

Alu elements and hominid phylogenetics

Abdel-Halim Salem^{*†‡}, David A. Ray^{*‡}, Jinchuan Xing^{*}, Pauline A. Callinan^{*}, Jeremy S. Myers^{*}, Dale J. Hedges^{*}, Randall K. Garber^{*}, David J. Witherspoon[§], Lynn B. Jorde[§], and Mark A. Batzer^{*†¶}

^{*}Department of Biological Sciences, Biological Computation and Visualization Center, Louisiana State University, 202 Life Sciences Building, Baton Rouge, LA 70803; [†]Department of Anatomy, Faculty of Medicine, Suez Canal University, Ismailia 4111, Egypt; and [§]Department of Human Genetics, University of Utah Health Sciences Center, Salt Lake City, UT 84112

Edited by Morris Goodman, Wayne State University School of Medicine, Detroit, MI, and approved August 26, 2003 (received for review June 18, 2003)

Alu elements have inserted in primate genomes throughout the evolution of the order. One particular Alu lineage (Ye) began amplifying relatively early in hominid evolution and continued propagating at a low level as many of its members are found in a variety of hominid genomes. This study represents the first conclusive application of short interspersed elements, which are considered nearly homoplasy-free, to elucidate the phylogeny of hominids. Phylogenetic analysis of Alu Ye5 elements and elements from several other subfamilies reveals high levels of support for monophyly of Hominidae, tribe Hominini and subtribe Hominina. Here we present the strongest evidence reported to date for a sister relationship between humans and chimpanzees while clearly distinguishing the chimpanzee and human lineages.

mobile elements | short interspersed elements | trichotomy | primates

Alu elements are short (≈ 300 bp) interspersed elements that amplify in primate genomes through a process termed retroposition (1–3). The proliferation of these elements has had a significant impact on the architecture of primate genomes (1). They comprise $>10\%$ of the human genome by mass and are the most abundant short interspersed element (SINE) in primate genomes (2). The majority of Alu amplification occurred early in primate evolution, and the current rate of Alu retroposition is at least 100-fold slower than the peak of amplification that occurred 30–50 million years ago (mya) (2–5). Alu elements are therefore a rich source of inter- and intraspecies primate genomic variation.

We have previously characterized $>2,500$ recently integrated Alu elements from the human genome that fall into six distinct subfamilies based on their diagnostic mutations (6–18). In the course of these investigations, we and others have found that individual loci from these subfamilies are informative for the study of primate systematics (19) as well as human population genetics and can be useful tools for resolving the remaining questions of hominid (siamang, orangutan, gorilla, chimpanzee, and human) phylogeny (20, 21).

Resolving the relationships among human (H), chimpanzee (C), and gorilla (G) (i.e., the trichotomy problem) has been particularly difficult. Which of the four possible relationships, ((H,C)G), ((H,G)C), ((C,G)H), and (H,C,G), reflects the true phylogeny of the three species? The consensus approach identifies the chimpanzee as the nearest living relative of humans, but the evidence supporting this conclusion is neither universal nor overwhelming (22–26). In a recent interpretation of primate taxonomy based on the application of modern cladistic methods to comparative biology data, Shoshani *et al.* (27) conclude: “we support with weak morphological data, the *Homo–Pan* clade, although some studies prefer the trichotomy hypothesis.”

DNA hybridization data support sister species relationships between humans and chimpanzees (28). Mitochondrial (mt) DNA studies based on restriction endonuclease digestion of 12S rRNA-encoding gene seemed to support either an equidistant relationship among the three lineages or a sister relationship between the chimpanzees and the gorillas (29). Additional mtDNA sequencing supported either the ((H,C)G) or the ((H,G)C) relationship depending on the genes analyzed (30, 31),

whereas complete mtDNA sequences supported the former relationship (32).

Some analyses of nuclear loci (e.g., the β -globin cluster) (21, 33) supported the ((H,C)G) relationship, and others (e.g., the involucrin gene) (34) supported the ((C,G)H) relationship. Satta *et al.* (24) studied sequences from 45 nuclear loci and found that 60% of the loci support the human–chimpanzee clade and the remaining 40% of loci support the two alternatives equally. Eleven of 14 DNA sequence data sets analyzed by Ruvolo (35) supported a human–chimpanzee clade, two supported a chimpanzee–gorilla clade and one supported a human–gorilla clade. The chimpanzees and their sibling species, the bonobos, differ from humans by an average of 0.6% at nonsynonymous sites of 97 studied human genes (25) and 1.2% in overall genomic DNA sequences (26), and they are estimated to have shared a common ancestor with humans 4.0–6.0 mya (22). Gorillas differ from humans by an average of 1.6% in genomic DNA sequences and are estimated to have shared a common ancestor with humans, chimpanzees, and bonobos 6.2–8.4 mya (26). Thus based on nucleotide divergence, chimpanzees and bonobos are the species most closely related to humans.

SINE elements represent a powerful new tool for systematic biology that can be strategically integrated with other conventional phylogenetic characters, most notably morphology and DNA sequences (36–38). There is no known mechanism for the specific removal of SINE elements from the human genome (2), and only one partial deletion of an Alu element has ever been identified (39). Because their mode of evolution is unidirectional (i.e., they do not revert to their ancestral state), individual SINE elements are generally thought to be nearly homoplasy-free characters and are thus useful for resolving phylogenetic and population genetic questions (2, 36–38, 40–43). For example, Shimamura *et al.* (42) successfully used SINE elements to support the hypothesis that cetaceans (whales, dolphins, and porpoises) form a clade within Artiodactyla (even-toed ungulates, including cows, camels, and pigs). Takahashi *et al.* (43) also used SINE elements to elucidate the relationships among the cichlid fishes in Lake Malawi. In each of these studies, the presence of a SINE in any particular lineage unambiguously joined the members of that node with only one case of potential homoplasy being introduced either by lineage sorting or via interspecies hybridization.

This does not mean that Alu elements and other SINEs are without problems with regard to phylogenetic analysis. It is known that insertion homoplasy can occur across distantly related taxa as a function of evolutionary time and that variable retroposition rates among species can limit the application of

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: SINE, short interspersed elements; mya, million years ago.

Data deposition: The sequences of the orthologous Alu insertion loci (bonobo, common chimpanzee, and gorilla) have been deposited in the GenBank database (accession nos. AY318829–AY318854).

[†]A.-H.S. and D.A.R. contributed equally to this work.

[¶]To whom correspondence should be addressed. E-mail: mbatzer@lsu.edu.

© 2003 by The National Academy of Sciences of the USA

```

Alu Y consensus  GGCCGGGCGCGGTGGCTCACGCCGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCGGATCACGAGGTCAGGAGATCGAGACCATCTCTG  90
Yc1 consensus    .....
Yd3 consensus    ..... T .....
Ye5 consensus    .....
Yi6 consensus    ..... T .....

Alu Y consensus  GCTAACACGGTGAACCCCGTCTCTACTAAAAATACAAAAA - TTAGCCGGGCGTGGTGGCGGGCCCTGTAGTCCCAGCTACTCGG - GA  178
Yc1 consensus    ..... A .....
Yd3 consensus    ..... A .....
Ye5 consensus    ..... A .....
Yi6 consensus    ..... A ..... T .....

Alu Y consensus  GGC TGAGGCAGGAGAATGGCGTGAACCC - GGGAGGCGGAGCTTGCACTGAGCCGAGATCGCGCCACTGCACCTCCAGCCTGGGCGCAGAG  267
Yc1 consensus    .....
Yd3 consensus    .....
Ye5 consensus    ..... C . G . C .....
Yi6 consensus    ..... T ..... C ..... C .....

Alu Y consensus  CGAGACTCCGTCTCA  282
Yc1 consensus    ..... 282
Yd3 consensus    ..... 271
Ye5 consensus    ..... 281
Yi6 consensus    ..... 273

```

Fig. 1. Alu consensus sequence alignment. Shown is the alignment of the consensus sequences of each Alu subfamily used in the study. The dots represent the same base present in the Y subfamily consensus sequence. Mutations are denoted by the appropriate base, and deletions are denoted by dashes.

SINEs to recently diverged phylogenetic relationships (37, 38, 44). However, it is important to note that no instances of Alu insertion homoplasy in hominids have been recovered from the analysis of >2,500 recently integrated human Alu insertions (13, 17, 18, 40, 45–47). Random sorting of the ancestral allelic lineages, sequence convergence, and sequence exchanges between alleles or duplicated loci have also been identified as likely factors confounding the interpretation of the interrelationships among species.

Despite these potential problems, Alu elements are novel genetic characters that lack most of the shortcomings encountered when employing sequence or restriction site data. The relatively low rate of retroposition of the Alu Y derived lineages in hominid genomes within the primate order and the relatively late radiation of the hominids (22) make these Alu subfamilies nearly ideal phylogenetic markers to resolve branching order in hominoids. Until now, only a few Alu loci that were informative with regard to the trichotomy issue had been identified (18, 45). A previous study (41) was successful in identifying a limited number of lineage-specific Alu elements; however, because of the limited number of Alu elements examined, questions about hominid phylogeny remained. The additional loci presented in this study facilitate a comprehensive analysis of the phylogeny of the entire group.

Here, we have identified and characterized a total of 153 Alu Ye subfamily members from the draft sequence of the human genome. One hundred seventeen of these loci, as well as 16 loci from other Alu Y lineage subfamilies (17, 18, 45) shown in Fig. 1, were screened using PCR assays to ascertain their relative insertion points in hominid evolution (humans and greater and lesser apes).

Materials and Methods

Alu Ye5 subfamily members were identified in the draft sequence of the human genome (August 6, 2001, University of California, Santa Cruz, GoldenPath assembly) by using BLAST (48) queries of the draft sequence to identify exact complements to the oligonucleotide 5'-GAACCCCGGGGGCGGAGCCT-GCAG-3', which is diagnostic for the Ye5 subfamily. We used this approach to identify 153 elements that comprise the Alu Ye5 subfamily. A complete list of all of the Alu elements identified in the searches can be found in Table 1, which is published as supporting information on the PNAS web site, www.pnas.org and at <http://batzerlab.lsu.edu>.

To aid in elucidating the relationship between the two chimpanzee species (*Pan paniscus* and *Pan troglodytes*) and to further clarify relationships among all of the great apes, we searched the available sequences from each of these species in GenBank by using the Alu Y subfamily consensus sequence as a BLAST query.

Sequences containing matches of 90% or greater were submitted to BLAT (<http://genome.ucsc.edu>; April 2003 assembly) in 20,000-bp batches to identify the homologous region of the human genome and examined computationally to determine whether the Alu insertion in *Pan* was also present in *Homo*. If the Alu appeared to be specific to chimpanzee, that locus was passed to the next step: primer design.

Oligonucleotide primers for PCR amplification of each of the Alu loci were designed as described (13, 15, 17). Using the primers and annealing temperatures for Alu Y lineage subfamily members shown in Table 2, which is published as supporting information on the PNAS web site and at <http://batzerlab.lsu.edu>, we amplified the loci from HeLa genomic DNA and from up to 12 individuals of each nonhuman primate species including *Pan troglodytes* (common chimpanzee), *Pan paniscus* (bonobo), *Gorilla gorilla* (lowland gorilla), *Pongo pygmaeus* (orangutan), *Chlorocebus aethiops sabaues* (green monkey), and *Aotus trivirgatus* (three-striped owl monkey), which were available as a primate phylogenetic panel (PRP00001) from the Coriell Institute for Medical Research, and from *Hylobates syndactylus* (siamang). For the Ye5 subfamily, 117 elements identified in the draft human genomic sequence were amenable to PCR analysis. Fourteen loci obtained from the BLAST search of *Pan* spp. sequences were amenable to PCR analysis. Examination of the orthologous regions of the various nonhuman primate genomes displayed a series of different PCR bands pattern indicative of the time of insertion of each element into the primate genomes (Fig. 2). The distribution of all elements in the various primate genomes tested can be found in Table 3, which is published as supporting information on the PNAS web site and at <http://batzerlab.lsu.edu>.

PAUP 4.0b10 (49) was used to perform a parsimony analysis on the 131 loci described above and two previously identified loci (Yi6AH36 and Yd3JX167) (18, 45) that were informative with regard to the human–chimpanzee–gorilla trichotomy issue. Dollo parsimony is most appropriately applied when handling presence/absence characters for which the ancestral state can safely be assumed (50). Thus, we implemented an exhaustive search in PAUP using Dollo parsimony and designating a New World monkey (*Aotus trivirgatus*) as the outgroup taxon. If no amplification was observed for a given locus in any taxon, the character state was coded as unknown. To determine branch support for each node, 10,000 bootstrap replicates were performed. The complete data set is available in Table 3.

Gel-purified PCR products amplified from genomic DNA templates (human, bonobo, common chimpanzee, and gorilla) were cloned by using the TOPO TA cloning vector (Invitrogen). Isolates from these cloning reactions were sequenced by using chain termination sequencing (51) on an Applied Biosystems

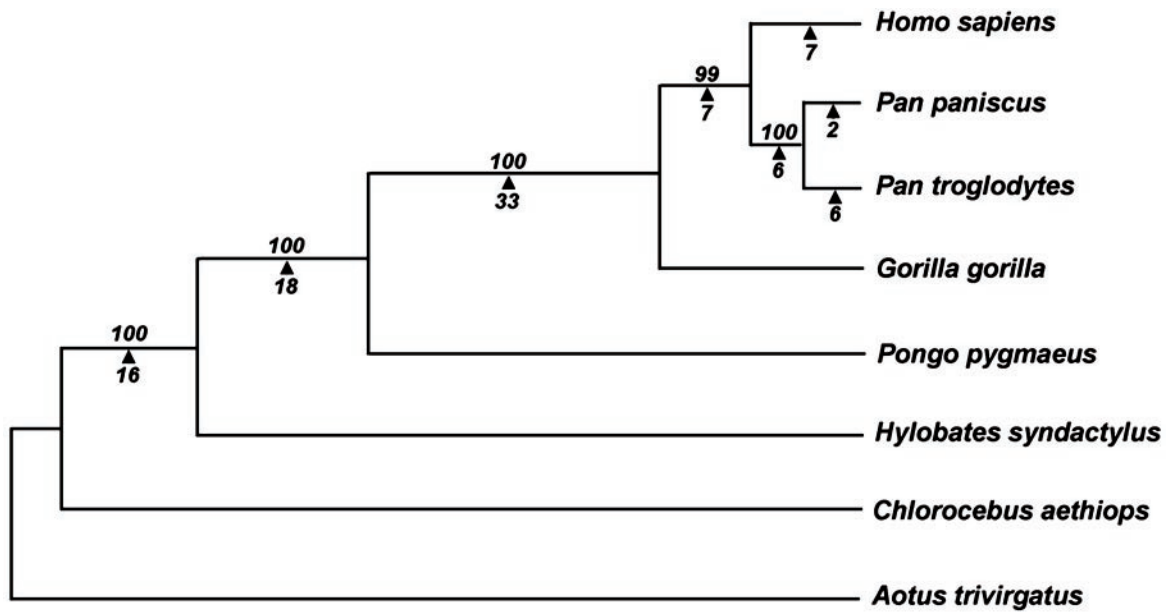


Fig. 3. Primate relationships reconstructed by using Dollo parsimony analysis of Alu elements. Primate relationships were derived from analysis of 133 Alu loci by using maximum parsimony criteria. The number of insertions observed along each branch of the tree is indicated, and bootstrap support values are placed above each node.

Homo–Gorilla clade, is interesting and may be explained by insertion homoplasy at first glance. However, sequence analysis of the locus indicated that it was the result of incomplete lineage sorting of an ancestral polymorphic locus (Fig. 4). Age estimates of this and the other seven Alu insertions at the *Homo–Pan* node also support this explanation. We calculated the relative ages of each Alu insertion by using CpG mutation densities. Estimates of insertion times for each of the seven loci supporting the *Homo–Pan* clade range from 2.74 to 5.48 mya (mean = 4.11 ± 1.12 mya). The Ye5AH137 insertion appears to be older than the others with an estimated age of ≈ 6.85 million years. Previously

reported sequence data suggests that gorillas shared a common ancestor with humans, chimpanzees, and bonobos 6.2–8.4 mya (26), around the time of the Ye5AH137 insertion. Thus, if the ancestral population were polymorphic for the presence of an Alu insertion at this locus during this period, then fixation of one or the other character state would have occurred randomly in the daughter lineages. In this case, it appears that the insertion was fixed in both human and gorilla lineages but not in chimpanzee. Furthermore, these data lend support to the hypothesis that humans, chimpanzees, and gorillas diverged during a relatively narrow evolutionary time frame (25). It should be noted that

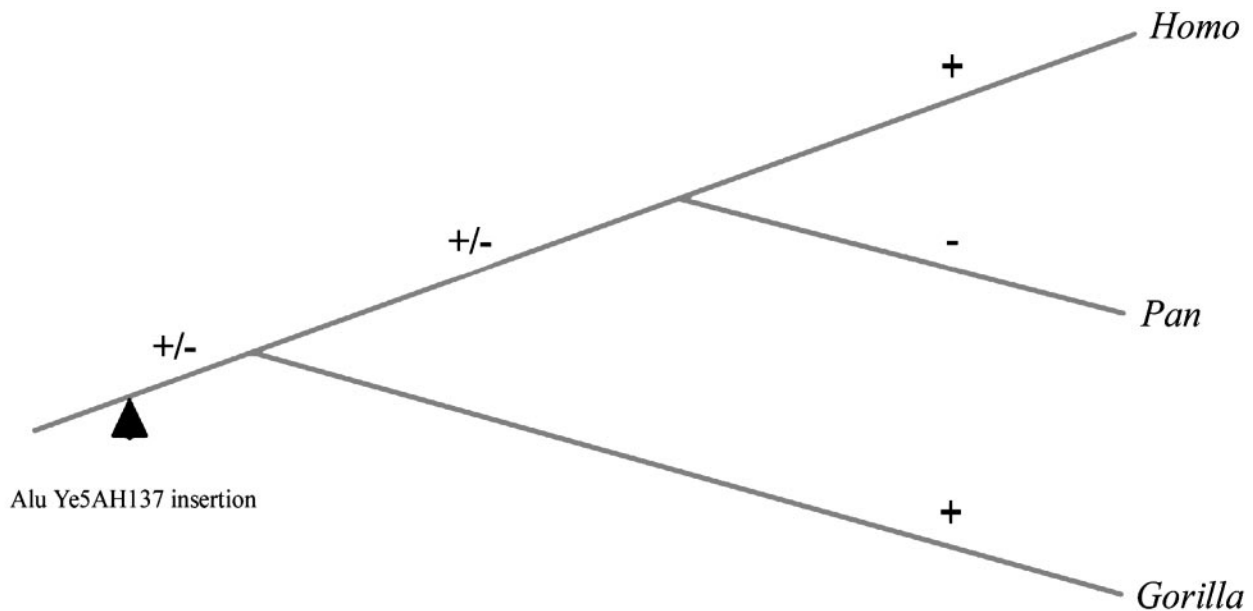


Fig. 4. Lineage sorting model proposed for the Alu Ya5AH137 locus. In this model, Alu Ye5AH137 is polymorphic (indicated as +/-) in the ancestral population that gave rise to humans, chimpanzees, and gorillas. Random fixation of the alternative alleles in each of the three subsequent lineages results in a pattern supporting the *Homo–Gorilla* association instead of *Homo–Pan*.

Takahashi *et al.* (43), in their study of cichlid phylogeny, also found a single SINE locus that may have been influenced by lineage sorting, although interspecies hybridization could also have been invoked in that case.

In contrast, two loci that join *Homo* and *Pan* (Ye5AH31 and Ye5AH93) had an estimated age of 2.74 million years, substantially younger than the estimated time of the human chimpanzee divergence (≈ 4 –6 mya). Although this is interesting in that these loci suggest very recent coancestry between *Homo* and *Pan*, it should be noted that age estimates were based on numbers of CpG mutations from the consensus sequence for the subfamily under examination. The stochastic nature of single nucleotide mutations and the short evolutionary times involved here produce large variances in age estimates for individual Alu elements (12). Therefore, age estimates based on the mean age of several Alu insertions are more accurate than estimates based on individual Alu repeats. Thus, the apparently young age estimates for these two insertions should not be taken as evidence for a more recent divergence between chimpanzees and humans when the mean age of all of the relevant insertions is ≈ 4.11 million years.

The utility of SINE insertions, and mobile elements in general, for phylogenetic analysis continues to be bolstered by studies

such as this one. Here, we present the first application of SINES to fully elucidate the phylogeny of the hominid lineage and present the strongest evidence to date for phylogenetic relationships among the hominid lineages. Of the 133 Alu insertion loci, 95 were unambiguously informative for determining the relative divergence of each of the major lineages. In addition, seven Alu insertions informative with regard to the *Homo*–*Pan*–*Gorilla* trichotomy unambiguously support monophyly of humans and chimpanzees. Finally, six loci discovered by using searches of the currently available chimpanzee genomic sequences unambiguously join the two *Pan* species to the exclusion of *Homo*. Only a single insertion from the entire data set showed any potential evidence of insertion homoplasy; however, the distribution of this element appears to be the result of lineage sorting of an ancestral polymorphism.

We thank Dr. Caro-Beth Stewart for providing Siamang DNA. This research was supported by Louisiana Board of Regents Millennium Trust Health Excellence Fund HEF (2000-05)-05, (2000-05)-01, and (2001-06)-02 (to M.A.B.), National Institutes of Health Grant RO1 GM59290 (to M.A.B. and L.B.J.), and National Science Foundation Grants BCS-0218338 (to M.A.B.) and BCS-0218370 (to L.B.J.).

- Deininger, P. L. & Batzer, M. A. (1993) *Evol. Biol.* **27**, 157–196.
- Batzer, M. A. & Deininger, P. L. (2002) *Nat. Rev. Genet.* **3**, 370–379.
- Kapitonov, V. & Jurka, J. (1996) *J. Mol. Evol.* **42**, 59–65.
- Labuda, D. & Striker, G. (1989) *Nucleic Acids Res.* **17**, 2477–2491.
- Shen, M. R., Batzer, M. A. & Deininger, P. L. (1991) *J. Mol. Evol.* **33**, 311–320.
- Arcot, S. S., Adamson, A. W., Lamerdin, J. E., Kanagy, B., Deininger, P. L., Carrano, A. V. & Batzer, M. A. (1996) *Genome Res.* **6**, 1084–1092.
- Arcot, S. S., Adamson, A. W., Risch, G. W., LaFleur, J., Robichaux, M. B., Lamerdin, J. E., Carrano, A. V. & Batzer, M. A. (1998) *J. Mol. Biol.* **281**, 843–856.
- Arcot, S. S., DeAngelis, M. M., Sherry, S. T., Adamson, A. W., Lamerdin, J. E., Deininger, P. L., Carrano, A. V. & Batzer, M. A. (1997) *Mutat. Res.* **382**, 5–11.
- Arcot, S. S., Shaikh, T. H., Kim, J., Bennett, L., Alegria-Hartman, M., Nelson, D. O., Deininger, P. L. & Batzer, M. A. (1995) *Gene* **163**, 273–278.
- Batzer, M. A. & Deininger, P. L. (1991) *Genomics* **9**, 481–487.
- Batzer, M. A., Gudi, V. A., Mena, J. C., Foltz, D. W., Herrera, R. J. & Deininger, P. L. (1991) *Nucleic Acids Res.* **19**, 3619–3623.
- Batzer, M. A., Kilroy, G. E., Richard, P. E., Shaikh, T. H., Desselle, T. D., Hoppens, C. L. & Deininger, P. L. (1990) *Nucleic Acids Res.* **18**, 6793–6798.
- Carroll, M. L., Roy-Engel, A. M., Nguyen, S. V., Salem, A. H., Vogel, E., Vincent, B., Myers, J., Ahmad, Z., Nguyen, L., Sammarco, M., *et al.* (2001) *J. Mol. Biol.* **311**, 17–40.
- Donaldson, C. J., Crapanzano, J. P., Watson, J. C., Levine, E. A. & Batzer, M. A. (2002) *Mutat. Res.* **501**, 137–141.
- Roy, A. M., Carroll, M. L., Kass, D. H., Nguyen, S. V., Salem, A. H., Batzer, M. A. & Deininger, P. L. (1999) *Genetica* **107**, 149–161.
- Roy, A. M., Carroll, M. L., Nguyen, S. V., Salem, A. H., Oldridge, M., Wilkie, A. O., Batzer, M. A. & Deininger, P. L. (2000) *Genome Res.* **10**, 1485–1495.
- Roy-Engel, A. M., Carroll, M. L., Vogel, E., Garber, R. K., Nguyen, S. V., Salem, A. H., Batzer, M. A. & Deininger, P. L. (2001) *Genetics* **159**, 279–290.
- Salem, A. H., Kilroy, G. E., Watkins, W. S., Jorde, L. B. & Batzer, M. A. (2003) *Mol. Biol. Evol.* **20**, 1349–1361.
- Schmitz, J., Ohme, M. & Zischler, H. (2001) *Genetics* **157**, 777–784.
- Stoneking, M., Fontius, J. J., Clifford, S. L., Soodyall, H., Arcot, S. S., Saha, N., Jenkins, T., Tahir, M. A., Deininger, P. L. & Batzer, M. A. (1997) *Genome Res.* **7**, 1061–1071.
- Singer, S. S., Schmitz, J., Schwiegk, C. & Zischler, H. (2003) *Mol. Phylogenet. Evol.* **26**, 490–501.
- Goodman, M., Porter, C. A., Czelusniak, J., Page, S. L., Schneider, H., Shoshani, J., Gunnell, G. & Groves, C. P. (1998) *Mol. Phylogenet. Evol.* **9**, 585–598.
- O'hUigin, C., Satta, Y., Takahata, N. & Klein, J. (2002) *Mol. Biol. Evol.* **19**, 1501–1513.
- Satta, Y., Klein, J. & Takahata, N. (2000) *Mol. Phylogenet. Evol.* **14**, 259–275.
- Wildman, D. E., Uddin, M., Liu, G., Grossman, L. I. & Goodman, M. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 7181–7188.
- Chen, F. C. & Li, W. H. (2001) *Am. J. Hum. Genet.* **68**, 444–456.
- Shoshani, J., Groves, C. P., Simons, E. L. & Gunnell, G. F. (1996) *Mol. Phylogenet. Evol.* **5**, 102–154.
- Sibley, C. G. & Ahlquist, J. E. (1984) *J. Mol. Evol.* **20**, 2–15.
- Hixon, J. E. & Brown, W. M. (1986) *Mol. Biol. Evol.* **3**, 1–18.
- Hasagawa, M. & Kishino, H. (1991) in *New Aspects of the Genetics of Molecular Evolution*, eds Kimura, M. & Takahata, N. (Japan Sci. Soc. Press, Tokyo), pp. 303–317.
- Ruvolo, M., Disotell, T. R., Allard, M. W., Brown, W. M. & Honeycutt, R. L. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1570–1574.
- Horai, S., Hayasaka, K., Kondo, R., Tsugane, K. & Takahata, N. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 532–536.
- Koop, B. F., Tagle, D. A., Goodman, M. & Slightom, J. L. (1989) *Mol. Biol. Evol.* **6**, 580–612.
- Djian, P. & Green, H. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8447–8451.
- Ruvolo, M. (1997) *Mol. Biol. Evol.* **14**, 248–265.
- Shedlock, A. M. & Okada, N. (2000) *BioEssays* **22**, 148–160.
- Miyamoto, M. M. (1999) *Curr. Biol.* **9**, R816–R819.
- Hillis, D. M. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 9979–9981.
- Edwards, M. C. & Gibbs, R. A. (1992) *Genomics* **14**, 590–597.
- Roy-Engel, A. M., Carroll, M. L., El-Sawy, M., Salem, A. H., Garber, R. K., Nguyen, S. V., Deininger, P. L. & Batzer, M. A. (2002) *J. Mol. Biol.* **316**, 1033–1040.
- Hamdi, H., Nishio, H., Zielinski, R. & Dugaiczky, A. (1999) *J. Mol. Biol.* **289**, 861–871.
- Shimamura, M., Yasue, H., Ohshima, K., Abe, H., Kato, H., Kishiro, T., Goto, M., Munekata, I. & Okada, N. (1997) *Nature* **388**, 666–670.
- Takahashi, K., Nishida, M., Yuma, M. & Okada, N. (2001) *J. Mol. Evol.* **53**, 496–507.
- Cantrell, M. A., Filanoski, B. J., Ingermann, A. R., Olsson, K., DiLuglio, N., Lister, Z. & Wichman, H. A. (2001) *Genetics* **158**, 769–777.
- Xing, J., Salem, A. H., Hedges, D. J., Kilroy, G. E., Watkins, W. S., Schienman, J. E., Stewart, C. B., Jurka, J., Jorde, L. B. & Batzer, M. A. (2003) *J. Mol. Evol.*, in press.
- Arcot, S. S., Adamson, A. W., Risch, G., LaFleur, J., Lamerdin, J. E., Carrano, A. V. & Batzer, M. A. (1998) *J. Mol. Biol.* **281**, 843–855.
- Callinan, P. A., Hedges, D. J., Salem, A.-H., Xing, J., Walker, J. A., Garber, R. K., Watkins, W. S., Bamshad, M. J., Jorde, L. B. & Batzer, M. A. (2003) *Gene*, in press.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410.
- Swofford, D. L. (2002) *PAUP: Phylogenetic Analysis Using Parsimony* (Sinauer, Sunderland, MA).
- Farris, J. S. (1977) *Syst. Zool.* **26**, 77–88.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- Bird, A. P. (1980) *Nucleic Acids Res.* **8**, 1499–1504.
- Jabbari, K. & Bernardi, G. (1998) *Gene* **224**, 123–127.