Selection Against Deleterious LINE-1-Containing Loci in the Human Lineage

Stéphane Boissinot, Ali Entezam, and Anthony V. Furano

Section on Genomic Structure and Function, Laboratory of Molecular and Cellular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland

We compared sex chromosomal and autosomal regions of similar GC contents and found that the human Y chromosome contains nine times as many full-length (FL) ancestral LINE-1 (L1) elements per megabase as do autosomes and that the X chromosome contains three times as many. In addition, both sex chromosomes contain a ca. twofold excess of elements that are >500 bp but not long enough to be capable of autonomous replication. In contrast, the autosomes are not deficient in short (<500 bp) L1 elements or SINE elements relative to the sex chromosomes. Since neither the Y nor the X chromosome, when present in males, can be cleared of deleterious genetic loci by recombination, we conclude that most FL L1s were deleterious and thus subject to purifying selection. Comparison between nonrecombining and recombining regions of autosome 21 supported this conclusion. We were able to identify a subset of loci in the human DNA database that once contained active L1 elements, and we found by using the polymerase chain reaction that 72% of them no longer contain L1 elements in a representative of each of eight different ethnic groups. Genetic damage produced by both L1 retrotransposition and ectopic (nonallelic) recombination between L1 elements could provide the basis for their negative selection.

Introduction

We recently showed that the human Ta L1 (LINE-1, long interspersed repeated DNA) family of retrotransposons (Skowronski, Fanning, and Singer 1988; also known as L1PA1 [Smit et al. 1995]) arose ~4 MYA in humans soon after the human and chimpanzee lineages split (Boissinot, Chevret, and Furano 2000). L1PA1 (Ta) has been amplifying in humans at about the same rate per generation as recently evolved active rodent L1 families (Cabot et al. 1997; DeBerardinis et al. 1998; Saxton and Martin 1998). Overall, about 50% of Ta insertions are polymorphic across human populations, and more than 90% of the inserts generated by Ta-1d, the youngest subfamily of Ta, are polymorphic (Boissinot, Chevret, and Furano 2000). These numbers attest to the recent and ongoing evolution and retrotransposition of the L1 Ta family in humans. Indeed, all 12 documented genetic defects caused by L1 insertions (some of which occurred in utero) are due to Ta, with 10 being caused by Ta-1d insertions (quoted in Kimberland et al. 1999; Boissinot, Chevret, and Furano 2000).

While novel L1PA1 (Ta) L1 insertions are increasing the genetic diversity of humans, a more far reaching question is that of how large a genetic load L1 activity imposes on its host. A clue that it does was our finding that about 35% of Ta elements are full length (FL L1), as compared with the previous estimates of only about 5%–10% FL for total genomic, and thus mostly ancestral, L1 elements (Fanning and Singer 1987). A possible explanation for this difference is that most of the ancestral FL L1-containing loci were deleterious and were soon lost as a result of negative selection. This implies that either insufficient time has elapsed to clear delete-rious FL Ta-containing loci or they are being generated as fast as they are cleared.

To determine if most of the ancestral FL L1-containing loci were cleared from the human genome, we compared the fractions of FL L1 elements on autosomes and on the X and Y chromosomes. If FL ancestral L1 elements were deleterious and subject to purifying selection, we would expect them to have been cleared from autosomes but not from the Y chromosome. This is because nonrecombining regions, which constitute most of the Y chromosome, accumulate deleterious mutations at a higher rate than do recombining regions. The inevitable increase in deleterious mutations in the absence of recombination has been termed Muller's ratchet (Felsenstein 1974) and has been verified experimentally (Rice 1994; reviewed by Hurst 1999). When present in the male, the X chromosome also cannot recombine, and thus the fraction of FL L1 would be intermediate between the Y chromosome and autosomes.

We show here that only ~8.5% of the autosomal L1 elements present in four ancestral non-Ta families (L1PA2–L1PA5; Smit et al. 1995) are FL. In contrast, ~30% of the members of these ancestral families residing on the Y chromosome are FL, and this value for the X chromosome is ~16.5%. Thus, it appears that FL L1-containing loci imposed a significant enough genetic load on its host to have been lost from the human genome.

Materials and Methods
Sequence Analysis

Human genomic sequences from chromosomes 21 (84 entries), 22 (156 entries), X (108 entries), and Y (64 entries) were obtained from GenBank. Every entry longer than 80 kb was analyzed using the program REPEAT-MASKER, version 3.0 (A. F. A. Smit and P. Green, http://ftp.genome.washington.edu/RM/RepeatMasker.html). Because the genomic distribution of L1 elements differs between regions with different GC contents (Smit 1999), we analyzed GenBank entries with similar GC contents. Since

analysis of entries that were either 30%–40% or 40%–45% G+C gave congruent results, we pooled these results. The total number of L1 elements and the number of FL L1s belonging to families L1PA1 (Ta) to L1PA5 were counted for each entry. These families have successively amplified beginning ~25 MYA. We limited our analysis to these five families because they were less likely to be interrupted by more recent internal insertions than older families. FL L1s were initially aligned using the GCG Pileup program (Wisconsin Package, version 10.0, Genetics Computer Group, Madison, Wis.), and the alignment was refined by hand using the GCG Seqlab editor. Phylogenetic analysis and pairwise distances were calculated using PAUP*, version 4.0b4 (Swofford 1998).

Analysis of L1-Containing Loci by PCR

The states of L1-containing loci in different human populations were determined as described earlier (Boissinot, Chevret, and Furano 2000).

Results

An Excess of FL L1 Elements on the Sex Chromosomes

Figures 1*A* and *B* show respectively, a diagram of an FL L1 element and that the modern human-specific Ta (L1PA1) L1 family is the latest version of a single lineage of L1 elements that extends back for at least 50 Myr in primate history. Figure 1*C* shows a more detailed phylogeny of the 3′ 500 bp of L1PA1–L1PA5 based on the members of these families from the selected regions of the sex chromosome and autosomes 21 and 22 (see *Materials and Methods*).

Table 1 shows a significant ($P < 0.05$) excess of L1 elements on the sex chromosomes compared with the autosomes. Overall, the Y chromosome contains ~2.2 more L1 elements per Mb than do autosomes, and the X chromosome contains ~1.6 as many. More importantly, there is a marked difference between the genomic distribution as a function of element length. Most striking is the excess of FL elements on the Y chromosome as compared with either the X chromosome or the autosomes; the Y chromosome contains three times as many FL elements per Mb as the X chromosome and nine times as many as the autosomes. All of these differences are statistically significant. Consequently, the fractions of L1s which are FL are very different between chromosomes: ~32% of L1 elements on the Y chromosome are FL, as compared with ~15% on the X chromosome and ~7% on the autosomes.

The 32% FL on the Y chromosome for the combined L1PA1–L1PA5 families shown in table 1 is similar to the value of about 35% FL for the currently amplifying L1PA1 (Ta) family (Boissinot, Chevret, and Furano 2000). Table 2 shows that the chromosomal regions examined contained 17 L1PA1 (Ta) elements (8 FL) of a total estimated genomic copy number of 700 L1PA1 members (Boissinot, Chevret, and Furano 2000). Therefore, the few L1PA1 elements included in table 1 have not materially skewed the results. Table 2 shows that the proportion of FL L1PA2–
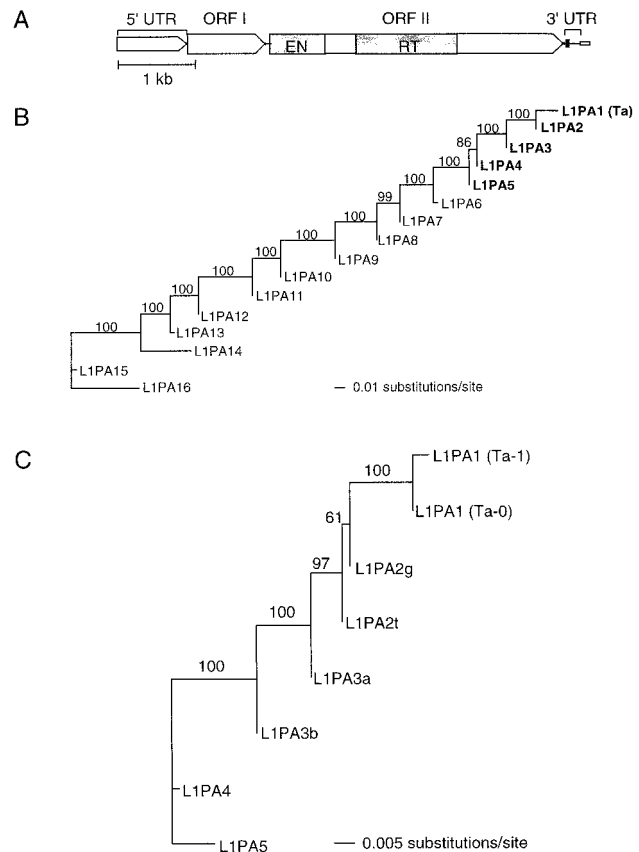

FIG. 1.—*A,* A typical human L1 element. The 5′ untranslated region (UTR) has a regulatory function; open reading frame (ORF) I encodes an RNA-binding protein; ORF II encodes the L1 cDNA replicase containing highly conserved endonuclease (EN) and reverse transcriptase (RT) domains; the 3′ UTR contains a conserved G-rich polypurine motif (black rectangle). Genomic copies of L1 usually end in an A-rich stretch (open rectangle, reviewed in Furano 2000). *B,* Maximum-likelihood tree (using the default settings in PAUP* version 4.0; see *Materials and Methods*) built on the consensus sequences of the 3′ UTR of 16 primate L1 families (Smit et al. 1995). The numbers indicate the percentages of times that the node was represented in 1,000 bootstrap replicates of the data. *C,* Maximum-likelihood tree built on the consensus sequences derived from the 3′ 500 bp of the L1PA1–L1PA5 elements (both full-length [FL] and long (>500 bp) non-FL elements) enumerated in tables 1 and 2. This region includes the last 294 bp of ORF II (which contains two of the characters that distinguish the Ta-0 and Ta-1 subfamilies; Boissinot, Chevret, and Furano 2000) and all of the 3′ UTR. The consensus sequences for L1PA2 and L1PA3 also included seven FL members of each of these families derived from other regions of the genome. The consensus sequences of the L1PA1 (Ta) subfamilies, Ta-0 and Ta-1 (Boissinot, Chevret, and Furano 2000), were indistinguishable from those derived from genomewide L1PA1 elements (Boissinot, Chevret, and Furano 2000). In this region, L1PA2t and L1PA2g differ by a single "t" (ancestral) to "g" change at position 6397 in the 3′ UTR. The numbers above the nodes indicate the percentages of the time the labeled node was present in 1,000 bootstrap replicates of the data.

L1PA5 elements on the Y chromosome ranges from 27.3% to 36.8%. These values are 2%–35% for the X chromosome and 4%–16% for the autosomes. The higher variance for the X chromosome and the autosomes is expected from the smaller numbers of FL elements on these chromosomes (table 2). The ~31% of Y L1PA2–L1PA5 elements that are FL presumably represents a minimal estimate of

Table 1
Copy Numbers of L1 Elements on the Sex Chromosomes and the Autosomes

	L1 COPIES					STATISTICAL SIGNIFICANCE							nr
			Autosomes			Y				X			
	Y	X	Total	nr	r	X	aut	nr	r	aut	nr	r	r
Entries	52	75	99	24	75								
Mb	8	10.6	29.7	8	21.7								
L1 (total)	180	175	310	122	188								
FL	58	26	23	11	12								
% FL	32	15	7.4	9	6.3								
L1 per Mb													
Total	22.5	16.5	10.4	15.3	8.6	0.02	0.000	0.011	0.000	0.000	NS	0.000	0.000
FL	7.2	2.5	0.8	1.4	0.6	0.002	0.000	0.004	0.000	0.004	NS	0.007	NS
non-FL	15.2	14	9.7	13.9	8.1	NS	0.001	NS	0.000	0.004	NS	0.000	0.000
>500 bp ...	10.6	7.8	4.5	7.8	3.3	NS*	0.000	NS**	0.000	0.000	NS	0.000	0.000
<500 bp	4.4	6.3	5.1	6.1	4.7	NS	NS	NS	NS	NS	NS	NS	NS
Alu (total)	2,800	2,996	9,944			0.000	NS						
Alu per Mb	350	282.6	334.8										

NOTE.—The L1 copies from families L1PA1–L1PA5 in regions of the Y and X chromosomes and autosomes 21 and 22 that ranged from 30% to 45% G+C were determined as described in *Materials and Methods* on the indicated number of GenBank entries (i.e., entries designated by different accession numbers). The average G+C contents of the DNA examined here were 39.5% for the Y and X chromosomes and 39.1% for the autosomes (aut). The autosomal entries have been divided into recombining (r) and nonrecombining (nr) regions, the latter from an 8-Mb centromeric region of chromosome 21 that recombines at low frequency in males (Hattori et al. 2000). The *F* statistics for the differences between the copies of L1 on the Y and X chromosomes and the autosomes was calculated as described in Erlandsson, Wilson, and Paalo (2000). NS indicates that differences were not significant at the 95% confidence level.

* *P* = 0.084.

** *P* = 0.105.

the basal efficiency at which FL elements were generated by these families.

There is also a moderate but statistically significant excess (~1.6) of non-FL L1s on the X and Y chromosomes compared with the autosomes. However, in contrast to the FL elements, there is no significant difference between the X and the Y chromosomes (table 1). The difference in the numbers of non-FL L1 elements between the sex chromosomes and the autosomes decreases with the size of the element; those that are >500 bp are ~2.4 times as abundant on the Y chromosome and ~1.7 times as abundant on the X chromosome as on the autosomes. In contrast, there is no difference between the sex chromosomes and autosomes for L1 elements that are <500 bp. In addition, the sex chromosomes do not contain an excess of Alu SINE elements (which are about 300 bp) relative to autosomes (table 1). In fact, the region of the X chromosome examined here contains a small but statistically significant deficit of Alu relative to both the Y chromosome and the autosomes.

The Sex-Chromosomal L1 Elements Are Typical Members of Their Subfamilies

The L1 elements enumerated in tables 1 and 2 were assigned to their respective L1PA subfamilies by Repeat Masker based on diagnostic characters in the 3′ untranslated region (UTR) (Smit et al. 1995). Although characters elsewhere in the element (e.g., ORF II) are confirmatory of these designations, the highest resolution of the primate L1 families is generally based on the 3′ UTR (Smit et al. 1995; Boissinot, Chevret, and Furano 2000; unpublished data). To confirm the Repeat Masker family assignments and to be sure that they did not include yet unrecognized subfamilies that might be unusually enriched in FL elements, we performed a phylogenetic analysis of the FL L1PA1–L1PA5 elements.

Figure 2 shows a single neighbor-joining tree built on the 3′ UTR of the FL L1PA1–L1PA5 elements. The FL Y elements were indistinguishable from their counterparts on the X chromosome and the autosomes except

Table 2
Fractions of Full-Length (FL) Elements in L1 Families of Different Ages

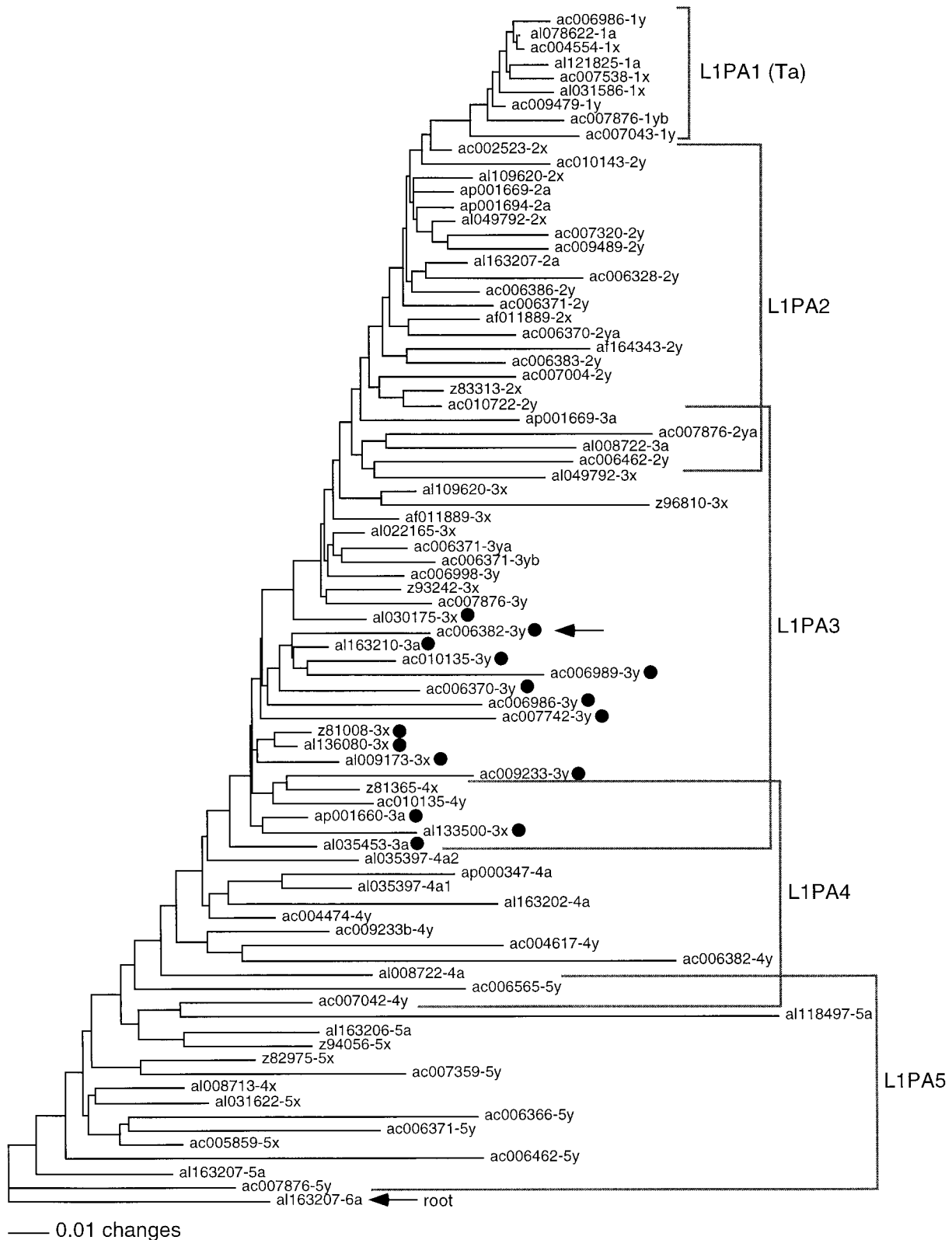|	L1 COPIES								
	Y			X			Autosomes		
	Total	FL	%	Total	Fl	%	Total	FL	%
LIPA1	4	4	100	5	3	60	8	1	12.5
LIPA2	38	14	36.8	17	6	35	49	8	16.3
LIPA3	39	12	30.7	57	11	19.3	62	5	8.1
LIPA4	55	15	27.3	45	1	2	92	5	5.4
LIPA5	44	13	29.6	51	5	9.8	99	4	4
LIPA2–LIPA5 (mean ± SE)			31.1 ± 2.0			16.5 ± 7.1			8.45 ± 2.8

FIG. 2.—Phylogenetic tree of full-length (FL) L1 elements. The 3′ untranslated region (UTR) of the FL elements that we identified on the examined regions of the sex chromosomes and autosomes 21 and 22 (see table 1) was used to construct a neighbor-joining tree (Saitou and Nei 1987). Pairwise distances were corrected for multiple substitutions using Kimura's 2ρ method (Kimura 1980). Each element is labeled in turn by its accession number; 1–5 for families L1PA1–L1PA5; "y," "x," or "a," for its location on the sex chromosomes or autosomes; and in some cases an added letter if more than one L1 element of a given family was present in the same GenBank record. The L1PA3 elements labeled with black dots are members of the L1PA3b subfamily that we identified here (see text). The L1PA3b element indicated by the arrow contains an L1PA3b "body" but an L1PA3a 5′ UTR (see text). The L1PA1 element, al078622-1a, is not included in table 1 because it was present in a region that was >45% G+C (see text and table 1). We included it here for comparison. The tree was rooted on an L1PA6 element (Smit et al. 1995).
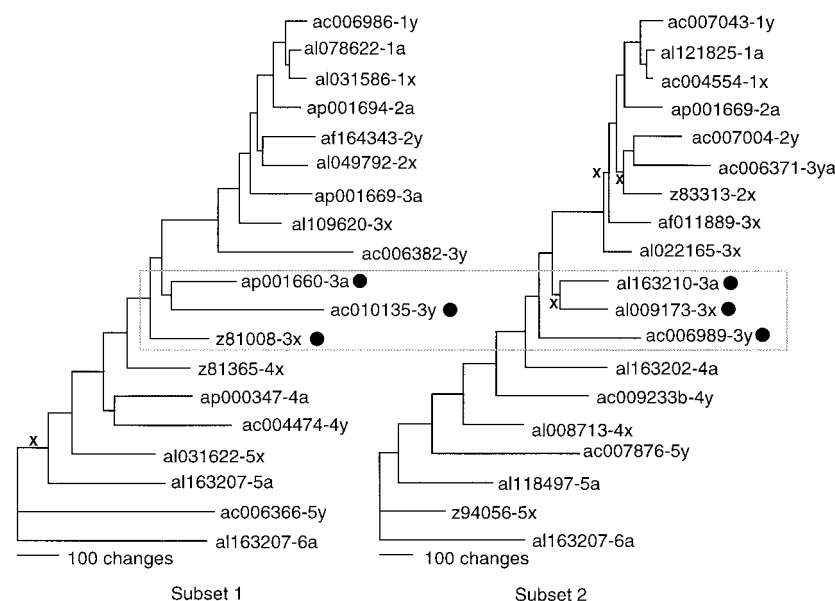
FIG. 3.—Phylogenetic analysis of full-length L1 elements. Maximum-parsimony trees were built on the entire length of two arbitrarily selected subsets of the elements shown in figure 2. All of the nodes except those indicated by an "x" were present in 76% or more of 1,000 bootstrap replicates of the data. The long branch separating L1PA3b (gray box) and older families from L1PA3a and younger families is due in large part to the 124-bp deletion in the 5′ untranslated region that the latter families share. Element designations are as in figure 2.

for their longer branch lengths. The latter reflects the greater divergence of L1 elements on the Y chromosome than on the other chromosomes (unpublished data). Maximum-parsimony analysis of the entire L1 sequence of arbitrarily selected subsets of FL L1PA1–L1PA5 elements produced similar results (fig. 3). Figures 1C, 2, and 3 also show that the L1PA3 family consists of two subfamilies: L1PA3b (the older subfamily) and L1PA3a. This division is supported by numerous characters, including an ancestral 124-bp region of the 5′ UTR that is present in L1PA3b and the older families but is deleted from the L1PA3a subfamily and the younger families (results not shown).

The congruence between the sex-chromosomal and autosomal members of the L1PA1–L1PA5 families (including the subfamily structure of L1PA3) was also revealed by analysis of the far larger data set of ≥500-bp elements (results not shown). Additionally, out of the ~400 (FL plus ≥500 bp) L1PA1–L1PA5 elements on the Y and X chromosomes and the autosomes, only 4 were misassigned by Repeat Masker, and these only to closely related families. Therefore, with the exception of our discernment of the L1PA3a and L1PA3b subfamilies, the L1PA1–L1PA5 sequences examined here displayed all of the diagnostic characters in the 3′ UTR that were used to originally define these subfamilies. Thus, the L1 elements on either sex chromosome, FL or not, are typical representatives of their counterparts on the autosomes.

The Density of FL L1 Elements on Autosomes Is Correlated with Recombination Rate

Since the sex chromosomes do not contain an excess of short (<500 bp) L1 elements, the excess of FL and long non-FL (>500 bp) L1s is not likely due to preferential insertion of L1 on the sex chromosomes. Perhaps longer or FL copies are more efficiently generated when sex chromosomes rather than autosomes are the targets for insertion. However, to account for our data, this process would have to be more efficient on the Y chromosome than on the X chromosome and more efficient on certain regions of the autosomes than on others (see below).

A more plausible explanation is that FL L1 elements, and, to a lesser extent, long non-FL (>500 bp) elements, are deleterious. Thus, they accumulated on the Y chromosome (and, to a lesser extent, on the X chromosome) because the removal of deleterious mutations by recombination is not available to the Y chromosome or to the X chromosome when it resides in males. On the other hand, the autosomal FL and long non-FL L1 would be more susceptible to removal by recombination than their counterparts on the sex chromosomes. If true, autosomal regions with abnormally low recombination rates should contain more of the putatively deleterious FL L1 inserts than regions with normal recombination rates. The 8 Mb of DNA located near the centromere of chromosome 21 has a very low recombination rate in males (fig. 5 in Hattori et al. 2000) but not in females. Therefore, this "pseudo-X" region of chromosome 21 should be similar to the X chromosome with respect to its L1 composition.

Table 1 shows that the L1 composition of the 8 Mb with a low recombination rate in males resembles the L1 composition of the X chromosome. These two regions of the genome do not significantly differ with respect to the number per megabase of FL, long (>500 bp), and short (<500 bp) L1 elements (table 1). Conversely, the X chromosome contains a statistically significant excess of FL and long non-FL elements over

Table 3
Loss of Active L1 Elements from Human Populations

L1-CONTAINING LOCUS	FAMILY	LENGTH (bp)	CHRO-MOSOME	LENGTH OF TRANSDUCED SEQUENCE	ORIGINATING LOCUS	CHRO-MOSOME	STATUS OF ORIGINATING LOCUS IN:		
							GenBank	Eight Human Populations	PCR PRIMER PAIR[a]
AC004499[b]	Ta-0	125	20	58	AC024387	14	Empty	Empty	Pr1
AL030998#2[b,c] . . .	Ta-0	1,277	X	157	AC022031	18	Empty	Empty	Pr2
AC004491[b]	Ta-0	1,502	7	132	AC022440	5	Empty	Empty	Pr3
AC004200[b,c]	Ta-1d	6,018	6	123	AP001005	18	Empty	Empty	Pr4
AC003689	Ta-1	1,986	11	44	AC004000	X	Occupied	Polymorphic	FL43
AL049588	Ta-0	2,893	X	94	AC008893	5	Empty	Empty	Pr5
AC002980[b,c]	Ta-1d	6,212	X	174	AL118519	6	Occupied	Fixed	Pr6
AC007966-B	LIPA2	1,320	2	31	AC022697	4	Occupied	Fixed	Pr7
AL021069[b,c]	pre-Ta[d]	6,016	1	178	AC007159	3	Empty	Empty	Pr8
AC006989	LIPA2	705	13	216	AL162378	13	Empty	Empty	Pr9
AL133386	Ta	159	6	26	AC022325	4	Empty	Empty	Pr10
AC003080[b,c,e]	Ta-0	6,016	7	458 (165)	Z97185	1	Empty	Empty	Pr11
AC010277	Ta-0	NA	5	948	AC003080	7	Occupied	Fixed	Ta0-2
AL049842	Ta-1	656	6	63	AC019165	?	Empty	Empty	Pr12
AL022153[c]	LIPA2	3,464	X	298	AC016550	5	Empty	Empty	Pr13
AC004976[c]	LIPA2	604	7	465	AL353652	13	Empty	Empty	Pr14
AC002069[b,c]	LIPA3	6,022	7	368	AC005019	7	Empty	Empty	Pr15
U09115	Ta-0	1,968	X	489	U09116	1	Occupied	Polymorphic	[f]
AC00060[c,e]	LIPA2	6,015	7	138 (33)	AC008965	5	Empty	[g]	
Z72001	LIPA3	6,019	X	514	AC027626	4	Empty	[g]	

[a] The sequences of these primer pairs are available in the file Primer-pairs.transd, available by anonymous ftp from helix.nih.gov in pub/avf.

[b] Goodier, Ostertag, and Kazazian (2000).

[c] Pickeral et al. (2000).

[d] Pre-Ta elements have an "ACg" in the 3′ untranslated region instead of an "ACA," diagnostic of Ta (Skowronski, Fanning, and Singer 1988; Boissinot, Chevret, and Furano 2000).

[e] These elements are the products of two sequential transpositions. The number in parentheses gives the size of the first transduced sequence, and it is the locus that donated this sequence that is empty in GenBank. The locus that donated the second transduced sequence is not represented in the current version of GenBank.

[f] Holmes et al. (1994).

[g] We were unable to design suitable PCR primers for these loci.

the recombining region of chromosome 21 (table 1). Therefore, a difference between recombination rates could explain the loss of potentially active FL L1 elements from most autosomal regions and their consequent relative enrichment on the sex chromosomes.

This conclusion is even more compelling when the low (in males) recombining 8-Mb region of chromosome 21 is compared with the adjacent 8 Mb. Not only do these regions share similar GC contents and densities of Alu and short L1 elements, but they also contain the same number of both known and predicted genes. At the DNA sequence level, they differ only in their numbers of FL L1 elements: the low recombining region contains 11, and the normal recombining region contains 2 (table 2 and Hattori et al. 2000).

The Absence of Once-Active L1-Containing Loci in Present Populations

One way to corroborate the elimination of active L1 elements is to demonstrate that loci which once contained them were subsequently lost from the population. About 10% of L1 insertions include adjacent 3′ non-L1 flanking sequence that was transduced during retrotransposition (Moran, DeBerardinis, and Kazazian 1999; Goodier, Ostertag, and Kazazian 2000; Pickeral et al. 2000). By determining the origin of a given transduced sequence, we could locate where the active progenitor of such an L1 insertion is, or was. We analyzed 29 elements that terminated in a 3′ transduced sequence identified either by others (Goodier, Ostertag, and Kazazian 2000; Pickeral et al. 2000) or us and found a match for 20 of them in GenBank (table 3). Seven of these elements belonged to the older L1PA2 and L1PA3 families, and surprisingly (given the results of table 1), four were FL. However, three of these four were identified by Pickeral et al. (2000), who limited their search to FL elements. Thus, this unexpectedly high yield of FL elements does not reflect any special feature of L1 elements that contain 3′ transduced sequences, but it represents an ascertainment bias.

Fifteen of the 20 GenBank loci that were identified as the source of the transduced sequence (table 3, column 6) lacked an L1 element 5′ of it. Thus, these loci, which at one time contained an active L1 element, were apparently lost from the population some time after they propagated the L1 insert listed in the first column of table 3. We extended this result by determining if these once-active L1-containing loci were also lost from humans other than the one(s) represented in GenBank. Figure 4 shows an example of two PCR reactions that reveal an empty site (left) and an occupied site (right) in the eight humans of different ethnic origins that we examined. Table 3 summarizes these data. Each of the 13 sites that were empty in GenBank were also empty in a representative of each of the eight populations.

Discussion

The autosome-specific loss of FL L1 elements reconciles the fact that while ~35% of the elements gen-
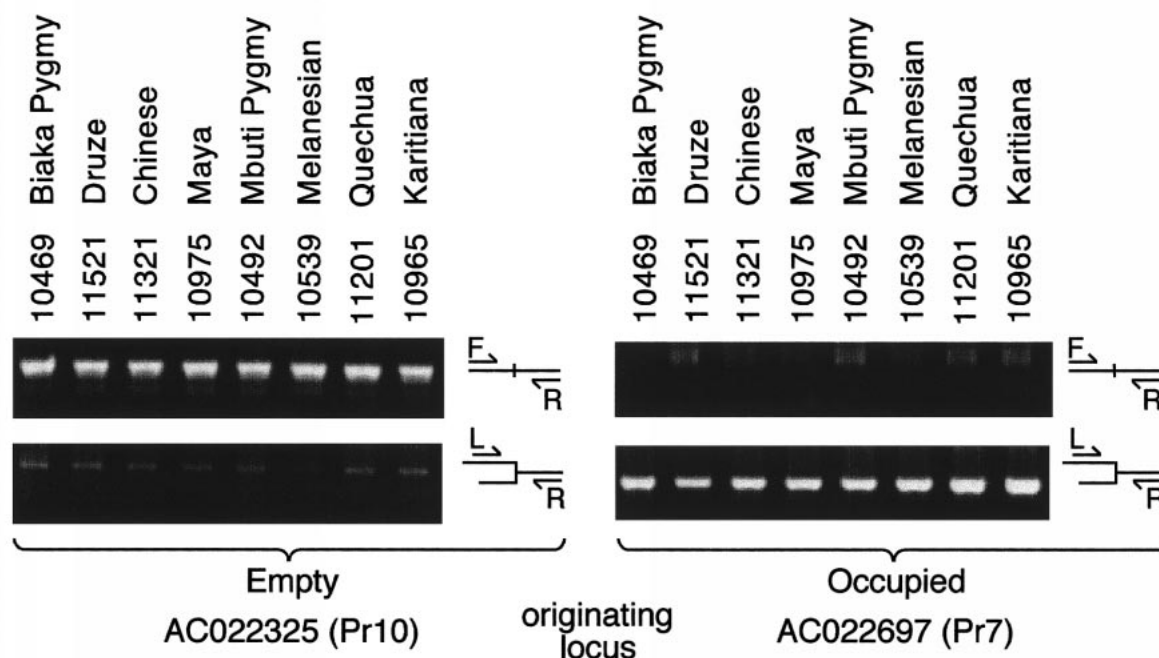
FIG. 4.—Population distribution of active L1-containing loci. Loci that contain (or once contained) active (retrotransposing) L1 elements were subjected to PCR using the primer pairs indicated in table 3. Each locus in the eight populations was examined by a pair of PCR reactions as described earlier (Boissinot, Chevret, and Furano 2000): one using PCR primers F and R, cognate to the non-L1 flanking region, and one using the appropriate flanking primer and an internal L1-specific primer. Representative results are shown here, and all are summarized in table 3.

erated by the currently expanding L1PA1 (Ta) family are FL, only about 5%–10% of total genomic L1 elements are (Fanning and Singer 1987; Scott et al. 1987). The L1PA1 (Ta) family would only represent a minority of the total genomic elements which are dominated by ancestral families such as L1PA2–LIPA5. Tables 1 and 2 show that the percentage of autosomal members of these ancestral families that are FL falls within the previously determined values for FL L1 elements (Fanning and Singer 1987).

The simplest explanation for our results is that FL L1 elements (and, to a lesser extent, long non-FL [>500 bp] elements) are deleterious and consequently are lost as a result of purifying selection. Thus, the relative enrichment of FL (and long non-FL) L1 elements on the Y and X chromosomes would represent yet another example of the accumulation of deleterious mutations on the sex chromosomes (Baker and Wichman 1990; Steinemann and Steinemann 1992; Wichman et al. 1992; Rice 1994; Kjellman, Sjogren, and Widegren 1995; Chalvet et al. 1998; Junakovic et al. 1998; Hurst 1999; Smit 1999; Erlandsson, Wilson, and Paabo 2000). If the L1 content of the Y chromosome represents the largely unselected products of past L1 family amplifications, then at least one third of the copies generated during the amplification of the ancestral L1PA2—L1PA5 L1 families were FL (tables 1 and 2), about the same proportion as that of the currently amplifying L1PA1 (Ta) family (Boissinot, Chevret, and Furano 2000).

Two bases for selection against (retro)transposable elements in eukaryotes have been proposed: the deleterious effects (genetic rearrangements) caused by ectopic (nonallelic) homologous recombination between elements and the deleterious effects (e.g., genetic damage) due to (retro)transposition. Although these effects are not mutually exclusive, arguments have been marshaled to support the predominance of one or the other (e.g., Biemont et al. 1997; Charlesworth, Langley, and Sniegowski 1997). Our results suggest that both effects may account for selection against FL and long non-FL elements.

Ectopic Recombination

Here, we distinguish between the recombination that facilitates the removal of deleterious L1-containing loci from the genome from the ectopic homologous recombination between nonallelic L1-containing loci that would make these loci deleterious.

Ectopic L1 recombination has caused genetic rearrangements in humans, including disease-producing genetic deletions (Burwinkel and Kilimann 1998; Segal et al. 1999). A presimian ancestral ectopic recombination produced a duplication of the gamma globin gene (Fitch et al. 1991), and a more recent one in hominids caused the inversion of a region of the Y chromosome (Schwartz et al. 1998). In addition, ectopic L1 recombination may remodel satellite DNA arrays of human centromeres (Laurent, Puechberty, and Roizes 1999).

These few cases of L1 ectopic homologous recombination contrast with the far greater numbers that have been documented for the 300-bp Alu SINE family members (e.g., Burwinkel and Kilimann 1998; Deininger and Batzer 1999). One explanation is that ectopic recombination between the more widely spaced L1 elements would lead to far more serious deleterious rearrangements (e.g., deletions) than the more closely spaced Alu elements (Burwinkel and Kilimann 1998; Deininger and

Batzer 1999). Thus, most ectopic L1 recombinants would be so deleterious that they would be rapidly lost from the population.

A direct correlation between the length of the homologous sequences and the rate of ectopic recombination in mammals has been demonstrated experimentally (e.g., Hasty, Rivera-Perez, and Bradley 1991; Cooper, Schimenti, and Schimenti 1998). These results are consistent with our finding that <500-bp L1 elements (less prone to ectopic recombination) are apparently not subject to purifying selection, whereas >500-bp L1 elements (more prone to ectopic recombination) are (table 1). However, the relatively frequent recombination between the 300-bp Alu sequences indicates that ectopic recombination between sequences that are <500 bp is possible, and some pairs of Alu elements have undergone repeated independent recombinations (e.g., Deininger and Batzer 1999; Hill et al. 2000).

Additionally, in contrast to ectopic homologous L1 recombination, ectopic nonhomologous (illegitimate) recombination involving L1 and non-L1 DNA is quite common even when an L1 partner is available (e.g., Inoue et al. 1997; McNaughton et al. 1998 and references therein). Fairly extensive deletions (e.g., 340 kb) can result, and the participation of L1 elements in these events is proportional only to their frequency in the sequences in question (McNaughton et al. 1998). Thus, one need not invoke ectopic homologous L1 recombination to produce deletions large enough to be seriously deleterious. In addition, the far more frequent occurrence of ectopic nonhomologous L1 recombination over ectopic homologous L1 recombination has been recapitulated experimentally (Richard et al. 1994).

Deleterious Effects of L1 Retrotransposition

If we assume that the content of FL L1 on the Y chromosome is a minimal estimate of the unselected products of L1 amplification, then >70% $(1 - (8.45/31.1))$ of the autosomal FL L1 elements generated by the older L1PA2–L1PA5 families have been lost. Thus, most of these FL L1-containing loci were deleterious. Although individual L1 insertions could be deleterious by directly inactivating a gene (e.g., Kazazian 2000) or altering the structural or regulatory properties of genic regions, for the most part these effects could be caused by both FL and non-FL elements. In addition, by whatever mechanism a single L1 insertion produces a deleterious effect, such L1-containing alleles should be more likely lost from the X chromosome than from autosomes because of the hemizygosity of the X chromosome in males. However, table 1 shows that this is the opposite of what we observed. In fact, the somewhat lower number of Alu sequences on the X chromosome than on the autosomes may reflect the deleteriousness of these insertions.

In addition to the above considerations, the randomness and sparseness (relative to the size of the genome) of insertions generated by any given L1 family suggest that selection against FL L1 elements is based on their having a global effect, i.e., by their retrotranspositional activity. If so, L1 retrotransposition must be sufficiently deleterious to be subject to purifying selection. An FL L1-containing locus that produced enough retrotransposition in germ line cells to lower fertility or enough retrotransposition in embryos to affect their viability would be rapidly lost from the population. These types of events constitute the phenomenon known as hybrid dysgenesis caused by transposable elements such as the L1-like I elements in *Drosophila melanogaster* (Busseau et al. 1994).

Since the long (>500 bp) non-FL L1 elements should not be capable of autonomous retrotransposition, their presumed participation in ectopic recombination could explain selection against these elements. However, perhaps some could produce L1 products, such as L1 RNA or an active reverse transcriptase, that could be deleterious. Regarding the latter, only a region beginning about 600 bp 5′ of the RT domain is required to efficiently generate cDNA from cellular RNA transcripts in vivo in a cell culture assay (Dhellin, Maestre, and Heidmann 1997). However, even >500-bp non-FL elements that were not long enough to encode a functional RT also exhibited a statistically significant sex-chromosomal bias (results not shown). Thus, the potential to produce an active RT does not account for the selection against these elements.

Whether or not something as drastic as a dysgenic phenomenon is the basis for selection against FL L1 elements, the repeated generation of deleterious FL L1-containing loci could nonetheless have affected (and may still affect) the genetic composition of primates, including humans. L1-containing loci that are both deleterious and linked to essential genetic loci could persist in the population, thereby decreasing its overall fitness. In addition, linkage of deleterious L1-containing loci to novel beneficial alleles could prevent the latter's subsequent fixation. Also, the presence of a particularly deleterious L1 on the Y chromosome could effectively eliminate that particular male lineage. How, or whether, such events affect a population may depend less on the spread of any particular deleterious L1-containing locus on the population than on the ability to repeatedly generate such loci.

Finally, the higher density of L1 elements on the X chromosome relative to the autosomes has been taken as evidence that L1 elements are involved in X inactivation (e.g., Bailey et al. 2000 and references therein). However, our results suggest that the excess of L1 elements on the X chromosome relative to autosomes can be explained by the selective loss of deleterious FL and long non-FL L1-containing loci from autosomes. This does not mean that a difference in the densities of L1 on the X chromosome and the autosomes has not been coopted for use in the X inactivation system. However, autosomal genes can be silenced if the autosome contains an active *Xist* gene. This suggests that an X-chromosome-like density of L1 elements may not be required for this process (Wutz and Jaenisch 2000).

Conclusions

We have shown that FL L1 elements have been under strong negative selection for at least the past 25

Myr of primate evolution. This selection is strong enough to account for the relative paucity of FL ancestral L1 elements in modern humans as compared with their number in the currently amplifying L1PA1 (Ta) family. A global deleterious effect of L1 retrotransposition, perhaps akin to retrotransposon-induced hybrid dysgenesis in insects, could be the basis, at least in part, for the negative selection on FL L1 elements.

LITERATURE CITED

BAILEY, J. A., L. CARREL, A. CHAKRAVARTI, and E. E. EICHLER. 2000. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. Proc. Natl. Acad. Sci. USA **97**:6634–6639.

BAKER, R. J., and H. A. WICHMAN. 1990. Retrotransposon *Mys* is concentrated on the sex chromosomes: implications for copy number containment. Evolution **44**:2083–2088.

BIEMONT, C., A. TSITRONE, C. VIEIRA, and C. HOOGLAND. 1997. Transposable element distribution in Drosophila. Genetics **147**:1997–1999.

BOISSINOT, S., P. CHEVRET, and A. V. FURANO. 2000. L1 (LINE-1) retrotransposon evolution and amplification in recent human history. Mol. Biol. Evol. **17**:915–928.

BURWINKEL, B., and M. W. KILIMANN. 1998. Unequal homologous recombination between LINE-1 elements as a mutational mechanism in human genetic disease. J. Mol. Biol. **277**:513–517.

BUSSEAU, I., M. C. CHABOISSIER, A. PELISSON, and A. BUCHETON. 1994. I factors in Drosophila melanogaster: transposition under control. Genetica **93**:101–116.

CABOT, E. L., B. ANGELETTI, K. USDIN, and A. V. FURANO. 1997. Rapid evolution of a young L1 (LINE-1) clade in recently speciated Rattus taxa. J. Mol. Evol. **45**:412–423.

CHALVET, F., C. DI FRANCO, A. TERRINONI, A. PELISSON, N. JUNAKOVIC, and A. BUCHETON. 1998. Potentially active copies of the gypsy retroelement are confined to the Y chromosome of some strains of Drosophila melanogaster possibly as the result of the female-specific effect of the flamenco gene. J. Mol. Evol. **46**:437–441.

CHARLESWORTH, B., C. H. LANGLEY, and P. D. SNIEGOWSKI. 1997. Transposable element distributions in Drosophila. Genetics **147**:1993–1995.

COOPER, D. M., K. J. SCHIMENTI, and J. C. SCHIMENTI. 1998. Factors affecting ectopic gene conversion in mice. Mamm. Genome **9**:355–360.

DEBERARDINIS, R. J., J. L. GOODIER, E. M. OSTERTAG, and H. H. KAZAZIAN JR. 1998. Rapid amplification of a retrotransposon subfamily is evolving the mouse genome. Nat. Genet. **20**:288–290.

DEININGER, P. L., and M. A. BATZER. 1999. Alu repeats and human disease. Mol. Genet. Metab. **67**:183–193.

DHELLIN, O., J. MAESTRE, and T. HEIDMANN. 1997. Functional differences between the human LINE retrotransposon and retroviral reverse transcriptases for in vivo mRNA reverse transcription. EMBO J. **16**:6590–6602.

ERLANDSSON, R., J. F. WILSON, and S. PAABO. 2000. Sex chromosomal transposable element accumulation and male-driven substitutional evolution in humans. Mol. Biol. Evol. **17**:804–812.

FANNING, T. G., and M. F. SINGER. 1987. LINE-1: a mammalian transposable element. Biochim. Biophys. Acta **910**:203–212.

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

FITCH, D. H., W. J. BAILEY, D. A. TAGLE, M. GOODMAN, L. SIEU, and J. L. SLIGHTOM. 1991. Duplication of the gamma-globin gene mediated by L1 long interspersed repetitive elements in an early ancestor of simian primates. Proc. Natl. Acad. Sci. USA **88**:7396–7400.

FURANO, A. V. 2000. The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons. Prog. Nucleic Acids Res. Mol. Biol. **64**:255–294.

GOODIER, J. L., E. M. OSTERTAG, and H. H. KAZAZIAN JR. 2000. Transduction of 3′-flanking sequences is common in L1 retrotransposition. Hum. Mol. Genet. **9**:653–657.

HASTY, P., J. RIVERA-PEREZ, and A. BRADLEY. 1991. The length of homology required for gene targeting in embryonic stem cells. Mol. Cell. Biol. **11**:5586–5591.

HATTORI, M., A. FUJIYAMA, T. D. TAYLOR et al. (26 co-authors). 2000. The DNA sequence of human chromosome 21. Nature **405**:311–319.

HILL, A. S., N. J. FOOT, T. L. CHAPLIN, and B. D. YOUNG. 2000. The most frequent constitutional translocation in humans, the t(11;22)(q23¼1) is due to a highly specific alu-mediated recombination. Hum. Mol. Genet. **9**:1525–1532.

HOLMES, S. E., B. A. DOMBROSKI, C. M. KREBS, C. D. BOEHM, and H. H. J. KAZAZIAN. 1994. A new retrotransposable human L1 element from the *LRE2* locus on chromosome 1q produces a chimaeric insertion. Nat. Genet. **7**:143–148.

HURST, L. D. 1999. The evolution of genomic anatomy. Trends Ecol. Evol. **14**:108–112.

INOUE, H., H. ISHII, H. ALDER, E. SNYDER, T. DRUCK, K. HUEBNER, and C. M. CROCE. 1997. Sequence of the FRA3B common fragile region: implications for the mechanism of FHIT deletion. Proc. Natl. Acad. Sci. USA **94**:14584–14589.

JUNAKOVIC, N., A. TERRINONI, C. DI FRANCO, C. VIEIRA, and C. LOEVENBRUCK. 1998. Accumulation of transposable elements in the heterochromatin and on the Y chromosome of Drosophila simulans and Drosophila melanogaster. J. Mol. Evol. **46**:661–668.

KAZAZIAN, H. H. JR. 2000. L1 Retrotransposons shape the mammalian genome. Science **289**:1152–1153.

KIMBERLAND, M. L., V. DIVOKY, J. PRCHAL, U. SCHWAHN, W. BERGER, and H. H. KAZAZIAN JR. 1999. Full-length human L1 insertions retain the capacity for high frequency retrotransposition in cultured cells. Hum. Mol. Genet. **8**:1557–1560.

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. SJOGREN, and B. WIDEGREN. 1995. The Y chromosome: a graveyard for endogenous retroviruses. Gene **161**:163–170.

LAURENT, A. M., J. PUECHBERTY, and G. ROIZES. 1999. Hypothesis: for the worst and for the best, L1Hs retrotransposons actively participate in the evolution of the human centromeric alphoid sequences. Chromosome Res. **7**:305–317.

MCNAUGHTON, J. C., D. J. COCKBURN, G. HUGHES, W. A. JONES, N. G. LAING, P. N. RAY, P. A. STOCKWELL, and G. B. PETERSEN. 1998. Is gene deletion in eukaryotes sequence-dependent? A study of nine deletion junctions and nineteen other deletion breakpoints in intron 7 of the human dystrophin gene. Gene **222**:41–51.

MORAN, J. V., R. J. DEBERARDINIS, and H. H. KAZAZIAN JR. 1999. Exon shuffling by L1 retrotransposition. Science **283**:1530–1534.

PICKERAL, O. K., W. MAKALOWSKI, M. S. BOGUSKI, and J. D. BOEKE. 2000. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. Genome Res. **10**:411–415.

Rice, W. R. 1994. Degeneration of a nonrecombining chromosome. Science **263**:230–232.

Richard, M., A. Belmaaza, N. Gusew, J. C. Wallenburg, and P. Chartrand. 1994. Integration of a vector containing a repetitive LINE-1 element in the human genome. Mol. Cell. Biol. **14**:6689–6695.

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

Saxton, J. A., and S. L. Martin. 1998. Recombination between subtypes creates a mosaic lineage of LINE-1 that is expressed and actively retrotransposing in the mouse genome. J. Mol. Biol. **280**:611–622.

Schwartz, A., D. C. Chan, L. G. Brown, R. Alagappan, D. Pettay, C. Disteche, B. McGillivray, A. de la Chapelle, and D. C. Page. 1998. Reconstructing hominid Y evolution: X-homologous block, created by X-Y transposition, was disrupted by Yp inversion through LINE-LINE recombination. Hum. Mol. Genet. **7**:1–11.

Scott, A. F., B. J. Schmeckpeper, M. Abdelrazik, C. T. Comey, B. O'Hara, J. P. Rossiter, T. Cooley, P. Heath, K. D. Smith, and L. Margolet. 1987. Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. Genomics **1**:113–125.

Segal, Y., B. Peissel, A. Renieri, M. de Marchi, A. Ballabio, Y. Pei, and J. Zhou. 1999. LINE-1 elements at the sites of molecular rearrangements in alport syndrome-diffuse leiomyomatosis. Am. J. Hum. Genet. **64**:62–69.

Skowronski, J., T. G. Fanning, and M. F. Singer. 1988. Unit-length line-1 transcripts in human teratocarcinoma cells. Mol. Cell. Biol. **8**:1385–1397.

Smit, A. F. A. 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr. Opin. Genet. Dev. **9**:657–663.

Smit, A. F. A., G. Tóth, A. D. Riggs, and J. Jurka. 1995. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. J. Mol. Biol. **246**:401–417.

Steinemann, M., and S. Steinemann. 1992. Degenerating Y chromosome of Drosophila miranda: a trap for retrotransposons. Proc. Natl. Acad. Sci. USA **89**:7591–7595.

Swofford, D. L. 1998. PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4. Sinauer, Sunderland, Mass.

Wichman, H. A., R. A. Van Den Bussche, M. J. Hamilton, and R. J. Baker. 1992. Transposable elements and the evolution of genome organization in mammals. Genetica **86**:287–293.

Wutz, A., and R. Jaenisch. 2000. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol. Cell **5**:695–705.