Primate numts and reticulate evolution of capped and golden leaf monkeys (Primates: Colobinae)

K Praveen Karanth

Centre for Ecological Sciences, Indian Institute of Science, Bangalore 560 012, India

Corresponding author (Fax, 91-80-23601428; Email,karanth@ces.iisc.ernet.in)

A recent phylogenetic study of langurs and leaf monkeys of South Asia suggested a reticulate evolution of capped and golden leaf monkeys through ancient hybridization between *Semnopithecus* and *Trachypithecus*. To test this hybridization scenario, I analysed nuclear copies of the mitochondrial cytochrome *b* gene (numts) from capped, golden and Phayre's leaf monkeys. These numts were aligned with mitochondrial cytochrome *b* sequences of various species belonging to the genera *Semnopithecus* and *Trachypithecus*. In the phylogenetic tree derived from this alignment, the numts fell into three distinct clades (A, B and C) suggesting three independent integration events. Clade A was basal to *Semnopithecus*, and clades B and C were basal to *Trachypithecus*. Among the numts in clades A and C were sequences derived from species not represented in their respective sister mitochondrial groups. This unusual placement of certain numts is taken as additional support for the hybridization scenario. Based on the molecular dating of these integration events, hybridization is estimated to have occurred around 7.1 to 3.4 million years ago. Capped and golden leaf monkeys might have to be assigned to a new genus to reconcile their unique evolutionary history. Additionally, northeast India appears to be a 'hot spot' for lineages that might have evolved through reticulate evolution.

[Karanth K P 2008 Primate numts and reticulate evolution of capped and golden leaf monkeys (Primates: Colobinae); *J. Biosci.* **33** 761–770]

1. Introduction

1.1 *Nuclear copies of mitochondrial DNA*

In the past two decades, many studies have reported the presence of nuclear copies of mitochondrial genes in a wide range of plant and animal systems (reviewed in Zhang and Hewitt 1996; Bensasson *et al* 2001). These mitochondrial-like DNA sequences in the nuclear genome are referred to as 'numt' (nuclear–mitochondrial) sequences (Lopez *et al* 1994). Numt sequences (numts) arise due to random transfer and incorporation of mitochondrial DNA (mtDNA) fragments into the nuclear genome. When transferred to the nucleus, most numts presumably lose their function (Bensasson *et al* 2001; Triant and DeWoody 2007) and thus become nuclear pseudogenes. This results in a dichotomy between mtDNA genes and their nuclear paralogs, i.e. while different mitochondrial genes and non-coding regions are selectively forced according to their functionality, numts evolve inoperably in a nuclear-specific manner and presumably without specific selection pressure (Schmitz *et al* 2005). Numts can confound the true evolutionary relationship between species or populations if erroneously used (instead of the real mitochondrial copy) in phylogenetic analysis. Therefore, the presence of numts is a major problem in phylogenetic and population genetic studies employing mitochondrial genes (Bensasson *et al* 2001).

Several methods can be employed to identify numts. As numts are nuclear pseudogenes, they evolve at a slower rate than the mtDNA gene (Collura *et al* 1996). Thus, a phylogenetic analysis of the sequences would show the numts as having shorter branch lengths than those of

Keywords. Hybridization; molecular phylogeny; mtDNA; nuclear markers; *Semnopithecus*; *Trachypithecus*

Abbreviations used: indel, insertion or deletion; ME, minimum evolution; ML, maximum likelihood; MP, maximum parsimony; mtDNA, mitochondrial DNA; mya, million years ago; numt, nuclear–mitochondrial; numts, numt sequences; PAUP*, phylogenetic analysis using parsimony and other methods; PCR, polymerase chain reaction; UPGMA, unweighted pair group method with arithmetic mean

mtDNA-encoded genes. The topology of a tree containing numts often appears unreasonable; for example, a particular species might fall basal to the ingroup, thereby causing one to suspect that the sequence is a numt. Some of these possibilities are discussed in greater detail in Collura *et al* (1996). Additionally, the pattern of nucleotide substitution among mitochondrial sequences will be quite different when compared with their nuclear paralogs due to higher overall rates of molecular evolution and a strong bias towards the third codon position in mtDNA (Mundy *et al* 2000). Numts often have insertions or deletions (indels) that would cause frameshifts, nonsense mutations and nucleotide substitutions which would result in unique amino acid replacements at sites that are highly conserved in the mtDNA-encoded proteins (Zhang and Hewitt 1996).

Numts can be used to our advantage if detected and analysed appropriately; for example, it has been used to root mtDNA trees (Zischler *et al* 1995; Hay *et al* 2004), infer ancestral state and determine the phylogenetic branching order (Bensasson *et al* 2001). Additionally, numts provide a unique and powerful tool to compare evolutionary processes of homologous sequences at paralogous locations (Schmitz *et al* 2005). Where they are numerous and selectively unconstrained, numts are ideal for the study of spontaneous mutation in nuclear genomes (Bensasson *et al* 2001). Numts have been reported in many primate species (Collura and Stewart 1995; van der Kuyl *et al* 1995; Collura *et al* 1996; Mundy *et al* 2000; Olson and Yoder 2002), but in none of these cases have they been used to study primate evolution. Here I report the presence of numts among colobine monkeys of Asia and explore their usefulness in studying the reticulate evolution of capped and golden langurs.

1.2 *Molecular phylogeny of langurs and leaf monkeys of South Asia*

According to Oates *et al* (1994), there are around five species of colobines in South Asia. These include the langurs found in the Indian subcontinent – Hanuman langur (*Semnopithecus entellus*), Nilgiri langur (*S. johnii*) and purple-faced langur (*S. vetulus*); the leaf monkeys found in the northeastern states of India, and in Bhutan and Bangladesh (referred to here as the Northeast) – golden leaf monkey (*Trachypithecus geei*) and capped leaf monkey (*T. pileatus*). Additionally, the genus *Trachypithecus* consists of at least five more species that are found predominantly in Southeast (SE) Asia – Phayre's leaf monkey (*T. phayrei*), ebony leaf monkey (*T. auratus*), silvered leaf monkey (*T. cristatus*), Francois' leaf monkey (*T. francoisi*) and dusky leaf monkey (*T. obscurus*). The biota of the Northeast has greater affinity with SE Asian biota than with the biota of the rest of the Indian subcontinent. Therefore, species distributed in the Northeast such as capped and golden langurs have not been included among those found in the Indian subcontinent. The generic level classification followed throughout this paper is as per Brandon-Jones *et al* (2003).

Recently, Karanth *et al* (2008) determined the molecular phylogeny of langurs and leaf monkeys based on the mitochondrial cytochrome *b* (*Cyt-b*) gene and two nuclear DNA-encoded genes, lysozyme (*Lzm*) and protamine P1 (*Prm1*). In this study, all three markers supported the clustering of the langurs of the Indian subcontinent, namely, Nilgiri, purple-faced and Hanuman langurs (genus *Semnopithecus*), while the *Cyt-b* and *Lzm* genes supported the clustering of the leaf monkeys of SE Asia (genus *Trachypithecus*) (figure 1). Interestingly, the *Cyt-b* and *Lzm* genes suggested a conflicting phylogenetic position of capped and golden leaf monkeys. The *Cyt-b* dataset placed these two species in the Indian clade (genus *Semnopithecus*), whereas the *Lzm* dataset placed them in the SE Asian clade (genus *Trachypithecus*) (figure 1). To reconcile this conflicting result, the authors invoked reticulate evolution of the capped-golden leaf monkey lineage through ancient hybridization between the *Semnopithecus* and *Trachypithecus* clades. In this paper, I analysed cytochrome *b* numts from capped, golden and Phayre's leaf monkeys to test this hybridization scenario.

2. Materials and methods

2.1 *Data compilation*

Sequences used in this study along with their accession numbers are listed in tables 1 and 2. For information on sample source, DNA extraction protocol, polymerase chain reaction (PCR) conditions, primer sequences and sequencing of PCR products, *see* Karanth *et al* (2008). When *Cyt-b* PCR products were directly sequenced, multiple overlapping peaks were sometimes observed for some species in the chromatograms produced by the automated sequencers. This indicated that more than one DNA sequence was present in the PCR product. Since mtDNA is haploid (Moore 1995), one expects to find only a single allele per sample (except in the rare case of heteroplasmy within an individual). Therefore, the presence of multiple sequences in a PCR product suggests that some of the sequences might be numts. To 'fish out' the true mitochondrial copy of the *Cyt-b* gene from the numts, PCR products were cloned using an Invitrogen TA cloning kit (cat. no. K2000-01 or K2000-40). For each cloned product, 4–10 positive clones were picked and grown overnight in LB medium. Plasmids were extracted using Qiagen mini-prep kits (cat. no. 27106). The insert was PCR-amplified with M13 forward and reverse primers. The first 825 bp of the product was sequenced using the *Cyt-b* primers listed in Karanth *et al* (2008). If all four inserts had different haplotypes then additional clones were sequenced. A combination of the

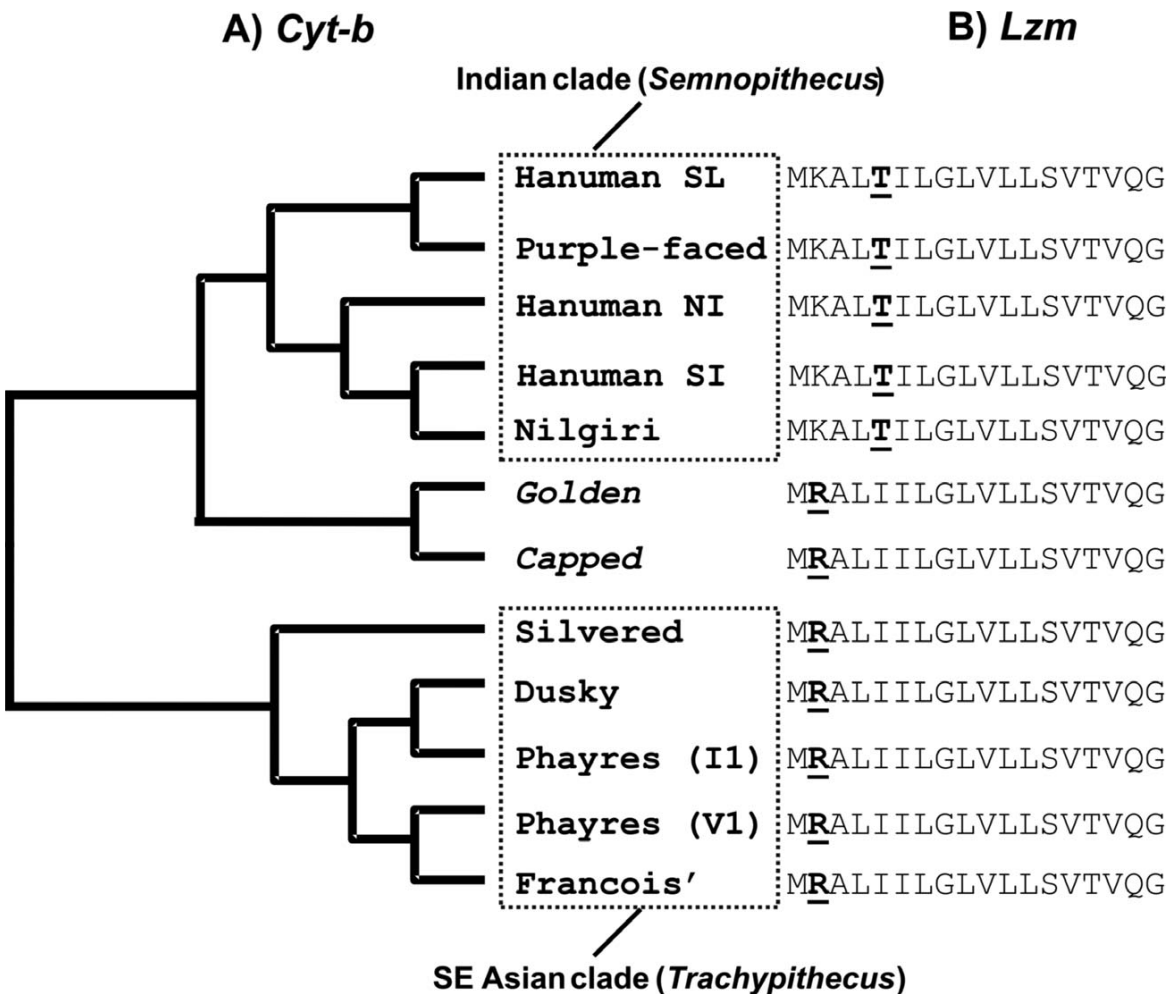

Figure 1. (**A**) Parsimony tree based on mitochondrial *Cyt-b* showing evolutionary relationships between langurs and leaf monkeys. Here, capped and golden leaf monkeys cluster with *Semnopithecus*. (**B**) Leader sequence of the nuclear-encoded Lzm protein from langurs and leaf monkeys. Underlined and bold amino acids are uniquely shared among the langurs and leaf monkeys. Here, capped and golden leaf monkeys share a unique amino acid (R) with *Trachypithecus*. I1, India; V1, Vietnam; SI, South India; NI, North India; SL, Sri Lanka.

methods described above was used to identify the mtDNA-encoded *Cyt-b* sequences (table 1) and numts (table 2). The presumed mtDNA-encoded genes have been used in Karanth *et al* (2008) in their phylogenetic analysis.

2.2 *Sequence analysis*

Numts were aligned with mtDNA-encoded *Cyt-b* sequences generated by Karanth *et al* (2008) and this combined alignment was subjected to phylogenetic analysis. Three different tree-building methods – maximum likelihood (ML), minimum evolution (ME) and maximum parsimony (MP) – were used in phylogenetic analysis using parsimony and other methods *(PAUP*)* (Swofford 2001) to determine the evolutionary relationship between numts and the mtDNA-encoded *Cyt-b* gene of the langurs and leaf monkeys of Asia. The program *MODELTEST* (Posada and Crandall 1998) was used to choose substitution models that best fit the dataset, to estimate the transition–transversion ratios, gamma shape parameters and base frequencies through likelihood ratio tests (Swofford *et al* 1996). The selected model, along with the estimated parameters, were used to derive ML and ME trees through heuristic searches. Support for various nodes for the ME tree was determined by executing 1000 bootstrap replications with full heuristic search in *PAUP**. For the MP analysis, heuristic search was performed with 10 replicates of the random addition option. Supports for various nodes were determined through 1000 bootstrap replications where each bootstrap replication did 10 additional replications with a different input order of the taxa. Gaps in the numt sequences were treated as 'missing data' for all analyses. The cercopithecine sequences (Rhesus monkey, baboon and Patas monkey) were used as outgroup to root these trees.

Table 1. Mitochondrial *Cyt-b* sequences used in this study

Species		Code	Acc. #
Langurs of Indian subcontinent			
Hanuman langur	*S. entellus*	S2	AF293952
		S5	AF293953
		N16	AF012470
		N17	AF295576
Nilgiri langur	*S. johnii*	NL1	AF294619
		NL2	AF294620
Purple-faced langur	*S. vetulus*	PF2	AF295577
Leaf monkeys of Northeast			
Golden leaf monkey	*(T.) geei*	GL2	AF294618
Capped leaf monkey	*(T.) pileatus*	CL2	AF294626
Leaf monkeys of SE Asia			
Phayre's leaf monkey	*T. phayrei*	I1	AF294621
		VI	AF294622
Francois' leaf monkey	*T. francoisi*		AF295578
Dusky leaf monkey	*T. obscurus*		AF295579
Silvered leaf monkey	*T. cristatus*		AF295580
African colobines			
Guereza colobus	*Colobus guereza*		U38264
Red colobus	*Procolobus badius*		AF294625
Cercopithecines			
Patas monkey	*Cercopithecus aethiops*		DQ069713
Rhesus monkey	*Macaca mulatta*		U38272
Baboon	*Papio hamadryas*		Y16590

S. Semnopithecus; *T. Trachypithecus*; *(T.),* tentatively placed in *Trachypithecus*; Acc. #, genbank accession number.

Table 2. Cytochrome *b* numts used in this study

Species	Code	Del.	Stop	Acc. #
Capped leaf monkey	CL2 C1	-	1	FJ042654
	CL2 C2	-	1	FJ042656
	CL2 A	-	-	FJ042655
Golden leaf monkey	GL2 C	-	-	FJ042657
Phayre's leaf monkey	I1 A1	2	-	FJ042661
	I1 A2	2	-	FJ042662
	I1 B	-	1	FJ042660
	V1 C	-	2	FJ042658
	I1 C1	5	1	FJ042659
	I1 C2	1	-	FJ042663

Del., deletion; Stop, stop codon; Acc.#, accession number.

Shimodaira–Hasegawa tests (Shimodaira and Hasegawa 1999) were performed for multiple comparisons of likelihood scores of various trees (derived through different tree-building methods) in *PAUP**. The program *MacClade* (Maddison and Maddison 1992) was used to build hypothetical phylogenetic trees based on the *Cyt-b* phylogeny of these monkeys (Karanth *et al* 2008) and the expected phylogenetic positions of numts. These hypothetical trees were used as

constraint trees in *PAUP** to derive alternative parsimony trees. The likelihood scores of these alternative parsimony trees were compared with the ML tree by implementing the Shimodaira–Hasegawa test in *PAUP**.

2.3 *Dating the nuclear integration event*

Two issues with respect to using mtDNA genes for dating nodes are heterogeneity in rates among branches in a phylogeny and among site rate variation (Yoder and Yang 2000). To mitigate these problems, I used two approaches. First, taxa exhibiting long branches from the immediate ancestor were removed from the alignment. This resulted in the removal of six mitochondrial *Cyt-b* sequences. Second, the third position of the codon was dropped from the analysis as they are more likely to attain saturation at deeper nodes. In the case of this pruned dataset, the molecular clock hypothesis that the rate of evolution is homogeneous among all branches in the phylogeny, could not be rejected (likelihood ratio test $P>0.05$). This pruned dataset was used to derive an unweighted pair group method with arithmetic mean (UPGMA) tree (with Kimura 2 parameter model) in *MEGA 3.1* (Kumar *et al* 2004). A rough estimate of the time of the nuclear integration event was determined by dating nodes that represented the common ancestors of *Cyt-b* sequences and their nuclear paralogs. Based on molecular data, Sterner *et al* (2006) inferred that the African–Asian colobine diverged around 10.8 million years ago (mya) with a 95% confidence interval of 9.8 and 11.8 mya. Thus, the molecular clock was calibrated by setting the African–Asian colobine split at 9.8 and 11.8 mya. The difference in date estimates served as the upper and lower limits for the ages of various nodes. The 'compute divergence time' option in *MEGA* was used to date these nodes.

3. Results

The hierarchical likelihood ratio test implemented in *MODELTEST* chose the HKY+G model of sequence evolution with the following parameter values: base frequencies (A = 0.3095, C = 0.3419, G = 0.1020, T = 0.2467); transition/transversion ratio = 8.7307; gamma distribution shape parameter = 0.3120. Accordingly, transversions were weighted 8:1 in parsimony analysis and the chosen model along with the parameters was used to derive ME and ML trees.

The overall topologies of the ME and MP trees were very similar. The ME tree is based on a model of sequence evolution, and is thus likely to give a better estimate of branch length than the MP tree. Therefore, the ME phylogram has been shown here (figure 2). As expected from an earlier study (Karanth *et al* 2008), the mitochondrial *Cyt-b* sequences fell into two distinct clusters, each representing the Indian (*Semnopithecus*) and SE Asian (*Trachypithecus*) clades (figure 2). A total of ten numts were identified in capped, golden and Phayre's leaf monkeys. All numts except golden GL2-C and capped CL2-A either had a deletion or nonsense mutation or both, strongly suggesting their nuclear pseudogene status (table 2). These numts were represented by three distinct clades (A, B and C) on the tree, with clade A being basal to *Semnopithecus* and clades B and C being basal to *Trachypithecus* (figure 2). Table 3 compares the pattern of nucleotide substitution within numt clades A and C with their corresponding mitochondrial counterparts (Indian and SE Asian clades). As expected, the mitochondrial sequences exhibit higher substitution rates with a strong bias towards substitution at the third codon position (Mundy *et al* 2000). Within the numt clades, substitutions are more evenly distributed among the three codon positions. Additionally, third position substitution rates are 3–5 times lower among numts when compared with mitochondrial sequences (table 3). Taken together, these results suggest that sequences in clades A and C represent numts. A similar analysis could not be undertaken for clade B due to lack of sequences. Given that the pattern of nucleotide substitutions at the third position varied significantly between the mitochondrial and numt clades (table 3), additional analyses were carried out wherein the third position was dropped. This pruned dataset was subjected to phylogenetic analysis in *PAUP**. All tree-building methods (ME, MP, ML) retrieved the three numt clades (tree not shown) and their overall topologies were similar to the topology shown in figure 2. Thus, all subsequent analyses are based on the complete dataset.

Interestingly, in figure 2, clade A has numts from species not represented in *Semnopithecus*, i.e. Phayre's I1 numts A1 and A2. Similarly, clade C also has numts from species not represented in *Trachypithecus*, such as those from capped and golden leaf monkeys. To test if this unexpected placement of certain numts in clades A and C could have occurred due to chance, two alternative parsimony tree topologies were generated. In the first tree (hypothesis 1), clade C was basal to *Semnopithecus* and *Trachypithecus*, and Phayre's numts in clade A were constrained to cluster with *Trachypithecus*. In the second tree (hypothesis 2), capped and golden numts in clade C were constrained to cluster with *Semnopithecus*, and Phayre's numts in clade A were constrained to cluster

Table 3. Nucleotide substitution pattern within clades at first, second, and third sites of the codon

		Mean p-distance			
Clade	*N*	1st	2nd	3rd	Total
A	3	0.017	0.007	0.054	0.026
Indian	9	0.042	0.023	0.167	0.077
C	6	0.023	0.016	0.032	0.031
SE Asian	5	0.027	0.009	0.181	0.072

N, sample size.

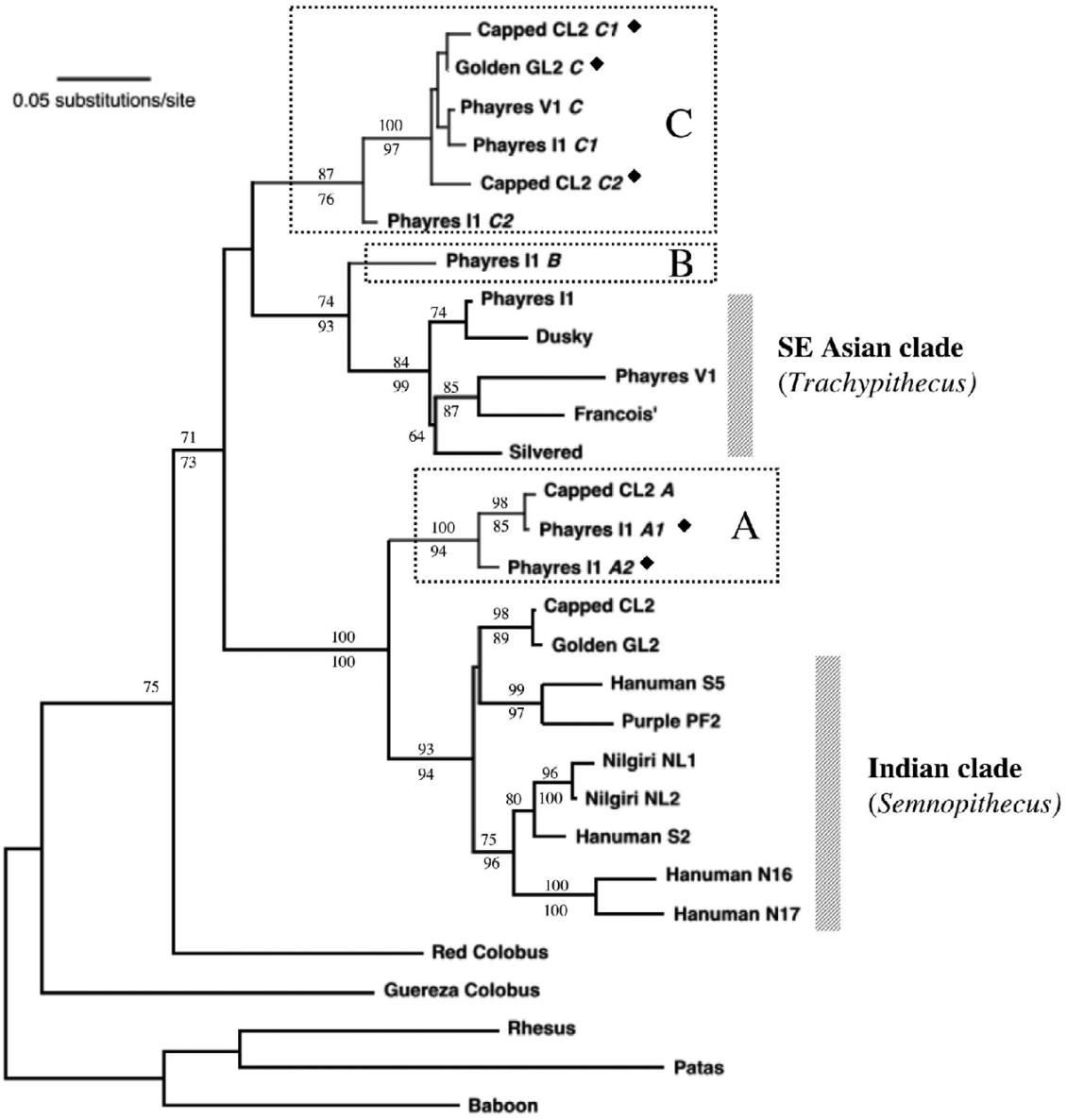

Figure 2. Evolutionary relationships between mitochondrial *Cyt-b* genes and their nuclear paralogs (numts) of langurs and leaf monkeys. This tree was generated using minimum evolution (ME) optimality criteria. Numbers above and below the nodes represent ME and maximum parsimony (MP) bootstrap supports, respectively, for 1000 replications. A, B and C represent the three numt clades. Diamonds indicate numts that have an 'unusual' phylogenetic position.

with *Trachypithecus*. The likelihood scores of the alternative parsimony trees were significantly lower than the best tree ($P<0.05$, Shimodaira–Hasegawa one-tail test) (table 4). The overall tree topology suggests that there have been at least three independent nuclear integration events represented by clades A, B and C. The divergence between numts clades (A, B and C) and their respective sister clades occurred around 6.2, 3.7 and 6.4 mya, respectively, after the split between *Semnopithecus* and *Trachypithecus* (figure 3). The ML tree (not shown) was different from the ME and MP trees in two respects. Here, the African colobines were monophyletic and numts in clade C were not monophyletic but nevertheless were basal to *Trachypithecus*. These changes do not alter the overall conclusions of the paper and are therefore not discussed further. The topologies generated by different tree-building methods were not significantly different from each other ($P>0.05$, Shimodaira–Hasegawa one-tail test) (table 4).

4. Discussion

Karanth *et al* (2008) reported that both nuclear *Lzm* and mitochondrial *Cyt-b* data support a close relationship

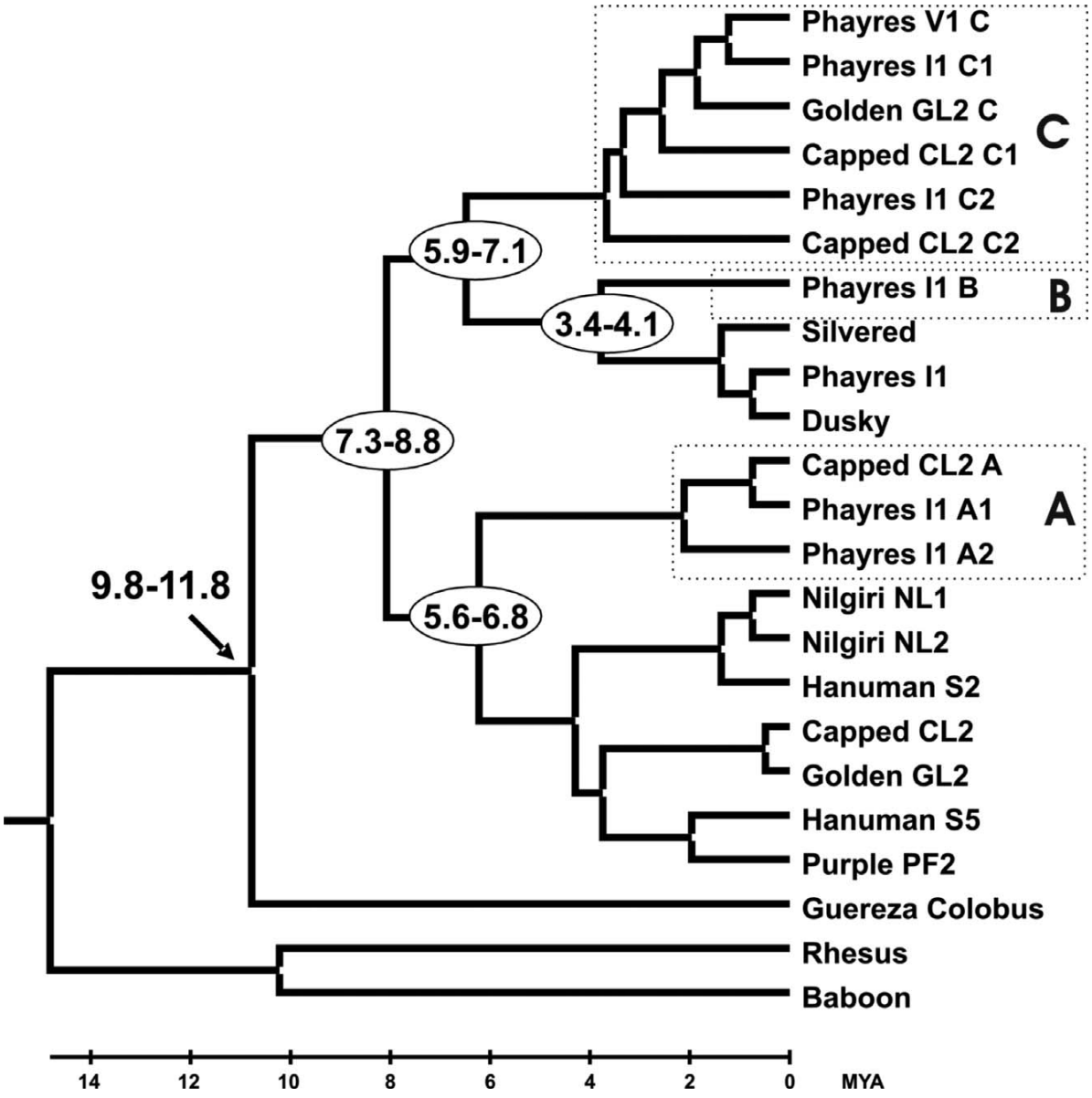

Figure 3. Molecular dating of the nuclear integration event. Arrow indicates the molecular clock calibration node. The numbers in oval are the upper and lower limits for the estimated divergence dates for various nodes (*see* text for details).

Table 4. Results of Shimodaira–Hasegawa tests

Tree	-ln L	Diff -ln L	*P*
MP	5965.66435	43.37439	0.098
ME	5964.08024	41.79029	0.098
ML	5922.28996	(best)	
H1	6041.89014	119.60018	0.000*
H2	6171.88953	249.59958	0.000*

MP, maximum parsimony; ME, minimum evolution; ML, maximum likelihood; H1 and 2, alternative parsimony trees, *see* text for details; $*P < 0.05$.

between capped and golden leaf monkeys. However, the mtDNA tree strongly suggests that they belong to the Indian clade (*Semnopithecus*), whereas the nuclear-encoded *Lzm* gene suggests that they may belong to the SE Asian clade (*Trachypithecus*) (figure 1). Interestingly, these two species are distributed in an area that is sandwiched between the distributions of *Semnopithecus* and *Trachypithecus* (figure 4). One explanation that can reconcile these two opposing results is that the capped-golden leaf monkey group might have evolved through past hybridization between the *Semnopithecus* and *Trachypithecus* (Karanth 2000). The results from the analyses of numts reported here

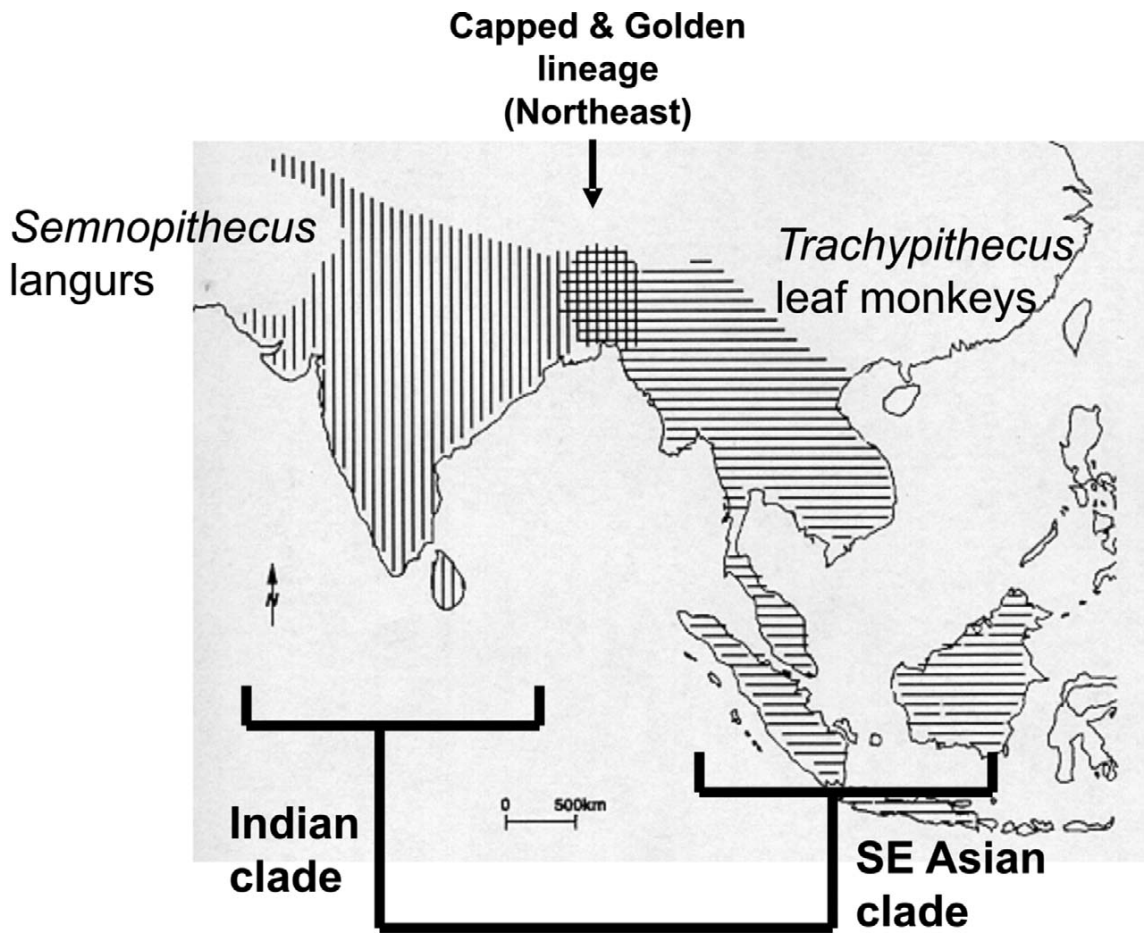

Figure 4. Biogeography of langurs and leaf monkeys of Asia. Here, the phylogeny of *Semnopithecus* and *Trachypithecus* has been overlaid onto their approximate distributions. The capped-golden lineage is distributed in areas where the distributions of *Semnopithecus* and *Trachypithecus* overlap.

further bolster this scenario. For example, the topology of the tree shown in figure 2 suggests that the numts in clade A represent a nuclear integration event that occurred on the lineage leading to *Semnopithecus* after *Semnopithecus* and *Trachypithecus* diverged from each other. Accordingly, one would predict clade A to consist of numts from species in the genus *Semnopithecus*. Thus, how do we explain the presence of Phayre's I1 (genus *Trachypithecus*) derived numt sequences in clade A? These Phayre's I1 numts might represent introgression of the *Semnopithecus* nuclear genome into *Trachypithecus*. Similarly, clade C, which is basal to *Trachypithecus*, has numts derived from capped and golden leaf monkeys, suggesting nuclear introgression in the opposite direction. The alternative tree topologies where the numts sequences were constrained to branch with their putative mtDNA clades were significantly different from the MP, ME and ML trees (table 4), suggesting that this result is not an artifact. Taken together, these findings lend additional support to the possibility of past hybridization between *Semnopithecus* and *Trachypithecus*. As mentioned above, the capped-golden lineage is distributed in an area where the distributions of *Semnopithecus* and *Trachypithecus* overlap, thus this hybridization scenario also makes biogeographical sense (figure 4). Recently, Ting *et al* (2008) reported a phylogenetic incongruence between nuclear and mitochondrial markers among Asian colobines. In their analyses, the mitochondrial dataset supported *Presbytis* +*Trachypithecus* group, whereas the nuclear dataset supported *Semnopithecus* +*Trachypithecus* group. Here again, the authors invoked hybridization as one of the possible reasons for this discordance. Thus, ancient hybridization might have occurred between multiple genera among the colobines. Indeed, introgression and hybrid speciation have caused a reticulate pattern that is still detectable in the often mosaic genomes of a number of primate groups (Arnold and Meyer 2006).

The divergence dates in figure 3 suggest that the sequences in clades A and C were transferred to the nuclear genome

around 7.1 to 5.6 mya after the split between *Semnopithecus* and *Trachypithecus*, which took place around 8 mya. The nuclear transfer of Phayre's I1 B (clade B) occurred around 4.1 to 3.4 mya. Interestingly, this clade does not contain any numts from capped and golden leaf monkeys. One possible reason for their absence is that PCR did not 'pick up' these numts in capped and golden leaf monkeys due to mutations at the primer annealing sites. Alternately, it is conceivable that by 3.4 mya the hybridization event between the *Semnopithecus* and *Trachypithecus* lineages had terminated and therefore introgression of *Trachypithecus* numts into *Semnopithecus* could no longer occur. Thus, these integration dates provide us with a broad window of 7.1 to 3.4 mya when hybridization might have occurred between *Semnopithecus* and *Trachypithecus*. Recently, Chakraborty *et al* (2007) determined the phylogenetic position of the newly described Arunachal macaque (*Macaca munzala*) based on mtDNA and Y-chromosome markers. Their study hinted at a hybrid origin of *M. munzala*, with mtDNA clustering it with *M. radiata* and Y-chromosome with *M. assamensis* and *M. thibetana*. Interestingly, this species is also distributed in the Northeast, indicating that this region might be a hot spot for the evolution of hybrid lineages.

Is it possible that any mechanism other than hybridization could have generated the topology reported here? Bensasson *et al* (2001) noted that in sexual species, numts can occur in individuals with mtDNA lineage that differs from lineages from which they are derived because they are not linked to the mtDNA lineages that generated them. The mtDNA sequence that was integrated might have belonged to a mitochondrial lineage that is now extinct or unsampled. In other words, unusual placement of some capped, golden and Phayre's numts might represent ancestral mtDNA polymorphism that has been lost since *Semnopithecus* and *Trachypithecus* diverged but have been retained in the nuclear genomes as numts. This scenario would predict a large divergence between *Semnopithecus* and *Trachypithecus* numts due to the antiquity of these sequences. The phylogeny in figure 2 suggests the contrary, for example, Phayre's I1 A1 and capped CL2 A sequences in clade A exhibit very little divergence. Thus, it is unlikely that the unusual placement of some numts reflects random sorting of ancestral mtDNA lineages.

Most taxonomic schemes published till date place capped and golden leaf monkeys in the genus *Trachypithecus*. A reticulate evolution of capped-golden lineage would imply that these two species cannot be placed in either *Semnopithecus* or *Trachypithecus*. Erecting a new genus might be warranted so that it reflects the unique evolutionary history of this lineage. Clearly, a robust test of the hybridization scenario needs to be undertaken by analysing nuclear autosomal as well as Y-chromosomal markers.

Acknowledgements

I would like to thank the members of my laboratory for their intellectual input. This work was partly funded by grants from the Department of Biotechnology and the Ministry of Environment and Forests, New Delhi.

References

Arnold M L and Meyer A 2006 Natural hybridization in primates: one evolutionary mechanism; *Zoology* **109** 261–276

Bensasson D, Zhang D-X, Hartl D L and Hewitt G M 2001 Mitochondrial pseudogenes: Evolution's misplaced witnesses; *Trends Ecol. Evol.* **16** 314–321

Brandon-Jones D, Eudey A A, Geissmann T, Groves C P, Melnick D J, Morales J C, Shekelle M and Stewart C-B 2004 Asian primate classification; *Int. J. Primatol.* **25** 97–162

Chakraborty D, Ramakrishnan U, Panor J, Mishra C and Sinha A 2007 Phylogenetic relationships and morphometric affinities of the Arunachal macaque *Macaca munzala*, a newly described primate from Arunachal Pradesh, northeastern India; *Mol. Phylogenet. Evol.* **44** 838–849

Collura R V, Auerbach M R and Stewart C-B 1996 A quick, direct method that can differentiate expressed mitochondrial genes from nuclear pseudogenes; *Curr. Biol.* **6** 1337–1339

Collura R V and Stewart C-B 1995 Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids; *Nature (London)* **378** 485–489

Hay J M, Sarre S D and Daugherty C H 2004 Nuclear mitochondrial pseudogenes as molecular outgroups for phylogenetically isolated taxa: a case study in *Sphenodon; Heredity* **93** 468–475

Karanth P 2000 *Molecular systematics and phylogeography of the langurs of Asia;* PhD dissertation, University at Albany, Albany, NY, USA

Karanth K P, Singh L, Collura R V and Stewart C-B 2008 Molecular phylogeny and biogeography of langurs and leaf monkeys of South Asia (Primates: Colobinae); *Mol. Phylogenet. Evol.* **46** 683–694

Kumar S, Tamura K and Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment; *Brief. Bioinform.* **5** 150–163

Lopez J V, Yuhki N, Masuda R, Modi W and O'Brien S J 1994 *Numt*, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat; *J. Mol. Evol.* **39** 174–190

Maddison W P and Maddison D R 1992 *MacClade: analysis of phylogeny and character evolution, Version 3* (Sunderland, MA: Sinauer Associates)

Moore W S 1995 Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees; *Evolution* **49** 718–726

Mundy N I, Pissinatti A and Woodruff D 2000 Multiple nuclear insertions of mitochondrial cytochrome *b* sequences in callitrichine primates; *Mol. Biol. Evol.* **17** 1075–1080

Oates J F, Davies A G and Delson E 1994 The diversity of living colobines; in *Colobine monkeys: their ecology, behaviour*

and evolution (eds) Davies A G and Oates J F (Cambridge: Cambridge University Press) pp 45–73

Olson L E and Yoder A D 2002 Using secondary structure to identify ribosomal numts: cautionary examples from the human genome; *Mol. Biol. Evol*. **19** 93–100

Posada D and Crandall K A 1998 MODELTEST: testing the model of DNA substitution; *Bioinformatics* **14** 817–818

Schmitz J, Piskurek O and Zischler H 2005 Forty million years of independent evolution: a mitochondrial gene and its corresponding nuclear pseudogene in primates; *J. Mol. Evol.* **61** 1–11

Shimodaira H and Hasegawa M 1999 Multiple comparison of log-likelihoods with applications to phylogenetic inference; *Mol. Biol. Evol.* **16** 1114–1116

Sterner K N, Raaum R L, Zhang Y-P, Beth C-B and Disotell T R 2006 Mitochondrial data support an odd-nosed colobine clade; *Mol. Phylogenet. Evol.* **40** 1–7

Swofford D 2001 *PAUP*– phylogenetic analysis using parsimony and other methods, version 4* (Sunderland, MA: Sinauer Associates)

Swofford D L, Olsen G J, Waddell P J and Hillis D M 1996 Phylogenetic inference; in *Molecular systematics* (eds) Hillis D M, Moritz C and Mable B K (Sunderland, MA: Sinauer Associates) pp 407–514

Ting N, Tosi A J, Li Y, Zhang Y-P and Disotell T R 2008 Phylogenetic incongruence between nuclear and mitochondrial markers in the Asian colobines and the evolution of the langurs and leaf monkeys; *Mol. Phylogenet. Evol.* **46** 466–474

Triant D A and DeWoody J A 2007 Extensive mitochondrial DNA transfer in a rapidly evolving rodent has been mediated by independent insertion events and by duplications; *Gene* **402** 61–70

van der Kuyl A C, Kuiken C L, Dekker J T, Perizonius W R K and Goudsmit J 1995 Nuclear counterparts of the cytoplasmic mitochondrial 12S rRNA gene: a problem of ancient DNA and molecular phylogenies; *J. Mol. Evol.* **40** 652–657

Yoder A D and Yang Z 2000 Estimation of primate speciation dates using local molecular clocks; *Mol. Biol. Evol*. **17** 1081–1090

Zhang D-X and Hewitt G 1996 Nuclear integrations: challenges for mitochondrial DNA markers; *Trends Ecol. Evol.* **11** 247–251

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

MS received 27 December 2007; accepted 13 August 2008

*e*Publication: 20 October 2008

Corresponding editor: VIDYANAND NANJUNDIAH