Molecular Ecology (2004) **13**, 321–335 doi: 10.1046/j.1365-294X.2003.02070.x

Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes

O. THALMANN, J. HEBLER, H. N. POINAR,* S. PÄÄBO and L. VIGILANT
Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, Leipzig 04103, Germany

Abstract

Analysis of mitochondrial DNA sequence variation has been used extensively to study the evolutionary relationships of individuals and populations, both within and across species. So ubiquitous and easily acquired are mtDNA data that it has been suggested that such data could serve as a taxonomic 'barcode' for an objective species classification scheme. However, there are technical pitfalls associated with the acquisition of mtDNA data. One problem is the presence of translocated pieces of mtDNA in the nuclear genome of many taxa that may be mistaken for authentic organellar mtDNA. We assessed the extent to which such 'numt' sequences may pose an overlooked problem in analyses of mtDNA from humans and apes. Using long-range polymerase chain reaction (PCR), we generated necessarily authentic mtDNA sequences for comparison with sequences obtained using typical methods for a segment of the mtDNA control region in humans, chimpanzees, bonobos, gorillas and orangutans. Results revealed that gorillas are notable for having such a variety of numt sequences bearing high similarity to authentic mtDNA that any analysis of mtDNA using standard approaches is rendered impossible. Studies on humans, chimpanzees, bonobos or orangutans are apparently less problematic. One implication is that explicit measures need to be taken to authenticate mtDNA sequences in newly studied taxa or when any irregularities arise. Furthermore, some taxa may not be amenable to analysis of mtDNA variation at all.

Keywords: barcode, gorilla, long-range PCR, numt, phylogeography, pseudogene

Received 25 July 2003; revision received 22 October 2003; accepted 22 October 2003

Introduction

Analysis of mitochondrial DNA (mtDNA) sequence variation has been used for more than a decade as a tool to understand the phylogeny of species as well as the geographical distribution of genetic variation and demographic history of populations (Brown & Wright 1979; Gemmell *et al*. 1996; Stanley *et al*. 1996; Talbot & Shields 1996; Vilà *et al*. 1997; Avise 2000). Some of the most intensively studied species are humans and our closest relatives, the great apes. Insights gained include the recent African origin of human mtDNA (Vigilant *et al*. 1991; Stoneking & Soodyall 1996; Ingman *et al*. 2000), the inference of a bottleneck event in the past of modern humans (Rogers & Harpending 1992) and the observation that humans possess notably less variation than do other great apes, a finding corroborated by analysis of nonmitochondrial DNA regions (Jensen-Seaman *et al*. 2001; Kaessmann *et al*. 2001; Stone *et al*. 2002).

Recently, the assumptions governing analysis of mtDNA sequence data, along with the accuracy of such data, have been much discussed. One advantage to using mtDNA for analysis is that it has been believed to represent a single, nonrecombining locus, so that genealogical trees depicting relationships among variants may be constructed and traced back to an ancestral type. Evidence suggesting that human mtDNA actually does recombine has been presented (Awadalla *et al*. 1999; Eyre-Walker *et al*. 1999a; Hagelberg *et al*. 1999). However, these reports were plagued with errors in the data (Eyre-Walker *et al*. 1999b; Hagelberg *et al*. 2000; Kivisild *et al*. 2000) and the methods of analysis have also been questioned (Macaulay *et al*. 1999), so that the occurrence of mtDNA recombination in humans remains

Correspondence: O. Thalmann. Fax: + 49 341 3550 299; E-mail: thalmann@eva.mpg.de
*Current address, McMaster University, Departments of Anthropoplogy and Geology, 1280 Main Street West, Hamilton L8S 4L9, Canada.

highly unlikely, although debate continues (Maynard Smith & Smith 2002; Hagelberg 2003).

The uniparental inheritance of animal mtDNA means that only one version of the mtDNA molecule should be present in each individual. The presence of multiple mtDNA versions, termed heteroplasmy, can plausibly arise as a result of mutations in somatic or oocyte cells (Petri *et al*. 1996) or paternal inheritance of mtDNA. Paternal inheritance of mtDNA in animals has been occasionally detected, as in some cases of crosses between mouse species (Gyllensten *et al*. 1985; Shitara *et al*. 1998) or inbred strains (Gyllensten *et al*. 1991), one case in hybrids between two bird subspecies (Kvist *et al*. 2003), one person with a severe disease phenotype (Schwartz & Vissing 2002) and in an animal which exhibits doubly uniparental inheritance (mussels: Zouros *et al*. 1994). In sum, there is little direct evidence that mtDNA recombination, heteroplasmy or paternal inheritance have had a major impact upon evolutionary analyses of animal mtDNA.

In contrast to the phenomena mentioned above, one source of significant, genuine problems in the analysis of mtDNA arises out of the presence of translocated copies of mtDNA in the nuclear genome (Bensasson *et al*. 2001). After transposition into the nucleus, these nuclear insertions or 'numts', evolve independently as paralogous copies of the original mtDNA segment (Smith *et al*. 1992; Arctander 1995) and sometimes represent useful molecular 'fossils' for phylogenetic comparisons (Zischler *et al*. 1995) or substitution rate computations (Lopez *et al*. 1997). Less usefully, during analysis numts may be amplified inadvertently by polymerase chain reaction (PCR) in addition to or even instead of the authentic target mtDNA and thus lead to incorrect results. The slower rate of sequence evolution in the nucleus as compared to the mitochondrion means that numts may be amplified in preference to authentic mtDNA when using 'universal' or 'conserved' primers designed using cross-species sequence comparisons (Sorenson & Fleischer 1996; Zhang & Hewitt 1996; Mirol *et al*. 2000). Hence, the key to accurate analysis of mtDNA sequences is to determine whether numts may have been inadvertently included.

Various approaches have been suggested for detecting and avoiding numts (Bensasson *et al*. 2001). While examining the individual sequences of multiple clones of PCR products reveals whether multiple sequences have been obtained in a PCR, this does not in itself distinguish authentic mtDNA from numts, and is of no use in cases where only the numt sequence has been amplified (e.g. Williams & Knowlton 2001). When coding segments are the targets of interest, it is possible to examine the sequences for frameshift or stop mutations (Collura & Stewart 1995) or, in the case of ribosomal RNAs, unexpected secondary structures (Olson & Yoder 2002). However, this approach cannot be applied to analysis of sequences of the control region, a noncoding segment useful for intraspecific studies, because of its particularly high mutation rate, nor can it be used to identify recently transposed numts with any efficiency.

Despite the considerable amount of work that has been conducted on mtDNA sequence analysis of humans and other great apes, no systematic assessment of potential inaccuracies due to numts has been carried out. This is particularly of concern in view of the heavy reliance upon noninvasive samples such as hair in studies of great apes (Morin *et al*. 1994; Goldberg & Ruvolo 1997; Saltonstall *et al*. 1998; Jensen-Seaman & Kidd 2001; Warren *et al*. 2001), as well as of humans (Vigilant *et al*. 1989; Vigilant *et al*. 1991; Allen *et al*. 1998), and indications that such a sample type can be more likely to produce numts (Greenwood & Pääbo 1999). The prevalence of numts in the draft human genome sequence has been recently assessed (Mourier *et al*. 2001; Tourmen *et al*. 2002; Woischnik & Moraes 2002). Numts originating from all parts of the mitochondrial genome have been found, with the largest reported segment encompassing some 88% of the mitochondrial genome (Tourmen *et al*. 2002), although the majority of the fragments are smaller than 5% of the length of the 16 569 base pairs (bp) mitochondrial genome (Woischnik & Moraes 2002). The resemblance of the numts to the mtDNA sequence varies, but segments of extremely high similarity are present (Woischnik & Moraes 2002).

Detailed information on the prevalence of numts in nonhuman great apes is almost entirely lacking, but it has been suggested that numts of the control region are particularly common in gorillas (Jensen-Seaman 2000). Factors including the presence of deep divergences in intraspecific phylogenies (Gagneux *et al*. 1999), the use of conserved primers designed from comparison of multiple taxa and the routine use of DNA obtained from noninvasive samples mean that the possibility that unrecognized great ape numt sequences contaminate analyses deserves scrutiny.

In this study, we used two different experimental approaches on samples from the same individuals in order to produce DNA sequences. The first approach, the logic of which is outlined in Fig. 1, used long-range PCR on an extract enriched for mtDNA and necessarily produced only authentic mtDNA sequences of the subsequently amplified first hypervariable segment (HV1). The second approach used conventional PCR from total genomic DNA followed by examination of multiple clones of the products obtained. Since the long-range approach produced authentic mtDNA sequences, comparison with clones obtained using the second, conventional approach revealed whether sequences derived from numts were amplified in addition to or in place of authentic mtDNA. Results indicate that current practices do not preclude inadvertent analysis of numts and imply that caution should be exercised in implementing the recent calls for the establishment of a DNA 'barcode' for

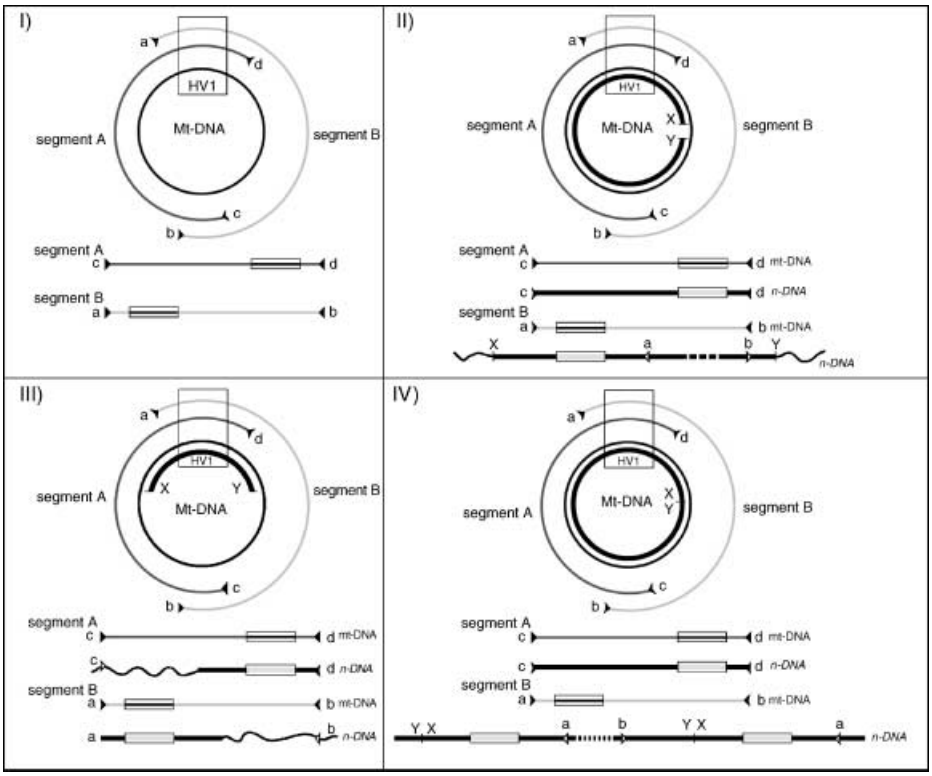

Fig. 1 All possible outcomes expected from the amplification of the HV1 using two long-range segments (segment A, segment B) are shown. The circle is the mtDNA molecule and the bold black circle fragments in II, III and IV are numts. X and Y label end-points of the numt while a, b, c and d represent long-range primers. For ease of comparison possible long-range products are depicted in linearized form in the lower part of each diagram, with the rectangle indicating the HV1 and wavy lines indicating nuclear DNA. Different scenarios: (I) Nuclear homologous sequences do not exist and therefore two identical, authentic mtDNA sequences would be amplified. (II) A transferred segment exists in the nucleus almost of the same length as the entire mitochondrial genome. Amplifications of one long-range target (segment A) could result in several different putative HV1 sequences, whereas amplifications from the other long-range segment would result exclusively in one authentic mtDNA sequence. (III) A fragment spanning the HV1 was transferred to the nucleus. Because one primer for each targeted long-range segment does not bind to the nuclear counterpart, only amplifications from the original mtDNA would succeed. (IV) The mtDNA molecule was completely transferred and duplicated in tandem after its arrival in the nuclear genome. Amplifications from both long-range segments would yield in several heterogeneous sequences, but since the inserted segments evolve independently only the original mtDNA should be reproducible from both segments.

taxonomy consisting of mtDNA sequences (Hebert *et al*. 2003; Mallet & Willmott 2003).

Materials and methods

Specimens and DNA extractions

We obtained frozen liver samples from two gorillas, two chimpanzees, two orangutans and one human. Gorilla samples included one western lowland gorilla, *Gorilla gorilla gorilla* (Rok, G0019) from the Yerkes National Primate Research Center and one eastern lowland gorilla, *G. beringei graueri* (Mukisi) from the Center for Research and Conservation of the Royal Zoological Society of Antwerp. Both chimpanzees were from Yerkes (Jenny, C0090 and Cookie, C0368) and are probably representatives of the West African subspecies *Pan troglodytes verus*. The two orangutan samples were one *Pongo pygmaeus pygmaeus* (Loklok, YN97-215, Yerkes) and one *P. p. abelii* (YN95-233, Yerkes) from Borneo and Sumatra, respectively. Finally, we acquired one human liver sample (internal ID 68788) and frozen leucocytes from one bonobo (*Pan paniscus*, Joey, internal ID: 68621, Leipzig Zoo).

We extracted genomic DNA from the samples in two ways. One procedure followed a standard DNA extraction method (Sambrook *et al*. 1989). In brief: we homogenized approximately 100 mg of frozen liver and digested the homogenate with proteinase K (10 mg/μL) and 10% sodium dodecyl sulphate (SDS) at 50 °C overnight, followed by a phenol/chloroform extraction and alcohol precipitation. A RNAse (10 mg/mL) treatment of the sample was carried out followed by an additional chloroform step. We diluted the final DNA pellet in 50 μL 1× TE buffer. A second extraction method used a commercial kit (mtDNA Extractor CT Kit,

Wako Pure Chemical Industries) according to the manual to produce from liver cells a DNA extract enriched for mtDNA.

Isolated bonobo leucocytes were kept in liquid nitrogen and extracted using the following procedure. We resuspended the frozen leucocyte pellet in TEN buffer (10 mM Tris/HCl, pH 8.2; 400 mM NaCl; 2 mM EDTA, pH 8.2) and digested the mixture at 37 °C overnight with 10% SDS and proteinase K (10 mg/mL). After adding saturated NaCl to the sample, the DNA was precipitated with ethanol. Due to the small amount of leucocytes available, a mitochondrial DNA enriching extraction procedure was not possible for the bonobo sample. We estimated the quantity of all extracted DNAs from the 260/280 absorbance using an UV spectrometer (GeneQuant pro, Amersham Bioscience).

PCR amplification, cloning and sequencing

We performed two separate long-range PCR amplifications on each DNA extract enriched for mtDNA in order to amplify two overlapping segments that together cover the entire mitochondrial DNA molecule (Fig. 1). Both segments include the control region of the molecule in the overlapping region. The approximately 9 kb segment A was amplified using Cytbf (5′-CACGAAACAGGATCAAATAACCC-3′) (termed ECB1 in (Muir *et al*. 2000)) and COIIrev592 (5′-TGGTTGGCTCCACAGATTTC-3′). Segment B was produced using 12So (5′-GTCGATTATAGGACAGGTTCCTCTA-3′) (Poinar *et al*. 2001) and COII28for (5′-AAGACGCTACTT-CTCCTATCATAGA-3′) and was about 10 kb in length. PCRs for both long-range segments were performed in 50 μL reactions using 1× High Fidelity PCR buffer; 2 mM $MgSO_4$, 0.2 μM of each primer, 0.2 mM each dNTP, 2.5 U Platinum® *Taq* DNA polymerase High Fidelity (Invitrogen) and approximately 150 ng DNA of the mtDNA-enriched extract. Reaction conditions were: denaturation at 94 °C for 30 s followed by 35 cycles of 30 s at 94 °C, 30 s at 60 °C and 11 min at 68 °C with a final cooling step at 4 °C in a Peltier thermal cycler, PTC 200 (MJ Research). Twenty μL of the PCR products were electrophoresed through 0.8% agarose gels (Seakem) containing ethidium bromide, visualized under UV light and bands representing the expected product length were excised. Gel slices were melted in 100 μL ddH_2O for 1 h at 80 °C, the mixture frozen at –20 °C and following centrifugation the supernatant containing DNA was transferred to another tube and used as a template for subsequent PCRs.

PCRs amplifying the HV1 of the mitochondrial control region were performed on two types of DNA templates, namely, the two amplification products resulting from the long-range PCRs as well as the total genomic extracts. A universal primer pair consisting of L15926, designed to be conserved in vertebrates (Kocher *et al*. 1989), and H16498 (Kocher & Wilson 1991) was used in all taxa. In addition, we used primer pairs applied previously in studies conducted on single ape taxa, which we refer to here as 'species-typical' primers. Table 1 lists the characteristics of the different primer pairs. PCRs were set up in 20 μL volumes consisting of 1× AmpliTaq Gold® PCR buffer, 2 mM $MgCl_2$, 0.2 μM of each primer, 0.1 mM each dNTP, 1 U AmpliTaq Gold® (Perkin Elmer) and 0.4 μg/μL BSA. A 3 μL volume of the purified long-range products or approximately 100 ng of total genomic DNA were used as template for the amplification. Except for the annealing temperatures, the PCR cycling conditions were similar for all primer sets: 3 min at 95 °C, then 35–40 cycles of 30 s at 95 °C, 30 s at X °C (see Table 1), 30 s at 72 °C followed by 72 °C for 30 min and 4 °C storage.

PCR products amplified with the different HV1 primer sets were cloned using the TOPO TA cloning kit (Invitrogen). We reamplified a minimum of 16 colonies from each PCR product in a colony PCR (Kilger & Pääbo 1997) using M13for and M13rev primers and analysed the products on an automated ABI 3700 capillary sequencer (Applied Biosystems).

Data analyses

Cloned sequences were aligned by eye using the sequence alignment editor BIOEDIT version 5.0.9 (Hall 1999). We

Table 1 Characteristics of the different primer sets used for the amplification of the HV1

Taxa	Primer name	Sequence	Annealing temp. (°C)	Ref.
Universal primer pair	L15926 H16498	5′-TCAAAGCTTACACCAGTCTTGTAAACC-3′ 5′-CCTGAAGTAGGAACCAGATG-3′	56	Kocher *et al*. (1989) Kocher & Wilson (1991)
Gorilla	L91-115 H402-427	5′-ACTAAGCTTTGTTCTTTCATGGGGAGAC-3′ 5′-ATTGAATTCACGGAGGATGGTGTTC-3′	53	Garner & Ryder (1996)
Pongo	D1 D5 D2 D4	5′-CAACATGAATATCACCC-3′ 5′-TGTGCGGGATATTGATTTCAC-3′ 5′-ACACAACAATCGCTTAAC-3′ 5′-GATGGTGAGYAAGGGATT-3′	Touch down procedure Max. 65 Min. 50	Warren *et al*. (2001)
Pan/Homo	L15996 H16498	5′-CTCCACCATTAGCACCCAAAGC-3′ 5′-CCTGAAGTAGGAACCAGATG-3′	50	Vigilant *et al*. (1989) Kocher & Wilson (1991)

confirmed polymorphic sites by examining the position in the electropherograms. The HV1 of gorillas contains an approximately 20 bp-long, highly variable polycytosine segment (Foran *et al*. 1988), which was excluded from analysis as alignment was not possible.

We conducted a χ^2 goodness-of-fit test to estimate if the number of observed singletons in the HV1 amplifications from the long-range products deviated significantly from the expected number of singletons incorporated due to polymerase errors alone. The *Taq* polymerase error rate is assumed to be 7.3×10^{-5} per bp per duplication (Kobayashi *et al*. 1999). We performed the same statistical test to examine whether or not two sequences amplified from the total genomic extract were significantly different from each other while taking into account the potential number of nucleotide changes incorporated due to *Taq* error. The statistical tests were done using StatXact-5 (Cytel Software). All unique sequences were compared to sequences in GenBank using BLAST search (Altschul *et al*. 1997).

Phylogenetic analyses were performed using maximum likelihood (ML), maximum parsimony (MP) and neighbour-joining methods [applying the Kimura 2-parameter substitution model (Kimura 1980)] using the computer programs TREEPUZZLE (Strimmer & von Haeseler 1996) and MEGA2 (Kumar *et al*. 2001), respectively. We ran 1000 puzzling and 1000 bootstrap steps, respectively, to assess the support for each branch. In addition to the sequences observed in this study, several sequences retrieved from the database were included as indicated in the phylogenetic analyses.

Results

Long-range mtDNA amplifications

The mtDNA is a circular molecule, whereas mitochondrial segments integrated in the nucleus are linear. Thus, if two overlapping long-range amplifications of mtDNA yield the same sequence of the HV1 of the control region in their overlaps, they are both highly likely to represent authentic circular mtDNA (Fig. 1). Our first step was to determine, for each of eight individuals representing the great apes, the HV1 sequence from each of the long-range segments A and B amplified from DNA enriched for mtDNA. We found that more than half the sequenced clones of each HV1 product from a single long-range segment were identical, whereas the remaining clones were not perfectly identical, but contained single-base changes (singletons) present in single clones. We tested whether the observed singletons could be attributed to polymerase error. Using an error rate for *Taq* polymerase of 7.3×10^{-5} per bp per duplication (Kobayashi *et al*. 1999) and assuming that every molecule replicates in each cycle, the expected number of errors over 12 000 bases amplified in 35 PCR cycles is 31. This is not significantly different from the 30 singletons we observed in a subset of 50 clones totalling 12 000 bp generated in 35 cycles from one representative individual (χ^2 goodness-of-fit: 0.0323; $P > 0.05$, 1 d.f.). Singletons were therefore regarded as being the result of polymerase error alone and were disregarded in making consensus sequences.

The sequences of the HV1 in the long-range segments A and B were amplified by PCR using both the universal primers and primers used frequently by other investigators in these species (Table 1). When the universal primers were applied to the long-range segment A, for all individuals a single consensus sequence of HV1 was obtained. The amplifications of HV1 from the same long-range segment using the species-typical primers also resulted in only one consensus sequence per individual. Similarly, for the second long-range segment (B), all sequenced clones produced one identical consensus sequence using either universal or species-typical primers for all individuals. Comparisons of the consensus sequences obtained from the long-range segments A and B from each individual revealed exactly the same sequence, indicating the successful amplification of the authentic mitochondrial HV1 sequences from each long-range segment. Consensus HV1 sequences for each individual are in supplementary Fig. 1, along with some variant sequences which differ by a very few changes that we attribute to errors in an early cycle of the PCR. The closest match for each HV1 consensus sequence found in a BLAST search was either another individual of the same species or, in the case of the gorillas, an identical sequence derived by other researchers using material from the same individuals [Rok, Accession no. X93348 (Xu & Arnason 1996) and Mukisi, Accession no. L76770 (Garner & Ryder 1996)]. The GenBank Accession nos for the other sequences are: chimpanzee C0368: AJ586557, chimpanzee C0090: AJ586556, orangutan YN97-215: AJ586559, orangutan YN95-233: AJ586558, human 68788: AJ586554 and bonobo 68621: AJ586555.

Amplifications from total genomic DNA

The same sets of primers used to amplify HV1 sequences from the long-range products were also used for amplifications directly from total genomic DNAs that were not enriched for mtDNA. This was performed in order to determine whether a single sequence identical to the mtDNA sequence obtained from the long-range approach was amplified, or whether additional sequences derived presumably from numts were also or instead amplified. Furthermore, it is of interest to determine whether non-organellar sequences are obtained more readily from universal primers designed to work in many taxa compared to primers designed to be used in particular species. As mentioned above, since multiple clones of the same sequence differ in singleton changes due to errors by the *Taq* polymerase, we determined the minimum number of differences

Table 2 List of sequences amplified using total genomic DNAs

	Individual	Sequence name	GenBank Accession no.	MtDNA	Amplification primers		No. of clones	Occurrence in two independent individual amplifications
					universal	typical		
Gorillas	Mukisi	Mukmt	L76770	X	X		17	X
		Muk1				X	4	X
		Muk2				X	1	
		Muk3				X	1	
		Muk4			X	X	23	X
		Muk5			X	X	20	X
		Muk6				X	1	
		Muk7				X	5	X
		Muk8				X	1	
		Muk9				X	1	
	Rok	Rokmt	X93348	X	X	X	14	X
		Rok1				X	2	
		Rok2				X	1	
		Rok3				X	1	
		Rok4			X	X	2	X
		Rok5			X	X	11	X
		Rok6				X	1	
		Rok7				X	2	
		Rok8			X	X	45	X
		Rok9			X		1	
		Rok10			X		1	
		Rok11			X		3	X
Chimpanzees	C0090	Chimp0090mt	AJ586556	X	X	X	56	X
		Chimp0090/1			X		7	X*
		Chimp0090/2			X		1	
	C0368	Chimp0368mt	AJ586557	X	X	X	41	X
		Chimp0368/1			X		10	X*
		Chimp0368/2			X		2	
		Chimp0368/3			X		1	

*Although these sequences did not occur in two independent PCRs of each single individual they have only a single substitutional difference and were therefore considered as separate occurrences of the same sequence.

necessary for sequences to be considered nonidentical. Assuming the aforementioned *Taq* error rate, we considered HV1 sequences unique only if they had more than five substitutions along the 478 bp (χ^2 goodness-of-fit test: 12.06; $P < 0.01$, 1 d.f.) amplified using the universal primers. Because sequences amplified with the *Pan*/*Homo* typical primers yielded a 474 bp segment, we also required a minimum of five substitutions to distinguish two sequences in these taxa. Due to a shorter amplification product (240 bp), a minimum of three substitutions was required to distinguish sequences obtained using primers for use in gorillas as nonidentical (χ^2 goodness-of-fit test: 9.624; $P < 0.01$, 1 d.f.).

When we performed two independent amplifications of the total genomic extract using the universal primer pair, only the authentic mitochondrial HV1 sequence was obtained from the bonobo, human and two orangutan samples, while multiple sequences were obtained from the chimpanzees and gorillas, as summarized in Table 2. We detected three sequences for one chimpanzee (C0090), including the authentic mtDNA sequence (Chimp0090mt) and two additional nonorganellar sequences (Chimp0090/1, C0090/2). For the other chimpanzee (C0368) we found four sequences, one of which was the authentic mtDNA HV1 sequence (Chimp0368mt). Each chimpanzee yielded a nonorganellar sequence (Chimp0090/1, Chimp0368/1) that differs by only one substitution, probably as a result of polymerase error. These sequences differ substantially (54 or 55 substitutions) from the corresponding mtDNA sequences (Table 3). For one gorilla sample (Rok) we observed the authentic mitochondrial sequence (Rokmt) and six additional nonorganellar sequences (Table 2). We found one

Table 3 Pairwise sequence differences between HV1 sequences obtained from total genomic DNAs estimated using the Kimura-2-parameter substitution model (above diagonal) and the total number of nucleotide differences (below diagonal). The sequence length were 238 bp and 470 bp for gorillas (Muk, Rok) and chimpanzees, respectively. Indels were excluded

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1	Mukmt		0.05	0.02	0.03	0.14	0.17	0.12	0.14	0.15	0.17	0.21	0.19	0.2	0.06	0.2	0.13	0.14	0.15	0.16	0.17	0.16	0.17							
2	Muk1	11		0.06	0.07	0.15	0.17	0.14	0.16	0.17	0.17	0.23	0.2	0.21	0.01	0.2	0.15	0.16	0.15	0.17	0.18	0.17	0.17							
3	Muk2	5	14		0.03	0.12	0.16	0.15	0.13	0.15	0.15	0.21	0.18	0.19	0.06	0.19	0.13	0.14	0.15	0.16	0.16	0.16	0.15							
4	Muk3	7	16	6		0.12	0.13	0.13	0.11	0.11	0.14	0.18	0.16	0.17	0.06	0.16	0.12	0.13	0.12	0.12	0.14	0.13	0.15							
5	Muk4	30	32	27	27		0.04	0.02	0.01	0.04	0.03	0.18	0.14	0.15	0.14	0.13	0.004	0.03	0.03	0.05	0.1	0.05	0.11							
6	Muk5	35	36	34	28	10		0.05	0.03	0.01	0.01	0.19	0.15	0.16	0.16	0.09	0.04	0.07	0.02	0.004	0.1	0.02	0.13							
7	Muk6	26	30	31	29	4	12		0.02	0.05	0.05	0.19	0.15	0.16	0.14	0.14	0.01	0.04	0.03	0.06	0.11	0.06	0.12							
8	Muk7	30	33	29	25	3	7	5		0.03	0.04	0.16	0.12	0.14	0.15	0.11	0.01	0.03	0.03	0.03	0.09	0.03	0.12							
9	Muk8	32	35	31	25	9	3	11	6		0.03	0.18	0.14	0.16	0.16	0.1	0.03	0.06	0.03	0.02	0.1	0.03	0.14							
10	Muk9	35	35	32	30	7	3	11	10	6		0.21	0.17	0.18	0.16	0.1	0.03	0.06	0.01	0.02	0.12	0.03	0.13							
11	Rokmt	43	45	42	38	36	38	38	33	37	41		0.03	0.02	0.21	0.1	0.17	0.19	0.2	0.19	0.09	0.18	0.21							
12	Rok1	39	41	38	34	29	31	31	26	30	34	7		0.01	0.19	0.13	0.13	0.17	0.16	0.15	0.06	0.14	0.19							
13	Rok2	41	43	40	36	32	34	34	29	33	37	4	3		0.2	0.12	0.15	0.18	0.18	0.17	0.07	0.16	0.2							
14	Rok3	13	2	14	14	30	34	30	31	33	33	43	39		41	0.19	0.14	0.15	0.14	0.16	0.17	0.16	0.15							
15	Rok4	40	41	39	33	27	19	29	24	22	22	21	28		25	39	0.12	0.14	0.1	0.08	0.19	0.1	0.21							
16	Rok5	29	31	28	26	1	9	3	2	8	8	35	28	31	29		26	0.03	0.02	0.04	0.1	0.04	0.11							
17	Rok6	31	33	30	28	7	15	9	8	14	14	38	34	36	31	29		6	0.05	0.07	0.13	0.07	0.09							
18	Rok7	32	32	31	27	6	4	8	7	7	3	40	33	36	30	21		5	11	0.02	0.12	0.04	0.11							
19	Rok8	34	35	33	27	11	1	13	8	4	4	39	32	35	33	18	10		16	5	0.1	0.02	0.14							
20	Rok9	35	37	34	30	22	22	24	19	21	25	20	13	16	35	39	21		27	26	21	0.09	0.14							
21	Rok10	34	35	33	29	11	5	13	8	8	8	37	30	33	33	22	10	16		9	4	19	0.13							
22	Rok11	36	35	33	33	24	28	26	25	29	27	42	38	40	33	42	23	21	24	29	29	27								
23	Chimp0090mt																								0.13	0.07	0.09	0.13	0.11	0.1
24	Chimp0090/1																							54		0.05	0.12	0.002	0.05	0.08
25	Chimp0090/2																							30	24		0.1	0.06	0.04	0.06
26	Chimp0368mt																							40	53	45		0.13	0.07	0.04
27	Chimp0368/1																							55	1	25	54		0.05	0.09
28	Chimp0368/2																							48	22	20	33	21		0.04
29	Chimp0368/3																							43	37	29	16	38	19	

sequence (Rok11) that had a 105 bp insertion relative to the authentic Rok HV1 sequence, located at the position where the gorilla HV1 sequence has a polycytosine stretch and a deletion of approximately 73 bp compared to the human, bonobo and chimpanzee sequence (Foran *et al*. 1988). The amplifications from the other gorilla (Mukisi) resulted in three sequences including the Mukisi HV1 sequence (Mukmt) and two additional nonorganellar sequences (Muk4, Muk5). The additional HV1 gorilla sequences differ by 21–42 substitutions from the authentic mtDNA sequences of those individuals (Table 3).

For all individuals except one chimpanzee (C0090) and the two gorillas, we did not find additional nonorganellar DNA sequences among the clones sequenced from amplification products obtained using species-typical primers. The three additional nonorganellar sequences derived from chimpanzee C0090 were completely unalignable to the authentic mtDNA sequence and a BLAST search of those sequences reveal close similarity (> 97%) to sequences of several different chromosomal sequences of the human genome. These sequences were not used in further analyses. For the gorillas, the final length of the HV1 segment obtained using the primer pair described in Garner & Ryder (1996) was only 240 bp after the removal of primer sequences and the unreadable polycytosine stretch. We observed nine different sequences from each gorilla (Table 2). The sequences from the western gorilla Rok included the authentic HV1 sequence (Rokmt) and eight additional nonorganellar sequences. Of these eight nonorganellar DNA sequences, three did not show any or only a single substitutional difference to some already generated from Rok using the universal primer pair and thus each unique sequence was only considered once in the subsequent analyses (Rok4, Rok5 and Rok8). Interestingly, for the gorilla Mukisi the authentic mitochondrial HV1 sequence was not observed among the sequences obtained from two independent amplifications using the species-typical primers on genomic DNA. Two sequences did not differ from sequences amplified from this individual with the universal primers, again indicating the occasional amplification of the same nonorganellar DNA sequence (Muk4, Muk5).

All sequences obtained from amplification of genomic DNAs are found in supplementary Fig. 1. The pairwise differences among the authentic mtDNA sequences and additional nonorganellar DNA sequences are shown in Table 3. We noted that some sequences obtained from the gorillas and the chimpanzees resemble combinations of other sequences detected in the same individual. Such sequences are probably attributable to *in vitro* recombination occurring between different templates during the PCR, also termed jumping PCR (Saiki *et al*. 1988; Pääbo *et al*. 1990). Such switching between templates has been reported to occur when the DNA template is damaged (Pääbo *et al*. 1990; Hofreiter *et al*. 2001), but it also occurs even when high-quality DNA is used (e.g. Zylstra *et al*. 1998; Thompson *et al*. 2002), particularly when a high number of PCR cycles depletes the available pool of oligonucleotide primers (Judo *et al*. 1998). If we follow the criterion of considering sequences valid (nonrecombinant) only when they have been obtained from two independent PCRs, four of the five nonorganellar DNA sequences from the two chimpanzees and 12 of the 20 nonorganellar DNA sequences from the two gorillas are not reproduced and represent possible recombinants. This leaves one sequence found in both chimpanzees and a total of eight sequences from the two gorillas that represent nonrecombinant, nonorganellar DNA sequences, and hence are considered numts.

We were interested in determining whether numts were detected in some species and not others due to differences in the extent of primer-template mismatches, as such mismatches reduce the specificity of the PCR and hence increase the chances of obtaining alternate products (Smith *et al*. 1992; Collura & Stewart 1995). The single replicated chimpanzee numt (Chimp0090/1, Chimp0368/1) was detected following amplification with universal primers, while the eight gorilla numts were found using universal, species-typical primers, or both (Table 2). We compared the sequences of the primers used with the mtDNA sequence of reference individuals of these species, and found no obvious differences in the number of mismatches between the universal primers and reference sequences for each species (Fig. 2).

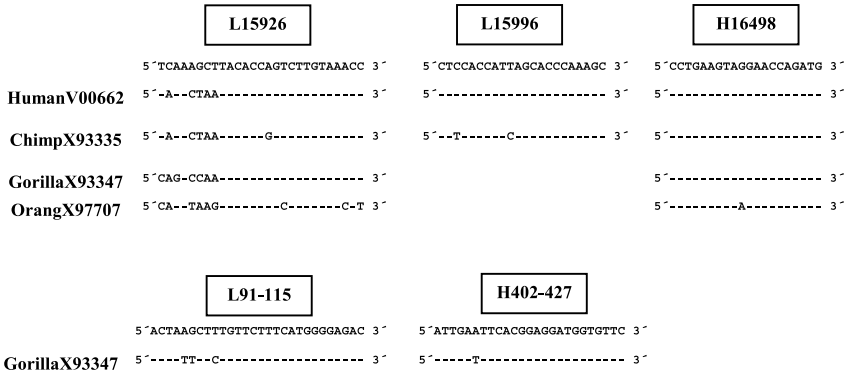

Fig. 2 Overview of the primer–template mismatches of primer pairs which amplified several nonorganellar DNA sequences in this study. As reference sequences we used sequences obtained from analysis of the entire mtDNA molecule. The sequences are indicated by their database accession numbers.

Phylogenetic analyses

Two topics were addressed using phylogenetic analysis: the relationship of all sequences obtained from all individuals and, more specifically, the relationship of the diverse sequences obtained from gorillas. We included all sequence variants obtained, i.e. both numts and recombinants, as replication of sequences cannot be assumed for comparative sequences from GenBank. However, the implications of excluding likely recombinants are also indicated. First, we reconstructed a tree of all great ape sequences found in this study and additional sequences from GenBank, including a macaque as an outgroup (Fig. 3). The tree topology of the authentic mtDNA sequences follows the branching patterns expected for ape evolution: first the separation of the orangutan, followed by the gorilla, the human and finally the bonobo and the chimpanzees. Chimpanzee sequences published previously elsewhere and available in databases were included in this analysis to evaluate both the arrangement of the authentic mtDNA sequences relative to sequences reported from various chimpanzee subspecies and the position of the nonorganellar DNA sequences. The mtDNA sequences from the chimpanzees C0368 and C0090 cluster together with GenBank mtDNA sequences from individuals of the subspecies *Pan troglodytes verus* and *P. t. troglodytes*, respectively. One nonorganellar DNA sequence found in the chimpanzee C0368 clusters together with the mtDNA sequence of this individual. Four additional nonorganellar sequences found in these chimpanzees, including the numt, fall in a basal position at a trifurcation also including a clade of west African chimpanzees (*P. t. verus*, *P. t. vellerosus*) sequences and a clade of sequences derived from central and east African chimpanzees (*P. t. troglodytes*, *P. t. schweinfurthii*). Similarly, the nonorganellar DNA sequences obtained from the gorillas Rok and Mukisi occupy various positions on the phylogenetic tree relating them to the authentic mtDNA sequences from these individuals. The relationship of the gorilla sequences was evaluated further in a second phylogenetic analysis.

All sequences obtained in this study from gorillas, along with gorilla HV1 sequences deposited in GenBank and one human sequence (Anderson *et al*. 1981), were used to produce a phylogenetic tree (Fig. 4). Two major clades are apparent within gorillas. Clade 1 contains the authentic mtDNA sequence from the eastern lowland gorilla Mukisi and gorilla sequences from the database that were generated from gorillas of East African origin. Within this clade sequences from eastern mountain gorillas cluster together (clade 1A), separately from the sequences of eastern lowland gorillas (clade 1B). Also present in clade 1 are four nonorganellar DNA sequences (Muk1, Muk2, Muk3 and Rok3). Clade 2 is comprised of sequences derived from gorillas of West African origin including the authentic mtDNA sequence of the gorilla Rok. A potential subdivision of clade 2 into separate clades would be possible but is not supported by the bootstrap values (Fig. 4), which are generally low at the deeper nodes of the tree. The majority of the nonorganellar DNA sequences identified from Rok and Mukisi as well as the putative numts identified and submitted directly to GenBank by Jensen-Seaman (2000) occur in this clade. Two of those nonorganellar DNA sequences cluster together with an authentic mtDNA sequence (Rok1, Rok2 and Rokmt), while other nonorganellar DNA sequences amplified from Rok intermingle with nonauthentic mtDNA sequences from Mukisi in clades 1 and 2. Overall, the nonorganellar DNA sequences found in this study do not fall within a single clade, and a separation of nonorganellar DNA sequences derived from the two species of gorillas from West and East Africa is not possible. The distribution in the tree of nonorganellar DNA sequences derived once (12 sequences) and those detected in two independent PCRs and hence attributed as numts (eight sequences) is similar (Fig. 4).

Discussion

Sequences of the mtDNA HV1 from apes

In this study we examined whether conventional methods produce accurate sequences of the mtDNA HV1 from representatives of all the great apes, including humans. While only authentic mtDNA sequences were obtained from humans, bonobos and orangutans, additional nonorganellar DNA sequences were found in the chimpanzees and gorillas. These additional sequences cannot be attributed to mitochondrial heteroplasmy as the long-range PCR approach identified only one mtDNA sequence for each of these individuals. Instead, these additional sequences appear to represent nuclear insertions of mtDNA (numts) as well as products of *in vitro* recombination of authentic and numt sequences during the PCR.

In retrospect, although it has not to our knowledge been previously raised as a potential complication of mtDNA analysis (Bensasson *et al*. 2001), it is not surprising that sequences generated by *in vitro* recombination of authentic HV1 and numt sequences were detected. PCR-generated recombinants can potentially be generated whenever partially similar templates are present in a single reaction, as in analysis of multigene families (Zylstra *et al*. 1998; Hugenholtz & Huber 2003), single loci from polyploid families (Cronn *et al*. 2002) or of mtDNA when numts are present, as in the work presented here. The presence of recombinants as well as numts complicates efforts to distinguish authentic mtDNA from artefactural sequences using typical suggestions for checking and avoiding numts (Bensasson *et al*. 2001).

In fact, our results imply that the only certain way to ensure analysis of authentic mtDNA is by designing

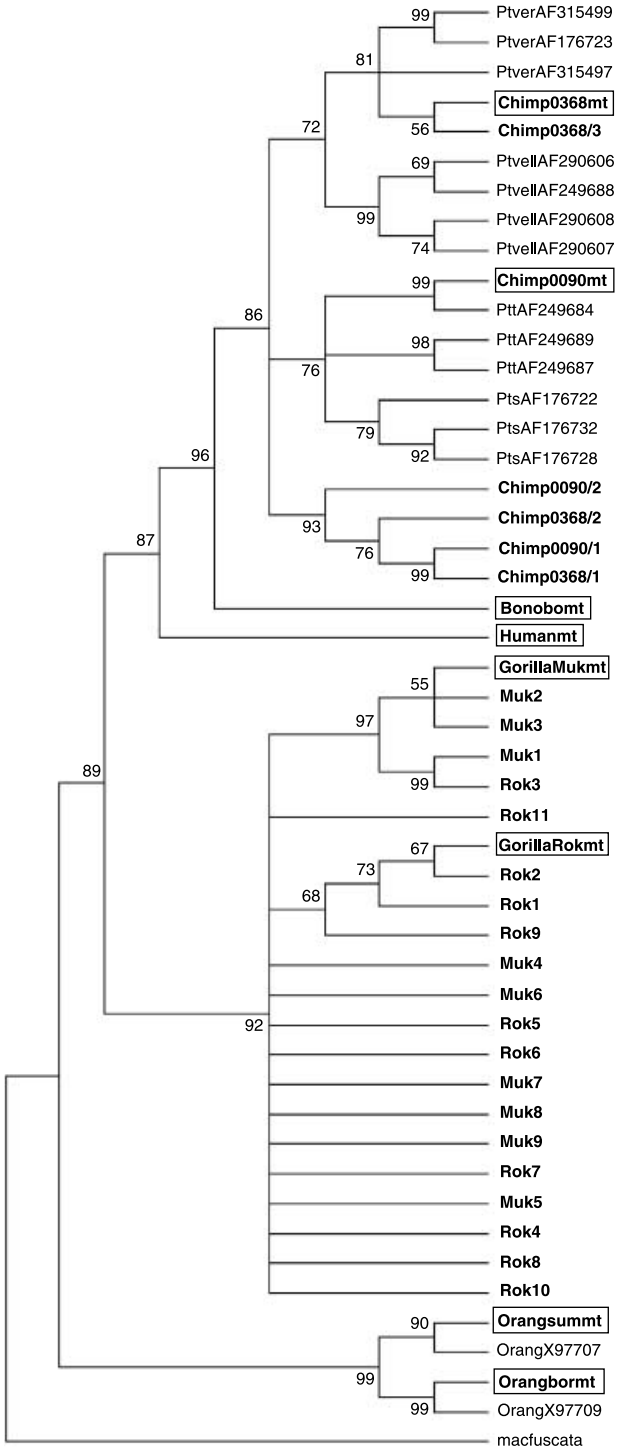

Fig. 3 Condensed maximum parsimony tree of hominoid sequences amplified in this study and additional database sequences showing only nodes that were supported by bootstrap values above 50. The sequences were cut to an equal length of 349 bp and equivalent results were obtained whether or not the large insert (105 bp) in the gorilla sequence Rok11 was included in the tree reconstruction. As an outgroup we used one sequence of *Macaca fuscata* (Accession no. AJ419864). Numbers at nodes indicate bootstrap support values greater than 50. Trees reconstructed using neighbour-joining or maximum-likelihood methods have the same branching pattern within taxa but the deepest nodes indicating relationships among the hominoid species were unresolved with weak bootstrap support. Authentic mtDNA sequences obtained in this study are indicated by the bold names in boxes, sequences labelled in bold are additional sequences identified in this study, and GenBank Accession IDs are prefaced with an indication of the relevant taxa (Ptvell, *Pan troglodytes vellerosus*; Ptver, *P. t. verus*; Ptt, *P. t. troglodytes*; Pts, *P. t. schweinfurthii*).

a long-range PCR strategy that precludes analysis of sequences integrated into the nuclear DNA or by other approaches where circular mtDNA is isolated. Such approaches yield only authentic mtDNA sequences, which can serve as a reference for comparison with results obtained using more conventional methods targeting the mitochondrial segment of interest. Our results illustrate the range of possible outcomes of such comparisons. In the human, bonobo and orangutans, despite mismatches between the universal primers and the template mtDNAs, only authentic mtDNA sequences were detected following amplification using either universal primers or primers used typically in those species. This suggests, although it does not guarantee, that analysis of additional individuals in these species using those methods would also yield authentic sequences. The results from the chimpanzees were that universal primers produced additional non-organellar DNA sequences, while primers used typically in chimpanzees would be more reliable, as only authentic mtDNA sequences were generated. In contrast, both universal and species-typical primers produced multiple nonorganellar DNA sequences from gorillas. This was particularly surprising, as only a small number of mismatches distinguish the species-typical primers and the gorilla reference sequence. The high similarity between the authentic mtDNA sequence and a numt from the same individual (11 differences, Mukmt and Muk1) compared to the authentic mtDNA sequence from the two gorillas (43 differences) implies that it would be extremely difficult to design primers that amplify only authentic gorilla HV1 mtDNA and exclude numts. While long-range PCR produces authentic gorilla mtDNA sequence, the process requires high-quality DNA (Barnes 1994; Cheng *et al*. 1994), and the DNA obtained from the noninvasive samples in common use in wildlife genetic studies is low in quality and quantity (Morin *et al*. 2001). It might be suggested that the low concentration of total DNA in extracts of noninvasive samples could actually improve the chances to amplify mtDNA instead of numts, because the nuclear DNA may be present in only a few copies and mtDNA is typically more abundant (Morin *et al*. 2001). However, experiments in our laboratory using species-typical primers to target the gorilla HV1 using extracts from faeces and shed hair roots from wild western gorillas found multiple sequences consistent with amplification of

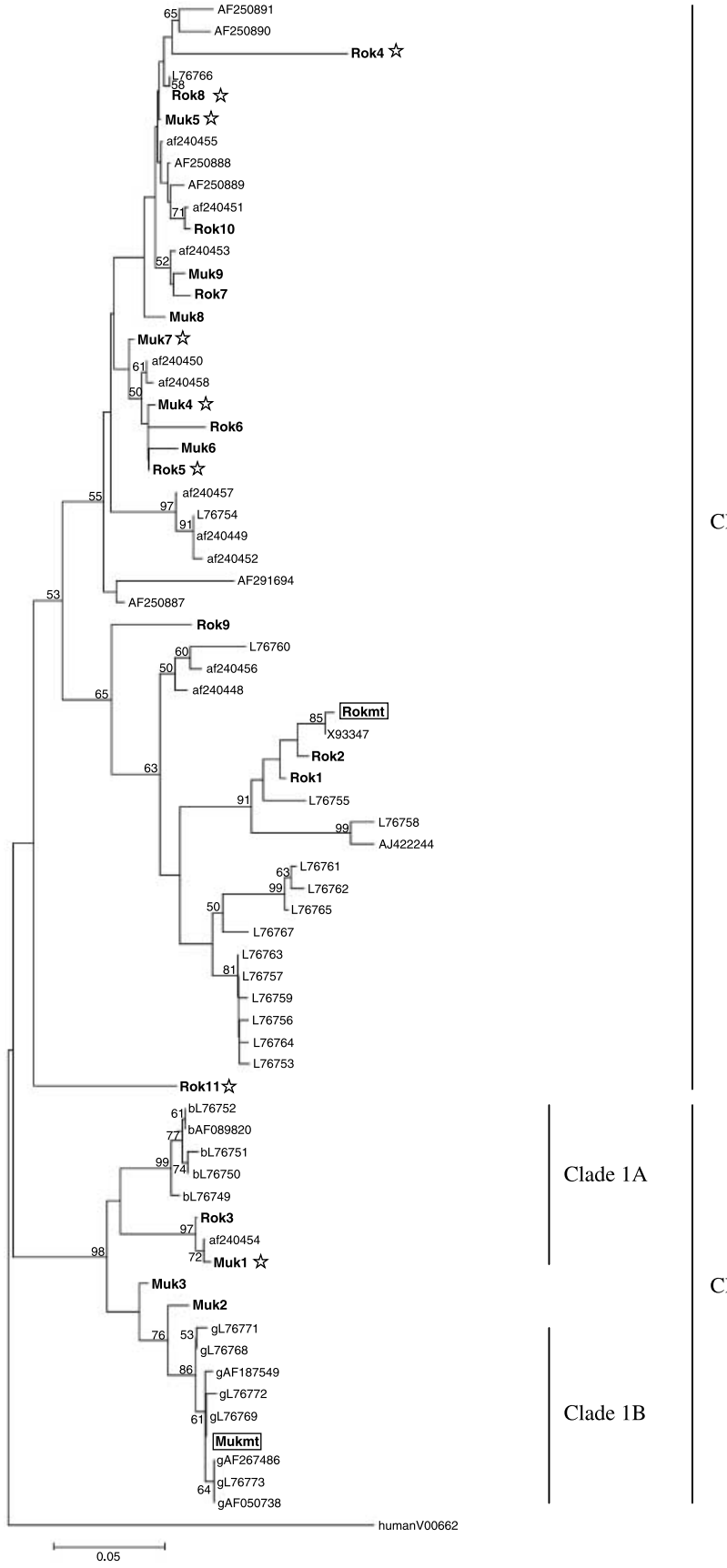

Fig. 4 Neighbour-joining tree of gorilla HV1 sequences reconstructed using 238 bp long sequences. The topology of the tree shown did not change when chimpanzee sequences or different human sequences were used as an outgroup. When we applied other tree-building methods (maximum-parsimony, maximum-likelihood) the grouping of gorilla sequences did not change but the relative arrangement of the clades varied slightly. Numbers at nodes indicate bootstrap support values greater than 50. The authentic mtDNA sequences for the western gorilla (Rokmt) and for the eastern gorilla (Mukmt) are indicated by boxes in this figure and are identical to sequences in GenBank (X93348 and L76770, Rok and Mukisi, respectively). Sequences from the database are labelled by Accession no. and the addition of 'b' indicates a sequence identified in an eastern mountain gorilla (clade 1A) and an additional 'g' indicates origin in an eastern lowland gorilla (clade 1B), while the absence of an additional letter indicates that the sequence was obtained from a western gorilla (clade 2). Additional sequences identified in this study are in bold. Sequences that we retrieved in two independent amplifications are marked with a star. Additional putative numt sequences amplified from one single tooth extraction and directly submitted to GenBank by Jensen-Seaman (2000) are indicated by lower case. In an additional experiment conducted in our lab several different sequences were observed from amplifications of DNA from a fecal sample of the gorilla whose previously published sequence has the accession number L76760 (data not shown).

nonorganellar DNA for one of two hair extracts and 13 of 33 fecal extracts (Bradley 2003).

If typical methods cannot exclude amplification of variant nonorganellar DNA sequences from gorillas, is it possible to devise a way to reliably distinguish authentic HV1 sequence from numt or recombinant sequences? By considering only sequences that had been observed in at least two amplifications, sporadic recombinants were probably excluded and the number of additional nonorganellar DNA sequences obtained from the two gorillas was reduced from 20 to eight. These eight sequences are, however, quite diverse and appeared in a variety of clades in the phylogenetic tree illustrating the relationships of sequences from gorillas. Included in this phylogenetic analysis were gorilla HV1 sequences deposited in GenBank. The majority of those sequences and others that have since been reported were generated using the conventional techniques shown here to produce artefacts and so they cannot be considered reliable (Garner & Ryder 1996; Saltonstall *et al*. 1998; Jensen-Seaman & Kidd 2001; Clifford *et al*. 2003; Hofreiter *et al*. 2003). Gorilla HV1 sequences that can be considered reliable include the two sequences (X93347, X93348) from western gorillas produced by using large overlapping clones (Xu & Arnason 1996), one of which we replicated, as well as one previously reported sequence (L76770) from an eastern gorilla (Garner & Ryder 1996) which we also replicated here. Thus, a total of three reliable gorilla HV1 sequences exist. We have considered whether it might be possible to make a catalogue of all gorilla numts against which any newly obtained sequences could be screened. Several problems with this approach immediately come to mind, including the difficulty of determining when a sufficient number of representative gorillas have been assayed and when a sufficient number of PCR products and clones have been analysed. Such a screening approach may be feasible in the future when the complete nuclear genome of gorillas, as well as other organisms, become available. We therefore conclude that because high-quality DNA cannot be currently derived from noninvasive samples, there is as yet no means to generate reliable HV1 sequences from such samples in order to determine the phylogeographical distribution of genetic variation in wild gorillas.

Implications for other studies

What are the implications of these results for studies of species other than the great apes? A recent review emphasized that the prevalence of reported numts varies among metazoans (Bensasson *et al*. 2001). Some well-studied species such as *Drosophila* do not have abundant numts (Blanchard & Schmidt 1996), whereas species such as grasshoppers (Bensasson *et al*. 2000), aphids (Sunnucks & Hales 1996), cats (Lopez *et al*. 1994), dogs (Ishiguro *et al*. 2002), macaques (Vartanian & Wain-Hobson 2002) as well as humans (Tourmen *et al*. 2002) have a great number of numts. Our results strongly suggest that published studies of mtDNA, particularly in novel species, should include details on how the authenticity of the sequences was determined. Possible methods include long-range PCR or reverse transcription (RT)-PCR to generate authentic sequences for at least some samples (Bensasson *et al*. 2001). This seems especially important in cases in which novel 'cryptic' species are suggested solely from presence of divergent mtDNA lineages in the absence of morphological or behavioural evidence. It has been suggested that DNA sequences could serve as 'barcodes' for a DNA-based taxonomy system (Hebert *et al*. 2003; Tautz *et al*. 2003). It is worth noting that in our study, although none of the eight identified numts were derived from both gorillas, in some cases (e.g. Muk 4 and Rok 5; Muk5 and Rok8) the sequences obtained from each of the two gorillas are extremely similar, differing only by a single substitution. Since western and eastern gorillas were formerly considered subspecies, but are now classified as separate species (Groves 2001), this indicates that the presence of highly similar or identical numts and potential for confusion exists even when comparing various subspecies or species of mammals. Thus, it is prudent that DNA-based taxonomic analyses include tests for validity of sequences, particularly as recommendations have been made for the use of mitochondrial genes which are known to have nuclear counterparts (Sunnucks & Hales 1996; Wu *et al*. 2000; Williams & Knowlton 2001) such as cytochrome oxidase genes (Hebert *et al*. 2003) as well as ribosomal DNA (Tautz *et al*. 2003), a nuclear gene family that can contain both functional genes and pseudogenes (Marquez *et al*. 2003).

Acknowledgements

We thank Dr K. Leus at the Center for Research and Conservation of the Royal Zoological Society of Antwerp for sample material, as well as the Yerkes Primate Centre, Atlanta. We thank C. Boesch, B. Bradley, J. Eriksson, M. Hofreiter, V. Jaenicke, D. Serre, H. Siedel and D. Stahl for helpful discussions and comments on the manuscript and C. Schwarz and B. Höffner for technical support. This work was supported by the Deutsche Forschungsgemeinschaft (VI 229/2–1) and the Max Planck Society.

Supplementary material

The following material is available from http://www.blackwellpublishing.com/products/journals/suppmat/MEC/MEC2070/MEC2070sm.htm

Figure S1. All sequences obtained from the great apes analyzed in this study. Sequences with the suffix mt indicate the authentic mtDNA sequence of an individual and two sequences with the suffix varmt indicate variants (see text). All other sequences represent non-organellar DNA sequences.

References

Allen M, Engström A, Meyers S *et al.* (1998) Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: Sensitivity and matching probabilities. *Journal of Forensic Science*, **43**, 453–464.

Altschul S, Madden T, Schaffer A *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389–3402.

Anderson S, Bankier AT, Barrell BG *et al.* (1981) Sequence and organization of the human mitochondrial genome. *Nature*, **290**, 457–465.

Arctander P (1995) Comparison of a mitochondrial gene and a corresponding nuclear pseudogene. *Proceedings of the Royal Society of London Series B, Biology Sciences*, **262**, 13–19.

Avise JC (2000) *Phylogeography*. Harvard University Press, Cambridge, MA.

Awadalla P, Eyre-Walker A, Smith JM (1999) Linkage disequilibrium and recombination in hominid mitochondrial DNA. *Science*, **286**, 2524–2525.

Barnes W (1994) PCR Amplification of up to 35-kb DNA with high fidelity and high yield from λ bacteriophage templates. *Proceedings of the National Academy of Sciences USA*, **91**, 2216–2220.

Bensasson D, Zhang D-X, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology and Evolution*, **16**, 314–321.

Bensasson D, Zhang D-X, Hewitt GM (2000) Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. *Molecular Biology and Evolution*, **17**, 406–415.

Blanchard J, Schmidt G (1996) Mitochondrial DNA migration events in yeast and humans: integration by a common end-joining mechanism and alternative perspectives on nucleotide substitution patterns. *Molecular Biology and Evolution*, **13**, 537–548.

Bradley B (2003) *Molecular ecology of wild gorillas*. PhD Thesis, University of Stony Brook, New York.

Brown WM, Wright JW (1979) Mitochondrial DNA analysis and the origin and relative age of parthenogentic lizards (genus *Cnemidophorus*). *Science*, **203**, 1247–1249.

Cheng S, Fockler C, Barnes W, Higuchi R (1994) Effective amplification of long targets from cloned inserts and human genomic DNA. *Proceedings of the National Academy of Sciences USA*, **91**, 5695–5699.

Clifford S, Abernethy K, White L *et al.* (2003) Genetic studies of western gorillas. In: *Gorilla Biology: a Multidisciplinary Perspective* (eds Taylor A, Goldsmith M), pp. 269–292. Cambridge University Press, Cambridge.

Collura RV, Stewart CB (1995) Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. *Nature*, **378**, 485–489.

Cronn R, Cedroni M, Haselkorn T, Grover C, Wendel JF (2002) PCR-mediated recombination in amplification products derived from polyploid cotton. *Theoretical and Applied Genetics*, **104**, 482–489.

Eyre-Walker A, Smith NH, Maynard Smith J (1999a) How clonal are human mitochondria? *Proceedings of the Royal Society of London Series B, Biology Sciences*, **266**, 477–483.

Eyre-Walker A, Smith NH, Maynard Smith J (1999b) Reply to Macaulay *et al.* (1999): mitochondrial DNA recombination-reasons to panic. *Proceedings of the Royal Society of London Series B, Biology Sciences*, **266**, 2041–2042.

Foran D, Hixson J, Brown W (1988) Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis. *Nucleic Acids Research*, **16**, 5841–5861.

Gagneux P, Wills C, Gerloff U *et al.* (1999) Mitochondrial sequences show diverse evolutionary histories of African hominoids. *Proceedings of the National Academy of Sciences USA*, **96**, 5077–5082.

Garner KJ, Ryder OA (1996) Mitochondrial DNA diversity in gorillas. *Molecular Phylogenetics and Evolution*, **6**, 39–48.

Gemmell N, Western P, Watson J, Graves J (1996) Evolution of the mammalian mitochondrial control region — comparisons of control region sequences between monotreme and therian mammals. *Molecular Biology and Evolution*, **13**, 798–808.

Goldberg TL, Ruvolo M (1997) The geographic apportionment of mitochondrial genetic diversity in east African chimpanzees, *Pan troglodytes schweinfurthii*. *Molecular Biology and Evolution*, **14**, 976–984.

Greenwood AD, Pääbo S (1999) Nuclear insertion sequences of mitochondrial DNA predominate in hair but not in blood of elephants. *Molecular Ecology*, **8**, 133–137.

Groves CP (2001) *Primate Taxonomy*, 1st edn. Smithsonian Institution Press, Washington.

Gyllensten U, Wharton D, Josefsson A, Wilson AC (1991) Paternal inheritance of mitochondrial DNA in mice. *Nature*, **352**, 255–257.

Gyllensten U, Wharton D, Wilson AC (1985) Maternal inheritance of mitochondrial DNA during backcrossing of two species of mice. *Journal of Heredity*, **76**, 321–324.

Hagelberg E (2003) Recombination or mutation rate heterogeneity? Implications for Mitochondrial Eve. *Trends in Genetics*, **19**, 84–90.

Hagelberg E, Goldman N, Lió P *et al.* (1999) Evidence for mitochondrial DNA recombination in a human population of island Melanesia. *Proceedings of the Royal Society of London Series B, Biology Sciences*, **266**, 485–492.

Hagelberg E, Goldman N, Lió P *et al.* (2000) Evidence for mitochondrial DNA recombination in a human population of island Melanesia: correction. *Proceedings of the Royal Society of London Series B, Biology Sciences*, **267**, 1595–1596.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Hebert PDN, Cywinska A, Ball SL, de Waard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B, Biology Sciences*, **270**, 313–321.

Hofreiter M, Jaenicke V, Serre D, von Haeseler A, Pääbo S (2001) DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research*, **29**, 4793–4799.

Hofreiter M, Siedel H, van Neer W, Vigilant L (2003) Mitochondrial DNA sequence from an enigmatic gorilla population (*Gorilla gorilla uellensis*). *American Journal of Physical Anthropology*, **121**, 361–368.

Hugenholtz P, Huber T (2003) Chimeric 16S rDNA sequences of diverse origin are accumulating in the public database. *International Journal of Systematic and Evolutionary Microbiology*, **53**, 289–293.

Ingman M, Kaessmann H, Pääbo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. *Nature*, **408**, 708–713.

Ishiguro N, Nakajima A, Horiuchi M, Shinagawa M (2002) Multiple nuclear pseudogenes of mitochondrial DNA exist in the canine genome. *Mammalian Genome*, **13**, 365–372.

Jensen-Seaman MI, Deinard AS, Kidd KK (2001) Modern African ape populations as genetic and demographic models of the last common ancestor of humans, chimpanzees and gorillas. *Journal of Heredity*, **92**, 475–480.

Jensen-Seaman M, Kidd K (2001) Mitochondrial DNA variation and biogeography of eastern gorillas. *Molecular Ecology*, **10**, 2241–2247.

Jensen-Seaman M (2000) *Evolutionary genetics of gorillas*. PhD Thesis, Yale University.

Judo M, Wedel A, Wilson C (1998) Stimulation and suppression of PCR-mediated recombination. *Nucleic Acids Research*, **26**, 1819–1825.

Kaessmann H, Wiebe V, Weiss G, Pääbo S (2001) Great ape DNA sequences reveal a reduced diversity and an expansion in humans. *Nature Genetics*, **27**, 155–156.

Kilger C, Pääbo S (1997) Direct DNA sequence determination from total genomic DNA. *Nucleic Acids Research*, **25**, 2032–2034.

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 11–120.

Kivisild T, Villems R, Jorde LB *et al.* (2000) Questioning evidence for recombination in human mitochondrial DNA. *Science*, **288**, 1931.

Kobayashi N, Tamura K, Aotsuka T (1999) PCR error and molecular population genetics. *Biochemical Genetics*, **37**, 317–321.

Kocher TD, Thomas WK, Meyer A *et al.* (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences USA*, **86**, 6196–6200.

Kocher TD, Wilson AC (1991) Sequence evolution of mitochondrial DNA in humans and chimpanzees: control region and a protein-coding region. In: *Evolution of Life: Fossils, Molecules and Culture* (ed. Osawa SHT), pp. 391–413. Springer Verlag, Tokyo.

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics*, **17**, 1244–1245.

Kvist L, Martens J, Nazarenko AA, Orell M (2003) Paternal leakage of mitochondrial DNA in the great tit (*Parus major*). *Molecular Biology and Evolution*, **20**, 243–247.

Lopez J, Culver M, Stephens J, Johnson W, O'Brien S (1997) Rates of nuclear and cytoplasmic mitochondrial DNA sequence divergence in mammals. *Molecular Biology and Evolution*, **14**, 277–286.

Lopez JV, Yuhki N, Masuda R, Modi W, O'Brien SJ (1994) Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *Journal of Molecular Evolution*, **39**, 174–190.

Macaulay V, Richards M, Sykes B (1999) Mitochondrial DNA recombination — no need to panic. *Proceedings of the Royal Society of London Series B, Biology Sciences*, **266**, 2037–2039.

Mallet J, Willmott K (2003) Taxonomy: renaissance or Tower of Babel? *Trends in Ecology and Evolution*, **18**, 57–59.

Marquez LM, Miller DJ, MacKenzie JB, van Oppen MJH (2003) Pseudogenes contribute to the extreme diversity of nuclear ribosomal DNA in the hard coral Acropora. *Molecular Biology and Evolution*, **20**, 1077–1086.

Maynard Smith J, Smith NH (2002) Recombination in animal mitochondrial DNA. *Molecular Biology and Evolution*, **19**, 2330–2332.

Mirol PM, Mascheretti S, Searle JB (2000) Multiple nuclear pseudogenes of mitochondrial cytochrome b in *Ctenomys* (Caviomorpha, Rodentia) with either great similarity to or high divergence from the true mitochondrial sequence. *Heredity*, **84**, 538–547.

Morin PA, Chambers KE, Boesch C, Vigilant L (2001) Quantitative PCR analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzes (*Pan troglodytes verus*). *Molecular Ecology*, **10**, 1835–1844.

Morin PA, Moore JJ, Chakraborty R *et al.* (1994) Kin selection, social structure, gene flow, and the evolution of chimpanzees. *Science*, **265**, 1193–1201.

Mourier T, Hansen AJ, Willerslev E, Arctander P (2001) The Human Genome Project reveals a continuous transfer of large mitochondrial fragments to the nucleus. *Molecular Biology and Evolution*, **18**, 1833–1837.

Muir C, Galdikas BMF, Beckenbach AT (2000) MtDNA sequence diversity of orangutans from the islands of Borneo and Sumatra. *Journal of Molecular Evolution*, **51**, 471–480.

Olson LE, Yoder AD (2002) Using secondary structure to identify ribosomal numts: cautionary examples from the human genome. *Molecular Biology and Evolution*, **19**, 93–100.

Pääbo S, Irwin D, Wilson A (1990) DNA damage promotes jumping between templates during enzymatic amplification. *Journal of Biological Chemistry*, **265**, 4718–4721.

Petri B, von Haeseler A, Pääbo S (1996) Extreme sequence heteroplasmy in bat mitochondrial DNA. *Biological Chemistry*, **377**, 661–667.

Poinar HN, Kuch M, Sobolik KD *et al.* (2001) A molecular analysis of dietary diversity for three archaic Native Americans. *Proceedings of the National Academy of Sciences USA*, **98**, 4317–4322.

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.

Saiki RK, Gelfand DH, Stoffel S *et al.* (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487–491.

Saltonstall K, Amato J, Powell J (1998) Mitochondrial DNA variability in Grauer's gorillas of Kahuzi-Biega National Park. *Journal of Heredity*, **89**, 129–135.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, New York.

Schwartz M, Vissing J (2002) Paternal inheritance of mitochondrial DNA. *New England Journal of Medicine*, **347**, 576–580.

Shitara H, Hayashi J-I, Takahama S, Kaneda H, Yonekawa H (1998) Maternal inheritance of mouse mtDNA in interspecific hybrids: segregation of the leaked paternal mtDNA followed by the prevention of subsequent paternal leakage. *Genetics*, **148**, 851–858.

Smith M, Thomas W, Patton J (1992) Mitochondrial DNA-like sequence in the nuclear genome of an akodontine rodent. *Molecular Biology and Evolution*, **9**, 204–215.

Sorenson MD, Fleischer RC (1996) Multiple independent transpositions of mitochondrial DNA control region sequences to the nucleus. *Proceedings of the National Academy of Sciences USA*, **93**, 15239–15243.

Stanley H, Casey S, Carnahan J *et al.* (1996) Worldwide patterns of mitochondrial DNA differentiation in the harbor seal (*Phoca vitulina*). *Molecular Biology and Evolution*, **13**, 368–382.

Stone AC, Griffiths RC, Zegura SL, Hammer MF (2002) High levels of Y-chromosome nucleotide diversity in the genus Pan. *Proceedings of the National Academy of Sciences USA*, **99**, 43–48.

Stoneking M, Soodyall H (1996) Human evolution and the mitochondrial genome. *Current Opinion in Genetics and Development*, **6**, 731–736.

Strimmer K, von Haeseler A (1996) Quartet puzzling: a quartet maximum-likelihood methods for reconstructing tree topologies. *Molecular Biology and Evolution*, **13**, 964–969.

Sunnucks P, Hales D (1996) Numerous transposed sequences of mitochondrial cytochrome oxidase I–II in aphids of the genus

Sitobion (Hemiptera: Aphididae). *Molecular Biology and Evolution*, **13**, 510–524.

Talbot SL, Shields GF (1996) Phylogeography of brown bears (*Ursus arctos*) of Alaska and paraphyly within the Ursidae. *Molecular Phylogenetics and Evolution*, **5**, 477–494.

Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. *Trends in Ecology and Evolution*, **18**, 70–74.

Thompson JR, Marcelino LA, Polz MF (2002) Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. *Nucleic Acids Research*, **30**, 2083–2088.

Tourmen Y, Baris O, Dessen P *et al.* (2002) Structure and chromosomal distribution of human mitochondrial pseudogenes. *Genomics*, **80**, 71–77.

Vartanian J-P, Wain-Hobson S (2002) Analysis of a library of macaque nuclear mitochondrial sequences confirms macaque origin of divergent sequences from old oral polio vaccine samples. *Proceedings of the National Academy of Sciences USA*, **99**, 7566–7569.

Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC (1989) Mitochondrial DNA sequences in single hairs from a southern African population. *Proceedings of the National Academy of Sciences USA*, **86**, 9350–9354.

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. *Science*, **253**, 1503–1507.

Vilà C, Savolainen P, Maldonado JE *et al.* (1997) Multiple and ancient origins of the domestic dog. *Science*, **276**, 1687–1689.

Warren KS, Verschoor EJ, Langenhuijzen S *et al.* (2001) Speciation and intrasubspecific variation of Bornean orangutans, *Pongo pygmaeus pygmaeus*. *Molecular Biology and Evolution*, **18**, 472–480.

Williams ST, Knowlton N (2001) Mitochondrial pseudogenes are pervasive and often insidious in the snapping shrimp genus Alpheus. *Molecular Biology and Evolution*, **18**, 1484–1493.

Woischnik M, Moraes CT (2002) Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. *Genome Research*, **12**, 885–893.

Wu W, Schmidt TR, Goodman M, Grossman LI (2000) Molecular evolution of cytochrome C oxidase subunit I in primates: is there coevolution between mitochondrial and nuclear genomes? *Molecular Phylogenetics and Evolution*, **17**, 294–304.

Xu X, Arnason U (1996) A complete sequence of the mitochondrial genome of the western lowland gorilla. *Molecular Biology and Evolution*, **13**, 691–698.

Zhang D-X, Hewitt GM (1996) Nuclear integrations: challenges for mitochodrial DNA markers. *Trends in Ecology and Evolution*, **11**, 247–251.

Zischler H, Geisert H, von Haeseler A, Pääbo S (1995) A nuclear 'fossil' of the mitochondrial ɒ-loop and the origin of modern humans. *Nature*, **378**, 489–492.

Zouros E, Oberhauser Ball A, Saavedra C, Freeman KR (1994) An annusual type of mitochondrial DNA inheritance in the blue mussel Mytilus. *Proceedings of the National Academy of Sciences USA*, **91**, 7463–7467.

Zylstra P, Rothenfluh H, Weiller GF, Blanden RV, Steele EJ (1998) PCR amplification of murine immunoglobin germline V genes: strategies for minimization of recombination artefacts. *Immunology and Cell Biology*, **76**, 395–405.

The authors are in the Primatology and Genetics departments the Max Planck Institute of Evolutionary Anthropology and share an interest in molecular evolution, particularly of primates. This work is part of Olaf Thalmann's dissertation research on genetic variation in gorillas and chimpanzees, and the analysis of orangutans was part of Juliane Hebler's Diplom research. Svante Pääbo is a director and Hendrik Poinar and Linda Vigilant are research scientists at the MPI.
