Short Communication

Mitochondrial and nuclear markers suggest Hanuman langur (Primates: Colobinae) polyphyly: Implications for their species status

K. Praveen Karanth a,c,*, Lalji Singh b, Caro-Beth Stewart c

a Centre for Ecological Sciences, Indian Institute of Science, Bangalore 560012, India
b Center for Cellular and Molecular Biology, Uppal Road, Hyderabad 500007, India
c Department of Biological Sciences, University at Albany, State University of New York, Albany, NY 12222, USA

1. Introduction

The Hanuman langur (Semnopithecus entellus, Subfamily Colobinae) is one of the most widely-distributed and morphologically variable species of colobe monkey (Newton, 1988). It is distributed throughout most parts of India and Sri Lanka, and is also found in parts of Pakistan, Nepal, and Bangladesh. According to Roonwal and Mohnot (1977), it is predominantly a deciduous woodland species. Roonwal (1984) recognizes two major forms of Hanuman langur based on tail morphology: the Northern type (here called as NT-Hanuman), which has a tail that loops forward, and the Southern type (here called as ST-Hanuman), which has a tail that loops backward (Fig. 1). The NT-Hanuman is found north of the Tapti and Godavari rivers, over the plains of northern India and to an altitude of 3000 m in the Himalayas. The ST-Hanuman is found south of the Tapti and Godavari rivers in southern India and in Sri Lanka (Roonwal, 1984). There is much disagreement in the literature regarding the species and subspecies status of various populations of the so-called Hanuman langurs. Most authors have considered the Hanuman langurs to be a single species, S. entellus, but have divided this species into as many as 14 (Pocock, 1939), 15 (Napier and Napier, 1967) and 16 (Roonwal and Mohnot, 1977) subspecies. Other authors have classified the Hanuman langurs into two (Brandon-Jones, 2004), four (Hill, 1939) and seven (Groves, 2001) distinct species. Hanuman langurs have also been extensively used as model system in a various biomedical (Nandi et al., 2003; Dube et al., 2004; Lohiya et al., 2005), ecological (Kurup, 1984; Kamilar et al., 2006), behavioral (Newton, 1988; Sterck, 1999) and evolutionary research (Karanth et al., 2008; Osterholz et al., 2008). Often these have been comparative studies wherein multiple populations of Hanuman langurs were studied from across the range. The decision on the nature of these comparative studies, i.e., inter vs. intra specific comparison, would depend on the species status of the various Hanuman langur populations. Thus for an appropriate interpretation of results from these studies, it is imperative that the taxonomic status of Hanuman langurs is resolved.

Within the overall range of the Hanuman langurs two other species of langurs are found, having more restricted distributions. These are Nilgiri langurs (S. johnii) found in the wet evergreen forests of south-west India and purple-faced langurs (S. vetulus) distributed in the wet zone of Sri Lanka (Fig. 1). Thus langur monkeys of the Indian subcontinent provide an interesting system to study speciation because there are several closely-related species found in different habitats and distributed in different geographical regions of this landmass.

Recent phylogenetic studies on the langurs of Asia suggested that the Hanuman langurs are polyphyletic with respect to Nilgiri and purple-faced langurs in their mitochondrial DNA (Karanth et al., 2008; Osterholz et al., 2008). These results are indicative of an interesting speciation model wherein certain ancestral populations of a
widely-distributed “proto” Hanuman langur might have diverged into purple-faced langurs in Sri Lanka and Nilgiri langurs in South India in the recent past. These diverged species might still share mitochondrial DNA (mtDNA) haplotypes or nuclear alleles with a subset of Hanuman langur populations resulting in their polyphyly. This speciation model has been reported for the *fascicularis* group of the macaques (genus *Macaca*) (Melnick and Hoelzer, 1993) as well as for Asian colobines (genus *Trachypithecus*) (Karanth et al., 2008).

The studies by Karanth et al. (2008) and Osterholz et al. (2008) were focused on resolving the phylogeny of Asian colobines and did not delve into the Hanuman langur polyphyly issue. Additionally, sampling of Hanuman, Nilgiri and purple-faced langurs in these studies were limited and derived mostly from zoos. To further investigate the Hanuman langur polyphyly issue, additional samples of this species were obtained from throughout its range. We have also used additional sequences of purple-faced and Nilgiri langurs downloaded from GenBank. Here we address two major questions. First, how are the various populations of Hanuman langurs related to each other and to the other langurs of the Indian subcontinent? Second, based on genetic data do the Hanuman langurs deserve being split into multiple species? To address these questions, two rapidly-evolving markers—the nuclear protamine 1 (*Prm1*) gene and the mitochondrial cytochrome *b* (*Cyt-b*) gene—were used.

### 2. Methods

Three kinds of samples—namely hair, blood, and muscle tissue—were collected from wild and captive animals. Samples of the widely-distributed Hanuman langurs were collected from five locations in the wild spanning the entire range of this species (Table 1 and Fig. 1). Blood samples were stored in digestion buffer (0.01 M Tris–HCl (pH 8), 0.01 M EDTA, 0.1 M NaCl, 10 μl 1 M DTT) in a 1:1 ratio, and tissue samples were stored in 95% ethanol. The methodology used for collecting these samples and for DNA extractions were the same as reported in Karanth et al. (2008).
2.1. PCR amplification of the Prm1 and Cyt-b gene

Protamines are arginine-rich proteins that replace histones and bind sperm DNA during spermatogenesis in vertebrates (Rooney and Zhang, 1999). The Prm1 gene has been used to resolve the phylogeny of primate (Retief and Dixon, 1993) and other mammalian groups (Retief et al., 1995; Krajewski et al., 1997; Blackett et al., 1999). In the case of Cyt-b gene, our studies have shown that this gene is highly variable among the langurs (Karanth et al., 2008). Therefore, we used these two markers to study the nuclear and mitochondrial genetic diversity of the langurs of the Indian subcontinent. The primers used for amplification, PCR conditions and sequencing strategy for Prm1 and Cyt-b genes were same as that in (Karanth et al., 2008).

All sequencing was performed using the Taq DyeDeoxy™ Terminator Cycle Sequencing Kit (Applied Biosystems), following the manufacturer’s instructions. Reactions were analyzed using either an ABI Model 373A or an ABI Model 377 Automated DNA Sequencer. Samples collected in India and Sri Lanka were sequenced in Dr. Lalji Singh’s laboratory. Sequences were edited using the program SeqEd (Applied Biosystems) and aligned manually in PAUP* (Swofford et al., 1996), and to estimate the transition–transversion rate ratio, gamma shape parameters, and base frequencies under the best supported model. These values, along with the selected model, were used to derive the maximum likelihood (ML) tree through a heuristic search with 10 replicates of the random addition and TBR branch swapping options in PAUP*. The above values were used again to derive a neighbor-joining tree through 10,000 bootstrap replications in PAUP*. In this dataset, transitions were empirically estimated to be around 11 times more frequent than transversions. To account for this difference in base substitution, an 11:1 weighting scheme was used. Published sequences of Southeast Asian leaf monkeys (Genus Trachypithecus) that are sister groups to Semnopithecus were also included for comparison and the African colobines were used to root these trees.

The program MacClade (Maddison and Maddison, 2000) was used to infer amino acid sequences and to construct a constraint tree, wherein representatives of each species were forced to be

2.2. Analysis of the Cyt-b gene sequences

A total of 11 Hanuman langur and four each of Nilgiri and purple-faced langur Cyt-b sequences were used to ascertain the evolutionary relationships between these species. The program MODELTEST (Posada and Crandall, 1998) was used to choose substitution models that best fit the dataset through hierarchical likelihood ratio tests (Swofford et al., 1996), and to estimate the transition–transversion rate ratio, gamma shape parameters, and base frequencies under the best supported model. These values, along with the selected model, were used to derive the maximum likelihood (ML) tree through a heuristic search with 10 replicates of the random addition and TBR branch swapping options in PAUP*, wherein a molecular clock was not enforced. The above values were used again to derive a neighbor-joining tree through 10,000 bootstrap replications in PAUP*. In this dataset, transitions were empirically estimated to be around 11 times more frequent than transversions. To account for this difference in base substitution, an 11:1 weighting scheme was used.

First, a branch-and-bound parsimony search was undertaken with ‘simple addition of sequences’ option. To ascertain support for various nodes, 1000 bootstrap replications were performed, using 10 replicates of the random addition option. For all these analyses, the incomplete regions of partial sequences were coded as missing data.

Published sequences of Southeast Asian leaf monkeys (Genus Trachypithecus) that are sister groups to Semnopithecus were also included for comparison and the African colobines were used to root these trees.
monophyletic. This constraint tree was used in PAUP* to derive a "species monophyly" tree as a null model for the data through a likelihood heuristic search. To examine support for the hypothesis of species monophyly, this constrained ML tree was compared with the unconstrained ML tree from above using the one-tailed Shimodaira–Hasegawa log-likelihood test (Shimodaira and Hasegawa, 1999) as implemented in PAUP*, using the re-sampling estimated log-likelihood (RELL) technique approximation with 10,000 bootstrap replications.

2.3. Analysis of the Prm1 gene sequences

The sequence based phylogenetic analysis of Prm1 gene did not resolve the evolutionary relationships between the langurs of the Indian subcontinent due to a lack of phylogenetically informative characters (Karanth et al., 2008). Nevertheless, this gene does exhibit high levels of allelic variation among the langurs of the Indian subcontinent. Therefore, we analyzed allelic data to determine if there were any fixed allelic differences between species or between populations of Hanuman langurs. Two different approaches were used to identify Prm1 alleles. First, the coding sequences from various samples were aligned and their amino acid sequences were inferred in the program MacClade (Maddison and Maddison, 2000). Alleles were classified on the basis of these amino acid sequences. For individuals that were heterozygous at this locus, the PCR products were cloned and multiple clones were sequenced to determine the sequence of the alleles. Heterozygous individuals were detected through the presence of multiple peaks at a site in the chromatograms. In the case of some Hanuman langurs, the 3′ non-coding region of this gene was found to have a tandem duplication of 40 bp (here called the 40 bp repeat) in some individuals. The presence or absence of this 40 bp repeat in a sample can be detected by electrophoresing the PCR product in a 2% agarose gel. Therefore, for many Hanuman langur samples this particular allele was identified on the basis of agarose gel electrophoresis. Four of these alleles with the 40 bp repeat were also sequenced to confirm the presence of the repeat and to determine if there were any variations in the sequence.

3. Results

3.1. The Cyt-b gene results

The program MODELTEST chose the HKY 85 (Hasegawa et al., 1985) + Γ substitution model with the following parameters: Base = (0.3231, 0.3018, 0.1111), TRatio = 11.0048, and Shape = 0.2472. The maximum likelihood search yielded a single tree (In L = −5599.5731) that is shown in Fig. 2A. In this tree, the langurs of the Indian subcontinent fall into three well-supported clades: clade 1, consisting of NT-Hanuman individuals; clade 2, consisting of ST-Hanuman individuals from South India together with the Nilgiri langur; and clade 3, consisting of ST-Hanuman individuals from Sri Lanka together with the purple-faced langur. Thus, the mtDNA of the so-called Hanuman langur is not monophyletic and encompasses mtDNA variation of two other species (purple-faced langur and Nilgiri langur). In addition, within the clades 2 and 3, the mtDNAs of the so-called Hanuman langurs are not monophyletic. For example, within clade 3, the ST-Hanuman from Sri Lanka and the purple-faced langur are polyphyletic. The likelihood score of the best tree (In L = −5830.8072) wherein the three species were constrained to be reciprocally monophyletic (P = 0.000, Shimodaira–Hasegawa test). The parsimony branch-and-bound search yielded three equally parsimonious trees (tree length = 2068). A strict consensus of these three trees (not shown) was very similar to the likelihood tree.

The overall tree topology was the same for various tree-building methods and there was no significant difference between them (P > 0.05, Shimodaira–Hasegawa tests). The three tree-building methods retrieved the three clades mentioned above and the relationships among these clades were also identical. The only difference between methods was with respect to the relationships between species in clades 2 and 3. These changes do not alter the conclusions of the paper.

The aligned sequences of the Cyt-b gene region had no stop codons or indels. Base frequencies at the three codon positions were in concordance with expectations for vertebrate Cyt-b gene (Johns and Avise, 1998; Prusak and Grzybowski, 2004). Additionally, third codon positions had a very low frequency of G's, compared to other mtDNA genes. Thus, these sequences were considered as true mitochondrial Cyt-b sequences rather than nuclear-mitochondrial sequences (Zhang and Hewitt, 1996; Bensasson et al., 2001). Further, it must be noted that the work done by Osterholz et al. (2008) also support Hanuman langur polyphyly. In their study Cyt-b sequences of Hanuman, Nilgiri and purple-faced langurs were derived from other sources wherein different primers sets were used. Therefore it is very unlikely that multiple labs independently amplified nuclear copies of Cyt-b from three different species using different primers.

3.2. The Prm1 gene results

A total of four Prm1 alleles (A1, A1b, A2, and A2b) were found among the three species of langurs. The amino acid sequences of these alleles are given in Fig. 2B. All of the five NT-Hanuman samples analyzed contained only allele A2b (Table 1), which has a 40 bp repeat and therefore could be resolved on a 2% agarose gel. Four of these samples that contained A2b alleles were sequenced to confirm the presence of the repeat and to detect any new allele (due to sequence difference). All of the four Prm1 A2b sequences were identical (data not shown). ST-Hanuman, on the contrary, showed a total of three alleles (A1, A2, and A2b) in six samples analyzed. Some of the individuals were heterozygous at this locus (Table 1).

The purple-faced langurs were found to have one allele (A2) in both the samples analyzed. In the case of Nilgiri langur, a total of three samples were analyzed and two alleles were detected (A1 and A1b). One of these alleles, A1 is also seen in ST-Hanuman langurs. Thus, as in the case of mtDNA diversity, the Nilgiri langur and purple-faced langur share at least one of their Prm1 alleles with ST-Hanuman langurs (see Fig. 2B).

4. Discussion

Results from this and earlier (Karathan et al., 2008; Osterholz et al. 2008) works indicate that the mtDNA of the so-called Hanuman langurs is not monophyletic. Instead, the mtDNA of the ST-Hanuman from Sri Lanka branches with that of purple-faced langur of Sri Lanka, and the mtDNA of the ST-Hanuman from South India groups with Nilgiri langur to the exclusion of NT-Hanuman (Fig. 2A). Furthermore, Prm1 allelic variation found in Nilgiri langurs, purple-faced langurs, and NT-Hanuman langurs is a subset of the variation seen in the ST-Hanuman langurs. These results suggest that the current day ST-Hanuman langur has maintained high levels of mitochondrial and nuclear variation some of which it shares with Nilgiri and purple-faced langurs.

A similar scenario has been reported for another Asian colobine of the genus Trachypithecus wherein the widely-distributed Phayre's leaf monkey (T. phayrei) is polyphyletic with respect to
Francois' leaf monkey (T. francoisi) and dusky leaf monkey (T. obscurus) in the mtDNA tree but the nuclear Prm1 gene support their monophyly (Karanth et al., 2008). This pattern is also observed among fascicularis group of the macaques (genus Macaca) wherein the widely-distributed rhesus macaque (M. mulatta) is polyphyletic with respect to Japanese (M. fuscata) and Taiwanese macaque (M. cyclopis) (Melnick and Hoelzer, 1993). Nuclear allozyme studies on this system, however, indicate that eastern and

---

**Fig. 2.** (A) Maximum likelihood tree based on Cyt-b gene of the langurs of the Indian subcontinent. The numbers by the nodes indicate likelihood, parsimony and neighbor-joining bootstrap supports respectively and the stars indicate 100% support. The arrows indicate the three major clades that were retrieved by all tree-building methods. (B) Prm1 amino acid sequences from the langurs. The site (amino acid 25) shown in bold was used to identify alleles in the coding sequence of the gene. +++ indicate 40 bp repeat; –, deletion. See Table 1 abbreviations.
western populations of rhesus macaques are members of one relatively homogeneous species (Melnick and Hoelzer, 1992). Thus in both the cases discussed above the nuclear markers suggest that these widely-distributed taxa constitute a single homogeneous species but their mtDNA have retained high levels of polymorphism that is shared with sister taxa. In contrast, the so-called Hanuman langurs show slightly different pattern in that the nuclear marker does not support a single homogeneous species. Here, NT-Hanuman is distinct from ST-Hanuman at both nuclear and mitochondrial loci, and preliminary data suggest that ST-Hanuman might have to be split into multiple species (see below).

4.1. Species status of NT- and ST-Hanuman langurs

In the mtDNA tree the NT-Hanuman is monophyletic (clade 1, Fig. 2A) and the average uncorrected pair-wise sequence differences between clade 1 and 2 (after excluding the partial sequences) is comparable to that of among-species comparisons within the genus Trachypithecus (around 7.9%). The mtDNA result is complemented by the nuclear DNA (nDNA) result, wherein all five NT-Hanuman samples collected from a wide area in northern India carry the A2b allele. This allele appeared to be rare in South Indian and Sri Lankan clade; only one sample of ST-Hanuman was found to have this allele (Table 1). Thus the NT-Hanuman appeared to be fixed for the A2b allele at the Prm1 locus. Additionally, the NT-Hanuman is morphologically distinct from ST-Hanuman and occupies a distinct geographical region north of the Tapti–Godavari river system (Fig. 1). These results suggest that NT-Hanuman might warrant species status, but more studies need to be undertaken to rigorously test this hypothesis.

In the case of the ST-Hanuman, due to limited sample size, we might have missed other divergent mtDNA haplotypes and Prm1 alleles. Therefore, more work needs to be done to determine the species status of various ST-Hanuman populations. Nevertheless it is interesting to note that within clades 2 and 3, the Hanuman langurs are polyphyletic with respect to Nilgiri and purple-faced langurs. This tree topology is retrieved by all the tree-building methods with minor rearrangements of the taxa.

Taxonomists have long considered Nilgiri and purple-faced langurs as distinct species (see Karanth et al., 2008, and references therein). These two species are very distinct from Hanuman langurs in their morphology and ecology and have until recently been assigned to genus Trachypithecus along with Southeast Asian langurs (Groves, 2001). Thus to reconcile our results with taxonomy it might be useful to tentatively split Hanuman langurs into three species namely; the NT-Hanuman, the ST-Hanuman from South India, and the ST-Hanuman from Sri Lanka as has been suggested by Osterholz et al. (2008). The divergence between ST-Hanuman and their sister taxa in South India and Sri Lanka might have occurred recently as suggested by the polyphyletic gene-tree. This is because shortly after speciation the probability is high that two sister taxa exhibit a polyphyletic gene-tree status (Avise, 1994) due to incomplete lineage sorting. Alternately, introgression of Hanuman mtDNA into Nilgiri and purple-faced langurs or vice versa could also produce such a topology. There is one record of hybridization between Hanuman and Nilgiri langurs reported from Anamalai Hills of South-west India (Hohmann, 1991). In this regard, it must be noted that the distance between the collection site of the ST-Hanuman samples from South India (Hyderabad) and the current Nilgiri langur distribution is around 1000 kms (Fig. 1). Given that langurs live in matrilineal society (Newton and Dunbar, 1994) and thus exhibit female natal philopatry it is unlikely that recent mtDNA introgression could have occurred over such a long distance. Nevertheless, with the limited data presented here we cannot rule out ancient hybridization.

Taken together these analyses and observations suggest that the so-called Hanuman langurs might have to be split into multiple species as has been done in some classification schemes (Hill, 1939; Groves, 2001; Brandon-Jones, 2004). But the results reported here are based on limited sampling of Hanuman langurs and does not include all the putative species/subspecies. Therefore, we have refrained from designating species names to the three Hanuman langur species discussed above. In the light of these findings, it is apparent that the results of comparative studies done on Hanuman langurs from across the range need to be reinterpreted.

To address the Hanuman langur species and lineage sorting issues, more samples of Hanuman langurs from South India, Bangladesh, and Sri Lanka need to be analyzed with additional markers (particularly a Y-chromosome marker and microsatellites). These analyses will help us better understand the factors that are responsible for the maintenance of intraspecific mtDNA and nDNA diversity in this species complex and also to determine the classification scheme that best fits the molecular data.

Acknowledgments

The authors thank R. Petriagoda, M. Singh, S. Siva, S.N. Nigam, C. Roos, K. Launhardt, P. Dolhinow, R.S. Harave, A. Chaudary, S.M. Mohnot, and N.K. Lohiya for helping us with sample collection, and A.P. Jason de Koning for proofreading the manuscript. Fieldwork for this study was supported through SUNY benevolent, SUNY GSO, and Primate Conservation Inc. grants to K.P.K.

Appendix A. Supplementary data


References


