protein evolution by using a great tit (*P. major*) transcriptome data set based on RNA extracted from eight different tissues (brain, heart, kidney, liver, muscle, pancreas, skin, and testis/ovary; Santure et al. 2011). We mapped the contigs assembled by Santure et al. (2011) to the seven bird species for which a reference genome is available (see Materials and Methods). The median d_s between zebra finch and great tit is ≈0.1 (fig. 1 and table 1; see also supplementary fig. S6, Supplementary Material online), comparable to estimates reported earlier between zebra finch and other passerine birds (Künstner et al. 2010; Backström et al. 2013; Balakrishnan et al. 2013). The median d_s values obtained from pairwise comparisons between great tit and each of the four nonpasserine birds are also shown in figure 1 and table 1 (see also supplementary fig. S6, Supplementary Material online). The observed levels of synonymous divergence are consistent with the phylogenetic relationship of these species (Hackett et al. 2008, table 1 and fig. 1). For ω (d_n/d_s) between zebra finch and great tit, the median and mean are ≈0.1 and ≈0.16, respectively (supplementary table S1 and fig. S6, Supplementary Material online), which is again fairly close to the values reported for other passerine birds (0.08–0.13, Künstner et al. 2010; Backström et al. 2013; Balakrishnan et al. 2013).

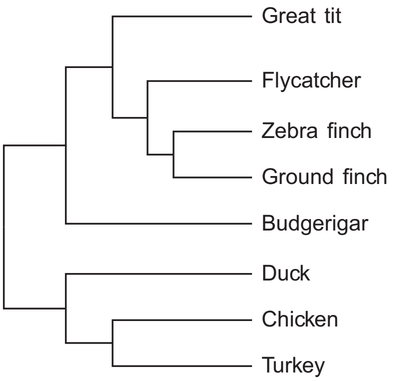

FIG. 1.—Phylogenetic relationship of great tit and seven bird species (Hackett et al. 2008).

Table 1
Median Estimates of d_s and d_n/d_s Based on Pairwise Alignment between the Great Tit Transcriptome and Each of the Seven Different Bird Species (fig. 1)

Genome Reference	d_s	d_n/d_s
Collared flycatcher (*Ficedula albicolis*)[a]	0.104	0.081
Zebra finch (*Taeniopygia guttata*)[a]	0.103	0.099
Medium ground finch (*Geospiza fortis*)[a]	0.111	0.083
Budgerigar (*Melopsittacus undulatus*)[b]	0.248	0.075
Mallard duck (*Anas platyrhynchos*)[b]	0.318	0.073
Chicken (*Gallus gallus*)[b]	0.339	0.074
Turkey (*Meleagris gallopavo*)[b]	0.346	0.071

[a] Passerine.
[b] Nonpasserine.

Because the quality of annotation is best for the zebra finch and chicken genomes, we used these two genomes as references. Specifically, we analyzed 8,294 two-way alignments between great tit and zebra finch; we were also able to obtain orthologous sequences from the chicken genome to construct three-way alignments for 5,460 genes. We focused on factors that may affect patterns of protein evolution between the two passerines, great tit and zebra finch.

Correlates of Variation in Rates of Evolution in Passerines

We explored pairwise relationship between several genomic features and evolutionary rates in passerines (ω and d_s) obtained from our two-way alignments, so as to identify predictors of evolutionary rates between the two passerines. Consistent with previous studies (Pál et al. 2006; Axelsson et al. 2008; Larracuente et al. 2008; Ekblom et al. 2010; Choi and Hannenhalli 2013), there is a highly significant positive correlation between ω and τ, a commonly used measure

Table 2

Pairwise Correlation Coefficients (Spearman's ρ) for Variables Covary with the Rates of Protein Evolution in Passerines

	d_n/d_s	GC3	τ	Intron Number	Intron Length	Protein Length	Chromosome Size	Gene Density
d_s	−0.04***	0.22***	0.10***	NS	−0.10***	NS	−0.10***	0.09***
$\omega = d_n/d_s$	–	−0.17***	0.06***	−0.07***	−0.14***	−0.05***	0.07***	NS
GC3		–	0.19***	−0.13***	−0.19***	−0.12***	−0.35***	0.29***
τ			–	0.06***	0.11***	0.10***	NS	NS
Intron number				–	0.66***	0.79***	−0.10***	0.13***
Intron length					–	0.57***	0.04***	−0.11***
Protein length						–	−0.09***	0.11***
Chromosome size							–	−0.55***

NOTE.–d_s and d_n/d_s were estimated using pairwise alignments between great tit and zebra finch. GC3, GC content at third codon position; τ, expression specificity; NS, not significant.

***$P < 0.001$.

Table 3

Partial Correlation Analyses Based on Kendall's τ to Investigate the Effect of Variation in Various Covariates on the Correlation between the Two Variables of Interest

Case	Variables	Covariates	Kendall's τ
1	ω, τ	GC3	0.047***
2	ω, τ	GC3, read depth	0.033***
3	ω, intron length	GC3, chromosome size	−0.031***
4	ω, intron number	GC3, chromosome size	−0.098***
5	ω, protein length	GC3, chromosome size	−0.028**
6	d_s, chromosome size	GC3	−0.031***
7	ω, genomic location[a]	GC3, gene density, τ, intron number, protein length	0.02*[1]
8	ω, genomic location[b]	GC3, gene density, τ, intron number, protein length	0.023[NS]
9	ω,[c] genomic location[a]	GC3, gene density, τ, intron number, protein length	0.075***
10	p,[d] genomic location[a]	GC3, gene density, τ, intron number, protein length	0.047***
11	Δln L,[e] genomic location[a]	GC3, gene density, τ, intron number, protein length	−0.13*[2]

NOTE.—NS, not significant.

[a] Outer parts of macrochromosomes and microchromosomes versus inner parts of macrochromosomes.

[b] Outer parts of macrochromosomes versus microchromosomes.

[c] ω at nonneutral sites (nearly neutral model M1a, fig. 3).

[d] Proportion of neutral sites (nearly neutral model M1a, fig. 3).

[e] Log-likelihood difference (model M7 vs. M8, test for positive selection, genes with $P < 1.0$).

***$P \leq 0.001$.

**$P \leq 0.01$.

*[1] $P = 0.032$.

*[2] $P = 0.027$

[NS] $P > 0.05$.

of tissue specificity (Yanai et al. 2005), which ranges from 0 (equal expression in all tissues) to 1 (highly biased expression with most transcripts found in only one tissue) (table 2). Because both τ and ω are significantly correlated with GC3 (GC content at 3rd codon positions), it is possible that the relationship between ω and τ is simply a by-product of these correlations. However, a partial correlation analysis suggests that this is not the case and that ω and τ are significantly positively correlated after variation in GC3 was controlled for (table 3, Case 1). To further rule out the possibility that the pattern is driven by a small number of genes with very high sequencing coverage (which may therefore have more accurate estimates of τ and potentially fewer assembling/sequencing errors), we introduced read depth as a second covariate and found that the positive correlation remains significant (table 3, Case 2).

The pairwise relationship between ω and several measures of gene compactness including the number of introns, total length of introns, and the length of the protein sequence are all significantly negative, in agreement with previous analysis of nonavian species (table 2, Larracuente et al. 2008; Choi and Hannenhalli 2013). Partial correlation analyses suggest (table 3, Cases 3–5) that none of these correlations were driven by of the apparent pairwise covariation between these gene features with GC3 (which covariates with ω) and chromosome size (which covariates with measure of gene compactness).

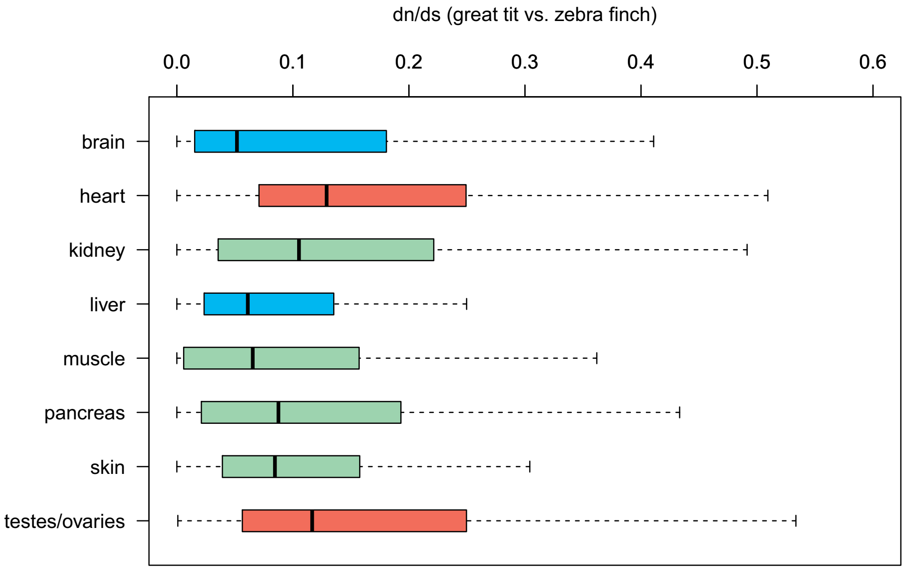

Fig. 2.—Boxplots of evolutionary rates for subsets of genes specifically expressed in certain tissues; boxes in blue and red denote significantly reduced and increased d_n/d_s values, respectively. Whiskers were drawn as implemented in the R-function boxplot (see Materials and Methods).

The positive relationship between intron length and chromosome size is consistent with the fact that microchromosomes are more compact (International Chicken Genome Sequencing Consortium 2004; Nam and Ellegren 2012). We also found a negative correlation between ω and the proportion of sequences near exon–intron boundaries (Spearman ρ = −0.06 and $P < 0.001$), as reported in humans (Parmley et al. 2007).

Our analysis of the two-way alignments also unearths the following patterns, which have been reported in earlier studies of other bird genomes, confirming the high quality of our data and the generality of these patterns (Axelsson et al. 2005; Nam et al. 2010; Künstner et al. 2010; Balakrishnan et al. 2013; reviewed by Ellegren 2013). First, smaller chromosomes tend to have higher divergence at synonymous sites (table 2, d_s versus chromosome size). Interestingly, the correlation between d_s and chromosome size remains significantly negatively correlated after controlling for GC3 (table 3, Case 6), implying that the correlation was not entirely due to the positive relationship between GC3 and substitution rates (table 2; Webster et al. 2006). It is possible that smaller chromosomes may have higher mutation rates (Axelsson et al. 2005; Nam et al. 2010; Künstner et al. 2010). Secondly, there is a significant positive relationship between ω and chromosome size (table 2). This pattern is unlikely to be driven by the fact that smaller chromosomes tend to have higher gene density, as ω and gene density are not statistically correlated (table 2); nor does it seem probable that gene expression specificity, which is uncorrelated with chromosome size (table 2), has played a major role. In a later section, we will go beyond previous studies and investigate whether the dramatic variation in recombination rate among different genomic regions has contributed to this correlation through the process of HRI.

Heterogeneity in ω between Genes Involved in Different Biological Functions

We have shown that the rate of molecular evolution varies substantially in passerine birds across the genome. It is, however, unclear whether genes with issue-specific expression also have different ω's. To answer this question, we extracted tissue-specific genes for the eight tissues included in the transcriptome sequencing, and compared their ω values (fig. 2). The median ω value for genes specifically expressed in the brain is 0.051, which is significantly lower than the genome-wide median of 0.1 (Mann–Whitney U test [MWU], $P = 0.005$) and the median value of other tissue-specific genes (MWU, $P = 0.021$). This could be explained by the theory put forward by Drummond and Wilke (2008, see Introduction). In contrast, testis/ovary-specific genes have significantly increased evolutionary rates when compared with other tissue-specific genes (MWU, $P = 0.027$), consistent with the intra- and interspecific arms race theory. Interestingly, genes specifically expressed in the heart had the highest median ω = 0.129, which is comparable to that of the testis/ovary-specific genes (MWU, $P = 0.15$). This is different from humans whose heart-specific genes have significantly lower median ω (0.07) than testis-specific genes (0.103) (Winter et al. 2004). The cause of this difference is unclear and warrants investigation in future.

We further tested whether genes with very low (smaller than the 10th percentile, likely to contain many genes under strong selective constraints) or very high (larger than the 90th percentile, likely to include genes either under relaxed constraints or fast-evolving genes driven by recurring episodes of positive selection) ω values are enriched for particular GO categories. Seven GO terms have significantly more low-ω genes than expected by random chance (Fisher's exact test with Bonferroni correction $P < 0.01$, supplementary table S2, Supplementary Material online); these include genes involved in core cellular functions such as ribosomal complexes or metabolic regulation. Genes associated with at least one of these seven GO terms tend to have lower expression specificity (MWU, $P < 0.001$), which is expected for housekeeping genes (Zhang and Li 2004). For genes with elevated ω, we did not observe significant over- or underrepresentation in any GO terms. However, in light of the potential problems of GO-based analysis (Pavlidis et al. 2012), these results should be regarded as exploratory.

Evidence of More Effective Natural Selection in Regions with Frequent Recombination

We first tested whether variation in local recombination rates contributes to the covariation between ω and chromosome size. Given the pronounced telomere effect observed in macrochromosomes and the substantial differences in average recombination rate between macro- and microchromosomes, we defined three sets of genes: 1) genes in central parts of macrochromosomes (low recombination frequency); 2) genes located near ends of macrochromosomes (i.e., subtelomeric regions, highly recombining); and 3) genes in microchromosomes (highly recombining). The ends of macrochromosomes and microchromosomes are often referred to as recombination jungles, and central parts of macrochromosomes as recombination deserts (e.g., Backström et al. 2010). Using the MWU, we found that the genes in recombinations deserts have significantly higher ω values than the other two sets (fig. 3*a*; $P = 0.002$). Because the MWU cannot control for the effects of covariates, we used a partial correlation method to test whether ω was positively correlated with a genomic location variable, which took the value of 0 or 1 for genes located in recombination jungles or deserts, respectively. We chose GC3, gene density, expression specificity, intron number, and protein length as covariates but did not consider chromosome size. This is because recombination jungles and deserts were defined mainly in light of the fact that microchromosomes have, on average, much higher recombination rates per base pair than macrochromosomes, but this relationship would disappear if chromosome size was held constant. After controlling for covariates, ω was found to be significantly lower in recombination jungles (table 3, Case 7). Interestingly, ω is not statistically different between the two high-recombination sets (fig. 3*a*; MWU, $P = 0.13$), and this remains the case when covariates were controlled for (table 3, Case 8).

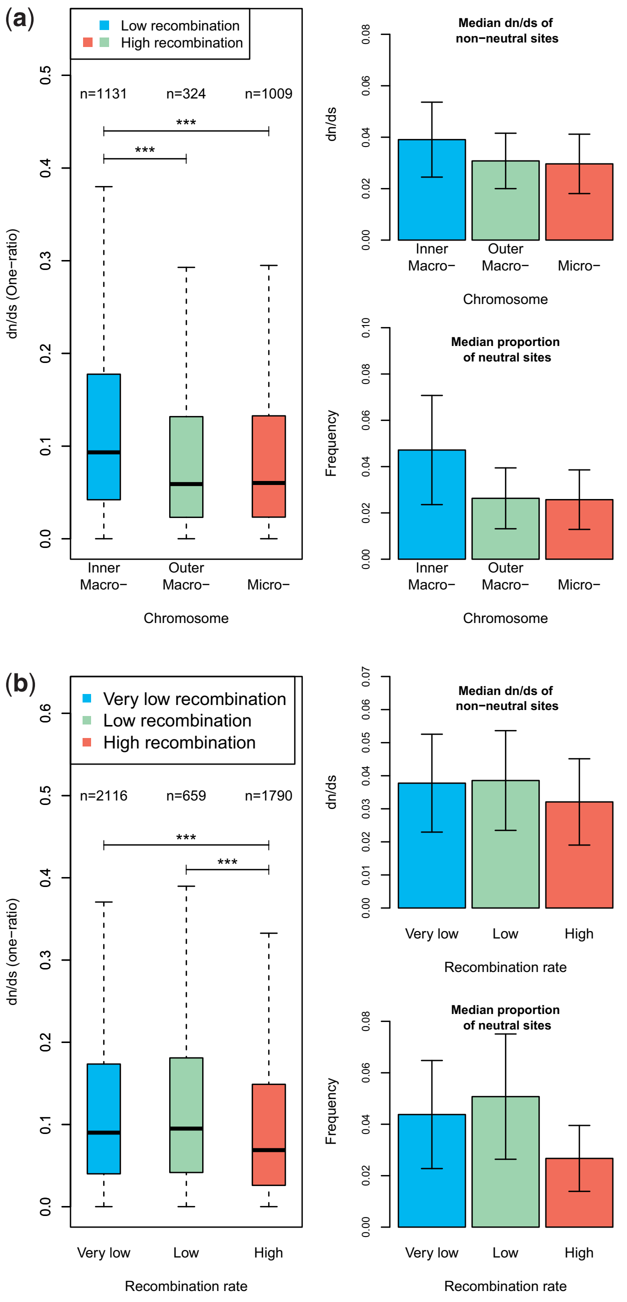

Fig. 3.—Box plots of evolutionary rates for subsets of genes according to their (*a*) chromosomal positions and (*b*) recombination rate estimates. Whiskers are drawn as implemented in the R-function box plot (see Materials and Methods). The bar plots show the proportion of sites estimated to be evolving under neutrality and the median d_n/d_s value for sites inferred to be evolving under purifying selection by the nearly neutral (model M1a) in PAML. Error bars indicate median absolute deviations (MDA). ***$P \leq 0.001$ under the MWU.

To check whether the above results are robust to how regions with frequent and infrequent recombination are defined, we estimated local recombination rates by comparing the zebra finch genetic map with its reference genome, and defined high-, low-, and very low-recombination regions (Backström et al. 2010; Warren et al. 2010, see Materials and Methods). Genes in high-recombination regions show a reduction regarding their median ω value when compared with either very low- or low-recombination regions (MWU, $P=0.001$ and $P=0.002$, respectively), which is consistent with the pattern found above. No significant difference in ω was found between regions with low and very low recombination rate estimates (MWU, $P=0.2$). Given the fact that the majority of the genes estimated to have low- and very low-recombination rates were not located at the ends of macrochromosomes or microchromosomes (2,607 out of 2,767, $\approx$94%), these results suggest that ω and local recombination rates are negatively correlated and that the overall difference in ω between macro- and microchromosomes (e.g., ω versus chromosome size in table 2) may be in part driven by high ω in low-recombination regions of macrochromosomes (i.e., recombination deserts).

Next we asked whether purifying selection is less effective in regions with infrequent recombination, as predicted by the HRI theory (see Introduction). To increase statistical power, we used three-way alignments including orthologous genes from the chicken genome in this analysis. We also excluded genes showing evidence of positive selection according to a "site model" implemented in PAML (at an FDR level of 10%; see Materials and Methods). We used PAML to estimate, for each locus, 1) the proportion of neutrally evolving sites and 2) ω at sites that are under purifying selection. As above, genes were assigned to three different groups according to their recombination rates. Compared with genes found in the recombination jungles, genes located in recombination deserts have a significantly higher proportion of neutral sites (fig. 3*a*; MWU, $P=0.001$ and $P<0.001$ for comparisons with ends of macrochromosomes and microchromosomes, respectively) and significantly higher ω at sites under purifying selection (fig. 3*a*; MWU, $P=0.002$ and $P<0.001$). After controlling for the covariates mentioned above, ω for nonneutral sites and the proportion of neutral sites were found to be significantly lower in recombination jungles (table 3, Cases 9 and 10)

Similar patterns can be seen when genes in either low- or very low-recombination regions were compared with those in high-recombination regions (fig. 3*b*; $P<0.005$ for all comparisons). These patterns are consistent with relaxed selective constraints in regions where recombination is infrequent. Interestingly, microchromosomes and ends of macrochromosomes have very similar median values of the proportion of neutral substitutions and ω for nonneutral substitutions (fig. 3*a*; MWU, $P>0.1$). Similarly, no statistically significant difference was found between low- and very low-recombination regions (fig. 3*b*; MWU, $P>0.1$).

Finally, we examine whether positive selection is also more efficient in regions with higher recombination frequencies. Support for positive selection was determined by using a site model implemented in PAML (see Materials and Methods). Among the 1,333 genes in recombination jungles, ten show evidence for positive selection, which is significantly more frequent than genes in recombination deserts where 2 out of 1,131 genes have experienced positive selection (table 4; with an FDR level of 10%; G test, $P=0.03$). Because the G test cannot take into account covariates, we carried out the following analyses. If high local recombination rates facilitate the fixation of beneficial variants, then the M8 model should, on average, fit the data better than the M7 model, and therefore difference in ln likelihood between the two models ($\Delta \ln L$) should be larger in high-recombination regions. Indeed, controlling for covariates, model M8, which includes positive selection, fitted the data from high-recombination regions better (table 3, Case 11).

Table 4

Location of Genes with Evidence for Positive Selection

FDR	Recombination Region	Positively Selected	Not Positively Selected	*P* (G Test)
10%	High	8	1,782	
	Low	0	659	0.02
10%	Jungle[a]	10	1,323	
	Desert[b]	2	1,129	0.03

Note.—Genes were classified according to their genomic locations (recombination jungles and deserts) or their estimated local recombination rate (high- and low/very low recombination, see Materials and Methods). Comparisons were conducted between genes located in different recombinational environments using G tests.

[a] Outer parts of macrochromosomes and microchromosomes.

[b] Inner parts of macrochromosomes.

A similar enrichment of positively selected loci is also found in high-recombination regions, relative to low recombination regions (table 4; G test, $P=0.02$). These results (as well as those presented earlier in this section) are robust to different definitions of regions with different recombination frequencies and different FDR thresholds (supplementary table S3, Supplementary Material online) and the use of a different combination of sequence aligner and alignment processing algorithm (PRANK and GUIDANCE; see Materials and Methods; supplementary table S4, Supplementary Material online) or a more robust but less powerful model comparison to detect positive selection (PAML; M1a vs. M2a, supplementary table S5, Supplementary Material online). Thus, in agreement with the HRI theory, elevated local recombination reduces interference between linked selected sites and facilitates both the spread of beneficial mutations and the removal of deleterious mutations.

Discussion

In this study, we used a multitissue transcriptome data set in great tits, together with the reference genomes of the zebra finch and the chicken, to study patterns of molecular evolution along the two passerine lineages. By contrasting patterns of sequence divergence between genes in high- and low-recombination regions of the genome and by analyzing genes with and without evidence of positive selection separately, we obtained evidence that the efficacy of both positive and negative selection is higher in regions with more frequent recombination, as predicted by the HRI theory (fig. 3 and table 4). We also showed that more compact genes with fewer introns, shorter introns, and shorter proteins tend to evolve faster (table 2) and that genes with a larger proportion of exon–intron boundaries have lower ω. The latter two results have not previously been examined in birds (Parmley et al. 2007; Choi and Hannenhalli 2013). These analyses demonstrate that transcriptome sequencing is a powerful way to address fundamental questions about genome evolution in organisms such as great tits where genomic resources are relatively limited.

Gene Expression Pattern as a Major Predictor of Protein Evolution

Gene expression pattern can be viewed as encompassing both expression level and tissue specificity of expression. Although it is well known that gene expression level is a key predictor of ω (reviewed by Pál et al. 2006 and Choi and Hannenhalli 2013), this factor was not considered in this study because our cDNA libraries were normalized prior to sequencing (Santure et al. 2011), which means that read depth is probably an unreliable measure of gene expression level. Nonetheless, when read depth was used as a proxy of expression level, we did find a significant negative correlation between ω and expression level (Spearman's $\rho = -0.06$, $P < 0.001$), which is in the same direction as reported previously in many organisms (reviewed by Pál et al. 2006 and Choi and Hannenhalli 2013).

Our finding of a positive correlation between tissue specificity in expression (τ) and ω should be conservative in the presence of library normalization, as this procedure suppresses differences in expression level, and is therefore expected to homogenize differences between tissues (Ekblom et al. 2012). This may explain why the pairwise correlation between τ and ω reported in table 2 is somewhat weaker than those presented in previous studies (e.g., 0.3 in *Drosophila*, 0.24 in mice, 0.12 in humans; Larracuente et al. 2008; Park et al. 2012). Evidence for this homogenization is provided by an analysis conducted on contigs assembled from at least 50 reads (combined across all tissues), the correlation increased to 0.11, which may reflect that estimates of tissue specificity were more accurate when more reads were available. A possible biological explanation of the relationship between τ and ω is increased pleiotropy: Proteins that are expressed in many tissues may tend to have more interacting partners, which lead to more constraints on the function and/or structure of the protein, and a corresponding reduction in evolutionary rate (Pál et al. 2006). There is evidence that τ and expression level are highly correlated in some species (Lercher et al. 2002; Subramanian and Kumar 2004) including birds (Ekblom et al. 2010). In mammals, tissue specificity seems to explain more of the variation in rates of protein evolution than expression level (Pál et al. 2006). It is possible that the positive correlation we observe here may be in part driven by variation in gene expression level. Better data are needed to establish the relative importance of the breadth and level of expression in determining evolutionary rates in passerines.

Although genes with high tissue specificity in expression tend to evolve faster as a whole, the distribution of ω is highly heterogeneous among tissues (fig. 2). Specifically, genes expressed mainly in the brain have, on average, the lowest ω, which is consistent with findings in metazoans (Axelsson et al. 2008; Drummond and Wilke 2008), and is probably due to the fact that neuronal tissues are particularly sensitive to the damaging effects of mistranslation-induced protein misfolding. Therefore, these genes are under strong selective constraints because of the rarity of well-adapted sequences with high translational accuracy and robustness (Drummond and Wilke 2008). On the other hand, genes expressed in testis/ovary have accelerated rates of molecular evolution. This pattern, which has been observed in other organisms such as humans (Nielsen et al. 2005) and *Drosophila* (Zhang et al. 2004), can potentially be caused, either individually or in some combination, by sperm competition, sexual selection, and sexual conflict (Swanson and Vacquier 2002), all of which are common in birds (Birkhead and Moller 1992). Additional research is needed to clarify the relative importance of these factors in birds.

HRI as a Determinant of Protein Evolution in Birds

Our analysis of patterns of protein evolution in the two passerine lineages suggests that HRI is likely to have played an important role in determining variation in ω. In particular, regions with reduced recombination tend to be more prone to the accumulation of slightly deleterious substitutions, whereas the fixation of beneficial mutations is more likely to take place in high-recombination regions (fig. 3 and table 4). These results are insensitive to different definitions of high- and low-recombination regions and FDR cutoffs. Intrachromosomal rearrangements, which have occurred between the three bird species considered (reviewed by Ellegren 2010, 2014; see also van Oers et al. 2014), should not make the test counterconservative. This is because, for macrochromosomes, shuffling genes between the ends and the central parts is expected to homogenize differences in recombination frequency, whereas for microchromosomes, genetic maps in all species studied to date suggest recombination rates are roughly constant along the length of the chromosome (Backström et al. 2010; Stapley et al. 2010; van Oers et al. 2014).

There is evidence that GC content is not at statistical equilibrium in multiple avian lineages and that GC-biased gene

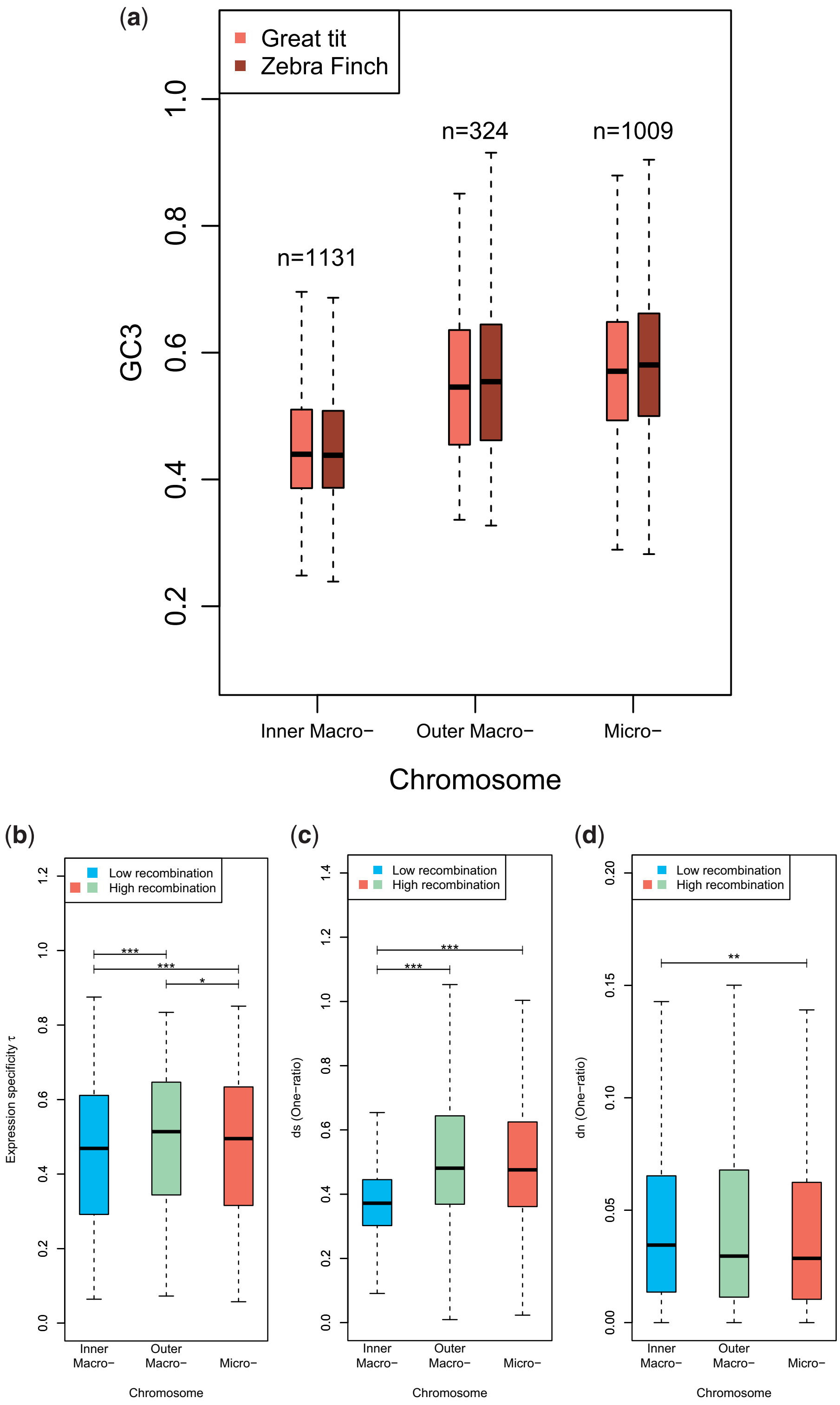

Fig. 4.—Box plots of (*a*) GC content at 3rd positions (GC3), (*b*) expression specificity τ, (*c*) d_s, and (*d*) d_n for subsets of genes according to their chromosomal positions. Whiskers are drawn as implemented in the R-function box plot (see Materials and Methods). $^{***}P \leq 0.001$, $^{**}P \leq 0.01$, $^{*}P \leq 0.05$.

conversion (gBGC) may have contributed to this (Webster et al. 2006; Nabholz et al. 2011; Mugal et al. 2013). However, our results should also be robust to variation in GC content and the action gBGC, which can upwardly bias estimates of ω and lead to false detection of targets of positive selection (reviewed by Duret and Galtier 2009). First, we used the site models in PAML, which analyzed substitution patterns over the entire phylogenetic tree in the search of positively selected genes. A recent analysis has shown that results produced by this approach are unlikely to be affected by gBGC (Ratnakumar et al. 2010). Second, if substitutions of slightly deleterious mutations were driven by gBGC (Galtier et al. 2009), we expect this effect to be stronger in regions with higher recombination rates and GC content, which are often used as proxies of the intensity of gBGC (Duret and Galtier 2009). Contrary to this prediction, in regions with reduced recombination, where ω is higher, GC content is lower (figs. 3 and 4*a*), and evidence of relaxed constraints on nonsynonymous sites also comes from these regions (fig. 4*d*).

There is little evidence that the systematic difference in ω displayed in figure 3 is driven by a similar difference in tissue specificity in gene expression (Case 7 in table 3). As shown in figure 4*b*, regions with frequent recombination actually have significantly higher τ than those with reduced recombination. A similar relationship between τ and recombination has been observed in humans (Necsulea et al. 2009) and *Drosophila* (Weber and Hurst 2011). We currently do not have data to ascertain whether gene expression level differ between these genomic regions. However, as mentioned above, if we were to assume that there is a strong negative correlation between τ and expression level (Lercher et al. 2002; Subramanian and Kumar 2004; Ekblom et al. 2010), then expression level would be lower in regions with frequent recombination and therefore would not be the main driver of the difference in ω.

Because ω is defined as the ratio of d_n to d_s, it may be inflated when d_s is unusually small either due to random chance or selective constraints on synonymous sites, resulting in false detection of positive selection. A recent analysis based on three-species alignments of chicken, turkey, and zebra finch has reported evidence that synonymous sites may be under significant constraints (Künstner et al. 2011). However, our results, which are based on comparisons of the number of selected genes detected by PAML between different classes of genes, should be robust. First, Künstner et al. (2011) found no evidence of regional variation in selective pressure on synonymous sites and suggested that difference in d_s between regions reflect variation in mutation rate. Second, the median d_s values of the positively selected genes (at an FDR level of 10%) and the other genes were 0.39 and 0.4, respectively, which were not significantly different (MWU, $P = 0.25$). Third, as shown in figure 4*c*, regions with high recombination rates tend to have much higher d_s than those with infrequent recombination. These three observations suggest that genes in high-recombination regions should not be more prone to false detection of positive selection. Hence, it seems unlikely that our results can be explained by selection at synonymous sites.

Our results therefore extend previous analyses of the negative correlation between recombination rate and ω in birds (Axelsson et al. 2005; Nam et al. 2010; Künstner et al. 2010; Balakrishnan et al. 2013) by showing that 1) higher ω in low-recombination regions is due to relaxed purifying selection rather than enrichment of fast-evolving genes driven by positive selection, 2) that frequent recombination facilitates the incorporation of new beneficial mutations, and 3) that microchromosomes and ends of macrochromosomes show very similar patterns of protein evolution as a consequence of frequent recombination. The positive relationship between the efficacy of selection and recombination rate appears to be consistent with the observation that diversity at putatively neutral sites (a proxy of local N_e; Charlesworth 2009) increases with local recombination rates in several birds (reviewed by Ellegren 2013). However, further research is needed to test whether evidence of HRI indeed exists both within and between species.

There are differences between the passerines and other organisms in terms of observations related to HRI. For instance, in *Drosophila*, differences in ω are most visible between regions that lack recombination (e.g., the fourth chromosome) and those where crossing-over occurs, whereas within the crossover regions, little difference in ω was found between regions with high, intermediate, and low crossover frequencies (Haddrill et al. 2007; Larracuente et al. 2008). However, in our case, ω appears to be more variable within the crossover regions, with regions with low, but nonzero, recombination rates having significantly higher ω than high-recombination regions (fig. 3*b*). The enrichment of targets of positive selection in high-recombination regions also contrasts with the lack of such enrichment in humans (Bullaughey et al. 2008). It is possible that the highly variable recombinational landscape in birds has made the effects of HRI more obvious across genomic regions. It is also possible that selection is more effective in birds than in humans, which may make it easier to detect HRI in the former (Bullaughey et al. 2008). Evidence of more effective selection in birds than in humans can be seen from the observation that a higher proportion of nonsynonymous substitutions in birds may be driven to fixation by positive selection than in humans (Eyre-Walker 2006; Axelsson and Ellegren 2009) and that birds have lower average ω (≈0.15) than humans (≈0.33; Zhang et al. 2013; supplementary table S1, Supplementary Material online). However, as pointed out by Cutter and Payseur (2013) (see also Connallon and Knowles 2007; Webster and Hurst 2012), predictions of HRI depend in a complicated way on parameters such as recombination rate, distribution of fitness effects of new mutations, and effective population size. More analysis is necessary to characterize these parameters in birds, which shall in turn facilitate comparisons with other species.

Supplementary Material

Supplementary tables S1–S4 and figures S1–S8 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors thank Judith Mank and three anonymous referees for their comments, which have helped to improve the quality of the manuscript. They thank Jessica Stapley for a helpful discussion on various aspects of bird genome evolution. This work was supported by a BBSRC grant (BB/K000209/1) awarded to K.Z. The transcriptome data collection was funded by an ERC Starting Grant, AvianEGG (grant number 202487) awarded to J.S.

Literature Cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215(3):403–410.

Axelsson E, Ellegren H. 2009. Quantification of adaptive evolution of genes expressed in avian brain and the population size effect on the efficacy of selection. Mol Biol Evol. 26(5):1073–1079.

Axelsson E, Webster MT, Smith NGC, Burt DW, Ellegren H. 2005. Comparison of the chicken and turkey genomes reveals a higher rate of nucleotide divergence on microchromosomes than macrochromosomes. Genome Res. 15(1):120–125.

Axelsson E, et al. 2008. Natural selection in avian protein-coding genes expressed in brain. Mol Ecol. 17(12):3008–3017.

Backström N, Zhang Q, Edwards SV. 2013. Evidence from a house finch (*Haemorhous mexicanus*) spleen transcriptome for adaptive evolution and biased gene conversion in passerine birds. Mol Biol Evol. 30(5): 1046–1050.

Backström N, et al. 2010. The recombination landscape of the zebra finch *Taeniopygia guttata* genome. Genome Res. 20(4):485–495.

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(9):130063.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B (Methodol). 57:289–300.

Birkhead T, Moller A. 1992. Sperm competition in birds: evolutionary causes and consequences. London: Academic Press.

Bouwhuis S, Charmantier A, Verhulst S, Sheldon BC. 2010. Trans-generational effects on ageing in a wild bird population. J Evol Biol. 23(3): 636–642.

Bullaughey K, Przeworski M, Coop G. 2008. No effect of recombination on the efficacy of natural selection in primates. Genome Res. 18(4): 544–554.

Cai JJ, Petrov DA. 2010. Relaxed purifying selection and possibly high rate of adaptation in primate lineage-specific genes. Genome Biol Evol. 2: 393–409.

Campos JL, Halligan DL, Haddrill PR, Charlesworth B. 2014. The relation between recombination rate and patterns of molecular evolution and variation in *Drosophila melanogaster*. Mol Biol Evol. 31:1010–1028.

Charlesworth B. 2009. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. Nat Rev Genet. 10(3):195–205.

Charlesworth B. 2012. The effects of deleterious mutations on evolution at linked sites. Genetics 190(1):5–22.

Charlesworth B, Charlesworth D. 2010. Elements of evolutionary genetics.

Choi SS, Hannenhalli S. 2013. Three independent determinants of protein evolutionary rate. J Mol Evol. 76(3):98–111.

Comeron JM, Williford A, Kliman RM. 2008. The Hill-Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. Heredity (Edinb) 100(1):19–31.

Connallon T, Knowles LL. 2007. Recombination rate and protein evolution in yeast. BMC Evol Biol. 7:235.

Cutter AD, Moses AM. 2011. Polymorphism, divergence, and the role of recombination in *Saccharomyces cerevisiae* genome evolution. Mol Biol Evol. 28(5):1745–1754.

Cutter AD, Payseur BA. 2013. Genomic signatures of selection at linked sites: unifying the disparity among species. Nat Rev Genet. 14(4): 262–274.

Drent PJ, van Oers K, van Noordwijk AJ. 2003. Realized heritability of personalities in the great tit (*Parus major*). Proc Biol Sci. 270(1510): 45–51.

Drummond DA, Wilke CO. 2008. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell 134((2):341–352.

Duret L, Galtier N. 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. Annu Rev Genomics Hum Genet. 10:285–311.

Duret L, Mouchiroud D. 2000. Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. Mol Biol Evol. 17(1):68–74.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5):1792–1797.

Ekblom R, Balakrishnan CN, Burke T, Slate J. 2010. Digital gene expression analysis of the zebra finch genome. BMC Genomics 11:219.

Ekblom R, Slate J, Horsburgh GJ, Birkhead T, Burke T. 2012. Comparison between normalised and unnormalised 454-sequencing libraries for small-scale RNA-seq studies. Comp Funct Genomics. 2012:281693.

Ellegren H. 2010. Evolutionary stasis: the stable chromosomes of birds. Trends Ecol Evol. 25(5):283–291.

Ellegren H. 2013. The evolutionary genomics of birds. Annu Rev Ecol Evol Syst. 44:239–259.

Eyre-Walker A. 2006. The genomic rate of adaptive evolution. Trends Ecol Evol. 21(10):569–575.

Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. Nat Rev Genet. 8(8):610–618.

Flicek P, et al. 2012. Ensembl 2012. Nucleic Acids Res. 40(Database issue): D84–D90.

Galtier N, Duret L, Glémin S, Ranwez V. 2009. GC-biased gene conversion promotes the fixation of deleterious amino acid changes in primates. Trends Genet. 25(1):1–5.

Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol. 11(5):725–736.

Griffin DK, Robertson LBW, Tempest HG, Skinner BM. 2007. The evolution of the avian genome as revealed by comparative molecular cytogenetics. Cytogenet Genome Res. 117(1–4):64–77.

Groenen MAM, et al. 2009. A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. Genome Res. 19(3):510–519.

Hackett SJ, et al. 2008. A phylogenomic study of birds reveals their evolutionary history. Science 320(5884):1763–1768.

Haddrill PR, Halligan DL, Tomaras D, Charlesworth B. 2007. Reduced efficacy of selection in regions of the *Drosophila* genome that lack crossing over. Genome Biol. 8(2):R18.

Haerty W, et al. 2007. Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. Genetics 177(3):1321–1335.

Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. Genet Res. 8(3):269–294.

Hodgkinson A, Eyre-Walker A. 2011. Variation in the mutation rate across mammalian genomes. Nat Rev Genet. 12(11):756–766.

International Chicken Genome Sequencing Consortium. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432(7018): 695–716.

Jensen-Seaman MI, et al. 2004. Comparative recombination rates in the rat, mouse, and human genomes. Genome Res. 14(4): 528–538.

Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO. 2012. The global diversity of birds in space and time. Nature 491(7424):444–448.

Kent WJ. 2002. BLAT–the BLAST-like alignment tool. Genome Res. 12(4): 656–664.

Kimura M. 1983. The neutral theory of molecular evolution. New York: Cambridge University Press.

Kosiol C, et al. 2008. Patterns of positive selection in six mammalian genomes. PLoS Genet. 4(8):e1000144.

Künstner A, Nabholz B, Ellegren H. 2011. Significant selective constraint at 4-fold degenerate sites in the avian genome and its consequence for detection of positive selection. Genome Biol Evol. 3:1381–1389.

Künstner 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(Suppl 1), 266–276.

Landan G, Graur D. 2008. Local reliability measures from sets of co-optimal multiple sequence alignments. Pac Symp Biocomput 15–24.

Larracuente AM, et al. 2008. Evolution of protein-coding genes in *Drosophila*. Trends Genet. 24(3):114–123.

Lercher MJ, Urrutia AO, Hurst LD. 2002. Clustering of housekeeping genes provides a unified model of gene order in the human genome. Nat Genet. 31(2):180–183.

Li WH. 1997. Molecular evolution. Sunderland (MA): Sinauer Associates Incorporated.

Löytynoja A, Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. Proc Natl Acad Sci U S A. 102(30):10557–10562.

Lynch M. 2010. Rate, molecular spectrum, and consequences of human mutation. Proc Natl Acad Sci U S A. 107(3):961–968.

Mank JE, Hultin-Rosenberg L, Zwahlen M, Ellegren H. 2008. Pleiotropic constraint hampers the resolution of sexual antagonism in vertebrate gene expression. Am Nat. 171(1):35–43.

Marais G, Nouvellet P, Keightley PD, Charlesworth B. 2005. Intron size and exon evolution in *Drosophila*. Genetics 170(1):481–485.

McVean GA, Charlesworth B. 2000. The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. Genetics 155(2):929–944.

Mugal CF, Arndt PF, Ellegren H. 2013. Twisted signatures of GC-biased gene conversion embedded in an evolutionary stable karyotype. Mol Biol Evol. 30(7):1700–1712.

Nabholz B, Künstner A, Wang R, Jarvis ED, Ellegren H. 2011. Dynamic evolution of base composition: causes and consequences in avian phylogenomics. Mol Biol Evol. 28(8):2197–2210.

Nam K, Ellegren H. 2012. Recombination drives vertebrate genome contraction. PLoS Genet. 8(5):e1002680.

Nam K, et al. 2010. Molecular evolution of genes in avian genomes. Genome Biol. 11(6):R68.

Necsulea A, Sémon M, Duret L, Hurst LD. 2009. Monoallelic expression and tissue specificity are associated with high crossover rates. Trends Genet. 25(12):519–522.

Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol. 3(5):418–426.

Nei M, Kumar S. 2000. Molecular evolution and phylogenetics. New York: Oxford University Press.

Nielsen R, et al. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol. 3(6):e170.

Obbard DJ, Welch JJ, Kim KW, Jiggins FM. 2009. Quantifying adaptive evolution in the *Drosophila* immune system. PLoS Genet. 5(10): e1000698.

Ostlund G, et al. 2010. InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. Nucleic Acids Res. 38(Database issue): D196–D203.

Park J, Xu K, Park T, Yi SV. 2012. What are the determinants of gene expression levels and breadths in the human genome? Hum Mol Genet. 21(1):46–56.

Parmley JL, Urrutia AO, Potrzebowski L, Kaessmann H, Hurst LD. 2007. Splicing and the evolution of proteins in mammals. PLoS Biol. 5(2):e14.

Pavlidis P, Jensen JD, Stephan W, Stamatakis A. 2012. A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans. Mol Biol Evol. 29(10):3237–3248.

Penn O, Privman E, Landan G, Graur D, Pupko T. 2010. An alignment confidence score capturing robustness to guide tree uncertainty. Mol Biol Evol. 27(8):1759–1767.

Pál C, Papp B, Hurst LD. 2001. Does the recombination rate affect the efficiency of purifying selection? The yeast genome provides a partial answer. Mol Biol Evol. 18:2323–2326.

Pál C, Papp B, Lercher MJ. 2006. An integrated view of protein evolution. Nat Rev Genet. 7(5):337–348.

Ratnakumar A, et al. 2010. Detecting positive selection within genomes: the problem of biased gene conversion. Philos Trans R Soc Lond B Biol Sci. 365(1552):2571–2580.

Remm M, Storm CE, Sonnhammer EL. 2001. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J Mol Biol. 314(5):1041–1052.

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.

Sella G, Petrov DA, Przeworski M, Andolfatto P. 2009. Pervasive natural selection in the *Drosophila* genome? PLoS Genet. 5(6): e1000495.

Skinner BM, Griffin DK. 2012. Intrachromosomal rearrangements in avian genome evolution: evidence for regions prone to breakpoints. Heredity (Edinb) 108(1):37–41.

Slotte T, et al. 2011. Genomic determinants of protein evolution and polymorphism in *Arabidopsis*. Genome Biol Evol. 3:1210–1219.

Smukowski CS, Noor MAF. 2011. Recombination rate variation in closely related species. Heredity (Edinb) 107(6):496–508.

Stapley J, Birkhead TR, Burke T, Slate J. 2008. A linkage map of the zebra finch *Taeniopygia guttata* provides new insights into avian genome evolution. Genetics 179(1):651–667.

Stapley J, Birkhead TR, Burke T, Slate J. 2010. Pronounced inter- and intrachromosomal variation in linkage disequilibrium across the zebra finch genome. Genome Res. 20(4):496–502.

Subramanian S, Kumar S. 2004. Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome. Genetics 168(1):373–381.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34(Web Server issue):W609–W612.

Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. Nat Rev Genet. 3(2):137–144.

van Oers K, et al. 2014. Replicated high-density genetic maps of two great tit populations reveal fine-scale genomic departures from sex-equal recombination rates. Heredity (Edinb) 112(3):307–316.

Visser ME, et al. 2003. Variable responses to large-scale climate change in European *Parus* populations. Proc Biol Sci. 270(1513): 367–372.

Warren WC, et al. 2010. The genome of a songbird. Nature 464(7289): 757–762.

Weber CC, Hurst LD. 2009. Protein rates of evolution are predicted by double-strand break events, independent of crossing-over rates. Genome Biol Evol. 1:340–349.

Weber CC, Hurst LD. 2011. Support for multiple classes of local expression clusters in *Drosophila melanogaster*, but no evidence for gene order conservation. Genome Biol. 12(3):R23.

Webster MT, Axelsson E, Ellegren H. 2006. Strong regional biases in nucleotide substitution in the chicken genome. Mol Biol Evol. 23(6): 1203–1216.

Webster MT, Hurst LD. 2012. Direct and indirect consequences of meiotic recombination: implications for genome evolution. Trends Genet. 28(3):101–109.

Winter EE, Goodstadt L, Ponting CP. 2004. Elevated rates of protein secretion, evolution, and disease among tissue-specific genes. Genome Res. 14(1):54–61.

Wu M, Chatterji S, Eisen JA. 2012. Accounting for alignment uncertainty in phylogenomics. PLoS One 7(1):e30288.

Yanai I, et al. 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics 21(5):650–659.

Yang Z. 2006. Computational molecular evolution, Vol. 284. Oxford: Oxford University Press.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586–1591.

Yang Z, Nielsen R, Goldman N, Pedersen AM. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155(1):431–449.

Yang Z, Nielsen R, Hasegawa M. 1998. Models of amino acid substitution and applications to mitochondrial protein evolution. Mol Biol Evol. 15(12):1600–1611.

Zhang C, Wang J, Long M, Fan C. 2013. gKaKs: the pipeline for genome-level Ka/Ks calculation. Bioinformatics 29(5):645–646.

Zhang G, Parker P, Li B, Li H, Wang J. 2012. The genome of Darwin's finch (*Geospiza fortis*). GigaScience, Available from: http://dx.doi.org/10.5524/100040.

Zhang J, Dean AM, Brunet F, Long M. 2004. Evolving protein functional diversity in new genes of *Drosophila*. Proc Natl Acad Sci U S A. 101(46):16246–16250.

Zhang L, Li WH. 2004. Mammalian housekeeping genes evolve more slowly than tissue-specific genes. Mol Biol Evol. 21(2): 236–239.

Associate editor: Judith Mank