Evolution of the *Zfx* and *Zfy* Genes: Rates and Interdependence between the Genes[1]

Pekka Pamilo and Nestor O. Bianchi†*

*Department of Genetics, Uppsala University; and †IMBICE, La Plata

A phylogenetic analysis of sex-chromosomal zinc-finger genes (*Zfx* and *Zfy*) indicates that the genes have not evolved completely independently since their initial separation. The sequence similarities suggest gene conversion in the last exon between the duplicated Y-chromosomal genes *Zfy-1* and *Zfy-2* in the mouse. There are also indications of conversion (or recombination) between the X- and Y-chromosomal genes in the crab-eating fox and in the mouse. The method for estimating synonymous and nonsynonymous substitutions is modified by incorporating the substitutions in the twofold-degenerate sites in a novel way. The estimates of synonymous substitutions support the generation-time hypothesis in that the obtained rates are higher in mice (by a factor of 4.7) than in humans and higher in the Y-chromosomal genes (by a factor of 1.9) than in the X-chromosomal genes.

Introduction

The eutherian X chromosomes include a zinc-finger gene (*ZFX* in humans and *Zfx* in other species), and a homologous gene (*ZFY* or *Zfy*) is found in the Y chromosomes. The genes have been detected in the sex chromosomes of all eutherian species studied (Page et al. 1987). Hybridization studies have identified a homologous gene in marsupials (Sinclair et al. 1988), birds (Page et al. 1987), and reptiles (Bull et al. 1988), but in none of them does it appear to be sex linked. It is therefore reasonable to assume that the *Zfx* and *Zfy* genes have become separated from each other after the split of the lineages leading to present marsupials and eutherian taxa and before the radiation of the eutherian species.

The eutherian X and Y chromosomes share a short region, a pseudoautosomal region, where recombination can take place (Ellis and Goodfellow 1989). The chromosomal location of the *Zfx* locus is different in humans and mice (Nagamine et al. 1989), but in both taxa the gene is located outside the pseudoautosomal region. Therefore, the X- and Y-chromosomal genes have been expected to evolve independently without recombination (Page et al. 1987).

Most eutherian species have the zinc-finger gene as a single copy in both the X chromosome and the Y chromosome. The only known exceptions come from rodents. Several species of mouse (genus *Mus*) have two Y-chromosomal genes (*Zfy-1* and *Zfy-2*) and an autosomal gene (*Zfa*) that appears to be an expressed retrogene (Mardon et al. 1989; Nagamine et al. 1989; Ashworth et al. 1990). A still stronger amplification of the *Zfy* genes has been detected in wood lemming (Lau et al. 1992) and in South American sigmodontine rodents (Bianchi et al. 1989).

The proteins produced by *Zfx* and *Zfy* have a conservative general structure

1. Key words: *Zfx, Zfy,* sex chromosome, gene conversion, nucleotide substitutions, evolutionary rate.

Address for correspondence and reprints: Pekka Pamilo, Department of Genetics, Uppsala University, Box 7003, 75007 Uppsala, Sweden.

Mol. Biol. Evol. 10(2):271–281. 1993.
© 1993 by The University of Chicago. All rights reserved.
0737-4038/93/1002-0001$02.00

Downloaded from https://academic.oup.com/mbe/article/10/2/271/1024... by guest on 16 August 2022

(Koopman et al. 1991). They consist of an acidic domain, a domain consisting of 13 zinc fingers, and a short basic chain connecting these two domains (Mardon and Page 1989). The zinc-finger domain consists of 393 amino acids and is coded by the last exon at the 3′ half of the gene. The acidic and basic chains result from several shorter exons (Schneider-Gädicke et al. 1989*a*) with a slight length variation among the genes (349 codons in *Zfa* and 390–412 codons in the others).

The existence of the zinc-finger genes in different chromosomes, and in different dosages within a chromosome (in some rodent Y chromosomes), allows sequence comparisons aiming to clarify the pace of evolution in the sex chromosomes. According to the neutral theory of molecular evolution, the long-term rate of nucleotide substitution is determined by the rate of selectively neutral mutations (Kimura 1983, pp. 46–47; Nei 1987, pp. 380–381). Calibrations of the molecular clock have generally indicated that the evolutionary distance between two lineages depends approximately linearly on the time of separation in years. It has, however, been proposed that the mutation rate should depend on the number of cell divisions rather than on the absolute time, because new mutations arise during DNA replication (Laird et al. 1969). Some recent analyses do suggest that the evolutionary rate in mammals could be faster in lineages with shorter generation times (Wu and Li 1985; Li et al. 1990). The number of cell generations also differs between oogenesis and spermatogenesis, there being more divisions in the latter. This predicts differences in the neutral evolutionary rates of autosomes and the sex chromosomes (Miyata et al. 1987), provided that the mutation rate correlates with the number of cell divisions. We can expect the highest rate in the Y chromosome, intermediate rates in autosomal genes, and the lowest rate in the X chromosome, because all Y chromosomes, half of the autosomes, and one-third of the X chromosomes are transmitted to the next generation in sperm. The sequence data analyzed by Miyata et al. (1987) support this prediction, the rates in the Y chromosome, the autosomes, and the X chromosome having a relationship of 1:0.5:0.33, the expected relationship if the mutations in spermatogenesis are vastly more frequent than those in oogenesis.

The aim of the present study is to use the sequence information from the zinc-finger genes to test the above predictions on the evolutionary rates. The data also allow examination of the evolutionary independence of homologous genes located either in different chromosomes (*Zfx, Zfy,* and *Zfa*) or as duplicated loci in the same chromosome (*Zfy-1* and *Zfy-2*). Lanfear and Holland (1991) concluded on the basis of the relative differences between human and murine Y- and X-chromosomal genes that the rate of synonymous substitutions in the Y chromosome has been twice that in the X chromosome. However, if the evolutionary changes in the two chromosomes are not independent of each other, this comparison does not reveal the true pattern.

Material and Methods

The sequences analyzed in this study are taken from the literature. The human and murine genes have been completely sequenced, and the sequences are given by Palmer et al. (1990) (human genes *ZFX* and *ZFY*), Mardon et al. (1990) [mouse (*Mus musculus*) gene *Zfx* (strain FVB/N)], Ashworth et al. (1989) (mouse gene *Zfy-1*), Mardon and Page (1989) [mouse gene *Zfy-2* (strain FVB/N)], and Ashworth et al. (1990) [mouse gene *Zfa* (strain 129/Sv)].

Lanfear and Holland (1991) amplified and sequenced a 360-bp-long sequence of *Zfx* from Chinese hamster, *Zfx* and *Zfy* from a crab-eating fox (*Dusicyon thous*), and related *Zf* genes from two bird species, the great tit (*Parus major*) and the lesser

Downloaded from https://academic.oup.com/mbe/article/10/2/271/1114325 by guest on 06 August 2022

black-backed gull (*Larus fuscus*). This sequence comes from the middle of the exon coding for the zinc fingers (the region of the seventh finger).

The phylogenies of the sequences are reconstructed by the parsimony method and the neighbor-joining method (Saitou and Nei 1987). For DNA parsimony we use the PHYLIP program that considers all nucleotide substitutions equally likely (Felsenstein 1986). The reliability of the obtained clusters is examined by bootstrapping over the nucleotides (Felsenstein 1985). The neighbor-joining trees are created by the program of Jin and Ferguson (1990).

The substitution differences between pairs of sequences are estimated with the two-parameter model (Kimura 1980), separately for nondegenerate, twofold-degenerate, and fourfold-degenerate sites (Li et al. 1985). This procedure results in separate estimates of transitional (A_0, A_2, and A_4) and transversional (B_0, B_2, and B_4) substitutions in each of the three categories. The total number of nucleotide substitutions per site can be given as the sum of transitions and transversions, $K_i = A_i + B_i$. As all the substitutions in the fourfold-degenerate sites are silent, K_4 could be used as an estimate for nucleotide substitution per synonymous site, and K_0 could be used as an estimate for nucleotide substitution per nonsynonymous site.

Li et al. (1985) proposed that the information from twofold-degenerate sites can also be used by dividing these sites into synonymous (one-third) and nonsynonymous (two-thirds), all transitions being practically synonymous and all transversions being nonsynonymous. When transitions are commoner than transversions, this method leads to estimates of synonymous differences, K_s, that can be much higher than K_4. We suggest two possibilities for correcting this bias. First, the twofold-degenerate sites can be divided into synonymous and nonsynonymous, according to the ratio of transitional and transversional changes (A_4:B_4), rather than according to the ratio 1:2. Second, we can estimate first both the average frequency of synonymous transitions over the twofold-degenerate and fourfold-degenerate sites and the average frequency of nonsynonymous transversions over the nondegenerate and twofold-degenerate sites. The final estimates for the synonymous (K_s) and nonsynonymous (K_a) substitutions are then calculated as

$$K_s = B_4 + (L_2 A_2 + L_4 A_4)/(L_2 + L_4) \quad (1)$$

and

$$K_a = A_0 + (L_0 B_0 + L_2 B_2)/(L_0 + L_2), \quad (2)$$

where L_0, L_2, and L_4 refer to the total number of nucleotides in the three degeneracy classes. We use formulas (1) and (2) for estimating nucleotide substitutions. Li (accepted) has independently derived the same estimators, and he also gives formulas for the variances that we use here:

$$V(K_s) = [L_2^2 V(A_2) + L_4^2 V(A_4)]/(L_2 + L_4)^2 + V(B_4) - b_4 Q_4[2a_4 P_4 - c_4(1 - Q_4)]/(L_2 + L_4) \quad (3)$$

Downloaded from https://academic.oup.com/mbe/article/12/2/271/1040324 by guest on 16 August 2022

and

$$V(K_a) = V(A_0)+[L_0^2V(B_0)+L_2^2V(B_2)]/(L_0 + L_2)^2 - b_0Q_0[2a_0P_0 - c_0(1 - Q_0)]/(L_0 + L_2), \quad (4)$$

where the variances $V(A_i)$ and $V(B_i)$ are as given by Li et al. (1985), P_i and Q_i are, respectively, the observed proportions of transitional and transversional differences in the degeneracy class i, and $a_i = 1/(1-2P_i-Q_i)$, $b_i = 1/(1-2Q_i)$, and $c_i = (a_i-b_i)/2$.

When the codons to be compared differ by two or three nucleotides, there are different possibilities concerning the order in which the nucleotides have changed. We have used two ways for estimating the substitutions in such codons. First, all possible paths can be equally likely (except those involving a stop codon). Second, the estimates can be calculated by maximizing the synonymous substitutional steps. The two methods give generally very similar values; the values reported here are based on the maximization of synonymous substitutions.

Results

The Gene Trees

The sequences produced by PCR amplification can be aligned without any gaps. Of the 312 nucleotides, 110 show a nucleotide difference in at least one of the sequences. When these variable nucleotides are used as characters, the putative branching pattern of the genes is reconstructed with the parsimony method (fig. 1). The minimum number of substitutions required in the best tree would be 126 if there were no homoplasy. The most parsimonious tree has 170 steps in it. When drawn with the two

Downloaded from https://academic.oup.com/mbe/article/10/2/271/1140324 by guest on 16 August 2022

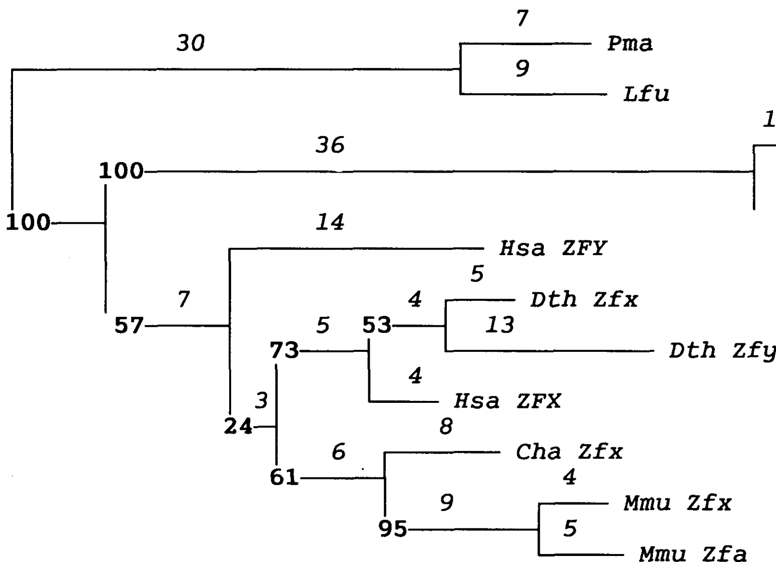

FIG. 1.—Parsimony tree of the PCR-amplified sequences (Lanfear and Holland 1991). The numbers at the roots of the branches show how often a cluster appeared when bootstrapping over nucleotides 100 times; the estimated numbers of nucleotide substitutions are given above the branches. The species are as follows: *Cha* = Chinese hamster; *Dth* = *Dusicyon thous; Hsa* = *Homo sapiens; Lfu* = *Larus fuscus; Mmu* = *Mus musculus;* and *Pma* = *Parus major.*

avian sequences as an outgroup, the tree suggests that there are not separate clusters of the X- and Y-chromosomal genes but that the clustering is affected by the taxonomy. The two fox genes cluster together, and the human *ZFY* does not cluster together with the mouse *Zfy* genes, although it remains outside the cluster of the X-chromosomal sequences. In the mouse, the two Y-chromosomal sequences (*Zfy-1* and *Zfy-2*) cluster closely together, as do the X-chromosomal (*Zfx*) and autosomal (*Zfa*) sequences. The topology of the neighbor-joining tree (based on the K_s estimates) differs from the parsimony tree only by clustering the human *ZFY* together with the hamster *Zfx*.

The clustering of the fox genes appears not statistically confident on the basis of bootstrapping (they stay together in 53% of the repeats). However, when the two human genes are removed from the data, the two fox genes cluster together in 88% of the bootstrapped trees. This tends to suggest that the two genes in the fox are indeed more similar to each other than either is to a murine gene.

A more detailed picture of the evolution of the zinc-finger genes can be obtained from the complete sequences of the coding region of the two human (*ZFX* and *ZFY*) and the four murine (*Zfx, Zfy-1, Zfy-2,* and *Zfa*) genes. Alignment of the sequences is somewhat problematic, as the sequence at the beginning of the *Zfa* gene (corresponding to a sequence coding for the first 20 amino acids following the methionine) is significantly different from those in all the other genes, and there are several gaps, particularly in the *Zfa* sequence. All the gaps map among the exons in the 5′ half. The alignment for the last exon, responsible for the finger structures, requires no gaps. As both the association of the *Zfy-1* and *Zfy-2* genes and that of the *Zfx* and *Zfa* genes are obvious, we include only four genes in the phylogenetic analysis—namely, *ZFX, ZFY, Zfx,* and *Zfy-2.* This avoids some of the problems of the alignment.

There are 703 nucleotide sites that show a difference among the four genes: 559 with one gene differing from the others; 86 with two character states (nucleotides), each shared by two sequences; 54 with three character states; and 4 in which each of the four genes has a unique nucleotide. The 86 phylogenetically informative nucleotides can be used to examine which one of the three possible clusterings they support. Thirty-eight support clustering on the chromosomal basis (X vs. Y), while 42 support clustering on the taxonomic basis (human vs. mouse). The remaining six sites support clustering that has no clear evolutionary interpretation (fig. 2).

Conflicting patterns arise when the analysis is done separately for different exons. The 5′ half of the gene, responsible for the acidic and basic parts of the protein, supports the clustering on the chromosomal basis rather than on the taxonomic basis (27 sites vs. 18 sites), whereas the opposite is true for the exon coding for the finger structures (11 vs. 24; the 2×2 contingency test shows a significant difference between the 5′ and 3′ halves, with $G = 6.44$, df = 1, $P = 0.011$). The distribution of the sites

Downloaded from https://academic.oup.com/mbe/article/10/2/271/1140324 by guest on 18 August 2022

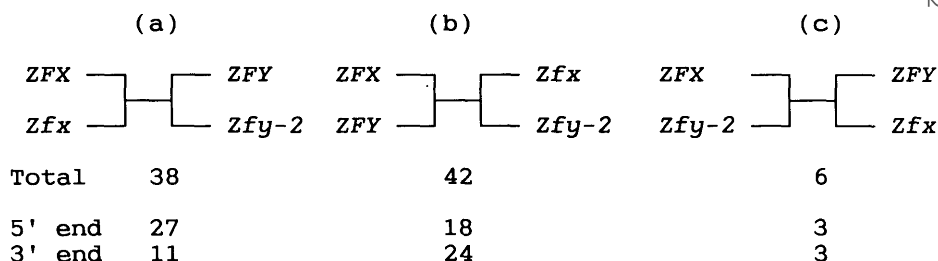

FIG. 2.—Possible clusterings of the human and murine genes, and the numbers of nucleotides supporting each of them, separately for the 5′ and the 3′ ends of the genes.

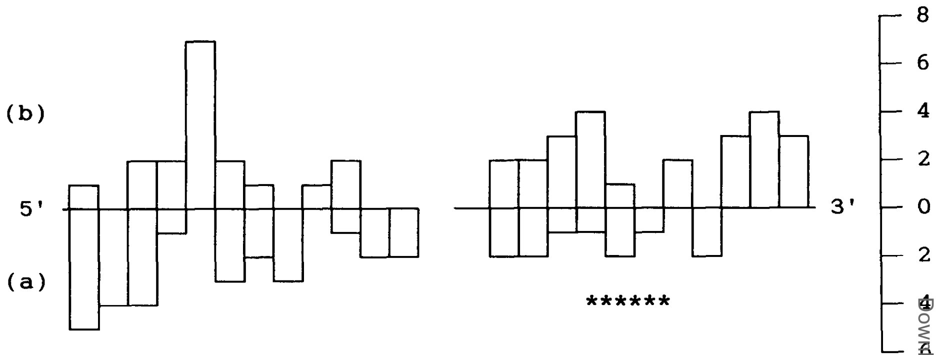

FIG. 3.—Distribution, along the genes, of the nucleotides supporting either clustering (a) (columns below the line) or clustering (b) (columns above the line), both of which are shown in fig. 2 (the columns are based on 100 nucleotides each, and the gap separates the last exon responsible for the finger structures). The asterisks indicate the location of the PCR-amplified sequence (Lanfear and Holland 1991).

supporting either clustering shows some aggregation along the gene, but there is also a large scatter (fig. 3).

Nucleotide Substitutions

Nucleotide substitutions between the complete sequences of the human and murine genes are calculated separately for the 5′ and the 3′ ends of the genes [the first 20 codons are neglected because of problems in aligning *Zfa* with the other genes (table 1)]. As expected, most substitutions are synonymous. The mean ratio of synonymous

Table 1
Estimates of Nucleotide Substitutions, per 100 Sites, between Human and Murine Genes

A. 5′ Half

	ZFX	*ZFY*	*Zfx*	*Zfy-1*	*Zfy-2*	*Zfa*
ZFX		20 (2)	31 (3)	56 (5)	56 (5)	31 (3)
ZFY	6 (1)		40 (3)	53 (5)	54 (5)	42 (4)
Zfx	4 (1)	8 (1)		76 (6)	80 (7)	8 (1)
Zfy-1	22 (1)	22 (1)	23 (1)		7 (1)	74 (7)
Zfy-2	23 (1)	23 (1)	23 (1)	4 (1)		78 (7)
Zfa	7 (1)	10 (1)	4 (1)	25 (2)	26 (2)	

B. 3′ Half

	ZFX	*ZFY*	*Zfx*	*Zfy-1*	*Zfy-2*	*Zfa*
ZFX		17 (2)	30 (3)	51 (4)	51 (4)	30 (3)
ZFY	1 (0)		36 (4)	54 (5)	54 (5)	37 (4)
Zfx	1 (0)	2 (0)		46 (4)	46 (4)	2 (1)
Zfy-1	10 (1)	11 (1)	10 (1)		1 (0)	47 (4)
Zfy-2	10 (1)	11 (1)	10 (1)	1 (0)		46 (4)
Zfa	2 (0)	3 (0)	2 (0)	11 (1)	11 (1)	

NOTE.—Data above the diagonals are the estimates for the synonymous differences (K_s), and data below the diagonals are the nonsynonymous differences (K_a). Standard errors (in parentheses) are calculated according to the method of Li (accepted) [our eqq. (3) and (4)]. The 3′ half is the last exon.

Downloaded from https://academic.oup.com/mbe/article/10/2/271/1140324 by guest on 16 August 2022

to nonsynonymous substitutions, K_s/K_a, is 3.4 (range 1.6–8) in the 5′ half of the genes. The comparisons of the 3′ sequences, between the murine Y-chromosomal genes and all other genes, produce a K_s/K_a ratio of 4.1–5.1, and in the comparisons of the 3′ parts of *ZFX, ZFY,* and *Zfx* the synonymous substitutions outnumber nonsynonymous substitutions, by a factor of 11–51. Similar high ratios of K_s/K_a are obtained from the partial sequences produced by PCR amplification, the ratio being generally >20, except that in comparisons involving *Zfy-1* or *Zfy-2* it is ~10.

Nonsynonymous substitutions are particularly rare in the last exon (coding for the finger domain). This is seen from the K_s/K_a ratios being higher in the 3′ end than in the 5′ end. Furthermore, the number of nucleotide differences per nonsynonymous sites is in the pairwise comparisons, higher in the 5′ end than in the 3′ end (table 1).

The comparison between *Zfx* and *Zfa* and that between *Zfy-1* and *Zfy-2* form exceptions in that there are equally many synonymous and nonsynonymous substitutions (per respective sites) in the 3′ halves of the genes. In both cases the estimates are based on small numbers of nucleotide differences.

The estimates of K_s are generally equal for the two halves of the genes, with a couple of exceptions. In the comparison between *Zfy-1* and *Zfy-2* the ratio $K_s(5')/K_s(3') = 13.6$, and in the comparison between *Zfx* and *Zfa* it is 3.6. These values are partly sensitive to both the close relationship and a small overall number of nucleotide differences, but they also indicate that there has been either a homogenizing factor in the 3′ end or a diversifying factor in the 5′ end. There are a total of 15 synonymous differences between the 5′ ends of the genes *Zfy-1* and *Zfy-2,* but there are only 2 differences in the 3′ end, the respective numbers being 17 and 8 between *Zfx* and *Zfa.* The expectation would be that there are about equally many differences in the two halves of the genes if synonymous substitutions are randomly distributed.

The 3′ halves of the X- and Y-chromosomal genes show also a greater similarity than do the 5′ halves, in both mouse and human. The comparison of the mouse genes *Zfy-1* and *Zfy-2* with *Zfx* gives ratios of $K_s(5')/K_s(3')$ of 1.6 and 1.7, respectively. (As *Zfa* is almost identical with *Zfx,* the same applies to the comparison between it and the Y-chromosomal genes.) The same ratio is 1.2 in the comparison between the human genes *ZFX* and *ZFY.*

The estimation method introduced in formulas (1) and (2) gives generally smaller values for K_s and larger values for K_a than does either the method of Li et al. (1985) or the method of Nei and Gojobori (1986). These two latter methods give estimates of K_s that are generally 10%–50% larger than those calculated here. However, the present estimates of K_s are much larger than those that Lanfear and Holland (1991) estimate from the PCR-amplified sequences.

Discussion

The sequence comparisons demonstrate that the rate of substitutions has not been equal in all nucleotides. There are many more synonymous than nonsynonymous differences, and there are more differences in the exons coding for the acidic and basic polypeptides than there are in the last exon coding for the finger domain. This indicates the conservatism of the zinc-finger structures, strongly emphasized in earlier sequence comparisons (Koopman et al. 1991).

The results cast some doubt on the independence of the evolution of different genes within the same species. The last exons of the murine genes *Zfy-1* and *Zfy-2* are almost identical, in contrast to many differences in the 5′ half. The scarcity of nonsynonymous substitutions in the last exon is explained by the constraints posed by the conservatism of the zinc-finger structures. The 3′ end of the gene is, however,

Downloaded from https://academic.oup.com/mbe/article/10/2/271/1140324 by guest on August 2022

not only conservative in nonsynonymous nucleotides, but the difference in synonymous sites is also only one-tenth of that seen in the 5′ half. Although the synonymous substitutions, in some cases, appear to be correlated with the rate of nonsynonymous substitutions in the surrounding nucleotides (Miyata et al. 1987; Li and Graur 1991, p. 77), the similarity of the last exons of *Zfy-1* and *Zfy-2* is probably best explained by a homogenizing event such as gene conversion since the duplication of the gene.

The high similarity of the fox X- and Y-chromosomal genes in the PCR results also indicates that the two genes have been homogenized after the radiation of the mammalian families. A similar tendency, although not so clear, is seen in the comparison of the human and murine sequences. The nucleotide substitutions in the last exon support clustering on the basis of taxonomy whereas the nucleotides in the 5′ half of the gene support clustering on the chromosomal basis (X- vs. Y-chromosomal genes).

The similarity of the X- and Y-chromosomal genes within a species hints of recombination or gene conversion in the history of the genes. Recombination of the sex chromosomes is restricted to the pseudoautosomal region that includes sequences homologous between the two chromosomes; the zinc-finger genes in both human and mouse have been mapped outside this region (Nagamine et al. 1989). Schneider-Gädicke et al. (1989*b*) proposed that gene conversion between the human *ZFX* and *ZFY* could explain their high similarity, which also extends outside the coding region.

The putative conversion (or recombination) between the X- and Y-chromosomal genes makes it difficult to estimate the rate of evolutionary change in them. Lanfear and Holland (1991) made such a comparison from the PCR-amplified sequences and concluded that the evolution has been faster in the Y chromosome than in the X chromosome, although the difference was less than the predicted threefold difference. Similarly, Bianchi et al. (1992) concluded, on the basis of restriction-fragment differences, that the *Zfy* genes in sigmodontine rodents have evolved faster than the *Zfx* genes. This comparison is a little problematic, because (*a*) the *Zfy* genes are amplified in these species and (*b*) the copy number and restriction-fragment number vary. However, comparisons of species with a similar level of amplification showed a similar elevation of differentiation in the *Zfy* genes. Rapid evolution of the murine Y chromosome was also inferred by Tucker et al. (1989).

The parsimony analysis of the human and murine genes (fig. 2) did not reveal a clear clustering of the genes. If we take into account the phylogenetically noninformative nucleotide differences, the total number of nucleotide differences separating the murine *Zfy-2* from the three other genes is much larger than the number of differences separating these three genes from each other, whether the clustering follows pattern (a) or pattern (b) in figure 2. It is likely that most nucleotide differences characterizing *Zfy-2* must have taken place in the mouse Y chromosome rather than in the common ancestor of the other genes. This implies that the murine *Zfy* has evolved faster than the other genes. The same is seen in the parsimony tree of the PCR-amplified sequences; there have been more changes in the Y chromosome than in the X chromosome, in both mouse and fox (fig. 1). The parsimony trees, however, include both synonymous and nonsynonymous differences. For testing the relative rates, we should only use the synonymous changes.

Assuming that the parsimony tree (a) (fig. 2) applies to the 5′ end of the gene and that the tree (b) applies to the 3′ end of the gene, we use the neighbor-joining method to estimate the branch lengths for synonymous substitutions (fig. 4). The trees are rooted by assuming that the rate difference between the X- and Y-chromo-

Downloaded from https://academic.oup.com/mbe/article/10/2/271/1044024 by guest on 16 August 2022

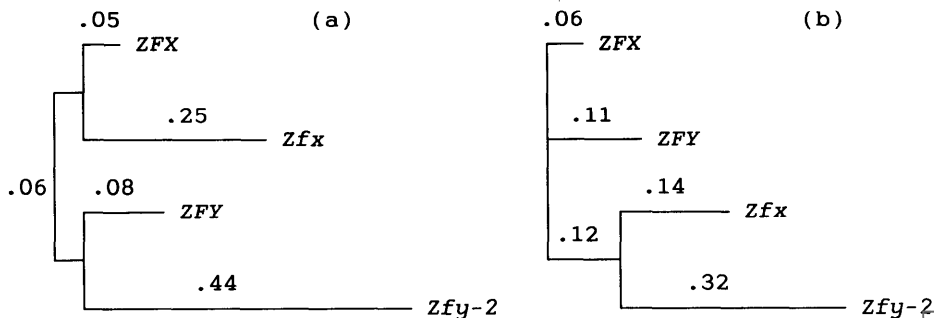

FIG. 4.—Gene trees produced by the neighbor-joining method, on the basis of synonymous substitutional differences (K_s). (a), 5′ Half of the gene. (b), 3′ Half (the last exon) of the gene.

somal genes, on the one hand, and between the human and murine genes, on the other hand, is the same in all exons.

The resulting trees suggest that the rate of nucleotide substitution in the Y-chromosomal genes has been about twice the rate in the X-chromosomal genes (mean 1.9, range 1.5–2.4) and that the murine genes have evolved at a rate 4.7 times as fast as that in the human genes (range 4.2–5.4). The trees also suggest that the X- and Y-chromosomal genes have separated only slightly before the separation of the rodent and primate lineages and that there has been recombination or conversion between the murine genes, homogenizing the 3′ ends. Table 1 shows that the estimate for synonymous differences between *Zfx* and *Zfy-2* is much smaller for the 3′ end (K_s = 0.46) than for the 5′ end (K_s = 0.80). The conclusion concerning the branching pattern of the 3′ end [figs. 2(b) and 4(b)] disagrees with the parsimony tree based on the PCR-amplified sequences from that part of the gene (fig. 1). Such a discrepancy could indicate sampling effects or true heterogeneity among different parts of the gene (see fig. 3).

The obtained rate difference between the X- and Y-chromosomal genes agrees well with the theoretical expectations and with the observations from other genes (Miyata et al. 1987) and is very close to that which Lanfear and Holland (1991) calculated from the partial data. The estimated rate difference between the murine and human genes is larger than the mean difference from several other genes. Although Li et al. (1987) estimated that the rate of synonymous substitutions in mice can be about six times higher than that in humans, the average long-term rate difference between the rodent and primate lineages is about twofold (Li et al. 1990). Bulmer et al. (1991) suggest that the rate of synonymous substitutions has been elevated in the rodents recently, after the separation of mice and rats, but there are also claims that no rate differences exist among the mammalian lineages (Easteal 1990).

If it is assumed that the rodent and primate lineages separated from each other ~80 Mya, it is possible to calculate rough estimates for synonymous substitutions per year. These estimates would then be 0.70×10^{-9} for *ZFX*, 1.16×10^{-9} for *ZFY*, 3.15×10^{-9} for *Zfx*, and 6.25×10^{-9} for *Zfy-2*. The estimates for humans are smaller than those generally obtained (Li et al. 1990). This is partly due to the new method introduced here for estimating synonymous substitutions, and the rate in the X-chromosomal sequences is predicted to be smaller than that in the autosomes (two-thirds).

Our interpretation of the results supports the hypothesis that rate differences depend on the mutation rates. The results also indicate past interactions—recombination or gene conversion—between different genes within a species. This makes the

Downloaded from https://academic.oup.com/mbe/article/10/2/271/1140249 by guest on 25 August 2022

inference indicated on the gene trees problematic. The 5′ and 3′ ends of the genes give consistent rate estimates, although they support different topologies of the gene tree.

Acknowledgments

We thank M. Nei and two reviewers for constructive criticisms on the manuscript, and we thank W.-H. Li for sending us his unpublished manuscript. This work has been supported by grants from the Natural Science Research Councils in Finland and Sweden (to P.P.) and from the Sigrid Juselius Foundation, the Academy of Finland, and the CONICET and CIC of Argentina (all to N.O.B.)

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.

BIANCHI, N. O., A. DE LA CHAPELLE, L. VIDAL-RIOJA, and S. MERANI. 1989. The sex determining zinc finger sequences in XY females of *Akodon azarae* (Rodentia Cricetidae). Cytogenet. Cell Genet. **52**:162–166.

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.

BULMER, M., K. H. WOLFE, and P. M. SHARP. 1991. Synonymous nucleotide substitution rates in mammalian genes: implications for the molecular clock and the relationship of mammalian orders. Proc. Natl. Acad. Sci. USA **88**:5974–5978.

EASTEAL, S. 1990. The pattern of mammalian evolution and the relative rate of molecular evolution. Genetics **124**:165–173.

ELLIS, N., and P. N. GOODFELLOW. 1989. The mammalian pseudoautosomal region. Trends Genet. **5**:406–410.

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

———. 1986. PHYLIP 2.8. Department of Genetics, University of Washington, Seattle.

JIN, L., and J. W. H. FERGUSON. 1990. NJTREE (version 2.0). Center for Population Genetics, University of Texas, Houston.

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

———. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

KOOPMAN, P., A. ASHWORTH, and R. LOVELL-BADGE. 1991. The ZFY gene family in humans and mice. Trends Genet. **7**:132–136.

LAIRD, C. D., B. L. MCCONAUGHTY, and B. J. MCCARTHY. 1969. Rate of fixation of nucleotide substitutions in evolution. Nature **224**:149–154.

LANFEAR, J., and P. W. H. HOLLAND. 1991. The molecular evolution of ZFY-related genes in birds and mammals. J. Mol. Evol. **32**:310–315.

LAU, Y.-F. C., 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.

LI, W.-H. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. (accepted).

LI, W.-H., M. GOUY, P. M. SHARP, C. O'HUIGIN, and Y.-W. YANG. 1990. Molecular phylogeny of Rodentia, Lagomorpha, Primates, Artiodactyla, and Carnivora and molecular clocks. Proc. Natl. Acad. Sci. USA **87**:6703–6707.

Downloaded from https://academic.oup.com/mbe/article/10/2/271/1140324 by guest on 16 August 2022

LI, W.-H., and D. GRAUR. 1991. Fundamentals of molecular evolution. Sinauer, Sunderland, Mass.

LI, W.-H., M. TANIMURA, and P. M. SHARP. 1987. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. J. Mol. Evol. **25**:330–342.

LI, W.-H., C.-I. WU, and C.-C. LUO. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. **2**:150–174.

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. 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.

MIYATA, T., H. HAYASHIDA, K. KUMA, K. MITSUYASU, and T. YASUNAGA. 1987. Male-driven molecular evolution: a model and nucleotide sequence analysis. Cold Spring Harb. Symp. Quant. Biol. **52**:863–867.

NAGAMINE, C. M., K. 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. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

NEI, M., and T. GOJOBORI. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**:418–426.

PAGE, D. C., R. MOSHER, E. M. SIMPSON, E. M. C. 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.

PALMER, M. S., P. BERTA, A. H. SINCLAIR, B. PYM, and P. N. GOODFELLOW. 1990. Comparison of human ZFY and ZFX transcripts. Proc. Natl. Acad. Sci. USA **87**:1681–1685.

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

SCHNEIDER-GÄDICKE, A., P. BEER-ROMERO, L. G. BROWN, G. MARDON, S.-W. LUOH, and D. C. PAGE. 1989*a*. Putative transcription activator with alternative isoforms encoded by human ZFX gene. Nature **342**:708–711.

SCHNEIDER-GÄDICKE, A., P. BEER-ROMERO, L. G. BROWN, R. NUSSBAUM, and D. C. PAGE. 1989*b*. ZFX has a gene structure similar to ZFY, the putative human sex determinant, and escapes X inactivation. Cell **57**:1247–1258.

SINCLAIR, A. H., J. W. FOSTER, J. A. SPENCER, D. C. PAGE, M. PALMER, P. N. GOODFELLOW, and J. A. M. GRAVES. 1988. Sequences homologous to ZFY, a candidate human sex-determining gene, are autosomal in marsupials. Nature **336**:780–783.

TUCKER, P. K., B. K. LEE, and E. M. EICHER. 1989. Y chromosome evolution in the subgenus *Mus* (genus *Mus*). Genetics **122**:169–179.

WU, C.-I., and W. H. LI. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. USA **82**:1741–1745.

MASATOSHI NEI, reviewing editor

Received January 28, 1992; revision received June 16, 1992

Accepted June 18, 1992

Downloaded from https://academic.oup.com/mbe/article/9/2/271/1403244 by guest on 16 August 2022