Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 4360–4364, April 1996
Evolution

DNA sequences of *Alu* elements indicate a recent replacement of the human autosomal genetic complement

(evolution/population genetics/nuclear alleles/nucleotide diversity)

ALEC KNIGHT*†, MARK A. BATZER‡§, MARK STONEKING¶, HEMANT K. TIWARI‖, W. DOUGLAS SCHEER§, RENE J. HERRERA**, AND PRESCOTT L. DEININGER*††

*Department of Biochemistry and Molecular Biology and the Center for Molecular and Human Genetics, §Department of Pathology, Stanley S. Scott Cancer Center, and ‖Department of Biometry and Genetics, Louisiana State University Medical Center, 1901 Perdido Street, New Orleans, LA 70112; ‡Human Genome Center, L-452, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, CA 94551; ¶Department of Anthropology, Pennsylvania State University, University Park, PA 16802; **Department of Biological Sciences, Florida International University, University Park Campus, Miami, FL 33199; and ††Laboratory of Molecular Genetics, Alton Ochsner Medical Foundation, New Orleans, LA 70121

Communicated by John C. Avise, University of Georgia, Athens, GA, January 2, 1996 (received for review June 16, 1995)

ABSTRACT **DNA sequences of neutral nuclear autosomal loci, compared across diverse human populations, provide a previously untapped perspective into the mode and tempo of the emergence of modern humans and a critical comparison with published clonally inherited mitochondrial DNA and Y chromosome measurements of human diversity. We obtained over 55 kilobases of sequence from three autosomal loci encompassing *Alu* repeats for representatives of diverse human populations as well as orthologous sequences for other hominoid species at one of these loci. Nucleotide diversity was exceedingly low. Most individuals and populations were identical. Only a single nucleotide difference distinguished presumed ancestral alleles from descendants. These results differ from those expected if alleles from divergent archaic populations were maintained through multiregional continuity. The observed virtual lack of sequence polymorphism is the signature of a recent single origin for modern humans, with general replacement of archaic populations.**

The fossil record documents the emergence of early modern human anatomy about 100,000 years ago (refs. 1; see various papers in ref. 2). Mitochondrial DNA (mtDNA) studies have revealed low levels of divergence among human populations and have been interpreted as indicative of a recent African origin for modern humans (3–5), providing evidence for the "Noah's Ark" hypothesis (6), a relatively recent worldwide replacement of archaic human populations by decendents of an original modern population. Alternatively, some paleontological evidence has been interpreted as supporting multiregional continuity of populations for a far greater time (7), spanning the relatively recent transition from archaic to modern forms.

Vertebrate mtDNA has been shown to sometimes introgress and become fixed in a pattern inconsistent with the evolution of populations (8–16), due to its mode of replication and inheritance as a single, clonal, maternal allele. Because mtDNA is inherited in this manner, its evolution is easily influenced by stochastic processes, its analysis is subject to stochastic errors (17, 18), and haplotype frequencies are vulnerable to selective sweeps. While extant mtDNA lineages trace to a single female, all other mtDNA lineages from the time of the mitochondrial ancestor have become extinct. The human mitochondrial genome is about 16.5 kb in length (19), with much of the variation limited to a hypervariable region of about 1 kb (20), providing useful, but limited, evolutionary information.

Conversely, nuclear DNA sequence data have the potential to complete the molecular perspective of human evolutionary history that was begun by mtDNA studies. The human nuclear genome is composed of about 3×10^9 bp (21) packaged into 23 chromosomes, providing essentially infinite variation for evolutionary reconstruction. All but a portion of the Y chromosome is biparentally inherited and subject to recombination. That small, haploid, paternally inherited, Y chromosome fraction of the nuclear genome evolves in a pattern similar to the mitochondrial genome and therefore has a similar recent theoretical coalescence date and is subject to similar problems of vulnerability to stochastic processes and selective sweeps. Recently, a survey of a paternally inherited Y chromosome intron locus (*ZFY*) among a sample of 38 genomes representing worldwide human populations revealed complete monomorphism (22). This finding helps to corroborate the mitochondrial evidence for a recent, single origin for modern humans. Low diversity in both mitochondrial-maternal and nuclear-paternal DNA sequences is best explained by the recent single origin hypothesis. However, due to the uniparental haploid nature of this nuclear locus, alternative explanations for the absence of polymorphism cannot be ruled out. These include "a recent selective sweep . . . recurrent male population bottlenecks, or historically small effective male population sizes" (22). The 4-fold slower fixation rate, recombination, and biparental transmission of autosomal nuclear DNA sequences allow many individuals who lived at the same time as the mtDNA or *ZFY* ancestor to have contributed to the autosomal alleles of modern humans. Also, these characteristics of inheritance inhibit the speed of selective sweeps in evolutionary time. Therefore, the presence of highly divergent alleles passed on from regional archaic populations might be detected at autosomal loci even though such alleles have not been found for uniparental loci. Human nuclear DNA variation has been assessed by indirect methods (23), providing valuable insights into human population phylogeny. To date, little is known regarding variation at the sequence level. The few reports of human autosomal nuclear nucleotide diversity were based on comparison of coding sequences among a very small number of individuals (24), a single geographic population (25), or loci which are subject to complex selection pressures (26, 27). Here we present a large autosomal nuclear DNA sequence data set of three neutral loci compared across diverse human populations and sequences of one of these loci for other hominoid species.

Alu is a primate-specific family of retroposable elements found at >500,000 loci in the human nuclear genome (28–30).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: Myr, million years; ACE, angiotensin I-converting enzyme; APO, apolipoprotein.

†Present address: Department of Biology, Sul Ross State University, Alpine, TX 79832.

The sequences reported in this paper have been deposited in the GenBank database (accession nos. L47222–L47229).

These elements average about 300 bp in length, are dispersed widely in the genome, and are thought to be generated by a small number of active master genes. *Alu* is ultimately derived from the 7SL RNA gene. In that sense, and in the sense that *Alu* repeats are generated by a small number of source, or master, genes, individual *Alu* repeats may be considered pseudogenes. *Alu* repeats are not generally subject to homogenization, but rather evolve independently once integrated. This is evident from comparison of nucleotide divergence at orthologous and paralogous loci (31, 32) and from the hierarchical subfamily structure of *Alu* sequences across primates that reflects the primate cladistic pattern.

The known antiquity of regional archaic human populations has recently been pushed back to well over 1.5 million years (Myr) by radiometric dates for two *Homo erectus* sites in Java (33) and by the startling discovery of an *H. erectus* mandible in southwestern Asia firmly dated at the Plio–Pleistocene boundary (34). Considering information available for evolutionary rate estimation from published comparisons of *Alu* repeats (31) and other noncoding genomic regions (35), we should expect to find considerable nucleotide differences in *Alu* alleles among the major human races if regional archaic alleles were maintained. That is, if human races evolved in widely separated geographic regions for over 1.5 Myr, *Alu* sequences would reveal substantial interallelic divergence and overall nucleotide diversity. On the other hand, if the human autosomal genetic complement has undergone a relatively recent worldwide replacement, we should expect only very low levels of differences among populations and individuals.

MATERIALS AND METHODS

PCR. Total cellular DNAs were on hand in our combined laboratory stocks. These had been purified by conventional methods for previous studies. These DNA samples were used as template for PCR amplifications using reagents supplied by Perkin–Elmer, including AmpliTaq cloned DNA polymerase. PCRs were carried out in a Perkin–Elmer GeneAmp PCR System 9600 thermal cycler. Thermal cycle parameters were denaturation at 94°C for 15 sec, annealing at 58°C for 15 sec, and extension at 72°C for 20 sec for 35 cycles with maximum ramp times. Cyclic amplification was initiated with a hot start at 94°C for 1 min and was followed by a 2-min terminal extension at 72°C. *Alu 1*, located in the α-globin 2 gene, was amplified by using the PCR primer pair 5′-GCATTGTTATTTCAACAGAAACAC-3′ and 5′-ACAGGAAAGAGAGACACTCTCC-3′. An *Alu* located in the angiotensin I-converting enzyme (ACE) gene (36) was amplified by using the flanking PCR primer pair 5′-GATGTGGCCATCACATTCGTCAGAT-3′ and 5′-CTGGAGACCACTCCCATCCTTTCT-3′. Another *Alu*, located in the apolipoprotein A1 (APO) gene (37), was amplified by using the flanking PCR primer pair 5′-AAGTGCTGTAGGCCATTTAGATTAG-3′ and 5′-AGTCTTCGATGACAGCGTATACAGA-3′ (38).

Sequencing. Direct linear amplification sequencing (39, 40) of double-stranded PCR products was accomplished by using exo$^-$ *Pfu* DNA polymerase and reagents supplied by Stratagene. Primers were end-labeled with [γ-^{32}P]ATP by using T4 polynucleotide kinase at 37°C for 11 min followed by heat inactivation at 55°C for 5 min. Sequencing reactions were carried out in a Hybaid (Middlesex, U.K.) OmniGene thermal cycler. Thermal cycle parameters were denaturation at 95°C for 30 sec, annealing at 60°C for 30 sec, and extension at 72°C for 1 min. Cyclic sequencing was initiated with a hot start at 95°C for 1 min. Both strands were sequenced. The α-globin 2 *Alu 1* locus was sequenced by using the internal primers 5′-TTGAGACGGAGTTTTGCTCT-3′ and 5′-CACGGTGGCTCACACCTGTAATC-3′. The ACE and APO loci were sequenced by using the internal primers 5′-TCACGCCTGTAATCCCAGCAC-3′ and 5′-TGAGACGGAGTCTCGCTCTG-3′. Direct sequencing of PCR products avoids visualization of PCR-induced sequence errors. We found that this method produced sequence with extremely low nonspecific background, and therefore heterozygous positions were unambiguously detected and confirmed in sequence reactions of both complementary strands. Sequencing reaction products were separated in 6% polyacrylamide/8 M urea gels and visualized by autoradiography using overnight exposure.

RESULTS

Three loci were studied. In a published comparison of divergence of seven *Alu* repeats between human and chimpanzee (31), one locus had about three times as many differences as the other six. We amplified and obtained the nucleotide sequence of this more rapidly evolving α-globin 2 *Alu 1* locus for 60 individuals (120 genomic sequences; 36,000 bp) representing diverse human populations. We also sequenced this α-globin 2 *Alu 1* locus for representatives of other hominoid species, including bonobo (*Pan paniscus*), common chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*), and orangutan (*Pongo pygmaeus*). This *Alu* repeat was absent from the single siamang (*Symphalangus syndactylus*) sampled, on the basis of PCR amplification and sequencing of the flanking regions. We obtained sequence for 36 individuals at the ACE locus (10,800 bp) and for 29 individuals at the APO locus (8700 bp). Each individual sampled was heterozygous for the presence of the *Alu* repeat at these loci. The ACE and APO *Alu* repeats are not present in other hominoids (38), but they occur in all diverse human populations sampled and therefore may be assumed to have been present prior to the radiation of modern human races.

Only three human alleles were found at the α-globin 2 *Alu 1* locus (Fig. 1). Allele *1* was present at a frequency of 0.96 and occurred in all populations (Table 1). Not a single base difference distinguished any individual among Australian Aborigines, Caucasians (Greek and Turkish Cypriots), Indonesians, Greenland Inuit, and New Guineans at this locus. Allele *2* was found in only one heterozygous individual (an African Pygmy). Allele *3* was present only within the Chimila of Columbia. A discrete character relative rate test (42), using orangutan as the outgroup taxon, found no significant difference between rates of evolution among human, bonobo, chimpanzee, and gorilla at the α-globin 2 *Alu 1* locus. Divergence (uncorrected) between human and bonobo was 3.3% and between human and common chimpanzee it was 4.0%. Corrected genetic distances calculated according to the substitution model of Kimura (43), with a transition-to-transversion ratio of 9:1 (the ratio observed in the bonobo–human sequence comparison) were 3.44% and 4.18%, respectively. Human nucleotide diversity (41) was 0.027 $\pm$ 0.011% at this locus. Human–bonobo corrected genetic distance, assuming a divergence date of 4.7 Myr (44), provides an estimate of substitution rate of 0.73%/Myr. At this rate, 0.027% nucleotide diversity yields an average age of allele divergence of about 37,000 yr for the α-globin 2 *Alu 1* locus. If we assume an older human–bonobo divergence date, 7.0 Myr (45), then the rate would be 0.49%/Myr, which suggests an average age of sequence divergence of 55,000 yr. Alternatively, if we use the human–common chimpanzee corrected genetic distance of 4.18%, and divergence dates of 4.7 and 7.0 Myr, average ages of sequence divergence are 30,000 and 45,000 yr, respectively.

Considering that allele divergence predates population divergence (46), and in light of the compelling new evidence for a wide geographic distribution of *H. erectus* in Asia at close to 2 Myr ago (33, 34), we consider 1.5 Myr as a conservative date for a model of archaic allele divergence. Assuming a molecular clock calibrated by the divergence of humans and bonobos, one can generate via computer simulation the expected distribution of observed nucleotide difference between sequences that

```
ORANG CGGACAGACACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGTGGGCGG
GORIL .C.....G...A.........A.........................C.T....C...A..
HUM-1 .C.....G.............A.............T............T....C.A.A..
HUM-2 .C.....G.............A.............T............T....C.A.A..
HUM-3 .C.....G.............A.............T............T....C.A.A..
CHIMP .C.....................A.....C.....................T....C...A..
BONOB .C.....................A.....C.....................T....C...A..

ORANG ATCACCTGAGGTTGGGAGTTCGAAACCAGCCTGACCAATATGGAGAAACCCCAGTTCTAC
GORIL ............C..........G.........................T.......A...
HUM-1 ............C.......T..G.................................A...
HUM-2 ............C.......T..G.................................C.A...
HUM-3 ............C.......T..G.T...............................A...
CHIMP ............C..........G.................................A...
BONOB ............C..........G.................................A...

ORANG TAAAAATACAAAATTAGCTGGGCGTGGTGGCGCATGCCTGTAATTCCAGCTACTAGGGAG
GORIL .......................T.G...................C..............
HUM-1 .......................T.....................C.T..........A..
HUM-2 .......................T.....................C.T..........A..
HUM-3 .......................T.....................C.T..........A..
CHIMP .......................T...C.................C..............
BONOB .......................T.....................C..............

ORANG GCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGTGGAGGTTGTGGTGAGCTGAGATCGCG
GORIL ..................A.............A.........TA..............A.A
HUM-1 .....................................................A..............A..
HUM-2 .....................................................A..............A..
HUM-3 .....................................................A..............A..
CHIMP ............A.....................................A..............A..
BONOB ............A.....................................A..............A..

ORANG CCATTGCACTCCAGCCTGGGCAACAAGAGCAAAACTCCATCTTAAAAAATAAATAAATAA
GORIL ..........................................C..................
HUM-1 ..........................................G...C..................
HUM-2 ..........................................G...C..................
HUM-3 ..........................................G...C..................
CHIMP ...C.....................................C............C....
BONOB ...C.....................................C..................
```

FIG. 1. DNA sequences of α-globin 2 *Alu 1* for orangutan, gorilla, human (alleles *1, 2*, and *3*), common chimpanzee, and bonobo. Percent divergences are as follows: orang/gorilla, 7.3%; orang/human, 6.7%; orang/chimp, 6.0%; orang/bonobo, 5.3%; gorilla/human, 4.7%; gorilla/chimp, 4.7%; gorilla/bonobo, 4.0%; human/chimp, 4.0%; human/bonobo, 3.3%; chimp/bonobo, 0.7%. By comparison, human nucleotide diversity, or average sequence divergence (41), is 0.027% ± 0.011%, reflecting the recent origin and expansion of modern humans and replacement of multiregional archaic populations.

diverged 1.5 Myr ago. Fig. 2 illustrates the number of observed differences among 10,000 independent pairs of 300-bp sequences that diverged 1.5 Myr ago. This distribution models differences expected for samples of independent pairs of sequences from different populations that diverged 1.5 Myr ago. Computer simulations were performed as follows: sequences of 300 bp were generated according to the substitution model of Kimura (43), assuming a substitution rate determined from comparison of fixed differences between bonobo and human α-globin 2 *Alu 1* sequences. The rate of transition to transversion substitutions was set at 9:1 in simulations. The distribution is useful as a model for providing a measure for recognition of alleles from regional archaic populations, although the assumption of no gene flow may not be met in reality (47). In other words, the distribution may not represent a realistic expectation of allele frequencies, but it serves as a model for expected divergence of alleles passed on to modern humans from regional archaic populations.

At the ACE locus, the sequences of all 36 individuals were identical to the published (36) sequence. At the APO locus two alleles were found. One was identical to the published (37) sequence with a C at position 164 and the other had a T at that position. The C allele was in high frequency in African Pygmies (0.91) and New Guineans (1.00), and the T allele was in high frequency in Australian Aborigines (1.00) and Indonesians (0.67).

DISCUSSION

Are these data consistent with continuity of regional alleles that diverged after the founding of regional archaic populations? Alternatively, do these data reveal the signature of a relatively recent, and largely complete, replacement of multiregional archaic populations by the descendants of an original early modern population, consistent with the Noah's Ark or recent African origin hypothesis?

For the entire data set of over 55 kb we see no evidence for an allele of ancient divergence. The expected distribution for the α-globin 2 *Alu 1* locus shown in Fig. 2 predicts an average of 3.7 nucleotide differences between alleles that diverged 1.5 Myr ago. In these simulations, identical alleles were found less than 4% of the time, and alleles with 7 differences were found at about the same frequency as alleles with no differences. According to the distribution, one would expect a pair of α-globin 2 *Alu 1* sequences sampled from different populations to have more than 2 differences between them >63% of the time. In contrast to these predictions, our data revealed that most individuals and populations were identical, and the maximum difference between any pair of sequences was 2.

Furthermore, the small amount of sequence polymorphism found in α-globin 2 *Alu 1* is best explained as of recent origin when viewed in a phylogenetic context. If we presume that the abundant allele is ancestral, then no descendant allele differs

Table 1. Number of genomes sampled and allele frequencies (f) for the ACE, APO, and α-globin 2 *Alu 1* loci

Population	ACE		APO		α-Globin 2 *Alu 1*	
	No.	Allele/f	No.	Allele/f	No.	Allele/f
African–American	2	*1*/1.0	0		0	
African Pygmy	8	*1*/1.0	11	*1*/0.91, *2*/0.09	18	*1*/0.94, *2*/0.06
Australian Aborigine	5	*1*/1.0	5	*2*/1.0	12	*1*/1.0
Caucasian–American	1	*1*/1.0	0		0	
Chimila (Amerind)	0		0		12	*1*/0.67, *3*/0.33
Greek Cypriot	4	*1*/1.0	0		12	*1*/1.0
Indonesian	6	*1*/1.0	6	*1*/0.33, *2*/0.67	12	*1*/1.0
Inuit	4	*1*/1.0	0		12	*1*/1.0
New Guinean	3	*1*/1.0	6	*1*/0.83, *2*/0.17	24	*1*/1.0
Turkish Cypriot	3	*1*/1.0	1	*1*/1.0	18	*1*/1.0

Allele *1* of ACE was identical to the published (36) sequence. Allele *1* of APO was identical to the published (37) sequence and allele *2* of APO differed by a T at position 164. α-Globin 2 alleles *1, 2*, and *3* differed from the published (31) sequence by having a T at position 231. In addition, allele *2* had a C at position 115 and allele *3* had a T at position 86.

from the ancestor by more than a single nucleotide difference. The alternative allele found in the Chimila is probably no greater than 20,000 years old. This Amerindian group represents a relatively recent branch of human population phylogeny, as Amerindians were derived from Asians after the establishment of most other major human regional populations. Parsimony then best explains this allele as having recently arisen in the New World rather than being lost in Asian, African, and Caucasian populations. The only other alternative allele, a single nucleotide difference in 1 of 18 African Pygmy genomes sampled, provides no information regarding its age other than its low divergence from the abundant allele, which indicates a probable young age on the basis of the expected distribution.

The estimated average age of sequence divergence of between 30,000 and 55,000 years provides further support for a recent worldwide replacement of the human nuclear genetic complement. Increases in divergence and differences among population allele frequencies bring about dramatic increases in nucleotide diversity and estimates of average age of divergence. We should expect substantial polymorphism if archaic alleles were maintained in diverse populations, whereas, given a replacement scenario, nucleotide diversity would be low, consistent with our data. Complete monomorphism at the ACE locus the presence of only two alternative alleles differing by a single nucleotide at the APO locus provide additional support for the hypothesis of a recent single origin for the modern human autosomal genetic complement with general replacement, rather than continuity, of archaic lineages.

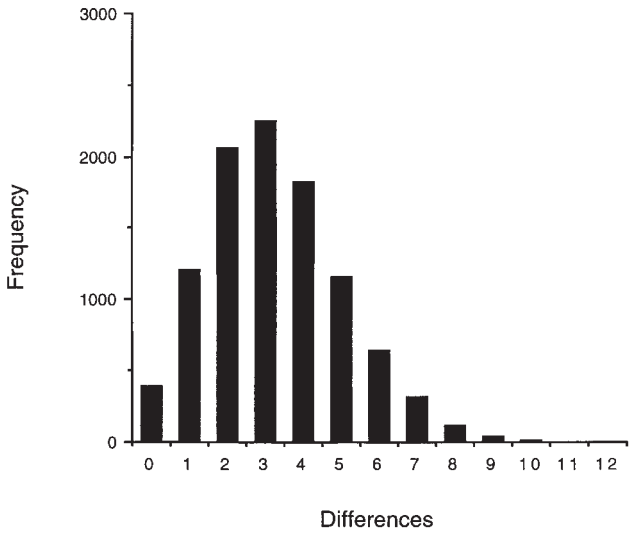

FIG. 2. Computer-generated distribution of the number of observable differences among 10,000 independent pairs of sequences 300 bp in length that diverged 1.5 Myr ago. The distribution approximates the observable nucleotide differences at the α-globin 2 *Alu 1* locus expected between two populations that were geographically isolated 1.5 Myr ago. It provides a means for recognition of alleles that diverged in different geographic regions following the wide dispersal of *Homo erectus* close to 2 Myr ago. Such alleles are expected to have more than one nucleotide difference more than 95% of the time.

We thank B. S. Gaut for providing the distribution in Fig. 2 and comments, B. Keats for helpful discussion, D. H. Kass and T. H. Shaikh for technical assistance, S. Sherry for comments, P. A. Ioannou for DNA samples, and E. Yunis and J. Yunis for collection of Chimila samples. This work was supported by National Institutes of Health Grant HG000770 to P.L.D., Department of Energy Laboratory Directed Research and Development Grant 94-LW-103 to M.A.B., and Public Health Service Grant RR08205 to R.J.H. Work by M.A.B. at Lawrence Livermore National Laboratory was conducted under the auspices of Department of Energy Contract W-7405-ENG-48.

1. Valladas, H., Reyss, J. L., Joron, J. L., Valladas, G., Bar-Yosef, O. & Vandermeersch, B. (1988) *Nature (London)* **331,** 614–616.
2. Mellars, P. & Stringer, C., eds. (1989) *The Human Revolution* (Princeton Univ. Press, Princeton, NJ).
3. Brown, W. M. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 3605–3609.
4. Cann, R. L., Stoneking, M. & Wilson, A. C. (1987) *Nature (London)* **325,** 31–36.
5. Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A. C. (1991) *Science* **253,** 1503–1507.
6. Howells, W. W. (1976) *J. Hum. Evol.* **5,** 477–495.
7. Wolpoff, M. H. (1989) in *The Human Revolution*, eds. Mellars, P. & Stringer, C. (Princeton Univ. Press, Princeton, NJ), pp. 62–108.
8. Powell, J. R. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 492–495.
9. Ferris, S. D., Sage, R. D., Huang, C.-M., Nielsen, J. T., Ritte, U. & Wilson, A. C. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 2290–2294.
10. Spolsky, C. & Uzell, T. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 5802–5805.
11. Carr, S. M., Ballinger, S. W., Derr, J. N., Blankenship, L. H. & Bickham, J. W. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 9576–9580.
12. Gyllensten, U. & Wilson, A. C. (1987) *Genet. Res.* **49,** 25–29.
13. Tegelström, H. (1987) *J. Mol. Evol.* **24,** 218–227.
14. Aubert, J. & Solinac, M. (1990) *Evolution* **44,** 1272–1282.
15. Kraus, F. & Miyamoto, M. M. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 2235–2238.
16. Knight, A., Densmore, L. D. & Rael, E. D. (1992) in *Biology of the Pitvipers*, eds. Campbell, J. A. & Brodie, E. D., Jr. (Selva, Tyler, TX), pp. 49–70.
17. Nei, M. & Livshits, G. (1989) *Hum. Hered.* **39,** 276–281.
18. Livshits, G. & Nei, M. (1990) *Ann. Hum. Biol.* **17,** 501–513.
19. Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R. & Young, I. G. (1981) *Nature (London)* **290,** 457–465.
20. Brown, W. M. (1985) in *Molecular Evolutionary Genetics*, ed. MacIntyre, R. J. (Plenum, New York), pp. 95–130.
21. Lewin, B. M. (1980) *Gene Expression* (Wiley, New York).

22. Dorit, R. L., Akashi, H. & Gilbert, W. (1995) *Science* **268,** 1183–1185.
23. Bowcock, A. M., Kidd, J. R., Mountain, J. L., Hebert, J. M., Carotenuto, L., Kidd, K. & Cavalli-Sforza, L. L. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 839–843.
24. Li, W.-H. & Sadler, L. A. (1991) *Genetics* **129,** 513–523.
25. Fullerton, S. M., Harding, R. M., Boyce, A. J. & Clegg, J. B. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 1805–1809.
26. Kurth, J. H., Mountain, J. L. & Cavalli-Sforza, L. L. (1993) *Genomics* **16,** 69–77.
27. Kurth, J. H. & Cavalli-Sforza, L. L. (1994) *Am. J. Hum. Genet.* **54,** 1037–1041.
28. Schmid, C. W. & Shen, C.-K. (1985) in *Molecular Evolutionary Genetics*, ed. MacIntyre, R. J. (Plenum, New York), pp. 323–358.
29. Deininger, P. L. (1989) in *Mobile DNA*, eds. Howe, M. M. & Berg, D. E. (Am. Soc. Microbiol., Washington, DC), pp. 619–636.
30. Deininger, P. L. & Batzer, M. A. (1993) *Evol. Biol.* **27,** 157–196.
31. Sawada, I., Willard, C., Shen, C.-K. J., Chapman, B., Wilson, A. C. & Schmid, C. W. (1985) *J. Mol. Evol.* **22,** 316–322.
32. Gonzalez, I. L., Peterson, R. & Sylvester, J. E. (1989) *Mol. Biol. Evol.* **6,** 413–423.
33. Swisher, C. C., III, Curtis, G. H., Jacob, T., Getty, A. G., Suprijo, A. & Widiasmoro (1994) *Science* **263,** 1118–1121.
34. Gabunia, L. & Vekua, A. (1995) *Nature (London)* **373,** 509–512.
35. Li, W.-H. & Graur, D. (1991) *Fundamentals of Molecular Evolution* (Sinauer, Sunderland, MA).
36. Tiret, L., Rigat, B., Visvikis, S., Breda, C., Corvol, P., Cambien, F. & Soubrier, F. (1992) *Am. J. Hum. Genet.* **51,** 197–205.
37. Antanorakis, S. E., Oettgen, P., Chakravarti, S. L., Halloran, S. L., Hudson, R. R., Feisee, L. & Karathanasis, S. K. (1988) *Hum. Genet.* **80,** 265–273.
38. Batzer, M. A., Stoneking, M., Alegria-Hartman, M., Bazan, H., Kass, D. H., Shaikh, T. H., Novick, G. E., Ioannou, P. A., Scheer, W. D., Herrera, R. J. & Deininger, P. L. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 12288–12292.
39. Carothers, A. M., Urlaub, G., Mucha, J., Grunberger, D. & Chasin, L. A. (1989) *BioTechniques* **7,** 494–499.
40. Saluz, H. & Jost, J.-P. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 2602–2606.
41. Nei, M. & Li, W.-H. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 5269–5273.
42. Mindell, D. P. & Honeycutt, R. L. (1990) *Annu. Rev. Ecol. Syst.* **21,** 541–566.
43. Kimura, M. (1980) *J. Mol. Evol.* **16,** 111–120.
44. Horai, S. (1993) *J. Mol. Evol.* **35,** 32–43.
45. Bailey, W. J., Hayasaka, K., Skinner, C. G., Kehoe, S., Sieu, L. C., Slightom, J. L. & Goodman, M. (1992) *Mol. Phylogenet. Evol.* **1,** 97–135.
46. Stoneking, M. (1993) *Evol. Anthropol.* **2,** 60–73.
47. Takahata, N. (1993) *Mol. Biol. Evol.* **10,** 2–22.