

Identification and characterization of two polymorphic Ya5 Alu repeats

Santosh S. Arcot,^{a,1} Margaret M. DeAngelis,^b Stephen T. Sherry,^b Aaron W. Adamson,^a Jane E. Lamerdin,^a Prescott L. Deininger,^{c,d} Anthony V. Carrano^a & Mark A. Batzer^{b,*}

^a Human Genome Center, L-452, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA 94551, USA

^b Department of Pathology, Department of Biometry and Genetics, Stanley S. Scott Cancer Center, Neuroscience Center of Excellence, Louisiana State University Medical Center, 1901 Perdido Street, New Orleans, LA 70112, USA

^c Department of Biochemistry and Molecular Biology, Neuroscience Center of Excellence, Louisiana State University Medical Center, New Orleans, LA 70112, USA

^d Laboratory of Molecular Genetics, Alton Ochsner Medical Foundation, New Orleans, LA 70121, USA

Accepted 4 February 1997

Two new polymorphic Alu elements (HS2.25 and HS4.14) belonging to the young (Ya5/8) subfamily of human-specific Alu repeats have been identified. DNA sequence analysis of both Alu repeats revealed that each Alu repeat had a long 3'-oligo-dA-rich tail (41 and 52 nucleotides in length) and a low level of random mutations. HS2.25 and HS4.14 were flanked by short precise direct repeats of 8 and 14 nucleotides in length, respectively. HS2.25 was located on human chromosome 13, and HS4.14 on chromosome 1. Both Alu elements were absent from the orthologous positions within the genomes of non-human primates, and were highly polymorphic in a survey of twelve geographically diverse human groups.

Keywords: Genetic variation; Mobile elements; Insertion polymorphism; Human-specific

1. Introduction

The most successful class of repeated DNA sequences in primate genomes is the Alu family of interspersed repeats which have arisen in the last 65 million years of evolution (reviewed in [1–3]). Alu repeats belong to a class of sequences defined as short interspersed elements (SINEs). Alu elements are approximately 300 nucleotides in length, dimeric in structure, and contain a middle A-rich region along with a 3'-oligo-dA-rich tail and short flanking direct repeats. Approximately 500,000 Alu SINEs exist within the human genome, representing about 5% of the genome by mass.

Alu repeats are ancestrally derived from the 7SL RNA gene and mobilize through an RNA polymerase-III-derived transcript in a process termed retroposition. The amplification of Alu elements has been dominated by a small number of 'master' genes as well as a few fortuitous source genes [4] resulting in the appearance of distinct subfamilies as defined by 'diagnostic' or subfamily-specific sequence variation (reviewed in [5]). The youngest Alu subfamilies termed Ya5/8 and Yb8 [5], appeared around the time humans diverged from other primates, therefore members of these subfamilies are largely restricted to humans [6–12]. A number of polymorphic Alu insertions have been identified [11–21]. Polymorphic members of the evolutionarily young Alu subfamilies are novel markers for the study of human evolution and forensics. In this study, we report the identification and analysis of two polymorphic members of the Ya5/8 subfamily of Alu repeats.

* Corresponding author. Tel.: +1 504 568-6075; Fax: +1 504 568-6037; E-mail: mbatze@lsu.edu

¹ Present address: Affymetrix, 3380 Central Expressway, Santa Clara, CA 95051, USA.

2. Methods

2.1. Cell lines and DNA samples

Cell lines used to isolate DNA samples were previously described [21]. Additional DNA samples from five individual chimpanzees (*Pan troglodytes*), one gorilla (*Gorilla gorilla*), three orangutans (*Pongo pygmaeus*), one macaque (*Macaca fascicularis*), and one tamarin (*Saguinus oedipus*) were obtained from BIOS laboratories (New Haven, CT). Human DNA samples from throughout the world were isolated from peripheral lymphocytes [22] available from previous studies [12,14].

2.2. Library construction and screening

Construction of a randomly sheared total genomic library in bacteriophage λ ZAP II (Stratagene, La Jolla, CA), screening with an oligonucleotide probe specific for the Ya5 subfamily of Alu elements, plaque purification and plasmid rescue have been described previously [18].

2.3. DNA sequence analysis

Plasmid templates for sequencing were prepared by the Qiagen mini-alkaline lysis method according to the manufacturer's instructions (Qiagen, Chatsworth, CA). Double-stranded plasmid templates corresponding to the positive clones from the HeLa genomic library were sequenced with internal Alu-specific primers as previously described [21]. Fluorescently labeled M13 forward and reverse primers and AB *Taq* cycle sequencing kits were used to sequence PCR products corresponding to each Alu element that had been cloned in the pCR II TA-cloning vector (Invitrogen, San Diego, CA). Sequencing reactions were fractionated on a 6% polyacrylamide gel, followed by data collection and analysis on an ABI 373A DNA sequencer. The sequences reported in this manuscript have been assigned GenBank accession numbers U67211 (HS2.25) and U67221 (HS4.14).

2.4. Primer design, PCR amplification and chromosomal localization

Sequences flanking the Alu elements were screened for the presence of human repetitive sequences and PCR primers were designed using the PRIMER software (Whitehead Institute for Biomedical Research, Cambridge, MA) as previously described [21]. PCR amplification was carried out as previously described [21]. PCR products were directly visualized following agarose gel electrophoresis using UV fluorescence. Alu family member HS2.25 was amplified using the primers HS2.25 5',5'-CAGCTAAGAGATGGAATCAACC-3' and HS2.25 3',5'-ATCCATTAACCAGCCTCTCC-3'. Amplification of the HS2.25 locus results in the production of a 658 bp fragment containing the Alu repeat or a 302 bp fragment without the Alu element. Alu family member HS4.14 was amplified using the primers HS4.14 5',5'-AAGAGAAGATCAGTAGG-3' and HS4.14 3',5'-TATGGCTCAGAATACAG-3'. Amplification of the HS4.14 locus results in the production of a 484 bp fragment containing the Alu repeat or a 151 bp fragment without the Alu element. The annealing temperatures for HS2.25 and HS4.14 were 67°C and 53°C, respectively. Phylogenetic analysis of the Alu elements was determined by PCR amplification of human and fifteen non-human primate DNA samples as previously described [21]. The chromosomal location of each Alu repeat was determined by PCR amplification of NIGMS human/rodent somatic cell hybrid mapping panels 1 and 2 (Coriell Institute for Medical Research, Camden, NJ).

3. Results and discussion

3.1. DNA sequence analysis

A randomly sheared total genomic human library was screened with an oligonucleotide probe harboring the diagnostic mutations specific for the Ya5/8 subfamily of Alu repeats as a part of a directed strategy to identify additional members of the subfamily [10]. Several hundred clones were randomly selected and arrayed into microtiter plates after screening. Two clones were sequenced using internal Alu-specific primers to obtain the sequence of the regions flanking the Alu repeat [10]. The flanking sequences of each Alu element were verified by comparing the sequences of the direct repeats of each Alu element, since the majority of young Alu elements are flanked by perfect direct repeats as compared to older

A

CON	GGCCGGGGCGC	GGTGGCTCAC	GCCTGTAATC	CCAGCACTTT	GGGAGGCCGA	50
HS2.25	50
HS4.14	50
CON	GGCGGGCGGA	TCACGAGGTC	AGGAGATCGA	GACCATCCCG	GCTAAAACGG	100
HS2.25	100
HS4.14A.....	100
CON	TGAAACCCCG	TCTCTACTAA	AAA---TACA	AAAA----T	TAGCCGGGCG	150
HS2.25AAA..	146
HS4.14AAA...AAAA.	150
CON	TAGTGGCGGG	CGCCTGTAGT	CCCAGCTACT	TGGGAGGCTG	AGGCAGGAGA	200
HS2.25	196
HS4.14	200
CON	ATGGCGTGAA	CCCGGGAGGC	GGAGCTTGCA	GTGAGCCGAG	ATCCCGCCAC	250
HS2.25A.....	A.....A..A....	246
HS4.14	250
CON	TGCACTCCAG	CCTGGGCGAC	AGAGCGAGAC	TCCGTCTC		288
HS2.25G...T.....	..T.....		284
HS4.14		288

B

HS2.25	<u>ATTCACCA</u> [ALU] (A) ₄₁	<u>ATTCACCA</u>
HS4.14	<u>GAAAGTAGAGTTGA</u> [ALU] (A) ₅₂	<u>GAAAGTAGAGTTGA</u>

Fig. 1. Sequences of the Alu elements and their flanking nucleotides. (A) Alignment of the sequences of HS2.25 and HS4.14 with the consensus sequence (CON) for the Ya5 subfamily. The five diagnostic nucleotides characteristic of the Ya5 subfamily are underlined in the consensus sequence. Nucleotide identities are indicated by dots while substitutions are indicated by appropriate nucleotides. (B) Flanking nucleotide sequences. The composition of the 3'-oligo-dA tail is denoted by the appropriate nucleotide in parentheses followed by the length in subscripts. The nucleotide sequences of the direct repeats of each Alu element are underlined.

Alu elements in which the flanking direct repeats have been subjected to random mutations after the genomic integration of these elements [7,10,23].

Oligonucleotide primers complementary to the 5' and 3' unique sequence regions flanking each Alu repeat were designed and optimized for their specificity in PCR reactions using total human or chimpanzee genomic DNAs as templates for the PCR. The PCR primers were used in a series of experiments to determine the phylogenetic distribution, chromosomal localization and insertion polymorphism of each Alu element (outlined below). Finished DNA sequences from the polymorphic Alu repeats were generated and compared to the subfamily consensus sequence. An alignment of the sequences of these two polymorphic Alu elements with the Ya5 subfamily consensus sequence is shown in Fig. 1A.

Both of the Alu elements had all five diagnostic mutations characteristic of the Ya5 subfamily (underlined in the consensus sequence) (Fig. 1A). Alu element HS4.14 only had a single additional nucleotide substitution at a CpG position, but also contained two insertions within the middle A-rich region of the element. The insertions/expansions in the middle A-rich region of HS4.14 are not surprising since the middle A-rich regions of Alu repeats serve as nuclei for the genesis of microsatellite elements [23]. Subfamily member HS2.25 had six mutations at CpG dinucleotides, one non-CpG mutation and an insertion within the middle A-rich region of the element in addition to the diagnostic mutations. The large number of CpG substitutions may reflect the fact that this element resides within a region of the genome that is subject to a large amount of methylation. The CpG positions within Alu repeats generally mutate at a rate which is 9.2-fold higher than the non-CpG positions as a result of the spontaneous deamination of 5-methyl cytosine [10,24]. A total of 1 out of 480 non-CpG nucleotides is mutated within the two dimorphic Alu repeats (excluding the three insertions) or an average of 0.5 ± 0.71 substitutions in each Alu repeat as compared to the subfamily consensus sequence. Using a neutral rate of evolution of 0.15% per million years [25], the average age of the two polymorphic Alu elements is approximately 1.39 ± 1.97 million years. This estimate compares well to the previously reported age of the Ya5 subfamily of 2.8 million years, since the previous estimate included both polymorphic and monomorphic Alu family members [10]. A direct comparison to the calculated age of several polymorphic Alu repeats of 1.63 million years [21] is also possible

since HS2.25 and HS4.14 are both polymorphic.

Both Alu elements HS2.25 and HS4.14 contained homopolymeric A-rich tails composed exclusively of adenine residues. Homopolymeric A-rich tails are characteristic of recent Alu insertions [10]. In addition, both of the elements were flanked by perfect direct repeats 8 and 14 bp in length for HS2.25 and HS4.14, respectively (Fig. 1B). Perfect direct repeats are a characteristic of young Alu repeats [10,11,13,21,23,26].

3.2. *Phylogenetic analyses and chromosomal locations*

Phylogenetic analyses were performed using DNA samples from humans as well as fifteen non-human primates as templates for the PCR. The Ya5 subfamily of Alu repeats began to amplify around the time humans diverged from non-human primates (approximately 5 million years ago) [8,10], therefore members of this subfamily are primarily restricted in their distribution to humans. However, a few Ya lineage Alu repeats have been detected in non-human primates [27,28]. Both HS2.25 and HS4.14 were found exclusively in humans, and were absent from the genomes of fifteen non-human primates (data not shown). These data are consistent with our previous observations that Ya5 subfamily members are predominantly restricted to the human genome [8,11–13,21]. Using DNA samples isolated from the NIGMS human/rodent somatic cell hybrid mapping panels 1 and 2 we determined the chromosomal location of HS2.25 and HS4.14 to be chromosomes 13 and 1, respectively (data not shown).

3.3. *Human genomic variation*

Polymorphic Alu elements are unique nuclear markers for the study of human genetic diversity [12,14,16,29,30]. The insertion of each Alu element within the human genome is a unique event which occurs in a single individual. Since the insertion of an Alu element into the genome is a unique event, the polymorphic Alu elements are identical by descent from a common ancestor [14]. Other commonly used polymorphic markers such as restriction fragment length polymorphisms (RFLP) and variable number of tandem repeats (VNTR) are only identical by state. In addition, the ancestral state of each polymorphic Alu element is known to be the absence of the Alu repeat. Knowledge about the ancestral state of each Alu repeat facilitates the rooting of trees of population relationships [12,14].

Previous studies have utilized polymorphic Alu elements to study human genetic diversity and evolution [12,14,16,29–31]. To ascertain the utility of these polymorphic Alu elements as tools for population studies, we determined the distribution of the two polymorphic elements in individuals from twelve different population groups using PCR-based assays for the presence/absence of each Alu repeat (Table 1). Each locus was in Hardy-Weinberg equilibrium for all populations except HS2.25 in French, French-Acadians, Syrians and U.S. Caucasians, as judged by independent chi-square tests for goodness of fit. One deviation from Hardy-Weinberg equilibrium would be expected by chance alone since 24 tests for goodness of fit were performed. In each case, the deviations were due to an excess of heterozygous individuals or absence of homozygous individuals which contained the Alu repeat. Both Alu elements were polymorphic in all twelve human populations tested.

The frequency of Alu element HS2.25 ranged from 0.129 in African-Americans to 0.362 in the French Caucasians while HS4.14 varied from 0.573 in Pakistani's to 0.865 in U.S. Caucasians. The lower frequencies of HS2.25 in the twelve populations suggests that it may have arisen more recently as a genomic fossil than HS4.14. The average heterozygosity of HS4.14 was 0.383 while HS2.25 was 0.350. This is quite high considering that these are bi-allelic polymorphisms with a maximum heterozygosity of 50%. F_{st} values, which are a measure of population differentiation, were 0.021 and 0.017 for HS2.25 and HS4.14, respectively. These F_{st} values are low as compared to previous studies of Alu insertion polymorphisms [12,14,29], presumably as a result of the absence of African and Asian populations from the analysis. Both HS2.25 and HS4.14 should prove to be very useful polymorphisms for the study of human population genetics.

Acknowledgements

This research was supported in part by a grant from the National Institutes of Health (HG 00770) to P.L.D. Work at Lawrence Livermore National Laboratory was performed under the auspices of the U.S. Department of Energy under contract No. W-7405-ENG-48.

References

- [1] Deininger, P.L. (1989) SINEs: short interspersed repeated DNA elements in higher eukaryotes. In: D.E. Berg and M.M. Howe (eds.), *Mobile DNA*, American Society for Microbiology, Washington, D.C., pp. 619–636.

Table 1. Frequencies and heterozygosities for Alu family members HS2.25 and HS4.14

Population	<i>n</i>	Frequency of (+)	No. +/+	No. +/-	No. -/-	<i>h</i> ±SE (<i>h</i>)
<i>HS 2.25</i>						
African American	31	0.129	0	8	23	0.228
Alaska Native	42	0.202	2	13	27	0.327
Breton	33	0.212	0	14	19	0.339
Cypriot	47	0.255	2	20	25	0.384
French ^a	47	0.362	0	34	13	0.457
French Acadian ^b	45	0.289	0	26	19	0.415
Greenland Native	36	0.181	1	11	24	0.300
Hispanic American	34	0.176	2	8	24	0.295
Pakistani	48	0.188	1	16	31	0.308
Swiss ^c	20	0.275	3	5	12	0.409
U.S. Caucasian ^d	46	0.239	0	22	24	0.368
Total	477	0.233	11	200	266	0.357
Avg. heterozygosity						0.350±0.179
Avg. allele frequency		0.229				
Fst		0.021				
<i>HS 4.14</i>						
African American	48	0.698	26	15	7	0.426
Alaska Native	40	0.738	23	13	4	0.392
Breton	38	0.750	23	11	4	0.380
Cypriot	48	0.677	24	17	7	0.442
French	37	0.622	17	12	8	0.477
French Acadian	48	0.823	33	13	2	0.295
Greenland Native	39	0.756	24	11	4	0.373
Hispanic American	44	0.852	33	9	2	0.255
Pakistani	41	0.573	16	15	10	0.495
Swiss	39	0.692	20	14	5	0.432
Syrian	47	0.734	23	23	1	0.395
U.S. Caucasian	48	0.865	37	9	2	0.237
Total	517	0.735	299	162	56	0.390
Avg. heterozygosity						0.383 ± 0.149
Avg. allele frequency		0.732				
Fst		0.017				

^a Deviates from Hardy-Weinberg equilibrium (HWE) proportions ($\chi^2 = 14.33$, d.f. = 1, $p < 0.005$).

^b Deviates from HWE proportions ($\chi^2 = 7.43$, d.f. = 1, $p < 0.01$).

^c Deviates from HWE proportions ($\chi^2 = 4.77$, d.f. = 1, $p < 0.05$).

^d Deviates from HWE proportions ($\chi^2 = 4.54$, d.f. = 1, $p < 0.05$).

- [2] Deininger, P.L. and M.A. Batzer (1993) Evolution of Retroposons. *Evol. Biol.*, 27, 157–196.
- [3] Schmid, C.W. and R. Maraia (1992) Transcriptional and transpositional selection of active SINE sequences. *Curr. Opin. Genet. Dev.*, 2, 874–882.
- [4] Deininger, P.L., M.A. Batzer, C.A. Hutchison III and M.H. Edgell (1992) Master genes in mammalian repetitive DNA amplification. *Trends Genet.*, 8, 307–311.
- [5] Batzer, M.A., P.L. Deininger, U. Hellmann-Blumberg, J. Jurka, D. Labuda, C.M. Rubin, C.W. Schmid, E. Zietkiewicz and E. Zukerkandl (1996) Standardized nomenclature for Alu repeats. *J. Mol. Evol.*, 42, 3–6.
- [6] Matera, A.G., U. Hellmann and C.W. Schmid (1990) A transpositionally and transcriptionally competent Alu subfamily. *Mol. Cell. Biol.*, 10, 5424–5432.
- [7] Matera, A.G., U. Hellmann, M.F. Hintz and C.W. Schmid (1990) Recently transposed Alu repeats result from multiple source genes. *Nucleic Acids Res.*, 18, 6019–6023.
- [8] Batzer, M.A. and P.L. Deininger (1991) A human-specific subfamily of Alu sequences. *Genomics*, 9, 481–487.
- [9] Jurka, J. (1993) A new subfamily of recently retroposed Alu repeats. *Nucleic Acids Res.*, 21, 2252.
- [10] Batzer, M.A., G.E. Kilroy, P.E. Richard, T.H. Shaikh, T.D. Desselle, C.L. Hoppens and P.L. Deininger (1990) Structure and variability of recently inserted Alu family members. *Nucleic Acids Res.*, 18, 6793–6798.
- [11] Batzer, M.A., C.M. Rubin, U. Hellmann-Blumberg, M. Alegria-Hartman, E.P. Leeflang, J.D. Stern, H.A. Bazan, T.H. Shaikh, P.L. Deininger and C.W. Schmid (1995) Dispersion and insertion polymorphism in two small subfamilies of recently amplified human Alu repeats. *J. Mol. Biol.*, 247, 418–427.
- [12] Batzer, M.A., S.S. Arcot, J.W. Phinney, M. Alegria-Hartman, D.H. Kass, S.M. Milligan, C. Kimpton, P. Gill, M. Hochmeister, P.A. Ioannou, R.J. Herrera, D.A. Boudreau, W.D. Scheer, B.J.B. Keats, P.L. Deininger and M. Stoneking (1996) Genetic variation of recent Alu insertions in human populations. *J. Mol. Evol.*, 42, 22–29.
- [13] Batzer, M.A., V.A. Gudi, J.C. Mena, D.W. Foltz, R.J. Herrera and P.L. Deininger (1991) Amplification dynamics of human-specific Alu family members. *Nucleic Acids Res.*, 19, 3619–3623.
- [14] Batzer, M.A., M. Stoneking, M. Alegria-Hartman, H. Bazan, D.H. Kass, T.H. Shaikh, G.E. Novick, P.A. Ioannou, W.D. Scheer, R.J. Herrera and P.L. Deininger (1994) African origin of human-specific polymorphic Alu insertions. *Proc. Natl. Acad. Sci. USA*, 91, 12288–12292.
- [15] Blondin, L.A.J., G.M. Terwindt, J.T. Den Dunnen and G.-J.B. Van Ommen (1994) A polymorphic STS in intron 44 of the dystrophin gene. *Hum. Genet.*, 93, 479–480.

- [16] Hammer, M.F. (1994) 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.
- [17] Milewicz, D.M., P.H. Byers, J. Reveille, A.L. Hughes and M. Duvic (1996) A dimorphic Alu Sb-like insertion is ethnic-specific. *J. Mol. Evol.*, 42, 117–123.
- [18] Muratani, K., T. Hada, Y. Yamamoto, T. Kaneko, Y. Shigeto, T. Ohue, J. Furuyama and K. Higashino (1991) Inactivation of the cholinesterase gene by Alu insertion: Possible mechanism for human gene transposition. *Proc. Natl. Acad. Sci. USA*, 88, 11315–11319.
- [19] Vidaud, D., M. Vidaud, B.R. Bahnak, V. Siguret, S.G. Sanchez, Y. Laurian, D. Meyer, M. Goossens and J.M. Lavergne (1993) Hemophilia B due to a *de novo* insertion of human-specific Alu subfamily member within the coding region of the factor IX gene. *Eur. J. Hum. Genet.*, 1, 30–36.
- [20] Wallace, M.R., L.B. Andersen, A.M. Saulino, P.E. Gregory, T.W. Glover and F.S. Collins (1991) A *de novo* Alu insertion results in neurofibromatosis type I. *Nature*, 353, 864–866.
- [21] Arcot, S.S., A.W. Adamson, J.E. Lamerdin, B. Kanagy, P.L. Deininger, A.V. Carrano and M.A. Batzer (1996) Alu fossil relics-distribution and insertion polymorphism. *Genome Res.*, 6, 1084–1092.
- [22] Ausabel, F.M., R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith and K. Struhl (1987) *Current protocols in molecular biology*. John Wiley and Sons, New York.
- [23] Arcot, S.S., T.H. Shaikh, J. Kim, L. Bennett, M. Alegria-Hartman, D.O. Nelson, P.L. Deininger and M.A. Batzer (1995) Sequence diversity and chromosomal distribution of ‘young’ Alu repeats. *Gene*, 163, 273–278.
- [24] Labuda, D. and G. Striker (1989) Sequence conservation in Alu evolution. *Nucleic Acids Res.*, 17, 2477–2491.
- [25] Miyamoto, M.M., J.L. Slightom and M. Goodman (1987) Phylogenetic relations of humans and African apes from DNA sequences in the pseudo- η -globin region. *Science*, 238, 369–373.
- [26] Arcot, S.S., J.J. Fontius, P.L. Deininger and M.A. Batzer (1995) Identification and analysis of a ‘young’ polymorphic Alu element. *Biochem. Biophys. Acta*, 1263, 99–102.
- [27] Leeftang, E.P., I. Chesnokov and C.W. Schmid (1993) Mobility of short interspersed repeats within the chimpanzee lineage. *J. Mol. Evol.*, 37, 566–572.
- [28] Leeftang, E.P., W.-M. Liu, C. Hashimoto, P.V. Choudary and C.W. Schmid (1992) Phylogenetic evidence for multiple Alu source genes. *J. Mol. Evol.*, 35, 7–16.
- [29] Perna, N.T., M.A. Batzer, P.L. Deininger and M. Stoneking (1992) Alu insertion polymorphism: a new type of marker for human population studies. *Hum. Biol.*, 64, 641–648.
- [30] Tishkoff, S.A., G. Ruano, J.R. Kidd and K.K. Kidd (1996) Distribution and frequency of a polymorphic Alu insertion at the plasminogen activator locus in humans. *Hum. Genet.*, 97, 759–764.
- [31] Tishkoff, S.A., E. Dietzsch, W. Speed, A.J. Pakstis, J.R. Kidd, K. Cheung, B. Bonne-Tamir, A.S. Santachiara-Benerecetti, P. Moral, M. Krings, S. Paabo, E. Watson, N. Risch, T. Jenkins and K.K. Kidd (1996) Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science*, 271, 1380–1397.