1 **Integrative taxonomy confirms the species status of the Himalayan langurs,**
2 ***Semnopithecus schistaceus* Hodgson 1840**

3 Kunal Arekar^{a,b,*}, S. Sathyakumar^b, K. Praveen Karanth^a

4 ^aCentre for Ecological Sciences, Indian Institute of Science, Bangalore 560012, India

5 ^bWildlife Institute of India, Dehradun 248001, India

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

7 **Abstract**

8 Taxonomy is replete with groups where the species identity and classification remain
9 unresolved. One such group is the widely distributed Hanuman langur (Colobinae:
10 *Semnopithecus*). For most part of the last century, Hanuman langur was considered to be a
11 single species with multiple subspecies. Nevertheless, recent studies using an integrative
12 taxonomy approach suggested that this taxon is a complex, with at least three species.
13 However, these studies did not include the Himalayan population of the Hanuman langur
14 whose taxonomic status remains unresolved. The Himalayan population of Hanuman langurs
15 have been classified as a distinct species with multiple subspecies or have been subsumed
16 into other species. These classification schemes are wholly based on morphological
17 characters and which are sometimes insufficient to delimit different species. Here, we have
18 integrated data from multiple sources viz. morphology, DNA, and ecology to resolve the
19 taxonomy of the Himalayan langur and to understand its distribution limit. Our results with
20 three lines of evidence corresponding to three different species concepts show that Himalayan
21 langur is a distinct species from *S. entellus* of the plains. Additionally, these results did not
22 show any support for splitting of the Himalayan langur into multiple subspecies. Our study
23 supports the classification proposed by Hill (1939) and we recommend *Semnopithecus*
24 *schistaceus*, Hodgson 1840 as species name for the Himalayan langur and subsume all the
25 known subspecies into it.

26

27 **Introduction:**

28 Species is the fundamental unit of study in many fields of biology such as systematics,
 29 ecology, evolution, behaviour and many more (de Queiroz, 2007). Thus, the first step in these
 30 studies is accurate and unambiguous identification of species for which it is imperative to
 31 have a well resolved taxonomy of the group of interest. Nevertheless, taxonomy is replete
 32 with “problematic groups” wherein species identity and classification remain unresolved.
 33 One such group are the colobine monkeys broadly referred to as Hanuman langur (Genus:
 34 *Semnopithecus*, Subfamily: Colobinae). Hanuman langur is a widely distributed primate in
 35 the Indian subcontinent (Bishop, 1977; Newton, 1988) which exhibits extensive
 36 morphological variation across its range. A multitude of classification schemes have been
 37 proposed mostly during early-mid 20th century to resolve the taxonomic status of Hanuman
 38 langurs (Brandon-jones, 2004; Ellerman, J. R. and Morrison-Scott, 1966; Groves & C, 2001;
 39 Hill, 1939; Napier & Napier, 1967; Pocock, 1928, 1939; Roonwal, 1984; Roonwal &
 40 Mohnot, 1977).

41 Hanuman langur is broadly divided into two categories; the northern type (NT) which
 42 is characterised by forward looping tail towards the head and southern type (ST) with the tail
 43 looping backwards away from the head (Roonwal, 1979, 1984). The Tapti-Godavari rivers in
 44 central India form the borderline with NT distributed to the north and ST distributed to the
 45 south of these rivers (Roonwal, 1979, 1984). Recent studies support the splitting of ST
 46 Hanuman langur into two species, namely *S. priam* and *S. hypoleucos*, based on an
 47 integrative approach wherein multiple lines of evidence from molecular, morphological and
 48 ecological data were used (Ashalakshmi, Nag, & Karanth, 2014; Nag, Karanth, & Vasudeva,
 49 2014; Nag, Pramod, & Karanth, 2011). Similarly, genetic and morphological data suggest
 50 that the plains population of NT Hanuman langur is a separate species, *S. entellus* (Karanth,
 51 Singh, & Stewart, 2010). However, taxonomy of the Himalayan population (hereafter

52 Himalayan langur) is still unresolved. In order to understand the exact number of species of
53 Hanuman langur in the Indian subcontinent, it is important to resolve the taxonomy of all the
54 populations of Hanuman langur.

55 Himalayan langur is the northernmost population of Hanuman langur (Sugiyama,
56 1976) distributed in the Himalayan region of India, Nepal and parts of Pakistan and Bhutan
57 (Blanford, 1888; Pocock, 1939). Altitudinally, they are distributed from the Himalayan
58 foothills up to 4270m asl (above sea level) (Bishop, 1977) and is one of the few colobine
59 monkey living in temperate climate (Nijman, 2013), while the rest are distributed
60 predominantly in tropical regions (Bishop, 1979). Himalayan langurs are classified in the NT
61 category based on the tail loop character (Roonwal, 1979, 1984). Behavioural characters
62 suggests that the Himalayan langurs might be distinct from *S. entellus* of the northern plains.
63 (Bishop, 1975; Bishop, 1979; Curtin, 1982; Sugiyama, 1976), Morphologically, the
64 Himalayan langur can be distinguished from the plains' population (*S. entellus*) by the tail
65 carriage pattern (discussed in methods section) and by their pelage – the langurs from the
66 Himalayas have a bushy white head which is very distinct from the darker grey-brown body
67 (Bishop, 1979).

68 A variety of classification schemes have been proposed to resolve the taxonomy of
69 the Himalayan langurs (Table 1). One of the earliest comprehensive classification of Indian
70 colobines was by Pocock (1928). He assigned the Himalayan langurs to five subspecies
71 *schistaceus*, *ajax*, *achilles*, *lanius* and *hector* under the species *Pithecus entellus*. Later
72 Pocock (1939) renamed *Pithecus* as *Semnopithecus* with three subspecies under it;
73 *schistaceus*, *ajax* and *achilles*. The subspecies *lanius* and *hector* from Pocock's (1928) earlier
74 classification were not included here. Hill (1939) considered the Himalayan langur to be a
75 single species *S. schistaceus* with five subspecies *hector*, *lanius*, *achilles*, *schistaceus* and
76 *ajax*.

77 Subsequent classification schemes (Ellerman and Morrison-Scott, 1966; Napier &
78 Napier, 1967; Roonwal, 1984; Roonwal & Mohnot, 1977) synonymised *Semnopithecus* with
79 *Presbytis* and subsumed all the Himalayan species into a single species *P. entellus* along with
80 the subspecies *entellus* from the northern plains. The subspecies *lanius* (Hill, 1939) was
81 changed to *lania* and the subspecies *hector* (Hill, 1939) was removed.

82 Later, Groves (2001) reverted back to using *Semnopithecus* for Hanuman langurs and
83 recognized three species of Himalayan langur *S. schistaceus*, *S. ajax* and *S. hector*. He
84 elevated the three subspecies from previous classification schemes to species level. Lastly,
85 Brandon-Jones (2004) included all the Himalayan species as subspecies of *S. entellus*, except
86 for *achilles* and *lanius* (Hill, 1939) which he did not include in the classification.

87 Thus, the Himalayan langur has had a convoluted taxonomic history which fall into
88 three broad groups of taxonomic schemes (TS). TS1 – Various populations of Himalayan
89 langurs are placed in multiple subspecies under *Pi. entellus*, *S. entellus* or *P. entellus*
90 (Pocock, 1928, 1939; Ellerman and Morrison-Scott, 1966; Napier and Napier, 1967; Roonwal
91 and Mohnot, 1977; Roonwal, 1984; Brandon-Jones et al., 2004). TS2 – Himalayan langurs is
92 considered a distinct species with multiple subspecies (Hill, 1939). TS3 – Himalayan langurs
93 is split into multiple species (Groves, 2001).

94 With the advent of molecular techniques many recent studies have used genetic data
95 to resolve taxonomic ambiguities in primates of the Indian subcontinent (Arekar, Parigi, &
96 Karanth, 2019; Ashalakshmi et al., 2014; Chakraborty, Ramakrishnan, Panor, Mishra, &
97 Sinha, 2007; Karanth, Singh, Collura, & Stewart, 2008; Karanth et al., 2010; Osterholz,
98 Walter, & Roos, 2008; Wangchuk, Inouye, & Hare, 2008). However, the use of molecular
99 data does not guarantee a robust description and identification (Will, Mishler, & Wheeler,
100 2005). Molecular data is often considered as another singular data type, such as the
101 morphological data, which can be used as a line of evidence to characterise and describe

species. In order to achieve a robust delineation of species, we need to integrate methods from different discipline (Dayrat, 2005)..

Given this background, we have used data from multiple sources viz. morphology, DNA and ecology to address the following questions 1) Is Himalayan langur a distinct species from *S. entellus*? 2) Does Himalayan langur occupy different niche than *S. entellus*? 3) Does Himalayan langur population consist of multiple species/subspecies? And 4) What is the potential distribution range of the Himalayan langurs.

Materials and Methods

Molecular data

Data collection –

We conducted field work in four Himalayan states of India – Jammu and Kashmir (J&K), Himachal Pradesh (HP), Uttarakhand and Sikkim. These states were chosen for field work based on distribution records from past studies (Pocock, 1939; Hill, 1939; Sugiyama, 1975; Bishop, 1979; Choudhury, 2001) as well as from IUCN website (www.iucnredlist.org). Furthermore, we included our field data and published sequences from Nepal.

Molecular data was based on 175 faecal samples collected from 46 locations (Table S1) from across the distribution range of Himalayan langur in India and which included multiple samples collected from each location. Additionally, five faecal samples of *S. entellus* were collected from one location in the northern plains (22.88220N, 88.39970E). Fresh faecal samples were collected by following the troops in the morning and the evening hours. Samples were collected by two different methods – First, as described in Kawamoto et al. (2013), a sterile cotton swab was rolled multiple times over the surface of the faeces and thoroughly rinsed in the lysis buffer (White & Densmore, 1992). Second method involved collection of the whole faeces which was then stored in absolute alcohol. These samples were stored at -20 °C in the laboratory until DNA extraction. Samples stored in lysis buffer were

first treated with starch to remove potential PCR inhibitors like bilirubin and bile salts (Kawamoto et al., 2013; Zhang, Li, Ma, & Wei, 2006) and then DNA was extracted by using Wizard® SV Gel and PCR Clean-Up System, Promega, Madison, WI, USA and stored in pure nuclease free water at -20 °C until further use. DNA from whole faecal 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, Reef Hardy, Anderson, Sanchez, & Buoncristiani, 2006;). 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).

PCR amplification and sequencing –

A 700 bp region of mitochondrial cytochrome *b* (Cyt-*b*) gene was PCR amplified using the primer pair Cytb_278F (5' – GCCTATTTCTACACGTAGGCCG – 3') and Cytb_1052R (5'– CCAATTGCAATGAAGGGTTGGT – 3'). A 25 µl of reaction was set with standard 1X reaction buffer premixed with 1.5 mM MgCl₂ (New England BioLabs® Inc.), 0.25mM of dNTPs (Bangalore Genei, Bangalore), 0.3 µM of each primer (Sigma-Aldrich Inc.), 1.5U Taq polymerase (New England BioLabs® Inc.) and 2 µl of template DNA (DNA concentration between samples varied from 20 ng/µl to 80 ng/µl). We also added 2 µl of BSA (Bovine Serum Albumin, Fisher Scientific) to augment the PCR reaction. Further, the template DNA was diluted to 1:5 (DNA extract: water) ratio and used for the reaction in order to reduce the amount of PCR inhibitors. The PCR cycling conditions were carried out with initial denaturation at 94 °C for 5 mins followed by 50 cycles with denaturation at 94 °C for 40 sec, annealing at 57.5 °C for 30 sec and elongation at 72 °C for 30 sec and final extension at 72

°C for 10 mins. The PCR products were outsourced for purification and sequencing to Medauxin, Bangalore.

Phylogenetic analysis –

The sequence files obtained were viewed and edited manually in ChromasLite v2.01 (Technelysium Pty Ltd). Sequences Himalayan langur as well as *S. entellus* were also downloaded from a previous molecular study (Supplementary material, Table S5) (Ashalakshmi et al., 2014; Khanal, Chalise, Wan, & Jiang, 2018). The sequences were aligned using MUSCLE algorithm (Edgar, 2004) incorporated in MEGA v7 (Kumar, Stecher, & Tamura, 2016).

We used jModelTest 2.1.3 (Darriba, Taboada, Doallo, & Posada, 2012) to pick the best model of sequence evolution. Phylogenetic reconstruction was performed using Maximum Likelihood (ML) and Bayesian methods. ML analysis was performed in RAxML7.4.2 incorporated in raxmlGUI v1.3 (Stamatakis, 2006). We used GTR+G model in RAxML as there is no provision to use other models. 1000 replicates were performed to assess support for different nodes. We used MrBayes 3.2.2 (Ronquist et al., 2012) to perform the Bayesian analysis with HKY+G nucleotide substitution model. Two parallel runs with four chains each were run for 10 million generations with sampling frequency every 1000 generations. Convergence between the two runs was determined based on standard deviation of split frequencies. The program Tracer v1.6 (Rambaut, A., Suchard, M.A., Xie, D., Drummond, 2013) was used to determine stationarity, an effective sample size (ESS) value of >200 for each parameter was used as a cut-off for run length. The first 25% of trees were discarded as burn-in.

Hypothesis testing –

We compared the Bayesian tree (Fig. 1), with an *a priori* hypothesis where the phylogeny was constrained to be consistent with Groves' (2001) three species of Himalayan langurs; *S. ajax*, *S. hector* and *S. schistaceus*. RAxML7.4.2 incorporated in raxmlGUI v1.3 (Stamatakis, 2006) was used to generate a constraint tree. The likelihood of the constraint tree was then compared with the best tree using the SH test (Shimodaira & Hasegawa, 1999) and AU test (Shimodaira, 2002) in PAUP* Version 4.0a (build 164) (Swofford, 2001). For both the SH and AU test, we performed 10000 bootstrap replicates and the nonparametric bootstrap with reestimated log likelihoods (RELL) approximation (Kishino, Miyata, & Hasegawa, 1990) was used for resampling the loglikelihoods.

Morphological data

Data collection –

Morphological data was collected by direct observations during field survey as well as from photographs. Based on past studies (Nag et al., 2011) and our field observations, six color-independent morphological characters were used to differentiate between *S. entellus* and the Himalayan langur. These included the four characters described by Nag et al. (2011) and two characters unique to Himalayan langurs. Characters specific to Himalayan langur included tail carriage (Roonwal, 1979, 1984) and demarcation between the head and the body (HBC) (Bishop, 1979; Groves, 2001; Hill, 1939; Pocock, 1939). Within the NT langurs, two forms of tail carriage are observed, in *S. entellus* from the northern plains the tail loops over the back and the tip of the tail hangs perpendicular to the ground, here onwards TC3 (Fig. 1b), whereas, in the Himalayan populations the tail loops well behind the back and the tip ends above the base of the tail, here onwards TC4 (Fig. 1a). We coded them TC3 and TC4 to be consistent with Nag et al.'s (2011) coding system. The tail carriage pattern was recorded when the individual was walking on a flat surface and not while climbing up or down the hill and nor while it was running or standing (as per Roonwal, 1984). The Himalayan langur has a

distinct demarcation between the head and the body, the head is bushy and white coloured distinct from the grey-brown body (Fig. 2a) (Bishop, 1979; Groves, 2001; Hill, 1939; Pocock, 1939). On the other hand, the langurs from the plains (*S. entellus*) have a uniform colour without any distinction between the head and the body (Fig. 2b). We call this character Head-Body contrast (HBC) character. Apart from these two characters we also used the four characters described in Nag et al. (2011), these characters are not seen in Himalayan langurs and therefore were coded '0'(absent). We recorded these characters by direct observations using 10 X 50 binoculars (Olympus) and through photographs taken from a digital camera (Canon PowerShot SX20 IS). These characters were scored for multiple adult individuals per location.

Analysis –

We typed 65 adult individuals from 29 locations (Supplementary material, Table S6). We also included one of the southern species from this complex, *S. hypoleucos* (St1 and St2 morphotypes in Nag et al., 2011) in the analysis. The characters were coded as described in Table 2. All the characters were coded in a way that they are consistent with Nag et al.'s (2011) coding system. The codes for *S. hypoleucos* were similar to ones used in Nag et al. (2011).

We prepared a character matrix with the terminal taxa in the rows and the six morphological characters in the columns. Using this matrix, first the mean character difference was calculated between the individuals in PAUP* Version 4.0a (build 164) (Swofford, 2001). Then a Neighbour Joining (NJ) tree was built using these distances, with the default settings. Mid-point rooting was used to root the tree.

Ecological data

Data collection –

For ecological data, we obtained 192 occurrence records of the Himalayan langurs (Available with authors on request). Out of these, 79 records were from the field surveys conducted for this study, 58 occurrence records were obtained from previous studies (Khanal et al., 2018; R.A. Minhas et al., 2018; Minhas et al., 2012) and 55 occurrence records were downloaded from GBIF (Global Biodiversity Information Facility) database (www.gbif.org). For *S. entellus*, 69 occurrence records were used, out of these, 21 records were obtained from Nag et al. (2014) and 48 records were downloaded from GBIF database (www.gbif.org). For the occurrence records downloaded from GBIF database, we plotted these occurrence records on the map and included only those records which were falling within the known distribution zones of the respective taxa. Further, we used 22 environmental layers, 19 were bioclimatic layers downloaded from www.worldclim.org, one altitude layer (USGS website), two more layers – slope and aspect were derived from the altitude layer in ArcGIS 10.2.1 (Table S2). All the layers were of 30 arcsec resolution, projected in WGS84 projection. These bioclimatic variables were clipped to the region from 68 °E to 97.4 °E and from 6.7 °N to 37 °N using ArcGIS 10.2.1. These clipped layers were then exported to ASCII format using QGIS 2.18.12. The 22 environmental layers were tested for multicollinearity by calculating Pearson's correlation coefficient (r). The layers with $r \leq 0.75$ were selected for further analysis (Supplementary material, Table S3).

Ecological Niche Modelling analysis –

Model selection:

We first performed model selection using the maximum entropy algorithm available in Maxent v3.4.1 (Phillips, Anderson, & Schapire, 2006). We tested 48 models, each for Himalayan langur and *S. entellus*, by employing different combinations of features and

regularisation multiplier (RM) (Supplementary material, Table S4) in MaxEnt v3.4.1 (Phillips et al., 2006).

MaxEnt algorithm:

Separate Maxent analyses for the Himalayan langur and for *S. entellus* were implemented in MaxEnt v3.4.1. Maxent was run with the following modifications. Random test percentage was set to 30%, maximum number of background points was set to 10000 and the replicates were set to 10 with replicated run type changed to Subsample. 5000 iterations were performed with the convergence threshold set to 1×10^{-5} . Jackknife test was used to estimate the contribution of each environmental variable. The feature type and RM value were based on the best model selected (Table S4). To overcome the sampling bias, a bias file was created in ArcGIS 10.2.1 by applying Gaussian kernel density function to 10000 background points (Elith, Kearney, & Phillips, 2010). The output format was chosen as Cloglog (Phillips, Anderson, Dudík, Schapire, & Blair, 2017). AUC values were examined to check for the predictive ability of the model.

Schoener's D (Schoener, 1968) was calculated as a measure of niche overlap between the two distribution models (Warren & Seifert, 2011) implemented in ENMTools (Warren, Glor, & Turelli, 2010)

Results

Phylogenetic analysis

Out of the total 182 samples subjected to DNA extraction we obtained sequences from 76 samples. We also performed DNA extraction and sequencing for one *S. hypoleucos* sample, it was used as an outgroup in this study. These 76 sequences contained multiple identical sequences, so for building the phylogenetic tree, we trimmed the dataset to one unique

sequence per location. Therefore, our final dataset contained 746 bp of Cyt-*b* sequence from 57 samples (Table S5), which included sequences from previous studies.

The Bayesian and ML analysis recovered two major clades, the *S. entellus* clade – containing the sequences from the northern plains; and the Himalayan clade – containing sequences from the Himalayan region (Fig 3). Within the Himalayan clade was a well-supported subclade consisting of haplotypes from the western Himalayas, whereas sequences from eastern Himalayas did not form a separate cluster. Both, the Bayesian and ML trees, showed similar topology where in all the major clades were retrieved. Furthermore, in the Bayesian tree (Fig. 1), two samples i.e. 128_Nepal and 129_Nepal were placed within the clade containing samples from the Indian Himalayan Region (IHR), whereas in the ML tree (Supplementary material, Fig. S1), these two samples were sister to the above mentioned clade.

Hypothesis testing

The likelihood score of the best tree was significantly higher than that of the constraint tree for both SH test and AU tests. Therefore, these trees based on the molecular data did not support the splitting of Himalayan langurs into three species (Table 3).

Morphological data

Our final NJ tree (Fig. 4) was based on morphological data collected from 64 adult individuals from 60 localities. The NJ dendrogram retrieved a distinct cluster consisting of Himalayan samples that was sister to *S. entellus*. The UPGMA method also generated a similar topology.

Ecological niche modelling

Model prediction –

The best model selected for the Himalayan langurs had features ‘LQPTH’ and a RM value of 2.5; whereas for *S. entellus*, the best model had features ‘Auto’ with a RM value of 3 (Table

S4). Based on these models and the environmental variables selected, we obtained different distribution maps for *S. entellus* and the Himalayan langur (Fig 5). In the distribution maps, the warmer colour indicates suitable area whereas the cooler colours indicate unsuitable areas. The AUC values for the training and test data for the Himalayan langur dataset were 0.9663 and 0.9621, respectively. And the AUC values for the training and test data for *S. entellus* were 0.889 and 0.87, respectively. These AUC values indicate that the potential distribution of these species fits well with our data.

Variable selection and their importance –

For the Himalayan langur, precipitation of driest quarter (Bio17) was the highest ranked variable (Table 5). Jackknife test also illustrates the importance of Bio17 (Supplementary material, Fig. S4). The response curve (Supplementary material, Fig. S2) shows that the habitat suitability increases with the precipitation of the driest quarter (Bio17) but eventually attains stationarity. Annual mean temperature (Bio1) and mean diurnal range (Bio2) were the next two contributing variables (Table 5).

Precipitation seasonality (Bio15) was the most contributing variable towards the distribution of *S. entellus* (Table 5). The response curve for Bio15 (Supplementary material, Fig S3) shows that the values for habitat suitability were high towards higher values of precipitation seasonality (Bio15), suggesting that *S. entellus* prefers areas with high variation in rainfall. Mean temperature of warmest quarter (Bio10) and annual mean temperature (Bio1) were the next two highly contributing variables (Table 5).

Niche overlap between *S. entellus* and Himalayan langur was 17 percent (Schoener's D value = 0.17). However, when we compare the areas with high probability of distribution (> 0.75), there was no overlap.

Discussion

The so-called Hanuman langur has been known to be a species complex since a long time. Nevertheless, recent studies have brought some clarity to their confused taxonomy. These studies suggest that the so-called Hanuman langur consists of at least three species: *S. entellus*, distributed in the plains of North India (Karanth et al., 2010); *S. priam* and *S. hypoleucos* distributed in peninsular India (Nag et al., 2011, Ashalakshmi et al., 2014, Nag et al., 2014). However, the taxonomic status of the Himalayan population of this complex remained unresolved. Here we address this issue by applying multiple lines of evidence to resolve the taxonomic status of the Himalayan langur.

Our molecular analysis based on mitochondrial Cyt-*b* gene retrieved a monophyletic Himalayan langur distinct from *S. entellus* of the plains. The Himalayan langur is also morphologically distinct and occupies a specific niche in the Himalayas. Thus, three lines of evidences, corresponding to different species concepts, suggest that the Himalayan langurs is separately evolving metapopulation lineage (de Queiroz, 1998). Phylogenetic species concept (PSC) II (de Queiroz, 1998; Donoghue, 1985; Mishler, 1985) identifies species as monophyletic groups on the basis of shared derived characters. Our genetic data shows that Himalayan langur and *S. entellus* are reciprocally monophyletic. Results from the morphological analysis suggests that the Himalayan langur is a separate species as per the Phenetic Species Concept (Sokal & Crovello, 1970). Our niche modelling analysis too shows that these two lineages occupy distinct ecological regions and thus show separation along the ecological axis (Van Valen, 1976).

Among the plethora of classification schemes only Hill (1939) placed the Himalayan langurs in a separate species, *S. schistaceus*, with multiple subspecies – *ajax*, *achilles*, *lanius* and *hector* (Table 1). Our study supports this classification, however there is no support for further splitting this species into multiple subspecies/species (see result under hypotheses testing). Therefore, we recommend *Semnopithecus schistaceus* Hodgson, 1840 as a species

name for all the Himalayan langurs and subsume all the earlier described species and subspecies within it until further detailed taxonomic studies address this issue.

Two of the characters used in this study – HBC and forms of tail carriage (TC3 and TC4), can be used as field identification characters for differentiating *S. schistaceus* and *S. entellus*. We do not recommend the use any of the external morphological characters listed in the earlier classification schemes (Table 1) for distinguishing between different subspecies of Himalayan langurs. The morphological characters used in earlier classification schemes to describe these subspecies, are highly plastic, variable, and subjective in nature (Nag et al., 2011). For instance, in the identification key Pocock (1939) describes *schistaceus* as follows “General colour paler, salty or greyish-buff; coat shorter and less woolly” and for *achilles* he writes, “General colour dark earthy brown; coat thick and woolly”. Hill (1939) describes *schistaceus* as “A slatey-grey race, with shorter, less woolly coat than those found at higher altitudes”. These characters tend to differ based on what month of the year the langur is being observed and what is the altitude at that location. Roonwal (1981) describes four types of tail carriage in the NT langurs, however we recorded only two tail carriage types that are used in this study, the other two tail carriage types have not been observed.

We used ecological niche modelling (ENM) to determine the distribution range of *S. schistaceus*. It predicted the distribution range of the *S. schistaceus* (Fig. 5a) with high accuracy (AUC = 0.96). The distribution was mainly governed by precipitation of the driest quarter (Bio17). Response curves (Supplementary material, Fig. S2) for the top three contributing variables (Table 4) suggests that *S. schistaceus* prefers areas with high precipitation but with moderate temperatures. Interestingly, slope and aspect did not contribute to the model prediction. A recent study (Khanal et al., 2018) also showed that precipitation plays a major role in the distribution of *S. schistaceus* in Nepal. Langurs in the Himalaya inhabit broad leaves subtropical forest at lower altitudes and temperate broadleaf

forest at higher altitudes (Bishop, 1979; Curtin, 1982; Minhas et al., 2013; Sayers & Norconk, 2008; Sugiyama, 1976). Occasionally, these langurs are found in temperate coniferous forests above 4000m altitude (Bishop, 1977). These forests receive high rainfall during the monsoon as well as precipitation in the form of snowfall in winter (Bhattarai & Vetaas, 2003; J. S. Singh & Singh, 1987; P. Singh, Ramasastri, & Kumar, 1995).

We also used ENM to determine if the ecological niches of *S. schistaceus* and *S. entellus* are separate. The ENM distinctly demarcated ecological niches of these two species. The AUC value for both the distribution models was significant, implying that the results greatly differ from the random predictions. Precipitation was the most important variable in demarcating these two species, with *S. entellus* requiring less precipitation and the Himalayan langurs need higher precipitation. Bishop (1979) also pointed out that climate is the factor that governs the distribution of the *S. schistaceus* and *S. entellus*. We also checked for niche overlap between these two species to determine if they are divergent in their ecological axis. The niche overlap between these two taxa was not significant suggesting that their ecological niches are separate.

This study provides a comprehensive evidence for elevating the Himalayan langur to a species that is distinct from *S. entellus* of the northern plains. Based on our results, we recommend that this species be assigned to *S. schistaceus* as per Hill (1939).

Acknowledgments

We would like to thank the Department of Biotechnology, Government of India for providing funds for laboratory work. We would like to thank Dr. Chetan Nag and Ms. Mehreen Khaleel for providing some of the samples used in this study. Ms. Himani Nautiyal provided photographs from Bhutan for the morphological analysis. Dr. Bilal Habib for help during sample collection and Ms. Neha Tiwari for help in lab work. We would also like to thank the

forest departments of Uttarakhand, Himachal Pradesh and Sikkim for providing necessary collection permits. Additionally, thanks to Jammu and Kashmir forest department for permits provided to Mehreen Khaleel. KA would like to thank the Rufford foundation and Primate Conservation Inc. (PCI) for providing funding for field work. Dr. Aravind Madhyastha and Mr. Bipin Charles for guidance in the niche modelling analysis.

References

- Arekar, K., Parigi, A., & Karanth, P. K. (2019). The convoluted evolutionary history of the capped-golden langur lineage (Cercopithecidae: Colobinae) – concatenation versus coalescent analyses. *BioRxiv*. <https://doi.org/http://dx.doi.org/10.1101/508929>
- 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(1939), 43–58. <https://doi.org/10.1007/s10592-014-0638-4>
- Bhattarai, K. R., & Vetaas, O. R. (2003). Variation in plant species richness of different life forms along a subtropical elevation gradients in the Himalayas, East Nepal. *Global Ecology and Biogeography*, 12(4), 327–340. <https://doi.org/10.1046/j.1466-822X.2003.00044.x>
- Bishop, N. H. (1975). *Vocal behavior of adult male langurs (presbytis entellus) in a high altitude environment*. PhD Thesis, University of California, Berkeley.
- Bishop, N. H. (1977). Langurs living at high altitudes. *Journal of The Bombay Natural History Society*, 74(3), 518–520.
- Bishop, N. H. (1979). Himalayan langurs: Temperate colobines. *Journal of Human Evolution*. [https://doi.org/10.1016/0047-2484\(79\)90092-7](https://doi.org/10.1016/0047-2484(79)90092-7)
- Blanford, W. T. (1888). *Mammalia*, in *The Fauna of British India*. London: Taylor & Francis.
- Brandon-jones, D. (2004). a Taxonomic Revision of the Langurs and Leaf Monkeys (Primates: Colobinae) of South Asia. *Zoo's Print Journal*, 19, 1552–1594.
- Chakraborty, D., Ramakrishnan, U., Panor, J., Mishra, C., & Sinha, A. (2007). Phylogenetic relationships and morphometric affinities of the Arunachal macaque *Macaca munzala*, a newly described primate from Arunachal Pradesh, northeastern India. *Molecular Phylogenetics and Evolution*, 44(2), 838–849. <https://doi.org/10.1016/j.ympev>

- 2007.04.007
- Choudhury, A. (2001). Primates in Northeast India: An Overview of their Distribution and Conservation Status. *Envis Bulletin: Wildlife and Protected Areas*, 1, 92–101.
- Curtin, R. . (1982). Range Use of Gray Langurs in Highland Nepal. *Folia Primatologica*, 38, 1–18.
- Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, 9(8), 772–772. <https://doi.org/10.1038/nmeth.2109>
- Dayrat, B. (2005). Towards integrative taxonomy. *Biological Journal of the Linnean Society*, 85(3), 407–415. <https://doi.org/10.1111/j.1095-8312.2005.00503.x>
- de Queiroz, K. (1998). The general lineage concept of species, and the process of speciation: A conceptual unification and terminological recommendations. In *Endless forms: Species and speciation* (pp. 57–75). Oxford University Press, New York.
- de Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, 56(6), 879–886. <https://doi.org/10.1080/10635150701701083>
- Donoghue, M. J. (1985). A Critique of the Biological Species Concept and Recommendations for a Phylogenetic Alternative. *The Bryologist*, 88(3), 172–181. <https://doi.org/10.2307/3243026>
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1(4), 330–342. <https://doi.org/10.1111/j.2041-210X.2010.00036.x>
- Ellerman, J. R. and Morrison-Scott, T. C. S. (1966). *Checklist of palaearctic and Indian mammals, 1758– 1946* (2nd ed.). London: Trustees of the British Museum (Natural History).
- Groves, C. P., & C, G. (2001). *Primate Taxonomy*. Smithsonian Institution Press.
- Hill, W. C . (1939). An Annotated Systematic List of the Leaf-monkeys. *Ceylon Journal of Science*, XXI.
- 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). *Molecular Phylogenetics and Evolution*, 46, 683–694.
- Karanth, K. P., Singh, L., & Stewart, C.-B. (2010). Mitochondrial and nuclear markers

suggest Hanuman langur (Primates: Colobinae) polyphyly: Implications for their species status. *Molecular Phylogenetics and Evolution*, 54(2), 627–633. <https://doi.org/10.1016/j.ympev.2009.10.034>

Kawamoto, Y., Takemoto, H., Higuchi, S., Sakamaki, T., Hart, J. a., Hart, T. B., ... Furuichi, T. (2013). Genetic Structure of Wild Bonobo Populations: Diversity of Mitochondrial DNA and Geographical Distribution. *PLoS ONE*, 8(3), 1–9. <https://doi.org/10.1371/journal.pone.0059660>

Khanal, L., Chalise, M. K., Wan, T., & Jiang, X. (2018). Riverine barrier effects on population genetic structure of the Hanuman langur (*Semnopithecus entellus*) in the Nepal Himalaya. *BMC Evolutionary Biology*, 18(159), 1–16.

Kishino, H., Miyata, T., & Hasegawa, M. (1990). Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *Journal of Molecular Evolution*, 31(2), 151–160. <https://doi.org/10.1007/BF02109483>

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(5), 1055–1061. <https://doi.org/10.1111/j.1556-4029.2006.00204.x>

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, 33(7), 1870–1874. <https://doi.org/10.1093/molbev/msw054>

Minhas, R. A., Ahmed, K. B., Awan, M. S., Zaman, Q., Dar, N. I., & Ali, H. (2012). Distribution patterns and population status of the Himalayan grey langur (*Semnopithecus ajax*) in Machiara National Park, Azad Jammu and Kashmir, Pakistan. *Pakistan Journal of Zoology*, 44(3), 869–877.

Minhas, R. A., Ali, U., Awan, M. S., Ahmed, K. B., Khan, M. N., Dar, N. I., ... Tsuji, Y. (2013). Ranging and foraging of Himalayan grey langurs (*Semnopithecus ajax*) in Machiara National Park, Pakistan. *Primates*, 54(2), 147–152. <https://doi.org/10.1007/s10329-013-0345-7>

Minhas, R. A., Khan, M. N., Awan, M. S., Ahmad, B., Bibi, S. S., Hanif, M., & Mian, A. (2018). RAPD based Genetic Diversity of Endangered Himalayan Gray Langur (*Semnopithecus ajax*) Populations of Pakistan. *Pakistan Journal of Zoology*, 50(6), 2059–2071. <https://doi.org/10.17582/journal.pjz/2018.50.6.2059.2071>

Mishler, B. D. (1985). The Morphological , Developmental, and Phylogenetic Basis of Species Concepts in Bryophytes. *The Bryologist*, 88(3), 207–214.

498 Mondol, S., Ullas Karanth, K., Samba Kumar, N., Gopalaswamy, A. M., Andheria, A., &
499 Ramakrishnan, U. (2009). Evaluation of non-invasive genetic sampling methods for
500 estimating tiger population size. *Biological Conservation*, 142(10), 2350–2360.
501 <https://doi.org/10.1016/j.biocon.2009.05.014>

502 Nag, C., Karanth, K. P., & Gururaja, K. V. (2014). Delineating Ecological Boundaries of
503 Hanuman Langur Species Complex in Peninsular India Using MaxEnt Modeling
504 Approach. *PLoS ONE*, 9(2), e87804. <https://doi.org/10.1371/journal.pone.0087804>

505 Nag, K. S. C., Pramod, P., & Karanth, K. P. (2011). Taxonomic Implications of a Field Study
506 of Morphotypes of Hanuman Langurs (*Semnopithecus entellus*) in Peninsular India.
507 *International Journal of Primatology*, 32(4), 830–848. [https://doi.org/10.1007/s10764-](https://doi.org/10.1007/s10764-011-9504-0)
508 011-9504-0

509 Napier, J. R., & Napier, P. H. (1967). *A handbook of living primates: morphology, ecology*
510 *and behaviour of nonhuman primates*. Academic Press.

511 Newton, P. N. (1988). The Variable Social Organization of Hanuman Langurs (*Presbytis*
512 *entellus*), Infanticide, and the Monopolization of Females. *International Journal of*
513 *Primatology*, 9(1), 59-77.

514 Nijman, V. (2013). Distribution and Ecology of the Most Tropical of the High-Elevation
515 Montane Colobines: The Ebony Langur of Java. In N. B. et al. Grow (Ed.), *High*
516 *Altitude Primates* (pp. 115–132). Springer. <https://doi.org/10.1007/978-1-4614-8175-1>

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

521 Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening
522 the black box: an open-source release of Maxent, (March), 887–893.
523 <https://doi.org/10.1111/ecog.03049>

524 Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of
525 species geographic distributions. *Ecological Modelling*, 190, 231–259.
526 <https://doi.org/10.1016/j.ecolmodel.2005.03.026>

527 Pocock, R. I. (1928). The Langurs, Or Leaf Monkeys, of British India. *Journal of The*
528 *Bombay Natural History Society*, 32, 472–504.

529 Pocock, R. I. (1939). *Primates and Carnivora (in part)* (Vol. 1). Taylor & Francis. Retrieved
530 from http://f.letmeprint.ru/14757241-7a331b5b/fragment_311854.pdf

531 Rambaut, A., Suchard, M.A., Xie, D., Drummond, A. J. (2013). Tracer v1.6. Available from

<http://beast.bio.ed.ac.uk/Tracer>.

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. *Systematic Biology*, 61(3), 539–542. <https://doi.org/10.1093/sysbio/sys029>

Roonwal, M. L. (1979). Field observations on distribution and tail carriage in the central Himalayan langur, *Presbytis entellus schistaceus* (Primates). *Proceedings of the Indian National Science Academy*, 45(B1), 43-55.

Roonwal, M. L. (1984). Tail form and carriage in Asian and other primates, and their behavioral and evolutionary significance. In N. S. Roonwal, M. L., Mohnot, S. M. and Rathore (Ed.), *Current primate research* (pp. 93–151). Jodhpur University Press, India.

Roonwal, M. L., & Mohnot, S. M. (1977). *Primates of South Asia: Ecology, Sociobiology, and Behavior*. Harvard University Press.

Sayers, K., & Norconk, M. A. (2008). Himalayan Semnopithecus entellus at Langtang National Park, Nepal: Diet, activity patterns, and resources. *International Journal of Primatology*, 29(2), 509–530. <https://doi.org/10.1007/s10764-008-9245-x>

Schoener, T. W. (1968). The Anolis lizards of Bimini: Resource partitioning in a complex fauna. *Ecology*, 49(4), 704–726.

Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. *Systematic Biology*, 51(3), 492–508. <https://doi.org/10.1080/10635150290069913>

Shimodaira, H., & Hasegawa, M. (1999). Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. *Molecular Biology and Evolution*, 16(8), 1114–1116. <https://doi.org/10.1093/oxfordjournals.molbev.a026201>

Singh, J. S., & Singh, S. P. (1987). Forest Vegetation of the Himalaya. *The Botanical Review*, 53(1), 80–192.

Singh, P., Ramasastri, K. S., & Kumar, N. (1995). Topographical Influence on Precipitation Distribution in Different Ranges of Western Himalayas. *Nordic Hydrology*, (26), 259–284.

Sokal, R. R., & Crovello, T. J. (1970). The Biological Species Concept: A Critical Evaluation. *The American Naturalist*, 104(936), 127–153.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22(21), 2688–2690. <https://doi.org/10.1093/bioinformatics/btl446>

Sugiyama, Y. (1976). Characteristics of the Ecology of the Himalayan Langurs. *Journal of*

- 566 *Human Evolution*, 5, 249–277.
- 567 Swofford, D. (2001). PAUP*—Phylogenetic Analysis Using Parsimony and Other Methods,
- 568 Version 4. Sunderland, MA: Sinauer Associates.
- 569 Van Valen, L. (1976). Ecological Species, Multispecies, and Oaks. *Taxon*, 25(2/3), 233–239.
- 570 Wangchuk, T., Inouye, D. W., & Hare, M. P. (2008). The Emergence of an Endangered
- 571 Species: Evolution and Phylogeny of the *Trachypithecus geei* of Bhutan. *International*
- 572 *Journal of Primatology*, 29, 565–582.
- 573 Warren, D. L., Glor, R. E., & Turelli, M. (2010). ENMTools: a toolbox for comparative
- 574 studies of environmental niche models. *Ecography*, 33, 607–611.
- 575 <https://doi.org/10.1111/j.1600-0587.2009.06142.x>
- 576 Warren, D. L., & Seifert, S. N. (2011). Ecological niche modeling in Maxent: the
- 577 importance of model complexity and the performance of model selection criteria.
- 578 *Ecological Applications*, 21(2), 335–342.
- 579 White, P. S., & Densmore III, L. D. (1992). Mitochondrial DNA isolation. In A. R. Hoelzel
- 580 (Ed.), *Molecular genetic analysis of populations: a practical approach* (pp. 29–58).
- 581 Oxford: IRL press, Oxford University Press.
- 582 Will, K. W., Mishler, B. D., & Wheeler, Q. D. (2005). The Perils of DNA Barcoding and the
- 583 Need for Integrative Taxonomy. *Systematic Biology*, 54(5), 844–851.
- 584 <https://doi.org/10.1080/10635150500354878>
- 585 Zhang, B.-W., Li, M., Ma, L.-C., & Wei, F.-W. (2006). A Widely Applicable Protocol for
- 586 DNA Isolation from Fecal Samples. *Biochemical Genetics*, 44(11–12), 494–503.
- 587 <https://doi.org/10.1007/s10528-006-9050-1>
- 588

Tables

Table 1: Different Taxonomic Schemes (TS) for Himalayan langur proposed by various

TS 1					TS 2	TS 3
Pocock (1928)	Pocock (1939)	Roonwal and Mohnot (1977)	Roonwal (1984)*	Brandon-Jones (2004)	Hill (1939)	Groves (2001)
<i>Pi. e. schistaceus</i>	<i>S. e. schistaceus</i>	<i>P. e. schistaceus</i>	<i>P. e. schistaceus</i>	<i>S. e. schistaceus</i>	<i>S. schistaceus</i>	<i>S. schistaceus</i>
<i>Pi. e. ajax</i>	<i>S. e. ajax</i>	<i>P. e. ajax</i>	<i>P. e. ajax</i>	<i>S. e. ajax</i>	<i>S. s. ajax</i>	<i>S. ajax</i>
<i>Pi. e. achilles</i>	<i>S. e. achilles</i>	<i>P. e. achilles</i>	<i>P. e. achilles</i>	—	<i>S. s. achilles</i>	—
<i>Pi. e. lanius</i>	—	<i>P. e. lania</i>	<i>P. e. lania</i>	—	<i>S. s. lanius</i>	—
<i>Pi. e. hector</i>	—	—	—	<i>S. e. hector</i>	<i>S. s. hector</i>	<i>S. hector</i>

authors

*Also includes Ellerman and Morison-Scott (1966) and Napier and Napier (1967)

Pi = *Pithecus*; *S* = *Semnopithecus*; *e.* = *entellus*

597

598 **Table 2:** Coding system used for the 6 morphological characters used in this study. The
599 coding system is designed to be consistent with Nag et al. (2011), unless mentioned
600 otherwise.

	Characters					
	Crest	Streak	EOB	Tail loop (TL)	Tail carriage (TC)	HBC
Himalayan langur	0	0	0	1	3	1
<i>Semnopithecus entellus</i> *	0	0	3	1	4 [#]	0
<i>Semnopithecus hypoleucos</i>	0	1	3/4 [†]	2	1	0
<i>Semnopithecus priam</i> **	1	0	1	2	2	0

601 **S. entellus* = Nt from Nag et al. (2011).

602 [#]Nag et al. (2011) coded Nt as 0, here we have coded it 4.

603 [†]*S. hypoleucos* contains two morphotypes (St1/St2; Nag et al., 2011).

604 ** We did not use this species in the analysis.

605 **Table 3: Topology test results**

Tree	-ln L	Diff -ln L	SH	AU	p-value
Best tree	2107.65046	(best)			
Constrained tree	2371.94530	264.29484	0.0000*	~0*	*P < 0.05

606

Table 4: Maxent results showing the most important variables ranked on the basis of the amount of variation they explain in predicted distribution of *Semnopithecus entellus* and the Himalayan langur along with the permutation importance for the variables.

<i>Semnopithecus entellus</i>			Himalayan langur		
Variable	Percent contribution	Permutation importance	Variable	Percent contribution	Permutation importance
Bio 15	35.5	39.6	Bio 17	42.1	5.9
Bio 10	26.5	7.2	Bio 1	20.6	48.2
Bio 1	20.7	19	Bio 2	15.5	16.5
Bio 12	11.7	19.9	Bio 4	13.7	23.7
Bio 18	2.5	5.7	Bio 12	8.1	5.7
Bio 3	1	5.4	Slope	0.1	0
Bio 8	0.8	0	Aspect	0	0
Slope	0.3	0.1			
Bio 9	0.3	1.7			
Aspect	0.3	0.7			
Bio 14	0.2	0.6			
Bio 19	0.1	0			
Bio 4	0.1	0			

Figure Captions

Fig. 1: Tail carriage pattern in the Himalayan langur (a) and *Semnopithecus entellus* (b).

Fig. 2: Contrast between the head and the dorsal region of the body in the Himalayan langur (a) and *Semnopithecus entellus* (b).

Fig. 3: Bayesian phylogeny of Himalayan langur for mitochondrial cytochrome *b* (Cyt-b) gene. Node support is indicated by black circles and black squares. Black circle indicates high Bayesian posterior probability support (> 0.75) as well as a high maximum likelihood bootstrap support (>75). Black square indicates only a high Bayesian posterior probability support (> 0.75). HP = Himachal Pradesh; JnK = Jammu & Kashmir; Uk = Uttarakhand. #Sequenced in this study.

Fig. 4: Neighbour-Joining tree based on 6 colour independent characters. *S. hypoleucos* is one of the two species of Hanuman langur from peninsular India

Fig. 5: Species distribution maps of Himalayan langur (a) and *S. entellus* (b).

a

b

a

b