

Identity by descent and DNA sequence variation of human SINE and LINE elements

A.-H. Salem,^{a,b} D.A. Ray^a and M.A. Batzer^a

^aDepartment of Biological Sciences, Biological Computation and Visualization Center, Louisiana State University, Baton Rouge, LA (USA); ^bDepartment of Anatomy, Faculty of Medicine, Suez Canal University, Ismailia (Egypt)

Abstract. To test the hypothesis that Alu and L1 elements are genetic characters that are essentially homoplasmy-free, we sequenced a total of five human L1 elements and eleven recently integrated Alu elements from 160 chromosomes (80 individuals representing four diverse human populations). Analysis of worldwide samples at L1 loci revealed 292 segregating sites and a nucleotide diversity of 0.0050. For Ya5 Alu loci, there were

129 segregating sites and nucleotide diversity was estimated at 0.0045. The Alu and L1 sequence diversity varied element to element. No completely or partially deleted Alu or L1 alleles were identified during the analysis. These data suggest that mobile element insertions are identical by descent characters for the study of human population genetics.

Copyright © 2005 S. Karger AG, Basel

Over the last 60 million years of evolution, the human genome has been bombarded by a variety of repetitive elements (Smit, 1999). Among these, LINE-1 (L1), a type of long interspersed element (LINE), and Alu, a type of short interspersed element (SINE), are most common. Combined, they account for an estimated 26% of the human genome (Lander et al., 2001; Venter et al., 2001). Each of these mobile elements is thought to have populated the human genome in successive waves of retrotransposition (Kazazian, 2001; Batzer and Deininger, 2002).

LINE elements have amplified over the last 100 million years of mammalian evolution and comprise approximately 15% of the human genome (~100,000 insertions; Prak and Kazazian, 2000). The majority of L1 elements within the

human genome are 5' truncated copies of a few active L1 elements that are capable of retrotransposition. Some young L1 elements have inserted into the human genome so recently that chromosomes are polymorphic for the presence or absence of the element at a particular chromosomal location (Sheen et al., 2000). The Alu family of repetitive DNA elements derives its name from a single recognition site for the restriction enzyme *AluI* located near the middle of the element. The human genome contains about one million copies of these elements, ~11% of the total genome. While unable to retrotranspose autonomously, Alu elements are thought to borrow the factors that are required for their amplification from LINE elements (Sinnott et al., 1992; Boeke, 1997; Kajikawa and Okada, 2002; Dewanieux et al., 2003), which encode a protein with endonuclease and reverse transcriptase activity (Feng et al., 1996; Jurka, 1997). Similar to L1 elements, many recently integrated Alu insertions are polymorphic with respect to presence/absence in the human genome (reviewed in Batzer and Deininger, 2002).

Alu and L1 insertion polymorphisms offer two important advantages over other nuclear based polymorphisms for systematic and population genetic studies. First, the presence of an element in an individual is thought to represent identity by descent (IBD), since the probability that two different young mobile elements would integrate independently in the same chromosomal location is negligible (Batzer and Deininger, 2002). Polymorphic mobile element insertions should thus more accurately reflect population relationships than many

Supported by the Louisiana Board of Regents Millennium Trust Health Excellence Fund HEF (2000-05)-05, (2000-05)-01, and the National Science Foundation BCS-0218338 (M.A.B.).

Received 13 October 2003; manuscript accepted 21 November 2003.

Request reprints from: Dr. Mark A. Batzer, Department of Biological Sciences
Biological Computation and Visualization Center, Louisiana State University
202 Life Sciences Building, Baton Rouge, LA 70803 (USA)
telephone: (225) 578-7102; fax: (225) 578-7113; e-mail: mbatzer@lsu.edu

Table 1. Locus names, original accession numbers, chromosomal locations and additional information for the L1HS loci sequenced for this study

Locus	Accession no.	Chr. loc.	Fragment size	Insert length	No. of clones sequenced	Total nucleotides sequenced
L1HS45	AC004865.1	1	326	183	438	80154
L1HS241	AP003112.1	8	366	200	415	83000
L1HS377	AC078856.11	3	575	415	410	170150
L1HS444	AC027315.3	5	463	138	538	74244
L1HS407	AC108076.2	4	412	172	394	67768
Total					2195	475316

other genetic markers (i.e., sequence data, restriction fragment length polymorphisms [RFLP], and microsatellites) in which the sharing of the same allele by two individuals may reflect identity by state only (i.e., homoplasy; Batzer et al., 1994). Once an Alu or L1 element inserts into a chromosomal locus it may generate new copies, but there is no evidence that it is ever completely excised or lost from that locus. Therefore, each mobile element insertion is thought to be stable through evolutionary time (Shedlock and Okada, 2000). Each new integration is a record of a unique retrotransposition event that occurred only once in primate evolution. A second advantage of these genetic markers is that the ancestral state of an insertion polymorphism is known to be the absence of the element at a particular genomic location (Batzer and Deininger, 1991; Perna et al., 1992; Batzer et al., 1994). Precise knowledge of the ancestral state of a genomic polymorphism allows us to draw trees of population relationships without making unnecessary assumptions (Perna et al., 1992; Batzer et al., 1994, 1996; Stoneking et al., 1997; Watkins et al., 2001, 2003; Batzer and Deininger, 2002).

Most L1 and Alu insertions are “fixed”, meaning that all individuals are homozygous for the insertion at a particular locus. However, a number of human-specific insertions are dimorphic – an insertion may be present or absent on each of the paired chromosomes of different individuals. Mobile elements at these loci are thought to have inserted within the last million years of modern human evolution (reviewed in Batzer and Deininger, 2002). These dimorphisms show differences in allele and genotype frequencies between extant populations and are tools that have proven useful for reconstructing human prehistory (Perna et al., 1992; Batzer et al., 1994; Hammer, 1994; Stoneking et al., 1997; Comas et al., 2000; Jorde et al., 2000; Nasidze et al., 2001; Watkins et al., 2001, 2003; Battilana et al., 2002; Romualdi et al., 2002).

Insertion homoplasy can occur across distantly related taxa as a function of evolutionary time and variable retroposition rates among species (Roy-Engel et al., 2002; Vincent et al., 2003). This limitation can inhibit the application of SINEs and LINEs to deeper evolutionary questions (Hillis, 1999; Cantrell et al., 2001). Several examples of SINE insertions that have occurred at or near the same genomic region have recently been reported (Cantrell et al., 2001; Roy-Engel et al., 2002a; Salem et al., 2003a) although no instances of LINE insertion homoplasy have ever been identified (Salem et al., 2003b; Vincent et al., 2003). However, the application of SINE and LINE elements to

the study of human population genetics is thought to be essentially homoplasy-free as a result of the short evolutionary time frame and the relatively low current rate of Alu and LINE retroposition in the human genome (Batzer and Deininger, 2002; Roy-Engel et al., 2002a). This is particularly true with respect to Alu-insertion polymorphisms, as the probability of two independent Alu insertions occurring in the same genomic region in the human population, given the current rate of Alu retrotransposition and the relatively short evolutionary time frame that is involved, is essentially zero (Batzer et al., 1994; Roy-Engel et al., 2002a). However, even though it is a generally accepted idea that Alu and L1 elements are identical by descent, no large sequence data sets have examined this question. The mobile element based sequence data also presents an opportunity to further examine human evolutionary history. Herein, we report a large scale sequence analysis of several Alu and LINE insertion loci.

Materials and methods

DNA sources and PCR amplification

Diverse human DNA samples were available from previous studies (Roy et al., 1999; Carroll et al., 2001; Roy-Engel et al., 2001; Salem et al., 2003a, b). We performed PCR amplification of all loci from 20 individuals from each of four geographically distinct populations (African American, Asian, European, and Egyptian for the Ya5 Alu loci; and African American, Asian, European, and South American for the L1HS loci), for a total of 160 chromosomes. We amplified five loci from the L1HS subfamily (L1HS45, L1HS241, L1HS377, L1HS407, and L1HS444; Table 1) and eleven loci from the Ya5 and Ya5a2 subfamilies (Ya5a2AD5, Ya5a2AD6, Ya5a2AD7, Ya5a2AD8, Ya5a2AD9, Ya5a2AD11, Ya5a2AD12, Ya5NBC220, Ya5NBC235a, Ya5NBC241, and Ya5NBC243; Table 2), hereafter collectively referred to as the Ya5 subfamily.

Alu Ya5 and L1HS subfamily PCR amplifications were carried out in 25- μ l reactions containing 20–100 ng of template DNA, 40 pM of each oligonucleotide primer (Myers et al., 2002; Roy-Engel et al., 2002), 200 μ M dNTPs, in 50 mM KCl, 1.5 mM MgCl₂, 10 mM Tris-HCl (pH 8.4) and Taq DNA polymerase (1.25 Units). Each sample was subjected to the following amplification for 32 cycles: an initial denaturation of 150 s at 94 °C, 1 min denaturation at 94 °C, 1 min at the annealing temperature (specific for each locus), and extension at 72 °C for 1 min. Following the cycles a final extension was performed at 72 °C for 10 min. For analysis, 20 μ l of each sample was fractionated on a 2% agarose gel with 0.05 μ g/ml ethidium bromide. PCR products were directly visualized using UV fluorescence.

Sequence analyses

DNA sequencing was performed on gel purified PCR products that had been cloned using the TOPO TA cloning vector (Invitrogen) using chain termination sequencing (Sanger et al., 1977) on an Applied Biosystems 3100 automated DNA sequencer. Five to ten clones from each individual were

Table 2. Locus name, original accession numbers, chromosomal locations and additional information for the Alu Ya5 loci sequenced for this study

Locus	Accession no.	Chr. loc	Fragment size	Insert length	No. of clones sequenced	Total nucleotides sequenced	Human Diversity
Ya5a2AD5	AC026839	18	460	282	58	16356	FP
Ya5a2AD6	AL355580.13	13	620	282	56	15792	FP
Ya5a2AD7	AL138681.17	13	419	282	41	11562	IF
Ya5a2AD8	AL162713.19	13	449	282	62	17484	IF
Ya5a2AD9	AC079456.27	12	450	282	39	10998	IF
Ya5a2AD11	AC090071.1	5	434	282	56	15792	FP
Ya5a2AD12	AC004057.1	4	451	282	58	16356	FP
Ya5NBC220	AC007611.5	16	467	282	65	18330	IF
Ya5NBC235a	AL357094	14	418	282	76	21432	HF
Ya5NBC241	AC018924.6	15	392	282	43	12126	IF
Ya5NBC243	AL442166.1	21	405	282	26	7332	IF
Total						164970	

sequenced. Sequence alignments for all of the Ya5 and L1HS subfamily loci were performed using MegAlign software (DNASTar version 3.1.7 for Windows 3.2). Multiple sequence alignments that contain all of the members of the Ya5 and L1HS subfamilies can be found on our website (<http://batzerlab.lsu.edu>) under publications. The poly-A tails of each element are highly variable in length and might have introduced bias in diversity estimates. This concern was addressed by eliminating that portion of each element prior to statistical analyses. Sequences for all Ya5 and L1HS sequences are available from GenBank using accession numbers AF504933–AF505511 and AY435515–AY436314, respectively.

Statistical analyses

Basic sequence statistics including nucleotide diversity (π), number of segregating sites (S), θ_S , θ_π , and Tajima's D (Tajima, 1996) were calculated using Arlequin v 2.0 (Schneider et al., 1997; <http://anthropologie.unige.ch/arlequin>). Additional statistics to test loci for the effects of natural selection were calculated using DnaSP v3.51 (Rozas and Rozas, 1999).

Results

Identity by descent

Both L1HS elements and Ya5 Alu family members are relatively young with estimated average ages of 1.99 million years (Myers et al., 2002) and 0.62 million years (Roy et al., 2000; Roy-Engel et al., 2002b), respectively. Comprehensive analyses of both groups of mobile elements have been reported previously (Roy et al., 2000; Myers et al., 2002; Roy-Engel et al., 2002b). All five L1HS loci examined in our study were fixed present (every individual tested had the L1HS element in both chromosomes). The eleven Ya5 subfamily loci examined were broken into three groups: four were fixed present, six were intermediate frequency insertion polymorphisms in human populations and one was a high frequency insertion polymorphism (Table 2). For the Alu Ya5 subfamily loci examined, we analyzed 580 clones resulting in 164,970 sequenced nucleotides. For the L1HS loci we sequenced and analyzed 2,195 clones, and a total of 475,316 nucleotides.

In each case, if all of the insertions being examined were fixed present and identical by descent, we would expect to see the same elements in every clone from each individual examined. Conversely, if they were not IBD, we would expect to recover some clones that contain sequences other than the

expected mobile element. We examined ~0.5 Mb sequences containing L1HS loci and did not find a single clone that contained non-L1HS and flanking sequence. Similarly, when we analyzed sequences containing Ya5 loci we did not find any non-Ya5 Alu sequences. We also did not recover any deletions, partial or complete, of any of the examined L1HS or Alu loci.

Human variability: L1HS

Our survey of a worldwide sample of 80 individuals (160 human haploid equivalents) revealed 292 sequence polymorphisms among all of the L1HS elements analyzed. Overall nucleotide diversity was 0.0050 for the world sample. Nearly equal variability was observed among the studied populations: African Americans ($S = 88$ and $\pi = 0.0054$), Asians ($S = 89$ and $\pi = 0.0050$), European ($S = 83$ and $\pi = 0.0046$), and South Americans ($S = 84$ and $\pi = 0.0044$) as shown in Table 3. Both quantities S and π can be used to estimate the parameter θ . The two estimates of θ are $\theta_S = S/\Sigma^{n-1}(1/i)$, where n is the number of samples, and θ_π , where Π represents the average pairwise sequence difference, equivalent to π when divided by the total number of sites experimentally examined. By comparing θ_S and θ_π estimates, we can test whether the analyzed segments have been evolving neutrally (Tajima, 1989). For the worldwide data set using all L1HS loci, the test statistic D had a negative value (Table 3), and the difference from zero was significant. We also performed two other tests of neutrality: Fu and Li's D^* and F^* tests (Fu and Li, 1993). Both tests produced negative values which were significantly different from zero. F_{ST} for all loci, divided into four continental populations, was 0.0401 indicating that only 4.01% of the variance is due to differences among these populations. Pairwise F_{ST} 's of the continental samples analyzed are also reported in Table 3.

The average heterozygosity per polymorphic site H_S was calculated for all polymorphisms and separately for the shared and specific sites (Table 3). Understandably, H_S for all polymorphic sites was lowest in the worldwide sample, where continent-specific polymorphisms counted little because of their low global frequency. All continental groups, including African Americans, are essentially homogenous for the number of shared segregating sites and for overall nucleotide diversity. Thus we

Table 3. Summary statistics for all L1HS loci

	World n = 160	Africa n = 40	Europe n = 40	Asia n = 40	Americas n = 40
S Total	292	88	83	89	84
Shared		23	25	23	25
Specific		65	58	66	59
Hs = $\theta_{II}/S(\%)$ Total	1.89	6.72	6.12	6.24	5.77
Shared		25.70	20.32	24.16	19.39
Specific		9.10	8.76	8.42	8.21
$\pi \pm SD$	0.005 \pm 0.003	0.005 \pm 0.003	0.005 \pm 0.003	0.005 \pm 0.003	0.004 \pm 0.002
$\theta_{II} \pm SD$	5.529 \pm 2.957	5.912 \pm 3.202	5.081 \pm 2.797	5.556 \pm 3.029	4.846 \pm 2.682
$\theta_s \pm SD$	51.688 \pm 11.798	20.689 \pm 6.295	19.513 \pm 5.958	20.924 \pm 6.363	19.748 \pm 6.025
D (Tajima)	-2.894	-2.605	-2.692	-2.68	-2.748
	-0.109	-0.129	-0.067	-0.128	-0.101
	$P = 0.001$	$P = 0.001$	$P = 0.001$	$P = 0.001$	$P = 0.001$
Fu and Li's D	-11.022	-5.503	-5.651	-5.627	-5.602
	$P = 0.02$	$P = 0.02$	$P = 0.02$	$P = 0.02$	$P = 0.02$
Fu and Li's F	-8.578	-5.327	-5.49	-5.449	-5.467
	$P = 0.02$	$P = 0.02$	$P = 0.02$	$P = 0.02$	$P = 0.02$
Population pairwise F_{ST} 's. Distance method: Pairwise difference					
	Africa	Asia	Europe		
Asia	0.04527 ^a				
Europe	0.01255 ^a	0.05741 ^a			
America	0.02183 ^a	0.09366 ^a	0.0037		
^a P value ≤ 0.5 .					
AMOVA					
	% of variation among populations	4.01			
	% of variation within populations	95.99			

were unable to distinguish between the African American population and the remaining populations.

With regard to individual L1HS loci, we were also unable to distinguish between African Americans and other groups using the number of segregating sites and/or nucleotide diversity on a consistent basis (data presented in Supplemental Tables 1–5 at our website <http://batzerlab.lsu.edu>). For example, in the case of locus L1HS241, we found the total number of segregating sites and the nucleotide diversity to be higher in the African Americans than in any of the other groups (Supplemental Table 2). Yet, at another locus L1HS444, we saw more segregating polymorphic sites and higher nucleotide diversity in Europeans, Asians and South Americans than in African Americans (Supplemental Table 5). Tajima's D and Fu and Li's D^* and F^* values were negative and significantly different from zero in all cases. High numbers of segregating sites were due to continent-specific polymorphisms. Differences among individuals within populations rather than among populations accounted for the major portion of F_{ST} variance.

When we examined the number and frequency of haplotypes (variants containing at least one mutation when compared to all other sequences at that locus) in different groups, we found that, in general, there is one main haplotype shared by all groups and the frequency of this haplotype ranged from 0.5 to 0.825 depending on the locus examined (Table 4). The one

exception to this rule is found at locus L1HS377. At this locus, four haplotypes occur at frequencies of 10% or greater in specific populations. One haplotype was specific to the African American population. A second haplotype was present in both African American and Asian populations at greater than 0.175. A third was found at frequencies of 0.1 and 0.05 in South American and European populations, respectively but not in Asians or African Americans. Finally, one haplotype was found at high frequencies in all four populations ≥ 0.175 . At each locus, the majority of haplotypes were population specific and present at low frequencies (≤ 0.05) in their respective populations.

Human variability: *Alu Ya5*

Our survey of a worldwide sample of 43 individuals or 86 human haploid equivalents ($n = 86$) revealed 129 polymorphisms when all eleven *AluYa5* loci are analyzed (Table 5). Nucleotide diversity for the world sample was similar to that found for the L1HS loci, 0.0045. Similar variability was observed among the studied groups. Egyptians showed the lowest nucleotide diversity and number of segregating sites when the loci are combined: African Americans ($S = 44$ and $\pi = 0.0042$), Asians ($S = 32$ and $p = 0.0046$), Europeans ($S = 50$ and $\pi = 0.0060$), and Egyptians ($S = 25$ and $\pi = 0.0036$) as shown in Supplemental Table 9, www.karger.com/doi/10.1159/000080803. Tajima's D was neg-

ative and significantly different from zero in the worldwide analysis, African Americans, and European groups but the difference value was not significant in Asians and Egyptians. The Fu and Li's D* and F* tests, also had significant negative values in same groups. The F_{ST} for the worldwide analysis is 0.0273 indicating that only 2.73% of the variance is due to differences among these populations. Pairwise F_{ST} 's of the continental samples analyzed are reported in Supplemental Table 10, www.karger.com/doi/10.1159/000080803.

The average per polymorphic site heterozygosity H_S was calculated for all polymorphisms and separately for the shared and specific sites (Table 5). Like the L1HS loci, H_S for all polymorphic sites was lowest in the world sample. Estimates of nucleotide diversity in African Americans and Egyptians were slightly lower than the other two groups. H_S in African Americans was lower than any of the other groups.

Sequence variability at individual loci was also similar to the patterns observed at L1HS loci (Supplemental Tables 6–16 at our website <http://batzerlab.lsu.edu>). For example, the number of segregating sites and the nucleotide diversity at locus Ya5NBC235a were both higher in African Americans than other groups. This was due to a larger number of continent-specific segregating sites. The number of segregating polymorphic sites was variable in the different studied groups. We found a negative Tajima D, Fu and Li's D* and F* value for each group and these values were significantly different from zero. Finally, the differences among individuals within populations rather than among populations were responsible for the major portion of F_{ST} variance.

With regard to the number and frequency of haplotypes present in each group, we again found results that mirrored the results found at L1HS loci (Table 6). Most of the haplotypes identified were population specific, but one main haplotype was usually shared by all groups and the frequency of this haplotype ranged from 0.2 to 1 depending on the locus. The exceptions to this rule were: Ya5NBC241, in which all four African American individuals containing an insertion exhibited unique haplotypes; Ya5NBC243, in which one haplotype was shared by three populations but was found in only one individual from the European sample; Ya5a2AD12, in which the most frequent haplotype in three populations was not found in the Asian sample; Ya5a2AD8, at which no single haplotype was found at high frequency; and Ya5a2AD6 where we found three haplotypes shared by all continents at relatively high frequencies.

Discussion

Identity by descent

Our data support the assertion that mobile element insertion polymorphisms are essentially homoplasmy free characters that can be used to study human population genetics. In addition, there is no evidence for any process that specifically removes mobile elements from the genome; even when a rare deletion occurs, it leaves behind a molecular signature (Edwards and Gibbs, 1992). By contrast, other types of genetic polymorphisms, such as microsatellites (Nakamura et al.,

Table 4. Haplotype frequencies for the L1HS loci

Haplotype	Africa	Asia	Europe	Americas
L1HS45				
1	0.475	0.6	0.625	0.75
2	0.025	0.025	0	0
3	0.05	0	0	0
4	0.025	0.025	0	0
5	0.05	0	0	0
6-21	0.025	0	0	0
22	0	0.025	0.025	0
23-35	0	0.025	0	0
36-49	0	0	0.025	0
50-59	0	0	0	0.025
L1HS241				
1	0.525	0.725	0.75	0.725
2	0	0.025	0	0.05
3-19	0.025	0	0	0
20-29	0	0.025	0	0
30-39	0	0	0.025	0
40-48	0	0	0	0.025
L1HS377				
1	0.2	0.175	0.225	0.275
2	0.175	0.025	0.05	0.025
3	0.175	0.225	0.075	0.025
4	0.025	0	0.025	0
5	0.025	0.025	0	0
6-21	0.025	0	0	0
22-43	0	0.025	0	0
44-66	0	0	0.025	0
67	0	0	0.05	0.1
68	0	0	0	0.05
69-89	0	0	0	0.025
L1HS407				
1	0.65	0.7	0.575	0.525
2	0.025	0	0	0.025
3-15	0.025	0	0	0
16	0	0.025	0.025	0.025
17-27	0	0.025	0	0
28-43	0	0	0.025	0
44	0	0	0	0.05
45-59	0	0	0	0.025
L1HS444				
1	0.825	0.75	0.65	0.725
2-7	0.025	0	0	0
8	0.025	0.025	0	0.025
9	0	0.025	0.025	0
10-17	0	0.025	0	0
18	0	0	0.05	0
19-28	0	0	0.025	0
29	0	0	0.025	0.025
30-38	0	0	0	0.025

1987), restriction fragment length polymorphisms (RFLPs) (Botstein et al., 1980) and single nucleotide polymorphisms (SNPs) (Chakravarti, 1998; Pennisi, 1998; Brookes 1999), are often merely identical by state; that is, they have arisen as a result of an independent parallel mutation(s).

The presence of specific mobile element insertions in a species can be used as a distinct character for phylogenetic analysis. If two species share a common retroelement insertion at a given location and a third species does not, the first two species are likely to be more closely related. If several such insertions are observed, this provides strong evidence of species related-

Table 5. Summary statistics for all Alu Ya5 loci

	World n = 43	Africa n = 14	Europe n = 11	Asia n = 9	Egypt n = 9
S Total	129	44	50	32	25
Shared		4	10	7	8
Specific		40	40	25	17
Hs = $\theta_{\pi}/S(\%)$ Total	6.52	16.21	23.16	24.67	24.57
Shared		178.30	115.78	112.76	76.79
Specific		17.83	28.95	31.57	36.14
$\pi \pm$ SD	0.005 \pm 0.002	0.004 \pm 0.0023	0.006 \pm 0.004	0.004 \pm 0.003	0.003 \pm 0.00
$\theta_{\pi} \pm$ SD	8.410 \pm 4.416	7.132 \pm 3.996	11.578 \pm 6.490	7.893 \pm 4.682	6.143 \pm 3.723
$\theta_s \pm$ SD	28.917 \pm 8.657	13.836 \pm 5.432	16.967 \pm 7.175	10.413 \pm 4.815	8.485 \pm 3.986
D (Tajima)	-2.63265 -0.1428 <i>P</i> = 0.001	-2.1154 -0.04481 <i>P</i> = 0.005	-1.70965 -0.07908 <i>P</i> = 0.032	-1.42192 -0.06686 <i>P</i> = 0.071	-1.36951 -0.03071 <i>P</i> = 0.078
Fu and Li's D	-5.753 <i>P</i> = 0.02	-2.771 <i>P</i> = 0.02	-2.118 <i>P</i> = 0.05	-1.669 <i>P</i> = 0.1	-1.829 <i>P</i> = 0.1
Fu and Li's F	-5.502 <i>P</i> = 0.02	-2.976 <i>P</i> = 0.02	-2.302 <i>P</i> = 0.05	-1.805 <i>P</i> = 0.1	-2 <i>P</i> = 0.1
Population pairwise FSTs. Distance method: Pairwise difference					
	Africa	Asia	Egypt		
Asia	0.09621 ^a				
Egypt	-0.00749	0.05335			
Europe	0.03111 ^a	-0.00275	-0.02929		
^a	<i>P</i> value \leq 0.5				
	% of variation among population		2.73		
	% of variation within population		97.27		

ness. This type of analysis has been used extensively to study speciation questions in various organisms (Lawrence et al., 1989; Nikaido et al., 2001b), including defining the whale as a close relative of the artiodactyls (Nikaido et al., 1999), determining the phylogenetic relationships among Pacific salmonids (Murata et al., 1993) and confirming the primate phylogeny (Ryan and Dugaiczky, 1989; Shen et al., 1991; Hamdi et al., 1999; Salem et al., 2003c). As previously stated, the primary advantage of retroelement insertions for such studies is the high likelihood that two genomes sharing a mobile element insertion at the same locus achieved that state through non-homoplasious means. There are rare examples of retroelements inserting independently in the same or nearly the same positions and, as one proceeds backwards along evolutionary lineages, the likelihood of observing such events increases (Arcot et al., 1998; Kass et al., 2000; Cantrell et al., 2001; Roy-Engel et al., 2002; Vincent et al., 2003). However, for most mobile elements, the rate of insertion is low enough and the number of potential insertion sites high enough that these events are relatively rare. We previously analyzed several hundred recently integrated human Alu and LINE element insertions throughout primate phylogeny and have found only a low level of parallel insertion in the New (Roy-Engel et al., 2002a; Salem et al., 2003a) and Old world monkeys (Salem et al., 2003a) and no parallel insertions involving LINE elements (Salem et al., 2003b; Vincent et al., 2003). Studies of many loci have made it clear that it is also very unlikely for a mobile element to be deleted from a genome.

Being identical by descent also makes retroelement insertions useful for the study of human population diversity and origins. Different numbers of such markers have been shown to provide robust measurements of the relations of various world populations to one another (Batzer et al., 1994; Stoneking et al., 1997; Watkins et al., 2001, 2003). It is possible that some of the frequency variation of specific Alu insertion alleles between populations is related to a relatively recent insert that occurred in one human population group and has only modestly spread to others. However, given the migration and demographics of most of the human populations, it is more likely that the allele frequency of different Alu inserts has changed through random population drift in the relatively small founding populations.

Human genetic diversity

The analysis of human Alu-insertion polymorphisms has been used previously to address questions about human origins and demography (Perna et al., 1992; Batzer et al., 1994, 1996; Hammer, 1994; Sherry et al., 1997; Stoneking et al., 1997; Novick et al., 1998; Comas et al., 2000; Jorde et al., 2000; Bamshad et al., 2001; Nasidze et al., 2001; Watkins et al., 2003). In several instances, many types of genetic variation (such as mitochondrial DNA sequences or RFLPs) have been examined in overlapping, diverse human populations and have provided largely congruent results with respect to the history of the human population (Jorde et al., 2000; Bamshad et al., 2001; Watkins et al., 2001).

In addition to our study of identity by descent, we analyzed our data to characterize human nucleotide diversity in a sample

sumes greater long-term effective population size in Africa that promoted a build up of extra diversity compared to other continents with populations of smaller size. The second scenario invokes the “out of Africa” bottlenecks that caused reduction of the previously accumulated diversity.

As predicted by the out-of-Africa model, we expected to observe consistently higher sequence diversity among African Americans when compared to the other sampled populations. However, this was not evident in our data set. The African American sample exhibited the highest nucleotide diversity at only two of the five L1HS loci and at only three of the eleven Alu Ya5 loci. It should also be noted that none of the other populations exhibited consistently higher diversity values. Instead the distribution of higher nucleotide diversity values appeared random throughout the various populations. We believe that this difference reflects the admixed nature of our Louisiana derived African American samples (Parra et al., 1998).

The mobile element based sequence diversity reported here should not be interpreted as refuting the “African Origin” hypothesis or a more general “recent-single-origin” hypothesis. Rather, the presence of one or two worldwide haplotypes at most loci continues to suggest a single origin for human populations. The present data merely fail to indicate which of the tester populations might be the original source. It should also be noted that this data set of nucleotide diversity is different from many data sets previously utilized in one major way. Many data sets examining sequence data used sequences from loci that existed well before the divergence between chimpanzees and humans and, for that matter, primates. This allows for diversity to accumulate in the source populations prior to dispersal from the source population. In the case of the Alu elements and L1 elements examined here, the loci examined were not present prior to the speciation event separating humans and chimpanzees. Instead the elements were inserted into the genome de novo and spread from an initial low frequency to a higher frequency via genetic drift. Therefore, these loci may not have had an opportunity to build up sufficient genetic diversity in the ancestral human population prior to giving rise to daughter populations. The sequence diversity within all of the populations (source and daughter) at those loci should therefore be affected only by processes that are population specific, with some showing greater diversity at some loci but not at others.

References

- Arcot SS, Adamson AW, Risch GW, LaFleur J, Robichaux MB, Lamerdin JE, Carrano AV, Batzer MA: High-resolution cartography of recently integrated human chromosome 19-specific Alu fossils. *J Mol Biol* 281:843–856 (1998).
- Aris-Brosou S, Excoffier L: The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. *Mol Biol Evol* 13:494–504 (1996).
- Bamshad M, Kivisild T, Watkins WS, Dixon ME, Ricker CE, Rao BB, Naidu JM, Prasad BVR, Reddy PG, Rasanayagam A, Papiha SS, Villems R, Carroll ML, Nguyen SV, Batzer MA, Jorde LB: Genetic evidence on the origins of Indian caste populations. *Genome Res* 11:994–1004 (2001).
- Battilana J, Bonatto SL, Freitas LB, Hutz MH, Weimer TA, Callegari-Jacques SM, Batzer MA, Hill K, Hurtado AM, Petzl-Erler ML, Salzano FM: Alu insertions versus blood group plus protein genetic variability in four Amerindian populations. *Ann Hum Biol* 29:334–347 (2002).
- Batzer MA, Deininger PL: A human-specific subfamily of Alu sequences. *Genomics* 9:481–487 (1991).
- Batzer MA, Deininger PL: Alu repeats and human genomic diversity. *Nat Rev Genet* 3:370–379 (2002).
- Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ, Deininger PL: African origin of human-specific polymorphic Alu insertions. *Proc Natl Acad Sci, USA* 91:12288–12292 (1994).
- Batzer MA, Arcot SS, Phinney JW, Alegria-Hartman M, Kass DH, Milligan SM, Kimpton C, Gill P, Hochmeister M, Ioannou PA, Herrera RJ, Boudreau DA, Scheer WD, Keats BJB, Deininger PL, Stoneking M: Genetic variation of recent Alu insertions in human populations. *J Mol Evol* 42:22–29 (1996).
- Boeke JD: LINES and Alus – the polyA connection. *Nat Genet* 16:6–7 (1997).
- The results of Tajima’s *D* and Fu and Li’s *D** and *F** tests show departure from neutrality. While it is possible that the loci are indeed under unknown selective pressures, the departures seen are more probably due to factors other than selective effects, such as an expansion from a very small population to a much larger one (a population bottleneck) or heterogeneity of mutation rates (Aris-Brosou and Excoffier, 1996; Tajima, 1996). The failure of the *D** and *F** tests to detect an excess of high frequency variants is probably due to the insensitivity that these tests have to such variants (Fu and Li, 1993). Tajima’s *D* statistic, which compares the number of segregating sites in samples with the mean pairwise difference (Tajima, 1989), is insensitive to the presence of very high frequency variants because they contribute little to mean pairwise differences. Fu and Li’s statistics, which compare the number of singletons with the number of nucleotide differences between sequences (Fu and Li, 1993), are insensitive to the presence of high frequency variants because they are scored simply as non-singletons. Thus, all three of these statistics are probably unaffected by extreme frequency spectra and represent tests with relatively low resolving power for data such as these. In addition, the hypotheses of population growth and natural selection are not mutually exclusive. It is possible for both to occur at the same time in the same populations and to attribute the observed patterns to one, the other, or both. Alternatively, gene conversion between mobile elements might also introduce sequence variation from other mobile elements.

In conclusion, we have presented a large data set whose primary purpose was to test the hypothesis that Alu and L1 insertions are essentially homoplasy free and suitable for use in studies of human population genetics and phylogenetics. Our examination of ~0.67 Mb of sequence supports this hypothesis. We observed no instances in which a parallel mobile element insertion had occurred in distinct human populations. In addition, there is no known mechanism for the specific removal of SINES or LINES from the genome and we have found no evidence for the removal of these elements from the genome. Thus, we can say that these elements are extremely well-suited to examination of human population genetics and primate phylogenetic relationships.

- Botstein D, White RL, Skolnick M, Davis RW: Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314–331 (1980).
- Brookes AJ: The essence of SNPs. *Gene* 234:177–186 (1999).
- Cantrell MA, Filanoski BJ, Ingermann AR, Olsson K, DiLuglio N, Lister Z, Wichman HA: An ancient retrovirus-like element contains hot spots for SINE insertion. *Genetics* 158:769–777 (2001).
- Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES: Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 22:231–238 (1999).
- Carroll ML, Roy-Engel AM, Nguyen SV, et al: Large-scale analysis of the Alu Ya5 and Yb8 subfamilies and their contribution to human genomic diversity. *J Mol Biol* 311:17–40 (2001).
- Cavalli-Sforza LL, Menozzi P, Piazza A: The history and geography of human genes (Princeton University Press, Princeton 1994).
- Chakravarti A: It's raining SNPs, hallelujah? *Nat Genet* 19:216–217 (1998).
- Chakravarti A: Population genetics – making sense out of sequence. *Nat Genet* 21:56–60 (1999).
- Clark AG, Weiss KM, Nickerson DA, Taylor SL, Buchanan A, Stengård J, Salomaa V, Vartiainen E, Perola M, Boerwinkle E, Sing CF: Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. *Am J Hum Genet* 63:595–612 (1998).
- Comas D, Calafell F, Benchemsi N, Helal A, Lefranc G, Stoneking M, Batzer MA, Bertranpetit J, Sajantila A: Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. *Hum Genet* 107:312–319 (2000).
- Dewannieux M, Esnault C, Heidmann T: LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* 35:41–48 (2003).
- Dorit RL, Akashi H, Gilbert W: Absence of polymorphism at the ZFY locus on the human Y chromosome. *Science* 268:1183–1185 (1995).
- Edwards MC, Gibbs RA: A human dimorphism resulting from loss of an Alu. *Genomics* 14:590–597 (1992).
- Feng Q, Moran JV, Kazazian HH, Jr, Boeke JD: Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 87:905–916 (1996).
- Fu YX, Li WH: Statistical tests of neutrality of mutations. *Genetics* 133:693–709 (1993).
- Fullerton SM, Bond J, Schneider JA, Hamilton B, Harding RM, Boyce AJ, Clegg JB: Polymorphism and divergence in the beta-globin replication origin initiation region. *Mol Biol Evol* 17:179–188 (2000).
- Hacia JG, Fan JB, Ryder O, Jin L, Edgemon K, Ghandour G, Mayer RA, Sun B, Hsie L, Robbins CM, Brody LC, Wang D, Lander ES, Lipshutz R, Fodor SP, Collins FS: Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. *Nat Genet* 22:164–167 (1999).
- Hamdi H, Nishio H, Zielinski R, Dugaiczky A: Origin and phylogenetic distribution of Alu DNA repeats: irreversible events in the evolution of primates. *J Mol Biol* 289:861–871 (1999).
- Hammer MF: A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies. *Mol Biol Evol* 11:749–761 (1994).
- Harding RM, Fullerton SM, Griffiths RC, Bond J, Cox MJ, Schneider JA, Moulin DS, Clegg JB: Archaic African and Asian lineages in the genetic ancestry of modern humans. *Am J Hum Genet* 60:772–789 (1997).
- Harris EE, Hey J: X chromosome evidence for ancient human histories. *Proc Natl Acad Sci USA* 96:3320–3324 (1999).
- Hillis DM: SINEs of the perfect character. *Proc Natl Acad Sci USA* 96:9979–9981 (1999).
- Jaruzelska J, Zietkiewicz E, Labuda D: Is selection responsible for the low level of variation in the last intron of the ZFY locus? *Mol Biol Evol* 16:1633–1640 (1999).
- Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA: The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am J Hum Genet* 66:979–988 (2000).
- Jurka J: Sequence patterns indicate an enzymatic involvement in integration of mammalian retrotransposons. *Proc Natl Acad Sci USA* 94:1872–1877 (1997).
- Kajikawa M, Okada N: LINES mobilize SINEs in the eel through a shared 3' sequence. *Cell* 111:433–444 (2002).
- Kass DH, Raynor ME, Williams TM: Evolutionary history of B1 retrotransposons in the genus *Mus*. *J Mol Evol* 51:256–264 (2000).
- Kazazian HH: Retrotransposon insertions in germ cells and somatic cells. *Dev Biol (Basel)* 106:307–329 (2001).
- Lander ES, Linton LM, Birren B, et al: Initial sequencing and analysis of the human genome. *Nature* 409:860–921 (2001).
- Lawrence JG, Dykhuizen DE, DuBose RF, Hartl DL: Phylogenetic analysis using insertion sequence fingerprinting in *Escherichia coli*. *Mol Biol Evol* 6:1–14 (1989).
- Li WH, Sadler LA: Low nucleotide diversity in man. *Genetics* 129:513–523 (1991).
- Murata S, Takasaki N, Saitoh M, Okada N: Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution. *Proc Natl Acad Sci USA* 90:6995–6999 (1993).
- Myers JS, Vincent BJ, Udall H, Watkins WS, Morrish TA, Kilroy GE, Swergold GD, Henke J, Henke L, Moran JV, Jorde LB, Batzer MA: A comprehensive analysis of recently integrated human Ta L1 elements. *Am J Hum Genet* 71:312–326 (2002).
- Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, Martin C, Fujimoto E, Hoff M, Kuhlman E, White R: Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* 235:1616–1622 (1987).
- Nasidze I, Risch GM, Robichaux M, Sherry ST, Batzer MA, Stoneking M: Alu insertion polymorphisms and the genetic structure of human populations from the Caucasus. *Eur J Hum Genet* 9:267–272 (2001).
- Nikaïdo M, Matsuno F, Abe H, Shimamura M, Hamilton H, Matsubayashi H, Okada N: Evolution of CHR-2 SINEs in cetartiodactyl genomes: possible evidence for the monophyletic origin of toothed whales. *Mamm Genome* 12:909–915 (2001a).
- Nikaïdo M, Matsuno F, Hamilton H, Cao Y, Brownell RL, Ding W, Zuoyan Z, Shedlock AM, Hasegawa M, Okada N: Retroposon analysis of major cetacean lineages: the monophyly of toothed whales and the paraphyly of river dolphins. *Proc Natl Acad Sci USA* 98:7384–7389 (2001b).
- Nikaïdo M, Rooney AP, Okada N: Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: hippopotamuses are the closest extant relatives of whales. *Proc Natl Acad Sci USA* 96:10261–10266 (1999).
- Novick GE, Novick CC, Yunis J, Yunis E, Antunez de Mayolo P, Scheer WD, Deininger PL, Stoneking M, York DS, Batzer MA, Herrera RJ: Polymorphic Alu insertions and the Asian origin of Native American populations. *Hum Biol* 70:23–39 (1998).
- Pennisi E: A closer look at SNPs suggests difficulties. *Science* 281:1787–1789 (1998).
- Parra E, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrell RE, Shriver MD: Estimating African-American admixture proportions by use of population-specific alleles. *Am J Hum Genet* 63:1839–1851 (1998).
- Perna NT, Batzer MA, Deininger PL, Stoneking M: Alu insertion polymorphism: a new type of marker for human population studies. *Hum Biol* 64:641–648 (1992).
- Prak ET, Kazazian HH: Mobile elements and the human genome. *Nat Rev Genet* 1:134–144 (2000).
- Romualdi C, Balding D, Nasidze IS, Risch G, Robichaux M, Sherry ST, Stoneking M, Batzer MA, Barbujani G: Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Res* 12:602–612 (2002).
- Roy AM, Carroll ML, Kass DH, Nguyen SV, Salem AH, Batzer MA, Deininger PL: Recently integrated human Alu repeats: finding needles in the haystack. *Genetica* 107:149–161 (1999).
- Roy AM, Carroll ML, Nguyen SV, Salem AH, Oldridge M, Wilkie AO, Batzer MA, Deininger PL: Potential gene conversion and source genes for recently integrated Alu elements. *Genome Res* 10:1485–1495 (2000).
- Roy-Engel AM, Carroll ML, Vogel E, Garber RK, Nguyen SV, Salem AH, Batzer MA, Deininger PL: Alu insertion polymorphisms for the study of human genomic diversity. *Genetics* 159:279–290 (2001).
- Roy-Engel AM, Carroll ML, El-Sawy M, Salem AH, Garber RK, Nguyen SV, Deininger PL, Batzer MA: Non-traditional Alu evolution and primate genomic diversity. *J Mol Biol* 316:1033–1040 (2002a).
- Roy-Engel AM, Salem AH, Oyeniran OO, Deininger L, Hedges DJ, Kilroy GE, Batzer MA, Deininger PL: Active Alu element "A-tails": size does matter. *Genome Res* 12:1333–1344 (2002b).
- Rozas J, Rozas R: DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15:174–175 (1999).
- Ryan SC, Dugaiczky A: Newly arisen DNA repeats in primate phylogeny. *Proc Natl Acad Sci USA* 86:9360–9364 (1989).
- Salem AH, Kilroy GE, Watkins WS, Jorde LB, Batzer MA: Recently integrated Alu elements and human genomic diversity. *Mol Biol Evol* 20:1349–1361 (2003a).
- Salem AH, Myers JS, Otieno AC, Watkins WS, Jorde LB, Batzer MA: LINE-1 pre-Ta elements in the human genome. *J Mol Biol* 326:1127–1146 (2003b).
- Salem AH, Ray DA, Xing J, Callinan PA, Myers JS, Hedges DJ, Garber RK, Witherspoon DJ, Jorde LB, Batzer MA: Alu elements and hominid phylogenetics. *Proc Natl Acad Sci USA* 100:12787–12791 (2003c).
- Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463–5467 (1977).
- Schiebel K, Meder J, Rump A, Rosenthal A, Winkelmann M, Fischer C, Bonk T, Humeny A, Rappold G: Elevated DNA sequence diversity in the genomic region of the phosphatase PPP2R3L gene in the human pseudoautosomal region. *Cytogenet Cell Genet* 91:224–230 (2000).

- Schneider S, Kueffer JM, Roessli D, Excoffier L: Arlequin: A software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland (1997).
- Shedlock AM, Okada N: SINE Insertions, powerful tools for molecular systematics. *Bioessays* 22:148–60 (2000).
- Sheen FM, Sherry ST, Risch GM, Robichaux M, Nasidze I, Stoneking M, Batzer MA, Swergold GD: Reading between the LINEs: human genomic variation induced by LINE-1 retrotransposition. *Genome Res* 10:1496–1508 (2000).
- Shen MR, Batzer MA, Deininger PL: Evolution of the master Alu gene(s). *J Mol Evol* 33:311–320 (1991).
- Sherry ST, Harpending HC, Batzer MA, Stoneking M: Alu evolution in human populations: using the coalescent to estimate effective population size. *Genetics* 147:1977–1982 (1997).
- Sinnett D, Richer C, Deragon JM, Labuda D: Alu RNA transcripts in human embryonal carcinoma cells. Model of post-transcriptional selection of master sequences. *J Mol Biol* 226:689–706 (1992).
- Smit AF: Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr Opin Genet Dev* 9:657–663 (1999).
- Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot SS, Saha N, Jenkins T, Tahir MA, Deininger PL, Batzer MA: Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res* 7:1061–1071 (1997).
- Tajima F: Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595 (1989).
- Tajima F: The amount of DNA polymorphism maintained in a finite population when the neutral mutation rate varies among sites. *Genetics* 143:1457–1465 (1996).
- Venter JC, Adams MD, Myers EW, et al: The sequence of the human genome. *Science* 291:1304–1351 (2001).
- Vincent BJ, Myers JS, Ho HJ, Kilroy GE, Walker JA, Watkins WS, Jorde LB, Batzer MA: Following the LINEs: an analysis of primate genomic variation at Human-specific LINE-1 insertion sites. *Mol Biol Evol* 20:1338–1348 (2003).
- Watkins WS, Ricker CE, Bamshad MJ, Carroll ML, Nguyen SV, Batzer MA, Harpending HC, Rogers AR, Jorde LB: Patterns of ancestral human diversity: an analysis of Alu-insertion and restriction-site polymorphisms. *Am J Hum Genet* 68:738–752 (2001).
- Watkins WS, Rogers AR, Ostler CT, Bamshad MJ, Brassington AME, Carroll ML, Nguyen SV, Walker JA, Prasad B, VR, Reddy PG, Das PK, Batzer MA, Jorde LB: Genetic variation among world populations: Inference from 100 Alu insertion polymorphisms. *Genome Res* 13:1607–1618 (2003).
- Wu X, Di Rienzo A, Ober C: A population genetics study of single nucleotide polymorphisms in the interleukin 4 receptor alpha (IL4RA) gene. *Genes Immun* 2:128–134 (2001).
- Zietkiewicz E, Yotova V, Jarnik M, Korab-Laskowska M, Kidd KK, Modiano D, Scozzari R, Stoneking M, Tishkoff S, Batzer M, Labuda D: Nuclear DNA diversity in worldwide distributed human populations. *Gene* 205:161–171 (1997).