A Hominoid-Specific Nuclear Insertion of the Mitochondrial D-Loop: Implications for Reconstructing Ancestral Mitochondrial Sequences

Hans Zischler,[1] *Helga Geisert, and Jose Castresana*

Zoologisches Institut der Universität München, Germany

A nuclear integration of a mitochondrial control region sequence on human chromosome 9 has been isolated. PCR analyses with primers specific for the respective insertion-flanking nuclear regions showed that the insertion took place on the lineage leading to Hominoidea (gibbon, orangutan, gorilla, chimpanzee, and human) after the Old World monkey–Hominoidea split. The sequences of the control region integrations were determined for humans, chimpanzees, gorillas, orangutans, and siamangs. These sequences were then used to construct phylogenetic trees with different methods, relating them with several hominoid, Old World monkey, and New World monkey mitochondrial control region sequences. Applying maximum-likelihood, neighbor-joining, and parsimony algorithms, the insertion clade was attached to the branch leading to the hominoid mitochondrial sequences as expected from the PCR-determined presence/absence of this integration. An unexpected long branch leading to the internal node that connects all insertion sequences was observed for the different phylogeny reconstruction procedures. This finding is not totally compatible with the lower evolutionary rate in the nucleus than in the mitochondrial compartment. We determined the unambiguous substitutions on the branch leading to the most recent common ancestor (MRCA) of the mitochondrial inserts according to the parsimony criterium. We propose that they are unlikely to have been caused by damage of the transposing nucleic acid and that they are probably due to a change in the evolutionary mode after the transposition.

Introduction

Several studies have demonstrated the ongoing interorganellar transfer and integration of mitochondrial nucleic acids into nuclear chromosomes in plant, invertebrate, and vertebrate taxa (Blanchard and Schmidt 1995; Zhang and Hewitt 1996). Assuming a lower evolutionary rate in the nuclear DNA of vertebrates than in the mitochondrial DNA (Brown et al. 1982), these mitochondrial integrations might preserve the ancestral state of the mitochondrial sequence that existed at the time of transposition and could therefore be regarded as ''molecular fossils.'' By isolating many of these insertions, the sequence history of the mitochondrial genome could be at least in part traced over long evolutionary timescales (Hu and Thilly 1994; Perna and Kocher 1996).

We herein describe the isolation and characterization of a D-loop insertion with its nuclear flanking sequences on human chromosome 9. The use of PCR primers specific for the insertion-flanking regions allowed us to check for the presence of this insertion in different species and to retrieve the homologous insertion sequences. Moreover, it was possible to amplify the respective noninsertion allele, thus identifying the target site and the sequence changes during integration. We show that this integration is present in all hominoid species but absent in Old World monkeys. Therefore, the DNA transfer from mitochondria to the nucleus can be estimated to have happened in the range of 17–30 MYA in a common ancestor of all hominoids. Based on the PCR-determined presence/absence of this integration and on the well-accepted phylogeny of Old World monkeys and different hominoid species, we related hominoid, Old World monkey, and New World monkey mitochondrial sequences with the hominoid nuclear counterparts in phylogenetic tree reconstructions and analyzed the sequence evolution of this insertion.

Materials and Methods

DNA extraction, restriction, electrophoresis, blotting, hybridization, ligation, sequencing, and transformation of bacterial cells were done according to standard protocols (Sambrook, Fritch, and Maniatis 1989). In the following, oligonucleotides are named H and L, designating the heavy- and light-strand sequences of the human mitochondrial genome, respectively. The number indicates the 3′ end of the oligonucleotide according to the human reference sequence (Anderson et al. 1981).

Isolation of Nuclear Sequences Flanking the Mitochondrial Insertion

Human nuclear DNA free of mtDNA was isolated by differential lysis of sperm cells (Zischler et al. 1995) and used for the determination of the integration-flanking nuclear sequences. On 20 ng of restricted and size-enriched nuclear DNA, the 5′ biotinylated primer L16420 (CACCATCCTCCGTGAAATCA, primer 1 in fig. 1) was extended unidirectionally in 50 cycles consisting of 1 min annealing at 60°C, 1.5 min extension at 72°C, and 40 s denaturation at 92° using *Taq* polymerase. Excess primers were removed by ultrafiltration, and extension products were captured by absorption on streptavidin-coated magnetic beads (Dynal). A poly-G tail was added by a terminal transferase catalyzed reaction following the supplier's instructions. After heat inactivation and buffer change, an unbiotinylated L16420 in combination with a C_{15}-mer was used to amplify the extension products. The amplification products

[1]Present address: Deutsches Primatenzentrum, Göttingen, Germany.

Key words: mitochondrial D-loop insertion, Hominoidea, Old World monkey, New World monkey.

Address for correspondence and reprints: Hans Zischler, Deutsches Primatenzentrum, Kellnerweg 4, D-37077 Göttingen, Germany. E-mail: hzischl@www.dpz.gwdg.de.

Mol. Biol. Evol. 15(4):463–469. 1998

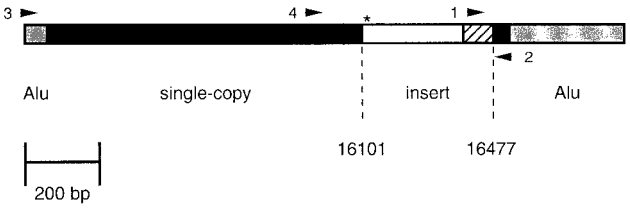

FIG. 1.—Diagram of the insertion locus in humans. Shaded boxes represent Alu elements, and black boxes represent single-copy nuclear sequences. The open box represents the inserted mtDNA sequence, with the hatched part being homologous to the central, conserved part of the mitochondrial D-loop (Saccone, Pesole, and Sbisa 1991). The boundaries coincide with the positions 16101 and 16477 of the human mitochondrial reference sequence (L-strand; Anderson et al. 1981). Arrowheads indicate the locations and directions of the PCR primers; the numbers beside them are referred to in the text. The star represents the 3′ end of the human 7S DNA.

were cloned into pCR-Script (Stratagene) as recommended by the supplier and electroporated into SURE strain bacteria (Stratagene). Colonies were screened with an internal human control-region-specific oligonucleotide (GCGGGATATTGATTTCAGGG). Positive plasmids were isolated and restricted, and the largest inserts were sequenced. In this way, the 3′ border of the insertion (based on the L-strand sequence) and 335 bp of nuclear flanking DNA were determined.

For isolation of the second flanking sequence, information of the 3′ flank was used to design primers for an interspersed repetitive sequence (IRS) PCR: an Alu-element-specific primer (GCTTGCAGTGAGCCGAGAT, primer 3 in fig. 1), modified on the basis of the TC-65 primer described by Nelson et al. (1989), was used in combination with a primer specific for the first nuclear flank (GCGAGCTGAAGGACTTTCTG, primer 2 in fig. 1) to PCR-amplify 400 ng of human nuclear DNA free of mitochondrial DNA (35 cycles with 1 min extension at 72°C, 40 s denaturation at 92°C, and 1 min primer annealing at 60°C). One thirtieth of the amplification product was subsequently subjected to 25 cycles of a second seminested PCR, carried out under the conditions described above. The same Alu-element-specific primer was used together with a flanking-region-specific primer annealing 3′ to the one used in the first round of PCR (GCTGAAGGACTTTCTGGCTG). The resulting products were again cloned as above, and the colonies were screened with H16438 (CCCGGAGCGAGGAGAGTAGC) at 58°C. Positive clones were sequenced.

Amplification of the Mitochondrial Insertions

For amplifying the mitochondrial insertions of the different species, primers located in the nuclear flanking regions were used: GATAACTGAGATTTGAGGATTTG (primer 4 in fig. 1) and GCGAGCTGAAGGACTTTCTG (primer 2 in fig. 1). Standard PCR amplifications were done on 300 ng of genomic DNA, predenatured at 92°C for 2 min. Thirty cycles with primer annealing at either 62°C (for humans and chimpanzees) or 57°C (for the other hominoid and Old World monkey DNA samples) followed by extension at 72°C for 1 min and denaturation at 92°C for 40 s were carried out. PCR products were resolved on 1% agarose gels and cloned as above. Five clones were sequenced for each amplification. The respective sequences are available in GenBank under the accession numbers AF035463 for human (HSA), AF035464 for chimpanzee (PTR), AF035465 for gorilla (GGO), AF035466 for orangutan (PPY), and AF035467 for siamang (HSY).

For mapping the insertion in human chromosomes, the same PCR protocol was applied on a set of rodent/human hybrid cell lines of DNA (Coriell mapping panel #2).

Amplification and Sequencing of the Mitochondrial Control Region

One member each of the Hylobatidae and Cercopithecoidea and one representative of the New World monkeys were sequenced. The primers L15997 (CACCATTAGCACCCAAAGCT) and H16498 (CCTGAAGTAGGAACCAGATG) were used to amplify the relevant region of the mitochondrial control region from 300 ng of total DNA. 25 cycles were done at 57°C annealing, 72°C extension (1 min), and 92°C denaturation (40 s). PCR products were cloned (see above), and five clones were sequenced for each reaction. The GenBank accession numbers are AF035462 for the common marmoset (CJA), AF035461 for the Celebes black ape (MNI), and AF035759 for the siamang (HSY).

Analysis of Sequence Data

The following published mitochondrial sequences (Horai et al. 1995) were included in the analysis: bonobo (accession number D38116), chimpanzee (D38113), and orangutan (D38115). In addition, one human mitochondrial sequence was taken from the data set described by Vigilant et al. (1991). The sequence corresponds to number 87 and was arbitrarily chosen from the sequences for which the complete mitochondrial control region was determined. Alignments were done with the CLUSTAL W program, version 1.7 (Thompson, Higgins, and Gibson 1994). Phylogenetic trees and branch lengths were computed by maximum likelihood assuming the HKY model of sequence evolution (Hasegawa, Kishino, and Yano 1985) using the PUZZLE program (Strimmer and von Haeseler 1996). Other programs used included SEQBOOT, DNADIST with maximum-likelihood distances, NEIGHBOR, DNAPARS, and DNAML of the PHYLIP package (Felsenstein 1993). Different models of sequence evolution were tested giving basically the same results. Changes accumulated in each branch of a tree were identified with MacClade (Maddison and Maddison 1992).

Results and Discussion

Isolation and Characterization of the Insertion

Various oligonucleotide probes specific for the human control region were used to hybridize restricted human nuclear DNA free of mitochondrial DNA (Zischler et al. 1995). A 1.8-kb signal after *Pvu* II digestion was generated by the oligonucleotide L16420. After size enrichment, the nuclear insertion was isolated by two rounds of PCR, first executing an anchored PCR and then applying a primer specific for human Alu elements

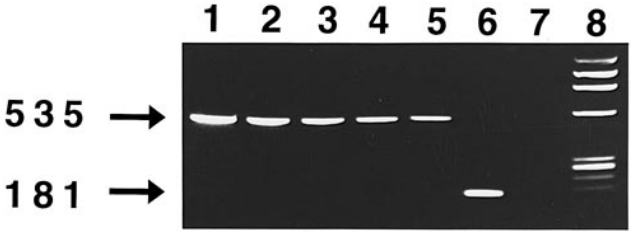

FIG. 2.—Electrophoresed and ethidium-bromide-stained PCR products obtained with primers flanking the insertion. Results with template DNA of each human (lane 1), chimpanzee (lane 2), gorilla (lane 3), orangutan (lane 4), siamang (lane 5), and rhesus monkey (lane 6) are shown. Lane 7 corresponds to the PCR control with no template DNA added, and lane 8 shows the size marker. The fragment sizes corresponding to the PCR products of the human and the rhesus monkey are indicated at the left in base pairs.

(Nelson et al. 1989). The two isolated overlapping PCR clones were sequenced in part to determine both the mitochondrial insert and its respective nuclear flanks. A map of the integration together with the flanks is shown in figure 1. The integration in the human nuclear genome comprises 365 bp with the ends coinciding with human mitochondrial DNA positions 16101 and 16477 (Anderson et al. 1981). Thus, the human nuclear integration is several base pairs shorter than its homologous region in the extant human mitochondrial sequence. One end of the insertion is close to the positions 16104, 16105, and 16106 (Anderson et al. 1981), which constitute the 3′ end of human 7S DNA, the third strand of the triple-stranded D-loop (indicated by a star in fig. 1; Doda, Wright, and Clayton 1981). The other insertion end extends 82 bp into the conserved part of the mitochondrial control region (Saccone, Pesole, and Sbisa 1991), and therefore ends about 200 bp upstream of one 5′ end of the 7S DNA (Tapper and Clayton 1981).

Primers specific for the flanking regions (primers 2 and 4 in fig. 1) were constructed and used to amplify DNA from a set of human/rodent cell lines, each carrying a single human chromosome. In this way, the insertion could be mapped to human chromosome 9 (data not shown). The same primers were used in PCR amplifications from 4 human individuals, as well as from 2 chimpanzees, 2 gorillas, 2 orangutans, and 2 individuals each of 14 different species belonging to the Hylobatidae (*Hylobates syndactylus,* siamang; *H. concolor,* crested gibbon; *H. moloch,* silvery gibbon; *H. lar agilis,* dark-handed gibbon; *H. pileatus,* capped gibbon) and Cercopithecoidea (*Macaca nemestrina,* pig-tailed macaque; *M. mulatta,* rhesus monkey; *M. nigra,* Celebes black ape; *Mandrillus sphinx,* mandrill; *Papio hamadryas,* sacred baboon; *Cercopithecus aethiops,* grivet; *C. albogularis,* white-throated guenon; *Cercocebus atterinus,* black mangabey; *Presbytis cristata,* silvered leaf monkey). The insertion could be amplified from all hominoid species, but only shorter fragments corresponding to the noninsertion allele were detected in all members of the Cercopithecoidea (fig. 2). The physical appearance of the insertion, as revealed by PCR, indicates that the transfer of this piece of mtDNA happened in a common hominoid ancestor after the lineages leading to the Hominoidea and Cercopithecoidea separated. Based on the fossil record, which dates the Cercopithecoidea–Hominoidea split at 30 MYA and the gibbon branching point at more than 17 MYA (Andrews 1986), this transfer occured in the range of 17–30 MYA.

The different hominoid PCR fragments representing the insertions were sequenced. For humans, the insertion sequences of two unrelated Europeans, one African, and one Japanese were determined. No intraspecific difference was observed. All hominoid insertion sequences were very similar to each other, with 96.2% identity between human and siamang, counting indels as differences. Moreover, the PCR fragment obtained from the rhesus monkey, representing the noninsertion allele, was also sequenced. The sequence comparison of the insertion and the noninsertion indicates an imprecise integration mechanism, as already observed for another control region insertion in humans (Zischler et al. 1995), where 11 bases were lost from the chromosomal sequence during the integration event.

Sequence Analysis of the Mitochondrial Insertions

In order to get more insight into the molecular evolution of such mtDNA integrations, we compared each insertion sequence from humans (HSA), chimpanzees (PTR), gorillas (GGO), orangutans (PPY), and siamangs (HSY) with published human (Vigilant et al. 1991), chimpanzee, bonobo (PPA), and orangutan (Horai et al. 1995) mitochondrial sequences. The gorilla mitochondrial sequence was not included, since this species shows a major deletion in this part of the control region when compared to other great apes. For completing the hominoid mitochondrial sequence set and getting both a mitochondrial sequence of a taxon that branched off before the integration took place and a more distant outgroup, we amplified mitochondrial DNA of the corresponding D-loop region from one representative each of the Hylobatidae (*Hylobates syndactylus,* HSY), the Cercopithecoidea (*Macaca nigra,* MNI), and the New World monkeys *(Callithrix jacchus,* CJA). Thus, altogether, five insertion sequences and seven mitochondrial D-loop sequences were aligned using the CLUSTAL W program (Thompson, Higgins, and Gibson 1994) without further manual changes.

The complete alignment (available from the authors on request) is 405 bp long. It allows a pairwise length comparison of this part of the D-loop with the respective nuclear integrations for the hominoid species for which both the mitochondrial D-loop and the corresponding integration sequences were determined: human (HSA), chimpanzee (PTR), orangutan (PPY), and siamang (HSY) (table 1). In total, 129 indels could be counted by summing up all pairwise comparisons in the mtDNA data set, and only 22 could be counted for the nuclear integrations. Thus, the mitochondrial mode of evolution leads to a higher sequence length heterogeneity than does the nuclear mode, in agreement with the results presented by Saitou and Ueda (1994). The range of indels spans from 1 to 7 bp in length. The latter represents an insertion on the lineage to the siamang mtDNA sequence. The major deletion in the gorilla D-loop se-

Table 1
Pairwise Comparisons of Indel Numbers Between Human (HSA), Chimpanzee (PTR), Orangutan (PPY), and Siamang (HSY) Mitochondrial D-Loop (above diagonal) and Integration Sequences (below diagonal)

Taxa	HSA	PTR	PPY	HSY
HSA	—	5	22	22
PTR.....	0	—	25	25
PPY.....	1	1	—	30
HSY	7	7	6	—

quence that precludes its use in the analysis also reflects the length heterogeneity in the mitochondrial sequences.

For further analyses, one variable part of the sequence alignment and all positions with gaps were removed, yielding a data set of 283 bp in length (fig. 3). First, we determined the number of substitutions for all pairwise comparisons (table 2). In total, 503 substitutions could be counted upon comparing the four mitochondrial sequences with each other, a value which is 14.4 times higher than the 35 substitutions calculated for the respective nuclear pseudogenes. Subdividing substitutions in transitions and transversions, a transition/transversion ratio of 0.9 can be calculated for the mitochondrial sequences, and one of 4.8 can be calculated for their nuclear homologs. Taking the transition bias observed upon comparing individuals of closely related species into account (Brown et al. 1982), this result indicates that the number of mtDNA transitions and, to a lesser extent, the number of mtDNA transversions are influenced by multiple hits, which is not the case for the nuclear pseudogenes.

Next, we related the insertion and mitochondrial sequences to each other using different phylogenetic tree reconstruction methods. In each case, the trees were rooted with the common marmoset mtDNA sequence. Expectedly, a maximum-likelihood-based algorithm (PUZZLE; Strimmer and von Haeseler 1996), maximum parsimony, and neighbor-joining grouped the nuclear sequences together. All three phylogenetic reconstruction methods placed the insertion clade correctly, in the lineage joining the *M. nigra* mitochondrial sequence with the most recent common ancestor (MRCA) of all hominoid mtDNA sequences (not shown). Thus, the presence/absence of the mtDNA insertion, as revealed by PCR, is reflected in the PUZZLE maximum-likelihood reconstruction with 84% support, as well as in the neighbor-joining and maximum-parsimony trees with 57% and 45% support, respectively, based on 1,000 bootstrap replications. Moreover, the mtDNA sequence subset always showed the widely accepted phylogeny (Shoshani et al. 1996; Ruvolo 1997). In contrast to this, the nuclear insertions were not related as expected due to the small number of differences among them.

Upon applying both the maximum-likelihood and neighbor-joining algorithms, a long branch leading to the ancestral node connecting the insertion sequences became apparent. To corroborate this observation, we gave the tree topology established from morphological and molecular studies (Shoshani et al. 1996; Ruvolo 1997) and the PCR evidence for presence/absence of the integration as a user tree and calculated the individual maximum-likelihood branch lengths with the PUZZLE program (fig. 4) following the HKY model of sequence evolution (Hasegawa, Kishino, and Yano 1985) with parameters estimated from the data set. As a result, the branch leading to the common ancestor of the insertion sequences accumulated 0.092 ± 0.023 substitutions per site, whereas the mitochondrial counterpart exhibited 0.056 ± 0.021. Similar results were obtained assuming a TN model of sequence evolution (Tamura and Nei 1993) and taking a gamma-distributed rate heterogeneity across sites into account for both models. The ratio of these two values is not explainable considering, e.g., the maximum-likelihood distances between human and siamang for both the mitochondrial and the respective nuclear sequences, which suggest a 20-fold slowdown of the evolutionary rate in the nucleus (see fig. 4 and Brown et al. 1982). Moreover, upon applying a DNAML routine, this tree shows a significantly higher likelihood than does a tree for which the branch length leading to the hominoid nuclear MRCA was set to $\frac{1}{20}$ of its mitochondrial counterpart. Thus, the MRCA of the nuclear insertions does not perfectly reflect the mitochondrial sequence at the time of transposition. In order to elucidate which positions most likely changed on both branches leading to either the hominoid mitochondrial or nuclear ancestral sequences, all substitutions that could be unambiguously assigned to one branch based on the principle of maximum parsimony were identified: 13 substitutions on the branch leading to the MRCA of the pseudogene sequences and 11 substitutions on the respective mtDNA branch were counted using the MacClade program (the former substitutions are indicated by stars in fig. 3). These values also reflect the unexpected branch length ratio already seen applying a maximum-likelihood as well as a distance-based branch length estimation.

To investigate whether the variable regions of the alignment affect the conclusions, we removed three blocks of higher sequence divergence (see the underlined regions in fig. 3) and recalculated the branch lengths: the observed amounts of substitutions per position were 0.048 ± 0.018 for the nuclear lineage and 0.064 ± 0.022 for the mitochondrial lineage leading to the respective hominoid MRCAs. Considering the ratio of the nuclear and mitochondrial human–siamang maximum-likelihood distances for this alignment, the branch leading to the MRCA of the nuclear integrations is about one order of magnitude longer than expected. On the basis of a maximum-parsimony criterion, six unambiguous changes were accumulated on the nuclear branch, and eight were accumulated on the comparable mitochondrial branch. Thus, a more conservative alignment also reveals a higher-than-expected postintegration accumulation of substitutions, although to a lesser extent. We cannot exclude, therefore, the possibility that parallel substitutions in the mitochondrial lineages might have led to erroneous assignments for some of the changes on the lineage to the MRCA of the insertions.

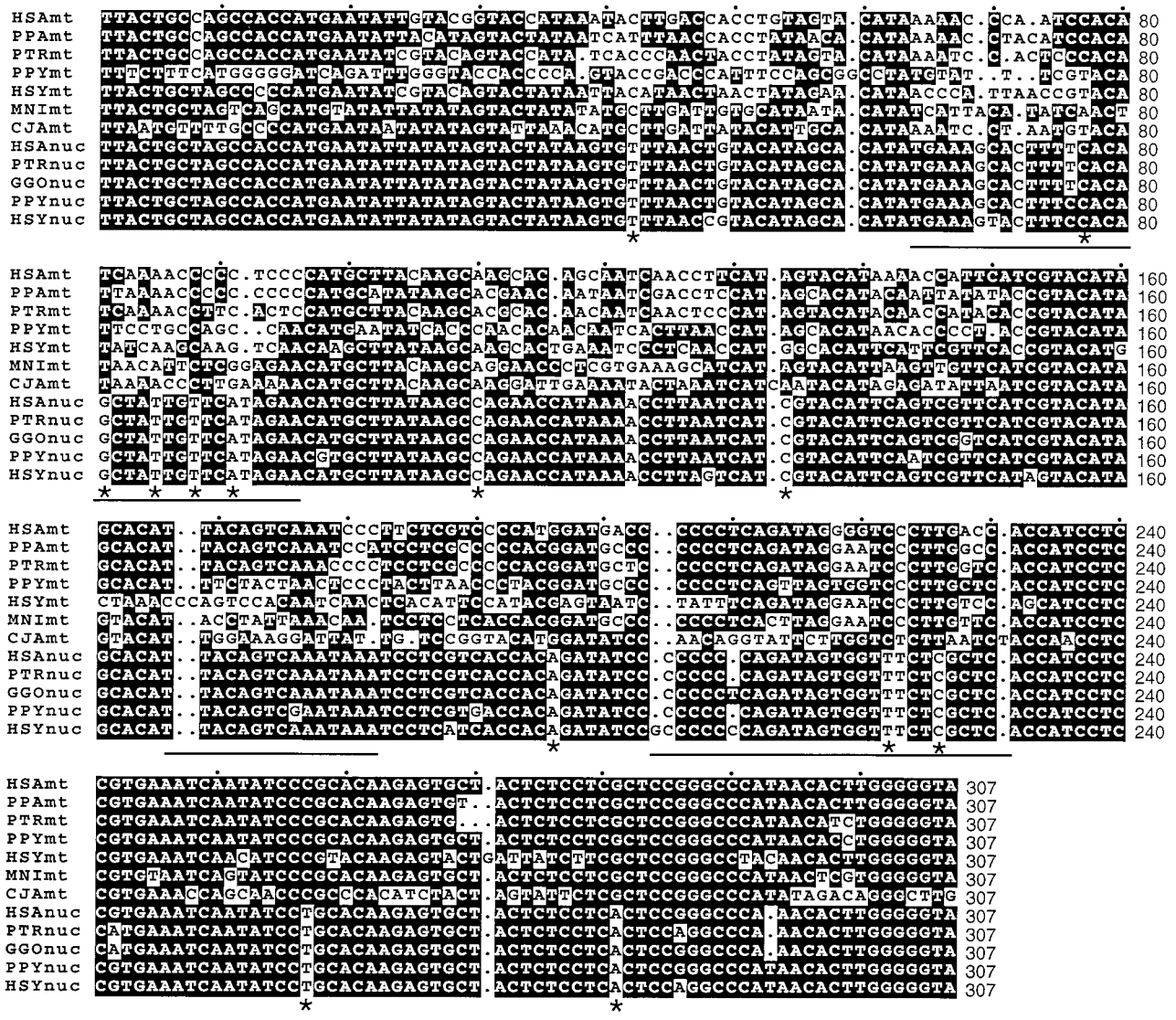

FIG. 3.—Alignment of five hominoid integration sequences and seven mitochondrial D-loop sequences after removing a variable block. Positions with gaps were excluded from the phylogenetic analyses. The species names are HSA for human, PPA for bonobo, PTR for chimpanzee, PPY for orangutan, HSY for siamang, MNI for Celebes black ape, CJA for common marmoset, and GGO for gorilla. Mt designates the mitochondrial D-loop sequence, and nuc designates the nuclear integration. Stars indicate unambiguous changes in the branch leading to the MRCA of the inserts according to parsimony criteria. Lines indicate blocks of higher sequence divergence that were removed to yield an alternative, more conservative alignment for branch length estimations. Dots above the sequences indicate intervals of 10 bp.

Table 2
Difference Matrix of Mitochondrial D-Loop Sequences (plain text) from Human (HSA), Chimpanzee (PTR), Orangutan (PPY), and Siamang (HSY) and their Nuclear Counterparts (in bold)

Taxa	HSA	PTR	PPY	HSY
HSA	—	28/**2**	39/**4**	44/**5**
PTR.....	7/**0**	—	37/**6**	37/**5**
PPY.....	52/**1**	55/**1**	—	51/**7**
HSY	46/**1**	45/**1**	62/**2**	—

NOTE.—Transitions are listed above the diagonal, and transversions are listed below the diagonal.

Damaged Inserts or a Change in the Mode of Evolution?

If the relatively long branch leading to the MRCA of the insertion sequences is attributed to damage during transposition (Collura and Stewart 1995), an accumulation of insert-specific substitutions should be observable right after the integration process, e.g., in recent integrations that are clearly confined to one species. Upon comparing 243 bp of a human-specific integration (Zischler et al. 1995) that corresponds to the conserved part of the mitochondrial D-loop of all available hominoid ($n = 14$) and human ($n = 261$) mtDNA sequences, only two insert-specific substitutions were traced, with both the ape mtDNA sequences and all human mtDNA sequences being identical at these positions. In another example of a recent mtDNA insertion (Fukuda et al.

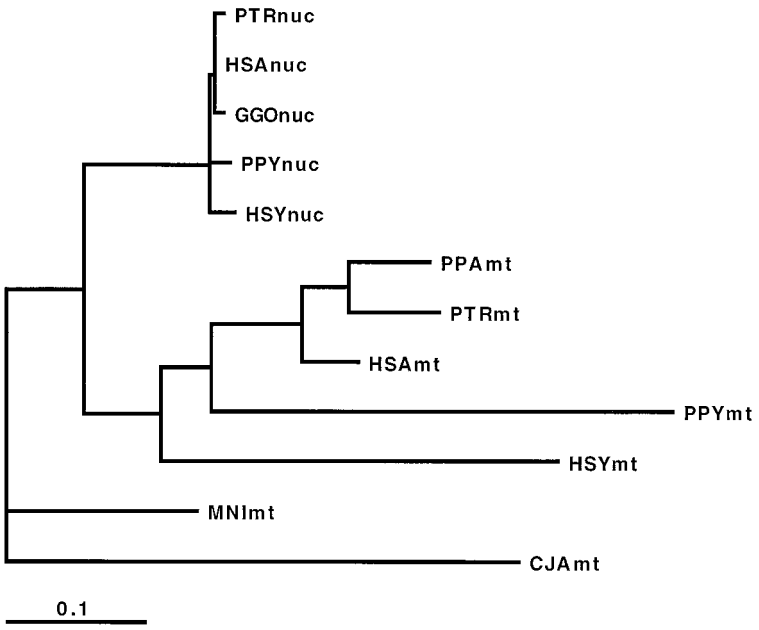

FIG. 4.—Maximum-likelihood calculation of branch lengths using the PUZZLE program with the preset tree: (CJAmt, (((((PTRnuc, HSAnuc), GGOnuc), PPYnuc), HSYnuc), ((((PPAmt, PTRmt), HSAmt), PPYmt), HSYmt), MNImt). The tree is rooted with the mitochondrial sequence of the common marmoset. Branch lengths were computed using the HKY model of sequence evolution (Hasegawa, Kishino, and Yano 1985) with parameters estimated from the data set. They are drawn according to the number of substitutions per position. The bar indicates 0.1 substitutions per position. See the legend to figure 3 for the abbreviations.

1985), a partial ND5-integration in humans exhibited no substitution over 215 bp upon comparison with the reference mtDNA sequence (Anderson et al. 1981). Thus, analyzing recent integrations, the number of insertion-specific substitutions is low, ruling out in these cases a damaging event during the DNA transfer from the mitochondria to the nucleus. It therefore appears that recent integrations faithfully reflect the mitochondrial sequence at the time of transposition.

On the other hand, for older integrations like the one described here, it might be speculated that a sudden change of the sequence-shaping forces that coevolved with a sequence over a long time span leads to a post-transpositional temporary increase in sequence change. On the molecular level, this can be due to a shift from the mitochondrial γ polymerase-based replication machinery to the nuclear α, β, δ, and ϵ polymerase system, which is involved in replication, repair, and recombination of nuclear DNA. Alternatively, the sequence changes could have been caused by DNA mutations specific for the nucleus (Hu and Thilly 1994). For example, there are two transitions (at positions 257 and 281; Fig. 3) which occur in the most conserved central part of the control region, where multiple hits on mitochondrial lineages should play a minor role. Both transitions happened in positions that are part of mitochondrial CpG-dinucleotides not only in the species presented herein, but also in other mamalian species (dolphin, cow, rat, and mouse; see Saccone, Pesole, and Sbisa 1991). Since mitochondrial DNA is not methylated in the mitochondria (Groot and Kroon 1979) but gets methylated de novo mainly in the CpG-dinucleotides (Yang et al. 1995) after transposition to the nucleus, the observed substitutions are in agreement with changes caused by spontaneous deamination reactions of methylated cytosines. Thus, they could reflect a nuclear-specific mutational pathway.

Therefore, the emerging picture on evolution of nuclear insertions is that older integrations (with transposition times in the range of several millions of years) accumulate more substitutions in relatively short time spans after the integration, as expected from the differences in the evolutionary dynamics of mitochondrial and nuclear sequences. DNA damage during transposition and integration is unlikely, but the detection and precise characterization of more recent integrations, together with their flanking nuclear sequences, will show if this is true for all mtDNA transpositions to the nucleus. Recent integrations will therefore exhibit a more pronounced fossil character than the ones that have been subject to a nuclear mode of evolution for several millions of years. The fossil character of the integration presented herein is not much affected by the substitutions taking place in the nucleus, since their relative number is low. Moreover, the inclusion of mitochondrial outgroup sequences helps to determine the ancestral mitochondrial states at these positions more accurately (Lopez et al. 1997). In the case of sequence length variability, the nuclear integrations offer the advantage of showing less heterogeneity than their mtDNA counterparts. Thus, the MRCA sequence deduced from homologous integrations in different species will represent the ancestral mtDNA sequence more reliably and with less sequence ambiguities than an ancestral sequence deduced from the fast-evolving mtDNA sequences. In this way, the isolation and analysis of mtDNA integrations will help to trace the history of especially fast evolving mtDNA sequences over long evolutionary timescales.

Acknowledgments

This work was carried out in the laboratory of Prof. S. Pääbo, to whom the authors are grateful for discussions, support, and providing facilities. Drs. R. Bichlmaier, I. Nanda, and R. D. Martin are thanked for DNA samples, and M. Storhas is thanked for experimental contributions during a practical course. We are indebted to Drs. A. von Haeseler, K. Strimmer, H. Poinar, and G. V. Boerner for help and discussions. Financial support was received from the DFG.

LITERATURE CITED

ANDERSON, S., A. T. BANKIER, B. G. BARRELL et al. (14 coauthors). 1981. Sequence and organization of the human mitochondrial genome. Nature **290**:457–465.

ANDREWS, P. 1986. Fossil evidence on human origins and dispersal. Cold Spring Harb. Symp. Quant. Biol. **51**:419–428.

BLANCHARD, J. L., and G. W. SCHMIDT. 1995. Pervasive migration of organellar DNA to the nucleus in plants. J. Mol. Evol. **41**:397–406.

BROWN, W. M., E. M. PRAGER, A. WANG, and A. C. WILSON. 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol. **18**:225–239.

COLLURA, R. V., and C.-B. STEWART. 1995. Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. Nature **378**:485–489.

DODA, J. N., C. T. WRIGHT, and D. A. CLAYTON. 1981. Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc. Natl. Acad. Sci. USA **78**:6116–6120.

FELSENSTEIN, J. 1993. PHYLIP (phylogeny inference package). Version 3.5c. Distributed by the author, Department of Genetics, University of Washington, Seattle.

FUKUDA, M., S. WAKASUGI, T. TSUZUKI, H. NOMIYAMA, K. SHIMADA, and T. MIYATA. 1985. Mitochondrial DNA-like sequences in the human nuclear genome. Characterization and implications in the evolution of mitochondrial DNA. J. Mol. Biol. **186**:257–266.

GROOT, G. S., and A. M. KROON. 1979. Mitochondrial DNA from various organisms does not contain internally methylated cytosine in -CCGG- sequences. Biochem. Biophys. Acta **564**:355–357.

HASEGAWA, M., H. KISHINO, and T. YANO. 1985. Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22**:160–174.

HORAI, S., K. HAYASAKA, R. KONDO, K. TSUGANE, and N. TAKAHATA. 1995. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc. Natl. Acad. Sci. USA **92**:532–536.

HU, G., and W. G. THILLY. 1994. Evolutionary trail of the mitochondrial genome as based on human 16S rDNA pseudogenes. Gene **147**:197–204.

LOPEZ, J. V., M. CULVER, J. C. STEPHENS, W. E. JOHNSON, and S. J. O'BRIEN. 1997. Rates of nuclear and cytoplasmic mitochondrial DNA sequence divergence in mammals. Mol. Biol. Evol. **14**:277–286.

MADDISON, W. P., and D. R. MADDISON. 1992. MacClade: analysis of phylogeny and character evolution. Version 3.0 Sinauer, Sunderland, Mass.

NELSON, D. L., S. A. LEDBETTER, L. CORBO, M. F. VICTORIA, R. RAMIREZ-SOLIS, T. D. WEBSTER, D. H. LEDBETTER, and C. T. CASKEY. 1989. Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. Proc. Natl. Acad. Sci. USA **86**:6686–6690.

PERNA, N. T., and T. D. KOCHER. 1996. Mitochondrial DNA: molecular fossils in the nucleus. Curr. Biol. **6**:128–129.

RUVOLO, M. 1997. Molecular phylogeny of the hominoids: inferences from multiple independent DNA sequence data sets. Mol. Biol. Evol. **14**:248–265.

SACCONE, C., G. PESOLE, and E. SBISA. 1991. The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary pattern. J. Mol. Evol. **33**:83–91.

SAITOU, N., and S. UEDA. 1994. Evolutionary rates of insertion and deletion in noncoding nucleotide sequences of primates. Mol. Biol. Evol. **11**:504–512.

SAMBROOK, J., E. F. FRITCH, and T. MANIATIS. 1989. Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

SHOSHANI, J., C. P. GROVES, E. L. SIMONS, and G. F. GUNNELL. 1996. Primate phylogeny: morphological vs molecular results. Mol. Phylogenet. Evol. **5**:102–154.

STRIMMER, K., and A. VON HAESELER. 1996. Quartet puzzling: a maximum likelihood method for reconstructing tree topologies. Mol. Biol. Evol. **13**:964–969.

TAMURA, K., and M. NEI. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. J. Mol. Evol. **10**: 512–526.

TAPPER, D. P., and D. A. CLAYTON. 1981. Mechanism of replication of human mitochondrial DNA. Localization of the 5′ ends of nascent daughter strands. J. Biol. Chem. **256**: 5109–5115.

THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.

VIGILANT, L., M. STONEKING, H. HARPENDING, K. HAWKES, and A. C. WILSON. 1991. African populations and the evolution of human mitochondrial DNA. Science **253**:1503–1507.

YANG, A. S., J.-C. SHEN, J.-M. ZINGG, S. MI, and P. A. JONES. 1995. HhaI and HpaII DNA methyltransferases bind DNA mismatches, methylate uracil and block DNA repair. Nucleic Acids Res. **23**:1380–1387.

ZHANG, D.-X., and G. M. HEWITT. 1996. Nuclear integrations: challenges for mitochondrial DNA markers. Trends Ecol. Evol. **11**:247–251.

ZISCHLER, H., H. GEISERT, A. VON HAESELER, and S. PÄÄBO. 1995. A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans. Nature **378**:489–492.

DAVID M. RAND, reviewing editor

Accepted January 9, 1998