Female-Biased Expression on the X Chromosome as a Key Step in Sex Chromosome Evolution in Threespine Sticklebacks

Erica H. Leder,*[1] Jose M. Cano,[2] Tuomas Leinonen,[2] Robert B. O'Hara,[3] Mikko Nikinmaa,[1] Craig R. Primmer,[1] and Juha Merilä[2]

[1]Division of Genetics and Physiology, Department of Biology (Vesilinnantie 5), University of Turku, Turku, Finland

[2]Ecological Genetics Research Unit, Department of Biosciences, University of Helsinki, Helsinki, Finland

[3]Biodiversity and Climate Research Centre, Frankfurt am Main, Germany

***Corresponding author:** E-mail: eriled@utu.fi.

Associate editor: Yoko Satta

Abstract

Given that the genome of males and females are almost identical with the exception of genes on the Y (or W) chromosome or sex-determining alleles (in organisms without sex chromosomes), it is likely that many downstream processes resulting in sexual dimorphism are produced by changes in regulation. In early stages of sex chromosome evolution, as the Y-chromosome degenerates, gene expression should be significantly impacted for genes residing on the sex chromosome pair as regulatory mutations accumulate. However, this has rarely been examined because most model organisms have clearly diverged sex chromosomes. Fish provide a unique opportunity to examine the evolution of sex chromosomes because genetic sex determination has evolved quite recently in some groups of fish. We compared sex-specific transcription in threespine stickleback (*Gasterosteus aculeatus*) liver tissue using a long-oligo microarray. Of the 1,268 genes that were differentially expressed between sexes, a highly significant proportion (23%) was concentrated on chromosome 19, corresponding to the recently described nascent sex chromosomes. The sex-biased genes are enriched for different functional categories in males and females, although there is no specific functional enrichment on the sex chromosomes. Female-biased genes are concentrated at one end of the sex chromosome, corresponding to a deletion in the Y, suggesting a lack of global dosage compensation. Prior research on threespine sticklebacks has demonstrated various degrees of dissimilarity in upstream regions of genes on the Y providing a potential mechanism for the observed patterns of female-biased expression. We hypothesize that degeneration of the Y chromosome results in regulatory mutations that create a sex-specific expression pattern and that this physical concentration of sex-biased expression on the nascent sex chromosome may be a key feature characterizing intermediate phases of sex chromosome evolution.

Key words: threespine stickleback, *Gasterosteus aculeatus*, sex chromosome evolution, sex-biased expression.

Introduction

Understanding the molecular mechanisms controlling sex determination and the evolution of sex-specific differences are fundamental problems in biology. Although dimorphic sex chromosomes have evolved repeatedly in animals (Charlesworth 1996), many vertebrates not only lack dimorphic sex chromosomes but also genetic sex determination (Ezaz et al. 2006). Much of the research on the evolution of sex chromosomes in nonmammalian vertebrates has come from cytogenetic studies and genetic mapping (reviewed in Ezaz et al. 2006), but little is known about the evolution of the genetic composition within sex chromosomes in these organisms. Fish provide a unique opportunity to examine the evolution of sex chromosomes since they exhibit large variation in sex determination and sex chromosome systems (Ezaz et al. 2006). Of the 1,700 fish species that have been karyotyped, only 176 have morphologically distinct sex chromosomes (Devlin and Nagahama 2002). In addition, genetic sex determination may have evolved quite recently in some groups of fish, including poeciliids (Volff and Schartl 2001), cichlids (Cnaani et al. 2008), salmonids (Woram et al. 2003), and sticklebacks (Ross et al. 2009).

It has been suggested that the early stages of vertebrate sex chromosome evolution begin with the development of barriers to recombination between the proto-Y (or W) and the X (or Z) in the region containing the sex-determining locus (Charlesworth et al. 2005; Bachtrog 2006a). This suppression of recombination leads to the degeneration of the Y (or W) resulting in morphologically distinct chromosomes (Rice 1987; Charlesworth B and Charlesworth D 2000). Additionally, hitchhiking of genes closely linked to the sex-determining locus and reduced recombination in this region favor the evolution of sexually dimorphic traits by allowing sexually antagonistic alleles to persist (Rice 1984; Bachtrog 2006a).

Regardless of whether or not an organism has genetic sex determination, once the decision to become male or female has been made, a sex-specific pattern of gene expression must be maintained to continue that trajectory.

Expression differences occur in tissues specifically involved in reproduction (e.g., gonads Santos et al. 2007; Mank et al. 2008) but also in other tissues (Yang et al. 2006; Mank et al. 2008), resulting in sex-specific morphology, physiology, and behavior. Because many human diseases exhibit sex-specific patterns of susceptibility, onset, and symptoms, understanding the evolution of sex-specific expression patterns thus has important implications for human health (e.g., Williams et al. 2003; Woods et al. 2003).

The evolution of sex-biased gene expression patterns may be tightly linked to the evolution of sex chromosomes in the same manner as morphological sexual dimorphism is suggested to be linked to sex chromosome evolution (Rice 1984). However, as previous studies of sex-biased expression have either focused on organisms with well-developed X/Y or W/Z chromosome systems (e.g., Ranz et al. 2003; Yang et al. 2006) or on organisms with no sex chromosomes (Santos et al. 2007, 2008), the potential importance of expression modification during the intermediate stages of sex chromosome evolution has rarely been investigated. In *Drosophila Miranda*, it has been observed that many genes on the neo-Y chromosome are downregulated regardless of whether or not the gene produces a functional product, suggesting random inactivation (Bachtrog 2006b). Further examination of organisms with newly emerging sex chromosomes will allow for a better understanding of the degeneration process and how it affects gene expression.

The threespine stickleback (*Gasterosteus aculeatus*) is an ideal target for such a study. In addition to the availability of a detailed linkage map (Peichel et al. 2001, 2004) and genome sequence information (http://www.ensembl.org/Gasterosteus_aculeatus/Info/Index), it has been recently suggested that one pair of chromosomes, chromosome group 19, are actually nascent sex chromosomes because the sex-determining locus maps to this chromosome, and this chromosome pair exhibits other characteristics typical of sex chromosomes, such as reduced recombination, an excess of repetitive elements, and lower sequence homology (Peichel et al. 2004). Through a combination of linkage mapping (Peichel et al. 2004; Ross and Peichel 2008) and cytogenetic (Ross and Peichel 2008) approaches, it has been revealed that these are distinct heteromorphic chromosomes. Additionally, cytogenetic evidence using fluorescent in situ hybridization (FISH) of BAC clones containing known markers have identified regions that are inverted in the Y chromosome as well as a large fragment that is missing from the Y (Ross and Peichel 2008). Given the above evidence supporting a recently evolved sex chromosome system in threespine sticklebacks, a significant feature is that the Y chromosome has begun to degenerate (Peichel et al. 2004; Ross and Peichel 2008). Hence, the question becomes: what is the impact of chromosome degeneration on the transcription pattern of genes on these chromosomes? A better understanding of such processes provides important information for understanding the intermediate phases of sex chromosome evolution and the evolution of sexual dimorphism.

Materials and Methods

Sampling of the Study Populations and Rearing of the Offspring

Fish from three study populations were included in this study, which was part of a study examining expression in different populations under thermal stress. The three populations were: Helsinki (Baltic Sea; 60°10′N, 25°00′E), Lake Pulmanki (Finnish Lapland; 69°58′N, 27°58′E), and Lake Vättern in Sweden (58°54′N, 14°24′E). Full-sib F1 families were created by crossing parental fish at the sampling sites, and fertilized eggs were transported to the laboratory. Initially, the offspring were maintained with water at 17 ± 1 °C and photoperiod of 18 h light 6 h dark, and six months after hatching, the environmental conditions were gradually changed to complete darkness (24-h dark) and 9 ± 1 °C to simulate wintering conditions. After five months, the environmental conditions were changed back to 18/6 h L/D photoperiod, 17 ± 1 °C, and new crosses were made to obtain F2 offspring from each population. At the time of the experiments, the F2 offspring were approximately 20 months old. They were adult fish although they were reproductively inactive at the time of the experiment.

Experimental Sampling

Twenty fish from each population were randomly selected. Equal numbers of fish from each population were allocated randomly to two identical tanks with recirculating water, and each fish was kept in a separate container that allowed water to flow through. Fish were acclimated overnight. In the morning, one tank was heated approximately 1 °C per hour for 6 h (from 17 °C to 24 °C). As fish were sampled from the tanks, they were immediately euthanized in a lethal dose of MS-222 anesthetic. Livers were removed from fish immediately, frozen in liquid nitrogen, and stored at −80 °C. Twelve fish per population were used for the microarrays (six per treatment). Because fish were not reproductively active, it was impossible to determine sex before the treatment. After treatment, fish were dissected and sex was determined by examining internal morphology. There were equal numbers of females in the treatment and control groups (four each), but the resulting sex ratio was skewed toward males with eight females and 28 males used for the arrays. The treated group contained one female and five males from Lake Pulmanki, one female and five males from Lake Vättern, and two females and four males from Helsinki. The control group contained 6 males from Lake Pulmanki, two females and four males from Lake Vättern, and two females and four males from Helsinki. As discussed below, there was no statistically significant interaction between treatment and sex or population and sex, so the results reported here are discussed only in terms of differences related to sex.

Microarray Design

Microarrays were designed using the custom gene expression 4 × 44 K platform from Agilent (Agilent, Santa Clara, CA) which consists of four arrays per slide with 45,220

features, 43,793 of which are user-designed 60-bp oligonucleotide probes (Leder et al. 2009).

Sample Preparation and Array Hybridization

Total RNA was isolated by means of Tri Reagent (Sigma, St Louis, MO) using the manufacturer's protocol. RNA was treated with DNase (Promega, Madison, WI) and reisolated using Tri Reagent. RNA concentration was quantified using a Nanodrop ND-1000, and RNA quality was assessed using an Experion automated electrophoresis system (Bio-Rad, Hercules, CA).

RNA labeling, hybridizations, and scanning were performed by an Agilent certified commercial service provider, the Finnish DNA Microarray Centre. Briefly, total RNA (400 ng) was amplified and Cy3-labeled with Agilent's Low RNA Input Linear Amplification Kit PLUS, One color (Agilent) along with Agilent's One-Color RNA Spike-in Kit following the manufacturer's protocols. After the labeling, the cRNA was examined with the Nanodrop ND-1000 and the Experion Automated Electrophoresis System cRNA to assess the concentration and quality of the labeling. Each sample (1.65 μg) was hybridized to the custom-designed stickleback array at 65 °C overnight (17 h) using Agilent's GE Hybridization Kit. Washes were conducted as recommended by the manufacturer using Agilent's Gene Expression Wash Pack without any stabilization or drying solution. Arrays were scanned with Agilent Technologies Scanner, model G2505B. Spot intensities and other quality control features were extracted with Agilent's Feature Extraction Software version 9.5.3.1.

Data Analysis

Array quality was assessed through the use of Agilent control features as well as spike-in controls (Agilent One-Color Spike-in Kit for RNA experiment). Due to poor hybridization, one array from the RNA experiment (one male) was removed from further analysis. Processed signals from the Feature Extraction Software (v 9.5.3.1) were used for the analysis in the Limma analysis package of R/Bioconductor (Smyth 2005). The Feature Extraction Software automatically flags features that are above background at the 99% confidence interval but to further filter the data within each population and within each sex, only probes with a background-corrected median intensity value of greater than 50 across the respective group were kept for analysis. Because probes with very low-expression levels may not be reliable, this is similar to keeping only the probes that are more than two times as intense as background levels because the median background for most features was less than 50.

Arrays were normalized using quantile normalization (Smyth and Speed 2003), and significance was determined using a linear model with a Bayesian method to moderate the standard errors (Smyth 2004). The variables in the model were population, treatment, and sex and the interactions sex by treatment and population by treatment. The interactions between the population and sex were not used because the sample sizes for females within each population were too low (see above). The interactions between sex and treatment and population and treatment were not significant at the 0.05 level with the Benjamini and Yekutieli (2001) false discovery rate correction for multiple tests, so the model was reapplied without the interaction terms.

Significant Gene Lists

Because many genes were represented by more than one unique probe, genes were deemed significant if at least one probe was significant with the correction for multiple tests, and the remaining probes were expressed in the same direction and significant at the 5% level without the correction for multiple tests. Thirty-nine genes were removed from further analysis because they did not meet these criteria. For sex-biased expression, the significance was based on the comparison between male expression and female expression, and it was considered female-biased if expression in females was higher than expression in males. Likewise, expression was considered male-biased when male expression was higher than female expression.

Ortholog Annotation

Threespine stickleback genes were initially matched to their human orthologs using Biomart from Ensembl (http://www.ensembl.org/biomart/martview/). Blast searches were also used to increase the annotation information. For genes with a human ortholog, an Entrez gene ID was obtained if possible to facilitate further analysis. Approximately 1,700 genes (out of 11,238) that had a significant intensity were not assigned a human Entrez identifier.

Functional Enrichment Analysis

The Database for Annotation, Visualization, and Integrated Discovery (DAVID) was used to determine if there was any significant overrepresentation of genes with particular functional categories, that is, functional enrichment, among sex-biased genes (http://david.abcc.ncifcrf.gov/) (Dennis et al. 2003). DAVID will accept several identifiers, but human Entrez gene identifiers gave the most comprehensive annotation lists for stickleback orthologs. DAVID uses Gene Ontology terms at several levels to cluster genes by related functional terms, and the significance of the cluster is assessed using a modified Fisher's exact test (Hosack et al. 2003). Enrichment is assessed by comparing the functional categories of the gene list (genes of interest) with theoretical proportions of gene functional categories which are determined from the functional categories of a background genome or population list. Several comparisons were conducted to determine if there were any patterns related to sex-biased expression and gene function on chromosome 19. The numbers of genes used for the comparisons in DAVID are less than the original numbers of significant genes due to genes with unmapped identifiers (i.e., no functional information in the database). All female-biased, all male-biased, and all sex-biased genes on chromosome 19 were each used as an input gene list with all active genes as the background population list. Female-biased and male-biased genes were

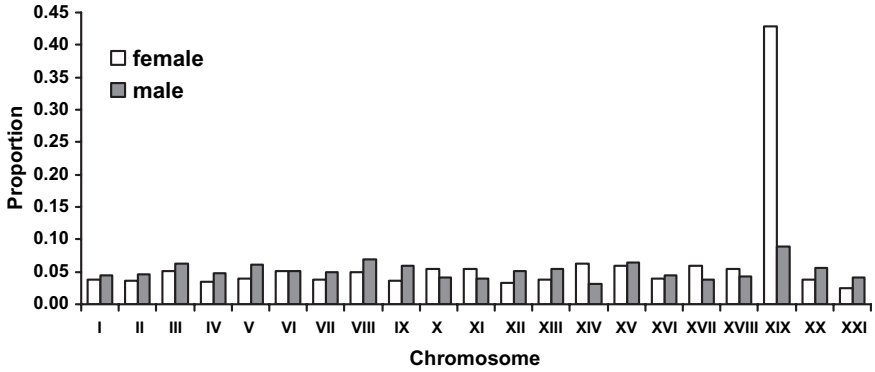

Fig. 1. Proportion of threespine stickleback genes upregulated by sex for a given chromosome. For example, there were 627 genes that were actively expressed on chromosome 1, of these, 24 were female-biased (0.038). For chromosome 19, the graph is presented with all the biased genes which has a significance of $P = 3.74 \times 10^{-265}$ for females and $P = 0.031$ for males for nonuniform distribution of biased genes among chromosomes. When the approximate deleted region of the Y chromosome (12–20.2 Mb) is removed from the analysis, there is still significant nonuniform distribution for sex-biased genes among chromosomes (χ^2, females $P = 1.72 \times 10^{-51}$ and males $P = 1.23 \times 10^{-4}$), as the proportion of female-biased genes is 0.26 and male-biased genes is 0.13 (not shown).

further examined by fold change. Additionally, sex-biased genes on chromosome 19 were tested for enrichment using all sex-biased genes as the background population.

Results

We compared female and male transcription patterns in threespine stickleback liver tissue using a custom-designed long-oligo DNA microarray representing 19,275 genes of the estimated 20,700 genes in the stickleback genome. Because these fish were also involved in a thermal stress experiment, we examined the interaction of thermal stress and sex for significant sex by treatment effects but found no significant interaction. Because the thermal stress treatment did not appear to differentially impact either sex, we proceeded with a separate analysis of sex-biased expression. Of the 11,239 genes with significant signal intensities, 1,268 (11.3%) were differentially transcribed between the sexes (supplementary table 1, Supplementary Material online), 704 of which showed upregulation in females relative to males (female-biased), and 564 showed upregulation in males relative to females (male-biased). Genes with sex-biased mRNA expression were assigned to chromosomes based on the currently available threespine stickleback genome sequence (http://www.ensembl.org/Gasterosteus_aculeatus/index.html). These sex-biased genes showed an uneven distribution with respect to the 21 known chromosome groups with almost a quarter (23.2%) of the genes concentrating on chromosome 19 (fig. 1) that corresponds to the nascent sex chromosomes (Peichel et al. 2004; Ross and Peichel 2008). Additionally, the majority of sex-biased genes on chromosome 19 exhibited female-biased expression (82.7%).

The distribution of sex-biased genes along chromosome 19 was examined to see if biased expression occurred at specific sites or regions along the chromosome (fig. 2).

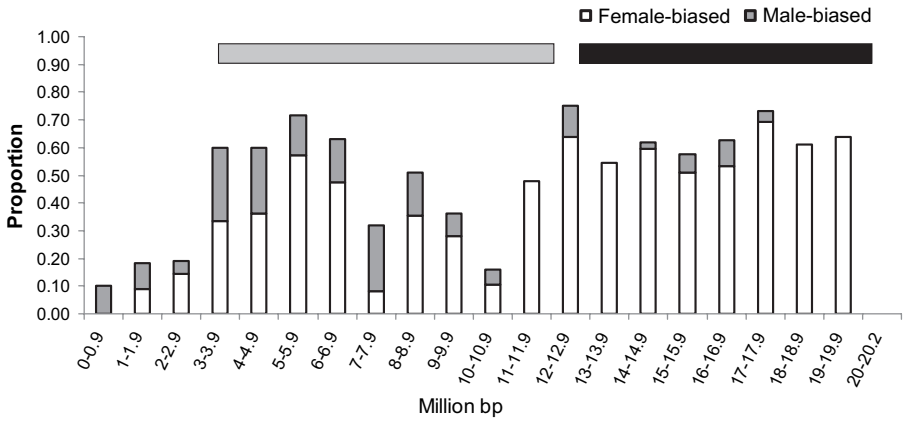

Fig. 2. Distribution of sex-biased genes on the sex chromosomes (chromosome 19) by sequence position. Proportions are the number of genes biased in one sex in a 1-Mb interval compared with the actively expressed genes in that interval. Base positions are from the published female sequence data but are modified according to Ross and Peichel (2008). The black box above the histogram indicates the approximate region absent from the Y chromosome (Ross and Peichel 2008). The gray box indicates the approximate region that is no longer recombining between the X and Y due to ancestral inversions (Ross and Peichel 2008). Numbers of genes expressed and numbers of sex-biased genes per interval are presented in supplementary table 3 (Supplementary Material online).

Table 1. Comparison of the Significance of Uneven Distribution of Sex-Biased Genes among Different Regions of the Sex Chromosomes.

	P for even distribution	
Chromosome (19) region	**Female-biased**	**Male-biased**
1–11.9 Mb	1.72×10^{-51}	1.23×10^{-4}
1–3.2 Mb	0.225	0.509
3.3–11.9 Mb	2.90×10^{-66}	5.70×10^{-5}
1–20.2 Mb (all)	3.74×10^{-265}	0.031

NOTE.—*P* values are for χ^2 tests comparing the expected genome-wide distribution of female-biased and male-biased genes using the proportion of sex-biased genes in that region of chromosome 19.

Because it was discovered that the orientation of markers on chromosome 19 of the genetic map do not correspond to the orientation of supercontig 3 of chromosome 19 in Ensembl (Ross and Peichel 2008), this supercontig was flipped and the gene positions renumbered accordingly for this analysis. With the new gene order, there appear to be very few male-biased genes in the second half of the chromosome (from 10 to 20.2 Mb), and the proportion of female-biased genes is significantly greater than in the beginning of the chromosome (χ^2, $P = 3.37 \times 10^{-10}$; fig. 2).

As it has been documented that there is a deletion at the end of the male copy of chromosome 19, the distribution of sex-biased genes among chromosomes was reanalyzed with these genes omitted to determine if these genes were the sole cause of the bias. It is unclear where the exact break point is located, but it is estimated that it is after the isocitrate dehydrogenase (IDH2) gene (at 11.2 Mb), so even though this corresponds to a larger deletion than suggested by Ross and Peichel (2008), 8.2 Mb instead of 6 Mb, we removed the genes from 12 Mb to the end of the chromosome (20.2 Mb) from the analysis. Although there were many female-biased genes in this deleted region, there was still a highly significant nonuniform distribution with respect to female-biased genes in the nondeleted region (table 1). There was also a significant nonuniform distribution of male-biased genes, although there were still significantly more female-biased genes on chromosome 19 (χ^2, $P = 3.30 \times 10^{-4}$). Additionally, if only the nonrecombining portion is used (3.2–11.9 Mb), there is a significant nonuniform distribution among chromosomes for both males and females, but if the recombining portion of the chromosome is used (1–3.2 Mb), there is no significant difference for the distribution of sex-biased genes among chromosomes for females or males.

Table 2. Chromosomes with Overrepresentation or Underrepresentation of Sex-Biased Genes by Fold Change.

Fold change	All sex-biased	Male-biased	Female-biased
>1.2	—	—	**19**
>1.5	**19**	—	**(1, 7, 8, 12) 19**
>2.0	**(4, 5, 9, 11, 14) 19**	**19**	**(4, 9, 11, 14, 20) 19**
>3.0	**(2, 4) 19**	**19**	**19**

NOTE.—Underrepresented chromosomes are in parentheses. Those that are still significant at the 0.05 level after Bonferroni correction for multiple comparisons are in bold.

The distribution of sex-biased genes among chromosomes was further categorized by the degree of fold change. For smaller fold changes, between 1.2 and 1.5, only chromosome 19 had a slightly significant overrepresentation of female-biased genes (table 2), but this was not significant after correcting for multiple tests using the Bonferonni correction. Considering all sex-biased genes, chromosome 19 had a significant overrepresentation for all other categories of fold change (>1.5, >2, >3). This was also the case for female-biased genes which is likely driving the significance for all sex-biased genes because the male-biased genes are only significantly overrepresented for the >3 fold-change category when using the multiple testing correction. Several chromosomes had an underrepresentation of sex-biased genes, but none of these were significant after Bonferonni corrections.

Because genes residing on the sex chromosomes have unique effects on the separate sexes due to dosage differences, we examined the functions of the genes that exhibited sex-biased expression. Functional categories for sex-biased genes were examined with DAVID functional annotation tool. Of the 11,239 genes that were actively expressed on the array, 9,484 had had human orthologs with Entrez identifiers, although only 8,264 of these were mapped to unique DAVID ids. Overall, female-biased genes and male-biased genes represented different functional categories (table 3, supplementary table 2, Supplementary Material online). Female-biased genes were enriched for genes involved in ribosome-related activities, translation, and intracellular activities, whereas male-biased genes were more involved in signaling and extracellular activities. Interestingly, despite containing almost 25% of the sex-biased genes, chromosome 19 had no specific functional enrichment when compared with either the active genes on all other chromosomes or the sex-biased genes on the other chromosomes.

Sex-biased genes were further divided into fold-change categories to determine if a particular level of differential expression (e.g., greater than 2-fold) was characterized by specific functional categories. In females, genes that exhibited higher fold changes compared with males (greater than 2-fold and greater than 3-fold) exhibited no significant functional enrichment, nor did genes that had low levels of differential expression (greater than 1.2-fold) and high levels of differential expression (greater than 3-fold) in males. The majority of functional categories that exhibited significant enrichment, when separated by fold-change, were a subset of the overall functional enrichment by sex. However, there were three categories that were enriched for female-biased genes at the greater than 1.2-fold level but were not enriched when all female-biased genes were considered: cellular protein metabolic process, cellular macromolecule metabolic process, and regulation of translational initiation. For male-biased genes, the biological process "organ development" was enriched for genes at the greater than 1.5-fold level but not when

Table 3. Representative Functional Terms for Genes That Exhibited Enrichment for Female- and Male-Biased Genes.

Female-biased genes	Male-biased genes
Biological process (432)	**Biological process (363)**
Ribosome biogenesis and assembly-35	Circulatory system process-16
Gene expression-152	Cell adhesion-35
rRNA metabolic process-25	Cell surface receptor linked signal transduction-53
rRNA processing-24	Blood circulation-16
Translation-59	Amino acid catabolic process-12
Cellular biosynthetic process-85	Amine catabolic process-13
Cellular metabolic process-317	Biological adhesion-35
Macromolecule biosynthetic process-63	Multicellular organismal process-117
Macromolecule metabolic process-269	Nitrogen compound catabolic process-13
Cellular component (442)	**Cellular component (393)**
Intracellular-386	Extracellular region-62
Intracellular non-membrane–bound organelle-108	Proteinaceous extracellular matrix-27
Cytosolic ribosome (sensu Eukaryota)-21	Extracellular space-30
Chaperonin-containing T-complex-6	Cytoskeleton-48
Ribonucleoprotein complex-56	Plasma membrane-69
Eukaryotic translation initiation factor 3 complex-7	
Molecular function (471)	**Molecular function (389)**
Structural constituent of ribosome-33	Calcium ion binding-46
Translation factor activity, nucleic acid binding-19	
Structural molecule activity-42	
Methyltransferase activity-19	
Nucleotidyltransferase activity-18	

NOTE.—The numbers after the headings are the number of genes that had annotation for that process from the input gene list. The numbers after the enrichment categories are the number of genes representing that category in the input gene list. Full lists of terms are provided in supplementary table 2 (Supplementary Material online).

all male-biased genes were considered (supplementary table 2, Supplementary Material online).

Discussion

Although sex-biased expression has been demonstrated in other organisms (Ranz et al. 2003; Yang et al. 2006; Santos et al. 2007, 2008; Mank et al. 2008), the threespine stickleback transcription pattern presented here is unique: There is an extreme concentration of female-biased genes on the sex chromosomes, it is from somatic tissue (liver), and it has no functional enrichment on the sex chromosomes (e.g., reproductive function). In contrast, zebra fish, a species that lacks heteromorphic sex chromosomes and exhibits environmental sex determination, female-biased genes were not concentrated on any specific chromosome (Santos et al. 2008). At the other end of the evolutionary scale, in species with dimorphic sex chromosomes, genes which exhibit sex-biased expression on sex chromosomes are not nearly so biased toward the homogametic sex (Ranz et al. 2003; Yang et al. 2006; Mank et al. 2008), and in mammals are mostly biased toward males (Lercher et al. 2003; Yang et al. 2006). Additionally, in the few cases, when sex-biased genes are biased toward the homogametic sex on the sex chromosomes, it is from mRNA expression in gonadal tissue and is therefore involved in reproductive function (Parisi et al. 2003; Mank et al. 2008).

In mice, female-biased expression on the X chromosome occurs in liver tissue but only when considering genes with less than a 1.2-fold increase (Yang et al. 2006). When considering genes with a stronger expression bias (>1.2 or more), there is no significant overrepresentation for female-biased genes on the X chromosome, although the Y chromosome has an overrepresentation of male-biased genes for all fold changes (Yang et al. 2006). It is likely that the lack of female-biased expression on the X chromosome in mouse liver tissue is due to dosage compensation that inactivates one of the X chromosomes in females in order to compensate for the hemizygous state in the males (Charlesworth 1996). Because stickleback sex chromosomes are in the early stages of evolution, this female-biased expression pattern may be indicative of a lack of dosage compensation.

It is known that sex chromosomes have evolved repeatedly as the genes on the sex chromosomes of chicken and human are not homologous (Nanda et al. 2000) and the gene complement on the pseudoautosomal portion of the mammalian Y chromosomes is not the same across mammalian lineages (Waters et al. 2007). Additionally, the sex-determining region is located on nonhomologous chromosomes in at least four species of salmonids (Woram et al. 2003), and there is evidence for the independent origin of sex chromosomes in the guppy (Tripathi et al. 2009) and medaka (Takehana et al. 2007). However, there is evidence for homologous chromosomes becoming the sex chromosome in two different stickleback species (Ross et al. 2009). Theoretical modeling has demonstrated how linkage disequilibrium between a sex-determining locus and a sexually antagonistic allele can drive the turnover of sex chromosomes, assuming nonheteromorphic chromosomes as well as recombining chromosomes (van Doorn and Kirkpatrick 2007). Therefore, it seems likely that within closely related groups that sex chromosomes could originate either from homologous chromosomes as in some stickleback species or from nonhomologous chromosomes

as in medaka. A main factor affecting which route is taken would be the timing of speciation in relation to the timing of sex chromosome divergence.

When considering current models of sex chromosome evolution (Charlesworth 1996; Charlesworth et al. 2005; Bachtrog 2006a), the lack of functional enrichment on the nascent sex chromosomes of threespine stickleback fits with predictions of the early stages of Y chromosome divergence. Specifically, the overall gene content on the chromosome in which the sex-determining locus resides when genetic sex determination first becomes fixed and recombination is suppressed may be considered somewhat random. As the evolution of sex chromosomes is thought to be quite dynamic, genes will migrate on and off the sex chromosomes depending on whether there is a selective advantage for a given gene to reside on that chromosome (Sturgill et al. 2007; Bachtrog et al. 2008). Hence, we hypothesize that, as observed in this study, genes that are present on the newly forming sex chromosomes as the X and Y begin to differentiate should be arbitrary with respect to function, and thus sex-biased expression patterns on those chromosomes would not initially show specific functional enrichment.

In sticklebacks, the majority of sex-biased transcription on chromosome 19 was biased toward upregulation in females relative to males, and this is consistent with expression differences resulting from Y chromosome degeneration. More specifically, lack of recombination and accumulation of repetitive DNA on the Y chromosome may result in mutations in regulatory regions. This could reduce male expression due to misregulated Y genes and give the appearance of female-biased expression. It has already been demonstrated from female- and male-specific BAC clones from chromosome 19 that intergenic regions have much reduced homology in threespine stickleback (Peichel et al. 2004). Areas upstream of the coding regions and introns contain large insertions and deletions as well as having reduced homology (Peichel et al. 2004), and it is likely that there are *cis*-regulatory elements present in these regions that would affect transcription in the different sexes. Furthermore, it has been demonstrated that sex-biased patterns of transcription are a result of regulation differences (Kopp et al. 2000; Yang et al. 2006; Williams et al. 2008), so it is feasible that regulatory mutations due to Y chromosome degeneration could create this pattern of sex-biased expression in the same manner as misregulation occurs between species hybrids, as a result of nonhomologous regulatory regions (Wittkopp et al. 2004).

Regulatory differences have been suggested to be a major source of expression variation between the sexes (Yang et al. 2006) as well as among individuals (Stranger et al. 2007). However, in the case of sex chromosomes, due to the lack of recombination and subsequent degeneration of the Y, it is possible that some of the expression differences between the sexes are due to sequence divergence within the mRNA or coding sequence. In *D. miranda*, 47% of 118 genes on the neo-Y were nonfunctional due to deletion, frameshift mutations, or premature stop codons (Bachtrog et al. 2008). In threespine sticklebacks, coding sequence mutations could impact the observed expression between the sexes in two different ways. First, insertion or deletion mutations could create frameshift mutations resulting in premature or overdue stop codons and the resulting abnormal transcripts would be degraded through nonsense-mediated decay. This would result in female-biased expression because only transcripts originating from the Y should be affected. However, this could not be the cause of male-biased expression which, although was much less than female-biased expression, was also significantly higher than expected in the 3.2–12 Mb region of chromosome 19.

A second possibility is that there are mutations in the mRNA that still produce functional products, such as synonymous substitutions or 3′UTR mutations, but cause reduced binding with the probes. However, due to the array design, it is unlikely that this could contribute to a large proportion of the genes involved. The majority of threespine stickleback genes were represented by at least two separate probes per gene on the array, although in some cases, the probes had a large overlap (Leder et al. 2009). Of the 294 significant genes on the sex chromosomes, 275 were represented by more than one probe, and 104 (37.8%) of these probes were not overlapping. Additionally, 43 genes (15.6%) had probes that overlapped 50% or less, therefore more than half of the probes would cover from 90 to 120 bases of the target gene. The probe sequence coverage is important if there is potential for sequence divergence. Without sequencing each significant gene, it is impossible to determine the sequence divergence, but based on data from Peichel et al. (2004), coding sequences are not as diverged as intronic or intergenic regions. For two genes examined that study, the divergence was 1.65% for 545 bp of exon sequence and 4.06% for 271 bp of 3′UTR sequence.

Based on studies of mismatch probes, there are many factors contributing to the hybridization signal of long oligonucleotide probes with the main factors being number of mismatches, position of mismatched bases within the probe sequence, G/C content of the probe sequence, and hybridization temperature (Letowski et al. 2004). A single-base mismatch within a 60-bp probe can reduce median intensity from 0% to 25% depending on the position of the mismatch, with centered mismatches having the greatest effect (Rennie et al. 2008). These data were from inbred strains of mice where the mismatches were homozygous. In the stickleback, any mismatch between X and Y chromosomes would be heterozygous so any reduction in intensity due to a mismatch is likely to be even lower. For individuals with multiple probes and little to no overlap, the array design would allow for a small degree of mismatch without any affect on the overall results. For instance, a one-base mismatch corresponds to a 1.67% divergence, and if multiple probes are representing a given gene, the number of tolerated mismatches could be even higher depending on how they are distributed within the probes.

Mutations are also occurring on the Y on a much larger scale and may be contributing to the observed expression

patterns. It has been determined that there has been a 6-Mb deletion in the Y after the IDH2 gene (located at 11.2 Mb) in threespine stickleback of Pacific and North American origin (Ross and Peichel 2008). The high proportion of female-biased genes and the paucity of male-biased genes in the last 10 Mb of chromosome 19 support these data and suggest that a global dosage compensation mechanism has yet to evolve in sticklebacks. However, there may be some local dosage compensation because not all the genes in the deleted region were upregulated in females, and there are even genes that are male-biased in this region.

It is further suggested that only the first 3.2 Mb of the sex chromosomes are recombining (Ross and Peichel 2008). In these data, there are much fewer sex-biased genes in the first 3 Mb than in any other 3 Mb region on the sex chromosomes which is more similar to the frequency of sex-biased genes in the autosomes. As such, when this region is analyzed separately, there is no significant difference in the proportion of sex-biased genes in this region compared with the proportion of sex-biased genes on the autosomes.

Cytogenetic experiments have revealed a different gene order for some genes on the sex chromosomes that may have been the result of several pericentric inversions in the Y chromosome (Ross and Peichel 2008). Overall it is estimated from 3.2 to 11.8 Mb (based on the female sequence) was affected by inversions, but multiple inversions in the same region may have reverted a portion of the Y chromosome to its original gene order. Without a Y-specific sequence, it is impossible to determine the exact boundaries. The impact of these inversions on gene expression does not seems to be obvious, but it is hypothesized that the whole area is not recombining (Ross and Peichel 2008) and thus genes on the X and the Y could be expressed differently due to regulatory mutations. It should also be noted that the positioning of the genes on the chromosome (fig. 2) is not accurate for the Y chromosome, however, there are similar numbers of male-biased genes throughout the affected region so it is likely that the pattern would be similar if we could determine the exact location of the genes on the Y chromosome.

Our data provide an important insight into the evolution of sex chromosomes and indicate that physical concentration of genes with sex-biased expression may be a key feature characterizing the intermediate phases of sex chromosome evolution. This study also provides the first whole-genome study linking intermediate phases of sex chromosome evolution, Y-chromosome degeneration, and sex-biased expression. To include these data into the current models of chromosome evolution, it follows that as recombination is reduced and repetitive DNA increases on the Y chromosome, mutations in the regulatory regions alter the expression patterns of the proto-Y chromosome resulting in a concentration of sex-biased expression in the region of reduced recombination. Additionally, coding sequence mutations could result in truncated products that get degraded and result in biased expression in the homogametic sex. It is predicted that further evolution of the sex chromosomes should result in the evolution of a dosage compensation mechanism as well as migration of male-biased genes off the X chromosome (Charlesworth 1996; Charlesworth et al. 2005; Bachtrog 2006a; Sturgill et al. 2007). Future examination of sex-biased gene regulation as well as sequence divergence on the Y chromosome in threespine sticklebacks and closely related species should allow for further insight into the specific mechanisms of sex-biased gene expression and its relationship to the intermediate phases of sex chromosome evolution.

Supplementary Material

Supplementary tables 1–3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank the Finnish Academy of Science (Centre of Excellence in Evolutionary Genetics and Physiology) for funding, Reija Venho and The Finnish DNA Microarray Centre—Turku Centre for Biotechnology for sample labeling and the microarray processing, and Wolfgang Wasser and Anti Vasemägi for assistance with sample collection. Additionally, we thank two anonymous reviewers for their very useful criticism, comments, and suggestions.

References

Bachtrog D. 2006a. A dynamic view of sex chromosome evolution. *Curr Opin Genet Dev.* 16:578–585.

Bachtrog D. 2006b. Expression profile of a degenerating neo-y chromosome in Drosophila. *Curr Biol.* 16:1694–1699.

Bachtrog D, Hom E, Wong KM, Maside X, de Jong P. 2008. Genomic degradation of a young Y chromosome in Drosophila miranda. *Genome Biol.* 9:R30.

Benjamini Y, Yekutieli D. 2001. The control of the false discovery rate in multiple testing under dependency. *Ann Stat.* 29:1165–1188.

Charlesworth B. 1996. The evolution of chromosomal sex determination and dosage compensation. *Curr Biol.* 6:149–162.

Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. *Philos Trans R Soc Lond B Biol Sci.* 355:1563–1572.

Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* 95:118–128.

Cnaani A, Lee BY, Zilberman N, et al. (15 co-authors). 2008. Genetics of sex determination in tilapiine species. *Sex Dev.* 2:43–54.

Dennis G Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. 2003. DAVID: database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 4:P3.

Devlin RH, Nagahama Y. 2002. Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture* 208:191–364.

Ezaz T, Stiglec R, Veyrunes F, Marshall Graves JA. 2006. Relationships between vertebrate ZW and XY sex chromosome systems. *Curr Biol.* 16:R736–R743.

Hosack D, Dennis G, Sherman B, Lane H, Lempicki R. 2003. Identifying biological themes within lists of genes with EASE. *Genome Biol.* 4:R70.

Kopp A, Duncan I, Godt D, Carroll SB. 2000. Genetic control and evolution of sexually dimorphic characters in Drosophila. *Nature* 408:553–559.

Leder EH, Merilä J, Primmer CR. 2009. A flexible whole-genome microarray for transcriptomics in three-spine sticklebacks (Gasterosteus aculeatus). *BMC Genomics*. 10:426.

Lercher MJ, Urrutia AO, Hurst LD. 2003. Evidence that the human X chromosome is enriched for male-specific but not female-specific genes. *Mol Biol Evol.* 20:1113–1116.

Letowski J, Brousseau R, Masson L. 2004. Designing better probes: effect of probe size, mismatch position and number on hybridization in DNA oligonucleotide microarrays. *J Microbiol Methods.* 57:269–278.

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.

Nanda I, Zend-Ajusch E, Shan Z, et al. (14 co-authors). 2000. Conserved synteny between the chicken Z sex chromosome and human chromosome 9 includes the male regulatory gene DMRT1: a comparative (re)view on avian sex determination. *Cytogenet Cell Genet.* 89:67–78.

Parisi M, Nuttall R, Naiman D, Bouffard G, Malley J, Andrews J, Eastman S, Oliver B. 2003. Paucity of genes on the Drosophila X chromosome showing male-biased expression. *Science* 299: 697–700.

Peichel CL, Nereng KS, Ohgi KA, Cole BL, Colosimo PF, Buerkle CA, Schluter D, Kingsley DM. 2001. The genetic architecture of divergence between threespine stickleback species. *Nature* 414: 901–905.

Peichel CL, Ross JA, Matson CK, Dickson M, Grimwood J, Schmutz J, Myers RM, Mori S, Schluter D, Kingsley DM. 2004. The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. *Curr Biol.* 14:1416–1424.

Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL. 2003. Sex-dependent gene expression and evolution of the Drosophila transcriptome. *Science* 300:1742–1745.

Rennie C, Noyes H, Kemp S, Hulme H, Brass A, Hoyle D. 2008. Strong position-dependent effects of sequence mismatches on signal ratios measured using long oligonucleotide microarrays. *BMC Genomics*. 9:317.

Rice WR. 1984. Sex chromosomes and the evolution of sexual dimorphism. *Evolution* 38:735–742.

Rice WR. 1987. Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. *Genetics* 116: 161–167.

Ross JA, Peichel CL. 2008. Molecular cytogenetic evidence of rearrangements on the Y chromosome of the threespine stickleback fish. *Genetics* 179:2173–2182.

Ross JA, Urton JR, Boland J, Shapiro MD, Peichel CL. 2009. Turnover of sex chromosomes in the stickleback fishes (gasterosteidae). *PLoS Genet.* 5:e1000391.

Santos EM, Kille P, Workman VL, Paull GC, Tyler CR. 2008. Sexually dimorphic gene expression in the brains of mature zebrafish. *Comp Biochem Physiol A Mol Integr Physiol.* 149:314–324.

Santos EM, Workman VL, Paull GC, Filby AL, Van Look KJ, Kille P, Tyler CR. 2007. Molecular basis of sex and reproductive status in breeding zebrafish. *Physiol Genomics.* 30:111–122.

Smyth GK. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol.* 3:Article 3.

Smyth GK. 2005. Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W, editors. Bioinformatics and computational biology solutions using R and bioconductor. New York: Springer. p. 397–420.

Smyth GK, Speed T. 2003. Normalization of cDNA microarray data. *Methods* 31:265–273.

Stranger BE, Nica AC, Forrest MS, et al. (14 co-authors). 2007. Population genomics of human gene expression. *Nat Genet.* 39:1217–1224.

Sturgill D, Zhang Y, Parisi M, Oliver B. 2007. Demasculinization of X chromosomes in the Drosophila genus. *Nature* 450:238–241.

Takehana Y, Demiyah D, Naruse K, Hamaguchi S, Sakaizumi M. 2007. Evolution of different Y chromosomes in two medaka species, Oryzias dancena and O. latipes. *Genetics* 175:1335–1340.

Tripathi N, Hoffmann M, Weigel D, Dreyer C. 2009. Linkage analysis reveals the independent origin of Poeciliid sex chromosomes and a case of atypical sex inheritance in the guppy (Poecilia reticulata). *Genetics* 182:365–374.

van Doorn GS, Kirkpatrick M. 2007. Turnover of sex chromosomes induced by sexual conflict. *Nature* 449:909–912.

Volff JN, Schartl M. 2001. Variability of genetic sex determination in poeciliid fishes. *Genetica* 111:101–110.

Waters PD, Wallis MC, Graves JAM. 2007. Mammalian sex–origin and evolution of the Y chromosome and SRY. *Semin Cell Dev Biol.* 18:389–400.

Williams AJ, Norcross AJ, Dix RJ, Gillespie KM, Gale EA, Bingley PJ. 2003. The prevalence of insulin autoantibodies at the onset of Type 1 diabetes is higher in males than females during adolescence. *Diabetologia* 46:1354–1356.

Williams TM, Selegue JE, Werner T, Gompel N, Kopp A, Carroll SB. 2008. The regulation and evolution of a genetic switch controlling sexually dimorphic traits in Drosophila. *Cell* 134:610–623.

Wittkopp PJ, Haerum BK, Clark AG. 2004. Evolutionary changes in *cis* and *trans* gene regulation. *Nature* 430:85–88.

Woods SC, Gotoh K, Clegg DJ. 2003. Gender differences in the control of energy homeostasis. *Exp Biol Med.* 228:1175–1180.

Woram RA, Gharbi K, Sakamoto T, et al. (16 co-authors). 2003. Comparative genome analysis of the primary sex-determining locus in salmonid fishes. *Genome Res.* 13:272–280.

Yang X, Schadt EE, Wang S, Wang H, Arnold AP, Ingram-Drake L, Drake TA, Lusis AJ. 2006. Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res.* 16:995–1004.