Published in final edited form as:

Int J Primatol. 2008 October 1; 29(5): 1341–1353. doi:10.1007/s10764-008-9303-4.

Mitochondrial *COII* Introgression into the Nuclear Genome of *Gorilla gorilla*

Wai Kwan Chung and Michael E. Steiper

W. K. Chung · M. E. Steiper; Department of Anthropology, Hunter College of the City University of New York, New York, NY 10065, USA, e-mail: msteiper@hunter.cuny.edu

M. E. Steiper; Department of Anthropology, Programs in Anthropology and Biology, The Graduate Center of the City University of New York, New York, NY 10016, USA

Abstract

Numts are nonfunctional mitochondrial sequences that have translocated into nuclear DNA, where they evolve independently from the original mitochondrial DNA (mtDNA) sequence. Numts can be unintentionally amplified in addition to authentic mtDNA, complicating both the analysis and interpretation of mtDNA-based studies. Amplification of numts creates particular issues for studies on the noncoding, *hypervariable 1* mtDNA region of gorillas. We provide data on putative numt sequences of the coding mitochondrial gene *cytochrome oxidase subunit II (COII)*. Via polymerase chain reaction (PCR) and cloning, we obtained *COII* sequences for gorilla, orangutan, and human high-quality DNA and also from a gorilla fecal DNA sample. Both gorilla and orangutan samples yielded putative numt sequences. Phylogenetically more anciently transferred numts were amplified with a greater incidence from the gorilla fecal DNA sample than from the high-quality gorilla sample. Data on phylogenetically more recently transferred numts are equivocal. We further demonstrate the need for additional investigations into the use of mtDNA markers for noninvasively collected samples from gorillas and other primates.

Keywords

gorilla; mtDNA; noninvasive samples; numts

Introduction

Researchers have applied mitochondrial DNA (mtDNA) analyses to a wide variety of questions concerning the evolutionary history of primates. Though the use of mtDNA for phylogenetic and population genetic analyses has provided information on many species, issues remain in its application. One important issue relates to numts, which are regions of mtDNA that have been incorporated into the nuclear genome. Within primates, numts occur, for example, in the *hypervariable 1* region (*HVI*) of great apes (Jensen-Seaman *et al.* 2004; Thalmann *et al.* 2004, 2005; Zischler *et al.* 1995, 1998) and the *cytochrome b* gene of Old World and New World monkeys (Collura and Stewart 1995; Mundy *et al.* 2000; Schmitz *et al.* 2005). Recent studies suggest that the gorilla genome has a particularly high incidence of numts (Clifford *et al.* 2004; Jensen-Seaman *et al.* 2004; Thalmann *et al.* 2004). The collection of numts has complicated the use of mtDNA as a population genetic and phylogenetic marker for gorillas

Correspondence to: Michael E. Steiper.

Present address: W. K. Chung, Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ 86011, USA

(Anthony *et al.* 2007; Clifford *et al.* 2004; Jensen-Seaman *et al.* 2004; Thalmann *et al.* 2004, 2005).

Researchers have suggested many explanations for the differential amplification and collection of numt sequences. At the *cytochrome b* gene of catarrhines, Collura and Stewart (1995) found that orangutan numt sequences are preferentially amplified when using universal primers (ones that can be used to amplify a particular DNA region across many different species). Similarly, in a study on gorillas exploring the presence of numts of *HVI*, a hypervariable noncoding segment of the mtDNA often used in population genetics studies, Thalmann *et al.* (2004) showed that the use of universal primers was likely to amplify numt sequences in chimpanzees and gorillas, regardless of precautions taken to avoid the problem. They also found that the use of species-specific *HVI* primers was helpful to avoid numt sequences in chimpanzees, but was ineffective in gorillas. The results are consistent with those of other researchers who have found many *HVI* numts in gorillas (Anthony *et al.* 2007; Clifford *et al.* 2004; Jensen-Seaman *et al.* 2004; Thalmann *et al.* 2005). Nonspecies-specific factors also contribute to the amplification of numts, such as extraction method, quality, and condition of samples (Bensasson *et al.* 2001). Previous studies to detect numts in gorillas (Jensen-Seaman *et al.* 2004; Thalmann *et al.* 2005) have utilized extracted DNA from a range of samples, including liver, blood, hair, feces, and teeth. DNA extracted from noninvasive samples may prove particularly likely to yield numts (Greenwood and Pääbo 1999; Jensen-Seaman *et al.* 2004; Thalmann *et al.* 2004).

We examined the differential incidence of numts and authentic mtDNA sequences from within polymerase chain reactions (PCRs) at the mtDNA gene *cytochrome oxidase subunit II* (*COII*) in gorillas, humans, and orangutans. We chose the *COII* gene to examine the PCR-based collection of numts from a more slowly evolving and coding mtDNA region than the oft-used noncoding *HVI* mtDNA region. First, we examined, via PCR, cloning, and sequencing of independent colonies from *COII* genes of human, gorilla, and orangutan, the differential incidence of both numt and authentic mtDNA sequences collected within PCRs. The methods determine whether there is a different incidence of numt amplification and collection within PCRs across these 3 species. Second, we evaluated the effect of sample quality on the rate of numt and authentic mtDNA sequences collected. We compared, via PCR, cloning, and DNA sequencing of individual clones, a high-quality gorilla DNA sample and a noninvasively collected fecal sample to determine whether sample type affects the incidence of numts vs. authentic mtDNA at the *COII* gene of gorillas.

Materials and Methods

We obtained genomic DNA purified from cell cultures of *Gorilla gorilla gorilla* (NG05251), *Pongo pygmaeus abelii* (NG12256), and *Homo sapiens* (NAIMR91) from the Coriell Institute for Medical Research. We obtained western gorilla (*Gorilla gorilla gorilla*) DNA from a <24 h-old fecal sample, collected in 70% ethanol and extracted via the QIAamp DNA Stool Mini Kit (sample provided by J. Satkoski and J. Dupain). We did not quantify the fecal DNA.

PCR Amplification

We performed PCR in 25- μ l reactions containing 1 $\times$ Mg-free thermophilic polymerase 10 $\times$ reaction buffer (Promega), 2.5 μ M MgCl₂, 0.5 μ M per primer (forward and reverse), 0.1 mM dNTPs, *ca.* 0.5 ng of DNA from the high-quality DNA or 1 μ l from the unquantified fecal DNA and 2.5 U of *Taq* DNA polymerase (Promega). Forward and reverse oligonucleotide primers specific for the mitochondrial *COII* gene were based on Ruvolo *et al.* (1991): A7552, 5'-AAC CAT TTC ATA ACT TTG TCA A-3' and B8321, 5'-CTC TTA ATC TTT AAC TTA AAA G-3'. Ruvolo *et al.* (1991, 1994) used the sequences to amplify *COII* across catarrhine primates. For reactions that had fecal DNA as template, we added 0.1 μ g/ μ l of bovine serum albumin (BSA) to the final master mix. We based cycling times and temperatures on Ruvolo

et al. (1991): 2 min at 95°C for the denaturation, followed by 35 cycles of 95°C for 1 min, 50°C for 1 min, and 72°C for 1 min. For the human and orangutan, we performed only 1 PCR. For both the gorilla high-quality and fecal samples, we performed 2 PCRs.

PCR products from the extracted fecal samples were less optimal for TA cloning, and we employed 2 additional steps to ensure successful insertion of the PCR product into the vector. First, after amplification, we used 8 µl of PCR product from the fecal samples in an additional reaction with *Taq* to produce sticky ends. Sticky ends are the single-nucleotide A overhang required for TA cloning. Subsequently, we purified 10 µl of the sticky end reaction per the Montage (Millipore) purification protocol.

Cloning and Sequencing

Cloning and insertion into pCR2.1-TOPO vector followed the Invitrogen TOPO TA cloning kit protocol. We transformed the plasmid chemically into One Shot TOP10 competent cells, plated it onto prewarmed kanamycin agar plates, and incubated overnight. We incubated isolated colonies from the plates overnight in LB and kanamycin. We extracted DNA from the cultures via the Eppendorf FastPlasmid Mini miniprep kit, per the manufacturer's instructions.

We selected 30 insert-positive samples at random, assessed via *EcoRI* digestion, from each species for sequencing with standard T7 and M13-reverse primers, thus sequencing both strands. We sent all samples to the Dana Farber/Harvard Cancer Center (DF/HCC) DNA Resource Core facility for sequencing.

We sent 6 sets of samples for sequencing. For each independent PCR, we sent *ca.* 30 clones for sequencing. Several human and orangutan sequences failed to produce usable data. After viewing the electropherograms for each sequence, we discarded ones that showed weak peaks with many blank nucleotides or were not long enough for analysis, or contained numerous ambiguous or unresolved nucleotides. We added all usable sequences to the previously published *COII* sequences (Ruvolo *et al.* 1993, 1994) available from GenBank (Table I). They included 3 western gorillas (*Gorilla gorilla gorilla*) Ggo1, Ggo3, and Ggo4; 1 Grauer's gorilla (*Gorilla gorilla graueri*) Ggo5; and 1 mountain gorilla (*Gorilla gorilla beringei*) Ggo6. We manually entered sequences listed in Table I without GenBank accession numbers (1 human, 1 common chimpanzee, 1 bonobo, 1 gorilla, and 1 orangutan) based on nucleotide differences in Ruvolo *et al.* (1993) and the parsimony consensus tree in Ruvolo *et al.* (1994). The published sequences have a high likelihood of being authentic mtDNA sequences because the DNA from ≥ 1 of the 2 studies of Ruvolo *et al.* (1991) was isolated via cesium chloride/propidium iodide gradient centrifugation, a method that preferentially isolates mtDNA. Further, the published sequences did not contain extra stop codons. Our clone sequences are available from GenBank (accession no. EU834957-EU835121).

Phylogenetic Analyses

We trimmed all sequences generated from clones via 4Peaks 1.7. We assembled forward and reverse sequences of all valid sequences in Vector NTI Contig Express to produce a consensus sequence from both forward and reverse strands. We aligned complete contigs from each colony to the previously published primate *COII* sequence alignment via CLUSTAL-X 1.8 (Thompson *et al.* 1997) under the default settings. We imported aligned sequences to PAUP* 4.0 (Swofford 2001) for phylogenetic analysis and further alignment.

Determination of Numt Sequences

Because numt sequences can be highly similar to authentic mtDNA sequences, we employed several methods to define a sequence from a particular clone as a numt. First, we analyzed the sequences phylogenetically with the GenBank *COII* mtDNA sequences. Using neighbor-

joining (Saitou and Nei 1987) via the HKY85 algorithm (Hasegawa *et al.* 1985), we defined sequences as putative numt sequences if they fell into separate phylogenetic groups, distinct from the mtDNA sequences already assigned to the species. We used neighbor-joining because the use of maximum parsimony on the data set was exceptionally computationally intensive and time consuming, without providing additional information. The method would best find numts that had integrated on a deeper time frame than within a species, referred to here as older numts. When phylogenetic analysis suggested that sequences were older numts, we analyzed the sequences further. We compared putative numt sequences via *blast* to decipher whether the highest match was to mtDNA or nuclear DNA. Because *COII* is translated, we also checked sequences for the presence of stop codons or frame-shift mutations, which would also indicate a numt.

We also used a phylogenetically based method to define putative younger numts within gorillas, which may have integrated on a more recent time scale. We analyzed the clone sequences from all 4 gorilla PCRs together via maximum parsimony without a bias toward transitions and transversions. We interpreted phylogenetic groups, based on the strict consensus tree of clone sequences from ≥ 2 independent PCRs as evidence that a numt had been amplified. The method essentially assumes that any SNP synapomorphy that was shared in clones from 2 separate PCRs indicates a numt amplification and not a homoplastic PCR error that affected the same site.

Results

We sequenced 165 clones from 6 PCRs: 2 each from high-quality DNA of a gorilla and from DNA extracted from a gorilla fecal sample, and 1 each from an orangutan DNA sample and a human DNA sample, resulting in a 690-base-pair alignment. Via the distance- and parsimony-based methods, the phylogenetic trees show the presence of sequences that represent potential *COII* numts, defined as sequences that form separate phylogenetic groups when compared to other clone sequences from the same sample (older numts) and also as unique clades within the 4 gorilla PCRs (younger numts).

High-quality Samples

An HKY-based neighbor-joining distance tree for the alignment of all sequenced clones and previously collected sequences is in Fig. 1. We found no phylogenetic outlier in any of the 30 human clone DNA sequences. All human clone sequences formed a phylogenetic group with published human *COII* sequences from GenBank, which are assumed to be authentic mtDNA. The average pairwise difference between the clone sequences is 0.2%. In orangutans, we recovered 2 phylogenetic outlier clone sequences: Orang_8 and Orang_9. Based on a *blast* search, the 2 clone sequences are most similar to a region of human chromosome 17 and had a number of stop codons (Table II). The two clone sequences are part of group A and are thereby classified as old numts. One orangutan sequence, Orang_22, is not a phylogenetic outlier, has a single stop codon, and has high similarity to orangutan mtDNA sequences, which we interpreted as a non-numt sequence with a base misincorporation. The remaining clone sequences formed a group with the orangutan mtDNA sequences from GenBank. Among all orangutan clones, the average pairwise difference is 3.2%; when we removed the phylogenetic outliers, it dropped to 0.2%.

The first high-quality gorilla PCR (GorillaPCR1) generated 2 colonies that yielded DNA sequences that are phylogenetic outliers: GorillaPCR1_34 and GorillaPCR1_12 (Fig. 1). The 34 clone sequence has *blast* similarity to region on human chromosome 17 (group B). The 12 clone sequence fell outside of the main group of clone sequences, but within other gorilla sequences. This was due to a particular presumed PCR error at position 82 that is shared with other species, such as *Pongo pygmaeus*. We analyzed further all 30 clone sequences from the

set for presence of stop codons, which occurred only in the 34 sequence (Table II). One gorilla clone sequence (GorillaPCR1_17) is not a phylogenetic outlier, but has a single stop codon and high similarity to gorilla mtDNA sequences, which we interpret as a non-numt sequence that had a base misincorporation. The remaining clone sequences formed a group with the gorilla mtDNA sequences from GenBank. The second high-quality gorilla PCR (GorillaPCR2) yielded 5 clones that contain outlier DNA sequences (Fig. 1). The clone sequences fell into 3 distinct phylogenetic groups: C, D, and E. The groups have affinities to different regions of the human and chimpanzee nuclear genomes (Table II) and they also contain multiple stop codons. The remaining clone sequences formed a group with the gorilla mtDNA sequences from GenBank.

Noninvasive Fecal Samples

The first fecal gorilla PCR (GorFecPCR1) yielded 9 colonies that contained DNA sequences that are phylogenetic outliers (Fig. 1). The clone sequences fell into 4 phylogenetic groups: A, B, D, and E. Most of the sequences are similar to other hominoid nuclear regions, though one colony's sequence has its closest matches to gorilla and orangutan mtDNA. However, the similarity is low (89%) and the sequence has a number of stop codons, as did the other outlier sequences. The remaining clone sequences form a group with the gorilla mtDNA sequences from GenBank. The second fecal gorilla PCR (GorFecPCR2) yielded 8 colonies that contained DNA sequences that are outliers (Fig. 1). Of these, 7 clone sequences fell into group E and had multiple stop codons (Table II). One additional sequence (GorFecPCR2_33) showed a 99% match to the human mitochondrial genome. This is the only gorilla sequence to match closely to a sequence of the human mitochondrial genome. The clone may have been caused by contamination and we did not include it in the statistical analyses. The remaining clone sequences formed a group with the gorilla mtDNA sequences from GenBank. When we analyzed all (high-quality and fecal) gorilla clones, the average pairwise difference was 7.9%. Removing the outliers dropped this to 0.9%.

Assessment of Younger Numts

The prior phylogenetic method would have best assessed cloned numt sequences that had transferred early within hominoid evolutionary history. However, because the method relies on the presence of stop codons and phylogenetic distinctiveness, it would have potentially missed clones derived from numts more recently transferred from the mtDNA genome. To find recent numts within gorillas, we used maximum parsimony to analyze all gorilla clones from all 4 independent PCRs in the same analysis (2 high-quality and 2 fecal) with the GenBank data. Here, we considered any grouping of clones from >1 independent PCR as representing a colony that derived from a more recently transferred numt because the probability of a PCR error affecting the same base position in 2 separate PCRs is low ($1/690 \text{ b.p.} \cdot 1/690 \text{ b.p.} = 2 \cdot 10^{-6}$), even discounting that the differences are changes to the same base. Here we assume that the role of other phenomena, such as the use of DNA from a cell line, is not a factor. Based on a consensus tree there are 9 such groups (Fig. 2), of which 9 clones were from high-quality DNA and 15 were from fecal PCRs (Fig. 2 and Table III). Most of the groups are defined by 1 shared base position. In all cases, they were transitions. None of the changes caused stop codons. Here we distinguish between the young numts and the older numts. Note that the old numt groups were also replicated in the parsimony tree. When we removed both old and young numts, the average pairwise difference among the gorilla clone sequences is 1.0%.

Statistical Analysis of Clone Sequences

The 2 high-quality gorilla PCRs each resulted in 30 clone sequences. In PCR1, there was 1 inferred older numt. In PCR2, there were 5 older inferred numts. The difference in the incidence of numt amplification is not significant (Fisher's exact test, $p=0.19$, 2-tailed). When we added

the inferred younger numts (0 from PCR1 and 9 from PCR2), the difference was significant (Fisher's exact test, $p < 0.001$, 2-tailed). The number of inferred older numts for humans was 0, and for orangutans 2. There is no significant difference between the 3 high-quality samples for older numts (human, orangutan, and gorilla [combined]; Fisher's exact test 2×3 , $p = 0.28$, 2-tailed). Because we did not use the same method to detect the collection of clones from younger numts in humans and orangutans, we did not perform the subsequent test. There is no significant difference between the number of clones from inferred older numts in the 2 fecal gorilla samples, 7 and 9 numts (Fisher's Exact Test, $p = 0.77$, 2-tailed), or clones from both younger and older numts (6 young from FecPCR1 and 9 young from FecPCR2; $p = 0.8$). There is a significant difference between the combined gorilla high-quality samples and the combined fecal gorilla samples for clones from older numts (Fisher's exact test, $p = 0.02$, 2-tailed). We assume that the difference is due to sample type and not to interindividual differences. However, when we combined clones that were from both older and younger numts, the high-quality PCRs were significantly different and could not be combined ($p < 0.001$). If we used PCR1 as a comparison, the difference is significant between high-quality and fecal samples, but if we used PCR2, it is not. Also, based on only the older numts, an exact test of the 4 samples (human, orangutan, gorilla high-quality combined, gorilla fecal combined) is significant (Fisher's exact test 2×4 , $p = 0.002$, 2-tailed) for the older numts.

Discussion

We found 7 clone sequences inferred to be older numt sequences out of 60 clone sequences from high-quality gorilla DNA (PCR1: 2/30, 6%; PCR2: 5/30, 17%), and 16 clones sequences from putative older numts out of 60 fecal DNA derived clone sequences (fecal 1: 9/30, 30%; fecal 2: 7/29, 24%). We inferred the clone sequences to be numts via phylogenetic analyses, *blast* searches, and the presence of stop codons. The groups correspond to particular numts that are present in the human genome. Group A and B numts correspond to a human chromosome 17 numt, group C to a human chromosome 5 numt, group D to a human chromosome 2 numt, and group E to a human chromosome 6 numt. Though there are ≥ 452 numts of varying length and age in the human genome (Hazkani-Covo and Graur 2007), few researchers have targeted the introgression history of particular mtDNA regions. Phylogenetic methods also recovered additional sequences that may potentially be younger numts among the clones sequenced from gorilla DNA. In this respect, we succeeded in providing data on the presence, incidence, and sequences of numts from the *COII* region of mtDNA among sequences generated from within a PCR based on 2 different types of DNA samples.

There is a significant difference among the sample types in the recovery of clones containing older numts relative to authentic mtDNA. The fecal-derived gorilla DNA yielded a significantly higher incidence of phylogenetic outliers, inferred to be older numts. Comparison of the 2 gorilla sample types suggests that fecal samples may be more prone to the amplification of older numts. For the younger numts, more numts were inferred from the fecal samples than from the high-quality samples, but the 2 high-quality samples themselves yielded different incidences of inferred younger numts. Generally, this is consistent with previously published data that show a higher ratio of numts in elephant hair vs. blood samples (Greenwood and Pääbo 1999). Though the incidence of older numts was greater overall in gorillas than in humans and orangutans, the difference is not statistically significant. However, our small sample size prevents one from drawing a definitive conclusion. Overall, it appears that the incidence of numts amplified from within a PCR is lower for *COII* than for the *HVI* region, where most analyzed clones were numts (Jensen-Seaman *et al.* 2004; Thalmann *et al.* 2004).

Our results demonstrate that additional information is needed to understand better the mechanisms of mtDNA transfer and to ascertain authentic mtDNA sequences, especially because it appears that taxon, sample-type, gene region, and numt age may all play a role in

the amplification, collection, and detection of numt sequences when using PCR. Additional studies are needed to determine if the same frequency of numts is present in other primates or if our results are specific only to gorillas. Though some researchers may be concerned that *COII* is not variable enough to ever be utilized as a population genetic marker, it may at least have methodological utility. Potentially, researchers can use the *COII* gene to develop methods that preferentially amplify authentic mtDNA in a more slowly evolving system in which codons can assist in detecting numts. If a particular method of PCR, cloning, and sequencing can demonstrate preferential amplification of mtDNA at *COII*, it could provide a valuable starting point for investigations of more variable regions, such as *HVI*. This would assist primatologists in continuing to utilize mtDNA as an effective marker in species including gorillas, in particular while utilizing noninvasive samples.

Acknowledgments

We thank J. Oates for his assistance and discussions regarding this project; R. Bergl, J. Satkoski, and the reviewers for providing numerous helpful comments; and J. Satkoski and J. Dupain and Projet Grands Singes for the fecal sample. M. E. Steiper acknowledges Research Centers in Minority Institutions award RR-03037 from the National Center for Research Resources of the National Institutes of Health, which supports the infrastructure of biological anthropology research at Hunter. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the NCRR/NIH.

References

- Anthony NM, Clifford SL, Bawe-Johnson M, Abernethy KA, Bruford MW, Wickings EJ. Distinguishing gorilla mitochondrial sequences from nuclear integrations and PCR recombinants: Guidelines for their diagnosis in complex sequence databases. *Molecular Phylogenetics and Evolution* 2007;43:553–566.10.1016/j.ympev.2006.09.013. [PubMed: 17084645]
- Bensasson D, Zhang D, Hartl DL, Hewitt GM. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology & Evolution* 2001;16:314–321.10.1016/S0169-5347(01)02151-6. [PubMed: 11369110]
- Clifford SL, Anthony NM, Bawe-Johnson M, Abernethy KA, Tutin CE, White LJ, et al. Mitochondrial DNA phylogeography of western lowland gorillas (*Gorilla gorilla gorilla*). *Molecular Ecology* 2004;13:1551–1565. 1567.10.1111/j.1365-294X.2004.02140.x. [PubMed: 15140097]
- Collura RV, Stewart CB. Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. *Nature* 1995;378:485–489.10.1038/378485a0. [PubMed: 7477403]
- Greenwood AD, Pääbo S. Nuclear insertion sequences of mitochondrial DNA predominate in hair but not in blood of elephants. *Molecular Ecology* 1999;8:133–137.10.1046/j.1365-294X.1999.00507.x. [PubMed: 9919702]
- Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* 1985;22:160–174.10.1007/BF02101694. [PubMed: 3934395]
- Hazkani-Covo E, Graur D. A comparative analysis of numt evolution in human and chimpanzee. *Molecular Biology and Evolution* 2007;24:13–18.10.1093/molbev/msl149. [PubMed: 17056643]
- Jensen-Seaman MI, Sarmiento EE, Deinard AS, Kidd KK. Nuclear integrations of mitochondrial DNA in gorillas. *American Journal of Primatology* 2004;63:139–147.10.1002/ajp.20047. [PubMed: 15258958]
- Mundy NI, Pissinatti A, Woodruff DS. Multiple nuclear insertions of mitochondrial cytochrome b sequences in callitrichine primates. *Molecular Biology and Evolution* 2000;17:1075–1080. [PubMed: 10889220]
- Ruvolo M, Disotell TR, Allard MW, Brown WM, Honeycutt RL. Resolution of the African hominoid trichotomy by use of a mitochondrial gene sequence. *Proceedings of the National Academy of Sciences of the United States of America* 1991;88:1570–1574.10.1073/pnas.88.4.1570. [PubMed: 1996358]

- Ruvolo M, Pan D, Zehr S, Goldberg T, Disotell TR, von Dornum M. Gene trees and hominoid phylogeny. *Proceedings of the National Academy of Sciences of the United States of America* 1994;91:8900–8904.10.1073/pnas.91.19.8900. [PubMed: 8090741]
- Ruvolo M, Zehr S, von Dornum M, Pan D, Chang B, Lin J. Mitochondrial COII sequences and modern human origins. *Molecular Biology and Evolution* 1993;10:1115–1135. [PubMed: 8277847]
- Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 1987;4:406–425. [PubMed: 3447015]
- Schmitz J, Piskurek O, Zischler H. Forty million years of independent evolution: a mitochondrial gene and its corresponding nuclear pseudogene in primates. *Journal of Molecular Evolution* 2005;61:1–11.10.1007/s00239-004-0293-3. [PubMed: 16007490]
- Swofford, DL. PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4. Sunderland, MA: Sinauer Associates; 2001.
- Thalmann O, Hebler J, Poinar HN, Pääbo S, Vigilant L. Unreliable mtDNA data due to nuclear insertions: A cautionary tale from analysis of humans and other great apes. *Molecular Ecology* 2004;13:321–335.10.1046/j.1365-294X.2003.02070.x. [PubMed: 14717890]
- Thalmann O, Serre D, Hofreiter M, Lukas D, Eriksson J, Vigilant L. Nuclear insertions help and hinder inference of the evolutionary history of gorilla mtDNA. *Molecular Ecology* 2005;14:179–188.10.1111/j.1365-294X.2004.02382.x. [PubMed: 15643961]
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 1997;25:4876–4882.10.1093/nar/25.24.4876. [PubMed: 9396791]
- Zischler H, Geisert H, Castresana J. A hominoid-specific nuclear insertion of the mitochondrial D-loop: Implications for reconstructing ancestral mitochondrial sequences. *Molecular Biology and Evolution* 1998;15:463–469. [PubMed: 9549097]
- Zischler H, Geisert H, von Haeseler A, Pääbo S. A nuclear ‘fossil’ of the mitochondrial D-loop and the origin of modern humans. *Nature* 1995;378:489–492.10.1038/378489a0. [PubMed: 7477404]

Fig. 1.

Neighbor joining tree based on a HKY distance matrix generated for all 165 clone sequences and available sequences. Note that the main gorilla group was moved to facilitate the legibility of the tree (the dotted line linking the asterisks shows where the gorilla group belongs in the original tree and is not a tree branch). Groups A, B, C, D, and E are described in the text. The dagger denotes a contaminant clone sequence described in the text.

Fig. 2. Maximum parsimony strict consensus tree of all gorilla clones and the GenBank *COII* sequences. The consensus is based on 196 trees of 927 steps. Clones that have a shared synapomorphy between 2 independent PCRs are indicated in a gray round-edged box, numbered to correspond with Table III. Note the vertical rearrangement of one of the numt groups in the tree to enhance legibility.

Table 1

Previously published sequences from GenBank

Taxon name	GenBank accession number
<i>Gorilla_Ggo1</i>	M58006, M58355
<i>Gorilla_Ggo2</i>	–
<i>Gorilla_Ggo3</i>	U12698
<i>Gorilla_Ggo4</i>	U12699
<i>Gorilla_Ggo5</i>	U12700
<i>Gorilla_Ggo6</i>	U12701
<i>P. troglodytes_Ptr1</i>	U12697
<i>P. troglodytes_Ptr2</i>	M58009, M58358
<i>P. troglodytes_Ptr3</i>	–
<i>P. troglodytes_Ptr4</i>	U12705
<i>P. troglodytes_Ptr5</i>	U12706
<i>P. paniscus_Ppa1</i>	U12695
<i>P. paniscus_Ppa2</i>	–
<i>P. paniscus_Ppa3</i>	U12696
<i>P. paniscus_Ppa4</i>	U12702
<i>Human_Hsa1</i>	–
<i>Human_Hsa2</i>	U12690, S67561
<i>Human_Hsa3</i>	U12691
<i>Human_Hsa4</i>	U12692
<i>Human_Hsa5</i>	U12693
<i>Human_Hsa6</i>	U12694
<i>P. pygmaeus_Ppy1</i>	–
<i>P. pygmaeus_Ppy2</i>	U12703
<i>P. pygmaeus_Ppy3</i>	U12704
<i>H. syndactylus</i>	M58007, M58356
<i>M. mulatta</i>	M74005

We aligned sequences not found in GenBank manually.

Table IIStop codons and *blast* results for all putative numt sequences found

Putative numt	No. of stop codons	Highest <i>blast</i> match (%)	Match sequence
GorillaPCR1_34	13	92	Human chromosome 17
GorillaPCR2_26	20	96	Human BAC clone 2
GorillaPCR2_31	20	96	Human BAC clone 2
GorillaPCR2_34	2	98	Human chromosome 5
GorillaPCR2_40	10	98	Human chromosome 6
GorillaPCR2_42	2	98	Human chromosome 5
GorFecPCR1_5	18	95	Human chromosome 17
GorFecPCR1_6	18	96	Human chromosome 17
GorFecPCR1_7	10	98	Human chromosome 6
GorFecPCR1_16	10	95	Human chromosome 6
GorFecPCR1_22	20	96	Human BAC clone 2
GorFecPCR1_31	20	96	Human BAC clone 2
GorFecPCR1_33	10	97	Human chromosome 6
GorFecPCR1_35	14	89	<i>Gorilla gorilla gorilla</i> and <i>Pongo pygmaeus abelii</i>
GorFecPCR1_43	20	96	Human BAC clone 2
GorFecPCR2_6	10	98	Human chromosome 6
GorFecPCR2_9	10	98	Human chromosome 6
GorFecPCR2_22	10	98	Human chromosome 6
GorFecPCR2_33	0	99	Human isolate mitochondrial genome
GorFecPCR2_39	10	98	Human chromosome 6
GorFecPCR2_40	10	98	Human chromosome 6
GorFecPCR2_45	10	98	Human chromosome 6
GorFecPCR2_46	10	98	Human chromosome 6
Orang_8	17	96	Human chromosome 17
Orang_9	7	96	Human chromosome 17

Table III

Young numt groups

Young numt group	Clones	Position	Substitution	Change
1	GorillaPCR2_45	278	A→G	Transition
	GorillafecPCR2_29			
	GorillafecPCR1_24			
2	GorillaPCR2_30	674	C→T	Transition
	GorillafecPCR1_10			
3	GorillaPCR2_19	83	C→T	Transition
	GorillafecPCR2_20			
4	GorillaPCR2_48	359	T→C	Transition
	GorillafecPCR1_4			
	GorillafecPCR2_47			
5	GorillaPCR2_7	185	A→G	Transition
	GorillafecPCR2_48			
	GorillafecPCR2_11			
6	GorillaPCR2_37	568	C→T	Transition
	GorillafecPCR1_28			
7	GorillaPCR2_35	337	A→G	Transition
	GorillafecPCR2_30			
	GorillafecPCR2_31	404	T→C	Transition
	GorillafecPCR1_25			
8	GorillaPCR2_38	423	C→T	Transition
	GorillafecPCR1_46			
	GorillafecPCR2_4			
9	GorillaPCR2_2	191	C→T	Transition
	GorillafecPCR2_5	195	T→C	Transition