The Effects of CpG Densities around Transcription Start Sites on Sex-Biased Gene Expression in *Poecilia reticulata*

Katsunori Tamagawa*, Takashi Makino, and Masakado Kawata*

Department of Ecology and Evolutionary Biology, Graduate School of Life Sciences, Tohoku University, Aoba-ku, Sendai, Japan

*Corresponding authors: E-mails: tamagawa.katsunori@gmail.com; kawata@m.tohoku.ac.jp.

Abstract

As most genes are shared between females and males, DNA methylation is assumed to play a crucial role in sex-biased gene expression. DNA methylation exclusively occurs at CpG dinucleotides, and therefore, we would expect that CpG density around transcription start sites (TSSs) relate to sex-biased gene expression. Here we investigated the relationship between CpG densities around TSSs and the ratio of gene expression levels between sexes in the guppy (*Poecilia reticulata*), which displays remarkable sexual dimorphisms. We found that genes with sex-biased gene expression had different CpG densities downstream of TSSs compared with genes lacking sex-biased gene expression. Intriguingly, male-biased expression genes with intermediate CpG density downstream of TSSs exhibited greater differences in gene expression between sexes in the gonad and tail. Our findings suggested the possibility that CpGs around TSSs, especially in the downstream regions, play a crucial role in sex-biased gene expression through DNA methylation.

Key words: sexual dimorphism, transcriptional regulation, CpG dinucleotides, DNA methylation, guppy.

Introduction

Sexually reproducing organisms exhibit morphological and behavioral dimorphisms. For example, male peacocks and male lions have spectacular, colorful tail feathers and shaggy manes, respectively, that are not present in their female counterparts (West and Packer 2002; Zi et al. 2003; Williams and Carroll 2009). Sequential hermaphrodites, such as the bluehead wrasse, also exhibit extensive sexual dimorphisms regarding morphology and behavior (Godwin 2009), although they possess identical genomic content. These sexual dimorphisms are mostly formed through differential gene expression between genders, and hence, many researchers have investigated a number of sex-biased genes and rates of gene expression between sexes (Yang et al. 2006; Tao et al. 2013; Sharma et al. 2014). Although males and females majorly share genomic content, sex-biased genes are located not only on sex chromosomes but also on autosomal chromosomes. For example, Kang et al. (2011) identified 159 sex-biased genes located on the Y (13 genes), X (9 genes), and autosomal (137 genes) chromosomes by analyzing gene expression in the human brain.

Epigenetic modifications, such as DNA methylation, play a crucial role in sex-biased gene expression (Wijchers and Festenstein 2011; Piferrer 2013). Previous studies found significant correlation between DNA methylation and sex-biased gene expression via transcriptome and methylome analyses (Shao et al. 2014; Xu et al. 2014). In particular, variable levels of DNA methylation at promoter regions influence sex-biased gene expression and sex differentiation in vertebrates (Navarro-Martín et al. 2011; Matsumoto et al. 2013) by directly (blocking transcription factors from binding to promoter regions) or indirectly (recruiting proteins related to chromatin formation) suppressing gene expression (Jones and Takai 2001; Bird 2002; Klose and Bird 2006).

In vertebrates, DNA methylation primarily occurs at cytosine-guanine dinucleotides (CpG) (Razin and Riggs 1980; Shao et al. 2014). CpGs are largely depleted in vertebrate genome and are primarily enriched around transcription start sites (TSSs) (Saxonov et al. 2006; Deaton and Bird 2011). Based on CpG density around TSSs, the promoter regions in vertebrates are assigned to two distinct classes: low CpG (low-CpG class) and high CpG content (high-CpG class) (Saxonov et al. 2006; Elango and Soojin 2008). These two classes may be distinct due to the evolutionary loss of CpGs in promoters with high DNA methylation owing to the high mutation rate of methylated cytosine (Antequera 2003; Weber et al. 2007). Promoters with suppressed gene expression in germline cells appear to be highly methylated, and thus, CpG density in the promoter regions might be depleted. Conversely, promoters without methylation in the germline cells may play a role in maintaining the CpG density (Saxonov et al. 2006). Furthermore, there is a

difference in function between low-CpG and high-CpG genes. Low-CpG genes tend to be expressed in a tissue-specific manner, and high-CpG genes tend to be broadly expressed (Vinogradov 2005; Saxonov et al. 2006). Therefore, the regulatory roles of DNA methylation may be gleaned from dissecting CpG density in promoters.

We hypothesized that CpG density around a TSS affects sex-biased gene expression since sex-biased gene expression and DNA methylation as well as DNA methylation and CpG distribution are linked. To identify the genomic regions around TSSs and the degree of CpG density that together influence sex-biased gene expression, we examined the relationship between CpG distribution around TSSs and sex-biased gene expression. In this study, we used the guppy, *Poecilia reticulata*, which displays remarkable sexual dimorphisms, such as color pattern, body size, and behavior (Griffiths and Magurran 1998; Brooks 2000; Brooks and Endler 2001; Hendry et al. 2006; Postma et al. 2011). Male-specific body ornamentations, especially body and tail color patterns, are important visual signals for the females' mating choice (Houde 1997). By examining the relationship between CpG density around promoter regions and sex-biased gene expression, we aimed to determine if low or high CpG regions are important for transcriptional regulation of sex-biased genes. Our results suggested that DNA methylation around TSSs may play a crucial role in sex-biased expression of genes with low-CpG rather than high-CpG promoters.

Materials and Methods

Genomic Sequences and Sex Ratio Gene Expression Data Sets

Whole genome sequences of *P. reticulata* (Guppy_female_ 1.0 + MT; BioProjects PRJNA232869, PRJNA238429) were obtained from RefSeq (Künstner et al. 2016). Data sets that detailed sexual dimorphism with regards to the gene expression in the brain (including eyes), tail (including the post-anal tissue up to the beginning of the tail fin and containing adult skin, skeletal muscle, dorsal cord, bone, and cartilage), and gonadal tissue were taken from Sharma et al. (2014). By comparing the different tissues, we checked whether there are underlying mechanisms that regulate the varying sex-biased gene expression across the tissue types.

Sharma et al. (2014) performed transcriptomic analysis using gonadal tissue and two types of somatic tissues (brain and tail) for elucidating the role of sex-biased gene expression in sexual dimorphism, particularly in the development of male ornamentations and the females' mating choice. A brief summary of the methods for identification of sex-biased gene expression in Sharma et al. (2014) is described as follows. Sample fish were selected from laboratory-reared fish derived from a wild population in the Quare River. Adults (12–14 individuals) between 5 and 6 months of age were killed and

their brain, eye, liver, spleen, skin, tail, and gonadal tissues were harvested; further, the best RNA samples obtained were used for library preparation. In some cases, the brain, tail, and gonadal tissue belonging to the same individuals was used, and for others, different individuals were used for quantitative analysis. To prevent mating, male and female fish were separated at the age of 3–4 weeks. Constructed nonbarcoded cDNA libraries were sequenced on HiSeq^{TM} 2000 (Illumina, 101 bp paired-end). Genome-guided and genome-independent assemblies were compiled using TOPHAT-CUFFLINKS-CUFFMERGE and TRINITY, respectively. These assemblies were then merged into a reference transcriptome data set. For quantitative analysis, barcoded cDNA libraries of the brain, tail, and gonadal tissues were constructed and sequenced in the same manner as nonbarcoded libraries. Differential expression analyses were performed using six biological replicates of each tissue and sex. Each library was mapped to the reference transcriptome data set using BOWTIE2 v2.0.04, and mapping reads were counted using eXpress v1.3.1. Read counts were normalized using TMM normalization, and differential gene expression between the sexes was detected with edgeR. Genes showing significantly different expression between the sexes (FDR < 0.1) were assigned as "sex-biased", and those without a significant difference were designated as "nonbiased". Furthermore, sex-biased genes with higher expression in females and those with higher expression in males were classified as "female-biased" genes and "male-biased" genes, respectively.

In the present study, we used the reference transcriptome data set and the list of sex-biased genes published by Sharma et al. (2014). Reciprocal BLAST analysis was performed to identify homology (e-value <1e−4) between the reference transcriptome and the RNA data set from RefSeq. We used only genes which are top-hit each other and obtained gene annotations, which resulted in 1,875 female-biased and 833 male-biased genes in the brain, 7,139 female-biased and 5,180 male-biased genes in the gonads, and 1,136 female-biased and 1,064 male-biased genes in the tail.

Calculation of CpG$_{O/E}$, GpC$_{O/E}$, and GC Skew

CpG$_{Observed/Expected}$ (CpG$_{O/E}$) is the index of the density of CpG dinucleotide normalized by the densities of G and C nucleotides. CpG$_{O/E}$ was calculated as $P_{CpG}/(P_G*P_C)$, where P_{CpG} is the proportion of CG fractions, P_G is the proportion of G fractions, and P_C is the proportion of C fractions, respectively. GpC$_{Observed/Expected}$ (GpC$_{O/E}$) was also calculated as $P_{GpC}/(P_G*P_C)$, where P_{GpC} is the proportion of GC fractions, since GC dinucleotides are hardly targeted by DNA methylation (Razin and Riggs 1980; Shao et al. 2014) but depend on G + C contents as well as CpG sites. We used CpG$_{O/E}$ and GpC$_{O/E}$ to assess the effect of nucleotide-composition bias on sex-biased gene expression. In addition, GC skew was calculated as the ratio of $(P_C − P_G)$ to $(P_C + P_G)$. We used the GC

skew for an index of strand asymmetry in the distribution of G fractions and C fractions, along with assessing the effect of CpG islands on sex-biased gene expression. Nonmethylated CpG islands in promoter regions were suggestive of significant GC skew (Ginno et al. 2012; Chen et al. 2014).

Tests for Bimodality

In vertebrates, the distribution of $CpG_{O/E}$ shows a bimodal pattern (Saxonov et al. 2006; Elango and Soojin 2008). $CpG_{O/E}$, $GpC_{O/E}$, and GC skew were calculated for 300, 500, 1,000, and 5,000 bp of sequences upstream and downstream of the TSSs of 24,641 annotated genes. The bimodal pattern was found to be more evident for the 500 bp sequence than the rest; therefore, we used a 500 bp window for the following analysis (see "Results"). The NOCOM program was used to test the bimodality of $CpG_{O/E}$ distribution (Ott 1992). NOCOM estimates whether the distribution is unimodal or bimodal by comparing the likelihood of fitting the data to either model. Regarding our observed bimodal model, the intersection of the two distinct distributions estimated by NOCOM was determined by R v3. 2. 2 ("uniroot" function). The Chi-square test was used for testing the deviation of density of sex-biased genes in each tissue.

Sliding Window Analysis around TSSs

$CpG_{O/E}$, $GpC_{O/E}$, and GC skew were calculated by sliding window analysis (200 bp window and 100 bp sliding) for 5,000 bp upstream and downstream sequences from TSSs. Number of genes used in the analysis are highlighted in table 1. The differences in $CpG_{O/E}$, $GpC_{O/E}$, and GC skew between sex-biased and nonbiased genes in each window were compared by using ANOVA. P-value was adjusted by false discovery rate with Benjamini–Hochberg method.

GAM Analyses

Generalized additive model (GAM) analyses were conducted to examine the effect of $CpG_{O/E}$ around TSSs on sex-biased gene expression. Number of genes used are presented in table 1. Genes containing ambiguous nucleic acid bases (i.e., "N") were removed from the analyses since this would result in incorrect calculation of $CpG_{O/E}$. For each gene, $CpG_{O/E}$ was determined using 1,000 bp sequences upstream and downstream of TSSs. For GAM analysis, $CpG_{O/E}$ was an independent variable, and an absolute value of fold change [$\log_2$ (gene expression in males/gene expression in females)] was used as a dependent variable.

Results

$CpG_{O/E}$ Bimodality Suggests Two Classes of Promoters

The density distribution of $CpG_{O/E}$ was calculated for 300, 500, 1,000, and 5,000 bp sequences upstream and

Table 1

Number of Genes in the Sliding Window and GAM Analyses

Analysis	Tissue	Sex-Bias	Region	No. of Genes
Sliding window analysis	Brain	Female	—	1,875
		Male		833
		Non		21,933
	Gonad	Female	—	7,139
		Male		5,180
		Non		12,322
	Tail	Female	—	1,136
		Male		1,064
		Non		22,441
GAM analysis	Brain	Female	Upstream	1,637
			Downstream	1,792
		Male	Upstream	714
			Downstream	796
	Gonad	Female	Upstream	6,314
			Downstream	6,888
		Male	Upstream	4,609
			Downstream	4,984
	Tail	Female	Upstream	992
			Downstream	1,084
		Male	Upstream	935
			Downstream	1,020

downstream from the TSSs, and we found that the bimodal pattern $CpG_{O/E}$ was more evident for the 500 bp sequence than the rest (fig. 1A and supplementary fig. S1, Supplementary Material online). Therefore, we used the density distribution of a 500 bp window for the following analyses. The bimodal pattern was observed neither for the distribution of $GpC_{O/E}$ nor for the GC skews (fig. 1B and C). The mean values of the higher and lower modes were 0.78 and 0.42, respectively. The intersection between the two modes was estimated as $CpG_{O/E} = 0.57$. Thus, we assigned a gene into the low-CpG class when $CpG_{O/E} < 0.57$ or into the high-CpG class when $CpG_{O/E} \geq 0.57$. Furthermore, we found that the proportion of sex-biased genes assigned to the low-CpG and high-CpG classes was different from that of all the genes (table 2). In the brain, the proportion of female-biased genes in the low-CpG class was significantly larger than that of all genes, and conversely, the proportion of male-biased genes in the low-CpG class was significantly smaller than that of all the genes. In the gonads, both the proportions of female-biased and male-biased genes in the high-CpG class were significantly larger than those were of all the genes. In the tail, the proportion of male-biased genes in the high-CpG class was significantly larger than that of all the genes, whereas that of the female-biased genes in the low-CpG and high-CpG classes was not significantly different from those of all the genes.

$CpG_{O/E}$ Patterns around TSSs

We analyzed the difference between sex-biased and nonbiased genes in distribution of $CpG_{O/E}$ calculated for 200 bp
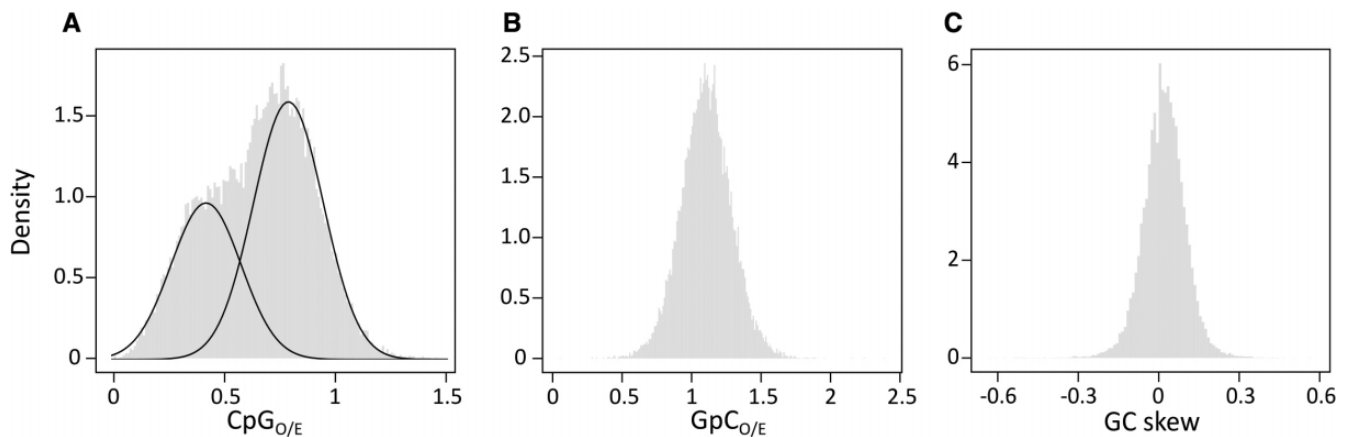
FIG. 1—Density distributions of (*A*) CpG$_{O/E}$, (*B*) GpC$_{O/E}$, and (*C*) GC skew calculated in 500 bp upstream and downstream regions from TSSs using 24,641 genes. Distribution curves in CpG$_{O/E}$ were estimated by NOCOM.

Table 2

The Number and Proportion of Genes in Each Tissue, with Sex-Bias and Low-CpG and High-CpG Classes Indicated

Tissue	No. of Genes	Sex-Bias	Gene Number	Low-CpG (%)	High-CpG(%)	Chi-squared Values
All genes	24,641	—	—	8,853 (36)	15,788 (64)	–
Brain	2,708	Female-biased	1,875	**826 (44)**	1,049 (56)	49.282 (<0.01)
		Male-biased	833	235 (28)	**598 (72)**	20.573 (<0.01)
Gonad	12,319	Female-biased	7,139	1,657 (23)	**5,482 (77)**	403.89 (<0.01)
		Male-biased	5,180	1,298 (25)	**3,882 (75)**	224.76 (<0.01)
Tail	2,200	Female-biased	1,136	424 (37)	712 (63)	0.85907 (0.354)
		Male-biased	1,064	259 (24)	**805 (76)**	59.327 (<0.01)

Note.—Differences in the proportion of low/high-CpG class genes were assessed using the chi-squared test. The proportions significantly higher than those of all genes are shown in boldface.

sliding window sequences in 5,000 bp upstream and downstream regions from TSSs. In all tissues and all biased genes examined, CpG$_{O/E}$ was largest near TSSs as previously found in other vertebrates (fig. 2). In the downstream region of TSSs, CpG$_{O/E}$ of sex-biased genes were either larger or smaller than those of nonbiased genes, depending on the tissue. In the brain, female-biased genes had lower CpG$_{O/E}$ values in the downstream regions of TSSs. In contrast, male-biased genes had larger CpG$_{O/E}$ values than nonbiased genes (fig. 2*A*). In the gonads, both female-biased and male-biased genes had extremely larger CpG$_{O/E}$ values in the downstream regions of TSSs than nonbiased genes (fig. 2*B*). In the tail, only male-biased genes had larger CpG$_{O/E}$ values in the downstream regions of TSSs than nonbiased genes (fig. 2*C*). Furthermore, in the brain and tail, there were no significant differences in GpC$_{O/E}$ and GC skew between sex-biased and nonbiased genes in almost all windows 1,000 bp downstream and upstream regions from TSSs (fig. 2*D, F, G,* and *I*). In contrast, in the gonads, GpC$_{O/E}$ and GC skew differed between sex-biased genes and nonbiased genes in wide regions around TSSs (fig. 2*E* and *H*).

The Effect of CpG$_{O/E}$ on Sex-Biased Gene Expression

GAM analysis was conducted to examine the effect of CpG distribution on sex-biased gene expression. In this analysis, the effects in the upstream and downstream regions of TSSs were analyzed separately. In several GAM regressions, a negative relationship between CpG$_{O/E}$ and log fold change of gene expression ratio was found (table 3 and fig. 3). The effect of CpG$_{O/E}$ on sex-biased gene expression differed depending on the tissues and biased gene types. In addition, the negative effects were more prominent in the downstream regions than upstream regions. In male-biased genes of the gonads and tail, when CpG$_{O/E}$ in the downstream regions of TSSs were around 0.2–0.5, male-biased gene expression was more prominent (fig. 3*D* and *F*).

Discussion

Herein, we demonstrated that the distribution of CpG$_{O/E}$ in promoter regions of the *P. reticulata* genome exhibited a bimodal pattern categorized into the low-CpG and high-CpG classes (fig. 1*A*). This observed pattern of CpG$_{O/E}$ distribution
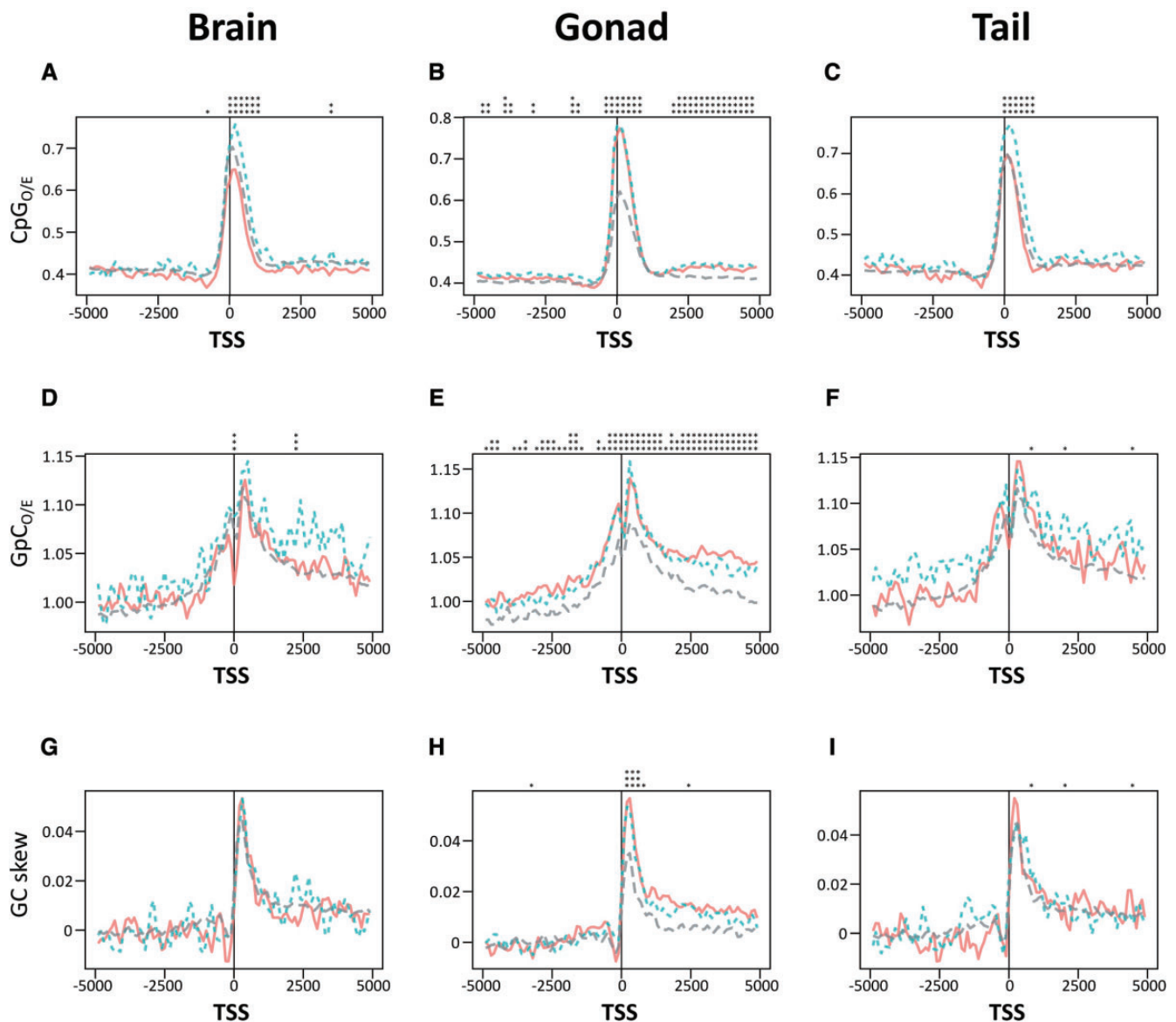
FIG. 2—Distribution patterns of $CpG_{O/E}$ (A, B, and C), $GpC_{O/E}$ (D, E, and F), and GC skew (G, H, and I) in 1,000 bp upstream and downstream regions from TSSs of brain (left), gonad (middle), and tail tissues (right). Red, blue, and gray lines indicate female-biased genes, male-biased genes, and nonbiased genes, respectively. ANOVA was conducted in each window to determine whether $CpG_{O/E}$, $GpC_{O/E}$, and GC skew values in a window differed between sex-biased and nonbiased genes. Significance is indicated for odd number windows: $*P < 0.01$; $**P < 0.001$; $***P < 0.0001$.

was similar to that found in zebrafish as highlighted in the work of Elango and Soojin (2008). As mentioned above, bimodal distribution of $CpG_{O/E}$ may be due to evolutionary loss of CpG, owing to the high mutation rate of methylated cytosines (Antequera 2003; Weber et al. 2007). In the present study, since the distribution of $GpC_{O/E}$ and GC skew exhibited unimodal distribution patterns, observed bimodality of $CpG_{O/E}$ might also be due to the variable degree of DNA methylation in different promoters in germline cells. Elango and Soojin (2008) suggested that the bimodal pattern of $CpG_{O/E}$ distribution, which is characteristic of vertebrate promoters, was

obtained gradually through the evolution of the vertebrate genome.

Our results showed that the proportions of sex-biased genes in the low-CpG and high-CpG classes differed significantly from those of all the genes, with the exception of the female-biased genes in the tail (table 2). The sliding window analysis showed that $CpG_{O/E}$ calculated in the regions of around TSSs of sex-biased genes was different from those of nonbiased gene (fig. 2). Both female-biased and male-biased genes showed extremely high $CpG_{O/E}$, especially in the gonads. In addition, $GpC_{O/E}$ and

Table 3

Summary of GAM Analysis Fitted to Absolute Value of $\log_2$(Gene Expression in Males/Gene Expression in Females) Variation by $CpG_{O/E}$ Calculated in 1,000 bp Upstream and Downstream Regions of TSSs

Tissue	Sex-Bias	Region	edf	F Statistics	P Value
Brain	Female	Upstream	2.367	8.904	<0.01
		Downstream	2.737	41.990	<0.01
	Male	Upstream	1.000	0.448	0.503
		Downstream	1.936	1.295	0.319
Gonad	Female	Upstream	1.626	27.040	<0.01
		Downstream	4.475	25.290	<0.01
	Male	Upstream	1.000	3.370	0.0665
		Downstream	5.241	35.360	<0.01
Tail	Female	Upstream	1.080	0.129	0.708
		Downstream	2.876	4.711	<0.01
	Male	Upstream	1.000	9.148	<0.01
		Downstream	6.734	12.600	<0.01

GC skews around the TSSs were also high (fig. 2B, E, and H). High $GpC_{O/E}$ values observed in the promoter sequences indicated that the sequences contain many G and C nucleotides (or CpG island). It has been shown that GC skew could be found in sequences with CpG islands, and sequences with high GC skew induced histone 3-lysine 4 trimethylation (H3K4 me3) and promote an open chromatin state (Ginno et al. 2012; Chen et al. 2014). Moreover, Sharma et al. (2014) found that the sex-biased genes in gonadal tissues showed a considerably nonrandom distribution among the chromosomes. This nonrandom distribution could be related to transcriptional regulation through the modification of chromatin state, similar to the X chromosome inactivation in mammals (Wutz 2011; Cotton et al. 2015). The present study utilized the data sets of sex differences in gene expression from Sharma et al. (2014), which had used mature gonads containing numerous germline cells. Thus, the observed patterns in this study might be related to the regulation mechanisms expressing genes homeostatically for gametogenesis in germline cells. In the brain and tail, $CpG_{O/E}$ distribution around the TSSs of male-biased genes was different from that of nonbiased genes, but neither $GpC_{O/E}$ distribution nor GC skew differed between the male-biased and nonbiased genes. Furthermore, the $CpG_{O/E}$ distribution around the TSSs of female-biased genes in the brain was lower than that of the nonbiased genes. In addition, sex-biased genes of the brain and tail distributed randomly among the chromosomes in contrast to the gonads (Sharma et al. 2014). These results suggested that the transcriptional regulation through CpGs around the TSSs is different for different tissue types.

GAM analysis showed that genes with lower $CpG_{O/E}$ around TSSs showed highly sex-biased gene expression than those with higher $CpG_{O/E}$. We speculate that the promoters with a low CpG density produce a highly differential gene expression between the two sexes, while those with a high CpG density do not. Especially in the gonads and tail, male-biased genes with intermediate $CpG_{O/E}$ (from 0.2 to 0.5) in the downstream regions of TSSs showed greater differences in gene expression between sexes than the other male-biased genes. Although the effects of CpG islands on transcriptional control has been well-studied (Han and Zhao 2008; Deaton and Bird 2011; Jones 2012), DNA methylation states in promoters without CpG islands and the effects of DNA methylation status on the transcriptional regulation in different tissue types remains unclear. Weber et al. (2007) suggested that, except for some genes, DNA methylation states in promoters without CpG islands could not be associated with gene regulation. However, recent studies (Han et al. 2011; Øster et al. 2013) found that DNA methylation in promoters without CpG islands could silence or weaken gene expression in human cells. These studies support the possibility that promoters maintaining low CpG density are associated with transcriptional control for differential gene expression between sexes in fish.

Our results also showed that the difference in $CpG_{O/E}$ between sex-biased and nonbiased genes was larger in the downstream regions of TSSs than those in upstream regions (fig. 2). Furthermore, the results of GAM analysis showed that the negative correlation between $CpG_{O/E}$ of downstream regions of TSSs and sex-biased gene expression was more prominent than that of upstream regions (fig. 3). These results suggest that transcriptional control, such as DNA methylation, along the downstream regions of the TSSs regulates sex-biased gene expression. In a temperature-dependent sex reversal fish, *Cynoglossus semilaevis*, found sex-biased methylation in downstream as well as upstream regions of TSSs (Shao et al. 2014). Brenet et al. (2011) showed that DNA methylation in not only regulatory sites of the gene but also sites in the gene body, particularly the first exon, affected transcriptional regulation. In addition, genes with higher CpG densities in the downstream regions of TSSs were expressed more (Krinner et al. 2014). These results support the hypothesis that CpGs and their methylated status along the downstream regions of TSSs affect transcriptional regulation.

The construction and maintenance of sex differences could be affected by not only genomic differences but also epigenetic modification, such as DNA methylation. In vertebrates, genomic regions may be methylated in accordance with CpG densities and distributions, and thus, CpG densities could be used as an index of the transcriptional control mediated via DNA methylation. The present study indicates that transcriptional regulation, such as DNA methylation, along the downstream regions of TSSs in low-CpG class genes plays a crucial role in large sex-biased gene expressions. These tendencies are supposed to differ according to the tissue type and developmental stages; therefore, gene expression under more varied conditions should be examined in the future. Further studies are needed to examine the genetic regulations
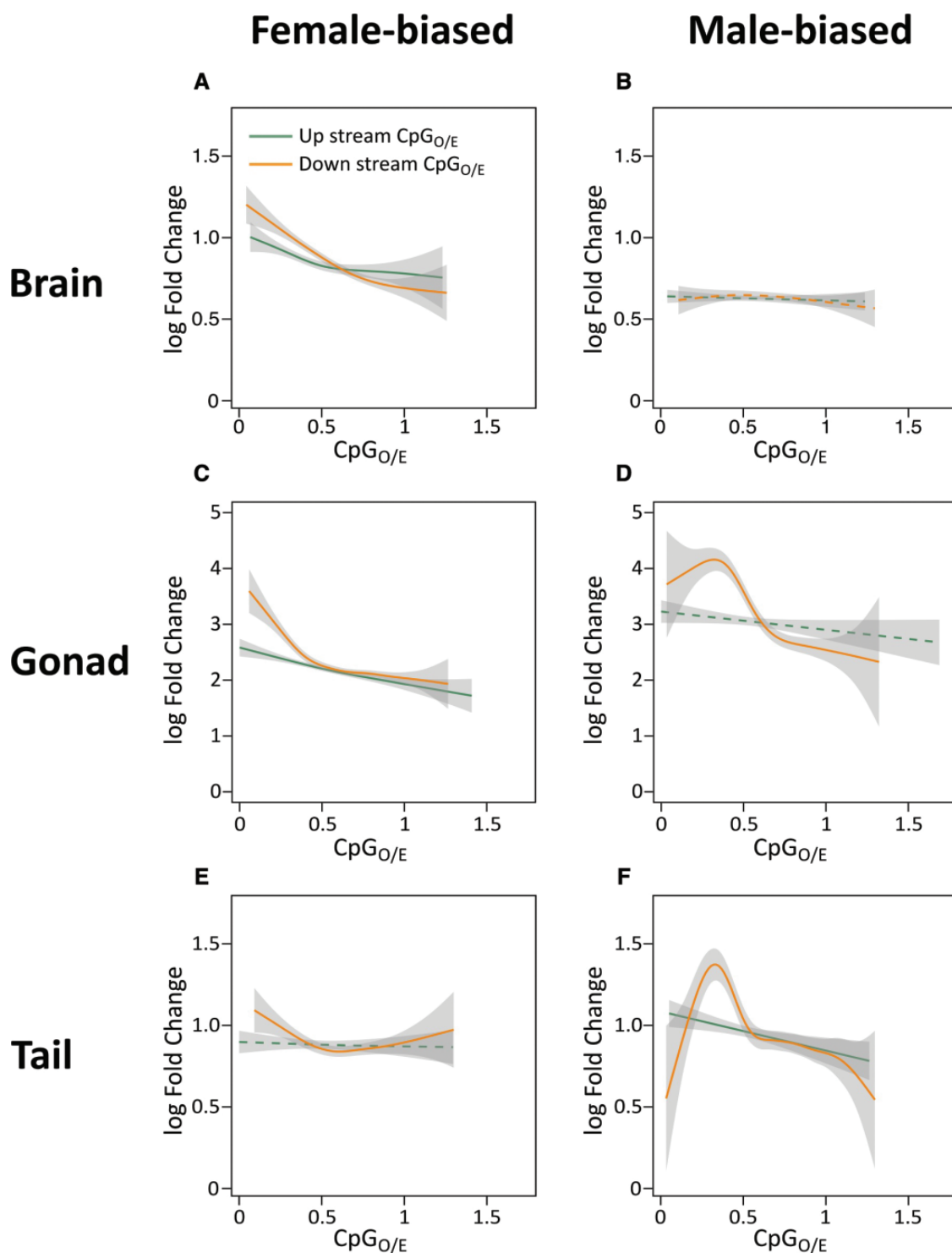
FIG. 3—GAM estimates for the effects of $CpG_{O/E}$ calculated in 1,000 bp upstream (green line) or downstream (orange line) regions from TSSs on the ratio of male to female expression of a specific gene (log Fold Change, absolute value of $\log_2$ [gene expression in male/gene expression in female]). The left column shows female-biased genes and the right column shows male-biased genes. Solid and dashed lines show that the estimated lines were significant ($P < 0.05$) and nonsignificant, respectively.

through DNA methylation in promoters without CpG islands and downstream of TSSs in order to clarify the mechanism governing sex-biased gene expression.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

Literature Cited

Antequera F. 2003. Structure, function and evolution of CpG island promoters. Cell Mol Life Sci. 60:1647–1658.

Bird A. 2002. DNA methylation patterns and epigenetic memory. Genes Dev. 16:6–21.

Brenet F, et al. 2011. DNA Methylation of the first exon is tightly linked to transcriptional silencing. PLoS One 6(1):e14524.

Brooks R. 2000. Negative genetic correlation between male sexual attractiveness and survival. Nature 406(6791):67–70.

Brooks R, Endler JA. 2001. Direct and indirect sexual selection and quantitative genetics of male traits in guppies (*Poecilia reticulata*). Evolution 55(5):1002–1015.

Chen RAJ, et al. 2014. Extreme HOT regions are CpG dense promoters in *C. elegans* and humans. Genome Res. 24:1138–1146.

Cotton AM, et al. 2015. Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. Hum Mol Genet. 24(6):1528–1539.

Deaton AM, Bird A. 2011. CpG islands and the regulation of transcription. Genes Dev. 25:1010–1022.

Elango N, Soojin YV. 2008. DNA methylation and structural and functional bimodality of vertebrate promoters. Mol Biol Evol. 25:1602–1608.

Ginno PA, Lott PL, Christensen HC, Korf I, Chédin F. 2012. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol Cell 45:814–825.

Godwin J. 2009. Social determination of sex in reef fishes. Semin Cell Dev Biol. 20:264–270.

Griffiths SW, Magurran AE. 1998. Sex and schooling behaviour in the Trinidadian guppy. Anim Behav. 56(3):689–693.

Han H, et al. 2011. DNA methylation directly silences genes with non-CpG island promoters and establishes a nucleosome occupied promoter. Hum Mol Genet. 20 (22):4299–4310.

Han L, Zhao Z. 2008. Comparative analysis of CpG islands in four fish genomes. Comp Funct Genomics 94:117–124.

Hendry AP, Kelly ML, Kinnison MT, Reznick DN. 2006. Parallel evolution of the sexes? Effects of predation and habitat features on the size and shape of wild guppies. J Evol Biol. 19:741–754.

Houde AE. 1997. Sex, color, and mate choice in guppies. New Jersey: Princeton University Press.

Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 13:484–492.

Jones PA, Takai D. 2001. The role of DNA methylation in mammalian epigenetics. Science 293(5532):1068–1070.

Kang HJ, et al. 2011. Spatio-temporal transcriptome of the human brain. Nature 478:483–489.

Klose RJ, Bird AP. 2006. Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci. 31:89–97.

Krinner S, et al. 2014. CpG domains downstream of TSSs promote high levels of gene expression. Nucleic Acids Res. 42(6):3551–3564.

Künstner A, et al. 2016. The Genome of the Trinidadian Guppy, Poecilia reticulata, and Variation in the Guanapo Population. PLoS One 11(12):e0169087.

Matsumoto Y, Buemio A, Chu R, Vafaee M, Crews D. 2013. Epigenetic control of gonadal aromatase (cyp19a1) in temperature-dependent sex determination of red-eared slider turtles. PLoS One 8(6):e63599.

Navarro-Martín L, et al. 2011. DNA methylation of the gonadal aromatase (cyp19a) promoter is involved in temperature-dependent sex ratio shifts in the European sea bass. PLoS Genet. 7(12):e1002447.

Øster B, et al. 2013. Non-CpG island promoter hypomethylation and miR-149 regulate the expression of SRPX2 in colorectal cancer. Int J Cancer 132:2303–2315.

Ott J. 1992. NOCOM and COMPMIX programs release. New York: Rockefeller University.

Piferrer F. 2013. Epigenetics of sex determination and gonadogenesis. Dev Dyn. 242:360–370.

Postma E, Spyrou N, Rollins LA, Brooks RC. 2011. Sex-dependent selection differentially shapes genetic variation on and off the guppy Y chromosome. Evolution 65(8):2145–2156.

Razin A, Riggs AD. 1980. DNA methylation and gene function. Science 210:604–610.

Saxonov S, Berg P, Brutlag DL. 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc Natl Acad Sci USA. 103:1412–1417.

Shao C, et al. 2014. Epigenetic modification and inheritance in sexual reversal of fish. Genome Res. 24:604–615.

Sharma E, et al. 2014. Transcriptome assemblies for studying sex-biased gene expression in the guppy, *Poecilia reticulata*. BMC Genomics 15:400.

Tao W, et al. 2013. Characterization of gonadal transcriptomes from Nile tilapia (*Oreochromis niloticus*) reveals differentially expressed genes. PLoS One 8:e63604.

Vinogradov AE. 2005. Dualism of gene GC content and CpG pattern in regard to expression in the human genome: magnitude versus breadth. Trends Genet. 21:639–643.

Weber M, et al. 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet. 39:457–466.

West PM, Packer C. 2002. Sexual selection, temperature, and the lion's mane. Science 297 (5585):1339–1343.

Wijchers PJ, Festenstein RJ. 2011. Epigenetic regulation of autosomal gene expression by sex chromosomes. Trends Genet. 27:132–140.

Williams TM, Carroll SB. 2009. Genetic and molecular insights into the development and evolution of sexual dimorphism. Nat Rev Genet. 10:797–804.

Wutz A. 2011. Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. Nat Rev. 12:542–553.

Xu H, et al. 2014. Sex-biased methylome and transcriptome in human prefrontal cortex. Hum Mol Genet. 23:1260–1270.

Yang X, et al. 2006. Tissue-specific expression and regulation of sexually dimorphic genes in mice. Genome Res. 16:995–1004.

Zi J, et al. 2003. Coloration strategies in peacock feathers. Proc Natl Acad Sci USA. 100:12576–12578.

Associate editor: Maria Costantini