The Master Sex-Determination Locus in Threespine Sticklebacks Is on a Nascent Y Chromosome

Catherine L. Peichel,[1,2,*] Joseph A. Ross,[2,3]
Clinton K. Matson,[2] Mark Dickson,[4]
Jane Grimwood,[4] Jeremy Schmutz,[4]
Richard M. Myers,[4] Seiichi Mori,[5] Dolph Schluter,[6]
and David M. Kingsley[1]
[1]Department of Developmental Biology and Howard
    Hughes Medical Institute
Stanford University School of Medicine
Palo Alto, California 94305
[2]Division of Human Biology
Fred Hutchinson Cancer Research Center
Seattle, Washington 98109
[3]Molecular and Cellular Biology Program
University of Washington
Seattle, Washington 98195
[4]Stanford Human Genome Center
Department of Genetics
Stanford University School of Medicine
Palo Alto, California 94304
[5]Biological Laboratory
Gifu Keizai University
Ogaki, Gifu 503-8550
Japan
[6]Department of Zoology
University of British Columbia
Vancouver V6T 1Z4
Canada

Summary

Background: Many different environmental and genetic sex-determination mechanisms are found in nature. Closely related species can use different master sex-determination switches, suggesting that these developmental pathways can evolve very rapidly. Previous cytological studies suggest that recently diverged species of stickleback fish have different sex chromosome complements. Here, we investigate the genetic and chromosomal mechanisms that underlie sex determination in the threespine stickleback (*Gasterosteus aculeatus*).
Results: Genome-wide linkage mapping identifies a single chromosome region at the distal end of linkage group (LG) 19, which controls male or female sexual development in threespine sticklebacks. Although sex chromosomes are not cytogenetically visible in this species, several lines of evidence suggest that LG 19 is an evolving sex chromosome system, similar to the XX female/XY male system in many other species: (1) males are consistently heterozygous for unique alleles in this region; (2) recombination between loci linked to the sex-determination region is reduced in male meiosis relative to female meiosis; (3) sequence analysis of X- and Y-specific bacterial artificial chromosome (BAC) clones from the sex-determination region reveals many sequence differences between the X- and Y-specific

clones; and (4) the Y chromosome has accumulated transposable elements and local duplications.
Conclusions: Taken together, our data suggest that threespine sticklebacks have a simple chromosomal mechanism for sex determination based on a nascent Y chromosome that is less than 10 million years old. Further analysis of the stickleback system will provide an exciting window into the evolution of sex-determination pathways and sex chromosomes in vertebrates.

Introduction

The decision to become a male or a female is one of the most fundamental in the life of an organism. However, the mechanisms that govern this choice are diverse, spanning myriad environmental and genetic sex-determination systems. Diverse sex-determination systems can be found in closely related species, suggesting that sex-determination mechanisms can evolve very rapidly. Genetic and molecular analyses of sex-determination pathways have identified many of the key regulatory molecules and pathways underlying sex-determination pathways in *C. elegans*, *D. melanogaster* and eutherian mammals [1, 2]. One important finding from this work is that master sex-determination genes are not conserved between divergent taxa [3–6]. This is surprising, given that many genes found in other developmental pathways are conserved between worms, flies, and mice [7]. However, recent data have provided evidence for conservation of some downstream genes in divergent organisms with very different mechanisms of sex determination [3, 5, 6, 8–9]. A theme emerging from these studies is that evolutionary lability at the top of the sex-determination hierarchy is coupled with stability toward the bottom of the hierarchy [10]. Comparisons between closely related species with divergent sex-determining mechanisms are now needed in order to understand how these developmental hierarchies evolve.

The evolution of genetic sex determination is often accompanied by the evolution of heteromorphic sex chromosomes. The first step in the evolution of a sex chromosome is the existence of an autosomal gene with two alleles, where homozygosity leads to the development of one sex and heterozygosity to the other sex. The evolution of heteromorphic sex chromosomes involves the suppression of recombination between homologous chromosomes, probably to reduce recombination between the sex-determination locus and linked genes with sex-specific fitness effects [11–13]. This suppression of recombination with the maintenance of one chromosome in a constant heterozygous state ultimately results in degeneration of sex-linked loci in the heterogametic sex [11, 13–14].

Several models have been put forth to explain the evolution of heteromorphic sex chromosomes [15–17], but empirical evidence that supports these theories is limited. While the complete sequence of the human Y chromosome provides an interesting picture of a highly

*Correspondence: cpeichel@fhcrc.org

evolved sex chromosome [18], the human Y chromosome is 300 million years old and does not provide insight into the mechanisms that underlie the initial stages of sex chromosome evolution. Empirical data from evolving Y chromosome systems in plant and *Drosophila* species have provided key insights into Y chromosome degeneration, but more independent examples are needed to provide a complete picture of this process [17, 19].

Fish are particularly attractive model systems in which to study the evolution of sex-determination mechanisms and sex chromosomes. Both environmental and genetic sex-determination mechanisms are represented in fish species [20]. In fish with environmentally determined sex, temperature or social interactions can be the primary sex determinant. Genetic mechanisms of sex determination in fish may be polygenic or simple and either associated with no cytogenetically visible sex chromosomes or with heteromorphic sex chromosomes in either males (XY systems) or females (ZW systems). This wide diversity of sex-determination mechanisms can be found even in closely related fish species, supporting the hypothesis that this developmental pathway is evolutionarily plastic and that sex-determination mechanisms and sex chromosomes can evolve very rapidly.

Stickleback fish (family Gasterosteidae) provide an excellent system in which to analyze the genetic and molecular mechanisms that underlie the evolution of sex determination and sex chromosomes. Threespine sticklebacks (*Gasterosteus aculeatus*) are small teleost fish found throughout the Northern hemisphere, and their behavior, ecology, and evolution have been extensively characterized [21]. Previous cytological studies suggest that threespine sticklebacks lack sexually dimorphic chromosomes, although closely related species do have distinguishable sex chromosomes in either males or females [22–23]. Some evidence has suggested that temperature and density of rearing could affect sex ratios in threespine sticklebacks [24]. However, an allozyme of isocitrate dehydrogenase (*IDH*) was found to be sexually dimorphic in multiple threespine stickleback populations in California and British Columbia [25–27], and DNA markers with sex-specific alleles have been reported [28]. These results suggest that there might be a genetic basis for sex determination in threespine sticklebacks despite the absence of visible sex chromosomes in this species.

Here, we take advantage of extensive new molecular tools [29–30] to map the genetic basis of sex determination in threespine sticklebacks, clone a chromosome region linked to the sex-determination locus, analyze the sex-linked regions for sequence characteristics typically found in evolving sex chromosomes, and study the evolution of sex-linked loci in different stickleback populations.

Results and Discussion

Male Sex Is Determined by a Single Chromosome Region in Threespine Sticklebacks

To determine if phenotypic sex in threespine sticklebacks is controlled primarily by genetic or environmental
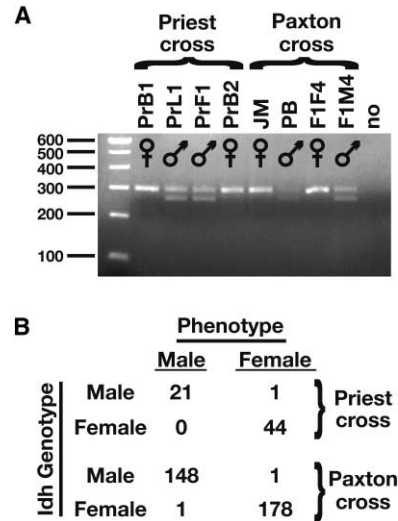

Figure 1. The Sex-Determination Locus Maps Near the *Idh* Gene in Threespine Sticklebacks

(A) Ethidium bromide stained agarose gel showing the pattern of the sex-specific polymorphism in the 3′UTR of the *Idh* gene in the grandparents and parents of both the Priest cross [37] and the Paxton cross [39–40]. The band at 302 bp is from the X chromosome and the band at 271 bp is from the Y chromosome. The following abbreviations are used: PrB1, Priest benthic female 1; PrL1, Priest limnetic male 1; PrF1, Priest F1 male 1; PrB2, Priest wild caught benthic female 1; JM, Japanese Pacific marine female; PB, Paxton benthic male; F1F4, F1 female parent of family 4; F1M4, F1 male parent of family 4; NO, no DNA control.

(B) Comparison of *Idh* genotype with phenotypic sex. Of 66 animals genotyped for *Idh* and phenotyped for sex in the Priest cross, one animal is recombinant. Of 328 animals genotyped for *Idh* and phenotyped for sex in the Paxton cross, two animals are recombinant.

factors, we analyzed the progeny of two independent genetic crosses: the Priest backcross [29] and the Paxton intercross [31–32]. The progeny of each cross were genotyped with the large panel of microsatellite markers developed for threespine sticklebacks [29]. In both crosses, development of testes showed nearly perfect concordance with the inheritance of markers at the distal end of linkage group (LG) 19.

Previous data identified a male-specific protein polymorphism in isocitrate dehydrogenase (IDH) in multiple populations from the west coast of North America [25–27]. We identified the *Idh* gene in a cDNA clone isolated in our original screen for microsatellite markers [29]. We designed PCR primers to the 3′UTR of *Idh* and found that the primers amplified a 302 bp band from females, but from males, they amplified both a 302 bp band and a 271 bp band (Figure 1). This polymorphism segregates nearly perfectly with phenotypic sex in both the Priest and Paxton crosses, suggesting that it is very tightly linked to a master sex-determination locus in threespine sticklebacks (Figure 1).

Recombination Rates Are Reduced around the Sex-Determination Locus

When the segregation of LG 19 markers from the male and female parents was analyzed in both the Paxton and Priest crosses, large sex-specific differences were
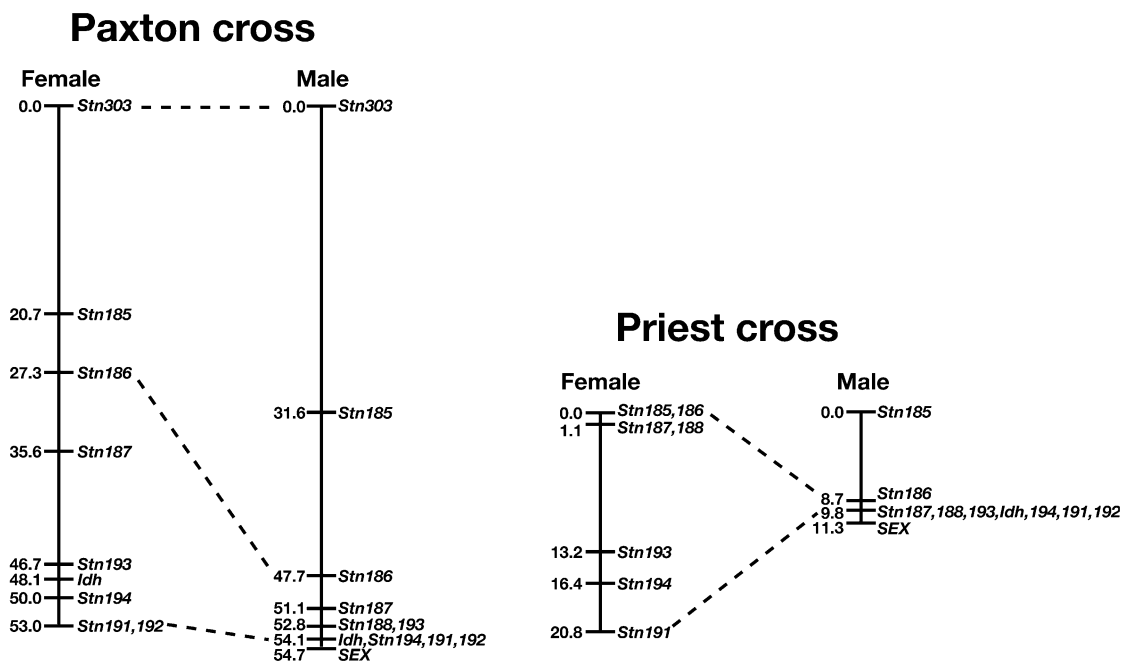
Paxton cross

Figure 2. Male and Female Recombination Rates Differ Near *SEX*

Female and male meiotic maps of LG 19 in the Paxton and Priest crosses are shown. Distance in centimorgans (cM) is indicated along each map. Microsatellite markers are designated with a prefix of *Stn* followed by a number. Marker *Stn303* could not be scored in the Priest cross, accounting for the difference in total map length between the two crosses. Markers that do not recombine with each other are listed on the same line. The sex-determination locus (*SEX*) is located at the end of LG 19 in both crosses.

seen in the calculated genetic distances along the linkage group (Figure 2). For example, the distance between *Stn186* and *Stn191* in the Paxton female meiotic map is 25.7 cM but only 6.4 cM in the Paxton male meiotic map (p < 0.001). A similar reduction in the male meiotic map in the Priest cross is seen in this region, where the 20.8 cM between *Stn186* and *Stn191* in the female meiotic map is reduced to a 1.1 cM interval in the male meiotic map (p < 0.001). This regional difference in recombination rate does not reflect a lower general rate of recombination in male meiosis, because the genetic distance between *Stn303* and *Stn186* in the Paxton cross is twice as large in the male meiotic map as it is in the female meiotic map (47.7 cM versus 27.3 cM; p < 0.005). These results suggest that recombination along this linkage group is suppressed, although not completely absent, in males around the sex-determination locus. The presence of rare recombinants in this interval is surprising; therefore, the genotype and phenotype of all putative recombinants was reconfirmed. Future work will reveal whether the observed suppression of recombination results from an inversion of this region in males relative to females or if it results from the accumulation of heterochromatin [15].

In both the Priest and the Paxton crosses, three of the sex-linked microsatellite markers (*Stn187*, *Stn191*, and *Stn192*) had null alleles segregating from the male parent (Table S1 in the Supplemental Data available with this article online). Null alleles can either result from a deletion of the microsatellite locus or from polymorphisms in the primer sites used to amplify the microsatellite. The high rate of male-specific null alleles on this linkage group suggests that the male-specific chromo-

some harbors significant differences from the homologous chromosome.

Our genetic data provide evidence that threespine stickleback have a single chromosome region that determines male sex, that males are heteromorphic for markers in this region, and that there is a male-specific reduction in recombination around these sex-linked markers. Based on these and the following results, we propose that threespine stickleback females have two X chromosomes and that males have an X and a Y chromosome.

Sequence Analysis Near the Sex-Determination Region

To analyze the molecular region around the sex-determination locus, a bacterial artificial chromosome (BAC) library made from a pool of DNA from 60 male and female sticklebacks [30] was screened with an *Idh* probe. Twenty-one BACs were identified with this probe, and the sex-specific polymorphism in the *Idh* 3′UTR was subsequently used to type them as being from the X or Y chromosome. Fifteen BACs were X chromosome specific and six BACs were Y chromosome specific. These BACs were oriented relative to each other by using PCR primers designed from BAC end sequences (data not shown). Two X-specific BACs (101E8 and 160O9) and two Y-specific BACs (169J23 and 119K16) that extended in opposite directions with the least amount of overlap were sequenced to completion.

A combination of BLAST and GENSCAN analysis revealed there are homologs of five known genes in this region on both the X and Y chromosomes: Semaphorin 4B (*Sema4B*), NADP-dependent isocitrate dehydroge-
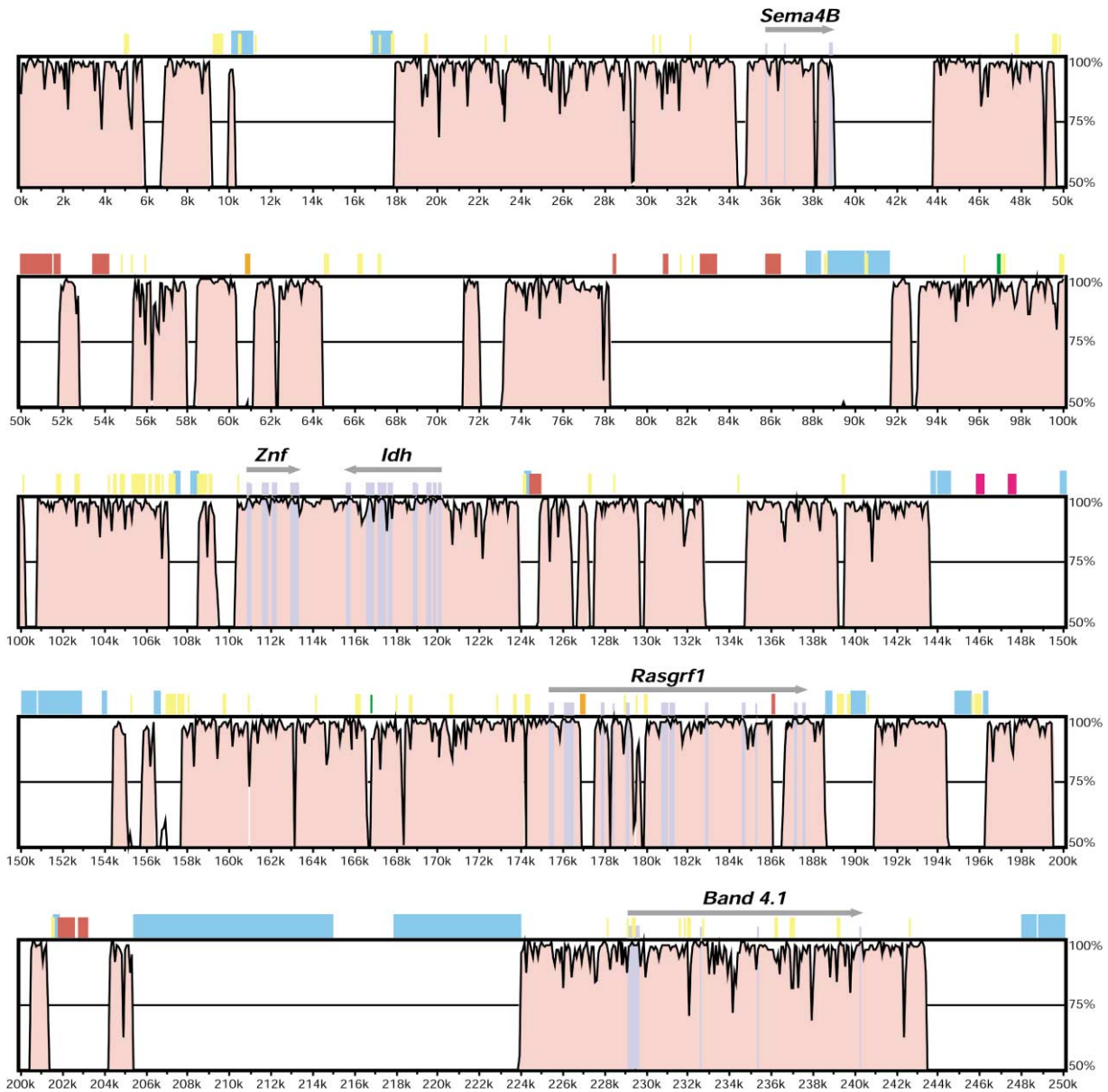
Figure 3. Poor Sequence Homology between X and Y Chromosome Sequence Contigs

VISTA plot comparing 250 kb of sequence from the X and Y chromosome sequence contigs. Shaded pink areas indicate regions of high sequence homology. Purple boxes indicate exons of putative genes, and the arrows indicate the direction of transcription. *Sema4B* is Semaphorin 4B; *Znf* is a zinc-finger-containing gene; *Idh* is NADP-dependent isocitrate dehydrogenase; *Rasgrf1* is Ras protein-specific guanine nucleotide-releasing factor 1; and *Band 4.1* is a protein containing a Band 4.1 domain. Red boxes indicate the position of long interspersed nuclear elements (LINEs), green boxes indicate the position of short interspersed nuclear elements (SINEs), magenta boxes indicate the position of LTR-containing retroviral elements, orange boxes indicate the position of DNA mobile elements, and yellow boxes indicate position of other repeats such as simple-sequence and low-complexity repeats. The blue boxes indicate the position of stickleback Y chromosome-specific repeats. The entire alignment used to generate Table 1 is not shown. Therefore, the numbers of repeats shown are a subset of those listed in Table 1. For this analysis, a 100 nt window size and 95% stringency were used.

nase (*Idh*), a zinc-finger-containing gene (*Znf*), Ras protein-specific guanine nucleotide-releasing factor 1 (*Rasgrf1*), and a protein containing a Band 4.1 domain (*Band 4.1*) (Figure 3). All five of these genes are found on human chromosome 15q24-26, suggesting that the distal end of LG 19 in sticklebacks has regions of conserved synteny with human chromosome 15. The *Idh* gene does not map to the sex chromosome of *Xiphophorus* [33], suggesting that the chromosomal basis of sex

determination has likely evolved independently in these different fish species.

**Poor Sequence Homology between
the X and Y Chromosome**

Suppression of recombination around a sex-determination locus with subsequent degeneration of sex-linked loci in the heterogametic sex is a hallmark of sex chromosomes [13]. Therefore, we expected to find poor ho-
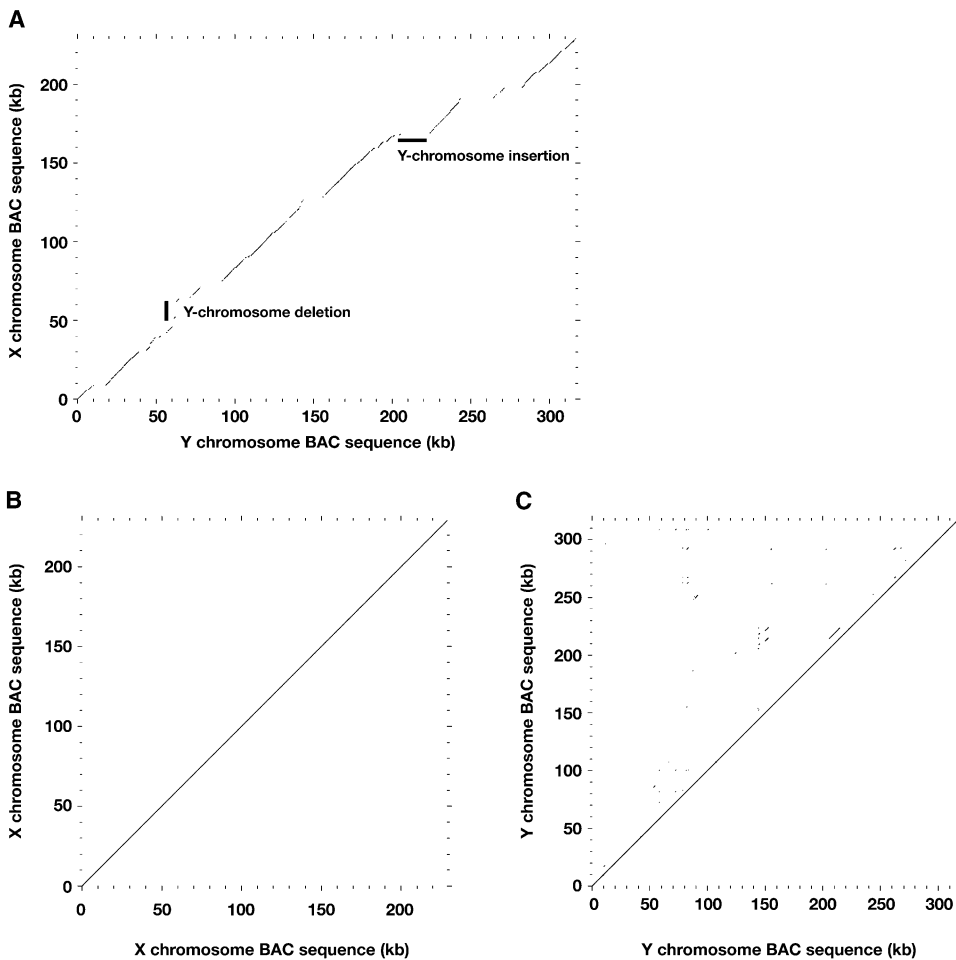
A

B

C

Figure 4. The Y Chromosome Contains Many Internal Repeats

Dot plots comparing (A) the X chromosome sequence contig to the Y chromosome sequence contig; (B) the X chromosome sequence contig to itself; (C) the Y chromosome sequence contig to itself.

(A) The broken lines in the X-Y comparison indicate that there are multiple insertions and a single major deletion in the Y chromosome sequence relative to the X chromosome sequence.

(B) The single diagonal line in the X-X comparison shows that the sequence aligns perfectly with itself over its entire length. The absence of other lines in the dot plot suggests that there are no internally repeated sequences in this region of the X chromosome.

(C) The single diagonal line in the Y-Y comparison shows that the sequence aligns over its entire length. However, the smaller lines above the diagonal indicate that there are multiple local duplications contained in this region of the Y chromosome. For this analysis, a 100 nt window size and 95% stringency were used.

mology between the X and Y chromosome sequences. Global alignment of the X-specific and Y-specific contigs revealed only 63.7% sequence identity over their length. When the global alignment between the X and Y chromosome sequences is visualized with VISTA, it is evident that this low-sequence identity results from blocks of very high homology that are interrupted by large gaps of poor homology (Figure 3). These gaps in homology are most likely due to insertions on the Y chromosome for four reasons. First, the X and Y chromosome sequence contigs are anchored with the same end sequences, but the Y chromosome sequence contig is 87 kb longer. Second, a dot plot comparing the X and Y chromosome sequence contigs shows that there are multiple insertions in the Y chromosome sequence relative to the X chromosome sequence (Figure 4A). Third, dot plot analysis reveals that there are numerous local

duplications on the Y chromosome (Figure 4C). Fourth, many of these gaps in homology contain repetitive DNA elements, particularly novel stickleback-specific repeats (Figure 3).

Accumulation of Transposable Elements and Duplications on the Y Chromosome

Accumulation of repetitive DNA and transposable elements is predicted to occur on a Y chromosome [13]. To determine whether previously characterized transposable elements exist on both the X and the Y chromosomes, RepeatMasker was used on the X and Y chromosome sequences. This analysis showed that the Y chromosome has more repetitive elements than the homologous region on the X chromosome (Table 1). The VISTA plot shows the positions of repetitive and transposable elements in the stickleback sex determination

Table 1. Repeat Content in X and Y Chromosome Contigs

Repeat Type	X Chromosome Sequence (229012 bp)			Y Chromosome Sequence (316654 bp)		
	Number	Length (bp)	% of Total Sequence	Number	Length (bp)	% of Total Sequence
SINE	0	0	0	4	210	0.07
LINE	0	0	0	13	12070	3.81
LTR	0	0	0	1	623	0.20
DNA	2	97	0.04	3	498	0.16
Simple Sequence	47	1884	0.82	60	3796	1.20
Total	49	1981	0.86	81	17197	5.44

The following abbreviations are used: SINE, short interspersed nuclear element; LINE, long interspersed nuclear element; LTR, LTR-containing retroviral element; and DNA, DNA-containing mobile element.

region (Figure 3). Dot plots were also used to compare the X chromosome sequence to itself and the Y chromosome sequence to itself (Figures 4B and 4C). This analysis revealed that the Y chromosome also has multiple local duplications, whereas the X chromosome has none. Many of these local duplications contain sequences that are not classified as known repetitive elements. These sequences are not contained on either X chromosome or autosomal BACs that have been sequenced to completion, suggesting that a class of stickleback-specific repetitive elements has been amplified on the Y chromosome. The Y-specific repeats account for a large number of the regions of discontinuity between the X and the Y chromosome sequence contigs (Figure 3).

Several other evolving Y chromosome systems in plants (*Silene latifolia*, *Marchantia polymorpha*, and *Carica papaya*) and animals (*Drosophila miranda*) have now been characterized, and many show similar patterns to the stickleback Y chromosome, such as the presence of duplicated sequences and the accumulation of repetitive DNA [34–38]. Similarly, the sex-determination regions of fish such as *Oryzias latipes* and *Xiphophorus maculatus* are characterized by duplications and the presence of multiple retrotransposable elements, including species- and Y-specific repeats [39–40]. On the highly evolved human Y chromosome, approximately 44% of the euchromatin in the male-specific region consists of ampliconic regions containing Y-specific repeats [18].

Sequence Divergence and the Origin of the Threespine Stickleback Y Chromosome

The BAC sequence contig analysis revealed characteristic differences between the X- and Y-specific alleles of the *Idh* and *Znf* genes. To examine how these genes differ in other stickleback populations, we amplified and sequenced the 3′UTR of the *Idh* gene and exon 2 of the *Znf* gene in sticklebacks from five distinct geographic locations around the world (Figure 5). We also sequenced these regions in two populations of *Gasterosteus wheatlandi*, a sister species to the threespine stickleback. For each gene, PCR products were amplified from several male and female fish, and the PCR products were cloned and sequenced so that several X-specific clones and several Y-specific clones could be analyzed (Table S2).

All five of the *G. aculeatus* populations showed char-

acteristic differences between the X- and Y-specific sequences in the *Idh* and *Znf* genes. In four of the five populations, including fish from both the Atlantic and Pacific Ocean, all of the sequence changes were identical to the sequence changes seen in the BAC contigs (derived from a Salmon River, BC population). These changes include a 32-bp deletion and 11 single-nucleotide differences in the 3′UTR of the *Idh* gene and nine synonymous nucleotide differences in exon 2 of the *Znf* gene on the Y chromosome relative to the X chromosome. No intrapopulation variability was seen in the X- or Y-linked alleles of either gene in the relatively small number of individuals studied (Table S2).

Threespine sticklebacks from the Japan Sea also showed characteristic sequence differences between X- and Y-linked alleles in the *Idh* and *Znf* genes. However, some of these changes were unique to the Japan Sea population, consistent with the long period of isolation between Japan Sea sticklebacks and other *Gasterosteus* populations. However, phylogenetic analysis showed very strong bootstrap support for both the clustering of Japan Sea X chromosome sequences with the X chromosome sequences of the other four populations and the clustering of Japan Sea Y chromosome sequences with the Y chromosome sequences of the other four populations (Figure 5). These results suggest that the Y chromosome of *Gasterosteus aculeatus* predates the separation of the Japan Sea and other stickleback populations approximately 2 million years ago [41].

The sex-linked genes *Idh* and *Znf* were also sequenced in *G. wheatlandi*, a sister species thought to have diverged from *G. aculeatus* around 10 million years ago [21]. This analysis revealed no sex-specific polymorphisms in either gene in *G. wheatlandi*, suggesting that neither gene is linked to a sex-determination region in *G. wheatlandi*. Because *G. wheatlandi* has a cytogenetically visible Y chromosome [22], our results suggest either that its XY chromosome system evolved independently from the XY system in *G. aculeatus* or that the linkage between the sex-determination region and the *Idh* and *Znf* genes has been broken up in *G. wheatlandi*. Taken together with the Japan Sea data, we believe that the threespine stickleback Y chromosome most likely arose after the split with *G. wheatlandi* 10 million years ago but before the origin of the *G. aculeatus* subpopulations in the Japan Sea and Pacific/Atlantic Oceans 2 million years ago.
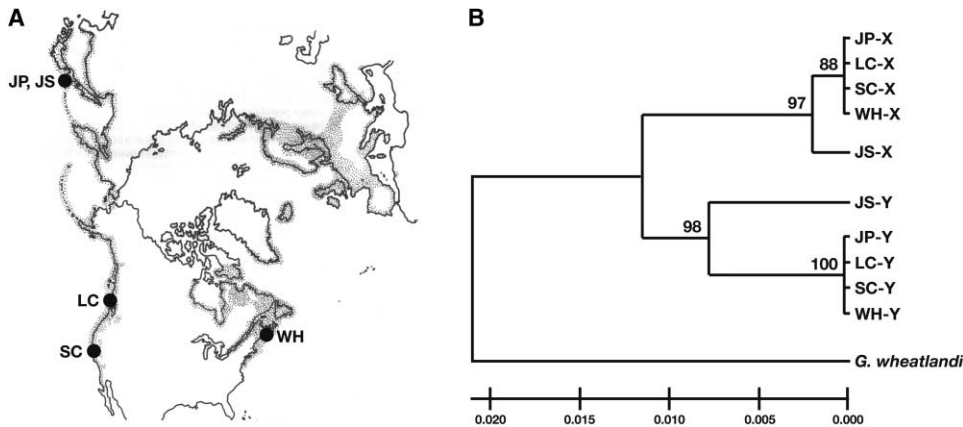
Figure 5. X and Y Chromosome Sequences Form Two Different Clusters in a Phylogenetic Tree

(A) Map of the distribution of populations used for phylogenetic analysis, modified from [29] by permission of Oxford University Press.
(B) A linearized tree was generated using the neighbor-joining method. Bootstrap values are shown for each branch. The scale below the tree represents the averaged pairwise distances between populations. The following abbreviations are used: Japan Pacific marine (JP), Japan Sea marine (JM), Little Campbell River (LC), Santa Clara River (SC), and White marines (WH). Details on populations and collections are given in Table S2.

Conclusions

These studies show that a single major chromosome region in *G. aculeatus* controls sex determination. This region has the properties of a nascent Y chromosome: heterozygosity in males, suppression of recombination, accumulation of repeat sequences, and substantial nucleotide divergence from the homologous region on the X chromosome. The existing sequence contigs already provide an initial glimpse of the sequence changes associated with young sex chromosomes that are not yet visibly different either at the light or electron microscope level [22–23]. Further extension of the X and Y sequence contigs should make it possible to identify the gene or genes responsible for sex determination in *G. aculeatus* as well as linked genes that may be involved in male- or female-specific functions. The stickleback system should also provide a useful vertebrate comparison to other evolving sex chromosome systems in plants and insects. Furthermore, the existence of closely related Gasterosteid species with both XY (*Gasterosteus wheatlandi*) and ZW (*Apeltes quadracus*) sex-determination systems [22] provides a unique opportunity to compare both the evolution of sex-determination pathways and the origin of sex chromosomes in closely related species that have diverged within the last 20 million years [21].

Experimental Procedures

Genetic Crosses

The Priest cross [29] and the Paxton cross [31–32] have been previously described. A single F2 family (Family 4), which generated 385 F2 progeny, from the Paxton cross was used in this analysis. In both crosses, sex was determined by visual examination of the gonads.

Genetic Mapping

Genotyping with microsatellites was performed as previously described [29]. One previously unpublished microsatellite (*Stn 303*) was used to genotype the Paxton cross. PCR genotyping for the *Idh* locus was performed using a forward primer (5′-GGGACGAG CAAGATTTATTG-3′) and a reverse primer (5′- TTATCGTTAGCCAG GAGATGG-3′) designed from the 3′UTR with PCR conditions as

previously described (Peichel et al. 2001). The *Idh* PCR products were visualized on a 2% agarose gel in 1× Tris-borate-EDTA (TBE). For mapping the *Idh* gene on the female meiotic map, forward primer (5′-AGCTCGTCCTCGTCCTTCC-3′) and reverse primer (5′-GAGGT CTGGCAATTTGAAGC-3′) were used with previously described PCR conditions [29], with the addition of 1 μCi [32P] dCTP per reaction to amplify a 150 bp product from the 5′UTR of *Idh*. Polymorphisms were detected using single-strand conformation polymorphism analysis with the mutation detection enhancement (MDE) gel solution (BioWhittaker Molecular Applications) and visualized with autoradiography.

To analyze female and male meiotic maps separately, F2 genotypes segregating from the F1 female and from the F1 male were entered into separate Map Manager v2.6.6 [42] files. Minimizing double recombination events determined the order of markers along the linkage group. Distance between markers was calculated in Map Manager using Backcross stats. A chi-square contingency test was used to determine if observed differences in recombination rates were significant.

BAC Library Screening

High density filters from a BAC library prepared from the DNA of sixty (male and female) anadromous Salmon River (British Columbia) sticklebacks (CHORI-213, Children's Hospital Oakland Research Institute) [30] were screened with radiolabeled overgo probes designed to the *Idh* 3′UTR using Overgo 1.02i (http://www.mouse-genome.bcm.tmc.edu/webovergo/OvergoInput.asp) and Overgo Maker (http://www.genome.wustl.edu/tools/index.php?overgo=1). Two overgo pairs were used: 5′-GGGACGAGCAAGATTTATTGG CAA-3′ and 5′-GGACTGTCAAACGTATTTGCCCAAT-3′; 5′-GATAGT CGGAAAGACATGAGGTGG-3′ and 5′-GTTGAGAGCTGTGCTTCCA CCTCA-3′. BACs were genotyped as X chromosome or Y chromosome specific using the *Idh* 3′UTR primers as described above.

BAC Sequencing

Two overlapping X chromosome BACs (101E8 and 160O9) and two overlapping Y chromosome BACs (169J23 and 119K16) were sequenced to completion. BAC DNA was hydrodynamically sheared using a Hydroshear Instrument (GeneMachines), size selected (3–4 kb) and subcloned into the plasmid pIK96 (http://shgc.stanford.edu). Universal primers and BigDye Terminator chemistry (Applied Biosystems) were used for sequencing randomly selected plasmid subclones to an average sequence depth of 10×. The Phred/Phrap/Consed suite of programs were used for assembling and editing the sequences [43–45]. After manual inspection of the assembled sequences, finishing was performed both by resequencing plasmid

subclones and by walking on plasmid subclones or the BAC clone using custom primers. All finishing reactions were performed using dGTP BigDye Terminator chemistry (Applied Biosystems). Finished clones contain no gaps and are estimated to contain less than one error per 100,000 bp.

X and Y Sequence Contig Assembly

The region of sequence overlap between the two X chromosome BACs and between the two Y chromosome BACs were removed from one sequence to generate single sequence contigs for the X and Y chromosomes. To facilitate subsequent analysis, the 5′ and 3′ ends of the X and Y chromosome contigs were defined by homologous X-Y anchor sequences having >5000 nt of X-Y alignment and containing no known transposable elements.

Global Alignment

A global pairwise alignment of the final X chromosome sequence contig and the final Y chromosome sequence contig was performed using VISTA (http://www-gsd.lbl.gov/vista) with 95% stringency, 100 nt window size, and "fugu" repetitive element masking.

Gene Identification

To identify genes on the X- and Y-specific contigs, nucleotide-nucleotide BLAST was performed against the nonredundant database in GenBank (http://www.ncbi.nlm.nih.gov/BLAST/). GENSCAN (http://genes.mit.eduGENSCAN.html) was also used on the X- and Y-specific contigs, and a protein-protein BLAST was performed with all predicted proteins against the nonredundant database in GenBank. In order to identify the positions of the exons of the five genes in the alignment, representative genes that contain the complete coding sequence from other organisms were used. The genes and accession numbers used are: human Semaphorin 4B (*Sema4B*; NM_020210.2), zebrafish C2H2 zinc finger protein (*Znf*; NM_199792.1), zebrafish NADP-dependent isocitrate dehydrogenase (*Idh*; NM_199564.1), human Ras protein-specific guanine nucleotide-releasing factor 1 (*Rasgrf1*; NM_002891.3), and rat Band 4.1 containing protein (*Band 4.1*; XM_230513.2). Exon boundaries were identified using the program sim4 [46] to align the mRNA sequences to the final Y contig. The sim4 report was used for annotating the VISTA plot with exon positions.

Repeat Analysis

To identify known repetitive elements in the X and Y chromosome contigs, the sequences were analyzed with RepeatMasker (http://www.repeatmasker.org) using the fugu library. To identify the internal duplications on the Y chromosome contig, the RepeatMasked Y chromosome sequence contig was BLASTed against itself. All multicopy sequences with no homology to known repetitive elements and greater than 200 bp in length were then mapped onto the VISTA plot.

Dot Plot Analysis

The final X and Y sequence contigs were subjected to analysis by "compare" and visualized using "dotplot" (Wisconsin package v.10.2-UNIX). The window size was 100 nt and the stringency was set at 95%.

Population Survey

The *Idh* 3′UTR was amplified from several individuals from five populations of *G. aculeatus* and two populations of *G. wheatlandi* using the primers and conditions described above. An 876 bp product containing exons 2 and 3 of the *G. aculeatus Znf* gene was amplified using the forward primer (5′-GAGGAGGAATTGGAAGAGGC-3′) and the reverse primer (5′-GATCGGTACCTTAAGGGCG-3′). A 547 bp product containing exon 2 of the *G. wheatlandi Znf* gene was amplified using the forward primer (5′-CGCTGGAAGTGCCGCATGTG-3′) and the reverse primer (5′-GGATCTGGACGAACTCCATGC-3′). PCR conditions were as described previously [29], with the addition of dimethyl sulfoxide (DMSO) to the reaction mix at a final concentration of 10%. PCR products were cloned into TA cloning vectors (Invitrogen) and multiple clones per individual were sequenced using the M13 forward and reverse primers. See Table S2 for collection sites, number of individuals used, and total number of clones sequenced.

Phylogenetic Analysis

A total of 545 bp from the *Znf* gene and 271 bp from the *Idh* gene were combined into an 816 bp sequence and used for this analysis. The male-specific deletion in *Idh* was represented by a single bp. Phylogenetic analysis was conducted with MEGA version 2.1 (http://www.megasoftware.net/). Pairwise distances were estimated using the Tajima-Nei distance method [47], but similar results were obtained using the Kimura 2-parameter, Tamura 3-parameter, Tamura-Nei, or Jukes-Cantor pairwise distance methods. The neighbor-joining method [48] and bootstrap analysis with 1000 replicates were used to generate a linearized tree.

Supplemental Data

Supplemental Data including two additional tables are available at http://www.current-biology.com/cgi/content/full/14/16/1416/DC1/.

References

1. Cline, T.W., and Meyer, B.J. (1996). Vive le difference: males vs females in flies vs worms. Annu. Rev. Genet. *30*, 637–702.
2. Swain, A., and Lovell-Badge, R. (1999). Mammalian sex determination: a molecular drama. Genes Dev. *13*, 755–767.
3. Marin, I., and Baker, B.S. (1998). The Evolutionary Dynamics of Sex Determination. Science *281*, 1990–1994.
4. Schutt, C., and Nothiger, R. (2000). Structure, function and evolution of sex-determining systems in Dipteran insects. Development *127*, 667–677.
5. Graham, P., Penn, J.K.M., and Schedl, P. (2002). Masters change, slaves remain. Bioessays *25*, 1–4.
6. Marshall-Graves, J.A., and Shetty, S. (2001). Sex from W to Z: Evolution of Vertebrate Sex Chromosomes and Sex Determining Genes. J. Exp. Zool. *290*, 449–462.
7. Carroll, S.B. (2000). Endless Forms: The Evolution of Gene Regulation and Morphological Diversity. Cell *101*, 577–580.
8. Raymond, C.S., Shamu, C.E., Shen, M.M., Seifert, K.J., Hirsch, B., Hodgkin, J., and Zarkower, D. (1998). Evidence for evolutionary conservation of sex-determining genes. Nature *391*, 691–695.
9. Beye, M., Hasselmann, M., Fondrik, M.K., Page, R.E., and Omholt, S.W. (2003). The Gene *csd* Is the Primary Signal for Sexual Development in the Honeybee and Encodes an SR-Type Protein. Cell *114*, 419–429.
10. Wilkins, A.S. (1995). Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway. Bioessays *17*, 71–77.
11. Bull, J.J. (1983). Evolution of Sex Determining Mechanisms, (Benjamin Cummings: Menlo Park).
12. Rice, W.R. (1987a). The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. Evolution *41*, 911–914.

13. Charlesworth, B. (1991). The evolution of sex chromosomes. Science *251*, 1030–1033.

14. Rice, W.R. (1987). Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. Genetics *116*, 161–167.

15. Jablonka, E., and Lamb, M.J. (1990). The evolution of heteromorphic sex chromosomes. Biol. Rev. *65*, 249–276.

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

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

18. Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntokova, T., Ali, J., Bierl, T., et al. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature *423*, 825–837.

19. Rice, W.R. (1996). Evolution of the Y Sex Chromosome in Animals. Bioscience *46*, 331–343.

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

21. Bell, M.A., and Foster, S.A. (1994). The Evolutionary Biology of the Threespine Stickleback. (Oxford: Oxford University Press).

22. Chen, T.-R., and Reisman, H.M. (1970). A comparative study of the North American species of sticklebacks (Teleostei:Gasterosteidae). Cytogenetics *9*, 321–332.

23. Cunado, N., Barrios, J., San Miguel, E., Amaro, R., Fernandez, C., Hermida, M., and Santos, J.L. (2002). Synaptonemal complex analysis in oocytes and spermatocytes of threespine stickleback *Gasterosteus aculeatus* (Teleostei, Gasterosteidae). Genetica *114*, 53–56.

24. Lindsey, C.C. (1962). Experimental study of meristic variation in a population of threespine stickleback, Gasterosteus aculeatus. Can. J. Zool. *40*, 271–312.

25. Avise, J.C. (1976). Genetics of plate morphology in an unusual population of threespine sticklebacks (*Gasterosteus aculeatus*). Genet. Res. *27*, 33–46.

26. Withler, R.E., and McPhail, J.D. (1985). Genetic variability in freshwater and anadromous sticklebacks (*Gasterosteus aculeatus*) of Southern British Columbia. Can. J. Zool. *63*, 528–533.

27. Withler, R.E., McPhail, J.D., and Devlin, R.H. (1986). Electrophoretic Polymorphism and Sexual Dimorphism in the Freshwater and Anadromous Threespine Sticklebacks (*Gasterosteus aculeatus*) of the Little Campbell River, British Columbia. Biochem. Genet. *24*, 701–713.

28. Griffiths, R., Orr, K.J., Adam, A., and Barber, I. (2000). DNA sex identification in the three-spined stickleback. J. Fish Biol. *57*, 1331–1334.

29. Peichel, C.L., Nereng, K.S., Ohgi, K.A., Cole, B.L.E., Colosimo, P.F., Buerkle, C.A., Schluter, D., and Kingsley, D.M. (2001). The genetic architecture of divergence between threespine stickleback species. Nature *414*, 901–905.

30. Kingsley, D., Zhu, B., Osoegawa, K., deJong, P.J., Schein, J., Marra, M., Peichel, C., Amemiya, C., Schluter, D., Balabhadra, S., et al. (2004). New genomic tools for molecular studies of evolutionary change in sticklebacks. Behavior, in press.

31. Shapiro, M.D., Marks, M.E., Peichel, C.L., Blackman, B.K., Nereng, K.S., Jonsson, B., Schluter, D., and Kingsley, D.M. (2004). Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature *428*, 717–723.

32. Colosimo, P.F., Peichel, C.L., Nereng, K., Blackman, B.K., Shapiro, M.D., Schluter, D., and Kingsley, D.M. (2004). The genetic architecture of parallel armor plate reduction in threespine sticklebacks. PLoS Biol. *2*, 0635–0641.

33. Morizot, D.C., McEntire, B.B., Coletta, L.D., Kazianis, S., Schartl, M., and Nairn, R.S. (1998). Mapping of Tyrosine Kinase Gene Family Members in a *Xiphophorus* Melanoma Model. Mol. Carcinog. *22*, 150–157.

34. Steinemann, M., and Steinemann, S. (1998). Enigma of Y chromosome degeneration: neo-Y and neo-X chromosomes of *Drosophila miranda* a model for sex chromosome evolution. Genetica *102/103*, 409–420.

35. Filatov, D.A., Moneger, F., Negrutiu, I., and Charlesworth, D. (2000). Low variability in a Y-linked plant gene and its implications for Y-chromosome evolution. Nature *404*, 388–390.

36. Okada, S., Sone, T., Fujisawa, M., Nakayama, S., Takenaka, M., Ishizaki, K., Kono, K., Shimizu-Ueda, T., Hanajiri, T., Yamoto, K.T., et al. (2001). The Y chromosome in the liverwort *Marchantia polymorpha* has accumulated unique repeat sequences harboring a male-specific gene. Proc. Natl. Acad. Sci. USA *98*, 9454–9459.

37. Bachtrog, D. (2003). Adaptation shapes patterns of genome evolution on sexual and asexual chromosomes in Drosophila. Nat. Genet. *34*, 215–219.

38. Liu, Z., Moore, P.H., Ma, H., Ackerman, C.M., Ragiba, M., Yu, Q., Pearl, H.M., Kim, M.S., Charlton, J.W., Stiles, J.I., et al. (2004). A primitive Y chromosome in papaya marks incipient sex chromosome evolution. Nature *427*, 348–352.

39. Nanda, I., Volff, J.-N., Weis, S., Korting, C., Froschauer, A., Schmid, M., and Schartl, M. (2000). Amplification of a long terminal repeat-like element on the Y chromosome of the platyfish, *Xiphophorus maculatus*. Chromosoma *109*, 173–180.

40. Volff-J.N., and Schartl, M. (2002). Sex determination and sex chromosome evolution in the medaka, *Oryzias latipes*, and the platyfish, *Xiphophorus maculatus*. Cytogenet. Genome Res. *99*, 170–177.

41. Higuchi, M., and Goto, A. (1996). Genetic evidence supporting the existence of two distinct species in the genus Gasterosteus around Japan. Environmental Biology of Fishes *47*, 1–16.

42. Manly, K. (1993). A Macintosh program for storage and analysis of experimental genetic mapping data. Mamm. Genome *1*, 123–126.

43. Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998). Base-calling of automated sequencer traces using Phred. I. accuracy assessment. Genome Res. *8*, 175–185.

44. Ewing, B., and Green, P. (1998). Base calling of automated sequencer traces using Phred. II. error probabilities. Genome Res. *8*, 186–194.

45. Gordon, D., Abajian, C., and Green, P. (1998). Consed: A graphical tool for sequence finishing. Genome Res. *8*, 195–202.

46. Florea, L., Hartzell, G., Zhang, Z., Rubin, G.M., and Miller, W. (1998). A computer program for aligning a cDNA sequence with a genomic DNA sequence. Genome Res. *8*, 967–974.

47. Tajima, F., and Nei, M. (1984). Estimation of evolutionary distance between nucleotide sequences. Mol. Biol. Evol. *1*, 269–285.

48. Saito, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. *4*, 406–425.

Accession Numbers