Genomic evidence for a large-Z effect

Hans Ellegren*

Department of Evolutionary Biology, Uppsala University, Norbyvägen 18D, 752 36 Uppsala, Sweden

The ‘large-X effect’ suggests that sex chromosomes play a disproportionate role in adaptive evolution. Theoretical work indicates that this effect may be most pronounced in genetic systems with female heterogamety under both good-genes and Fisher’s runaway models of sexual selection (males ZZ, females ZW). Here, I use a comparative genomic approach (alignments of several thousands of chicken–zebra finch–human–mouse–opossum orthologues) to show that avian Z-linked genes are highly overrepresented among those bird–mammalian orthologues that show evidence of accelerated rate of functional evolution in birds relative to mammals; the data suggest a twofold excess of such genes on the Z chromosome. A reciprocal analysis of genes accelerated in mammals found no evidence for an excess of X-linkage. This would be compatible with theoretical expectations for differential selection on sex-linked genes under male and female heterogamety, although the power in this case was not sufficient to statistically show that ‘large-Z’ was more pronounced than ‘large-X’. Accelerated Z-linked genes include a variety of functional categories and are characterized by higher non-synonymous to synonymous substitution rate ratios than both accelerated autosomal and non-accelerated genes. This points at a genomic ‘large-Z effect’, which is widespread and of general significance for adaptive divergence in birds.

Keywords: large-Z; adaptive evolution; sex chromosomes

1. INTRODUCTION

It has been suggested that sex chromosomes play a disproportionately large role in adaptive evolution, including traits involved with reproduction and speciation (Dobzhansky 1974; Templeton 1977; Rice 1984; Coyne 1985; Charlesworth et al. 1987; Coyne & Orr 1989). The theoretical underpinnings to this so-called ‘large-X effect’ (or ‘Coyne’s rule’) are not clear (Presgraves 2008), although it may relate to the faster rate of sequence evolution of X-linked genes if mutations are favourable and partially or fully recessive (the ‘fast-X effect’). The hemizygous exposure of such mutations in males will imply that they are immediately subject to positive selection (Charlesworth et al. 1987). However, this expectation is not valid for selection on standing genetic variation (Orr & Betancourt 2001; Betancourt et al. 2002) and empirical studies have recently also challenged the fast-X effect in Drosophila species (Betancourt et al. 2002; Thornton & Long 2002, 2005; Torgerson & Singh 2003, 2006; Lu & Wu 2005; Richards et al. 2005). Moreover, the lower effective population size of sex chromosomes gives an increased probability of fixation of dominant slightly deleterious mutations due to drift, providing a neutral explanation to fast-X (although the opposite is expected for recessive deleterious mutations; Charlesworth et al. 1987). Furthermore, it is not known whether the large-X is a genome-(chromosome)-wide phenomenon or whether sex-linked genes contribute disproportionaly to adaptation over evolutionary time scales.

The large-X effect is likely to apply as well to organisms with female heterogamety (males ZZ, females ZW) and as discussed further below, there are theoretical reasons, and some empirical data in support thereof, to believe that the effect may be even more pronounced in such systems (Hastings 1994; Reinhold 1998; Iyengar et al. 2002). However, a critical examination of the large-X effect, be that in systems of male or female heterogamety, requires large-scale data.

Comparative genomics offers a means to study the role of genes, and different classes of genes, to long-term divergence among evolutionary lineages. Specifically, the rate of functional evolution manifested by the ratio of non-synonymous ($d_N$) to synonymous ($d_S$) substitution ($d_N/d_S$) is indicative of to what extent genes have contributed to adaptive evolution over longer time scales. With the availability of the chicken (Gallus gallus) genome sequence (International Chicken Genome Sequencing Consortium 2004) and upcoming genomic resources for a second bird species, the zebra finch (Taeniopygia guttata; Wade et al. 2004; Wada et al. 2006; Li et al. 2007; Replogle et al. 2008), large-scale molecular evolutionary analysis in birds is now possible. In a recent study (Mank et al. 2007) we found that mean $d_N/d_S$ for genes on avian Z chromosome is 15–30 per cent higher than for autosomal genes. One important question that arises from this observation is whether Z-linked genes are overrepresented among those genes that have actually accelerated their rate of sequence evolution in bird lineages, compared with sister groups such as mammals. If so, this would provide more direct evidence for the importance of sex-linkage in adaptive evolution. Using a comparative genomic approach, I here address the overall importance of sex-linked genes during avian evolution.

2. MATERIAL AND METHODS

I used data from Axelsson et al. (2008) which reports on a large-scale molecular evolutionary analysis of more than 5400 1 : 1 orthologous genes in the chicken and zebra finch, along with a comparison of sequence evolution for more than 2600...
of these genes in three mammals (human, mouse *Mus musculus* and opossum *Monodelphis domestica*). Details are given in Axelsson et al. (2008); briefly, this analysis used cDNA sequences from zebra finches obtained from embryonic or adult brain which were aligned to known and *ab initio* predicted protein-coding genes in the chicken genome identified by ENSEMBL (http://www.ensembl.org/Gallus_gallus/index.html). Orthology was established using the principle of best reciprocal hit. Subsequently, 1 : 1 : 1 : 1 : 1 bird–mammal orthologues were identified by alignment to the human, mouse and opossum transcriptomes. Maximum likelihood (COI, PAML package) was used to test for evolutionary rate differences between birds and mammals. As depicted in figure 1, the null model assumed the same $d_{S}/d_{S}$ ratio across all branches of a tree of the two birds and the three mammals, while the alternative model allowed the two avian branches to attain a separate $d_{S}/d_{S}$ ratio. In the absence of large-scale genomic sequence data for a non-avian reptilian, this test has to contrast $d_{S}/d_{S}$ in the two bird lineages with that of the rest of the tree, including the long branch connecting birds and mammals. Genes that were significantly accelerated in birds at $p < 0.05$ with a false discovery rate correction (see Benjamini & Hochberg 1995) of less than 0.1 were selected for further analysis. Fisher’s exact tests were used to test whether the fraction of sex-linked genes in the accelerated set deviated from the proportion of sex-linked genes in the total list of genes analysed by Axelsson et al. (2008).

Information on genomic location of genes was taken from ENSEMBL. Although one has to rely on data on chromosomal location from the chicken genome, available information suggests that the Z chromosomes of the chicken and zebra finch (Itoh et al. 2006), as well as of other Passeriform birds, are entirely syntenic (Backström et al. 2006). By contrast, genes that are Z-linked genes in birds are mainly located on human chromosomes 5 (HSA5), 9 (HSA9) and 18 (HSA18; all Z-linked genes found to be accelerated in birds are on HSA5 and HSA9). There are therefore good reasons to believe that genes sex-linked in the chicken have been so during avian evolution. The overall proportion of sex-linked genes in the chicken genome was taken from within the dataset itself, to avoid any potential bias caused by the proportion of such genes differing between genes expressed in brain and the whole genome (cf. Nguyen & Distèche 2006).

The UniGene EST assemblage (ftp://ncbi.nlm.nih.gov/repository/UniGene/) was used to calculate an index of tissue specificity (τ), which is a measure of expression breadth that ranges from 1 (expression restricted to a single tissue) to 0 (expression levels completely equal among all tissues investigated). I used the EMBL-EBI database for Gene Ontology (http://www.ebi.ac.uk/ego/) to obtain manually curated information on biological processes and molecular functions associated with genes.

### 3. RESULTS

Data from 2686 1 : 1 : 1 : 1 : 1 orthologous genes shared between the zebra finch, chicken, human, mouse and opossum were available. From alignments of these gene sequences followed by maximum-likelihood analysis, a set of 228 genes showed statistical evidence for a higher rate of functional evolution in the two avian lineages than in other branches of an unrooted phylogenetic tree of the five species (figure 1). These genes will be referred to as accelerated bird genes.

**Figure 1.** A phylogenetic tree over the five vertebrate species from which orthologous gene sequences were aligned. A maximum-likelihood approach was used to test a null model that assumed the same $d_{S}/d_{S}$ ratio across all branches of the tree against an alternative model that allowed the two bird branches shown to the left to attain a separate $d_{S}/d_{S}$ ratio from the rest of the tree.

**Figure 2.** The percentage of Z-linked and autosomal genes, respectively, identified as evolving faster in birds than in mammals. Error bars represent 95% confidence intervals of the proportions.

Z-linked genes were found to be significantly over-represented among accelerated bird genes. The frequency of genes from the Z chromosome was 0.065, which is two times higher than the genomic proportion of Z-linked genes (0.032; $p = 0.013$, Fisher’s exact test). For genes on the Z chromosome, more than 25 per cent are identified as fast evolving in birds, whereas less than 10 per cent of all autosomal genes belong to this category ($p = 0.00003$; figure 2). This indicates that, on a genomic scale, the Z chromosome has had a major influence on avian-specific adaptations and that the large-Z effect is a widespread phenomenon.

Accelerated bird genes on the Z chromosome are not confined to reproduction since the zebra finch gene sequences used for comparative analyses were obtained from cDNA libraries prepared from embryonic or adult brain. Moreover, an inspection of the biological processes and molecular functions identified for these genes indicates that they represent a broad range of functions (table 1). The large-Z effect is thus likely to involve many different traits, for example, given the neural origin of zebra finch sequence data, including behaviour.

Accelerated bird genes on the Z chromosome show several unique genomic features. They evolve more rapidly (mean $d_{S}/d_{S} = 0.219$ in the avian lineages) than both autosomal accelerated genes (0.153; $p = 0.0001$, t-test) and the genomic average (0.097, $p = 0.000008$). This is a combined consequence of an elevated rate of non-synonymous substitution ($d_{N}$ of 0.085 versus 0.058 and 0.041) and decreased rate of synonymous substitution ($d_{S}$ of 0.395 versus 0.436 and 0.498). They tend to show a more tissue-specific expression pattern in the chicken
Table 1. A list of Z-linked genes that show evidence of accelerated sequence evolution in birds compared with mammals.

<table>
<thead>
<tr>
<th>gene ID</th>
<th>gene</th>
<th>biological process</th>
<th>molecular function</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSGALG00000001801</td>
<td>intraflagellar transport 74 homologue (2009)</td>
<td>smooth muscle cell proliferation</td>
<td>ATP-binding activity, DNA-binding activity, transferase activity</td>
</tr>
<tr>
<td>ENSGALG00000002064</td>
<td>stomatin-like protein 2 (SLP-2)</td>
<td>smooth muscle cell proliferation</td>
<td>ATP-binding activity, DNA-binding activity, transferase activity</td>
</tr>
<tr>
<td>ENSGALG00000012598</td>
<td>corneal wound healing-related protein</td>
<td>smooth muscle cell proliferation</td>
<td>ATP-binding activity, DNA-binding activity, transferase activity</td>
</tr>
<tr>
<td>ENSGALG00000014713</td>
<td>DEAD (Asp-Glu-Ala-Asp) box polypeptide 4</td>
<td>smooth muscle cell proliferation</td>
<td>ATP-binding activity, DNA-binding activity, transferase activity</td>
</tr>
<tr>
<td>ENSGALG00000014737</td>
<td>kinesin-like protein KIF2A</td>
<td>smooth muscle cell proliferation</td>
<td>ATP-binding activity, DNA-binding activity, transferase activity</td>
</tr>
<tr>
<td>ENSGALG00000014768</td>
<td>similar to small glutamine-rich tetra-tripeptide repeat (TPR)-containing beta</td>
<td>smooth muscle cell proliferation</td>
<td>ATP-binding activity, DNA-binding activity, transferase activity</td>
</tr>
<tr>
<td>ENSGALG00000014999</td>
<td>claustrin</td>
<td>smooth muscle cell proliferation</td>
<td>ATP-binding activity, DNA-binding activity, transferase activity</td>
</tr>
<tr>
<td>ENSGALG00000015107</td>
<td>phosphatidylinositol-4-phosphate 5-kinase type I</td>
<td>smooth muscle cell proliferation</td>
<td>ATP-binding activity, DNA-binding activity, transferase activity</td>
</tr>
<tr>
<td>ENSGALG00000015265</td>
<td>clathrin light chain A (Lca)</td>
<td>smooth muscle cell proliferation</td>
<td>ATP-binding activity, DNA-binding activity, transferase activity</td>
</tr>
<tr>
<td>ENSGALG00000015548</td>
<td>similar to ring finger protein 20</td>
<td>smooth muscle cell proliferation</td>
<td>ATP-binding activity, DNA-binding activity, transferase activity</td>
</tr>
<tr>
<td>ENSGALG00000015641</td>
<td>cyclin-H (MO15-associated protein)</td>
<td>smooth muscle cell proliferation</td>
<td>ATP-binding activity, DNA-binding activity, transferase activity</td>
</tr>
<tr>
<td>ENSGALG00000015669</td>
<td>regulator of differentiation 1 (Rod1)</td>
<td>smooth muscle cell proliferation</td>
<td>ATP-binding activity, DNA-binding activity, transferase activity</td>
</tr>
</tbody>
</table>

(mean $r = 0.573$ versus $0.546$ for all genes; $p = 0.15$) and can be expressed at lower levels (mean number of transcripts per million = 216.7 versus 355.3 for all genes; $p = 0.23$), although this is not statistically significant.

For comparison, I used the same dataset to analyse those genes that showed evidence of being accelerated in mammals compared with the other branches of the tree depicted in figure 1. The proportion of genes X-linked in humans in the whole dataset is 3.2 per cent. There were 69 accelerated mammalian genes, of which four were X-linked in humans (5.7%), which is not significantly different from the proportion in the whole dataset ($p = 0.19$, Fisher’s exact test). However, this test has limited power due to the relatively few genes found to be accelerated in mammals. Moreover, the proportion of sex-linked genes in the accelerated datasets does not differ significantly between birds and mammals ($\chi^2 = 0.82$).

### 4. DISCUSSION

This study shows that sex-linked genes are significantly overrepresented among those genes that show an accelerated rate of function evolution (for which increased $dS/dd$ is a strong indicator) in two divergent bird lineages when compared with several mammals. The two avian lineages—Galliformes and Passeriformes—represent two orders from the earliest divergence within Neognathae (the large group of birds including all extant species but ratites and tinamous), leading to Galloanserae (including Galliformes) and Neoaves (including Passeriformes). This split is estimated to have occurred 100 Myr ago (Van Tuinen et al. 2000). The evidence for accelerated evolution is thus likely to reflect long-term trends common to phenotypically divergent birds. Moreover, as the model contrasts the molecular evolution of genes in two bird lineages with that in the rest of an unrooted tree including the long branch connecting early avian and mammalian ancestors, this design tests for accelerated genes in the terminal bird lineages, not in the lineage connecting the most recent common ancestor of birds and mammals and an early Neognathae ancestor. Whether these genes have also been important during the evolution of Dinosauria (dinosaurs and birds), Theropoda (bipedal dinosaurs including birds) and early Aves cannot be revealed.

Although high $dS/dd$ estimates are usually interpreted as evidence for adaptive evolution (driven by positive selection), they may also be suggestive of relaxed constraint (reduced purifying selection). In the latter, the overall fixation rate of slightly deleterious mutation is expected to be negatively correlated with the effective population size, thereby representing a neutral explanation to increased $dS/dd$. However, mean $dS/dd$ in the total list of bird-mammal orthologues used in this study is similar in birds and mammals (Axelsson et al. 2008). Moreover, Mank et al. (2007), by contrasting data on intraspecific polymorphism (chicken) with interspecific divergence (chicken–zebra finch), found that high rates of molecular evolution of the avian Z chromosome mainly derives from an increased rate of adaptive evolution rather than relaxed constraint. I therefore conclude that the accelerated bird genes observed here are most easily conceived under a scenario of representing a suite of genes important to adaptive evolution.

Proc. R. Soc. B (2009)
This study extends and strengthens the conclusions of Mank et al. (2007) in several respects. For example, by focusing on genes showing an accelerated rate of sequence evolution (compared with an outgroup) rather than comparing the mean $d_S/d_o$ of chromosomal categories, a more direct role for sex-linked genes in adaptive avian evolution is ascertained. Moreover, data presented here suggest that the large-Z effect is more pronounced than previously indicated, with an observed twofold excess of accelerated genes on the Z chromosome. Perhaps most importantly, since the current study used only genes present as 1:1 orthologues in birds and mammals, while Mank et al. 2007 also included chicken–zebra finch orthologues not found in mammals, it can be ascertained that it is the chromosomal location (sex-linkage) in the avian genome per se that is important to adaptive evolution. The bird-specific genes included in Mank et al. (2007) evolve faster than those present as 1:1 orthologues in birds and mammals (mean $d_S/d_o$ of 0.135 versus 0.106). Moreover, bird-specific genes evolving new adaptive functions may be prone to sex-linkage due to the higher likelihood of evolution of new sexually antagonistic genes on sex chromosomes (Rice 1984; Charlesworth et al. 1987).

While a large-Z effect was detected in the present study, a reciprocal analysis made using the same dataset did not provide evidence for an excess of X-linkage among genes accelerated in mammals compared to birds. However, since there were relatively few mammalian accelerated genes (69), there is not sufficient power in these data to statistically demonstrate that the large-Z effect is more pronounced than the large-X effect. The genome sequence of the zebra finch is soon to become available and will allow a genome-wide analysis of the genomic distribution of genes accelerated in avian and mammalian lineages, to test the hypothesis that male- and female-heterogametic taxa differ with respect to the importance of sex-linkage in adaptive evolution. Theoretical work suggests that ZW systems are more conducive to selection on sex-linked genes under both good-genes and Fisher's runaway models of sexual selection (Reeve & Pfennig 2003; Servedio & Saetre 2003; Kirkpatrick & Hall 2004; Albert & Otto 2005; Hall & Kirkpatrick 2006). For example, when a female preference gene is Z-linked, at least half of mothers' sons will inherit both the preference gene and an autosomal trait gene that is the target for female preference. As a consequence, exaggerated male ornaments (flashy displays) may be more likely to evolve in ZW systems (Albert & Otto 2005). There is circumstantial empirical support for this inference from female heterogametic birds, butterflies and fishes (Hastings 1994; Prowell 1998; Reinhold 1998; Iyengar et al. 2002; Lindholm & Breden 2002; but see Mank et al. 2006). Moreover, Saether et al. (2007) have recently shown that sex chromosomes are likely to represent a hot spot for adaptive speciation in a flycatcher model system, with evidence for Z-linkage of species recognition, species-specific male plumage traits and genes causing low hybrid fitness. Interestingly, an analysis of more than 10 000 orthologues in human, chimpanzee and rhesus macaque did not find any evidence for a non-random genomic distribution of positively selected genes (Rhesus Macaque Genome Sequencing and Analysis Consortium 2007).

The large-X effect was originally perceived as the disproportionate contribution of the X chromosome in causing heterogametic $F_1$ hybrid inviability/sterility (Coyne 1985, 1992; Coyne & Orr 1989). Breeding experiments have provided support for this phenomenon in the speciation process, although data from backcrosses suffer from the limitation of not being able to distinguish an overrepresentation of speciation genes on X from the exposure of recessive speciation genes in hybrid males (Presgraves 2008). However, introgression analyses in Drosophila species have confirmed an excess of male sterility factors on the X (Tao et al. 2003; Masly & Presgraves 2007). It seems highly unlikely that the excess of sex-linkage among adaptively evolving bird genes found in this study would be entirely due to genes involved in speciation. It rather indicates that large-Z is widespread and of general significance for adaptive divergence in birds. This would be compatible with a fast-Z effect that affects all kinds of genes on the Z chromosome, not only those involved with pre- or postzygotic isolation. The existence of fast-X in Drosophila is controversial (Betancourt et al. 2002; Thornton et al. 2006; Begun et al. 2007), but it has recently been demonstrated for several mammals (Nielsen et al. 2005; Torgerson & Singh 2006; Baines & Harr 2007).

The finding of a genomic large-Z effect is of particular interest in relation to how the expression of Z-linked genes is dealt with by male and female birds. Surprisingly, it has recently been found that birds do not have a sex chromosome-wide mechanism for dosage compensation, meaning that males consistently express higher levels of Z-linked genes compared with females (Ellegren et al. 2007; Itoh et al. 2007). Given the ubiquity of dosage compensation across the animal kingdom, it remains to be understood how birds can cope with unequal levels of gene expression from sex chromosomes in males and females. However, relevant in the present context, it may be that the higher levels of sex-linked gene expression in males favour Z-linkage of genes under sexual selection, augmenting the large-Z effect. This novel hypothesis could be tested by investigating sexual antagonism of Z-linked genes, for example by comparing the amplitude of sex-biased gene expression and sequence evolution of genes on the avian Z chromosome. Moreover, since theoretical work indicates that the dominance coefficient has to average less than 0.25, instead of less than 0.5, to get fast-X evolution without dosage compensation (Charlesworth et al. 1987), a better understanding of the distribution of dominance coefficient for sex-linked genes is very much needed.

This work was supported by grants from the Swedish Research Council. Comments from two anonymous reviewers are acknowledged.

REFERENCES


Backström, N., Brandström, M., Gustafsson, L., Qvarnström, A., Cheng, H. & Ellegren, H. 2006 Genetic


Itoh, Y. et al. 2007 Dosage compensation is less effective in birds than in mammals. *J. Biol.* 6, 2. (doi:10.1186/jbiol53)


