Both Male-Biased and Female-Biased Genes Evolve Faster in Fish Genomes

Liandong Yang[1], Zhaolei Zhang[2,3], and Shunping He[1,*]

[1]The Key Laboratory of Aquatic Biodiversity and Conservation of Chinese Academy of Sciences, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, People's Republic of China

[2]Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada

[3]Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada

*Corresponding author: E-mail: clad@ihb.ac.cn.

Accepted: September 24, 2016

Abstract

Males and females often display extensive phenotypic differences, and many of these sexual dimorphisms are thought to result from differences between males and females in expression of genes present in both sexes. Sex-biased genes have been shown to exhibit accelerated rates of evolution in a wide array of species, however the cause of this remains enigmatic. In this study, we investigate the extent and evolutionary dynamics of sex-biased gene expression in zebrafish. Our results indicate that both male-biased genes and female-biased genes exhibit accelerated evolution at the protein level. In order to differentiate between adaptive and nonadaptive causes, we tested for codon usage bias and signatures of different selective regimes in our sequence data. Our results show that both male- and female-biased genes show signatures consistent with adaptive evolution. In order to test the generality of our findings across fish, we also analyzed publicly available data on sticklebacks, and found results consistent with our findings in zebrafish.

Key words: sex-biased gene expression, sex-specific selection, adaptive evolution.

Introduction

Phenotypic differences between males and females (sexual dimorphism) are a major source of intra-specific variation (Darwin 1871), particularly in animals, where sex differences can include behavioral (Breedlove 1992), physiological (Bardin and Catterall 1981), and morphological (Darwin 1871) dimorphisms. It is often assumed that the majority of sexually dimorphic traits arise from differences in expression for genes present in both sexes (Connallon and Knowles 2005; Rinn and Snyder 2005). In line with this, studies in a wide array of organisms have attempted to determine the differences in gene expression between males and females (Parsch and Ellegren 2013), including *Drosophila* (Assis et al. 2012; Perry et al. 2014), birds (Pointer et al. 2013), nematodes (Albritton et al. 2014), and brown alga (Lipinska et al. 2015). These works have revealed that a significant fraction of genes in the genome exhibit differential expression between males and females (Rinn and Snyder 2005; Mank, Hultin-Rosenberg, Webster, et al. 2008; Reinius et al. 2008; Jiang and Machado 2009), which suggests that the expression of sexual dimorphism is related to marked genetic reprogramming (Lipinska et al. 2015).

Sex-biased genes often show elevated rates of evolution, although there is substantial variation among organisms in whether male-biased genes, female-biased genes, or both expression classes show elevated rates of evolution. In *Drosophila* and mammals, male-biased genes show elevated rates of evolution (Khaitovich et al. 2005; Assis et al. 2012), however fungi show elevated rates of evolution for female-biased genes (Whittle and Johannesson 2013). Both male- and female-biased genes in brown alga show elevated rates of evolution (Lipinska et al. 2015). Finally, in birds, genes that are female-biased in late development show higher rates of evolution than male-biased adult-expressed genes, although both categories show higher rates of evolution than unbiased genes (Mank et al. 2010).

There are many potential causes of rapid rates of sequence evolution for sex-biased genes, including natural selection,

sexual selection, and relaxed purifying selection (reduced functional pleiotropy) (Ellegren and Parsch 2007; Mank and Ellegren 2009; Parsch and Ellegren 2013), and there is considerable debate regarding whether elevated rates of evolution observed for sex-biased genes are due to adaptive (Proschel et al. 2006) or nonadaptive processes (Gershoni and Pietrokovski 2014; Harrison et al. 2015). Although work in *Drosophila* has shown that male-biased genes more often exhibit a signature of adaptive evolution (Proschel et al. 2006), evidence from birds (Harrison et al. 2015) and humans (Gershoni and Pietrokovski 2014) indicate relaxed constraint might be driving rapid rates of evolution. Additionally, there have been recent concerns about how expression-bias might alter mutation-selection dynamics (Dapper and Wade 2016), leading to the fixation of mildly deleterious alleles in strongly male-biased genes (Gershoni and Pietrokovski 2014). Data from more species are needed to resolve this debate.

Although there have been several studies of sex-biased genes expression in zebrafish and other fish species (Small et al. 2009; Wong et al. 2014; Liu et al. 2015), few of them detailed the evolutionary dynamics of sex-biased genes. Hence, a comprehensive analysis of the expression and evolution of sex-biased genes in fishes is needed. Here, we study sex-biased genes in zebrafish in order to determine rates of evolution for sex-biased genes, as well as determine the relative importance of adaptive versus nonadaptive processes. The zebrafish, *Danio rerio*, is an important model organism in biomedicine, neurophysiology, and developmental genetics (Kinkel and Prince 2009). Although the species does not exhibit much sexual dimorphism in morphology, previous studies on its mating patterns have shown that high variance exists in male and female mating success (Spence et al. 2008). Additionally, domestic zebrafish, relative to its wild ancestor, have been shown to have recently lost sex chromosomes (Wilson et al. 2014), and previous studies have shown that many factors, including environment, hormones, and genetic factors, can influence sex differentiation in the lab (Liew and Orban 2014). This provides a unique opportunity to measure the degree of sex-bias in a species without distinct sex chromosomes.

We observe a greater number of male-biased than female-biased genes, as well as higher expression levels for male-biased genes. Interestingly, our results also indicate that both male-biased and female-biased genes exhibit accelerated evolution, and a greater proportion of sites that experienced positive selection, suggesting that their faster evolution appears to be partly driven by adaptive evolution. We used publicly available data from stickleback to test the generality of these findings across fish, and find consistent signatures of adaptive evolution in both female- and male-biased genes in this species. Taken together, our results add important insight into the role of adaptive versus nonadaptive processes underlying rates of evolution for sex-biased genes.

Materials and Methods

Identification of Sex-Biased Genes from RNA-Seq Datasets

To identify sex-biased genes in zebrafish (*Danio rerio*), paired-end RNA-seq datasets from zebrafish adult testis, adult ovary, male head, female head, whole male body without head or testis, whole female body without head or ovary were collected from NCBI's SRA database (Collins et al. 2012) and one testis sample was generated by our lab at *Novogene* (Beijing, China) (Zhong et al. 2016). Detailed information about these RNA-seq data can be found in supplementary table S1, Supplementary Material online. Data quality was assessed using FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/) and reads were filtered using Trim galore (version 0.3.7) (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) to remove any potential residual Illumina adaptor sequence and trim bases with a Phred quality score <20. Only paired-end reads for which either read was longer than 25 bp after trimming were retained for subsequent analysis.

Filtered paired-end reads from each sample were aligned to transcript sequences of zebrafish downloaded from Ensembl (release 78) (Flicek et al. 2014) using Bowtie (version 1.1.1) (Langmead et al. 2009), and transcript abundances were estimated using RSEM program (v1.2.20) (Li and Dewey 2011). Gene expression level (FPKM) were determined by RSEM and only genes with FPKM > 1 in at least half of the individuals for each tissue were considered as transcriptionally active genes and were used for the subsequent analysis. We extracted raw read counts for each gene from RSEM, normalized them to control for differences in sequencing depth across each tissue separately using TMM method and identified differentially expressed genes with the edgeR package (Robinson et al. 2010) using a minimal fold-change of 2 and an adjusted *P* value cutoff of 0.001. Full lists of sex-biased genes can be found in supplementary table S3, Supplementary Material online.

Measurement of Nonsynonymous and Synonymous Substitution Rates

To estimate the evolutionary rates of sex-biased gene sequences, we downloaded protein-coding sequences from the zebrafish in Ensembl (release 78), and grass carp (*Ctenopharyngodonidellus*) genome (Wang et al. 2015), and retained the longest transcript for each gene for this analysis. Grass carp was chosen to compare with zebrafish because it is the closest species, whose genome has been sequenced, to zebrafish (Wang et al. 2015). One-to-one orthologs between zebrafish and grass carp were determined using Inparanoid 4.1 (Ostlund et al. 2010) with default parameters. A total of 16,612 pairwise 1:1 orthologs were identified and aligned using PRANK (v.140603) (Loytynoja and Goldman 2005) at the codon level with the option "-codon". SWAMP (Version 31-03-14) was used to filter regions with poor alignment with

a cutoff of 4 in a window size of 5, and a minimum length of 75 bp (Harrison et al. 2014). We also removed all positions having gaps and "N" from the alignments and excluded the alignment shorter than 100 bp from analysis. Using this method, 16,577 1:1 orthologs were identified. We then calculated the number of nonsynonymous substitutions per nonsynonymous site (dN), the number of synonymous substitutions per synonymous site (dS), and their ratio (dN/dS) for each ortholog pairwise using codeml program in PAML 4.7 package (runmode = −2 and model = 1) (Yang 2007). Genes with saturated synonymous substitution values (dS > 2) or N*dN or S*dS < 1 were excluded from further analysis.

The above approaches using two species measure the functional divergence that has occurred between zebrafish and grass carp, which shared a last common ancestor roughly 50 Ma (Wang et al. 2015). Considering that our gene expression datasets were taken from zebrafish, we thus were also interested in the patterns of substitution that has occurred on the zebrafish lineage alone. To this end, we expanded the pairwise dN/dS analysis to include another closely related species: cave fish (*Astyanax mexicanus*). We obtained one-to-one orthologs between zebrafish and cave fish using Biomart and combined them with grass carp to yield 11,958 1:1:1 orthologs among these three species. Orthologs alignment and trimming were performed in the same way as the pairwise analysis. Lineage specific evolutionary rates (dN, dS, and dN/dS) were determined using codeml (PAML 4.7 package) with the free-ratio model (runmode = 0 and model = 1) and orthologs with dS > 2 or N*dN or S*dS < 1 in zebrafish lineage were excluded from analysis.

The effective number of codons (ENCs) for all sex-biased and unbiased genes in this study was calculated using CodonW (version 1.4.2) (http://codonw.sourceforge.net/). Lower ENCs value indicates stronger synonymous codon usage bias (Hambuch and Parsch 2005).

Expression Breadth

RNA-seq data from different zebrafish tissues (including liver, muscle, eye, spleen, intestine and pancreas) (Kelkar et al. 2014) and the tissues used to identified sex-biased genes were combined to estimate breadth of gene expression. Transcript abundances (FPKM) were calculated by RSEM. The specificity index (τ) (Yanai et al. 2005) was used as a measure of breadth of gene expression for each gene, using the following formula:

$$T = \frac{\sum_{i=1}^{N}(1 - \times_i)}{N - 1},$$

where N is the number of tissues, x_i is the expression value in the given tissue i normalized by the highest expression value of the gene in all analyzed tissues (N). τ index values range from 0 to 1, with higher τ values corresponding to stronger tissue specificity (low expression breadth).

Positive Selection Analysis

We obtained the one-to-one orthologs among zebrafish, cod, fugu, and stickleback from Ensembl using Biomart. Then we integrated these orthologs with the above 11,958 1:1:1 orthologs among zebrafish, grass carp and cave fish to test for the signatures of positive selection. To detect the evidence of positive selection acting on a subset of sites of orthologs, we employed the paired nested site models (M1a, M2a; M7, M8) (Yang 2000, 2007) in the codeml program (PAML 4.7 package). These models allow the dN/dS ratio to vary among sites but not across lineages. The likelihood ratio test (LRT) was used to compare twice the log likelihood value differences with a chi2 distribution with df = 2 between the two nested models (M1a vs. M2a; M7 vs. M8) for each ortholog and genes with P value <0.05 adjusted by multiple testing from one of the two LRT tests were identified as positively selected genes.

Considering that the site model in codeml only allow ω ratio to vary among sites but not across lineages, we further used the branch-site models (model = 2, NS sites = 2) in codeml program to detect genes under positive selection only in zebrafish lineage. By setting the zebrafish as the foreground branch, we compared a selection model that allowed a class of codons on the zebrafish branch to have ω > 1 (Model A2, fix_omega = 0, omega = 1.5) with a neutral model that constrained this additional class of sites to have ω = 1 (Model A1, fix_omega = 1, omega = 1). The LRT test was used to obtain a P value from the statistics (twice the log likelihood value differences) and genes with FDR-adjusted P value <0.05 from LRT tests were identified as positively selected genes.

Deviations from neutrality for sex-biased genes were further assessed using polymorphism data, which were downloaded from Ensembl. The number of nonsynonymous substitutions (Dn) and synonymous substitutions (Ds) in zebrafish were extracted for each ortholog from outputs of codeml. Polymorphism data (SNPs) in zebrafish was downloaded using BIOMART and the number of nonsynonymous polymorphism (Pn) and synonymous polymorphism (Ps) were determined by SnpEff (Cingolani et al. 2012). These data were employed by DoFE to calculate the direction of selection statistic, DoS = Dn/(Dn + Ds)−Pn/(Pn + Ps), which is a measure of the difference in the proportions of nonsynonymous fixed differences and polymorphisms according to previous study (Stoletzki and Eyre-Walker 2011; Perry et al. 2014).

Results

Sex-Biased Gene Expression Patterns in Zebrafish Different Tissues

We collected and compared data from 12 RNA-seq experiments, including two biological replicates for the following different tissues: adult testis, adult ovary, male head, female head, whole male body without head or testis, whole female body without head or ovary (supplementary table S1, Supplementary Material online). Gene expression abundances, which were measured as FPKM, were strongly and positively correlated between biological replicates of each tissue, with r ranging from 0.86 to 0.99 ($P < 2.2 \times 10^{-16}$). We recovered a total of 19,908, 21,002, and 19,120 genes that were expressed in gonad, head, and whole body, respectively. Figure 1A exhibited the overall correlation matrix among the expression profiles for all experiments analyzed, indicating that the male and female tissue pairs of profiles were also highly correlated, despite some were coming from independent experiments. Therefore, these pairs of male and female expression profiles can be appropriately used as biological replicates to identify differentially expressed genes between sexes.

Based on these RNA-seq experiments, we found that almost 30% (7,785 out of 26,459) of protein-coding genes in zebrafish showed sex-biased expression in one of the tissues analyzed (supplementary tables S2 and S3, Supplementary Material online), which is generally consistent with previous reports in zebrafish (Small et al. 2009) and other species, such as *Drosophila* (Jiang and Machado 2009) and birds (Pointer et al. 2013). Just as expected, the number of genes exhibiting differential expression between males and females was significantly higher in adult gonad than in other somatic tissues (fig. 1B–D), suggesting that the majority of cases of sex-biased gene expression resulted from the gonad (Parisi et al. 2004). Male-biased genes were found to be more numerous than female-biased genes in adult gonad (fig. 1B), although more female-biased genes were found in somatic tissues (fig. 1C and D). Furthermore, male-biased genes were, on an average, expressed at significantly higher levels than female-biased genes in gonad (Wilcoxon rank sum test, $P < 2.2 \times 10^{-16}$) (fig. 1E). On the other hand, the magnitude of differential expression between the sexes was also significantly greater for male-biased genes than female-biased genes in gonad (Wilcoxon rank sum test, $P < 2.2 \times 10^{-16}$) (fig. 1F). Taken together, these results suggested that gene expression in zebrafish is also "masculinized", consistent with other taxa (Zhang et al. 2007; Small et al. 2009; Pointer et al. 2013). We also noted that the majority of sex-biased genes exhibited significant sex-biased expression in only one of the tissues analyzed (fig. 1G and H), indicating that sex-biased genes were generally expressed in a sex-biased manner in one tissue.

To examine the relationship between gene expression level (FPKM) and magnitude of sex-biased expression, we grouped the sex-biased genes according to the fold-change difference between male and female and plotted the mean expression level in male and female samples for each group (supplementary fig. S1, Supplementary Material online). This analysis suggested that high degree of sex bias arises in different ways for male- and female-biased genes. When genes showed a higher degree of male-biased expression, this was predominantly the product of decreased expression of these genes in females (supplementary fig. S1A, C, and E, Supplementary Material online). These patterns were broadly similar in both gonad and somatic tissue. The results obtained for female-biased genes were relatively complex. In head, the situation was same as that observed for the male-biased genes in that a higher degree of female-biased expression appeared to be correlated with down regulation in males (supplementary fig. S1D, Supplementary Material online). However, in both gonad (supplementary fig. S1B, Supplementary Material online) and whole body (supplementary fig. S1F, Supplementary Material online), when genes showed a higher degree of female-biased expression, this was probably the product of both decreased expression in males and increased expression in females.

Accelerated Protein Evolution in Both Male- and Female-Biased Genes

Consistent with previous studies in other species (Yang et al. 2006; Mank, Hultin-Rosenberg, Webster, et al. 2008; Small et al. 2009; Pointer et al. 2013), our above analysis of sex-biased expression indicated that the gonad is the most dimorphic tissue transcriptionally in zebrafish and thus we performed evolutionary analysis primarily on this tissue. To test for differences in the rate of evolutionary divergence between sex-biased and unbiased genes, we first compared levels of nonsynonymous (dN), synonymous (dS) substitution, and the ratio of nonsynonymous to synonymous substitutions (dN/dS) for sex-biased and unbiased genes using pairwise comparisons with orthologs between zebrafish and grass carp (table 1). The results of this analysis were generally consistent with previous reports in other taxa. That is, male-biased genes evolved significantly faster (i.e., had higher dN/dS ratios) than unbiased genes (Wilcoxon rank sum test, $P < 2.2 \times 10^{-16}$). Interestingly, our results also show that female-biased genes evolve more rapidly than unbiased genes (Wilcoxon rank sum test, $P < 6.3 \times 10^{-4}$) (fig. 2C). In addition, the dN values for both male- and female-biased genes were significantly higher than unbiased genes (Wilcoxon rank sum test, $P < 2.2 \times 10^{-16}$ and $P = 0.018$, respectively) (fig. 2A), but no differences were observed for dS values (Wilcoxon rank sum test, $P = 0.3$ and $P = 0.22$, respectively) (fig. 2B). These findings demonstrated that the higher dN/dS ratios for both male- and female-biased genes compared with unbiased genes were the product of significantly elevated rates of nonsynonymous substitution, rather than a

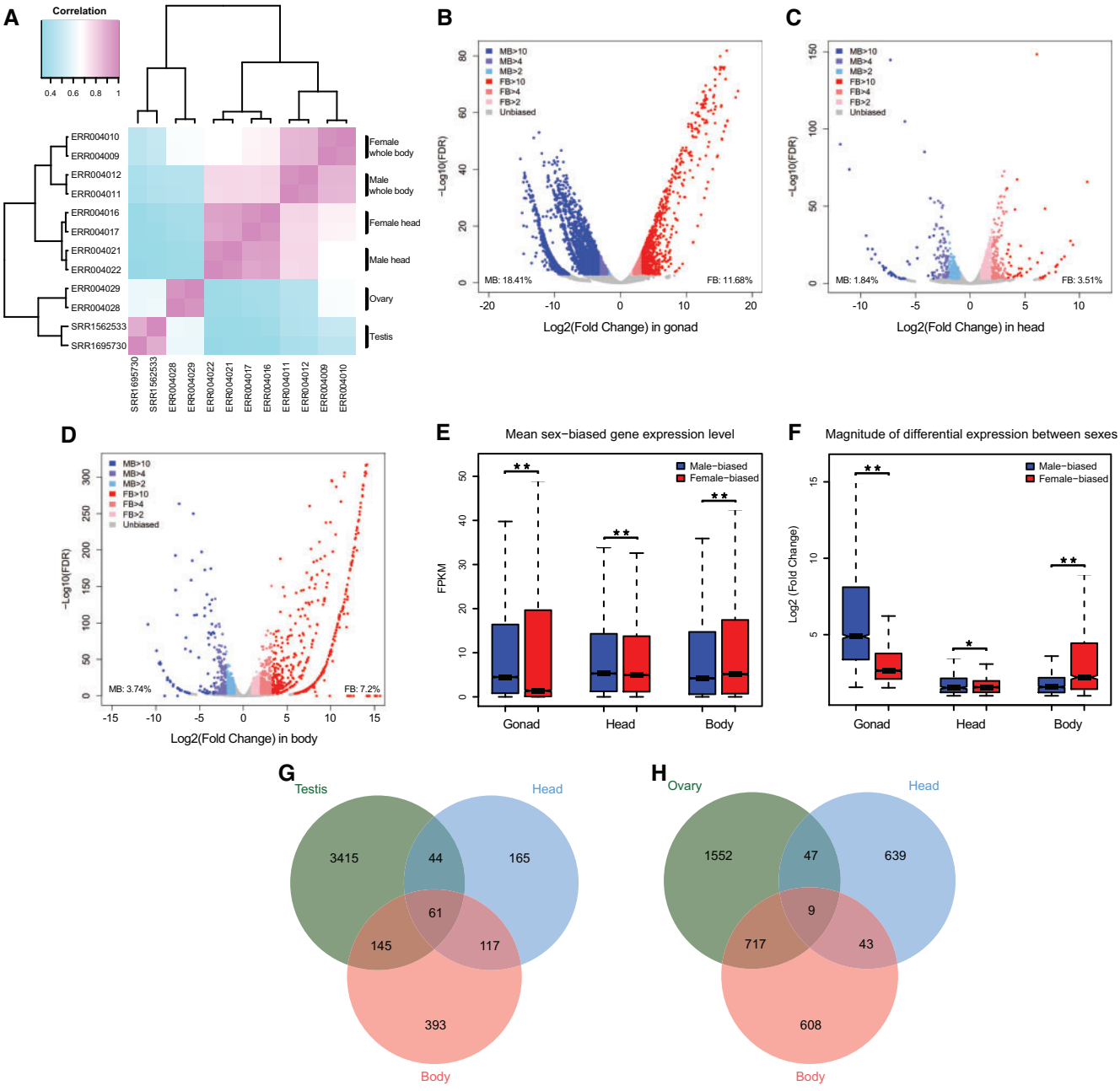

Fig. 1.—Sex-biased gene expression in zebrafish. (*A*) Heatmap of cross-correlations of all expression datasets analyzed with different experiments being hierarchically clustered. (*B–D*) Pairwise comparisons of gene expression abundances between male and female in gonad (*B*), head (*C*), and whole body without gonad or head (*D*). Genes that exhibited significantly different levels of transcript abundance (sex-biased genes) were shown with colored dots. Percentages in each panel indicate the percentage of all genes expressed in that tissue that were at least 2-fold male-biased (MB; upper left) and female-biased (FB; lower right). (*E*) Comparison of the gene expression level (FPKM) for male-biased genes and female-biased genes in different tissues. (*F*) Comparison of magnitude of differential expression (Fold Change) between male-biased and female-biased genes in different tissues. (*G* and *H*) Venn diagram showing tissue specificity for male-biased genes (*G*) and female-biased genes (*H*). Significant differences are indicated by the asterisks, based on Wilcoxon rank sum test, $*P < 0.05$, $**P < 0.001$.

reduction for the synonymous substitution rate, which were consistent with enhanced rates of protein evolution in both male- and female-biased genes. Furthermore, the frequency distribution of dN/dS ratios also indicated that both male- and female-biased genes tended to be enriched in genes with higher dN/dS ratios and to contain fewer genes under strong selective constraint (dN/dS < 0.1) compared with unbiased genes (fig. 2G).

Table 1
Divergence Estimates for Sex-Biased and Unbiased Genes in Zebrafish Gonad

	Male-biased	Female-biased	Unbiased	P_{MU}[b]	P_{FU}[b]	P_{MF}[b]
Zebrafish-grass carp comparison						
n Gene	2,487	1,637	12,145			
dN[a]	0.0712	0.0637	0.06	2.2×10^{-16}	0.0181	6.2×10^{-5}
dS[a]	0.4379	0.4395	0.4443	0.2974	0.2181	0.7565
dN/dS[a]	0.1541	0.1407	0.1302	2.2×10^{-16}	6.4×10^{-4}	1.2×10^{-5}
Zebrafish lineage-specific						
n Genes	1736	1,195	8331			
dN[a]	0.0285	0.0252	0.0237	1.7×10^{-12}	0.0273	9.3×10^{-4}
dS[a]	0.2445	0.2426	0.2478	0.8683	0.234	0.0638
dN/dS[a]	0.1112	0.1023	0.094	2.3×10^{-10}	0.0015	0.0441

[a]The values provided are median values for each category of genes.
[b]*P* value of Wilcoxon rank sum test for comparisons among male-biased (M), female-biased (F), and unbiased (U) genes.

For all the above analysis, we estimated the functional divergence that has occurred between zebrafish and grass carp, which shared a common ancestor roughly 50 Ma (Wang et al. 2015). As our expression data were collected from zebrafish, it is meaningful to detect the patterns of functional divergence in the zebrafish lineage alone. To this end, we compiled 11,958 1:1:1 three-species alignments (zebrafish, grass carp, and cave fish) to examine the divergence pattern leading to the zebrafish lineage since its split from the common ancestor with grass carp (table 1). The results of this analysis indicated that the divergence data for the zebrafish lineage based on the three-species data were broadly consistent with the two-species data, with significantly higher dN/dS ratios for both male- and female-biased genes compared with unbiased genes (Wilcoxon rank sum test, $P < 2.3 \times 10^{-10}$ and $P = 0.0015$, respectively) (fig. 2F). Meanwhile, the dN values were also significantly higher for both male- and female-biased genes than unbiased genes (Wilcoxon rank sum test, $P < 1.7 \times 10^{-12}$ and $P = 0.027$, respectively) (fig. 2D), but no differences were observed for dS values (Wilcoxon rank sum test, $P = 0.8683$ and $P = 0.234$, respectively) (fig. 2E). Analysis of the relative frequency distribution of dN/dS ratios further showed that both male- and female-biased genes tended to be enriched in genes with higher dN/dS ratios and to contain fewer genes under strong selective constraint (dN/dS < 0.1) compared with unbiased genes (fig. 2H). Taken together, our analyses, based on both zebrafish-grass carp pairwise comparison and the zebrafish lineage-specific divergence data, provides convergent evidence that both male- and female-biased genes exhibit accelerated evolution rates in zebrafish.

In order to provide an independent validation on our above findings and to test whether they are specific to specific lineage or the same case in other fish groups, we further analyzed the evolutionary dynamics of sex-biased genes in another independent lineage and with independent datasets. We obtained the sex-biased genes (564 male-biased and 704 female-biased genes) identified in threespine stickleback liver tissue (Leder et al. 2010) and the dN, dS, and dN/dS values calculated between threespine and ninespine sticklebacks (3091 orthologs) (Guo et al. 2013) from previous studies directly. Interestingly, we also found that the dN/dS ratios for both male- and female-biased genes were statistically significantly higher than unbiased genes in stickleback lineage (Wilcoxon rank sum test, $P = 0.0143$ and $P = 0.0005$, respectively) (146 male-biased and 283 female-biased genes were used for this analysis). In addition, the dN values were significantly higher in both male- and female-biased genes compared with unbiased genes (Wilcoxon rank sum test, $P = 0.0095$ and $P = 0.013$, respectively), but no differences were found for dS values. Furthermore, the relative frequency distribution of dN/dS ratios also indicated that both male- and female-biased genes tend to be enriched in genes with higher dN/dS ratios and to contain fewer genes under strong selective constraint relative to unbiased genes in stickleback (supplementary fig. S2, Supplementary Material online). Considering that 23% of the sex-biased genes identified in stickleback were concentrated on the nascent sex chromosomes and that genes located on the sex chromosomes evolve more rapidly, we further compared the evolutionary rates for autosomal genes only. This analysis also showed that both male- and female-biased genes evolved faster in stickleback (supplementary fig. S3, Supplementary Material online). Taken together, our results indicate that the faster evolution for both male- and female-biased genes is maintained in fish lineages across different evolutionary timescales, and is independent of molecular techniques (RNA-seq vs. Microarray).

Adaptive versus Nonadaptive Signatures

In order to differentiate adaptive from nonadaptive causes of the accelerated rate of evolution for sex-biased genes, we first assessed codon bias, which has been shown to be reduced in male-biased genes of *Drosophila* (Hambuch and Parsch 2005),

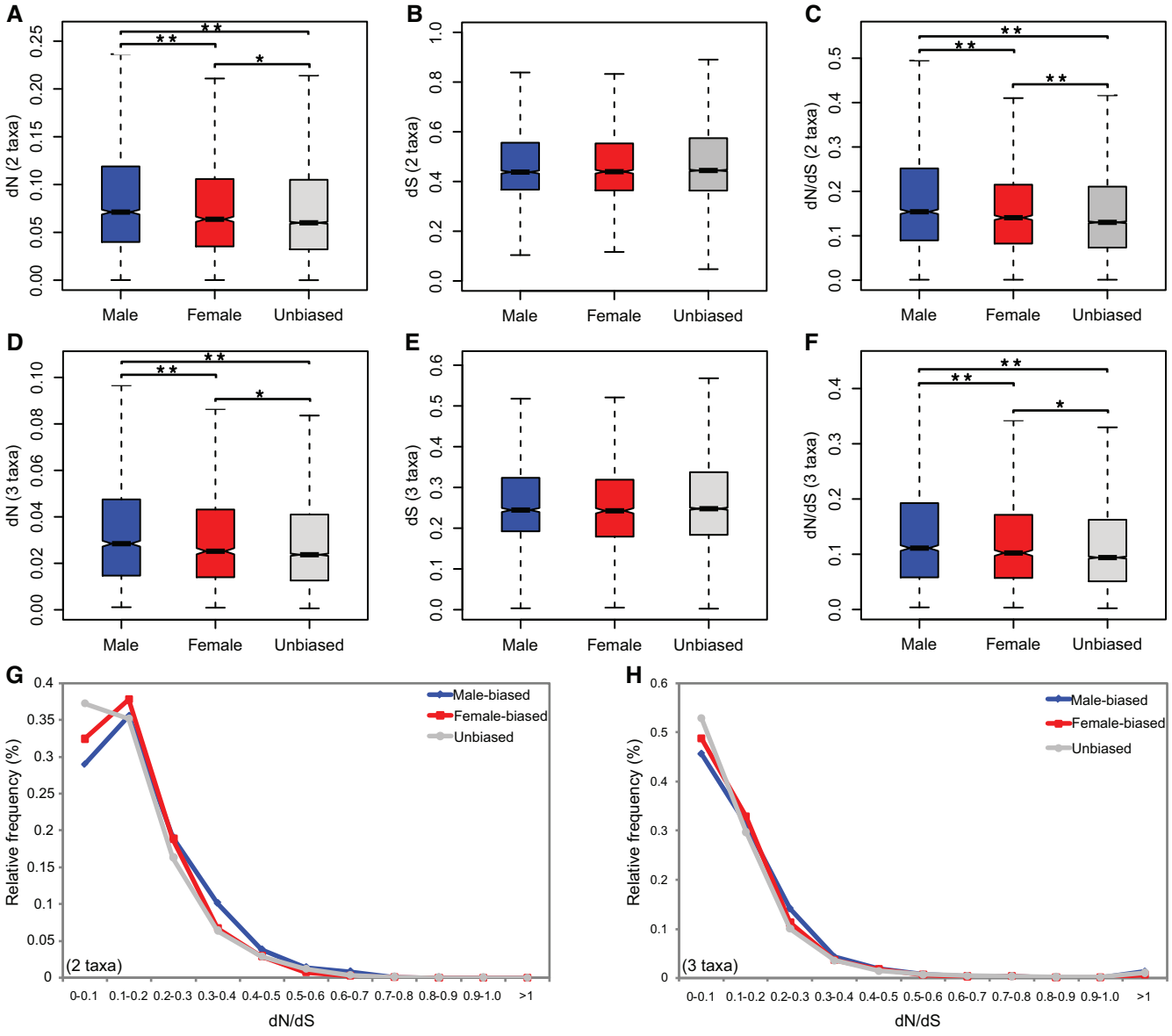

Fig. 2.—Divergence estimates for male-biased, female-biased, and unbiased genes. Pairwise dN (*A*), dS (*B*), and dN/dS (*C*) values were estimated by comparing putative orthologs between zebrafish and grass carp. Zebrafish lineage-specific dN (*D*), dS (*E*), and dN/dS (*F*) values were calculated using three-species alignments among zebrafish, grass carp, and cave fish. The relative frequency distribution of dN/dS ratios for each category of genes were calculated from zebrafish-grass carp comparison (*G*) and zebrafish lineage alone (*H*). Outliers were removed from the boxplot. Significant differences are indicated by the asterisks, based on Wilcoxon rank sum test, $*P < 0.05$, $**P < 0.001$.

possibly because adaptive protein evolution has influenced selection on codon usage. In order to examine levels of codon bias in genes with sex-biased expression in zebrafish, we investigated synonymous codon usage in genes with complete coding sequences and sex-biased expression. As shown in figure 3A, there were significant differences in levels of codon bias between sex-biased and unbiased genes in zebrafish. Consistent with the result in *Drosophila*, male-biased genes in zebrafish exhibited significantly less codon bias than unbiased genes, defined as higher values for the ENCs (Wilcoxon rank sum test, $P < 4.5 \times 10^{-11}$). However, our results further demonstrated that female-biased genes in zebrafish also exhibited significantly less codon bias compared with unbiased genes (Wilcoxon rank sum test, $P < 2.2 \times 10^{-16}$).

We also investigated the ENCs in sex-biased genes identified in threespine stickleback (fig. 3B). Consistent with the phenomenon in zebrafish, this analysis also showed that both male- and female-biased genes in threespine stickleback

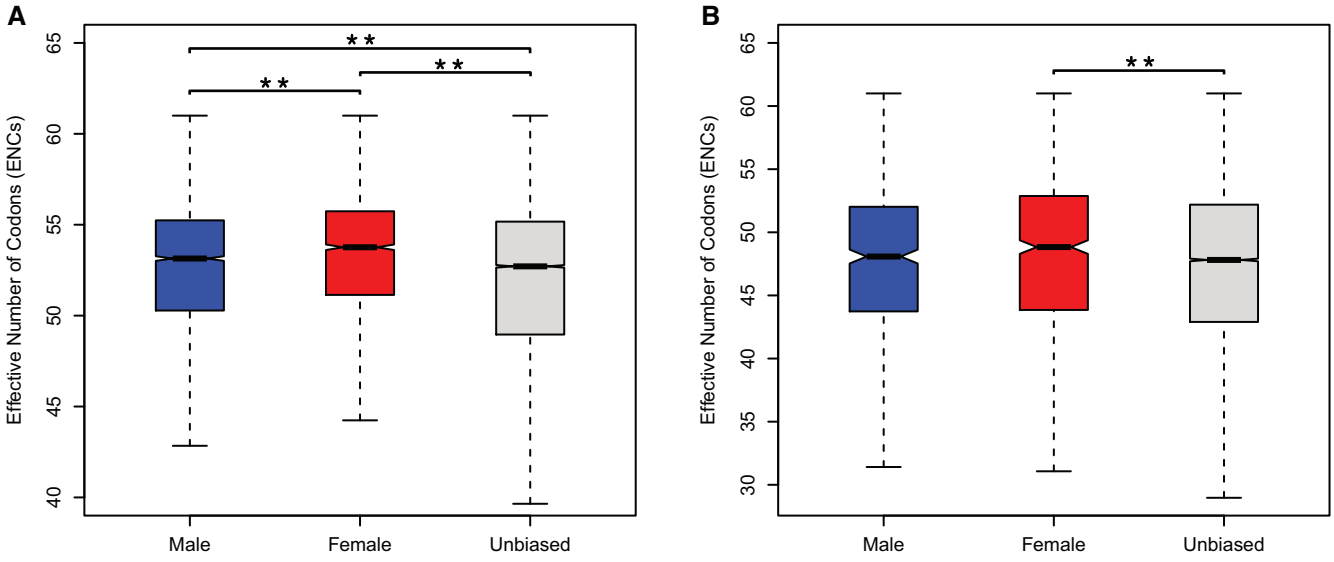

Fig. 3.—Codon usage bias in sex-biased versus unbiased genes in zebrafish (*A*) and threespine stickleback (*B*). Significant differences were indicated by the asterisks, based on Wilcoxon rank sum test, $*P < 0.05$, $**P < 0.001$.

have higher ENCs values compared with unbiased genes, although the statistics test is not significant for male-biased genes (Wilcoxon rank sum test, $P = 0.2639$ and $P = 0.0002$ for male- and female-biased genes, respectively), which suggests that sex-biased genes in threespine stickleback also exhibit less codon bias than unbiased genes. Taken together, these data indicate that both male- and female-biased genes show less codon bias than unbiased genes in fish genomes, potentially indicating reduced efficacy of selection on optimal codon usage for these loci.

We also tested for tissue-specificity of expression, which has shown to be correlated with evolutionary rate (Zhang et al. 2004). Previous work has shown that sex-biased genes tend to be more tissue specific in their expression, possibly indicating relaxed constraint of their function (Mank, Hultin-Rosenberg, Zwahlen et al. 2008; Meisel 2011). We therefore compared tissue specificity (τ) for male-biased, female-biased and unbiased genes (fig. 4A). Our results showed that both male-biased and female-biased genes had significantly higher τ values compared with unbiased genes (Wilcoxon rank sum test, $P < 2.2 \times 10^{-16}$ and $P < 2.2 \times 10^{-16}$ for male- and female-biased genes, respectively), implying that sex-biased genes tend to be more tissue specific in expression. Additionally, the specificity index of male-biased genes was significantly higher than that of female-biased genes (Wilcoxon rank sum test, $P < 2.2 \times 10^{-16}$). This suggests that faster rates of evolution for sex-biased genes may be partly driven by relaxed functional constraints. However, it is worth noting that the increase in the specificity index values of genes was not significantly correlated with the values of dN/dS (Pearson $r = 0.1$, fig. 4B) or ENCs ($r = 0.007$, fig. 4C), suggesting that the breadth of expression of genes was not a sole determinant of their speed of evolution.

We tested our sequence data for signatures of positive versus nonadaptive evolution, as both can lead to elevated dN/dS (Zhang et al. 2007). In order to assess the contribution of positive selection, we used both sequence divergence and polymorphism data to detect whether positive selection is more effective on sex-biased genes. Using the paired nested site models (M1a, M2a; M7, M8) implemented in codeml, we found significant signatures of positive selection acting on 184/3,856 male-biased genes, 128/2,325 female-biased genes, and 800/20,278 unbiased genes, respectively (supplementary table S4, Supplementary Material online). The proportion of genes exhibiting evidence of positive selection for both male- and female-biased genes were significantly higher than that for unbiased genes based on site model in PAML (chi-squared test, $P = 0.017$, $P = 3.3 \times 10^{-4}$, and $P = 1.6 \times 10^{-4}$ for male-, female-, and sex-biased genes, respectively) (fig. 5A). Positively selected genes identified by branch-site model also suggests that the proportion of genes exhibiting evidence of positive selection for sex-biased genes were significantly higher than that for unbiased genes (chi-squared test, $P = 0.023$).

We further used polymorphism data on standing variation in protein-coding sequences to test for deviations from neutrality by the statistic of direction of selection (DoS), a measurement of the difference between the proportion of divergent and polymorphic nonsynonymous substitutions (Stoletzki and Eyre-Walker 2011). Here, negative DoS values suggest slightly deleterious mutations segregating, whereas zero values suggest only neutral evolution and positive

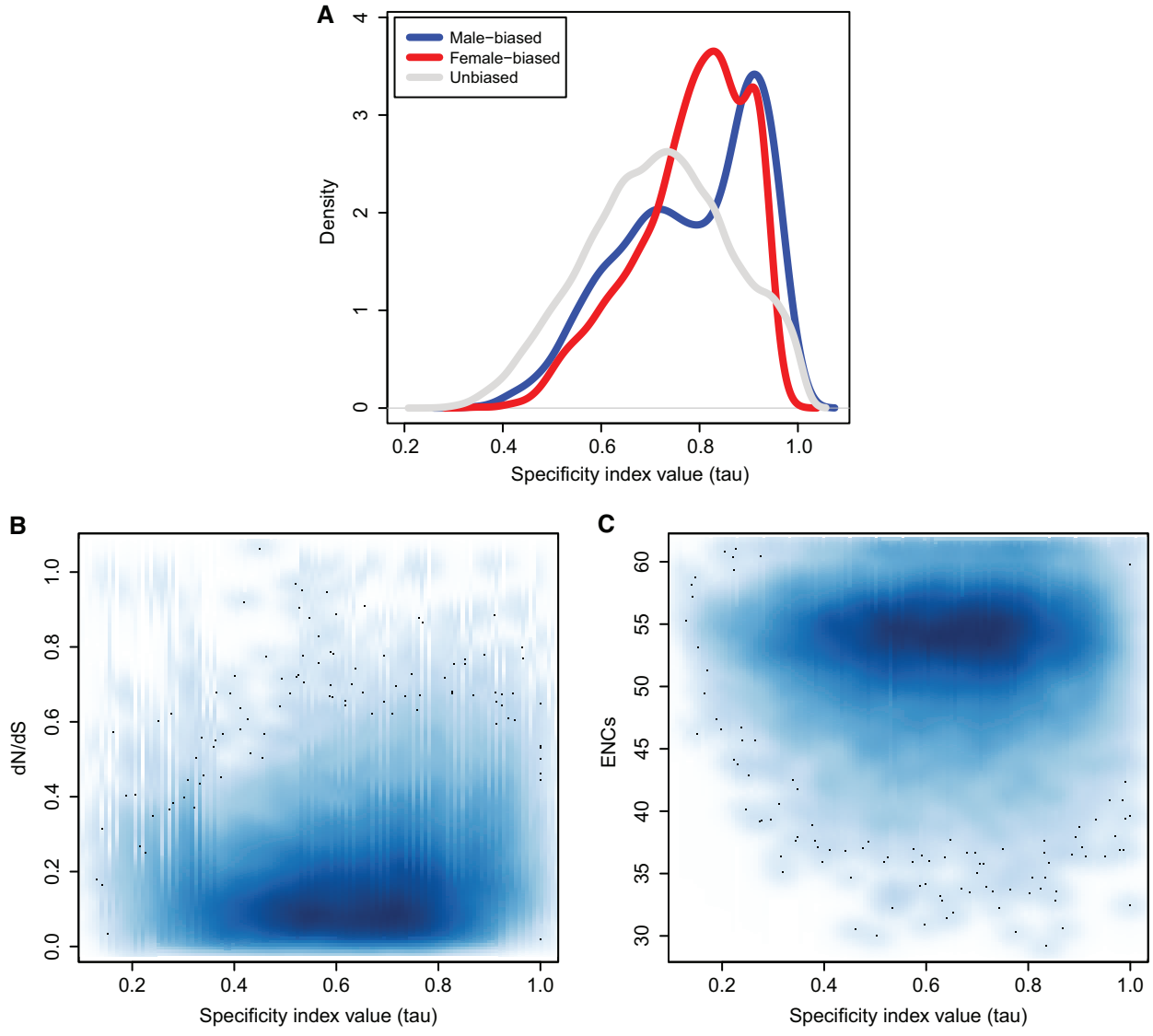

Fig. 4.—Breadth of expression among male-biased, female-biased, and unbiased genes measured by the specificity index (τ) (*A*). Correlation between specificity index values and dN/dS (*B*) or ENC (*C*).

values suggest there is evidence of adaptive evolution (Stoletzki and Eyre-Walker 2011). Our results showed that the DoS values for both sex-biased and unbiased genes were higher than zero (fig. 5B), suggesting broad effects of positive selection on the evolution of genes in fish. As expected, comparison between sex-biased and unbiased genes revealed that the DoS values for sex-biased genes were higher than that for unbiased genes, with a significantly higher DoS values for male-biased genes (Wilcoxon rank sum test, $P = 0.013$) (fig. 5B). This analysis further confirmed that faster protein evolution in sex-biased genes were more likely to be driven by stronger positive selection. In sum, our analysis based on both sequence divergence and polymorphism data consistently suggested a stronger effect of positive selection on the evolution of sex-biased genes in zebrafish.

Discussion

Analyses of sex-biased gene expression in a range of species, including *Drosophila* (Ranz et al. 2003; Assis et al. 2012; Perry et al. 2014), birds (Mank et al. 2010; Pointer et al. 2013), and mammals (Yang et al. 2006; Reinius et al. 2012), have shown that a large percentage of the transcriptome displays differential gene expression between the two sexes (Ellegren and Parsch 2007; Parsch and Ellegren 2013). Although there were also some studies focusing on sex-biased genes in gonad or

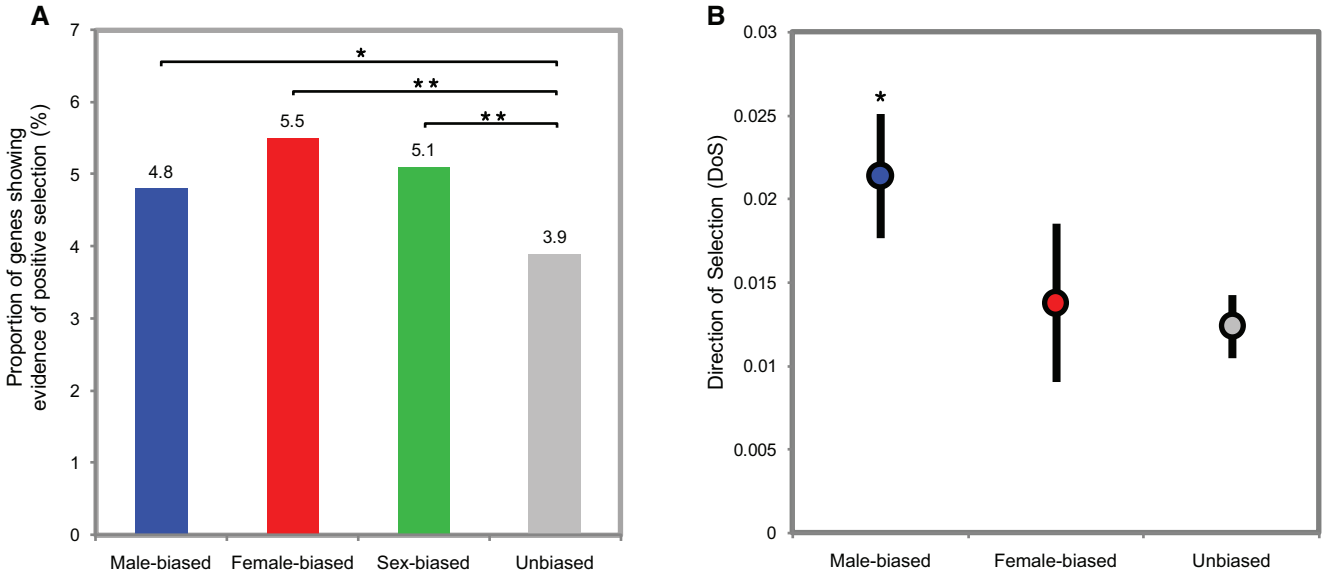

Fig. 5.—Adaptive evolution of sex-biased genes in zebrafish. (*A*) Comparison of the proportions of genes showing signature for positive selection between sex-biased and unbiased genes. The percentage of positively selected genes were given on top of the bar. (*B*) Comparison of the values of direction of selection among male-, female-biased, and unbiased genes. Significant differences were indicated by the asterisks, based on Wilcoxon rank sum test, $*P < 0.05$, $**P < 0.001$.

brain in zebrafish and other fishes (Small et al. 2009; Wong et al. 2014; Liu et al. 2015), few of them have investigated the evolutionary dynamics of sex-biased genes and their driving forces. Our results in zebrafish are consistent with this from other model organisms. Nevertheless, zebrafish are interesting in that they do not exhibit ostentatious sexual dimorphisms (Spence et al. 2008), thus the sex-biased genes we identified are likely largely to do with differences between the female and male gonad in gametic production and delivery, rather than somatic sexual dimorphisms related to sexual selection. Within the gonad, our results show that the zebrafish transcriptome is masculinized, both in terms of the greater proportion of male-biased genes, as well as the greater level of expression in males. This is consistent with other taxa, including *Drosophila* (Parisi et al. 2003; Ranz et al. 2003), mice (Yang et al. 2006), and frogs (Malone et al. 2006).

Most importantly, we found that both male- and female-biased genes in zebrafish show elevated nonsynonymous-to-synonymous divergence ratios compared to unbiased genes. Although this is somewhat different than work in *Drosophila* and mammals, where only male-biased genes show elevated rates of evolution (Ellegren and Parsch 2007; Zhang et al. 2007; Parsch and Ellegren 2013), it does add to the diversity of patterns observed in other taxa. For example, female-biased genes show more rapid rates of evolution in *Neurospora crassa* (Whittle and Johannesson 2013), and both male- and female-biased genes show elevated rates of evolution in brown alga (Lipinska et al. 2015).

In order to confirm the generality of this pattern across fish, we tested the evolutionary rate of sex-biased genes in the stickleback lineage by comparing threespine and ninespine sticklebacks which diverged roughly 13 Mya (Bell et al. 2009). The convergence in pattern of evolutionary rates for both female- and male-biased genes between zebrafish and stickleback lineages, which last shared a common ancestor roughly 290 Mya (Steinke et al. 2006), suggests that this pattern may be representative for fishes in general. This is also supported by convergent results from a small-scale study in the guppy (Sharma et al. 2014).

The current evidence is unclear whether rapid rates of evolution for sex-biased genes observed in other organisms is due to positive selection (Proschel et al. 2006), perhaps related to sexual selection or sperm competition (Ellegren and Parsch 2007), or is due to relaxed constraint and genetic drift (Harrison et al. 2015; Dapper and Wade 2016). We used codon-bias, tissue-specificity and signatures of selection on coding sequence to differentiate these two potential forces. We observed that both male-biased genes and female-biased genes exhibiting reduced optimal codon usage compared to unbiased genes. The preferential use of a subset of synonymous codons (codon bias) is a prevalent phenomenon in a wide range of species (Akashi and Eyre-Walker 1998; Akashi 2001; Duret 2002), including vertebrates (Doherty and McInerney 2013; Ma et al. 2014). The "preferred codons" are generally assumed to arise as a result of natural selection favoring efficient and accurate translation (Duret and

Mouchiroud 1999; Duret 2000). However, a previous study found that male-biased genes in *Drosophila*, which evolved faster than female-biased and unbiased genes, have significantly less codon bias than both female-biased and unbiased genes (Hambuch and Parsch 2005), which indicated either reduced efficacy of selection for optimal codon usage or recent disturbance of optimal codon usage due to positive selection.

Additionally, we find that both male- and female-biased genes tend to have more tissue-specific expression relative to unbiased genes. In general, broadly expressed genes tend to have more complex functional roles and undergo stronger functional constraints, thus evolve slower than genes with limited expression (Duret and Mouchiroud 2000). Therefore, this parallel reduction in expression breadth for both male- and female-biased genes in fishes may also serve as one of the possible reasons underlying their faster evolutionary rates. However, it should be noted that only weak correlation exists between expression breadth and evolutionary rate, implying that relaxed functional constraints may not be the unique driving force of evolutionary rates.

Finally, our analysis of divergence and polymorphism data indicate that a greater proportion of male- and female-biased genes show evidence of positive selection when compared to unbiased genes. The potential role of positive selection is also indicated by our DoS values for sex-biased genes, which were higher than for unbiased genes.

There are several reasons that may contribute to this symmetry of accelerated evolution for both male- and female-biased genes. First, zebrafish does not exhibit much morphological sexual dimorphism (Spence et al. 2008), which may either result in less sexual selection for both male-biased and female-biased genes simultaneously, or lead to more strong sexual selection for both male- and female-biased genes simultaneously. Considering the faster evolution for both male- and female-biased genes, we think the latter would be more logical. Another possible reason for the faster evolution rates for both male- and female-biased genes in fishes may be that both male- and female-biased genes tend to have restricted breadth of expression pattern relative to unbiased genes. In general, broadly expressed genes tend to have more complex functional roles and undergo stronger functional constraints, thus to evolve slower than genes with limited expression (Duret and Mouchiroud 2000). Therefore, this parallel reduction in expression breadth for both male- and female-biased genes in fishes may also serve as one of the possible reasons underlying their faster evolutionary rates. Furthermore, this symmetry of accelerated evolution for both male- and female-biased genes may also be driven by external fertilization in fishes. Compared with internal fertilization in mammals, *Drosophila*, and birds, fishes need to produce much more viable eggs, thus female may be undergone stronger selection, which may drive faster evolution for female-biased genes.

The codon usage, tissue-specificity and signatures of selection tests reveal a complex pattern, where both positive selection and relaxed constraints may both contribute to the accelerated rates of evolution observed for sex-biased genes. This complex pattern may vary across distantly related groups, explaining why work in *Drosophila* has recovered stronger signatures of selection in male-biased genes (Proschel et al. 2006), whereas studies in adult birds (Harrison et al. 2015) and humans (Gershoni and Pietrokovski 2014) reveals patterns more consistent with relaxed evolutionary constraint and genetic drift. Given this complex, clade-specific pattern, as well as recent concerns about how expression-bias might alter mutation-selection dynamics (Dapper and Wade 2016), data from more species are needed to resolve this debate.

Supplementary Material

Supplementary figures S1–S3 and tables S1–S4 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors thank the Associate Editor Prof Judith Mank and three anonymous reviewers for their valuable comments. The authors also acknowledge the financial support from the China Scholarship Council to Liandong Yang. This work was supported by the Pilot projects (grant no. XDB13020100).

Literature Cited

Akashi H. 2001. Gene expression and molecular evolution. Curr Opin Genet Dev. 11:660–666.
Akashi H, Eyre-Walker A. 1998. Translational selection and molecular evolution. Curr Opin Genet Dev. 8:688–693.
Albritton SE, et al. 2014. Sex-biased gene expression and evolution of the x chromosome in nematodes. Genetics 197:865–883.
Assis R, Zhou Q, Bachtrog D. 2012. Sex-biased transcriptome evolution in Drosophila. Genome Biol Evol. 4:1189–1200.
Bardin CW, Catterall JF. 1981. Testosterone: a major determinant of extragenital sexual dimorphism. Science 211:1285–1294.
Bell MA, Stewart JD, Park PJ. 2009. The World's Oldest Fossil Threespine Stickleback Fish. Copeia 256–265.
Breedlove SM. 1992. Sexual Dimorphism in the Vertebrate Nervous-System. J Neurosci. 12:4133–4142.
Cingolani P, et al. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6:80–92.
Collins JE, White S, Searle SM, Stemple DL. 2012. Incorporating RNA-seq data into the zebrafish Ensembl genebuild. Genome Res. 22: 2067–2078.
Connallon T, Knowles LL. 2005. Intergenomic conflict revealed by patterns of sex-biased gene expression. Trends Genet. 21:495–499.
Dapper AL, Wade MJ. 2016. The evolution of sperm competition genes: the effect of mating system on levels of genetic variation within and between species. Evolution 70:502–511.
Darwin CR. 1871. The descent of man, and selection in relation to sex. 2nd ed. London: Murray.

Doherty A, McInerney JO. 2013. Translational selection frequently overcomes genetic drift in shaping synonymous codon usage patterns in vertebrates. Mol Biol Evol. 30:2263–2267.
Duret L. 2000. tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. Trends Genet. 16:287–289.
Duret L. 2002. Evolution of synonymous codon usage in metazoans. Curr Opin Genet Dev. 12:640–649.
Duret L, Mouchiroud D. 1999. Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, and Arabidopsis. Proc Natl Acad Sci U S A. 96:4482–4487.
Duret L, Mouchiroud D. 2000. Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. Mol Biol Evol. 17:68–74.
Ellegren H, Parsch J. 2007. The evolution of sex-biased genes and sex-biased gene expression. Nat Rev Genet. 8:689–698.
Flicek P, et al. 2014. Ensembl 2014. Nucleic Acids Res. 42:D749–D755.
Gershoni M, Pietrokovski S. 2014. Reduced selection and accumulation of deleterious mutations in genes exclusively expressed in men. Nat. Commun. 5:4438.
Guo B, Chain FJ, Bornberg-Bauer E, Leder EH, Merila J. 2013. Genomic divergence between nine- and three-spined sticklebacks. BMC Genomics 14:756.
Hambuch TM, Parsch J. 2005. Patterns of synonymous codon usage in *Drosophila melanogaster* genes with sex-biased expression. Genetics 170:1691–1700.
Harrison PW, et al. 2015. Sexual selection drives evolution and rapid turnover of male gene expression. Proc Natl Acad Sci U S A. 112:4393–4398.
Harrison PW, Jordan GE, Montgomery SH. 2014. SWAMP: Sliding Window Alignment Masker for PAML. Evol Bioinform Online 10:197–204.
Jiang ZF, Machado CA. 2009. Evolution of sex-dependent gene expression in three recently diverged species of Drosophila. Genetics 183: 1175–1185.
Kelkar DS, et al. 2014. Annotation of the zebrafish genome through an integrated transcriptomic and proteomic analysis. Mol Cell Proteomics 13:3184–3198.
Khaitovich P, et al. 2005. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. Science 309: 1850–1854.
Kinkel MD, Prince VE. 2009. On the diabetic menu: zebrafish as a model for pancreas development and function. Bioessays 31:139–152.
Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10:R25.
Leder EH, et al. 2010. Female-biased expression on the X chromosome as a key step in sex chromosome evolution in threespine sticklebacks. Mol Biol Evol. 27:1495–1503.
Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323.
Liew WC, Orban L. 2014. Zebrafish sex: a complicated affair. Brief Funct Genomics 13:172–187.
Lipinska A, et al. 2015. Sexual dimorphism and the evolution of sex-biased gene expression in the brown alga ectocarpus. Mol Biol Evol. 32:1581–1597.
Liu H, et al. 2015. Large-scale transcriptome sequencing reveals novel expression patterns for key sex-related genes in a sex-changing fish. Biol Sex Differ. 6:26.
Loytynoja A, Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. Proc Natl Acad Sci U S A. 102:10557–10562.
Ma L, Cui P, Zhu J, Zhang Z. 2014. Translational selection in human: more pronounced in housekeeping genes. Biol Direct. 9:17.
Malone JH, Hawkins DL, Jr, Michalak P. 2006. Sex-biased gene expression in a ZW sex determination system. J Mol Evol. 63:427–436.
Mank JE, Ellegren H. 2009. Are sex-biased genes more dispensable?. Biol Lett. 5:409–412.
Mank JE, Hultin-Rosenberg L, Webster MT, Ellegren H. 2008. The unique genomic properties of sex-biased genes: insights from avian microarray data. BMC Genomics 9:148.
Mank JE, Hultin-Rosenberg L, Zwahlen M, Ellegren H. 2008. Pleiotropic constraint hampers the resolution of sexual antagonism in vertebrate gene expression. Am Nat. 171:35–43.
Mank JE, Nam K, Brunstrom B, Ellegren H. 2010. Ontogenetic complexity of sexual dimorphism and sex-specific selection. Mol Biol Evol. 27:1570–1578.
Meisel RP. 2011. Towards a more nuanced understanding of the relationship between sex-biased gene expression and rates of protein-coding sequence evolution. Mol Biol Evol. 28:1893–1900.
Ostlund G, et al. 2010. InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. Nucleic Acids Res. 38:D196–D203.
Parisi M, et al. 2003. Paucity of genes on the Drosophila X chromosome showing male-biased expression. Science 299:697–700.
Parisi M, et al. 2004. A survey of ovary-, testis-, and soma-biased gene expression in *Drosophila melanogaster* adults. Genome Biol. 5:R40.
Parsch J, Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene expression. Nat Rev Genet. 14:83–87.
Perry JC, Harrison PW, Mank JE. 2014. The ontogeny and evolution of sex-biased gene expression in *Drosophila melanogaster*. Mol Biol Evol. 31:1206–1219.
Pointer MA, Harrison PW, Wright AE, Mank JE. 2013. Masculinization of gene expression is associated with exaggeration of male sexual dimorphism. Plos Genet. 9:e1003697.
Proschel M, Zhang Z, Parsch J. 2006. Widespread adaptive evolution of drosophila genes with sex-biased expression. Genetics 174: 893–900.
Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL. 2003. Sex-dependent gene expression and evolution of the Drosophila transcriptome. Science 300:1742–1745.
Reinius B, et al. 2008. An evolutionarily conserved sexual signature in the primate brain. Plos Genet. 4:e1000100.
Reinius B, et al. 2012. Abundance of female-biased and paucity of male-biased somatically expressed genes on the mouse X-chromosome. BMC Genomics 13:607.
Rinn JL, Snyder M. 2005. Sexual dimorphism in mammalian gene expression. Trends Genet. 21:298–305.
Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140.
Sharma E, et al. 2014. Transcriptome assemblies for studying sex-biased gene expression in the guppy, *Poecilia reticulata*. BMC Genomics 15:400.
Small CM, Carney GE, Mo Q, Vannucci M, Jones AG. 2009. A microarray analysis of sex- and gonad-biased gene expression in the zebrafish: evidence for masculinization of the transcriptome. BMC Genomics 10:579.
Spence R, Gerlach G, Lawrence C, Smith C. 2008. The behaviour and ecology of the zebrafish, *Danio rerio*. Biol Rev Camb Philos Soc. 83:13–34.
Steinke D, Salzburger W, Meyer A. 2006. Novel relationships among ten fish model species revealed based on a phylogenomic analysis using ESTs. J Mol Evol. 62:772–784.
Stoletzki N, Eyre-Walker A. 2011. Estimation of the neutrality index. Mol Biol Evol. 28:63–70.

Wang Y, et al. 2015. The draft genome of the grass carp (*Ctenopharyngodon idellus*) provides insights into its evolution and vegetarian adaptation. Nat Genet. 47:625–631.
Whittle CA, Johannesson H. 2013. Evolutionary dynamics of sex-biased genes in a hermaphrodite fungus. Mol Biol Evol. 30:2435–2446.
Wilson CA, et al. 2014. Wild sex in zebrafish: loss of the natural sex determinant in domesticated strains. Genetics 198:1291–1308.
Wong RY, McLeod MM, Godwin J. 2014. Limited sex-biased neural gene expression patterns across strains in Zebrafish (Danio rerio). BMC Genomics 15:905.
Yanai I, et al. 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics 21:650–659.
Yang X, et al. 2006. Tissue-specific expression and regulation of sexually dimorphic genes in mice. Genome Res. 16:995–1004.
Yang Z. 2000. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. J Mol Evol. 51:423–432.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.
Zhang Y, Sturgill D, Parisi M, Kumar S, Oliver B. 2007. Constraint and turnover in sex-biased gene expression in the genus Drosophila. Nature 450:233–237.
Zhang Z, Hambuch TM, Parsch J. 2004. Molecular evolution of sex-biased genes in Drosophila. Mol Biol Evol. 21:2130–2139.
Zhong Z, Yang L, Zhang YE, Xue Y, He S. 2016. Correlated expression of retrocopies and parental genes in zebrafish. Mol Genet Genomics. 291:723–737.

Associate editor: Judith Mank