Evolution of *DMY*, a Newly Emergent Male Sex-Determination Gene of Medaka Fish

Jianzhi Zhang[1]

Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan 48109

Manuscript received October 11, 2003
Accepted for publication January 7, 2004

ABSTRACT

The Japanese medaka fish *Oryzias latipes* has an XX/XY sex-determination system. The Y-linked sex-determination gene *DMY* is a duplicate of the autosomal gene *DMRT1*, which encodes a DM-domain-containing transcriptional factor. *DMY* appears to have originated recently within Oryzias, allowing a detailed evolutionary study of the initial steps that led to the new gene and new sex-determination system. Here I analyze the publicly available *DMRT1* and *DMY* gene sequences of Oryzias species and report the following findings. First, the synonymous substitution rate in *DMY* is 1.73 times that in *DMRT1*, consistent with the male-driven evolution hypothesis. Second, the ratio of the rate of nonsynonymous nucleotide substitution (d_N) to that of synonymous substitution (d_S) is significantly higher in *DMY* than in *DMRT1*. Third, in *DMRT1*, the d_N/d_S ratio for the DM domain is lower than that for non-DM regions, as expected from the functional importance of the DM domain. But in *DMY*, the opposite is observed and the DM domain is likely under positive Darwinian selection. Fourth, only one characteristic amino acid distinguishes all *DMY* sequences from all *DMRT1* sequences, suggesting that a single amino acid change may be largely responsible for the establishment of *DMY* as the male sex-determination gene in medaka fish.

MOST animals have two sexes. However, whether an undifferentiated embryonic gonad eventually develops into a testis or an ovary is determined by genetic, environmental, or both factors, depending on the species concerned. For placental mammals, in which females have two X chromosomes and males have one X and one Y chromosome, male sex is determined by a Y-linked gene *SRY* (Berta *et al.* 1990; Sinclair *et al.* 1990). It has been proposed that *SRY* originated by divergence from its X-linked homologous gene *SOX3*, probably due to the lack of recombination between the two genes (Stevanovic *et al.* 1993; Foster and Graves 1994). It is difficult to delineate the initial steps that led to the emergence of *SRY*, because *SRY* likely originated before the separation of placental and marsupial mammals (Foster *et al.* 1992; Foster and Graves 1994) about 170 million years ago (MYA; Kumar and Hedges 1998). The identification of the male sex-determination gene *DMY* in the Japanese medaka fish *Oryzias latipes* (Matsuda *et al.* 2002; Nanda *et al.* 2002) provides an excellent opportunity for such a study of the origin of sex-determination genes. Similar to the situation in mammals, *O. latipes* has XX/XY sex chromosomes. But, in contrast to the small size of mammalian Y, medaka Y is identical in morphology to X. Because the Y chromosome usually undergoes evolutionary degeneration when it no longer recombines with X (Nei 1970; Charlesworth and Charlesworth 2000), the observation in medaka suggests a recent origin of its Y as well as its sex-determination system. In fact, the Y-specific region in medaka originated from transposition of an autosomal region and it spans ~280 kb, containing only one functional gene, *DMY* (Nanda *et al.* 2002). A naturally occurring insertion in *DMY* that leads to the truncation of the gene results in XY females (Matsuda *et al.* 2002). A second mutation that reduces the level of *DMY* expression also results in a high frequency of XY females (Matsuda *et al.* 2002). These observations demonstrate the necessity for *DMY* in male sex determination and strongly suggest it to be the medaka counterpart of the mammalian *SRY* gene (Matsuda *et al.* 2002; Nanda *et al.* 2002). A recent survey of several different strains, however, identified some male medaka fish that do not possess *DMY*, suggesting the possibility of other contributing factors in male sex determination (Nanda *et al.* 2003). DNA sequence analysis shows that *DMY* is a paralog of an autosomal gene *DMRT1*, which encodes a conserved zinc-finger transcription factor that is found in mammals, birds, reptiles, fruitflies, and nematodes (Zarkower 2001; Matsuda *et al.* 2002; Nanda *et al.* 2002). Phylogenetic and evolutionary analyses further indicated that the duplication event was recent, as the *DMY* gene is found only in two closely related species (*O. latipes* and *O. curvinotus*), but is absent from other species of the genus Oryzias so far examined (Kondo *et al.* 2003; Matsuda *et al.* 2003). In this work, I analyzed

[1]*Address for correspondence:* Department of Ecology and Evolutionary Biology, University of Michigan, 3003 Natural Science Bldg., 830 North University Ave., Ann Arbor, MI 48109.
E-mail: jianzhi@umich.edu

the publicly available DNA sequences of the *DMY* and *DMRT1* genes from Oryzias in an attempt to address the following questions. First, when did the duplication occur? Second, is the mutation rate in *DMY* higher than that in *DMRT1*? This question is raised because *DMY* is located on Y, which is always in males, while *DMRT1* is in males half the time and in females half the time. The hypothesis of male-driven evolution asserts that the mutation rate is higher in males than in females due to more rounds of cell divisions in the male germ line (Li *et al.* 2002). Third, does *DMY* exhibit a high rate of protein evolution? I raise this question because the mammalian male sex-determination gene *SRY* shows a rapid pace of sequence evolution among species of primates and rodents (Tucker and Lundrigan 1993; Whitfield *et al.* 1993). Fourth, is there any difference in amino acid substitution pattern between *DMY* and *DMRT1*, as such a difference may provide information on the critical regions of the protein that confers the new function of *DMY*? Finally, how many amino acid substitutions may be responsible for *DMY*'s new role in sex determination?

MATERIALS AND METHODS

The *DMY* and *DMRT1* gene sequences from *O. latipes* HNI strain, *O. latipes* Carbio strain, *O. curvinotus*, and the *DMRT1* sequence of *O. celebensis* were obtained from GenBank. The GenBank accession numbers are listed in the legend to Figure 1. The protein sequences were aligned using CLUSTAL X (Thompson *et al.* 1997) and the DNA sequences were then aligned following the protein alignment. It was found that there is a one-nucleotide deletion near the end of the coding region in the *O. latipes* HNI *DMY* gene, which results in a frameshift and earlier termination of the open reading frame (see Figure 1). The nucleotide sequence alignment was thus manually adjusted by incorporating this deletion, and the sequence region that is downstream of this deletion is not used in the evolutionary rate analysis because of the difficulty in assigning synonymous and nonsynonymous differences. Gene trees were reconstructed using the neighbor-joining method (Saitou and Nei 1987) implemented in MEGA2 (Kumar *et al.* 2001), as well as the likelihood method (Felsenstein 1981) implemented in PAUP* (Swofford 1998). Several different distance measures or substitution models (Jukes-Cantor, Kimura's two parameter, Tajima-Nei, and Tamura-Nei; Nei and Kumar 2000) were used. The bootstrap test (Felsenstein 1985; 2000 replications for neighbor joining and 200 replications for likelihood) was used to examine the reliability of the reconstructed trees. Four-cluster analysis was performed by the PHYLTEST program (Rzhetsky *et al.* 1995). Ancestral gene sequences were reconstructed for interior nodes by the distance-based Bayesian method (Zhang and Nei 1997). Numbers of synonymous (s) and nonsynonymous (n) substitutions on tree branches were counted as in Zhang *et al.* (1997). A maximum likelihood method (Yang 1998) was used to analyze changes in $w = d_N/d_S$ in the evolution of *DMY* and *DMRT1* genes. Here d_N and d_S refer to the numbers of nonsynonymous and synonymous substitutions per site, respectively. The likelihood method (Yang *et al.* 2000) was also used for testing positive selection on individual sites.

RESULTS

Phylogenetic relationships of *DMRT1* and *DMY* genes of Oryzias species: The *DMY* gene was found only in *O. latipes* (HNI and Carbio strains) and *O. curvinotus*, but not in other species of the Oryzias genus so far examined (Kondo *et al.* 2003; Matsuda *et al.* 2003). This would suggest that the gene duplication that gave rise to *DMY* occurred in the common ancestor of *O. latipes* and *O. curvinotus* after this ancestor diverged from other Oryzias species (Matsuda *et al.* 2003). To verify this hypothesis, I first aligned the *DMY* and *DMRT1* sequences (Figure 1) and then reconstructed a gene tree by the neighbor-joining method (Figure 2a). The tree shows the clustering of *DMY* sequences from the two strains of *O. latipes*. However, the *DMY* gene of *O. curvinotus* does not cluster with those of *O. latipes*. Rather, it clusters with *O. curvinotus DMRT1*. Nevertheless, this grouping has only 50% bootstrap support, suggesting a possibility that it may not reflect the true relationships among the sequences. A likelihood analysis resulted in the same tree with slightly different bootstrap numbers (Figure 2a). From these analyses, it appears that the *DMY* genes from the two strains of *O. latipes* form a cluster (with 100% bootstrap support) and the *DMRT1* genes from *O. latipes* form another cluster (with >97% bootstrap support). It is the relations of these two clusters and the *DMY* and *DMRT1* genes of *O. curvinotus* that are in question. I then used a distance-based four-cluster analysis, which is particularly suitable for evaluating the relationships among four monophyletic groups (Rzhetsky *et al.* 1995). This analysis shows that none of the three possible groupings of the above four clusters is significantly better than the other groupings ($P > 0.2$; Figure 2 legend). That is, there is not sufficient phylogenetic information in the data that can resolve the relationships among the four clusters. It is well known that sex-determination mechanisms are highly variable among animals (Zarkower 2001). It is thus very unlikely that an identical sex-determination system evolved twice in two closely related species by independent duplications of the same gene. Therefore, it is parsimonious to assume that *DMY* originated only once in Oryzias, as depicted in Figure 2b. Further analyses are based mainly on this assumption. However, to examine whether the results obtained are robust, I also repeated all the analyses without using the two sequences of *O. curvinotus*, as the tree becomes well resolved when these sequences are excluded (Figure 2a).

A higher rate of synonymous substitution in *DMY* than in *DMRT1*: To study the rate of nucleotide substitution in the evolution of *DMY* and *DMRT1*, I inferred the ancestral gene sequences on all interior nodes of the tree shown in Figure 3, which has the same topology as the tree of Figure 2b. Because the sequences are relatively closely related, this inference is expected to be

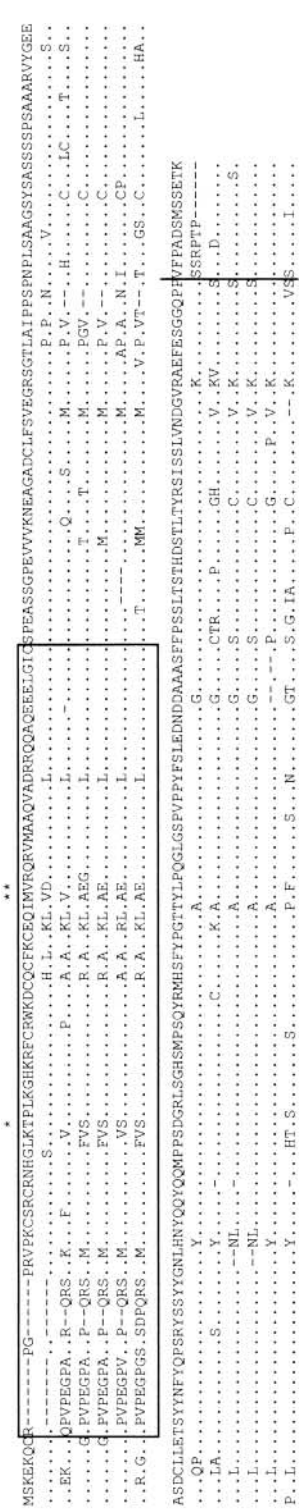

FIGURE 1.—Protein sequence alignment of *DMY* and *DMRT1* genes of medaka fish (Oryzias species). The DM domain is boxed. Amino acid substitutions in *DMY* that occurred in the common ancestor of *O. latipes* and *O. curvinotus* (see tree b of Figure 2) are indicated by an asterisk. Sequences downstream of the vertical line are not used in the evolutionary rate analyses because of a frameshifting deletion in *O. latipes* (HNY) *DMY*. Carbio and HNI are two strains of *O. latipes*. The sequences are obtained from GenBank, with the accession numbers as follows: *O. latipes* (Carbio) *DMY*, AY129240; *O. latipes* (HNI) *DMY*, AY129241; *O. curvinotus DMY*, AB091695; *O. latipes* (Carbio) *DMRT1*, AF319994; *O. latipes* (HNI) *DMRT1*, AY157712; *O. curvinotus DMRT1*, AB091696; and *O. celebensis DMRT1*, AY239587.

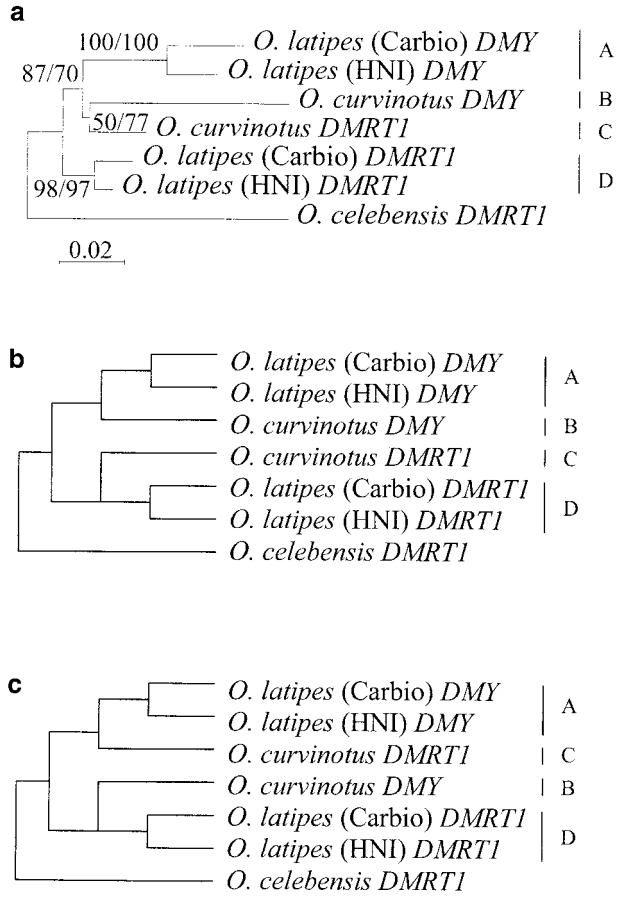

FIGURE 2.—Phylogenetic relationships of seven *DMY* and *DMRT1* gene sequences of medaka fish. (a) The same tree generated by the neighbor-joining and likelihood methods. The neighbor-joining tree was reconstructed with KIMURA's (1980) two-parameter distance, whereas the likelihood tree was reconstructed under the general-reversible model with variable rates of substitutions across sites (plus a class of invariant sites). Bootstrap percentages from neighbor-joining and likelihood analyses are shown on interior branches. (b and c) Alternative trees when all possible relations among groups A, B, C, and D are considered. In a four-cluster analysis (see text), tree a is found to be better than tree c, which is in turn better than tree b; but none of the differences among the three trees are statistically significant ($P > 0.2$).

reliable. In fact, the computed average posterior probability is >99% for all the ancestral sequences. I then counted the numbers of synonymous and nonsynonymous substitutions on each tree branch (Figure 3a). There are a total of 31.6 synonymous substitutions in *DMY*, but only 18.3 in *DMRT1* for all the branches since the duplication event. The difference between the two numbers is statistically significant ($P = 0.03$, one-tailed z-test), indicating a higher rate of synonymous substitution in *DMY* than in *DMRT1*. Because synonymous changes are more or less neutral and are generally immune to selection (except in the case of weak selection on codon usage in certain organisms), the above result suggests a higher mutation rate in *DMY* than in *DMRT1*. Since *DMY* is on the Y chromosome, which is always in

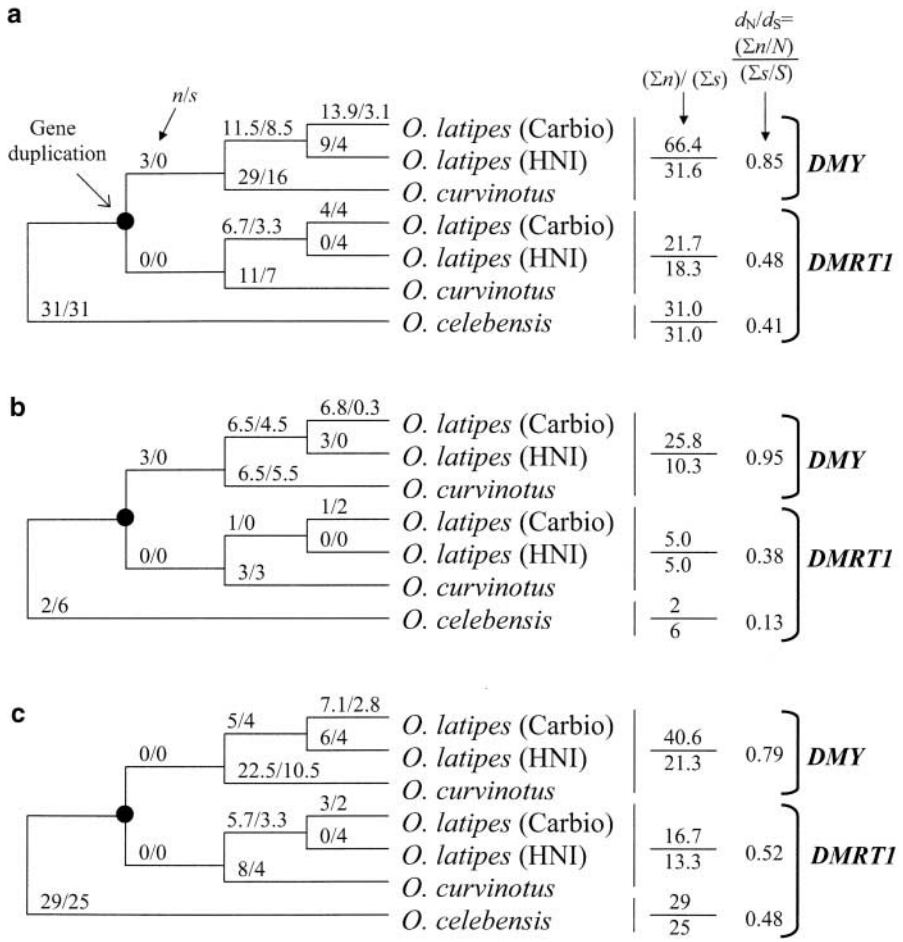

FIGURE 3.—Nucleotide substitutions in the evolution of *DMY* and *DMRT1* of medaka fish for (a) the entire protein, (b) the DM domain, and (c) the non-DM regions. The tree with a single origin of *DMY*, as depicted in Figure 2b, is used. The numbers of synonymous (s) and nonsynonymous (n) substitutions for each tree branch are shown on branches. The sums of n for all branches involving *DMY* (after gene duplication), *DMRT1* (after duplication), and *DMRT1* (before duplication) are indicated by respective Σn values. Σs values are similarly obtained. N and S are potential numbers of nonsynonymous and synonymous sites of the sequences, respectively. For the entire protein, $N = 523.0$ and $S = 212.0$. For the DM domain, $N = 143.7$ and $S = 54.3$. For non-DM regions, $N = 379.3$ and $S = 157.7$.

males, and *DMRT1* is on an autosome, which is in males half the time and in females half the time, my observation suggests a higher rate of mutation in males than in females. Note that this inference is not affected by the fact that some males do not have *DMY* (NANDA *et al.* 2003) as long as *DMY* is always in males, which has never been violated. The ratio of the mutation rate in males to that in females may be estimated by $\alpha_m = 1/(2A/Y - 1)$ (MIYATA *et al.* 1987), where A/Y is the mutation rate ratio between autosome and Y chromosome and is $18.3/31.6 = 0.579$ for this data set. Thus, α_m is 6.32. This result supports the male-driven evolution hypothesis, which asserts that the mutation rate is higher in males than in females (MIYATA *et al.* 1987; LI *et al.* 2002). When the two sequences of *O. curvinotus* are excluded from the analysis (Figure 4), $A/Y = 8.3/15.1 = 0.550$ and α_m is 10.1.

It is important to know when *DMY* originated. The divergence times for the species concerned are unknown, making it difficult to use *DMRT1* or *DMY* data for time estimation directly. It has been estimated that the nucleotide mutation rate in mammals is on average 2.2×10^{-9}/site/year (KUMAR and SUBRAMANIAN 2002). If the mutation rate in medaka *DMRT1* is similar to the above rate, it can be estimated that the duplication occurred $(7.15/212.0)/(2.2 \times 10^{-9}) = 15.3$ MYA. Here 7.15 is the average number of synonymous substitutions in *DMRT1* from duplication to present, and 212.0 is the number of synonymous sites in the gene, estimated by the modified Nei-Gojobori method (ZHANG *et al.* 1998). ELLEGREN and FRIDOLFSSON (2003) estimated the synonymous substitution rates for four autosomal genes from salmonid fish *Oncorhynchus kisutch* and *O. tshawytscha*, which were estimated to be separated ~5 MYA (OOHARA *et al.* 1997). Weighted by the sequence length, these rates averaged 3.6×10^{-9}/site/year. If the synonymous substitution rate in medaka *DMRT1* is similar to this rate, I estimated that *DMY* originated $(7.15/212.0)/3.6 \times 10^{-9} = 9.4$ MYA. When I exclude the two sequences from *O. curvinotus* and repeat the above estimation, the date of duplication becomes 5.7–9.2 MYA. Although these estimates may have large errors, it appears that the age of *DMY* is on the order of 10 MY.

A higher rate of nonsynonymous substitution in *DMY1* than in *DMRT1*: The evolutionary analysis also reveals a higher rate of nonsynonymous substitution in *DMY* than in *DMRT1* (Figure 3a), as the total number of nonsynonymous substitutions since gene duplication is

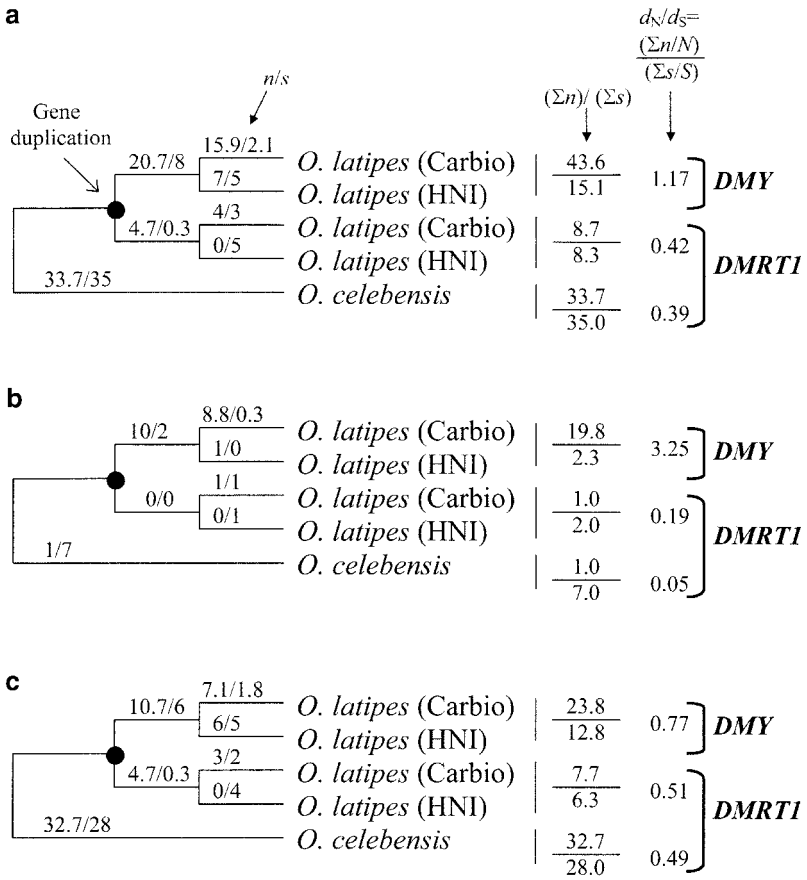

FIGURE 4.—Nucleotide substitutions in the evolution of *DMY* and *DMRT1* of medaka fish for (a) the entire protein, (b) the DM domain, and (c) the non-DM regions. The *DMY* and *DMRT1* genes from *O. curvinotus* are excluded from the analysis because of their uncertain phylogenetic positions (Figure 2). The numbers of synonymous (s) and nonsynonymous (n) substitutions for each tree branch are shown on branches. The sums of n for all branches involving *DMY* (after gene duplication), *DMRT1* (after duplication), and *DMRT1* (before duplication) are indicated by respective Σn values. Σs values are similarly obtained. N and S are potential numbers of nonsynonymous and synonymous sites of the sequences, respectively. For the entire protein, $N = 523.0$ and $S = 212.0$. For the DM domain, $N = 143.7$ and $S = 54.3$. For non-DM regions, $N = 379.3$ and $S = 157.7$.

significantly greater in *DMY* (66.4) than in *DMRT1* (21.7; $P < 10^{-5}$, two-tailed z-test). This phenomenon is due in part to the elevation of mutation rate in *DMY*. It may also arise from an alteration in natural selection on *DMY*, compared to that in *DMRT1*. The n/s ratio is greater in *DMY* (66.4/31.6 = 2.10) than in *DMRT1* (21.7/18.3 = 1.19) since gene duplication (Figure 3a). However, the difference is statistically insignificant ($P = 0.12$, Fisher's test, ZHANG *et al.* 1997). I noted that the n/s ratio for *DMRT1* after gene duplication (1.19) is similar to that before the duplication (31/31 = 1.00), and thus computed an average ratio [(21.7 + 31)/(18.3 + 31) = 1.07] for *DMRT1*. When this number is compared to the n/s ratio for *DMY*, I found that the former is significantly smaller than the latter ($P = 0.02$, Fisher's test). These results suggest that the acceleration in nonsynonymous substitutions of *DMY* is due to the elevation of mutation rate and a change in natural selection on the gene. To further confirm this result, I performed a likelihood analysis (YANG 1998). The log likelihood value under the model of equal w ($= d_N/d_S$) among all branches of the tree in Figure 3a is −1929.68, significantly lower than that (−1926.38) under the model of one w for *DMY* and a different w for *DMRT1* ($P = 0.01$, likelihood ratio test). This two-w model, however, is not significantly worse in fitting the data than a three-w model (log likelihood = −1926.27) that assigns one w for *DMY*, one for *DMRT1* after gene duplication, and one for *DMRT1* before duplication ($P = 0.64$, likelihood ratio test). These results are consistent with the above nonlikelihood-based analyses. Similar results are obtained when the two sequences from *O. curvinotus* are excluded, as shown in Figure 4a.

Altered rates of nonsynonymous substitution in the DM domain: The DM domain (Figure 1) is the DNA-binding domain in *DMRT1* and *DMY*. DNA-binding domains are usually the most conserved parts of transcriptional factors. For example, the DNA-binding homeodomain is conserved among 13 cognate Hox genes of vertebrates, some of which diverged over 600 MYA (ZHANG and NEI 1996). *DMRT1* appears to follow this rule as well. There are 0.049 nonsynonymous substitutions per nonsynonymous site in the *DM* domain on all the branches linking *DMRT1*, in comparison to 0.120 in the non-DM regions (Figure 3, b and c). The former is significantly smaller than the latter ($P = 0.008$, Fisher's test). Surprisingly, for *DMY*, the opposite is observed. There are 0.180 nonsynonymous substitutions per nonsynonymous site in the DM domain of *DMY*, in comparison to 0.107 in the non-DM regions (Figure 3, b and

c), the difference between them being statistically significant ($P = 0.02$, Fisher's test). A direct comparison between *DMY* and *DMRT1* can be made by a 2×2 table with the numbers of nonsynonymous substitutions in the DM domain of *DMY* (25.8), DM domain of *DMRT1* (7), non-DM regions of *DMY* (40.6), and non-DM regions of *DMRT1* (45.7), respectively. This comparison reveals a significant excess of nonsynonymous substitutions in the DM domain of *DMY* ($P = 0.001$, Fisher's test). However, the d_N/d_S ratio for the DM domain of *DMY* is 0.95, not significantly different from 1 ($P > 0.5$, Fisher's exact test). Thus, it is unclear whether the accelerated nonsynonymous substitution in the DM domain of *DMY* is due to a relaxation of functional constraints or positive Darwinian selection. To distinguish the two hypotheses, I used a likelihood method, which may be more powerful in detecting positive selection at individual sites (Yang *et al.* 2000). In this analysis, I compared the likelihoods for the data set of three *DMY* sequences under models 1 and 2. Here model 1 assumes that the d_N/d_S ratio for a given site is either 0 or 1, while model 2 adds an extra class of sites to model 1. The result showed that model 2 fits the data significantly better than model 1 ($\chi^2 = 14.3$, 2 d.f., $P = 0.0008$), with an additional class of $d_N/d_S = 9.8$. Three codons were identified to be under positive selection with posterior probabilities >95%. To examine whether this result is robust against possible violations of assumptions made in models 1 and 2, I conducted a second test by comparing models 7 and 8. Model 7 assumes that the d_N/d_S ratio follows a β-distribution between 0 and 1, while model 8 adds an extra class of sites to model 7. Model 8 was found to fit the data significantly better than model 7 ($\chi^2 = 18.54$, 2 d.f., $P = 0.0001$), with an additional class of $d_N/d_S = 7.3$. Here, the above three codons and two additional codons were found to be under positive selection with >95% posterior probability. Thus, the positive selection hypothesis is supported by likelihood. Although both tests favor the positive selection hypothesis, the results should be interpreted with caution as false detections of positive selection by likelihood have been reported recently (Suzuki and Nei 2001, 2002).

When the two sequences from *O. curvinotus* are excluded from the analysis (Figure 4), qualitatively similar results are obtained. In particular, the d_N/d_S ratio (3.25) for the DM domain of *DMY* is significantly >1 ($P = 0.03$, Fisher's exact test). This finding, based on a more conservative test (Zhang *et al.* 1997), provides additional and strong evidence for the action of positive selection on the DM domain.

Characteristic amino acids of *DMY*: There were no synonymous substitutions in *DMY* after its origin from duplication but before the split of *O. latipes* and *O. curvinotus* (Figure 3), suggesting that the duplication occurred shortly before the species separation (Matsuda *et al.* 2003). Interestingly, there were three amino acid substitutions during this short period of time, all located in the DM domain (Figures 1 and 3). Because *O. latipes* and *O. curvinotus* have the same sex-determination mechanism, it may be inferred that one or more of the three substitutions are important to the establishment of *DMY* as a sex-determination gene in medaka. Among the three sites, only one (position 26 in *O. latipes* Carbio *DMY*) has remained invariant among orthologous *DMY* sequences. Thus, this site may be most critical to the function of *DMY*. This site is Thr in the three *DMY* sequences and Ser in the four *DMRT1* sequences shown in Figure 1. I examined this site in all fish *DMRT1* sequences available in GenBank and found that all of them have Ser. These sequences include green spotted pufferfish *Tetraodon nigroviridis* (AY152820), pufferfish *Takifugu rubripes* (AJ295039), platyfish *Xiphophorus maculates* (AF529187), tilapia *Oreochromis niloticus* (AF203489), trout *Oncorhynchus mykiss* (AF209095), halibut *Hippoglossus hippoglossus* (CAD44607), sturgeon *Acipenser transmontanus* (AY057061), cod *Gadus morhua* (CAD44608), zebrafish *Danio rerio* (AF080622), and wrasse *Halichoeres tenuispinis* (AAO18650). Furthermore, I examined this site in *DMRT1* sequences of all nonfish vertebrates that are available in GenBank, including human *Homo sapiens* (NM_021951), mouse *Mus musculus* (NM_015826), rat *Rattus rattus* (AAK57706), pig *Sus scrofa* (AF216651), and chicken *Gallus gallus* (AAF19666), and found that all have Ser. Thus, the Ser at this position has been conserved among all vertebrate *DMRT1* sequences so far examined, suggesting its functional importance. The substitution from Ser to Thr in *DMY* may therefore have significant functional consequences.

DISCUSSION

In this article, I analyzed the rate and pattern of nucleotide substitutions in *DMY*, a newly emergent sex-determination gene in medaka fish, and made several observations. First, there is evidence for an elevation of the mutation rate in the Y-linked *DMY*, compared to its autosomal mother gene *DMRT1*. This result is in support of the male-driven evolution hypothesis, with the estimated male/female mutation rate ratio α_m being about ~6–10. The male-driven evolution hypothesis has gained substantial support from evidence in mammals (Miyata *et al.* 1987; Shimmin *et al.* 1993; Chang *et al.* 1994; Nachman and Crowell 2000; Li *et al.* 2002; Makova and Li 2002; but see McVean and Hurst 1997; Bohossian *et al.* 2000) and some evidence in birds (Ellegren and Fridolfsson 1997; Kahn and Quinn 1999; Carmichael *et al.* 2000). It has also gained support from a comparison of the sequence data of the Y-linked *GH-2Y* gene and autosomal *GH-2* gene of salmonid fish (Ellegren and Fridolfsson 2003). Because the estimated α_m values have large errors from either the salmon or the medaka data set, I combined the two

data sets and estimated that $\alpha_m = 7.7$, with the 95% confidence interval of (1.36, ∞). This confidence interval was determined using computer simulation as follows. The observed total number of synonymous substitutions on the two Y-linked genes (*DMY* and *GH-2Y*) is 52.6 and the corresponding number in the homologous autosomal genes (*DMRT1* and *GH-2*) is 29.7. Because nucleotide substitutions may be considered a Poisson process, the number of synonymous substitutions on the two Y-linked genes is a Poisson variable with a mean of 52.6 and the corresponding number for the autosomal genes is an independent Poisson variable with a mean of 29.7. I generated 10,000 pairs of Poisson random variables and then estimated the 95% confidence interval for α_m. As discussed by Ellegren and Fridolfsson (2003), it is currently unknown in fish how many cell divisions per generation occur in the male and female germ lines, but males produce much more sperm than females produce eggs and it is probable that there are more rounds of cell divisions in the male germ line. However, it should be stressed that this conclusion on male-driven evolution in fishes is tentative, as it is derived from only two genes. There appears to be substantial variation in mutation rate across genomic regions and more data are necessary to confirm this finding.

Second, a higher d_N/d_S value (0.85–1.17, depending on whether *O. curvinotus* is used) is found in *DMY*, in comparison to that (0.42–0.48) in *DMRT1*. This elevation is particularly prominent in the DM domain (0.95–3.25 *vs.* 0.24–0.19). It is interesting to note that *SRY*, the mammalian sex-determination gene on the Y chromosome, is also known to have a high d_N/d_S ratio in primates and rodents (Tucker and Lundrigan 1993; Whitfield *et al.* 1993; Pamilo and O'Neill 1997; Wang *et al.* 2002). However, the rapid evolution is limited to regions that are outside the DNA-binding HMG domain. It is still debatable whether the high d_N/d_S ratio in *SRY* is due to positive selection or relaxation of functional constraints, and *SRY* genes of different species may be under different forms of selection (Wang *et al.* 2002). Recent studies favor the view that Y-linked genes generally have elevated d_N/d_S in comparison to their homologous genes on X or autosomes (Sandstedt 2003; Tucker *et al.* 2003). This may be due to several reasons, such as changes in gene function or expression and alleviated intensity of natural selection caused by lack of recombination or reduced effective population size for Y-linked genes. In this case of *DMY*, however, it is the DNA-binding DM domain that shows high d_N/d_S values. Both likelihood- and nonlikelihood-based methods suggest the action of positive selection. The selective agent, however, is not immediately clear. It is interesting to relate this finding to the observation that the DNA-binding homeodomains of several homeobox genes of mammals and Drosophila evolve rapidly by positive selection (Sutton and Wilkinson 1997; Ting *et al.* 1998; Wang and Zhang 2004) and all these homeobox genes are located in the X chromosome and expressed in the testis (Wang and Zhang 2004). The rapid evolution of these otherwise conserved DNA-binding (DM and homeobox) domains is intriguing. Their involvement in sexual differentiation and reproduction suggests the possibility that the Darwinian selection that acts on them is related to differential reproductive success and possibly speciation (Ting *et al.* 1998).

Third, I estimated that the duplication that generated the *DMY* gene took place ~10 MYA. The split of *O. latipes* and *O. curvinotus* followed shortly (assuming the tree of Figure 2b). There is only one amino acid position that is conserved among *DMY* sequences but differs between *DMY* and *DMRT1*. It is possible that this single amino acid substitution (Ser26Thr) played a major role in the establishment of *DMY* as the primary sex-determination gene in medaka fish. In this context, it is interesting to note the high plasticity of sex-determination mechanisms in animals (Zarkower 2001). For example, single mutations in the *tra-1* gene of the nematode *Caenorhabditis elegans* can change the sex-determination mechanism entirely (Hodgkin 1983) and strains of *C. elegans* with different sex-determination mechanisms have been constructed in the laboratory (Hodgkin 2002). Several placental mammals are known to lack the *SRY* gene (Just *et al.* 1995) or have multiple *SRY* genes (Lundrigan and Tucker 1997). In the future, it will be interesting to examine the functional effect of Ser26Thr in the *DMY* gene of medaka fish and test whether this single amino acid substitution was the key to the origin of a new sex-determination gene.

Among vertebrates, *SRY* and *DMY* are the only sex-determination genes so far identified. A number of features are surprisingly similar between them, including the existence of close paralogs in the genome, higher d_N/d_S ratios than those of their paralogs, possible actions of positive selection, and occasional disappearances of the genes in males (Just *et al.* 1995; Nanda *et al.* 2003). These features probably reflect the low functional and developmental constraints on the master controllers of sex-determination pathways and the easiness for natural selection to modify them. Studying additional species may reveal whether these observations are general to all master controllers of sex-determination pathways.

On a final note, the late geneticist Susumo Ohno pioneered evolutionary studies of sex chromosomes and gene duplication (Ohno 1967, 1970). He would be pleased to see the simple, yet extraordinary path to the emergence of the medaka sex chromosome and sex-determination mechanism, which vividly demonstrates the theory of evolution by gene duplication that he championed (reviewed in Zhang 2003).

I thank David Webb for valuable comments on the manuscript. This work was supported by a startup fund from the University of Michigan and a research grant from the National Institutes of Health (GM-67030).

LITERATURE CITED

Berta, P., J. R. Hawkins, A. H. Sinclair, A. Taylor, B. L. Griffiths *et al.*, 1990 Genetic evidence equating SRY and the testis-determining factor. Nature **348:** 448–450.

Bohossian, H. B., H. Skaletsky and D. C. Page, 2000 Unexpectedly similar rates of nucleotide substitution found in male and female hominids. Nature **406:** 622–625.

Carmichael, A. N., A. K. Fridolfsson, J. Halverson and H. Ellegren, 2000 Male-biased mutation rates revealed from Z and W chromosome-linked ATP synthase alpha-subunit (ATP5A1) sequences in birds. J. Mol. Evol. **50:** 443–447.

Chang, B. H., L. C. Shimmin, S. K. Shyue, D. Hewett-Emmett and W. H. Li, 1994 Weak male-driven molecular evolution in rodents. Proc. Natl. Acad. Sci. USA **91:** 827–831.

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

Ellegren, H., and A. K. Fridolfsson, 1997 Male-driven evolution of DNA sequences in birds. Nat. Genet. **17:** 182–184.

Ellegren, H., and A. K. Fridolfsson, 2003 Sex-specific mutation rates in salmonid fish. J. Mol. Evol. **56:** 458–463.

Felsenstein, J., 1981 Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17:** 368–376.

Felsenstein, J., 1985 Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:** 783–791.

Foster, J. W., and J. A. Graves, 1994 An SRY-related sequence on the marsupial X chromosome: implications for the evolution of the mammalian testis-determining gene. Proc. Natl. Acad. Sci. USA **91:** 1927–1931.

Foster, J. W., F. E. Brennan, G. K. Hampikian, P. N. Goodfellow, A. H. Sinclair *et al.*, 1992 Evolution of sex determination and the Y chromosome: SRY-related sequences in marsupials. Nature **359:** 531–533.

Hodgkin, J., 1983 Two types of sex determination in a nematode. Nature **304:** 267–268.

Hodgkin, J., 2002 Exploring the envelope: systematic alteration in the sex-determination system of the nematode *Caenorhabditis elegans*. Genetics **162:** 767–780.

Just, W., W. Rau, W. Vogel, M. Akhverdian, K. Fredga *et al.*, 1995 Absence of Sry in species of the vole Ellobius. Nat. Genet. **11:** 117–118.

Kahn, N. W., and T. W. Quinn, 1999 Male-driven evolution among Eoaves? A test of the replicative division hypothesis in a heterogametic female (ZW) system. J. Mol. Evol. **49:** 750–759.

Kimura, M., 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:** 111–120.

Kondo, M., I. Nanda, U. Hornung, S. Asakawa, N. Shimizu *et al.*, 2003 Absence of the candidate male sex-determining gene dmrt1b(Y) of medaka from other fish species. Curr. Biol. **13:** 416–420.

Kumar, S., and S. B. Hedges, 1998 A molecular timescale for vertebrate evolution. Nature **392:** 917–920.

Kumar, S., and S. Subramanian, 2002 Mutation rates in mammalian genomes. Proc. Natl. Acad. Sci. USA **99:** 803–808.

Kumar, S., K. Tamura, I. B. Jakobsen and M. Nei, 2001 MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17:** 1244–1245.

Li, W. H., S. Yi and K. Makova, 2002 Male-driven evolution. Curr. Opin. Genet. Dev. **12:** 650–656.

Lundrigan, B. L., and P. K. Tucker, 1997 Evidence for multiple functional copies of the male sex-determining locus, Sry, in African murine rodents. J. Mol. Evol. **45:** 60–65.

Makova, K. D., and W. H. Li, 2002 Strong male-driven evolution of DNA sequences in humans and apes. Nature **416:** 624–626.

Matsuda, M., Y. Nagahama, A. Shinomiya, T. Sato, C. Matsuda *et al.*, 2002 DMY is a Y-specific DM-domain gene required for male development in the medaka fish. Nature **417:** 559–563.

Matsuda, M., T. Sato, Y. Toyazaki, Y. Nagahama, S. Hamaguchi *et al.*, 2003 Oryzias curvinotus has DMY, a gene that is required for male development in the medaka, O. latipes. Zool. Sci. **20:** 159–161.

McVean, G. T., and L. D. Hurst, 1997 Evidence for a selectively favourable reduction in the mutation rate of the X chromosome. Nature **386:** 388–392.

Miyata, T., H. Hayashida, K. Kuma, K. Mitsuyasu and T. Yasunag, 1987 Male-driven molecular evolution: a model and a nucleotide sequence analysis. Cold Spring Harbor Symp. Quant. Biol. **52:** 863–867.

Nachman, M. W., and S. L. Crowell, 2000 Estimate of the mutation rate per nucleotide in humans. Genetics **156:** 297–304.

Nanda, I., M. Kondo, U. Hornung, S. Asakawa, C. Winkler *et al.*, 2002 A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka, Oryzias latipes. Proc. Natl. Acad. Sci. USA **99:** 11778–11783.

Nanda, I., U. Hornung, M. Kondo, M. Schmid and M. Schartl, 2003 Common spontaneous sex-reversed XX males of the medaka *Oryzias latipes*. Genetics **163:** 245–251.

Nei, M., 1970 Accumulation of nonfunctional genes on sheltered chromosomes. Am. Nat. **104:** 311–322.

Nei, M., and S. Kumar, 2000 *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.

Ohno, S., 1967 *Sex Chromosomes and Sex-Linked Genes*. Springer-Verlag, Berlin.

Ohno, S., 1970 *Evolution by Gene Duplication*. Springer-Verlag, Berlin.

Oohara, I., K. Sawano and T. Okazaki, 1997 Mitochondrial DNA sequence analysis of the Masu Salmon-phylogeny in the genus Oncorhynchus. Mol. Phylogenet. Evol. **7:** 71–78.

Pamilo, P., and R. J. O'Neill, 1997 Evolution of the Sry genes. Mol. Biol. Evol. **14:** 49–55.

Rzhetsky, A., S. Kumar and M. Nei, 1995 Four-cluster analysis: a simple method to test phylogenetic hypotheses. Mol. Biol. Evol. **12:** 163–167.

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

Sandstedt, S. A., 2003 Evolution of sex chromosomal genes, *Smcy* and *Smcx*, in the mouse genus *Mus*. Ph.D. Thesis, University of Michigan, Ann Arbor, MI.

Shimmin, L. C., B. H. Chang and W. H. Li, 1993 Male-driven evolution of DNA sequences. Nature **362:** 745–747.

Sinclair, A. H., P. Berta, M. S. Palmer, J. R. Hawkins, B. L. Griffiths *et al.*, 1990 A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature **346:** 240–244.

Stevanovic, M., R. Lovell-Badge, J. Collignon and P. N. Goodfellow, 1993 SOX3 is an X-linked gene related to SRY. Hum. Mol. Genet. **2:** 2013–2018.

Sutton, K. A., and M. F. Wilkinson, 1997 Rapid evolution of a homeodomain: evidence for positive selection. J. Mol. Evol. **45:** 579–588.

Suzuki, Y., and M. Nei, 2001 Reliabilities of parsimony-based and likelihood-based methods for detecting positive selection at single amino acid sites. Mol. Biol. Evol. **18:** 2179–2185.

Suzuki, Y., and M. Nei, 2002 Simulation study of the reliability and robustness of the statistical methods for detecting positive selection at single amino acid sites. Mol. Biol. Evol. **19:** 1865–1869.

Swofford, D. L., 1998 *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4.0. Sinauer, Sunderland, MA.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin and D. G. Higgins, 1997 The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:** 4876–4882.

Ting, C., S. C. Tsaur, M. L. Wu and C.-I Wu, 1998 A rapidly evolving homeobox at the site of a hybrid sterility gene. Science **282:** 1501–1504.

Tucker, P. K., and B. L. Lundrigan, 1993 Rapid evolution of the sex determining locus in Old World mice and rats. Nature **364:** 715–717.

Tucker, P. K., R. M. Adkins and J. S. Rest, 2003 Differential rates of evolution for the ZFY-related zinc finger genes, Zfy, Zfx, and Zfa in the mouse genus Mus. Mol. Biol. Evol. **20:** 999–1005.

Wang, X., and J. Zhang, 2004 Rapid evolution of mammalian X-linked testis-expressed homeobox genes. Genetics **167** (in press).

Wang, X., J. Zhang and Y. P. Zhang, 2002 Erratic evolution of SRY in higher primates. Mol. Biol. Evol. **19:** 582–584.

Whitfield, L. S., R. Lovell-Badge and P. N. Goodfellow, 1993 Rapid sequence evolution of the mammalian sex-determining gene SRY. Nature **364:** 713–715.

Yang, Z., 1998 Likelihood ratio tests for detecting positive selection

and application to primate lysozyme evolution. Mol. Biol. Evol. **15:** 568–573.

Yang, Z., R. Nielsen, N. Goldman and A. M. Pedersen, 2000 Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155:** 431–449.

Zarkower, D., 2001 Establishing sexual dimorphism: conservation amidst diversity? Nat. Rev. Genet. **2:** 175–185.

Zhang, J., 2003 Evolution by gene duplication: an update. Trends Ecol. Evol. **18:** 292–298.

Zhang, J., and M. Nei, 1996 Evolution of Antennapedia-class homeobox genes. Genetics **142:** 295–303.

Zhang, J., and M. Nei, 1997 Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. J. Mol. Evol. **44** (Suppl. 1): S139–S146.

Zhang, J., S. Kumar and M. Nei, 1997 Small-sample tests of episodic adaptive evolution: a case study of primate lysozymes. Mol. Biol. Evol. **14:** 1335–1338.

Zhang, J., H. F. Rosenberg and M. Nei, 1998 Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc. Natl. Acad. Sci. USA **95:** 3708–3713.

Communicating editor: S. Yokoyama