Differential Rates of Evolution for the *ZFY*-Related Zinc Finger Genes, *Zfy*, *Zfx*, and *Zfa* in the Mouse Genus *Mus*

Priscilla K. Tucker, Ronald M. Adkins,[1] *and Joshua S. Rest*

Department of Ecology and Evolutionary Biology, and Museum of Zoology, University of Michigan, Ann Arbor

A comparative study of the last exon of the zinc finger genes *Zfx*, *Zfy*, and *Zfa* from species of mice in the genus *Mus* was conducted to assess the extent of gene-specific and chromosome-specific effects on the evolutionary patterns among related X-, Y-, and autosomal-linked genes. Phylogenetic analyses of 29 sequences from *Zfx*, *Zfa*, and *Zfy* from 10 taxa were performed to infer relatedness among the zinc finger loci, and codon-based maximum likelihood analyses were conducted to assess evolutionary pattern among genes. Five models of nucleotide sequence evolution were applied and compared using a likelihood ratio test. Estimates of nonsynonymous to synonymous changes (d_N/d_S) for these genes suggest that amino acid substitutions are occurring at a more rapid rate across the autosomal- and Y-specific lineages compared to the X-specific lineage, with the Y-specific lineage showing the highest rate under certain models. The data suggest the action of gene-specific effects on evolutionary pattern. In particular, *Zfa* and *Zfy* genes, both with presumed restricted expression, appear less functionally constrained relative to ubiquitously expressed *Zfx*. Slightly elevated d_N/d_S for *Zfy* genes in comparison to *Zfa* also suggest Y-specific effects.

Introduction

The ZFY-related zinc finger genes comprise a highly conserved vertebrate gene family (Page et al. 1987; Bull, Hillis, and O'Steen 1988; Sinclair et al. 1988; Zimmerer and Threlkeld 1995). They are characterized by an amino-terminal acidic domain, a putative localizing signal, and a carboxy-terminal DNA binding domain containing 13 zinc fingers (reviewed in Luoh et al. 1995) and are thought to function as transcription activators. In eutherian mammals these genes are sex-linked, having been incorporated into the sex chromosomes after the divergence of the lineages giving rise to marsupials and eutherian mammals well over 100 MYA (Kumar and Hedges 1998; Eizirik, Murphy, and O'Brien 2001; Ji et al. 2002).

Typically in eutherians there are two copies present, one on the X chromosome (*Zfx*) and one on the non-recombining portion of the Y chromosome (*Zfy*) (Page et al. 1987). However, multiple Y-chromosome–linked copies have been detected in a diverse group of rodents in the family Muridae. These include wood lemmings (subfamily Arvicolinae: Lau et al. 1992), South American oryzomyne-akodontine mice (subfamily Cricitinae: Bianchi et al. 1992) and mice belonging to the genus *Mus* (subfamily Murinae). In several species of *Mus* there are two copies on the Y chromosome (*Zfy-1* and *Zfy-2*) resulting from a recent intrachromosomal duplication, as well as an autosomal copy (*Zfa*) on chromosome 10, resulting from a recent retroposition of a processed *Zfx* transcript (Page et al. 1987; Ashworth, Swift, and Affara 1989; Mardon and Page 1989; Mardon et al. 1989; Mitchell et al. 1989; Nagamine et al. 1989; Mardon et al. 1990; Page et al. 1990).

Expression studies in laboratory mice demonstrate that these putatively functional genes have different patterns of expression. *Zfx* is ubiquitously expressed (Mardon et al. 1990). *Zfa* is expressed only in adult testes (Ashworth et al. 1990). Both *Zfy-1* and *Zfy-2* are expressed in the germ cells of adult testes. *Zfy-1* and possibly *Zfy-2* are expressed in somatic cells of the genital ridge and possibly at low levels in other tissues during development. *Zfy-1*, alone, is expressed in mouse embryonic stem cells and blastocysts (Koopman et al. 1989; Nagamine et al. 1989, 1990; Su and Lau 1992; Zwingman et al. 1993; Zambrowicz et al. 1994).

The presence of related and putatively functional genes located on the X and Y chromosomes as well as on an autosome in the genus *Mus* provides an opportunity to explore whether the evolution of specific genes is influenced by chromosomal location. Of particular interest is whether linkage to the Y chromosome can affect the pattern of gene evolution. Because the Y chromosome in mammals does not recombine with the X chromosome and is, thus, clonally inherited from father to son, Y-linked genes are potentially subject to a variety of phenomena that may result in higher rates of amino acid change relative to related genes elsewhere in the genome (reviewed in Tucker and Lundrigan 1995; Charlesworth and Charlesworth 2000).

Amino acid changes on the Y chromosome can reflect chromosome-specific effects such as an increased fixation of deleterious mutations caused by processes associated with the degeneration of the Y chromosome (Charlesworth and Charlesworth 2000). Specifically, an increase in the fixation of slightly deleterious mutations can result from the following phenomena: (1) genetic drift (Nei 1970) as a result of the Y chromosome having a smaller effective population size relative to the X chromosome and to the autosomes; e.g., when the male-to-female breeding sex ratio is one, Y-linked genes are only one-quarter as numerous as autosomes and one-third as numerous as X chromosomes; (2) Muller's ratchet (Muller 1964; Felsenstein 1974), a strictly stochastic process whereby the class of nonrecombining chromosomes with the fewest number of mutations is lost from the population; (3) background selection (Charlesworth, Morgan, and Charlesworth 1993; Charlesworth 1994), a process whereby nonrecombining chromosomes carrying strongly selected deleterious mutations are eliminated from the population, resulting in a reduced effective population of nonrecombining chromosomes, an increase in the fixation of

[1] Present address: Department of Pediatrics, and Center of Genomics and Bioinformatics, University of Tennessee-Memphis.

Key words: *Mus*, sex chromosomes, retrosequences, zinc finger genes, *Zfx*, *Zfy*, *Zfa*.

E-mail: ptuck@umich.edu.

Mol. Biol. Evol. 20(6):999–1005. 2003
DOI: 10.1093/molbev/msg112

Table 1
Taxonomic Source of Original Data Collected and Analyzed for the Last Exon of the *Zfx*, *Zfa*, and the *Zfy* Genes

	Origin	*Zfx*	*Zfa*	*Zfy-1*	*Zfy-2*
Mus domesticus	Maryland, USA	X	X	X	X
	Molise, Italy	X	X	X	X
	Lazio, Italy	X	X	X	X
	Zalende, Switzerland	X	X	X	X
	Tafilalt Oasis, Morocco			X	X
	Sondrio, Italy			X	X
Mus musculus musculus	Slovakia	X	X	X	X
	Viborg, Denmark	X	X	X	X
	Kyushu, Japan	X	X	X	X
Mus musculus castaneus	Chonburi Prov., Thailand	X	X	X	X
Mus spretus	Azrou, Morocco	X	X	X	X[b]
Mus spicilegus	Halbturn, Austria	X	X		X
Mus macedonicus[a]	Gradsko, Macedonia	X	X		X
Mus cookii	Tak Prov., Thailand	X	X		X
Mus caroli	Chonburi Prov., Thailand	X	X		X

[a] Data from two individuals collected from the same locality.

[b] *Zfy-2* from *M. spretus* is possibly not orthologous to *Zfy-2* from *M. musculus*.

slightly deleterious mutations, and a decrease in the fixation of mildly advantageous mutations (Charlesworth and Charlesworth 2000); (4) hitchhiking effects (Maynard Smith and Haigh 1974) in a nonrecombining genome where slightly deleterious alleles linked to a favorable mutation can become fixed; and (5) the Hill-Robertson effect (Hill and Robertson 1966; Felsenstein 1974, 1988; Birky and Walsh 1988) where alleles under selection interfere with selection at linked sites. All of these phenomena can have the effect of increasing the fixation of amino-acid changes at weakly constrained sites (see table 1 in Charlesworth and Charlesworth 2000).

Alternatively, amino acid changes on the Y chromosome can result from gene-specific effects such as positive Darwinian selection acting on Y-linked genes evolving independent of related X-linked genes. These can include sexually antagonistic genes such as genes involved in male mating success (reviewed in Rice 1996) or selfish growth-promoting Y-linked genes associated with embryonic development (Hurst 1994).

Finally, differences in functional constraint between related genes, also a gene-specific effect, may result in different evolutionary patterns. For example, expression patterns have been correlated with degree of functional constraint where ubiquitously expressed genes are under greater functional constraint than genes with restricted tissue expression (Hastings 1996; Duret and Mouchiroud 2000).

Here we compare the last zinc finger–encoding exon of *Zfx*, *Zfy-1*, *Zfy-2*, and *Zfa*, among species in the mouse genus *Mus* to determine whether these genes display different patterns of evolution and, if so, whether the patterns are consistent with chromosome-specific and/or gene-specific effects as described above. We provide evidence for distinctly elevated rates of amino acid change for the *Zfa* and *Zfy* genes in comparison to *Zfx*, and for slightly higher rates of amino acid change for the *Zfy* genes in comparison to the *Zfa* genes. We suggest that both gene-specific and chromosome-specific effects play a role in the differential evolution of the zinc finger genes.

Materials and Methods

Sampling

The taxonomic source for all original data collected is given in table 1. DNA sequence was obtained from *Zfx*, *Zfa*, and the *Zfy* genes corresponding to positions 1580–2575 from inbred mouse strain FVB/N *Zfy-2* mRNA (GenBank accession number NM_009571; Mardon and Page 1989). *Zfx*/*Zfy* sequences from three additional taxa were taken from the literature. These include *Zfx* and *Zfy* sequences from the Norway rat, *Rattus norvegicus* (Murinae) (GenBank accession numbers X75171, X75172; Shimmin, Chang, and Li 1994), a putative *Zfx* sequence from the Japanese spinous country-rat, *Tokudaia osimensis* (Murinae) (D83489; Xiao, Tsuchiya, and Sutou 1998), and an alligator (*Alligator mississippiensis*) zinc finger gene, *Azf-1* (X61714; Valleley et al., 1992).

PCR Amplification of Targeted Sequences

The targeted region was amplified from genomic DNA by polymerase chain reaction (PCR) using gene-specific primers and sequenced directly in both directions by the dideoxynucleotide chain termination method in a thermocycling reaction (GIBCO BRL), using single primers kinased with ^{32}P ATP. The last exon of *Zfy-1* differs from that of *Zfy-2* by the deletion of the fifth codon preceding the termination codon. This difference was used to design a primer specific for *Zfy-1* (5′-TTAG-GGCAGGCCAACTTT-3′). Other loci were amplified specifically by using unique primers at their 5′ end (*Zfa*: 5′-GCTTATGGTAATAATTCTGATGGA-3′; *Zfy*: 5′-GGCCCTGATGGACATCCTTTGAC-3′; *Zfx*: 5′-ACTA-AATCAGCATGTTTTGATCAC-3′) and a common primer at the 3′ end (5′-TTAGGGCAGGCCAACTT-CTTT-3′). *Zfy* sequences were verified as being male-specific by the failure to produce a product in PCR reactions from female *M. musculus* and *M. domesticus* in the same PCR experiment. To guarantee that only one copy of *Zfx* was sequenced per individual, *Zfx* sequences were obtained from males. *Zfa* sequences were from the same males and are present in two copies. There was no evidence of heterozygosity at this locus. GenBank accession numbers for original data compared in this study include AY159976 through AY160025.

Alignment

The conceptually translated amino acid sequences for all genes were aligned in ClustalX (Thompson et al. 1997) and forced back onto the nucleotide sequence. Identical sequences were combined in the data matrix, leaving a total of 29 sequences for subsequent analyses.

Phylogeny Reconstruction

A parsimony analysis was performed using the parsimony ratchet as implemented in WinClada 1.00.08 (Nixon 2002) with 500 iterations per replication, 10 trees

held per iteration, and 10 sequential ratchet runs. To assess support for the topology in the data set, a parsimony bootstrap (Felsenstein 1985) was performed using NONA (Goloboff 1999) with 1,000 replications, 10 search replicates, and one starting tree per replication.

For all parsimony searches, nucleotide characters were unordered and equally weighted, and gaps were treated as missing data. The alligator *Azf-1* sequence was used as an outgroup.

A Bayesian phylogenetic analysis was conducted with MrBayes 2.1 (Huelsenbeck and Ronquist 2001). A general time reversible (GTR) model with a gamma rate distribution and a proportion of invariable sites was used (Yang 1994). The analysis was initiated with random starting trees and was run for 1×10^6 generations, sampling every 100th generation. Four continuous chains were run with the initial 50,000 generations discarded as burn-in. To check that stationarity had been reached, the fluctuating value of the likelihood was checked graphically. The simulation was conducted twice.

Codon Based Likelihood Analyses

Codon based maximum likelihood analyses were conducted using PAML 3.12 (Yang 1999) to assess evolutionary patterns within and among genes. Two methods were used. The first method allows for analysis of lineage-specific d_N/d_S ratios within a phylogeny (Goldman and Yang 1994) where d_N is the number of nonsynonymous changes per nonsynonymous site and d_S is the number of synonymous changes per synonymous site. Five models of nucleotide sequence evolution were applied to the data set. They include (1) a ''one ratio'' model where the d_N/d_S ratio is assumed to hold for all lineages; (2) a ''two ratio'' model where the d_N/d_S ratios are allowed to vary between *Zfx*, *Zfa*, and *Azf-1* lineages (the X and autosomal genes) and *Zfy* lineages (the Y-linked genes); (3) a ''three ratio'' model where the d_N/d_S ratios are allowed to vary among *Azf-1* (a gene whose expression pattern is not fully determined [Valleley et al. 1992]), the *Zfy*/*Zfa* lineages (genes with putatively restricted expression), and the *Zfx* lineages (putative ubiquitously expressed genes); (4) a ''four ratio'' model where the d_N/d_S ratios are allowed to vary among the *Zfx*, *Zfy*, *Zfa*, and *Azf-1* gene lineages; and (5) a ''free ratio'' model where the d_N/d_S ratios are allowed to vary across all lineages. Given the data set and the inferred phylogeny, the fit of these four different models to the data was statistically compared using the likelihood ratio test. The test statistic compares twice the difference in log likelihood values with a χ^2 distribution with degrees of freedom equal to the difference in free parameters between the models.

The second method allows for the detection of positively selected sites within genes. The NSsites models in PAML (Yang 1999) were used to evaluate whether positive selection has acted on particular sites in *Zfy*, *Zfa*, and *Zfx* (Nielsen and Yang 1998). Two types of models were applied to the data set, one that allows for only neutral and negatively selected sites and one that allows for positively selected sites in addition to neutral and negatively selected sites. The test was also performed with a model that allows for heterogeneity of the d_N/d_S rate ratio among sites. As described above, the fit of the data to each of the models was statistically compared using the likelihood ratio test where the test statistic compares twice the difference in log likelihood values with a χ^2 distribution with degrees of freedom equal to the difference in free parameters between the models.

Results

Variation Within Species

No within species variation in *Zfx* sequences was found in either the *M. domesticus* or *M. musculus* sampled (table 1). In fact, *Zfx* sequences from these two species were identical to each other. No within-species variation in *Zfa* sequences was found within *M. domesticus* or within *M. musculus*. No within-species variation in *Zfy-1* sequences from *M. domesticus* was found. However, two variants of *Zfy-1* were found in *M. musculus*. The *M. musculus* from Japan differed from the other *M. musculus* samples at two nucleotide positions involving one synonymous and one nonsynonymous change. No within-species variation was found in *Zfy-2* sequences from *M. musculus*; however, two variants of *Zfy-2* were found in *M. domesticus*. The two *M. domesticus* from Molise and Lazio, Italy, respectively, differed from the remaining samples at a single site involving one synonymous change. Identical sequences were pooled for subsequent analyses.

Variation Within and Between Genes

Percent identity within the genus *Mus* among *Zfx* nucleotide sequences ranges from 99 to 100; among *Zfa* sequences, from 97.5 to 100; among *Zfy-1* sequences, from 98.5 to 100; and among *Zfy-2* sequences, from 95.8 to 100. Percent identity between all *Zfy* and all *Zfa* sequences ranges from 82.1 to 85.7; between all *Zfy* and *Zfx* sequences, from 82.8 to 85.8; and between all *Zfx* and *Zfa* sequences, from 96.9 and 98.7.

Phylogenetic Analyses

One-hundred and fifty-three equally most parsimonious trees were recovered from the heuristic search of 29 taxa. A strict consensus of these trees (not shown) is consistent with a 50% majority rule bootstrap consensus tree (fig. 1). Bayesian analyses yielded a compatible topology, the only differences being the resolution of two nodes that are collapsed in the parsimony analyses. Two distinct clades exist when the tree is rooted with the alligator zinc finger gene (*Azf-1*). One clade comprises the *Zfx* and *Zfa* sequences. The other clade comprises the *Zfy* sequences and one sequence from *Tokudaia osimensis* spp. labeled in GenBank as a *Zfx* gene (Xiao, Tsuchiya, and Sutou 1998). There is a lack of resolution among *Zfx* sequences after the *Mus–Rattus* split but varying degrees of resolution among *Zfa* sequences and *Zfy* sequences.

Evolutionary Patterns

Likelihood ratio tests (Goldman and Yang 1994; Yang 1998) to assess the fit of the five evolutionary models given the data set and phylogeny (the 50%

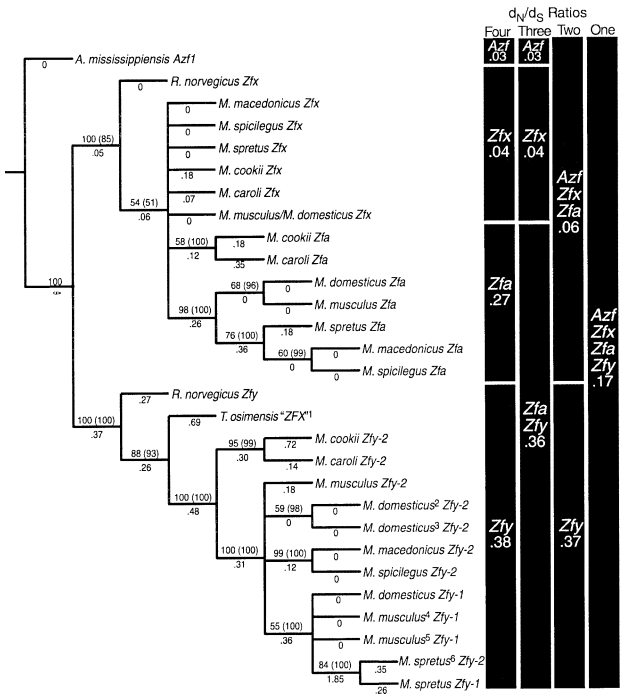

FIG. 1.—Parsimony bootstrap majority rule consensus tree for *Zfx*, *Zfy*, and *Zfa* genes from species of mice in the genus *Mus*, *Rattus norvegicus*, and *Tokudaia osimensis*. The autosomal-linked gene *Azf-1* from *Alligator mississippiensis* was used as an outgroup. Numbers above the branches are bootstrap values from the parsimony analysis. Numbers in parentheses are the posterior probabilities from a Bayesian likelihood analysis. Numbers below the branches are the d_N/d_S ratios under a free ratio model of codon sequence evolution across lineages. Zeros reflect no nonsynonymous changes. The ∞ represents nonsynonymous change with no synonymous change. To the right, d_N/d_S are given for four additional models of codon sequence evolution across lineages as described in the methods. [1]This sequence was entered into GenBank as a *Zfx* sequence because of the absence of a Y chromosome in this taxon although, clearly, it is of *Zfy* origin. [2]Representing *M. domesticus* from Molise and Lazio, Italy. [3]Representing *M. domesticus* from USA, Switzerland, Morocco, and Sondrio, Italy. [4]Representing *M. musculus* from Japan. [5]Representing *M. musculus* from Slovakia, Denmark, Austria, and Thailand. [6]*Zfy-2* from *M. spretus* is not orthologous to *Zfy-2* from other taxa (see text).

majority rule consensus tree described above) indicate that the free ratio model, the two ratio model (the d_N/d_S value for the Y-specific gene lineage varies independently from all other lineages), the three ratio model (the d_N/d_S value for the *Zfx* gene lineage varies independently from the d_N/d_S for *Zfa*/*Zfy* gene lineages and from *Azf-1*), and the four ratio model (d_N/d_S values for the X-, Y-, autosomal-specific, and *Azf-1* lineages vary independently from each other) are all a significantly better fit to the data than the one ratio model (d_N/d_S are constant across all gene lineages) (table 2). The three, four, and free ratio models are significantly better ($P < 0.05$) than the two ratio model. The free and four ratio models are not significantly better than the three ratio model, and the free ratio model is not significantly better than the four ratio model. Taken together, these analyses suggest that the three ratio model best describes the data and indicate that there is not significant variation in d_N/d_S within the four zinc finger gene lineages. Rather, the zinc finger genes are evolving in a gene-specific manner that is possibly associated with tissue expression patterns.

Estimates of d_N/d_S under the three, four, and free ratio models suggest that the amino acid substitutions are occurring at a more rapid rate across Y-specific and autosomal lineages compared to the X-specific lineages (fig. 1). Under the three ratio model the d_N/d_S for *Zfy*/*Zfa* is nine times greater than that for *Zfx*. Under the four ratio model, the d_N/d_S for *Zfy* is also nine times greater than that for *Zfx* but only slightly greater than that for *Zfa*, and the d_N/d_S for *Zfa* is six times greater than that for *Zfx*. The d_N/d_S ratios under the free ratio model (also shown in fig. 1) show a similar pattern.

Likelihood analyses for detecting positively selected amino acid sites yielded varying results. For *Zfx*, the model allowing positive selection performs significantly better ($P < 0.05$) than the model allowing for only neutral or negatively selected sites, both when the d_N/d_S rate ratio is constant and when it is allowed to vary across sites according to a beta distribution (data not shown). According to these models, a single codon (position 185 in our alignment) is positively selected ($d_N/d_S > 1$). However, the meaning of positive selection at this site is difficult to interpret without more structural and/or functional information on the *Zfx* protein. For *Zfy* and *Zfa*, models that allow for positive selection do not perform significantly better than models that allow for only neutral or negatively selected sites.

Discussion

Phylogenetic Analyses

The existence of two distinct clades, one comprising the *Zfx* and *Zfa* sequences and the other comprising the *Zfy* sequences, supports the hypothesis that *Zfa* originated as a retroposition of a processed *Zfx* transcript. Monophyly of the *Zfx* genes and, with one anomaly (see below), the *Zfy* genes suggest that gene conversion has not played a major role in the evolution of these two genes in the murine rodents sampled. This is in contrast to evidence for gene conversion in human *ZFX*/*ZFY* and in *Zfx*/*Zfy* from the crab-eating fox (*Dusicyon thous*) (Pamilo and Bianchi 1993). A lack of resolution among the *Zfx* sequences after the *Mus–Rattus* split is a result of a lack of phylogenetic signal as a result of the conserved nature of these sequences.

The clade comprising the *Zfy* sequences includes a sequence from *Tokudaia osimensis* spp. labeled in GenBank as a *Zfx* gene (Xiao, Tsuchiya, and Sutou 1998). This taxon has a 2N = 45 XO karyotype with no distinguishable differences in karyotype between males and females (Honda et al., 1978). Because of the absence of the Y chromosome in this taxon, the gene was labeled *Zfx*. However, based on our phylogenetic analyses, this sequence is clearly related to *Zfy*. In all likelihood this gene is the result of a ''recent'' translocation of *Zfy* from the Y to the X chromosome associated with the loss of the Y chromosome from this taxon. A closely related taxon, *T. osimensis muenninki*, in addition to the closely related genus *Rattus* has the standard XX/XY sex-dermining system (Sutou, Mitsui, and Tsuchiya 2001), indicating that the lack of the Y chromosome is a recently derived condition. Although the translocation of *Zfy* to the X chromosome in populations of *Tokudaia osimensis* could in theory

Table 2
χ^2 Significance Values for Likelihood Ratio Tests Comparing Four Different Models for d_N/d_S Among Lineages

			One Ratio			Two Ratio			Three Ratio			Four Ratio		
	L	Free Parameter	2ΔL	df	*P*	2ΔL	df	*P*	2ΔL	df	*P*	2ΔL	df	*P*
One ratio	−3672	1												
Two ratio	−3637	2	70	1	<0.01									
Three ratio	−3621	3	101	2	<0.01	31	1	<0.01						
Four ratio	−3621	4	102	3	<0.01	32	2	<0.01	1	1	0.30			
Free ratio	−3605	46	136	45	<0.01	66	44	0.02	35	43	0.81	34	42	0.82

NOTE.—One ratio model: d_N/d_S ratio is constant for all lineages; two ratio model: d_N/d_S ratios are allowed to vary between *Zfx*, *Zfa* and *Azf-1* (the X and autosomal copies) on the one hand and *Zfy* on the other; three ratio model: d_N/d_S ratios are allowed to vary among *Zfx*, *Zfa*/*Zfy*, and *Azf-1*; four ratio model: d_N/d_S ratios are allowed to vary among *Zfx*, *Zfy*, *Zfa*, and *Azf-1* lineages; and free ratio model: d_N/d_S ratios are allowed to vary across all lineages.

provide a test for the effects of chromosome location on gene evolution, the translocation event probably occurred too recently to detect changes in evolutionary pattern.

The clade comprising the *Zfy-1* genes includes a sequence labeled *Zfy-2* from *M. spretus*. This sequence is more closely related to *Zfy-1* than to other *Zfy-2* sequences and suggests either that the sequence is a recent intrachromosomal duplication of *Zfy-1*—i.e., it is nonhomologus to other *Zfy-2* genes—or a *Zfy-2* sequence has undergone gene conversion. An analysis of the 3′ portion of the gene may shed light on these alternatives.

Evolutionary Patterns

Variation in d_N/d_S ratios across genes over time is generally attributed to differential selective constraints. Typically, genes appear to be highly selectively constrained and d_N/d_S values are low. For example, Wolfe and Sharp (1993) estimated the average d_N/d_S for 363 genes compared between mouse and rat to be 0.14. The elevated rates for *Zfy* (0.38 under the four ratio model) and *Zfa* (0.27) may indicate such gene-specific effects as the action of positive Darwinian selection, reduced selective constraint, or, in the case of *Zfy*, its position in the nonrecombining region of the Y chromosome.

When d_N/d_S ratios are greater than one for a gene, positive selection is hypothesized to account for the rapid amino acid divergence. However, d_N/d_S ratios greater than one are a conservative estimate of positive selection, especially when made over long periods of time because positive selection can be episodic and followed by purifying selection (Zhang, Rosenberg, and Nei 1998). This can result in a muted signal for positive selection (Schaner et al. 2001), i.e., elevated d_N/d_S that are not greater than 1. While ratios of d_N/d_S for the *Zfy* genes are generally not greater than 1, the exception being the lineage giving rise to the *M. spretus Zfy* genes (fig. 1), they are relatively high for a functional gene, especially compared to the related gene, *Zfx*, which presumably shares a similar function (i.e., transcription activation).

This pattern is similar to what was found for another Y-chromosome–linked functional gene, the male sex–determining locus, *Sry* (Whitfield, Lovell-Badge, and Goodfellow 1993; Tucker and Lundrigan 1993; Pamilo and O'Neill 1997; Wang, Zhang, and Zhang 2002; Jansa, Lundrigan, and Tucker in press). In comparative studies of primates and rodents d_N/d_S values for *Sry* ranged from 0.47 to 1.88 for primates and from 0.33 to 0.45 for rodents, with values being especially high in the C-terminal region of the gene (Tucker and Lundrigan 1993; Whitfield, Lovell-Badge, and Goodfellow 1993; Pamilo and O'Neill 1997). In addition, variation in the elevated d_N/d_S was observed among lineages (Wang, Zhang, and Zhang 2002). Both weak positive Darwinian selection and purifying selection have been offered as explanations for these patterns (reviewed in O'Neill and O'Neill 1999; Wang, Zhang, and Zhang 2002; Jansa, Lundrigan, and Tucker 2003). However, studies of 12 closely related species of rock wallaby (*Petrogale*) indicated that *Sry* evolution was not rapid over a short evolutionary time (O'Neill et al. 1997). In O'Neill et al. (1997), the d_N/d_S ratios were quite low with 71% being less than 0.03 (reviewed in O'Neill and O'Neill 1999). Furthermore, a population level study of *M. domesticus* (Nachman and Aquadro 1994), in which levels of polymorphism to divergence were compared for *Sry* flanking sequence, revealed no evidence for selection acting on the nonrecombining portion of the Y chromosome. Taken together, these data suggest that positive Darwinian selection may not account for the elevated d_N/d_S in *Sry*, at least not over short evolutionary time periods. This may also be the case for *Zfy*, as there is no other evidence for positive selection acting on this gene; e.g., no positively selected sites were detected in *Zfy* using the codon-based likelihood analysis.

Assuming that expression patterns are similar within gene lineages, the elevated d_N/d_S ratios for *Zfy* and *Zfa* genes in comparison to the *Zfx* genes is consistent with the prediction that ubiquitously expressed genes are under greater functional constraint than genes with limited tissue expression (Hastings 1996; Duret and Mouchiroud 2000). The more slowly evolving *Zfx* is ubiquitously expressed in inbred mouse strains (Mardon et al. 1990) that are of *M. domesticus*/*M. musculus* origin (Tucker et al. 1992) in contrast to *Zfy* and *Zfa* tissue expression, which is limited primarily to the adult testis in inbred mouse strains (Koopman et al. 1989; Nagamine et al. 1989, 1990; Ashworth et al. 1990; Su and Lau 1992; Zwingman et al. 1993; Zambrowicz et al., 1994).

The even higher d_N/d_S for *Zfy* genes in comparison to *Zfa* genes under the four and free ratios models, however, may reflect inefficient selection on weakly constrained sites, which is an outcome of several of the hypotheses proposed to explain the degeneration of the Y chromosome over evolutionary time (Charlesworth and

Charlesworth 2000). Indeed, in a mouse–rat comparison of 834 genes the d_N/d_S for genes with the most restricted tissue expression is lower (avg. = 0.16; Duret and Mouchiroud 2000) than was found for *Zfy*. A study of the evolution of an X- and Y-linked gene pair exhibiting similar tissue expression would provide a more straightforward interpretation of the role of chromosome-specific effects on gene evolution.

Acknowledgments

We thank M. Potter (National Cancer Institute contract N01-CB-71085), E. Eicher (The Jackson Laboratory), R. Sage (University of California Santa Barbara), and M. Nachman (University of Arizona) for providing frozen tissue or DNA from various stocks and strains of wild mice. J. Zhang, S. Sandstedt, and M. Nachman provided numerous suggestions that improved the manuscript, and A. Eyre-Walker and two anonymous reviewers provided comments that improved a revision. R.M.A. was supported by the Alfred P. Sloan Foundation; P.K.T., by the National Science Foundation, DEB-9209950; and J.S.R., by National Institutes of Health grant T32-HG00040.

Literature Cited

Ashworth, A., B. Skene, S. Swift, and R. Lovell-Badge. 1990. *Zfa* is an expressed retroposon derived from an alternative transcript of the *Zfx* gene. Embo. J. **9**:1529–1534.

Ashworth, A., S. Swift, and N. Affara. 1989. Sequence of cDNA for murine *Zfy-1*, a candidate for *Tdy*. Nucleic Acids Res. **17**: 2864.

Bianchi, N. O., M. S. Bianchi, P. Pamilo, L. Vidal-Rioja, and A. de la Chapelle. 1992. Evolution of zinc finger-Y and zinc finger-X genes in oryzomyne-akodontine rodents (Cricetidae). J. Mol. Evol. **34**:54–61.

Birky, C. W., Jr., and J. B. Walsh. 1988. Effects of linkage on rates of molecular evolution. Proc. Natl. Acad. Sci. USA **85**: 6414–6418.

Bull, J. J., D. M. Hillis, and S. O'Steen. 1988. Mammalian *ZFY* sequences exist in reptiles regardless of sex-determining mechanism. Science **242**:567–569.

Charlesworth, B. 1994. The effect of background selection against deleterious mutations on weakly selected, linked variants. Genet. Res. **63**:213–227.

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

Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics **134**:1289–1303.

Duret, L., and D. Mouchiroud. 2000. Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. Mol. Biol. Evol. **17**:68–74.

Eizirik, E., W. J. Murphy, and S. J. O'Brien. 2001. Molecular dating and biogeography of the early placental mammal radiation. J. Hered. **92**:212–219.

Felsenstein, J. 1974. The evolutionary advantage of recombination. Genetics **78**:737–756.

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

———. 1988. Sex and the evolution of recombination. Pp. 74–86 *in* R. E. Michod and B. R. Levin, eds. The evolution of sex. Sinauer Associates, Sunderland, Mass.

Goldman, N., and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. **11**:725–736.

Goloboff, P. 1999. NONA (NO NAME) ver. 2. Published by the author, Tucuman Argentina.

Hastings, K. E. M. 1996. Strong evolutionary conservation of broadly expressed protein isoforms in the troponin I gene family and other vertebrate gene families. J. Mol. Evol. **42**: 631–640.

Hill, W. G., and A. Robertson. 1966. Linkage disequilibrium in finite populations. Theor. Appl. Genet. **38**:226–231.

Honda, T., H. Suzuki, M. Itoh, and K. Hayashi. 1978. Karyotypical differences of the Amami spinous country-rats, *Tokudaia osimensis osimensis* obtained from two neighboring islands. Jpn. J. Genet. **53**:297–299.

Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17**:754–755.

Hurst, L. D. 1994. Embryonic growth and the evolution of the mammalian Y chromosome. I. The Y as an attractor for selfish growth factors. Heredity **73**:223–232.

Jansa, S. A., B. L. Lundrigan, and P. K. Tucker. 2003. Tests for positive selection on immune and reproductive genes in closely related species of the murine genus *Mus*. J. Mol. Evol. **56**:294–307.

Ji, Q., Z. X. Luo, C. X. Yuan, J. R. Wible, J. P. Zhang, and J. A. Georgi. 2002. The earliest known eutherian mammal. Nature **416**:816–822.

Koopman, P., J. Gubbay, J. Collignon, and R. Lovell-Badge. 1989. *Zfy* gene expression patterns are not compatible with a primary role in mouse sex determination. Nature **342**: 940–942.

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

Lau, Y. F., T. L. Yang-Feng, B. Elder, K. Fredga, and U. H. Wiberg. 1992. Unusual distribution of *Zfy* and *Zfx* sequences on the sex chromosomes of the wood lemming, a species exhibiting XY sex reversal. Cytogenet. Cell Genet **60**:48–54.

Luoh, S. W., K. Jegalian, A. Lee, E. Y. Chen, A. Ridley, and D. C. Page. 1995. CpG islands in human *ZFX* and *ZFY* and mouse *Zfx* genes: sequence similarities and methylation differences. Genomics **29**:353–363.

Mardon, G., S. W. Luoh, E. M. Simpson, G. Gill, L. G. Brown, and D. C. Page. 1990. Mouse *Zfx* protein is similar to *Zfy-2*: each contains an acidic activating domain and 13 zinc fingers. Mol. Cell Biol. **10**:681–688.

Mardon, G., R. Mosher, C. M. Disteche, Y. Nishioka, A. McLaren, and D. C. Page. 1989. Duplication, deletion and polymorphism in the sex-determining region of the mouse Y chromosome. Science **243**:78–80.

Mardon, G., and D. C. Page. 1989. The sex-determining region of the mouse Y chromosome encodes a protein with a highly acidic domain and 13 zinc fingers. Cell **56**:765–770.

Maynard Smith, J., and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. Genet. Res. **23**:23–35.

Mitchell, M., D. Simon, N. Affara, M. Ferguson-Smith, P. Avner, and C. Bishop. 1989. Localization of murine X and autosomal sequences homologous to the human Y located testis-determining region. Genetics **121**:803–809.

Muller, H. J. 1964. The relation of recombination to mutational advance. Mutat. Res. **1**:2–9.

Nachman, M. W., and C. F. Aquadro. 1994. Polymorphism and divergence at the 5′ flanking region of the sex-determining locus, *Sry*, in mice. Mol. Biol. Evol. **11**:539–547.

Nagamine, C. M., K. Chan, L. E. Hake, and Y. F. Lau. 1990. The

two candidate testis-determining Y genes (*Zfy-1* and *Zfy-2*) are differentially expressed in fetal and adult mouse tissues. Genes Dev. **4**:63–74.

Nagamine, C. M., K. M. Chan, C. A. Kozak, and Y. F. Lau. 1989. Chromosome mapping and expression of a putative testis-determining gene in mouse. Science **243**:80–83.

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

Nielsen, R., and Z. Yang. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics **148**:929–936.

Nixon, K.C. 2002. WinClada ver. 1.00.08. Published by the author, Ithaca, N.Y.

O'Neill, M. J., and R. J. W. O'Neill. 1999. Whatever happened to *SRY*? Cell Mol. Life Sci. **56**:883–893.

O'Neill, R. J. W., M. D. B. Eldridge, R. H. Crozier, and J. A. M. Graves. 1997. Low levels of sequence divergence in rock wallabies (*Petrogale*) suggest a lack of positive directional selection in *SRY*. Mol. Biol. Evol. **14**:350–353.

Page, D. C., C. M. Disteche, E. M. Simpson, A. de la Chapelle, M. Andersson, T. Alitalo, L. G. Brown, P. Green, and G. Akots. 1990. Chromosomal localization of *ZFX*—a human gene that escapes X inactivation—and its murine homologs. Genomics **7**:37–46.

Page, D. C., R. Mosher, E. M. Simpson, E. M. Fisher, G. Mardon, J. Pollack, B. McGillivray, A. de la Chapelle, and L. G. Brown. 1987. The sex-determining region of the human Y chromosome encodes a finger protein. Cell **51**: 1091–1104.

Pamilo, P., and N. O. Bianchi. 1993. Evolution of the *Zfx* and *Zfy* genes: rates and interdependence between the genes. Mol. Biol. Evol. **10**:271–281.

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

Rice, W. R. 1996. Evolution of the Y sex chromosome in animals. BioScience **46**:331–343.

Schaner, P., N. Richards, A. Wadhwa, I. Aksentijevich, D. Kastner, P. Tucker, and D. Gumucio. 2001. Episodic evolution of pyrin in primates: human mutations recapitulate ancestral amino acid states. Nat. Genet. **27**:318–321.

Shimmin, L. C., B. H. Chang, and W. H. Li. 1994. Contrasting rates of nucleotide substitution in the X-linked and Y-linked zinc finger genes. J. Mol. Evol. **39**:569–578.

Sinclair, A. H., J. W. Foster, J. A. Spencer, D. C. Page, M. Palmer, P. N. Goodfellow, and J. A. Graves. 1988. Sequences homologous to *ZFY*, a candidate human sex-determining gene, are autosomal in marsupials. Nature **336**:780–783.

Su, H., and Y. F. Lau. 1992. Demonstration of a stage-specific expression of the *ZFY* protein in fetal mouse testis using anti-peptide antibodies. Mol. Reprod. Dev. **33**: 252–258.

Sutou, S., Y. Mitsui, and K. Tsuchiya. 2001. Sex determination without the Y chromosome in two Japanese rodents, *Tokudaia osimensis osimensis* and *Tokudaia osimensis* spp. Mamm. Genome **12**:17–21.

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

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.

———. 1995. The nature of gene evolution on the mammalian Y chromosome: lessons from *Sry*. Philos. Trans. R. Soc. Lond. B **350**:221–227.

Tucker, P. K., B. K. Lee, B. L. Lundrigan, and E. M. Eicher, 1992. Geographic origin of the Y chromosomes in ''old'' inbred strains of mice. Mamm. Genome **3**:254–261.

Valleley, E. M., U. Muller, M. W. Ferguson, and P. T. Sharpe. 1992. Cloning and expression analysis of two *ZFY*-related zinc finger genes from *Alligator mississippiensis*, a species with temperature-dependent sex determination. Gene **119**: 221–228.

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.

Wolfe, P. M., and P. M. Sharp. 1993. Mammalian gene evolution: nucleotide sequence divergence between mouse and rat. J. Mol. Evol. **37**:441–456.

Xiao, C., K. Tsuchiya, and S. Sutou. 1998. Cloning and mapping of bovine *ZFX* gene to the long arm of the X-chromosome (Xq34) and homologous mapping of *ZFY* gene to the distal region of the short arm of the bovine (Yp13), ovine (Yp12-p13), and caprine (Yp12-p13) Y chromosome. Mamm. Genome **9**:125–130.

Yang, Z. 1994. Estimating the pattern of nucleotide substitution. J. Mol. Evol. **39**:105–111.

———. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. **15**:568–573.

———. 1999. PAML: phylogenetic analysis by maximum likelihood. University College, London.

Zambrowicz, B. P., J. W. Zimmermann, C. J. Harendza, E. M. Simpson, D. C. Page, R. L. Brinster, and R. D. Palmiter. 1994. Expression of a mouse *Zfy-1/lacZ* transgene in the somatic cells of the embryonic gonad and germ cells of the adult testis. Development **120**:1549–1559.

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.

Zimmerer, E. J., and L. Threlkeld. 1995. A *ZFY*-like sequence in fish, with comments on the evolution of the *ZFY* family of genes in vertebrates. Biochem. Genet. **33**:227–235.

Zwingman, T., R. P. Erickson, T. Boyer, and A. Ao. 1993. Transcription of the sex-determining region genes *Sry* and *Zfy* in the mouse preimplantation embryo. Proc. Natl. Acad. Sci. USA **90**:814–817.

Adam Eyre-Walker, Associate Editor

Accepted February 19, 2003