

# Sex Chromosomal Transposable Element Accumulation and Male-Driven Substitutional Evolution in Humans

Rikard Erlandsson,<sup>1</sup> James F. Wilson,<sup>2</sup> and Svante Pääbo

Max-Planck-Institute for Evolutionary Anthropology, Leipzig, Germany

We sequenced the genomic region containing the human Y-linked zinc finger gene (ZFY). Comparison of ZFY to the related region on the X chromosome (ZFX) and to autosomal sequences reveals a significant accumulation of transposable elements on the sex chromosomes. In addition, five times as many retroviruslike elements (RLEs) are present in the ZFY region as in the ZFX region. Thus, transposable elements accumulate more rapidly on the sex chromosomes, and the insertion of RLEs may occur more frequently in the male than in the female germ line. When the accumulation of substitutions in Alu elements was analyzed, it was found that the Alu elements at the Y-chromosomal locus diverged significantly faster than those at the X-chromosomal locus, whereas the divergence of autosomal Alu elements was intermediate. The male-to-female mutation rate ratio was estimated to be 2.5.

## Introduction

In mammals, sex is determined by the presence or absence of the Y chromosome, which encodes the SRY gene necessary for testis development (Whitfield, Lovell-Badge, and Goodfellow 1993). The acquisition by the Y chromosome of the sex-determining function early in mammalian evolution is likely to have been associated with the suppression of recombination between large regions of the X and Y chromosomes (Charlesworth 1991; Graves 1995). Due to this absence of recombination, most DNA sequences on the Y chromosome are confined to the male germ line, whereas their counterparts on the X chromosome spend one third of their time in the male germ line, and two thirds in the female germ line. The comparison of related, strictly sex-linked sequences therefore allows the study of evolutionary forces that affect the genome in the male and female germ lines.

To date, comparisons of homologous sequences on the sex chromosomes in humans include intronic (Shimmin, Chang, and Li 1993) and exonic sequences (Shimmin, Chang, and Li 1994) in the X- and Y-linked zinc finger genes (ZFX and ZFY, respectively) and intronic sequences in the amelogenin genes (Huang et al. 1997), as well as exonic (Agulnik et al. 1997) and intronic (Chang, Hewett-Emmett, and Li 1996) sequences in the SMCX/SMCY genes. Although these sequences add up to less than 10 kb, they have revealed that more nucleotide substitutions occur per unit time on the Y chromosome than on the X chromosome. When extrapolated to the mutation rate in the germ lines, the data indicate that two to six times more mutations occur in the male germ line (Shimmin, Chang, and Li 1993, 1994; Huang et al. 1997). Therefore, it has been suggested that sub-

stitutional evolution in the nuclear genome is "male-driven" (see Hurst and Ellegren [1998] for a review). However, some data seem to challenge this. For example, comparison of the rate of silent substitution on the autosomes and sex chromosomes of rodents has revealed a reduced rate on the X chromosome relative to the autosomes, rather than a higher rate on the Y chromosome (McVean and Hurst 1997).

ZFY and ZFX are located on the nonrecombining parts of the sex chromosomes, ZFY at Yp11.32, approximately 200 kb from the pseudoautosomal boundary (Page et al. 1987), and ZFX at Xp22.12, approximately 23 Mb proximal to this boundary (Nelson et al. 1995). Their common ancestor transposed to the sex chromosomes before the divergence of placental mammals (Page et al. 1987; Sinclair et al. 1988; Watson et al. 1991), approximately 115–130 MYA (Carroll 1988; Janke et al. 1994). In humans, both genes are ubiquitously expressed in adult tissues, and ZFX is not subject to X-inactivation (Palmer et al. 1990).

We determined 136 kb of sequence at the ZFY locus. By comparing this sequence with that of the related region at the ZFX locus and with autosomal sequences of similar GC content, we examined rates of nucleotide substitutions, insertions/deletions, and transpositions in the male and female germ lines. The results allow an estimate of the ratio of the male-to-female substitution rates based on large numbers of Alu repeats. The data furthermore reveal that a larger fraction of the sex-chromosomal sequences is made up of interspersed repeats and that more retroviral-derived elements accumulate on the Y chromosome than on the autosomes and, possibly, the X chromosome.

## Materials and Methods

### Sequencing of the ZFY Region

Filters spotted with clones from a human genomic PAC library (RPC11) obtained from the German Genome Project Resource Center (Zehetner and Leh-rach 1994) were screened with a PCR-amplified random-labeled probe from the last exon of ZFX. Several positive clones were identified. Two of them contained ZFX and one contained ZFY, as determined by specific PCR amplification of the last exons. The ZFY

<sup>1</sup> Present address: Department of Biotechnology, Royal Institute of Technology, Stockholm, Sweden.

<sup>2</sup> Present address: Department of Zoology, University of Oxford, Oxford, England.

Key words: ZFY, ZFX, human sex chromosomes, Alu repeats, retroviral elements, retroviruslike elements.

Address for correspondence and reprints: Rikard Erlandsson, Department of Biotechnology, Royal Institute of Technology (KTH), S-10044 Stockholm, Sweden. E-mail: rikerl@biochem.kth.se.

*Mol. Biol. Evol.* 17(5):804–812. 2000

© 2000 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

**Table 1**  
**Comparison of Repetitive DNA in the ZFX and ZFY Regions and in the GenBank**

REPEAT FAMILY	ZFX			ZFY			GENBANK 40% GC			SIGNIFICANCE		
	No.	No./kb	%	No.	No./kb	%	No.	No./kb	%	X-A	A-Y	X-Y
SINEs	110	0.99	25.9	106	0.78	19.3	1,379	0.34	7.3	0.000	0.000	NS
Alu's	101	0.91	24.8	103	0.76	19.0	885	0.22	5.7	0.000	0.000	NS
MIRs	9	0.08	1.1	3	0.02	0.3	494	0.12	1.6	NS	NS	NS
LINEs	34	0.31	14.5	36	0.27	24.9	1,290	0.34	22.6	NS	NS	NS
RLEs	3	0.03	1.5	21	0.16	8.9	389	0.10	5.6	NS	0.034	NS
MaLRs	1	0.01	0.3	7	0.05	2.2	261	0.06	3.1	NS	NS	NS
HERVs	1	0.01	0.3	8	0.06	4.2	80	0.02	1.7	NS	0.003	NS
MER4 group	1	0.01	0.9	6	0.04	2.1	48	0.01	0.8	NS	0.002	NS
DNA elements	9	0.08	1.8	6	0.04	1.2	289	0.07	1.9	NS	NS	NS
Unclassified	0	0.00	0	1	0.01	0.1	80	0.02	1.0	NS	NS	NS
Total interspersed repeats	156	1.41	43.7	170	1.25	54.4	3,436	0.84	38.5	0.000	0.000	NS
Microsatellites	11	0.10	0.5	9	0.07	0.4	—	—	—	—	—	NS

NOTE.—Abbreviations: SINE, short interspersed nuclear element; MIR, mammalian-wide interspersed repeat; LINE, long interspersed nuclear element; RLE, retroviruslike element (containing long terminal repeats); MaLR, mammalian-wide apparent LTR transposon; HERV, human endogenous retrovirus; MER4, medium reiteration repetitive element 4 family, data for all nonredundant human sequences around 40% GC in GenBank from Smit (1996); NS, nonsignificant.

clone (LLNLP704G05242Q13) had an insert of 135,849 bp, spanning the entire gene. Fluorescence in situ hybridization (FISH) analysis confirmed that the clone was located at Yp11 and was not chimeric (data not shown). The purified P1 artificial chromosome (PAC) DNA was partially digested with Tsp509I and size-separated on an agarose gel after treatment with Plasmid-Safe (Biozym). Fragments of sizes 1–1.5 kb, 1.5–2.5 kb, and 2.5–5 kb were subcloned into pUC18/*EcoRI* and introduced XL1-Blue into *Escherichia coli* by electroporation. Ampicillin selection was performed on standard LB-agar plates. Either templates for sequencing were amplified directly by colony PCR (Hultman et al. 1991), or plasmids were prepared by an Autogene 740 robot from overnight LB-ampicillin cultures of positive clones. Templates were sequenced in both directions using Cy5- and FITC (fluorescein isothiocyanate)-labeled universal (5'-GCCA-GTGCCAAGCTTGCA) and reverse (5'-CAGCTAT-GACCATGATTACGA) primers and electrophoresed on ARAKIS (EMBL, Heidelberg) and ALF/ALFexpress sequencers (Amersham Pharmacia Biotech).

The pCPAC2 vector sequence was included in contig building by SeqmanII (DNASTar) to evaluate coverage. When approximately fivefold coverage was achieved, it was possible to order all sequences in two contigs. Primers designed from the ends of the contigs generated two PCR products (GAP1, 300 bp; GAP2, 1,400 bp) covering the two gaps. These were cloned into the pGEM-T vector (Stratagene), and several independent clones were sequenced. A minimum coverage of each base once on both strands, or sequencing with dye primers and dye terminators (Amersham Pharmacia Biotech), was achieved by primer walking on plasmids. The sequences were ordered into a single contig using SeqmanII (DNASTar), and the contig was confirmed by long-range PCR; 9.7 kb of this region have previously been sequenced (GenBank accession numbers M30607, U24118, U00242, and AF026807 and sequence in North et al. [1991]).

## Analysis

Dot plots were drawn using the DOTTER program (Sonnhammer and Durbin 1995). Repetitive DNA was identified by RepeatMasker (Smit and Green, 1995–1998, <http://ftp.genome.washington.edu/cgi-bin/RepeatMasker>).

The numbers of repeats in the sequences were assumed to have a Poisson distribution, and differences in numbers were tested for significance accordingly. To compare X and Y, the *F* statistic was used. Using  $n(X)$  and  $n(Y)$  as the numbers of repeats on X and Y and  $L(X)$  and  $L(Y)$  as the lengths of the respective sequences, we calculated  $F(X, Y) = (n(X) + 0.5) L(Y) / (n(Y) + 0.5) L(X)$ , given that  $n(X)/L(X) > n(Y)/L(Y)$ . The critical values are found in the appropriate *F* table (at [http://www.ruf.rice.edu/~lane/hyperstat/F\\_table.html](http://www.ruf.rice.edu/~lane/hyperstat/F_table.html)), with  $n(X)$  being the numerator and  $n(Y)$  the denominator. When retroviruslike elements (RLEs) were interrupted by Alu and other elements, we counted the RLE fragments as one. This reduced the number of RLEs from 6 to 3 in the ZFX region and from 27 to 21 in the ZFY region (table 1).

RepeatMasker assigned each Alu element to a subfamily, and the number of transitions and transversions from the subfamily consensus sequences (Jurka and Milosavljevic 1991; Batzer et al. 1996) were counted. AluYa5, AluYa8, AluYb8, and elements that could not be assigned to any subfamily were excluded. The Kimura (1980) corrected divergence was calculated for each Alu element. Calculations were performed separately for ancestral CpG and non-CpG nucleotides; insertions/deletions and polyadenine tails were excluded from the analyses. The significance of differences between mean divergences was tested by unpaired *t*-tests. For example,  $t = |Y - X| / \sqrt{(s^2(1/n(X) + 1/n(Y)))}$ , where  $Y$  and  $X$  are the mean divergences on Y and X,  $n(X)$  and  $n(Y)$  are the numbers of Alu elements on X and Y, and  $s^2$  is a weighted estimate of the variance;  $s^2 = ((n(X) - 1)s(X)^2 + (n(Y) - 1)s(Y)^2) / ((n(X) - 1) + (n(Y) - 1))$ , where  $s(X)^2$  and  $s(Y)^2$  are the variances of the mean divergences on X and Y.

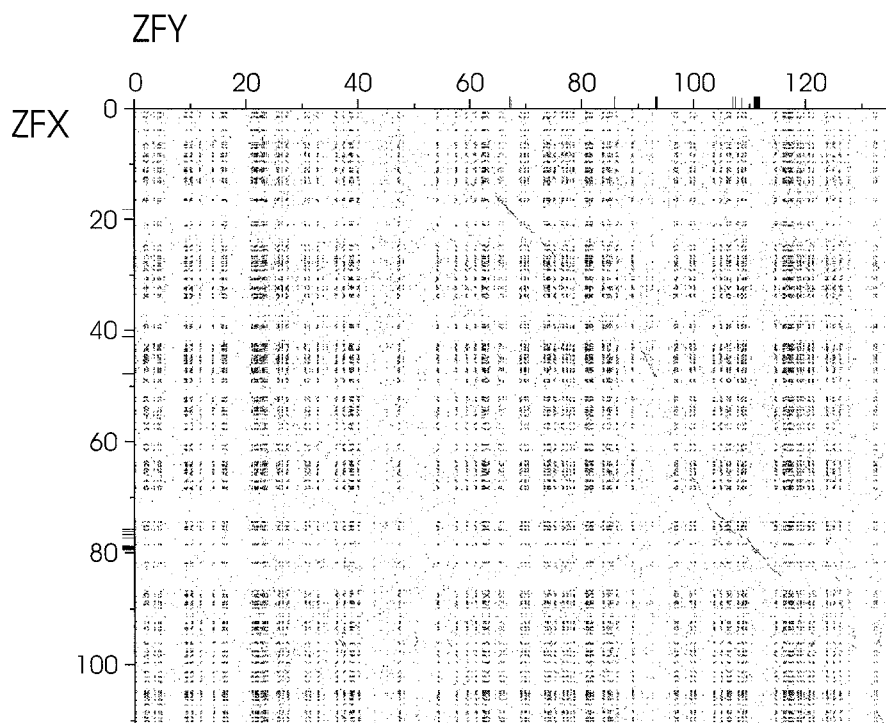


FIG. 1.—Dot plot comparison of the ZFY sequence versus the ZFX sequence. The length of the sequences (in kb) are given on the axes, with the CpG islands and exons marked as black boxes. Window size was 25 nt, and match scores were 85%–100% (white to black).

The male-to-female mutation rate ratio,  $\alpha$ , was calculated according to the following formulas ( $Y$ ,  $X$ , and  $A$  are the mean divergences of Alu elements on ZFY, ZFX, and autosomes, respectively): for  $R = A/X$ ,  $\alpha = (4R - 3)/(3 - 2R)$ , whose variance was approximated by the formula  $S_z^2 = S_x^2 (-6A/(3X - 2A))^2 + S_a^2 (6X/(3X - 2A))^2$ ; for  $R = Y/A$ ,  $\alpha = R/(2 - R)$  and  $S_z^2 = S_y^2 (2A/(2A - Y))^2 + S_a^2 (-2Y/(2A - Y))^2$ ; for  $R = Y/X$ ,  $\alpha = 2R/(3 - R)$  and  $S_z^2 = S_x^2 (-6Y/(3X - Y))^2 + S_y^2 (6X/(3X - Y))^2$ . The 95% confidence intervals were calculated as  $\alpha \pm 1.96S_z$  assuming that base changes are normally distributed.

The sizes of the deletions and insertions in the Alu sequences were counted from the RepeatMasker alignments, but without exclusion of any bases. The numbers of deletions and insertions per Alu were counted with the Alu normalized to a length of 300 bp (i.e., the number of deletions was divided by the length of the aligned sequence and multiplied by 300) and tested for significant differences under the assumption that they had a Poisson distribution.

## Results

### Features of the ZFY and ZFX Regions

A PAC carrying the human ZFY gene was identified using a probe encompassing the last exon of ZFX. Sequencing revealed that the 46-kb ZFY transcription unit, as well as 66 kb and 23 kb located 5' and 3' of the gene, respectively (135,849 bp in total), were contained in the PAC. The sequence of the human ZFX region (110,816 bp) was retrieved from GenBank (AC002404).

Exon-intron boundaries were determined by comparison with cDNA sequences and mouse *Zfx*, *Zfy-1*, and *Zfy-2* boundaries (Luoh and Page 1994; Mahaffey et al. 1997). All splice sites in the human genes adhere to the GT-AG rule. Two CpG islands exist around ZFY (at positions 67000 and 130000) (Page et al. 1987), and one is present in the ZFX sequence (around position 18000) (Schneider-Gädicke et al. 1989) (figs. 1 and 2). Analysis using the computer program GRAIL detected no significant open reading frames except those associated with the two known genes.

The GC contents for the ZFY and ZFX sequences are 39% and 41%, respectively. A dot plot comparison (fig. 1) shows that the strongest similarities between the two regions occur within and close to the exons and CpG islands. The 3' ends of the genes including the four last introns are, however, alignable after removal of repetitive elements (not shown). In contrast, only shorter sections in the larger 5' introns could be aligned.

### Numbers of Repetitive Elements

Using the RepeatMasker program (<http://ftp.genome.washington.edu/cgi-bin/RepeatMasker>), we identified 170 fragments of interspersed repeat elements and 9 microsatellites in the ZFY region. Together, they make up 55% of the sequence. In the ZFX region, 156 interspersed repeats and 11 microsatellites account for 44% of the sequence (fig. 2 and table 1).

In order to compare the occurrence of repeated elements at the sex-chromosomal loci with that of the genome in general, we used over 4 Mb of sequence (of a GC content similar to that of the ZFY and ZFX loci)

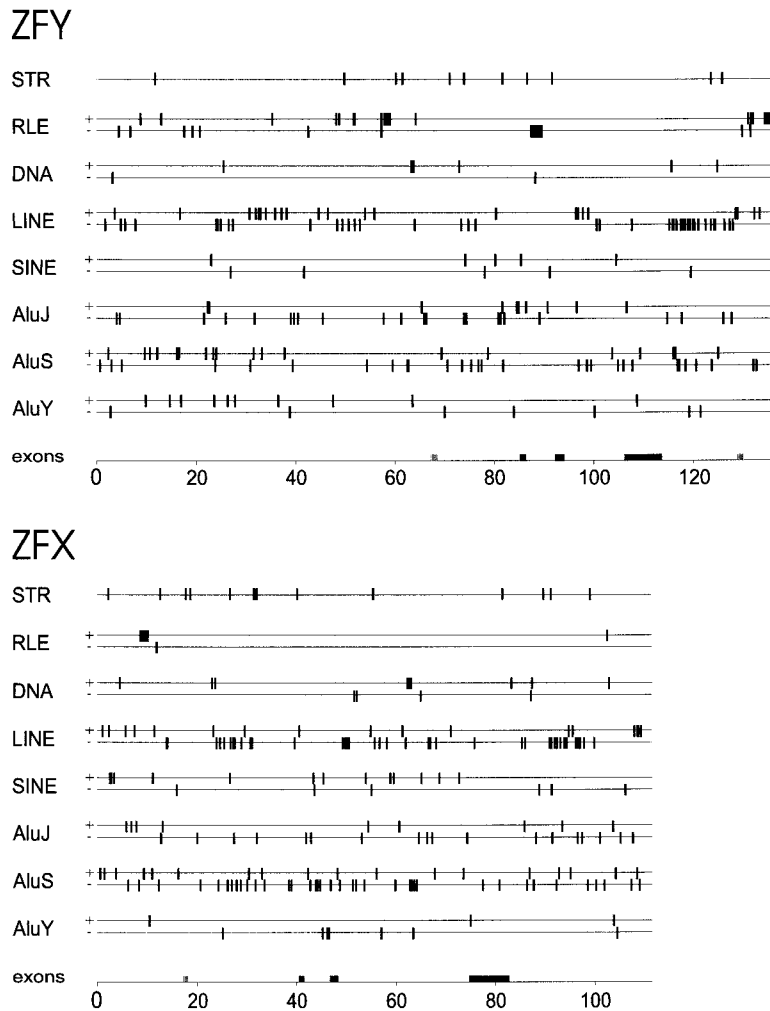


FIG. 2.—Repetitive elements in the ZFY and ZFX regions. Abbreviations: STR, short tandem repeat; RLE, retroviruslike element (containing long terminal repeats); DNA, DNA elements; LINE, long interspersed nuclear element; SINE, short interspersed nuclear element. AluJ, AluS, and AluY refer to Alu classes. The approximate locations of the exons (black boxes) and CpG islands (gray boxes) are indicated. Sizes are in kb.

retrieved from GenBank by Smit (1996) (table 1). When all interspersed elements are pooled, both sex-chromosomal sequences are found to harbor significantly more such elements than the autosomes ( $P < 0.001$ ). This is due to the occurrence of more Alu elements and more RLEs on the Y chromosome. The percentage of sequence composed of Alu elements and the density of Alu repeats increase from the autosomes (5.7%, 0.22 elements per kilobase) to ZFY (19%, 0.76/kb) to ZFX (25%, 0.91/kb). Furthermore, significantly more RLEs are present at ZFY than on the autosomes ( $P < 0.05$ ). This is due to increased numbers of the HERV and MER4 subfamilies (see Paulson et al. 1985; Smit 1993; Wilkinson, Mager, and Leong 1994; Erickson and Maeda 1995 for a description of the repeats). RLEs are about five times as numerous at ZFY (0.16/kb) as at ZFX (0.03/kb) and make up a correspondingly greater proportion of the sequence in the former (8.9%) compared with the latter (1.5%). However, this difference is not significant, presumably due to the low number of elements observed, particularly on the X

chromosome. The autosomal sequences are intermediate between the X and Y loci, with 0.10 RLEs/kb making up 5.6% of the sequence (table 1).

Although the densities of long interspersed nuclear elements (LINEs) are similar in the sequences compared (table 1), 25% of ZFY and 23% of the autosomes, but only 14% of ZFX, are made up of LINEs. This is because the mean length of LINEs on the autosomes (920 bp) is similar to that at ZFY (939 bp), whereas at ZFX, LINEs average only 474 bp in length ( $P < 0.05$ , *t*-test).

#### Alu Element Divergence

In order to compare the point mutational changes that have affected the ZFX and ZFY loci in spite of the fact that large parts of the sequences cannot be aligned, we made use of Alu elements. Alu elements have been divided into classes (and, further, into subfamilies) based on diagnostic positions within their sequences (Batzer et al. 1996). Each class is thought to be composed of retrotransposed copies of a master gene, the sequence of

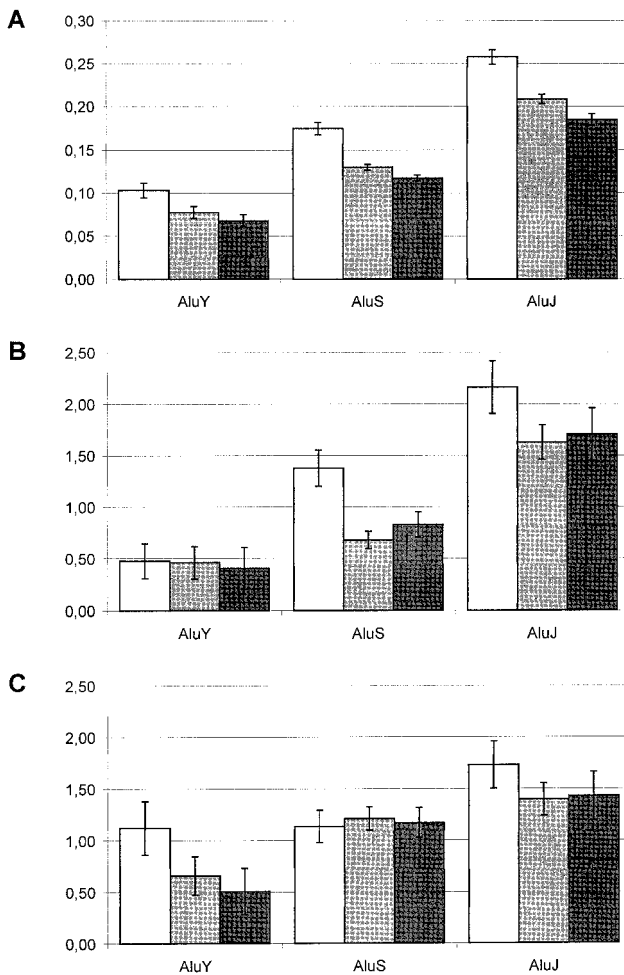


FIG. 3.—Mutational events in Alu repeats in the ZFY, ZFX, and autosomal sequences. Comparisons of different mutational events between the ZFY (white bars), autosomal (gray bars), and ZFX (dark gray bars) regions. The autosomal sequences were GenBank entries HS874C20, AC004804, AC001226, and AC005208 (GC content 38%–41%), from chromosomes 6, 12, 13, and 17, respectively, covering a total of 436,925 nt. Altogether, 17 AluY, 46 AluS, and 33 AluJ from ZFY; 19 AluY, 96 AluS, and 57 AluJ from the autosomes; and 10 AluY, 56 AluS, and 26 AluJ from ZFX were included. A, The Kimura-corrected mean Alu divergences ( $\pm$ SE) calculated as number of substitutions per base of aligned sequence, excluding polyA tails. The mean numbers of deletions and insertions per Alu (normalized to 300 bp length) are shown in B and C, respectively.

which is approximated by the consensus of that subfamily. Since different master genes were active at different times during primate evolution, each Alu class is of a different age (Deininger et al. 1992). The oldest class is AluJ (about 80 Myr), whereas AluS is of intermediate age, and AluY is the youngest (<20 MYA) (Batzer et al. 1996; Kapitanov and Jurka 1996; Mighell, Markham, and Robinson 1997).

The numbers of transitional and transversional differences from the consensus sequence of each Alu class were counted, and divergences were corrected for multiple substitutions (Kimura 1980). To allow autosomal comparison, 172 Alu elements from four autosomal regions were identified, and the divergences were similarly calculated (fig. 3A). As expected, the oldest class, AluJ, has the highest level of divergence, while the youngest class, AluY, has the lowest. Furthermore, Alu elements of all classes located at the ZFY locus have a higher mean sequence divergence than those on the autosomes or at the ZFX locus. For the AluJ and AluS classes, these differences are significant ( $P < 0.01$ , *t*-test). For each Alu class, the divergence for Alu elements located on the autosomes is intermediate to those at the ZFY and ZFX loci. Thus, Alu elements at the ZFY locus accumulate substitutions significantly faster than those on the autosomes and at ZFX.

Mutations at CpG dinucleotides occur at a higher rate than other substitutions due to methylation of many of these sites in mammalian genomes. Therefore, differences in the methylation patterns between the two chromosomes could conceivably account for the observed differences in the mutation rates. For example, Driscoll and Migeon (1990) suggested that the female germ line is hypomethylated compared with the male germ line. If the observed Y- to X-chromosomal ratio of divergences were due to such a difference in methylation, we would expect this ratio to decrease when CpG sites were removed from the analysis. However, the ratio of divergence between ZFY and ZFX Alu repeats increases when CpG dinucleotides present in the consensus sequences are excluded from the comparisons (table 2). Furthermore, when CpG sites are considered alone, they show higher levels of substitution at ZFY than at ZFX but to a lesser extent than the non-CpG sites (not shown). Thus, the difference in the accumulation of mutations affects both CpG and other sites.

**Table 2**  
Calculation of the Male-to-Female Substitution Rate Ratio ( $\alpha$ )

COMPARISON NUCLEOTIDES		ALUJ		ALUS		ALUY	
		Ratio	$\alpha$ (CI)	Ratio	$\alpha$ (CI)	Ratio	$\alpha$ (CI)
Y/X	Non-CpG	1.51	2.0 (1.5–2.6)	1.75	2.8 (1.6–4.0)	2.10	4.6 (–2.4–12)
	All	1.39	1.7 (1.4–2.1)	1.49	2.0 (1.6–2.3)	1.52	2.0 (1.0–3.1)
Y/A	Non-CpG	1.28	1.8 (1.2–2.4)	1.54	3.4 (1.0–5.7)	1.59	3.9 (–3.2–11)
	All	1.24	1.6 (1.3–2.0)	1.34	2.0 (1.5–2.6)	1.33	2.0 (0.6–3.4)
A/X	Non-CpG	1.18	2.7 (0.6–4.7)	1.13	2.1 (0.4–3.7)	1.32	6.3 (–25–38)
	All	1.13	2.0 (0.9–3.1)	1.11	1.8 (1.0–2.6)	1.14	2.2 (–1.4–5.8)

NOTE.—CI = 95% confidence interval. For Y/X,  $\alpha = 2R/(3 - R)$ ; for Y/A,  $\alpha = R/(2 - R)$ ; for A/X,  $\alpha = (4R - 3)/(3 - 2R)$  (Miyata et al. 1987); *R* = the ratio of the divergences of the Alu elements from their ancestral consensus sequences. Non-CpG and all nucleotide divergences were corrected for multiple hits according to the Kimura (1980) two-parameter model.

Older Alu elements have accumulated more deletions and insertions than younger ones, both at the ZFY and ZFX loci (fig. 3B and C). While the data are suggestive of more deletions and insertions at the ZFY locus than at the ZFX locus, the differences are not statistically significant, probably due to the small number of events scored. Furthermore, no significant difference in the lengths of the deletions or insertions could be observed (data not shown).

## Discussion

### Description of the ZFX and ZFY Regions

The nucleotide sequence of a PAC clone containing an insert of 136 kb encompassing the complete human ZFY gene was determined. This sequence was compared with 111 kb of sequence surrounding the X-linked homolog, which is likely to have diverged from ZFY at least 115 MYA (Janke et al. 1994). Although the genes are located in the strictly sex-linked parts of the chromosomes and are therefore not expected to undergo recombination, rare gene conversion has been suggested (Hayashida, Kuma, and Miyata 1992; Pamilo and Bianchi 1993). However, we were unable to identify transposable elements in homologous positions, suggesting that there has not been any recent recombination or gene conversion between ZFX and ZFY involving areas carrying transposable elements.

### Transposable Element Accumulation on the Sex Chromosomes

A build-up of interspersed repeats on the Y chromosome and in regions with low recombination was suggested by Charlesworth (1991) and has been observed in *Drosophila* (Charlesworth, Sniegowski, and Stephan 1994). In line with this suggestion, we see a significantly higher density of interspersed repeats in the ZFY (1.25/kb) and ZFX (1.41/kb) regions (table 1) compared with DNA sequences present in GenBank (0.84/kb) (Smit 1996) ( $P < 0.001$ ). However, the density of mobile elements in these regions does not correlate inversely with the opportunity for recombination, since ZFX displays a higher density of repeat elements than ZFY. Thus, in addition to chromosomal recombination frequencies, factors specific to the various types of repeat seem to be at work. For example, the ZFX locus carries a large number of Alu elements, while the average length of LINES at this locus is smaller than those in the other regions studied.

An interesting pattern is seen for RLEs. Twenty-one RLEs make up 8.9% of the ZFY locus, while only three elements compose 1.5% of the ZFX locus. Sequences collected from the entire human genome contain 5.6% RLEs (table 1). Although the proportion of DNA sequences made up of RLEs varies with GC content (Smit 1996), this cannot explain the difference observed, since the ZFY and ZFX regions both contain 39%–41% GC and the data representing the entire genome contain 36%–43% GC (Smit 1996). Thus, the ZFY region contains significantly more RLEs than does the genome in general ( $P < 0.001$ ).

RLEs are the result of retroviral germ line infections and are, after integration, transmitted in a Mendelian manner from generation to generation (Wilkinson, Mager, and Leong 1994). Evidence of an increased number of particular types of retroviral elements has previously been found on both the murine (Phillips et al. 1982; Eicher et al. 1989) and the human Y chromosomes (Kjellman, Sjögren, and Widegren 1995) by Southern analyses. It is interesting to consider possible reasons for the differential distributions of RLEs and other elements on the sex chromosomes and autosomes.

First, it has been suggested that the genes situated on the Y chromosome are disposable and in various stages of decay unless directly involved in sex determination (Charlesworth 1991; Graves 1995). This supposed lack of functional constraint has been assumed to allow more transposable elements to accumulate on the Y chromosome (Charlesworth 1991). However, it has recently been shown that the Y chromosome contains several housekeeping genes and genes expressed in the testes (Lahn and Page 1997). Thus, the Y chromosome may not be as devoid of functional constraint as previously supposed. Furthermore, since interspersed elements show no tendency to be more abundant in the ZFY region than in the ZFX region (table 1), functional constraint is an unlikely explanation for the difference in RLE abundance.

Second, as the strictly sex-linked region of the Y chromosome never recombines, it has no opportunity to rid itself of transposable elements through allelic gene conversion. However, the X chromosome and the autosomes can recombine during two thirds and all of their germ line passages, respectively. Therefore, if recombinational elimination were the sole factor explaining the differences in RLE number, we would expect more elements in both the ZFY and the ZFX regions compared with the autosomes. However, we observe more (albeit not significantly more) RLEs on the autosomes than on the X chromosome. Furthermore, recombination would again be expected to affect all transposable elements equally and thus cannot explain the absence of a difference in content observed for the other elements. It therefore seems that differences in recombination rate cannot fully explain the observed RLE distribution. They might, however, partly explain why more transposable elements occur at both ZFX and ZFY compared with the autosomes.

Third, since retroviruses (with the exception of lentiviruses) require host cell replication in order to integrate into the host genome (Brown 1997), the larger number of cell divisions in the male germ line may explain the accumulation of RLEs on the Y chromosome, as well as the fact that they are less numerous at the ZFX locus than on the autosomes (table 1). Thus, the majority of the RLEs in the human genome may have integrated during male germ cell development.

### Male-Driven Substitutional Evolution

In order to investigate the rate of point mutational evolution on the sex chromosomes, it is necessary to

compare sequences, which, as far as is known, are selectively neutral and ancestrally related and which have not undergone gene conversion (Shimmin et al. 1993). Alu elements are primate-specific SINES which fulfill these criteria. Furthermore, Alu elements are found on all chromosomes in sufficient numbers for statistical comparisons to be performed. Only LINEs are present at copy numbers similar to those of Alu elements. However, a large portion of LINEs is deleted during integration, and their time of integration is less well defined than those of Alu classes, rendering them less than optimal for comparisons among elements.

We compared the sequences of Alu elements belonging to each of three Alu classes between the ZFY and ZFX loci, as well as autosomal regions. The divergence of Alu elements from the relevant consensus sequences indicates that the rate of substitution on the Y chromosome is greater than that on the autosomes, which, in turn, is greater than that on the X chromosome (fig. 3A). When we calculate the male-to-female mutation rate ratio,  $\alpha$ , for each class of Alu element, using comparisons between both of the sex chromosomes and the autosomes, the estimates fall between 1.6 and 2.2 (table 2). When CpG sites in the consensus sequences are excluded,  $\alpha$  values increase. Thus, there is no evidence for the idea that differences in methylation between the male and female germ lines would be responsible for the differences in mutational rates. This is supported by a recent study of the methylation status at two loci in mature male and female gametes, which were found to be equally and highly methylated (El-Maarri et al. 1998). In fact, the near saturation of substitutions at CpG sites obscures to some degree the differences in rates between the different sequences. Thus, the exclusion of CpG sites can be expected to yield a more reliable estimate of the male-to-female mutation rate ratio. Moreover, the estimates for the AluY elements vary substantially, probably due to the low number of such elements observed and a larger relative variation in their time of retrotransposition. Therefore, in order to arrive at an overall estimate of  $\alpha$ , non-CpG sites in the AluJ and AluS elements were used. This resulted in an  $\alpha$  of 2.5 (weighted for the numbers of Alu's in each class). This value is lower than previous estimates for primates, which vary from 3 to 6 (Shimmin, Chang, and Li 1993, 1994; Huang et al. 1997; Agulnik et al. 1998; Anagnostopoulos et al. 1999). It is hard to judge the importance of this, since estimates of  $\alpha$  seem to vary substantially between loci (Hurst and Ellegren 1998). In principle, our calculations improve on others by comparing a larger number of nucleotides. Furthermore, whereas other methods have generally only compared the two sex chromosomes with each other, our method allows us to include comparisons between the autosomes and both sex chromosomes and thus to potentially better gauge the ratio. However, the fact that Alu elements of the same class have integrated at slightly different times will add to the uncertainty of our estimate. Eventually, only more sequence data, preferably from

closely related species, will allow more accurate estimates of the relative mutation rates in the male and female germ lines.

In summary, a significantly lower density of transposable elements is observed on the autosomes than on the sex chromosomes. The fact that the substitution rates of Alu elements on the sex chromosomes and the autosomes correlate with the times the chromosomes spend in the male germ line suggests that the number of DNA replications experienced is a major factor contributing to the mutation rate differences. Interestingly, the comparison between the ZFY and ZFX loci shows that this may apply not only to the accumulation of point mutations, but also to the accumulation of RLEs in the human genome.

### Supplementary Material

The sequence data described in this paper have been submitted to GenBank data library under accession number AF114156.

### Acknowledgments

We thank V. Benes, D. Thomas, and H. Kaessmann for help with shotgun libraries; T. Haaf for FISH; C. Kilger for library screening; K. Bauer, C. Baumann, B. Stiening, S. Ridgway, and V. Wiebe for technical assistance; and G. Weiss and especially S. Zöllner for statistical advice. This work was supported by the Deutsche Forschungsgemeinschaft and the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie. R.E. received a fellowship from Stiftelsen för internationalisering av högre utbildning och forskning (STINT, Stockholm). R.E. and J.F.W. contributed equally to this work.

### LITERATURE CITED

- AGULNIK, A. I., C. E. BISHOP, J. L. LERNER, S. I. AGULNIK, and V. V. SOLOVYEV. 1997. Analysis of mutation rates in the SMCY/SMCX genes shows that mammalian evolution is male driven. *Mamm. Genome* **8**:134–138.
- AGULNIK, A. I., A. ZHARKIKH, H. BOETTGER-TONG, T. BOURGERON, K. MCELREAVEY, and C. E. BISHOP. 1998. Evolution of the DAZ gene family suggests that Y-linked DAZ plays little, or a limited, role in spermatogenesis but underlines a recent African origin for human populations. *Hum. Mol. Genet.* **7**:1371–1377.
- ANAGNOSTOPOULOS, T., P. M. GREEN, G. ROWLEY, C. M. LEWIS, and F. GIANNELLI. 1999. DNA variation in a 5-Mb region of the X chromosome and estimates of sex-specific/type-specific mutation Rates. *Am. J. Hum. Genet.* **64**:508–517.
- BATZER, M. A., P. L. DEININGER, U. HELLMANN-BLUMBERG, J. JURKA, D. LABUDA, C. M. RUBIN, C. W. SCHMID, E. ZIETKIEWICZ, and E. ZUCKERKANDL. 1996. Standardised nomenclature for Alu repeats. *J. Mol. Evol.* **42**:3–6.
- BROWN, P. O. 1997. Integration. Pp. 161–204 in J. M. COFFIN, S. H. HUGHES, and H. E. VARMUS, eds. *Retroviruses*. Cold Spring Harbor Laboratory Press, New York.
- CARROLL, R. L. 1988. *Vertebrate paleontology and evolution*. W. H. Freeman, New York.
- CHANG, B. H.-J., D. HEWETT-EMMETT, and W.-H. LI. 1996.

- Male-to-female ratios of mutation rate in higher primates estimated from intron sequences. *Zool. Stud.* **35**:36–48.
- CHARLESWORTH, B. 1991. The evolution of sex chromosomes. *Science* **125**:1030–1033.
- CHARLESWORTH, B., P. SNIĘGOWSKI, and W. STEPHAN. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**:215–220.
- DEININGER, P. L., M. A. BATZER, C. A. I. HUTCHISON, and M. H. EDGELL. 1992. Master genes in mammalian repetitive DNA amplification. *Trends Genet.* **8**:307–311.
- DRISCOLL, D. J., and B. R. MIGEON. 1990. Sex difference in methylation of single-copy genes in human meiotic germ cells: implications for X chromosome inactivation, parental imprinting, and origin of CpG mutations. *Somat. Cell Mol. Genet.* **16**:267–282.
- EICHER, E. M., K. W. HUTCHISON, S. J. PHILLIPS, P. K. TUCKER, and B. K. LEE. 1989. A repeated segment on the mouse Y-chromosome is composed of retroviral-related, Y-enriched and Y-specific sequences. *Genetics* **122**:181–192.
- EL-MAARRI, O., A. OLEK, B. BALABAN, M. MONTAG, H. VAN DER VEN, B. URMAN, K. OLEK, S. H. CAGLAYAN, J. WALTER, and J. OLDENBURG. 1998. Methylation levels at selected CpG sites in the factor VIII and FGFR3 genes, in mature female and male germ cells: implications for male-driven evolution. *Am. J. Hum. Genet.* **63**:1001–1008.
- ERICKSON, L. M., and N. MAEDA. 1995. A new family of retroviral long terminal repeat elements in the human genome identified by their homologies to an element 5' to the spider monkey haptoglobin gene. *Genomics* **27**:531–534.
- GRAVES, J. A. M. 1995. The origin and function of the mammalian Y chromosome and Y-borne genes—an evolving understanding. *Bioessays* **17**:311–320.
- HAYASHIDA, H., K. KUMA, and T. MIYATA. 1992. Interchromosomal gene conversion as a possible mechanism for explaining divergence patterns of ZFY-related genes. *J. Mol. Evol.* **35**:181–183.
- HUANG, W., B. H. J. CHANG, X. GU, D. HEWETT-EMMETT, and W. LI. 1997. Sex differences in mutation rate in higher primates estimated from AMG intron sequences. *J. Mol. Evol.* **44**:463–465.
- HULTMAN, T., S. BERGH, T. MOKS, and M. UHLEN. 1991. Bidirectional solid-phase sequencing of in vitro-amplified plasmid DNA. *Biotechniques* **10**:84–93.
- HURST, L. D., and H. ELLEGREN. 1998. Sex biases in the mutation rate. *Trends Genet.* **14**:446–452.
- JANKE, A., G. FELDMAIER-FUCHS, W. K. THOMAS, A. VON HAESELER, and S. PÄÄBO. 1994. The marsupial mitochondrial genome—the evolution of placental mammals. *Genetics* **137**:243–256.
- JURKA, J., and A. MILOSAVLJEVIC. 1991. Reconstruction and analysis of human Alu genes. *J. Mol. Evol.* **32**:105–121.
- KAPITANOV, V., and J. JURKA. 1996. The age of Alu subfamilies. *J. Mol. Evol.* **42**:59–65.
- KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**:111–120.
- KJELLMAN, C., H.-O. SJÖGREN, and B. WIDEGREN. 1995. The Y chromosome: a graveyard for endogenous retroviruses. *Gene* **161**:163–170.
- LAHN, B. T., and D. C. PAGE. 1997. Functional coherence of the human Y chromosome. *Science* **278**:675–680.
- LUOH, S. W., and D. C. PAGE. 1994. The structure of the Zfx gene on the mouse X chromosome. *Genomics* **19**:310–319.
- MCVEAN, T. M., and L. D. HURST. 1997. Evidence for a selectively favourable reduction in the mutation rate of the X chromosome. *Nature* **386**:388–392.
- MAHAFFEY, C. L., J. K. BAYLERAN, G. Y. YEH, T. C. LEE, D. C. PAGE, and E. M. SIMPSON. 1997. Intron/exon structure confirms that mouse zfy1 and zfy2 are members of the zfy gene family. *Genomics* **41**:123–127.
- MIGHELL, A. J., A. F. MARKHAM, and P. A. ROBINSON. 1997. Alu sequences. *FEBS Lett.* **417**:1–5.
- MIYATA, T., H. HAYASHIDA, K. KUMA, K. MITSUYASA, and T. YASUNAGA. 1987. Male-driven molecular evolution: a model and nucleotide sequence analysis. *Cold Spring Harb. Symp. Quant. Biol.* **52**:863–867.
- NELSON, D. L., A. BALLABIO, F. CREMERS, A. P. MONACO, and D. SCHLESSINGER. 1995. Report of the sixth international workshop on X chromosome mapping. *Cytogenet. Cell Genet.* **71**:307–342.
- NORTH, M., C. SARGENT, J. O'BRIEN, K. TAYLOR, J. WOLFE, N. A. AFFARA, and M. A. FERGUSON-SMITH. 1991. Comparison of ZFY and ZFX gene structure and analysis of alternative 3' untranslated regions of ZFY. *Nucleic Acids Res.* **19**:2579–2586.
- PAGE, D. C., R. MOSHER, E. M. SIMPSON, E. M. FISHER, G. MARDON, J. POLLACK, B. MCGILLIVRAY, A. DE LA CHAPPELLE, 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.
- 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.
- PAULSON, K. E., N. DEKA, C. W. SCHMID, R. MISRA, C. W. SCHNIDLER, M. G. RUSH, L. KADYK, and L. LEINWAND. 1985. A transposon-like element in human DNA. *Nature* **316**:359–361.
- PHILLIPS, S. J., E. H. BIRKENMEIER, R. CALLAHAN, and E. M. EICHER. 1982. Male and female mouse DNAs can be discriminated using retroviral probes. *Nature* **297**:241–243.
- SCHNEIDER-GÄDICKE, A., P. BEER-ROMERO, L. G. BROWN, R. NUSSBAUM, and D. C. PAGE. 1989. ZFX has a gene structure similar to ZFY, the putative human sex determinant, and escapes X inactivation. *Cell* **57**:1247–1258.
- SHIMMIN, L. C., B. H.-J. CHANG, D. HEWETT-EMMETT, and W.-H. LI. 1993. Potential problems in estimating the male-to-female mutation rate ratio from DNA sequence data. *J. Mol. Evol.* **37**:160–166.
- SHIMMIN, L. C., B. H.-J. CHANG, and W.-H. LI. 1993. Male-driven evolution of DNA sequences. *Nature* **362**:745–747.
- . 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.
- SMIT, A. F. A. 1993. Identification of a new, abundant superfamily of mammalian LTR-transposons. *Nucleic Acids Res.* **21**:1863–1872.
- . 1996. The origin of interspersed repeats in the human genome. *Curr. Opin. Genet. Dev.* **6**:743–748.
- SONNHAMMER, E. L. L., and R. DURBIN. 1995. A dot-matrix program with dynamic threshold control suited for ge-

- omic DNA and protein sequence analysis. *Gene* **167**:1–10.
- WATSON, J. M., J. A. SPENCER, A. D. RIGGS, and J. A. GRAVES. 1991. Sex chromosome evolution: platypus gene mapping suggests that part of the human X chromosome was originally autosomal. *Proc. Natl. Acad. Sci. USA* **88**:11256–11260.
- 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.
- WILKINSON, D. A., D. L. MAGER, and J. C. LEONG. 1994. Endogenous human retroviruses. Pp. 465–535 *in* J. A. LEVY, ed. *The Retroviridae*. Vol. 3. Plenum, New York.
- ZEHETNER, G., and H. LEHRACH. 1994. The reference library system—sharing biological material and experimental data. *Nature* **367**:489–491.

FUMIO TAJIMA, reviewing editor

Accepted January 28, 2000