# Species identification for small mammals of Western Himalayas using genetics and morphometrics


A thesis submitted towards partial fulfilment of

BS-MS Dual Degree

By

Mukta Suhas Joshi

20131115

Indian Institute of Science Education and Research, Pune

Under guidance of

Dr. Uma Ramakrishnan

National Centre for Biological Sciences, Bangalore

## Certificate

This is to certify that this dissertation entitled "**Species identification for small mammals of Western Himalayas using genetics and morphometrics**" towards the partial fulfilment of the BS-MS dual degree programme at Indian Institute of Science Education and Research, Pune represents original research carried out by **Mukta Suhas Joshi** at **National Centre for Biological sciences**, **Bangalore** under supervision of **Dr. Uma Ramakrishnan**, Associate Professor at Department of Ecology and Evolution during the academic year **2017-2018**

**Signature of Supervisor** **Date**: 30/03/2018

## Declaration

I hereby declare that the matter embodied in the report entitled "**Species identification for small mammals of Western Himalayas using genetics and morphometrics**" are results of investigations carried out by me at Department of Ecology and Evolution, **National Centre for Biological sciences**, under the supervision of **Dr. Uma Ramakrishnan** and the same has not been submitted elsewhere for any other degree.

**Signature of the Student** **Date:** 30/03/2018

# Abstract

Rodents are diverse yet one of the most neglected mammalian group. The Himalayas, which are known as a biodiversity hotspot remain poorly studied in terms of rodent species assemblages. We used Cytochrome b phylogeny and Multivariate statistics together to identify the species of forest mice of genus *Apodemus* and voles of genus *Alticola* from Western Himalayas. From genetic data, we identified *Apodemus* species from all the sampling locations as *Apodemus pallipes.* Some of the individuals of *Apodemus* were identified as *Mus musculus castaneus* from both genetic and morphometric data. We suspect that some of the individuals identified as *A.pallipes* individuals might be of *Apodemus rusiges* because ranges of these two species broadly overlap in the Western Himalayas. We identified two species for *Alticola* genus, *Alticola barakshin, Alticola montosa* and one from *Hyperacrius* genus, *Hyperacrius fertilis.* The two clusters from our genetic data are not sister to any of the known species. Ideally, we would identify exact species status by comparing our genetic data to those from voucher specimens. We attempted to generate the latter with specimens from BNHS, but preliminary results did not yield any target sequences. As a result, for the purpose of present study, we identified them as *Neodon sp*., and *Alticola sp*. We also built species distribution models to identify suitable habitats for the species with ranges in Western Himalayas. Our distribution models revealed suitable habitats outside existing known ranges, which was consistent with our field sampling. Future work will include work on museum specimens and sequencing nuclear genes to infer the colonisation history of these two genera.

# List of tables



# List of figures

# Acknowledgements

I would like to thank my thesis supervisor Dr. Uma Ramakrishnan for making me part of this project which proved to be really amazing learning experience for me. One year of thesis period gave me some valuable experiences which helped in my growth both academically and personally. I would like to thank my TAC member Prof. Milind Watve from IISER Pune for giving me opportunity to work with him on various interesting projects at IISER which definitely aided in my development as a researcher. I would like to thank my mentor in this project Dr. Vivek Ramachandran for my first ever unforgettable field work experience and for constantly guiding me throughout the thesis period. This project work was carried out under his DST grant number YSS/2015/000339. I would like to thank Mayuresh Ambekar for helping out with the labwork, Nishma Dahal for helping out with the alignments in geneious, Aparna Lajmi for useful inputs regarding the data analysis and all other lab-3 members for their constant help and support. I'm grateful to the members of sequencing facility at NCBS and Dr. Vicki Thomson from University of Adelaide for helping me out in troubleshooting in the initial part of this project. I would like to thank my friends and family for being my support system. Last but not the least, I would like to thank IISER Pune for giving me this opportunity to work at such a vibrant research institution and overall shaping my attitude towards research as a career.

# Introduction

Small mammals are the most abundant, globally diverse and least known mammalian group (Nowak R.M., 1999). The three small mammal orders- Rodentia (rodents), Scandentia (tree shrews) and eulipotyphlans (consisting shrews, moles, hedgehogs and solenodons) together contain more than 2800 species (About 40% of mammalian fauna) out of which 437 are threatened with extinction as per assessment by International Union for Conservation of Nature (IUCN, 2011). Small mammals have adapted to the wide range of habitats and are often crucial to the healthy functioning of ecosystems (Carleton M.D., 1984). Yet, they are not very well known in terms of their ecology, evolutionary history and diversity (Nowak R.M., 1999).

The order Rodentia is the largest mammalian order which encompasses 33 families, 63 genera and 2277 species (Musser and Carleton 2005). All the rodents are characterized by a single pair of upper and lower constantly growing incisors (Carleton M.D., 1984). Within Rodentia, the rodents of family Muridae are the most speciose group of mammals (Michaux J. et al, 2001). They consist nearly one third of total species diversity of Rodentia with many taxonomically diverse groups (Musser and Carleton, 1993). They are known to occur on every major landmass in the world (Steppan et al, 2004). Many muroid species are viral reservoirs and vectors of human diseases which makes studies on their ecology and phylogeny important (Jansa et al, 2003). Molecular studies on small mammals are also useful in understanding how past environmental changes affected the movements of lineages across continents (H. Suzuki et al, 2008).

Species of genus *Apodemus,* which belong to Murinae subfamily (commonly known as wood mouse) are widely distributed rodents inhabiting broad-leaf forests in the temperate zone of palaeartic region (Sakka et al, 2010). Phylogeography of this genus has received considerable attention during last few years (JR Michaux et al., 2005; Sakka et al., 2010, H.Suzuki et al., 2008). About 22 *Apodemus* species are known, half of which are from Europe and remaining from Asia (Musser and Carleton, 2005). *Apodemus* species are found in variety of habitats such as forests, grasslands and they usually feed on acorns, insects and other small invertebrates (H. Suzuki et al, 2008). Recent studies on molecular phylogeny of *Apodemus* which are extensively based on mitochondrial Cytochrome b gene and IRBP (Interphotoreceptor Retinoid

Binding Protein) gene (Serizawa et al., 2000; Michaux et al., 2002; Suzuki et al., 2003; Liu et al., 2004;) have revealed four distinct lineages of *Apodemus*; the *Sylvaemus*, the *Apodemus*, *A. argenteus* and *A. gurkha* groups. However, the systematics of sub-genus *Sylvaemus* (cosisting of species *A. alpicola, A. uralensis, A. fulvipectus, A. hermonensis, A. flavicollis, A rusiges and A pallipes*) is complicated (H.Suzuki et al, 2008) and taxonomic problems are rooted in misidentification of specimens resulting in confusions regarding species identifications (Hoofer et al, 2007). The three species of sub-genus *Sylvaemus, Apodemus pallipes*, *Apodemus rusiges* and *Apodemus uralensis* are known to have ranges in India. The phylogenetic relationships among species of this subgenus are not very well established (H. Suzuki et al., 2008) and there is considerable gap in knowledge about their taxonomic status (Liu et al, 2004).

The name *Alticola* (Blanford, 1881) comes from Latin *cola* meaning inhabiting and *altus* meaning high, i.e. the mouse that is found in high altitudes (Krystufek et al, 2016). These are also known as Central Asian high mountain or rock voles and are one of the least known groups of voles in terms of their evolutionary history (Lebedev et al, 2007). Species of *Alticola* usually inhabit steppe, rocky montane and alpine habitats (Kohli et al, 2015). This genus is included in the tribe Clethrionomyini (Arvicolinae, Cricetidae, Rodentia) but its position within this tribe remains problematic (Lebedev et al, 2007). The genus *Alticola* is known to have 12 species and is grouped into 3 subgenera: *Alticola* s.str., *Platycranius* Kastschenko, 1899 and *Aschizomys* Miller, 1899 (Musser and Carleton, 2005). The current taxonomic knowledge about subgenus *Alticola* s.str*.,* is based on revision by Rossolimo and Pavlinov (1992) who identified eight species within this subgenus:1) *Alticola argentatus* 2) *Alticola montosus* 3) *Alticola albicaudus* 4) *Alticola semicanus* 5) *Alticola tuvinicus* 6) *Alticola Olchonensis* 7) *Alticola stolikzcanus* (stoliczka's mountain vole), 8) *Alticola roylei*

Table (1) summarises the list of species of these two genera that are known to have ranges in India and their type locality. The only information available about these species is the sampling records from early 1900s which were largely collected from Western parts of Himalayan ranges.

| Species | Type Locality |
|---|---|
| *Apodemus pallipes* | Eastern Tajikistan, Pamir Altai |
| *Apodemus rusiges* | Northern India, Central Kashmir |
| *Apodemus uralensis* | Russia, South Ural mountains |
| *Alticola montosa* | Central Kashmir, India |
| *Alticola roylei* | Kumaon, India |
| *Alticola stolikzcanus* | North-West India, Ladakh, Nubra valley |
| *Alticola argentatus* | Tajikistan, Pamir mountains |
| *Alticola albicaudus* | India, Baltistan, Braldu Valley |

**Table (1):** Species of *Apodemus* and *Alticola* known to have ranges in Western Himalayan part of India (as per IUCN Redlist 2017) and their type localities (Musser and Carleton, 2005)

**Objectives and scope of the project**

The objectives of this work can be summarised as follows:

1) To integrate the genetic and morphometric data in our study to identify and delimit the species of *Apodemus* and *Alticola*

2) Species distribution modelling for selected species to identify the suitable habitats.

As discussed above, the taxonomy species of the two genera *Apodemus* and *Alticola* from Western Himalayas remain poorly resolved. However, this issue has not been addressed in this thesis due to time constraints.

## Part 1: Genetics and Morphometrics

Defining species has been an unsolved problem in biology since a long time (Pereria et al., 2008). Currently, there are close to 26 definitions of species in the literature most important of which include Biological Species Concept (BSC), Phylogenetic Species Concept (PhSC), Phenetic Species concept (PSC), Evolutionary species concept, Genotypic Cluster Definition (Housdorf, 2011). However, there is little agreement between these concepts regarding the criteria employed by them to delimit the species. Accurately identifying and delimiting the species is important for understanding many evolutionary patterns and processes (Sites et al, 2003). Since species are commonly used as fundamental units of analysis in biogeography, ecology, macroevolution and conservation biology (Sites et al, 2003), it is important to employ the methods that delimit the species objectively and rigorously.

For the past 250 years, phenotypic characteristics or morphological features of an organism have been the base of taxonomy and species identification (Herbert et al, 2005). The term “morphology” refers to the external features or a structure of an organism. The Morphology based taxonomy only identifies ‘morphospecies’ (Cain, 1954), that is, species exclusively established on morphology. However, there is considerable amount of morphological plasticity that exists between organisms. This poses limitations to traditional morphology based identification since it can become time consuming (Pereria et al., 2008). Use of morphology can also become a problem in case of sister species complexes and recently diverged lineages since there could be significant amount of morphological similarity in spite of being reproductively isolated (M. Pfenninger et al., 2007). This also holds true for cryptic species complexes, which are genetically divergent but are considered as single species due to their morphological similarity (M. Pfenninger et al., 2007). Thus, morphology based species identification can easily lead to field misidentifications.

Molecular genetic methods of species identification are based on neutral theory of molecular evolution (F.Pereria, 2008). According to this theory, molecular changes are accumulated over time which leads to divergence of different lineages. These changes are assumed to be neutral (Kimura, 1968). DNA based methods have advantages over the morphology based identification because morphological characters represent only a small fraction of the species genome (Hilis, 1987). DNA based techniques identify the species on the basis of single gene sequence similarity. These group of techniques are collectively called “DNA barcoding” (Herbert et al, 2003). The species identification methods that come under this term are:

1) BLAST: BLAST which is short form for Basic Local Alignment Search tool assigns the query (unknown) sequence to a set of reference sequences (which have already been identified) on the basis of similarity (Ross et al, 2008). BLAST is easy-to-use and informal method of identification. NCBI reference database (www.ncbi.nlm.nih.gov/BLAST) gives the best alignments to all or a part of the query sequence. However, if a reference sequence is not present in the database for a particular species, BLAST tend to show incorrect identity for a query sequence which makes it unreliable for accurate species identification (Agarwal and States, 1998).

2) Distance-based methods: Genetic distance is a measure of divergence in the sequences that have evolved from common ancestor. Genetic distance methods are

commonly used in barcoding studies to identify the species based on a generally accepted threshold (Herbert et al, 2003). P-distances, or pairwise distance is one the simplest approaches which is calculated as number of nucleotide differences per site. However, this approach has shortcomings: if degree of divergence is high, then p-distances are generally not very informative with regard to number of substitutions that actually occurred (Salemi and Vandamme, 2009).

3) Tree based methods: Tree based methods delimit the species on the basis of principals of phylogenetics (Sites et al, 2003). Phylogenetic trees are routinely used to infer the evolutionary relationships between different biological entities. According to the evolutionary theory, all the organisms have evolved from one common ancestor. There are different mechanisms of acquiring variation which include mutations, duplication of genes, genetic exchanges such as lateral gene transfer. Phylogenetic methods are based on similarity among the genes, assuming that they are homologous. Higher the similarity between the sequences under investigation, more are the chances that the sequences are derived from common ancestor. It is known from comparative taxonomic studies that the genes of closely related species usually differ only by a limited number of point mutations (Salemi and Vandamme, 2009). Phylogenetics infers the common history of gene fragments and establishes the relationships between them. Usually the genes that code for catalytic sites or the core of the proteins are more conserved than the others. Such gene fragments are mostly the target of phylogenetic studies. Mitochondrial Cytochrome b is one of such most targeted gene fragment. Mitochondrial genes are usually used to resolve the relationships within recently evolved groups (Galewski et al, 2006). The mitochondrial DNA codes for the proteins involved in electron transport chain of mitochondria (Pereria et al., 2008). Since the animal mitochondrial DNA usually has a high mutation rate, there is a large amount of genetic variation that is present in closely related species (Pereria et al., 2008). This can be useful for species identification procedures. Mitochondrial DNA is uniparentally inherited without recombination and is also easier to obtain from degraded or low-quality DNA samples it has lots of copies present in cells (Pereria et al., 2008).

However, since mitochondrial genome is maternally inherited in some species, using mtDNA information has limitations. Also, due to possible existence of mitochondrial DNA nuclear copies, known as pseudogenes or numts, nuclear genes are increasingly being used as markers in addition to mitochondrial markers in phylogenetic studies

(Barbosa, 2013). Nuclear genes are usually used to investigate the phylogenies at deeper level (Galewski et al, 2006). Inclusion of nuclear markers can help in detection of hybridization, incomplete lineage sorting, etc (Alves et al, 2006; Heckman et al, 2007).

**Need for integrating multiple approaches-Integrative taxonomy**

As discussed above, traditional morphology based approach is increasingly being replaced by molecular genetic methods of species identification (DNA barcoding). However, there doesn't necessarily have to be discrepancy between these two methods. Different approaches can be used for particular cases depending on the results they provide. DNA-based system of identification can only work if all the species have their unique diagnostic sequence in the database (B. Dayrat, 2004). The incomplete database might not lead to actual identification, but will identify the unknown entity as a member of new species. Thus, employing methods that will delimit the species from multiple perspectives is necessary.

## Part 2: Species Distribution Modelling

'Ecological Niche models' or 'Species distribution models' are widely used modelling methods that predict geographic distribution of species by relating the known occurrences of species to environmental features to infer ecological properties and predict geographic occurrences (Peterson AT, 2006). These modelling methods are based on the principle that an organism can sustain in an environment that suits its physiological setting (Hirzel & Lay 2008). This environmental setting is often referred as a niche of a species. Niche is multi-dimensional ecological construct which defines optimum environment for growth, reproduction and survival of species (Hutchinson, 1957). However, in nature species rarely fill their entire fundamental niche due to factors such as a dispersal barriers and biotic interactions that limit their range. The realised niche space is therefore always smaller than the fundamental niche. Species Distribution modelling (SDM) methods relate the species presence to the environmental predictors to extrapolate fundamental niche outside its realised niche, i.e., the locations where a species is present (Parolo et al, 2008).

Due to climate change, many species are rapidly shifting their distributions to adapt to the changing environments. The impact of climatic change is not uniform across the globe. Montane regions, Himalayas in particular, are considered as most vulnerable

region as the rate of warming is shown to be higher than global average (Shrestha et al., 2012). Therefore, the biodiversity in the Himalayan region could be at the risk of habitat loss.

Some of the species of paleartic *Apodemus* and *Alticola* which have ranges in Western Himalayas are narrowly distributed and are endemic to Western Himalayas. *Alticola montosa*, which is endemic to northwestern parts of Himalayan ranges is assessed as vulnerable by IUCN (2017) with the two known populations, from Jammu Kashmir and North-West frontier province in Pakistan being fragmented. While *Apodemus pallipes* is fairly broad-ranging species, *Apodemus rusiges* is endemic to Himalayas of North-Eastern Pakistan and Himalayas in northwestern India (Jammu and Kashmir). SDM could prove to be useful tool to identify the possibility of existence of suitable habitats outside the known ranges of these species. In the present study, we modelled contemporary distribution of two narrow-ranging and one wide ranging species of these two genera to elucidate their suitable habitats.

# Materials and Methods

### *1) Sample collection*

Sample collection was in two field seasons (2016 and 2017) at four different regions that span the Western part of Himalayan ranges. Geographically, Himalayas is divided into four parallel zones: (1) Outer Himalayas- the Siwalik ranges (2) Middle or lesser Himalayas- Pir Panjal and Dhauladhar ranges (3) the Great Himalayas (4) trans-Himalayas. The Western Himalayas consist of Zanskar range, Pir Panjal range, part of Siwalik range and the Greater Himalayas (Karan, 1996). Our sampling locations included: Dachigam National park, Overa-Aaru wildlife Sanctuary and Great Himalayan National Park (GHNP) which falls in Greater Himalayan ranges and Ladakh and parts of Karakoram range, Changtang (Tibetan plateau) and Spiti valley which are part of the trans-Himalayan ranges. The time periods and elevational transects covered for each of these regions are summarised in table 2.

Samples were collected covering the altitudinal range from about 1500 metres to 6000 metres. Each elevational trascect was sampled for the period of 5 days. This included setting up traps on first day and sampling over the period of next 4 days. Sampling protocol followed was: Setting up Sherman and Tomahawk traps (see appendix) in 10

metres *10 metres grid. Traps were baited and placed at the distance of 10 metres from each other. Animal captured overnight were processed next day for getting hair, ear punch and tail tissue samples. Morphometric measurements such as Tail Length (TL), Hind-leg Length (HL), Ear Length, and Body Length (HBL) were taken. Trapped individuals were released after taking ear punch and body measurements. The samples were then taken to lab and preserved at -20°C.

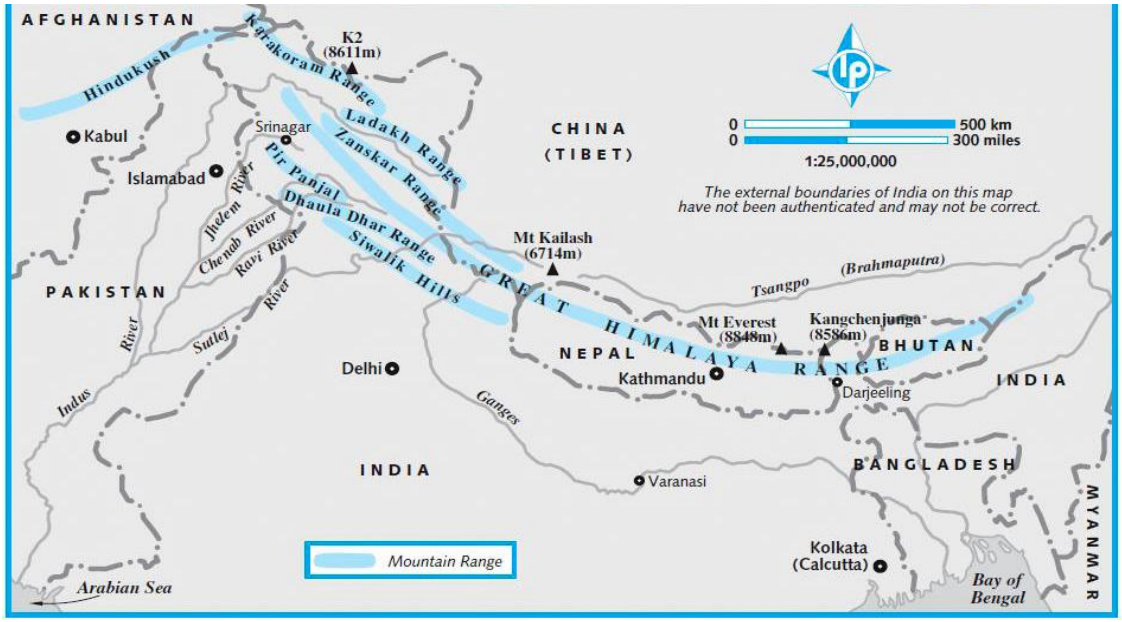


**Figure (1):** The four parallel ranges of Himalayas

| Field season | Region in Western Himalaya | Elevational transect covered | Total no of individuals captured |
|---|---|---|---|
| May-June, 2016 | Kashmir valley, J & K | 1500- 4200 m | 60 |
| August-September, 2016 | Ladakh, J& K | 3200-5500 m | 162 |
| April-June, 2017 | Great Himalayan National Park, Himachal Pradesh | 1500-5000 m | 87 |
| August-September, 2017 | Spiti valley, Himachal Pradesh | 3500-6000 m | 40 |

**Table (2):** Sampling details

Since the rodents of Western Himalayas are not very well studied, no genetic data has ever been generated for some of the species of these two genera such as *Apdoemus rusiges, Alticola roylei, Alticola stoliczkanus, Hyperacrius wynnei.* Hence, we visited

Bombay Natural History Society (BNHS) to study museum collections of these species which were recorded in early 1900s. We collected the crusties (the pieces of tissue that remain after cleaning the skull) from these museum specimens.

## Part 1: Genetics and Morphometrics

*1) DNA extraction, amplification and sequencing*

Genomic DNA was extracted for 278 samples collected from 4 locations (mentioned in table 2) from small ear pieces preserved in absolute ethanol using Qiagen DNAeasy blood and tissue kit following the protocol given by manufacturer. 1140 base pairs fragment of cytochrome b gene was amplified by PCR with primer pairs L14724_hk3 and H15915_hk3 (Zhang et al, 2016, Kocher et al, 1989). All the PCR's were carried out in 20 µl reaction mixture including 6 µls of DNAase/RNAase free PCR water, QIAGEN taq master mix, 1 µls of each forward and reverse primer with concentration ranging from 2 µM to 7 µM . PCR was performed in Eppendorf thermal cycler gradient under following parameters: a pre-denaturation step at 95°C for 10 min; 40 cycles of denaturation at 94°C for 45 seconds, annealing at 50°C for 50 seconds, extension at 72°C for 1 min 20 seconds; plus a final extension at 72°C for 5 minutes. PCR products were electrophoresed on a 1% agarose gel, visualized with Gel Red/Orange G staining to verify polymerase chain reaction quality. PCR products were further purified using QIAquick PCR purification kit to obtain clean sequences following manufacturer's instructions. Sequences were obtained at Sanger sequencing facility at National Centre for Biological Sciences and at Chromegene Biotek Private company.

| | Total no of individuals processed | Samples that worked | Samples that failed | Nuclear copies amplified |
|---|---|---|---|---|
| Apodemus | 108 | 101 | 7 | - |
| Alticola | 170 | 114 | 18 | 38 |

**Table (3):** Success rate of individuals processed

### *2) Sequence editing and alignment*

Out of 278 extracted tissue samples, about 215 individuals of *Apodemus* and *Alticola* and *Microtus* (as per field ID) were processed further for species identification (See table 3). The raw forward and reverse sequences obtained were assembled in Geneious 6.1.8 (http://www.geneious.com/) using Geneious assembler with default parameters. The complete assembly including contig region was checked for mismatches and gaps. Bad quality bases at the ends were trimmed. Consensus sequences obtained after editing raw sequences were further used for constructing phylogenetic trees.

For both the *Apodemus* and *Alticola* groups, the mitochondrial cytochrome b sequences from literature were retrived from Genbank. These sequences were obtained from the papers that describe most recent classification of these two genera (Liu et al, 2004 for *Apodemus* and Lebedev et al, 2007 for *Alticola*). Newly acquired sequences were aligned with previously published sequences using muscle alignment in Geneious. The alignment was manually checked for insertions, deletions and stop codons. These sequences were translated according to vertebrate mitochondrial genetic code. Final alignment was adjusted to a length of 1140 base pairs.

- ***Brief description of terminologies used in phylogenetics***

1) Models of molecular evolution: These are basically sets of assumptions about the process of nucleotide substitution. Different models consist of parameters that describe the rates at which one nucleotide replaces another. These parameters define a rate matrix that is used to calculate the probability of evolving from one base to another. In general, more complex model fits the data better because they have more parameters (Salemi and Vandamme, 2009)

2) Maximum likelihood framework and Bayesian framework: In Bayesian inference, the probability of a hypothesis (tree) or a model conditioned on observed data (sequences) is estimated. While Maximum Likelihood (ML) approach estimates the probability of data (sequences) given the model (tree). ML approach is a point estimate while Bayesian approach estimates the distribution.

3) Bayesian Posterior Probability: In Bayesian statistics, prior is the probability distribution that represents uncertainty before you have sampled any data. Posterior

is the probability distribution representing your uncertainty after you have sampled data. It is a parameter of confidence or support for branches while inferring a tree in Bayesian framework. The probability value above 0.95 is considered as a very good support.

4) Bootstrapping/ Bootstrap Support: Bootstrap support values indicate how many times out of 100 the same branch was observed when repeating a phylogenetic reconstruction on a re-sampled set of data. Bootstrap support above 95% is very good and very well accepted. Bootstrap support between 75% to 95% is reasonably good, anything below 75% is a poor bootstrap support. Booststrap support less than 50% is usually not shown on a phylogenetic tree.

5) Outgroup: Out group is an organism or group of organisms that serves as a reference group which inferring the evolutionary relationships of an ingroup (organisms under study). Outgroup is usually used to root the unrooted networks or to infer hypothetical ancestral states (Luo A-Rong et al, 2010). Appropriate choice of outgroup is critical because topology of ingroup tree can vary with the choice of outgroup taxa (Luo A-Rong et al, 2010).

6) Polytomies: Polytomies are unresolved branches on phylogenetic tree.

***3) Data partitioning and Model Selection***

Several different approaches have been proposed in last few years, to select the best-fit model of evolution. Some examples are: 1) Hierarchical Likelihood Ratio tests (hLRT) 2) Information criteria 3) Bayesian approaches. In the present study, PartitionFinder version 2.1.1 (Lanfear R., 2016) was used to select the appropriate model of molecular evolution under the Akaike Information Criterion (AIC) value and to select the best partitioning scheme for each dataset. Data partitioning essentially involves splitting site in the alignment into the sets that have evolved under similar models. Because of the triplet structure of the genetic code, different codon positions tend to evolve at different rates and experience different substitutional process (Lanfear R., 2016).

For both Maximum Likelihood and Bayesian trees, two separate runs were performed in PartitionFinder. Table 4 summarises the partitioning schemes and best-fit models for each partition as described by PartitionFinder.

| Group | Mode of tree construction | Partitioning scheme | Best model chosen under AIC |
|---|---|---|---|
| *Apodemus* | RaxML | 1) Gene1_pos2, Gene1_pos1<br>2) Gene1_pos3 | GTR+I+G<br>GTR+I+G |
| | MrBayes | 1) Gene1_pos1, Gene1_pos2<br>2) Gene1_pos3 | HKY+I+G<br>GTR+I+G |
| *Alticola* | RaxML | 1)Gene1_pos1,Gene1_pos2, Gene1_pos3 | GTR+I+G |
| | MrBayes | 1) Gene1_pos1, Gene1_pos2, Gene1_pos3 | GTR+I+G |

**Table (4)**: Best partitioning schemes and best-fit models as described by PartitionFinder

### *4) Phylogenetic reconstruction*

Maximum likelihood trees (ML) trees were constructed using RaxML (version 1.5b2) with rapid bootstrap covering 100 replicates. Bayesian Posterior Probabilities (BPP) were calculated in MrBayes (Version 3.2.6) with two independent runs consisting of four heated chains and one cold chain and burning fraction of 0.30. Trees and parameters were sampled for every 1000 generations with total of 20 million generations.

### *5) Morphometric Analysis*

Principal Component Analysis (PCA) is a multivariate dimensionality reduction technique used to extract important variables from a data set consisting of variables available in the data set. It produces linear combination original variables to generate the axes which are a set of orthogonal variables known as Principal Components, or PCs. The first principal component usually retain most of the information present in the dataset. Principal Components are usually interpreted based on finding which variables are most strongly related with each component. First principal component usually determines the direction of highest variability in the data. Linear Discriminant analysis (LDA) or Discriminant Function Analysis (DFA) on the other hand maximizes the separation between multiple classes. In other words, this method maximizes the ratio of between-class variance to the within class variance. DFA is used to determine which variables discriminate between two or more naturally occurring groups.

| | |
|---|---|
| Ear | Length of the ear |
| HBL | Head body Length from snout to anus |
| HL | Hind-leg length |
| Weight | Weight of an individual |
| TL | Tail Length |

**Table (5):** Details of morphometric measurements taken

The morphometric measurements that were taken (summarised in table 5) were used as variables in PCA and DFA. Plots were obtained after removing the juvenile individuals and weight variable.

## Part 2: Species Distribution Modelling

The model for predicting current distribution of *Apodemus pallipes, Apodemus rusiges* and *Alticola montosa* was developed in MaxEnt Version 3.3.3 (Philip et al, 2006). MaxEnt takes the presence records of species (known as Presence Only data) and a set of environmental variables as input and estimates the probability of presence of species (Merow et al, 2013). Environmental predictor variables are divided into grid cells across user defined landscape. A sample of background locations is extracted from this landscape and contrasted against the presence locations (Merow et al, 2013).

SDM procedure involved 3 steps: (1) Data preparation (2) Variable selection (3) MaxEnt modelling

### Data Preparation

We used the occurrence data obtained from sampling localities (after their correct identification through phylogenies), museum records (provided by Bombay Natural History society), occurrences from GBIF database (http://www.gbif.org/) for modelling the distribution. The occurrence data was thinned using spatial thinning procedure for those species having high number of closely distributed sample points in order to account for spatial sampling biases. For thinning, we randomly removed occurrences that were within 1 km. There were 45 and 27 occurrences for *Apodemus pallipes* and *Alticola montosa* before filtering which were then filtered to 36 and 19 unique

occurrences, respectively. For *Apodemus rusiges*, we selected 20 occurrence points out of 25.

**Predictor variables**

The elevation data was acquired from the Shuttle Radar Topographic Mission (SRTM v. 3) digital elevation model (DEM) at a resolution of 30 arc-seconds. Total of 19 bioclimatic variables downloaded from Worldclim database (www.worldclim.org) at a resolution of 30 arc-seconds. All the variables were masked to include only 55° to 85° East and 26° to 43° North. For each location, bioclimatic variables were extracted using QGIS ver.2.14.2. Apart from 19 bioclimatic variables, several variables were derived which included Terrain Roughness Index, slope, and eastness. The details of derived variables are provided below.

Terrain Roughness Index (TRI): TRI was calculated to account for the topographic heterogeneity represented by Himalayas. It was derived from elevation layer in QGIS 2.18.9. It is essentially the amount of elevation difference between the value of a cell and the mean of an Eight-cell neighbourhood of surrounding cells.

Slope: Slope is derived from elevation in QGIS 2.18.9

Eastness: Eastness is calculated as sin aspect in degrees. The value ranges from -1 to 1. A value of one indicates east facing slope.

All 19 bioclimatic variables and derived variables were tested for collinearity by examining pairwise Pearson's' correlation coefficient between them. It is important to remove variables that are correlated because it can result in wrong identification of relevant predictor. The cut-off of pairwise Pearson's coefficient was set to 0.8, to exclude the correlated variables.

After testing for collinearity, the variables selected for *Apodemus pallipes* included Slope, TRI, Mean Diurnal Range (Bio2), Isothermality (Bio3), Mean Temperature of wettest quarter (Bio8), Mean temperature of driest quarter (Bio9), Precipitation seasonality (Bio15) and altitude. The selected variables' contribution to the model was assessed by using jack-knife test (systematically leaving out one variable at a time and a regularized gain change) in MaxEnt, following which, the contribution of each variable was assessed hierarchically. Temperature of wettest quarter (Bio8) and

Eastness variables were dropped because of their low contribution to the model. Thus, final set of variables included Slope, TRI, Bio2, Bio3, Bio9 and Bio15 and altitude. Similarly for *Alticola montosa*, selected variables included Slope, Mean diurnal range (Bio2), Annual precipitation (Bio12), Precipitation of Driest month (Bio14), Precipitation seasonality (Bio15) and altitude. For *Apodemus rusiges*, the selected variables included Slope, Mean diurnal range (Bio2), Precipitation of driest month (Bio14), Precipitation seasonality (Bio15), Precipitation of coldest quarter (Bio19) and Temperature Annual Range (Bio7).

**MaxEnt modelling**

The habitat suitability map was produced using logistic output in MaxEnt. Logistic output gives the probability of species' presence in the form of values which range from 0 to 1. Suitable habitats are distinguished based on these probability values. Distinction between suitability of different areas is made based on these probability values. Jackknife procedure was used to test the contribution of each bioclimatic variable to the model. Model was run with 5000 iterations and 10 replicates for each species and model performance was evaluated with Area under Receiver Operating Characteristic Curve (also known as AUC).

# Results

## Part 1: Genetics and Morphometrics

1) Species identification for species of genus *Apodemus*

The individuals are coded according to their grid IDs. The letters correspond to the name of the locality from which the individual was collected, while the number correspond to the day of capture and serial number respectively. List of the individuals and localities from where they were collected is given in Appendix table A1. The position of the individual in the tree and the bootstrap support (>75) of the relationship is used as the criteria for species identification.

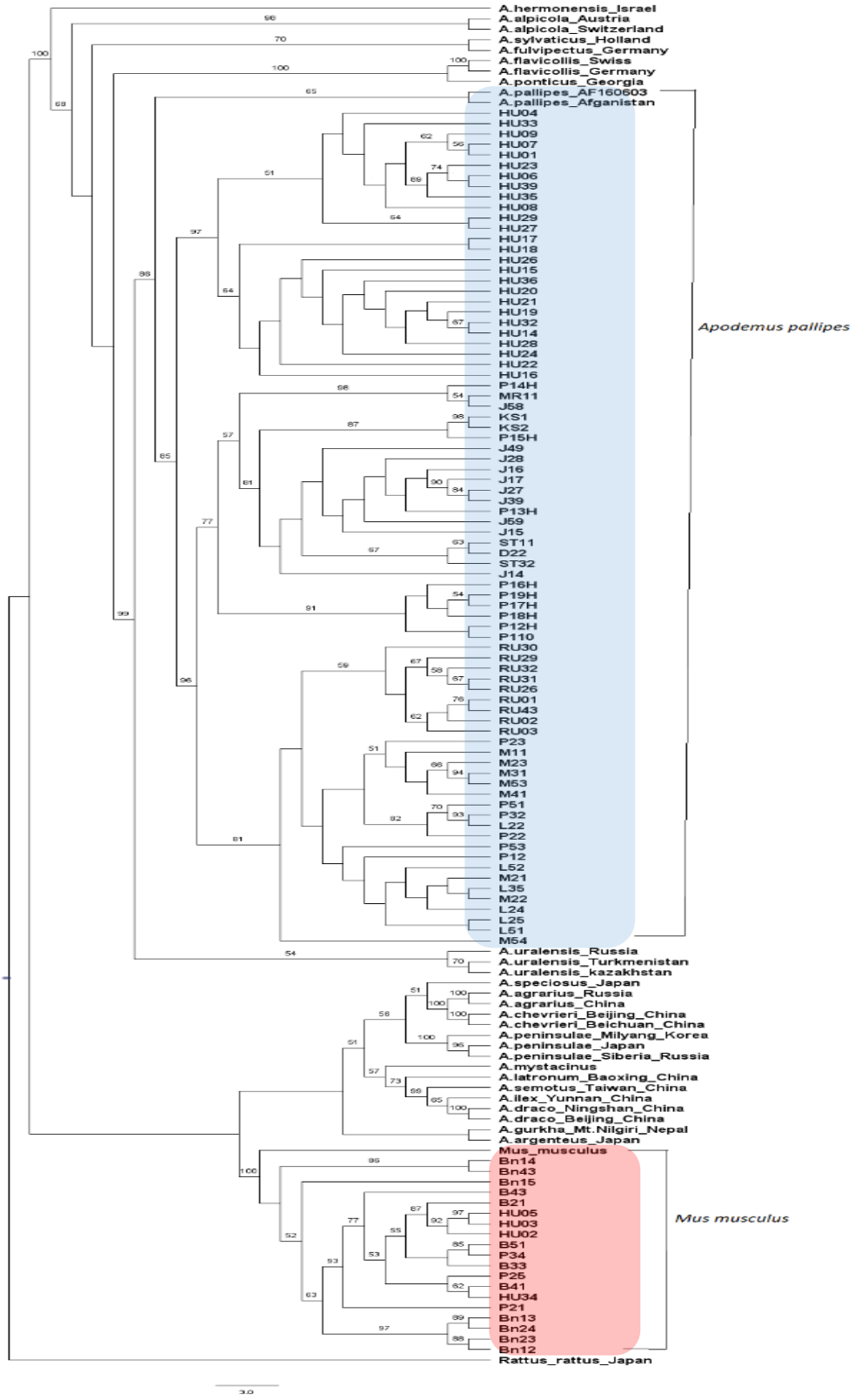


**Figure (2):**

Maximum-likelihood tree using RaxML. Numbers above the branches represent bootstrap support based on 100 replicates. Bootstrap values below 50% are not shown on the tree.

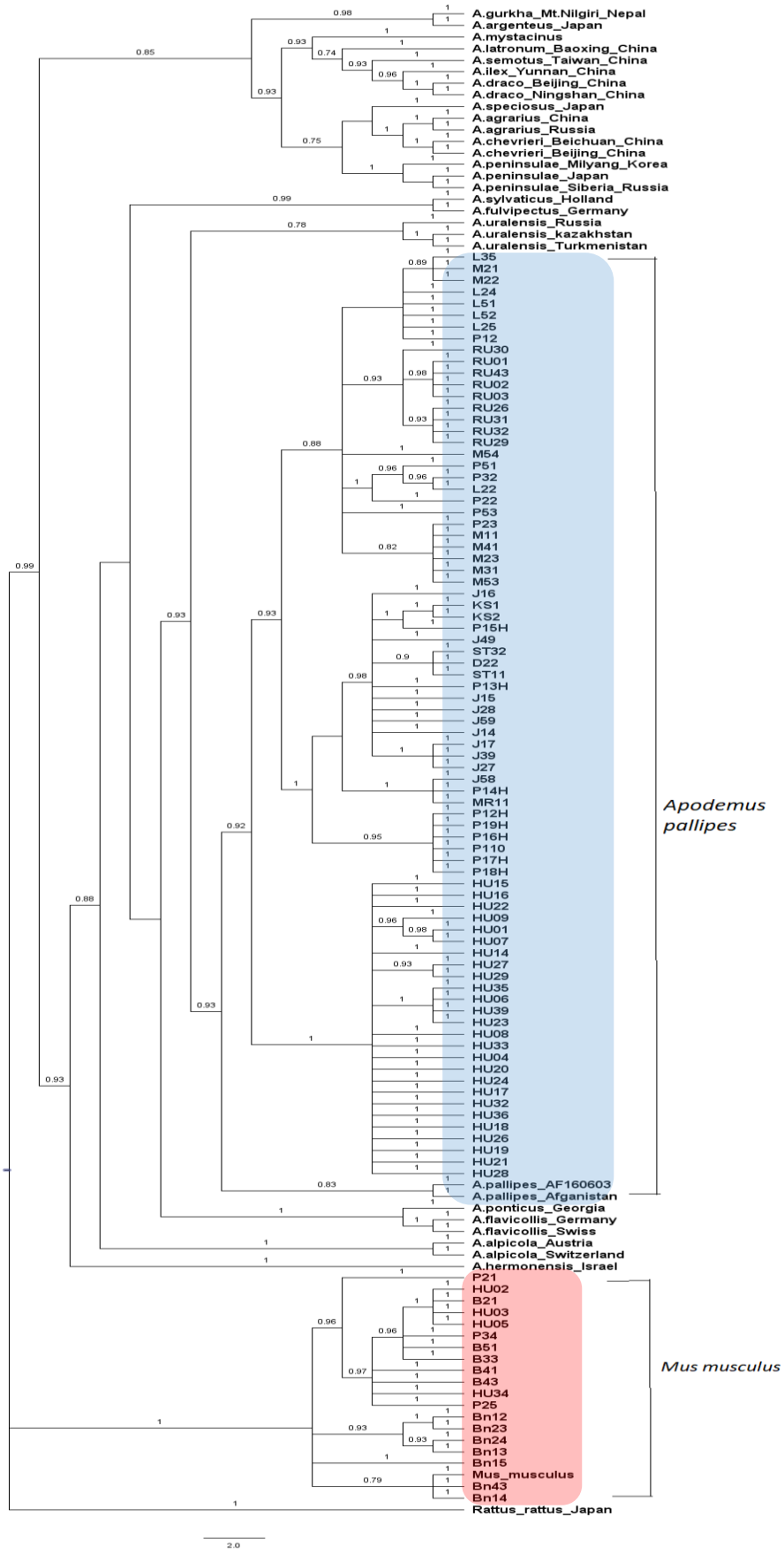
A.gurkha_Mt.Nilgiri_Nepal
A.argenteus_Japan
A.mystacinus
A.latronum_Baoxing_China
A.semotus_Taiwan_China
A.ilex_Yunnan_China
A.draco_Beijing_China
A.draco_Ningshan_China
A.speciosus_Japan
A.agrarius_China
A.agrarius_Russia
A.chevrieri_Beichuan_China
A.chevrieri_Beijing_China
A.peninsulae_Milyang_Korea
A.peninsulae_Japan
A.peninsulae_Siberia_Russia
A.sylvaticus_Holland
A.fulvipectus_Germany
A.uralensis_Russia
A.uralensis_kazakhstan
A.uralensis_Turkmenistan
L35
M21
M22
L24
L51
L52
L25
P12
RU30
RU01
RU43
RU02
RU03
RU26
RU31
RU32
RU29
M54
P51
P32
L22
P22
P53
P23
M11
M41
M23
M31
M53
J16
KS1
KS2
P15H
J49
ST32
D22
ST11
P13H
J15
J28
J59
J14
J17
J39
J27
J58
P14H
MR11
P12H
P19H
P16H
P110
P17H
P18H
HU15
HU16
HU22
HU09
HU01
HU07
HU14
HU27
HU29
HU35
HU06
HU39
HU23
HU08
HU33
HU04
HU20
HU24
HU17
HU32
HU36
HU18
HU26
HU19
HU21
HU28
A.pallipes_AF160603
A.pallipes_Afganistan
A.ponticus_Georgia
A.flavicollis_Germany
A.flavicollis_Swiss
A.alpicola_Austria
A.alpicola_Switzerland
A.hermonensis_Israel
P21
HU02
B21
HU03
HU05
P34
B51
B33
B41
B43
HU34
P25
Bn12
Bn23
Bn24
Bn13
Bn15
Mus_musculus
Bn43
Bn14
Rattus_rattus_Japan
Apodemus pallipes
Mus musculus
2.0

**Figure (3)**: Tree generated from MrBayes with Bayesian Posterior Probabilities (BPP) shown above the branches. BPPs below 0.7 are not shown in the tree.

The phylogenetic trees represented in figure 2 and 3 are based on classification of genus *Apodemus* as described by Liu et al 2004 and Suzuki et al, 2008. Out of 101 processed individuals (which were identified as species of *Apodemus* on field), about 19 individuals formed part of *Mus musculus* (common house mouse) clade with bootstrap support of 100%. There are 3 known sub-species of *Mus musculus*, *Mus musculus castaneus, Mus musculus domesticus and Mus musculus musculus*. To further identify which sub-species of Mus musculus our individuals are, we constructed a sub-species level tree of *Mus musculus* as described by Suzuki et al, 2013 and identified the individuals as *Mus musculus castaneus.* Rest of the 82 individuals from all 4 sampled regions (Kashmir, Ladakh, Himachal and spiti) formed part of *Apodemus pallipes* clade with bootstrap support of 86% (>75%) and with Bayesian Posterior Probability of 0.93 (>0.70). Thus, these individuals were identified as *Apodemus pallipes* by both modes of tree construction.

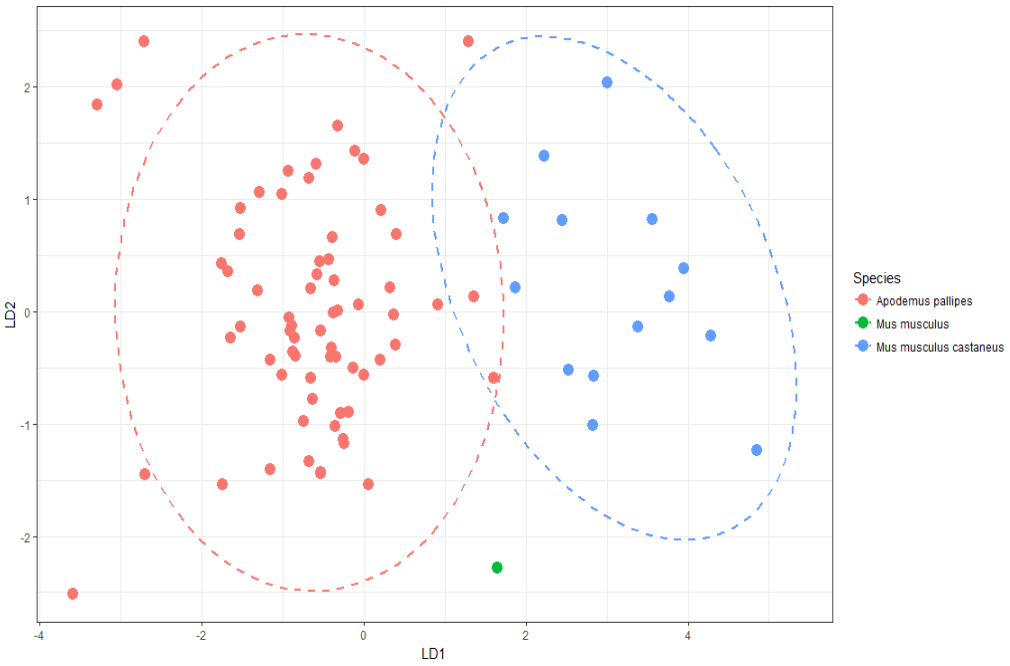


**Figure (4):** Discriminant Function Analysis (DFA) plot for *Apodemus.*

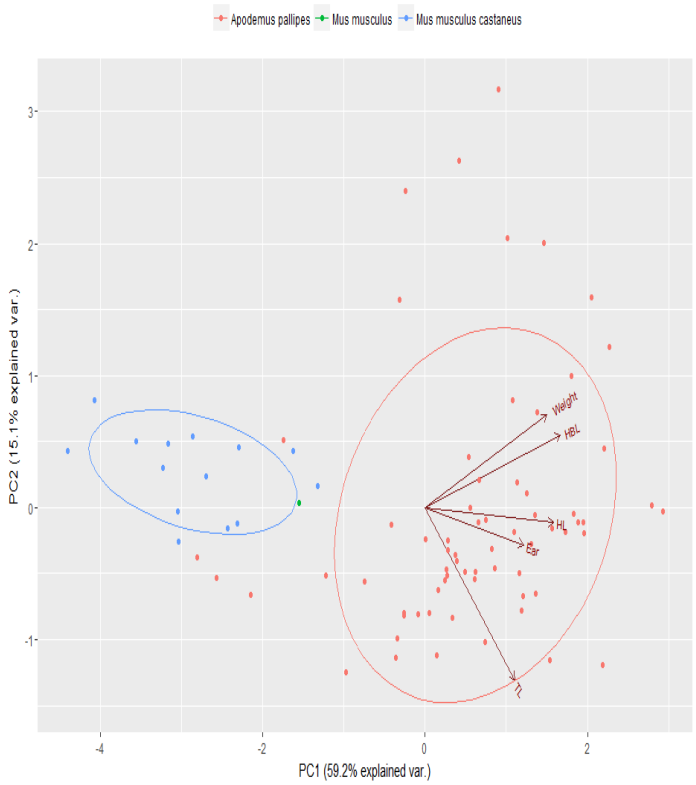


**Figure (5):** Principal Component Analysis (PCA) plot for *Apodemus*.

Both PCA and DFA plots revealed two distinct morphological clusters, one cluster of *Apodemus pallipes* and one cluster of *Mus musculus castaneus* (Figure 4 and Figure 5). Hind leg length (HL) and Ear were the variables majorly contributing to discrimination between groups (Appendix Table A3). Our genetic data also identified two separate clades which suggests that out genetic data was consistent with morphometric data.

2) Species identification for species of genus *Alticola*

The individuals are coded according to their grid IDs. The letters correspond to the name of the locality from which the individual was collected, while the number correspond to the day of capture and serial number respectively. List of the individuals and localities from where they were collected is given in Appendix table A2. The position of the individual in the tree and the bootstrap support (>75) of the relationship is used as the criteria for species identification

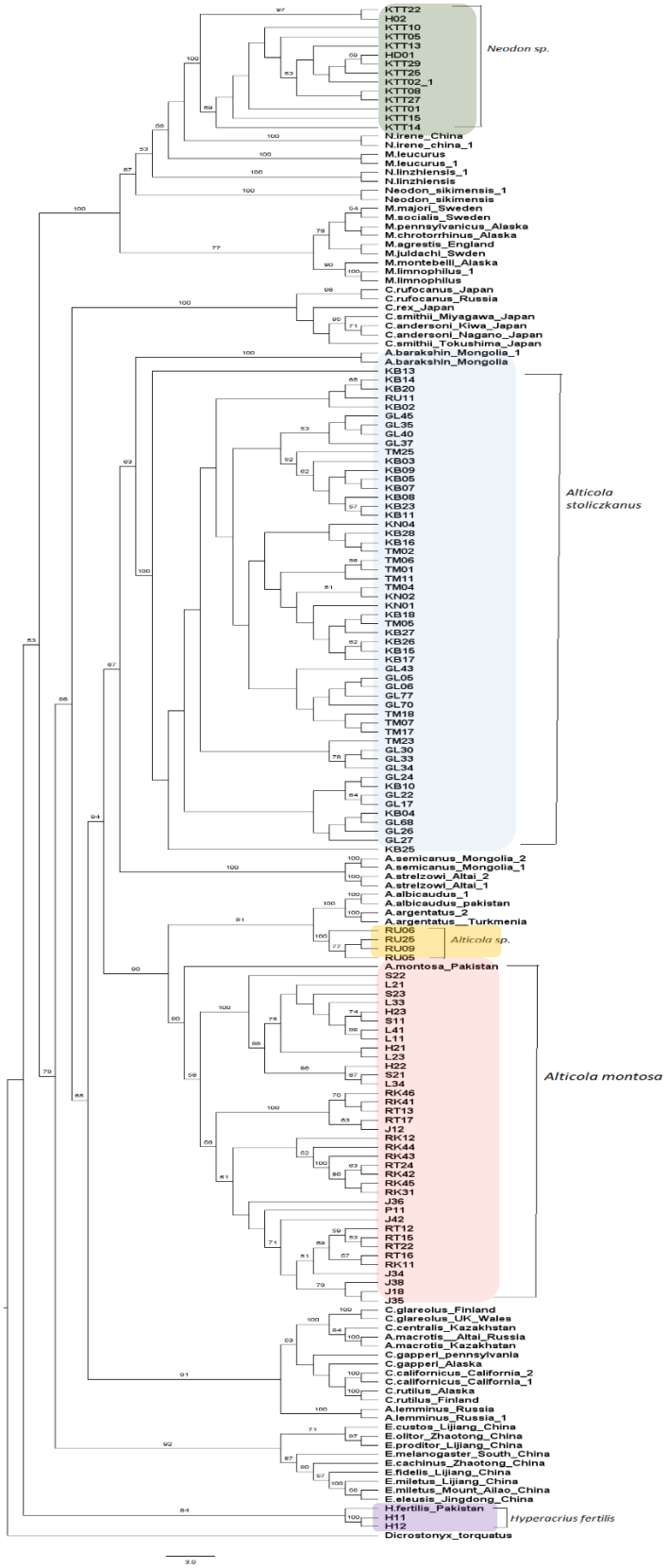


**Figure (6):** Maximum-likelihood tree using RaxML. Numbers above the branches represent bootstrap support based on 100 replicates. Bootstrap values below 50% are not shown on the tree.

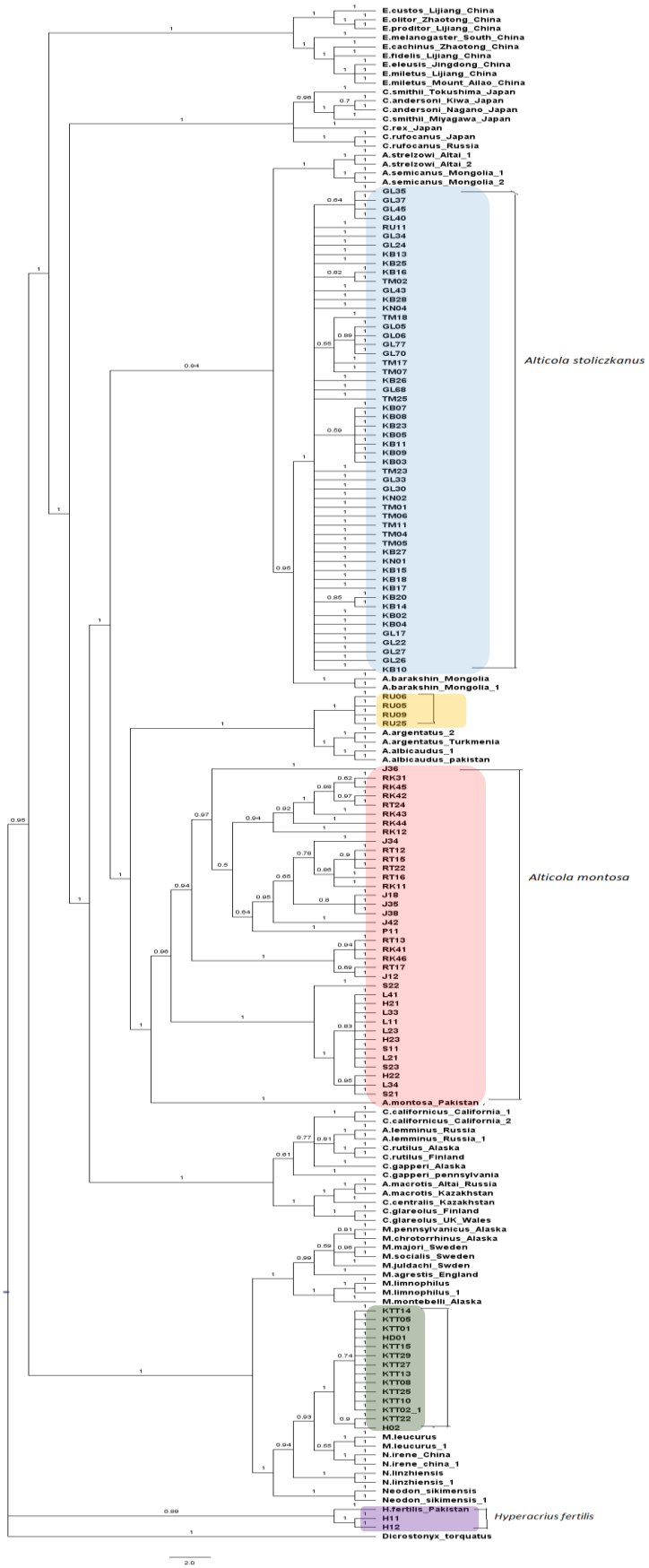
E.custos_Lijiang_China
E.olitor_Zhaotong_China
E.proditor_Lijiang_China
E.melanogaster_South_China
E.cachinus_Zhaotong_China
E.fidelis_Lijiang_China
E.eleusis_Jingdong_China
E.miletus_Lijiang_China
E.miletus_Mount_Allao_China
C.smithii_Tokushima_Japan
C.andersoni_Kiwa_Japan
C.andersoni_Nagano_Japan
C.smithii_Miyagawa_Japan
C.rex_Japan
C.rufocanus_Japan
C.rufocanus_Russia
A.strelzowi_Altai_1
A.strelzowi_Altai_2
A.semicanus_Mongolia_1
A.semicanus_Mongolia_2
GL35
GL37
GL45
GL40
RU11
GL34
GL24
KB13
KB25
KB16
TM02
GL43
KB28
KN04
TM18
GL05
GL06
GL77
GL70
TM17
TM07
KB26
GL68
TM25
KB07
KB08
KB23
KB05
KB11
KB09
KB03
TM23
GL33
GL30
KN02
TM01
TM06
TM11
TM04
TM05
KB27
KN01
KB15
KB18
KB17
KB20
KB14
KB02
KB04
GL17
GL22
GL27
GL26
KB10
Alticola stoliczkanus
A.barakshin_Mongolia
A.barakshin_Mongolia_1
RU06
RU05
RU09
RU25
A.argentatus_2
A.argentatus_Turkmenia
A.albicaudus_1
A.albicaudus_pakistan
J36
RK31
RK45
RK42
RT24
RK43
RK44
RK12
J34
RT12
RT15
RT22
RT16
RK11
J18
J35
J38
J42
P11
RT13
RK41
RK46
RT17
J12
S22
L41
H21
L33
L11
L23
H23
S11
L21
S23
H22
L34
S21
Alticola montosa
A.montosa_Pakistan
C.californicus_California_1
C.californicus_California_2
A.lemminus_Russia
A.lemminus_Russia_1
C.rutilus_Alaska
C.rutilus_Finland
C.gapperi_Alaska
C.gapperi_pennsylvania
A.macrotis_Altai_Russia
A.macrotis_Kazakhstan
C.centralis_Kazakhstan
C.glareolus_Finland
C.glareolus_UK_Wales
M.pennsylvanicus_Alaska
M.chrotorrhinus_Alaska
M.majori_Sweden
M.socialis_Sweden
M.juldachi_Swden
M.agrestis_England
M.limnophilus
M.limnophilus_1
M.montebelli_Alaska
KTT14
KTT05
KTT01
HD01
KTT15
KTT29
KTT27
KTT13
KTT08
KTT25
KTT10
KTT02_1
KTT22
H02
M.leucurus
M.leucurus_1
N.irene_China
N.irene_china_1
N.linzhiensis
N.linzhiensis_1
Neodon_sikimensis
Neodon_sikimensis_1
H.fertilis_Pakistan
H11
H12
Hyperacrius fertilis
Dicrostonyx_torquatus
2.0

**Figure (7):** Tree generated from MrBayes with Bayesian Posterior Probabilities (BPP) shown above the branches. BPPs below 0.7 are not shown in the tree.

The phylogenetic reconstruction for genus *Alticola* (Figure 5 and 6) is based on classification described by Lebedev et al, 2007. Total 114 processed individuals segregated themselves into 5 distinct clades of different species: *Alticola montosa, Alticola barakshin*, *Hyperacrius fertilis* and two unknown separate clades which did not cluster together with any of the identified species clade (yellow and green boxes on the tree respectively). The branch supports for each of these clades were as follows: for *montosa* clade 90% and BPP 1, for *barakshin* clade 63% and 0.95, for *Hyperacrius fertilis* 84% and 0.99 and for unknown clade within *Neodon* and *Alticola* 100% and 1 respectively. Except for *barakshin* clade, all the other clades showed greater than 75% bootstrap support and greater than 0.70 BPP confirming their identification as respective species with whom they clade together.

We calculated p-distance to see to what known species these unknown clades were close to (Table 7). In both cases, the p-distances between the known species and unknown clades were in range of 5% to 11%. Thus based on a criteria given by Bradley and Baker (2001), we identified these clades as *Alticola sp.* and *Neodon sp.* respectively.

| | Known Species | P-distance (avg) |
|---|---|---|
| *Alticola sp.* | *A.montosa* | 7% |
| | *A.argentatus* | 6.5% |
| | *A.albicaudus* | 6.8% |
| *Neodon sp.* | *Neodon leucurus* | 11.2% |
| | *Neodon irene* | 9.4% |
| | *Neodon linzhiensis* | 11.8% |
| | *Neodon sikimensis* | 9.4% |
| | | |

**Table (7)**: Pairwise genetic distance between unknown *Alticola* and *Neodon* clade and known species

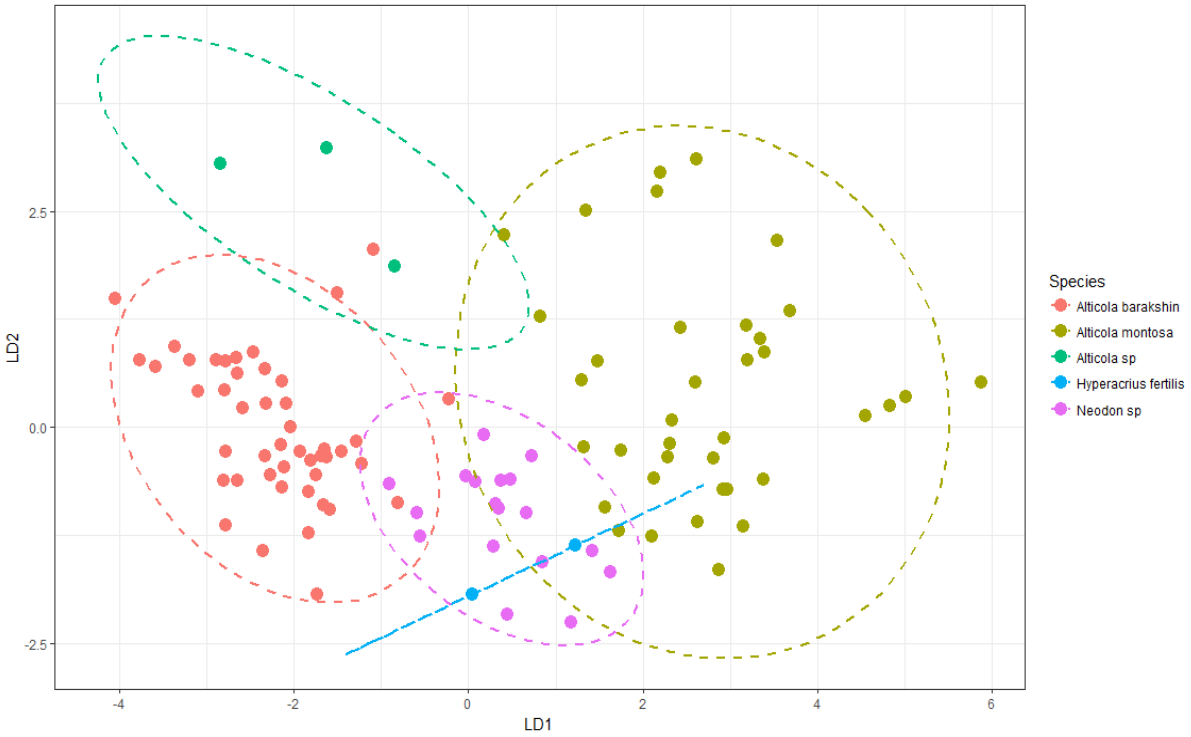


**Figure (8):** Discriminant Function Analysis (DFA) plot for *Alticola*

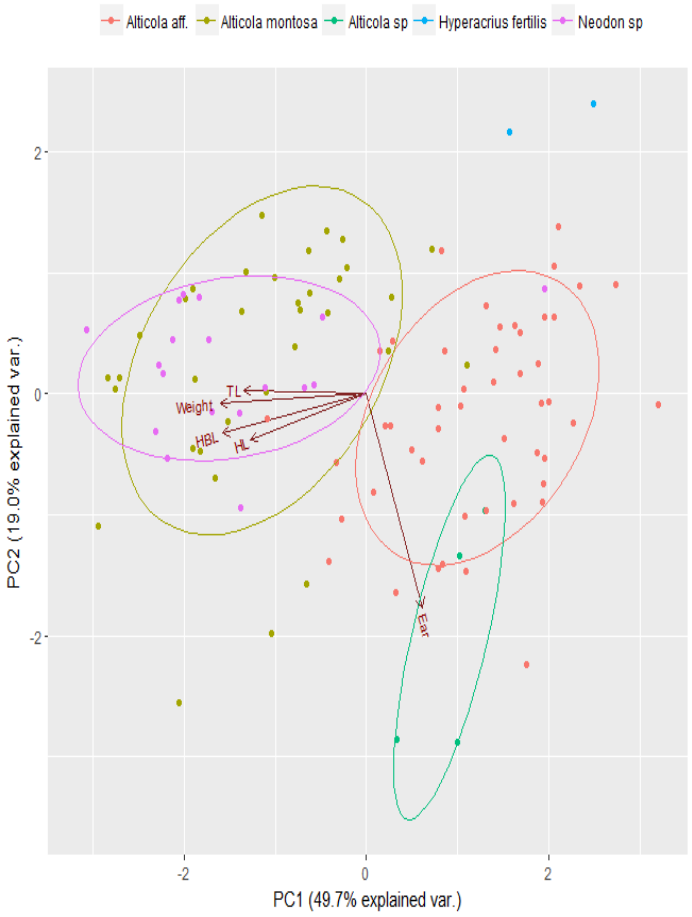


**Figure (9):** Principal Component Analysis (PCA) plot for *Alticola*

In Principal Component analysis, TL (Tail Length) and Length of Ear of an individual were the variables contributing most to the discrimination between groups (Appendix, table A4). Both PCA and DFA plots showed four different clusters and two Hyperacrius individuals which were completely separate from all the other four clusters in PCA plot. (Blue points). This is consistent with our genetic data on *Alticola* which shows 5 different clades. The clusters representing species *Alticola barakshin* and *Alticola montosa* are completely separate with no overlap in both plots. Cluster representing *Alticola sp.* shows little overlap with *barakshin* and *montosa* cluster in DFA plot and no overlap with *montosa* in PCA plot.

## Part 2: Species Distribution modelling

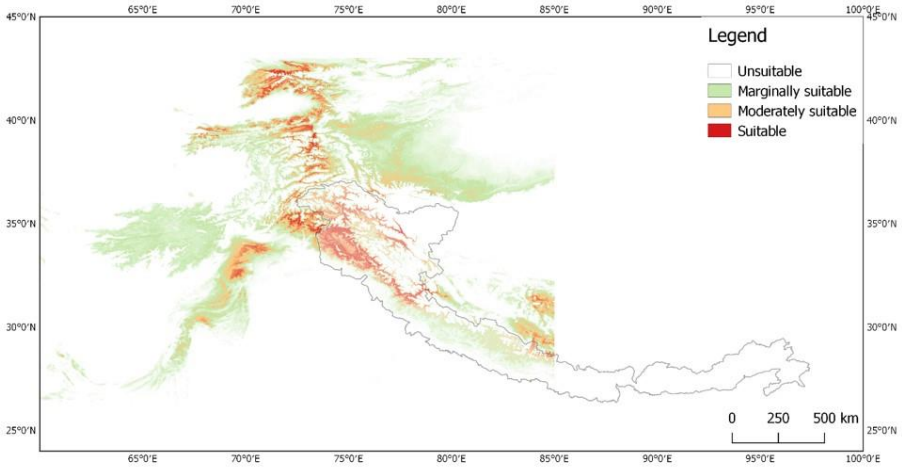


**Figure (10):** Potential distribution of *Apodemus pallipes* in Western Himalayan region. The different colours in the map depict predicted suitable area, moderately suitable, marginally suitable and unsuitable area

From figure 10, *Apodemus pallipes* shows suitable habitat throughout the greater and some parts of trans-Himalayan ranges. Moderately suitable to suitable habitats are also shown in Northern Pakistan, parts of Tajikistan and Kyrgyzstan. This is consistent with known range of *A.pallipes*. Marginally suitable habitats are shown in parts of lesser Himalayas, in China and in parts of Afghanistan.

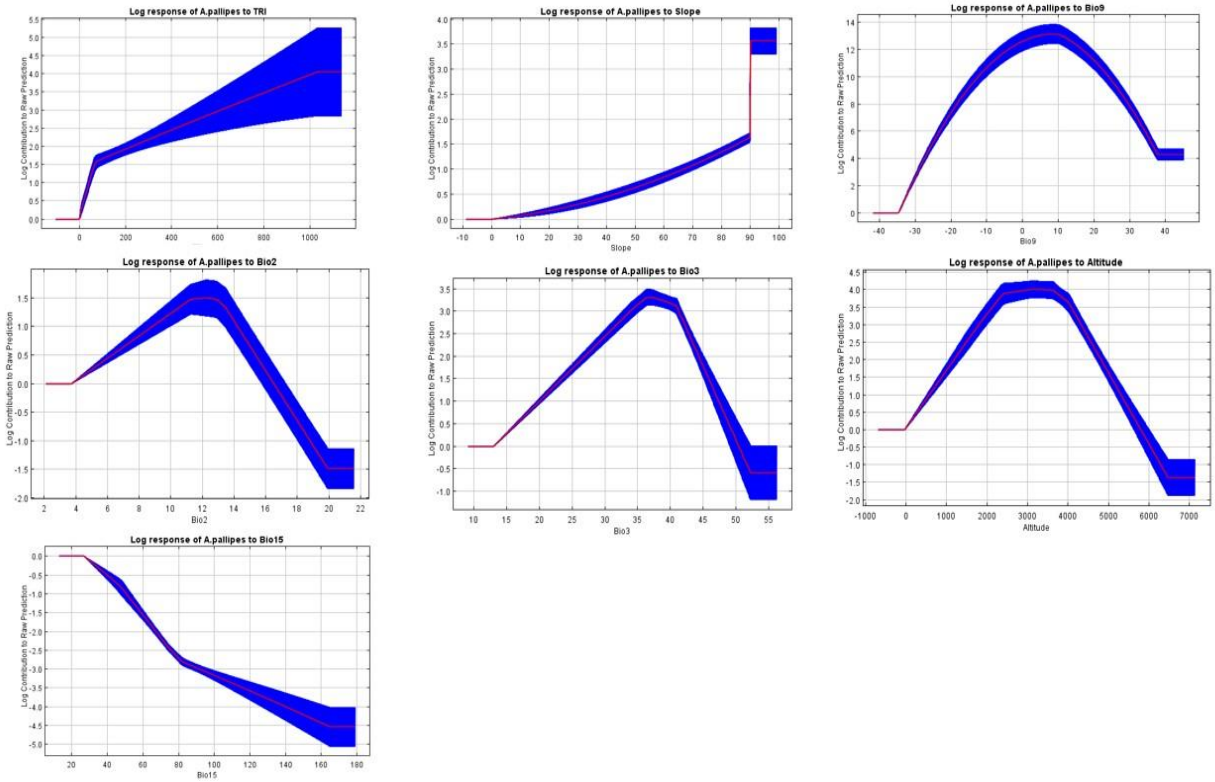


**Figure (11)**: Response curves indicating log contribution of each environmental variable in predicting distribution of *Apodemus pallipes*

Suitability of *A.pallipes* increased with TRI and Slope, increased till certain threshold with Bio8, Bio2, Bio3 and Altitude and decreased with Bio15 (Figure 11).

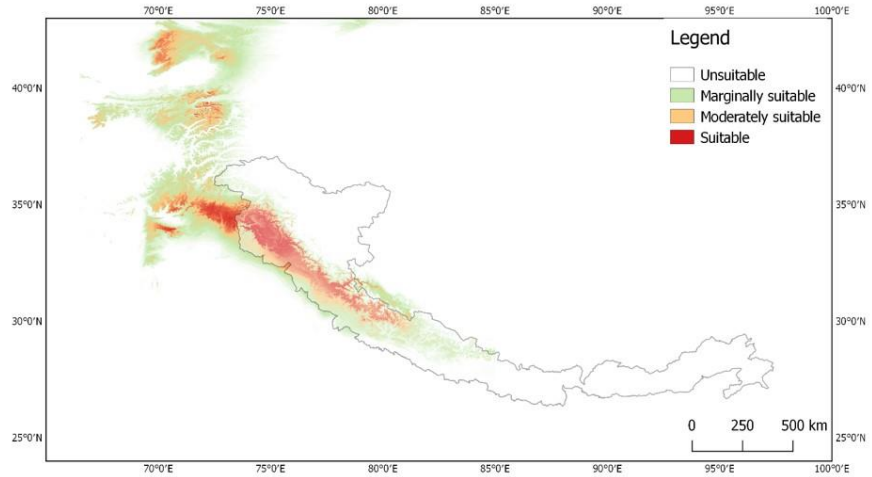


**Figure (12):** Potential distribution of *Apodemus rusiges* in Western Himalayan region. The different colours in the map depict predicted suitable area, moderately suitable, marginally suitable and unsuitable area

*Apodemus rusiges* is known to be endemic to Northeastern Pakistan and Northwestern India (Jammu and Kashmir). Habitat suitability map for *Apodemus rusiges,* shows moderately suitable to suitable habitat throughout the greater Himalayan ranges. Marginally suitable habitats are shown in adjacent ranges of greater himalayas, It is known that the ranges of *Apodemus pallipes* and *Apodemus rusiges* overlap. From comparison between suitability map for *Apodemus pallipes* and *rusiges*, both these species show suitable habitats throughout the greater Himalayan ranges extending from Northwesten parts of Jammu and Kashmir till Himachal pradesh.

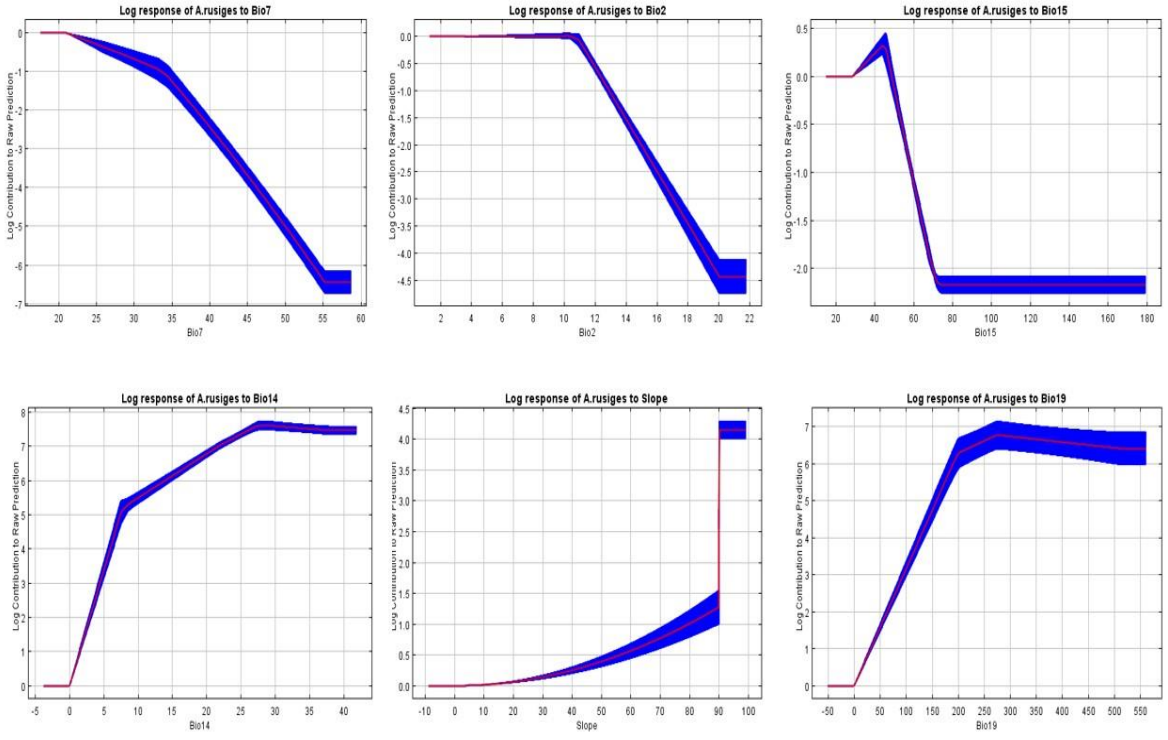


**Figure 13**: Response curves indicating log contribution of each environmental variable in predicting distribution of *Apodemus rusiges*

From response curves, habitat Suitability for *Apodemus rusiges* increased with Bio14, Slope, and Bio19 and decreased with Bio2, Bio7 and Bio15.

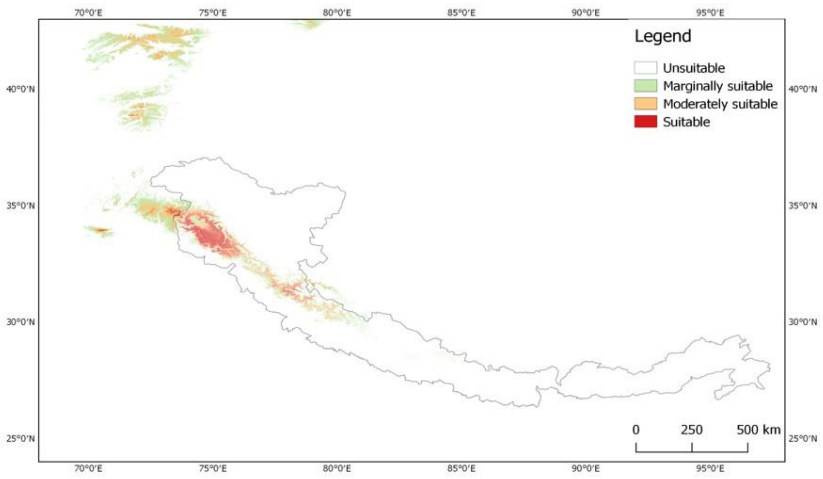


**Figure (14):** Potential distribution of *Alticola montosa* in Western Himalayan region. The different colours in the map depict predicted suitable area, moderately suitable, marginally suitable and unsuitable area

From figure 14, *Alticola montosa* shows the suitable habitats in Northwestern part of Himalayan ranges in Jammu and Kashmir. Marginally suitable and moderately suitable habitats are shown in Northern areas of Pakistan and moderately suitable to suitable habitats are shown in parts of greater himalayan ranges in Himachal Pradesh. This is consistent with our results from genetics part, where we have identified voles from Kashmir and Himachal Pradesh as *Alticola montosa*. No *Alticola montosa* has been identified from Ladakh and Spiti which is consistent with habitat suitability map where *Alticola montosa* doesn't show suitable habitat in trans-himalayan ranges in Eastern part of Kashmir.

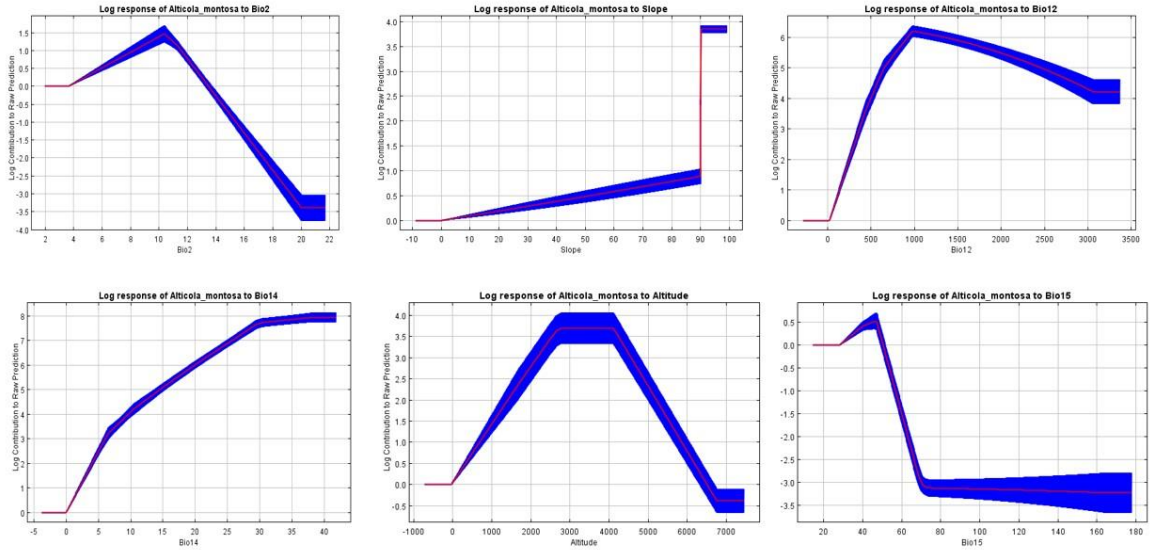


**Figure (15):** Response curves indicating log contribution of each environmental variable in predicting distribution of *Alticola montosa*

From response curves, habitat suitability for *A.montosa* increased with Slope, Bio14, increased till certain threshold for Altitude, Bio2 and Bio12 and decreased for Bio15.

# Discussion:

As discussed in the introduction, the rodents of Western Himalayas remain poorly studied. Last sampling records of rodent species from Western Himalayas date back to early 1900s. Since then, there have been multiple revisions of taxonomic positions of species of *Apodemus* and *Alticola*. For these genera which have co-occurring species within same genus (H. Suzuki et al., 2008; Krystufek et al, 2016) the intraspecific differences between the species are small (Krystufek et al, 2016). Hence, species identification in such cases can pose a challenge and can lead to misidentification in field.

Fairly large number of studies have focused on using genetics to classify the species within *Apodemus* (Liu et al, 2004, Suzuki et al, 2008, Filippucci et al., 2002; Michaux et al., 2002; Serizawa et al., 2000, Hoofer et al., 2007). Same is not the case for *Alticola*, where there have been comparatively less number of studies focusing on

phylogeny of genus *Alticola*. Species identification and delimitation for species within subgenus *Alticola* has been exclusively based on skull and teeth characters and general habitus (Rossolimo and Pavlinov., 1992; Krystufek et al., 2016; Bodrov et al., 2016).

We generated the genetic data using mitochondrial cytochrome b gene to first identify the species considering which species are known to occur in Western Himalayas. We did phylogenetic reconstruction referring to the publications that describe most recent classification of these two genera. We used techniques from multivariate statistics to see what morphometric data that we have collected from field suggests. From our results, 19 individuals which were identified as *Apodemus* on field turned out to be species of *Mus*. Species that were identified as *Microtus* on field formed part of *Alticola montosa* clade (L11, L21, L33, L41, H21, H22, H23). The species that were identified as *Alticola stolikzcanus* taking into account its external features and geographic range appear to be part of *Neodon* genus (All KTTs, H02, HD01). However, they formed a separate clade within genus *Neodon*.

Out of 22 known species of *Apodemus, A.pallipes* and *A.rusiges* are the only species of *Apodemus* recorded from Himalayas of North Pakistan and Northwest India and their ranges broadly overlap (Musser and Carleton, 2005). These two species are known to co-occur in this part of Himalayan ranges. *Apodemus rusiges* is larger in body size than *Apodemus pallipes* and is known to have a longer tail relative to head and body (Carleton, 2005). Since there's no genetic data available for *Apodemus rusiges*, there's no way to find out if some of the individuals are indeed *rusiges* until we successfully sequence the museum specimens.

We identified two species of *Alticola*, *Alticola montosa* and *Alticola barakshin* and one species of genus *Hyperacrius*, *Hyperacrius feritilis*. *Alticola barakshin*, which is commonly known as Gobi Altai Mountain Vole, is known to occur in Russia and Mongolia. *Alticola barakshin* was initially included in *A. stoliczkanus*, but later revisions by Rossolimo and Pavlinov (1992, 1994) separated them as a species on the basis of morphology. Also, the p-distance calculation revealed high (avg) genetic distance of 5.8% between our identified species and known *Alticola barakshin*. Thus based on taxonomic history, p-distance and the fact that Alticola barakshin is not known to occur in India, we suspect that these individuals might be species of *Alticola stoliczkanus*.

Initially, the name *Alticola roylei* was used for what are now identified as six different species (Krystufek et al, 2016) within sub-genus *Alticola.* As per Musser and Carleton (2005), *Alticola roylei* is occurs in Western Himalayas from Kullu valley In Himachal Pradesh to Uttarakhand. In Alticola tree, all the Kashmir individuals within *Alticola montosa* clade clustered together with a bootstrap support of 100%. While all the Himachal individuals clustered together with a poor bootstrap support of 56%. Also, there is only single cluster of *Alticola montosa* revealed by both PCA and DFA plots. Thus based on morphometrics and bootstrap support for phylogenetic tree, we identified all these individuals from both Kashmir and Himachal as *Alticola montosa*, regardless of the possibility that Himachal individuals might be of species *Alticola roylei.*

The genus *Hyperacrius* is hypothesized to be derived from *Alticola* (Kohli et al, 2015). The two species of this genus *Hyperacrius fertilis* and *Hyperacrius wynnei* are found in high altitude forests and alpine meadows of Northern India and Northern Pakistan (Kohli et al, 2015). Even though the identification of H11 and H12 individual as *Hyperacrius fertilis* is supported by bootstrap value of 84% and posterior probability of 0.99, due to lack of sequence data for *Hyperacrius wynnei*, the identification is still uncertain.

Our results show why generating genetic data is important in species identification and delimitation. However, even genetics has limitations in case of identifying species for which no genetic data has ever been generated. If the actual species is not present on phylogenetic tree, then individuals might show affinity to its most closely related species. Hence, integrating different approaches is useful. In our study, we have used genetics and multivariate morphometrics together to identify the species. Combining different perspectives is common for groups such as mammals, where majority of species are already known or in case of groups where morphological features are limited (Dayrat, 2004). Some recent studies have used combination of both DNA-based and morphology-based methods to resolve cryptic species complexes (e.g., Lumley et al, 2010).

We also did Species Distribution Modelling (SDM) using MaxEnt for species of *Apodemus* and *Alticola* to see the possibility having suitable habitats occurring elsewhere than their current known range (For future sampling). We did Maximun

Entropy species distribution modelling, which is known to be performing better than currently known other species distribution modelling methods (Elith et al., 2006). Taking into account the topographic heterogeneity of Himalayas, we incorporated the derived variables such as Terrain Roughness Index (TRI), Eastness into our model (See Materials and Methods section for details). We used 30 arc-seconds (approx. 1 km), the highest resolution available to model the distribution to account for the occurrence points that were very close. The regions that are currently known range were predicted as suitable habitats for all 3 species. Moreover, our results were consistent with our field sampling.

**Limitations and caveats**

For the species for which genetic data has not yet been generated, we collected the museum specimens (crustis). But, usually the DNA that is present in these specimens can be very degraded and contaminated due to preservations over the years. Since we were unable to extract the DNA from museum specimens, we still don't have complete genetic data yet. In case of species where no genetic data has ever been generated, studying diagnostic traits could be one possible way of delimiting the species.

Our Bayesian trees showed the presence of polytomies which can usually mean two things: (1) Incomplete taxon sampling, there's no enough data to find out how lineages are related. (2) Rapid speciation: Sometimes polytomies can also mean multiple speciation events happened at same time. Polytomies could also be a result of rapid diversification or splitting in a very short period of time. If they have evolved recently, the sequences will simply be similar and lack variation.

Since the mitochondrial genes evolve faster and saturate the changes quickly, generating genetic data from multiple loci has become preferred approach. However, since the purpose of this study was to identify the species, we have not included the sequence data from nuclear genes. Sequence data from multiple loci will be useful while inferring the colonisation history of these two genera.

In broader perspective, if one is interested in pursuing the issue of unresolved taxonomy further, the best approach would be to visit the museums which house the collections of specimens of these species.

**Species Distribution Modelling- pros and cons**

The very fundamental limitation of SDM is that it is influenced by number of data points. Thus, the species with limited data cannot be modelled effectively using SDM. The occurrence data might possess sampling biases which can result in spatial autocorrelation influencing the model performance and resulting in over-fitting (A.Aryal et al, 2016). Also, the fact that species realised niche is affected by biotic interactions such as prey-predator interactions, competition, etc. and ability to colonise the available habitat, should be considered while drawing inferences from Species Distribution models. Essentially, the range predicted by SDMs comprises of the range close to the fundamental niche. The realised niche space could be much smaller as the species may not be able to occupy all the localities it could potentially survive and reproduce in.

Even with above mentioned limitations, SDM is still a powerful tool for predicting current distribution of species and identifying where suitable environments are likely to occur under climate change. This is very important from point of view of conservation. Recent human-induced environmental changes are already causing shifts in species' ranges and also extinctions in some cases (Priti et al, 2016). Understanding how species will respond to future climate change is important to develop effective conservation strategies and reducing the risk of future biodiversity losses.

# Future Directions

The future work will include (1) projecting distributions of species under future climate change scenarios. (2) Generating sequence data from nuclear genes to infer evolutionary history of *Apodemus* and *Alticola.* (3) Implementing better ways to successfully generate genetic data from museum specimens.

# References

1) Abdo, Z., and B. Golding. (2007). A step toward barcoding life: A model-based, decision- theoretic method to assign genes to pre-existing species groups. Syst. Biol. 56:44–56.

2) Agarwal, P., and D. J. States. (1998). Comparative accuracy of methods for protein sequence similarity search. Bioinformatics 14:40–47.

3) Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R.P. (2015). spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, *38*(5), 541–545. https://doi.org/10.1111/ecog.01132

4) Alexandre H. Hirzel and Gwenaelle Le Lay. (2008). Habitat suitability modelling and niche theory. *Journal of Applied Ecology* 45, 1372-1381

5) Alves PC, Harris DJ, Melo-Ferreira J et al. (2006) Hares on thin ice: introgression of mitochondrial DNA in hares and its implications for recent phylogenetic analyses. *Molecular Phylogenetics and Evolution*. 40, 640–641.

6) Aryal, A., Shrestha, U. B., Ji, W., Ale, S. B., Shrestha, S., Ingty, T.,Raubenheimer, D. (2016). Predicting the distributions of predator (snow leopard) and prey (blue sheep) under climate change in the Himalaya. *Ecology and Evolution*, *6*(12), 4065–4075. https://doi.org/10.1002/ece3.2196

7) Barbosa, S., Pauperio, J., & Searle, J. B. (2013). Genetic identification of Iberian rodent species using both mitochondrial and nuclear loci : application to noninvasive sampling, 43–56. https://doi.org/10.1111/1755-0998.12024

8) Benoit Dayrat. (2005). Towards integrative taxonomy. *Biological Journal of Linnean Society.* 5, 407-415.

9) Bernhard Hausdorf (2011). Progress toward a general species concept *Evolution*. doi:10.1111/j.1558-5646.2011.01231.x

10) Bin Zhang, Kai He, Tao Wan, Peng Chen, Guozheng Sun, Shaoying Liu, Truong Son Nguyen, Liangkong Lin and Xuelong Jiang. (2016). Multi-locus phylogeny using topotype specimens shed light on the systematics of Niviventer (Rodentia, Muridae) in China. *BMC Evol. Biol*. 16. 261.

11) Bodrov, S. Y., Kostygov, A. Y., Rudneva, L. V., & Abramson, N. I. (2016). Revision of the taxonomic position of the Olkhon mountain vole (Rodentia, Cricetidae). *Biology Bulletin*, *43*(2), 136–145. https://doi.org/10.1134/S1062359016020035

12) Cain AJ. (1954)**.** *Animal Species and their evolution*. New York: Harper & Row.

13) Carleton, M. D. (1984). Introduction to rodents. Pp. 255-265 in Anderson, S. and J. K. Jones, Jr. (eds). Orders and Families of Recent Mammals of the World. John Wiley and Sons, N.Y. xii+686 pp.

14) Carleton, M. D., Stanley, W. T., & Graves, G. R. (2005). Review of the Hylomyscus denniae complex (Rodentia: Muridae) in Tanzania, with a description of a new species. *Proceedings of the Biological Society of Washington*, *118*(3), 619–646. https://doi.org/10.2988/0006-324X(2005)118[619:ROTHDC]2.0.CO;2

15) Darvish, J., Mohammadi, Z., Ghorbani, F., Mahmoudi, A., & Dubey, S. (2015). Phylogenetic Relationships of Apodemus Kaup, 1829 (Rodentia: Muridae) Species in the Eastern Mediterranean Inferred from Mitochondrial DNA, with Emphasis on Iranian Species. *Journal of Mammalian Evolution*, *22*(4), 583–595. https://doi.org/10.1007/s10914-015-9294-9

16) Filipe Pereira, Joao Carneiro and Antonio Amorim. Identification of species with DNA based technology: Current Progress and Challenges. (2008). *Recent Patents on DNA and Gene Sequences, 2*, 187-200

17) Filippucci M.G., Macholan M., Michaux J.R. (2002). Genetic variation and evolution in the genus Apodemus (Muridae, Rodentia). *Bio. J. Linn. Soc*. 75, 395-419.

18) Galewski, T., Tilak, M. K., Sanchez, S., Chevret, P., Paradis, E., & Douzery, E. J. P. (2006). The evolutionary radiation of Arvicolinae rodents (voles and lemmings): Relative contribution of nuclear and mitochondrial DNA phylogenies. *BMC Evolutionary Biology*, *6*(i), 1–17. https://doi.org/10.1186/1471-2148-6-80

19) Gilberto Parolo, Graziano Rossi, and Alessandro Ferrarini (2008). Toward improved species niche modelling: *Arnica Montana* in the Alps as a case study. *Journal of Applied Ecology*, 45, 1410–1418

20) Hartenberger, J.-L. (1985). The order Rodentia: Major questions on their evolutionary origin, relationships, and suprafamilial systematics. *In* “Evolutionary Relationships among Rodents: A Multidisciplinary Analysis” (W. P. Luckett and J.-L. Hartenberger, Eds.), pp. 1–33. Plenum, New York.

21) Hausdorf, B. (2011). Progress toward a general species concept. *Evolution*, *65*(4), 923–931. https://doi.org/10.1111/j.1558-5646.2011.01231.x

22) Hebert, Paul, and T. Ryan Gregory. (2005). 'The Promise Of DNA Barcoding For Taxonomy'. *Systematic Biol.* 54.5: 852-859.

23) Hebert, P. D. N., A. Cywinska, S. L. Ball, and J. R. deWaard. (2003). Biological identifications through DNAbarcodes. Proc. R. Soc. Lond. B 270:313–321.

24) Heckman KL, Mariani CL, Rasoloarison R, Yoder AD. (2007). Multiple nuclear loci reveal patterns of incomplete lineage sorting and complex species history within western mouse lemurs (Microcebus). *Molecular Phylogenetics and Evolution*. 43, 353–367.

25) Hela Sakka, Jean Pierre Quere, Irina Kartavtseva, Marina Pavlenko, Galina Chelomina, Dmitry Atopkin, Aleksey Bogdanov and Johan Michaux. (2010). Comparative phylogeography of four *Apodemus* species (Mammalia:Rodentia) in the Asian Far East: Evidence of Quaternary climatic changes in their genetic structure. *Bio. Journal of Linnean Society*. 797-82

26) Hillis, D M. (1987). "Molecular Versus Morphological Approaches to Systematics". *Annu. Rev. Ecol. Syst.* 18.1: 23-42.

27) Hirzel, A. H., & Le Lay, G. (2008). Habitat suitability modelling and niche theory. *Journal of Applied Ecology*, *45*(5), 1372–1381. https://doi.org/10.1111/j.1365-2664.2008.01524.x

28) HITOSHI SUZUKI, JUN J. SATO, KIMIYUKI TSUCHIYA, JING LUO, YA-PING ZHANG, YING-XIANG WANG, Xue-Long Jiang. (2003). Molecular phylogeny of wood mice (*Apodemus*, Muridae) in East Asia, *Biological Journal of the Linnean Society*, Volume 80, Issue 3, Pages 469–481, https://doi.org/10.1046/j.1095-8312.2003.00253.x

29) Hitoshi Suzuki, Maria Grazia Filippucci, Galina N. Chelomina, Jun J. Sato, Keiko Serizawa, Eviatar Nevo. (2008). A biogeographic view of *Apodemus* in Asia and Europe inferred from Nuclear and Mitochondrial Gene Sequences. *Biochem Genet*. 46:329-346

30) Hoofer, S. R., Gaschak, S., Dunina-Barkovskaya, Y., Makluk, J., Meeks, H. N., Wickliffe, J. K., & Baker, R. J. (2007). New Information for Systematics, Taxonomy,

and Phylogeography of the Rodent Genus Apodemus (Sylvaemus) in Ukraine. *Journal of Mammalogy*, *88*(2), 330–342. https://doi.org/10.1644/06-MAMM-A-218R1.1

31) Howard A Ross, Sumathi Murugan, Wai Lok Sibon Li. (2008). Testing the reliability of genetic methods of species identification via Simulation. *Syst. Biol.* 57(2):216–230

32) Hutchinson, G. E. (1957). Concluding remarks. *Cold Spring Harbor Symp. Quant. Biol*. 22:415-27

33) Jack W Sites Jr and Jonathon C Marshall. (2003). Delimiting species: a Renaissance issue in systematic biology. *TRENDS in Ecology and Evoluton.* Vol.18. No.9

34) Jansa, S. A., & Weksler, M. (2004). Phylogeny of muroid rodents: Relationships within and among major lineages as determined by IRBP gene sequences. *Molecular Phylogenetics and Evolution*, *31*(1), 256–276. https://doi.org/10.1016/j.ympev.2003.07.002

35) Karan, P.P. (1996). Geographic regions of the Himalaya.

36) Kevin De Queiroz. (2007). Species Concepts and Species Delimitation. *Syst. Biol*. 56(6): 879-886.

37) Kimura M. (1968). Evolutionary rate at the molecular level. Nature; 217: 624-626

38) Kohli, B. A., Speer, K. A., Kilpatrick, C. W., Batsaikhan, N., Damdinbaza, D., & Cook, J. A. (2014). Multilocus systematics and non-punctuated evolution of Holarctic Myodini (Rodentia: Arvicolinae). *Molecular Phylogenetics and Evolution*, *76*(1), 18–29. https://doi.org/10.1016/j.ympev.2014.02.019

39) Kryštufek, B., Shenbrot, G. I., Gregori, J., Benda, P., & Hutterer, R. (2017). Taxonomic and geographic setting of Royle's mountain vole Alticola roylei revisited. *Mammalia*, *81*(5), 503–511. https://doi.org/10.1515/mammalia-2016-0102

40) Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., Calcott, B. (2016) PartitionFinder 2: new methods for selecting partitioned models of evolution

formolecular and morphological phylogenetic analyses. *Molecular biology and evolution*. DOI: dx.doi.org/10.1093/molbev/msw260

41) Lebedev, V. S., Bannikova, A. A., Tesakov, A. S., & Abramson, N. I. (2007). Molecular phylogeny of the genus Alticola (Cricetidae, Rodentia) as inferred from the sequence of the cytochrome b gene. *Zoologica Scripta*, *36*(6), 547–563. https://doi.org/10.1111/j.1463-6409.2007.00300.x

42) Liu, X., Wei, F., Li, M., Jiang, X., Feng, Z., & Hu, J. (2004). Molecular phylogeny and taxonomy of wood mice (genus Apodemus Kaup, 1829) based on complete mtDNA cytochrome b sequences, with emphasis on Chinese species. *Molecular Phylogenetics and Evolution*, *33*(1), 1–15. https://doi.org/10.1016/j.ympev.2004.05.011

43) LUMLEY, L. M., & SPERLING, F. A. H. (2010). Integrating morphology and mitochondrial DNA for species delimitation within the spruce budworm (Choristoneura fumiferana) cryptic species complex (Lepidoptera: Tortricidae). *Systematic Entomology*, *35*(3), 416–428. https://doi.org/10.1111/j.1365-3113.2009.00514.x

44) LUO A-Rong, ZHANG Yan-Zhou, QIAO Hui-Jie, SHI Wei-Feng, Robert W. MURPHY, ZHU Chao-Dong. (2010). Outgroup selection in tree reconstruction: a case study of the family Halictidae (Hymenoptera: Apoidea) *Acta Entomologica Sinica*. 53(2): 192-201

45) Matz, M. V., & Nielsen, R. (2005). A likelihood ratio test for species membership based on DNA sequence data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1462), 1969–1974. https://doi.org/10.1098/rstb.2005.1728

46) Merow et al (2013). A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography 36: 1058–1069

47) Michaux J.R., Chevret P., Filippucci M. G., Macholan. (2002). Molecular Phylogeny of genus *Apodemus* with a special emphasis to the subgenus *Sylvaemus* using nuclear IRBP gene and two mitochondrial markers: cytochrome *b* and 12s rRNA. *Mol. Phylogenet. Evol*. 23, 123-

48) Michaux, J. R., & Filippucci, M. G. (2002). Genetic variation and evolution in the genus *Apodemus*. *Biological Journal of the Linnean Society*, *75*, 395–419. https://doi.org/10.1046/j.1095-8312.2002.00032.x

49) Michaux, J., & Reyes, A. (2017). Evolutionary History of the Most Speciose Mammals : Molecular Phylogeny of Muroid Rodents, *18*(11), 2017–2031.

50) Musser, G.G., Carleton, M.D. (1993). Family Muridae. In: Wilson, D.E., Reeder, D.M. (Eds.), Mammal Species of the World, seconded. Smithsonian Institution Press, Washington, DC, pp. 501–753.

51) Musser, G.G. and Carleton, M.D. (2005). Superfamily Muroidea. In: Wilson, D.E. and Reeder, D.M., Eds., Mammal Species of the World: A Taxonomic and Geographic Reference, The Johns Hopkins University Press, Baltimore, 2142.

52) Nielsen, R., and M. V. Matz. 2006. Statistical approaches for DNA barcoding. Syst. Biol. 55:162-169.

53) Nowak, R.M., (1999). Walker's Mammals of the World, sixth ed. The John Hopkins University Press, Baltimore and London.

54) Parolo, G., Rossi, G., & Ferrarini, A. (2008). Toward improved species niche modelling : Arnica montana in the Alps as a case study, 1410–1418. https://doi.org/10.1111/j.1365-2664.2007.0

55) Paul O. Lewis, Mark T. Holder and Kent E. Holsinger. (2005). Polytomies and Bayesian Phylogenetic inference. *Syst. Biol*. 54(2):241-253.

56) Peterson AT. (2006). Use and requirements of Ecological Niche models and related Distributional models. *Biodiversity informatics*, 3, pp. 59-72.

57) Pfenninger, M., & Schwenk, K. (2007). Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC Evolutionary Biology*, *7*, 1–6. https://doi.org/10.1186/1471-2148-7-121

58) Phillipe Lemey, Mark Salemi and Anne-Mieke Vandamme. (2009). The Phylogenetic Handbook: A practical approach to DNA and Protein Phylogeny. Second Edition. Cambridge University Press.

59) Phillips SJ, Anderson RP, Schapire RE. (2006). Maximum entropy modelling of species geographic distributions, *Ecological Modelling*; 190: 231–259.

60) Priti, H., Aravind, N.A., Uma Shaanker, R., Ravikanth, G. (2016). Modelling impacts of future climate on the distribution of Myristicaceae species in the Western Ghats, India. *Ecol. Engg*. 89, 14e23.

61) Rossolimo, O.L. and I.Ya. Pavlinov. (1992). Species and subspecies of *Alticola* s. str. (Rodentia: Arvicolinae). In: (I. Horáček and V. Vohralík, eds.) Prague Studies in Mammalogy. Collection of papers dedicated to Prof. Dr. Vladimir Hanák on occasion of his 60th birthday. Charles University Press, Praha. pp. 149–176

62) S. Barbosa, J. Pauperio, J.B. Searle, P.C. Alves. (2013). Genetic identification of Iberian rodent species using both mitochondrial and nuclear loci: application to non-invasive sampling. *Molecular Ecology Resources.* 13. 43-56

63) Serizawa K., Suzuki H., Iwasa M.A., Tsuchiya K., Pavlenko M.V., Kartavtseva I.V., Chelomina G. N., Dokuchaev N.E., Han S.H. (2002). A spatial aspect on mitochondrial DNA genealogy in *Apodemus Peninsulae* from East Asia. *Biochem. Genet*. 40, 149-161.

64) Serizawa, K., Suzuki, H., & Tsuchiya, K. (2000). A phylogenetic view on species radiation in Apodemus inferred from variation of nuclear and mitochondrial genes. *Biochemical Genetics*, *38*(1–2), 27–40. https://doi.org/10.1023/a:1001828203201

65) Shrestha, U. B., S. Gautam, and K. S. Bawa. 2012. Widespread climate change in the Himalayas and associated changes in local ecosystems. PLoS One 7:e36741

66) Sites, J. W., & Marshall, J. C. (2003). Delimiting species: A Renaissance issue in systematic biology. *Trends in Ecology and Evolution*, *18*(9), 462–470. https://doi.org/10.1016/S0169-5347(03)00184-8

67) Sorte, F. A. La, & Jetz, W. (2010). Projected range contractions of montane biodiversity under global warming. https://doi.org/10.1098/rspb.2010.0612

68) Soumyajit Banerjee, Gautam Aditya , Goutam K. Saha (2015). Household Wastes as Larval Habitats of Dengue vectors: Comparison between Urban and Rural areas of Calcutta, India. PLoS ONE10 (10):e0138082.doi:10.1371/journal.pone.0138082

69) Steppan, S. J., Adkins, R. M., & Anderson, J. (2004). Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes.

*Systematic Biology*, *53*(4), 533–553. https://doi.org/10.1080/10635150490468701

70) Steven R Hoofer, Sergey Gaschak, Yelena Dunina-Barkovskaya, Julia Makluk, Heather N. Meeks, Jeffrey K. Wickliffe, and Robert J. Baker. (2007). New information for systematics, taxonomy, and phylogeography of the rodent genus *Apodemus (Sylvaemus)* in Ukraine. *Journal of mammalogy*. 88(2):330-342.

71) Suzuki, H., Filippucci, M. G., Chelomina, G. N., Sato, J. J., Serizawa, K., & Nevo, E. (2008). A biogeographic view of Apodemus in Asia and Europe inferred from nuclear and mitochondrial gene sequences. *Biochemical Genetics*, *46*(5–6), 329–346. https://doi.org/10.1007/s10528-008-9149-7

72) Suzuki, Y., Tomozawa, M., Koizumi, Y., Tsuchiya, K., & Suzuki, H. (2015). Estimating the molecular evolutionary rates of mitochondrial genes referring to Quaternary ice age events with inferred population expansions and dispersals in Japanese Apodemus. *BMC Evolutionary Biology*, *15*(1), 2–4. https://doi.org/10.1186/s12862-015-0463-5

73) T.D. Kocher, W.K. Thomas, A. Meyer, S.V. Edwards, S. Paabo, F.X. Villablanca, and A. C. Wilson. (1989). Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. *Proc. Nati. Acad. Sci. USA*. Vol. 86, pp. 6196-6200, Evolution

74) Vladimir S. Lebedev, Anna A. Bannikova, Alexey S. Tesakov and Natalia I. Abramson. (2007). Molecular phylogeny of the genus *Alticola* (Cricetidae, Rodentia) as inferred from the sequence of the cytochrome *b* gene. *Zoologica Scripta*. 36, pp547-563

75) Wiens, J. A., Stralberg, D., Jongsomjit, D., Howell, C. A., & Snyder, M. A. (2009). Niches , models , and climate change : Assessing the assumptions and uncertainties, *106*.

76) Wilson, D. E., and Reeder, D. M. (1993). “Mammal Species of the World: A Taxonomic and Geographic Reference,” Smithsonian Institute, Washington, D.C.

77) Wilson, D. E., and D. M. Reeder. (2005). Mammal Species of the World, A Taxonomic and Geographic Reference. 3rd edition. Smithsonian Institution Press,

Washington.

78) Xiaoming Liu, Fuwen Wei, Ming Li, Xuelong Jiang, Zuojian Feng, Jinchu Hu. (2004). Molecular phylogeny and taxonomy of wood mice based on complete mtDNA cytochrome b sequences, with emphasis on Chinese species. *Mol. Phylogenetics and Evolution* 33. 1

# Appendix

| **Individual code** | **Collection Locality** | **Elevational transcect (in metres)** |
|---|---|---|
| B21, B33, B41, B43, B51 | Badin, Kashmir | 1500-2000 |
| P12, P21, P22, P23, P25, P32, P34, P51, P53 | Pehlipura, Kashmir | 2000-2500 |
| M11, M21, M22, M23, M31, M41, M53, M54 | Mamnate, Kashmir | 2500-3000 |
| L22, L24, L25, L35, L51, L52 | Ledhwas, Kashmir | 3000-3500 |
| HU01, HU02, HU03, HU04, HU05, HU06, HU07, HU08, HU09, HU14, HU15, HU16, HU17, HU18, HU19, HU21, HU22, HU23, HU24, HU26, HU27, HU28, HU29, HU32, HU33. HU34, HU35, HU36, HU39 | Hunder, Ladakh | 3200 |
| RU01, RU02, RU03, RU26, RU29, RU30, RU32, RU43 | Rumbak, Ladakh | 4000-4500 |
| Bn12, Bn13, Bn14, Bn15, Bn23, Bn24, Bn43 | Manhar, GHNP, Himachal Pradesh | 1500-2000 |
| D22 | Durna, GHNP, Himachal Pradesh | 2500-3000 |
| ST11 | Shilt top, GHNP, Himachal Pradesh | 3000-3500 |
| P12H, P13H, P14H, P15H, P16H, P17H, P18H, P19H, P110 | Sainj Valley, GHNP, Himachal Pradesh | 3000-3500 |
| J14, J15, J16, J17, J27, J28, J36, J39, J44, J49, J58, J59 | Jorah Thatch, GHNP, Himachal Pradesh | 3500-4000 |

| | | |
|---|---|---|
| KS1, KS2 | Kibber, Spiti valley | |

Table (A1): Individuals that were identified as Apodemus on field and their collection locality

| **Individual code** | **Collection Locality** | **Elevational transect in metres** |
|---|---|---|
| L34, L33, L41, L11, L23, L21 | Ledhwas, Kashmir | 3000-3500 |
| H11, H12, H21, H22, H23 | Hoksar, Kashmir | 3500-4000 |
| S21, S22, S11, S23 | Serebal (Latan lake) | 3900 and above |
| RU05, RU06, RU09, RU11, RU25 | Rumbak | 4000-4500 |
| GL37, GL35, GL34, GL33, GL24, GL43, GL45, GL68, GL40, GL30, GL26, GL17, GL22, GL27, GL70, GL05, GL06, GL77 | Gandala | 4500-5000 |
| KTT01, KTT02, KTT13, KTT22, KTT05, KTT14, KTT29, KTT15, KTT13, KTT08, KTT27, KTT25, KTT10, | Kalaktatar plateau | 4900 |
| H02 | Hanle | 4800 |
| HD01 | Hunder Dock | 4000 |
| P11 | Sainj valley | 3000-3500 |
| J34, J42, J18, J35, J38, J12, J36 | Jorah Thatch | 3500-4000 |
| RK42, RK11, RK12, RK41, RK44, RK43, RK45, RK31, RK46 | Raktisar | 4000-4500 |
| RT24, RT12, RT15, RT16, RT22, RT17, RT13 | Rakti top | 4500-5000 |
| TM26, TM17, TM07, TM03, TM04, TM11, TM06, TM01, TM23, TM02, TM18, TM05, TM25 | Timoreso, Spiti valley | 4500-5000 |

| | | |
|---|---|---|
| KB16, KB02, KB09, KB14, KB20, KB17, KB18, KB15, KB26, KB27, KB04, KB07, KB10, KB11, KB25, KB13, KB05, KB08, KB23, KB03, KB28 | Kibber, Spiti valley | |
| KN01, KN02, KN04 | Kanamo, Spiti valley | |

Table (A2): Individuals that were identified as Alticola/Microtus on field and their collection locality

| | LD1 | LD2 | LD3 |
|---|---|---|---|
| TL | -0.2083636 | -0.01775903 | 0.01293476 |
| HBL | -0.2379655 | -0.37329734 | -1.10702276 |
| HL | **-3.4680855** | 3.97852765 | 1.99581976 |
| Ear | **-2.6051911** | -4.19389483 | 2.57030994 |

Table (A3): Coef. Of eigenvectors/Eigenvalue matrix showing contribution of variables in discrimination between groups for *Apodemus*

| | LD1 | LD2 | LD3 | LD4 |
|---|---|---|---|---|
| TL | **1.70271680** | 0.6258676 | -0.4325994 | -0.05078039 |
| HBL | 0.07646516 | -0.5473054 | 0.6031303 | -0.83293634 |
| HL | -0.04459627 | 0.9249855 | 4.3218874 | 3.89236728 |
| Ear | **-2.78090879** | 3.5658448 | 0.9472031 | -1.27011493 |

Table (A4): Coef. Of eigenvectors/Eigenvalue matrix showing contribution of variables in discrimination between groups for *Alticola*

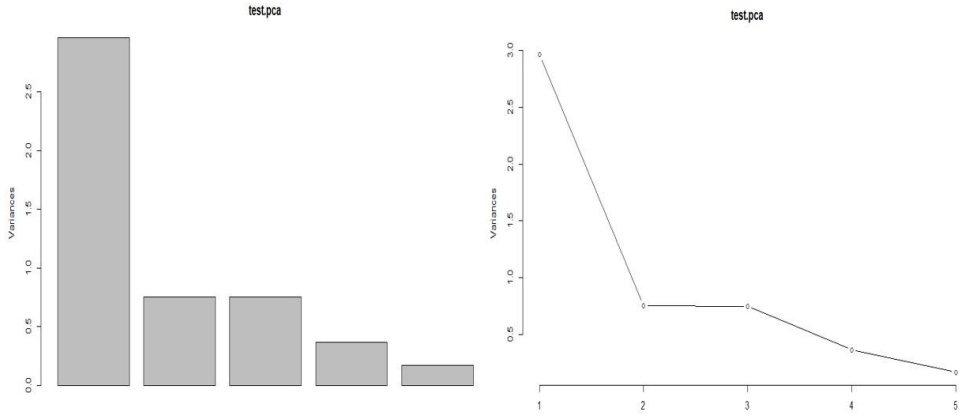


Figure (B1): Screeplot for *Apodemus* dataset showing fraction of total variance in the data explained by each Principal Component

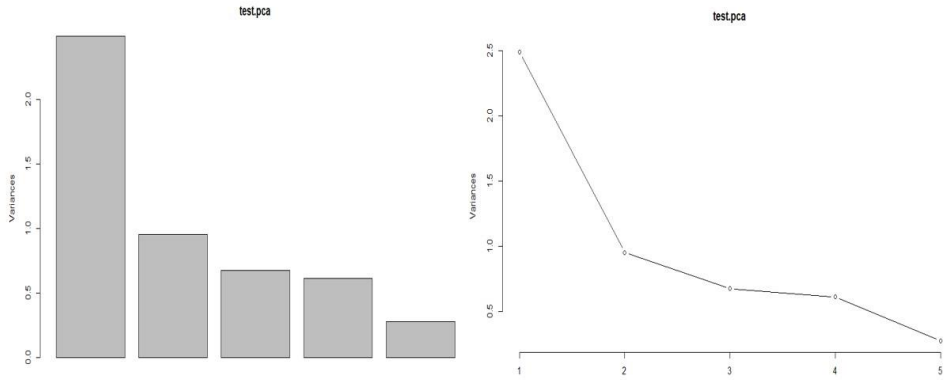


Figure (B2): Screeplot for *Alticola* dataset showing fraction of total variance in the data explained by each Principal Component

Sherman (On right) and Tomahawk (On left) traps used for rodent trapping