Hybridisation or Incomplete Lineage Sorting: The convoluted story about the origin of the capped-golden langur lineage (Cercopithecidae: Colobinae)

Kunal Arekar[a, *], Abhijna Parigi[a, b], K. Praveen Karanth[a]

[a] Evolving-Phylo Lab, Centre for Ecological Sciences, Indian Institute of Science, Bangalore 560012, India

[b]presently at Integrative Biology Department, Michigan State University, East Lansing, Michigan 48824, USA

*Corresponding author. Email – kunal.cyan@gmail.com

Abstract

Hybridisation and incomplete lineage sorting (ILS) are two of the many naturally occurring evolutionary processes that contribute to variability in individual gene histories. However, distinguishing between hybridisation and ILS is challenging and these processes can lead to serious difficulties in inferring phylogenies. These processes have been invoked as a possible explanation for the evolution of the endangered golden langur and its sister species, the capped langur. Previous molecular studies have retrieved conflicting phylogenies, with mitochondrial tree grouping capped and golden langur (CG) lineage with a largely Indian genus *Semnopithecus* while nuclear markers supporting their affiliation with a Southeast Asian genus, *Trachypithecus*. However, as pointed by others, the use of nuclear copies of mitochondrial DNA in the above studies might have generated the discordance. Because of this discordance, the phylogenetic position of CG lineage has been much debated in recent times. In this study, we have used nine nuclear and eight mitochondrial markers in conjunction with coalescent based species tree approach to better understand the evolutionary origin of CG lineage. Concatenated nuclear as well as the mitochondrial dataset recovered congruent relationships where CG lineage was sister to *Trachypithecus*. However nuclear species tree estimated using different multispecies coalescent methods were incongruent with the above result, suggesting presence of ILS/hybridisation. Furthermore, CG lineage is morphologically intermediate between *Semnopithecus* and *Trachypithecus* with respect to skull and body measurements. Based on these evidences, we argue that CG lineage evolved through hybridisation between *Semnopithecus* and *Trachypithecus*. Taxonomic and conservation implications of these results are also discussed.

Keywords: CG lineage; ancient hybridization; incomplete lineage sorting; *Trachypithecus*; *Semnopithecus*.

1. Introduction

With advances in molecular and analytical techniques, there is an enormous amount of genetic data available to resolve evolutionary relationships between species. Nevertheless, for certain groups determining their evolutionary histories has been challenging due to the presence of evolutionary processes like hybridisation, incomplete lineage sorting (ILS), gene duplication and loss and horizontal gene transfer (Doyle, 1992; Maddison, 1997; Chou et al., 2015). These processes can lead to discrepancy between the gene trees and the true species tree (Maddison 1997; Degnan and Rosenberg, 2006). Of these processes, two of the most studied are ILS and hybridisation (Posada, 2002; Gauthier and Lapointe, 2007; Liu and Pearl, 2007; Joly et al., 2009; Kubatko, 2009; Kubatko et al., 2009; Meng and Kubatko, 2009; Heled and Drummond, 2010; Liu et al., 2010; Bryant et al., 2012; Chifman and Kubatko, 2014; Mirarab et al., 2014; Yu et al., 2014; Mirarab and Warnow, 2015). ILS arises when two or more gene trees fail to coalesce back in time in the most recent common ancestor (Maddison, 1997) and this is more likely to occur in lineages where not enough time has passed since speciation (Leache and Rannala, 2011) whereas, hybridisation is a process that results in mixing of previously isolated gene pools. The interbreeding of individuals from different gene pools results in a hybrid species that shares genetic information with both the parent species (Mallet, 2005, 2007; Zinner et al., 2011). There have been many methods developed to estimate species tree in presence of ILS (Edwards et al., 2007; Liu and Pearl, 2007; Mossel and Roch, 2008; Kubatko et al., 2009; Heled and Drummond, 2010; Liu et al., 2010; Bryant et al., 2012; Chifman and Kubatko, 2014; Mirarab and Warnow, 2015) and hybridisation (Xu 2000; Posada, 2002; Jin et al., 2006; Gauthier and Lapointe, 2007; Olave et al., 2017). However, distinguishing between these two processes is difficult and very few methods have been developed that incorporates both hybridisation and ILS (Buckley et al., 2006; Than et al., 2007; Joly et al., 2009; Meng and Kubatko, 2009). Here we have used multiple nuclear and mitochondrial markers in conjunction with different multispecies coalescent methods to understand the processes that gave rise to a lineage of Asian colobine monkeys – the capped and golden langur (CG) lineage.

Colobines are predominantly leaf eating monkeys (Chivers and Hladik, 1980) distributed in the tropical old world. They are a diverse group consisting of 10 genera and over 50 species (Groves, 2001; Brandon-Jones et al., 2004). Among the Asian colobines, *Semnopithecus* and *Trachypithecus*, are the most species rich genera distributed in the Indian subcontinent and Southeast Asia (SEA) respectively. There has been much ambiguity in the taxonomy of these two genera (Oates et al., 1994; Groves, 2001; Brandon-Jones et al., 2004);

however, with the advent of molecular tools many of these issues have been resolved (Karanth et al., 2008; Karanth, 2008; Ashalakshmi et al., 2014). Nevertheless, the taxonomic placement of two species in the genus *Trachypithecus* continues to be in debate; these include the endangered golden langur (*Trachypithecus geei*) and capped langur (*Trachypithecus pileatus*). Golden langur is distributed in parts of Bhutan and adjoining Indian state of Assam in Northeast India (Wangchuk et al., 2008; Ram et al., 2016) whereas, Capped langur is more widespread and is distributed in northeast India, parts of Bhutan, Bangladesh, northwest Myanmar and southern China (Zhang et al., 1981; Srivastava and Mohnot, 2001; Kumar and Solanki, 2008). Molecular data suggests that these two species are closely related however their phylogenetic position in relation to *Semnopithecus* and *Trachypithecus* remains unresolved. In the molecular studies undertaken by two separate groups, the mitochondrial Cytochrome *b* (Cyt-*b*) phylogenetic tree placed Capped-Golden (CG) lineage in the *Semnopithecus* clade whereas nuclear autosomal and Y chromosomal markers suggested that they were sister to *Trachypithecus* (Karanth et al., 2008; Osterholz et al., 2008). This discordance led the authors to hypothesise reticulate evolution of CG lineage due to past hybridisation between *Semnopithecus* and *Trachypithecus*. Osterholz et al. (2008) pointed out that incomplete lineage sorting in either nuclear or mitochondrial markers could also generate this discordance. Interestingly the distribution of CG lineage is flanked by *Semnopithecus* on the west and by *Trachypithecus* in the east (Groves 2001; Karanth et al., 2008; Osterholz et al., 2008). The same year another study looked at the phylogenetic position of CG lineage using Cyt-*b* marker (Wangchuk et al., 2008). Surprisingly in their mitochondrial tree CG lineage branched within the *Trachypithecus* clade. Thus, Wangchuk et al. (2008) results were consistent with the retrieved topology using nuclear markers (Karanth et al., 2008; Osterholtz et al., 2008) of the aforementioned studies and did not support the mitochondrial-nuclear discordance scenario (Karanth et al., 2008; Osterholtz et al., 2008). Accordingly, they rejected the hybridization hypothesis as a possible reason for the origin of CG lineage.

These studies suggested that CG lineage had at least two types of Cyt-*b* sequences, one of which cluster them with *Semnopithecus* and the other with *Trachypithecus*. Multiple mitochondrial sequences when amplified from the same species suggest the presence of nuclear mitochondrial DNA sequences (numts). Numts are mitochondrial sequences incorporated into the nuclear genome (Doolittle, 1998; Blanchard and Lynch, 2000; Bensasson et al., 2001) and behave like nuclear pseudogenes. These nuclear copies of mitochondrial sequences when considered as true mitochondrial sequences and used in analysis can lead to erroneous

phylogenetic topologies (Collura and Stewart, 1996; Bensasson et al., 2001). Numts have been reported from both *Semnopithecus* and *Trachypithecus* (Karanth, 2008). A recent study reported the isolation of full length numts from capped langur and *Trachypithecus shortridgei* (Wang et al., 2015). These numt sequences branched with the genus *Semnopithecus* (Karanth et al., 2008; Osterholz et al., 2008), but the mitochondrial sequences from capped langur and *T. shortridgei* places them sister to *Trachypithecus* genus consistent with the Cyt-*b* phylogeny of Wangchuk et al., (2008). Thus, the authors concluded that "mitochondrial" Cyt-*b* sequences of CG lineage used by Karanth et al. (2008) and Osterholz et al. (2008) were numts.

Taken together, these studies would suggest that CG lineage indeed belongs to *Trachypithecus* clade. Nevertheless, studies on numts by Karanth (2008) and Wang et al. (2015) suggest that hybridisation hypothesis for the origin of CG lineage cannot be rejected. Karanth (2008) isolated numt sequences from capped langur, golden langur and Phayre's leaf monkey (*T. phayrei*). Assuming that CG lineage belongs to *Trachypithecus*, these numts are predicted to be related to the *Trachypithecus* clade or expected to be sister to the clade consisting of both *Semnopithecus* and *Trachypithecus*. Interestingly in their analysis some numts isolated from both capped and golden langurs were also sister to *Semnopithecus*. This unusual placement of some CG numts thus suggested hybridization. Based on molecular dating analysis Karanth (2008) estimated the hybridisation to have occurred between 7.1 mya to 3.4 mya (million years ago). Similarly, Wang et al. (2015) proposed that the numts originated in the genus *Semnopithecus* and hypothesised a unidirectional introgression from *Semnopithecus* into the ancestors of capped langur and *T. shortridgei*. They dated the time of hybridisation to be no later than 0.26 mya and no earlier than 3.47 mya. Hybridisation in primates has been reported between species and subspecies (Rabarivola et al., 1991; Vasey and Tatersall, 2002; Pastorini et al., 2009), but there are a few studies that have observed it between genera (Dunbar and Dunbar, 1974; Jones et al., 2005; Davenport et al., 2006). Earlier works are mostly based on morphological evidences of hybridisation (Nagel, 1973; Dunbar and Dunbar, 1974; Rabarivola et al., 1991), but the advances made in the field of molecular tools can be applied to understand the time of occurrence of a hybridisation event and its fitness consequences (Arnold and Meyer, 2006; Zinner et al 2011).

Thus, nuclear and mitochondrial data support placing CG lineage in *Trachypithecus*. Nevertheless, the numts data also suggests past hybridization between *Trachypithecus* and *Semnopithecus*. Given this background, the goal of this study is to resolve the phylogenetic position of CG lineage using multiple nuclear and mitochondrial markers.

2. Materials and Methods

2.1. Taxon sampling and DNA extraction

A non-invasive sampling technique was employed for this study. Two fecal samples from golden langur and four samples (two fecal and two tissue) from capped langurs were used in this study. All these samples were collected from different zoos across India (Appendix A). Care was taken to collect only samples of wild caught captive animals and zoo born individuals were excluded. Fresh faeces were collected in all cases, samples were collected using plant twigs to avoid contamination from humans and preserved in sterile vials with absolute alcohol (Merck, Germany) and stored in -20°C.

For DNA extraction from tissue samples, we used the commercially available DNeasy® Blood & Tissue Kit (QIAGEN Inc.), we followed the manufacturer's protocol. DNA from fecal samples was extracted using the commercially available QIAamp DNA stool mini kit (QIAGEN Inc.), following the manufacturer's protocol with slight modifications as mentioned in Mondol et al. (2009), however, we did not add the carrier RNA (Poly A) (Kishore et al., 2006; Mondol et al., 2009). Each extraction had a negative control to monitor contamination. The quantity of extracted DNA was measured using a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific Inc).

2.2. PCR Amplification and Sequencing

Nine nuclear markers were selected for this study (Table 1) (Perelman et al., 2011; Karanth et al., 2008). The nuclear markers were amplified using the following conditions. A 25 µl PCR reaction volume was set with 2 µl (total DNA concentration varied between samples ranging from 20 ng/µl to 80 ng/µl) of template DNA, 0.25 mM of dNTPs (Bangalore Genei, Bangalore), 0.2 µM of each primer (Amnion Biosciences, Bangalore), 1U Taq polymerase (New England BioLabs® Inc.) and a standard 1X reaction buffer premixed with 1.5 mM $MgCl_2$ (New England BioLabs® Inc.). A 'touchdown' PCR program was used with an initial annealing temperature of 60 °C with subsequent decrease by 1 °C per cycle till the specified annealing temperature, 50 °C, is reached. Each reaction was performed for 95 °C for 15 sec, 60-50 °C for 30 sec and 72 °C for 60 sec, with initial denaturation at 95 °C for 10 min. PCR products were purified using the QIAquick® PCR purification kit (QIAGEN Inc.) following the manufacturer's protocol, for some samples we used ExoSAP-IT (Affymetrix®, USB Products). The purified product was outsourced for sequencing to Amnion Biosciences Pvt. Ltd., Bangalore; some of the samples were outsourced to Medauxin, Bangalore.

Additionally, eight mitochondrial markers were amplified from a single capped langur tissue sample (Table 2). First, we used a long-range PCR kit (Expand Long Template PCR System, Roche) to amplify a long fragment of mitochondrial DNA (mtDNA) genome in order to minimise the chances of amplifying a numt sequence. Second, we identified these eight genes by comparing our sequence with the whole mitochondrial (mt) genome of capped langur available in GenBank. Following primer pair was used to amplify the long fragment; 5275F (5' – ACCYCTGTCTTTAGGTTTACAGCCTAATG – 3') and 11718R (5' – CCAATGGATAGCTGTTATCCTTTAAAAGTTGAG – 3') (Raaum et al., 2005). The PCR cycle conditions followed were as per the manufacturer's protocol (Roche). PCR product was outsourced for purification and sequencing to Amnion Biosciences Pvt. Ltd., Bangalore. We were not able to amplify the long mitochondrial fragment from fecal samples as the amount of DNA in fecal samples is very less and therefore not suitable for amplification of longer fragments. We did not attempt to amplify small mitochondrial fragments for single markers from the fecal samples, to avoid accidental amplification of numt sequences.

2.3. Phylogenetic analysis

The sequence chromatogram file was viewed and edited manually using ChromasLite v2.01 (Technelysium Pty Ltd). We generated a long fragment of mitochondrial genome for one capped langur tissue sample (CES11/299, Appendix A). This long fragment of mtDNA consists of eight protein coding loci – Cox1, Cox2, ATP8, ATP6, Cox3, ND3, ND4L, ND4 (Table 2). All the sequences are deposited in GenBank (Table #; sequences will be deposited after the acceptance of manuscript). Each sequence was checked by comparing with orthologous sequences available in GenBank by using Nucleotide BLAST search tool.

The mitochondrial fragment sequenced in this study was aligned with other mitochondrial orthologs available in GenBank (Table 3) and furthermore, the six non-coding tRNAs were excluded from the analysis. Nuclear dataset generated in this study were aligned with sequences of 13 other Asian and African colobines available in Genbank (Table 4). The two data sets were aligned separately using Muscle algorithm (Edgar, 2004) incorporated in MEGA v5.2 (Tamura et al., 2011) with default settings. We used p-distance algorithm in MEGA to estimate the evolutionary divergences between sequences.

PartitionFinder 1.1.0 (Lanfear et al., 2012) was used to pick the best partition scheme as well as the model of sequence evolution. Partitioning schemes used for analysis of our data are listed in Table 5 [(A) and (B)]. Phylogenetic reconstruction was performed using Maximum

Likelihood (ML) and Bayesian methods. ML analysis was performed in RAxML 7.4.2 incorporated in raxmlGUIversion 1.3 (Stamatakis, 2006). As there is no provision in RAxML for applying multiple models across partitions, we used GTR + G as the model of substitution for all partitions in our data set (Table 5). We performed 1000 replicates to assess support for different nodes. The Bayesian analysis was done in MrBayes 3.2.2 (Ronquist et al., 2012). Nuclear data analysis was run for 15 million generations with sampling frequency at 500 and the mitochondrial dataset was run for 15 million generations and sampling frequency at 2000. Analyses had two parallel runs with four chains, convergence of the runs was determined if the standard deviation of split frequencies was <0.01. Using Tracer v1.6 (Rambaut et al., 2013) we plotted each parameter against generation time to determine the effective sample size (ESS) value of >200. The first 25% trees were discarded as burn-in.

2.4. Species tree estimation

Evolutionary processes like hybridisation, incomplete lineage sorting (ILS), gene duplication and gene loss make it difficult to estimate species trees from multi-locus nuclear data. These processes make the gene trees incongruous and different from the overall species tree (Maddison, 1997). The individual gene trees from the nine nuclear markers were uninformative due to lack of phylogenetically informative sites; therefore, to infer a nuclear marker based species tree we used two different approaches. First the program ASTRAL II (Mirarab and Warnow, 2015) was used to build species tree for the nuclear dataset. It uses individual nuclear gene trees as input file and estimates a phylogeny that agrees with the maximum quartet trees induced by the gene tree set. The command "java -jar astral.4.10.11.jar –I in.tree -o out.tre" was used to run the input file and to get an output tree file. Input tree file was generated by combining individual gene tree outputs from RAxML.

ASTRAL is a 'summary statistic' method which uses input gene trees to estimate species tree. However, summary methods are known to be sensitive to gene tree estimation error, specifically on short alignments (Chou et al., 2015). The program SVDquartets is a 'single site' method that first builds a quartet tree for all possible quartet combinations; for all the possible trees in a quartet, SVD (single value decomposition) score is calculated and the relationship with the lowest score is picked up as the best tree. It then uses a quartet assembly technique, like QFM (Reaz et al., 2014) to build a tree. We ran the program with 100000 quartets to evaluate with the option to check all the possible quartets in PAUP* 4.0a157 (Swofford, 2002). Two approaches were used; first species tree analysis was undertaken with

"taxa partition" option wherein individual members were assigned to major lineages recovered from the concatenation analysis (Fig. 1 & Fig. 2) and secondly, the analysis was run without taxon partitions where the tips were designated as separate populations rather than grouping them in major lineages. For both the analysis, we used the multispecies coalescent tree model and the analyses were subjected to 100 bootstrap replicates.

2.5. Divergence dating

We used the species tree ancestral reconstruction - *BEAST (Heled et al., 2013) incorporated in BEAST v2.4.7 (Bouckaert et al., 2014) for divergence time estimation for our nuclear data set. Input files were created in BEAUti v2.4.7 (Bouckaert et al., 2014). The selected partitioning scheme is described in Table 5(A). We used an uncorrelated relaxed lognormal clock for each gene with default values for all the priors. Yule model was used as species tree prior. It is important to set the correct ploidy levels as differences in effective population size influences the coalescence time (Charlesworth, 2009). All the genes in our analysis are autosomal nuclear except for FAM123B, which is an X-linked locus; so, the ploidy levels were selected accordingly. We *a priori* grouped the tips into their respective genera; individuals from CG lineage were assigned to a separate group (Appendix B). Two independent analyses were run for 80 million generations, sampling every 5000 generations. Stationarity was assessed in Tracer v1.6 (Rambaut et al., 2013) based on the plots of parameters vs. state or generation and the effective sample sizes (>200). The tree files were combined using LogCombiner 2.4.7 and constructed maximum clade credibility tree with median height using TreeAnnotator v2.4.7 after discarding first 25% trees as burn-in.

We used two fossil calibrations and one secondary calibration for dating the phylogeny. A previous study (Perelman et al., 2011) inferred the most recent common ancestor (MRCA) for Asian colobine at 8.8mya with 95% CI of 6.5 – 11.2 mya. Therefore, the molecular clock prior was set to normal with mean of 8.8. One of the fossils we used for calibration is a primate from middle to late Miocene Eurasia, *Mesopithecus pentelicus* (Pan et al., 2004). To the best of our knowledge, it has never been used as a calibration for any primate phylogeny. According to Pan et al. (2004), this fossil is related to the odd-nosed colobine genus *Rhinopithecus* and *Nasalis*. According to our ML and Bayesian phylogenies, *Pygathrix* is placed sister to the *Rhinopithecus* + *Nasalis* clade. We constrained the *Rhinopithecus* + *Nasalis* + *Pygathrix* clade as monophyletic and calibrated this node using lognormal prior distribution with mean = 9.0, offset = 6. Another fossil we used is a late Pliocene *Semnopithecus gwebinensis* fossil

discovered from the Irrawaddy sediments of central Myanmar (Takai et al., 2015). For this calibration, we constrained *Semnopithecus* + *Trachypithecus* + CG lineage as monophyletic and used a lognormal distribution prior with mean = 12.0 and offset = 2.58.

3. Results

3.1. Phylogenetic Analysis

3.1.1. Nuclear phylogeny

Our final concatenated nuclear dataset contained 3307 bp from nine nuclear genes: ABCA1 (324 bp), BCHE (244 bp), BDNF (468bp), DMRT1 (530 bp), ERC2 (333 bp), FAM123B (369 bp), FES (479 bp), LZM (136) and MAPKAP (396 bp) for 20 individuals. Out of these, seven individuals belonged to CG lineage and one to *Semnopithecus hypoleucos*. Remaining sequences were downloaded from GenBank (see Table 4).

In the Bayesian phylogeny (Fig. 1), the CG lineage was sister to *Trachypithecus* with a very high support (PP = 1). The *S. hypoleucos* specimen from this study was branching with rest of the *Semnopithecus* individuals. Furthermore, *Semnopithecus* and *Trachypithecus* were sister to each other to the exclusion of the odd-nosed monkey group. Relationships within these major clades are not completely resolved and lack good support. The ML tree (not shown) also retrieved similar relationships with bootstrap support (BS = 98) for the CG + *Trachypithecus* clade.

3.1.2. Mitochondrial Phylogeny

The mitochondrial dataset consists of a large fragment (5933 bp) containing eight mitochondrial genes (Table 2) from 23 individuals. Out of this, one belonged to capped langur (CG lineage) generated from tissue sample in this study; remaining sequences were downloaded from GenBank (Table 3). Downloaded sequences consisted of representatives from odd-nosed monkeys, *Trachypithecus* group, *Semnopithecus* group and outgroup from African colobines. In the Bayesian tree (Fig. 2), capped langur (CG lineage) was sister to *T. shortridgei* and together these two species are sister to other *Trachypithecus* species from Southeast Asia (PP=1). Furthermore, *Presbytis melalophos*, instead of *Semnopithecus* was sister to *Trachypithecus*. The topology in the likelihood tree (BS = 100; not shown) was similar to the Bayesian tree. Thus, both nuclear and mitochondrial markers suggested sister relationship between CG lineage and *Trachypithecus*. However, as previously described by Roos et al. (2011), there was discordance between mtDNA and nuclear dataset with respect to

relationships between *Semnopithecus*, *Trachypithecus* and *Presbytis* (For further discussion see Roos et al., 2011). We also conducted a separate analysis of the mtDNA dataset that included the previously reported numts by Wang et al. (2015). In this tree, the mtDNA fragment generated in this study did not branch with the numts (tree not shown). Additionally, our sequence did not have any of the typical changes associated with numts, such as indels that cause frame shift, nonsense mutations etc.

3.2. Species tree estimation

Our coalescent based multilocus analysis generated contradictory results. In the SVDquartet analysis, where multispecies coalescent model was used without taxon partitions, CG lineage was sister to *Trachypithecus* (BS = 36) with the exclusion of *Semnopithecus*. However, when "taxon partitions" option was used, the position of CG lineage was unresolved with respect to *Trachypithecus* and *Semnopithecus*. Whereas, in the result from the ASTRAL tree, *Semnopithecus* was sister to *Trachypithecus* with the exclusion of CG lineage with low branch support (PP = 0.42). (Figure 3).

3.3. Divergence Dating

Molecular dates are shown in Fig. 4. These age estimates show the split of African and Asian colobines to be 9.53 my (million years) with credible interval (CI) of 12.66 – 7.09my. Within the Asian colobines, the odd-nosed monkeys are estimated to have diverged from the langurs in late Miocene between 9.75 and 6.7mya. The split between *Semnopithecus* and *Trachypithecus* is estimated to be between 5.2 and 2.7mya. The date estimates in this study are younger than estimated in an older study using nuclear concatenated dataset (Perelman et al., 2011). Divergence of CG lineage from the rest of *Trachypithecus* occurred 2.5mya with CI of 4.0 -1.2 my. The ESS values for all the parameters were >>200 for the two independent BEAST runs suggesting stationarity.

4. Discussion

Previous studies that have attempted to resolve the phylogenetic placement of CG lineage were based on single nuclear and / or mitochondrial markers (Karanth et al., 2008; Osterholtz et al., 2008; Wangchuk et al., 2008). Among these studies, Karanth et al. (2008) and Osterholtz et al. (2008) reported incongruence between nuclear and mitochondrial markers which was taken as evidence for hybrid origin of CG lineage. However, the study by Wang et al. (2015) suggested that this incongruence might be due to the use of numts instead of true mitochondrial

sequences. In the present study, we used multiple nuclear and mitochondrial markers to resolve the phylogenetic position of CG lineage. Additionally, we also implemented coalescent based species tree building methods to better understand the origin of CG lineage (Fig. 3).

Phylogenetic tree obtained from the nuclear dataset showed the CG lineage to be monophyletic. However, within CG lineage, neither the capped langur nor the golden langur was found to be monophyletic. This could be because these are recently diverged taxa and the nuclear DNA does not contain enough information to separate out these two lineages. The relationship within CG lineage can further be resolved by using fast evolving nuclear markers, such as microsatellites. Additionally, adding more samples to the phylogeny might also help resolve the relationship within the CG lineage.

The nuclear as well as the mitochondrial concatenated analyses placed CG lineage sister to *Trachypithecus* with high support. However, the nuclear species trees supported conflicting topologies. Among them only SVDquartet analysis (without taxon partitions) recovered the above relationship but with very low support (Fig 3, A). Whereas ASTRAL and SVDquartet (with taxon partitions) supported alternate scenarios again with low support. Thus, the species trees are unable to resolve the relationship between *Trachypithecus*, *Semnopithecus* and CG lineage. The low support for relevant nodes in the species trees is suggestive of different evolutionary histories of these nuclear markers due to processes like incomplete lineage sorting. Whereas the high support for CG lineage + *Trachypithecus* monophyly in the nuclear concatenated analysis might be due to few highly variable markers overwhelming the dataset. Additionally, under conditions of low ILS, the species trees are identical to trees based on concatenated analysis (Chou et al., 2015) which is not the case with our results. However, it must be noted that multispecies coalescent models incorporate ILS in species tree estimation (Chou et al., 2015) but these methods cannot distinguish between hybridisation and ILS (Kubatko, 2009).

Collectively, these results suggest that the phylogenetic position of CG lineage remains unresolved. The convoluted evolutionary history of the CG nuclear makers, we believe, is due to hybridization rather than ILS. There are three additional lines of evidence that support this scenario. First the numts analyses by Karanth (2008) and Wang et al. (2015) strongly support past hybridization between *Semnopithecus* and *Trachypithecus*. Secondly, the CG lineage is distributed in an area that abuts the eastern and western distributional limits of *Semnopithecus* and *Trachypithecus* respectively. Recent discovery of the late Pliocene fossil of *Semnopithecus*

gwebinensis from central Myanmar (Takai et al., 2015) suggests that *Semnopithecus* langurs were distributed till central Myanmar in late Pliocene, whereas today the eastern most distribution of *Semnopithecus* is Bangladesh and Bhutan. This implies that there was an overlap in the distribution of *Semnopithecus* and *Trachypithecus* in the past. Thirdly, the mean body weight (Wang et al., 2015) and body size and skull measurements of CG lineage is intermediate between *Semnopithecus* and *Trachypithecus* (Table 6). Presence of morphologically intermediate forms can be an indication of possible hybridization (Zinner et al., 2011). Such intermediate forms have been seen in other primates like baboons where hybridization has been reported (Phillips-Conroy and Jolly, 1981). Wang et al. (2015) suggest that this hybridization could have been initiated by the bigger *Semnopithecus* males which have an advantage over the smaller *Trachypithecus* males for access to *Trachypithecus* females. The genetic data is consistent with this scenario wherein the mtDNA of CG lineage is related to *Trachypithecus* whereas the nuclear DNA appears to be of both *Semnopithecus* and *Trachypithecus* origin. Our dating analysis (Fig. 4) shows the divergence between *Semnopithecus* and the *Trachypithecus* + CG clade to be around early to mid-Pliocene (3.85 mya; CI – 5.24 mya to 2.73 mya) and the divergence between *Trachypithecus* and CG lineage occurred in late Pliocene (2.55 mya; CI – 4.04 mya to 1.22 mya). Thus, this hybridisation event between lineages of *Semnopithecus* and *Trachypithecus* might have occurred during the Pliocene to early Pleistocene period (5.24 mya to 1.22 mya).

5. Conclusion

Multiple lines of evidence suggest that the CG lineage has evolved as a result of hybridisation between *Semnopithecus* and *Trachypithecus* lineages. The genetic evidence for the reticulate evolution of CG lineage needs to be further fortified with more nuclear markers. Nevertheless, our study suggests that the CG lineage has a unique evolutionary history. Therefore, unlike most of the taxonomic schemes that place CG lineage in the genus *Trachypithecus*, we recommend that this lineage should be assign to a new genus. Erecting a new genus will also warrant a revisit of the conservation status of capped and golden langurs (Karanth, 2010).

Acknowledgments

We thank the Department of Biotechnology, Govt. of India for the funding to carry out this research. We also thank the Ministry of Environment Forest and Climate Change for the funds that covered a part of fieldwork for this study. The forest department of Assam provided the permits to collect samples. We thank Guwahati zoo for allowing us to collect the samples. Dr.

Senthil Kumar and Dr. H.T. Lalremsanga from Mizoram University for providing samples from Tripura zoo and Mizoram zoo. Dr. G. Umapathy from LaCONES, CCMB for help in sample collection from Hyderabad zoo. KA would like to thank Mr. Rajani Deka and Mr. Firoz Ahmed for support during field work. Aniruddha Datta-Roy and V. Deepak for comments on the manuscript. Bhavani from CES for help in accounting.

References

Arnold, M. L., Meyer, A., 2006. Natural hybridization in primates: One evolutionary mechanism. Zoology. 109, 261 – 276.

Ashalakshmi, N. C., Nag, K. S. C., Karanth, K. P., 2014. Molecules support morphology: species status of South Indian populations of the widely distributed Hanuman langur. Conservation Genetics. 16, 43–58. https://doi.org/10.1007/s10592-014-0638-4.

Bensasson, D., Zhang, D. X., Hartl, D. L., Hewitt, G. M., 2001. Mitochondrial pseudogenes: Evolution's misplaced witnesses. Trends in Ecology and Evolution. 16, 314–321. https://doi.org/10.1016/S0169-5347(01)02151-6.

Blanchard, J. L., Lynch, M., 2000. Organellar genes why do they end up in the nucleus? Trends in Genetics. 16, 315–320. https://doi.org/10.1016/S0168-9525(00)02053-9.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., … Drummond, A. J., 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Computational Biology. 10, 1–6. https://doi.org/10.1371/journal.pcbi.1003537.

Brandon-Jones, D. 1996. The Asian Colobinae (Mammalia:Cercopithecidae) as indicators of Quaternary climatic change. Biological Journal of Linnean Society. 59, 327-350.

Brandon-Jones, D., Eudey, A. A., Geissmann, T., Groves, C. P., Melnick, D. J., Morales, J. C., … Stewart, C.-B. B., 2004. Asian Primate Classification. International Journal of Primatology. 25, 97–164. https://doi.org/10.1023/B:IJOP.0000014647.18720.32.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., Roychoudhury, A., 2012. Inferring Species Trees Directly from Biallelic Genetic Markers: Bypassing Gene Trees in a Full Coalescent Analysis. Mol. Biol. Evol. 29, 1917–1932. https://doi.org/10.1093/molbev/mss086.

Buckley, T. R., Cordeiro, M., Marshall, D. C., Simon, C., 2006. Differentiating between Hypotheses of Lineage Sorting and Introgression in New Zealand Alpine Cicadas (*Maoricicada* Dugdale). Syst. Biol. 55, 411–425.

Charlesworth, B., 2009. Effective population size and patterns of molecular evolution and variation. Nature Reviews Genetics. 10, 195–205. https://doi.org/10.1038/nrg2526.

Chifman, J., Kubatko, L., 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics. 30, 3317–3324. https://doi.org/10.1093/bioinformatics/btu530.

Chivers, D. J., Hladik, C. M., 1980. Morphology of the gastrointestinal tract in primates: Comparisons with other mammals in relation to diet. Journal of Morphology. 166, 337–386. https://doi.org/10.1002/jmor.1051660306.

Chou, J., Gupta, A., Yaduvanshi, S., Davidson, R., Nute, M., Mirarab, S., Warnow, T., 2015.

A comparative study of SVDquartets and other coalescent-based species tree estimation methods. BMC Genomics. 16, 2–11. https://doi.org/10.1186/1471-2164-16-S10-S2.

Collura, R. V., Stewart, C.-B., 1995. Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. Nature. 378, 485–489. Retrieved from http://dx.doi.org/10.1038/378485a0.

Davenport, T. R. B., Stanley, W. T., Sargis, E. J., De Luca, D. W., Mpunga, N. E., Machaga, S. J., Olson, L. E., 2006. A new genus of African monkey, *Rungwecebus*: morphology, ecology, and molecular phylogenetics. Science (New York, N.Y.). 312, 1378–1381. https://doi.org/10.1126/science.1125631.

Davies, G., Oates, J., 1994. Colobine Monkeys: Their Ecology, Behaviour and Evolution. Cambridge University Press, New York.

Degnan James H., Rosenberg, N. A., 2006. Discordance of species trees with their most likely gene trees: A unifying principle. PLoS Genetics. 3, 762–768. https://doi.org/10.1093/molbev/mst160.

Doolittle, F. W., 1998. You are what you eat: A gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends in Genetics. 14, 307–311. https://doi.org/10.1016/S0168-9525(98)01494-2.

Doyle, J. J., 1992. Gene Trees and Species Trees: Molecular Systematics as One-Character Taxonomy. Systematic Botany. 17, 144–163.

Dunbar, R. I. M., Dunbar, P., 1974. On hybridization between *Theropithecus gelada* and *Papio anubis* in the wild. Journal of Human Evolution. 3, 187–192. https://doi.org/10.1016/0047-2484(74)90176-6.

Edgar, R. C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 32, 1792–1797. https://doi.org/10.1093/nar/gkh340.

Edwards, S. V., Liu, L., Pearl, D. K., 2007. High-resolution species trees without concatenation. Proceedings of the National Academy of Sciences. 104, 5936–5941. https://doi.org/10.1073/pnas.0607004104.

Gauthier, O., Lapointe, F., 2007. Hybrids and Phylogenetics Revisited : A Statistical Test of Hybridization Using Quartets. Systematic Botany. 32, 8–15.

Groves, C. P., 1993. Primates: Cercopithecidae: Colobinae, pp. 269-273 in Mammal Species of the World. A Taxonomic and Geographic Reference, D.E. Wilson and D.M. Reedes, eds. Washington DC: Smithsonian Institution Press.

Groves, C. P., 2001. Primate Taxonomy. Smithsonian Institution Press, Washington.

Heled, J., Bouckaert, R., Drummond, A. J., Xie, W., 2013. * BEAST in BEAST 2 . 0 Estimating

Species Trees from Multilocus Data. http://journals.plos.org/ploscompbiol/article/file?type=supplementary&id=info:doi/10.1371/journal.pcbi.1003537.s006.

Heled, J., Drummond, A. J., 2010. Bayesian Inference of Species Trees from Multilocus Data. Mol. Biol. Evol. 27, 570–580. https://doi.org/10.1093/molbev/msp274.

Jin, G., Nakhleh, L., Snir, S., Tuller, T., 2006. Maximum likelihood of phylogenetic networks. Bioinformatics. 22, 2604–2611. https://doi.org/10.1093/bioinformatics/btl452.

Joly, S., McLenachan, A.P., Lockhart, P. J., 2009. A Statistical Approach for Distinguishing Hybridization and Incomplete Lineage Sorting. The American Naturalists. 174, 54–70.

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

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

Karanth, K. P., 2010. Molecular systematics and conservation of the langurs and leaf monkeys of South Asia. Journal of Genetics. 89, 393–399. https://doi.org/10.1007/s12041-010-0057-3.

Khajuria H. A new langur (Primates: Colobidae) from Goalpara District, Assam. Annals and Magazine of Natural History. 1956; 12(9):86–8.

Kishore, R., Reef Hardy, W., Anderson, V. J., Sanchez, N. A., Buoncristiani, M. R., 2006. Optimization of DNA extraction from low-yield and degraded samples using the BioRobot® EZ1 and BioRobot® M48. Journal of Forensic Sciences. 51, 1055–1061. https://doi.org/10.1111/j.1556-4029.2006.00204.x.

Kubatko, L. S., 2009. Identifying Hybridization Events in the Presence of Coalescence via Model Selection. Syst. Biol. 58, 478–488. https://doi.org/DOI:10.1093/sysbio/syp055.

Kubatko, L. S., Carstens, B. C., Knowles, L. L., 2009. STEM: Species tree estimation using maximum likelihood for gene trees under coalescence. Bioinformatics. 25, 971–973. https://doi.org/10.1093/bioinformatics/btp079.

Kumar, A., Solanki, G. S., 2008. Population Status and Conservation of Capped Langurs (*Trachypithecus pileatus*) in and around Pakke Wildlife Sanctuary, Arunachal Pradesh, India. Primate Conservation. 23, 97 – 105.

Lanfear, R., Calcott, B., Ho, S. Y. W., Guindon, S., 2012. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29, 1695–1701. https://doi.org/10.1093/molbev/mss020.

Leache, A. D., Rannala, B., 2011. The Accuracy of Species Tree Estimation under Simulation : A Comparison of Methods. Syst. Biol. 60, 126–137.

Liu, L., Pearl, D. K., 2007. Species Trees from Gene Trees : Reconstructing Bayesian Posterior Distributions of a Species Phylogeny Using Estimated Gene Tree Distributions. Syst. Biol. 56, 504–514.

Liu, L., Yu, L., Edwards, S. V., 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evolutionary Biology. 10:302.

Maddison, W., 1997. Gene Trees in Species Trees. Syst. Biol. 46, 523–536. https://doi.org/10.2307/2413694.

Mallet, J., 2005. Hybridization as an invasion of the genome. Trends in Ecology and Evolution. 20, 229–237. https://doi.org/10.1016/j.tree.2005.02.010.

Mallet, J., 2007. Hybrid speciation. Nature, 446, 279–283.

Meng, C., Kubatko, L. S., 2009. Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: A model. Theoretical Population Biology. 75, 35–45. https://doi.org/10.1016/j.tpb.2008.10.004.

Mirarab, S., Bayzid, S. M., Boussau, B., Warnow, T., 2014. Statistical binning enables an accurate coalescent-based estimation of the avian tree. Science. 346, 12504631–12504639. https://doi.org/10.1126/science.1250463.

Mirarab, S., Warnow, T., 2015. ASTRAL-II : coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics. 31, 44–52. https://doi.org/10.1093/bioinformatics/btv234.

Mondol, S., Ullas Karanth, K., Samba Kumar, N., Gopalaswamy, A. M., Andheria, A., Ramakrishnan, U., 2009. Evaluation of non-invasive genetic sampling methods for estimating tiger population size. Biological Conservation. 142, 2350–2360. https://doi.org/10.1016/j.biocon.2009.05.014.

Mossel, E., Roch, S., 2008. Incomplete lineage sorting: Consistent phylogeny estimation from multiple loci. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 7, 166–171. https://doi.org/10.1109/TCBB.2008.66.

Nagel, U., 1971. Social organization in baboon hybrid zone. Proceeding of the Third International Congress on Primatology (Zurich). 3, 48–57.

Olave, M., Avila, L. J., Sites, J. W., Morando, M., 2017. Detecting hybridization by likelihood calculation of gene tree extra lineages given explicit models. Methods in Ecology and Evolution. 1–13. https://doi.org/10.1111/2041-210X.12846.

Osterholz, M., Walter, L., Roos, C., 2008. Phylogenetic position of the langur genera *Semnopithecus* and *Trachypithecus* among Asian colobines, and genus affiliations of their species groups. BMC Evolutionary Biology. 8, 1–12. https://doi.org/10.1186/1471-2148-8-58.

Pan, R., Groves, C., Oxnard, C. 2004. Relationships between the Fossil Colobine *Mesopithecus pentelicus* and Extant Cercopithecoids, Based on Dental Metrics. American Journal of Primatology. 62, 287–299. https://doi.org/10.1002/ajp.20022.

Pastorini, J., Zaramody, A., Curtis, D. J., Nievergelt, C. M., Mundy, N. I., 2009. Genetic analysis of hybridization and introgression between wild mongoose and brown lemurs. BMC Evolutionary Biology. 9, 1–13. https://doi.org/10.1186/1471-2148-9-32.

Perelman, P., Johnson, W. E., Roos, C., Seuánez, H. N., Horvath, J. E., Moreira, M. A. M., … Pecon-Slattery, J., 2011. A Molecular Phylogeny of Living Primates. PLoS Genetics. 7, e1001342. https://doi.org/10.1371/journal.pgen.1001342

Pocock, R. I. (1939). Primates and carnivora (in part) (pp. 83–140) in The Fauna of British India, Vol. I. London: Taylor and Francis

Posada, D., 2002. Evaluation of Methods for Detecting Recombination from DNA Sequences: Empirical Data. Mol. Biol. Evol. 19, 708–717. https://doi.org/10.1093/oxfordjournals.molbev.a004129.

Raaum, R. L., Sterner, K. N., Noviello, C. M., Stewart, C.-B., Disotell, T. R., 2005. Catarrhine primate divergence dates estimated from complete mitochondrial genomes: concordance with fossil and nuclear DNA evidence. Journal of Human Evolution. 48, 237–257. https://doi.org/10.1016/j.jhevol.2004.11.007.

Rabarivola, C., Meyers, D., Rumpler, Y., 1991. Distribution and morphological characters of intermediate forms between the black lemur (*Eulemur macaco macaco*) and the Sclater's lemur (*E. m. flavifrons*). Primates. 32, 269–273. https://doi.org/10.1007/BF02381186.

Ram, M. S., Kittur, S. M., Biswas, J., Nag, S., Shil, J., Umapathy, G., 2016. Genetic Diversity and Structure among Isolated Populations of the Endangered Gees Golden Langur in Assam , India. PLoS ONE. 11, e0161866. https://doi.org/10.1371/journal.pone.0161866

Reaz, R., Bayzid, M. S., Rahman, M. S., 2014. Accurate phylogenetic tree reconstruction from quartets: A heuristic approach. PLoS ONE. 9, e104008.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., … Huelsenbeck, J. P., 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542. https://doi.org/10.1093/sysbio/sys029.

Roos, C., Zinner, D., Kubatko, L. S., Schwarz, C., Yang, M., Meyer, D., … Ziegler, T., 2011. Nuclear versus mitochondrial DNA : evidence for hybridization in colobine monkeys. BMC Evolutionary Biology. 11, 1–13. https://doi.org/10.1186/1471-2148-11-77.

Silvestro, D., Michalak, I., 2012. RaxmlGUI: A graphical front-end for RAxML. Organisms Diversity and Evolution. 12, 335–337. https://doi.org/10.1007/s13127-011-0056-0.

Srivastava, A., Mohnot, S. M., 2001. Distribution, Conservation Status and Priorities for Primates in Northeast India. Envis Bulletin: Wildlife and Protected Areas. 1, 102 – 108.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 22, 2688–2690. https://doi.org/10.1093/bioinformatics/btl446.

Takai, M., Nishioka, Y., Thaung-Htike, Maung, M., Khaing, K., Zin-Maung-Maung-Thein, … Egi, N., 2016. Late Pliocene *Semnopithecus* fossils from central Myanmar: rethinking of the evolutionary history of cercopithecid monkeys in Southeast Asia. Historical Biology. 28, 172–188. https://doi.org/10.1080/08912963.2015.1018018.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol, 28, 2731–2739. https://doi.org/10.1093/molbev/msr121.

Than, C., Ruths, D., Innan, H., Nakhleh, L., 2007. Confounding Factors in HGT Detection: Statistical Error, Coalescent Effects, and Multiple Solutions. Journal of Computational Biology. 14, 517–535. https://doi.org/10.1089/cmb.2007.A010.

Vasey, N., Tattersall, I., 2002. Do ruffed lemurs form a hybrid zone? Distribution and discovery of *Varecia*, with systematic and conservation implications. American Museum Novitates. 3376, 1–26.

Wang, B., Zhou, X., Shi, F., Liu, Z., Roos, C., Garber, P. A., … Pan, H., 2015. Full-length Numt analysis provides evidence for hybridization between the Asian colobine genera *Trachypithecus* and *Semnopithecus*. American Journal of Primatology. 77, 901–910. https://doi.org/10.1002/ajp.22419.

Wangchuk, T., Inouye, D. W., Hare, M. P., 2008. The Emergence of an Endangered Species: Evolution and Phylogeny of the *Trachypithecus geei* of Bhutan. International Journal of Primatology. 29, 565 – 582.

Xu, S., 2000. Phylogenetic Analysis Under Reticulate Evolution. Mol. Biol. Evol. 17, 897–907. http://mbe.oxfordjournals.org/cgi/content/abstract/17/6/897.

Yu, Y., Dong, J., Liu, K. J., Nakhleh, L., 2014. Maximum likelihood inference of reticulate

evolutionary histories. Proceedings of the National Academy of Sciences. 111, 16448–16453. https://doi.org/10.1073/pnas.1407950111.

Zhang, Y. Z., Wang, S., Quan, G. Q., 1981. On the geographical distribution of primates in China. Journal of Human Evolution, 10, 219–226. https://doi.org/10.1016/S0047-2484(81)80060-7.

Zinner, D., Arnold, M. L., Roos, C., 2011. The strange blood: Natural hybridization in primates. Evolutionary Anthropology. 20, 96 – 103.

605 **Tables**

606 **Table 1:** Nuclear markers used in this study with primers and source. * indicates primer pairs that have not been redesigned.

Marker	Name	Product Size (bp)	Forward Primer (5' to 3')	Reverse Primer (5' to 3')	Source
ABCA1	ATP-binding cassette, sub-family A (ABC1), member 1	400	CCTCCATCTTTTCAGCTCTACCTAC	GGGTTCTAGAGCCTCAATCACAG	Perelman et al., 2011
BCHE	butyrylcholinesterase	244	TGGCTTCAGCAAAGACAACA	GGCATTATTTCCCCATTCTG	Perelman et al., 2011
BDNF*	brain-derived neurotrophic factor	570	CATCCTTTTCCTTACTATGGTT	TTCCAGTGCCTTTTGTCTATG	Perelman et al., 2011
DMRT1*	doublesex and mab-3 related transcription factor 1	560	ATCCCTTGTTCTGAGTGCCA	ACATTGCAAAGACCCCTGAC	Perelman et al., 2011
ERC2	ELKS/RAB6-interacting/CAST family member 2	400	TGTCTTAGGTCTGCTTGTCACAT	ATTCGGTCTGCAAACTCAGG	Perelman et al., 2011
FAM123B	family with sequence similarity 123B	388	TGGTGGGAAAAGCTGAGGTA	AACGGCTAGTGACCATCCAG	Perelman et al., 2011
FES*	feline sarcoma oncogene	500	GGGGAACTTTGGCGAAGTGTT	TCCATGACGATGTAGATGGG	Perelman et al., 2011
LZM	lysozyme	500	GAATCAGTAGATCAATACACAGT	CCTAGCACTAGAACAGTACA	Karanth et al., 2008
MAPKAP	mitogen-activated protein kinase associated protein 1	400	TCAAGTGATAACGGCTGTAGCTT	GAGAGATAGAGCTCCTGACATCAA	Perelman et al., 2011

Table 2: Eight genes that are amplified in the long mitochondrial DNA (mtDNA) fragment (5933 bp).

Gene	Name	Length (bp)
COX1	Cytochrome c oxidase subunit 1	1546
COX2	Cytochrome c oxidase subunit 2	688
ATP8	ATP synthase protein 8	211
ATP6	ATP synthase protein 6	681
COX3	Cytochrome c oxidase subunit 3	785
ND3	NADH dehydrogenase 3	346
ND4L	NADH-ubiquinone oxidoreductase chain 4L	297
ND4	NADH dehydrogenase 4	1379

Table 3: List of specimens used in the mtDNA phylogeny.

Name	GenBank Accession numbers
Trachypithecus shortridgei	HQ149048
Trachypithecus germaini	HQ149047
Trachypithecus hatinhensis	HQ149046
Trachypithecus cristatus	KJ174503
Trachypithecus obscurus	AY863425
Trachypithecus francoisi	KJ174502
Trachypithecus shortridgei (wang2015)	KP834334
Trachypithecus pileatus	KF680163
Semnopithecus entellus	DQ355297
Semnopithecus johnii	HQ149050
Semnopithecus vetulus	HQ149049
Presbytis melalophos	DQ355299
‡*Rhinopithecus roxellana*	JQ821835
Rhinopithecus avunculus	HM125578
‡*Pygathrix roxellana*	DQ355300
Pygathrix nigripes	JQ821840
Pygathrix nemeaus	DQ355302
Nasalis larvatus	JF293094
Nasalis concolor	JF293095
Colobus guereza	AY863427
†Capped langur (CES 11/299)	-

† - Sequenced in this study; ‡ - *R. roxellana* was previously called *Py. Roxellana* (Brandon-Jones 1996, Groves 1993)

Table 4: List of specimens used in the nuclear data set.

		Genes								
Sr. No.	**Species**	**ABCA1**	**BCHE**	**BDNF**	**DMRT1**	**ERC2**	**FAM123B**	**FES**	**LZM**	**MAPKAP**
1	*S. e. entellus*	HM765389	HM764193	HM763811	HM762646	HM762230	HM762051	HM761726	AF294862	HM760733
3 †	*S. hypoleucos* CES08/333									
4	*S. vetulus*	HM765412	HM764189	HM763819	HM762672	HM762231	HM762061	HM761734	AH004926	HM760757
5	*T. phayrei*	HM765410	HM764161	HM763818	HM762670	HM762237	HM762060	HM761733	AF294867	HM760755
6	*T. obscurus*	HM765409	HM764066	HM763817	HM762669	HM762248	HM762059	HM761732	U76917	HM760754
7	*P.melalophos*	HM765375	HM764081	HM763806	HM762629	HM762234	HM762046	HM761722	*****	HM760717
8	*R. brelichi*	HM765387	HM764153	HM763810	HM762642	HM762235	HM762050	HM761725	*****	HM760729
9	*N. larvatus*	HM765362	HM764154	HM763800	HM762616	HM762243	HM762040	*****	U76945	HM760705
10	*Py.nigripes*	HM765377	HM764167	HM763809	HM762631	HM762219	HM762049	*****	*****	HM760719
11	*Py. nemaeus*	HM765376	HM764159	HM763808	HM762630	HM762229	HM762048	HM761724	U76941	HM760718
12	*Py. cinerea*	HM765369	HM764165	HM763807	HM762623	HM762251	HM762047	HM761723	*****	HM760711
13	*C. guereza*	HM765292	HM764095	HM763784	HM762539	HM762317	HM762019	HM761695	U76916	HM760637
14 †	*S. hypoleucos* CES09/401									
15 †	Capped langur CES 12/300									

16 †	Capped langur CES 11/299	
17 †	Golden langur CES 12/303	
18 †	Golden langur CES12/305	
19 †	Capped langur CES12/308	
20 †	Capped langur CES12/309	

† = Sequences from this study; ***** = no sequence available.

Table 5: Partition schemes and best-fit models of sequence evolution selected in PartitionFinder for reconstructions of phylogeny using Bayesian inference (MrBayes), maximum likelihood (RAxML) and divergence dating in BEAST. Codon positions are denoted by cp1, cp2 and cp3. E1 = Exon 1; E2 = Exon 2; **(A)** = Nuclear markers; **(B)** = Mitochondrial markers.

(A)

	Partition no.	**Partition name**	**Best substitution model**	
			MrBayes	**RAxML**
	1	ABCA1_intron	HKY	GTR + G
	2	BCHE cp1	HKY + I	GTR + G
	3	BCHE cp2	F81	GTR + G
	4	BCHE cp3	HKY	GTR + G
	5	BDNF cp1	F81	GTR + G
	6	BDNF cp2	JC69	GTR + G
	7	BDNF cp3	HKY + I	GTR + G
	8	DMRT1_intron	HKY	GTR + G
	9	ERC2_intron	HKY	GTR + G
	10	FAM123B cp1	HKY + I	GTR + G
	11	FAM123B cp2	HKY + I	GTR + G
	12	FAM123B cp3	HKY + I	GTR + G
	13	FESE1 cp1	HKY + I	GTR + G
	14	FESE1 cp2	F81	GTR + G
	15	FESE1 cp3	HKY + I	GTR + G
	16	FES_intron	HKY + I	GTR + G
	17	FESE2 cp1	HKY + I	GTR + G
	18	FESE2 cp2	JC69	GTR + G
	19	FESE2 cp3	HKY	GTR + G
	20	LZM	HKY + I	GTR + G
	21	MAPKAP_intron	HKY	GTR + G
BEAST	1	ABCA1	HKY+ I	
	2	BCHE	HKY+ I	

3	BDNF	HKY
4	DMRT1	HKY+ I
5	ERC2	HKY+ I
6	FAM123B	HKY + I
7	FES	HKY
8	LZM	HKY
9	MAPKAP	HKY+ I

(B)

Partition no.	**Partition name**	**Best substitution model**	
		MrBayes	**RAxML**
1	COX1_cp1	HKY+I+G	GTR+I+G
2	COX1_cp2	HKY+I+G	GTR+I+G
3	COX1_cp3	HKY+G	GTR+I+G
4	COX2_cp1	HKY+G	GTR+I+G
5	COX2_cp2	HKY+G	GTR+I+G
6	COX2_cp3	HKY+G	GTR+I+G
7	ATP8_cp1	HKY+I+G	GTR+I+G
8	ATP8_cp2	HKY+G	GTR+I+G
9	ATP8_cp3	HKY+I+G	GTR+I+G
10	ATP6_cp1	HKY+G	GTR+I+G
11	ATP6_cp2	HKY+I+G	GTR+I+G
12	ATP6_cp3	GTR+I	GTR+I+G
13	COX3_cp1	HKY+I+G	GTR+I+G
14	COX3_cp2	HKY+I+G	GTR+I+G
15	COX3_cp3	HKY+G	GTR+I+G
16	ND3_cp1	HKY+G	GTR+I+G
17	ND3_cp2	HKY+I+G	GTR+I+G
18	ND3_cp3	GTR+I	GTR+I+G
19	ND4L_cp1	HKY+G	GTR+I+G

20	ND4L_cp2	HKY+I+G	GTR+I+G
21	ND4L_cp3	GTR+I	GTR+I+G
22	ND4_cp1	HKY+I+G	GTR+I+G
23	ND4_cp2	HKY+G	GTR+I+G
24	ND4_cp3	HKY+I+G	GTR+I+G

Table 6: Skull measurements and body size measurements (Pocock 1939, Khajuria 1956).

Genus	Species	Skull measurments (in mm)			Body length (cm)	Tail length (cm)
		Total length	Condylobasal length	Zygomatic width		
Semnopithecus	*Semnopithecus schistaceus*	141	113	108	69.8	99
	Semnopithecus ajax	144	113	110	76.2	96.5
	Semnopithecus entellus	130	106	106	64.7	107.9
	Semnopithecus hypoleucos	120	91	92	68.5	109.2
	Semnopithecus priam	123	99	96	64.2	95.2
CG lineage	Capped langur	116	92	88	69.5	99
	Golden langur	110.5	88.1	89.1	72	90
Trachypithecus	*Trachypithecus phayrei*	106	87	80	57.1	78.7
	Trachypithecus obscurus	106	84	78	59.6	71.1

Appendices

Appendix A: Samples used in this study.

Sample ID	Species	Sample collected from	Location in wild	Sample Type
CES11/299	Capped langur	Hyderabad Zoo	Nagaland	Tissue
CES12/300	Capped langur	Hyderabad zoo	Unknown	Tissue
CES12/303	Golden langur	Guwahati Zoo	Kachugaon, Assam	Fecal
CES12/305	Golden langur	Guwahati Zoo	Bongaigaon, Assam	Fecal
CES12/308	Capped langur	Guwahati Zoo	Guwahati, Assam	Fecal
CES12/309	Capped langur	Guwahati Zoo	Amchang, Assam	Fecal

Appendix B: *A priori* species assignment in the respective genera for the dating analysis using starBEAST in BEAST v2.4.7.

***Lineage**	**Species**
CG lineage	CL_CES11/299
	GL_CES12/303
Colobus	*Colobus polykomos*
	Colobus guereza
Presbytis	*Presbytis comata*
	Presbytis melalophos
Pygathrix	*Pygathrix cinerea*
	Pygathrix nemaeus
Semnopithecus	*Semnopithecus entellus*
	Semnopithecus vetulus
Trachypithecus	*Trachypithecus obscurus*
	Trachypithecus phayrei

Figures

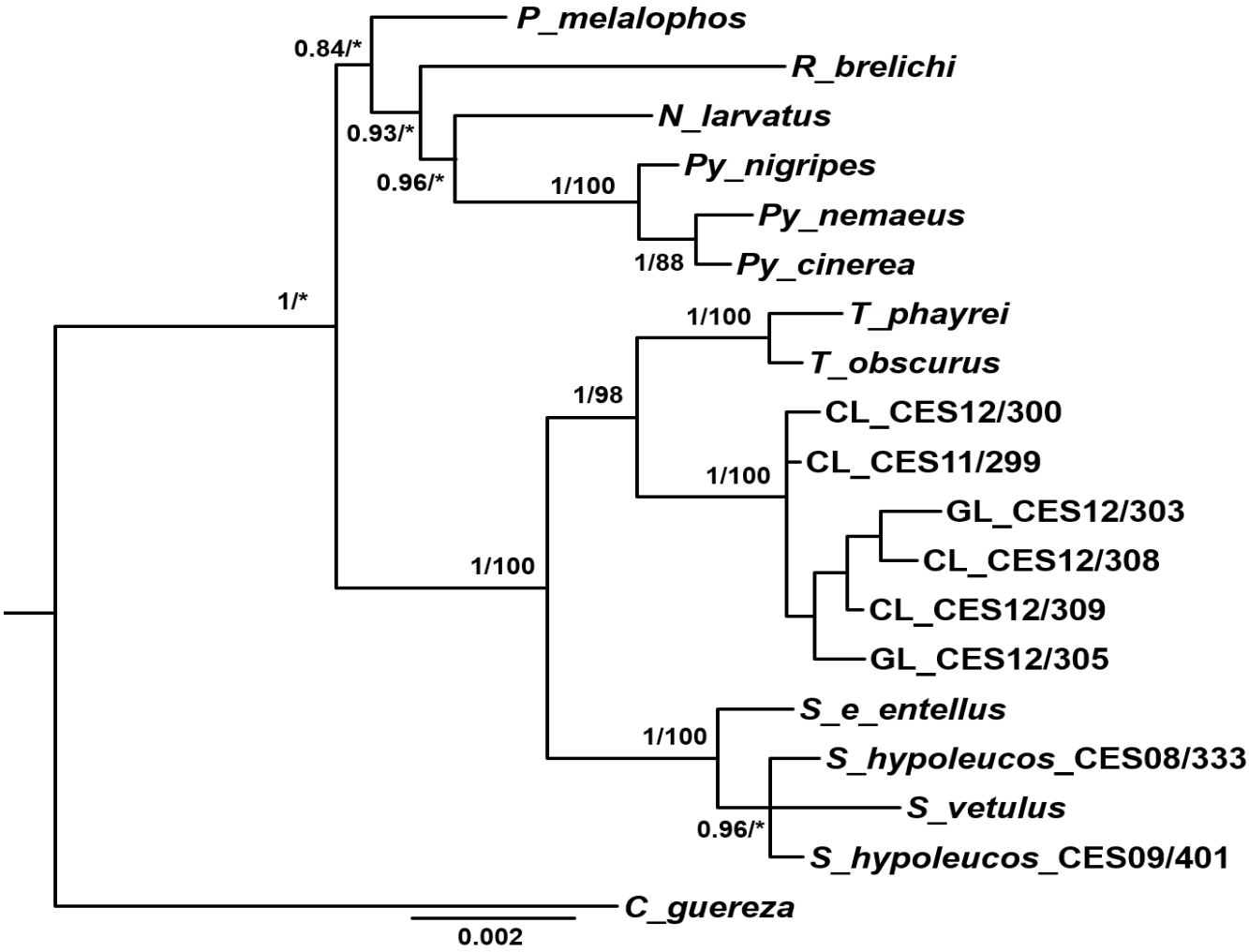

Fig. 1: Bayesian phylogeny of Asian colobines for the concatenated nuclear data set. The numbers on the nodes are the posterior probability values/Maximum-likelihood bootstrap support. Posterior probability and ML bootstrap support less than 0.75 and 75 respectively are not shown. (* indicates Posterior probability or ML bootstrap values less than 0.75 or 75 respectively for that particular node; CL = capped langur; GL = golden langur). Support for all the nodes within the CL and GL clade was below 0.75/75.

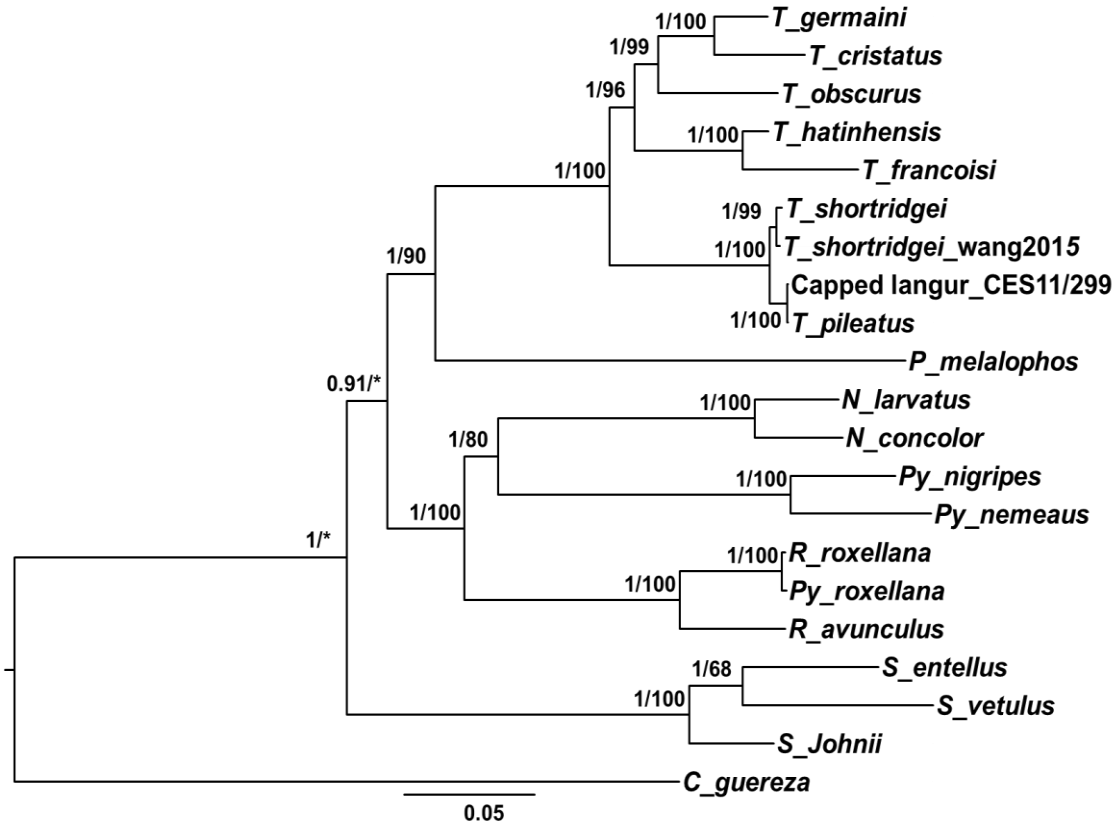

Fig. 2: Bayesian phylogeny of Asian colobines for the mitochondrial concatenated data set. The values at the nodes denote the posterior probability values/Maximum-likelihood bootstrap support. Posterior probability and ML bootstrap support below 0.75 and 75 respectively are not shown. (* indicates Posterior probability or ML bootstrap values less than 0.75 or 75 respectively for that particular node).

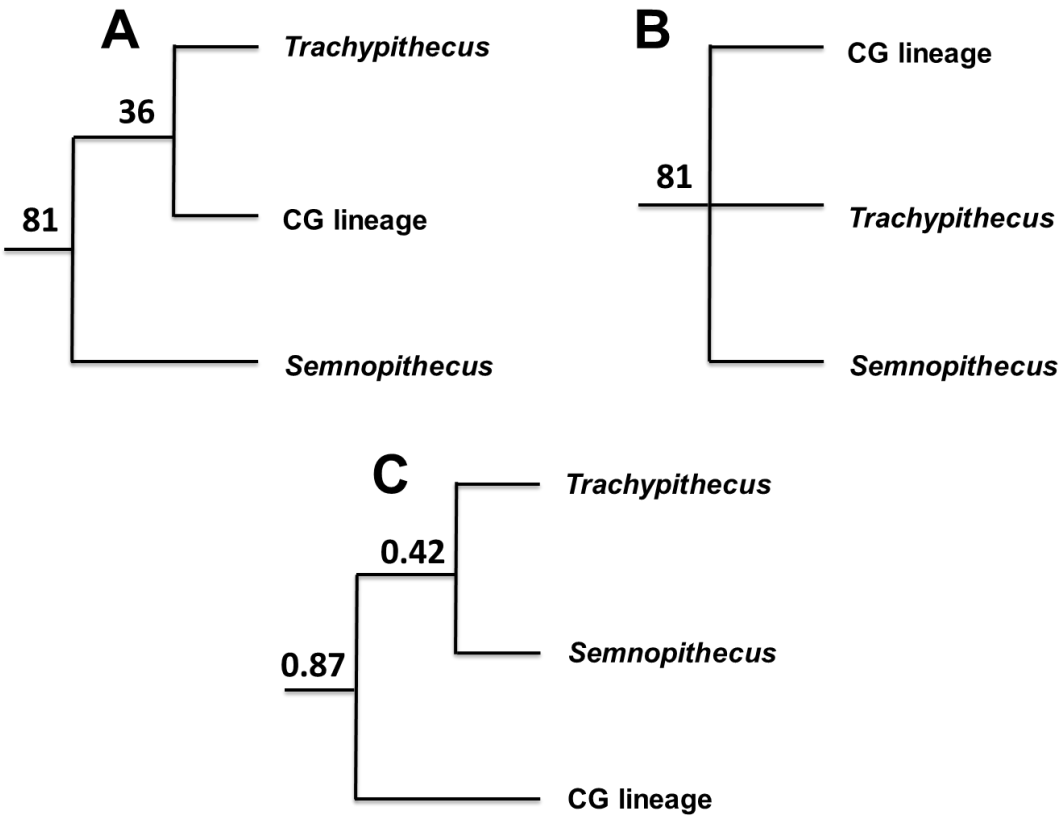

Fig. 3: Comparing the phylogenetic position of CG lineage with respect to *Semnopithecus* and *Trachypithecus*, from different analyses. (A) Species-tree topology from SVDq without taxa partitions (for details, see text), (B) Topology estimated with SVDq with taxa partitions (for details, see text), (C) Species-tree topology estimated with ASTRAL-II. Numbers at the node are support for that node (support values only for node of interest are shown). Panel A & B shows the BS values; panel C shows PP values.

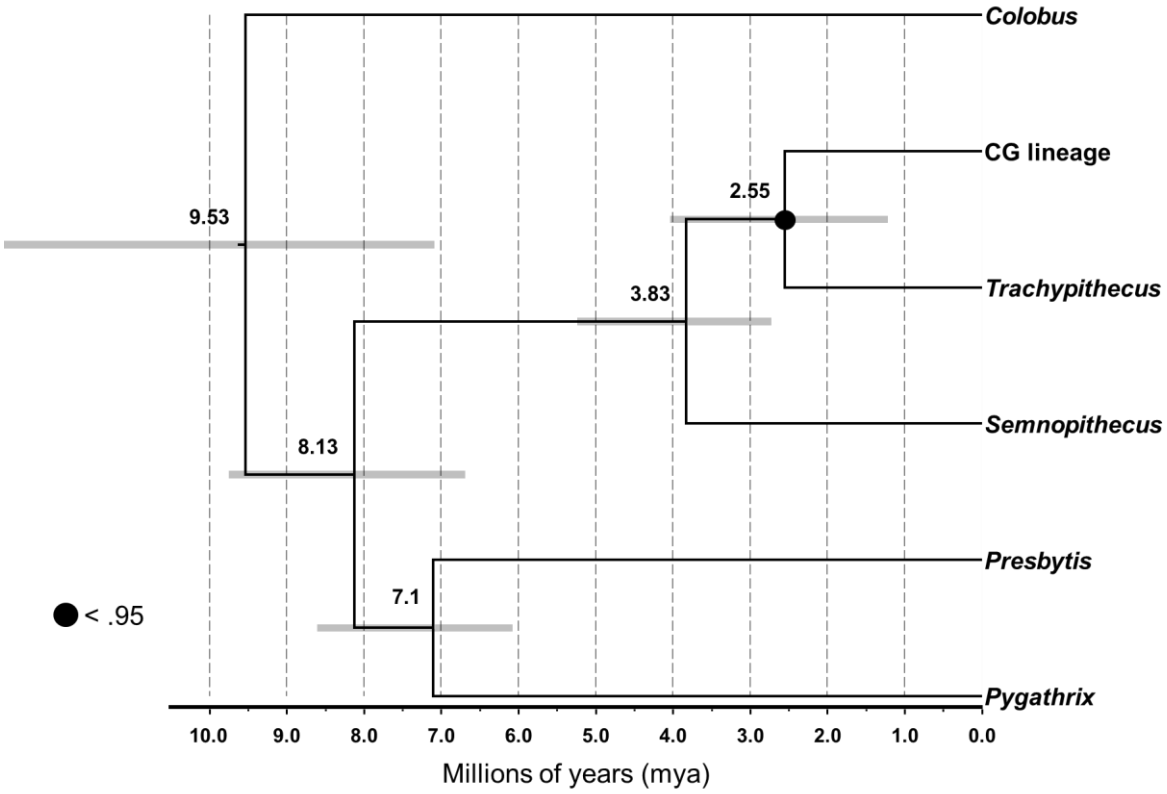

Fig. 4: Divergence time tree constructed using coalescent approach in *BEAST using the nuclear data set. The value above each node indicates the mean divergence time and the grey bars indicate the 95% credible intervals. PP support for all the nodes was above 0.95 (except for the one mentioned).