J Mol Evol (1994) 39:569–578

Journal of Molecular Evolution

Contrasting Rates of Nucleotide Substitution in the X-Linked and Y-Linked Zinc Finger Genes

Lawrence C. Shimmin, Benny H.-J. Chang, Wen-Hsiung Li

Human Genetics Center, University of Texas Houston Health Science Center, P.O. Box 20334, Houston, TX 77225, USA

Received: 11 March 1994 / Accepted: 17 June 1994

Abstract. We have sequenced the entire exon (~1,180 bp) encoding the zinc finger domain of the X-linked and Y-linked zinc finger genes (*ZFX* and *ZFY*, respectively) in the orangutan, the baboon, the squirrel monkey, and the rat; a total of 9,442 bp were sequenced. The ratio of the rates of synonymous substitution in the *ZFY* and *ZFX* genes is estimated to be 2.1 in primates. This is close to the ratio of 2.3 estimated from primate *ZFY* and *ZFX* intron sequences and supports the view that the male-to-female ratio of mutation rate in humans is considerably higher than 1 but not extremely large. The ratio of synonymous substitution rates in *ZFY* and *ZFX* is estimated to be 1.3 in the rat lineage but 4.2 in the mouse lineage. The former is close to the estimate (1.4) from introns. The much higher ratio in the mouse lineage (not statistically significant) might have arisen from relaxation of selective constraints. The synonymous divergence between mouse and rat *ZFX* is considerably lower than that between mouse and rat autosomal genes, agreeing with previous observations and providing some evidence for stronger selective constraints on synonymous changes in X-linked genes than in autosomal genes. At the protein level *ZFX* has been highly conserved in all placental mammals studied while *ZFY* has been well conserved in primates and foxes but has evolved rapidly in mice and rats, possibly due to relaxation of functional constraints as a result of the development of X-inactivation of *ZFX* in rodents. The long persistence of the *ZFY-ZFX* gene pair in mammals provides some insight into the process of degeneration of Y-linked genes.

Key words: Sex ratio of mutation rate — Silent substitution — Amino acid substitution — Selective constraints — Loss of Y-linked genes

Correspondence to: W.-H. Li

Introduction

Taking advantage of the existence of a homologous pair of zinc finger genes on the Y and X chromosomes (denoted *ZFY* and *ZFX*, respectively) in all placental mammals investigated (Mardon et al. 1990; Lanfear and Holland 1991), we (Shimmin et al. 1993a; Chang et al. 1993) have recently studied the male/female ratio of mutation rate (α_m) in higher primates and rodents. We sequenced the last introns of the *ZFY* and *ZFX* genes in humans, orangutans, baboons, squirrel monkeys, mice, and rats and used Miyata et al.'s (1987) theory to obtain an estimate of $\alpha_m = 6$ for higher primates and an estimate of $\alpha_m = 2$ for mice and rats. Our estimate of $\alpha_m = 6$ in higher primates and Ketterling et al.'s (1993) estimate of $\alpha_m = 3.5$ for humans from hemophilia B families, which are not significantly different from each other, provide support for Haldane's (1947) view that the rate of mutation is higher in males than in females because the number of germ cell divisions is much higher in males than in females.

In the present study we have sequenced the last exons of the *ZFY* and *ZFX* genes in organgutans, baboons,

squirrel monkeys, and rats; the corresponding sequences in humans and mice have already been determined (Schneider-Gädicke et al. 1989; Page et al. 1987; Mardon et al. 1990; Ashworth et al. 1989; Mardon and Page 1989). Our purpose is twofold. First, we wish to see whether the ratio α_m estimated from synonymous substitutions in *ZFY* and *ZFX* is similar to that estimated from introns. Miyata et al. (1987) compared the average rate of synonymous substitution in X-linked genes with that in autosomal genes from humans and rodents and concluded that α_m is infinitely large. The same conclusion was reached in a study of rodent genes (Wolfe and Sharp 1993). These results seem to imply that synonymous changes are on average subject to stronger selective constraints in X-linked genes than in autosomal genes (Charlesworth 1993; Shimmin et al. 1993a,b). However, the putative difference in selective constraints might be reduced by using a homologous pair such as *ZFY* and *ZFX*. Indeed, previous estimates of α_m based on synonymous substitutions from limited *ZFY* and *ZFX* sequence data (Lanfear and Holland 1991; Hayashida et al. 1992; Pamilo and Bianchi 1993) were similar to our estimates from *ZFY* and *ZFX* introns. We wish to know whether this observation is supported by further data and whether synonymous changes are indeed subject to ssignificant selective constraints. Second, we wish to see whether *ZFY* has evolved much faster than *ZFX* at the amino acid sequence level. While the *ZFX* gene appears to serve some important function in both females and males and would not be dispensable, the *ZFY* gene is carried by males only and so might be dispensable and be free to change its function or to become nonfunctional. There is, therefore, a good possibility that *ZFY* has evolved faster than *ZFX* in some mammalian lineages (Lanfear and Holland 1991). We chose to study the last exon of *ZFY* and *ZFX* because it is fairly long (~1,180 bp) and becuase it encodes 13 zinc finger DNA binding motifs, which appear to serve the DNA recognition/binding function of the two gene products.

Materials and Methods

Sample Sources. Genomic DNA was isolated from 50 mg of baboon liver (*Papio cyanocephalus*) or rat kidney tissue (*Rattus norvegicus* strain F334) by the method of Ellsworth et al. (1993), or was received as a gift—orangutan (*Pongo pygmaeus*) from M. Schriver and squirrel monkey (*Siamiri boliviensis*) from S.-K. Shyue, both of the University of Texas at Houston. All genomic DNA samples were isolated from males.

PCR Amplification. Amplification of the entire ZF domain exon by the polymerase chain reaction (PCR) in high yield proved difficult; hence, it was amplified in two overlapping fragments. The sequence of the oligonucleotides used in amplifying the partial genes are listed in Table 1, along with the gene and species which they amplified. In each case the PCR product containing the 5′ end of a ZF-domain exon utilized an X- or Y-specific 5′ PCR primer located in the previously sequenced upstream intron (Shimmin et al. 1993a; Chang et al. 1993) and a 3′ PCR primer (oLW111) matched to the consensus of the available human and mouse ZF exon seuqences located at positions 1031–1008 (Fig. 1). These amplifications yielded fragments approximately 1,070–1,210 bp long depending upon the gene and species. To obtain the 3′ end of the ZF-domain exon, PCR reactions utilizing a consensus 5′ PCR primer (oLW142) at position 834–862 (Fig. 1) and X- and Y-linkage-specific 3′ PCR primers matched to the unique 3′ untranslated regions successfully yielded products approximately 550 bp long for all exons except the rat Y-linked gene. For this latter fragment a separate oligonucleotide pair with both primers unique to the Y-linked gene in positions 267–302 (Fig. 1, taking advantage of a deletion in the rat Y-linked gene with respect to the rat X-linked homologoue) and in the 3′-untranslated region resulted in amplification of the 3′ fragment of the Y-linked gene. All PCR reactions contained 10 ng μl^{-1} genomic DNA, 0.04 U μl^{-1} Taq DNA polymerase, 1 μM each of the appropriate oligonucleotide primers, 10 mM Tris-HCl pH 8.8, 50 mM KCl, 0.1% Triton X-100, 200 μM deoxynucleotide triphosphates, and 2.5 mM $MgCl_2$. Target genes were amplified in a Perkin-Elmer Cetus 480 DNA thermocycler with an initial denaturation step of 5 min at 94°C; 30 cycles consisting of denaturation 94°C for 1 min, annealing 55°C to 58°C for 1 min, and extension 72°C for 2 min + 4 s autoextension; followed by a single final extension at 72°C for 5 min. The conditions were tuned to yield the majority incorporation into a single product of the expected length.

Sequencing. PCR products were purified with Magic PCR Preps directly from the reaction mix, without gel purification, in accordance with the manufacturer's methodology (Promega, Madison WI). The use of males as the source of genomic DNA eliminated the possibility of polymorphisms being present in the final X- and Y-linked PCR products. To avoid the problem of artifacts present in clones of PCR products generated by the low fidelity of Taq DNA polymerase, we sequenced the purified PCR products directly by enzymatic methods (Sanger et al. 1977) using the Promega *fmol* sequencing kit essentially as described by the manufactuer (Promega, Madison WI). Sequencing reactions contained 10–25 fmol of template, 0.4 pmol [^{32}P]-end-labeled oligonucleotide primer (listed in Table 1, synthesized on an Applied Biosystems 391 Synthesizer), and subjected to a cycle-sequencing regime of a single initial denaturation cycle of 94°C for 100 s; 25 cycles of denaturation 30 s at 94°C and annealing/extension 30 s at ~55–59°C (typically 5°C above calculated T_m of the specific oligonucleotide primer). Every position of both DNA strands was determined.

Linkage Analyis. The 5′ ends of the zinc finger exon PCR products in this study overlap the 3′ ends of the upstream zinc finger intron PCR products (which included 185 bp of the 5′ end of the zinc finger exon) we had previously characterized and for which X- or Y-linkage had been confirmed (Shimmin et al. 1993a; Chang et al. 1993). The linkage of the exon products in this study was checked by direct comparison of the overlapping intron and exon sequences; matching X- or Y-specific sequences confirmed that the PCR primer pairs used in this study had yielded the specific genes that they were designed to amplify.

Results

Characterization of the Exons

We have isolated the exon encoding the zinc finger domain of the X-linked *ZFX* and Y-linked *ZFY* genes by PCR amplification from genomic DNA isolated from males of an orangutan, a baboon, a squirrel monkey, and a rat. The entire exon (1,175 bp to 1,181 bp depending

Table 1. Oligonucleotides

PCR Primers Oligo	Sequence: 5′-3′	Specificity[a]	Position[b]	*ZFX*[c]	*ZFY*[c]
oLW108	TTA(TC)ATTCGCAAAGAAACTG(GC)AAC	Unique	~−60 to −44	OB	
oLW109	TTACATTCATGAAGAAACTGGAAC	Unique	−69 to −45	S	
oLW122	AGCATGTTTTGATCACTTCTGCTCC	Unique	−36 to −12	R	
oLW110	ATTCATGAGGA(AG)ACCAGAAGTTTG	Unique	−60 to −37		OBS
oLW124	ACCCTGTCATAGTGGGTCAGATTC	Unique	−175 to −152		R
oLW142	AATGCACCA(AG)TGTAGACATTGTGACTTTAA	Consensus	834 to 862	OBSR	OBS
oLW152	GGGACTTTGTGTACTCACAAAAAAGAA	Unique	267 to 302		R
oLW111	ATA(CT)TCACAGTACTCACACTGATA	Consensus	1031 to 1008	OBSR	OBS
oLW143	AACAAACTGAATAGAATTCAGAACACAC	Unique	~+195 to +165	OBSR	
oLW144	AGAAACTGAATAGGATTCAGAACAGCG	Unique	~+195 to +165		OBS
oLW148	AATGAATTTACTGTCTTCAATGCCA	Unique	~+55 to +30		R

PCR Amplifications Primer pair	Coding region amplified[b]	Product size (bp)		
oLW108/oLW111	1–1,031	~1,100	OB	
oLW109/oLW111	1–1,031	~1,100	S	
oLW122/oLW111	1–1,031	~1,070	R	
oLW110/oLW111	1–1,031	~1,090	OBS	
oLW124/oLW111	1–1,031	~1,210	R	
oLW142/oLW143	834–1,184	~550	OBSR	
oLW142/oLW144	834–1,184	~550		OBS
oLW152/oLW148	267–1,184	~975		R

[a] "Unique" indicates that the primer is chromosome specific; "consensus" indicates that primer matches both the X- and Y-linked genes

[b] Negative values indicate positions within the upstream intron; values preceded by "+" indicate positions within the 3′-untranslated region; otherwise the position is indicated with respect to sequence alignment of Fig. 1. The position of some of the primers located within the upstream intron and within the 3′ untranslated region varies depending upon the species

[c] Species for which the primer was designed: O, orangutan; B, baboon; S, squirrel monkey and R, rat

upon species) was sequenced using direct sequencing of the PCR products, thus eliminating the introduction of artifacts due to the low fidelity of Taq DNA polymerases and, by virtue of selecting males as sources of genomic DNA, eliminating the possibility of polymorphisms generating sequence ambiguities during direct sequencing. Figure 1 illustrates the alignment of the nucleotide sequences of these genes with the nucleotide sequences of available homologues, i.e., the entire zinc finger domain of human *ZFX* and *ZFY* (Schneider-Gädicke et al. 1989; Page et al. 1987) and of mouse *Zfx, Zfy1,* and *Zfy2* (*Mus musculus,* Mardon et al. 1990; Ashworth et al. 1989; Mardon and Page 1989) and partial sequences of the *ZFX* and *ZFY* genes of the crab-eating fox *Dusicyon thous* (Lanfear and Holland 1991). The alignment of the 15 sequences was made by eye and is unambiguous, requiring only three gaps: a deletion of 9 bp (three codons) and an insertion of 3 bp (1 codon) at positions 288–296 and 309–311, respectively, in the rat *Zfy* gene, and a deletion of 3 bp (one codon) at position 1170–1172 in the mouse *Zfy1* gene.

Ratio of Substitution Rates

From the alignment in Fig. 1 we have computed the number of substitutions per synonymous site (K_S) and per nonsynonymous site (K_A) for each pair of sequences using the method of Li (1993) (results not shown). From the K_S values we have inferred a phylogenetic tree (Fig. 2a) by using the neighbor-joining method (Saitou and Nei 1987). Since the *ZFY* and *ZFX* genes are both present in all placental mammals studied (Mardon et al. 1990; Lanfear and Holland 1991; Shimmin et al. 1993a), they were probably derived from a gene duplication event prior to the radiation of the placental mammals; and this was supported by the phylogenetic tree inferred from the sequences of the last intron in primate and rodents (Chang et al. 1993). Contrary to this expectation, however, in Fig. 2a the primate *ZFY* genes are clustered with the primate *ZFX* genes rather than with the rodent *ZFY* genes. As will be explained later, the clustering of the primate *ZFY* genes with the *ZFX* genes is probably due to conversion of the *ZFY* gene by the *ZFX* gene prior to the divergence of the four higher primate species studied. Despite the putative gene conversion events, however, Fig. 2a does show the correct relationships among the organisms, e.g., human and orangutan (a great ape) are clustered in one clade, and joined earlier by the baboon (an Old World Monkey) and even earlier by the squirrel monkey (A New World Monkey).

We now consider the ratio of the rates of synonymous substitution in *ZFY* and *ZFX;* we shall denote this ratio

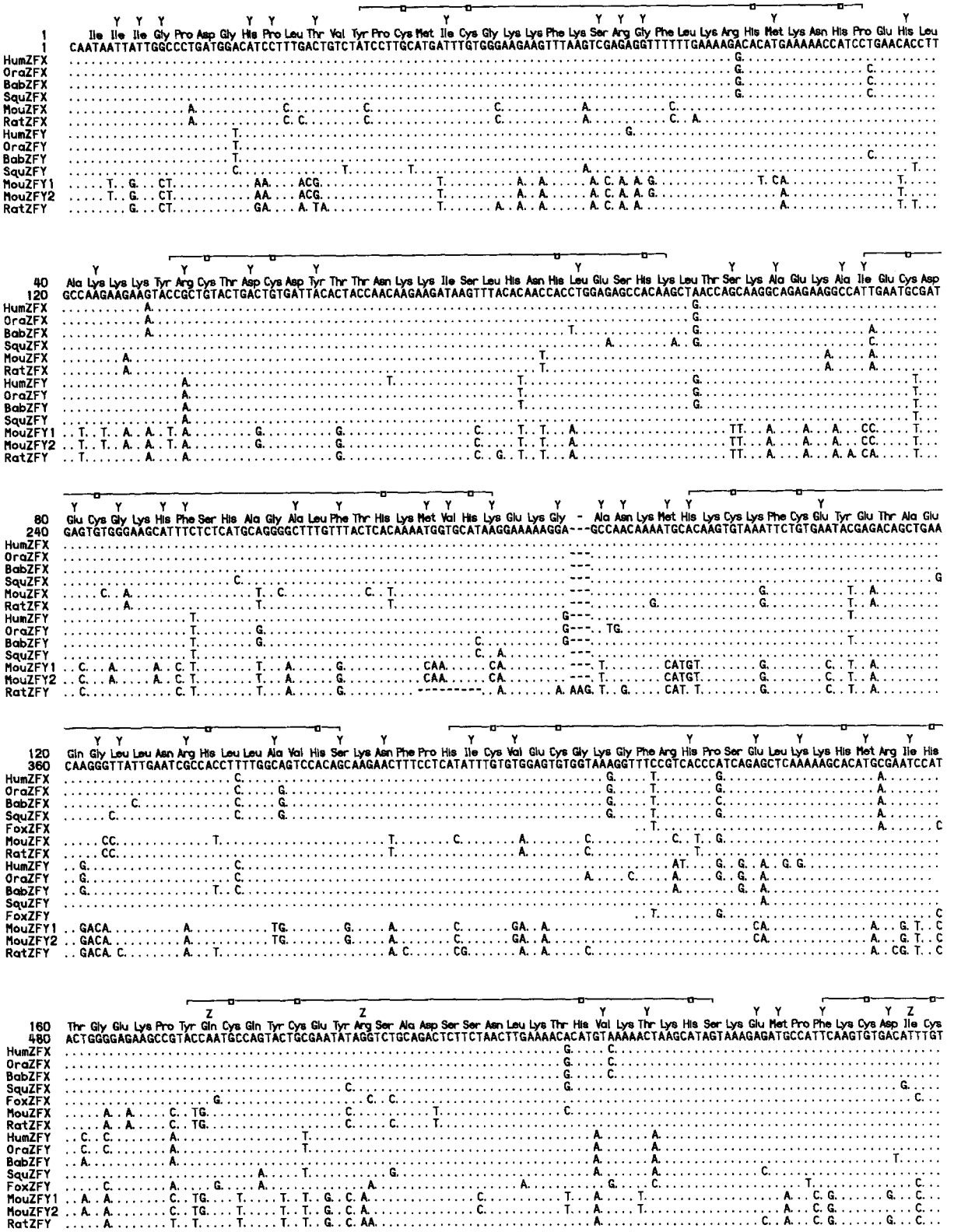

Fig. 1. Nucleotide sequences of the entire X-linked (*ZFX*) and Y-linked (*ZFY*) zinc finger domain exons of primates and rodents (alignment positions 1–1,184) and partial sequences of the zinc finger domain exon of the crab-eating fox (alignment positions 438–749) aligned below the consensus nucleotide sequence. A *dot (.)* indicates identity to the consensus sequence; *dashes (-)* indicate insertions or deletions required for the alignment. Species abbreviations are: *Hum, Homo sapiens; Ora,* the orangutan *Pongo pygmaeus; Bab,* the baboon *Papio cyanocephalus; Squ,* the squirrel monkey *Siamiri boliviensis; Fox,* the crab-eating fox *Dusicyon thous; Mou,* the mouse *Mus musculus; Rat,* the rat *Rattus norvegicus.* The translation of the consensus sequence is provided with residues for which nonsynonymous substitutions occur within the Y-linked genes as "Y" and within both X- and Y-linked genes as "Z." (Note that there are no uniquely X-linked nonsynonymous substitutions.) The 13 C2H2-type zinc finger domains are illustrated by line diagrams, with the *boxes* (□) indicating the positions of the conserved cysteine and histidine residues. The positions of the first residue in each line of the consensus sequence and its translation are provided at left and the total nucleotide length of each gene follows the nucleotide sequence. Accession numbers for the orangutan, baboon, squirrel monkey, and rat, respectively, are: *ZFX:* X75169, X75174, X75175, X75171; *ZFY:* X75176, X75173, X75170, X75172.

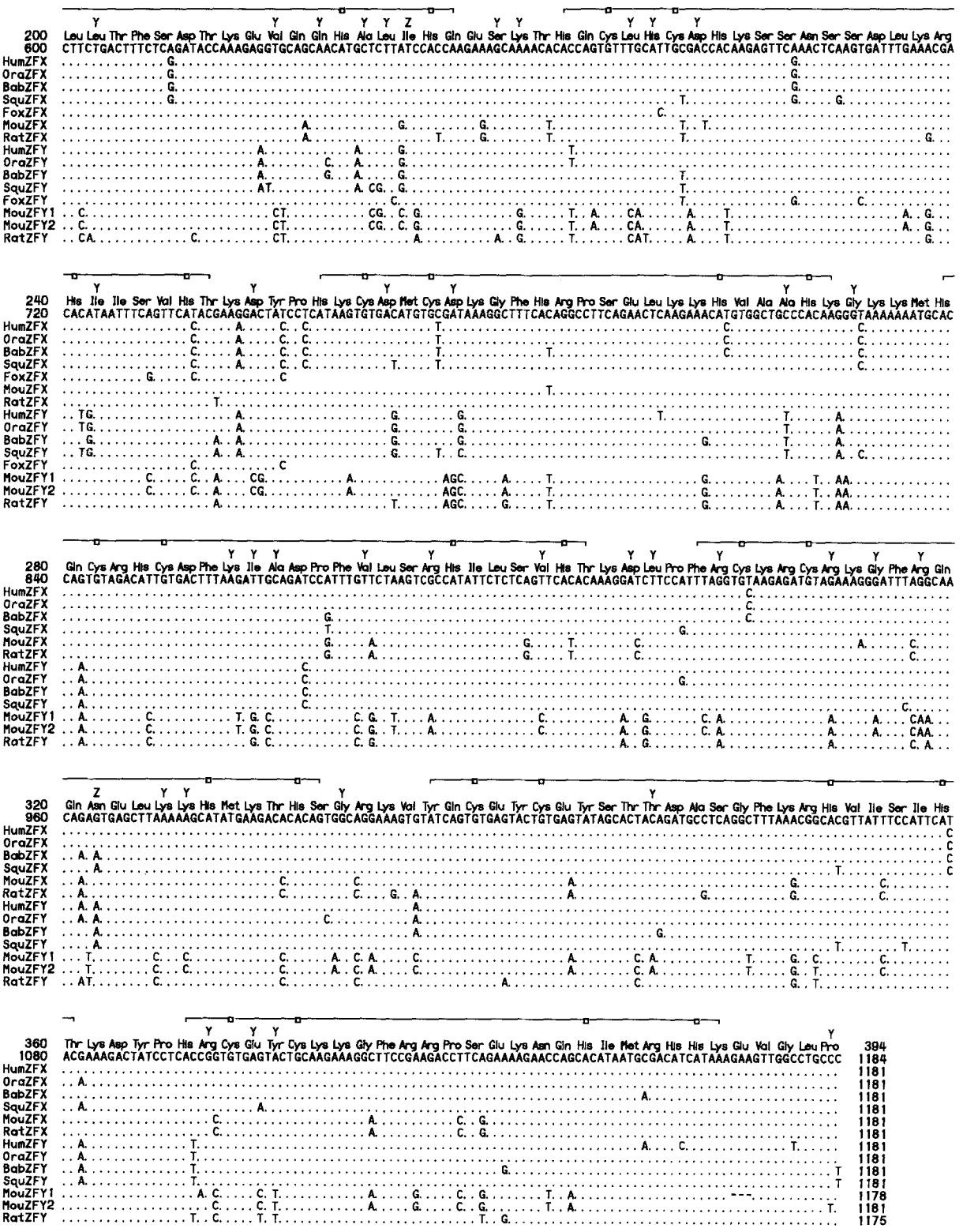

Fig. 1. Continued.

by Y/X. It is clear from Fig. 2a that in each lineage the rate of synonymous substitution is higher in *ZFY* than in *ZFX*. For example, the length of the branch leading to baboon *ZFX* is 0.023 whereas the length of the branch leading to baboon *ZFY* is 0.034, the ratio being 1.5. Because the number of substitutions in each primate lineage is small, estimates of the branch lengths are subject to large statistical errors, as are estimates of the ratio Y/X; for example, the ratio for the human lineage is 0.024/0.004 = 6, which is much higher than that for the baboon lineage. To reduce the effect of statistical fluctuations, we obtain an overall estimate for the four pri-

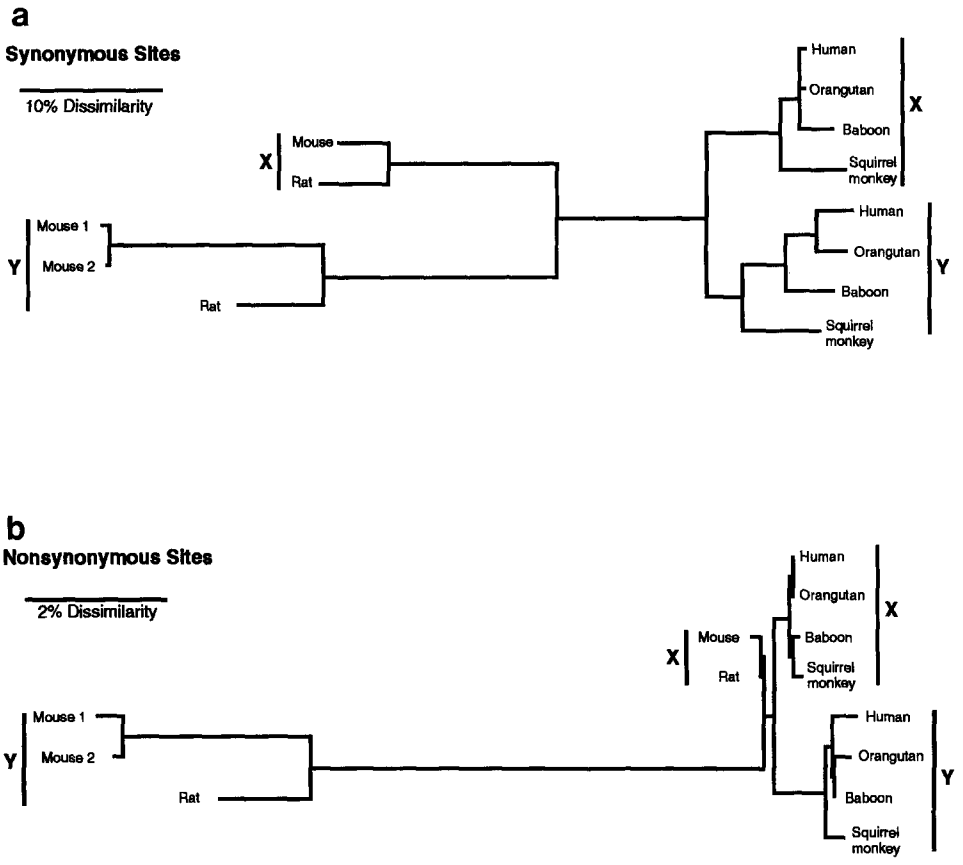

Fig. 2. Phylogenetic trees of the zinc finger domain exon of the *ZFX* and *ZFY* genes in rodents and primates based on synonymous substitution sites (upper tree) and nonsynonymous sites (lower tree). Note that the scale of the nonsynonymous substitution phylogram is expanded by a factor of approximately five compared to the synonymous substitution phylogram.

mate lineages: the total length of the six branches for the four primate species is Y = 0.024 + 0.021 + 0.022 + 0.034 + 0.029 + 0.053 = 0.183 for ZFY and X = 0.004 + 0.003 + 0.001 + 0.023 + 0.013 + 0.045 = 0.089 for *ZFX*, and so Y/X = 0.183/0.089 = 2.1 and the standard error is approximately 0.6 from a formula in Shimmin et al. (1993a). This estimate is not significantly different from the ratio of 2.3 obtained from intron sequences. For the rat lineage we obtain Y/X = 0.061/0.048 = 1.3 ± 0.55, which is close to the value 1.4 obtained from intron sequences. For the mouse lineage, we first obtain an average Y = (0.153 + 0.150)/2 = 0.1515 for mouse *Zfy1* and *Zfy2* and then the ratio Y/X = 0.1515/0.036 = 4.2 ± 1.66, which is much higher than but not statistically different from that (1.4) obtained from intron sequences. When the mouse and rat lineages are considered together, Y/X is 0.2125/0.084 = 2.53 ± 0.80, which is also not statistically different from the estimate from introns (Chang et al. 1993).

We have used also the nonsynonymous substitutions (i.e., the K_A values) to construct a tree (Fig. 2b). Because the K_A values for the primate lineages are small, some distortion of the branching order have occurred; for example, baboon *ZFX* and squirrel monkey *ZFX* are clustered together. (A parsimony analysis also led to some errors in the branching order of the four primates.) Despite the branching errors, the overall features of Fig. 2b are similar to those of Fig. 2a; for example, the primate *ZFY* genes are clustered with the primate *ZFX* genes and as a result of the putative gene conversion events, the branch lengths for the primate *ZFY* gene are very short. However, the differences in branch lengths between the rodent *Zfy* lineages and all the *ZFX* lineages are much more conspicuous than those in Fig. 2a. Since there are errors in the branching order for the four primates in Fig. 2b, the branch lengths estimated are not reliable. Using the tree in Fig. 2a and a parsimony analysis (Fitch 1977) we found only one nonsynonymous substitution in mouse *Zfx* and no nonsynonymous substitution in rat *Zfx* since their separation but 26.5 and 11.5 nonsynonymous substitutions in mouse *Zfy* (average for mouse *Zfy1* and *Zfy2*) and rat *Zfy*, respectively, since their separation. These estimates indicate a very high Y/X ratio in both the mouse and rat lineages, though the exact ratio cannot be obtained because of the small number (i.e., one of nonsynonymous substitutions between mose and rat *Zfx*. In contrast, there are only two nonsynonymous substitutions among the four primate *ZFX* genes and seven nonsynonymous substitutions among the four primate *ZFY* genes, giving a Y/X ratio of only 3.5. Obviously, *ZFX*

Table 2. Mean number of synonymous (plain text) and nonsynonymous (italicized) nucleotide substitutions per nucleotide site within the three regions 1–437 (upper), 438–749 (middle), and 750–1,184 (lower) of the zinc finger domain exon of eutherian ZFX and ZFY

1–437	HumX[a]	OraX	BabX	SquX	MouX	RatX	HumY	OraY	BabY	SquY	MouY1	MouY2	RatY
HumX	0	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.004*	*0.007*	*0.004*	*0.004*	*0.142*	*0.139*	*0.125*
OraX	0.014	0	*0.000*	*0.000*	*0.000*	*0.000*	*0.004*	*0.007*	*0.004*	*0.004*	*0.142*	*0.139*	*0.125*
BabX	0.035	0.021	0	*0.000*	*0.000*	*0.000*	*0.004*	*0.007*	*0.004*	*0.004*	*0.142*	*0.139*	*0.125*
SquX	0.064	0.050	0.066	0	*0.000*	*0.000*	*0.004*	*0.007*	*0.004*	*0.004*	*0.142*	*0.139*	*0.126*
MouX	0.358	0.383	0.393	0.403	0	*0.000*	*0.004*	*0.007*	*0.004*	*0.004*	*0.142*	*0.139*	*0.126*
RatX	0.337	0.360	0.371	0.380	0.091	0	*0.004*	*0.007*	*0.004*	*0.004*	*0.142*	*0.139*	*0.126*
HumY	0.107	0.122	0.147	0.165	0.428	0.406	0	*0.031*	*0.016*	*0.030*	*0.083*	*0.080*	*0.085*
OraY	0.134	0.153	0.177	0.193	0.468	0.444	0.069	0	*0.031*	*0.042*	*0.100*	*0.095*	*0.095*
BabY	0.114	0.114	0.139	0.156	0.409	0.407	0.035	0.062	0	*0.033*	*0.086*	*0.082*	*0.086*
SquY	0.114	0.131	0.156	0.173	0.393	0.373	0.105	0.139	0.116	0	*0.083*	*0.081*	*0.082*
MouY1	0.387	0.387	0.400	0.460	0.497	0.485	0.386	0.453	0.393	0.393	0	*0.007*	*0.049*
MouY2	0.366	0.366	0.379	0.436	0.476	0.465	0.366	0.429	0.372	0.373	0.007	0	*0.049*
RatY	0.408	0.444	0.441	0.527	0.460	0.481	0.407	0.451	0.391	0.389	0.141	0.133	0

438–749	HumX[a]	OraX	BabX	SquX	FoxX	MouX	RatX	HumY	OraY	BabY	SquY	FoxY	MouY1	MouY2	RatY
HumX	0	*0.000*	*0.000*	*0.005*	*0.004*	*0.009*	*0.004*	*0.031*	*0.024*	*0.020*	*0.031*	*0.005*	*0.079*	*0.073*	*0.071*
OraX	0.000	0	*0.000*	*0.005*	*0.004*	*0.009*	*0.004*	*0.031*	*0.024*	*0.020*	*0.031*	*0.005*	*0.079*	*0.073*	*0.071*
BabX	0.034	0.034	0	*0.005*	*0.004*	*0.009*	*0.004*	*0.031*	*0.024*	*0.020*	*0.031*	*0.005*	*0.079*	*0.073*	*0.071*
SquX	0.062	0.062	0.099	0	*0.009*	*0.014*	*0.009*	*0.036*	*0.029*	*0.025*	*0.036*	*0.010*	*0.085*	*0.079*	*0.077*
FoxX	0.158	0.158	0.182	0.162	0	*0.012*	*0.007*	*0.034*	*0.027*	*0.024*	*0.035*	*0.009*	*0.083*	*0.077*	*0.075*
MouX	0.357	0.357	0.403	0.299	0.380	0	*0.005*	*0.029*	*0.022*	*0.019*	*0.030*	*0.014*	*0.074*	*0.068*	*0.081*
RatX	0.408	0.408	0.439	0.350	0.357	0.129	0	*0.034*	*0.027*	*0.024*	*0.035*	*0.009*	*0.075*	*0.069*	*0.075*
HumY	0.338	0.338	0.388	0.363	0.390	0.478	0.530	0	*0.014*	*0.010*	*0.021*	*0.026*	*0.096*	*0.090*	*0.095*
OraY	0.296	0.296	0.341	0.322	0.351	0.432	0.481	0.032	0	*0.004*	*0.015*	*0.030*	*0.088*	*0.082*	*0.087*
BabY	0.315	0.315	0.340	0.317	0.331	0.429	0.433	0.183	0.137	0	*0.011*	*0.026*	*0.084*	*0.078*	*0.083*
SquY	0.269	0.269	0.294	0.270	0.284	0.419	0.393	0.255	0.212	0.188	0	*0.037*	*0.090*	*0.084*	*0.089*
FoxY	0.278	0.278	0.323	0.272	0.247	0.479	0.519	0.429	0.391	0.466	0.428	0	*0.086*	*0.080*	*0.078*
MouY1	0.733	0.733	0.757	0.756	0.625	0.653	0.661	0.847	0.774	0.689	0.674	0.933	0	*0.005*	*0.042*
MouY2	0.743	0.743	0.767	0.769	0.632	0.658	0.668	0.862	0.786	0.694	0.681	0.950	0.000	0	*0.047*
RatY	0.557	0.557	0.581	0.561	0.458	0.531	0.480	0.625	0.576	0.609	0.518	0.646	0.283	0.286	0

750–1184	HumX[a]	OraX	BabX	SquX	MouX	RatX	HumY	OraY	BabY	SquY	MouY1	MouY2	RatY
HumX	0	*0.000*	*0.004*	*0.004*	*0.000*	*0.000*	*0.013*	*0.013*	*0.013*	*0.010*	*0.079*	*0.080*	*0.047*
OraX	0.007	0	*0.004*	*0.004*	*0.000*	*0.000*	*0.013*	*0.013*	*0.013*	*0.010*	*0.079*	*0.080*	*0.047*
BabX	0.027	0.034	0	*0.000*	*0.004*	*0.004*	*0.009*	*0.009*	*0.009*	*0.006*	*0.084*	*0.084*	*0.051*
SquX	0.071	0.064	0.093	0	*0.004*	*0.004*	*0.009*	*0.009*	*0.009*	*0.006*	*0.084*	*0.084*	*0.051*
MouX	0.251	0.261	0.234	0.299	0	*0.000*	*0.013*	*0.013*	*0.013*	*0.010*	*0.079*	*0.079*	*0.047*
RatX	0.281	0.293	0.281	0.332	0.049	0	*0.013*	*0.013*	*0.013*	*0.010*	*0.079*	*0.079*	*0.047*
HumY	0.127	0.118	0.128	0.151	0.285	0.294	0	*0.000*	*0.000*	*0.003*	*0.093*	*0.094*	*0.061*
OraY	0.110	0.102	0.127	0.117	0.264	0.272	0.041	0	*0.000*	*0.003*	*0.093*	*0.094*	*0.061*
BabY	0.118	0.110	0.153	0.142	0.293	0.301	0.063	0.048	0	*0.003*	*0.093*	*0.094*	*0.061*
SquY	0.101	0.093	0.134	0.109	0.331	0.343	0.087	0.071	0.063	0	*0.090*	*0.091*	*0.058*
MouY1	0.478	0.492	0.508	0.548	0.283	0.323	0.479	0.450	0.441	0.487	0	*0.007*	*0.030*
MouY2	0.487	0.501	0.516	0.557	0.282	0.321	0.488	0.459	0.451	0.496	0.007	0	*0.030*
RatY	0.343	0.352	0.344	0.373	0.263	0.310	0.299	0.278	0.271	0.314	0.245	0.233	0

[a] Species abbreviations are as for Fig. 1.

has been well conserved in both the primate and the rodent lineages but *ZFY* has evolved rapidly in the rat lineage and even faster in the mouse lineage.

Gene Conversion Events

The human *ZFX* and *ZFY* genes show anomalously high similarity in the last exon. This has been suggested to be due to gene conversion (Hayashida et al. 1992; Pamilo and Bianchi 1993). A detailed analysis by dividing the last exon into three regions has provided strong evidence for at least two gene conversion events, one in the first and the other in the third region (Shimmin et al. 1993b). To see whether the two putative gene conversion events occurred before the divergence of the four primate species studied, we have again divided the last exon into three regions: from nucleotides 1 to 437, 438 to 749, and 750 to 1184 (Table 2). We note that in the middle region both the K_S values and the K_A values are quite uniform between primate *ZFX* and *ZFY,* averaging 0.32 and 0.027, respectively. In contrast, in the first and third regions the K_S values between primate *ZFX* and *ZFY* av-

Table 3. Number of substitutions with respect to putative function

Functional category[a]	*ZFX* type[b]	Rodent *Zfy*	Total substitutions	Number of sites[c]	Substitutions per site
1) Zn^{2+} coordinating residues	0	0	0	52	0
2) Invariant/conserved residues	1	7	8	39	0.21
3) DNA binding residues	2	9	11	52	0.21
4) Base recognition residues	6	11	17	52	0.33
5) Noncritical residues within zinc finger	9	40	49	122	0.40
6) Residues linking/flanking zinc fingers	2	26	28	76	0.37
Total	20	93	113	393	0.29

[a] Consensus zinc finger motif (Lee et al. 1989) has 24 residues: InZnnZnnnInBDBnIBDZBDnZD divided into five categories with suggested functionality: Z, Zn binding; I, invariant; D, DNA binding; B, base recognition; and n, nonfunctional. The sixth category includes all residues linking and flanking the 13 zinc finger consensus motifs composing the zinc finger domain exon

[b] Includes primate *ZFX* and *ZFY* and rodent *Zfx*

[c] Total number of sites in the zinc finger domain exon with the putative function—e.g., there are four residues coordinating the Zn^{2+} per zinc finger and the exon is comprised of 13 zinc fingers, which therefore totals 52 sites

erage 0.14 and 0.12, respectively, and the K_A values average 0.005 and 0.010, respectively, which are considerably lower than the corresponding values for the middle region. Therefore, it appears that both putative conversion events had occurred before the divergence among the four primate species.

Table 2 also includes the partial sequences (the middle region) for the last exon of fox *ZFX* and *ZFY* (Lanfear and Holland 1991). We note that the K_A values between fox *ZFY* and primate *ZFX*s average 0.006 (cf. primate *ZFX* vs *ZFY* average 0.027 above) and that the K_S value between fox *ZFY* and *ZFX* is only 0.25, which is lower than the values between fox *ZFY* and primate *ZFX*s (average 0.29) or fox *ZFX* and primate *ZFY* (average 0.34). A simple explanation for these observations is that the middle part of the last exon of *ZFY* was converted by *ZFX* within the carnivore lineage. A similar suggestion was made earlier by Hayashida et al. (1992). There exists no obvious boundary within the available partial sequences and thus it is impossible to delimit the boundaries of the conversion event.

Discussion

We have seen that *ZFX* has been highly conserved in both primates and rodents. This fact suggests that the structural requirement for zinc finger domains is stringent. In contrast, *ZFY* has evolved rather rapidly in rodents, though it too has evolved very slowly in primates. In an effort to understand the effects of functional constraints on the evolution of *ZFX* and *ZFY*, we have categorized each position of the zinc finger domain according to putative function (after Lee et al. 1989) and compared the amino acid substitution rates for the six categories (Table 3): (1) the cysteine/histidine Zn^{2+} coordinating residues (2) other invariant/highly conserved residues, (3) putative DNA binding residues (4) putative DNA site recognition residues (5) all other residues within the zinc finger motif, and (6) residues flanking or linking the 13 zinc finger motifs. This analysis yielded four observations. First, strict conservation of the cysteine/histidine Zn^{2+} coordinating residues is observed within the 176 zinc finger motifs encoded by the 15 exons (Table 3), with the only exception being the 9-bp deletion in the rat *Zfy* which deletes the second histidine residue of the third zinc finger motif (Fig. 1). Second, base recognition residues (category 4) have not been well conserved in either *ZFX* or *ZFY* and have evolved only somewhat slower than noncritical residues (categories 5 and 6). Indeed, if we put categories 4, 5, and 6 in one class and categories 2 and 3 in another class, the former class has evolved significantly faster than the latter. Third, there is no evidence that the pattern of amino acid substitutions within the *ZFX* and primate *ZFY* is qualitatively different from that in rodent *Zfy*, although with only 20 amino acid changes within the former groups this result is not definitive. Fourth, it has been previously noted (Palmer et al. 1990) that of the 11 amino acid substitutions between the human *ZFX* and *ZFY* within the zinc finger domains, five of these are residues that Lee et al. (1989) have suggested are involved in DNA binding and site recognition. Based on this high rate of change within putatively critical amino acid residues Palmer et al. (1990) suggested that human *ZFX* and *ZFY* have diverged functionally. However, the high substitution rate is not reflected in the other primate *ZFY* exons, with two of the five putative critical site substitutions (His_{148} to Tyr_{148} and Lys_{153} to Arg_{153}) being unique to the human *ZFY* gene product. Assessment of the functional significance of these two unique changes awaits elucidation of the function of the zinc finger exon domain.

The quantitatively different patterns or evolution with respect to the substitution rate of *ZFY* in primates and rodents have been suggested to be due to their different patterns of expression (Lanfear and Holland 1991). Human *ZFY* and *ZFX* are expressed ubiquitously in adult tissues and *ZFX* does not seem to be subject to X-inactivation (Palmer et al. 1990; Schneider-Gädicke et al. 1989). This is probably also the case in other higher

primates. Thus, in higher primates *ZFY* is probably subject to a similar degree of functional constraint as is *ZFX* and so, like *ZFX*, it has evolved slowly. On the other hand, mouse *Zfy1* and *Zfy2* genes are expressed only in fetal and adult testes, respectively (Palmer et al. 1990; Koopman et al. 1989), and mousse *Zfx* may have adopted X-inactivation (Adler et al. 1991), so equal gene dosage between the two sexes is fulfilled by *Zfx* alone. This would allow mouse *Zfy1* and *Zfy2* to change rapidly if they play no fundamental role in testes or if they have been in the process of adapting to a new role. Our new sequence data suggest that rat *Zfy* has also evolved rapidly but at a rate less than half that in mouse *Zfy1* and *Zfy2;* it will be interesting to learn the expression pattern of rat *Zfy*. Figure 2b suggests that the ancestral rodent *Zfy* had already been evolving at a fast rate before the divergence of the mouse and rat lineages. According to our parsimony analysis (Fitch 1977), the ancestral rodent *Zfy* gained 52 amino acid substitutions from the point of separation between rodent *Zfy* and *Zfx* to the point of divergence between the mouse and rat lineages, compared to 11.5 substitutions in the rat lineage and 26.5 substitutions in the mouse lineage since the mouse–rat split.

In summary, the X- and Y-linked zinc finger genes appear functionally conserved, but the regulatory region of the rodent *Zfy* may have undergone a functional change, allowing developmental regulation which may, in turn, have permitted an accelerated rate of amino acid substitution. At present, none of the regulatory regions of any of these zinc finger genes has been identified.

In primate *ZFY* and *ZFX* zinc finger domain exons the Y/X ratio (2.0) estimated from synonymous substitutions is very close to that (2.3) estimated from introns. Under the assumption of selective neutrality we can use Miyata et al.'s (1987) theory to obtain an estimate of $\alpha_m = 4$ for the male-to-female ratio of mutation rate. This is between Ketterling et al.'s (1993) estimate of $\alpha_m = 3.5$ from hemophilia B families and our estimate of $\alpha_m = 6$ from the last intron sequences of *ZFY* and *ZFX*. These estimates strongly suggest that the α_m value is much smaller than the infinitely large value obtained by Miyata et al. (1987).

The relatively small value of $\alpha_m = 4$ suggests that the effect of selective constraints on Y/X is weak or negligible in primates. If the effect is significant, then even under the same rate of mutation in males and females the synonymous rate in Y-linked genes is likely to be faster than that in X-linked genes because the effective population size for a Y-linked gene is smaller, only one-third of that for an X-linked gene (Charlesworth et al. 1987). Note, however, that a weak effect may not necessarily imply that synonymous changes are free of selective constraints because it can be partly owing to the same degree of functional constraints in primate *ZFY* and *ZFX* and partly owing to a small effective size in higher primates.

The Y/X ratio in the rat lineage (1.3) is very close to that (1.4) estimated from introns whereas that in the mouse lineage (4.2) is much higher, though the difference is not significant because the estimate (4.2) has a large standard error (i.e., 1.66). Since it is even considerably beyond the possible maximum theoretical value (i.e., 3.0), it will be interesting to see if further data support the high ratio. If so, then this would imply a relaxation in selective constraints on synonymous changes in mouse *ZFY* genes, as in the case of nonsynonymous changes.

There is some evidence that synonymous changes in mouse and rat *ZFX* genes are subject to selective constraints. The synonymous divergence between these two genes is 0.083 ± 0.022 whereas the average synonymous divergence estimated from 14 pairs of mousse and rat autosomal genes is 0.144 ± 0.008 (Li 1993). Therefore, the ratio of synonymous rates in X-linked and autosomal genes in X/A = 0.58. This is close to the estimate (0.60) by Miyata et al. (1987) and Wolfe and Sharp (1993). Under the assumption of no selective constraint this implies $\alpha_m = \infty$, because the ratio X/A is lower than the theoretical minimum 0.67. The extremely large α_m value implies that the synonymous rate in rodent *ZFX* genes is lower than expected under selective neutrality; the α_m value estimated from introns is only two (Chang et al. 1993). Thus, it seems that synonymous changes in rodent *ZFX* genes are subject to stronger selective constraints than are autosomal genes. (For the theoretical basis of this, see Charlesworth et al. 1987; Charlesworth 1994.)

The evolution of *ZFX* and *ZFY* in mammals may provide some insight into the process of loss of functional genes of the Y chromosome. There are two different views of this issue. Muller's (1914) view is that loss of function for a Y-linked gene occurs by accumulation of recessive lethal or deleterious mutations—such mutations are effectively neutral because of the sheltering effect of the functional gene on the X chromosome. Charlesworth (1978) noted some difficulties with Muller's view. First, as noted by Fisher (1935), the probability of fixation of a Y-linked deleterious mutant is negligibly small in a large population because selection against a Y mutant occurs whenever the individual carrying it also contains a deleterious, allelic mutant gene on the X chromosome. Although Nei (1970) showed that Fisher's conclusion may not be true for a small population, the effective population size of a species may not be as small as Nei assumed (i.e., the order of 10,000). Second, the assumption of recessivity of deleterious mutations is not supported by data from *Drosophila.* Third, under Muller's hypothesis it is difficult to explain the phenomenon of dosage compensation. Charlesworth (1978, 1991) therefore proposed the following alternative hypothesis. In the absence of recombination the number of mutant loci on a Y chromosome can only increase because it cannot get rid of a mutant allele by recombination with a chromosome having the wild-type

allele; this is known as Muller's (1964) ratchet. If those Y chromosomes without mutant loci become lost by random drift, then every Y chromosome in the population will carry at least one deleterious mutation. This provides a favorable condition for the development of dosage compensation. Once dosage compensation develops, loss of functional genes on the Y chromosome is accelerated. As noted above, the *ZFX-ZFY* pair was apparently created before the radiation of placental mammals, yet they are both functional in all species examined. The long persistence of *ZFY* might have occurred as follows. The stringent structural requirement of zinc finger domains has retarded the accumulation of amino acid substitutions in *ZFY*. Note that in the absence of amino acid substitutions in *ZFY* there would be no advantage for the development of dosage compensation of this locus and this would in turn not provide favorable conditions for amino acid substitutions in *ZFY*. Moreover, conversion of *ZFY* by *ZFX* would retard the evolution of *ZFY* and thus also the development of dosage compensation. This might explain the slow evolution of primate *ZFY* and suggests that under Muller's view loss of Y-linked genes may occur at a very slow rate. On the other hand, in mice and rats *Zfy* has evolved relatively rapidly, possibly as a consequence of the development of dosage compensation. This is in line with Charlesworth's view. Of course, it is not clear how the development of dosage compensation had occurred in the first place. There are two possible scenarios. First, fixation of one or more deleterious mutations in the Y chromosome might have occurred by chance in an ancestral species, and this then provided a favorable condition for the development of dosage compensation. Note that chance fixation could occur if a population bottleneck occurred, so Nei's (1970) condition for fixation was met. Second, it might have been a regional or whole chromosome phenomenon rather than due to fixation of a deleterious mutation. Further data from other rodent species might indicate which scenario is more plausible.

Acknowledgments. We thank Drs. B. Charlesworth, D. Hewett-Emmett, and A. Zharkikh for discussions and help. This study was supported by NIH grants.

References

Adler DA, Bressler SL, Chapman VM, Page DC, Disteche CM (1991) Inactivation of the Zfx gene on the mouse X chromosome. Proc Natl Acad Sci U S A 88:4592–4595

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

Chang BH-J, Shimmin LC, Shyue S-K, Hewett-Emmett D, Li W-H (1993) Weak male-driven evolution in rodents. Proc Natl Acad Sci U S A 91:827–831

Charlesworth B (1978) Model for evolution of Y chromosomes and dosage compensation. Proc Natl Acad Sci U S A 75:5618–5622

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

Charlesworth B (1993) More mutations in males? Curr Biol 3:466–467

Charlesworth B (1994) The effect of selection against deleterious mutations on weakly-selected, linked variants. Genet Res (in press)

Charlesworth B, Coyne JA, Barton NH (1987) The relative rates of evolution of sex chromosomes and autosomes. Am Nat 130:113–146

Ellsworth DL, Rittenhouse KD, Honeycutt RL (1993) Artifactual variation in randomly amplified polymorphic DNA banding patterns. Biotechniques 14:214–217

Fisher RA (1935) The sheltering of lethals. Am Nat 69:446–455

Fitch WM (1977) Problem of discovering the most parsimonious tree. Am Nat 111:223–257

Haldane JBS (1947) The mutation rate of the gene for hemophilia, and its segregation ratios in males and females. Ann Eugen 13:262–271

Hayashida H, Kuma K, Miyata T (1992) Interchromosomal gene conversion as a possible mechanism for explaining divergence patterns of ZFY-related genes. J Mol Evol 35:181–183

Ketterling RP, Vielhaber E, Bottema CDK, Schaid DJ, Cohen MP, Sexauer CL, Sommer SS (1993) Germ-line origins of mutation in families with hemophilia B: the sex ratio varies with the type of mutation. Am J Hum Genet 52:152–166

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

Lanfear J, Holland PWH (1991) The molecular evolution of ZFY-related genes in birds and mammals. J Mol Evol 32:310–315

Lee MS, Gippert GP, Soman KS, Case DA, Wright PE (1989) Three-dimensional solution structure of a single zinc finger DNA-binding domain. Science 245:635–637

Li W-H (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J Mol Evol 36:96–99

Mardon G, Page DC (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

Mardon G, Luoh S-W, Simpson EM, Gill G, Brown LG, Page DC (1990) Mouse Zfx is similar to Zfy-2: each contains an acidic activating domain and 13 zinc fingers. Mol Cell Biol 10:681–688

Miyata T, Hayashida H, Kuma K, Mitsuyasa K, Yasunaga T (1987). Male-driven molecular evolution. Cold Spring Harbor Symp Quant Biol 52:863–867

Muller HJ (1914) A gene for the fourth chromosome of Drosophila. J Exp Zool 17:325–336

Muller HJ (1964) The relation of recombination to mutational advance. Mutat Res 1:2–9

Nei M (1970) Accumulation of nonfunctional genes on sheltered chromosomes. Am Natur 104:311–322

Page DC, Mosher R, Simpson EM, Fisher EMC, Mardon G, Pollack J, McGillivray B, de la Chapelle A, Brown LG (1987) The sex-determining region of the human Y chromosome encodes a finger protein. Cell 51:1091–11104

Palmer MS, Berta P, Sinclair AH, Pym B, Goodfellow PN (1990) Comparison of human ZFY and ZFX transcripts. Proc Natl Acad Sci U S A 87:1681–1685

Pamilo P, Bianchi NO (1993) Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. Mol Biol Evol 10:271–281

Saitou N, Nei M (1987) The neighbour joining method: a new method for constructing phylogenetic trees. Mol Biol Evol 4:406–425

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci U S A 74:5463–5467

Schneider-Gadicke A, Beer-Romero P, Brown LG, Mardon G, Luoh S-W, Page DC (1989) Putative transcription activator with alternative isoforms encoded by human ZFX gene. Nature 342:708–711

Shimmin LC, Chang BH-J, Li W-H (1993a) Male-driven evolution of DNA sequences. Nature 362:745–747

Shimmin LC, Chang BH-J, Hewett-Emmett D, Li W-H (1993b) Potential problems in estimating the male-to-female mutation rate ratio from DNA sequence data. J Mol Evol 37:160–166

Wolfe KH, Sharp PM (1993) Mammalian gene evolution: nucleotide sequence divergence between mouse and rat. J Mol Evol 37:441–446