Altitudinal Variation in the Communities of Small Mammals in the Central Himalaya of Nepal

Simon Maurice Charles Poulton

A thesis submitted for the degree of
Doctor of Philosophy

University of East Anglia,
School of Biological Sciences,
Norwich, UK

January 2019

Dedication

This thesis is dedicated to the people of Nepal; the villagers and city dwellers, the hoteliers, restauranteurs, cooks and cleaners, the bus drivers, taxi drivers and jeep drivers, the porters and Sherpas, who, together, strive through unbelievable obstacles such as earthquakes, landslides, floods and avalanches to build their extraordinary country.

I also dedicate this thesis to my father, for the vivid tales of his life in India and the far east, which set me on my road, and to my mother who kept us both in check.

ABSTRACT

Altitudinal variation in small mammal communities has been studied in many parts of the world, such as the Rocky Mountains in USA, the Andes of South America and a number of gradients in Tanzania, China, Malaysia and the Philippines. However, no specific elevational studies have been carried out in the Himalaya of Nepal – the region with the largest altitudinal gradients in the world. This six year study sets out to redress this shortfall, and its geographical and temporal scope make it the largest ecological small mammal survey ever undertaken in the Nepali Himalaya. Fieldwork was undertaken over three years in four regions on 147 live-trapping grids at altitudes between 1300m and 4400m, comprising 8046 trap events. 792 animals were caught belonging to 18 putative taxa, with a dominance of shrews not recorded in previous studies. Habitat data was used to characterise the grids, with a combination of cluster analysis and a novel genetic algorithm. The abundance of various small mammal species showed strong associations with different grid characteristics.

To aid species identification, two additional data sources were explored. Firstly, over 8000 historical records of small mammals were compiled from museum records and used to generate predictive models of distributions in Nepal. Secondly, 720 tissue samples were collected in the field, from which a subsample yielded 94 successful sequences of the cytochrome-b gene. Detailed phylogenetic analyses of 200 sequences from GenBank combined with the field-derived sequences, consolidated their identifications into 17 taxa and suggested previously unrecorded cryptic species of voles and shrews.

The main ecological findings of this study were that a) overall species abundance showed a significant decline with altitude and a significant variation between regions, b) there was compelling evidence for competitive exclusion between two species of mouse, c) there was strong evidence of altitudinal zonation for many species and d) both α and β–diversity showed strong elevational and regional variation.

ACKNOWLEDGEMENTS

This study would have been impossible to complete without the help and encouragement of a huge number of people. My primary supervisor, Dr Diana Bell, gave unstinting wisdom, knowledge, encouragement and friendship throughout the whole six-year period of my PhD. I thank my supervisory team of Prof Carlos Peres and Prof Cock van Oosterhout, who gave crucial support and stimulating advice at all times. I am also grateful to other members of staff at UEA for their sound advice; Dr Kevin Tyler (genetics), Dr Iain Barr (PCR), Prof Andrew Lovett (GIS), Dr Jeff Price (Maxent). Prof Justin O'Grady and Michael Strinden at the Bob Champion Institute gave invaluable assistance with PCR and sequencing and colleagues at the Earlham Institute; Dr Amanda Chong (PCR), Dr David Thybert (phylogenetic analysis) and Dr David Wright (genetics) gave helpful advice. I thank David Harrison, Dr Malcolm Pearch and Dr Paul Bates at the Harrison Institute, Kent, for access to their specimen collection and stimulating discussion. My thanks also go to Dr Mike Gilchrist, Prof Jane Hurst, Prof Andrew King and Prof Hugh Montgomery for advice on the mysteries of constructing a PhD thesis and to Simon Baker for valuable comments on the final draft.

In Nepal, I must thank Dr Maheshwar Dakal of the Department of National Parks and Wildlife Conservation and Dr Krishna Chandra Poudel of the Ministry of Forests and Soil Conservation for permission to carry out this study in Nepal. It would have been impossible to complete the fieldwork without the help of members of the Small Mammals Conservation and Research Foundation. Four of them accompanied me on six arduous field expeditions with unstinting good humour, enthusiasm and friendship; Hem Katuwal, Dipendra Adikhari, Sagar Dahal and especially Hari Basnet, who contributed to three expeditions. We were supported on the expeditions by many porters from the village of Keruwa near Pokhara. I would especially like to thank Khadga Ghole (the best camp cook in Nepal) and Padam Paudel for his remarkable field skills. Himalayan Nature provided essential logistical support in Kathmandu and I thank Prof Karan Shah, Dr Hem Sagar Baral and Mr Sharad Singh for their support and encouragement. Laboratory services were provided by the Centre for Molecular Dynamics Nepal and I thank the Director, Dr Dibesh Karmacharya for the use of the laboratory and his friendship. Many staff were involved in my study and I would especially like to thank Sulochana Manandhar, Adarsh Serchan, Ajay Sharma, Jyoti Joshi and Hemanta Chaudhary for their expertise and comradeship. Finally, but most importantly, my dear friends Laxman Poudyal (Chief Ecologist at DNPWC) and Sujan Maharjan (MoFaSC) guided me through many problems and pitfalls during six years – I will never be able to thank them adequately.

On several trips to India I received help and advice from many people. I would especially like to thank Dr D. Apte, Director of the Bombay Natural History Society, for unlimited access to the

society's collection and to Rahul Khot and Sameer Bajaru for their assistance and helpful discussions. Dr K. Venkatraman kindly gave permission for me to access the Zoological Survey of India collection in Kolkata, where I was ably assisted by Dr G. Sharma. At the Wildlife Institute of India in Dehra Dun, Prof Y. V. Jhala, Dr B. S. Adhikari, Dr S. P. Goyal and Dr S. A. Hussain all gave valuable guidance at the outset of my study. Prof Uma Ramakrishnan and Krishnapriya Tamma at the National Centre for Biological Sciences in Bangalore also provided much stimulating discussion and encouragement during my visit.

Funding for this project and the supply of small mammal traps was provided by BioEcoSS Ltd. In addition, Mr Ken Simms at Thrigby Hall Wildlife Gardens generously gave two research grants, for which I am very grateful.

My wife, Wendy, was the most important person throughout this period of study, for which I will be eternally grateful. Not once did she complain about the nine long trips to Nepal and she was always there with love and encouragement when my spirits flagged.

THESIS STRUCTURE

Detailed tables of contents, tables and figures are given at the start of each chapter.

Supplementary Information is generally appended to the relevant chapter. However, four additional PDF files are appended to the electronic version of the thesis.

Chapter 1 GENERAL INTRODUCTION

Contents

Figures

1.1 An Introduction to Aspects of Small Mammal Ecology

According to the current edition of Wilson and Reeder (2005), there are 5416 species of mammals recognised in the world, which fall into 29 orders and 153 families. This study focuses on two orders; Rodentia, which is by far most speciose with 2277 species (42%) and Soricomorpha which has 428 species (7.9%), exceeded in number only by the bats (Chiroptera). Both are ancient orders with the Soricomorpha, in particular, considered to be near the basal types of all mammals, with one genus (*Solenodon*) having diverged at least 76m years ago (Derbridge *et al.*, 2015). Within the rodents, I have concentrated on the two most diverse families, Muridae (rats and mice) with 730 species (13.5%) and Cricetidae (voles) with 681 species (12.6%). The third family of interest is the major sub-division of Soricomorpha, the Soricidae (shrews), which comprises 376 species (6.9% of all mammalian species).

Not only are these three families very diverse, accounting for one third of all known mammal species, they are also the most widespread geographically. Murids are indigenous throughout Eurasia, Africa and Australia (Myers, 2001), although they have become commensal in every continent except Antarctica. The cricetids are also very widespread, being found in both North and South America, and most of the palearctic region. However, they are absent from Africa and Australasia (Poor, 2005) and in the oriental region they are only found at high elevation (Corbet and Hill, 1992). The family Soricidae is distributed throughout most major landmasses of the world, excluding Australia and Antarctica, and their distribution in South America is limited to the northern zones near the Isthmus of Panama (Nowak, 1991). Shrews are also naturally absent from New Zealand, Tasmania, the Pacific Islands, the arctic islands (including Greenland and Iceland), the Ungava region in Canada and the West Indies.

The species belonging to these three families occupy a vast range of habitats. Even within the Indo-Malayan region alone (Corbet and Hill, 1992), examples include arid and semi-arid deserts (*e.g. Acomys*), grasslands (*e.g. Millardia* & *Mus*), rain forests (*e.g. Crateromys* & *Maxomys*), forested hills (*e.g. Episoriculus* & *Niviventer*) and high mountains (*e.g. Alticola*). Several species are known to occupy arboreal or semi-arboreal habitats, such as *Micromys* and *Vandeleuria*, or have fossorial (*Soriculus*) or semi-aquatic habits (*Nectogale* & *Neomys*).

A few species are highly commensal (Hulme-Beaman *et al.*, 2016) and have become the most ubiquitous of terrestrial mammals. In particular the black rat (*Rattus rattus*), the brown rat (*Rattus norvegicus*) and the house mouse (*Mus domesticus/musculus*) have spread to all continents except Antarctica, with the assistance of man (Aplin *et al.*, 2011). Recent DNA studies have shown that *M. musculus*

has colonized many islands, following human trade routes, presumably as stowaways on trade ships (Gabriel *et al.*, 2015). The common musk or house shrew, *Suncus murinus*, of south and south-east Asia is strongly commensal and has been introduced to several oceanic islands (Nakomoto and Nakanishi, 2013).

In a number of cases, the commensalism of rodents and shrews has developed into less benign relationships and several rodent species have become universal pests of agriculture and food (Myllymäki, 1979; Stenseth *et al.*, 2003; Wood and Singleton, 2014). Although only a small proportion of the known numbers of murid and cricetid species have been recorded as agricultural pests (Southern, 1979), pest species can become spectacularly numerous. For example, *M. musculus* reached densities of >800 ha^{-1} in outbreaks in Australia (Brown *et al.*, 2007) and the vole *Microtus agrestis* has been recorded at over 2000 ha^{-1} in grassland in NW Europe (Wood and Singleton, 2014). Even species which we now consider to be endangered and in need of conservation, such as the Eurasian harvest mouse *Micromys minutus*, have been recorded as pest species at densities above 200 ha^{-1} (Trout, 1978). These large outbreaks can cause immense damage, such as 30 million tonnes of rice production per annum in Asia (Singleton, 2003) and Stenseth *et al.* (2003) quote similar losses in Africa and South America.

Some species of murids and soricids can have a more direct and adverse affect on humans as vectors of emerging infectious diseases (Morand *et al.*, 2015). To cite just a few; hantavirus in *Peromyscus* sp. (Parmenter *et al.*, 1998), scrub typhus from *Rattus exulans* in Taiwan (Kuo *et al.*, 2011) and leptospirosis in SE Asia found especially in *Berylmys berdmorei*, *Mus cookii*, *Rattus argentiventer* and *R. losea* (Cosson *et al.*, 2014). Of particular relevance to this study was *Bartonella* found in *Bandicota bengalensis*, *R. rattus* and *S. murinus* from Nepal (Gundi *et al.*, 2010). The most infamous example of a zoonotic disease with a rodent vector (*Rattus rattus*) is undoubtedly the plague *Yersinia pestis* (Audoin-Rouzeau, 1999; Brouat *et al.*, 2013).

From an ecological perspective small mammals can often play a vital role in the functioning of ecosystems. Although individual species of murids, cricetids and soricids may not constitute keystone species (Delibes-Mateos *et al.*, 2011; Mills *et al.*, 1993), together they undoubtedly provide important ecosystem services. Most importantly, they are prey for many species of small mammalian predators especially mustelids (Korpela *et al.*, 2014; Norrdahl and Korpimäki, 2000), felids (Loss *et al.*, 2013; Woods *et al.*, 2003) and canids (Baker *et al.*, 2006; Macdonald, 1977). Avian predators of small mammals include owls (Paspali *et al.*, 2013) and raptors (Norrdahl and Korpimäki, 1996; Olson *et al.*, 2017). Small mammals can also provide a service as seed dispersers (Sunyer *et al.*, 2013) although

they may have adverse effects on plant populations as seed predators (Bricker *et al.*, 2010; Flowerdew, 1976; Flowerdew and Gardner, 1978; Maron *et al.*, 2018). Shrews are known to be preferential predators of invertebrates, especially insect larvae (Hamilton, 1930; Rudge, 1968; Rychlik and Jancewicz, 2002) as well as coleopteran adults (Churchfield *et al.*, 1999). Buckner (1966) showed that masked shrews *Sorex cinereus* attained 90% predation rate on sawfly pupae in Manitoba and up to 50% in Newfoundland, providing a potentially valuable pest control service.

1.2 Ecological Studies of Mammals in the Indian Sub-Continent

Historically, ecological surveys and studies of mammals in the Indian sub-continent have tended to focus on large, charismatic, keystone species. In particular, observational studies were undertaken on large mammals such as the Indian elephant *Elephas maximus* (Eisenberg, 1980; Sukumar, 1989), the Indian one-horned rhinoceros *Rhinoceros unicornis* (Gee, 1963; Laurie *et al.*, 1983), the tiger *Panthera tigris* (McDougal, 1977; Schaller, 1967) and deer such as the sambar *Rusa unicolor* (Leslie, 2011). Other species of conservation interest have been studied in Nepal, such as the hispid hare *Caprolagus hispidus* (Bell, 1986; Khadka *et al.*, 2017) and the snow leopard *Panthera uncia* (Oli *et al.*, 1993; Schaller, 1977). In recent years ecological studies have been combined with genetic techniques to explore the taxonomy of species such as the Himalayan wolf *Canis lupus himalayensis* (Werhahn *et al.*, 2017) and the snow leopard (Karmacharya *et al.*, 2011).

With such a number of truly large mammals in the Indian sub-continent, the term "small mammal" is applied to species that would not be considered small in Europe. For example, species of particular conservation interest are the Indian pangolin *Manis crassicaudata* (Mahmood *et al.*, 2014), the red panda *Ailurus fulgens* (Choudhury, 2001) and the fishing cat *Prionailurus viverrinus* (Mukherjee *et al.*, 2012). Studies of small mammals *sensu stricta*, *i.e.* rodents and shrews generally less than 200g in body weight, are far less common considering the number of species in the subcontinent. Pradhan and Talmale (2009) list 103 species of rodent in India, although this includes six species of giant squirrels and porcupines. Of these, two murids are listed by the IUCN as critically endangered; *Cremnomys elvira* and *Millardia kondana* and three soricids; *Crocidura andamanensis*, *C. jenkinsi* and *C. nicobarica* (Anon., 2011). There are few ecological studies of murids or shrews, *e.g. Mus booduga* in Pakistan (Khan *et al.*, 2017) and a recent study of *M. kondana* (Bajaru, *pers. comm.*). In Chapter 2, I introduce and discuss the previous surveys of small mammals in Nepal and highlight the dearth of systematic, ecological studies.

1.3 An Introduction to the Geography and Habitat of Nepal

Nepal is a small country, approximately 147,000 km^2 in area, landlocked between India and China (Figure 1.1). Geographically, it is probably most famous as the location (jointly with China) of the highest mountain in the world at 8848m; known in the west as Mt. Everest but in Nepal as Sagarmatha, "goddess of the sky". As the lowest altitude in the country is approx. 60m, the elevational range is huge, but occurs over a horizontal distance of less than 200km. The country runs for approx. 800km along the Himalayan chain which is cut by a number of major river systems notably, from west to east, the Ghaghara, the Gandaki/Narayani, and the Kosi/Arun. One of the main tributaries of the second of these, the Kali Gandaki, is said to form the deepest gorge in the world, where it flows between the Dhaulagiri and Annapurna massifs, producing a valley over 6500m deep. Nepal can be divided into three major elevational zones; the lowlands (< 500m) along the border with India, known as the Terai, the mid-hills (500 – 3000m) encompassing the Siwalik and Mahabharat ranges and the high Himalaya (> 3000m) which also includes small regions of trans-Himalaya or Tibetan plateau.

Figure 1.1 Relief map of Nepal. (© ezilon.com.)

The climate of Nepal is driven largely by its sub-tropical latitude (26.4°N to 30.5°N) and its proximity to the Indian sub-continent. Consequently, it has three primary seasons; a pronounced monsoon period from June to early-September, followed by a cool dry season from October to March, and a short hot dry season in April and May. Temperatures and rainfall are further influenced by the extreme altitudinal range. Temperatures in the lowland Terai can reach 40°C in the hot season and rarely fall below 15°C. In contrast, the highest inhabited regions (approx.. 4500m) have sub-zero

day-time temperatures during the winter (and can fall to –20°C at night) but can reach 25°C in the hot season. Rainfall during the monsoon season can be very heavy, with Nepalganj in the Terai receiving over 1000mm and Pokhara in the mid-hills up to 3000mm annually. In the dry season, rainfall is sparse in most parts of the country, usually below 20mm per month.

The distribution of the major land cover zones of Nepal is driven primarily by these extremes of climate and altitude (Figure 1.2). Much of the lowland Terai has been converted to agriculture, as have many of the valleys in the mid-hills, albeit on a very small scale. The natural habitat of much of the lowlands and mid-hills was broad-leaved closed forest, with needle-leaved open and closed forest above 2500m. Alpine grassland occurs in a narrow altitudinal band between 3500m and 4500m, above which barren ground and rock, snow and glacier predominate.

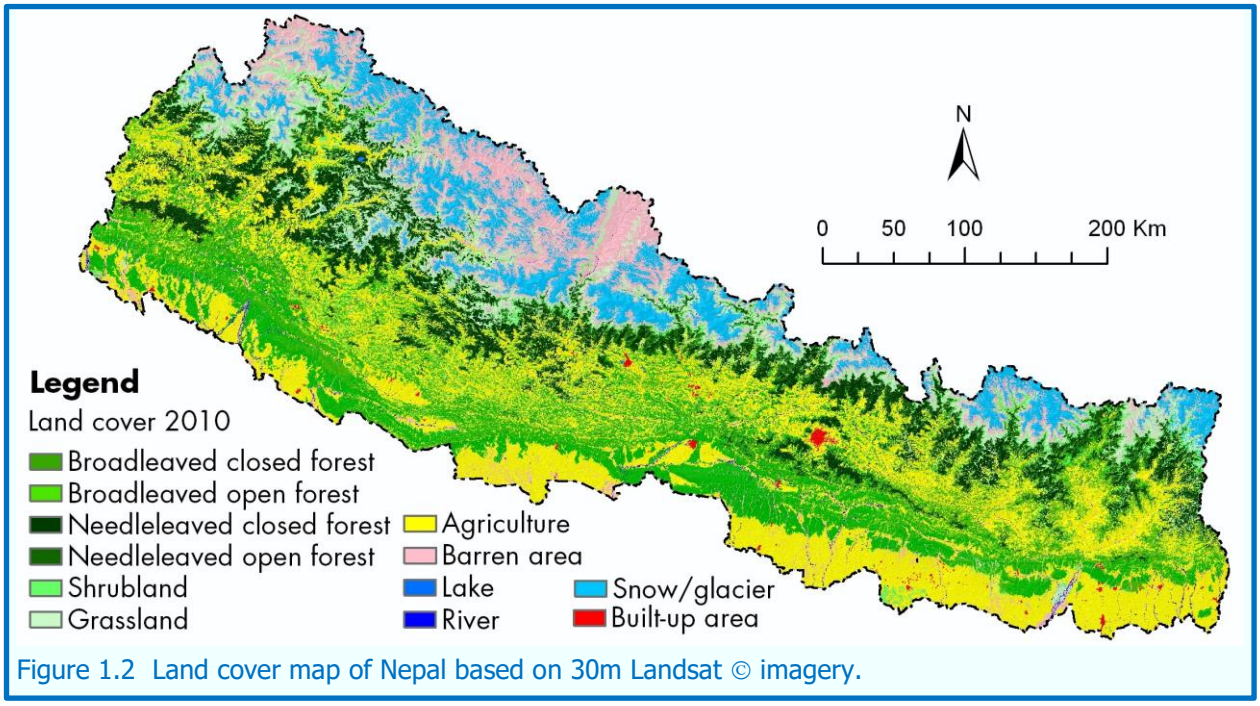


Figure 1.2 Land cover map of Nepal based on 30m Landsat © imagery.

The population of Nepal currently stands at around 29m. However, with a median age of approx. 22 and an annual growth rate of 2.3%, the population is set to rise over the next two decades. This pressure, plus the added effect of climate change is forcing human habitation to encroach further up the altitudinal range. Small scale farming can be found above 3000m in parts of the Solu Khumbu region and livestock are grazed at higher altitudes. This is placing greater pressure on the natural habitats, especially the *Rhododendron* woodland and alpine grasslands, and is opening up more areas for hunting and poaching.

1.4 The Scope of this Thesis

Following this introduction, the thesis is divided into five chapters. These have different research questions and are based on different datasets, the results of which are brought together in the final chapter.

In Chapter 2, I introduce the target species and describe the fieldwork undertaken during the current survey. I explain the survey design and field methodology employed, as well as the collection of other types of data, including topographic and weather data. The collection and preparation of specimens and tissue samples is also briefly introduced, as a precursor to the full description given in Chapter 5. Basic trapping results are presented with analyses of capture rates, void-trap rates and mortality rates. An assessment of recapture rates and population estimates is also made, as well as movements demonstrated by marked animals. Finally, I give a full appraisal of the field identifications, with descriptions of the putative species caught. It is important to emphasise that the analyses presented in this chapter are mostly anonymous with respect to species identities, as detailed analyses by species are given in Chapter 6.

Chapter 3 is also based on field data collected during the current survey. Here, I describe the methods used to collect a suite of environmental variables, including habitat structure, proximity to features such as human habitation and botanical composition. In turn, these are used to examine and characterise the trapping grids, employing a number of multivariate analytical procedures, including cluster analysis. I also introduce a novel, computer-intensive genetic algorithm which is used to explore the characterisations of the trapping grids. Finally, in this chapter, I analyse the spatial and altitudinal relationships with these grid characteristics.

In Chapter 4, I move from field data that I collected personally to the collation and analysis of historical data. This enables me to set my own study into the context of small mammal surveys undertaken over a larger geographical region and over a much longer time span. In this chapter I present two datasets; the specimen collection of the Bombay Natural History Society (BNHS), numbering over 7200 individuals, and the monograph by Pearch (2011), which collates records of small mammals in Nepal from a range of online museum catalogues and other sources. These two datasets complement each other in space and time. The BNHS collection covers a wide geographical area of approx. 5.25 million square kilometres, which encompasses seven modern countries, and the majority of the collection was made around 100 years ago. For these reasons, the proportion of records from Nepal is small. In contrast, the Pearch monograph only covers Nepal, but provides nearly 1000 records, mostly collected between 1950 and 2000. In this chapter, I use a Digital Elevation Model

(DEM), in combination with these two source datasets, to derive predictive models for approximately 50 species, using two different modelling methods. I then project the models onto the DEM for Nepal to predict likely presence of species of interest to this study.

Chapter 5 takes another new concept; the use of genetic methods to corroborate species identities. Here, I describe the field methods used to collect tissue samples and the laboratory methods employed to extract DNA, carry out polymerase chain reactions (PCR) and obtain sequences of the cytochrome-b gene from approx. 100 animals. In addition, 200 sequences were obtained from the NCBI GenBank database, which I analysed phylogenetically to derive average conspecific and congeneric thresholds of genetic distances, and to uncover a number of anomalous identifications. By combining the sequences from my tissue samples with selected GenBank sequences, I present an insight into the identities of my specimens which would not have been possible with morphological field identifications alone.

Finally, in Chapter 6, I consolidate the previous chapters to give a comprehensive analysis of species identifications. This brings together the original field identifications (Chapter 2), with the geographical predictions from the historical datasets (Chapter 4) and the genetic analyses described above (Chapter 5). These are then used to address the main theme of this thesis; the altitudinal variation in distribution and abundance of small mammal species and communities. I accomplish this in four stages. Firstly, I analyse the relationship between the overall abundance of small mammals with geographic and altitudinal variation. Secondly, I compare these relationships between taxa, such as the two main families (mice and shrews) and between individual species. Thirdly, I present an analysis of the relationships between the habitat variables derived in Chapter 3 and the distribution of individual small mammal species. Finally, I calculate two measures of species diversity, α and β–diversity, which are analysed against geographic range and altitudinal variation. I conclude the chapter with a discussion of opportunities for future research.

1.5 References

Anon. (2011) *Critically Endangered Animal Species of India.* Zoological Survey of India, Kolkatta, India. pp 26.

Aplin, K.P., Suzuki, H., Chinen, A.A., Chesser, R.T., Ten Have, J., Donnellan, S.C., Austin, J., Frost, A., Gonzalez, J.P., Herbreteau, V., Catzeflis, F., Soubrier, J., Fang, Y.P., Robins, J., Matisoo-Smith, E., Bastos, A.D., Maryanto, I., Sinaga, M.H., Denys, C., Van Den Bussche, R.A., Conroy, C., Rowe, K., and Cooper, A. (2011) Multiple geographic origins of commensalism and complex dispersal history of Black Rats. *PLoS One* **6**: e26357.

Audoin-Rouzeau, F. (1999) [The black rat (*Rattus rattus*) and the plague in ancient and medieval western Europe]. *Bulletin De La Societe De Pathologie Exotique (1990)* **92**: 422-426.

Baker, P., Furlong, M., Southern, S., and Harris, S. (2006) The potential impact of red fox *Vulpes vulpes* predation in agricultural landscapes in lowland Britain. *Wildlife Biology* **12**: 39-50.

Bell, D. (1986) A study of the hispid hare *Caprolagus hispidus* in Royal Suklaphanta Wildlife Reserve, western Nepal: a summary report. *Dodo: Journal of the Jersey Wildlife Preservation Trust* **23**: 24-31.

Bricker, M., Pearson, D., and Maron, J. (2010) Small-mammal seed predation limits the recruitment and abundance of two perennial grassland forbs. *Ecology* **91**: 85-92.

Brouat, C., Rahelinirina, S., Loiseau, A., Rahalison, L., Rajerison, M., Laffly, D., Handschumacher, P., and Duplantier, J.-M. (2013) Plague circulation and population genetics of the reservoir Rattus rattus: the influence of topographic relief on the distribution of the disease within the Madagascan focus. *PLoS neglected tropical diseases* **7**: e2266.

Brown, P., I. Huth, N., Banks, P., and Singleton, G. (2007) *Relationship between abudance of rodents and damage to agricultural crops*. pp 405-415.

Buckner, C.H. (1966) The Role of Vertebrate Predators in the Biological Control of Forest Insects. *Annual Review of Entomology* **11**: 449-470.

Choudhury, A. (2001) An overview of the status and conservation of the red panda *Ailurus fulgens* in India, with reference to its global status. *Oryx* **35**: 250-259.

Churchfield, S., Nesterenko, V.A., and Shvarts, E.A. (1999) Food niche overlap and ecological separation amongst six species of coexisting forest shrews (Insectivora: Soricidae) in the Russian Far East. *Journal of Zoology* **248**: 349-359.

Corbet, G.B. and Hill, J.E. (1992) *The Mammals of the Indomalayan Region: A systematic Review.* Oxford University Press. pp 488.

Cosson, J.-F., Picardeau, M., Mielcarek, M., Tatard, C., Chaval, Y., Suputtamongkol, Y., Buchy, P., Jittapalapong, S., Herbreteau, V., and Morand, S. (2014) Epidemiology of leptospira transmitted by rodents in southeast Asia. *PLoS neglected tropical diseases* **8**: e2902-e2902.

Delibes-Mateos, M., Smith, A.T., Slobodchikoff, C.N., and Swenson, J.E. (2011) The paradox of keystone species persecuted as pests: A call for the conservation of abundant small mammals in their native range. *Biological Conservation* **144**: 1335-1346.

Derbridge, J.J., Posthumus, E.E., Chen, H.L., and Koprowski, J.L. (2015) *Solenodon paradoxus* (Soricomorpha: Solenodontidae). *Mammalian Species* **47**: 100-106.

Eisenberg, J.F. (1980) Ecology and Behaviour of the Asian Elephant. *Digital Commons* **1**: 36-56.

Flowerdew, J.R. (1976) The effect of a local increase in food supply on the distribution of woodland mice and voles. *Journal of Zoology* **180**: 509-513.

Flowerdew, J.R. and Gardner, G. (1978) Small rodent populations and food supply in a Derbyshire ashwood. *Journal of Animal Ecology* **47**: 725-740.

Gabriel, S.I., Mathias, M.L., and Searle, J.B. (2015) Of mice and the 'Age of Discovery': the complex history of colonization of the Azorean archipelago by the house mouse (*Mus musculus*) as revealed by mitochondrial DNA variation. *Journal of Evolutionary Biology* **28**: 130-145.

Gee, E.P. (1963) Report on a Brief Survey of the Wild Life Resources of Nepal, Including the Rhinoceros. *Oryx* **7**: 67-76.

Gundi, V.A., Kosoy, M.Y., Myint, K.S., Shrestha, S.K., Shrestha, M.P., Pavlin, J.A., and Gibbons, R.V. (2010) Prevalence and genetic diversity of bartonella species detected in different tissues of small mammals in Nepal. *Applied and Environmental Microbiology* **76**: 8247-8254.

Hamilton, W.J. (1930) The Food of the Soricidae. *Journal of Mammalogy* **11**: 26-39.

Hulme-Beaman, A., Dobney, K., Cucchi, T., and Searle, J.B. (2016) An Ecological and Evolutionary Framework for Commensalism in Anthropogenic Environments. *Trends in Ecology & Evolution* **31**: 633-645.

Karmacharya, D.B., Thapa, K., Shrestha, R., Dhakal, M., and Janecka, J.E. (2011) Noninvasive genetic population survey of snow leopards (*Panthera uncia*) in Kangchenjunga conservation area, Shey Phoksundo National Park and surrounding buffer zones of Nepal. *BMC Res Notes* **4**: 516.

Khadka, B., Prasad Yadav, B., Aryal, N., and Aryal, A. (2017) *Rediscovery of the hispid hare (Caprolagus hispidus) in Chitwan National Park, Nepal after three decades*. pp 10-13.

Khan, A.A., Khan, W.A., and Khan, N.-H. (2017) Abundant Distribution of Little Indian Field Mouse, *Mus booduga*, in the Wetlands of Salt Range, Punjab, Pakistan. *Pakistan Journal of Zoology* **49**: 481-485.

Korpela, K., Helle, P., Henttonen, H., Korpimäki, E., Koskela, E., Ovaskainen, O., Pietiäinen, H., Sundell, J., Valkama, J., and Huitu, O. (2014) Predator-vole interactions in Northern Europe: the role of small mustelids revised. *Proc Biol Sci* **281**: 20142119.

Kuo, C.-C., Wang, H.-C., and Huang, C.-L. (2011) The potential effect of exotic Pacific rats *Rattus exulans* on vectors of scrub typhus. *Journal of Applied Ecology* **48**: 192-198.

Laurie, W.A., Lang, E.M., and Groves, C.P. (1983) Rhinoceros unicornis. *Mammalian Species*: 1-6.

Leslie, J.D.M. (2011) *Rusa unicolor* (Artiodactyla: Cervidae). *Mammalian Species* **43**: 1-30.

Loss, S.R., Will, T., and Marra, P.P. (2013) The impact of free-ranging domestic cats on wildlife of the United States. *Nature Communications* **4**: 1396.

Macdonald, D.W. (1977) On food preferences in the red fox. *Mammal Review* **7**: 7-23.

Mahmood, T., Irshad, N., and Hussain, R. (2014) Habitat preference and population estimates of Indian pangolin (*Manis crassicaudata*) in district Chakwal of Potohar Plateau, Pakistan. *Russian Journal of Ecology* **45**: 70-75.

Maron, J., L. Hajek, K., Hahn, P., and Pearson, D. (2018) Rodent seed predators and a dominant grass competitor affect coexistence of co-occurring forb species that vary in seed size. *Journal of Ecology* **106**: 1795-1805.

McDougal, C. (1977) *The Face of the Tiger*. André Deutsch. pp 180.

Mills, L.S., Soul, xe, E., M., and Doak, D.F. (1993) The Keystone-Species Concept in Ecology and Conservation. *Bioscience* **43**: 219-224.

Morand, S., Jittapalapong, S., and Kosoy, M. (2015) Rodents as hosts of infectious diseases: biological and ecological characteristics. *Vector borne and zoonotic diseases (Larchmont, N.Y.)* **15**: 1-2.

Mukherjee, S., Adhya, T., Thatte, P., and Ramakrishnan, U. (2012) Survey of the Fishing Cat *Prionailurus viverrinus* Bennett, 1833 (Carnivora: Felidae) and some aspects impacting its conservation in India. *Journal of Threatened Taxa* **4**: 3355-3361.

Myers, P. (2001) "Muridae" (On-line), Animal Diversity Web. Accessed December 10, 2018 at https://animaldiversity.org/accounts/Muridae/.

Myllymäki, A. (1979) *Importance of small mammals as pests in agriculture and stored products*. Pages 239-279 *in* D.M. Stoddart, editor. *Ecology of Small Mammals*. Chapman & Hall, London.

Nakomoto, A. and Nakanishi, N. (2013) Home Range, Habitat Selection, and Activity of Male Asian House Shrews, *Suncus murinus*, on Okinawa-Jima Island. *Mammal Study* **38**: 147-153.

Norrdahl, K. and Korpimäki, E. (1996) *Do nomadic avian predators synchronize population fluctuations of small mammal? A field experiment*. pp 478-483.

Norrdahl, K. and Korpimäki, E. (2000) The impact of predation risk from small mustelids on prey populations. *Mammal Review* **30**: 147-156.

Nowak, R.M. (1991) *Walker's Mammals of the World*. 5th edition. Johns Hopkins University Press, Baltimore, USA. pp 1629.

Oli, M.K., Taylor, I.R., and Rogers, D.M.E. (1993) Diet of the snow leopard (*Panthera uncia*) in the Annapurna Conservation Area, Nepal. *Journal of Zoology* **231**: 365-370.

Olson, L.E., Squires, J.R., Oakleaf, R.J., Wallace, Z.P., and Kennedy, P.L. (2017) Predicting above-ground density and distribution of small mammal prey species at large spatial scales. *PLoS One* **12**: e0177165-e0177165.

Parmenter, C., Yates, T., Parmenter, R., Mills, J., Childs, J., Campbell, M., Dunnum, J., and Milner, J. (1998) Small mammal survival and trapability in mark-recapture monitoring progams for hantavirus. *Journal of Wildlife Diseases* **34**: 1-12.

Paspali, G., Oruci, S., Koni, M., Frank Wilson, I., Krystufek, B., and Bego, F. (2013) Seasonal variation of small mammals in the diet of the barn owl (*Tyto alba*) in the Drinos River valley, southern Albania. *Turkish Journal of Zoology* **37**: 97-105.

Pearch, M.J. (2011) A review of the biological diversity and distribution of small mammal taxa in the terrestrial ecoregions and protected areas of Nepal. *Zootaxa* **3072** 1-286.

Poor, A. (2005) "Cricetidae" (On-line), Animal Diversity Web. Accessed December 10, 2018 at https://animaldiversity.org/accounts/Cricetidae/.

Pradhan, M.S. and Talmale, S.S. (2009) *A Checklist of Valid Indian Rodent Taxa (Mammalia: Rodentia).* Zoological Survey of India, Pune, India. pp 13.

Rudge, M.R. (1968) The Food of the Common Shrew *Sorex araneus* L. (Insectivora: Soricidae) in Britain. *Journal of Animal Ecology*: 565-581.

Rychlik, L. and Jancewicz, E. (2002) Prey size, prey nutrition, and food handling by shrews of different body sizes. *Behavioral Ecology* **13**: 216-223.

Schaller, G.B. (1967) *The Deer and the Tiger*. University of Chicago Press, Chicago, USA. pp 370.

Schaller, G.B. (1977) *Mountain Monarchs: Wild Sheep and Goats of the Himalaya*. University of Chicago Press, London. pp 425.

Singleton, G. (2003) *Impacts of rodents on rice production in Asia. IRRI Discussion Paper Series No. 45.* International Rice Research Insitute, Los Baños, Philippines. pp 30.

Southern, H.N. (1979) *Population processes in small mammals*. Pages 63-101 *in* D.M. Stoddart, editor. *Ecology of Small Mammals*. Chapman & Hall, London.

Stenseth, N.C., Leirs, H., Skonhoft, A., Davis, S.A., Pech, R.P., Andreassen, H.P., Singleton, G.R., Lima, M., Robert, S.M.u., Makundi, R.H., Zhang, Z., Brown, P.R., Shi, D., and Wan, X. (2003) Mice, Rats, and People: The Bio-Economics of Agricultural Rodent Pests. *Frontiers in Ecology and the Environment* **1**: 367-375.

Sukumar, R. (1989) Ecology of the Asian Elephant in Southern India. I. Movement and Habitat Utilization Patterns. *Journal of Tropical Ecology* **5**: 1-18.

Sunyer, P., Muñoz, A., Bonal, R., and Espelta, J.M. (2013) The ecology of seed dispersal by small rodents: a role for predator and conspecific scents. *Functional Ecology* **27**: 1313-1321.

Trout, R.C. (1978) A review of studies on populations of wild harvest mice (*Micromys minutus* (Pallas)). *Mammal Review* **8**: 159-175.

Werhahn, G., Senn, H., Kaden, J., Joshi, J., Bhattarai, S., Kusi, N., Sillero-Zubiri, C., and Macdonald, D.W. (2017) Phylogenetic evidence for the ancient Himalayan wolf: towards a clarification of its taxonomic status based on genetic sampling from western Nepal. *Royal Society Open Science* **4**: 170186.

Wilson, D.E. and Reeder, D.M. (2005) *Mammal Species of the World. A Taxonomic and Geographic Reference*. 3rd edition. Johns Hopkins University Press. pp 2142.

Wood, B.J. and Singleton, G. (2014) *Rodents in agriculture and forestry*. Pages 33-80 *in* A.P. Buckle and Smith, R.H., editors. *Rodent Pests and Their Control: 2nd Edition*. CAB International.

Woods, M., McDonald, R.A., and Harris, S. (2003) Predation of wildlife by domestic cats *Felis catus* in Great Britain. *Mammal Review* **33**: 174-188.

Chapter 2 The Small Mammal Community of the Central Himalaya, Nepal

Abstract

The geographical and temporal scope of this study makes it the largest ecological small mammal survey ever undertaken in the Himalayan region of Nepal, requiring 720 staff-days in the field. Furthermore, the systematic survey design, based on altitudinal transects, with regularly placed sites and using a standard grid layout makes this study unique. The fieldwork was undertaken over three seasons, on four altitudinal transects in the Annapurna, Langtang and Sagarmatha regions of Nepal. 147 trapping grids were laid, using newly-designed live-catch traps, at altitudes between 1300m and 4400m, comprising 8046 trap events. This yielded a total of 890 captures, representing a basic capture rate of 11.4%. A full analysis of the factors affecting capture rates showed that the main effect was higher rates in dawn sessions (14.2%) compared to dusk sessions (1.7%). Mortality rates and recapture rates were also analysed and discussed. A total of 792 animals were caught which were identified in the field to 18 putative species or higher taxa. Three species dominated the captures; a mouse identified as Mus booduga *and two shrews;* Episoriculus caudatus *and* Soriculus nigrescens, *representing 80% of all animals caught. All species are described and illustrated and morphological analyses are presented which show a very high degree of discrimination between species, based on only two or three characteristics. The results from this study are compared with historic studies of small mammals, all of which used killing traps. The much greater number of very small shrews captured in this study are contrasted with historical studies and the reasons for this are discussed.*

Contents

Tables

Figures

2.1 Introduction

2.1.1 Background to Ecological Research on Small Mammals in Nepal

There is a long history of small mammal collecting in Nepal, essentially for taxonomic purposes. This is described eloquently by Pearch (2011) in his monograph of the small mammals of Nepal and will be discussed further in Chapter 4. Much of the early work in the 19th and the first half of the 20th centuries was carried out under the auspices of the Bombay Natural History Society (BNHS), which will also be discussed in more detail in Chapter 4.

There was considerable interest in collection expeditions to the Himalaya in the 20 year period, starting in the mid–1960s. In particular, Abe (1971, 1977) carried out collection surveys in the Annapurna and Langtang regions, collecting 14 species of non-volant small mammals. From 1966 to 1970 a total of 3695 mammal records were collected as part of a nation-wide survey of ectoparasites and their mammalian hosts (Mitchell and Punzo, 1976; Mitchell, 1977). Other important surveys during this period focused on the taxonomy of the genera *Apodemus*, *Rattus* and *Niviventer* (Martens and Niethammer, 1972; Niethammer and Martens, 1975) or were part of wider ecological/geological expeditions (Daniel, 2015). Two university expeditions took place in the late 1970s; the Durham University expedition to Langtang (Green, 1981) and the University of East Anglia expedition (Rands *et al.*, 1979). Around twenty years later, a Japanese team collected 131 specimens of cricetids, murids and soricids from the Annapurna region and the vicinity of Kathmandu (Mekada *et al.*, 2001). This appears to be the last large scale field survey specifically for small mammals undertaken in Nepal before the current study. However, in 2009 the Small Mammals Conservation and Research Foundation (SMCRF) was established in Kathmandu. It is a non-profit NGO dedicated to the conservation and research of biodiversity in Nepal with particular reference to diversity and conservation of small mammals. Over the last nine years it has become the focus for research into small mammals with a number of publication of local relevance, *e.g.* Acharya *et al.* (2010). During this study SMCRF collaborated closely with me on the fieldwork and logistics.

In 2005 a Conservation Assessment and Management Plan (C.A.M.P.) workshop on south Asian non-volant small mammals was held in Coimbatore, southern India (Molur *et al.*, 2005). This provided a comprehensive basis for assessing the research requirements relevant to the current study. In particular, the International Union for the Conservation of Nature (IUCN) status current at the time was given for all small mammal species known to exist within the proposed study area. Two further publications in 2011 provided additional information. Jnawali *et al.* (2011) gave assessments of the IUCN Red List status for Nepal. Pearch (2011) provided a comprehensive compilation of all

records of small mammals from Nepal, based on specimens held in museums that gave on-line access to their records and other published sources. The recent checklist compiled by Thapa (2014) brings together many of these sources and is the most up-to-date.

These sources have been used to compile a list of all small mammal species recorded in Nepal, based on three criteria;

- Species distributions must be known to exceed 1000m in altitude.
- Species should be terrestrial in habit, *e.g.* they do not include members of the family Squiridae.
- Specimens should commonly not exceed 100g body mass.

The first criterion was demanded by the survey design for the current study (see Section 2.2.1). The last two criteria were required to ensure that species could be trapped using the traps and field methods described in Section2.2.2. This compilation generated a list of 45 species belonging to three families and 15 genera (Table 2.1). However, there was disagreement among the three sources as to the presence of certain species in Nepal, so that the maximum number from any single source was 38. Six species of *Rattus* have been included in the table, despite their maximum published weights being over 100g (Corbet and Hill, 1992). The justification for this was that this is an important genus and sub-adults or juveniles would meet the size criterion. This list formed the basis for the identifications presented in sub-section 2.3.2 and discussed in sub-section 2.4.2.

Of particular interest for these species is the IUCN status presented by Molur *et al.* (2005) and Jnawali *et al.* (2011) in Table 2.1. There is considerable disagreement between these two sources, despite there being only six years between their publication. Furthermore, Molur *et al.* (2005) classify 11 of their 37 species as Data Deficient (DD), with a further seven as Vulnerable (V) or Endangered (E), and Jnawali *et al.* (2011) defined 19 of their 38 species as DD, with one as E. This all testifies to the urgent need for systematic field studies of small mammals across a wide geographic and altitudinal range.

Table 2.1 Compilation of the identity and status of species of non-volant small mammals weighing less than approx. 100g. from regions of Nepal > 1000m altitude. The first two columns giving the Status in Nepal are IUCN country statuses. The third column (Pearch) lists the number of discrete locations recorded for each species. Numbers marked * are unconfirmed identities.

Scientific Name	Vernacular Name	Status in Nepal			Habitat	Altitude	
		Molur *et al.* (2005)	Jnawali *et al.* (2011)	Pearch (2011)		Min	Max
a) Cricetidae							
Alticola roylei	Royle's vole		Data deficient		Coniferous forest & alpine grasslands	2500	4300
Alticola stoliczkanus	Stoliczka's vole	Vulnerable, near threatened	Data deficient	11	Temperate forest & alpine grasslands	4000	
Alticola stracheyi	Thomas' short-tailed vole	Least concern			Temperate forest & alpine grasslands	4000	
Cricetulus alticola	Ladakh hamster	Data deficient	Data deficient	2	Coniferous forest & alpine grasslands	4000	
Microtus leucurus	Blyth's vole	Least concern	Data deficient	4	Temperate forest & mountain	4500	
Microtus sikimensis	Sikkim vole	Least concern	Data deficient	35	Rhododendron forest and meadow	2500	
b) Muridae							
Apodemus gurkha	Himalayan wood mouse	Endemic to Nepal, Endangered (all records from Annapurna)	Endangered	29	*Rhododendron* forest	2400	3500
Apodemus pallipes	Ward's field mouse	Data deficient	Data deficient	14	Coniferous and *Rhododendron* forest	3000	
Apodemus sylvaticus	Eurasian wood mouse	Vulnerable, near threatened	Least concern		Montane forest	1850	3400
Mus booduga	Common Indian field mouse	Least concern	Least concern	14	Deciduous forest and fields		4000
Mus cervicolor	Fawn-coloured mouse	Least concern	Least concern	27	All		1000
Mus cookii	Cook's mouse	Data deficient	Data deficient	12	Temperate, broadleaf and coniferous forest	100	2500
Mus musculus	House mouse	Least concern	Least concern	59	All		3000
Mus pahari	Sikkim mouse	Sikkim - Least concern		1	Temperate forest		2000
Mus phillipsi	Wroughton's Small Spiny mouse	Data deficient	Data deficient		Arid scrub near fields	500	1500
Mus platythrix	Brown spiny mouse		Data deficient		Dry deciduous forest		2000
Mus saxicola	Brown spiny mouse	Data deficient	Least concern	3	Agricultural		1000
Mus terricolor	Earth-coloured mouse	Least concern	Least concern	4	Agricultural		1000
Niviventer eha	Little Himalayan rat	Least concern	Least concern	41	*Rhododendron* & coniferous forest and bamboo	2250	3700
Niviventer fulvescens	Chestnut rat	Least concern	Data deficient	33	Temperate, broadleaf and coniferous forest	180	2700
Niviventer niviventer	Himalayan white-bellied rat	Data deficient	Least concern	32	Temperate, broadleaf and coniferous forest		3600
Rattus andamanensis	Sikkim rat		Data deficient	1*			
Rattus nitidus	White-footed Himalayan rat	Least concern	Least concern	22	Montane temperate forest	1000	3000
Rattus rattus	Black rat	Least concern	Least concern	42	All		3500
Rattus pyctoris	Himalayan rat		Least concern	46	Montane temperate forest	1000	3000
Rattus sikkimensis	Sikkim rat	Least concern			Montane temperate forest		2000
Rattus tanezumi	Oriental house rat	Not evaluated		46 (4*)	Montane temperate forest		3000
Vandeleuria oleracea	Indian long-tailed tree mouse	Least concern	Least concern	3	Deciduous forest	200	1500

Table 2.1 (Cont.) Compilation of the identity and status of species of non-volant small mammals weighing less than approx. 100g. from regions of Nepal > 1000m altitude. The first two columns giving the Status in Nepal are IUCN country statuses. The third column (Pearch) lists the number of discrete locations recorded for each species. Numbers marked * are unconfirmed identities. Generic names beginning (Epi) indicate that whilst Molur *et al.* refer to the genus *Soriculus*, this is now generally recognised as the genus *Episoriculus*.

Scientific Name	Vernacular Name	Status in Nepal			Habitat	Altitude	
		Molur *et al.* (2005)	Jnawali *et al.* (2011)	Pearch (2011)		Min	Max
c) Soricidae							
Chimmarogale himalayica	Himalayan water shrew	Endangered, vulnerable	Endangered	2	Temperate forests near streams	800	1500
Crocidura attenuata	Grey shrew	Endangered, vulnerable	Least concern	5	Foothill Terai		1000
Crocidura horsfieldi	Horsfield's shrew	Vulnerable	Data deficient	2*	Subtropical montane forest	100	4000
Crocidura perigrisea	Pale grey shrew		Data deficient		Temperate montane forest	4000	
Nectogale elegans	Elegant water shrew	Vulnerable, near threatened	Data deficient	2	Montane temp forest with streams	900	2270
Sorex araneus	Common shrew			2 (1*)			
Sorex bedfordiae	Lesser stripe-backed shrew	Data deficient	Data deficient	9	Montane forest	4000	
Sorex cylindricauda	Stripe-backed shrew			3*			
Sorex excelsus	Highland shrew	Data deficient	Data deficient	1*			
Sorex minutus	Pygmy shrew	Data deficient	Data deficient	8	Temperate, broadleaf and coniferous forest	2280	3380
Sorex thibetanus	Tibetan shrew			3	Sub-alpine coniferous forest	2000	4000
(Epi)soriculus caudatus	Hodgson's brown-toothed shrew	Least concern	Least concern	58	Temperate, broadleaf and coniferous forest	1800	3600
(Epi)soriculus leucops	Indian long-tailed shrew	Least concern	Least concern	9	Temperate, broadleaf and coniferous forest		2990
(Epi)soriculus macrurus	Arboreal brown-toothed shrew	Data deficient	Data deficient	11	Temperate, broadleaf and coniferous forest	1700	1700
Soriculus nigriscens	Sikkim large-clawed shrew	Data deficient	Least concern	70	Temperate, broadleaf and coniferous forest	1500	4300
Suncus etruscus	Savi's pygmy shrew	Least concern	Data deficient	7	Temperate, broadleaf and coniferous forest		5000
Suncus murinus	House shrew	Least concern	Least concern	43	Montane forest and agriculture		3700
Species	**45**	**37**	**38**	**38**			

2.1.2 Aims and Objectives of the Current Field Survey

The main aim of the work covered by this chapter was to undertake a field survey of the small mammals of the mountains of central Nepal, incorporating a robust and repeatable survey design, with geographic and temporal replication. The specific objectives were:

- To establish a geo-located series of standard trapping grids.
- To use repeatable field methods to trap small mammals over at least two nights trapping.
- To identify the animals captured and record biological information.
- To compile a species list for grids, sites and transects
- To obtain estimates of relative or absolute species abundance.
- If possible, to analyse individual movements of animals to indicate directional or non-random movements.

2.2 Methods

All fieldwork was undertaken with the permission of the Department of National Parks and Wildlife conservation (DNPWC) (see Appendix 1; Sub-section 2.6.1). In addition, specific permission was required from the National Trust for Nature Conservation (NTNC) to work in the Annapurna Conservation Area. My proposal to DNPWC specified the use of live-catch traps and no explicit permission was granted to collect voucher specimens.

In June 2013, prior to the first full field season, a two-week pilot study was undertaken in the Langtang region. The main aims were to test field methodologies, explore suitable types of habitat and gain experience of the logistics necessary to complete the main field survey. A full description of the methods employed and the results are given in Supplementary Information 2.A.

2.2.1 Survey Design

The survey design comprised three regions; Annapurna, Langtang and Sagarmatha, approximately 200km apart in central and eastern Nepal (Figure 2.1). Four altitudinal transects were established, two in the Annapurna region (known as Pipar and Annapurna) and one each in Langtang and Sagarmatha. The length and altitudinal range of each transect varied according to local trails and human habitation but, overall, they encompassed an altitudinal range from 1300m to 4400m. Each transect had five or six sites at approximately 500m altitudinal intervals (Figure 2.2). Two transects had short horizontal distances between sites; in the case of the Pipar transect, these totalled just over

Figure 2.1 Relief map of Nepal showing the three survey regions; Annapurna (blue), Langtang (magenta) & Sagarmatha (green).

5km, making it the steepest transect. The longest transect (Sagarmatha) had six sites, with a horizontal distance between the first and last sites of nearly 24km. Detailed maps and elevational profiles are given in Supplementary Information 2.B and the habitat of the sites are described in more detail in Chapter 3.

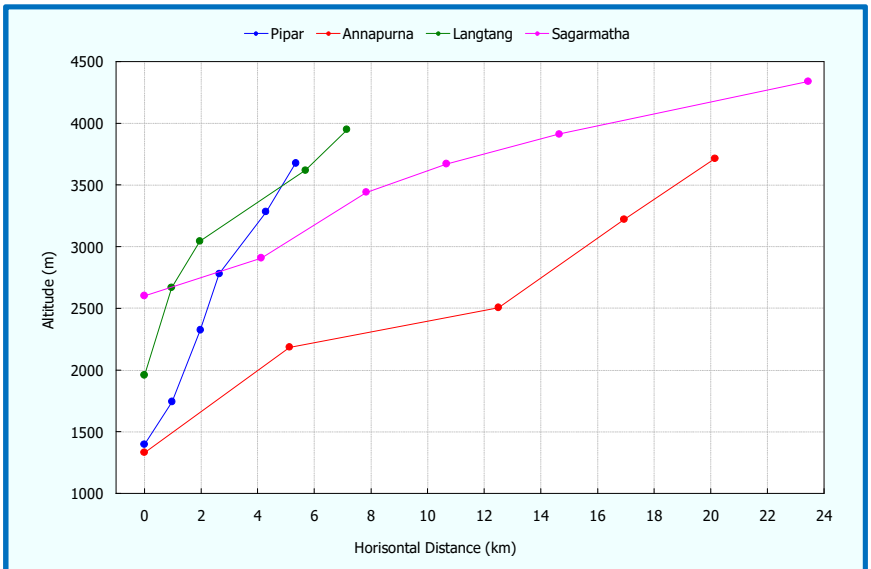


Figure 2.2 Schematic representation of the four transects showing the straight-line horizontal distance between sites and the mean altitude of the grids for each site.

Most sites comprised six trapping grids, although four sites only had four or five grids (Table 2.2). Where possible, these were set entirely within a single habitat and were placed a minimum of 100m, usually at least 200m, apart. The trapping grid formed the primary sampling unit of the survey design.

Transects were visited in 2013, 2014 and 2015. The Pipar transect was visited twice; in 2013 and 2014, while the other three were visited once (Table 2.2). All visits took place in the 10 week window between the end of the monsoon (usually the first two weeks in September) and the onset of winter snows (the first week in December). In 2013 and 2014, two transects were completed in one season but in 2015, only one transect was possible due to its remote location.

Each transect took between 17 and 23 days to complete. This usually commenced with a two to four hour trek from the road-head to the first site on the transect. Trekking between sites took anything from four to eight hours. The Pipar and Annapurna transects were undertaken early in the season, so sites were visited on the ascent of the transect. The descent from the top of the transects only took one, albeit extremely arduous, day. In contrast, the Langtang and Sagarmatha transects were carried out later in the season and were generally higher. To avoid problems of snow and low temperatures at the top of the transect, the sites were completed on the descent. This meant

Table 2.2 Numbers of trapping grids within each region, transect and site, visited by year.

Region	Transect	Site	Trapping Grids				
			2013	2014	2015	Visits	Locations
Annapurna	1 Pipar Trail	1 Keruwa	6	6		12	6
		2 Thulo Kobang	4	4		8	4
		3 Pilicho	5	5		10	5
		4 Pipar Pond	6	6		12	6
		5 Sano Kobang		6		6	6
		6 Bunga Ledu		5		5	5
		Total	21	32		53	32
	2 Annapurna Base Camp Trail	1 Kyumi	6			6	6
		2 Chomrong	6			6	6
		3 Dhorban	6			6	6
		4 Deurali	4			4	4
		5 MBC	6			6	6
		Total	28			28	28
Langtang	3 Gosain Kund Trail	1 Laurabina		6		6	6
		2 Cholong Pati		6		6	6
		3 Dimsa		6		6	6
		4 Deurali		6		6	6
		5 Ghhata Kola		6		6	6
		Total		30		30	30
Sagarmatha	4 Everest Base Camp Trail	1 Phakding			6	6	6
		2 Monjo			6	6	6
		3 Namche Bazaar			6	6	6
		4 Kyangjuma			6	6	6
		5 Tangboche			6	6	6
		6 Dingboche			6	6	6
		Total			36	36	36
		Survey Total	49	62	36	147	126

that two or four days were required to make the initial ascent, although the treks between sites, being descents, were much easier and quicker.

Trekking distances and ascents were a major consideration in the design of the survey. Approximately 259km were covered trekking between sites and on the ascent or descent of the transects, which required 21,250m of ascent and descent (Table 2.3). In addition, 360 km and 21,750m of

Table 2.3 Distances walked and meters of ascent during trekking between sites, trapping sessions and during the whole study.

Year	Transect	Trekking		Trapping			Total		Annual Totals	
		Distance (km)	Ascent (m)	Grid Visits	Distance (km)	Ascent (m)	Distance (km)	Ascent (m)	Distance (km)	Ascent (m)
2013	Pilot	35	2806	18	76	3825	111	6631	184	14156
	1: Pipar	18	2296	24	48	2575	66	4871		
	2: Annapurna	71	5485	22	47	3800	118	9285		
2014	1: Pipar	27	3277	30	58	3675	84	6952	167	12402
	3: Langtang	27	2325	25	55	3125	82	5450		
2015	4: Sagarmatha	80	5060	30	78	4750	157	9810	157	9810
	Total	259	21,249		360	21,750	618	42,999		

ascent were completed during the trapping sessions. In total, including the pilot study, we completed 112 days in the field during which we trekked approx. 620km and ascended/descended 43,000m.

2.2.2 The Field Methods

Each trapping grid comprised 18 live-catch traps set in a cross-shaped design (Figure 2.3). Traps were laid in pairs within 1m of each other, with two pairs in each of the four arms of the cross and one pair at the centre. The pairs were spaced 10m apart, so the overall dimensions of the trapping grid were 40m × 40m. The first arm (holding trap numbers 3 – 6 ran up the slope, with the third arm (traps 11 – 14) running down the slope (the Y dimension). The other two arms ran along the contour of the slope where possible (the X dimension). Consequently, grids were oriented according to the slope, rather than compass bearings. Some grids had a slope of over 30°, so the top pair of traps could have been >20m vertically above the bottom pair. On each grid, one photograph was taken from the centre point along each of the arms to record the topography and habitat.

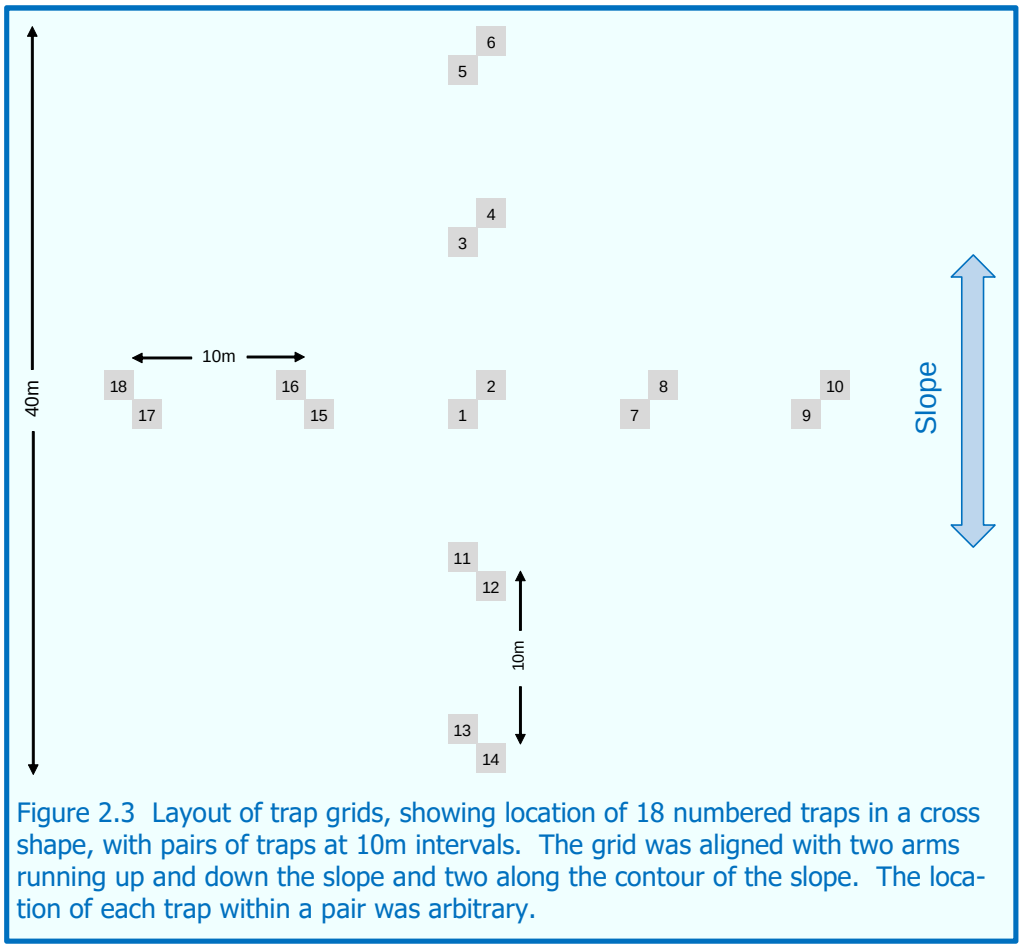


Figure 2.3 Layout of trap grids, showing location of 18 numbered traps in a cross shape, with pairs of traps at 10m intervals. The grid was aligned with two arms running up and down the slope and two along the contour of the slope. The location of each trap within a pair was arbitrary.

The BioEcoSS TubeTrap (Poulton, 2018) was the only type of trap employed in the study (Figure 2.4). Field trials carried out at UEA have shown that the TubeTrap has a significantly higher capture-rate than Longworth traps (Le-Moine, 2017; Murgatroyd, 2017) and is also superior to metal traps in a number of ways;

- It is a small, light-weight trap of plastic construction and its cylindrical design makes it extremely strong for its weight.
- It is especially sensitive to small animals, such as shrews, weighing <5g, but is also suitable for animals up to 100g.
- It has a double-walled nestbox which gives additional insulation, making is especially suitable for high altitude, low temperature habitats.
- It is extremely easy to repair in the field as all parts are snap-fit and the lack of internal corners makes it extremely easy to clean.

Figure 2.4 The BioEcoSS TubeTrap.

All traps were placed on the ground, within cover as far as possible, to shade them from sun and/or frost. In farmland, this was often in the vegetation at the base of terraces or in the crops themselves. In grassland, they were placed deep inside tussocks of vegetation or within shrubs. In woodland, traps were placed at the base of trees, within fallen rotting tree trunks or deep within leaf litter and moss. When rocky outcrops were present, they were placed at the base of boulders or, if available, within crevasses or under piles of stones. Trapping grids were occasionally located near riparian habitats, usually very small streams less than 1m across, but larger rivers on occasion. In this case, traps were placed as close to the water as possible (c. 1m), without risking loss due to flash floods.

In montane habitats, traps were located under whatever cover was available; usually dwarf shrubs or tussock grass.

Traps were filled with dry grass or moss for bedding and supplied with *ad libitum* food in the form of grain (rolled oats, rice or barley) and chopped vegetables for rodents. Food for shrews was difficult to obtain in Nepal, but tiny dried fish were readily available, which were rehydrated to make them more palatable. This was augmented with chicken skin for the first few days of trapping. In addition a variety of potential food types (dried mealworms, fat-balls with insects, *etc*.) were brought from UK.

All trapping grids within a site were laid at the same time. In most sites, traps were set for two nights, with no pre-baiting. In three sites, adverse weather restricted trapping to one night. In one site (Pipar Site 4 in 2013) two grids were set for three nights and four grids were set for four nights. Traps were set at dusk and revisited at dawn the next day. In total 447 grid trapping sessions were completed – 294 dawn and 153 dusk sessions. Any traps containing animals were cleaned, bedding replaced, food replenished and reset in exactly the same location. Traps were inspected again at dusk on the second day and finally at dawn on the third day. After this they were collected, cleaned thoroughly and packed ready for use at the next site.

Captured animals were removed from traps into heavy-duty polythene bags. They were identified to species level where possible, or to higher level taxa where identification was uncertain. Animals were sexed, aged and their breeding condition recorded, although this was often not possible for the small shrews. They were transferred to small zip-lock plastic bags and weighed using one of three appropriately scaled Pesola balances. In addition, measurements were taken of head and body length (HB), tail length, hind foot length (HF) and ear length (from the base to the tip). They were marked with a unique pattern of fur clips (Gurnell and Flowerdew, 2006) allowing individual animals to be identified on recapture. All this information was recorded on paper field forms (see Supplementary Information 2.E). Where possible, photographs were taken of the dorsal, ventral and lateral views to aid identification. In most cases, tissue samples were collected (see Section 2.2.5). Animals were then released at the point of capture.

During the trapping sessions a record was kept of all void-traps. These generally fell into four categories; a) closed but empty, b) disturbed with no capture, c) failed to trip but with obvious activity and d) damaged, destroyed or missing. In the latter event, traps were repaired where possible or replaced if further trapping sessions were pending.

Weather records were collected at each site for the period prior to each trapping session. These included a) minimum temperature, b) rainfall (binary), c) snowfall (binary) and d) a general description including wind direction and strength. As only two nights of snow were recorded, these were amalgamated with rainfall as a general "precipitation" binary variable.

2.2.3 Logistics

One member of the Small Mammals Conservation and Research Foundation (SMCRF) joined the field team for each transect. All participants were masters graduates from Kathmandu universities and all had experience of fieldwork. They were also invaluable as interpreters and coordinators of local staff and equipment.

A team of at least four porters was required for each transect, two to carry the traps and two for other equipment. When camping, at least two more porters (including a cook) were required to carry camping and cooking equipment as well as food and kerosene. Porters were also required to collect water and to assist in fieldwork, by carrying traps up to 1km from the hotel or camp site to the trapping grids. In total, the whole fieldwork project required approx. 720 man/days.

Sites were located around suitable camp-sites or trekking lodges. The site schedule took three nights to complete, with the last morning spent in one site becoming the first morning of the next site (Table 2.4). Generally, this meant that in each site five trips to the grids were required; one to explore for grid locations, one to set out the traps and three actual trapping sessions. Traps were collected on the last trapping session.

Table 2.4 3-night site schedule showing trapping sessions and repeating pattern between sites.

Day 1	am:	Trek to next site
	pm:	Manage tissue samples from previous site Explore site for grid locations
Day 2	am:	Explore site for grid locations Clean and prepare traps
	pm:	Set out trap grids
Day 3	am:	Trapping Session 1 Manage tissue samples and collate data
	pm:	Trapping Session 2
Day 4	am:	Trapping Session 3 Collect, clean and pack traps Strike camp and pack equipment

2.2.4 The Field Identification Handbook

Prior to the first field season in 2013, a field identification handbook was created, based on museum specimens held at the Bombay Natural History society (BNHS), the Zoological Survey of India (ZSI) in Kolkata and the Harrison Museum in UK. The bulk of the specimens came from the BNHS collection, which took eight days during three visits to compile. The handbook comprised 14 photographs of four species of Cricetidae, 64 photographs of 11 species of Muridae and 66 photographs of 13 species of Soricidae. The complete handbook (comprising 144 photographs of 28 species of interest) is presented in Supplementary Information 2.C.

2.2.5 Tissue Sample Collection and Management

Tissue samples were collected in the field from live animals using a 2mm ear punch (Harvard Apparatus Ltd.). One punch was taken from the ear of each animal and stored in a 2ml Eppendorf or screw-cap tube, holding 1ml of 99% ethanol. To avoid leakage and drying, tubes were sealed with a short strip of parafilm. Each tube was labelled or indelibly marked with a unique, five-element code:

YY_T_S_G_AA

where YY = two-digit year, T = transect, S = site, G = grid and AA = animal number.

During the field trips, samples were stored in sealed polythene bags. Although it was impossible to ensure stable, cool storage conditions, every precaution was taken to avoid exposure to direct sunlight. Some samples were inevitably held in field conditions for up to 20 days. After each field trip, all samples were checked to ensure adequate ethanol and stored at CMDN in an air-conditioned laboratory at ambient temperature (see Chapter 5 for more details).

Trap fatalities were collected. If time permitted, larger specimens (large mice, voles and rats) were skinned and prepared with borax. Skulls were removed, cleaned and dried as well as possible. Small mice and shrews were stored whole in 99% ethanol in screw-top plastic jars. If time permitted, tissue samples were taken from these specimens in the field to ensure consistency, but some specimens had tissue samples taken in the laboratory on return to Kathmandu. All skins, skulls and whole specimens were labelled in the same way as tissue samples using pencil on water-proof paper tied to the sample with nylon thread. After each field trip, whole specimens were sorted, the ethanol refreshed and stored at ambient temperature in the laboratory at CMDN.

2.2.6 Collection of Topographical, Habitat, Environmental and Weather Variables

A range of environmental variables were recorded at each trapping grid. These were broadly grouped into four types;

- Topographical – including altitude, slope and aspect, obtained from a SatMap Active 10 GPS (ref).
- Proximity – estimated distance in broad categories to human habitation, crops, water, woodland, scrub and grassland.
- Habitat – estimated percentage cover and height for canopy, shrub and ground layers, plus leaf litter and rock cover.
- The presence within the trapping grid of general botanical taxa. The main types of dominant trees, such as *Rhododendron, Quercus* or *Abies* were recorded as well as dominant shrubs and ground flora such as dwarf *Azalea, Potentilla, Primula*, labiates, nettles, course grasses and sedges.

At each site, before each trapping session, brief weather conditions were recorded for the previous period. This included the minimum temperature, rain or snow and a general description.

2.2.7 Data Handling and Analysis Methods

All field data were stored in a custom-built Microsoft Access 2016 database (Anon., 2016). This enabled site and grid data to be recorded along with trapping results. It was also used to store photographs of grid habitats and animals captured, as well as records of all tissue samples, DNA extractions, PCR results and the final DNA sequences.

Generalized Linear Mixed Models (GLMM) were carried out in SPSS v25 (Anon., 2017a) using the GENLINMIXED procedure. These used trap session as a repeated-measures variable, with diagonal covariance structure, and site as a random factor. For most models, the response variable was binary, so a logit link function was used. Due to the unbalanced design of many of these models, the Satterthwaite method was used to calculate degrees-of-freedom with robust covariance estimation. Where possible, I followed the guidance given in Bolker *et al.* (2009) in designing and interpreting these models.

Other statistical analyses were carried out in Statistica v13 (Anon., 2017b). General Discriminant Analysis (GDA) was used to explore the relationships between morphometric variables and species membership.

Non-parametric tests such as contingency χ^2 and Kruskal-Wallis one-way ANOVA by ranks were undertaken when data were highly non-normal. Finally, due to the large number of analyses carried out, to reduce the possibility of Type I errors, I set an *a priori* α level of 0.01 for all statistical testing.

2.3 Results

The results presented in this section are divided into two sub-sections. Firstly, I have analysed capture and recapture rates and how they were influenced by methodological factors such as geographical location, altitude and field methods. The full analysis is given in Supplementary Information 2.D, so a summary is presented here. Secondly, I have summarised of the numbers of animals caught during the study with their putative identifications and brief descriptions of the species. A full analysis of species distributions by topographical factors and habitat characterisation will be presented in Chapter 6.

2.3.1 Analysis of Capture Rates by Methodological Factors

The three years of fieldwork comprised 147 grid visits in 26 site visits to four transects (Table 2.2). During this period, 8046 trap events were undertaken; a trap event being defined as a single trap laid for one session. 66% of trap events took place in dawn sessions and 34% at dusk. The standard number of trap events per grid visit was 54 which equated to 324 trap events per site. There was a small amount of variation as a result of adverse weather effects (Table 2.8), or a reduced number of grids per site.

Void-trap rates were very low. In total 235 trap events (2.9%) were void, of which 69% were closed with no capture. The only significant methodological factor affecting void-trap rates was the type of session; dawn or dusk ($F_{(1, 207)} = 11.9$, $p < 0.001$). Dawn sessions had a mean predicted void rate of 3.0%, whereas for dusk sessions it was 1.5%. Although this effect was small, all subsequent analyses were based on effective trap events (ETE), calculated as total trap events minus void-traps, resulting in a total of 7811 ETE.

In total, there were 890 captures during the survey, which gave an 11.4% capture rate based on ETE. Every one of the 26 site visits achieved captures, with a minimum of three and a maximum of 114 (38% capture rate). 137 of the 147 grid visits (93.2%) achieved at last one capture. The median capture rate was 8.9%, reflecting a high degree of overdispersion in the statistical distribution of captures. Over 94% of captures were in dawn sessions and this was the only methodological factor that affected capture rates ($F_{(1, 185)} = 34.7$, $p < 0.001$). Predicted mean capture rates at dawn were 14.2% and at dusk they were 1.7%.

Overall mortality rates were 34% and a GLMM showed that three factors were significant. There were significantly higher rates on the Annapurna transect (45%) and lower in Sagarmatha (12%). There was also a significantly lower rate in dusk sessions (18%) than dawn sessions (40%). However,

the strongest effect was that mortality rates increased with altitude ($F_{(1, 19)} = 17.8$, $p < 0.001$; Figure 2.35).

A total of 792 individual animals was caught during the survey, meaning that only 98 captures (11%) were recaptures. These were used in two ways; a) to investigate movements between recaptures and b) to estimate population size using capture-mark-recapture (CMR) methods. 58 recaptures (59%) occurred within the same trap pair, so exhibited zero movement, and the remaining recaptures showed no significant difference between movements along the X and Y axes of the grid. Peterson Index values were extremely variable compared to observed numbers (minimum number alive) and their confidence intervals meant that the estimates were effectively unusable.

Minimum overnight temperatures were –6°C and the maximum was 23°C, although the modal minimum was –1°C. Minimum daytime temperatures were 8°C, rising to a maximum of 25°C in six trapping sessions. Temperature had a significant negative effect on mortality rate ($F_{(1, 48)} = 6.308$, $p < 0.02$) and precipitation caused a highly significant increase in void-trap rate ($F_{(1, 288)} = 11.04$, $p = 0.001$).

2.3.2 Species Identification and Descriptions

The 792 animals were identified to species level where possible or to higher level taxa, where identification was unclear (Table 2.5). The majority were photographed, usually showing dorsal and ventral pelages and other details when relevant; 863 images have been catalogued for identification purposes. A total of 18 different taxa was distinguished, although one was only to genus level and a further four were classified as different groups, at a family level. Three orders and four families were represented, although two families only had one species each and were very rarely recorded. The Muridae had four confirmed genera and potentially nine species, dominated by *Mus booduga* with 209 (69%) animals. There were three unidentified taxa, one *Rattus* and two unconfirmed murid types, numbering seven animals (2.3%) in total. Furthermore, three species of *Niviventer* were distinguished, although the specific identifications were uncertain, so are marked with ? in Table 2.5. There were four confirmed genera of Soricidae with potentially seven species. This was the most abundant family caught during the survey and was dominated by two species; *Episoriculus caudatus* and *Soriculus nigrescens* with 428 animals (89%) caught in total. There were only 12 (2.5%) unidentified specimens of shrews, belonging to two distinct taxa.

Table 2.5 Field Identifications of 792 animals. Latin and common names are taken primarily from Baral and Shah (2008), with confirmations from Pearch (2011) and Molur *et al.* (2005).

Order	Family	Genus	Taxon	Common Name	Animals	
Lagomorpha	Ochotonidae (Pikas)	*Ochotona*	*Ochotona tibetana*	Moupin pika	1	1
Rodentia	Cricetidae (Voles)	*Neodon*	*Neodon sikkimensis*	Sikkim vole	5	5
	Muridae (Mice & rats)		Murid sp1		3	305
			Murid sp2		2	
		Apodemus	*Apodemus gurkha*	Himalayan wood mouse	41	
		Mus	*Mus booduga*	Little Indian field mouse	209	
		Niviventer	*Niviventer fulvescens*?		26	
			Niviventer eha?	White-bellied rats	10	
			Niviventer niviventer?		5	
		Rattus	*Rattus nitidus*	Himalayan field rat	7	
			Rattus sp		2	
Soricomorpha	Soricidae (Shrews)		Soricid sp1		11	481
			Soricid sp2		1	
		Episoriculus	*Episoriculus caudatus*	Hodgson's brown-toothed shrew	240	
			Episoriculus leucops	Long-tailed brown-toothed shrew	20	
		Sorex	*Sorex bedfordiae*	Lesser striper shrew	1	
		Soriculus	*Soriculus nigrescens*	Himalayan shrew	187	
		Suncus	*Suncus murinus*	Asian house shrew	21	
Total						792

2.3.2.1 Ochotona thibetana

One juvenile specimen of Moupin pika (Figure 2.5) was captured in 2014 in Site 3 on the Pipar transect. Its tiny size (45g), extremely short hairy tail and blunt head indicated the species which was confirmed by Smith (pers. comm.). It was caught in dense *Rhododendron arboreum* woodland, with very little ground cover.

Figure 2.5 *Ochotona thibetana.*

2.3.2.2 *Neodon sikimensis*

The five specimens of the only cricetid species were all caught at the highest sites on three different transects, all in alpine grassland. Although there was some variation in pelage colour, they were generally dark brown with dark grey ventral fur. The median length of the head & body was 97mm (min; 90mm – max; 108mm) and the median weight was 24g (20g – 32g). However the characteristic measurement for this family is their short tail length (36mm; 24mm – 48mm).

Figure 2.6 *Neodon sikimensis*.

2.3.2.3 *Unidentified Murid Species*

The two unidentified murid taxa were both medium sized mice, but with quite distinct colouration. There were three specimens classified as sp. 1, which were characterised by a brown dorsal pelage, with a light-grey ventral fur, which was generally short (Figure 2.7a). The two specimens of sp. 2 had a much greyer pelage overall, which was also much thicker and luxuriant (Figure 2.7b). They also had bulkier bodies, giving the impression of a distinct difference from sp. 1 in the hand.

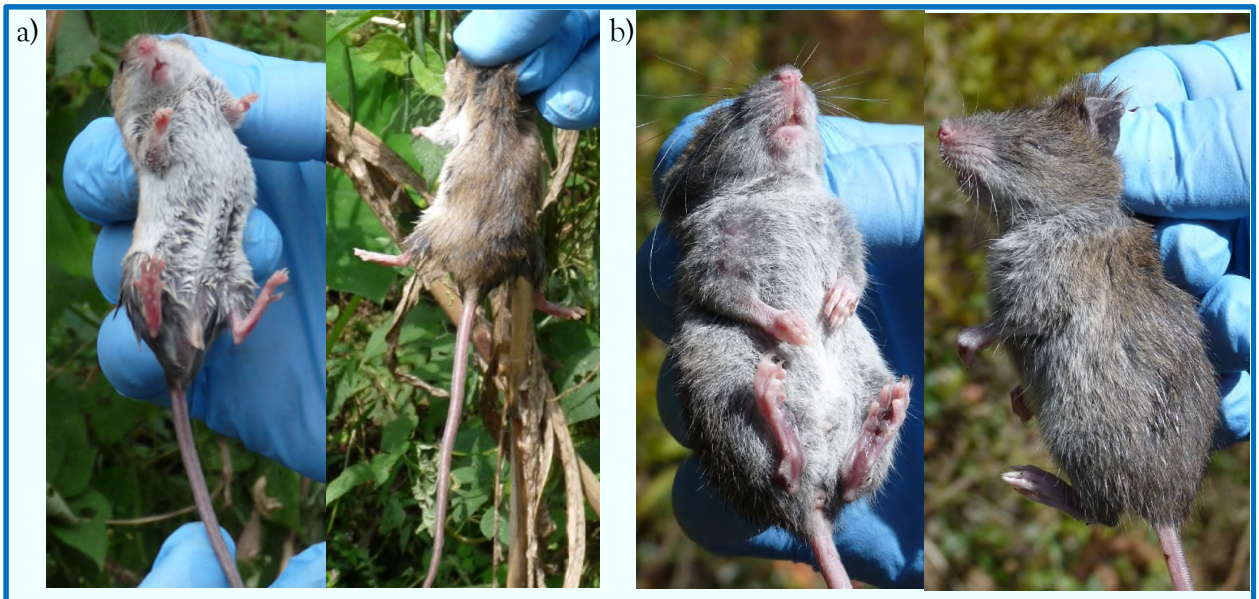


Figure 2.7 Unidentified murid species. a) sp. 1 – light brown with short fur and b) sp. 2 – grey-brown with denser fur and bulkier body.

2.3.2.4 *Apodemus gurkha*

41 specimens were identified as *A. gurkha*, although their general colouration varied considerably. They were medium-sized mice, with a grey or grey-brown dorsal pelage, sometimes quite dark, and a pale grey ventral pelage, with no clear dividing line between them (Figure 2.8). Both dorsal and ventral fur was usually very long and luxuriant, although some specimens had shorter hair, rather like murid sp. 1. These mice had longer tails and larger hind feet than the previous species.

Figure 2.8 *Apodemus gurkha*, including one specimen that died in the trap.

2.3.2.5 *Mus booduga*

This tiny mouse had a mean weight of approximately 12g and was the most abundant rodent caught during the survey. The pelage was shorter and sandy-brown on the dorsal surface with a pale yellow-grey belly (Figure 2.9). The hind feet were smaller than *A. gurkha* and the tail shorter, about the same length as the head and body.

Figure 2.9 *Mus booduga*.

2.3.2.6 *Niviventer spp.*

These small-sized rats all had similar morphological features, such as a very long tail, large hind feet and ears, and a clear demarcation between the dorsal and ventral pelage. The three putative species differed in their general colouration. The taxon identified as *N. fulvescens* (Figure 2.10a) had snowy-white ventral fur, a rufous dorsal pelage and often had a rufous spot in the centre of the chest. In contrast, the putative *N. eha* (Figure 2.10b) had a smoky-grey ventral hair, with dark shafts, but still had the rufous dorsal pelage. The third putative species was identified as *N. niviventer* and had a snowy-white ventral pelage, but had metallic grey dorsal fur (Figure 2.10c).

Figure 2.10 Niviventer spp. a) *N. fulvescens*? with rufous and white colouration, showing the clear chest spot, b) *N. eha*? with rufous and smoky-grey, with no chest spot, and c) *N. niviventer*? with grey and white colouration.

2.3.2.7 *Rattus nitidus and unidentified Species*

A small number of large rats were captured weighing, on average, c. 100g. Seven individuals were identified as *Rattus nitidus* as they had a course pelage with distinctive colouration (Figure 2.11a). The dorsal fur was dense and had dark hair shafts with light yellow-brown tips, giving a grizzled appearance. The ventral fur also had grey shafts, but the tips were almost white. There was no clear demarcation between the two zones (as with *Niviventer*), and the chin, upper chest and axial regions often had a yellow ochre or buff tinge. In contrast, two specimens were distinguished by having a much greyer overall colouration, with no hint of the ochre chin and chest (Figure 2.11b).

Figure 2.11 *Rattus* spp. a) *R. nitidus* with overall brown and yellow colouration and b) sp. with greyer dorsal pelage and no hint of ochre chest.

2.3.2.8 Exploratory Analysis of Murid Morphological Data

Morphometric data for these putative species comprised four linear measurements and weight (see Section 2.2.2 for full details). These variables were recorded for almost every specimen, although in a few cases of handling difficulties they had to be released before measurements could be taken. A general discriminant analysis (GDA) using all five variables was run to predict species membership. As this approach required case-wise missing data, only 252 specimens (83%) could be included. Overall 95% of specimens were correctly classified (Table 2.6), in particular, all 182 *M. booduga* specimens. In contrast, the putative *Niviventer fulvescens* had only 13 of the 20 individuals (65%) correctly classified. There was a degree of cross-classification of the three *Niviventer* spp. although only 2 specimens out of 30 were classified as other genera.

Table 2.6 Classification matrix from GDA of nine putative murid species using five morphometric variables (n = 252). The correct classifications are given in the green diagonal cells and misclassified individuals are shown in pink cells.

Species	Correct (%)	Murid sp 1	Murid sp 2	*A. gurkha*	*M. booduga*	*N. fulvescens?*	*N.* eha?	*N. niviventer?*	*R. nitidus*	*Rattus* sp.
Murid sp 1	0.00	0			1					
Murid sp 2	100.00		2							
A. gurkha	93.75			30		1	1			
M. booduga	100.00				182					
N. fulvescens?	65.00			1		13	4	1		1
N. eha?	88.89					1	8			
N. niviventer?	100.00							1		
R. nitidus	75.00								3	1
Rattus sp.	100.00									2

The clearest distinction uncovered by the GDA can be seen in the relationship between tail length and hind foot length (Figure 2.12). *M. booduga* is clearly defined by small feet and short tail. *A. gurkha* forms a distinct cluster, although there is slight overlap with the *N. fulvescens*. Finally, the two *Rattus* species are clearly the largest, although there is also some overlap with *N. fulvescens*. The misclassified specimen of Murid sp. 1, shown in Table 2.6, clearly falls into the *M. booduga* cluster and the two specimens of Murid sp. 2 do seem to have short tails relative to other metrics. The reference line in the figure represents a TL:HFL ratio of 5:1, which shows that the *Niviventer* species generally have longer tails than this ratio, whereas *M. booduga* has a ratio nearer to 4:1.

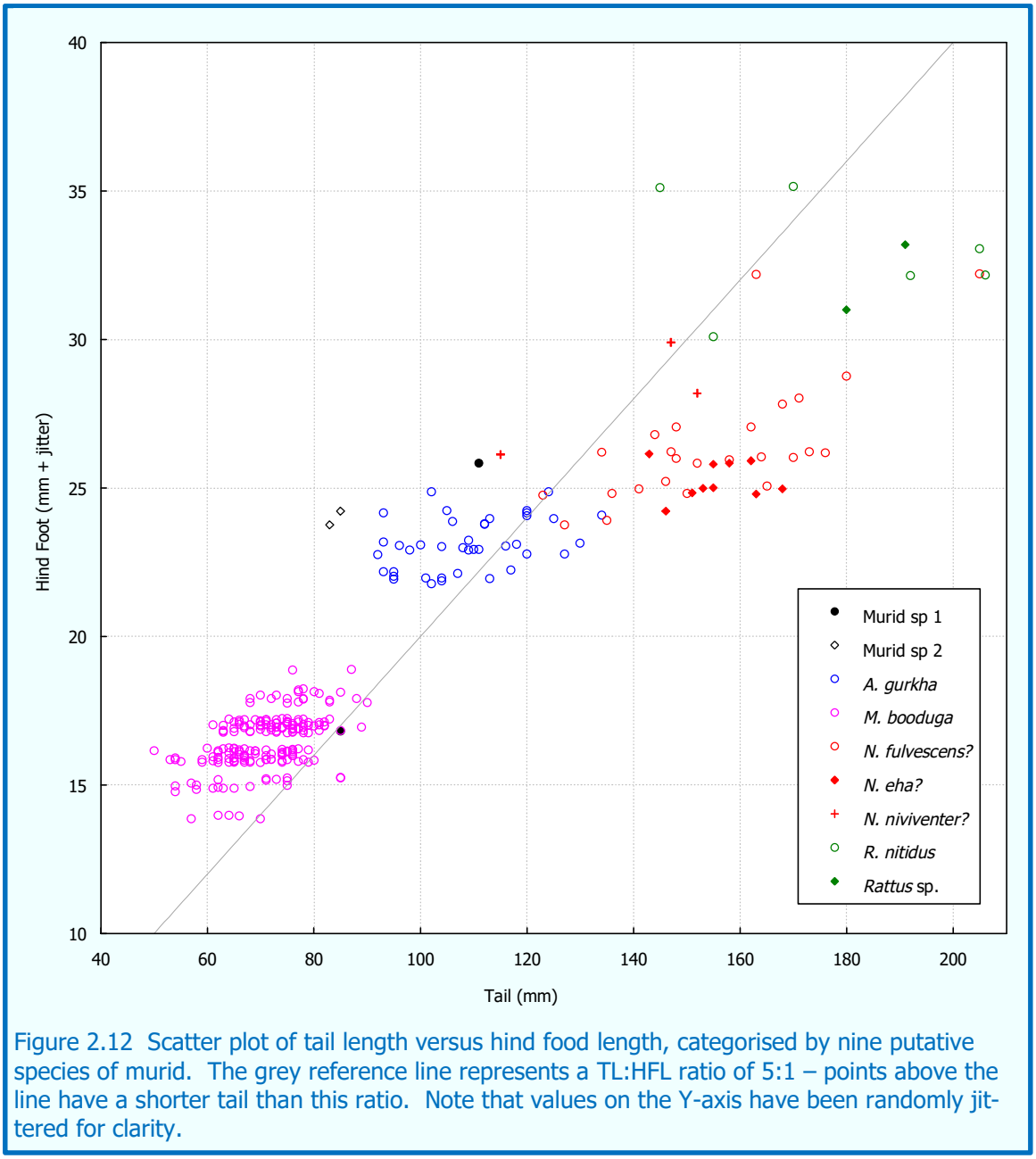


Figure 2.12 Scatter plot of tail length versus hind food length, categorised by nine putative species of murid. The grey reference line represents a TL:HFL ratio of 5:1 – points above the line have a shorter tail than this ratio. Note that values on the Y-axis have been randomly jittered for clarity.

2.3.2.9 Unidentified Soricid Species

Two different unidentified soricid types were distinguished; eleven specimens of one type and a single specimen of the other. They were all very small; approximately 60mm HBL with approx. 52mm TL. All weights were between 4g and 5g, with one exception that weighed only 3.4g. The main distinguishing features of the first group were the very dense and pale pelage, with mid-brown dorsal fur and creamy-white ventral fur (Figure 2.13). These specimens also had a short, furry tail, with a distinct difference between the dorsal and ventral colouration and, in some specimens, fore feet with silvery-white hairs on the upper surface. One specimen had an very faint dorsal stripe, but this was possibly just an effect of the extremely dense pelage. The single specimen of sp. 2 was similar in size and coloration, but had a completely naked tail.

Figure 2.13 Soricid sp. 1, specimen 14_1_6_1_01, (found dead in the trap) showing light coloured dense fur and hairy tail.

2.3.2.10 Episoriculus caudatus

The species was the most frequently caught in the whole survey. It was distinguished by its overall dark-brown pelage, especially the ventral fur (Figure 2.14), although some specimens were lighter in colour. This species had a mean weight of 5.1g, and a tail that was approximately the same length as the head and body (tail 55.8mm & HB 59.5mm). The tail appeared longer and thinner than the unknown species described above and was usually naked.

Figure 2.14 Two specimens of *Episoriculus caudatus*.

2.3.2.11 *Episoriculus leucops*

Twenty specimens of this species were caught. In general appearance and size they were very similar to *E. caudatus*, but they were distinguished by their extremely long tails (Figure 2.15) which averaged 85mm, and larger hind feet. Although they showed a strong arboreal habit (ascending straight up a *R. arboreum* trunk on release), the tail did not appear to be prehensile.

Figure 2.15 *Episoriculus leucops*.

2.3.2.12 *Sorex bedfordiae*

Only one specimen of this species was recorded. It was similar to *E. caudatus* in coloration, but had a distinct dorsal stripe, running from the back of the head almost to the rump (Figure 2.16). Its size fell into the general weight and tail length ranges for *E. caudatus*, but was towards the upper extreme of HB length. It had a naked tail and small feet and claws.

Figure 2.16 *Sorex bedfordiae*.

2.3.2.13 *Soriculus nigrescens*

This was the second-most frequently caught shrew with a highly distinctive appearance (Figure 2.17). It was a medium-sized shrew with a mean weight of 14g and a HB length of 84mm and a relatively short tail, averaging 45mm. The pelage was generally very dense, dark chocolate brown on both dorsal and ventral surfaces, although some specimens had a more slate-grey colouration. They had relatively small hind feet, but extremely well developed fore feet with very long claws.

Figure 2.17 Two specimens of *Soriculus nigrescens*.

2.3.2.14 *Suncus murinus*

21 specimens of the Asian house shrew (musk shrew) were caught during the survey. They were highly distinctive having a very large body size (mean weight 35.6g and HB of 120mm) and a relatively short but thick and strong tail (mean length 67mm) which had long, sparse guard hairs. They also had large, almost naked ears that were exposed from the pelage. In the hand, they were extremely aggressive and difficult to handle, often biting and exuding the musk that their generic vernacular name indicates.

Figure 2.18 *Suncus murinus*.

2.3.2.15 Exploratory Analysis of Soricid Morphometric Data

A similar general discriminant analysis on morphometric data was carried out for shrews, although due to the large number of missing values, ear length was not included and the two taxa with single records (soricid sp. 1 and *S. bedfordiae*) were also excluded. The classification matrix for the remaining five taxa was unequivocal; the four identified species were all correctly classified (Table 2.7), but the 11 unidentified specimens were all classified as *E. caudatus*.

Table 2.7 Classification matrix from GDA of five putative soricid species using four morphometric variables (n = 430). The correct classifications are given in the green diagonal cells and misclassified individuals are shown in pink cells.

	Correct (%)	Soricid sp. 1	*E. caudatus*	*E. leucops*	*S. nigrescens*	*S. murinus*
Soricid sp.1	0.00	0	11	0	0	0
Episoriculus caudatus	100.00	0	226	0	0	0
Episoriculus leucops	100.00	0	0	18	0	0
Soriculus nigrescens	100.00	0	0	0	158	0
Suncus murinus	100.00	0	0	0	0	17

The reason for this clear result can be seen by plotting, for example, head and body length against tail length (Figure 2.19). The four major species formed quite distinct clusters within this two-dimensional space, although there were two outlying specimens of *S. nigrescens*. HB length almost perfectly discriminated between *E. caudatus*, *S. nigrescens* and *S. murinus*, and the tail length perfectly and very strongly distinguished the two *Episoriculus* species. However, the single specimen of *S. bedfordiae* fell on the edge of the ellipse defining *E. caudatus* and the single specimen of soricid sp. 2 fell in the centre of this zone. The eleven specimens of sp. 1 had a slightly ambiguous distribution in this two-dimensional space, with the majority falling into the *E. caudatus* ellipse, but three of them, including the specimen shown in Figure 2.13, appeared to have distinctly shorter tails. The three reference lines in the figure show a clear distinction in the ratios between tail length and head and body length, where *E. leucops* had a ratio of 1.5:1, *E. caudatus* had a 1:1 ratio and the *S. nigrescens* ratio was nearly 0.5:1.

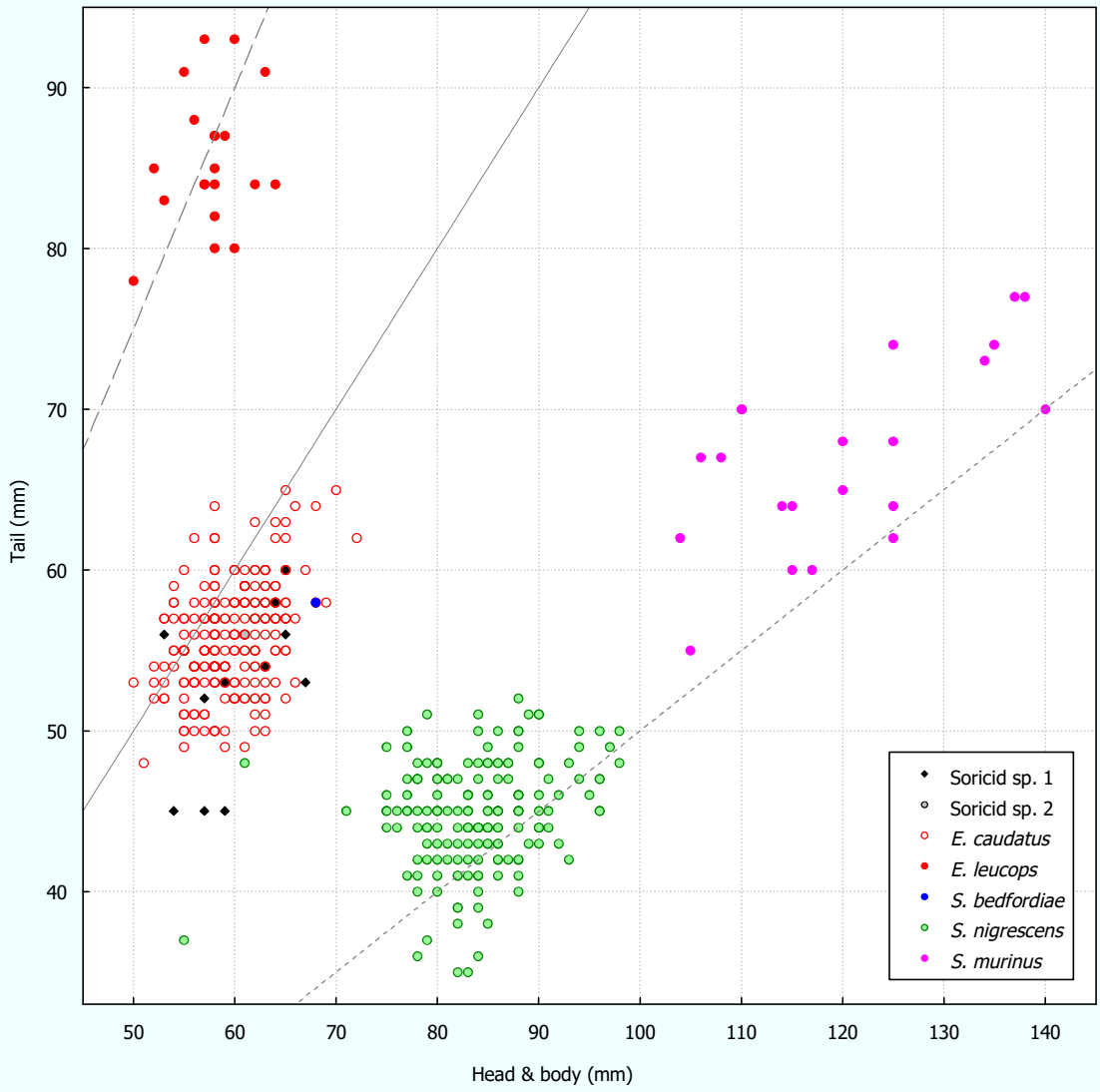


Figure 2.19 Scatter plot of head & body versus tail lengths, categorised by seven putative species of soricid. The three reference lines represent tail:HB ratios of the solid line; 1:1, dashed line; 1.5:1 and dotted line; 0.5:1.

2.4 Discussion

2.4.1 The Scope of this Study and Small Mammal Capture Rates

The fieldwork for this study took place over three years (2013 – 2015) in four altitudinal transects located in the central and eastern parts of Nepal. Twenty-six sites were established at altitudes ranging from 1300m to 4400m and, within these, 147 visits to 126 trapping grids were made. The survey comprised a total of 8046 trap events, during which 890 captures of 792 small mammals were achieved. These were identified in the field as 18 putative species (including five unidentified taxa), which will be analysed in detail in Chapter 6.

The geographical and temporal scope of this study makes it the largest ecological small mammal survey ever undertaken in the Himalayan region of Nepal, requiring 720 staff/days in the field. Furthermore, the systematic survey design, based on altitudinal transects, with regularly placed sites and using a standard grid layout makes this study unique. Previous studies of small mammals in Nepal have not used such a rigorous and systematic sampling strategy. In his seminal study in the 1970s, Abe (1971) visited 22 sites in two regions and collected 196 animals belonging to 14 species (Abe, 1982). The first of these included ten sites in the Annapurna region including the town of Pokhara, so was within 20km of my two transects. The altitudinal range of the sites was 800m to 2700m, although there was no attempt to select sites on a continuous altitudinal gradient. The second region had 12 sites, which ranged from the terai at 300m altitude to Gosain Khund at 4300m and included the city of Kathmandu and its suburbs. Three of his sites were almost exactly on my Langtang transect and he must have followed virtually the same trekking route. He used approximately 20 to 50 traps, for between one and seven nights at each site. As part of a large, country-wide study of mammals and ectoparasites over four years, Mitchell and Punzo (1976) collected 3695 records of mammals including three soricids new to Nepal. Mitchell (1977) described the use of snap-traps for this study, but did not list numbers or methods of deployment. In 1969 and 1970 Martens and Niethammer (1972) collected 44 specimens of *A. gurkha* and 25 of *A. wardi* (*A. pallipes*) from the Kali Gandaki and Phoksumdo regions west of Pokhara and Niethammer and Martens (1975) report several species of *Rattus* and *Niviventer* from the central and eastern regions. These were primarily taxonomic studies, so they give no details of field methods. During the 1973 Czechoslovak mountaineering expedition to the Makulu-Barun region, east of Sagarmatha, Daniel (2015) collected 139 small mammals from 23 sites ranging in altitude from 3450m to 5950m. Again, no detailed description is given of survey design or field methods. In 1978 the UEA Nepal Expedition (Ingles *et al.*, 1980; Newton *et al.*, 1990; Rands *et al.*, 1979), working largely in the terai, chose their trapping sites subjectively. They collected 73 specimens of small mammals from 11 species. Finally, there are

several DNA sequences from murids and soricids lodged on Genbank, which come from specimens collected in Nepal (see Chapter 5 for a full analysis of these data). Most of these have no published reference, but the metadata indicates that they were often collected as part of wider, multi-national collaborations.

One crucial aspect of this study which sets it apart from previous surveys in Nepal is the use of live-catch traps. I can find no records of systematic live trapping in Nepal prior to the start of this study. All the studies reported above either explicitly mention the use of break-back traps (Abe, 1971; Mitchell, 1977) or imply that some form of killing trap was used to obtain specimens. Many other recent studies have involved the use of snap-traps or other methods of specimen collection, even when the primary purpose of the study has been ecological (Krystufek *et al.*, 2010; Li *et al.*, 2003; Rajabi-Maham *et al.*, 2012; Rowe *et al.*, 2015). Break-back traps have the advantage that they are generally smaller, lighter and cheaper, and so can be deployed in greater numbers in remote terrain compared to live traps. They also do not necessitate frequent visits to release captured animals. However, few modern ecological studies would encourage their use from an ethical or welfare point-of-view (Baker *et al.*, 2012; Powell and Proulx, 2003). Live-catch traps have been recommended for long-term monitoring programmes of small mammals (Flowerdew *et al.*, 2004; Sibbald *et al.*, 2006) and it is axiomatic that live trapping is necessary to undertake Capture-Mark-Recapture studies (Toms *et al.*, 1999), excepting non-invasive techniques.

Another aspect of using highly sensitive live-catch traps has become apparent from the ratios of small shrews to larger murids and shrews caught in this study, compared to those that used break-back traps. I collated the numbers of each species categorised into seven weight categories, from the current study and three other sources; Abe (1971), Daniel (2015) and the Bombay Natural History

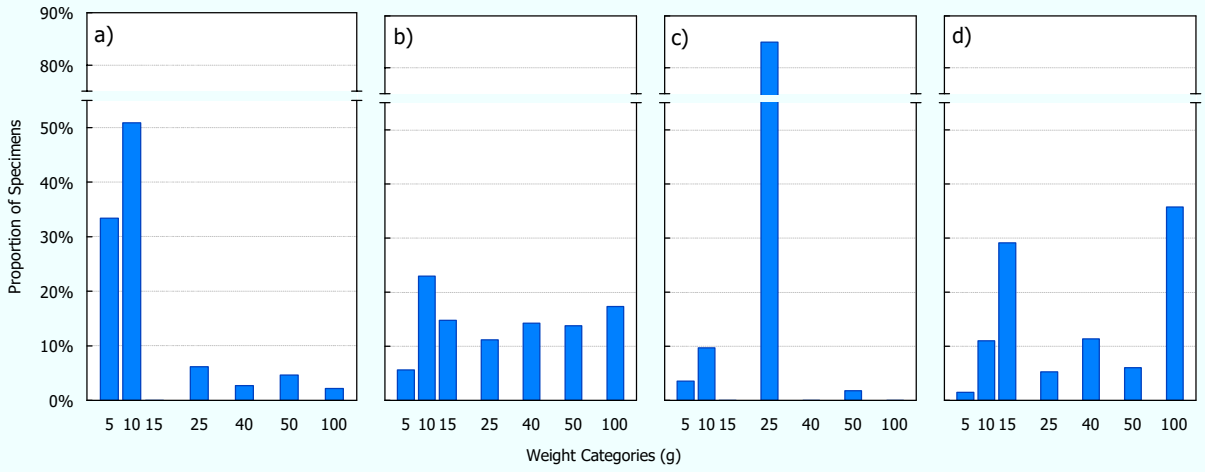


Figure 2.20 Weight profiles using approximate weight categories for specimens from a) the current study, b) Abe (1971), c) Daniel (2005) and d) the BNHS collection. Note the scale-break on the Y-axes.

Society (BNHS) collection The percentage of specimens falling into weight categories <15g. was 84% for the current study, compared to 29%, 13% and 12% respectively (Figure 2.20). Although this is only circumstantial evidence, it strongly suggests that live trapping with sensitive traps would reveal different proportions of small mammal species present within a community compared to break-back traps. I am confident that the current study has exposed a fauna of micro-mammals that has been virtually invisible to previous investigators.

Mortality rates were unexpectedly high, although not as high (76% for shrews and 52% for mice) as recorded by Edwards and Jones (2014) using pitfall traps in the USA. The analyses in Supplementary Information 2.D showed that there were significantly, higher mortality rates in dawn sessions, which lasted approx. 16 hours, compared to the eight hour dusk sessions. There was also a significant positive relationship with altitude and a negative relationship with overnight temperatures, which were confounded effects. These factors all point towards exposure as the functional cause. Furthermore, 82% of fatalities were shrews, which are known to be especially susceptible to starvation due to their high metabolic requirements (Churchfield *et al.*, 2012; Ochocinska and Taylor, 2005; Rychlik *et al.*, 2012). In the UK and other temperate countries, it is recommended that live invertebrate food, such *Calliphora* pupae (readily available as fisherman's casters) be provided for shrews (Gurnell and Flowerdew, 2006; Sibbald *et al.*, 2006). Unfortunately, these items were unavailable in Nepal and, despite testing a range of other food types such as rehydrated dried fish, chicken skin, dried mealworms and fatballs brought from the UK, none were effective in reducing shrew mortality. This should be a subject for future research to facilitate future small mammal trapping studies in Nepal.

The basic recapture rate (animals / animals that were recaptured at least once) obtained in this study (11%) was considerably lower than expected. In a nationwide study in the UK of yellow-necked mice *Apodemus flavicollis*, the equivalent mean recapture rate on the second morning was 30% (Marsh *et al.*, 2001). During a pilot study for a UK small mammal monitoring scheme, Poulton and Stone (2008) achieved a recapture rate of 42%, although their field methodology included an additional dusk trapping session. It is not clear why the recapture rate was so low during this study as the mean number of animals caught per grid was 2.1, With 18 traps available per grid, trap saturation was not the explanation, but the high mortality rate would obviously have had an effect.

This relatively low recapture rate mitigated against the last two objectives of this survey. The Petersen Index estimates of abundance were unusable due to high confidence limits, so all subsequent analyses have utilised minimum number alive (MNA) as their response variable. CMR studies would

require a combination of increased survival rates and additional trapping sessions to be effective. It should be noted though, that the pilot study and Site 4 in Pipar in 2013, both of which trapped for four nights, did not substantially increase recapture rates. Finally, no systematic movement on re-capture could be determined.

2.4.2 Species Identifications

Field identifications were based on the field handbook that I compiled from museum specimens (Supplementary Information 2.C) and published identification guides such as Baral and Shah (2008) and Smith and Xie (2008). The majority of animals were photographed, usually showing dorsal and ventral pelages and other details when relevant; 863 images have been catalogued for identification purposes. In Chapter 1, I extracted the small mammal fauna likely to be encountered in my study areas from a number of references, especially Molur *et al.* (2005), Jnawali *et al.* (2011) and Pearch (2011), which formed the underlying basis for the identifications. The 13 named species in Table 2.5 totalled 773 specimens, which comprised 97.7% of all specimens recorded. In general, the identifications made in this study concur with the previously cited authors.

Neodon sikimensis was the only cricetid species that I captured, with five specimens. Both Abe (1971) (as *Pitymys)* and Daniel (2015) recorded this species. Indeed it was by far the most common species recorded by Daniel (98 animals) in the altitudinal range 3450m to 5950m, but it was not recorded by Rands *et al.* (1979).

Apodemus gurkha is the only endemic non-volant species of mammal in Nepal (Molur *et al.*, 2005) of which I identified 41 specimens. The BNHS collection has only four specimens of this species and I found two specimens in the Zoological Survey Museum, Kolkata. Abe recorded eight specimens and it was not recorded by Daniel or Rands. Despite the fact that Pearch (2011) identified 129 other museum specimens (held in Canada, Germany, Hungary and USA) it would appear that my records form a substantial contribution to known localities of this species. This is discussed further in Chapter 6.

Mus booduga was a difficult species to identify, owing to its similarity to two congeners, *M. cervicolor* and *M. musculus*, known to exist in this part of Nepal. However, the maximum recorded altitudes presented in Table 2.1 indicated that the former was not found above 1000m and the latter, 3000m, compared to *M. booduga* with 4000m maximum elevation. Small size is the main distinguishing characteristic of this species and the first of these animals that I caught during the pilot study, and in Pipar in 2013, were all very small. One adult male (with scrotal testes) weighed 10.5g and I caught

a perforate female that weighed 10g. Ten juveniles weighed less than 6g and the median weight of the 209 animals caught was 11g. Abe recorded 29 *M. cervicolor* and no other *Mus* species throughout his study However, he noted that animals from lower altitude or forest habitats were much smaller than expected and approached the size of *M. booduga*, and he conceded that the relationship between the two species was open to question. Daniel recorded no *Mus* at all. Rands recorded two *M. booduga* (in the terai at 200m altitude), two *M. musculus* in the Kathmandu valley and 24 *M. cervicolor*, mostly below 1500m. In the light of these published records, my identifications may need revision.

My identification of the three *Niviventer* species was based on size and colouration. The rufous dorsal fur and pure white ventral fur of *N. fulvescens*, combined with its extremely long tail, made this species quite characteristic. Most specimens also had a small yellow patch on the upper chest region, which became a clear diagnostic feature (Figure 2.10). The vernacular name for *N. eha* is smoke-bellied rat (which rather contradicts the generic vernacular of "snow-bellied") and grey ventral pelage made it quite easy to identify. I was less confident of the identification of five specimens of *N. niviventer*. However, one behavioural characteristic of all three species became apparent with experience, which is that they were all very docile and never attempted to bite – unlike *Apodemus*. Abe recorded both *N. fulvescens* (20) and *N. eha* (7), whilst Daniel only recorded two *N. eha*. Rands recorded one *N. fulvescens* and five *N. niviventer* at 1800m.

Two species dominated the captures of soricids with 89% of specimens (54% of all animals caught). *Episoriculus caudatus* was the most common animal caught during the survey with 240 specimens and I captured a further 187 specimens of *Soriculus nigrescens*. The BNHS collection only contained 40 and 51, respectively, of these two species. This means I collected exactly six times the number of *E. caudatus* and 3.67 times the number of *S. nigrescens* as were collected throughout the whole of the Indian sub-continent during the first half of the 20th century (see Chapter 4). Furthermore, Abe caught 9 *E. caudatus* and 45 *S. nigrescens*, Daniel caught 7 and 11 respectively and Rands only caught 3 *S. nigrescens*. I have discussed the likely reason for my extremely high capture rate of these small animals in the previous section.

The final shrew species of particular interest was *Episoriculus leucops*. Its extremely long tail and its small size made this species easy to identify; 20 specimens were recorded. However, there is one other long-tailed shrew which is recorded from Nepal; *Episoriculus macrurus*, with which there is considerable confusion over identity. Both are recorded as present in Nepal (Baral and Shah, 2008; Pearch, 2011; Smith and Xie, 2008), but there are no specimens of *E. macrurus* in the BNHS or ZSI

collections (and only three of *E. leucops* in the BNHS). Abe recorded one specimen of *E. leucops*, but neither were recorded by Daniel or Rands.

2.5 References

Abe, H. (1971) Small Mammals of Central Nepal. *Journal of the Faculty of Agriculture, Hokkaido University* **56**: 367-423.

Abe, H. (1977) Variation and taxonomy of some small mammals from central Nepal. *J. Mamm. Soc. Japan* **7**: 63-73.

Abe, H. (1982) Ecological distribution and faunal structure of small mammals in central Nepal. *Mammalia* **46**: 477-504.

Acharya, P.R., Adhikari, H., Dahal, S., Thapa, A., and Thapa, S., editors. (2010) Bats of Nepal, A field guide. SMCRF, Kathmandu.

Anon. (2016) Access 2016. Microsoft.

Anon. (2017a) SPSS Statistics v. 25. IBM Corporation. https://www.ibm.com/support/knowledgecenter/en/SSLVMB_25.0.0/

Anon. (2017b) Statistica (data analysis software system) v.13. TIBCO Software Inc. http://statistica.io.

Baker, S.E., Ellwood, S.A., Tagarielli, V.L., and Macdonald, D.W. (2012) Mechanical performance of rat, mouse and mole spring traps, and possible implications for welfare performance. *PLoS One* **7**: e39334.

Baral, H.S. and Shah, K.B. (2008) *Wild Mammals of Nepal*. Himalayan Nature, Kathmandu. pp 188.

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H., and White, J.S. (2009) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution* **24**: 127-135.

Churchfield, S., Rychlik, L., and Taylor, J.R.E. (2012) Food resources and foraging habits of the common shrew, *Sorex araneus*: does winter food shortage explain Dehnel's phenomenon? *Oikos* **121**: 1593-1602.

Corbet, G.B. and Hill, J.E. (1992) *The Mammals of the Indomalayan Region: A systematic Review*. Oxford University Press. pp 488.

Daniel, M. (2015) Expansion of small terrestrial mammals and their parasites into the Barun Valley (Makalu Mt. Region, Nepal Himalaya) linked with changes in glaciation and human activities. *Journal of Mountain Science* **12**: 14-29.

Edwards, K.E. and Jones, J.C. (2014) Trapping efficiency and associated mortality of incidentally captured small mammals during herpetofaunal surveys of temporary wetlands. *Wildlife Society Bulletin* **38**: 530-535.

Flowerdew, J.R., Shore, R.F., Poulton, S., and Sparks, T.H. (2004) Live trapping to monitor small mammals in Britain. *Mammal Review* **34**: 31-50.

Green, M.J.B. (1981) A check-list and some notes concerning the mammals of the Langtang National Park, Nepal. *Journal of the Bombay Natural History Society* **78**: 77-87.

Gurnell, J. and Flowerdew, J.R. (2006) *Live trapping small mammals. A practical guide*. 4 edition. The Mammal Scociety. pp 48.

Ingles, J., Newton, P., Rands, M., and Bowden, C. (1980) The first record of a rare murine rodent *Diomys* and further records of three shrew species from Nepal. *Bulletin of the British Museum (Natural History), Zoology Series* **39**: 205-211.

Jnawali, S.R., Baral, H.S., Lee, S., Acharya, K.P., Upadhyay, G.P., Pandey, M., Shrestha, R., Joshi, D., Laminchhane, B.R., Griffths, J., Khatiwada, A.P., Subedi, N., and Amin, R. (2011) *The Status of Nepal's Mammals: The National Red List Series*. IUCN, Kathmandu, Nepal. pp 266.

Krystufek, B., Donev, N.R., and Skok, J. (2010) Species richness and distribution of non-volant small mammals along an elevational gradient on a Mediterranean mountain. *Mammalia* **75**: 3-11.

Le-Moine, N. (2017) *Small Mammal Live-Capture Trapping Methods: A Comparative Study of Longworth Traps and BioEcoSSII Tube Traps in Woodland and Reedbed Habitats in Norfolk, UK*. BSc Thesis. UEA, Norwich.

Li, J.S., Song, Y.L., and Zeng, Z.G. (2003) Elevational gradients of small mammal diversity on the northern slopes of Mt. Qilian, China. *Global Ecology & Biogeography* **12**: 449-460.

Marsh, A.C., Poulton, S., and Harris, S. (2001) The Yellow-necked Mouse *Apodemus flavicollis* in Britain: status and analysis of factors affecting distribution. *Mammal Review* **31**: 203-227.

Martens, J. and Niethammer, J. (1972) Die Waldmäuse (*Apodemus*) Nepals. *Zeitschrift für Säugetierkunde* **37**: 144-154.

Mekada, K., Koyasu, K., Harada, M., Narita, Y., Shrestha, K.C., and Oda, S.-I. (2001) Faunal survey of small mammals in central Nepal, with reference to the distribution of the genus *Soriculus* (Insectivora, Mammalia). *Biogeography* **3**: 33-40.

Mitchell, R. and Punzo, F. (1976) New Mammal Records from Nepal. *Journal of the Bombay Natural History Society* **73**: 54-58.

Mitchell, R.M. (1977) *Accounts of Nepalese mammals and analysis of the host-ectoparasite data by computer techniques*. PhD Thesis. Iowa State University, Ames, Iowa, USA.

Molur, S., Srinivasulu, C., Srinivasulu, B., Walker, S., Nameer, P., and Ravikumar, L. (2005) *Status of South Asian Non-volant Small Mammals: Conservation Assessment and Management Plan (C.A.M.P.) Workshop Report*. Zoo Outreach Organisation / CBSG-South Asia, Coimbatore, India, . pp 618.

Murgatroyd, E. (2017) *The Comparison of Small Mammal Capture Rates between Longworth and BioEcoSS Tube Traps Used at Varying Heights*. BSc thesis. UEA, Norwich.

Newton, P., Rands, M., and Bowden, C. (1990) A collection of small mammals from eastern Nepal. *Mammalia* **54**: 239-244.

Niethammer, J. and Martens, J. (1975) Die Gattungen *Rattus* and *Maxomys* in Afghanistan und Nepal. *Zeitschrift für Säugetierkunde* **40**: 325-355.

Ochocinska, D. and Taylor, J.R. (2005) Living at the physiological limits: field and maximum metabolic rates of the common shrew (Sorex araneus). *Physiological and Biochemical Zoology* **78**: 808-818.

Pearch, M.J. (2011) A review of the biological diversity and distribution of small mammal taxa in the terrestrial ecoregions and protected areas of Nepal. *Zootaxa* **3072** 1-286.

Poulton, S. (2018) The BioEcoSS TubeTrap. BioEcoSS Ltd. http://bioecoss.azurewebsites.net/Content/Products/TubeTrap/TubeTrap.aspx

Poulton, S. and Stone, E. (2008) *Pilot Study for a National Monitoring Scheme for Small Mammals in the United Kingdom and the Republic of Ireland.* T.M. Society. pp 70.

Powell, R.A. and Proulx, G. (2003) Trapping and Marking Terrestrial Mammals for Research: Integrating Ethics, Performance Criteria, Techniques, and Common Sense. *ILAR Journal* **44**: 259-276.

Rajabi-Maham, H., Orth, A., Siahsarvie, R., Boursot, P., Darvish, J., and Bonhomme, F. (2012) The south-eastern house mouse *Mus musculus* castaneus (Rodentia: Muridae) is a polytypic subspecies. *Biological Journal of the Linnean Society* **107**: 295-306.

Rands, M., Bowden, C., Brown, P., and Newton, P. (1979) *UEA Nepal Expedition 1978/79.* UEA. pp 92.

Rowe, R.J., Heaney, L.R., and Rickart, E.A. (2015) Scale effects on the pattern and predictors of small mammal diversity along a local elevational gradient in the Great Basin. *Journal of Biogeography* **42**: 1964-1974.

Rychlik, L., Churchfield, S., Taylor, J., Kardynia, P., and Oldakowski, L. (2012) Live trapping small mammals under snow: a protocol for maximising captures and minimising mortality. *Polish Journal of Ecology* **60**: 621-628.

Seber, G.A.F. (1973) *The estimation of animal abundance and related parameters*. Griffin, London. pp 506.

Sibbald, S., Carter, P., and Poulton, S. (2006) *Proposal for a National Monitoring Scheme for Small Mammals in the United Kingdon and the Republic of Eire.* 6, T.M. Society. pp 90.

A.T. Smith and Xie, Y. *Eds.* (2008) *A Guide to the Mammals of China*. pp 544. Princeton University Press.

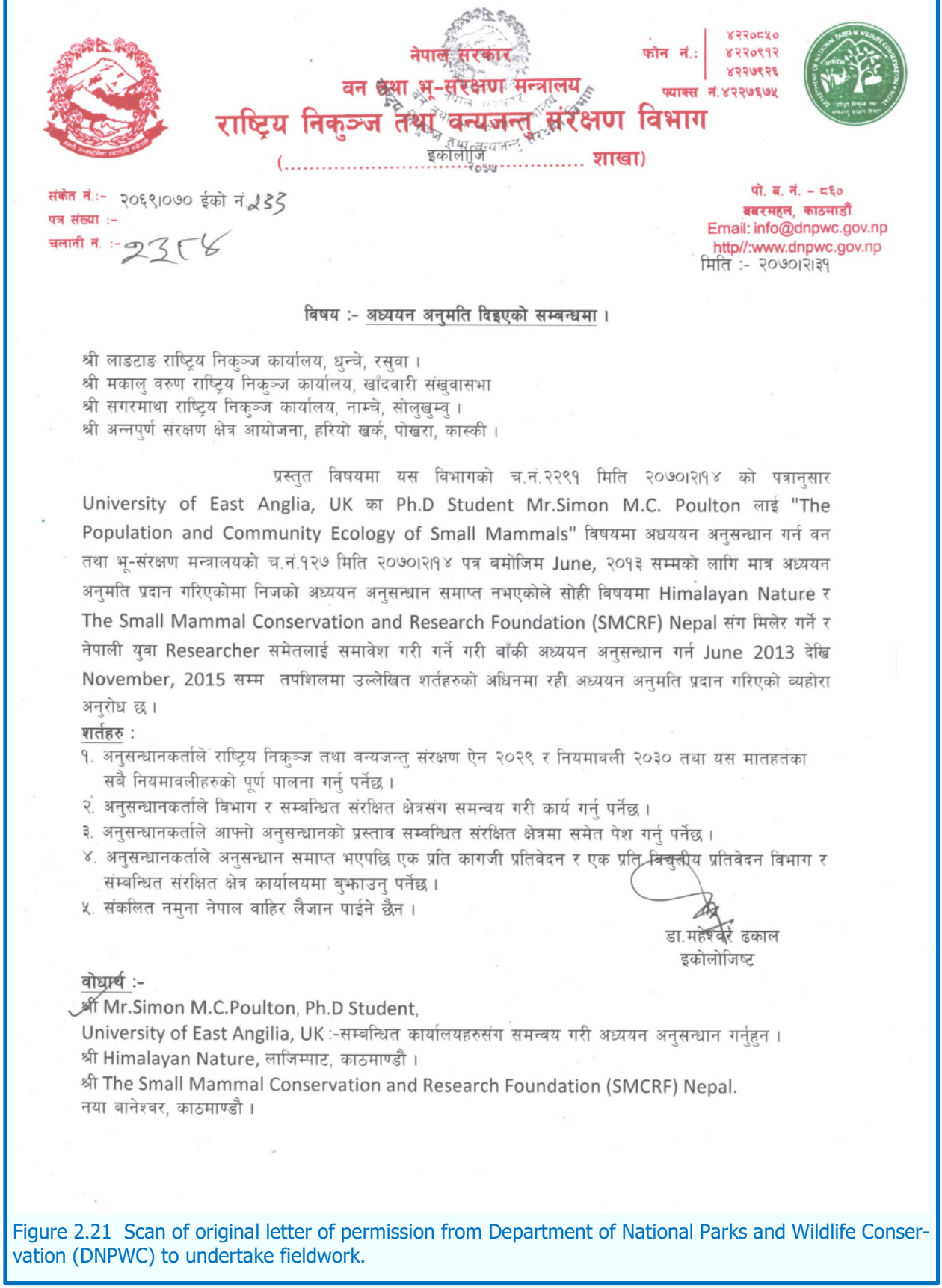
Thapa, S. (2014) A checklist of mammals of Nepal. *Journal of Threatened Taxa* **6**: 6061-6072.

Toms, M.P., Siriwardena, G.M., and Greenwood, J.J.D. (1999) *Developing a mammal monitoring program for the UK.* BTO. pp 313.

Zar, J.H. (1984) *Biostatistical Analysis*. Second edition. Prentice-Hall, New Jersey.

2.6 Appendices

2.6.1 Appendix I: Letters of Permission from DNPWC to Undertake Fieldwork



नेपाल सरकार
वन तथा भू-संरक्षण मन्त्रालय
राष्ट्रिय निकुञ्ज तथा वन्यजन्तु संरक्षण विभाग
(..... इकोलोजि ..... शाखा)

फोन नं.: ४२२०८५०, ४२२०९१२, ४२२७९२६
फ्याक्स नं.४२२७६७५

संकेत नं.:- २०६९।०७० ईको नं.२३५
पत्र संख्या :-
चलानी नं. :- २३५६

पो. ब. नं. - ८६०
बबरमहल, काठमाडौं
Email: info@dnpwc.gov.np
http//:www.dnpwc.gov.np
मिति :- २०७०।२।३१

विषय :- अध्ययन अनुमति दिइएको सम्बन्धमा ।

श्री लाङटाङ राष्ट्रिय निकुञ्ज कार्यालय, धुन्चे, रसुवा ।
श्री मकालु वरुण राष्ट्रिय निकुञ्ज कार्यालय, खाँदवारी संखुवासभा
श्री सगरमाथा राष्ट्रिय निकुञ्ज कार्यालय, नाम्चे, सोलुखुम्बु ।
श्री अन्नपुर्ण संरक्षण क्षेत्र आयोजना, हरियो खर्क, पोखरा, कास्की ।

प्रस्तुत विषयमा यस विभागको च.नं.२२९१ मिति २०७०।२।१४ को पत्रानुसार University of East Anglia, UK का Ph.D Student Mr.Simon M.C. Poulton लाई "The Population and Community Ecology of Small Mammals" विषयमा अधययन अनुसन्धान गर्न वन तथा भू-संरक्षण मन्त्रालयको च.नं.१२७ मिति २०७०।२।१४ पत्र बमोजिम June, २०१३ सम्मको लागि मात्र अध्ययन अनुमति प्रदान गरिएकोमा निजको अध्ययन अनुसन्धान समाप्त नभएकोले सोही विषयमा Himalayan Nature र The Small Mammal Conservation and Research Foundation (SMCRF) Nepal संग मिलेर गर्ने र नेपाली युवा Researcher समेतलाई समावेश गरी गर्ने गरी बाँकी अध्ययन अनुसन्धान गर्न June 2013 देखि November, 2015 सम्म तपशिलमा उल्लेखित शर्तहरुको अधिनमा रही अध्ययन अनुमति प्रदान गरिएको व्यहोरा अनुरोध छ ।

शर्तहरु :

१. अनुसन्धानकर्ताले राष्ट्रिय निकुञ्ज तथा वन्यजन्तु संरक्षण ऐन २०२९ र नियमावली २०३० तथा यस मातहतका सबै नियमावलीहरुको पूर्ण पालना गर्नु पर्नेछ ।
२. अनुसन्धानकर्ताले विभाग र सम्बन्धित संरक्षित क्षेत्रसंग समन्वय गरी कार्य गर्नु पर्नेछ ।
३. अनुसन्धानकर्ताले आफ्नो अनुसन्धानको प्रस्ताव सम्बन्धित संरक्षित क्षेत्रमा समेत पेश गर्नु पर्नेछ ।
४. अनुसन्धानकर्ताले अनुसन्धान समाप्त भएपछि एक प्रति कागजी प्रतिवेदन र एक प्रति विद्युतीय प्रतिवेदन विभाग र संम्बन्धित संरक्षित क्षेत्र कार्यालयमा बुझाउनु पर्नेछ ।
५. संकलित नमुना नेपाल बाहिर लैजान पाईने छैन ।

डा.महेश्वर ढकाल
इकोलोजिष्ट

बोधार्थ :-
श्री Mr.Simon M.C.Poulton, Ph.D Student,
University of East Angilia, UK :-सम्बन्धित कार्यालयहरुसंग समन्वय गरी अध्ययन अनुसन्धान गर्नुहुन ।
श्री Himalayan Nature, लाजिम्पाट, काठमाण्डौ ।
श्री The Small Mammal Conservation and Research Foundation (SMCRF) Nepal.
नया बानेश्वर, काठमाण्डौ ।

Figure 2.21 Scan of original letter of permission from Department of National Parks and Wildlife Conservation (DNPWC) to undertake fieldwork.

Government of Nepal
Ministry of Forest and Soil Conservation
Department of National Park and Wildlife Conservation

Registration no: 2384

Date 2070/2/31

Subject: Study Permission Approval

Langtang National Park, Dhunche, Rasuwa
Makalu Barun National Park, Khadbari, Shankhubashava
Sagarmatha National Park, Namche, Sholukhumbu
Annapurna Conservation Area, Hariyo Kharka, Pokhara, Kaski

Department of National Park and Wildlife Conservation, Nepal has given study permission to the student of University of East Anglia, UK, Mr Simon M.C. Poulton for his Ph. D study entitled "The Population and Community Ecology of Small Mammals" from June 2013 to November 2015 in collaboration with Himalayan Nature and The Small Mammals Conservation and Research Foundation (SMCRF) Nepal with the involvement of youth researcher under the following circumstances.

Conditions:

1. Researcher has to fully work under the Wildlife Protection Act 2029 and 2030.
2. Researcher has to work with collaboration with Department of National Park and Wildlife Conservation and concerned protected areas.
3. Researcher has to submit research proposal to the concerned protected area.
4. After completion of the research, researcher has to submit both hard and electronic copy/one to Department of National park and Wildlife Conservation and concerned Protected Areas
5. Collected specimen should not be taken out of Nepal

Dr. Maheshwor Dhakal
Ecologist

Cc:
Mr. Simon M.C. Poulton, Ph. D student
University of East Anglia, UK
Himalayan Nature, Lazimpat, Kathmandu
Small Mammals Conservation and Research Foundation (SMCRF) Nepal, New Baneshwor, Kathmandu

Figure 2.22 Translation of letter of permission to undertake fieldwork.

SUPPLEMENTARY INFORMATION

2.A The Pilot study

2.A.a Introduction

The pilot study was undertaken between 30th May and 11th June 2013. It was based in Dhunche, a small town approximately 120km by road north of Kathmandu. It was also the location of the headquarters of the Department of National Parks and Wildlife conservation (DNPWC) for Langtang National Park, where permission was granted for the pilot study. Three personnel were involved in the fieldwork; SMCP, Mr Sujan Maharjan who was a forest ranger for DNPWC and Mr Hem Katuwal, who was a founder member of SMCRF. In addition, two porters were hired to carry 70 traps and equipment to the second site.

The pilot study had five main objectives;

- To explore a sample of habitats, at different altitudes, that would be covered by the full survey and to investigate their suitability for small mammal surveys.
- To trial three different designs of trapping grid.
- To test the effects of pre-baiting on capture rates.
- To test the efficacy of four different types of shrew food.
- To gain experience of the logistics necessary to complete the main field survey. In particular;
- How long would it take to lay out the trapping grids, check the traps and collect them at the end of the trapping session?
- What distances and altitudes could be trekked between sites, which would indicate how many sites and transects could be completed?
- What options were available for using local porters and what would they be able to carry?
- Would trekking/pilgrim hostels be available at suitable locations, at what cost and what provisions wold be available?

2.A.b Methods

Two sites were selected for the trials;

- Terraced farmland and grazing meadows about 2km from Dhunche at 1900m altitude.

- Abies grandis and Rhododendron campanulatum woodland about 10km along the Gosain Khund trail from Dhunche at 3600m altitude.

Within each site, three different trapping grids were set out with different numbers of traps and different inter-trap distances. Two layouts used a "two-scale" design of 25 traps in five clusters. The traps in each cluster were laid out in a cross shape (2m or 5m apart), with two arms of the cross running up and down the slope and two along the contour. The five clusters were set out in a similar, but larger pattern, 10m or 25m apart. The third layout used 18 traps in a single cross shape, with two traps at each of nine locations, 10m apart).

Traps in one of the two-scale grids were pre-baited for two nights and then trapped for two-nights. The other two grids were trapped for four nights. One of four different types of food for shrews was randomly allocated to each trap, along with grain and vegetable food for rodents. Dry grass was used for bedding and the traps were laid in dense ground vegetation or shrubs in the first site, and in moss and under fallen timber in the woodland at the second site.

2.A.c Results

A total of 752 trap events across the two sites were undertaken. Six animals were captured in total. Five were caught in Site 1, which were identified as four *Mus booduga* and one *Apodemus gurkha*. Three of the *M. booduga* were recaptured within the same trapping grid. Only one animal was caught (once) in Site 2 – a large rodent, identified at the time as *Apodemus* sp. but more likely to be a *Niviventer* sp.

2.A.d Discussion

The pilot was very useful in indicating the small size of micro-habitat patches. This mitigated against the use of the two larger grids as the aim was to locate grids entirely within single habitats. The final decision to use the smaller grid of 18 traps also meant that more grids could be laid in a single site. The pilot also indicated how arduous the fieldwork was and encouraged a revision to the original survey design.

However, the capture-rates during this pilot were extremely low, 1.3% in the first site and 0.3% in the higher site. This meant that the effects of trapping grid design and pre-baiting could not be evaluated. Furthermore, the total absence of shrews meant that no assessment of different shrew foods could be made. The reason for this low trapping rate is probably the season, being at the beginning of the monsoon period. This was the start of the breeding season, so populations would have been intrinsically low, comprising mainly adult breeding animals. Vegetation was very dense

so natural food and shelter was abundant, which might have made the traps less attractive. Nevertheless, four animals were caught after the first night of trapping, which indicated that the traps were suitable for small mammals in this part of the world.

2.B Transect Maps and Elevational Profiles

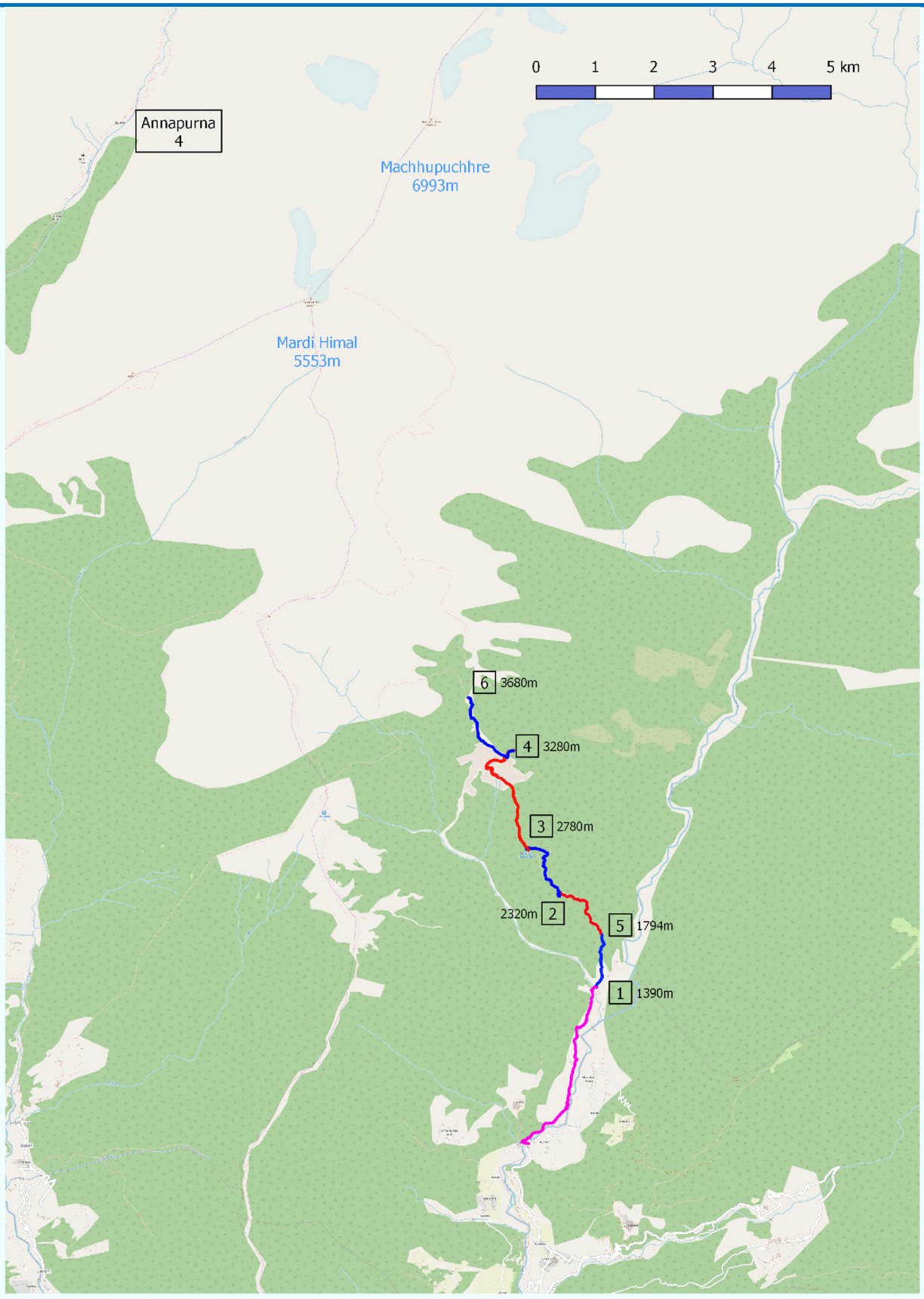


Figure 2.23 Route of Transect 1; Pipar. Numbers in squares are site codes (note out-of-sequence Site 5). Trekking routes between sites are shown in alternating blue and red, with magenta line showing route from roadhead to Site 1. Note location of Annapurna transect Site 4.

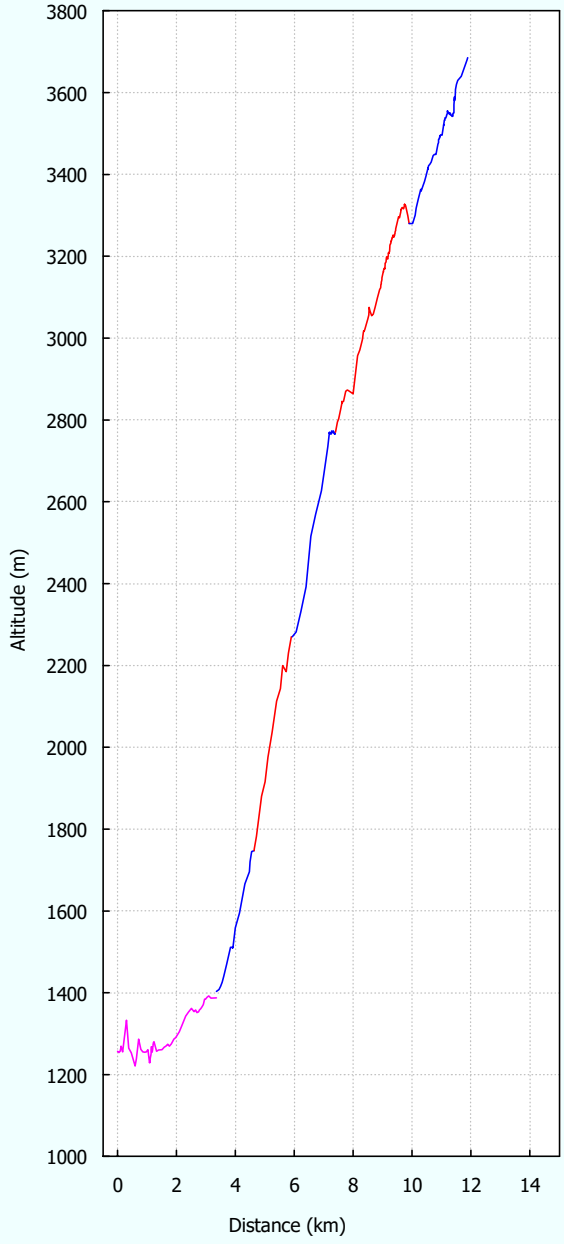


Figure 2.24 Profile for Transect 1, Pipar. Colour-coding for each route as in Figure 2.23.

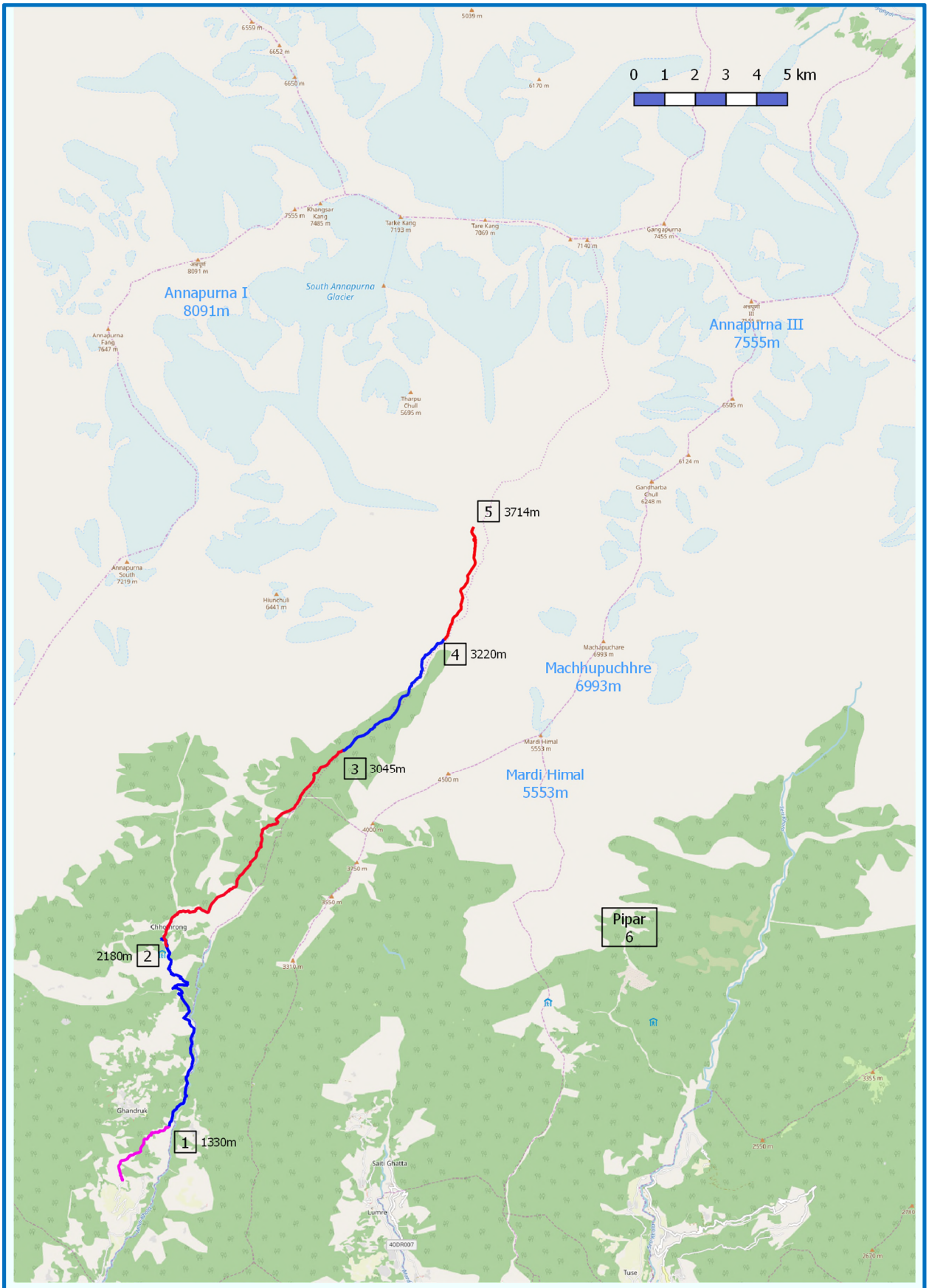


Figure 2.25 Route of Transect 2; Annapurna. Numbers in squares are site codes. Trekking routes between sites are shown in alternating blue and red, with magenta line showing route from roadhead to Site 1. Note location of Pipar transect Site 6.

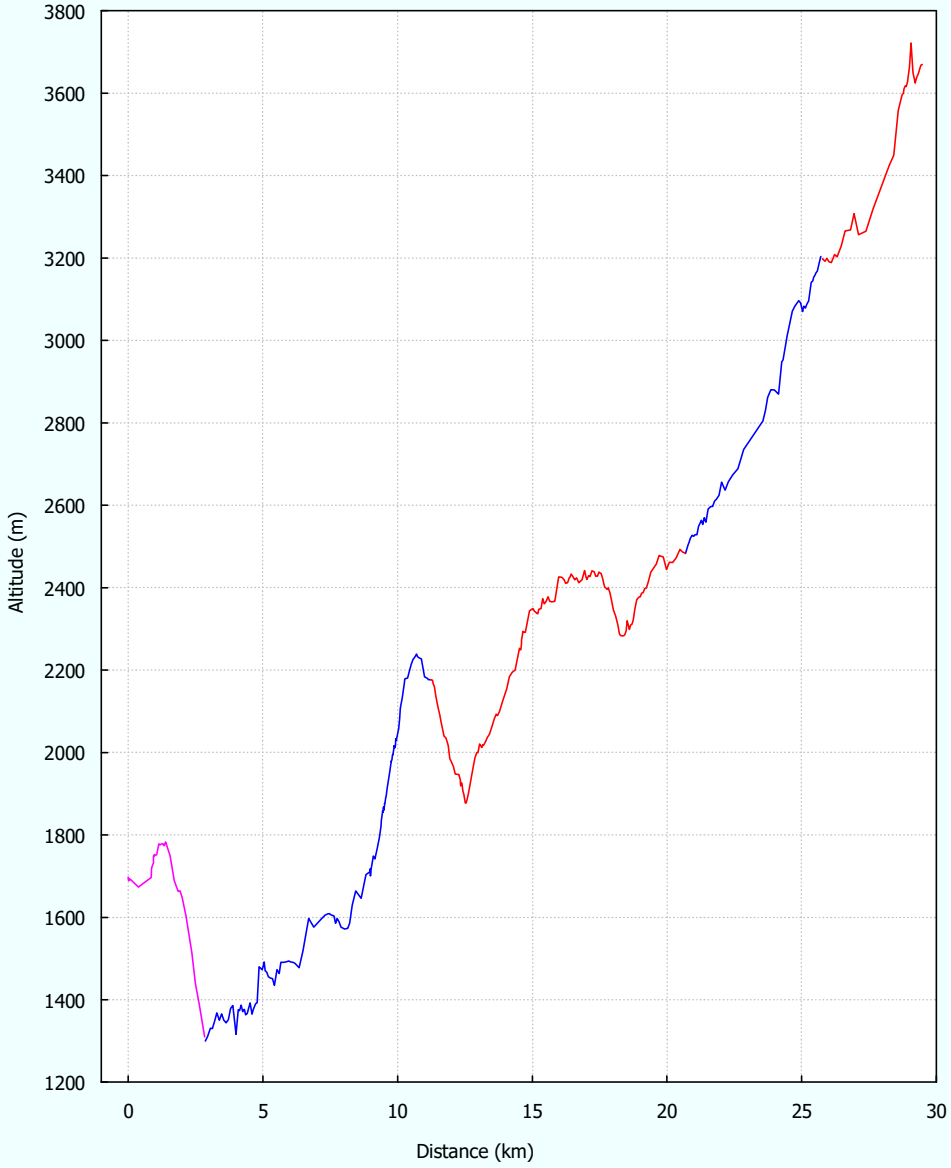


Figure 2.26 Profile for Transect 2, Annapurna. Colour-coding for each route as in Figure 2.25.

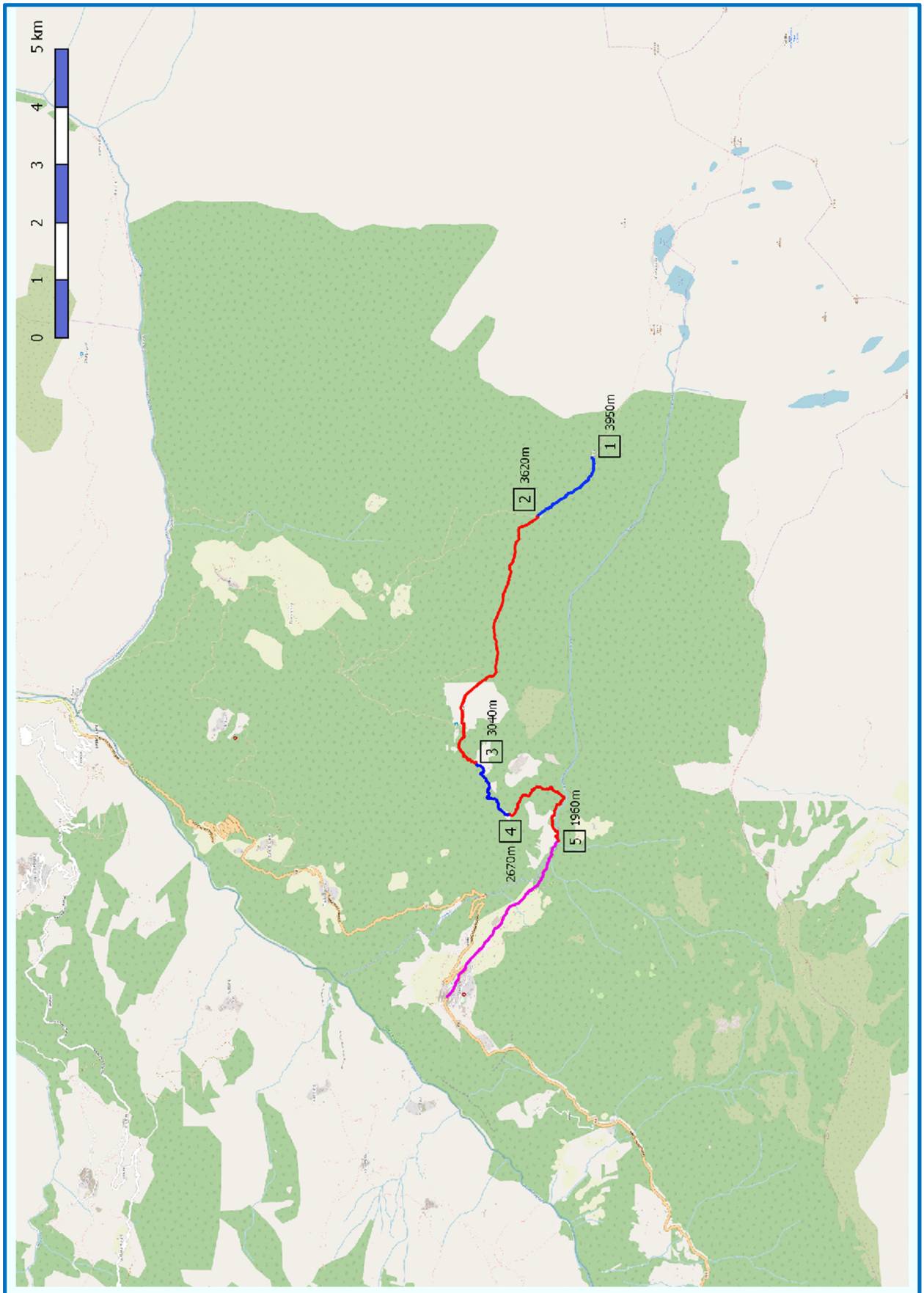


Figure 2.27 Route of Transect 3; Langtang. Numbers in squares are site codes (note sites in descending order). Trekking routes between sites are shown in alternating blue and red, with magenta line showing route from roadhead to Site 5.

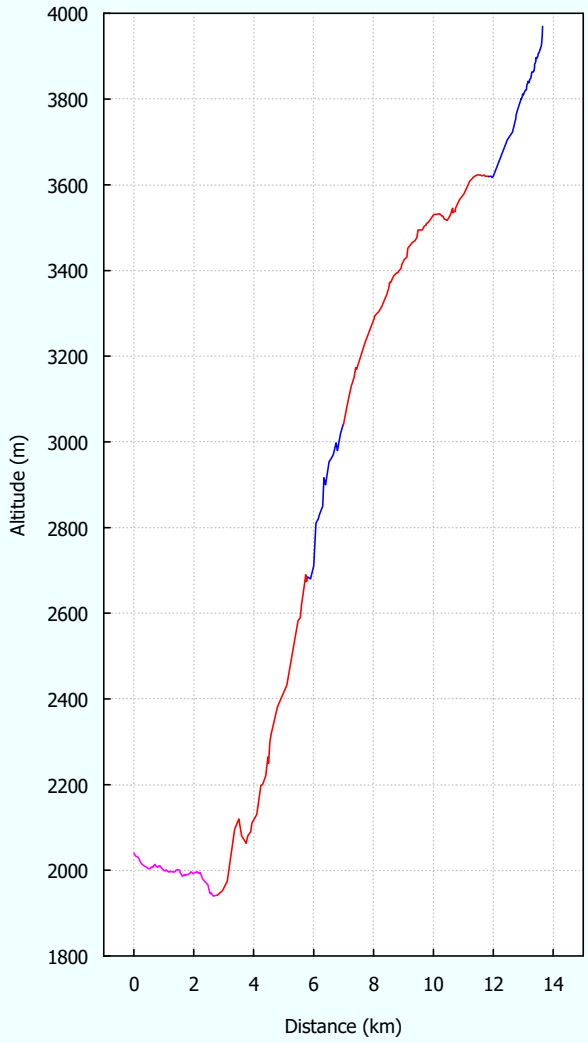


Figure 2.28 Profile for Transect 3, Langtang. Colour-coding for each route as in Figure 2.27.

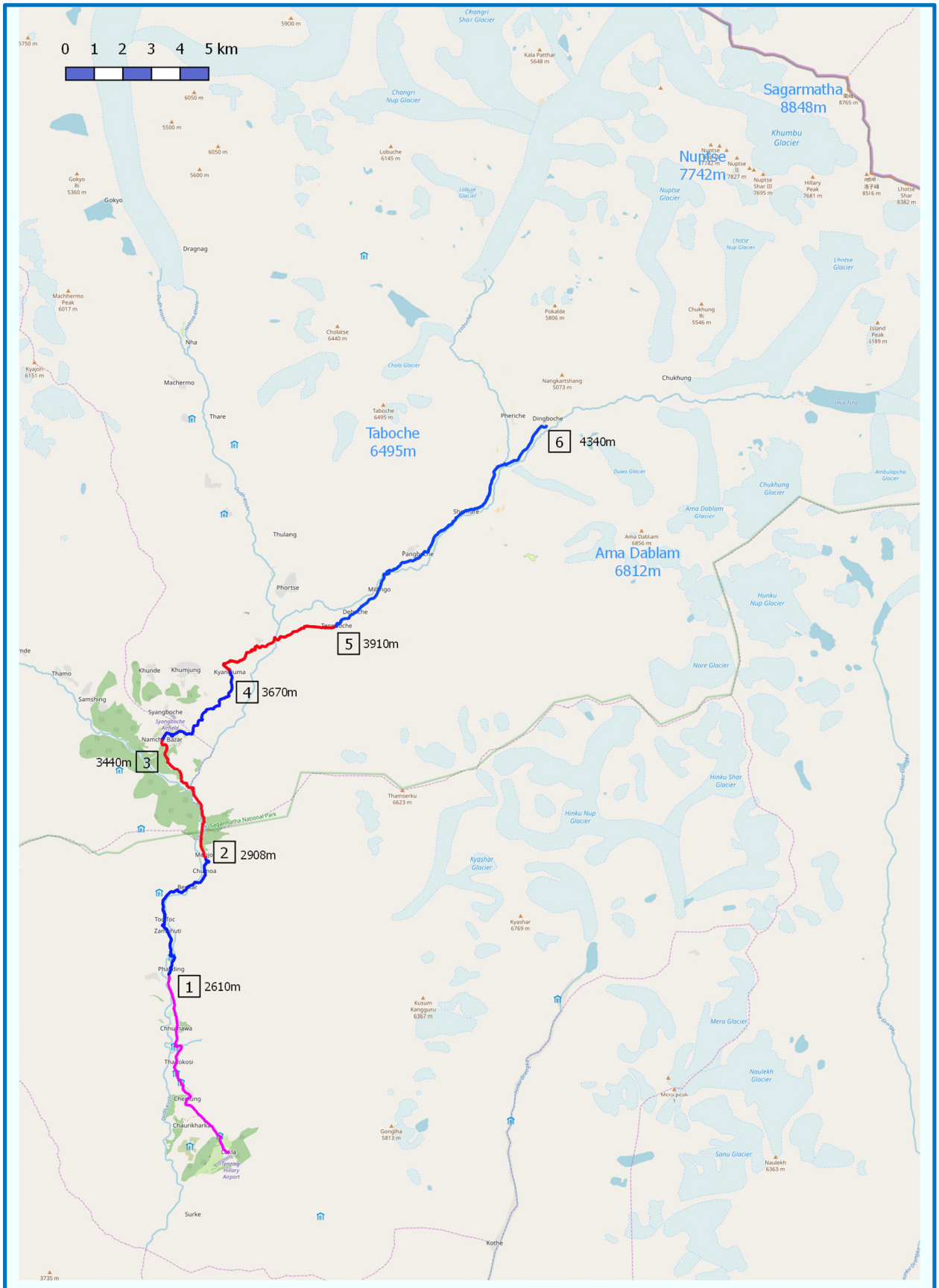


Figure 2.29 Route of Transect 4; Sagarmatha. Numbers in squares are site codes. Trekking routes between sites are shown in alternating blue and red, with magenta line showing route from roadhead to Site 1.

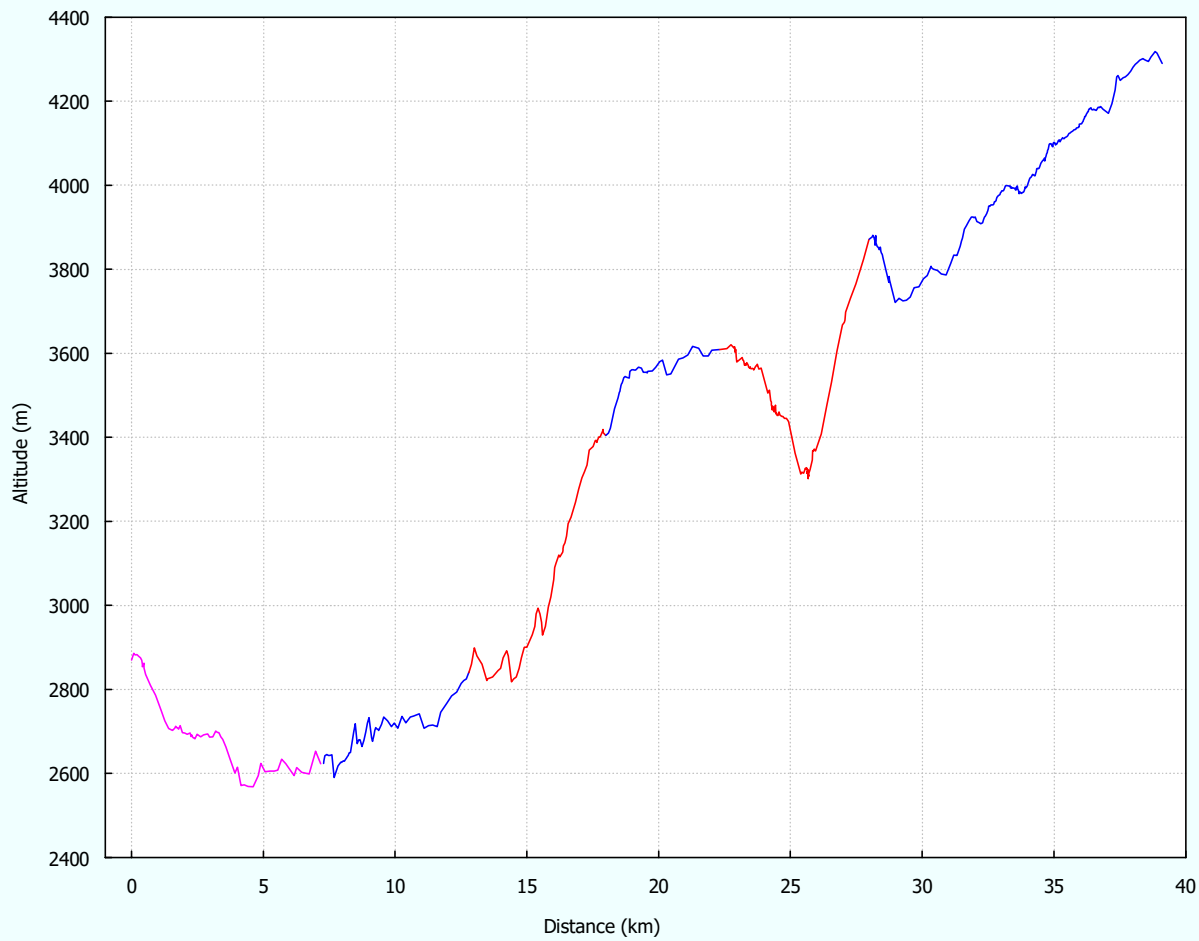


Figure 2.30 Profile for Transect 4, Sagarmatha. Colour-coding for each route as in Figure 2.29.

2.C Field Identification Handbook

Field Handbook.pdf

Figure 2.31 Embedded PDF – double-click to open.

2.D Analysis of Methodological Factors

2.D.a Trap Events

The three years of fieldwork (excluding the pilot study) comprised 147 grid visits in 26 site visits to four transects (Table 2.2). During this period, 8046 trap-events were set (Table 2.8a). Of these, 66% were dawn and 34% were dusk events. The standard number of trap events per site visit was 324, which represented 6 grids × 18 traps (108) for three sessions. This was achieved in three of the six sites in Pipar (transect 1), the first three sites in Annapurna (transect 2) and all the sites in Langtang and Sagarmatha (transects 3 & 4 respectively).

Table 2.8 The numbers of a) trap events, summarised by type (dawn or dusk) and b) void-trap events by type (see Section 2.2.2). The proportion of void-traps is given in the right-hand column.

			a) Trap Events				b) Void-Trap Events					
Year	Transect	Site	Dawn	Dusk	Total		Closed	Disturbed	Void	Damaged /Missing	Total	
2013	Pipar	1	216	108	324	1494	6	5			47	3.1%
		2	144	72	216		6	10				
		3	180	90	270		5					
		4	396	288	684		7		8			
	Annapurna	1	216	108	324	1260	12	2		1	47	3.7%
		2	216	108	324		7	1				
		3	216	108	324		4					
		4	72	0	72		4					
		5	108	108	216		16					
2014	Pipar	1	216	108	324	1728	7	4			64	3.7%
		2	144	72	216		9		1			
		3	180	90	270		6		4			
		4	216	108	324		1	1	2			
		5	216	108	324		6	4	2			
		6	180	90	270		14	1	2			
	Langtang	1	216	108	324	1620	2				15	0.9%
		2	216	108	324							
		3	216	108	324							
		4	216	108	324		2		1			
		5	216	108	324		4		2	4		
2015	Sagarmatha	1	216	108	324	1944	17		3	4	62	3.2%
		2	216	108	324		2	1				
		3	216	108	324		5	1	1	2		
		4	216	108	324		3					
		5	216	108	324		2	3	1			
		6	216	108	324		15	1	1			
		Total	5292	2754		8046	162	34	28	11	235	2.9%

2.D.b Void-trap Rates

The four types of void-trap events described in Section 2.2.2 were recorded for each site visit (Table 2.8b). Overall, void-trap rates were very low; 2.9% of trap events, with over two-thirds of these due to traps being closed with no capture. Void-trap rates were analysed as a binomial response variable with logit link function, using a series of GLMM models. As Year and Transect were strongly confounded, a combined factor with five levels, Visit, was used as the single fixed effects factor, Altitude as a fixed covariate and Session as the within-subjects (repeated-measure) fixed effect. SiteID was used as a random-effects factor using variance components for the covariance matrix. The first model included both two-way interactions of the between-subjects variables with Session, which were then sequentially removed according to their p-value. AIC_C was used to select the final model, which included only the Visit × Session interaction. Overall, this model was highly significant (Table 2.9) and had a predicted mean void-trap rate of 0.021 (0.014 – 0.031). However, the only term that was significant was the main effect of Session, with a significantly higher predicted void-trap rate at dawn (0.030) compared to dusk (0.015). Although the overall effect of Visit was not significant, individual deviance contrasts showed that the predicted mean void-trap rate in Langtang (Transect 3), 0.007 which is close to the observed rate shown in Table 2.8b, was significantly lower than the overall predicted mean ($t_{(18)} = -3.008$, $p < 0.01$).

Table 2.9 ANCOVA table of fixed effects from a GLMM of void-trap rate (void-trap / events) with binomial probability distribution and logit link function. Effect *p*-values shown in red are significant at $\alpha < 0.01$.

Effect	F	df1	df2	*p*
Corrected Model	3.153	10	47	0.0037
Visit	2.340	4	26	0.0815
Alt	0.421	1	13	0.5274
Session	11.89	1	207	0.0007
Visit × Session	1.645	4	205	0.1642

Due to the significance of this model, all subsequent analyses were based on effective trap events (ETE), calculated as actual trap events minus void-traps. Overall, there were 7811 (8046 – 235; see Table 2.8) ETE.

2.D.c Capture Rates

During the three-year period of this survey 890 captures were made (Table 2.10). Allowing for the void-trap rates described above, this equated to an overall capture rate of 11.4%. Every Site visit achieved some captures, with a minimum of three in Langtang in Site 2 and a maximum of 114 in Sagarmatha in Site 1. The highest capture rate (captures / ETE) was 38% in Sagarmatha Site 1. The

vast majority of captures took place in dawn sessions (94%), although only 5 out of 25 sites failed to capture during dusk sessions.

At the trap-grid level, 10 of the 147 grid visits (6.8%) failed to make any captures. The maximum captures in a single grid visit was 23, which occurred on three occasions. On one of these trap-visits (Sagarmatha, Site 1, Grid 4), this represented a 50% capture rate. The median capture rate at this scale was 8.9%.

The basic unit of recording captures was the grid-session (see 2.2.2). At this scale the maximum number of captures was 12, representing a 66% capture rate (Figure 2.32). There were 162 grid sessions out of 447 (36.2%) with no captures. This represents significant over-dispersion ($\chi^2_{(4)} = 425$, $p \approx 0$. However, the majority of these (112; 69%) were dusk sessions, which is reflected in the figures given in Table 2.10.

Table 2.10 Captures and capture rates (captures / ETE) by Year, Transect, Site and Session.

Year	Transect	Site	Dawn		Dusk		Total			
2013	1	1	21	10.1%	2	1.9%	23	7.3%	131	9.1%
		2	21	15.8%	0	0.0%	21	10.5%		
		3	19	10.8%	0	0.0%	19	7.2%		
		4	58	15.1%	10	3.5%	68	10.2%		
	2	1	14	6.8%	0	0.0%	14	4.5%	120	9.9%
		2	23	11.0%	1	0.9%	24	7.6%		
		3	47	22.2%	2	1.9%	49	15.3%		
		4	22	32.4%			22	32.4%		
		5	9	9.2%	2	2.0%	11	5.5%		
2014	1	1	60	29.1%	2	1.9%	62	19.8%	261	15.7%
		2	40	29.6%	2	2.8%	42	20.4%		
		3	15	8.8%	2	2.2%	17	6.5%		
		4	31	14.6%	1	0.9%	32	10.0%		
		5	77	37.7%	3	2.8%	80	25.6%		
		6	26	15.8%	2	2.3%	28	11.1%		
	3	1	7	3.3%	1	0.9%	8	2.5%	129	8.0%
		2	2	0.9%	1	0.9%	3	0.9%		
		3	13	6.0%	0	0.0%	13	4.0%		
		4	42	19.6%	3	2.8%	45	14.0%		
		5	57	27.3%	3	2.9%	60	19.1%		
2015	4	1	105	53.3%	9	8.7%	114	38.0%	249	13.2%
		2	27	12.7%	0	0.0%	27	8.4%		
		3	33	15.6%	2	1.9%	35	11.1%		
		4	13	6.1%	1	0.9%	14	4.4%		
		5	16	7.5%	2	1.9%	18	5.7%		
		6	38	18.9%	3	2.8%	41	13.4%		
		Total	836	16.4%	54	2.0%	890	11.4%		

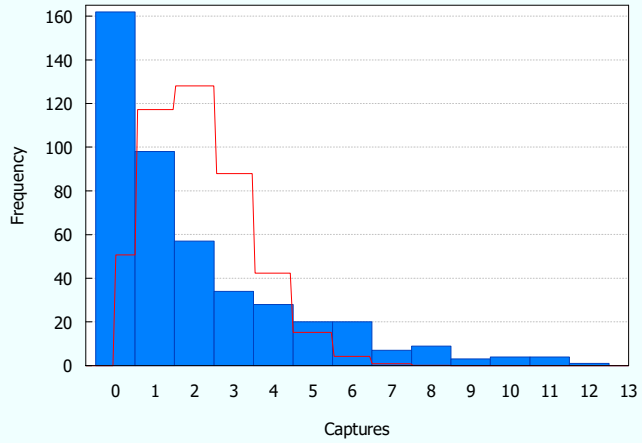


Figure 2.32 Frequency histogram of captures per grid-session. The expected binomial distribution is shown by the red line.

The modelling process described in sub-section 2.D.b was used to analyse capture rates by the structural variables and Altitude. The best overall model included both interaction terms and was very highly significant (Table 2.11). However, it was not a good fit, with significant heteroscedasticity and non-normality of residuals (K-S $D_{(290)} = 0.222$, $p \approx 0$). The model gave an overall predicted mean capture rate of 0.051, which reflected the overdispersion shown in Figure 2.32. Nevertheless, there was one highly significant main effect; Session, with predicted mean capture rates of 0.142 at dawn and 0.017 at dusk (Figure 2.33a). In addition, the interaction between Altitude and Session was marginally significant (Figure 2.33b), indicating that the negative effect of altitude was only significant for dawn capture rates, as the dusk captures were dominated by zeros.

Table 2.11 ANCOVA table of fixed effects from a GLMM of capture-rate (captures / ETE) with binomial probability distribution and logit link function. Effect *p*-values shown in red are significant at $\alpha < 0.01$.

Effect	F	df1	df2	*p*
Corrected Model	21.509	11	51	0.0000
Visit	0.760	4	28	0.5602
Alt	0.891	1	20	0.3560
Session	34.73	1	185	0.0000
Visit × Session	0.676	4	191	0.6091
Alt × Session	5.332	1	191	0.0220

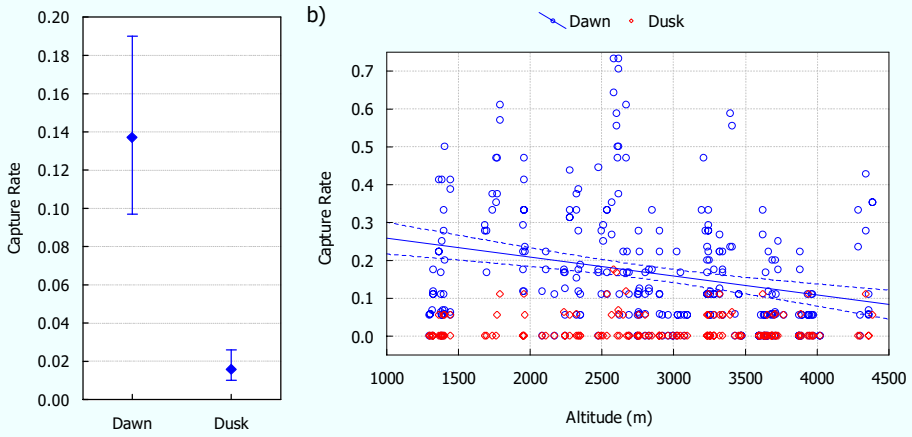


Figure 2.33 a) Predicted means plot (plus 95% C. I.) of capture rates for dusk and dawn sessions from the GLMM. b) Scatterplot showing interaction of dawn and dusk capture rates with altitude from the GLMM. Raw data points are shown, with the best-fit regression line (plus 95% C. I. of the slope) for the dawn relationship only.

2.D.d Mortality Rates

Mortality rates were calculated as the number of animals dead in the trap divided by the number of captures (Table 2.12), which gave an overall rate of 34%. Only one site achieved zero mortality which was, ironically, the first site in the first transect in 2013. Apart from this occasion, the lowest

Table 2.12 Mortality rates (dead in trap / captures), summarised by session type, site and year.

Year	Transect	Site	Dawn		Dusk		Total			
2013	1	1	0	0.0%	0	0.0%	0	0.0%	37	28.2%
		2	10	47.6%			10	47.6%		
		3	14	73.7%			14	73.7%		
		4	12	20.7%	1	10.0%	13	19.1%		
	2	1	4	28.6%			4	28.6%	64	53.3%
		2	14	60.9%	1	100.0%	15	62.5%		
		3	23	48.9%	2	100.0%	25	51.0%		
		4	15	68.2%			15	68.2%		
		5	5	55.6%	0	0.0%	5	45.5%		
2014	1	1	9	15.0%	0	0.0%	9	14.5%	90	34.5%
		2	15	37.5%	0	0.0%	15	35.7%		
		3	8	53.3%	1	50.0%	9	52.9%		
		4	16	51.6%	0	0.0%	16	50.0%		
		5	18	23.4%	1	33.3%	19	23.8%		
		6	22	84.6%	0	0.0%	22	78.6%		
	3	1	4	57.1%	0	0.0%	4	50.0%	49	38.0%
		2	1	50.0%	1	100.0%	2	66.7%		
		3	6	46.2%			6	46.2%		
		4	17	40.5%	0	0.0%	17	37.8%		
		5	20	35.1%	0	0.0%	20	33.3%		
2015	4	1	19	18.1%	0	0.0%	19	16.7%	62	24.9%
		2	8	29.6%			8	29.6%		
		3	9	27.3%	1	50.0%	10	28.6%		
		4	5	38.5%	0	0.0%	5	35.7%		
		5	5	31.3%	1	50.0%	6	33.3%		
		6	14	36.8%	0	0.0%	14	34.1%		
		Total	293	35.0%	9	16.7%	302	33.9%		

mortality rate was 14.5%, which was recorded in the same site as the zero rate, but in 2014. The highest mortality rate was 78.6% in Pipar Site 6 in 2014. Apart from the single site mentioned, dawn sessions always had some mortality, but of the 20 sites that had dusk captures, 12 (60%) had no mortality.

A GLMM model was used to analyse mortality rates by visit, session and altitude. Neither of the two-way interactions with session were significant and the AIC_C criteria indicated that the main effects model was the best fit (Table 2.13). The residuals from this model were reasonably well distributed with only moderate right skew. Consequently, the overall predicted mean mortality was 27.5%, which was not radically different from the observed value.

Table 2.13 ANCOVA table of fixed effects from a GLMM of mortality-rate (deaths / captures) with binomial probability distribution and logit link function. Effect *p*-values shown in red are significant at $\alpha < 0.01$.

Effect	F	df1	df2	*p*
Corrected Model	5.518	6	26	0.0009
Visit	5.257	4	26	0.0030
Session	7.411	1	41	0.0095
Alt	17.796	1	19	0.0005

All three main effects were significant at $\alpha = 0.01$. Visit was highly significant, with the greatest contrast between Annapurna in 2013 and Sagarmatha in 2015; the other three visits were not significantly different from each other (Figure 2.34a). The observed means for dawn and dusk were significantly different from each other, despite the larger standard error for dusk mortality rates (Figure 2.34b). The most significant effect was a positive logistic relationship with altitude (Figure 2.35). This predicted that mortality in the grid with the lowest altitude (approx. 1300m) was 14% and at the highest elevation (approx. 4400m) it was 64%. Extending the elevational range predicted that the mortality rate at sea level would be about 7% and it would exceed 95% at approximately 7500m.

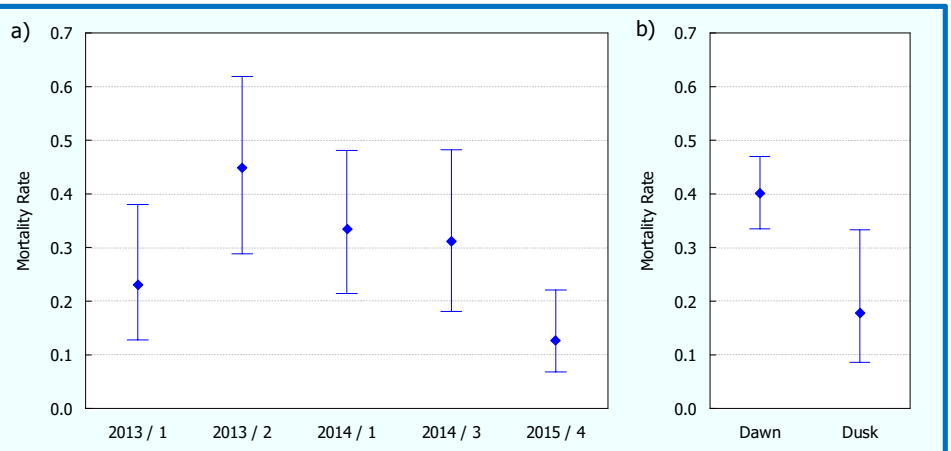


Figure 2.34 Mortality rates (mean + 95% C. I.) from a GLMM with binomial probability distribution and logit link function for a) Visits (Year / Transect Code) and b) Sessions.

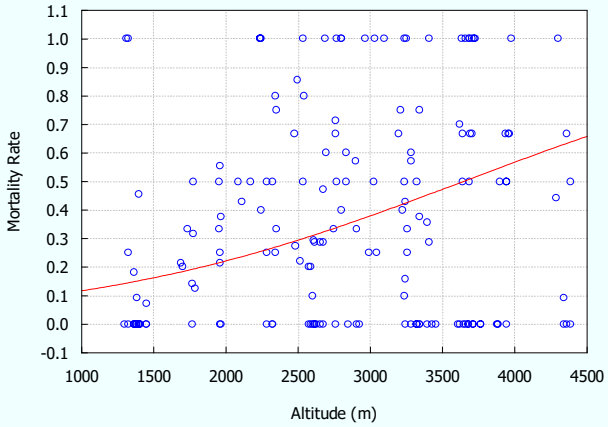


Figure 2.35 Relationship between altitude and raw mortality rates. The red line shows the logistic relationship derived from a GLMM with binomial probability distribution and logit link function

2.D.e Captures of Animals and Recapture Rates

A total of 792 individual animals was caught during the three-year survey, excluding the pilot. Thus only 98 of the 890 captures (11%) were recaptures. A complete calendar of captures for these animals is given in Supplementary Information 2.F.

Although this basic recapture rate was unexpectedly low, 291 animals were found dead on first capture. This resulted in 501 animals alive on their first capture. However, of these, seven were caught in Session 1 when it was the only session, two were caught in Session 2 when it was the last session, 203 were caught in Session 3 when it was the last and a further three animals were first caught in Session 7 which was the last session. Consequently, only 286 animals were actually caught in a condition to be recaptured. The revised recapture rate, taking account of mortality, based on 85 recaptured animals was 29.7%.

The table in Supplementary Information 2.F shows that in 2013 in Transect 1, Site 4, all grids were set for more than two nights; two grids were set for three nights and four grids were set for four nights. To summarise true recapture rate, it was necessary to sub-divide the dataset into two, slightly overlapping, cohorts, for which frequency histograms have been derived;

- The 267 animals that were first caught alive in Sessions 1 or 2 with a chance of recapture (Figure 2.36a),
- The 31 animals first caught alive with a chance of recapture in Transect 1, Site 4 in 2013 (Figure 2.36b).

Up to and including the first two recaptures, these two cohorts had virtually identical distributions (Contingency $\chi^2_{(2)} = 0.722$, $p > 0.5$). Combining these cohorts gave a final recapture rate of 28.9% of animals.

Mortality rates differed significantly between first captures and recaptures (37% and 11% respectively; Contingency $\chi^2_{(1)} = 25.33$, $p < 10^{-5}$).

Recapture rates were analysed by GLMM including Visit and Altitude as fixed effects and Site as a random effect. The response variable was the number of animals recaptured in Session 3 in a trapping grid, with the total number of animals caught alive in Sessions 1 or 2 as the trials variable. These excluded the nine grids in Transect 2 which had no opportunity for recapture, leaving 138 grid visits. Of these, 41 had zero live captures so were automatically excluded from the GLMM, leaving 97 valid cases. The model was not significant overall ($F_{(5, 14)} = 2.599$, $p > 0.05$) and had a predicted overall mean recapture rate of 14.2%, which indicated a large degree of right-skew in the data. There was no significant variation across the random factor. Altitude showed no significant effect on recapture rates ($F_{(1, 11)} = 0.250$, $p > 0.5$). However, although the main effect of visit was not

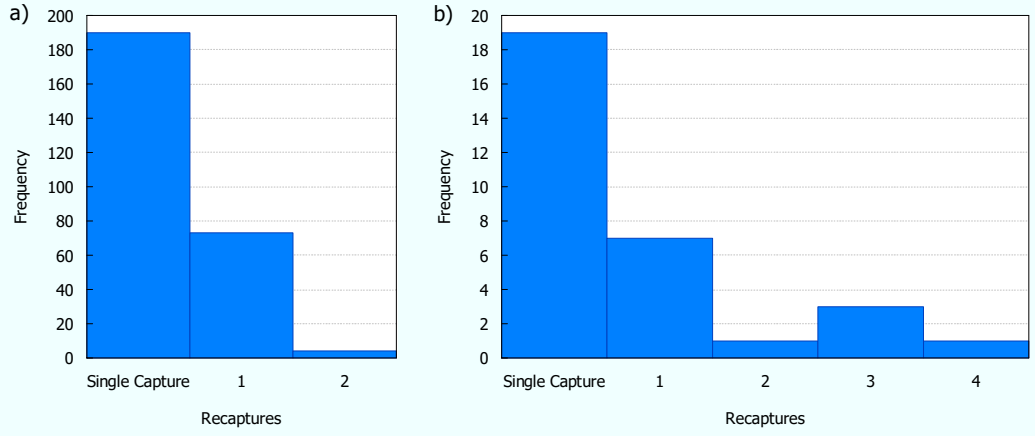


Figure 2.36 Frequency histograms of recapture rates for a) the cohort from Sessions 1 & 2 with a chance of recapture and b) the cohort from Transect 1, Site 4 in 2013 where grids had up to seven trapping sessions.

significant overall ($F_{(4, 16)} = 2.550$, $p > 0.05$), deviation contrasts indicated that the predicted recapture rates in Annapurna Transect 2 (3.4%) were significantly below the overall mean ($t_{(19)} = -3.434$, $p < 0.005$) and the predicted rate in Sagarmatha (43%) was marginally significantly above the overall mean ($t_{(7)} = 2.692$, $p < 0.05$).

2.D.e.i Population estimates using the Petersen Index

The low number of recaptures mitigated against the use of Capture-Mark-Recapture (CMR) methods to estimate abundance. However, to confirm this, 86 grid visits had live captures in Sessions 1 or 2, which were marked and released, plus at least one capture in Session 3. For these, the unbiased method recommended by Seber (1973) was used to calculate Petersen Index values, with 95% confidence limits calculated according to the formulae given in Zar (1984). Although the estimated values were reasonable, especially when the number of recaptures was relatively high, the confidence limits were extremely wide (Figure 2.37). the most extreme example was in Langtang, Site 4, Grid

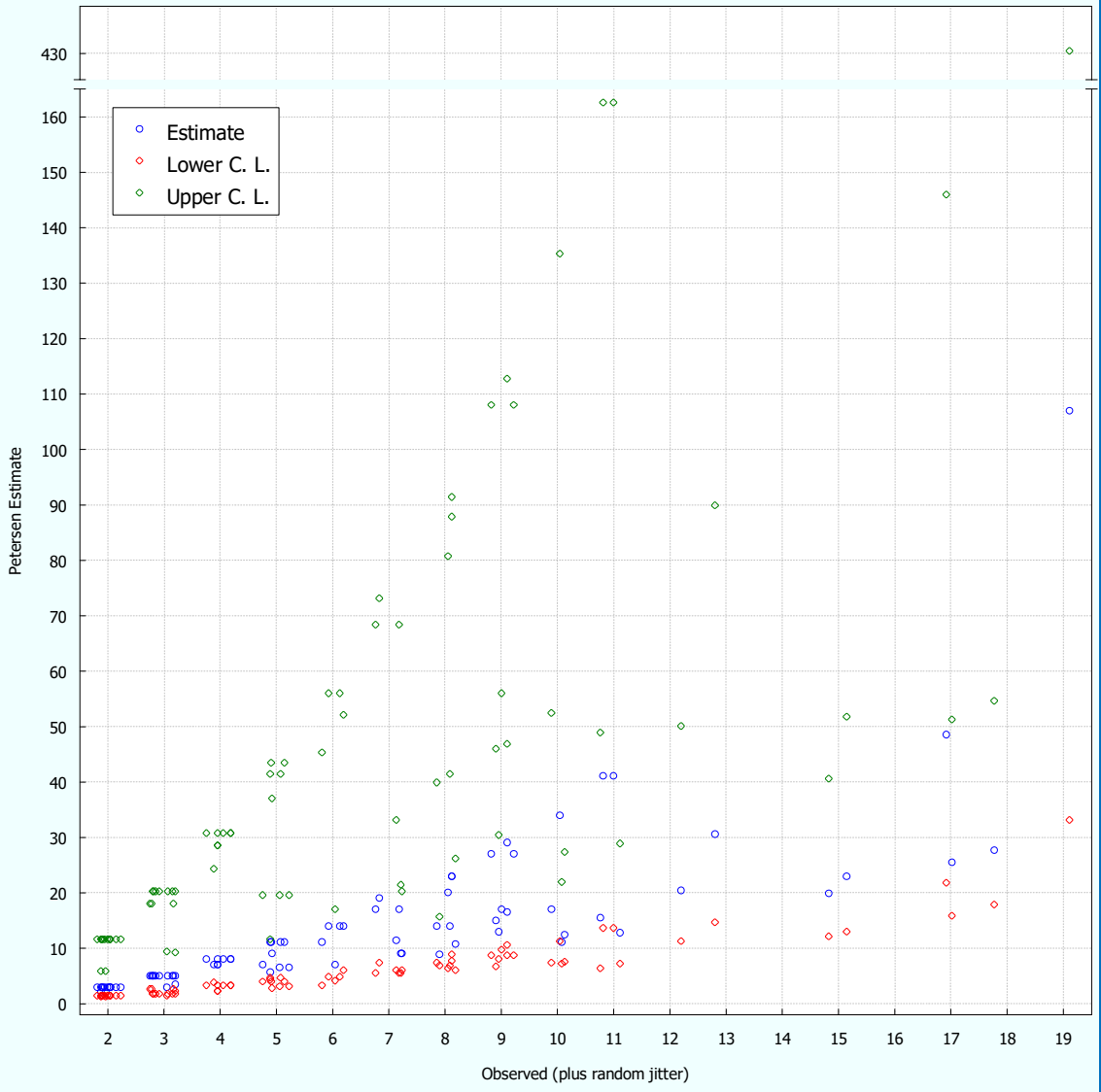


Figure 2.37 Scatterplot of Petersen Index estimates against observed number of animals. Note the scale break on the y-axis at 170 and the horizontal jitter on the x-axis.

6, when eight animals were marked and released and then 11 animals captured in Session 3, none of which were marked. This gave a Petersen estimate of 107 animals, with confidence limits of 33 and 430. At the other extreme, grids with one marked animal and only one new capture in Session 3 gave estimates of 3 animals with limits of 1.4 and 11.6. One grid with 10 animals gave a Petersen estimate of 34 with limits of 11 and 135!

2.D.e.ii Location of captures within the trapping grids

Overall captures were not distributed evenly within the 18 traps of the grid (Figure 2.38). There was a clear pattern of more captures in the traps at the ends of the arms than in the inner traps (Goodness-of-fit $\chi^2_{(2)} = 24.4$, $p < 10^{-5}$). There was no significant difference between captures at the centre and the inner ring of traps. Despite the greater captures in the extremities of the top (132) and right-hand arms (136), there was no significant difference between all four arms (G-o-f $\chi^2_{(3)} = 5.56$, $p > 0.1$). Furthermore, the interaction between the rings and arms of the grid (excluding the centre traps) was not significant (Contingency $\chi^2 = 6.84$, $p > 0.05$). These results were confirmed with a Poisson regression model, excluding the centre pair of traps; *p*-values were almost exactly the same as from the non-parametric analyses.

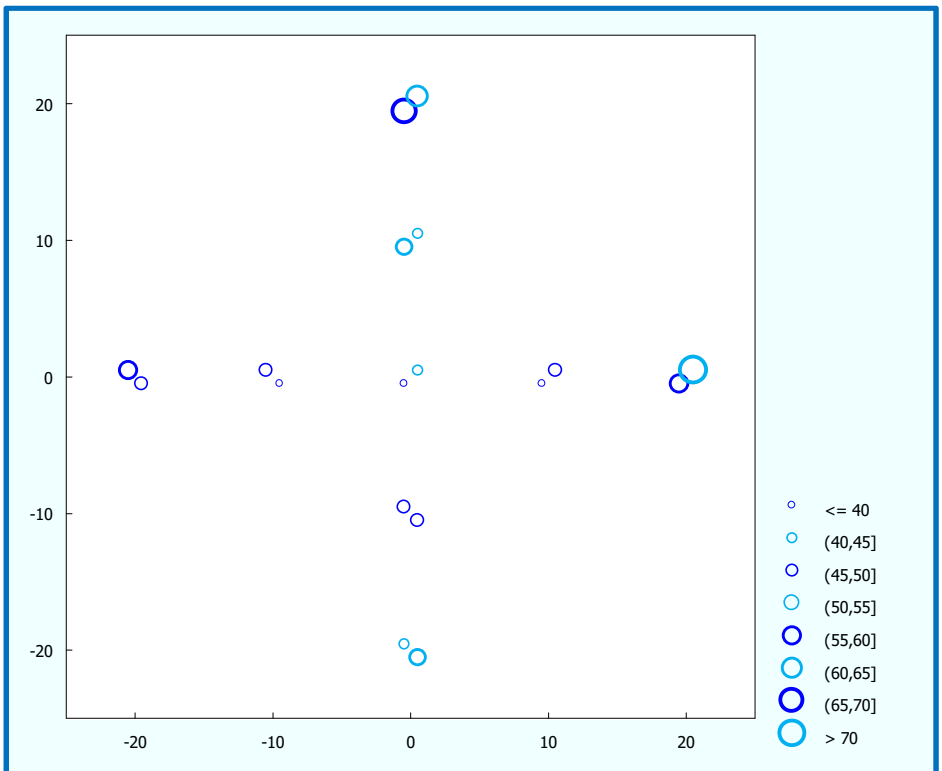


Figure 2.38 Frequency scatterplot of the spatial distribution of total captures within the 18 traps tallied across all grids. Trap positions match the codes given in Figure 2.3. Axes show distances in metres from the centre of the

2.D.e.iii Movements of animals between recaptures

In total, 85 animals were recaptured 98 times. No animals were recaptured in different grids within sites. On 40 occasions, the animal was recaptured in the same trap, and on 18 occasions it was recaptured in the other trap of the pair (Figure 2.39). Thus, 59% of recaptures showed no effective movement. One animal moved 40m between recaptures and the mean distance between recaptures in different trap-pairs was 18.3m.

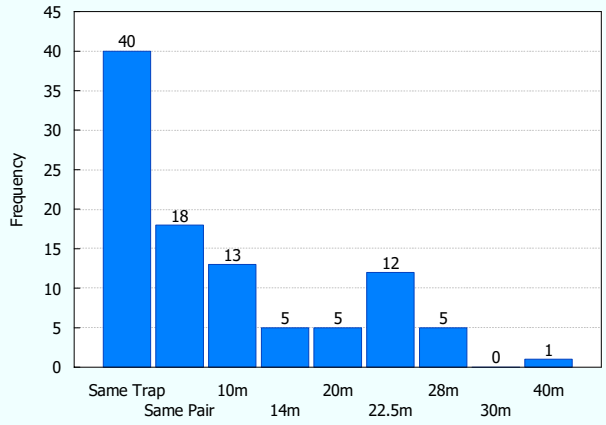


Figure 2.39 The number of recaptures at different distances within grids. The values which are not multiples of 10 are approximate Euclidean distances between points on different arms of the grid.

Movements along the X and Y axes of the trapping grids were compared, both for directional and total absolute movements (Figure 2.40). Despite these vectors being independent, 63 recaptures showed zero distance moved in each dimension, although these were tallied from different recaptures. To test for directional movement along each axis, one-sample t-tests were run against a hypothesised mean movement of zero. Neither of these were significant, as is evident from the symmetry of the graphs in Figure 2.40. To compare the amount of absolute movement along the axes, a paired-sample t-test was run, which gave $t_{(97)} = 1.118$, $p > 0.25$. However, the ratio of the variances was significant (X-axis var. / Y-Axis var. = 1.54, $p < 0.05$) which gave an indication that there was more movement along the X-axis than the Y-axis.

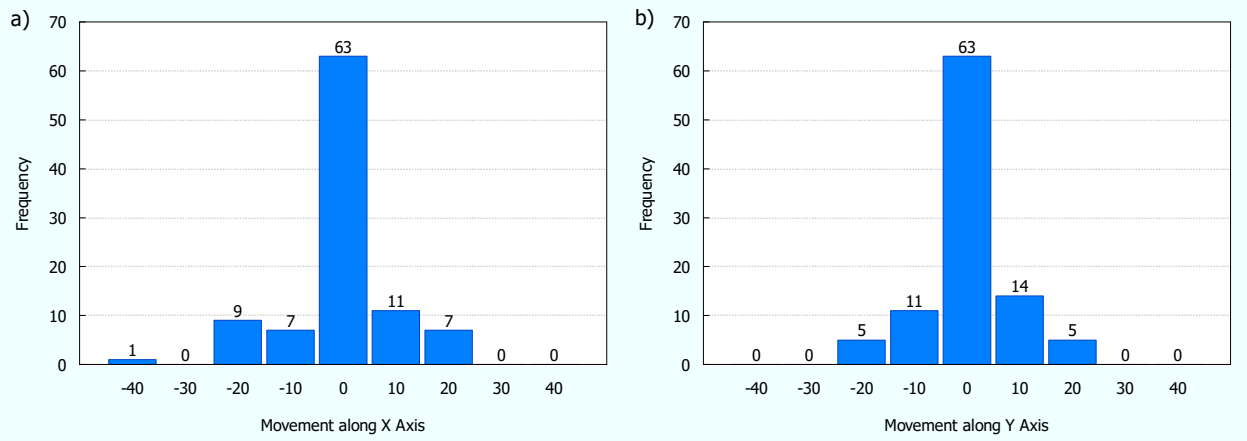


Figure 2.40 Frequency of distances moved between recaptures on a) the X axis and b) the Y axis of the trapping grids.

2.D.f Weather Effects

The two weather variables were collected for 421 trapping grid sessions. The mean temperature (of the minima prior to the trapping session) was 10°C, with a minimum of –6°C and a maximum of +25°C (Figure 2.41a). There were 94 trapping sessions where the temperature dropped to freezing or below during the previous period, all of which were dawn sessions. Precipitation occurred prior to 94 (22%) grid trapping sessions (Figure 2.41b).

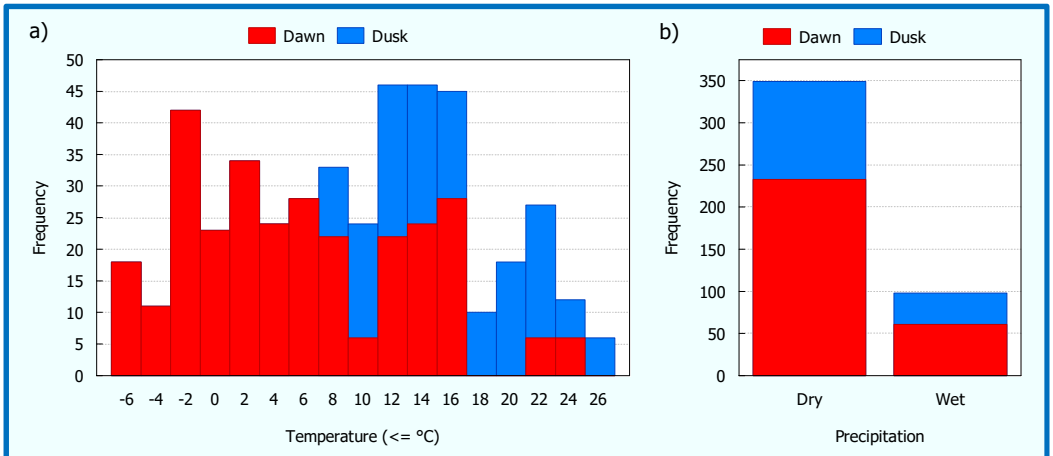


Figure 2.41 Frequency histograms of a) minimum temperature in the period prior to the trapping session and b) precipitation in the period, categorised by dawn and dusk

These variables were analysed using a simplified GLMM, because they were recorded at the trapping session level, which was a repeated-measure. Three predictor variables were included in the models; session (categorical with three levels), temperature (continuous) and precipitation (binary). Three different response variables were analysed, corresponding to the previous analyses in Sections 2.D.b to 2.D.d; void-traps, capture rates and mortality.

The overall model for void-traps was highly significant, with a significant effect of Session ($F_{(2, 206)}$ = 7.01, $p < 0.002$). Temperature had no influence on void-traps, but precipitation was highly significant ($F_{(1, 288)}$ = 11.04, $p = 0.001$). On sessions with no prior precipitation, the mean rate of void-traps was 0.020, whereas when there had been some precipitation, the mean void-trap rate was 0.040.

Neither of the weather variables were significant in the model for capture rates.

However, the best model for mortality included the interaction between temperature and precipitation (Figure 2.42). The main effect of temperature showed a significant negative effect ($F_{(1, 48)}$ = 6.308, $p < 0.02$), but the main effect for precipitation was not significant. Although the interaction was not significant in itself, it is interesting that the logistic regression lines of mortality on temperature for each category cross at exactly temperature = 0°C.

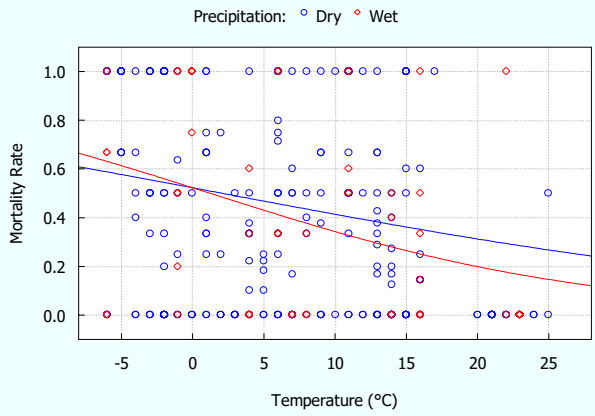


Figure 2.42 The relationship between mortality rate and temperature prior to the trapping session, categorised by precipitation level, with logistic regression lines for each.

2.E Field forms for small mammal recording

Location:		Transect:		Site No:		Grid No:		Alt:	
Date:		Recorders:				Lat:		Long:	

Session:	Start:	Finish:	Temp: °	Rain:✓	Snow:✓	General Weather:
1						
2						
3						
4						
5						

Animal	Mark	Species?	Wt	Sex	Age	Breed	H & B	Tail	H F	Ear	Photo No	Sample?	Specimen?
1													
2													
3													
4													
5													
6													
7													
8													
9													
10													
11													
12													

Session	Trap	Animal	Dead	Failure	Session	Trap	Animal	Dead	Failure

Figure 2.43 Field form printed on water-proof paper. One or more field forms were used in each grid to record times and weather for each session, a list of animals caught with morphometric data and descriptions, and a record of which trap each animal was captured in during each session.

2.F Calendar of Captures

Double-click the embedded workbook to open it. The calendar of captures shows the Year, Transect, Site and Grid for each animal that was captured. Trapping sessions are numbered 1 to 7; odd numbered sessions are dawn and even numbered are dusk sessions. Capture status for each animal is coded as;

C = Capture, R = Recapture, D = Dead, . = No Capture.

Only grids in 2013, Transect 1, Site 4 have trapping sessions after 3, as shown by the pattern of capture status codes.

CalendarOfCaptures.xlsx

Chapter 3 THE HABITAT SURVEY AND CHARACTERISATION OF THE TRAPPING TRANSECTS AND SITES

Abstract

During the small mammal trapping survey detailed habitat and botanical data were recorded for all 126 trapping grids. The habitat variables included three topographical variables (altitude, slope and aspect), and two categories of semi-quantitative variables; structural, (e.g. canopy height, percentage ground cover, etc.) and proximity to features (e.g. human habitation, crops, etc.). The botanical data comprised binary, presence/absence records for 81 taxa. These variables were used in three separate cluster analyses (structural, proximity and botanical) to group trapping grids into homogeneous types. I then used a novel genetic algorithm to characterize the final clusters in terms of their constituent variables. This generated logical rules, such as "Canopy Height > 10m", that predicted cluster membership with non-linear relationships and parsimonious use of variables. The total proportions of successful classifications were 98.2% for structural and proximity, and 96.7% for the botanical analysis. The distributions of the clusters between transects and along the altitudinal gradients was analysed. This showed that all three categories had highly significant altitudinal variation and the structural and botanical components varied significantly between transects. These three topographical variables and three categorical habitat variables, which were largely independent of each other, were subsequently used (in Chapter 6) in predictive models of small mammal abundance.

Contents

Tables

Figures

3.1 Introduction

Despite its relatively small area (147,181 km²; approx. 0.1% of global land area), Nepal has a highly diverse topography and 112 reported forest ecosystems (Anon., 2014a). Straddling the highest mountain range on earth, its altitude ranges from 59m on the Gangetic plain to 8848m at the summit of Sagarmatha (Mount Everest). This immense altitudinal range and the effects of a monsoon-dominated weather system (Sigdel and Ikeda, 2012) results in a highly diverse range of environmental conditions, which encourage extremely high biodiversity (Uddin *et al.*, 2015). Nepal contains 12 of the 867 terrestrial ecoregions in the world (Anon., 2014b; Olson *et al.*, 2001), four of which are included in the Global 200 network of priority ecoregions for global conservation (Olson and Dinerstein, 2002; Pearch, 2011).

The flora of Nepal is especially diverse with approx. 3.2% of the world's angiosperm species, 5.1% of the gymnosperms and pteridophytes and 8.2% of the world's known bryophyte species (Chaudhary *et al.*, 2016). Endemism is especially high with 284 endemic species of flowering plants (Anon., 2014a). On a national scale the flora of Nepal has been described in detail by Hara *et al.* (1978), Hara and Williams (1979) and Hara *et al.* (1982). Reflecting the importance of pteridophytes and bryophyte, several authors have enumerated national species collections for Nepal (Bhagat and Shrestha, 2010; Rajbhandary, 2016). At a local level, high botanical diversity has been described for the Manaslu Conservation Area (Sapkota *et al.*, 2017), where 276 species of flowering plants were recorded, and the Kanchenjunga region (Rai *et al.*, 2017). In the Sagarmatha National Park Bhattarai and Upadhyay (2013) recorded 168 species of vascular and non-vascular plants in only 160 1m² quadrats and (Paudel *et al.*, 2010) recorded 180 species of herbaceous angiosperms in a study area ranging between 3400m and 4650m in altitude.

The large altitudinal range in a small geographical area has been highly attractive for studies of elevational gradients. Species richness, in particular, has been explored for all plant taxa (Bhattarai and Vetaas, 2003; Carpenter, 2005), trees (Bhattarai and Vetaas, 2006), pteridophytes (Bhattarai *et al.*, 2004) and orchids (Acharya *et al.*, 2011). The peak of distribution in endemic species was shown to be at a higher elevation (3800m – 4200m) than that for total species richness (1500m – 2500m) by Vetaas and Grytnes (2002). In contrast, Acharya *et al.* (2009) showed that the maximum richness of medicinal plant species was found around 1100m, which was poorly matched by the altitudinal distribution of protected areas (3000m – 3500m).

The primary focus of the fieldwork during the current study was small mammal trapping. However, as many of the citations above indicate, the topography and habitats within my study area were very

diverse. This made it essential to record a range of environmental variables, including altitude, aspect and slope, structural and proximity features and a botanical description. These were collected in a systematic way for all trapping grids, using rapid field techniques, including estimation. In this chapter, I describe the habitats encountered, the methods employed to collect these data, as well as the analytical procedures used. Furthermore, the chapter forms an essential stage in the preparation of environmental data used in Chapter 6 for the analysis of distribution and abundance of the small mammals recorded during the current survey.

3.1.1 Aims and Objectives

The main aim of the work presented in this chapter was to collect a suite of environmental variables that could be used to characterise the trapping grids. The imperative was to obtain these as quickly as possible to avoid compromising the small mammal trapping programme. Specific objectives were;

- To record three topographical variables (altitude, aspect and slope) for all trapping grids.
- To record three categories of habitat variables (structural, proximity and botanical) for all trapping grids.
- To use exploratory analysis techniques to reduce the dimensionality of these datasets.
- To provide a descriptive analysis of the trapping grids using the derived variables and to test the relationships between them.
- To compile a suite of predictive variables for use in the analysis of distribution and abundance of small mammals presented in Chapter 6.

3.2 Methods

3.2.1 Field Data

Four basic categories of habitat data were collected at each trapping grid:

- Three topographical variables. These were recorded according to the methods of (Carpenter, 2005). Firstly, altitude was recorded using a Satmap 10 GPS (Satmap Systems Ltd, Leatherhead, UK), usually accurate to within 10m. Next, to calculate the slope of the grid, recall that two arms of the grid were aligned up and down the slope (Section 2.2.2). I estimated the vertical height difference between the top of the up-slope arm (traps 5 & 6) and the bottom of the down-slope arm (traps 13 & 14). The slope in degrees was calculated as the sine of the vertical difference / 40. Thirdly, the GPS was used to measure the aspect of the grid in degrees, which I then converted to one of eight 45° segments (octants).

- Seven **structural** variables. Estimates were made for i) canopy height in metres, ii) canopy cover (%), iii) shrub height (m), iv) shrub cover (%), v) vegetative ground cover (%), vi) leaf litter cover (%) and vii) rock cover (%). The estimates were similar to those recorded by Bateman *et al.* (2010), although they distinguished a sub-canopy layer between canopy and shrub levels. Estimates were semi-quantitative, for example small cover values would be estimated to 2% or 5% intervals, larger cover values at 10% or 25% intervals. Similarly for canopy and shrub heights, using single integer intervals for lower values and 5m intervals for tall trees.

- Six **proximity** variables. Estimates were made of the distance to the nearest features; i) human habitation, ii) crops, iii) woodland, iv) scrub, v) grassland and vi) water. I recorded distances in metres using an approximately geometric scale; *e.g.* 0 for present in the grid, 25, 50, 100, 200 and >500m. Occasionally, intermediate distances were recorded, where these were known accurately.

- General **botanical taxa**. I made a short written description of the vegetation of each grid. This listed the major groups of plants present within the grid, such as grasses, rushes, mosses, *etc.* Some plants could be identified to genus, such as *Berberis*, *Artemisia* or *Quercus*, and a few could be reliably recorded to species, *e.g. Rhododendron arboream* or *R. campanulatum*. Species identification was aided by Polunin and Stainton (1984) and Stainton (1988). After the completion of the fieldwork, I collated these into a matrix of presence/absence for 81 taxonomic groups by 126 trapping grids.

3.2.2 Analytical methods

I analysed the four categories of habitat data separately. I summarised the three topographic variables, using their raw data values. I also used the Pearson correlation coefficient to analyse the relationship between altitude and slope and one-way ANOVA to investigate, separately, the relationship between these two continuous variables and aspect. All these analyses were undertaken in Statistica v13 (Anon., 2017) The other three categories of habitat data were analysed using a four-stage procedure for each.

3.2.2.1 Cluster Analysis

Firstly, I carried out a cluster analysis to reduce dimensionality of the datasets from *n* pseudo-quantitative variables to one categorical variable. I used the joining (tree-clustering) algorithm with Ward's method for amalgamation, based on percent disagreement as the linkage distance. The seven structural variables were recorded on different intrinsic scales, especially the percentage cover compared to shrub height which only ranged between 0 and 5. I standardised this dataset (to a mean of 0 and standard deviation of 1) before clustering. The other two datasets were recorded either on a constant scale such as proximity, from 0 to 500, or as binary data for the botanical dataset, so the raw data were used for clustering. All cluster analyses were carried out in Statistica v13.

3.2.2.2 Cluster Selection

Secondly, to derive an objective threshold for selecting clusters, I used an algorithm based on the linkage distance. For each cluster analysis, I extracted a full matrix of pair-wise distances and classified these into nine different clustering patterns from two clusters up to ten. I then calculated the total within-cluster distance as a proportion of the overall total pair-wise distance (which remained constant) for all clustering patterns from two to ten. This allowed me to select the number of clusters for each category where the within-cluster distance fell below 10%. The distance matrices were generated in Statistica v13 and imported to Microsoft SQL Server for manipulation.

3.2.2.3 Cluster Description

In the final stage, I used a genetic algorithm (GA) which I had previously developed as a predictive modelling application, to characterise each cluster, based on the raw data. A full description of this GA is given in Supplementary Information 3.A. In summary, the GA generated logical rules of the form

"If canopy height > 10m then the grid belongs to Cluster 1, else another cluster."

Up to three of these logical rules were defined, combined with AND or OR conjunctions to describe each cluster. This resulted in an expression which defined a 2 × 2 confusion matrix of predicted against actual cluster membership. The ϕ (phi) coefficient was used as the test statistic for the degree of association between two binary variables. I ran the algorithm on each cluster in turn, which identified unique variable/value combinations which best predicted membership of each cluster. The GA was implemented in Microsoft SQL Server using Asp.net as a front-end.

3.2.2.4 Geographical and Altitudinal Distributions of Clusters

I carried out two contingency table χ^2 analyses for each category of clusters. Firstly, a cross-tabulation of Altitude × Cluster was carried out using three categories of altitude (< 2000m, < 3000m and > 3000m). The overall contingency χ^2 was calculated to test for differences in cluster distribution by altitude. If this test was significant, then a heterogeneity χ^2 analysis was carried out, testing each cluster in turn. Secondly, I repeated this process for the Transect × Cluster cross-tabulation. Due to the potential problem of multiple testing, I set a conservative a priori α-level of 0.005.

3.3 Results

3.3.1 Habitat Descriptions

A total of 126 trapping grids were established at altitudes ranging from 1300m to 4400m (see Chapter 2 and Supplementary information 3.B for illustrations of a sample of habitats). At the lower elevations, grids were placed in farmland; often tiny fields, less than 0.1ha in area, with terraces 0.5m – 1m high. In a few cases grids were placed in abandoned farmland, although the degree of scrub encroachment suggested that this was a temporary feature. Between 2000m and 3000m the majority of grids were set in *Quercus* or *Rhododendron arboreum* woodland, which included other tree species such as *Betulus* and *Alnus*, often with an understorey of bamboo and pteridophytes. In Sagarmatha and Langtang, *Pinus* woodland was common on sandy soils. Grids were also placed in pasture or adjacent to streams and rivers where these occurred. Above 3000m *Abies spectabilis* woodland with dense moss ground cover was most common on wet soils with open *Juniperus* woodland in xerophytic habitats. At or above the treeline *Berberus* scrub in Pipar or stands of bamboo in Annapurna dominated. Above 3500m alpine grassland dominated in Annapurna and Pipar. At 4000m and above in Langtang and Sagarmatha, *Rhododenron campanulatum* dwarf woodland or xerophytic dwarf shrubs such as *Azalea*, *Juniperus* and *Salix* were the main vegetation type.

3.3.2 Topographical Variables

Three topographical variables were recorded for each grid (Figure 3.1). The minimum altitude was 1297m and the maximum was 4388m, with a median of 2980m. 50% of grids lay between 2324m and 3658m. Slopes ranged between 0° (six grids) and a maximum of 39°, with a median of 14°. Although there was a significant correlation between slope and altitude ($r_{(124)} = 0.2408$, $p < 0.01$), R^2 was less than 6%. There was a significant difference in the number of sites by aspect ($\chi^2_{(7)} = 17.05$, $p < 0.02$), ignoring the six flat sites and assuming a null-hypothesis of uniform distribution. This was largely due to the greater than expected number of south-facing slopes and the fewer than expected grids facing NW. However, there were no significant relationships between aspect and the two continuous variables (altitude: $F_{(7, 112)} = 1.986$, $p < 0.1$ and slope: $F_{(7, 112)} = 1.022$, $p < 0.5$). These three variables were treated as independent in the analyses carried out in Chapter 6.

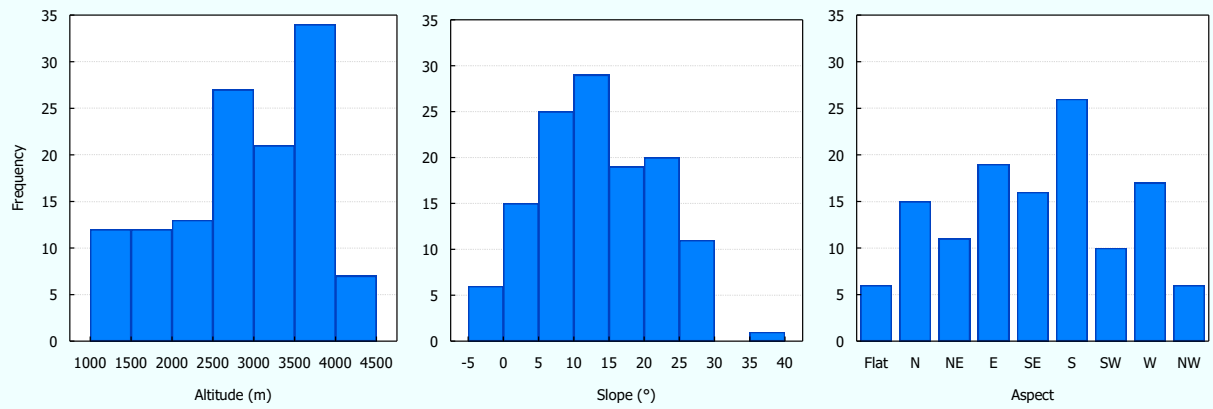


Figure 3.1 Three topographical profiles for 126 trapping grids.

3.3.3 Structural Variables

The cluster analysis was applied to the seven structural variables from the 126 trapping grids (Figure 3.2). This showed a major split, with a linkage distance of 10.85, into two groups. Firstly, there was a cluster of 52 grids, characterised by virtually no tree canopy (cover < 1% and height ≈ 1m), a low shrub layer (approx. 1.7m) and low litter cover (< 9%). This represented a cluster of grassland or agricultural land, possibly with scattered shrubs. The other cluster of 74 grids was characterised by tree canopy height ≈ 12m and cover ≈ 30%, an understory shrub height > 3m and leaf litter cover approx. 30%. This cluster included all types of woodland.

Clustering below this level had much lower linkage distances, so that the discrimination between clusters was less distinct. For example, at a linkage distance of approx. 3.6, the "woodland" cluster split into a small cluster of 18 grids which had canopy cover and leaf litter cover both > 40%. This represented dense *Abies*, *Rhododendron* or *Quercus* woodland. The other cluster had 56 grids where both canopy and litter cover were < 25%, representing more open woodland or wood pasture, with *Betulus* or *Juniperus*.

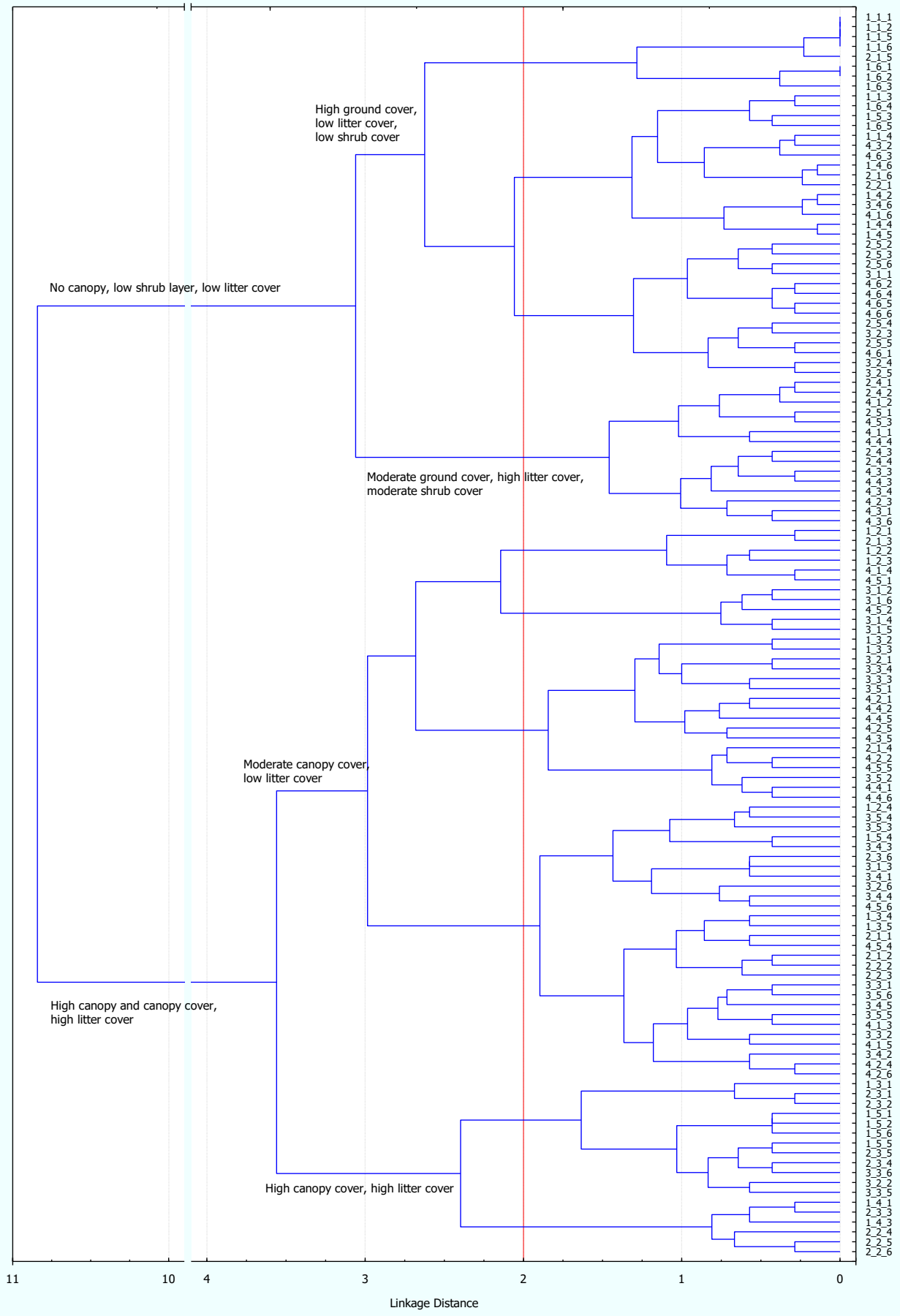


Figure 3.2 Dendrogram from the cluster analysis of the structural variables. The labels on the dendrogram leaves are trapping grid codes in the form Transect_Site_Grid. The red reference line at linkage distance = 2 divides the tree into ten clusters. Note the scale break for linkage distance between 4 & 10.

The "grassland" cluster split, with a linkage distance of approx. 3.1, into a small homogeneous cluster of 15 grids and a more heterogeneous cluster of 37 grids. The former represented dwarf shrubs such as *Azalea* and *Rhododendron*, with ground cover < 40% and litter cover > 15%. The latter had no canopy and shrub cover < 17%, ground cover > 70% and very low litter cover, representing arable, lowland pasture or alpine grassland.

The cluster selection algorithm was applied to the nine levels of clustering (from two to ten) and reached the 10% threshold described in Sub-section 3.2.2.2 at ten clusters (Figure 3.3). This corresponded to a linkage distance of approx. 2%, which is shown as a reference line in Figure 3.2.

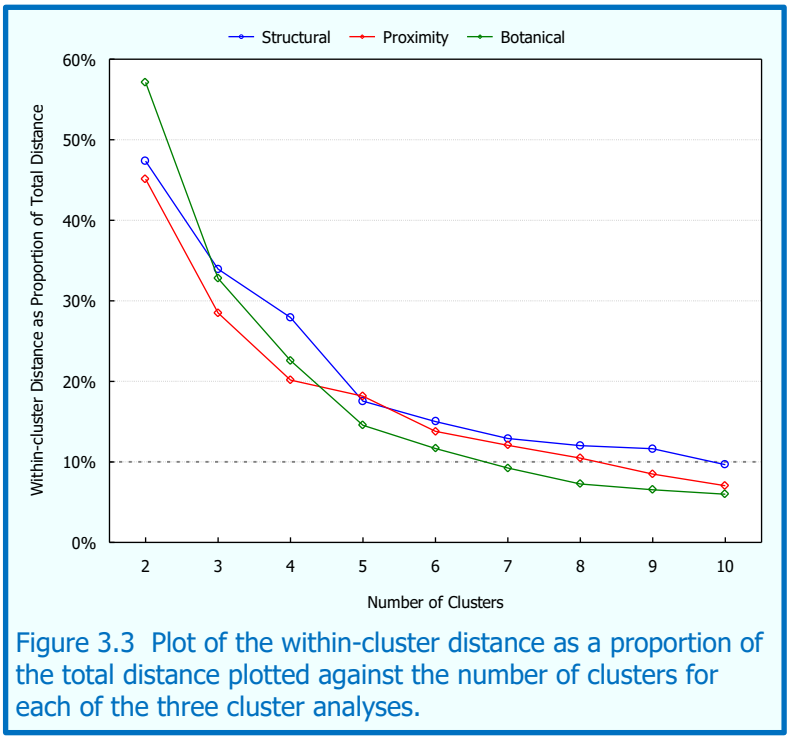


Figure 3.3 Plot of the within-cluster distance as a proportion of the total distance plotted against the number of clusters for each of the three cluster analyses.

The genetic algorithm was run using each cluster in turn as the target, against all other clusters combined. Membership of four of the ten clusters was predicted perfectly, and the overall correct prediction rate was 98.2% (Table 3.1). For example, Cluster 1 contained five grids, all of which were correctly identified (true positives) by the logical rule;

"CanopyHt BETWEEN 3 & 4 AND LitterCover >= 20"

Furthermore, no grids belonging to any other clusters were classified into this cluster, so they were all true negatives. There were no false positives or false negatives (Table 3.7 in Supplementary information 3.C.a).

Table 3.1 Summary of classification of trapping grids into three type of habitat clusters

Dataset	Clusters	With 100% Class. Rate		Grids Classified Total	Correct	
Structural	10	4		1260	1237	98.2%
Proximity	8	4		1008	990	98.2%
Botanical	7	0		882	853	96.7%
			Total	3150	3080	97.8%

Table 3.7 presents the predictive rules (together with a more easily understood interpretation) for all ten clusters, along with the number of trapping grids that they contained. The geographical locations of the grids, in terms of their transect and altitude, are shown in Figure 3.4. This shows the proportion of grids within each of the 26 sites that belonged to each (colour-coded) structural cluster.

The Altitude × Cluster contingency χ^2 analysis was highly significant (Table 3.2), which indicated that the structural clusters were not distributed randomly by altitude. The heterogeneity analysis showed that only two clusters (5 and 7) were individually significant. All grids classified as Cluster 5 were found above 3000m and 12 of the 15 grids falling into Cluster 7 were also found above 3000m. The five grids in Cluster 1 were also only found above 3000m, but this effect was marginally significant. The apparent mid-elevational predominance of Clusters 2 & 8 was not significant, as was the mid-elevational absence of Cluster 4.

Table 3.2 Summary of contingency table χ^2 analyses and heterogeneity χ^2 analyses of individual structural clusters. *P*-values highlighted in red are $p < 0.005$ and those highlighted in blue are $p < 0.05$. Non-significant results are not displayed.

			Clusters									
		Contingency	1	2	3	4	5	6	7	8	9	10
Altitude	χ^2	57.5	6.7				28		15.6			
	D.F.	18	2				2		2			
	p	5.1×10^{-6}	0.007				8×10^{-7}		4×10^{-4}			
Transect	χ^2	88.7	8.6			17		10.33	18.33		9.12	
	D.F.	27	3			3		3	3		3	
	p	1.7×10^{-8}	0.035			7×10^{-4}		0.016	4×10^{-4}		0.028	

The Transect × Cluster analysis was also very highly significant and two clusters were individually significant. Seven of the eight grids classified as Cluster 4 were found in Pipar. The majority of grids in Cluster 7 were found in Sagarmatha and their complete absence in Pipar and Langtang was highly significant.

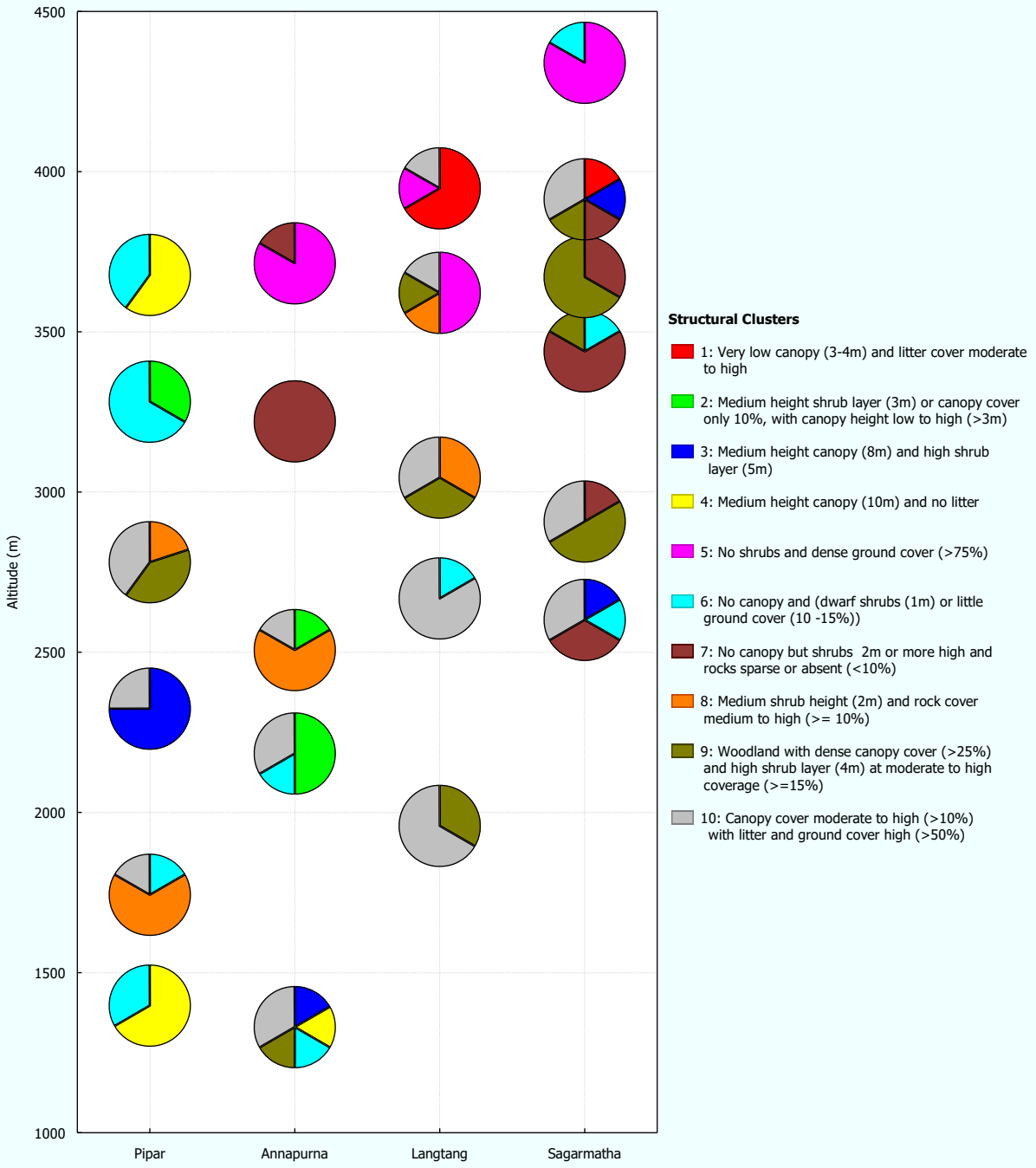


Figure 3.4 The proportion of trapping grids within each site belonging to the ten structural clusters, plotted by transect and mean site altitude.

3.3.4 Proximity Variables

The same analytical procedure was run on the proximity variables. The cluster analysis generated 8 clusters at the 10% linkage distance threshold (Figure 3.5). The first major distinction was driven by the proximity of woodland, scrub and grassland. Within these two clusters, proximity to human habitation, crops and water were the main distinguishing factors. The genetic algorithm generated descriptions with an overall correct prediction rate of 98.2% (Table 3.1). Individually, four clusters were described perfectly (Table 3.8 in Supplementary information 3.C.b).

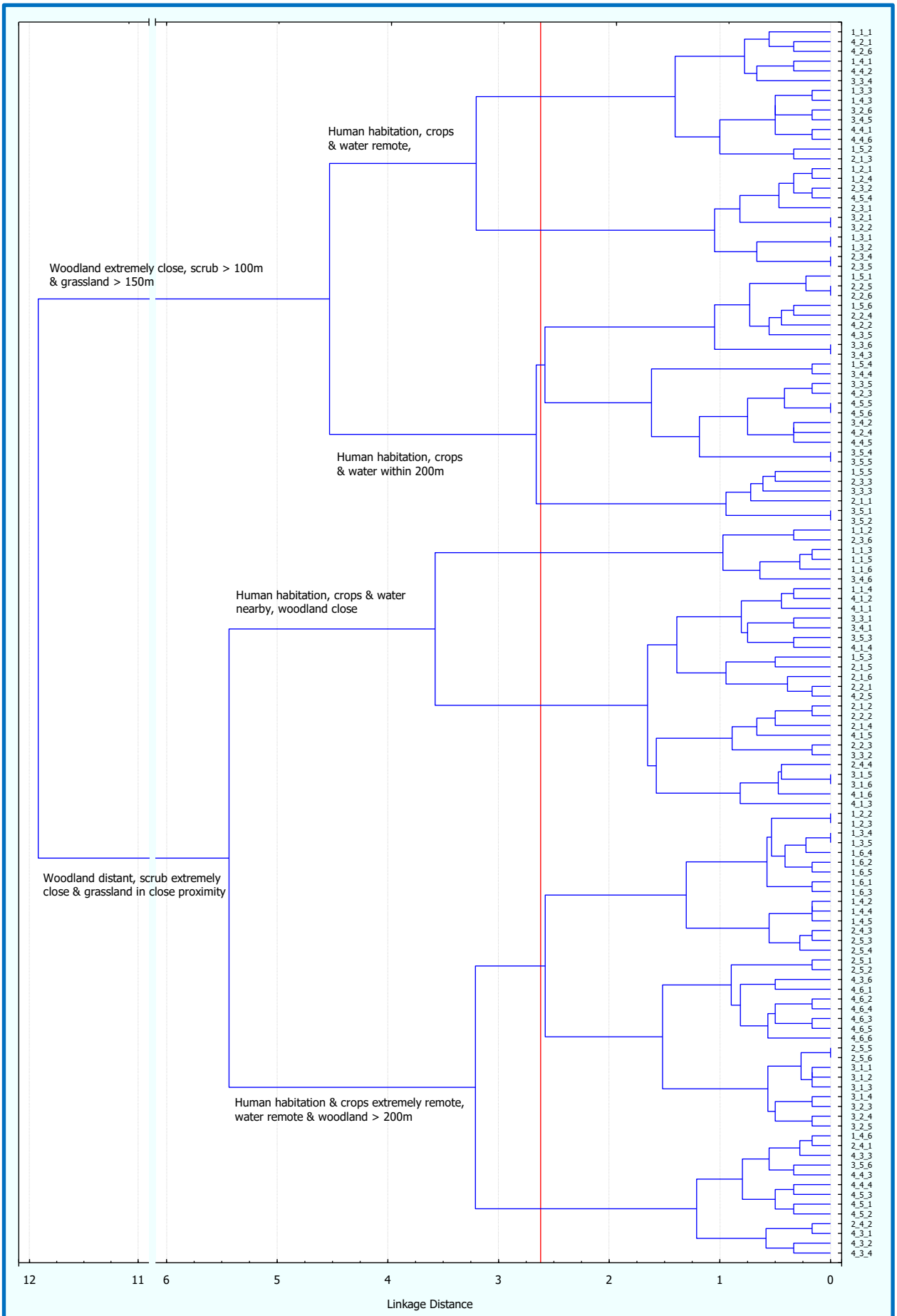


Figure 3.5 Dendrogram from the cluster analysis of the proximity variables. The labels on the dendrogram leaves are grid codes in the form Transect_Site_Grid. The red reference line at linkage distance = 2.62 divides the tree into eight clusters. Note the scale break for linkage distance between 6 & 10.

The Altitude × Cluster contingency table analysis was very highly significant (Table 3.4). Two clusters were individually highly significant; 12 of the 13 grids in Cluster 4 were above 3000m and 29 of the 33 grids in Cluster 7 were also found in this altitudinal category. The mid-elevation concentration of grids in Cluster 3 was only marginally significant (Figure 3.6).

The distribution of proximity clusters between transects was only marginally significant overall, with all clusters except 1 and 2 being found in all transects. However, individually, Cluster 4 had a significantly greater number of grids in Sagarmatha.

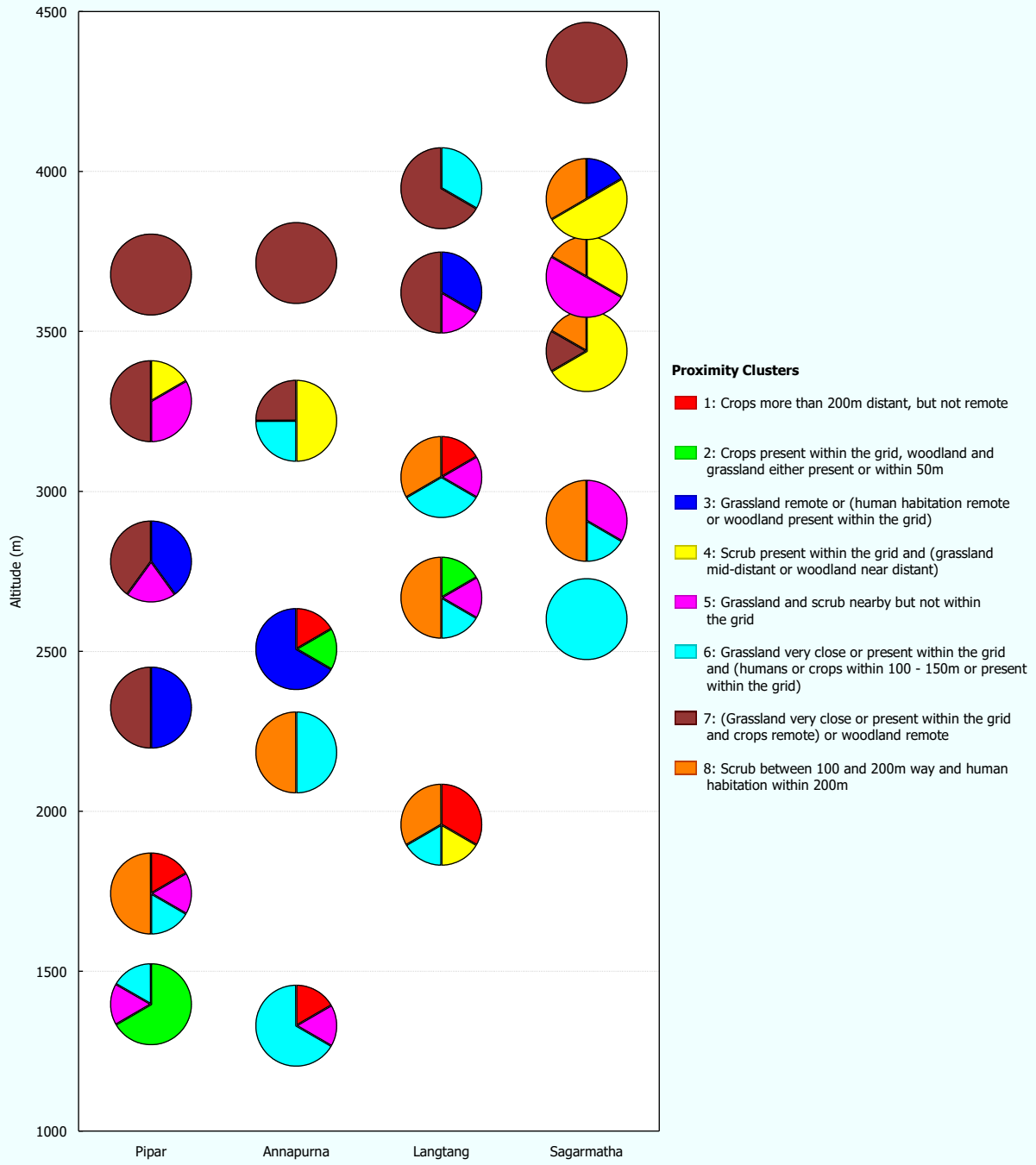


Figure 3.6 The proportion of trapping grids within each site belonging to the eight proximity clusters, plotted by transect and mean site altitude.

Table 3.3 Summary of contingency table χ^2 analyses and heterogeneity χ^2 analyses of individual proximity clusters. *P*-values highlighted in red are $p < 0.005$ and those highlighted in blue are $p < 0.05$. Non-significant results are not displayed.

			Clusters							
		Contingency	1	2	3	4	5	6	7	8
Altitude	χ^2	68.4			8.91	20.5			44.9	
	D.F.	14			2	2			2	
	p	$3.8\times10\text{-}9$			0.012	4×10^{-5}			2×10^{-10}	
Transect	χ^2	35.3				13.8				
	D.F.	21				3				
	p	0.026				3×10^{-3}				

3.3.5 Botanical Variables

The cluster analysis on the 81 botanical variables generated seven clusters at the 10% linkage distance threshold (Figure 3.7). The genetic algorithm described these clusters with 96.7% overall accuracy, although no clusters were perfectly described (Table 3.9 in Supplementary information 3.C.c). The absolute linkage distances were much lower for this dataset because the variables were binary, and were absent in most trapping grids, so that the percentage differences were lower. The first split generated clusters with 94 and 32 grids. The consistent descriptors were that the smaller cluster had *Rhododendron campanulatum* present and Compositae absent, with the logical NOT condition describing the larger cluster.

The geographical and altitudinal distribution of the botanical clusters (Figure 3.8) showed a very highly significant departure from random. The Altitude × Cluster contingency (Table 3.4) indicated that botanical clusters were found in particular altitudinal zones. Clusters 3 and 6 (with only one exception) were found in grids above 3000m. In contrast, grids classified as Cluster 1 or Cluster 5 were generally found between 2000m and 3000m. Although Cluster 2 was predominantly found below 2500m this effect was only marginally significant.

Table 3.4 Summary of contingency table χ^2 analyses and heterogeneity χ^2 analyses of individual botanical clusters. *P*-values highlighted in red are $p < 0.005$ and those highlighted in blue are $p < 0.05$. Non-significant results are not displayed.

			Clusters						
		Contingency	1	2	3	4	5	6	7
Altitude	χ^2	102.0	12.7	7.3	35.0	8.7	19.6	34.0	
	D.F.	12	2	2	2	2	2	2	
	p	2.0×10^{-16}	2×10^{-3}	0.026	3×10^{-8}	0.013	6×10^{-5}	4×10^{-8}	
Transect	χ^2	91.6	15.7	11.3	40.0	14.7	26.6	42.0	
	D.F.	18	3	3	3	3	3	3	
	p	3.0×10^{-12}	2×10^{-4}	0.028	4×10^{-4}	6×10^{-6}	0.020	0.034	

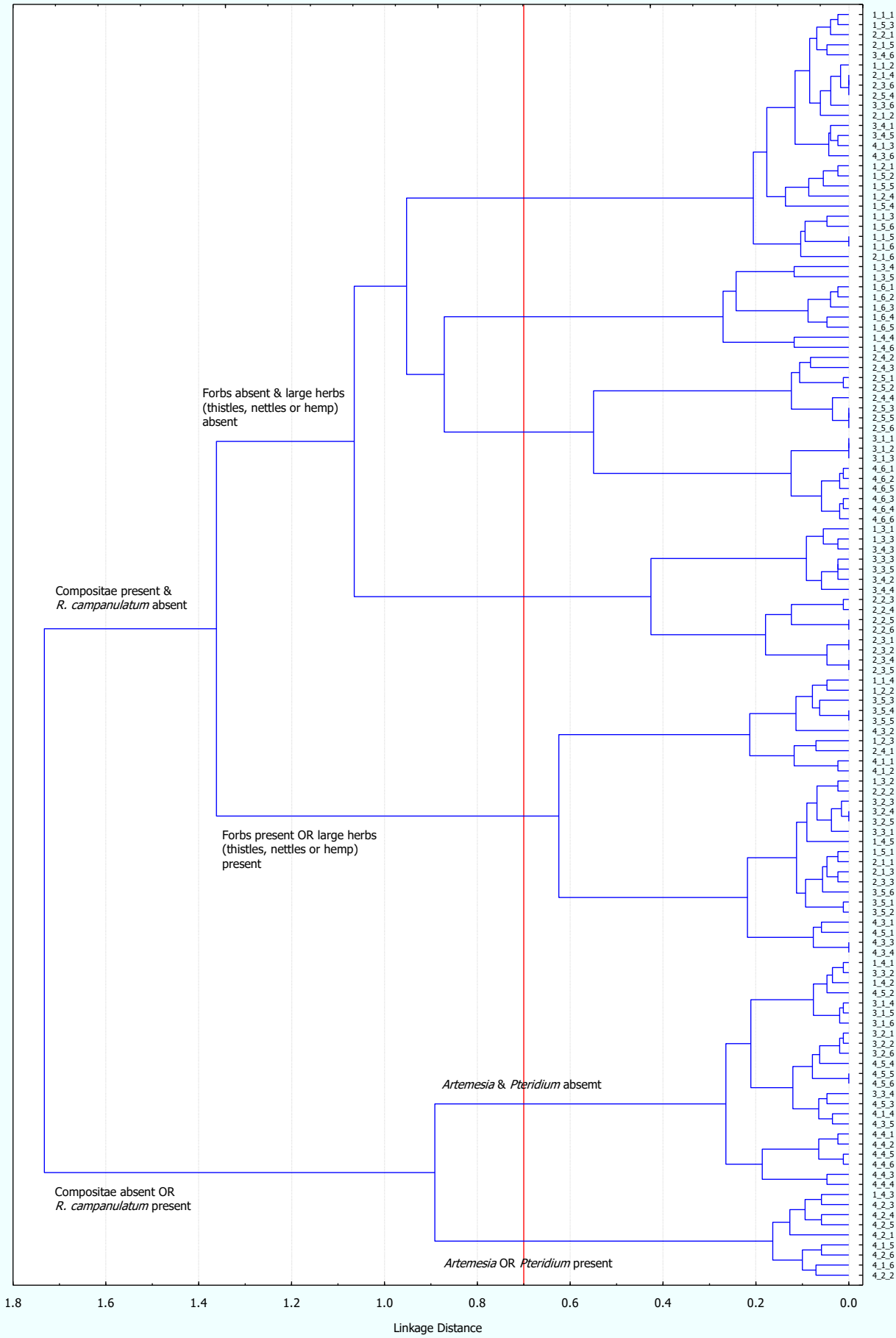


Figure 3.7 Dendrogram from the cluster analysis of the botanical variables. The labels on the dendrogram leaves are grid codes in the form Transect_Site_Grid. The red reference line at linkage distance = 0.7 divides the tree into seven clusters.

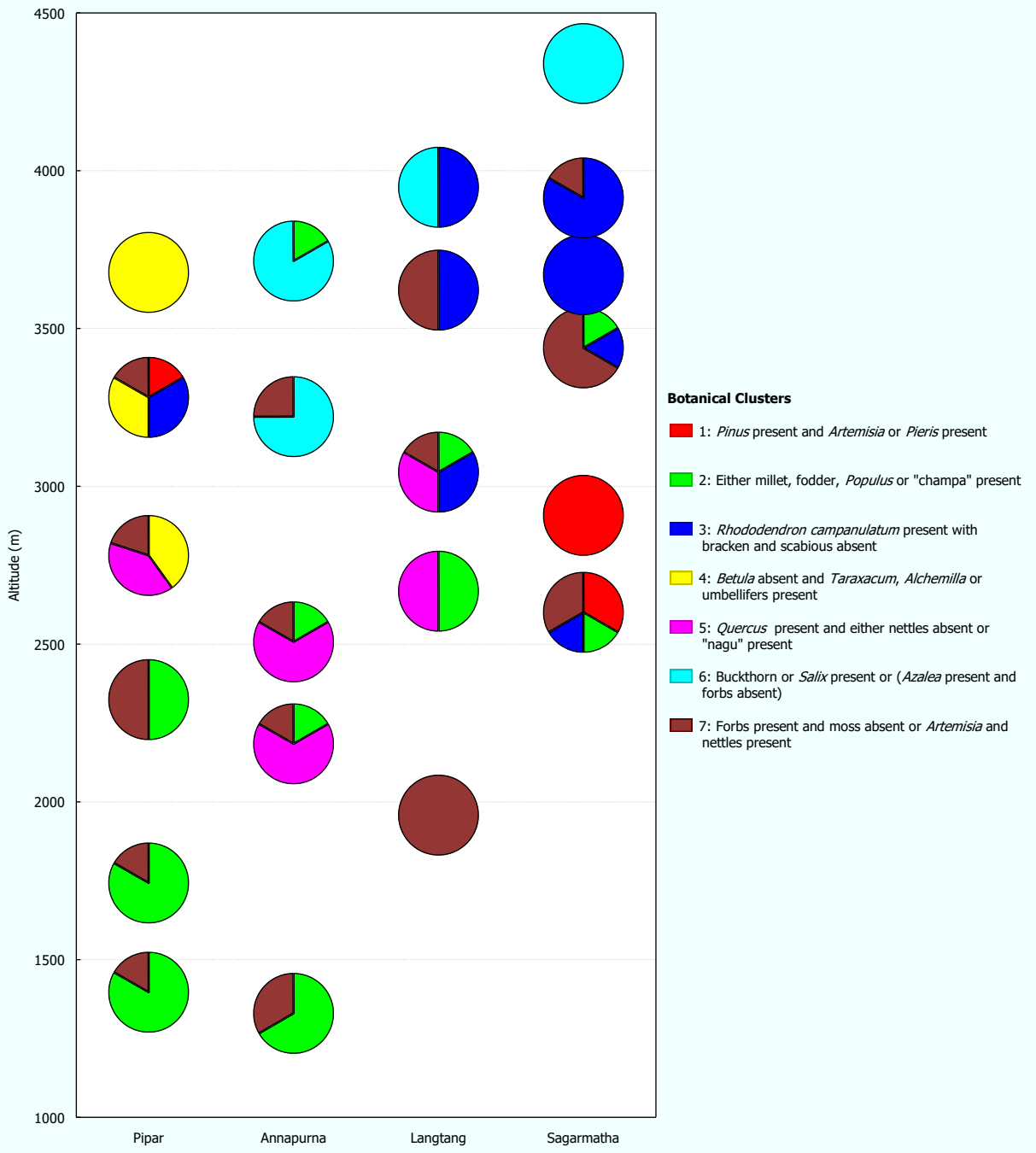


Figure 3.8 The proportion of trapping grids within each site belonging to the seven botanical clusters, plotted by transect and mean site altitude.

The overall distribution of botanical clusters between transects also showed a very highly significant departure from random and three clusters were individually significant. Eight out of nine grids classified as Cluster 1 were found in Sagarmatha. 57% of the grids classified as Cluster 3 were found in Sagarmatha, and their absence in Annapurna was highly significant. Finally, all nine grids classified as Cluster 4 were found in Pipar. Three other botanical clusters (2, 5 and 6) had marginally significant departures from expected distributions ($p > 0.01$) and Cluster 7 was ubiquitous.

3.3.6 Cluster Associations

The degree of association between categories of clusters was tested using Cramer's V. Three pair-wise comparisons were made with values of 0.465 (Figure 3.9a), 0.470 (Figure 3.9b) and 0.438 (Figure 3.9c). There was a strong association between proximity cluster 7, structural cluster 5 and botanical cluster 2, but only ten trapping grids (8%) had this combination of clusters. Otherwise, there was a generally low degree of association between clusters.

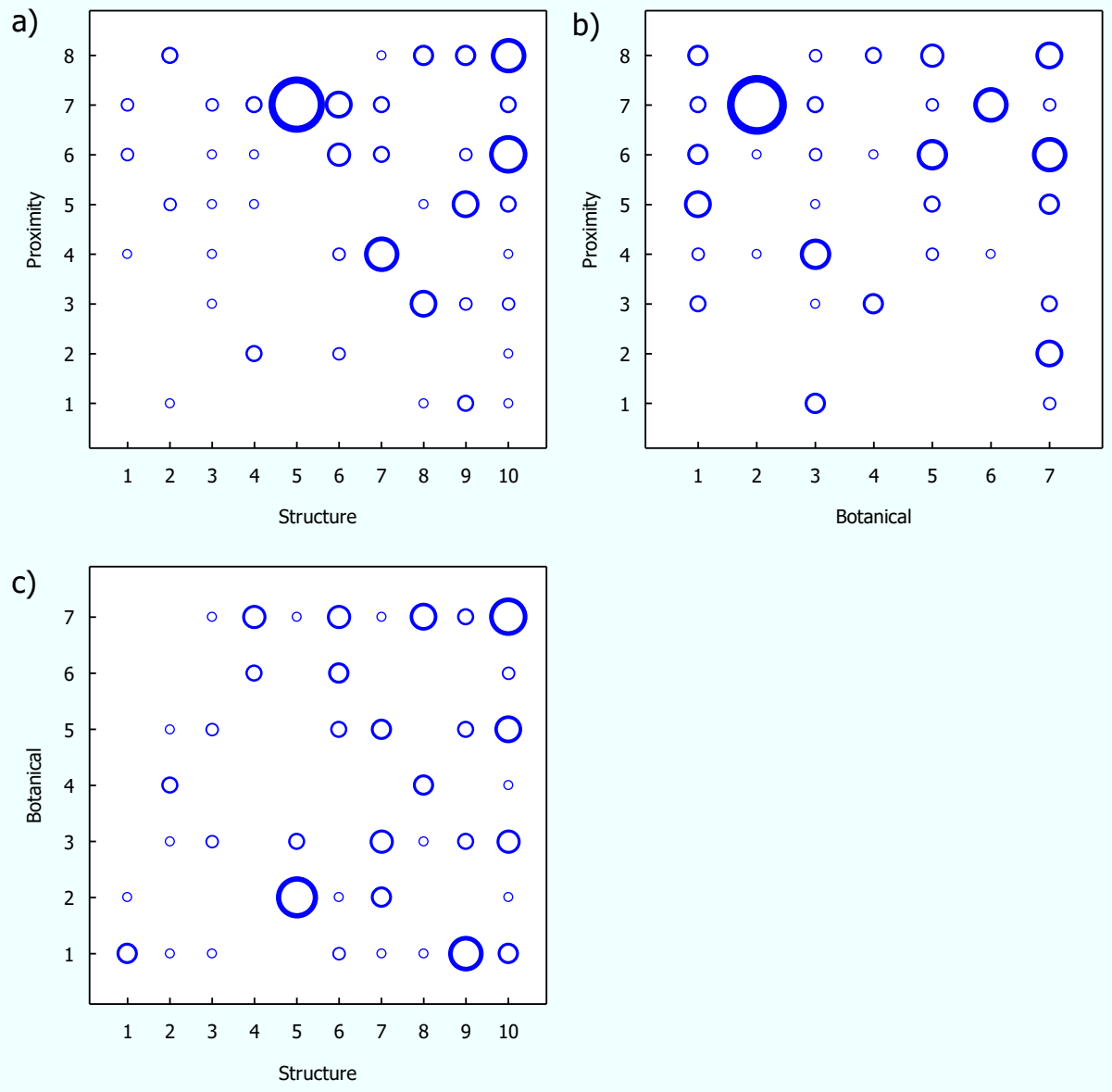


Figure 3.9 Bubble plots displaying the three two-way relationships between the category clusters. The size of the bubbles is proportional to the number of trapping grids in each cluster combination, scaled independently for each chart. The largest bubbles represent a) 14, b) 10 and c) 15 grids.

3.4 Discussion

A total of 16 habitat variables and 81 binary botanical variables were recorded for all 126 trapping grids. Together, these indicated that the habitats in which the trapping grids were placed were very diverse. This diversity occurred at all spatial levels; between transects, at different altitudes (*i.e.* between sites) and, at a small scale, between trapping grids within sites. The habitat variables comprised two continuous measurements (altitude and slope) and aspect which was categorised on eight points of the compass. Seven structural variables, including tree canopy height and cover, shrub cover and ground cover were recorded on an ordinal scale, as were six proximity variables, such as distance human habitation, crops and grassland. The botanical variables were a distillation of the presence within the trapping grids of general taxonomic categories. This generated a comprehensive matrix (126 rows × 97 columns) which could be used to characterise the trapping grids, both in terms of properties that were intrinsic to the grids themselves and within the wider environmental context.

Variables such as these have been used intrinsically in vegetation studies. Woodland canopy height is an important ecological variable and has been the focus of many studies. At a global scale Tao *et al.* (2016) used satellite-based LiDAR methods to identify precipitation and potential evapotranspiration as the major factors influencing canopy height. Also at a global scale, Gatti *et al.* (2017) used NASA's Global Forest Canopy Height map to suggest that forest volume, rather than just area, be considered in ecological studies and conservation planning. At a local level, Salas-Morales *et al.* (2018) showed that soil variables, water stress and elevation were the major drivers of canopy height in the tropical dry forest of southern Mexico. Mao *et al.* (2017) used airborne laser scanning methods to obtain upper decile values for boreal forest heights in Alberta, Canada. They showed that canopy height was positively correlated with mean annual temperature, slope and depth-to-water, although together these factors only accounted for 33% of variation. Canopy structure, rather than just height was measured using a portable LiDAR system by Hardiman *et al.* (2018), who showed that canopy structure was heterogeneous at a scale of <100m.

Several studies have used measures of proximity as predictor variables in autecological studies of mammals. Hargis *et al.* (1999) used a proximity index of open spaces in their study of forest fragmentation on American martens *Martes americana*, showing that marten captures were negatively correlated with increasing proximity to open areas. To investigate human activity influence on grizzly bear *Ursos artos* distribution, Ladle *et al.* (2018) recorded proximity to trails, roads, forest edges

and streams. Boughey *et al.* (2011) recorded the proximity to woodland and water as explanatory factors for the presence of bat species in the UK.

In the Introduction to this chapter I illustrated the extent to which the botany of Nepal has been documented, and specific concepts such as altitudinal gradients in plant species richness explored. The overall aim of the current study was to investigate the populations and communities of small mammals in the central Himalaya and to analyse their altitudinal relationships. As explained in Chapter 2, the fieldwork schedule undertaken in each site was very intensive, combining the rigours of a systematic live-trapping programme, tissue and specimen preparation with trekking and camping logistics. Within this schedule, it was not possible to undertake detailed botanical surveys or record physical measurements of features of the trapping grids. Instead, I used rapid field methods, including visual estimates, to record the structural properties of the habitat and, knowing the local geography, the proximity to major environmental features. These data were used to characterise the trapping grids, although their primary purpose was as predictor variables in the analysis of small mammal distribution and abundance presented in Chapter 6.

In this chapter, therefore, I have focussed on reducing the dimensionality of the data matrix to facilitate analysis and interpretation. My intention was to derive a small number of variables for each of the three habitat categories that explained the majority of variation across the 126 trapping grids. Consequently, my first approach was to use principal components analysis (PCA), the method used by Salas-Morales *et al.* (2018) to reduce the dimensionality of soil variables. However, this proved to have little value as, for all three categories, the eigenvalues were small. Firstly, the structural variables were not highly correlated, so the first three components only explained 76% of overall variation and the factor scores for the higher order components made them difficult to interpret. A similar pattern emerged for the proximity variables, with only 78% of overall variability being explained in the first three components and with generally low factor scores after the first two components. Finally, the 81 binary botanical variables were not amenable to this approach at all; the first component only explained 7.7% of overall botanical variation.

The alternative approach that I chose to adopt was the use of cluster analysis. I used a Joining (Tree Clustering) procedure (Manley, 1986) as I had no *a priori* expected number of clusters, which would be required by the k-means Clustering method. I used Ward's (1963) method of variance minimisation on percent disagreement as the linkage distance measure because the raw data matrices were categorical or ordinal (Hill and Lewicki, 2013).

The use of a genetic algorithm (GA) to describe the properties of the cluster is a novel approach, which proved to be very successful. Previously, I used the GA developed by Forsyth (1987) in a predictive way to identify the climatic, soil and woodland variables associated with the distribution of yellow-necked mice *Apodemus flavicollis* in the UK (Marsh *et al.*, 2001). I have used my own implementation of a GA (see Supplementary information 3.A) in a descriptive way in this chapter. This approach gave three big advantages over traditional methods; a) cluster descriptions are based on non-linear, threshold values, b) only sufficient dimensions are included in the descriptions and c) the use of the phi coefficient provides an intrinsic test of the accuracy of the description. For example, Cluster 8 of the proximity analysis contained 20 trapping grids (Table 3.8). The simple expression "Scrub between 100m and 200m away and human habitation within 200m" perfectly described this cluster; all 20 grids in the cluster conformed to this rule and the rule evaluated to false for all 106 grids not in the cluster. The more traditional method of associating means of the raw variables with cluster membership is far less useful, especially when the predictors are ordinal or categorical.

The development and use of genetic algorithms has accelerated over recent years. These tend to be highly complex and written specifically to carry out a particular task. For example, Cui *et al.* (2011) developed a GA specifically to reduce dimensionality in a large remote sensing dataset. Babu and Gopi (2015) developed a GA to overcome a similar problem of dimensionality reduction in five different medical datasets. In an industrial setting, Ribeiro *et al.* (2015) developed two GA models to reduce dimensionality in an investigation of bio-diesel adulteration. It is important to emphasise that all three examples used GAs only to reduce dimensionality – they did not use them for the actual classification analysis, which usually involved more traditional discriminant function models.

There appear to be fewer examples of the use of GAs in ecological studies. Hamblin (2013) provided a valuable overview of their use in behavioural and evolutionary studies, but the examples he gave also tended to be models designed for specific computational problems such as the "Prisoner's Dilemma" or the "Hawk-Dove" game. A tool with a more ecological focus is provided by GARP (Genetic Algorithm for Rule-set Prediction) (Stockwell and David, 1999), which applies a GA to the problem of predicting species distributions from environmental datasets (see Chapter 4 for more details). The value of the GA that I have developed here is its generality. I have used it as described in this chapter on three different types of habitat datasets. I also used it successfully as a preliminary analysis of morphological data from the small mammal collection of the Bombay Natural History Society (see Chapter 4), not included in this thesis. The GA has recently been applied to nucleotide

and amino acid sequences from guppy (*Poecilia reticulata*) datasets to predict membership of "super-type" categories (Smallbone *et al.*, in prep).

The final stage of these analyses was to test for the distribution of the clusters between transects and by altitude. The parametric approach of multinomial logistic regression models failed to resolve in two out of the three categories, resulting in dubious parameter estimation. I then undertook log-linear analysis (Fienberg, 1994) on each clustering in turn, using cluster code, transect and the three-level categorisation of altitude as the factors. Even these reduced models resulted in too many empty cells and failed to resolve.

My final solution was to resort to contingency χ^2 analysis of cluster membership versus altitudinal categories and transect separately. This showed that the botanical clusters had the most significant variation by altitude and across transects (Table 3.4). Both these effects were to be expected as the transects traversed at least four of the terrestrial ecoregions of Nepal (Pearch, 2011; WWF, 2018a, b, c, d). Although the data recorded here do not lend themselves to analysis of species richness or density, other studies in Nepal (Bhattarai and Vetaas, 2003; Bhattarai and Vetaas, 2006; Bhattarai *et al.*, 2004; Carpenter, 2005) have shown a mid-elevational "hump" in plant species richness within the altitudinal ranges in which the current study took place. In the Swat valley in Pakistan, Akhtar and Bergmeier (2015) showed that total plant species richness (S) exhibited a quadratic relationship with altitude, with a maximum predicted S at approx. 2100m. This might have been better described as a piecewise regression (Neter *et al.*, 1996) with zero relationship between 1700m and 2500m followed by a steep linear decline to 3400m. In contrast, they showed a complex cubic relationship for tree species richness with a local minimum S at 2200m. Intriguingly, this pattern repeated the one described by Carpenter (2005) in eastern Nepal as a bimodal distribution of tree species richness, with a local minimum between 1750m and 2250m. His tentative explanation was that this zone represents the interface between intensive agriculture and forest. At a lower range of altitudes in eastern China Zhang *et al.* (2016) showed a mid-elevational effect for tree species richness (1100m and 600m on two mountains), whereas most other vegetation groups showed a monotonic decline. They also showed an increase in β-diversity between plots with increased elevational differences. These studies provide corroboration of the habitat clusters which I obtained, although they do not directly inform their identities.

For future fieldwork, I would recommend a greater focus on estimates or measurements of physical characteristics of the habitat, rather than attempting to record detailed botanical data which requires a high degree of expertise. After all, it is more likely that small mammals would be responding to

characteristics such as depth and availability of leaf litter or the presence of rock cavities, which would provide shelter from adverse weather or predators. It would still be possible to include quantitative biotic variables, such as abundance of berries or tree seeds, both of which have been shown to influence small mammal abundance (Flowerdew, 1976; Flowerdew and Gardner, 1978; Poulton, 1994).

The analyses that I have presented in this chapter have generated three topographical variables and three categorical variables for different components of the habitat, which can be used to characterise the trapping grids. They have been shown to be largely independent of each other, and so constitute six useful predictor variables for the analysis of small mammal distribution presented in Chapter 6.

3.5 References

Acharya, K.P., Chaudhary, R.P., and Vetaas, O.R. (2009) Medicinal plants of Nepal: Distribution pattern along an elevational gradient and effectiveness of existing protected areas for their conservation. *Banko Janakari* **19**: 16-22.

Acharya, K.P., Vetaas, O.R., and Birks, H.J.B. (2011) Orchid species richness along Himalayan elevational gradients. *Journal of Biogeography* **38**: 1821-1833.

Akhtar, N. and Bergmeier, E. (2015) Species richness, alpha and beta diversity of trees, shrubs and herbaceous plants in the woodlands of Swat, Pakistan. *Pakistan Journal of Botany* **47**: 2107-2113.

Anon. (2014a) *Nepal Fifth National Report to Convention on Biological Diversity*. M.o.F.a.S.C. Government of Nepal, Kathmandu, Nepal. pp 77.

Anon. (2014b) *Nepal National Biodiversity Strategy and Action Plan 2014-2020*. M.o.F.a.S.C. Government of Nepal, Kathmandu, Nepal. pp 226.

Anon. (2017) Statistica (data analysis software system) v.13. TIBCO Software Inc. http://statistica.io.

Babu, P.H. and Gopi, E.S. (2015) Medical Data Classifications Using Genetic Algorithm Based Generalized Kernel Linear Discriminant Analysis. *Procedia Computer Science* **57**: 868-875.

Bateman, B.L., Kutt, A.S., Vanderduys, E.P., and Kemp, J.E. (2010) Small-mammal species richness and abundance along a tropical altitudinal gradient: an Australian example. *Journal of Tropical Ecology* **26**: 139-149.

Bhagat, I.M. and Shrestha, S. (2010) Ferns and Fern-allies of Eastern Terai, Nepal. *Our Nature* **8**.

Bhattarai, K.R. and Upadhyay, T.P. (2013) Rangeland Management in Sagarmatha (Mount Everest) National Park and Buffer Zone, Nepal: An Ecological Perspective. *Mountain Research and Development* **33**: 19-28.

Bhattarai, K.R. and Vetaas, O.R. (2003) Variation in plant species richness of different life forms along a subtropical elevation gradient in the Himalayas, east Nepal. *Global Ecology & Biogeography* **12**: 327-340.

Bhattarai, K.R. and Vetaas, O.R. (2006) Can Rapoport's rule explain tree species richness along the Himalayan elevation gradient, Nepal? *Diversity and Distributions* **12**: 373-378.

Bhattarai, K.R., Vetaas, O.R., and Grutnes, J.A. (2004) Fern species richness along a central Himalayan elevational gradient, Nepal. *Journal of Biogeography* **31**: 389-400.

Boughey, K.L., Lake, I.R., Haysom, K.A., and Dolman, P.M. (2011) Improving the biodiversity benefits of hedgerows: How physical characteristics and the proximity of foraging habitat affect the use of linear features by bats. *Biological Conservation* **144**: 1790-1798.

Carpenter, C. (2005) The environmental control of plant species density on a Himalayan elevation gradient. *Journal of Biogeography* **32**: 999-1018.

Chaudhary, R.P., Uprety, Y., and Joshi, S.P. (2016) *Plant Biodiversity in Nepal: Conservation and Legal Status*. Pages 224-268 *in* P.K. Jha, Siwakoti, M., and Rajbhandary, S., editors. *Frontiers of Botany*. Central Department of Botany, Tribhuvan University, Kathmandu, Nepal.

Cui, M., Prasad, S., Mahrooghy, M., Bruce, L.M., and Aanstoos, J. (2011) Genetic algorithms and Linear Discriminant Analysis based dimensionality reduction for remotely sensed image analysis. 2373-2376.

Fienberg, S.E. (1994) *The Analysis of Cross-Classified Categorical Data*. 2nd edition. MIT Press. pp 198.

Flowerdew, J.R. (1976) The effect of a local increase in food supply on the distribution of woodland mice and voles. *Journal of Zoology* **180**: 509-513.

Flowerdew, J.R. and Gardner, G. (1978) Small rodent populations and food supply in a Derbyshire ashwood. *Journal of Animal Ecology* **47**: 725-740.

Forsyth, R. (1987) *PC | BEAGLE User Guide*. Pathway Research Ltd. Nottingham, UK.

Gatti, R.C., Di Paola, A., Bombelli, A., Noce, S., and Valentini, R. (2017) Exploring the relationship between canopy height and terrestrial plant diversity. *Plant Ecology* **1**: 899-908.

Hamblin, S. (2013) On the practical usage of genetic algorithms in ecology and evolution. *Methods Ecol Evol* **4**: 184-194.

Hara, H., Chater, A.O., and Williams, H.J. (1982) *An Enumeration of the Flowering Plants of Nepal. III*. British Museum (Natural History), London.

Hara, H., Stearn, W.T., and Williams, L.H.J. (1978) *An Enumeration of the Flowering Plants of Nepal. I*. British Museum (Natural History).

Hara, H. and Williams, H.J. (1979) *An Enumeration of the Flowering Plants of Nepal. II*. British Museum (Natural History), London.

Hardiman, B.S., LaRue, E.A., Atkins, J.W., Fahey, R.T., Wagner, F.W., and Gough, C.M. (2018) Spatial Variation in Canopy Structure across Forest Landscapes. *Forests* **9**: 474-485.

Hargis, C.D., Bissonette, J.A., and Turner, D.L. (1999) The influence of forest fragmentation and landscape pattern on American martens. *Journal of Applied Ecology* **36**: 157-172.

Hill, T. and Lewicki, P. (2013) *Statistics: Methods and Applications. An Electronic Statistics Textbook: http://www.statsoft.com/textbook/*. Statsoft, Tulsa, Oklahoma, USA.

Ladle, A., Avgar, T., Wheatley, M., Stenhouse, G.B., Nielsen, S.E., and Boyce, M.S. (2018) Grizzly bear response to spatio-temporal variability in human recreational activity. *Journal of Applied Ecology* **00**: 1-12.

Manley, B.F.J. (1986) *Multivariate Statistical Methods*. Chapman & Hall, London. pp 159.

Mao, L., Bater, C.W., Stadt, J.J., White, B., Tompalski, P., Coops, N.C., and Nielsen, S.E. (2017) Environmental landscape determinants of maximum forest canopy height of boreal forests. *Journal of Plant Ecology* **rtx071**: 1-7.

Marsh, A.C., Poulton, S., and Harris, S. (2001) The Yellow-necked Mouse *Apodemus flavicollis* in Britain: status and analysis of factors affecting distribution. *Mammal Review* **31**: 203-227.

Neter, J., Kutner, M., H., Nachtsheim, C.J., and Wasserman, W. (1996) *Applied Linear Statistical Models*. Irwin, Chicago & London. pp 1408.

Olson, D.M. and Dinerstein, E. (2002) The Global 200: Priority ecoregions for global conservation. *Annals of the Missouri Botanical Garden* **89**: 199-224.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'Amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P., and Kassem, K.R. (2001) Terrestrial ecoregions of the world: a new map of life on Earth. *Bioscience* **51**: 933-938.

Paudel, E.N., Shrestha, K.K., and Bhuja, D. (2010) *Enumeration of Herbaceous Plants in Imja Valley, Sagarmatha National Park, Nepal*. Pages 173-188 *in* P.K. Jha and Khanal, I.P., editors. *Contemporary Research in Sagarmatha (Mt. Everest) Region, Nepal*. Nepal Academy of Science and Technology, Lalitpur, Kathmandu, Nepal.

Pearch, M.J. (2011) A review of the biological diversity and distribution of small mammal taxa in the terrestrial ecoregions and protected areas of Nepal. *Zootaxa* **3072** 1-286.

Polunin, O. and Stainton, A. (1984) *Flowers of the Himalaya*. OUP, New Delhi, India. pp 580.

Poulton, S.M.C. (1994) *Small mammal populations in hedgerows: the relationship with seed and berry production*. Pages 133-138 *in* N. Boatman, editor. *Field Margins: Integrating Agriculture and Conservation*. British Crop Protection Council, University of Warwick, UK.

Rai, S.K., Tamang, R., Gajurel, J.P., Devkota, S., Shrestha, B., Nobis, M., and Scheidegger, C. (2017) Environmental covariates of species richness and composition of vascular plants of Olangchung Gola and Ghunsa valleys of Eastern Nepal. *Asian journal of Conservation Biology* **6**: 94-104.

Rajbhandary, S. (2016) Fern and Fern Allies of Nepal. *Frontiers of Botany*: 124-150.

Ribeiro, L.d.A., Soares, A.d.S., Lima, T.W.d., Jorge, C.A.C., Costa, R.M.d., Salvini, R.L., Coelho, C.J., Federson, F.M., and Gabriel, P.H.R. (2015) Multi-objective Genetic Algorithm for Variable Selection in Multivariate Classification Problems: A Case Study in Verification of Biodiesel Adulteration. *Procedia Computer Science* **51**: 346-355.

Salas-Morales, S.H., González, E.J., and Meave, J.A. (2018) Canopy height variation and environmental heterogeneity in the tropical dry forests ofcoastal Oaxaca, Mexico. *Biotropica* **50**: 26-38.

Sapkota, S., Pandey, B., and Shrestha, K.K. (2017) Diversity of Flowering Plants in Nubri Valley, Manaslu Conservation Area, Central Nepal. *American Journal of Plant Sciences* **8**: 1484-1498.

Sigdel, M. and Ikeda, M. (2012) Summer Monsoon Rainfall over Nepal Related with Large-Scale Atmospheric Circulations. *Journal of Earth Science & Climatic Change* **3**: 112.

Smallbone, W., Ellison, A., Poulton, S., Van Oosterhout, C., and Cable, J. (in prep) Depletion of MHC supertype during domestication can compromise immunocompetence. *Molecular Ecology*.

Stainton, A. (1988) *Flowers of the Himalaya: A Supplement*. OUP, New Delhi, India. pp 86.

Stockwell, D.R.B. and David, P. (1999) The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science* **13**: 143-158.

Tao, S., Guo, Q., Li, C., Wang, Z., and Fang, J. (2016) Global patterns and determinants of forest canopy height. *Ecology* **97**: 3265-3270.

Uddin, K., Shrestha, H.L., Murthy, M.S., Bajracharya, B., Shrestha, B., Gilani, H., Pradhan, S., and Dangol, B. (2015) Development of 2010 national land cover database for the Nepal. *Journal of Environmental Management* **148**: 82-90.

Vetaas, O.R. and Grytnes, J.A. (2002) Distribution of Vascular Plant Species Richness and Endemic Richness along the Himalayan Elevation Gradient in Nepal. *Global Ecology & Biogeography* **11**: 291-301.

Ward, J.H.J. (1963) Hierarchical Grouping to Optimize an Objective Function. *Journal of American Statistical Association* **58**: 236-244.

WWF. (2018a) The Eastern Himalayan Alpine Shrub and Meadows [PA1003]. https://www.worldwildlife.org/ecoregions/pa1003

WWF. (2018b) The Eastern Himalayan Broadleaf Forests [IM0401]. https://www.worldwildlife.org/ecoregions/im0401

WWF. (2018c) The Eastern Himalayan Sub-Alpine Conifer Forests [IM0501]. https://www.worldwildlife.org/ecoregions/im0501

WWF. (2018d) The Himalayan Subtropical Pine Forests [IM0301]. https://www.worldwildlife.org/ecoregions/im0301

Zhang, W., Huang, D., Wang, R., Liu, J., and Du, N. (2016) Altitudinal Patterns of Species Diversity and Phylogenetic Diversity across Temperate Mountain Forests of Northern China. *PLoS One* **11**: e0159995.

SUPPLEMENTARY INFORMATION

3.A A Genetic Algorithm for Uncovering Patterns in Multivariate Data

I developed this genetic algorithm (GA) for the analysis of a range of datasets, from the concepts of Richard Forsyth (1987). The algorithm predicts category membership for cases from a matrix of predictor variables, using logical expressions. In this sense, it is similar to discriminant function analysis (DFA), but differs in three important respects;

- It is not constrained to use continuous predictor variables, rather it can use any combination of binary, categorical, ordinal and continuous variables.
- It can use non-linear functions of the independent variables to predict membership of the target condition.
- It automatically builds threshold, interaction and polynomial terms – these do not have to be explicitly defined, *a priori*.

The application employed here uses a MS SQL Server database engine with Transact-SQL views, stored procedures and functions to run the GA. It also has a graphical user interface developed in ASP.NET, with both running on a MS Azure cloud platform. This will allow future development as a publicly-available cloud application.

3.A.a The Logical Model

1) Take a dataset with cases or records in rows and variables in columns. Define one binary target variable which classifies the "true" and "false" conditions. This can be an intrinsically binary variable, such as sex, one of a set of multinomial categories or a continuous variable split into two groups by a threshold value.
2) Define which of the remaining variables in the dataset are to be used as predictor variables. These can be binary, categorical, ordinal or continuous.

The target and predictor variables together define the hypothetical question. For example, "Can we predict membership of a habitat category from a suite of structural variables?". An example dataset of this type is shown in Table 3.5.

Table 3.5 Partial dataset (first 20 rows) of the structural variables used by the genetic algorithm to predict habitat cluster membership of the 126 trapping grids. The first column is the grid ID, simply used for identification. This is followed by seven columns (green) used as predictor variables. Finally, the (blue) column is the target variable, which is used to partition the rows into Target-True (*e.g.* C1) and Target-False (C2) categories.

Grid ID	Canopy Cover (%)	Canopy Height (m)	Shrub Cover (%)	Shrub Height (m)	Ground Cover (%)	Litter Cover (%)	Rock Cover (%)	Habitat Cluster
G1_1_1	0	0	0	0	100	0	0	C1
G1_1_2	0	0	0	0	100	0	0	C1
G1_1_3	0	0	5	5	90	0	0	C1
G1_1_4	0	0	25	3	75	0	0	C1
G1_1_5	0	0	0	0	100	0	0	C1
G1_1_6	0	0	0	0	100	0	0	C1
G1_2_1	10	10	0	0	50	0	10	C2
G1_2_2	25	10	20	5	33	0	5	C2
G1_2_3	5	10	25	4	50	0	5	C2
G1_2_4	40	15	10	5	75	10	3	C2
G1_3_1	75	12	0	0	10	90	0	C2
G1_3_2	65	10	10	4	25	75	10	C2
G1_3_3	75	10	5	4	25	75	0	C2
G1_3_4	5	6	25	3	100	10	2	C2
G1_3_5	5	8	10	3	100	5	0	C2
G1_4_1	5	8	80	4	75	10	2	C2
G1_4_2	0	0	10	2	100	0	5	C1
G1_4_3	5	8	75	4	60	40	2	C2
G1_4_4	0	0	2	2	100	0	0	C1
G1_4_5	0	0	2	2	100	10	0	C1
…	…	…	…	…	…	…	…	…

3) The algorithm starts by generating a "founder population" of logical expressions (predictive rules). The size of the founder population is user-defined, but is usually in the order of 10 – 20. The expressions are generated from a randomly chosen predictor variable, a randomly chosen arithmetic operator and a randomly chosen value, of the form; "Canopy Height ≤ 10m".

4) Each of the predictive rules is tested against the target expression by tallying a 2 × 2 contingency table. This creates four cells; Target True/Prediction True, Target False/Prediction False, both of which constitute correct predictions, and two cells representing incorrect predictions (Table 3.6).

Table 3.6 2 × 2 contingency table of target = C1 and a predictive rule for the example dataset shown in Table 3.5. Cells shaded green represent correct predictions, cells shaded red are incorrect predictions.

		Predictive Rule (Canopy Height ≤ 10m)		
		FALSE	TRUE	TOTAL
Target (C1)	FALSE	7	4	11
	TRUE	0	9	9
	TOTAL	7	13	20

5) The phi (ϕ) coefficient is used to score this contingency table:

$$\phi = \frac{n_{11}n_{00} - n_{10}n_{01}}{\sqrt{n_{1*}n_{0*}n_{*0}n_{*1}}}$$

$$\phi = \frac{9 \times 7 - 0 \times 4}{\sqrt{13 \times 7 \times 11 \times 9}} = \frac{63}{94.92} \cong 0.6637$$

6) The n top scoring predictive rules (n is user-defined – usually 10) are saved as the "breeding population" and any remaining rules are killed off.

7) Each rule in the breeding population is then used to breed a new rule through a number of different operations;

   - Altering the numeric value(s) of the rule.
   - Changing the operator.
   - Continuous variables and values can be combined arithmetically.
   - Combining rules using AND or OR conjunctions.
   - Modifying existing combinations by changing the conjunction or removing elements.
   - Stochastically creating entirely new predictive rules.

8) All new rules are scored using the process defined in stages 4 & 5. The phi coefficient is penalised by the number of variables/values employed in each rule to ensure the most parsimonious rule structure. These child rules are then combined with the parent rules to give a new generation of fully scored rules.

9) Stages 6 to 8 are then repeated for a user-defined number of generations, usually between 1,000 and 10,000. A user-defined stopping rule can be applied to halt execution after a given number of generations with no improvement in the top score.

10) Multiple populations can be run to overcome local bottle-necks or "extinctions", and explore different regions of multi-dimensional space.

3.B Illustrations of Habitats

The following figures provide examples of the habitats encountered during the trapping programme. They are ordered by ascending altitude, with the transect name and trapping grid ID.

Figure 3.10 1345m. Transect: Pipar. Grid: 1_1_1.
Farmland growing millet, beans and fodder crops, with scattered trees and shrubs.

Figure 3.11 1365m. Transect: Annapurna. Grid: 2_1_4
Riparian woodland with dense ground cover and riverside boulders.

Figure 3.12 2220m. Transect: Pipar. Grid: 1_2_1
Dense *Quercus* woodland with shrub layer and high levels of ground cover.

Figure 3.13 2475m. Transect: Annapurna. Grid: 2_3_1
Dense *Rhododendron* woodland with bamboo.

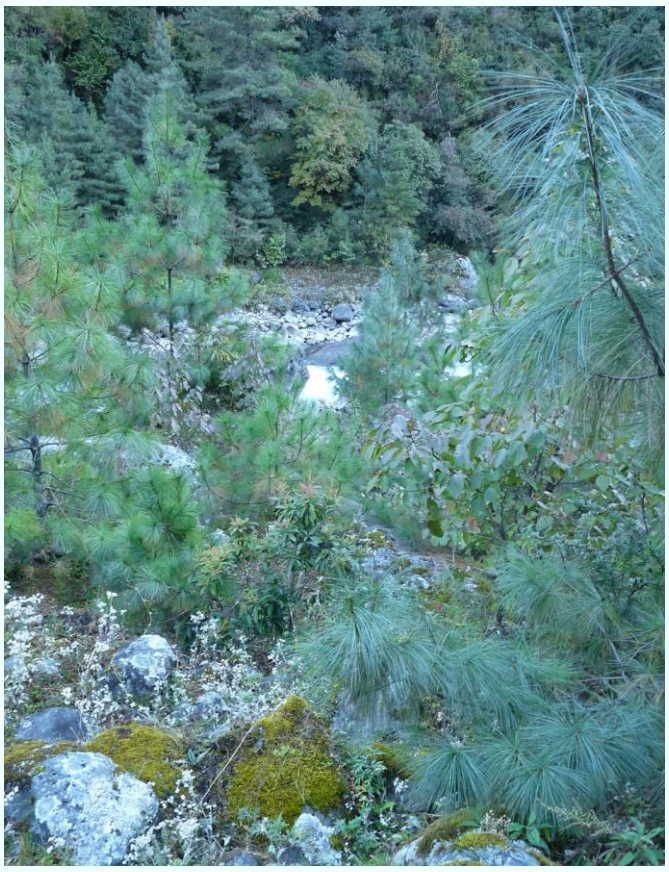


Figure 3.14 2572m. Transect: Sagarmatha. Grid: 4_1_4
Open *Pinus* woodland with rocks, lichens and mosses.

Figure 3.15 2749m. Transect: Pipar. Grid: 1_3_1
Rhododendron arboreum woodland with bamboo, dense leaf litter and pteridophytes.

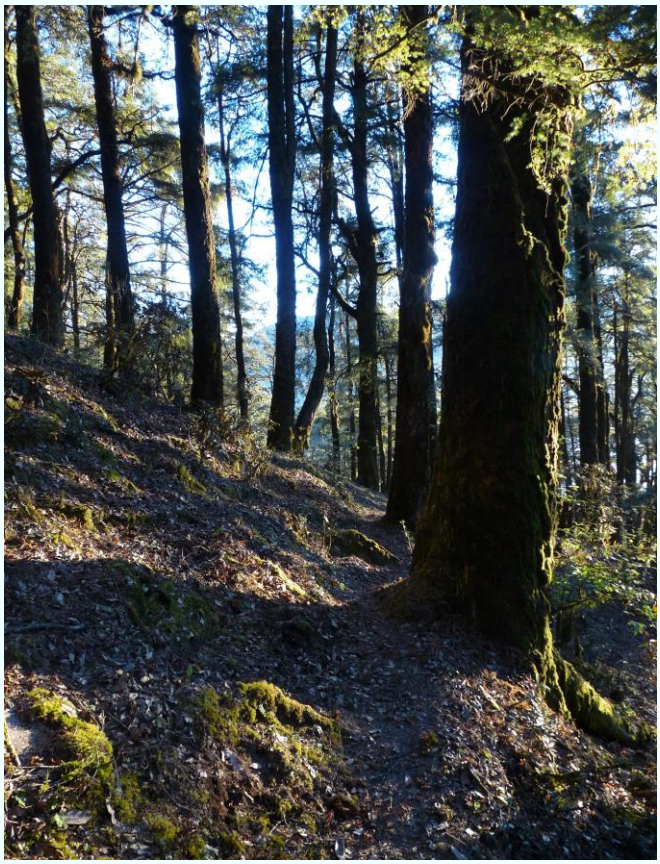


Figure 3.16 3029m. Transect: Langtang. Grid: 3_3_5
Mature *Abies* woodland with *Rhododendron* understorey.

Figure 3.17 3225m. Transect: Annapurna. Grid: 2_4_4
Bamboo scrub and grassland.

Figure 3.18 3284m. Transect: Pipar. Grid: 1_4_4
Grassland with *Berberis* and *Cotoneaster* scrub.

Figure 3.19 3630m. Transect: Langtang. Grids: 3_2_1 to 3_2_6
Xerophytic *Juniperus* woodland on south-facing slope (left) with wet *Abies* woodland on north-facing slope of a steep ridge on the Gosain Kund trail.

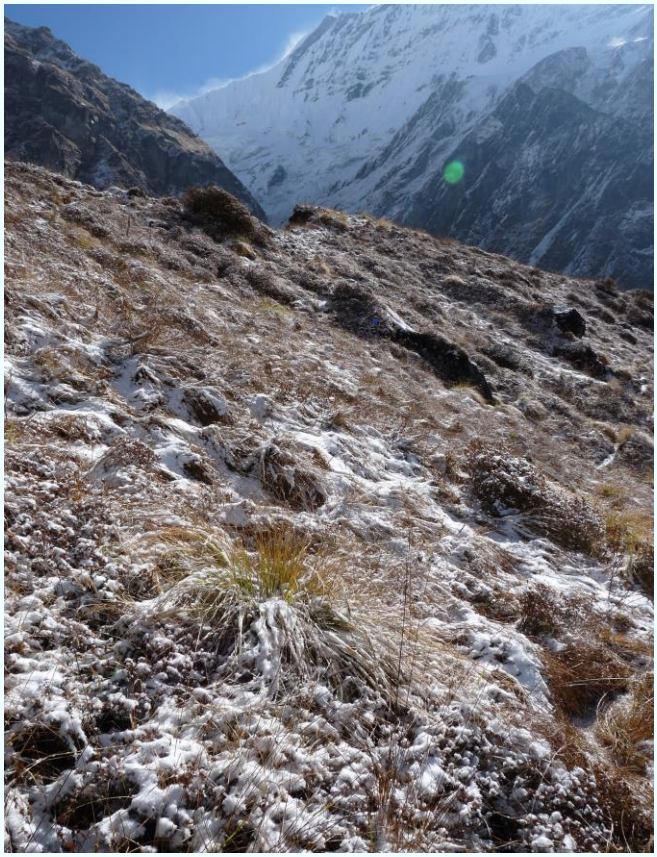


Figure 3.20 3769m. Transect: Annapurna. Grid: 2_5_6
Alpine grassland and dwarf shrubs (*Azalea*).

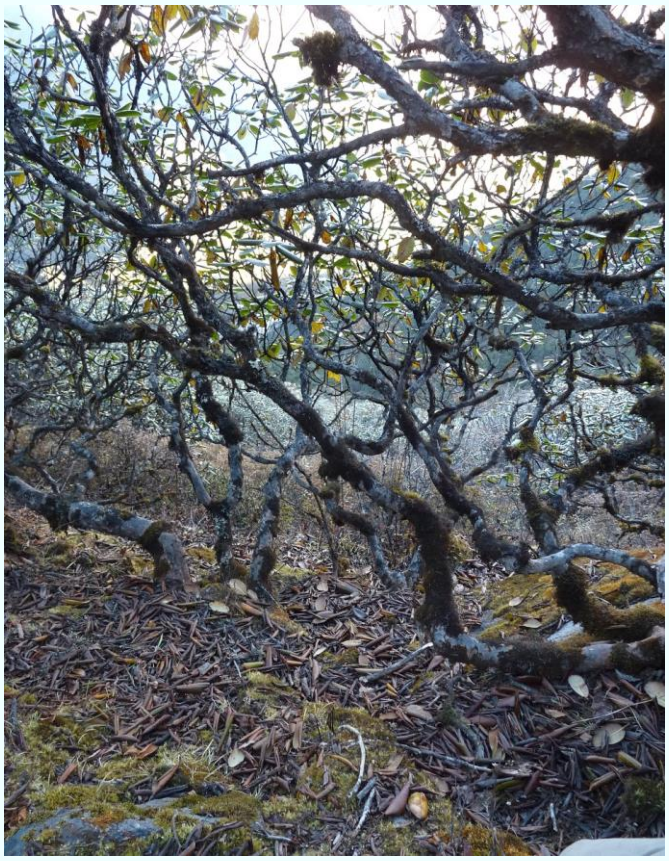


Figure 3.21 3966m. Transect: Sagarmatha. Grid: 3_5_3
Rhododendron campanulatum scrub woodland.

Figure 3.22 4362m. Transect: Sagarmatha. Grid: 4_6_5
Xerophytic dwarf shrubs (*Azalea*, *Juniperis* and *Salix*) in trans-montane habitat.

3.C Results from Running the Genetic Algorithm on Habitat Datasets

3.C.a Structural Variables

Table 3.7 Predictive rules (upper row) and interpreted rules (lower row) for each of ten structural clusters, with numbers of grids and numbers of predictions correctly made by the predictive rules.

Cluster	Ruleset	Within Cluster?	Grids	Correctly Predicted	
1	CanopyHt BETWEEN 3 & 4 AND LitterCover >= 20	Yes	5	5	100.0%
		No	121	121	100.0%
	Very low canopy (3-4m) and litter cover moderate to high	Total	126	126	100.0%
2	(ShrubHt = 3 OR CanopyCover = 10) AND CanopyHt > 3	Yes	28	26	92.9%
		No	98	94	95.9%
	Medium height shrub layer (3m) or canopy cover only 10%, with canopy height low to high (>3m)	Total	126	120	95.2%
3	CanopyHt = 8 AND ShrubHt >= 4	Yes	6	6	100.0%
		No	120	120	100.0%
	Medium height canopy (8m) and high shrub layer (5m)	Total	126	126	100.0%
4	Littercover < 2 AND CanopyHt = 10	Yes	6	5	83.3%
		No	120	120	100.0%
	Medium height canopy (10m) and no litter	Total	126	125	99.2%
5	GroundCover > 75 AND ShrubHt = 0	Yes	8	8	100.0%
		No	118	118	100.0%
	No shrubs and dense ground cover (>75%)	Total	126	126	100.0%
6	CanopyHt < 3 AND (ShrubHt = 1 OR GroundCover < 20)	Yes	14	14	100.0%
		No	112	108	96.4%
	No canopy and (dwarf shrubs (1m) or little ground cover (10 - 15%))	Total	126	122	96.8%
7	Canopycover = 0 AND ShrubHt >= 2 AND RockCover < 10	Yes	15	13	86.7%
		No	111	111	100.0%
	No canopy but shrubs 2m or more high and rocks sparse or absent (<10%)	Total	126	124	98.4%
8	ShrubHt = 2 AND RockCover >= 10	Yes	15	12	80.0%
		No	111	111	100.0%
	Medium shrub height (2m) and rock cover medium to high (>= 10%)	Total	126	123	97.6%
9	ShrubHt = 4 AND ShrubCover >= 15 AND CanopyCover > 25	Yes	12	11	91.7%
		No	114	113	99.1%
	Woodland with dense canopy cover (>25%) and high shrub layer (4m) at moderate to high coverage (>=15%)	Total	126	124	98.4%
10	LitterCover > 25 AND GroundCover >= 25 AND CanopyCover > 10	Yes	17	12	70.6%
		No	109	109	100.0%
	Canopy cover moderate to high (>10%) with litter and ground cover high (>50%)	Total	126	121	96.0%
Total Classifications			1260	1237	98.17%

3.C.b Proximity Variables

Table 3.8 Predictive rules (upper row) and interpreted rules (lower row) for each of eight proximity clusters, with numbers of grids and numbers of predictions correctly made by the predictive rules.

Cluster	Ruleset	Within Cluster?	Grids	Correctly Predicted	
1	Crops = 250	Yes	6	6	100.0%
		No	120	120	100.0%
	Crops more than 200m distant, but not remote	Total	126	126	100.0%
2	Crops < 20 AND Wood < 100 AND Grass <= 50	Yes	6	6	100.0%
		No	120	120	100.0%
	Crops present within the grid, woodland and grassland either present or within 50m	Total	126	126	100.0%
3	Grass >= 500 OR (Human = 500 AND Wood = 0)	Yes	11	11	100.0%
		No	115	112	97.4%
	Grassland remote or (human habitation remote or woodland present within the grid)	Total	126	123	97.6%
4	Scrub = 0 AND (Grass = 200 OR Wood = 100)	Yes	13	13	100.0%
		No	113	113	100.0%
	Scrub present within the grid and (grassland mid-distant or woodland near distant)	Total	126	126	100.0%
5	Grass <= 100 AND Grass > 0 AND Scrub = 50	Yes	14	13	92.9%
		No	112	110	98.2%
	Grassland and scrub nearby but not within the grid	Total	126	123	97.6%
6	Grass <= 20 AND (Human < 150 OR Crops < 200)	Yes	23	21	91.3%
		No	103	101	98.1%
	Grassland very close or present within the grid and (humans or crops within 100 - 150m or present within the grid)	Total	126	122	96.8%
7	(Grass <= 20 AND Crops >= 500) OR Wood > 250	Yes	33	32	97.0%
		No	93	86	92.5%
	(Grassland very close or present within the grid and crops remote) OR woodland remote	Total	126	118	93.7%
8	Scrub > 50 AND Scrub <= 200 AND Human < 250	Yes	20	20	100.0%
		No	106	106	100.0%
	Scrub between 100 and 200m away and human habitation within 200m	Total	126	126	100.0%
Total Classifications			1008	990	98.21%

3.C.c Botanical Variables

Table 3.9 Predictive rules (upper row) and interpreted rules (lower row) for each of seven botanical clusters, with numbers of grids and numbers of predictions correctly made by the predictive rules.

Cluster	Ruleset	Within Cluster?	Grids	Correctly Predicted	
1	Pinus = 1 AND (Artemisia > 0 OR Pieris >=1)	Yes	9	8	88.9%
		No	117	116	99.1%
	Pinus present and Artemisia or Pieris present	Total	126	124	98.4%
2	millet = 1 OR Champa >0 OR Populus <> 0 OR fodder <> 0	Yes	25	16	64.0%
		No	101	98	97.0%
	Either millet, fodder, Populus or "champa" present	Total	126	114	90.5%
3	Rcampanulatum > 0 AND bracken < 1 AND scabious = 0	Yes	23	23	100.0%
		No	103	100	97.1%
	Rhododendron campanulatum present with bracken and scabious absent	Total	126	123	97.6%
4	Betula <> 1 AND (Alchemilla > 0 OR Umbellifer >=1 OR Taraxacum > 0)	Yes	9	8	88.9%
		No	117	117	100.0%
	Betula absent and Taraxacum, Alchemilla or umbellifers present	Total	126	125	99.2%
5	Quercus > 0 AND (Nettles = 0 OR Nagu = 1)	Yes	15	15	100.0%
		No	111	109	98.2%
	Quercus present and either nettles absent or "nagu" present	Total	126	124	98.4%
6	Buckthorn > 0 OR Salix > 0 OR (Azalea = 1 and forbs < 1)	Yes	17	17	100.0%
		No	109	105	96.3%
	Buckthorn or Salix present or (Azalea present and forbs absent)	Total	126	122	96.8%
7	(Forbs > 0 AND Moss = 0) OR (Artemisia > 0 AND nettles = 1)	Yes	28	27	96.4%
		No	98	94	95.9%
	Forbs present and moss absent or Artemisia and nettles present	Total	126	121	96.0%
Total Classifications			882	853	96.7%

Chapter 4 The Use of Two Small Mammal Reference Collections to Predict Species Distributions in Nepal

Abstract

Museum collections provide a valuable, independent, historical and geographical context for ecological studies. Here, I present two datasets; the small mammal collection of the Bombay Natural History Society and data derived from a monograph of the small mammals of Nepal. The former comprised over 7200 specimens of small mammals from the whole Indian sub-continent, mostly collected 100 years ago. The monograph complemented these data with 1000 records from a variety of sources, specifically from Nepal and generally collected in the second half of the 20th century. These unique resources are used to compare the taxonomic composition of the current study with historical field surveys, both geographically and temporally. I also analyse the geographical representativeness of the source data and give an insight into collector bias in species identification. The source datasets are used in combination with the Landsat 30m DEM of the Indian sub-continent to derive predictive models for 56 species of small mammals. Logistic regression and Maxent models are generated for each species and then projected onto the DEM for Nepal. The two modelling processes are compared, showing that Maxent almost always predicts a greater probability of occurrence, but logistic regression models tend to resolve less frequently. The projections are used specifically on the ten 0.1° grid squares encompassing the field sites of the current study, to predict likely presence of the putative species described in Chapter 2. These are carried forward to Chapter 6 to inform the final species identities.

Contents

Tables

Figures

4.1 Introduction

The Bombay Natural History Society (BNHS) was founded in 1883 and is amongst the premier conservation NGOs in India. Its headquarters and collections are located at Hornbill House in Mumbai, from where it publishes The Journal of the Bombay Natural History Society. World renowned ecologists and conservationists who were involved with the society include Sálim Ali, S. Dillon Ripley and S. H. Prater. In 1911 the BNHS began a survey of the mammals of India under the leadership of R. C. Wroughton, involving many of its members and other volunteers. This was thought to be the first collaborative biodiversity study in the world (Wroughton, 1912). Over 12 years contributors to the project collected 50,000 specimens, many of which are still available in the society's collection. The survey began work in Nepal in 1920 and carried out three expeditions until 1923. This chapter is based, in part, on data from a subset of 7285 small mammals specimens, which are in the process of being electronically catalogued (Bajaru, pers. comm.). The dataset is a unique resource, with records from a huge collecting area of over 5,245,000 km^2, representing the whole of British India, Nepal and Bhutan. The majority of specimens were collected approximately 100 years ago (1913, 1914 and 1915), exactly one century before my own study.

The other primary source of data for this chapter is the eloquent monograph by Pearch (2011) His history of mammalogy in Nepal describes the central position of Brian Houghton Hodgson (1800-1894) during the 21 years he spent in Nepal. During this time, he collated 373 mammal specimens, of which 40 were described as new species. The history of small mammal collecting expeditions since then is described in detail in this monograph so will not be repeated here. In summary, the data complement those from the BNHS collection in space and time. The majority of the BNHS records were collected between 1912 and 1923, and from the whole of British India, covering the full extent of seven modern countries; Bangladesh, Bhutan, India, Myanmar, Nepal, Pakistan and Sri Lanka. In contrast, the records obtained from the Pearch monograph, were only collected in Nepal, generally between 1950 and 2000, although a few records may refer to specimens from the BNHS surveys in 1920s, held in the British Museum (NH).

Museum specimens are being recognised as a valuable source of data to explore evolutionary and ecological themes. For example, Pergams and Lawler (2009) investigated morphological change in rodent species from 28 museum collections. Holmes *et al.* (2016) advocated the use of museum specimens as sources for comparative studies, *e.g.* small mammal communities in Yosemite National Park from the 1920s compared to the present day (Moritz *et al.*, 2008). Arnold *et al.* (2017) was able to extract DNA from feathers taken from museum specimens of cryptic songbirds from Arizona to aid identification of recent field specimens.

The general method of analysis that I have employed in this chapter is known as Species Distribution Modelling (SDM). These have been widely used since the 1990s (Guisan and Zimmermann, 2000) and developed enormously in the last 10 – 15 years, with the advent of powerful desktop computers and statistical and GIS software. Many broad types of SDM have been advocated, including parametric statistical modelling, such as logistic regression (GLM) and generalized additive models (GAM), genetic algorithms and maximum entropy models. The former were used by Pearce and Ferrier (2000) to model vascular plant, bird and reptile species distributions in New South Wales, Australia. Anderson *et al.* (2002) and Anderson (2003) used the Genetic Algorithm for Rule-set Prediction (GARP) (Stockwell and David, 1999) in their empirical study of the distribution of *Oryzomys* and *Heteromys* rodent species in South America. In the first decade of this century Phillips *et al.* (2004), Phillips *et al.* (2006) and (Phillips *et al.*, 2004), *etc.* developed Maxent, which has now been widely used and calibrated. It has been employed to model, *e.g.*, the distribution of plant species richness on elevational gradients in the Andes (Mateo *et al.*, 2012) and the tsetse fly in Gambia (Dicko *et al.*, 2014).

In this study, I have concentrated on the use of logistic regression and Maxent models. I used the two source datasets to develop predictive models for 56 species and then projected them onto a digital elevation map (DEM) of Nepal to give probabilistic distributions. I assessed the quality of the models and compared their predictions for each species. Finally, I have focused the results of the predictive models on the locations that I used for fieldwork in the current study (Chapter 2). This enabled me to confirm species identifications based on their likely predicted presence.

4.1.1 Aims and Objectives

The main aim of reviewing these two datasets was to provide an independent taxonomic and ecological framework for the current survey. By using them it was possible to obtain a historic perspective through collections made up to 100 years ago, plus a wider geographic view than would be possible from a single three-year study.

There were six main objectives for analysing these datasets;

- To collate the small mammal species present in the collections, relevant to the current study, and to explore their taxonomic identities and relative abundance.
- To explore the spatial and topographical distribution of collecting effort within the geographical boundary encompassed by the datasets, to determine their representativeness.

- To use the BNHS collection records to analyse a) the historical patterns of specimen collection during the 20th century and b) the potential collector bias in species identifications.
- To combine the two reference datasets with a publicly available digital elevation model (DEM) to build predictive models of species probability distributions.
- To project these models onto the DEM of Nepal to generate probabilistic distribution maps for relevant small mammal species in this country.
- To extract the predictions for the grid squares where the current study was located to seek confirmation of the species identifications made in the field, from the likely presence of species within those squares.

4.2 Methods

4.2.1 Data Sources and Data Manipulation

The raw data sources required considerable manipulation to make them suitable for quantitative analysis. Once the required data had been extracted from each source, they were stored in a single Microsoft SQL Server (Anon., 2018b) database for further manipulation.

4.2.1.1 The BNHS Reference Collection

The primary data source was a Microsoft Excel spreadsheet of the complete mammal collection of the BNHS (Bajaru, pers. comm.), holding 13,455 records. This was compiled from the card index of all specimens, over a period of several years, by Sameer Bajaru and volunteers at the society's head-quarters. In turn, the card index was compiled over many years from the original labels attached to the specimens in the collection. These were written in the field (or soon after) by the collecting team, but many had subsequent, uncatalogued amendments, such as taxonomic alterations and lo-cality corrections. Furthermore, as many of the small mammal specimens were approximately 100 years old, the state of the labels was not good. The consequence of these issues, especially the double transcription, meant that the data in the spreadsheet required considerable validation and verifica-tion before analysis.

The columns in the spreadsheet fell into six logical groups:

- Taxonomic information comprising OLD SCIENTIFIC NAMES (recorded by the original col-lector) and NEW SCIENTIFIC NAMES (the final taxon derived from all corrections).
- Date of collection – often just a month or year.
- Collector details.
- Geographical information, which was generally just a descriptive locality, State / Country and altitude, usually recorded in feet.
- Specimen details including morphological data, sex and whether skins and skulls were collected.
- Collection information, such as cabinet and draw numbers.

The old and new scientific names were an attempt by BNHS to rationalise the original names given by the collectors into a more recent taxonomy. There was no indication when this was done, or by whom. I have made no attempt to verify the new names, but they were various combinations of synonyms, sub-genera and sub-species. From these I extracted a unique species list of 372 names, 64 of which belonged to the families Cricetidae, Muridae and Soricidae.

Dates were rationalised into three fields in the database; Day, Month and Year.

There were 845 unique collector values. However, there was a massive amount of redundancy due to spelling variations in the name and initials, as well as basic transcription errors from the original specimen labels. I rationalised a list of 107 collectors involved in the collection of the three relevant families of small mammals. These were based only on the first collector on a label and excluded generic records such as "BNHS Collection" or "Supdt. Victoria Garden".

Geographical information was the most difficult to rationalise as it was almost entirely descriptive. Firstly, I extracted 1196 unique location names. Only five of these actually had latitude and longitude values, although 267 had altitudes. I used a combination of Google (https://www.google.co.uk/maps/) and Bing Maps (https://www.bing.com/maps/), plus other on-line applications to geo-reference the remaining records. I was able to supply 1028 records with latitude and longitude to the nearest 0.1°, giving 1033 records in total. However, many of the original locations were synonyms and typographic errors, so I aggregated these into 385 unique grid squares, based on 0.1° of latitude and longitude. These represented approximate rectangles with areas of 110km^2 (11km N–S × 10km E–W). In a similar way I was able to apply approximate altitudes to 790 records, giving a total of 1057 records with altitude data. I aggregated these to the nearest 10m to give 204 unique altitudes from sea level to 3930m.

The final stage of data rationalisation was to extract a 7285 × 10 matrix of specimen records by taxonomic geographic, temporal and collector variables.

4.2.1.2 The Pearch Monograph

The primary data source was a published monograph (Pearch, 2011) provided to me as a 286-page pdf file by the author. This was a review of all non-volant small mammal species recorded for Nepal from all available online museum databases via the Mammal Networked Information System (Wieczorek, 2015). Most importantly, it did not include any data from the BNHS collection because their catalogue was not published online. The major sources included in this review were the British Museum (Natural History), the American Museum of Natural History, the Field Museum Chicago, the Smithsonian Natural History Museum and the Royal Ontario Museum Canada, plus a further 14 museums and collections.

Extracting quantitative data was a laborious task, requiring several intermediate stages using MS Word and MS Excel. Fortunately, the document was well structured and the author had already rationalised taxonomies and localities of the records. The species of interest were extracted from

the contents page and used to reference individual species accounts. Each species account contained a list of localities × sources (collectors) which were extracted to the database. All localities extracted in this way were then linked to the gazetteer compiled by the author. This contained 516 records which either had explicit latitudes and longitudes or were referenced to other explicit records. These were recorded as degrees and minutes, which enabled conversion to degrees with two decimal places. Of the remaining 81 records, I geo-referenced 16, using the same methods as for the BNHS dataset, giving a total of 532 geo-referenced localities. Similarly, 463 records had explicit altitudes (or ranges) or were cross-referenced to those that did. A further 52 records had altitudes obtained from on-line sources, giving a total of 515 localities with altitudes accurate to 10m resolution.

The final dataset comprised 966 unique taxon / collector / locality records, where a locality was represented by a 0.1° grid square with a 100m altitudinal zone, known throughout the remainder of this chapter as a "geo-cell". It must be emphasised that these were not individual specimens, rather they were records of one or more specimens, belonging to a specific species, from an individual collection, from a particular geo-cell.

4.2.1.3 The DEM

The USGS GTOPO30 digital elevation map was compiled at a resolution of 30 seconds of arc. This was divided into 27 "tiles" each covering 50 degrees of latitude by 40 degrees of longitude. Fortunately, a single tile (file E060N40.DEM) completely covered the Indian sub-continent from 60°E – 100°E by 10°S – 40°N. The file comprised a matrix of 6000 rows by 4800 columns, representing altitude (in metres) recorded at roughly 1km resolution.

The open-source software package QGIS v2.13 (Anon., 2018a) was used to convert the DEM format file into an ASCII matrix file and imported to SQL Server. This was unravelled into a linear file, excluding records with null altitudes (sea), giving 12,501,969 records. These were then aggregated into 71,473 0.1° grid squares. Consequently, all entirely land-based squares comprised 144 elevation values, with coastal squares having any number of values from 1 to 143.

Digital boundary maps were obtained from the internet for the seven extant countries comprising the Indian sub-continent. These are Bangladesh, Bhutan, India, Myanmar, Nepal, Pakistan and Sri Lanka. I used QGIS to convert these to ASCII format and imported them to the SQL Server database. By utilising the geography data types and functions, I selected all the 0.1° grid squares from the DEM that were wholly or partially within the seven boundaries, giving a subset of 46,500 squares.

Using SQL Server's set-based logic, I was able to calculate a number of aggregate statistics for each grid square – firstly, four altitudinally based statistics (plus N);

- N The number of raster points in the grid square (Inland squares always have 144).
- Alt The mean altitude
- AltMin The minimum altitude
- AltMax The maximum altitude
- AltV The coefficient-of-variation of the altitudes in the grid square.

I also developed an algorithm to calculate slope and aspect statistics for each grid square, with the following logical code;

- Loop through all the 0.1° grid squares in the seven countries.
- Each inland grid square had 10,296 possible pairwise combinations of raster points, with coastal points containing fewer. For each square, select a sample of 200 random pairs of raster points (with replacement) and loop through them.
- For each pair, use basic trigonometry to calculate the horizontal (latitude & longitude) and vertical differences and store these three statistics in a table variable.
- From the sample of 200 statistics, calculate the mean slope, coefficient-of-variation (CoV) of slope and the mean aspect.
- Store these in the 0.1° square record.

The calculation of the mean and CoV of the slopes was straightforward, being the arithmetic mean and the standard deviation divided by the mean of the 200 random slopes. However, calculation of aspect was more complex as the orientation of the slopes is radial rather than linear. Although Zar (1984) gives a number of equations for calculating the "sample mean" aspect, set-based logic was much easier and quicker to use for the 46,500 grid squares. I used the signed differences of eastings and northings for each pair of points, to give a mean easting and northing relevant to an arbitrary point zero. I then used the arc-tan of the ratio of the two means to give the mean angular direction of the slope, taking account of the quadrant defined by the signs of eastings and northings. From this I also calculated the octant (the zones encompassing the eight points of the compass) for the grid square as a whole.

I tested the accuracy and efficacy of this algorithm on data from a random square; 28.0N, 86.0E (Figure 4.1). This shows a general decline in altitude from east to west, although a high ridge runs in a north-westerly direction from the south-eastern corner, which creates slopes in a north-easterly and easterly direction. I ran ten trials each on all combinations of three different random sample sizes (200, 100 & 50) and three different proportions of the sample (100%, top 50% and top 25% of slopes) to calculate aspect (Table 4.1). The mean aspect calculated from these trials was very consistent across the different parameters, ranging from 268° to 274°, which confirms the general pattern of west-facing slope shown in the map. However, decreasing sample size caused an increase in the CoV and range of aspects. But, interestingly, decreasing the proportion of the random sample, generally reduced the CoV and range of aspects. Subsequent data exploration showed that the statistics based on the top 25% of the sample had better distribution and less cross-correlation with the other topographical variables. For this reason I have calculated slopes and aspects from a sample of 200 (as the penalty for the increased calculations was low), using the top 25% to calculate the final slope and aspect statistics.

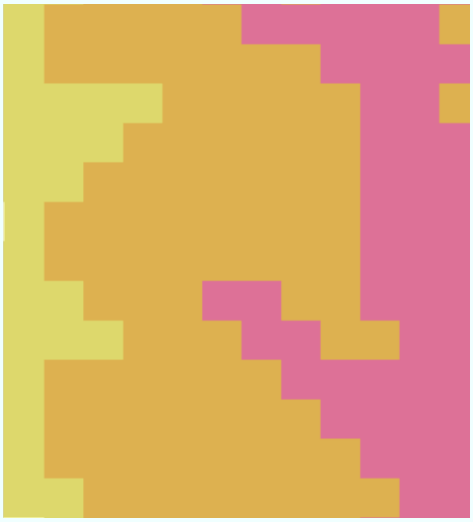


Figure 4.1 Grid square 28.0N 86.0E showing 144 pixels colour-coded into 1000m elevations (yellow 3000 – 4000, orange 4000 – 5000, pink > 5000m.

Table 4.1 The aspect in degrees from 10 trials of every combination of three random sample sizes and three percentages of the sample, used in calculations of aspect for grid square 28.0E 86.0N.

Sample Size	200			100			50		
Top %	100	50	25	100	50	25	100	50	25
Trial									
1	276	273	266	277	263	267	271	281	273
2	274	265	276	268	266	274	263	267	265
3	270	276	273	276	276	277	282	274	270
4	260	268	275	266	260	264	286	254	282
5	271	271	268	276	264	267	276	283	268
6	273	262	272	269	276	276	285	263	273
7	262	266	269	260	271	267	251	275	277
8	280	271	273	281	271	269	273	284	269
9	268	276	269	262	263	275	245	275	248
10	269	265	273	271	271	274	263	282	252
Mean	270.3	269.3	271.4	270.6	268.1	271	269.5	273.8	267.7
SD	6.1	4.9	3.2	6.9	5.7	4.7	14.0	9.8	10.5
CoV	2.2%	1.8%	1.2%	2.5%	2.1%	1.7%	5.2%	3.3%	3.9%
Range	20	14	10	21	16	13	41	30	34

4.2.2 Generation of Predictive Models from the Source Datasets

The data from the two reference collections were combined with the DEM at a 0.1° resolution. The analysis proceeded in four stages;

- The datasets were used to generate predictive models based on individual species distributions for the whole of the Indian sub-continent.
- The models were then projected onto the DEM of Nepal to give maps of predicted probabilities of occurrence for each species.
- These were then used to generate species richness maps at different thresholds of probability.
- Finally, the predictions were used to investigate likely probabilities of occurrence of the species identified from this study (see Chapter 2) in the grid squares where the field work took place.

Models were built separately for each of the 56 species listed in the Supplementary Information Section 4.A, that were present in at least three grid squares. Eight predictor variables were used in the models – seven continuous and one categorical;

- Latitude and Longitude
- Altitude, Co-efficient-of-variation of Altitude and Altitudinal Range
- Slope and Co-efficient-of-variation of Slope
- Aspect (octants).

Two model-building procedures were used; logistic regression, as implemented in Statistica v13 Anon. (2017), and Maxent (Phillips *et al.*, 2004). The logistic regression models used the presence or absence in the grid square as the binary response variable. The model structure had a second-order surface design, which included all eight main effects, seven squared effects for the linear variables, and all two-way interactions of the linear variables. This gave a total of 36 terms in the model. These were analysed using a backwards step-wise approach with *p*-to-enter and *p*-to-leave both set to 0.01, to give a conservative model. The coefficients from the model were extracted and stored in the SQL server database.

The Maxent models used two datasets. The samples were the set of grid references for each square in which a species was present, which may have had as few as three records. The environmental sample was the same topographical dataset as used in the logistic regressions. The model used the logistic output option to generate predicted probabilities of occurrence. The input features were linear, quadratic and products, which equated to the same set of predictors as used in the logistic

regression. All other Maxent settings were applied as defaults. I conducted a series of runs with a regularization multiplier value of 5 (rather than the default of 1) as advised by Merow *et al.* (2013). This reduced the AUC for all species, and the maps displayed severely reduced sensitivity, so the default value was employed for all.

4.3 Results

4.3.1 Taxonomic Breakdown of the Reference Datasets

Data for three families of small mammals were extracted from the two source datasets. In total, there were 7285 specimen records in the BNHS dataset and 966 geo-cells (see Section 4.2.1.2) in the Pearch dataset (Table 4.2a). These were collated into presence in 0.1° grid squares, giving 375 and 143 for the two sources respectively (Table 4.2b). When the source datasets were combined, the total number of grid squares was 510, indicating that eight grid squares were common to both datasets.

Table 4.2 a) Counts of BNHS specimens and Pearch geo-cells by family and genus. b) Tallies of the presence in 0.1° grid squares of BNHS specimens, Pearch records and both datasets combined by family and genus.

		a) Data Records		b) Tallies of Presence in 0.1° Squares		
		BNHS Specimens	Pearch Geo-cells	BNHS	Pearch	Combined
Cricetidae	*Alticola*	40	11	7	8	15
	Clethrionomys	3	0	0	0	0
	Cremnomys	154	0	18	0	18
	Cricetulus	148	0	1	0	1
	Eothenomys	5	0	1	0	1
	Microtus	15	0	0	0	0
	Neodon	14	67	3	32	35
	Total	**379**	**78**	**28**	**37**	**65**
Muridae	*Apodemus*	261	59	23	25	46
	Bandicota	618	0	108	0	108
	Berylmys	44	0	5	0	5
	Diomys	0	1	0	1	1
	Golunda	206	3	40	2	42
	Leopoldamys	15	0	11	0	11
	Maxomys	146	0	16	0	16
	Millardia	0	6	0	5	5
	Mus	2500	146	191	81	272
	Nesokia	50	4	8	3	11
	Niviventer	314	146	36	56	90
	Rattus	1861	190	200	82	277
	Vandeleuria	0	4	0	3	3
	Total	**6015**	**559**	**340**	**128**	**460**
Soricidae	*Anourosorex*	49	0	5	0	5
	Chimarrogale	2	2	2	2	4
	Crocidura	91	8	27	6	33
	Episoriculus	43	117	6	42	47
	Nectogale	10	2	4	2	6
	Sorex	0	29	0	20	20
	Soriculus	51	106	11	45	55
	Suncus	645	65	130	40	170
	Total	**891**	**329**	**158**	**93**	**249**
Total		**7285**	**966**	**375**	**143**	**510**

The BNHS specimen records were dominated by Muridae (approx. 83%), followed by Soricidae (12%), with only 5% of the specimens belonging to Cricetidae. This pattern was repeated in the Pearch dataset, although the proportions were less extreme (58%, 34% and 8% respectively). These data have been further broken down into genera, showing the huge dominance of two genera (*Mus* and *Rattus*) in the BNHS dataset – these represent 73% of murid records and 60% of all records in the dataset. Two genera also dominate the cricetid records comprising 80% of all specimens for that family. The Soricidae are dominated by one genus (*Suncus*), comprising 72% of the family records.

A further breakdown to species level is provided in Supplementary Information 4.A. This shows a total of 73 species across both datasets, with 62 in the BNHS dataset and 40 in the Pearch dataset. Only 29 species were common to both. The two dominant species in the BNHS dataset were *Mus musculus* (20% of specimens) and *Rattus tanezumi* (13%). Three other species, *Bandicota bengalensis*, *Mus booduga* and *Suncus murinus* each comprised between 7% and 8% of all specimens. The most frequently recorded species in the Pearch dataset was *Soriculus nigrescens*, with 11% of geocells, followed by *Episoriculus caudatus* (9%) and *Mus musculus* (8%).

4.3.2 Analysis of Collecting Locations

The DEM provided 46,500 0.1° grid squares within the boundary of the seven countries. Of these, 375 were used for collection of the BNHS specimens (Figure 4.2), an overall rate of 0.8%. For the purposes of this analysis each square had five variables; two exact values (latitude & longitude) and three derived values (altitude, slope & aspect) as described in Sub-section 4.2.1.3. The null hypothesis tested was that the distributions of the grid squares visited by the BNHS collectors was not significantly different from the overall distributions of the grid squares within the five-dimensional space defined by these variables. In other words, the spatial distribution of the collecting locations was not biased. There was no assumption in this hypothesis of a linear relationship between the proportion of squares visited and the predictor variables, so they have all been coded into convenient categories (Table 4.3).

Table 4.3 Categorisation of continuous predictor variables for the analysis of representativeness of BNHS grid squares.

Variable	Category Description	N
Latitude	3° intervals from <9°N to >=33°N	10
Longitude	3° intervals from <66°E to >=96°E	11
Altitude	<100m, <500m and then 500m intervals to >5000m	12
Slope	<0.5°, <1°, <2°, <5°, <10°, <15° & > 15°	7
Aspect	Octants	8

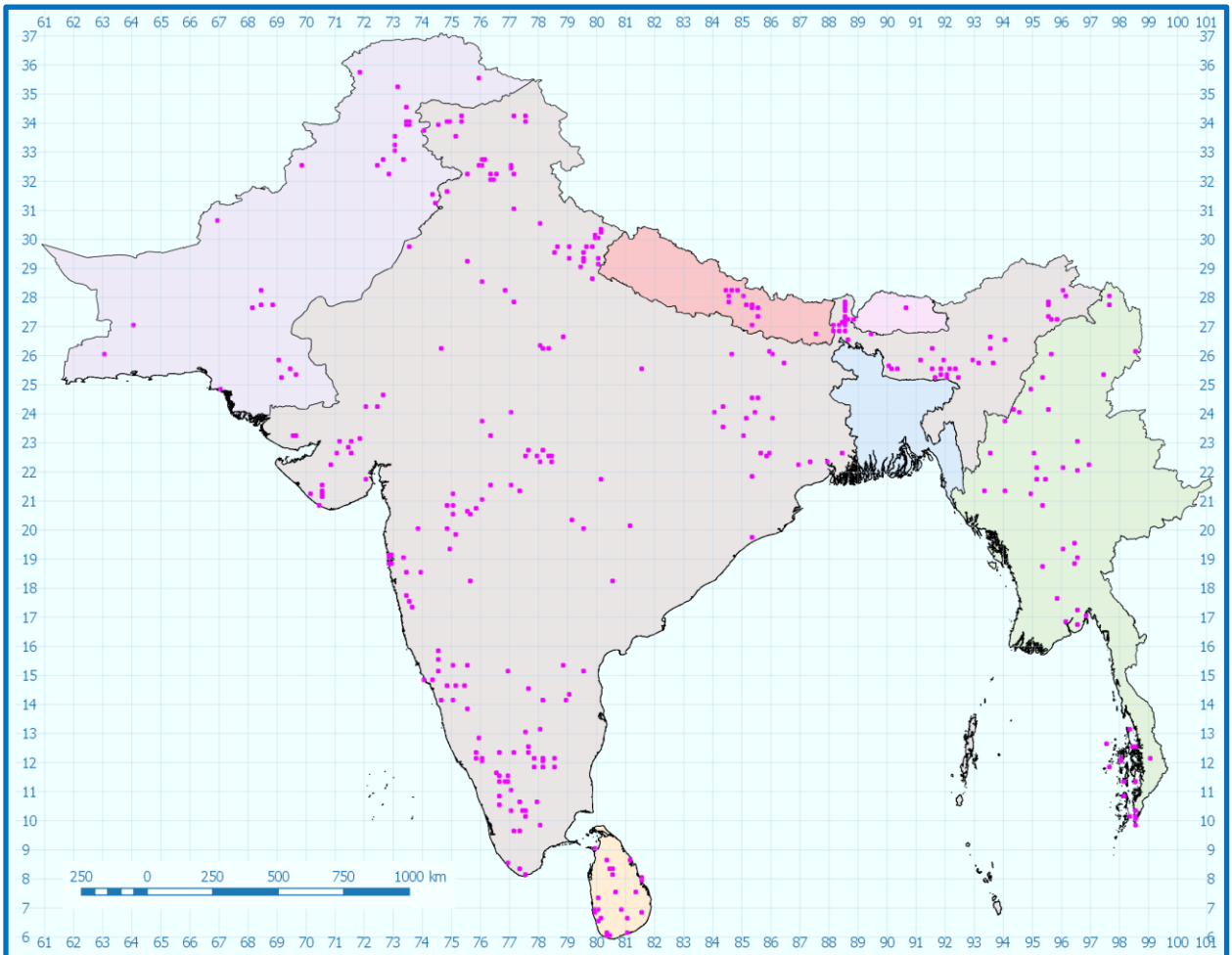


Figure 4.2 The locations of the 375 0.1° grid squares visited in the BNHS collection within the Indian sub-continent. The seven existing countries in this region are colour coded.

A five-way ANOVA model was constructed using a generalized linear model with a binomial response variable and logit link-function. Only the main effects were included and Type III log-likelihood sums-of-squares were used for significance testing. Three of the five predictor variables were very highly significant (Table 4.4), indicating that there were significant differences in the proportions of squares visited at different latitudes, longitudes and altitudes (Figure 4.3). There was a distinct bias towards the south of the region, with significantly more grid squares being covered in the southern tip of India and Sri Lanka, but there was also a cluster in islands of southern Myanmar. There was a very low proportion of grid squares visited in the far west of Pakistan, but also in the zone between 81°E and 84°E, which corresponds to the western half of Nepal and a large block of central India around the modern states of Uttar Pradesh, Chhattisgarh and Odisha. However, the most important ecological effect was the higher proportion of squares visited at mid altitudes, between approx. 1000 and 2500m.

Table 4.4 Five-Way logistic ANOVA for proportions of 0.1° grid squares used for collection by the BNHS.

Effect	DF	Log-Likelihood	$\chi 2$	p
Latitude	9	-2084.20	70.64	$<10^{-10}$
Longitude	10	-2088.46	79.15	$<10^{-10}$
Altitude	11	-2070.46	43.15	$<10^{-10}$
Slope	6	-2055.35	12.93	0.0442
Aspect	7	-2056.81	15.86	0.0264

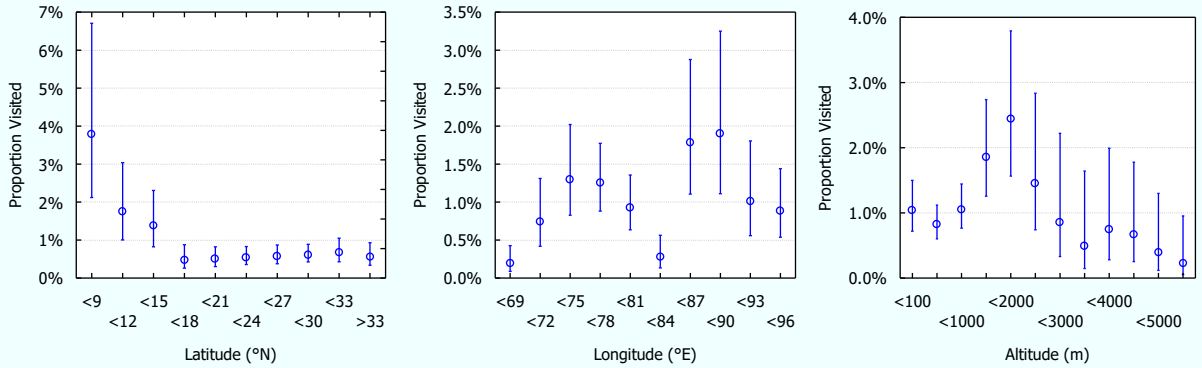


Figure 4.3 Means plots (+ 95% C. I.) for three significant predictor variables from a logistic ANOVA of proportion of grid squares visited by BNHS for collection.

There were 1529 0.1° grid squares in the DEM dataset within the boundaries of Nepal. Of these, 143 squares (9.4%) had records in the Pearch dataset (Figure 4.4), representing over ten times the recording intensity of the BNHS survey. The same analysis was carried out using the five main effects in a logistic regression model, although the categorisation for latitude and longitude was 0.5° and 1° intervals respectively. The model showed similar results, with latitude, longitude and altitude effects all being highly significant, with *p*-values of $<10^{-4}$, $<10^{-7}$ and $<10^{-6}$ respectively (Figure 4.5). Slope and aspect were not significant. There was a significantly greater proportion of grid squares visited between 27.5°N and 28°N, which corresponds to the Kathmandu valley and the Gauri Shankar Conservation Area, and the Sagarmatha and Makulu Barun National Parks. In contrast, squares north of 29°N were significantly less likely to have been visited. The strongest longitudinal effect was a general increase in proportion of squares visited towards the east of the country, although the zone between 85°E and 86°E, corresponding to the Kathmandu valley and Langtang National Park, had the highest proportion of squares visited. The most important altitudinal effect was that the mid-range "hump" was at a considerably higher altitude than for the BNHS dataset – around 3000m to 4000m.

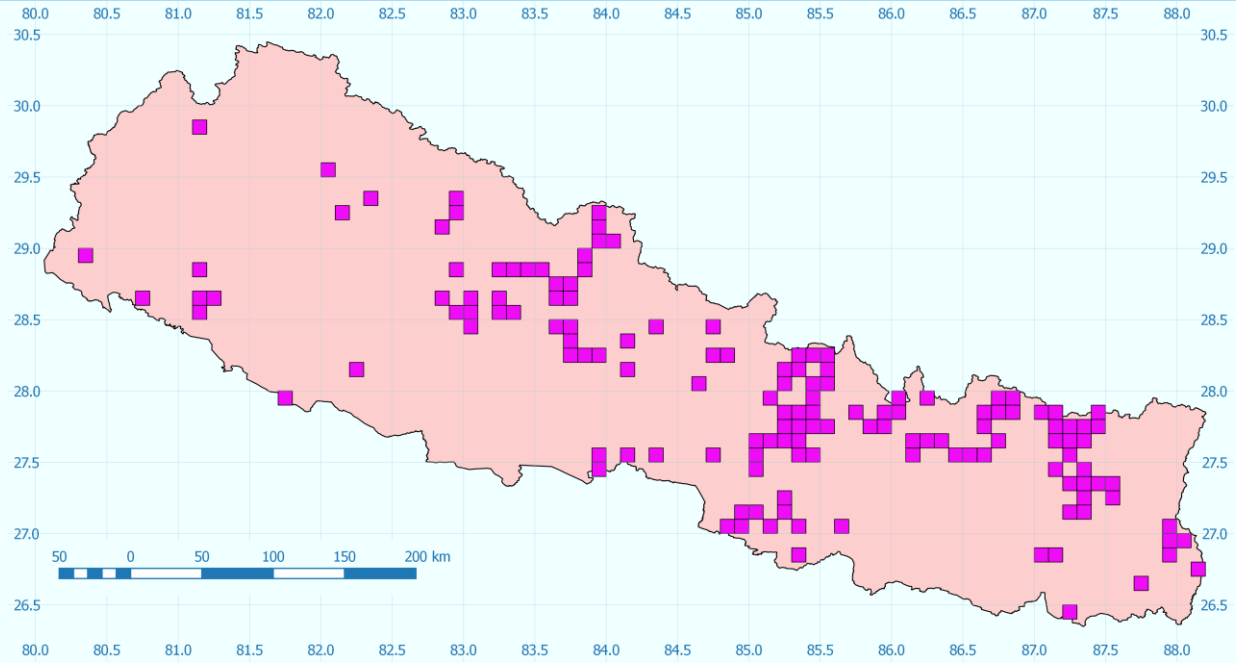


Figure 4.4 Location of the 143 0.1° grid squares within Nepal from the Pearch dataset.

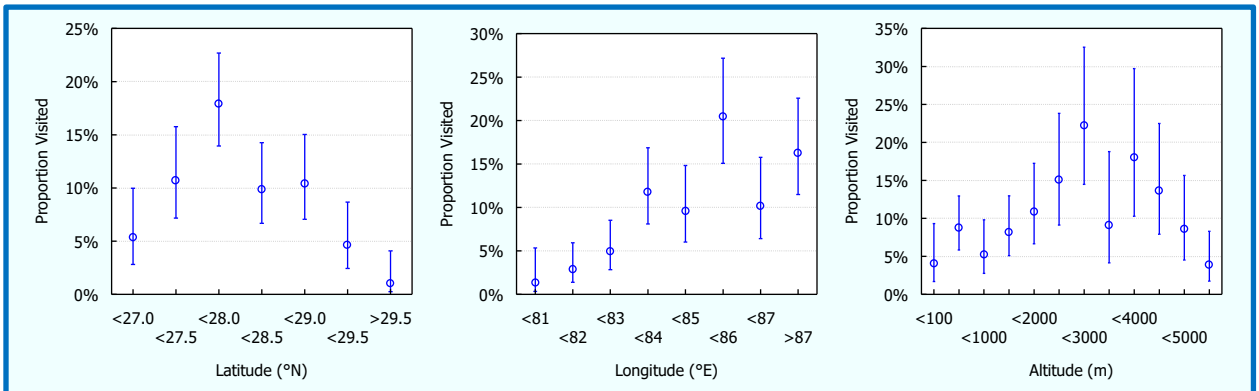


Figure 4.5 Means plots (+ 95% C. I.) for three significant predictor variables from a logistic ANOVA of proportion of grid squares recorded in the Pearch monograph.

4.3.3 Analysis of Collector Effects in the BNHS Dataset

As discussed in Chapter 2, accurate identification to species level is often difficult in the field. In particular, the genera *Apodemus*, *Mus* and *Rattus* often require detailed anatomical inspection of skulls and dentition. This analysis attempted to quantify the degree of variation in species identifications between the major collectors of the BNHS specimens.

Of the 7285 specimen records, 7010 had collector information, from which 106 unique collectors were identified. However, 6198 of these records (88.4%) were recorded by only ten collectors and three of them collected 4068 specimens (58%). These data were used to analyse potential collector biases in the identification of ubiquitous species – the full analysis is presented in Supplementary Information 4.B.

In summary, three analyses were carried out, one for each of the three genera indicated above. *Apodemus* had specimens belonging to four species, but only one of them (*A. rusiges*) was recorded in sufficient number (224) for analysis. There was a marginally significant collector effect for this species ($p < 0.02$), with one collector recording only this species from the genus, another never recording this species and a third recording this and other con-generics. *Rattus* had four species which were recorded frequently by nine collectors. Significant preferential identification for three of the species was shown, especially for *R. pyctoris*, ($p < 10^{-8}$) and *R. tanezumi*; ($p < 0.0005$), although not for *R. nitidus*, the species recorded in this study. Finally, three species of *Mus* were collected in large numbers, two of which showed significant collector effects; the species recorded in the current study, *M. booduga* ($p < 10^{-5}$) and *M. musculus* ($p < 0{\cdot}0002$).

4.3.4 Generation of Predictive Models from the Source Datasets

Predictive models were run on 56 species, of which 42 (75%) generated successful models from both procedures. Thirteen of the logistic regression models failed to resolve. In five of these cases the step-wise procedures removed all effects leaving an intercept-only model which, therefore, predicted a constant value. The other seven only included terms with longitude and/or latitude, which meant that they tended to predict constant gradients in one or two dimensions. Seven Maxent models were considered to be over-parameterized, with variables being forced into the model that had little or no predictive power. The models were compared and contrasted in two ways; a) goodness-of-fit and b) variable contribution.

4.3.4.1 Goodness-of-fit of the Models

The goodness-of-fit of models were evaluated using the Area under the ROC curve (AUC) following the protocol of Dicko *et al.* (2014). A high value of AUC (approaching 1) indicated a greater departure from the null model and, therefore, a better fit. Overall, the performance of each model type was very similar (Figure 4.6). Logistic regression models from 27 species had a greater AUC than the equivalent maxent models, with the converse being true for 14 species. However, there was a marginally significant difference between these contrasts in the three families. Logistic regression models had greater AUC for 64% of murid species, whereas the proportions for Cricetidae and Soricidae were 40% and 24% respectively (Contingency $\chi^2_{(2)} = 7.404$, $p < 0.025$).

The relationship between AUC and the number of grid squares in which the species was recorded showed two significant effects;

- Firstly, excluding the null models, there was a significant negative correlation ($r = -0.781$, $p \approx 0$; Figure 4.7a). For species present in fewer than 20 grid squares, the models generally

achieved an AUC of > 0.9, whereas the two species found in more than 140 grid squares had AUCs from both models below 0.75.

- For each species in which both models resolved, I calculated the ratio between the two AUCs such that when this value was > 1, the logistic regression model had a higher AUC than the maxent model. There was a significant positive relationship between the number of grid squares in which a species was recorded and the AUC ratios ($r = 0.539$, $p < 0.0005$; Figure 4.7b). Every species that was recorded in more than 10 grid squares (with one exception) had a ratio >1. In contrast, of the 22 species recorded in fewer than 10 grid squares, 20 had a ratio < 1.

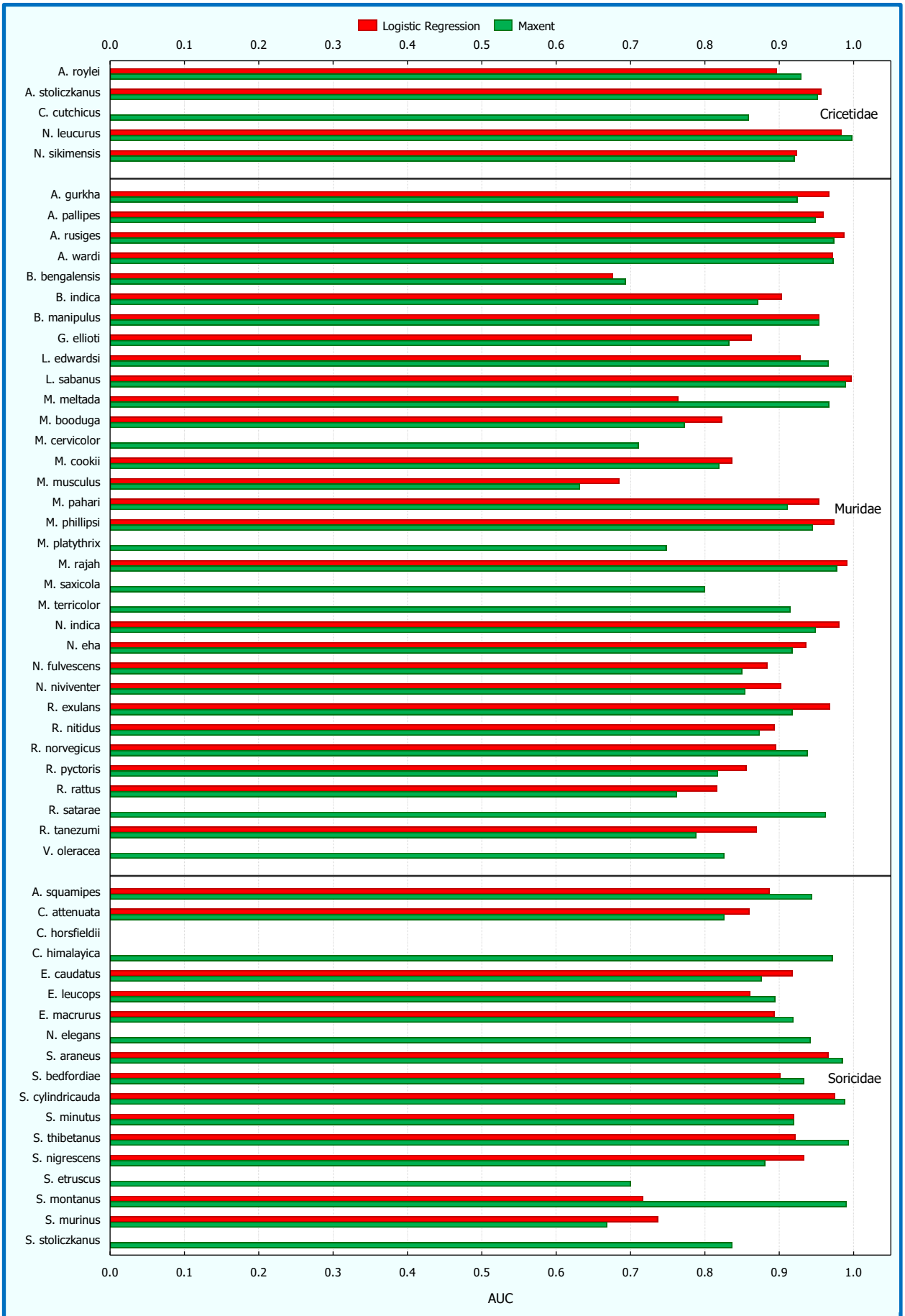


Figure 4.6 Goodness-of-fit as defined by the AUC from logistic regression and maxent model-building procedures for the 56 species recorded in >3 grid squares.

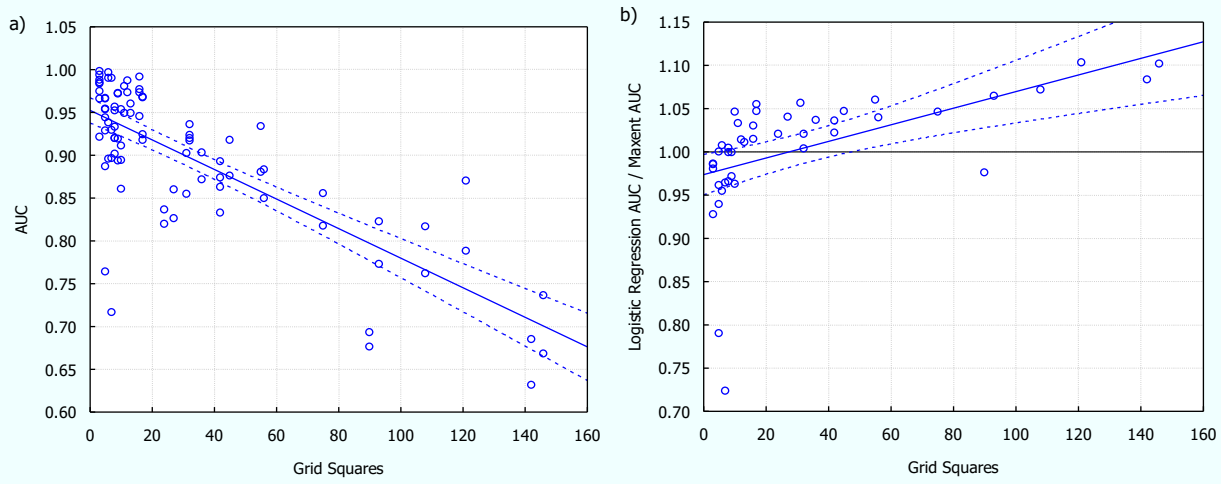


Figure 4.7 The relationship between the number of grid squares in which a species was recorded and a) the absolute AUC for all non-failed models and b) the ratio between the AUC from the logistic regression model and the AUC from the maxent model for each species. The best-fit regression line with 95% confidence intervals is shown in each graph. The reference line in (b) represents a ratio of 1.

4.3.4.2 Variable Contributions to the Models

Variable contributions were calculated automatically by the maxent process and scaled from 0 to 1. The variable contributions in the logistic regression models were calculated from the Wald statistic for each term in the model. Within each model, for each variable, the Wald statistics were tallied across all the terms in which it was included. Then each variable was standardised using the total Wald statistic for the model, so that they scaled from 0 to 1. The justification for using this procedure was that, with the exception of aspect (which was only incorporated in one model), the degrees-of-freedom for all terms in the models were always 1.

There was a significant difference in the variable contributions to the two different models (Figure 4.8). The logistic regression models tended to use fewer variables, usually only two or three, although *Golunda ellitoti* incorporated six variables. In contrast, 23 of the maxent models incorporated all eight variables and the most parsimonious model was for *Apodemus wardi*, with only three. Furthermore, all the maxent models except two incorporated aspect, which was unused by the logistic regression models, except for the failed model for *R. norvegicus*. The other important result from these models was the preponderance of longitude and latitude effects, indicating that much of the variation in the distribution of these species is likely to be geographic rather than topographic.

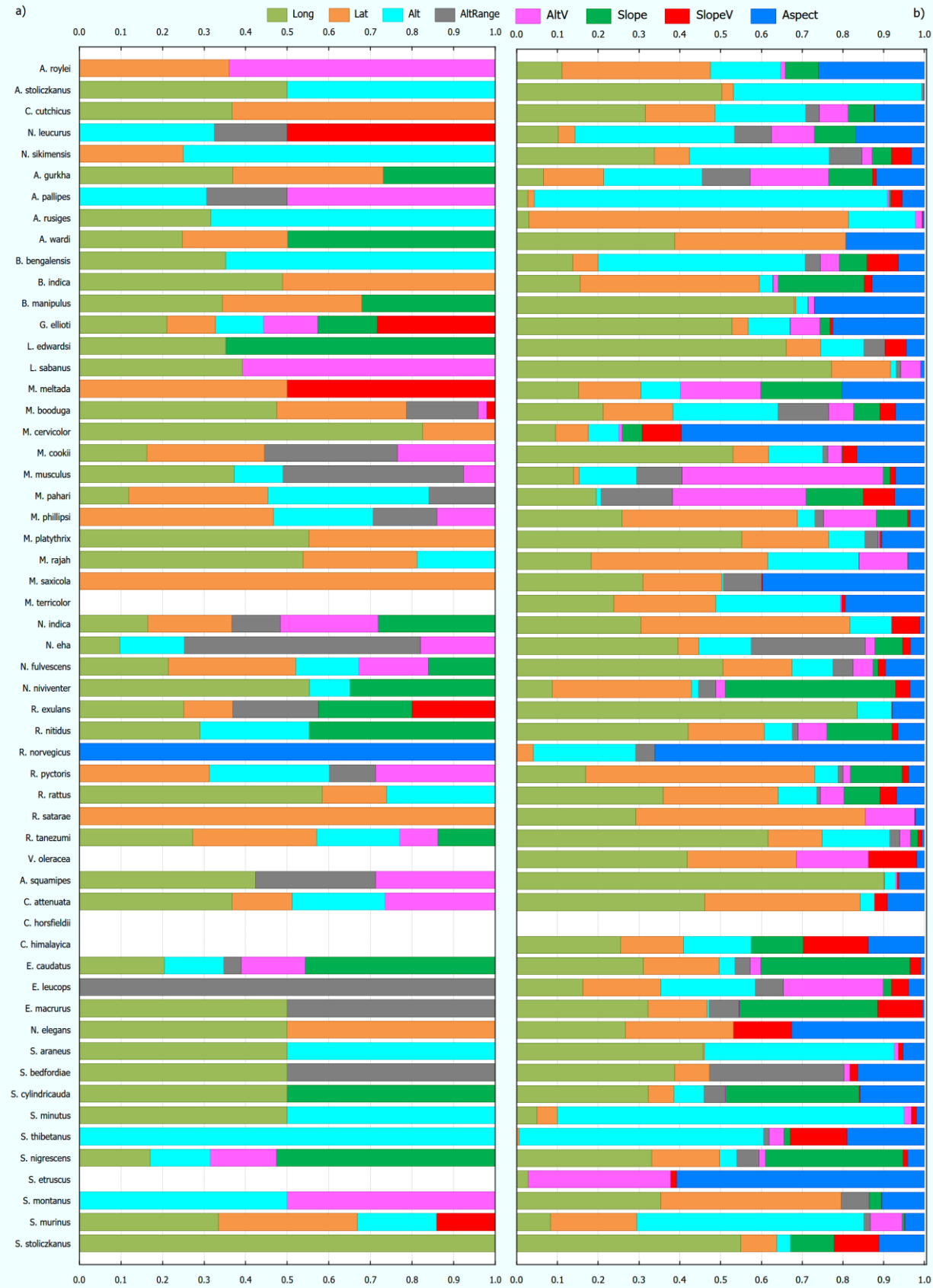


Figure 4.8 The contribution of variables to the a) logistic regression and b) maxent models. The five logistic regression models and the single maxent model that failed to resolve are shown as blank spaces. In addition, the seven logistic regression models that only included longitude or latitude effects can be identified – *e.g.*, *Mus Saxicola,* which only had latitude effects and *Bandicota indica,* which only had longitude and latitude. No maxent models were restricted to these two effects.

4.3.5 Exploration of the DEM Dataset for Nepal

The distributions of the six primary variables within the 1529 0.1° grid squares of the DEM for Nepal are presented in Figure 4.9 to Figure 4.14. Note that each figure uses its own classification for the colour coding. Most importantly for the efficacy of these models, the spatial distributions of the high and low values were very different for all six variables, although there was a degree of similarity between altitudinal range (Figure 4.10) and slope (Figure 4.12).

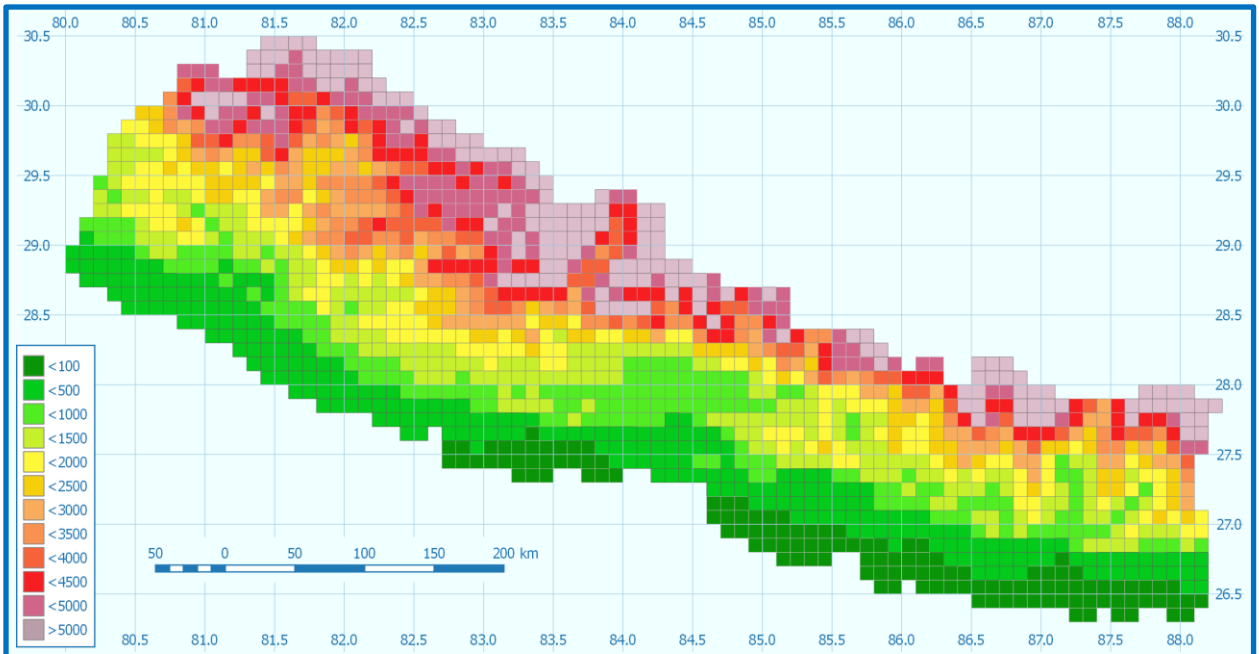


Figure 4.9 The distribution of mean altitude (m) in 0.1° grid squares in Nepal

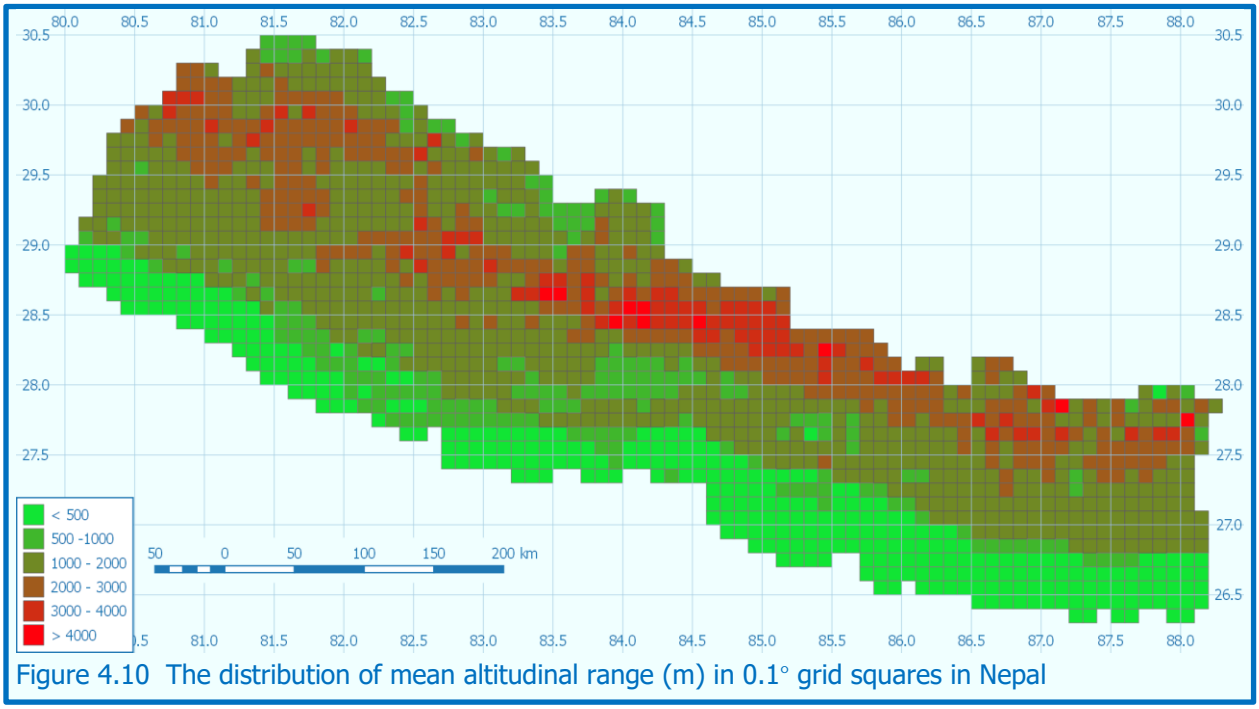


Figure 4.10 The distribution of mean altitudinal range (m) in 0.1° grid squares in Nepal

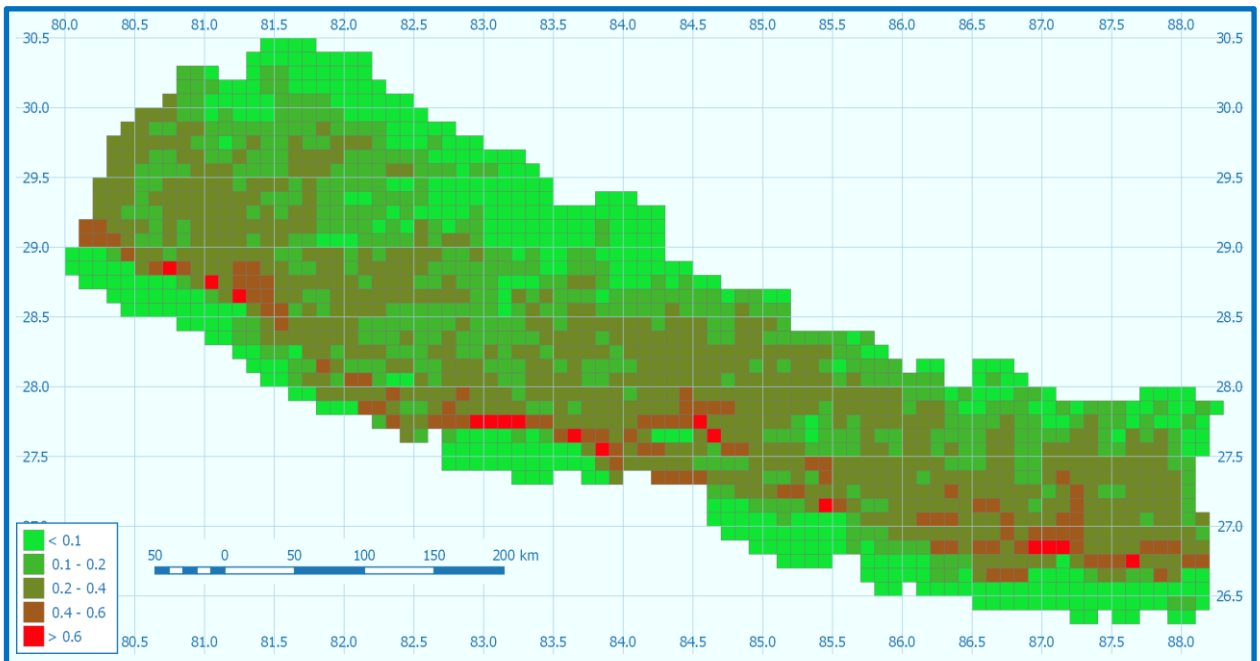


Figure 4.11 The distribution of mean coefficient-of-variation of altitude in 0.1° grid squares in Nepal

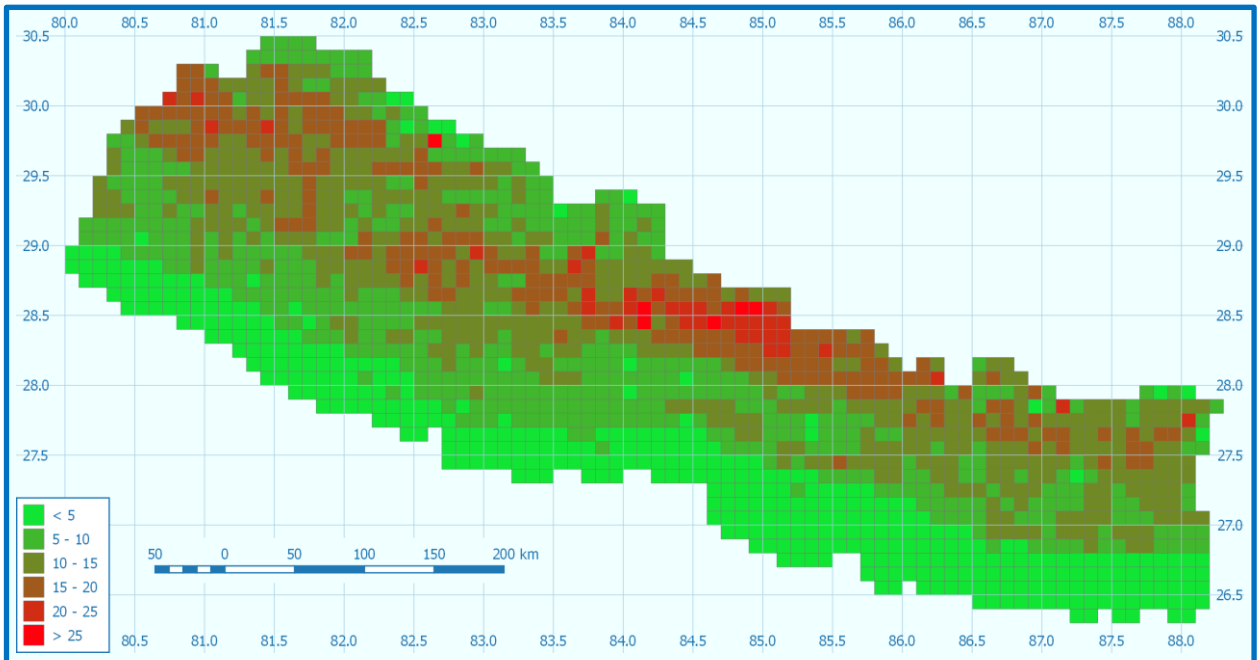


Figure 4.12 The distribution of mean slope (°) in 0.1° grid squares in Nepal

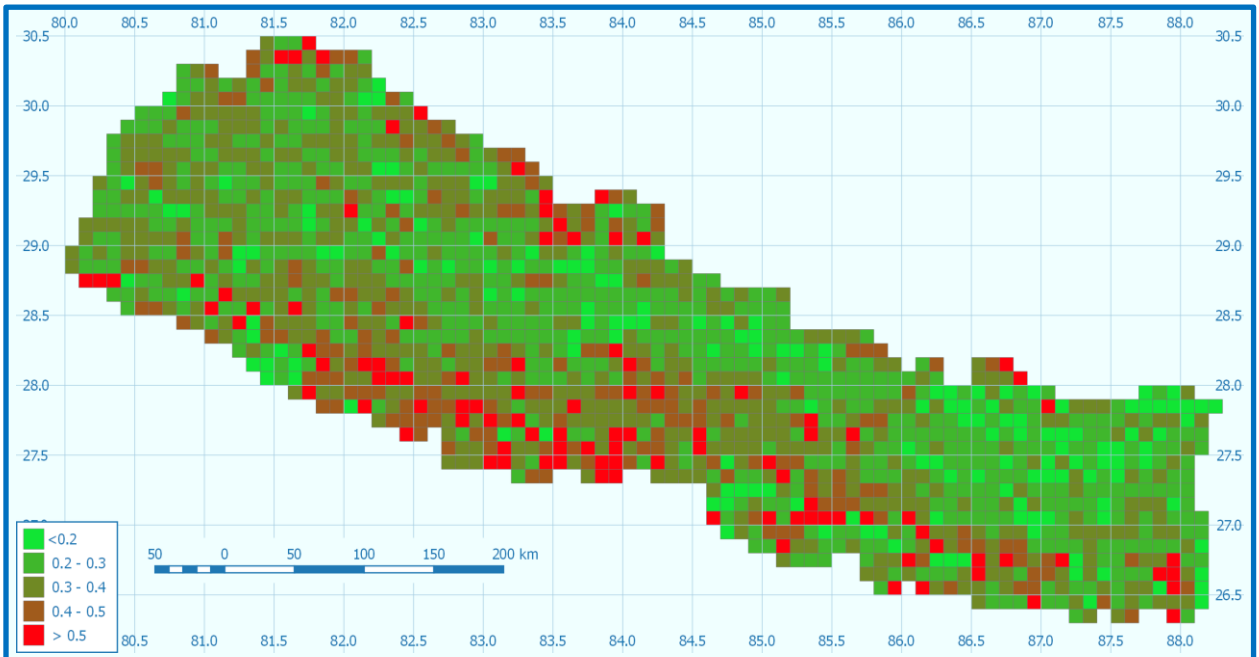


Figure 4.13 The distribution of mean coefficient-of-variation of slope in 0.1° grid squares in Nepal

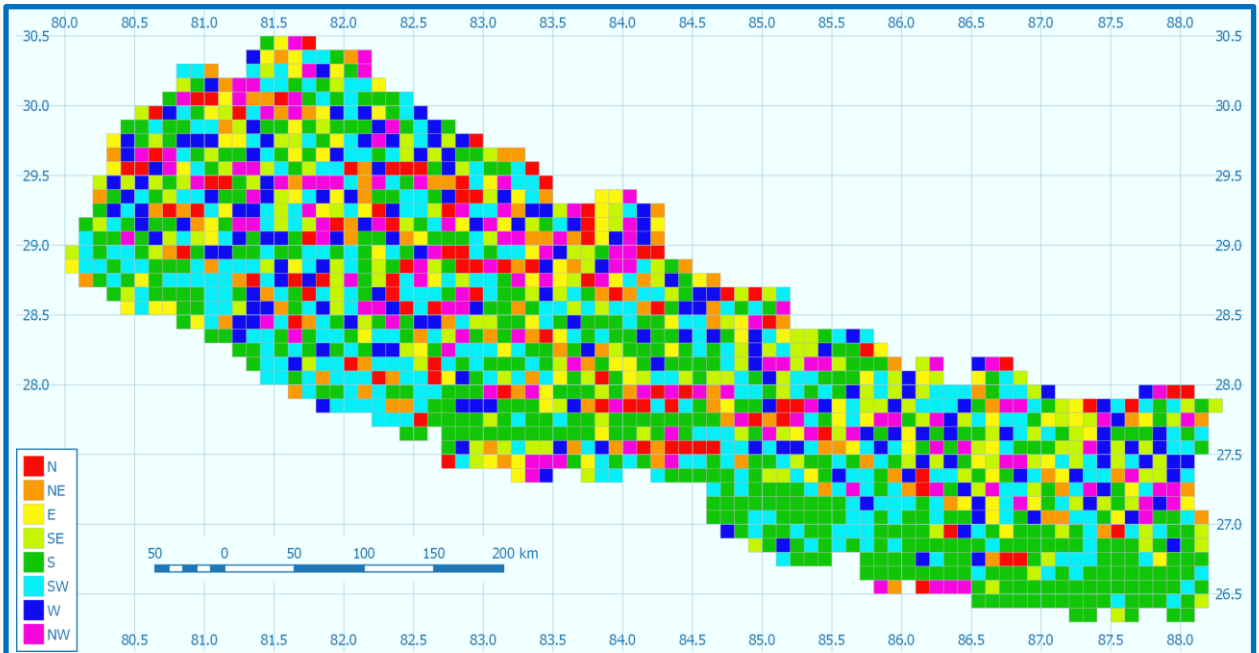


Figure 4.14 The distribution of eight octants in 0.1° grid squares in Nepal

4.3.6 Individual Species Projections for Nepal

The outputs from the models generated in Section 4.3.4 were applied to the DEM for Nepal described in Section 4.3.5. This gave 55 pairs of maps predicting the probability of occurrence of each species. These are presented in Supplementary Information Section 4.B and specific examples discussed in detail below.

The probabilities of occurrence derived from each model for each species have been categorised using a non-linear scale, expressed as percentages, which is common across all species and model types. For some species, *e.g. Rattus tanezumi*, the two models generated very similar maps (Figure 4.15). Both showed a predicted probability of distribution of between 50% and 75% in large parts of the eastern mid-hills, between 25% and 50% in the rest of mid and lowland Nepal, and general absence (< 5%) from the mountainous region bordering Tibet. Both maps showed interesting features, such as the higher probabilities in the (very deep and wide) Kali Gandaki valley running SW from Tibet at 84.0°E 29.4°N. There were some small differences, such as the lower probability predicted by the logistic regression model in the western terai region.

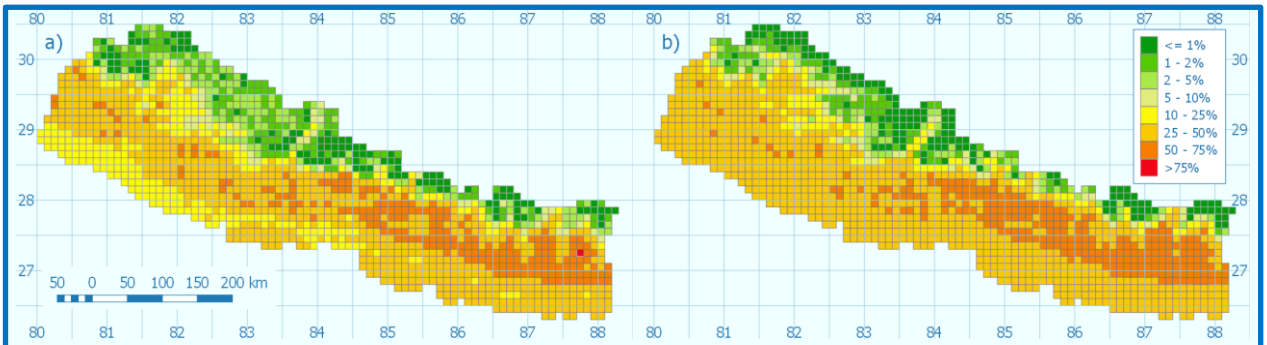


Figure 4.15 Predicted distributions in Nepal for *Rattus tanezumi* by a) logistic regression and b)

In contrast, the maps for *Mus pahari* showed a graphic difference between the two models (Figure 4.16). The logistic regression model indicated absence from most of the country, with a zone of probability between 5% and 25% in the mid-hills of the eastern region and, to a lesser extent, in the central and far western regions. However, the maxent model predicted >10% through almost the whole of the country, with large swathes of the mid-hills and main Himalayan slopes with >50% probability of occurrence.

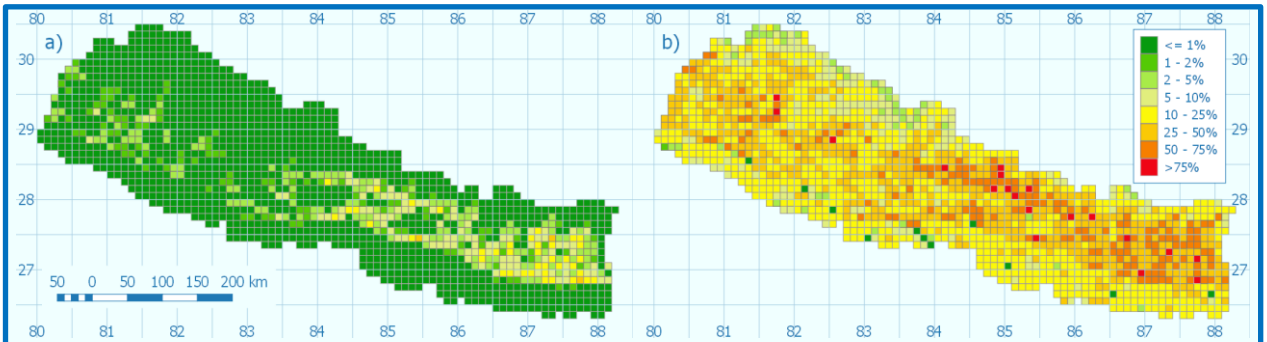


Figure 4.16 Predicted distributions in Nepal for *Mus pahari* by a) logistic regression and b) Maxent.

The final difference between the two model types was shown by, *e.g.*, *Mus cervicolor* (Figure 4.13). The logistic model effectively failed to resolve; it only included three terms with longitude, latitude and their interaction, which resulted in two zones of uniform probabilities separated by an artificial

diagonal threshold. In contrast, the maxent model appeared to give a functional prediction map. However, the main contributing variable to the model was Aspect (see Figure 4.8 & Figure 4.14), which resulted in a fairly random spatial pattern. Although the model had an AUC of 0.711, it was classified as over-parameterized because maxent was forcing variables into the model which gave little predictive ability.

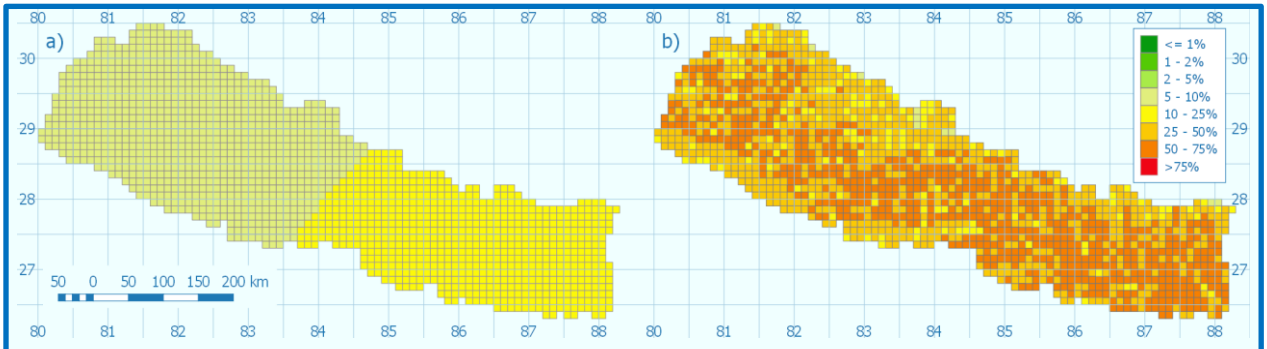


Figure 4.17 Predicted distributions in Nepal for *Mus cervicolor* by a) logistic regression and b) Maxent.

The most striking overall difference between the models was that the maxent models consistently produced higher probabilities than the logistic regression models. Although the projections provided values for the whole population of 0.1° grid squares in Nepal, the models which generated them were based on sparse samples of grid squares from the Indian sub-continent. On this basis, I tested the populations of predictions from the two different models as if they were samples from a population of all possible predictions. Two statistical procedures were used to test each species in turn;

- Parametric tests using paired-sample t-tests on the log-odds. The null-hypothesis was that the mean of the differences in the log-odds from the two models, across all 1529 squares, was zero.
- χ^2 goodness-of-fit tests using a null-hypothesis that the number of grid squares where the logistic regression prediction was greater than the maxent prediction was the same as the converse condition. This was equivalent to a large sample approximation of the sign test (Zar, 1984).

Every species showed a very highly significant difference in the χ^2 tests and only one (*Mus booduga*) showed no difference under the t-test (Table 4.5). Forty of the 43 species showed significantly greater mean probabilities from the maxent model than the logistic regression model. The other three (*L. sabanus, M. meltada* & *S. montanus*) all had very low prediction probabilities (with five of the six models predicting less than 1%).

Table 4.5 Statistical Comparison of Projections by the Logistic Regression and Maxent Models. The three species highlighted in pale blue had significantly greater predictions from the logistic regression models – all others had significantly greater maxent model predictions. The species highlighted in pale green showed no significant difference in the paired-sample t-test.

Family	Genus	Species	Mean Projected Probability				Log. Reg. > Maxent		
			Log. Reg.	Maxent	t	*p*	N	χ^2	*p*
Cricetidae	*Alticola*	*A. roylei*	0.0%	29.5%	28.6	0	0	1529	0
		A. stoliczkanus	0.5%	20.6%	134.8	0	0	1529	0
	Neodon	*N. leucurus*	0.0%	1.2%	51.2	0	126	1067	0
		N. sikimensis	5.7%	14.2%	18.6	0	239	722	0
Muridae	*Apodemus*	*A. gurkha*	0.1%	28.3%	36.7	0	96	1169	0
		A. pallipes	0.5%	13.3%	73.5	0	17	1462	0
		A. rusiges	0.5%	5.9%	92.1	0	44	1358	0
		A. wardi	0.1%	13.3%	129.3	0	1	1525	0
	Bandicota	*B. bengalensis*	5.0%	14.3%	59.2	0	141	1017	0
	Berylmys	*B. manipulus*	0.0%	9.2%	74.4	0	0	1529	0
	Golunda	*G. ellioti*	1.4%	12.8%	48.9	0	19	1454	0
	Leopoldamys	*L. edwardsi*	0.5%	26.5%	129.7	0	5	1509	0
		L. sabanus	0.0%	0.0%	-27.3	0	1200	496	0
	Maxomys	*M. rajah*	0.0%	0.0%	51.8	0	0	1529	0
	Millardia	*M. meltada*	0.8%	0.1%	-13.9	0	922	65	0
	Mus	*M. booduga*	10.2%	10.3%	0.4	0.70	522	154	0
		M. cookii	2.4%	35.0%	113.8	0	3	1517	0
		M. musculus	31.9%	45.0%	49.8	0	97	1166	0
		M. pahari	0.0%	23.6%	70.4	0	3	1517	0
		M. phillipsi	0.0%	0.0%	155.8	0	1	1525	0
	Nesokia	*N. indica*	0.0%	18.2%	90.7	0	13	1477	0
	Niviventer	*N. eha*	1.9%	22.7%	60.3	0	37	1385	0
		N. fulvescens	7.3%	25.6%	56.2	0	97	1166	0
		N. niviventer	9.5%	34.9%	70.7	0	20	1450	0
	Rattus	*R. exulans*	0.2%	2.9%	51.9	0	165	940	0
		R. nitidus	7.2%	33.6%	65.1	0	54	1321	0
		R. norvegicus	0.0%	0.1%	37.4	0	258	671	0
		R. pyctoris	15.4%	38.3%	61.5	0	15	1470	0
		R. rattus	16.8%	31.0%	50.1	0	90	1190	0
		R. tanezumi	17.1%	21.1%	14.8	0	370	407	0
Soricidae	*Anourosorex*	*A. squamipes*	0.2%	13.5%	173.8	0	0	1529	0
	Crocidura	*C. attenuata*	4.9%	38.2%	155.5	0	1	1525	0
	Episoriculus	*E. caudatus*	5.6%	34.9%	68.7	0	11	1485	0
		E. leucops	1.8%	35.5%	203.1	0	4	1513	0
		E. macrurus	1.3%	22.3%	61.4	0	36	1388	0
	Sorex	*S. araneus*	0.2%	9.9%	193.6	0	1	1525	0
		S. bedfordiae	1.2%	26.0%	179.3	0	4	1513	0
		S. cylindricauda	0.2%	8.0%	144.7	0	0	1529	0
		S. minutus	1.3%	32.5%	189.2	0	0	1529	0
		S. thibetanus	0.4%	11.5%	68.8	0	31	1408	0
	Soriculus	*S. nigrescens*	12.9%	32.7%	47.5	0	140	1020	0
	Suncus	*S. montanus*	1.5%	0.1%	-93.4	0	1496	1400	0
		S. murinus	28.2%	31.3%	8.7	0	576	93	0

4.3.7 Predicted Species Richness

The models presented in the previous section were used to generate species richness maps. The 43 individual species projections were recast as present or absent in each grid square at four different thresholds (20%, 30%, 40% & 50%). These were then tallied across all species, separately for the

two different model types (Figure 4.18). The effect of a lower threshold was that, for each species, a greater number of grid squares exceeded the threshold and were classified as likely present. When tallied across all species, this was manifest as a higher predicted species richness – as shown in the top pair of maps. At higher thresholds, fewer species were predicted as present, so species richness was lower, as shown in the bottom row.

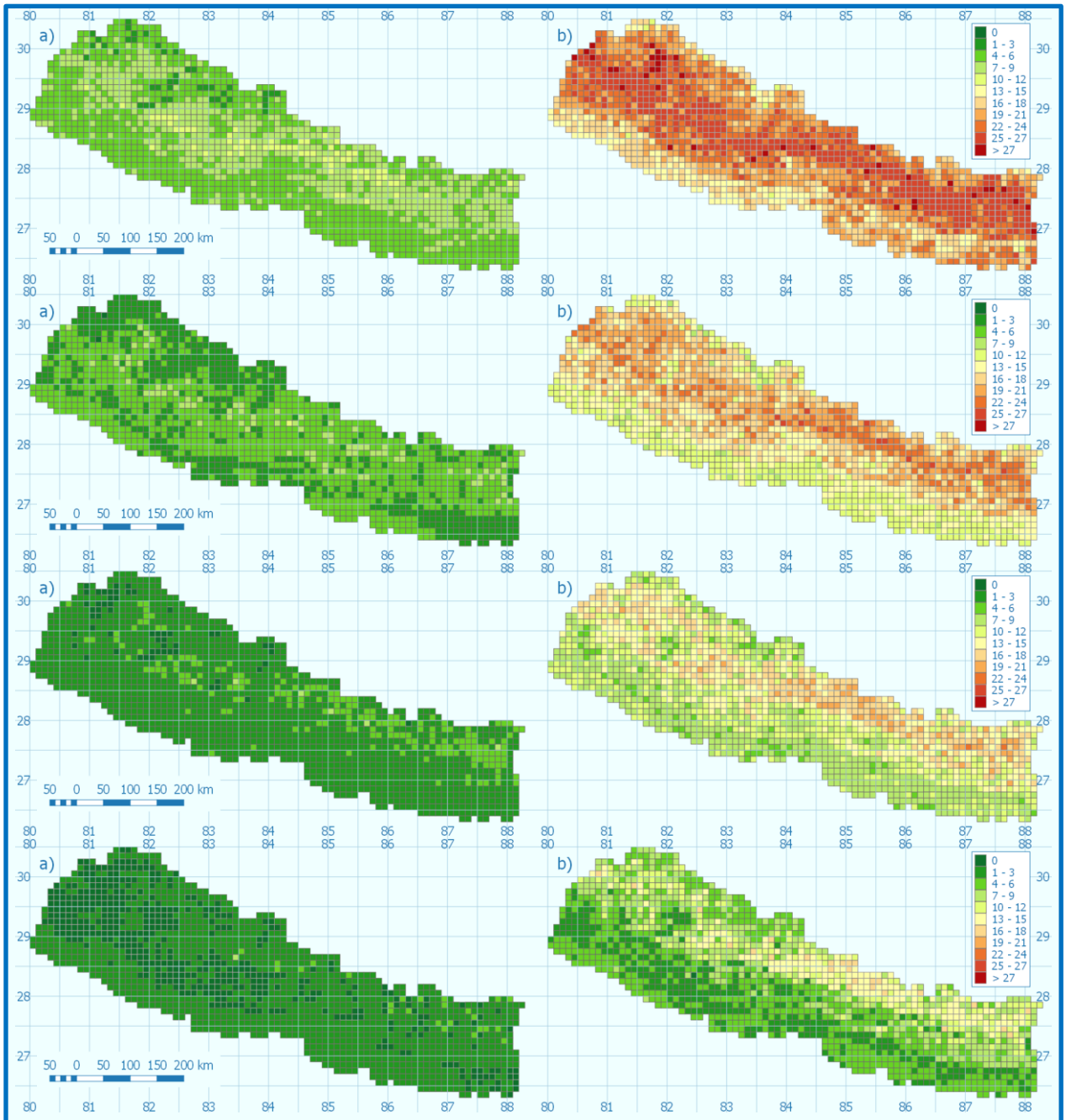


Figure 4.18 Predicted species richness based on 43 individual species projections from a) logistic regression models and b) maxent models. The rows represent four different thresholds of probability of occurrence – from the top – 20%, 30%, 40% & 50%.

4.3.8 Predictions for Species of Interest to this Study in the Fieldwork Locations

Chapter 2 described the four transects in which the trapping programme took place. They were located in ten 0.1° grid squares (Figure 4.19). Note that one square (83.9E 28.4N) encompassed two transects; Pipar and Annapurna, reflecting the fact that, at their nearest, they were less than 10km apart, although isolated by a 6000m ridge (see Figures 2.23 and 2.25).

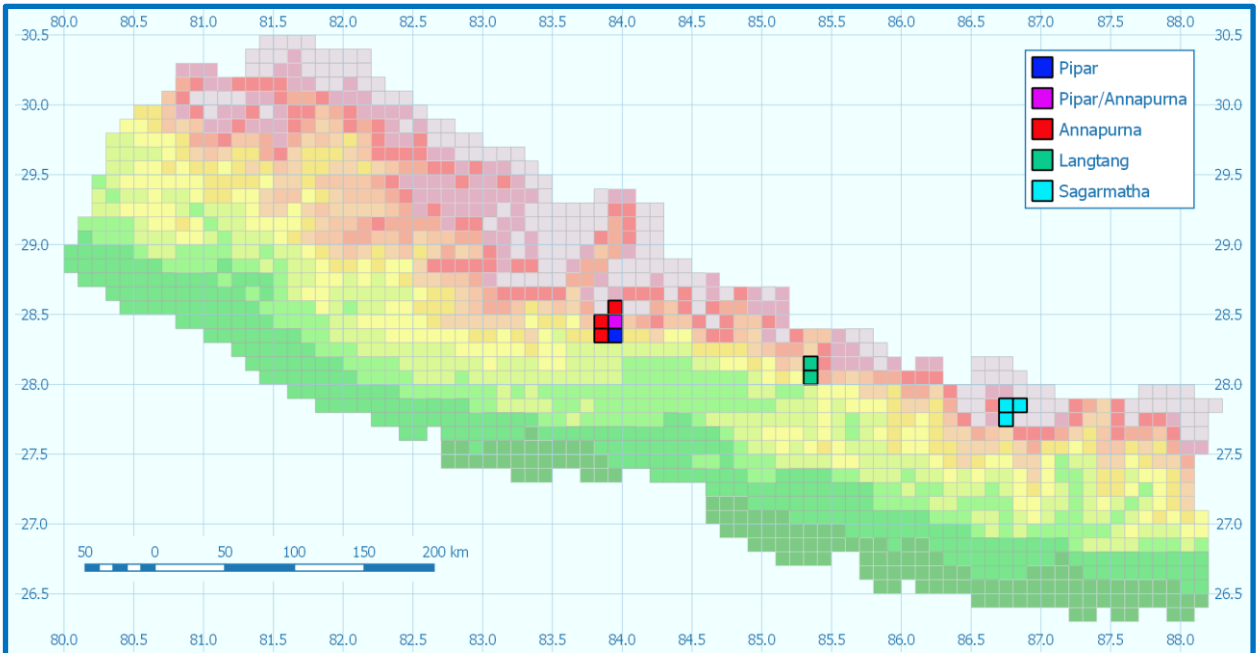


Figure 4.19 Location of the ten 0.1° grid squares encompassing the four transects from the current study, superimposed on the map of altitude given in Figure 4.9.

Chapter 2 also presented the trapping results from the current study, in which 18 putative species were identified (Table 2.5). The records for all 792 animals were tallied by transect and 0.1° grid square, giving 11 combinations (Table 4.6). The large number of animals recorded in the two Pipar grid squares was a consequence of two factors. Firstly, it was a very steep transect with small horizontal distances, so a large amount of trapping effort was concentrated in a small area. Secondly it was the only transect that was visited twice. The most ubiquitous species was *E. caudatus*, which was found in every grid square, followed by *S. nigrescens* and *M. booduga*, found in ten and nine squares respectively. Two species (*N. eha*? And *E. leucops*) was recorded from every transect, but not in all grid squares.

Table 4.6 Number of animals from each putative species caught in each 0.1° grid square.

		Pipar		Annapurna				Langtang		Sagarmatha			
Order	Species	83.9E 28.3N	83.9E 28.4N	83.8E 28.3N	83.8E 28.4N	83.9E 28.4N	83.9E 28.5N	85.3E 28.0N	85.3E 28.1N	86.7E 27.7N	86.7E 27.8N	86.8E 27.8N	Total
Lago-morpha	*Ochotona tibetanus*		1										1
Rodentia	*Neodon sikimensis*		2					1				2	5
	Murid sp1	2	1										3
	Murid sp2									2			2
	Apodemus gurkha	2	29		6	1	3						41
	Mus booduga	70	3	2	1			18	17	48	25	25	209
	Niviventer fulvescens?	14	1		8			2	1				26
	Niviventer eha?		1			1			1	3	4		10
	Niviventer niviventer?	4		1									5
	Rattus nitidus	6						1					7
	Rattus sp									2			2
Sorico-morpha	*Soricid* sp1		3					4	2		1	1	11
	Soricid sp2										1		1
	Episoriculus caudatus	40	45	2	42	1	5	26	28	27	21	3	240
	Episoriculus leucops		4		7				1	6	2		20
	Sorex bedfordiae											1	1
	Soriculus nigrescens	68	40	5	24	3	3	9	7	26	2		187
	Suncus murinus	17		4									21
	Total	223	130	14	88	6	11	61	57	114	56	32	792

Here the results of the predictive models for the species of interest (and closely related species – see Section 4.3.7 and Supplementary Information 4.C) have been collated for the ten grid squares. For each species, the predicted probabilities from the logistic regression and maxent models were averaged by transforming each to log-odds, taking the mean of the two values, and then back transforming to probabilities. These were then used to indicate the likelihood that the putative species of interest were present in the grid squares in which they were captured (Table 4.7).

The table is divided into eight groups which broadly represent taxonomic groupings. The first six groups each have one species that was identified in the field (shaded in pale blue) and a number of other congeneric (or closely related) species. Within each grid square, the species with the highest probability of occurrence is highlighted in red. Squares with both shading and highlighting provide strong evidence that the field identification was probably correct. For example, the four squares in which *A. gurkha* was recorded all had this species as the most likely species of *Apodemus*. Furthermore, the general distribution of highest probabilities across all ten squares, gave an overall indication of the presence of the species in the study area. Thus, *N. sikimensis* was predicted the most likely species of cricetid in nine out of ten squares, although it was only recorded in three. However, the opposite situation occurred with the four species of *Mus* (Table 4.7c), where the highest probability

in every square was predicted to be M. *musculus*, yet the species recorded in nine of the ten squares was M. *booduga*. The last two groups (*Niviventer* and soricids) contained six species all recorded in the field, so no attempt was made to rank them with the highest probabilities. Instead, the squares in which they were recorded are shaded indicating, for example, that N. *eha* was recorded in four squares, three of which were the highest probabilities for that species. Similarly, the square in which S. *nigrescens* was not recorded was the only square which had a probability of occurrence of < 40% for this species. And the two squares where S. *murinus* was recorded were the only ones with a predicted probability of > 40%.

Table 4.7 Average predicted probability of occurrence (%) in the ten 0.1° grid squares covered by the current study for eight groups of species. The first six groups (a to f), illustrate the locations of six putative species identified in the field, shaded in pale blue. Within each group, the species with the highest probability in each grid square is highlighted with red text. Group (g) shows all three putative species of *Niviventer*, with no highlighting of the probabilities, as they were all identified in the field. Similarly, group (h) contains three species of shews for which there was no doubt about their field identifications.

Species	83.9E 28.3N	83.9E 28.4N	83.8E 28.3N	83.8E 28.4N	83.9E 28.5N	85.3E 28.0N	85.3E 28.1N	86.7E 27.7N	86.7E 27.8N	86.8E 27.8N
a)										
Alticola roylei	1.3	5.7	15.5	3.8	28.3	9.2	3.4	13.4	14.2	14.3
Alticola stoliczkanus	2.1	16.5	2.3	9.0	44.7	8.7	5.2	16.8	26.2	68.8
Neodon leucurus	0	0	0.1	0	0.1	0	0	0	0.1	1.6
Neodon sikimensis	12.1	41.9	9.2	32.9	58.6	26.6	23.1	39.2	53.1	76.5
b)										
Apodemus gurkha	32.6	35.1	17.5	45.0	60.3	40.9	29.4	21.7	17.1	20.1
Apodemus pallipes	1.2	2.8	2.4	4.0	37.0	2.8	2.7	4.2	20.2	74.6
Apodemus rusiges	1.3	3.8	1.8	2.8	8.2	1.8	1.5	1.4	1.7	2.3
Apodemus wardi	2.1	4.3	2.5	3.6	5.6	5.0	2.9	7.9	3.6	3.2
c)										
Mus booduga	12.3	0.4	15.4	1.7	0.5	2.4	4.4	0.8	1.2	0.6
Mus cookii	12.3	1.2	12.7	3.6	1.1	6.5	6.5	6.1	4.4	2.4
Mus musculus	49.2	31.0	54.7	38.5	26.2	51.7	50.6	51.9	48.4	22.2
Mus pahari	17.9	2.1	4.8	7.7	0.3	10.2	22.6	6.0	2.6	0.3
d)										
Rattus nitidus	38.8	9.8	33.2	23.5	4.2	31.7	50.0	23.1	27.6	6.7
Rattus pyctoris	32.8	33.9	39.0	46.3	32.6	44.4	45.5	39.1	37.8	24.8
Rattus rattus	30.6	15.8	31.1	24.0	8.4	19.9	25.9	16.4	18.0	5.5
Rattus tanezumi	49.2	7.1	46.7	16.6	0.8	24.9	38.9	13.1	6.1	0.2
e)										
Episoriculus leucops	15.1	53.8	10.4	33.6	37.2	30.4	26.9	42.1	28.7	30.2
Episoriculus macrurus	17.0	59.9	7.9	36.4	34.2	31.5	25.1	45.5	18.0	24.5
f)										
Sorex araneus	0.8	6.7	1.0	3.6	24.5	3.2	2.7	6.2	18.2	54.9
Sorex bedfordiae	10.8	57.3	6.5	30.5	35.7	28.4	32.2	44.7	37.5	39.0
Sorex cylindricauda	4.7	44.1	1.0	28.5	12.9	8.0	18.0	16.6	19.7	14.8
Sorex minutus	6.6	24.2	5.5	15.8	37.8	12.4	8.9	18.6	24	52.1
Sorex thibetanus	4.1	13.3	1.9	8.2	16.9	4.1	2.6	6.1	6.2	28.1
g)										
Niviventer fulvescens?	38.7	8.5	38.1	19.8	1.7	29.0	44.5	20.7	18.1	1.4
Niviventer eha?	22.1	62	20.0	43.2	48.3	50.6	41.3	68.6	69.2	33.5
Niviventer niviventer?	42.5	59.4	31.7	62.6	34.6	47.4	56.9	42.5	35.4	12.5
h)										
Episoriculus caudatus	31.4	58.8	34.1	48.8	51.8	56.5	51.3	69.6	70.8	33.9
Soriculus nigrescens	41.9	49.6	46.5	52.6	44.9	59.2	64.9	64.7	74.3	34.7
Suncus murinus	44.4	18.2	41.5	26.2	5.2	25.2	35.8	16.7	13.3	2.1

4.4 Discussion

The BNHS dataset is an exceptionally valuable resource that has remained virtually un-analysed up until now. The data collation, geo-referencing and data structuring that I have carried out gives this dataset much added value, making it amenable for additional analysis. In a similar way, the Pearch (2011) monograph is a very comprehensive study, but the data are encapsulated in a way that makes them totally unavailable for statistical analysis. The SQL Server database that now holds these two datasets can be interrogated taxonomically, geographically or by collector, with summaries and cross-tabulations easily extracted.

The BNHS dataset in particular has two unique characteristics. Firstly, it covered a huge collecting area of over 5,245,000 km², representing the whole of British India, Nepal and Bhutan which, today, encompasses seven countries. Secondly, the majority of collections were made almost exactly 100 years ago – approx. 35% of the records used in this analysis were collected in 1913, 1914 and 1915, exactly one century before my own study.

4.4.1 Taxonomic Breakdown of the Reference Datasets

The taxonomic breakdown of these datasets showed some interesting contrasts with the current study (Figure 4.20). Firstly, the records from the current study were dominated by three genera; *Mus*, *Episoriculus* and *Soriculus*, with 636 animals in total, representing 80.3% of all specimens caught. These three species formed a cohort of extremely small animals with the mean body weights being 11.2g, 5.0g and 14.0g respectively. The genera with the next highest mean weights were *Apodemus* (23.6g) and *Neodon*, (25.6g), with the remaining three all averaging over 35g. In contrast, the total records for *Episoriculus* (43) and *Soriculus* (51) only constituted 1.3% of the BNHS collection. The consequences of very light weights for trapping efficacy has already been alluded to in Chapter 2 (Figure 2.20). The BNHS collection was dominated by two genera, *Mus* and *Rattus*, which together comprised 59.9% of the specimens. This was almost certainly due to the geographical coverage of the BNHS collection, with many specimens of commensal species collected from areas with relatively high human habitation (Parshad, 1999). This was supported by the third most commonly recorded genus (*Suncus*) dominated by the highly commensal *S. murinus* (7.6%). The more equitable distribution of the eight genera in the Pearch dataset probably better reflects the relative abundances of these genera in Nepal, given the range of sources from which they were compiled (Abe, 1971; Daniel, 2015; Mitchell, 1977; Niethammer and Martens, 1975).

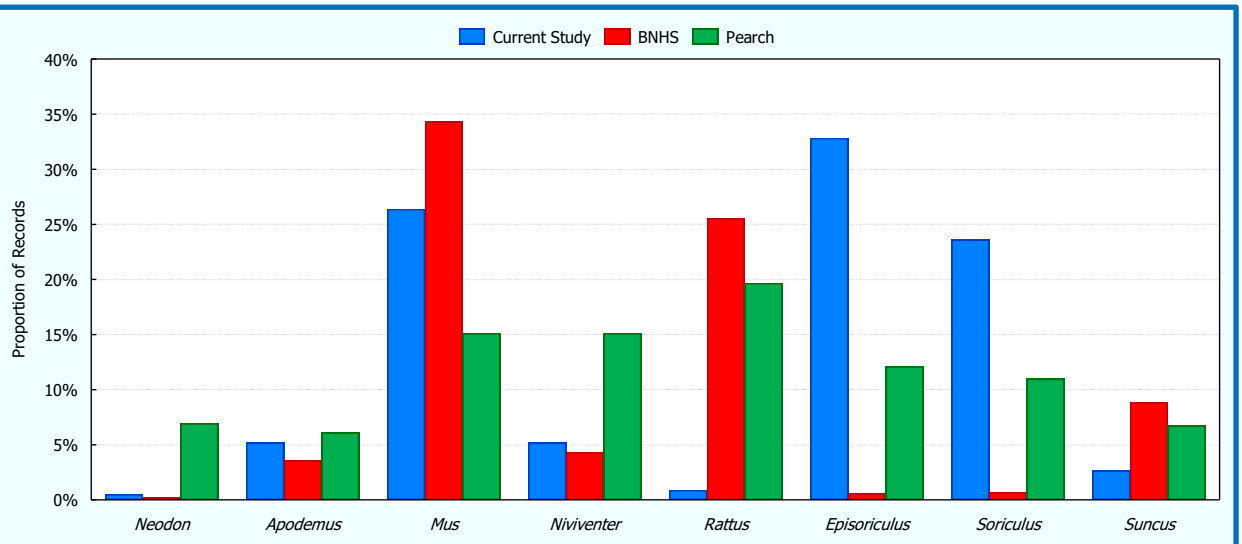


Figure 4.20 The proportion of records belonging to eight genera from the current study, the BNHS collection and collated from the Pearch (2011) monograph.

4.4.2 Analysis of Collecting Locations

The analysis of geographic and topographic distributions of sampling locations revealed significant departure from random in both datasets. This is important because maxent modelling in particular makes assumptions about spatial sampling bias (Merow *et al.*, 2013; Phillips *et al.*, 2004), which is discussed further below. However, the stringent analytical criteria that I applied to the ANOVA model for distribution of BNHS locations within "British" India (reported in Table 4.4 and Figure 4.3) should be interpreted with care. Overall, these sampling sites were distributed across all latitudes, longitudes and altitudes. In particular, the range of latitudes extended from the southern tip of Sri Lanka to within approx. 120km of the northern extent of current-day Pakistan. The non-random latitudinal distribution was only caused by the higher than expected location of sites in the southern tip of India, Sri Lanka and Myanmar. Similarly, the range of longitudes covered was 36°, over 3500km, with only the far western region of Pakistan and a zone between 81°E and 84°E showing a lower than expected number of sites. The altitudinal distribution gives some of the explanation for this distribution, with much higher proportion of sites visited in the 1000m – 2500m range. This corresponds to many of the hill stations used by the British Raj in the early part of the 20th century, such as Almora (29.5°N, 79.5°E), Darjeeling (27°N, 88.5°E), Murree (34°N, 73.5°E) and Ootacamund (11.5°N, 76.7°E), as well as areas such as Kashmir, the Mudumalai Hills, central Sri Lanka and the Meghalya hills. Given the extreme topography of Nepal, the locations catalogued in Pearch (2011) cover virtually the whole extent of the country. However, the analysis showed a lower than expected number of locations in the western third of the country (east of 83°E). The most important distinction between the altitudinal distributions in the Pearch dataset was that the mid-elevation "hump" in locations was higher (2500m – 4000m) than in the BNHS dataset. This is an

interesting result because the main centres of population are Kathmandu (1400m), Pokhara (900m) and Biratnagar, Birganj and Lumbini (all in the Terai at 100m – 200m), well below this zone. The topography of Nepal has clearly drawn fieldworkers to the mid-hills and Himalayan regions preferentially.

4.4.3 Species Distribution Models

Many different types of species distribution models (SDM) have been proposed. Guisan and Zimmermann (2000) gave a comprehensive review of a number of these, including generalized linear models (GLM *e.g.* logistic regression), generalized additive models (GAM), environmental envelopes (*e.g.* models BIOCLIM, DOMAIN and HABITAT), canonical correspondence analysis and neural networks. Of particular relevance to this study was the work by Ponder *et al.* (2001), who focused on museum specimens as their source material. Anderson *et al.* (2002) and Anderson (2003) used the Genetic Algorithm for Rule-set Prediction (GARP) (Stockwell and David, 1999) in their empirical study of the distribution of *Oryzomys* and *Heteromys* rodent species in South America. More recently, the maximum entropy (MaxEnt) approach (Phillips *et al.*, 2018) has found strong support (Merow *et al.*, 2013; Phillips *et al.*, 2006; Phillips *et al.*, 2004).

Two major concepts have to be considered for the successful employment of SDMs. The first is sample size, that is, the absolute number of recorded localities of a species required to generate an adequate model. Pearce and Ferrier (2000) tested three different sample sizes (50, 250 & 500) using Monte Carlo methods on GLM and GAM models. They showed that sample size was the most important of five main properties of these models, with a significant increase in accuracy of prediction occurring between samples of 50 and 250. Stockwell and Peterson (2002) explored sample sizes required by four different modelling procedures on Mexican bird species and showed that GARP models could achieve 90% of their final accuracy with only 10 presence locations. In contrast, logistic regression models required a sample size of 20 to achieve this level of relative accuracy, although both converged at approximately 68% absolute accuracy with samples of 50. The effect of sample sizes as small as five was assessed by (Bean *et al.*, 2012) for MaxEnt models and recently Liu *et al.* (2018) recommended creating ensemble models from multiple modelling techniques to overcome some of these problems.

The other issue is biased sampling, which is where the species location records are not distributed representatively across the environmental space. This was described as biased distribution of survey effort by Phillips *et al.* (2009). This has a number of undesirable consequences; for example, Nelson

et al. (1990) showed that supposed centres of endemism of plant species in Brazil were spatially correlated with areas of high sampling effort.

To discuss biased sampling we need to address the concepts of "background samples" or "pseudo-absences", about which there seems to be some confusion in the literature, even if just of terminology. MaxEnt aims to model presence only, but still requires a suite of geo-referenced predictor variables, known as the "background data" (Merow *et al.*, 2013). There is no assumption that these data points represent true absences of a particular species; the TI or "Taxon of Interest" of Ponder *et al.* (2001). In contrast, methods such as logistic regression modelling require absence as well as presence records to constitute a binary response variable. In reality, absences are not usually known, especially from museum specimens, but several authors have proposed the creation of "pseudo-absence" records. Ponder *et al.* (2001) suggest that "...Survey sites where background sampling has taken place but the TI is absent can be used as 'pseudo-absences'...". Anderson (2003) defined a target group of species and gave, as a relevant example, the group of non-volant small mammals collected using a particular type of trap. His assumption was that if a similar field method was used to collect presence records for all these species, then locations where the TI was absent could be treated as pseudo-absences. He defined this as his "Locality Model", which is similar to the "broad set of similar species" called a target group by Phillips and Dudík (2008). Elith *et al.* (2006) described these concepts generically as community-based methods. The confusion arises because a number of authors (Barbet-Massin *et al.*, 2012; Bean *et al.*, 2012; Phillips *et al.*, 2009) appear to use the terms inter-changeably. I propose that for presence-only models (such as MaxEnt) the term used should be "background data" and for models that use a binary response variable (*e.g.* logistic regression, GAM, *etc.*) the term should be "pseudo-absence".

Yackulic *et al.* (2013) point out that the assumption of random, or at least representative, sampling is far more important for presence-only methods such as MaxEnt, compared to presence-absence models. In particular, they emphasise that any sampling biases should not be correlated with the environmental variables used in the models. For example, I have already shown the spatial sampling bias in both source datasets; the probability of a grid square being sampled was clearly related to its latitude, longitude or altitude. In turn, these variables were often selected as major predictors in the models, which is the correlation they refer to. However, Phillips *et al.* (2009) point out that presence-only modelling is more robust to sample selection bias as long as the background points have the same bias. By using the "pseudo-absence" squares as the background points, rather than just selecting them at random from the DEM, I was able to ensure this preferable situation. Nevertheless, they caution that other implicit assumptions, such as all species having the same probability of being

recorded at each site, could still weaken the models if violated. For example, we do not know if the collectors of the BNHS dataset recorded and collected every animal they caught, or whether they ceased recording particular species after an a priori number of specimens had been collected.

Yackulic *et al.* (2013) also raise another criticism of MaxEnt which is that the logistic output is often treated as an occurrence probability, even when essential assumptions (random or representative sampling and a constant detection probability) are not met. Furthermore, the distribution maps thus produced are based on an assumption that the occupancy probability at an "average" site is 0.5 – which is highly unlikely. However, they do not say what the logistic output actually does represent.

Renner and Warton (2013) introduce a different problem with MaxEnt models, which is that of scale-dependence. They show that MaxEnt models are mathematically equivalent to Poisson point process models (PPM) and, in doing so, reveal a number of diagnostic methods available in PPM that are absent from MaxEnt models. In addition, they show that the spatial resolution at which a PPM should be applied can be ascertained from the data themselves using log-likelihood estimates. In contrast, MaxEnt models require an arbitrary, a priori spatial resolution to be selected to represent the size of the grid cells in which presence is recorded. Thus when presence locations are known accurately (modern GPS equipment can record locations to within a few metres), their allocation to grid cells that may be several kilometres in size results in a severe loss of information. However, the current datasets, especially that from the BNHS, did not have this degree of intrinsic accuracy. At best the records could be assigned to 0.1° grid squares, representing an area of approximately 110km2, and some of those may still have been mis-located. In this context the scale-dependency of the MaxEnt models was not arbitrary, but determined by the best available accuracy of the records themselves. Furthermore, the same scale-dependency was applied to the logistic regression models.

For this study, I used the BNHS and Pearch datasets as the locality data with two of these modelling methods to provide a degree of cross-verification. I chose to use logistic regression (LR) and MaxEnt following the recommendations of (Elith *et al.*, 2006) who suggested that group discrimination models (such as logistic regression) had higher performance than profile models (Lobo and Tognelli, 2011), with the exception of MaxEnt. I included species with as few as three location records, following the procedure of Ponder *et al.* (2001), although Bean *et al.* (2012) reported that accuracy levels of MaxEnt models may be artificially inflated for small sample sizes. This might explain the very high AUC values that I obtained for some of my species (Figure 4.6) and negative relationship shown in Figure 4.7a.

The LR models used a second-order surface design which included all eight main effects, seven squared effects (quadratics) for the linear variables, and all two-way interactions of the linear variables. Pearce and Ferrier (2000) showed that third-order polynomial (cubic) models performed less well than simpler models. These authors also showed that the stepwise method of model building had superior discrimination to forward-inclusion (with no removal) methods. For the MaxEnt models I chose the default values for model parameters after testing an increased regularization multiplier of 5 as advised by Merow *et al.* (2013), which gave severely reduced sensitivity. Exhaustive testing of all combinations of user-defined parameters and values was beyond the scope of this analysis.

The final models projected onto the DEM of Nepal are presented in Supplementary Information 4.C. Clearly, some of these were not functional, largely due to the small number of location records. This particularly affected the LR models, with eleven failing to resolve usefully. The MaxEnt models usually resolved, but seven of them appeared to be over-parameterized, in the sense that they forced variables into the model when they had little predictive value. This situation may have been improved by alteration of user-defined parameter values.

Subjectively, the logistic regression models appeared to give more realistic distributions of most species in Nepal. For example, the LR model for *Apodemus gurkha* (Figure 4.21a) showed a predicted focus in the mid-hills and high Himalaya region in the western half of the country. This conforms well with the distribution of the 27 actual locations given by Pearch (2011) (Figure 4.21c). Furthermore, the model predicted that the distribution extends to the western border of Nepal and possibly into the zone of tropical pine forest. In contrast, the MaxEnt model (Figure 4.21b) predicted presence of *A. gurkha* throughout the country, including the terai, with up to 25% probability – a distinctly unlikely result. The models for *Leopoldamys edwardsi* (Figure 4.39) showed a huge difference between models. The LR model predicted, with very low probability (< 5%), presence in the far north-eastern region in the Kanchenjunga massif. This is not surprising as Corbet and Hill (1992) report the distribution of this species extending east from Darjeeling through the Himalayan foothills to south China and north Vietnam. However, the MaxEnt model predicted a ubiquitous distribution throughout the country with > 25% probability in over 75% of grid squares. Overall, two different statistical methods (paired t-tests and χ^2 goodness-of-fit) both showed that MaxEnt models had significantly greater probabilities of occurrence in all but four of the species. LR models were clearly more conservative.

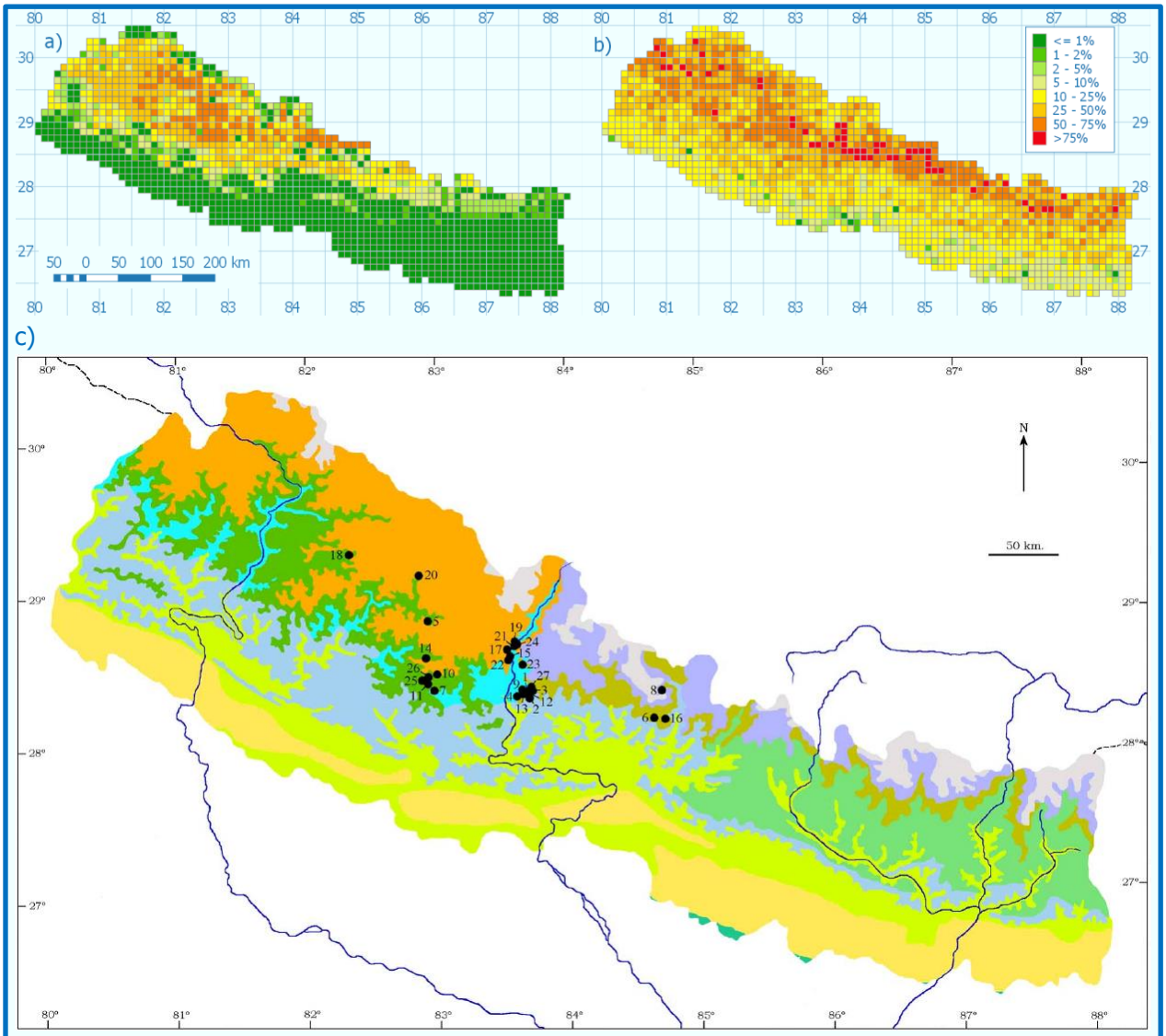


Figure 4.21 Predicted distributions in Nepal for *Apodemus gurkha* derived from the BNHS and Pearch datasets by a) logistic regression and b) Maxent models, and c) the 27 actual locations after Pearch (2011).

One other potential source of variation that could have affected both modelling methods was the underlying identities of the museum specimens. In Supplementary Information 4.B.b, I presented evidence of collector variation in species identification, after accounting for variation in geographical and topographical collecting effort. Goodwin *et al.* (2015) reported that 58% of 4500 specimens of African gingers in 40 herbaria had incorrect names and they concluded that, worldwide, more than 50% of tropical specimens were incorrectly named. If some of the specimens used in these models were incorrectly identified, then the predictive maps generated from them could have large errors, especially for allopatric species. This issue could be explored profitably using simulation models, by randomly or systematically misidentifying specimens.

The final stage of analysis presented in this chapter was to derive predictions of likely presence of the species of interest in the current study (Table 4.7). Despite the reservations expressed above

regarding the MaxEnt models, I decided to combine the two models for each species to create ensemble models (Grenouillet *et al.*, 2011); Liu *et al.* (2018). For six of the species of interest, these were used to compare probabilities with other con-generic species, and for six others they were used to compare absolute probabilities between squares where I did and did not recorded them. In both cases, therefore, any aberrations could be considered relative. The final probabilities of occurrence of the species shown in Table 4.7 are described below.

4.4.3.1 *Neodon sikimensis*

The first group in the table shows *N. sikimensis* and three other species of Cricetidae for which the predictive models resolved. In nine out of the ten squares, this species had the highest predicted probability of occurrence and it was actually recorded in three of these squares. From the predicted distributions, it seems highly likely that the five vole specimens recorded in this study were *N. sikimensis*.

4.4.3.2 *Apodemus gurkha*

A similar situation occurred in this group, with eight grid squares having the highest predicted probability of occurrence for this species. *A. gurkha* was recorded in four squares in the current study, all from the Pipar or Annapurna transects, and all with this species as the highest predicted probabilities. *A. gurkha* seems to be a highly likely recorded species.

4.4.3.3 *Mus booduga*

Four species of *Mus* had good predictive models and the highest probabilities for all squares was unequivocally *M. musculus*. The species identified in the field was *M. booduga*. The highest predicted likelihood of occurrence for this species was only 15.4% and in eight squares it had probabilities of <5%. Based on geographic and topographic variables, it seems unlikely that the small *Mus* recorded in this study was *M. booduga*.

4.4.3.4 *Rattus nitidus*

This group had four species with good models, three of which had the highest probability in at least one grid square. However, seven squares predicted *R. pyctoris*, two predicted *R. tanezumi* and only one had *R. nitidus* as the higher probability. However, it was either the first or second highest probability in seven squares, two of which it was recorded in in this study. The evidence is equivocal for *R. nitidus*.

4.4.3.5 *Episoriculus leucops*

Only 20 specimens of this species were recorded during this study in five different grid squares. However, a very similar congeneric (*E. macrurus*) was known to exist in the Himalayan region, both of which had similar predictive models. Consequently, each species had the highest probability of occurrence in five out of the ten squares, although in all cases they were very similar. In the three squares where *E. leucops* was found when *E. macrurus* had a higher probability, the differences in predictions were minimal. I consider the identification of *E. leucops* is likely to be correct.

4.4.3.6 *Sorex bedfordiae*

Five species of *Sorex* had good predictive maps, and in eight of the ten squares, *S. bedfordiae* had the highest probability of occurrence. The single specimen of this species was recorded in Sagarmatha, in one of the two squares where it did not have the highest probability.

4.4.3.7 *Niviventer*

This group included three species of *Niviventer*, all of which were putatively identified in the field. Consequently, there is no highlighting of the species with the highest predicted probabilities. The most important result here is that, with only two exceptions, the eleven grid square / species combinations all had >25% probability of occurrence, so there is no reason to reject the probability that all three species were present in the current study.

4.4.3.8 *Confirmed Soricid species*

The final group presents three species for which there is no doubt about the identifications. *E. caudatus* was recorded in every square and the lowest predicted probability of occurrence was 31%. *S. nigrescens* was found in nine out of ten grid squares. The only square where it was not recorded was the one with the lowest probability of occurrence. Finally, *S. murinus* was only recorded in two squares, but these were the two with by far the highest predicted probability of occurrence. The results from these three species in particular indicate that this method is robust.

4.5 References

Abe, H. (1971) Small Mammals of Central Nepal. *Journal of the Faculty of Agriculture, Hokkaido University* **56**: 367-423.

Anderson, R.P. (2003) Real vs. artefactual absences in species distributions: tests for *Oryzomys albigularis* (Rodentia: Muridae) in Venezuela. *Journal of Biogeography* **30**: 591-605.

Anderson, R.P., Gómez-Laverde, M., and Peterson, A.T. (2002) Geographical distributions of spiny pocket mice in South America: insights from predictive models. *Global Ecology & Biogeography* **11**: 131-141.

Anon. (2017) Statistica (data analysis software system) v.13. TIBCO Software Inc. http://statistica.io.

Anon. (2018a) QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

Anon. (2018b) SQL Server Express 2017. Microsoft Corp.

Arnold, E.A., Andersen, E.M., Taylor, M.J., and Steidl, r.J. (2017) Using cytochrome b to identify nests and museum specimens of cryptic songbirds. *Conservation Genetic Resources* **9**: 451-458.

Bajaru, S. (pers. comm.) Catalogue of the mammal collection of the BNHS.(Eds.).

Barbet-Massin, M., Jiguet, F., Albert, C.H., and Thuiller, W. (2012) Selecting pseudo-absences for species distribution models: how, where and how many? *Methods Ecol Evol* **3**: 327-338.

Bean, W.T., Stafford, R., and Brashares, J.S. (2012) The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. *Ecography* **35**: 250-258.

Corbet, G.B. and Hill, J.E. (1992) *The Mammals of the Indomalayan Region: A systematic Review.* Oxford University Press. pp 488.

Daniel, M. (2015) Expansion of small terrestrial mammals and their parasites into the Barun Valley (Makalu Mt. Region, Nepal Himalaya) linked with changes in glaciation and human activities. *Journal of Mountain Science* **12**: 14-29.

Dicko, A.H., Lancelot, R., Seck, M.T., Guerrini, L., Sall, B., Lo, M., Vreysen, M.J., Lefrancois, T., Fonta, W.M., Peck, S.L., and Bouyer, J. (2014) Using species distribution models to optimize vector control in the framework of the tsetse eradication campaign in Senegal. *Proc Natl Acad Sci U S A* **111**: 10149-10154.

Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M., and E. Zimmermann, N. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**: 129-151.

Goodwin, Z.A., Harris, D.J., Filer, D., Wood, J.R.I., and Scotland, R.W. (2015) Widespread mistaken identity in tropical plant collections. *Current Biology* **25**: R1066-R1067.

Grenouillet, G., Buisson, L., Casajus, N., and Lek, S. (2011) Ensemble modelling of species distribution: the effects of geographical and environmental ranges. *Ecography* **34**: 9-17.

Guisan, A. and Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling* **135**: 147-186.

Holmes, M.W., Hammond, T.T., Wogan, G.O.U., Walsh, R.E., LaBarbera, K., Wommack, E.A., Martins, F.M., Crawford, J.C., Mack, K.L., Bloch, L.M., and Nachman, M.W. (2016) Natural history collections as windows on evolutionary processes. *Molecular Ecology* **25**: 864-881.

Liu, C., Newell, G., and White, M. (2018) The effect of sample size on the accuracy of species distribution models: considering both presences and pseudo-absences or background sites. *Ecography* **In Press**.

Lobo, J.M. and Tognelli, M.F. (2011) Exploring the effects of quantity and location of pseudo-absences and sampling biases on the performance of distribution models with limited point occurrence data. *Journal for Nature Conservation* **19**: 1-7.

Mateo, R.G., Felicísimo, Á.M., Pottier, J., Guisan, A., and Muñoz, J. (2012) Do Stacked Species Distribution Models Reflect Altitudinal Diversity Patterns? *PLoS One* **7**: e32586.

Merow, C., Smith, M.J., and Silander, J.A. (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* **36**: 1058-1069.

Mitchell, R.M. (1977) *Accounts of Nepalese mammals and analysis of the host-ectoparasite data by computer techniques*. PhD Thesis. Iowa State University, Ames, Iowa, USA.

Moritz, C., Patton, J., Conroy, C., Parra, J., White, G., and Beissinger, S. (2008) Impact of a Century of Climate Change on Small-Mammal Communities in Yosemite National Park, USA. *Science* **322**: 261-264.

Nelson, B., A. C. Ferreira, C., F. da Silva, M., and L. Kawasaki, M. (1990) *Endemism centres, refugia and botanical collection density in Brazilian Amazonia*. pp 714-716.

Niethammer, J. and Martens, J. (1975) Die Gattungen *Rattus* and *Maxomys* in Afghanistan und Nepal. *Zeitschrift für Säugetierkunde* **40**: 325-355.

Parshad, V.R. (1999) Rodent Control in India. *Integrated Pest Management Reviews* **4**: 97-126.

Pearce, J. and Ferrier, S. (2000) An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling* **128**: 127-147.

Pearch, M.J. (2011) A review of the biological diversity and distribution of small mammal taxa in the terrestrial ecoregions and protected areas of Nepal. *Zootaxa* **3072** 1-286.

Pergams, O.R.W. and Lawler, J.J. (2009) Recent and widespread rapid morphological change in rodents. *PLoS One* **4**: e6452-e6452.

Phillips, S.J., Anderson, R.P., and Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**: 231-259.

Phillips, S.J. and Dudík, M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**: 161-175.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., and Ferrier, S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications* **19**: 181-197.

Phillips, S.J., Dudík, M., and Schapire, R.E. (2004) A Maximum Entropy Approach to Species Distribution Modeling pp 655-662 *in Proceedings of 21st Internation Conference on Machine Learning*, Banff, Canada.

Phillips, S.J., Dudík, M., and Schapire, R.E. (2018) Maxent software for modeling species niches and distributions (Version 3.4.1). Available from url: http://biodiversityinformatics.amnh.org/open_source/maxent/.

Ponder, W.F., Carter, G.A., Flemons, P., and Chapman, R.R. (2001) Evaluation of Museum Collection Data for Use in Biodiversity Assessment. *Conservation Biology* **15**: 648-657.

Renner, I.W. and Warton, D.I. (2013) Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics* **69**: 274-281.

Stockwell, D.R.B. and David, P. (1999) The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science* **13**: 143-158.

Stockwell, D.R.B. and Peterson, A.T. (2002) Effects of sample size on accuracy of species distribution models. *Ecological Modelling* **148**: 1-13.

Wieczorek, J. (2015) Mammal Networked Information System: MaNIS. http://manisnet.org/

Wroughton, R.C. (1912) Bombay Natural History Society's Mammal Survey of India: Report 3. *Journal of the Bombay Natural History Society* **21**: 826-844.

Yackulic, C.B., Chandler, R., Zipkin, E.F., Royle, J.A., Nichols, J.D., Campbell Grant, E.H., and Veran, S. (2013) Presence-only modelling using MAXENT: when can we trust the inferences? *Methods Ecol Evol* **4**: 236-243.

Zar, J.H. (1984) *Biostatistical Analysis*. Second edition. Prentice-Hall, New Jersey.

SUPPLEMENTARY INFORMATION

4.A Species Breakdown of BNHS and Pearch Datasets

Table 4.8a Species breakdown of BNHS and Pearch Datasets for Cricetidae and Muridae.

			Data Records		Presence in 0.1° Grid Squares		
			BNHS Specimens	Pearch Geo-cells	BNHS	Pearch	Combined
Cricetidae	*Alticola*	*A. roylei*	37	0	7	0	7
		A. stoliczkanus	3	11	0	8	8
	Clethrionomys	*C. rufocanus*	3	0	0	0	0
	Cremnomys	*C. cutchicus*	154	0	18	0	18
	Cricetulus	*C. barabensis*	2	0	0	0	0
		C. migratorius	146	0	1	0	1
	Eothenomys	*E. melanogaster*	5	0	1	0	1
	Microtus	*M. arvalis*	2	0	0	0	0
		M. gregalis	9	0	0	0	0
		M. Irani	2	0	0	0	0
		M. montebelli	2	0	0	0	0
	Neodon	*N. leucurus*	2	5	0	3	3
		N. sikimensis	12	62	3	29	32
Muridae	*Apodemus*	*A. gurkha*	3	44	2	16	17
		A. pallipes	0	15	0	13	13
		A. rusiges	224	0	12	0	12
		A. sylvaticus	11	0	2	0	2
		A. wardi	23	0	9	0	9
	Bandicota	*B. bengalensis*	512	0	90	0	90
		B. indica	99	0	36	0	36
		B. savilei	7	0	1	0	1
	Berylmys	*B. bowersi*	5	0	0	0	0
		B. mackenziei	9	0	0	0	0
		B. manipulus	30	0	5	0	5
	Diomys	*D. crumpi*	0	1	0	1	1
	Golunda	*G. ellioti*	206	3	40	2	42
	Leopoldamys	*L. edwardsi*	7	0	5	0	5
		L. sabanus	8	0	6	0	6
	Maxomys	*M. rajah*	146	0	16	0	16
	Millardia	*M. meltada*	0	6	0	5	5
	Mus	*M. booduga*	521	17	80	13	93
		M. cervicolor	57	29	14	23	37
		M. cookii	81	12	12	12	24
		M. musculus	1460	78	95	47	142
		M. pahari	24	1	9	1	10
		M. phillipsi	67	0	16	0	16
		M. platythrix	190	0	27	0	27
		M. saxicola	85	5	13	3	16
		M. shortridgei	15	0	2	0	2
		M. terricolor	0	4	0	4	4
	Nesokia	*N. indica*	50	4	8	3	11
	Niviventer	*N. confucianus*	23	0	2	0	2
		N. eha	10	66	2	30	32
		N. fulvescens	236	39	34	23	56
		N. niviventer	45	41	5	27	31
	Rattus	*R. exulans*	99	0	17	0	17
		R. nitidus	163	25	24	18	42
		R. norvegicus	30	0	6	0	6
		R. pyctoris	225	56	38	37	75
		R. rattus	371	47	71	37	108
		R. satarae	4	0	3	0	3
		R. tanezumi	969	62	90	37	122
	Vandeleuria	*V. oleracea*	0	4	0	3	3

Table 4.8b Species breakdown of BNHS and Pearch Datasets for Soricidae

			Data Records		Presence in 0.1° Grid Squares		
			BNHS Specimens	Pearch Geosquares	BNHS	Pearch	Combined
Soricidae	*Anourosorex*	*A. squamipes*	49	0	5	0	5
	Chimarrogale	*C. himalayica*	2	2	2	2	4
	Crocidura	*C. attenuata*	85	6	23	4	27
		C. fuliginosa	3	0	1	0	1
		C. horsfieldii	2	2	2	2	4
		C. pergrisea	1	0	1	0	1
	Episoriculus	*E. caudatus*	40	89	5	41	45
		E. leucops	3	15	2	8	10
		E. macrurus	0	13	0	9	9
	Nectogale	*N. elegans*	10	2	4	2	6
	Sorex	*S. araneus*	0	3	0	3	3
		S. bedfordiae	0	12	0	8	8
		S. cylindricauda	0	3	0	3	3
		S. minutus	0	8	0	8	8
		S. thibetanus	0	3	0	3	3
	Soriculus	*S. nigrescens*	51	106	11	45	55
	Suncus	*S. etruscus*	34	8	11	6	17
		S. montanus	30	0	7	0	7
		S. murinus	556	54	111	35	146
		S. stoliczkanus	25	3	14	3	17
	Species Counts	73	62	40			

4.B Analysis of Collector Effects on the BNHS Specimens

As discussed in Chapter 2, accurate identification to species level is often difficult in the field. In particular, the genera *Apodemus*, *Mus* and *Rattus* often require detailed anatomical inspection of skulls and dentition. This analysis attempted to quantify the degree of variation in species identifications between the major collectors of the BNHS specimens.

Of the 7285 specimen records, 7010 had collector identities. However, only 10 of the 106 collectors had more than 100 records (known hereafter as "super-collectors") and only three, NA Baptista, CA Crump and HW Wells had more than 1000.

4.B.a Collectors' Efforts by Year

Firstly, I have cross-tabulated the super-collectors by year, to indicate when they made their major collection efforts (Table 4.9). From this it is clear that Crump and, to a lesser extent Shortridge, were the major contributors to the 1912–15 national survey. Although Baptista was the second most prolific collector, he worked primarily in three different periods. Despite this he collected over 1000 specimens in the two years 1915–6. Hotson worked between 1917 and 1920, but only in Iran. Wells was the major collector in the 1920s.

4.B.b Species Identifications by Collectors

A preliminary analysis of the data suggested that different species were recorded by different collectors. Of course, this could have been because they were working in different areas. For example, the four major collectors worked in almost entirely exclusive areas (Figure 4.22). Indeed, the only square that all four worked in was 27N 88E (Darjeeling, *e.g.* the blue square on the eastern border of Nepal), although Wells and Shortridge only collected 5 and 1 specimen respectively. However, all collectors had wide geographical spreads with a large degree of overlap, so it was possible to account for latitude, longitude and altitude to investigate residual variation in species identifications. I have analysed three genera of murids; *Apodemus*, *Mus* and *Rattus*, as these were recorded in sufficient numbers in the BNHS dataset and were particularly relevant to the putative species caught in the current study

To achieve this all records from nine super-collectors were selected. (Hotson, who only worked in two squares in Iran, was excluded.) For each of the three genera separately, I have tallied the number of specimens of a particular species and all the other species in the genus, by collector, latitude, longitude and altitude. Latitude and longitude were tallied at 1° grid square resolution and altitude at 100m intervals. For each collector, within each geo-cell, I have expressed the numbers of

specimens as the log-odds (n + 1) between the species of interest and the others. For example, in geo-cell 26N 85E 0A, NA Baptista collected two M. *musculus* and nine other *Mus* specimens, which gave a log-odds of

$$Log_n \frac{(2+1)}{(9+1)} = -1.204$$

In contrast, in geo-cell 26N 88E 1200A, he caught three M. *musculus* and no other species which gave a value of 1.386. So, these values are a unit-less index of the relative abundance (RAI) of the species of interest against all the other species in the genus. This has been used as the response variable in the analyses described below.

Table 4.9 Cross-tabulation of ten super-collectors by year. (Numbers of specimens > 25 are high-lighted in red).

Year	Baptista NA	Crump CA	Hotson JEB	Mackenzie JMD	Mayor EW	O'Brien JR	Prater SH	Primrose CM	Shortridge GC	Wells HW	Totals
1903	0	0	0	0	0	0	1	0	0	0	1
1906	0	2	0	0	0	0	0	0	0	0	2
1907	0	3	0	0	0	0	0	0	1	0	4
1908	0	0	0	0	0	0	4	0	1	0	5
1909	0	0	0	0	0	0	1	0	0	0	1
1910	0	1	0	0	0	0	0	0	0	0	1
1911	1	18	0	0	7	0	0	0	60	0	86
1912	2	248	0	0	0	0	1	0	208	2	461
1913	0	524	0	14	98	0	0	0	269	0	905
1914	7	500	0	5	88	0	69	0	90	3	762
1915	572	150	0	38	0	0	23	0	2	5	790
1916	443	3	0	20	0	0	0	0	1	0	467
1917	0	0	4	0	0	0	19	0	3	0	26
1918	0	0	90	0	0	0	0	0	0	0	90
1919	0	1	373	0	0	0	0	0	0	25	399
1920	5	1	113	0	0	1	0	0	0	278	398
1921	38	0	0	0	0	61	0	94	2	296	491
1922	55	3	0	0	0	188	9	90	0	394	739
1923	18	4	0	0	0	0	0	2	0	297	321
1924	0	0	0	0	0	0	0	3	0	0	3
1925	1	0	0	0	0	0	0	0	0	4	5
1926	1	0	0	0	0	0	0	0	0	0	1
1927	0	0	0	57	0	0	0	0	1	0	58
1928	1	0	0	4	0	0	0	0	0	0	5
1929	86	0	0	1	0	0	0	0	0	0	87
1930	89	0	0	0	0	0	0	0	0	1	90
All Years	1319	1458	580	139	193	250	127	189	638	1305	6198

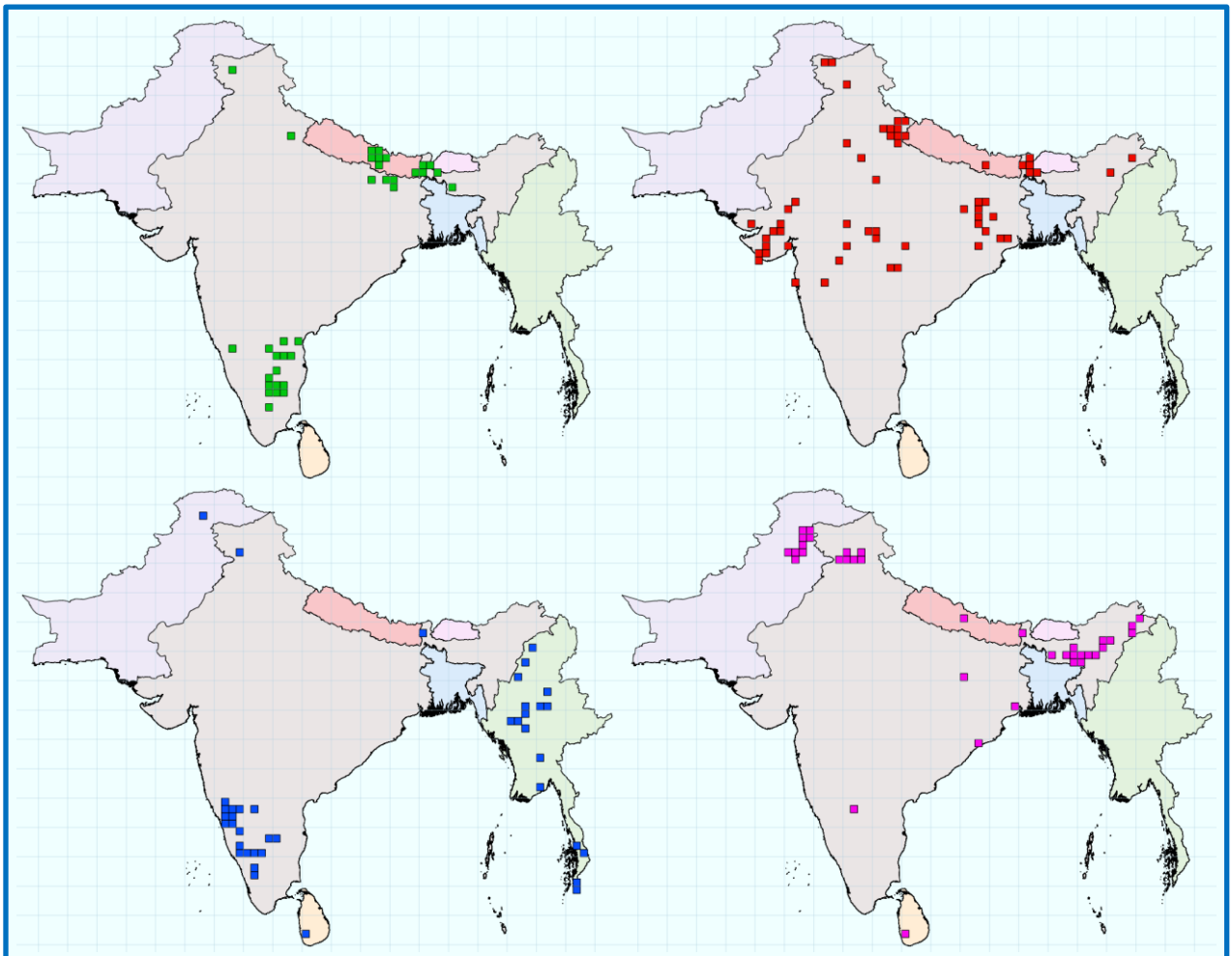


Figure 4.22 Location of effort by four major collectors; NA Baptista (green), CA Crump (red), GC Shortridge (blue) and HW Wells (magenta). Squares are 0.5° latitude / latitude.

Apodemus

This genus only had 261 specimens and four species, dominated by *A. rusiges* with 224 specimens (86%), so this was the only species analysed within the genus. An ANCOVA model with three continuous predictors (latitude, longitude and altitude) and one categorical predictor (collector) was used with the RAI for *A. rusiges* as the response variable. Type I sums-of-squares, which partitions the overall model sums-of-squares (SS) sequentially, was used to calculate F-ratios. In this way, it was possible to account for the three geographical covariates first and then calculate the SS for collector from the remaining SS, *i.e.* I removed the geographical effects first and then tested for the specific effect of collector.

The genus was only collected by Baptista, Crump and Wells, which yielded only 18 collector / geo-cell combinations, so the model was not especially robust. However, it did give a marginally significant collector effect ($F_{(2, 12)} = 5.57$, $p \approx 0.019$). The means plot (Figure 4.23) can be explained by

reference to the actual captures; Baptista never described this species, Wells only recorded this species (189 specimens) of *Mus* and Crump collected both this species and other con-generics.

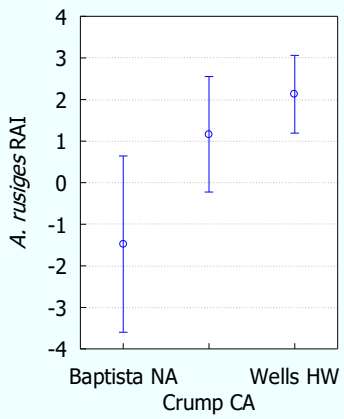


Figure 4.23 Collector identification effect for *A. rusiges*. Least squares means and 95% C. I.

Rattus

This genus had four species of interest; *nitidus* (163), *pyctoris* (225), *rattus* (371) and *tanezumi* (969), which together account for 93% of *Rattus* specimens. The RAI was calculated separately for each species, based on records from 173 collector / geo-cell combinations. These were then standardised within species and stacked into a single dataset with a species factor identifying them. A more complex model was used to analyse these four species together, although it was intrinsically the same as that used for *Apodemus*. In this case, however, species was the only main effect included in the model, plus the interaction between species and the other four effects. Now, the only term of interest was the interaction between species and collector, so by using Type I SS, the variances attributable to the other terms in the model were removed first.

Because the RAI was standardised within species, the Intercept and species terms in this model had F-ratios of 0.0 – they had no variance. However, all the other four terms were very highly significant. (Table 4.10). Most importantly, even after removing the species × geographic effects, the 2-way interaction of species × collector were still very highly significant (Figure 4.24). This indicated that the tendency for different species to be recorded differed between observers. I used planned deviance contrasts, within collectors, to assess the significance of the individual species deviances from the observer mean. In this case, only four collectors showed significant effects between species (Table 4.11), although the strongest effects were seen in Baptista and Crump. They both had a significantly lower than average RAI for *R. pyctoris*, but Baptista had a significantly higher than average RAI for

R. tanezumi whilst Crump recorded a higher than average RAI for *R. rattus*. In contrast Primrose and Wells both recorded a higher than average RAI for *R. pyctoris*.

Table 4.10 ANCOVA table for four species of *Rattus*. The response variable is the standardised RAI for each species within the genus.

Effect	SS	DF	MS	F	*p*
Intercept	0.00	1	0.00	0.00	1.0
Sp	0.00	3	0.00	0.00	1.0
Sp x Lat	66.12	3	22.04	33.88	0.0
Sp x Long	120.35	3	40.12	61.66	0.0
Sp x Alt	16.11	3	5.37	8.25	$<10^{-4}$
Sp x Collector	59.29	24	2.47	3.80	$<10^{-8}$
Error	426.13	655	0.65		

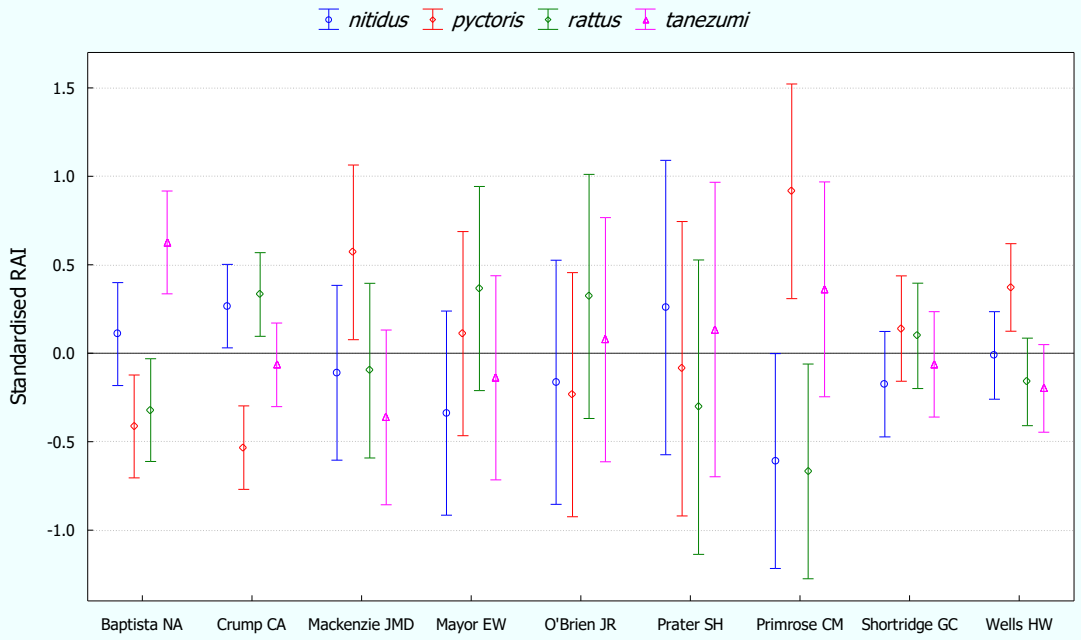


Figure 4.24 The 2-way interaction of four *Rattus* sp. by collector. The response variable was standardised RAI – LS means and 95% C. I. are shown.

Table 4.11 Overall deviance contrasts within collectors between species of the genus *Rattus*.

Collector	SS	DF	MS	F	*p*
Baptista	15.77	3	5.26	8.08	$<10^{-4}$
Crump	16.96	3	5.65	8.69	$<10^{-4}$
Primrose	9.21	3	3.07	4.72	<0.005
Wells	6.67	3	2.22	3.42	<0.02
Error	426.13	655	0.65		

Interpreting the interaction in the other dimension, three species showed a significant preferential collector effect (*R. pyctoris*, $F_{(8, 655)} = 7.17$, $p < 10^{-8}$, *R. rattus*; $F_{(8, 655)} = 3.17$, $p < 0.002$, *R. tanezumi*; $F_{(8, 655)} = 2.55$, $p < 0.0005$). There was no tendency for collectors to preferentially record *R. nitidus*.

Mus

This genus had three species of interest; M. *booduga* (521 specimens), *domesticus* (296) and *musculus* (1164). Within the constraints (of collectors and geography) this yielded 149 collector / geo-cell combinations. The same model was used as for *Rattus*, with the intercept and species each having zero variance. However, after accounting for marginally significant latitude and longitude effects, and a highly significant altitudinal effect, the collector × species interaction was very highly significant ($F_{(16, 422)} = 3.68$, $p < 10^{-5}$; Figure 4.25). Deviance contrasts for individual collectors yielded only two overall significances; Crump ($F_{(2, 422)} = 10.78$, $p < 0{\cdot}0001$) and Mayor ($F_{(2, 422)} = 11.33$, $p < 0{\cdot}0001$). Crump showed significant individual species deviances for M. *booduga* with a significantly greater tendency to be recorded ($p < 0{\cdot}0001$) and a significantly lower tendency for M. *musculus* ($p < 0{\cdot}0005$). Mayor showed the reverse tendencies; lower for M. *booduga* ($p < 10^{-4}$) and higher for M. *musculus* ($p < 10^{-4}$). Two within-species contrasts were significant; M. *booduga* ($p < 10^{-5}$) and M. *musculus* ($p < 0{\cdot}0002$), although there was no tendency for M. *domesticus* to be collected preferentially by any observer.

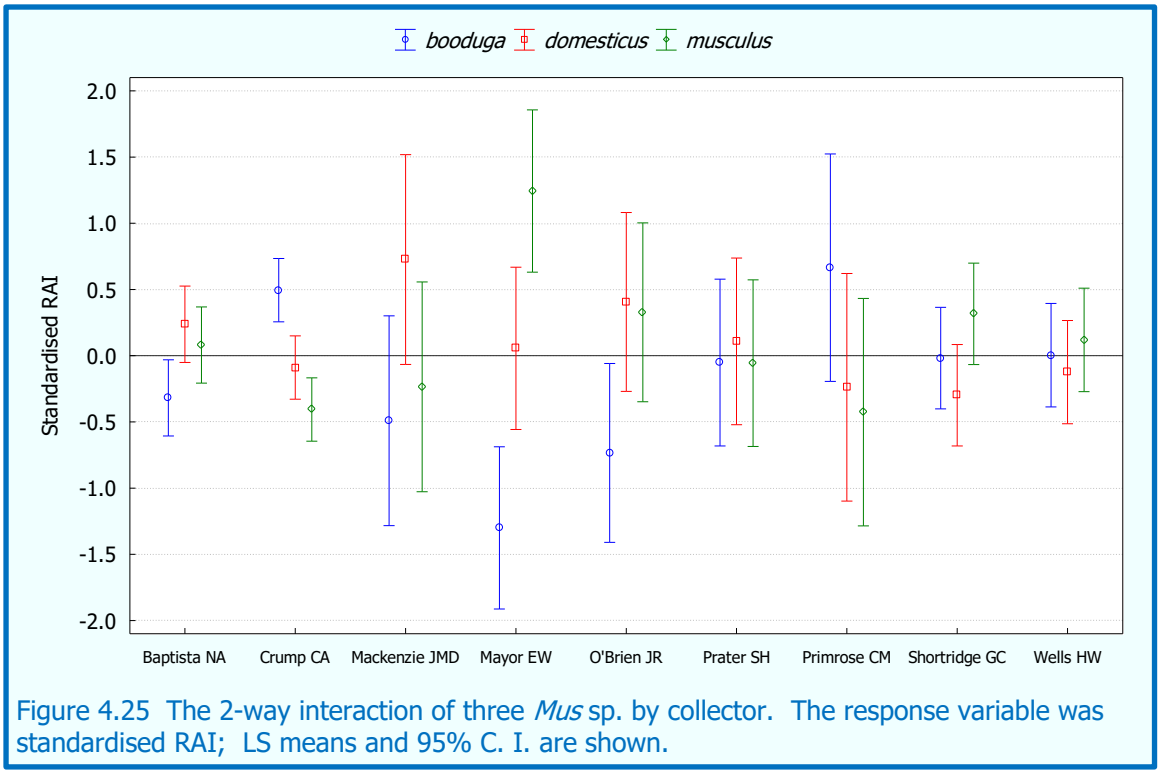


Figure 4.25 The 2-way interaction of three *Mus* sp. by collector. The response variable was standardised RAI; LS means and 95% C. I. are shown.

4.C Projections of Logistic Regression and MaxEnt Models onto the DEM of Nepal

The projections are displayed as maps with 0.1° latitude and longitude cells. Probability of occurrence is colour coded on a non-linear eight-point scale. Projections derived from the logistic regression models are shown on the left-hand side of the pair. Eleven of these models were classified as unresolved; either the models actually did not resolve at all, or they resolved with only latitude and/or longitude effects. Projections derived from the MaxEnt models are shown on the right-hand side of the pair. Seven of these models were classified as over-parameterized because, although they all resolved, the projections just generated maps with a random spatial pattern.

4.C.a Cricetidae

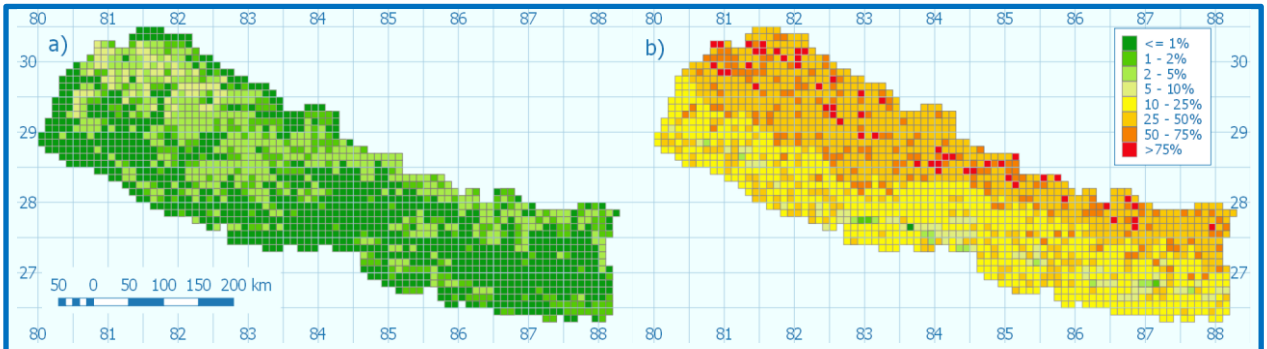


Figure 4.26 Predicted distributions in Nepal for *Alticola roylei* by a) logistic regression and b) Maxent.

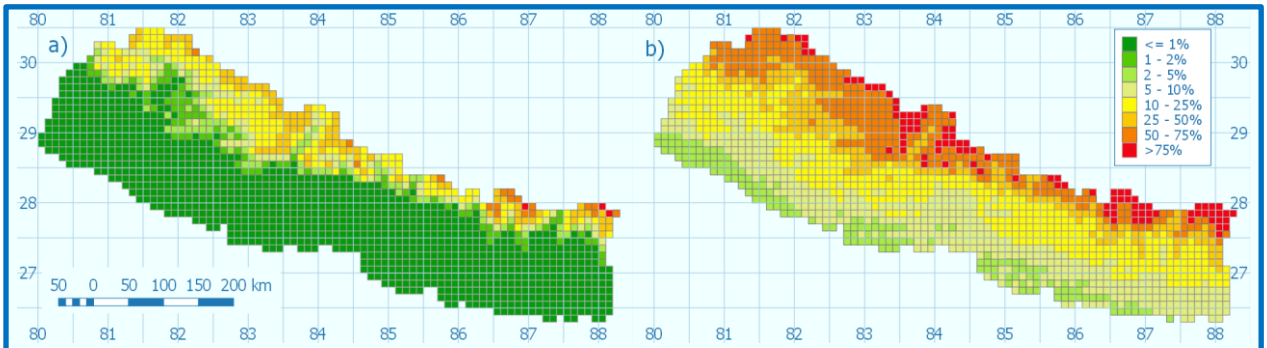


Figure 4.27 Predicted distributions in Nepal for *Alticola stoliczkanus* by a) logistic regression and b) Maxent.

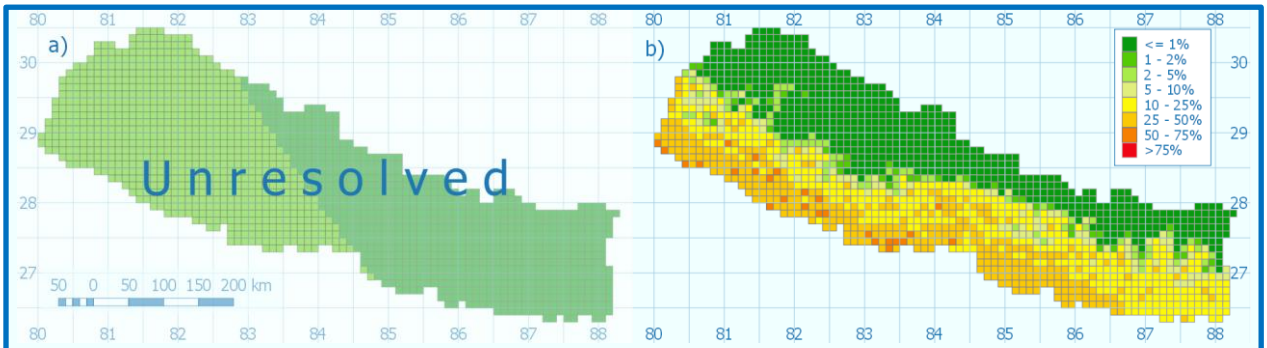


Figure 4.28 Predicted distributions in Nepal for *Cremnomys cutchicus* by a) logistic regression and b) Maxent.

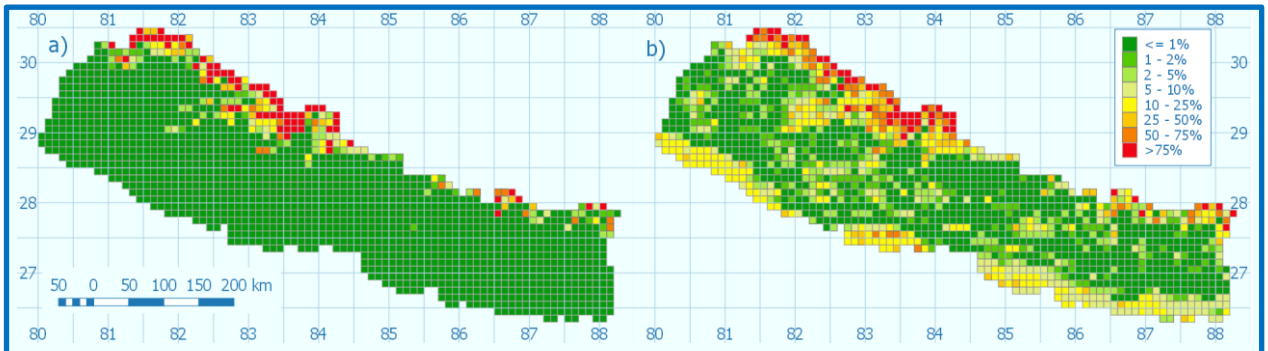


Figure 4.29 Predicted distributions in Nepal for *Neodon leucurus* by a) logistic regression and b) Maxent.

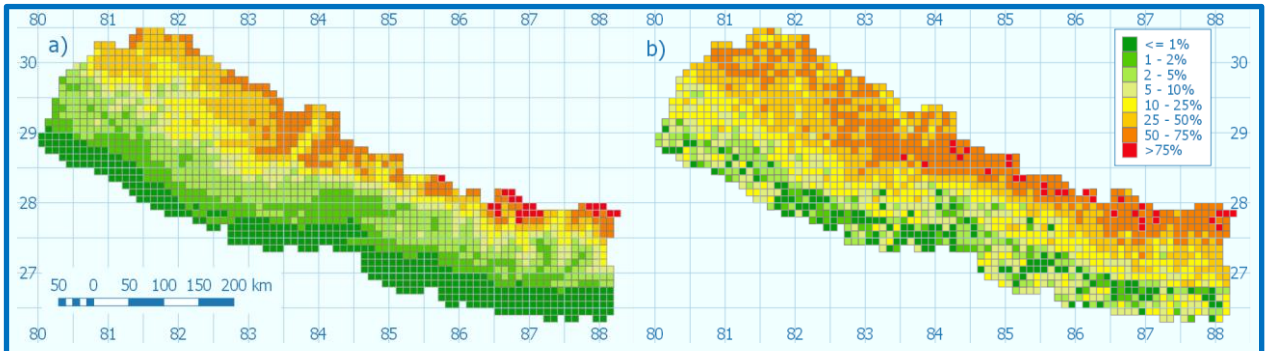


Figure 4.30 Predicted distributions in Nepal for *Neodon sikimensis* by a) logistic regression and b) Maxent.

4.C.b Muridae

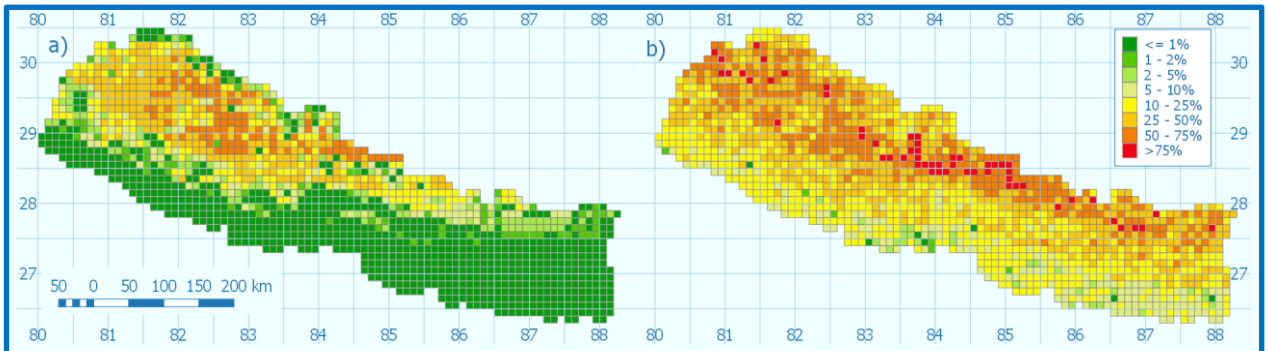


Figure 4.31 Predicted distributions in Nepal for *Apodemus gurkha* by a) logistic regression and b) Maxent.

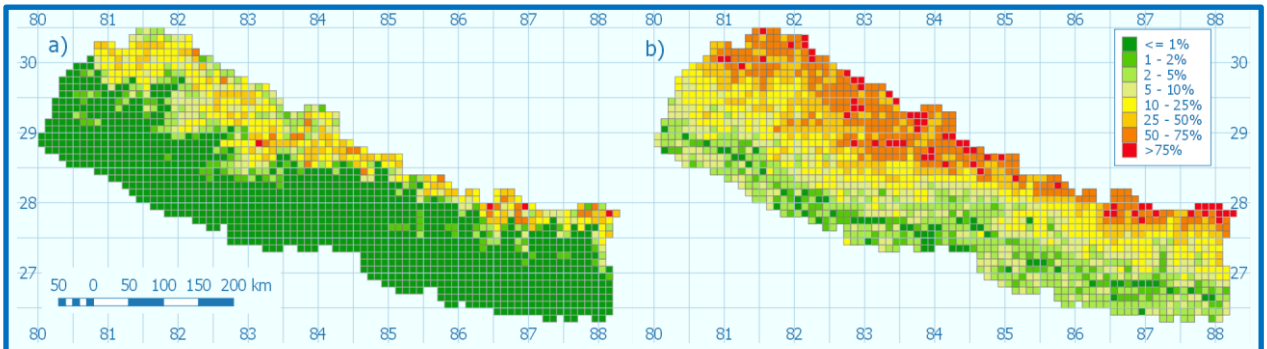


Figure 4.32 Predicted distributions in Nepal for *Apodemus pallipes* by a) logistic regression and b) Maxent.

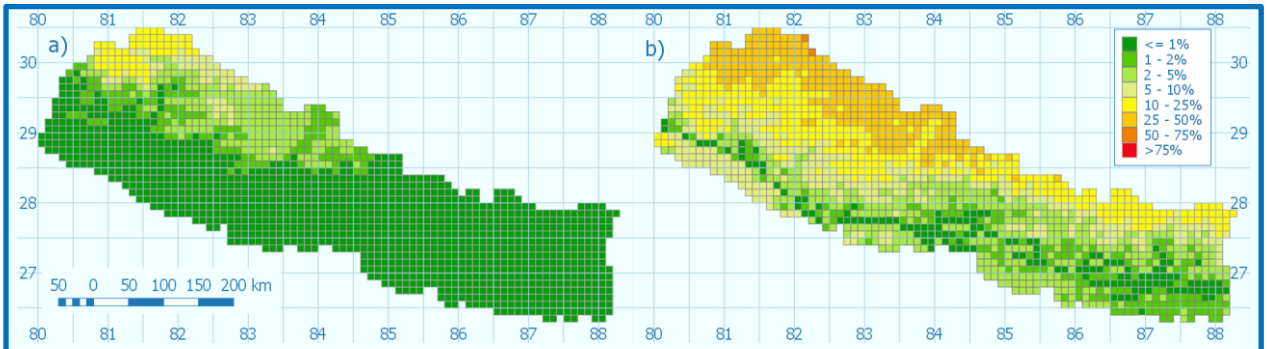


Figure 4.33 Predicted distributions in Nepal for *Apodemus rusiges* by a) logistic regression and b) Maxent.

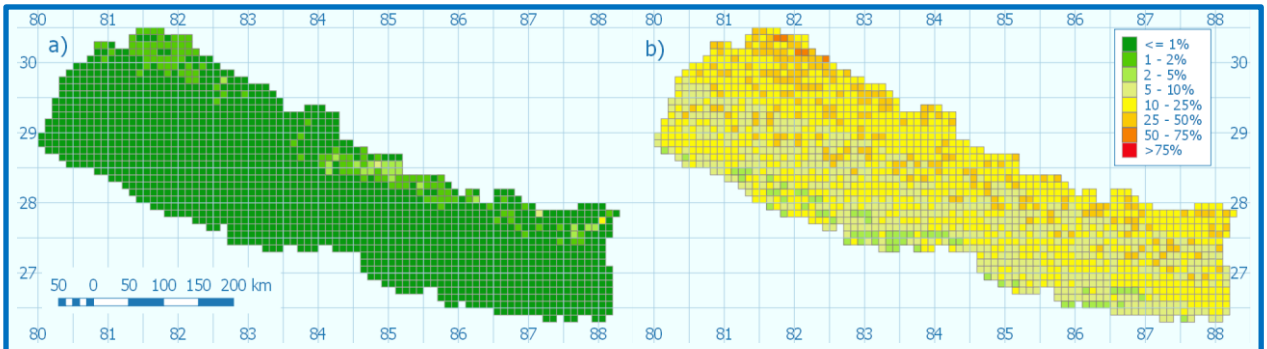


Figure 4.34 Predicted distributions in Nepal for *Apodemus wardii* by a) logistic regression and b) Maxent.

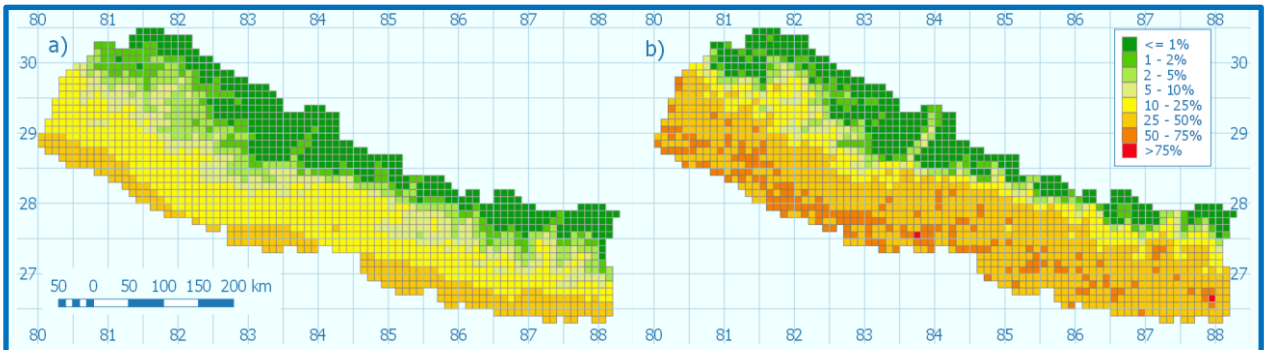


Figure 4.35 Predicted distributions in Nepal for *Bandicota bengalensis* by a) logistic regression and b) Maxent.

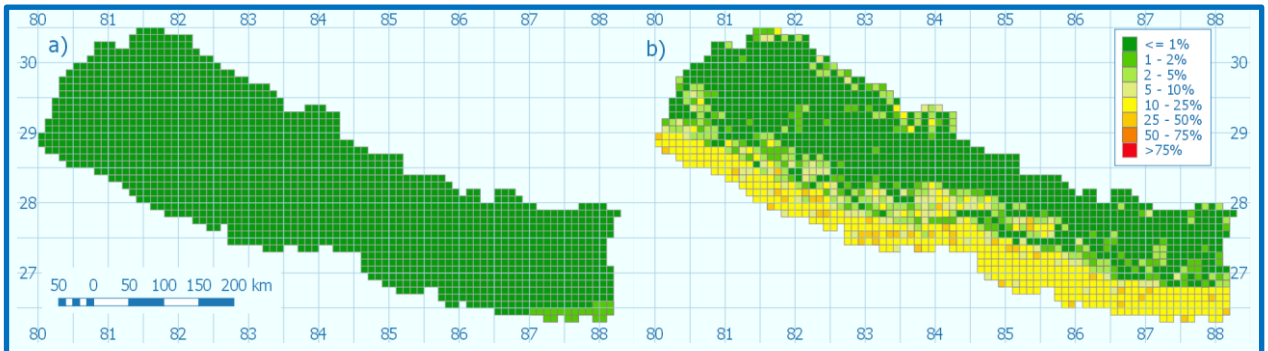


Figure 4.36 Predicted distributions in Nepal for *Bandicota indica* by a) logistic regression and b) Maxent.

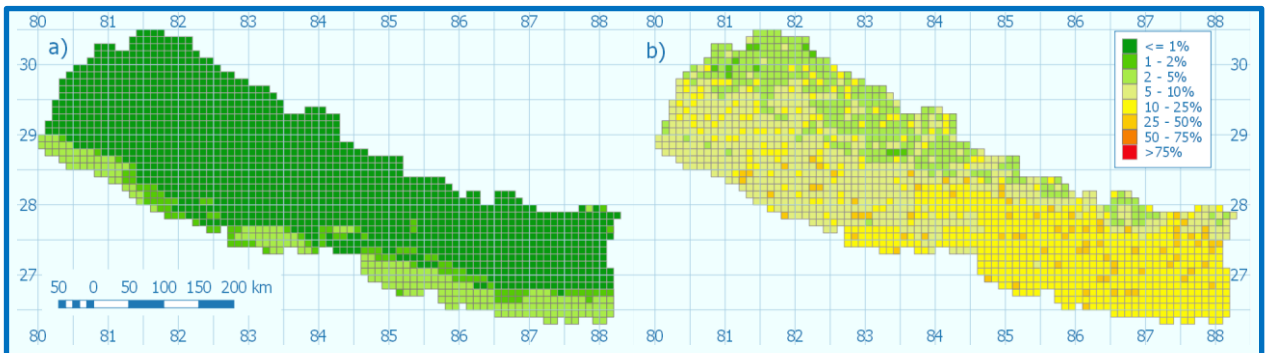


Figure 4.37 Predicted distributions in Nepal for *Berylmys manipulus* by a) logistic regression and b) Maxent.

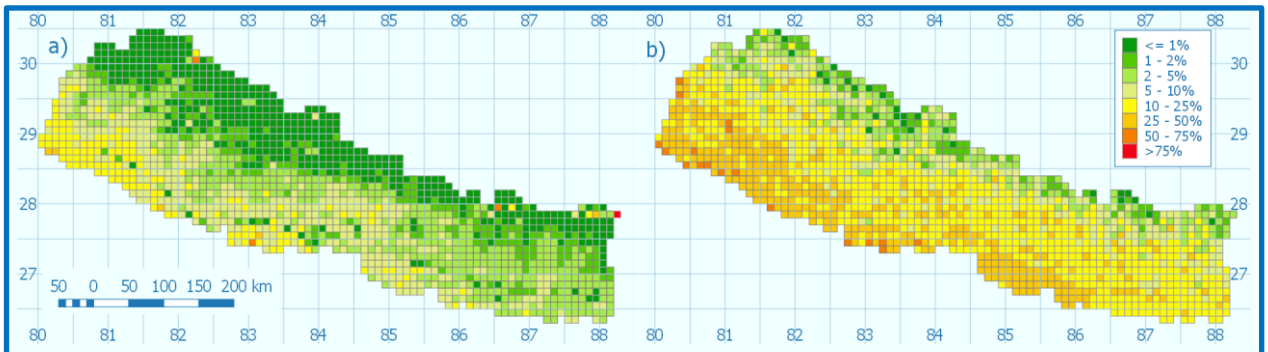


Figure 4.38 Predicted distributions in Nepal for *Golunca ellioti* by a) logistic regression and b) Maxent.

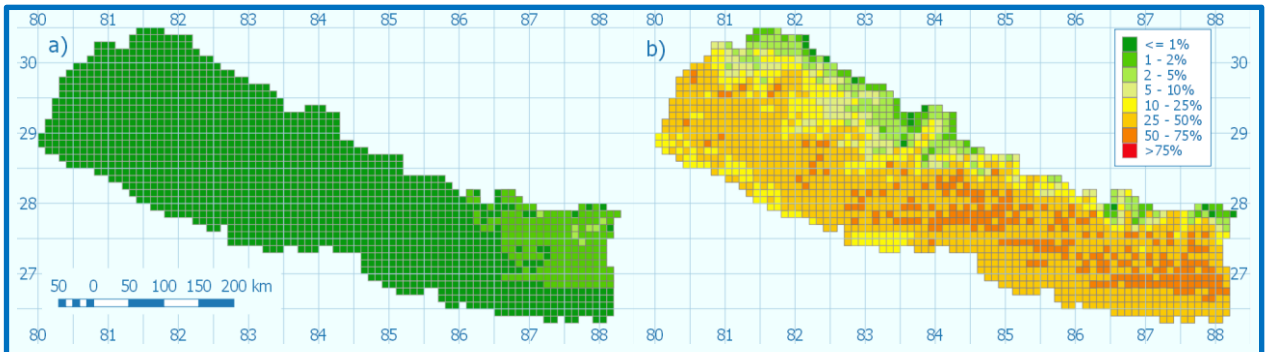


Figure 4.39 Predicted distributions in Nepal for *Leopoldamys edwardsi* by a) logistic regression and b) Maxent.

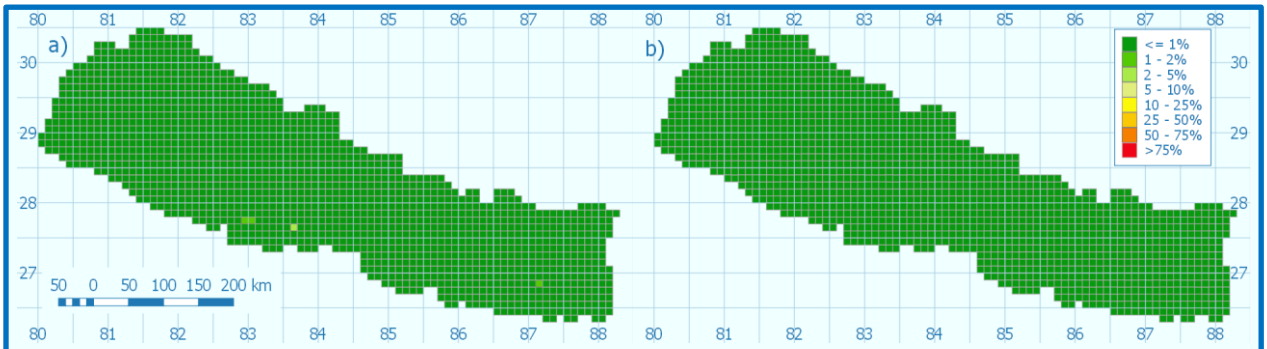


Figure 4.40 Predicted distributions in Nepal for *Leopoldamys sabanus* by a) logistic regression and b) Maxent.

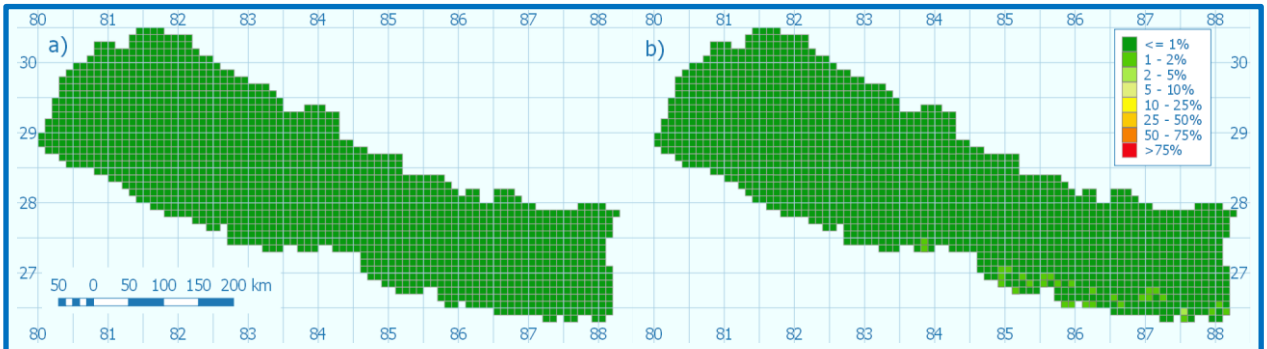


Figure 4.41 Predicted distributions in Nepal for *Maxomys rajah* by a) logistic regression and b) Maxent.

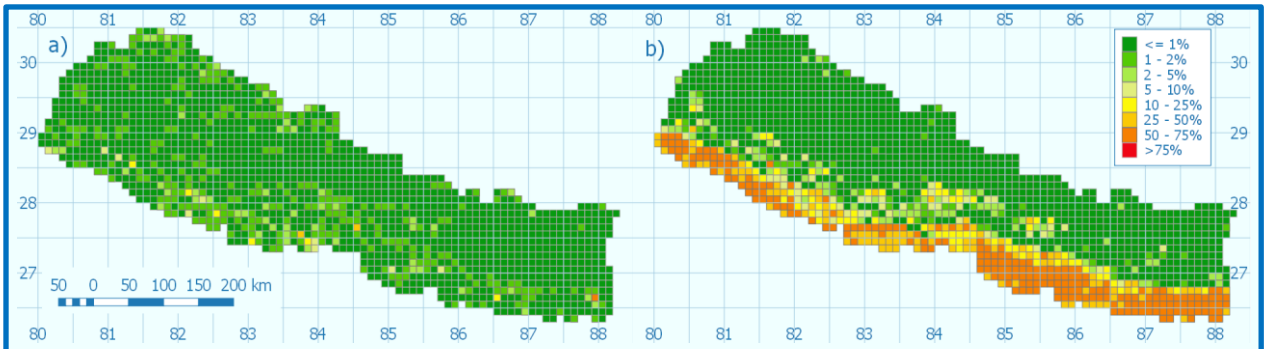


Figure 4.42 Predicted distributions in Nepal for *Millardia meltada* by a) logistic regression and b) Maxent.

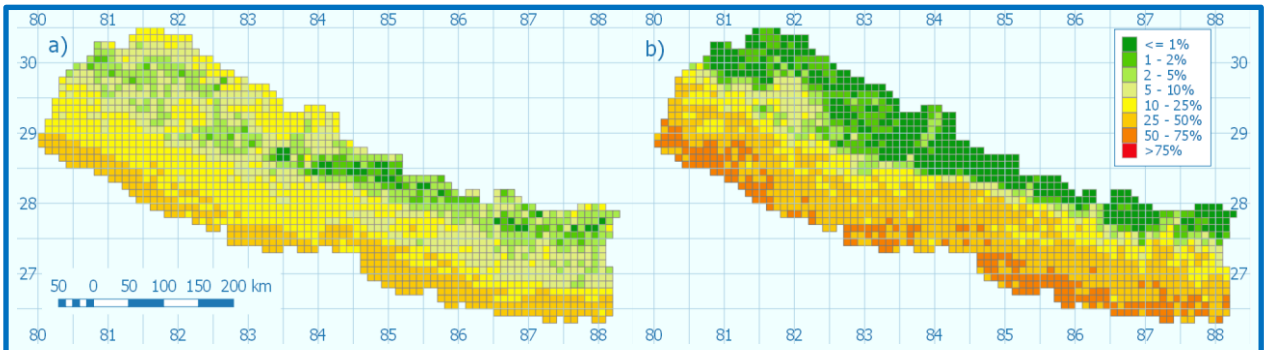


Figure 4.43 Predicted distributions in Nepal for *Mus booduga* by a) logistic regression and b) Maxent.

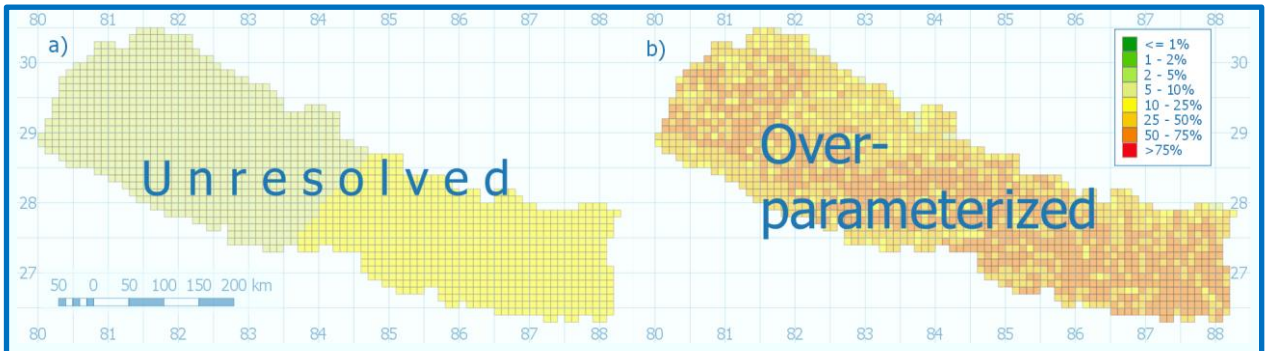


Figure 4.44 Predicted distributions in Nepal for *Mus cervicolor* by a) logistic regression and b) Maxent.

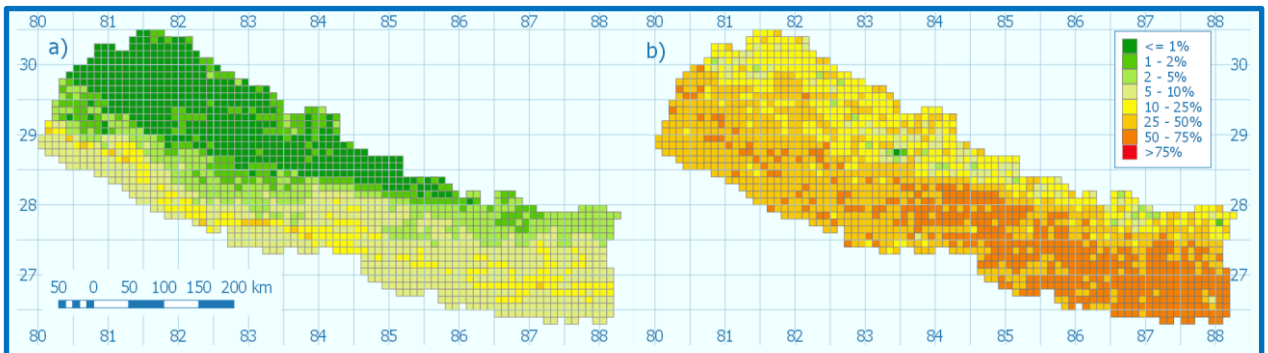


Figure 4.45 Predicted distributions in Nepal for *Mus cookii* by a) logistic regression and b) Maxent.

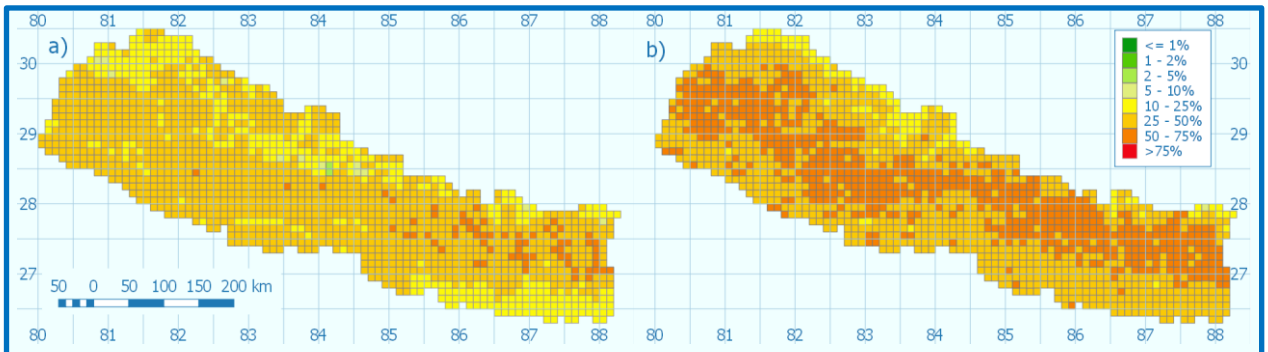


Figure 4.46 Predicted distributions in Nepal for *Mus musculus* by a) logistic regression and b) Maxent.

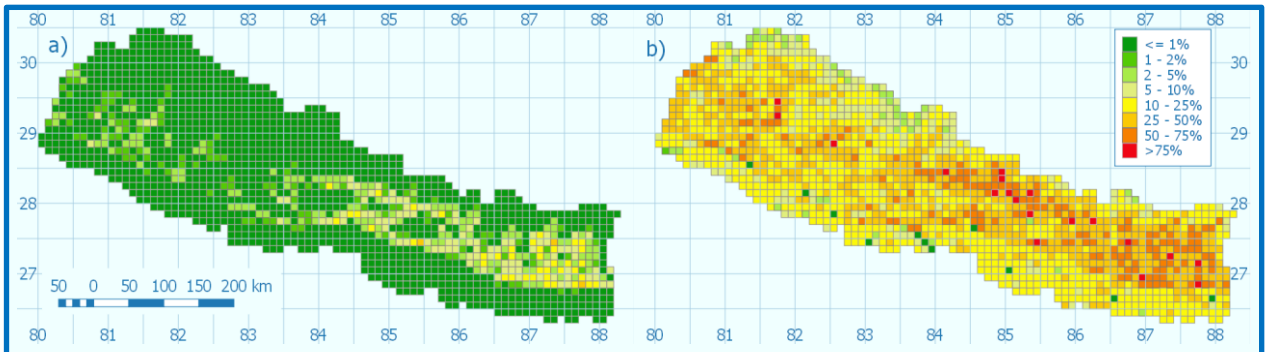


Figure 4.47 Predicted distributions in Nepal for *Mus pahari* by a) logistic regression and b) Maxent.

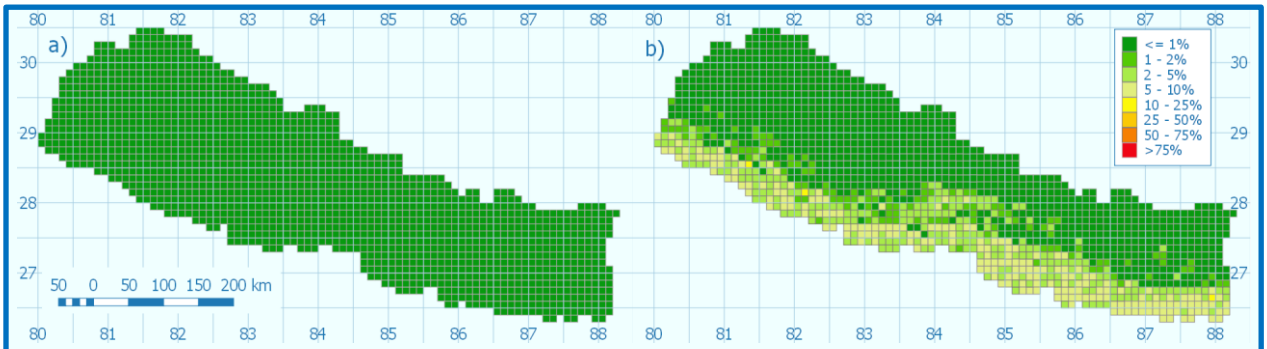


Figure 4.48 Predicted distributions in Nepal for *Mus phillisi* by a) logistic regression and b) Maxent.

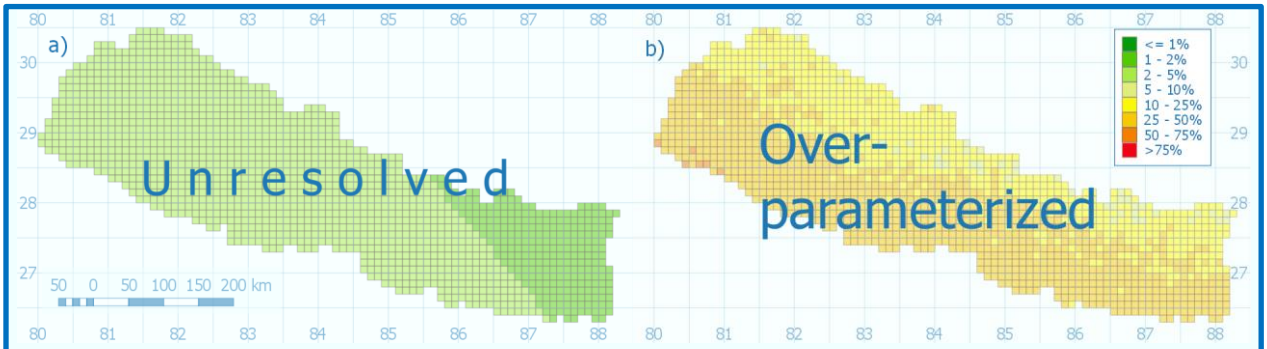


Figure 4.49 Predicted distributions in Nepal for *Mus platythrix* by a) logistic regression and b) Maxent.

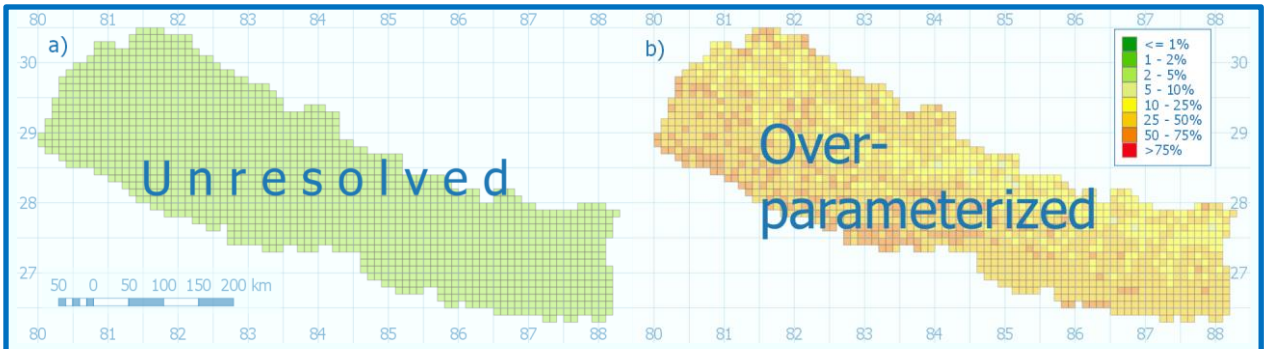


Figure 4.50 Predicted distributions in Nepal for *Mus saxicola* by a) logistic regression and b) Maxent.

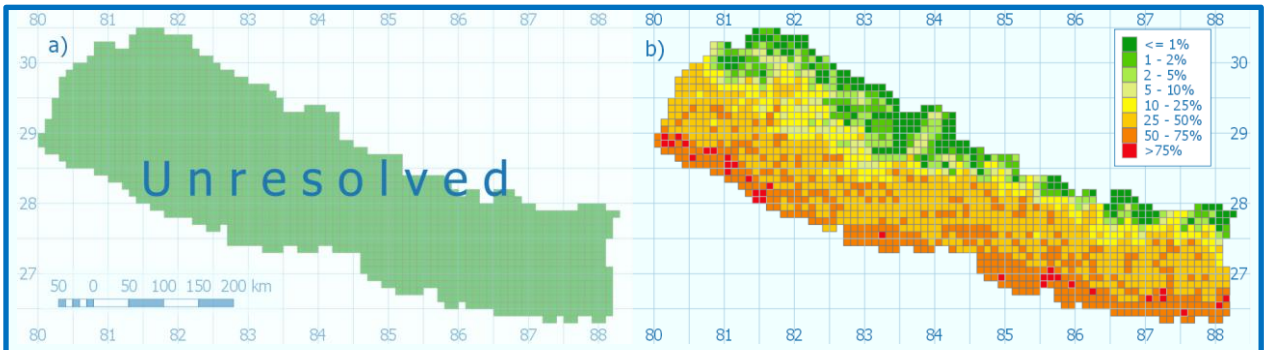


Figure 4.51 Predicted distributions in Nepal for *Mus terricolor* by a) logistic regression and b) Maxent.

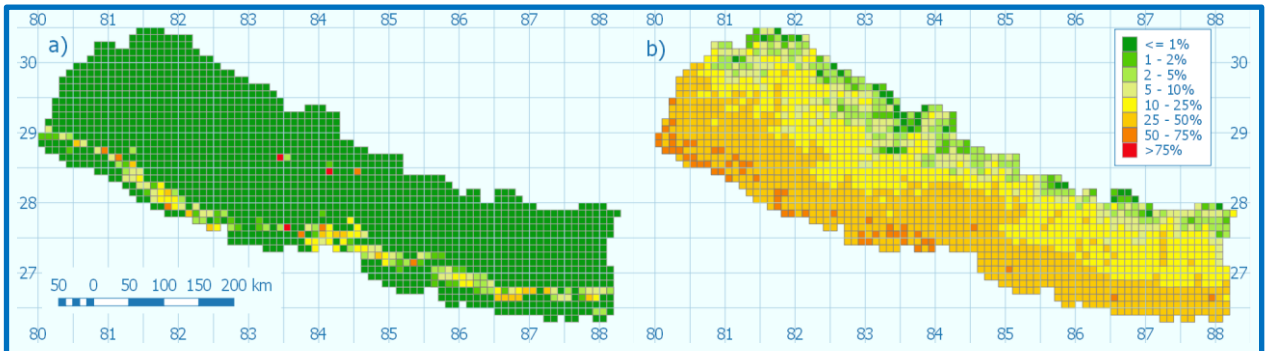


Figure 4.52 Predicted distributions in Nepal for *Nesokia indica* by a) logistic regression and b) Maxent.

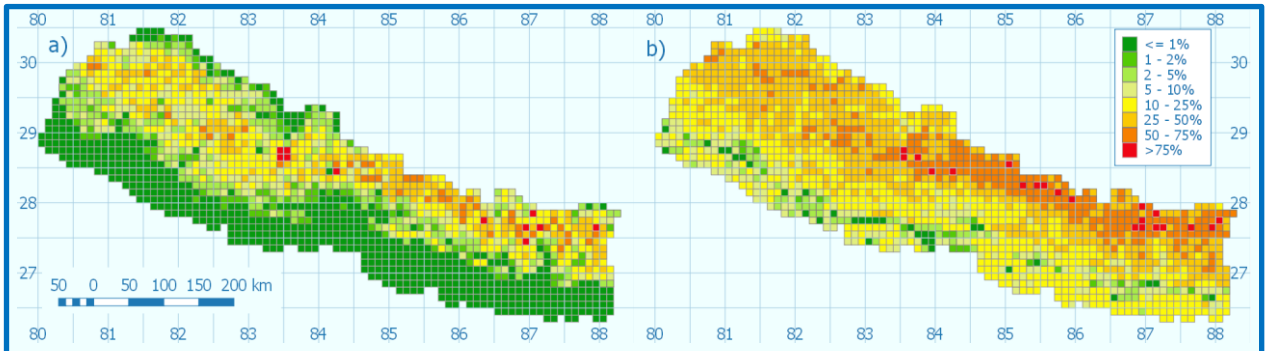


Figure 4.53 Predicted distributions in Nepal for *Niviventer eha* by a) logistic regression and b) Maxent.

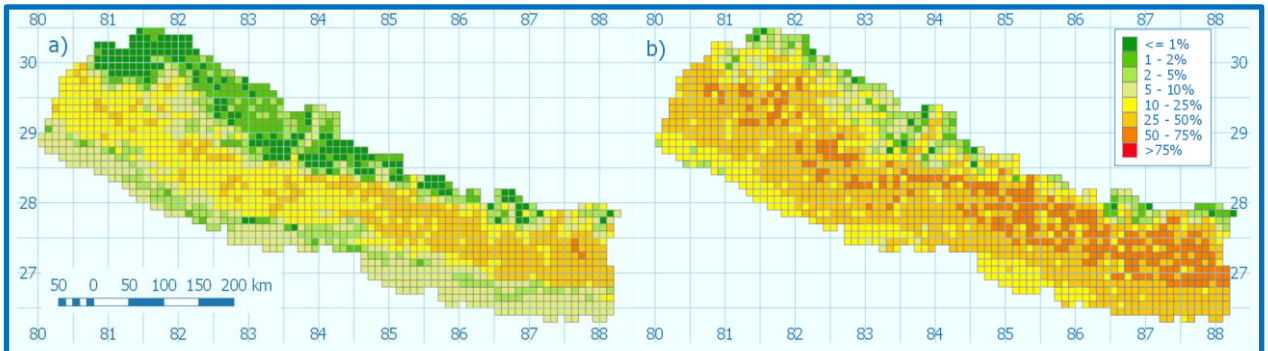


Figure 4.54 Predicted distributions in Nepal for *Niviventer fulvescens* by a) logistic regression and b) Maxent.

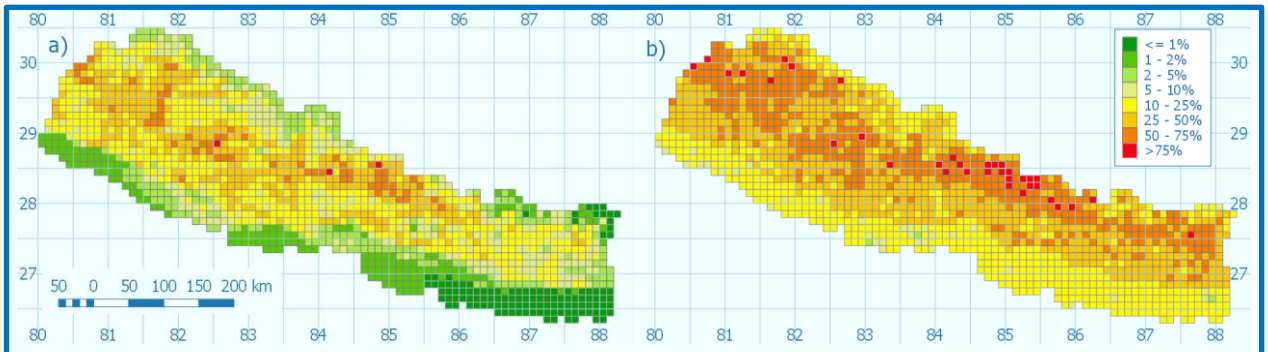


Figure 4.55 Predicted distributions in Nepal for *Nivivener niviventer* by a) logistic regression and b) Maxent.

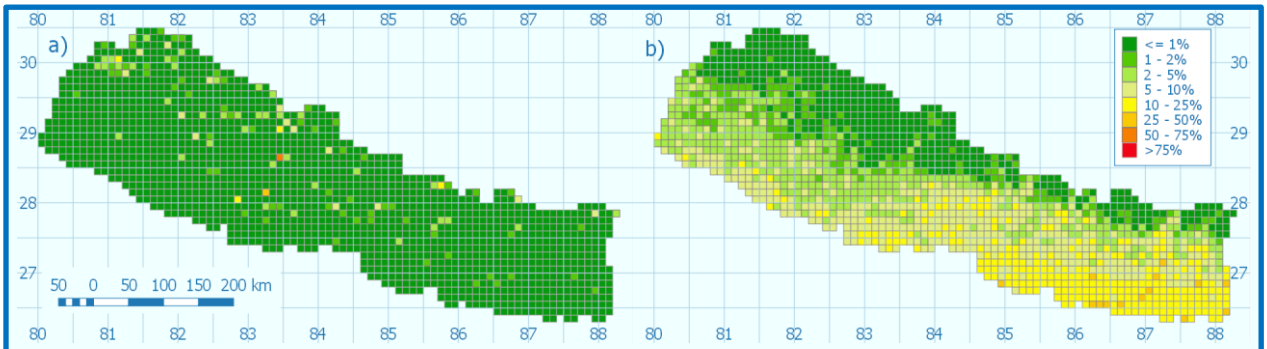


Figure 4.56 Predicted distributions in Nepal for *Rattus exulans* by a) logistic regression and b) Maxent.

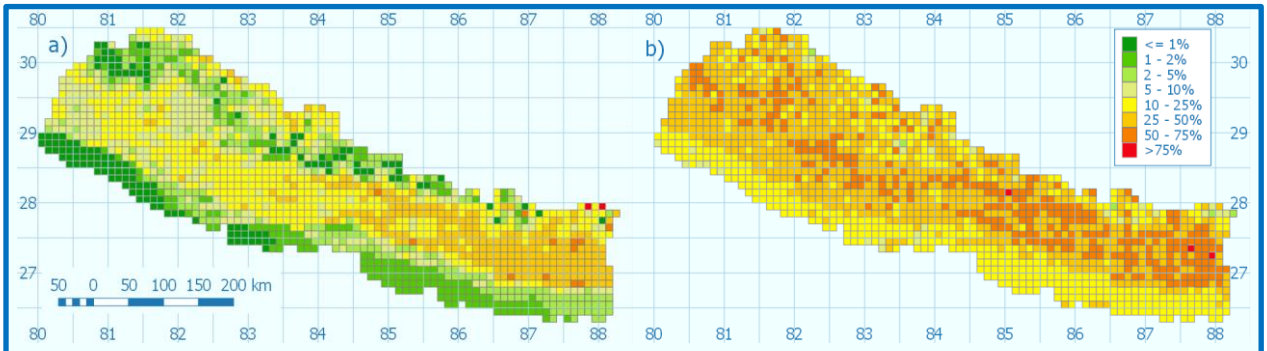


Figure 4.57 Predicted distributions in Nepal for *Rattus nitidus* by a) logistic regression and b) Maxent.

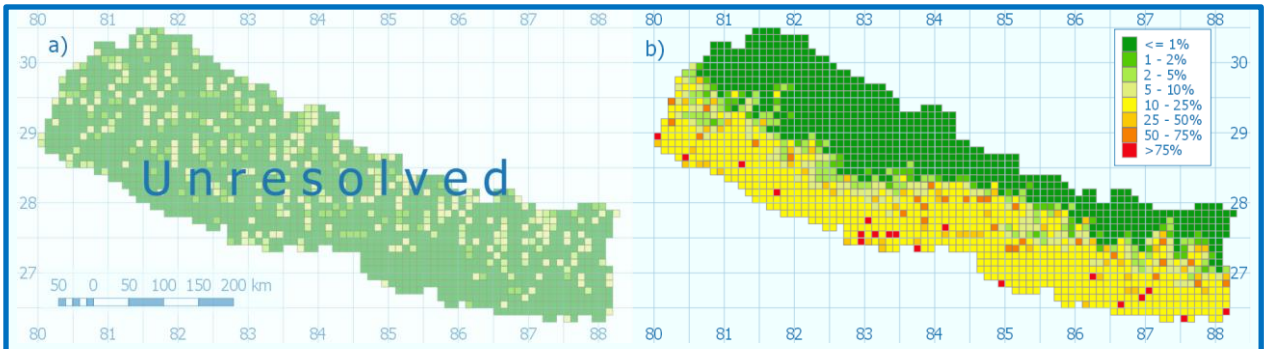


Figure 4.58 Predicted distributions in Nepal for *Rattus norvegicus* by a) logistic regression and b) Maxent.

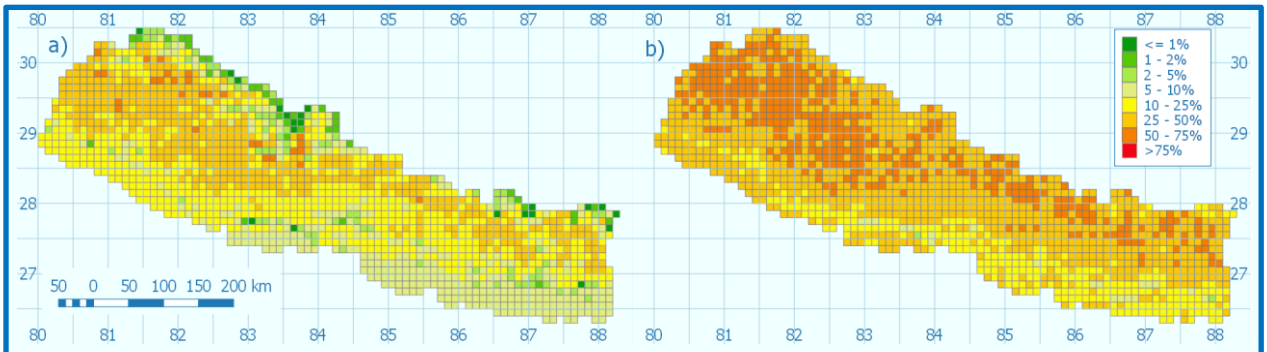


Figure 4.59 Predicted distributions in Nepal for *Rattus pyctoris* by a) logistic regression and b) Maxent.

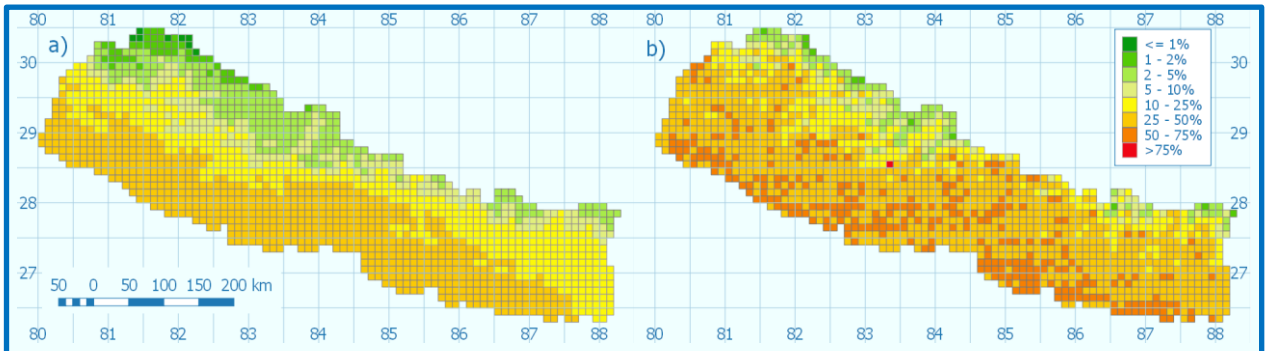


Figure 4.60 Predicted distributions in Nepal for *Rattus rattus* by a) logistic regression and b) Maxent.

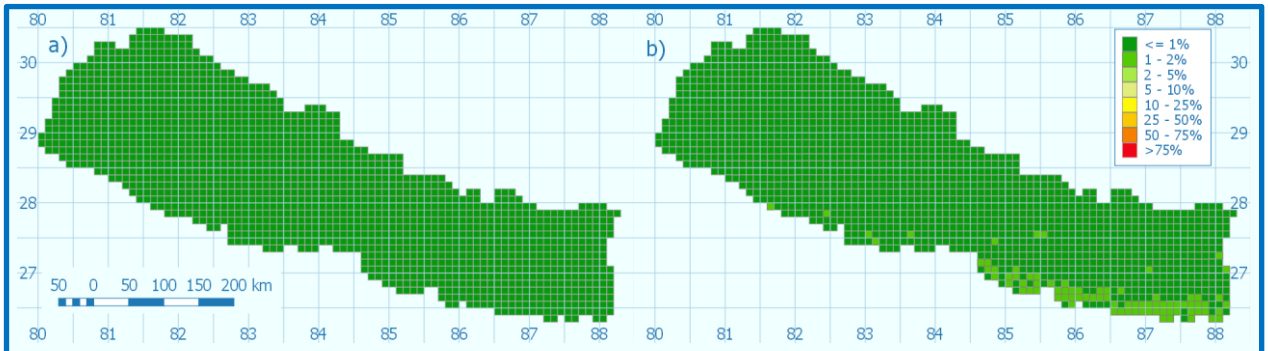


Figure 4.61 Predicted distributions in Nepal for *Rattus satarae* by a) logistic regression and b) Maxent.

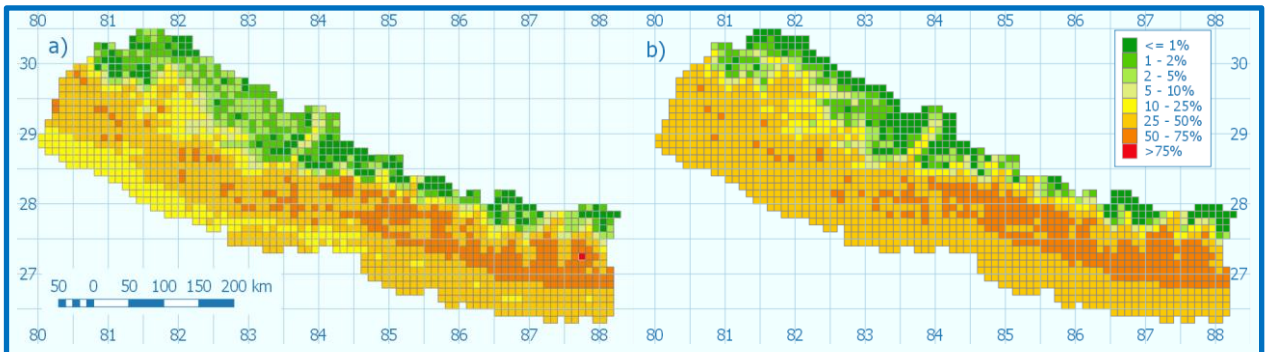


Figure 4.62 Predicted distributions in Nepal for *Rattus tanezumi* by a) logistic regression and b) Maxent.

4.C.c Soricidae

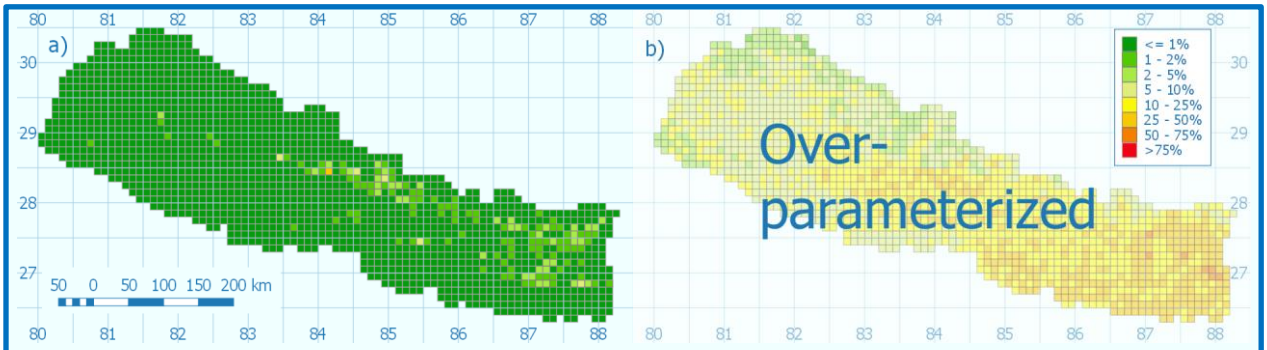


Figure 4.63 Predicted distributions in Nepal for *Anousorex squamipes* by a) logistic regression and b) Maxent.

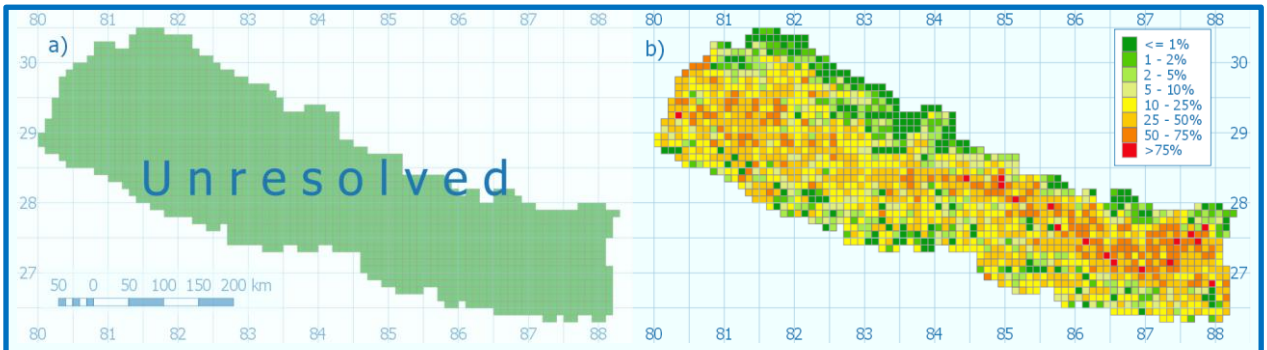


Figure 4.64 Predicted distributions in Nepal for *Chamarrogale himalayica* by a) logistic regression and b) Maxent.

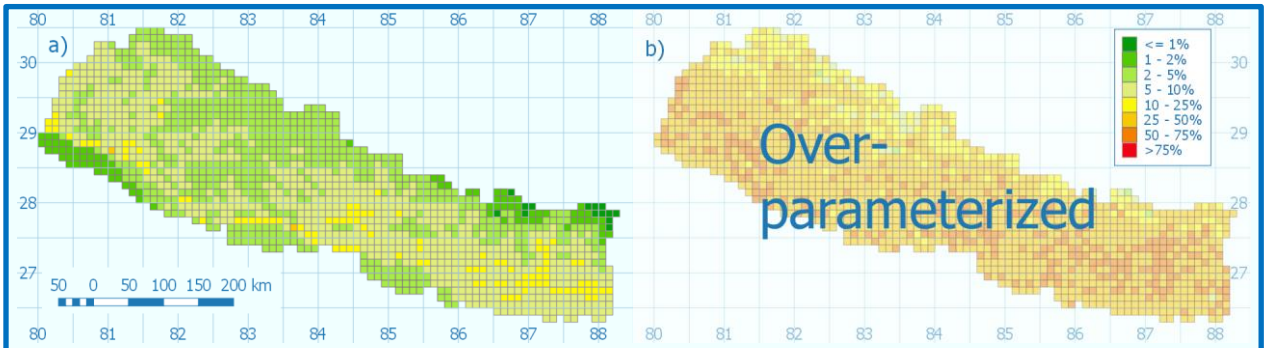


Figure 4.65 Predicted distributions in Nepal for *Crocidura attenuata* by a) logistic regression and b) Maxent.

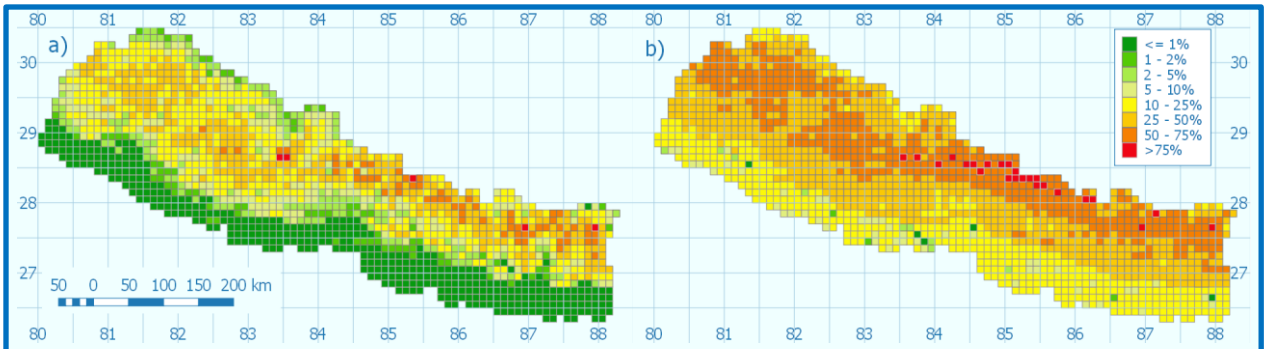


Figure 4.66 Predicted distributions in Nepal for *Episoriculus caudatus* by a) logistic regression and b) Maxent.

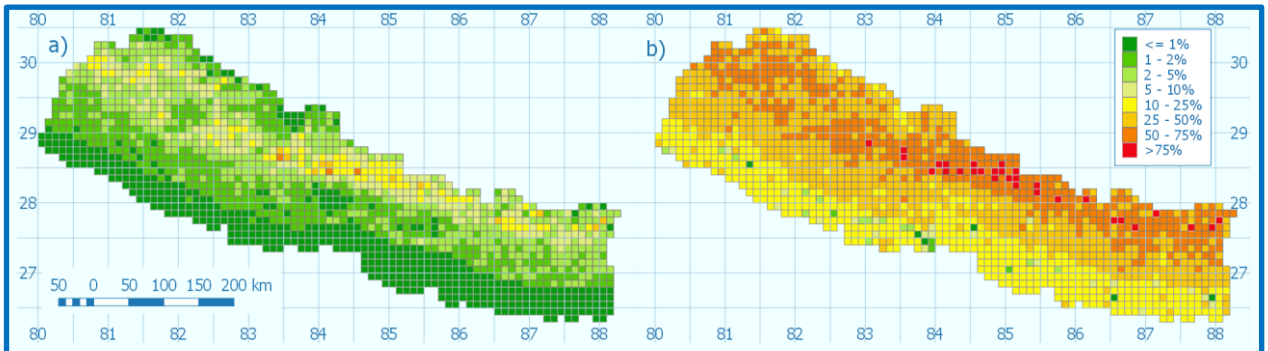


Figure 4.67 Predicted distributions in Nepal for *Episoriculus leucops* by a) logistic regression and b) Maxent.

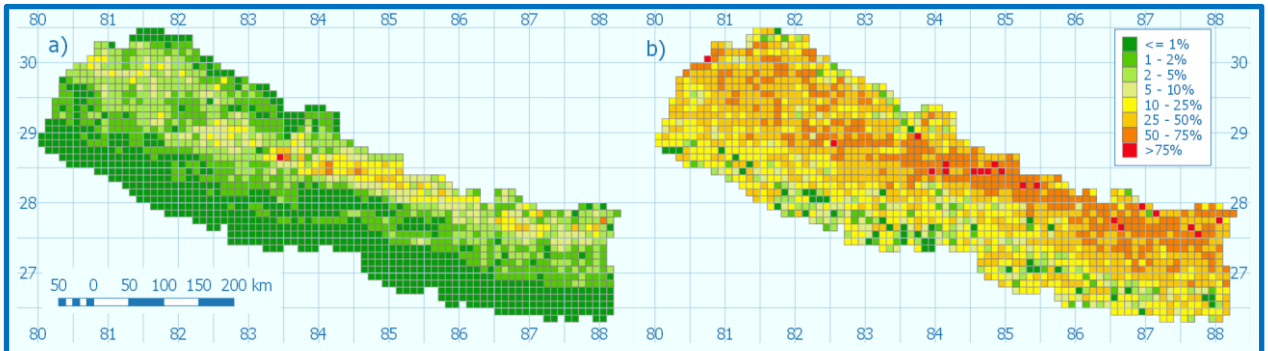


Figure 4.68 Predicted distributions in Nepal for *Episoriculus macrurus* by a) logistic regression and b) Maxent.

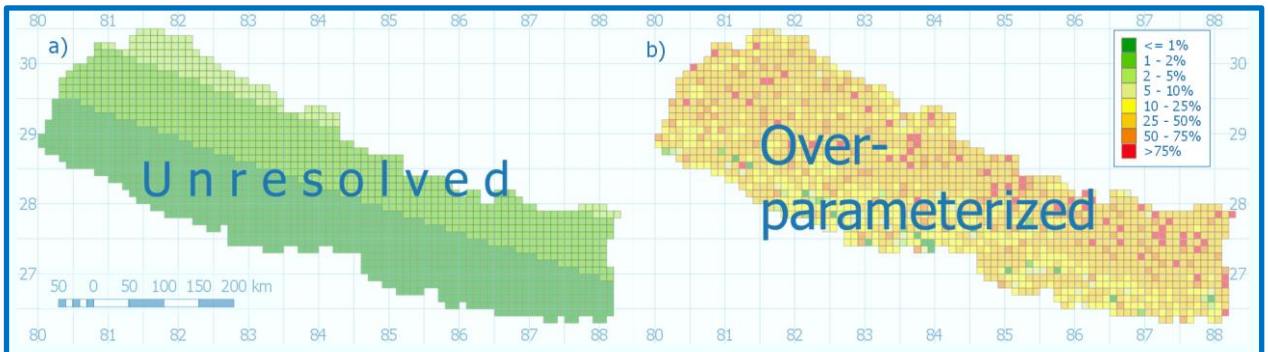


Figure 4.69 Predicted distributions in Nepal for *Nectogale elegans* by a) logistic regression and b) Maxent.

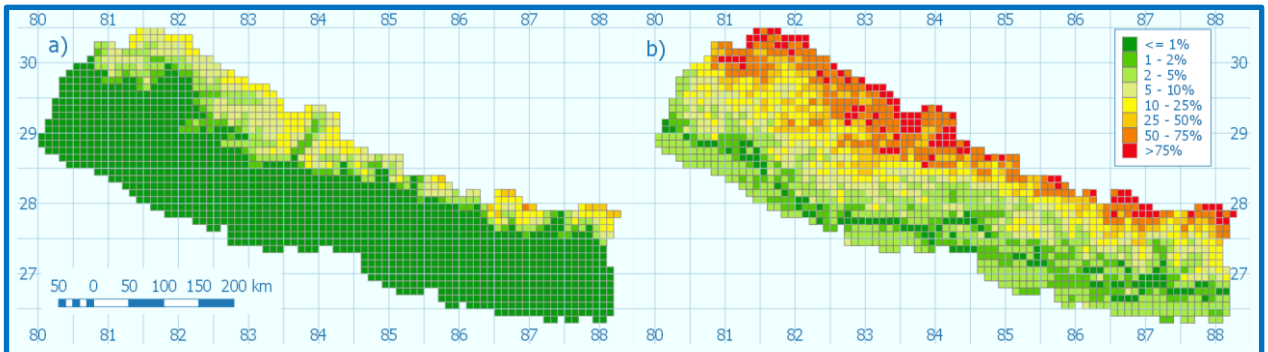


Figure 4.70 Predicted distributions in Nepal for *Sorex araneus* by a) logistic regression and b) Maxent.

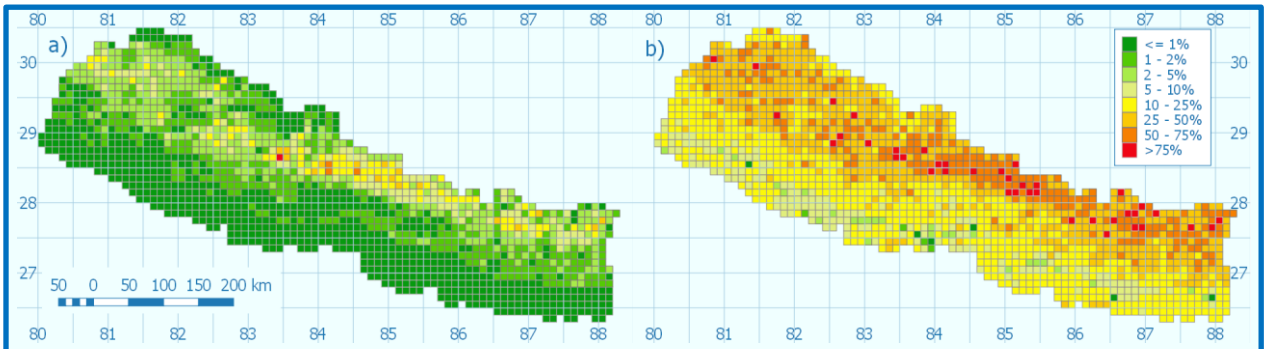


Figure 4.71 Predicted distributions in Nepal for *Sorex bedfordiae* by a) logistic regression and b) Maxent.

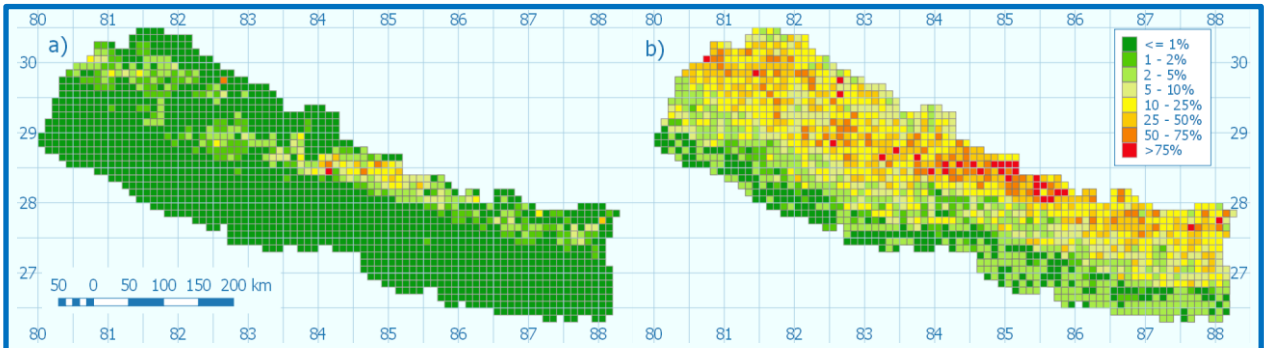


Figure 4.72 Predicted distributions in Nepal for *Sorex cylindricauda* by a) logistic regression and b) Maxent.

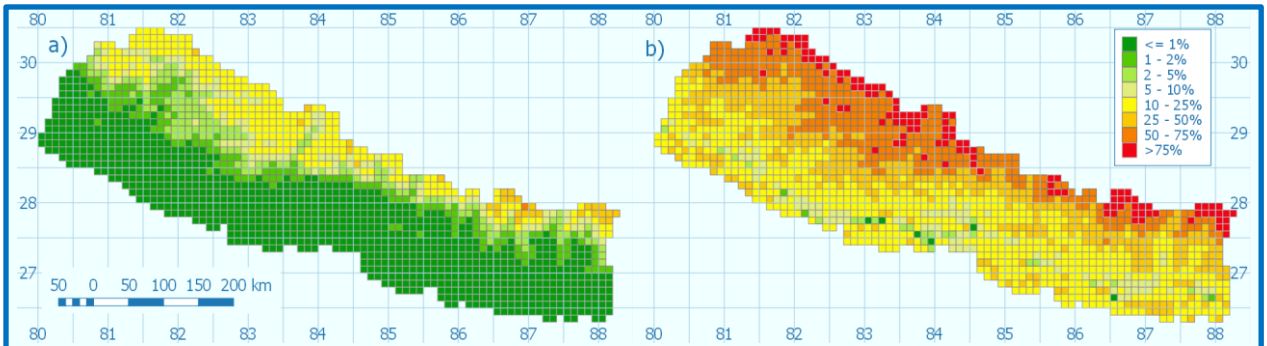


Figure 4.73 Predicted distributions in Nepal for *Sorex minutus* by a) logistic regression and b) Maxent.

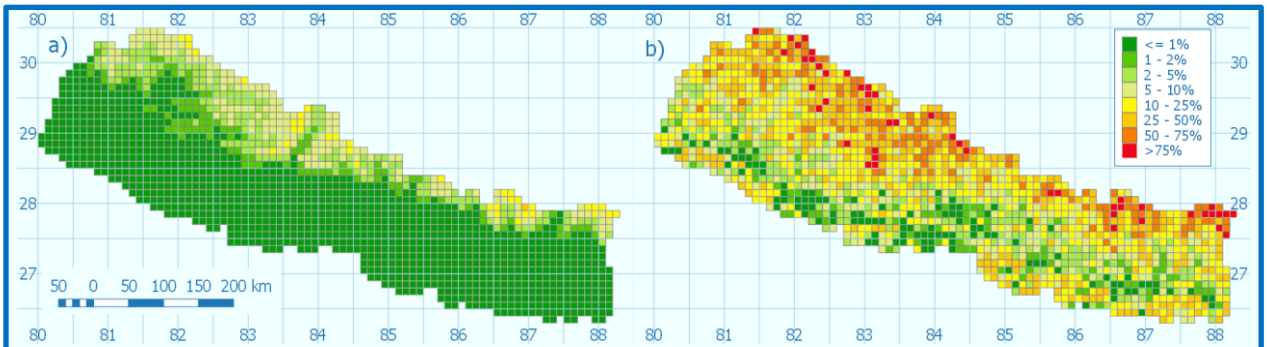


Figure 4.74 Predicted distributions in Nepal for *Sorex thibetanus* by a) logistic regression and b) Maxent.

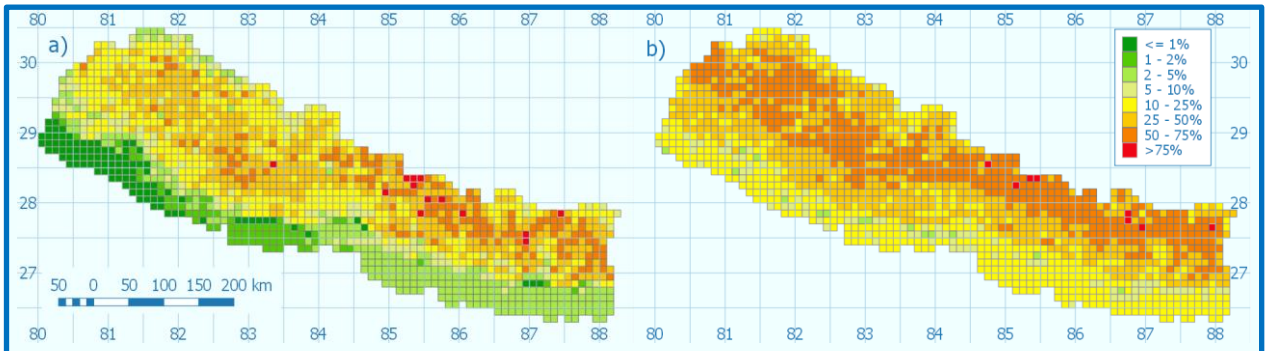


Figure 4.75 Predicted distributions in Nepal for *Soriculus nigrescens* by a) logistic regression and b) Maxent.

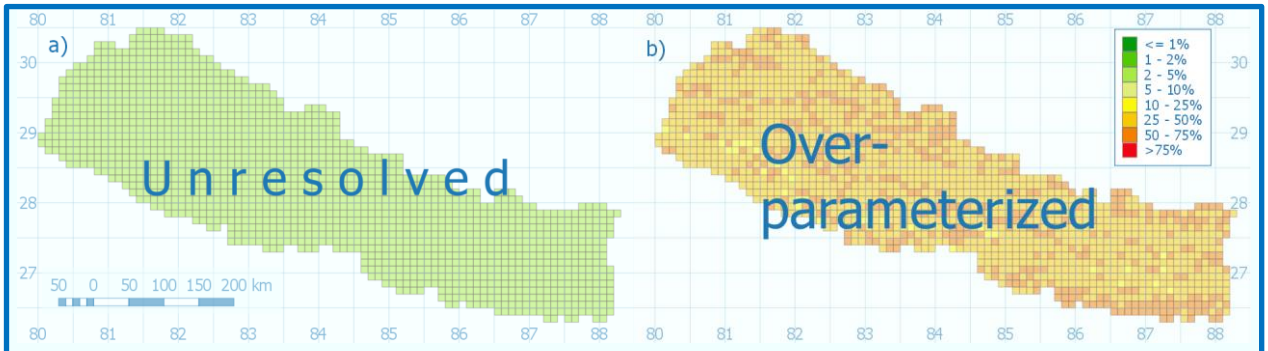


Figure 4.76 Predicted distributions in Nepal for *Suncus etruscus* by a) logistic regression and b) Maxent.

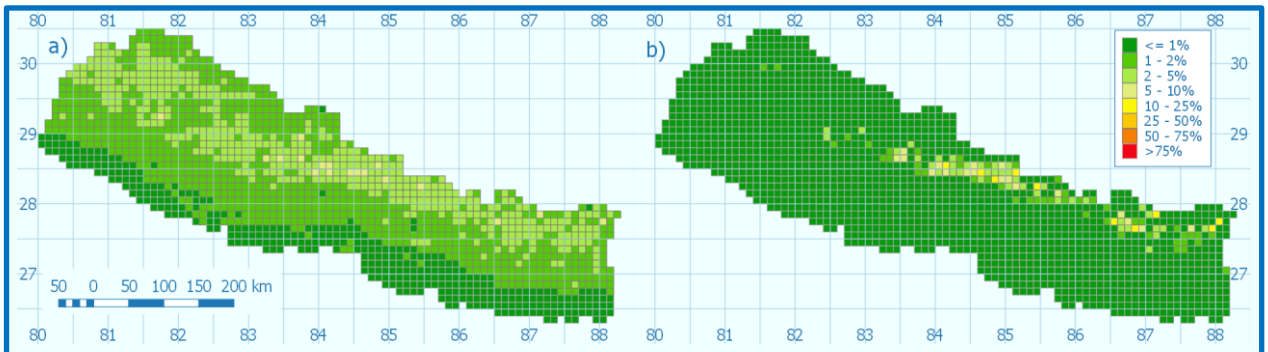


Figure 4.77 Predicted distributions in Nepal for *Suncus montanus* by a) logistic regression and b) Maxent.

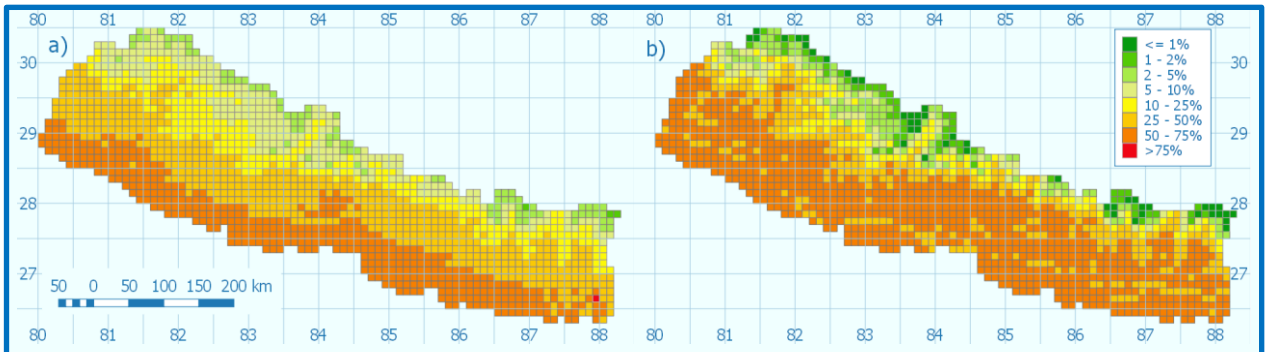


Figure 4.78 Predicted distributions in Nepal for *Suncus murinus* by a) logistic regression and b) Maxent.

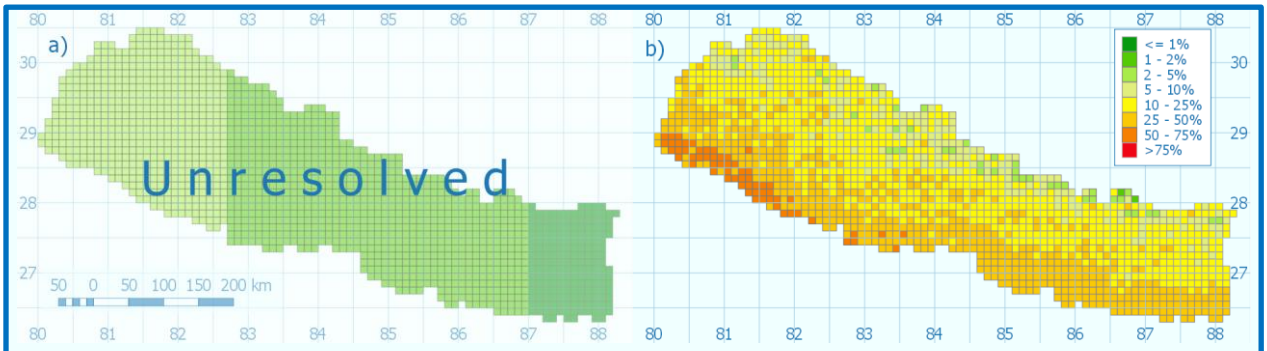


Figure 4.79 Predicted distributions in Nepal for *Suncus stoliczkanus* by a) logistic regression and b) Maxent.

Chapter 5 The Use of Genetic Methods to Aid Identification of the Small Mammals Collected from the Current Study

Abstract

In this chapter I describe the genetic analysis of DNA sequences obtained from the 720 tissue samples that I collected in the field. DNA was extracted from 501 samples using Qiagen DNeasy kits in the CMDN laboratory in Kathmandu. PCR of the cytochrome-b gene from approx. 200 extractions yielded 115 successful products which were sequenced by Source Bioscience to give 94 good quality sample sequences belonging to 16 putative species. In parallel, I analysed 200 sequences from GenBank belonging to 32 species, from which 45 reference sequences were selected. This enabled the quantification of conspecific and congeneric thresholds of p-distances (approx. 0.08 and 0.15 respectively). I created three separate phylogenies, one for each of the families, Cricetidae, Muridae and Soricidae, combining the 94 sample and 45 reference sequences. From these it was possible to confirm the identity of nine species, reassign one species and identify 11 unknown specimens. Furthermore, this analysis indicated the possibility of cryptic species of voles and revealed that three of my good species had no close alignments with sequences on GenBank. The problems of using GenBank uncritically are discussed and indicated a number of apparent misidentifications and a potential cryptic species of shrew. This is the largest genetic sampling survey of small mammals carried out in Nepal. The analyses presented here utilized only 13% of the available tissue samples and highlights the considerable potential for further questions to be addressed through genetic analyses of this unique resource.

Contents

Tables

Figures

5.1 Introduction

5.1.1 Background

Very few studies have explored the conservation genetics of small mammals in Nepal. Those that have, tended to focus on single species or sub-species identification (Adhikari *et al.*, 2018a; Adhikari *et al.*, 2018b). There have been some recent studies on karyotypes of three shrew species in Nepal (Motokawa *et al.*, 2008), and there is ongoing research into the only endemic small mammal, *Apodemus gurkha* (Karmacharya, pers. comm.). Not surprisingly, given the limited facilities for DNA analysis in Nepal, attention has focussed on charismatic species such as tigers (*Panthera tigris*) (Thapa *et al.*, 2018), snow leopards (*Panthera uncia*) (Karmacharya *et al.*, 2011) and Himalayan wolf (*Canis lupus*) (Werhahn *et al.*, 2017). Genetic studies of small mammals are more prevalent in neighbouring countries such as *Niviventer* in China (Chen *et al.*, 2011), voles (*Alexandromys*) in central Asia (Lissovsky *et al.*, 2018) and *Mus* in Myanmar (Shimada *et al.*, 2010). Researchers from China and Japan have explored the genetics of many species of rodent and shrew in Sichuan, Xinjiang and Tibet, bordering Nepal (Ge *et al.*, 2017; Liu *et al.*, 2018; Liu *et al.*, 2012; Suzuki *et al.*, 2003; Zhang *et al.*, 2016).

5.1.2 Aims and Objectives of the DNA Analysis

In this chapter I analyse the tissue samples collected from animals captured during the current study. The overall aim of the DNA Analysis was to utilise species barcoding techniques to confirm the field identifications. The specific objectives were:

- To collect tissue samples from as many of the animals caught as possible.
- To apply current state-of-the-art laboratory techniques to obtain good quality DNA sequences.
- To utilise modern bioinformatics in conjunction with NCBI databases to obtain correct identifications of the animals captured.

5.2 Methods

Permission to collect tissue samples and undertake their analysis was granted by the Department of National Parks and Wildlife Conservation (DNPWC) through their MOU with the Centre for Molecular Dynamics Nepal (CMDN) (see Appendix I. Letters of Permission to Undertake Tissue Sample Collection and Laboratory Analysis).

The majority of the DNA-based work was carried out by CMDN in Kathmandu. The main collaborator for this project was Dibesh Karmacharya (DK) and three members of staff were directly involved in laboratory procedures; Ajay Sharma (AjS), Adarsh Serchan (AdS) and Hemanta Choudhary (HC). Facilities were provided for me (SMCP) to carry out DNA extraction. The PCR products were sent to UEA for sequencing and a small number of additional PCRs were carried out by Michael Strinden (MS).

5.2.1 Tissue Sample Handling and Storage

All the DNA analyses described in this chapter were based on tissue samples or specimens collected in the field (see Chapter 2 for full details). Tissue samples were taken using a 1.5mm ear punch (Harvard Apparatus Ltd., UK) which was cleaned with 5% bleach and 75% ethanol between animals. Samples were preserved in 99% ethanol in 2ml Eppendorf tubes, sealed with Parafilm M. Each tube was labelled with a unique, five-element code:

YY_T_S_G_AA

where YY = two-digit year, T = transect, S = site, G = grid and AA = animal number.

On return to the laboratory in Kathmandu, all samples were inspected for adequate ethanol and label legibility. Before DNA extraction, I divided most tissue samples into two. This created two series of samples to increase security and allow for a second extraction on failure of the first. Sample codes were supplemented with _1 or _2 to represent the two series. A minority of samples (approx. 13%) were so small that they could not be divided. The two series were stored in separate locations in Kathmandu for security against flooding, earthquake, *etc*. Series 1 was stored in the offices of Himalayan Nature (see Chapter 2) and series 2 remained in CMDN. A preliminary batch of 37 samples was sub-divided for analysis before the main divisions were made (see 5.2.2 below). These were classified as Series 0 and the residuals were further sub-divided into the two series.

After sub-division, all tissue samples were rinsed in 99% ethanol to remove any DNA contamination from the field, and placed in new Eppendorf tubes with 1ml of fresh ethanol. All sample tubes were labelled with self-adhesive bar-codes based on the six-element code and stored in 100 compartment

tube-boxes, at ambient temperature, in their respective locations. Full details of the box number, row and column codes were stored in the MS Access database (see Chapter 2) to aid selection of samples for subsequent DNA extraction.

5.2.2 DNA Extraction and Storage

All DNA extractions were carried out on samples from series 0 and 2, at CMDN during three separate periods.

- An extraction was carried out by AjS in September 2014. This comprised 37 samples of series 0 from Transects 1 and 2 in 2013. It was used as a trial of PCR and sequencing techniques (see 5.2.3 & 5.2.4 below).
- Twenty-two batches of 15 samples each were extracted in February 2017. The first was carried out by AdS with SMCP, as a training exercise, and the remainder by SMCP alone.
- A further seven batches of 15 samples and one of 21 samples were extracted by SMCP in September 2017.

Tissue samples were selected to include a full range of putative species, beginning with the less abundant and least confidently identified animals. They were also selected sequentially to provide at least one sample from each trapping grid and year. DNA was extracted using proprietary kits (Qiagen DNeasy Blood and Tissue Kits). The standard protocol (Anon., 2006) was followed exactly, although I carried out a double final elution using half volumes, to ensure extraction of as much DNA as possible. One negative extraction control (two in the first batch) was included with each batch. All samples were double-labelled with the self-adhesive bar codes and the batch and tube number (in the form B/TT) written on the tube lid with indelible marker. Final DNA samples were stored in 2ml Eppendorf tubes at –20°C in the CMDN laboratory.

I tested all extractions after the first batch, including the negative extraction control (NEC), for DNA concentration on a Nanodrop machine (Series 2000). This was recalibrated before every batch and after the eighth sample in each batch. The first and last samples in each batch were tested three times to investigate consistency whilst all others were measured once. DNA concentration in ng/ml plus the A260:A280 and A260:A230 ratios were recorded for each sample.

5.2.3 PCR

PCR was undertaken at CMDN in four stages;

- In September 2014 AjS carried out a trial PCR of 37 samples for the full CO1 mitochondrial gene.
- In November 2014 AjS reran four of these samples for the cytochrome-b mitochondrial gene. He used primers that defined a region of the cytochrome-b gene up to 389 bp in length.
- In February 2017 AdS ran five batches and HC ran one batch for the full length (1190 bp) cytochrome-b gene.
- From July to September 2017 HC ran six batches using primers designed specifically for rodent taxonomy as part of the PREDICT research programme (www.ecohealthalliance.org/program/predict), which isolated a 457 bp section of the cytochrome-b gene. At the outset of this period HC also ran four optimisation routines to improve the results obtained in the February 2017 runs.

The primers used in Stages 1 – 3 were provided by CMDN, supplied from Macrogen (South Korea). Following the protocol of Dubey *et al.* (2007), the Stage 3 primers were;

L14724 (5' – CGAAGCTTGATATAAAAACCATCGTTG – 3') and

H15915 (5' - AACTGCAGTCATCTCCGGTTACAAGAC – 3')

as the forward primer and reverse primers respectively for the full mammalian cytochrome-b gene.

Before the fourth stage in autumn 2017, I carried out an analysis of suitable primers. In particular, I compared the ability of a short-length cytochrome-b sequence derived from the PREDICT primers with the full gene sequence to determine species identity. This is described fully in Supplementary Information 5.A: The comparison of short and full-length primers. The result of this analysis indicated that short-length cytochrome-b sequences were equally powerful at distinguishing species identities, so I proceeded to develop a set of primers based on the PREDICT design. In summary, I used Primer–Blast to test the efficacy of the PREDICT primers on 37 sequences of Soricidae and 225 sequences of Muridae on GenBank. I also ran this procedure on two other primers; a generic mammalian full-length cytochrome-b pair as defined by Naidu *et al.* (2012) and a pair for sex determination in shrews (Matubara *et al.*, 2001). This exercise indicated that both primers for the cytochrome-b gene needed certain sites to be converted to degenerate codes to accommodate the full range of species (Table 5.1).

Table 5.1 The primers used in Stage 4 of the PCRs run at CMDN. Degenerate loci are marked in red.

Source	Primer name	Sequence (5′ – 3′)
PREDICT Cyt-B short-length	CytB_F	GAGGMCAAATATCATTCTGAGG
	CytB_R	TAGGGCVAGGACTCCTCCTAGT
Generic Cyt-B for mammals	MTCB-F	CCHCCATAAATAGGNGAAGG
	MTCB-R	WAGAAYWTCAGCTTTGGG
Sex determination for shrews	F1	CATGGTGTGGGCTCGCAATC
	R1	CTGCCTGTAGTCTCTGTGCC

All PCR reactions used 2 or 3µl of DNA template and 0.5 or 0.75 µl of primers. The total volume of 25µl was made up with Qiagen mastermix (MM-206145), Qiagen Q-solution and water. Negative extraction controls from all relevant extractions were included and every PCR included one negative template control (NTC) and at least one positive template control (PTC) using a known concentration of generic mouse DNA.

Following PCR, all samples were checked by CMDN staff for products using agarose gel electrophoresis with 1.5% ethidium bromide. Each reaction was scored 0 for failed, 1 for faint band and 2 for strong band. The gel plates were photographed, and resulting images stored in the MS Access database (see 5.2.7 below), and the plates then destroyed. All PCR products were stored at –20°C in the CMDN laboratory.

A final stage of PCR was undertaken at UEA by MS on PCR products from CMDN. There were 71 failed reactions from PCR runs 3 to 8 that used the full-length cytochrome-b primers. These were re-run using PREDICT primers (Table 5.1) for the partial cytochrome-b gene (457 bp). The reactions were tested using Qubit fluorometer (ThermoFisher Scientific).

5.2.4 Sequencing

The four products obtained in November 2014 were cleaned and sequenced by Macrogen (Seoul, South Korea), soon after PCR. Sequence strings were returned, but no quality data or spectrograms were provided. Cleaning and sequencing of all other PCR products was carried out by Source BioScience (Nottingham, UK) in June 2018. All PCR products were sequenced using the relevant forward primers only.

5.2.5 Bioinformatics

Sequences were returned as individual .ab1 files and were inspected using Chromas (Anon., 2018). Automatic trimming options were set to trim from each end of the sequence until the average "phred" score across the next ten bases exceeded 20. In some cases additional manual trimming was carried out when sequences clearly included poorly defined structures. Manual editing of N-bases

was undertaken where the call was clear and a very small number of deletions and insertions were made. In some cases, the sequence traces had extremely poor structure, so they were classified as "Poor Sequences" and not used in the scoring and matching procedures described below.

Individual sequences were output at .fastq format files, which included a quality score string as well as the sequence string itself. These were both imported into the MS Access database (see 5.2.7 and 5.2.6 below), allowing the sequence length, the numbers of each nucleotide, the number of N-bases and the mean Phred score for the whole sequence to be calculated.

BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to identify matches to the NCBI GenBank database, using the following parameters;

- The BlastN suite. was used to find nucleotide sequences in the "nr/nt" collection.
- Individual sequences were pasted manually into the sequence textbox.
- Organism filters: Soricidae (taxid:9376
  OR Muridae (taxid:10066)
  OR Cricetidae (taxid:337677).
- Max target sequences: 50.
- Expected threshold: 20.
- Word Size: 11.
- Match/Mismatch scores: 1, –2.
- Gap Costs Existence: 2 Extension 1.

Details of the top match for each sequence was stored in the MS Access database. Details included the Accession number, length of sequence, species or sub-species, country and locality where available. Match statistics, including the score, length of matched sequence, number of identities and gaps and the e-value were also recorded. Optionally, up to three more matches were recorded. These were subjectively chosen when the lower matches were different, but plausible, species and when the GenBank "pop-set" was different.

All alignments were carried out in Mega 7 (Kumar *et al.*, 2016) using the MUSCLE algorithm (Edgar, 2004) with default penalty settings, but using the UPGMA clustering method. Comparisons were made using alternative settings in MUSCLE, and with the Clustal W algorithm (Ge *et al.*, 2017). This showed that altering the gap-opening extension penalties generally cause the number and size of gaps to be increased. As many of the sequences had poor distal portions, the alignments were

trimmed, reversed and realigned twice to remove anomalous alignments. This also improved the patterns of gap insertions, which were often inexplicable when only one or two sequences required insertions in the others. All sequences were checked visually and, if needed, further manual editing was undertaken.

All phylogenies were constructed using Mega 7. I use the term "phylogeny" to refer to gene genealogy trees, *sensu* Kroken and Taylor (2001), based on genetic distances between samples. I do not infer any evolutionary relationships, but am simply using the display of "phylogenetic" trees to illustrate molecular species delimitation, *sensu* Zhang *et al.* (2016) and Liu *et al.* (2012). In particular, I have used this term in the captions of all figures generated by the Mega 7 software. Since sequence-pair lengths varied considerably, pair-wise exclusion of gaps and missing data was used. For the same reason, evolutionary distances were calculated using p-distances between nucleotides (Nei and Kumar, 2000). The primary purpose of constructing the phylogenies was to infer species identities for the unknown sequences by clustering with known sequences, based on their similarities. Consequently, the Neighbour-joining method (Saitou and Nei, 1987) was used for all phylogenies (Suzuki *et al.*, 2003). N-J trees included all substitutions and uniform rates, and used 1000 bootstrap trees to test for phylogeny consistency Felsenstein (1985).

Matrices of p-distances were calculated in a similar way using distances between nucleotides with pair-wise deletion of gaps and missing data, including all substitutions and uniform rates. Bootstrap estimations of p-distance variation were not calculated.

5.2.6 Procedures for Phylogenetic Analysis

The sequences obtained from this study, hereafter known as "sample sequences", were separated into the three families as identification at this level was absolutely certain. For each family I identified reference species based on the putative field identifications and published sources (Corbet and Hill, 1992; Molur *et al.*, 2005; Pearch, 2011). I then subjectively selected cytochrome-b sequences from GenBank, hereafter known as the "reference sequences". Where possible full-length sequences were selected, but for some species with few specimens, I accepted shorter sequences when available. Generally, all available sequences from each species were selected, unless they were very common, in which case I selected a variable number reflecting their provenance and the number of pop-sets (as indicated by the first two letters of their accession number), up to a maximum of ten. These sequences were labelled with the abbreviated taxon, the GenBank accession number and the location from where the specimen was collected, if known.

The sequences were then aligned in Mega 7, separately for each family to give a reference phylogeny. These were annotated with colour-coded square symbols to indicate species and genus identities, plus branch lengths and bootstrap percentages for each node. They were assessed in four stages;

- Firstly, the identification of anomalies in the phylogenies, where members of one species clustered separately with other species or formed outliers.
- The investigation of the similarity of con-specific sequences, based on the p-distance matrices.
- The selection of suitable con-specific and con-generic p-distance thresholds, either for the family as a whole or for individual genera or species where necessary.
- Using these thresholds, one sequence was selected subjectively to represent phylogenetic clusters in subsequent analyses. Occasionally more than one sequence was selected to represent different pop-sets or to include sequences of particular relevance from Nepal.

Full details of the methods employed in analysing the reference phylogenies is given in Supplementary Information 5.C: Analysis of Cytochrome-b Sequences Obtained From GenBank.

The selected reference sequences were then combined with the sample sequences for each family and p-distance matrices and phylogenies reconstructed. Two distance matrices were extracted;

- A half-square matrix which displayed the p-distances between all combinations of sample sequences.
- A rectangular matrix which displayed the p-distances between every sample sequence (rows) and every reference sequence (columns).

The p-distance values in both matrices were colour-coded using the relevant threshold values obtained from the analysis of reference sequences. The combined phylogenies were annotated as described above for the reference phylogenies, with the addition of colour-coded diamond symbols to indicate the field identifications for sample sequences. Using both these outputs field identifications were either confirmed or reassigned.

5.2.7 Data Handling and Analysis Methods

The main data storage medium for all DNA, PCR and sequence data was the MS Access database described in Chapter 2. It was expanded to include tables, queries and forms linked to those holding the field data. It also included tables and queries to hold and manage sequences down-loaded from GenBank.

All statistical analysis was carried out in Statistica v13 (Anon., 2017). Where possible, parametric analyses were undertaken by building the most complete models possible given the unbalanced nature of the datasets. Generally, the GLM module was used with raw response variables and categorical and/or continuous predictor variables building a variety of ANOVA, ANCOVA or regression models. When raw response variables were severely non-normal, transformations were applied or non-linear analyses were carried out using the GLZ module. These were used for binary response variables (using a logit link-function) or for count data (using a log link-function). On occasion, when the number of predictor variables in the models was large with a high degree of collinearity, forward step-wise model-building was utilised. In a few cases it was necessary to resort to non-parametric tests such as contingency χ^2 and Kruskal-Wallis one-way ANOVA by ranks. Finally, due to the large number of analyses carried out, to reduce the possibility of Type I errors, an *a priori* α level of 0.01 was used for all statistical testing.

5.3 Results

5.3.1 Summary of Laboratory Analysis of Tissue Samples from the Current Study

A full statistical analysis of the procedures and results of the tissue sample collection and laboratory analysis is given in Supplementary Information 5.B. In summary, tissue samples were collected from 720 of the 792 animals caught during the study. From these, DNA was extracted from 501 samples. A total of 250 PCRs was carried out, of which 126 were successful. These yielded 115 good quality sequences. However, initial inspection through BLAST revealed that 16 sequences appeared to be contaminated by human DNA and one sequence appeared to have an error in its coding. Furthermore, four sequences from PCR run 2 could not be used because they were short sections not related to the PREDICT sequences. This resulted in 94 sequences that could be used in taxonomic analysis to confirm species identity. These were known as the sample sequences and were labelled with the abbreviated taxon (the first four letters of genus and species), their five-part animal code (see 5.2.1) and their three-part PCR code (see 5.2.3). For ease of interpretation of the phylogenies, recall that the first part of the animal codes represented year, and the second and third parts represented the transect and site numbers.

5.3.2 Summary of Analysis of Cytochrome-b Sequences obtained from GenBank

Full details of the analysis of the reference sequences are given in Supplementary Information 5.C. The outcome of this analysis was the distillation of the 200 sequences originally selected into 45 distinctive reference sequences (Table 5.2). Of the 32 species, 20 were represented by a single sequence and 11 species had two sequences each, representing different pop-sets, geographic locations and/or distinct phylogenetic relationships. One species (*N. fulvescens*) required three sequences to fully represent its phylogenetic diversity.

This analysis also generated three thresholds of p-distances; conspecific, congeneric and confamilial (equivalent to cross-generic). For example, of the 112 sequences from the family Muridae, 33 were from the genus *Mus*. Two of these species (*M. booduga* and *M. musculus*) had six and eight sequences respectively. This meant that they had 30 and 56 pair-wise p-distances to sequences of the same species (conspecific p-distances). Similarly, they had 162 and 200 congeneric p-distances, and 480 and 640 confamilial p-distances respectively.

Table 5.2 Summary of reference sequences obtained from GenBank, selected for subsequent analysis with the sample sequences. The calculation of the con-specific and con-generic thresholds is explained in the text.

Family	Genus	Species	Acc. No.	Location (where known)	Thresholds	
					Conspecific	Congeneric
Cricetidae	*Alticola*	*A. stoliczkanus*	MF040961	China	0.090	0.124
	Neodon	*N. irene*	KU214684	China		
			GU908336	China		
		N. leucurus	KP190222	China		
		N. sikimensis	JF906124	China		
			HQ123606	China		
Muridae	*Apodemus*	*A. gurkha*	AB032852	Nepal	0.070	0.160
		A. pallipes	MG748222	China		
		A. sylvaticus	AB033695	Japan		
			AF159395			
	Mus	*M. booduga*	KY587424	Nepal		
		M. cervicolor	AB125766	Cambodia		
		M. cookii	KX790791	Pakistan		
		M. musculus	AB649506	Nepal, Tukuche		
		M. pahari	AY057814			
			EU349767			
		M. platythrix	AB125782	India		
		M. terricolor	AB125777	Nepal		
	Niviventer	*N. andersoni*	KY068735	China	0.095	0.150
		N. brahma	KY068773	China		
		N. eha	KY068779	China		
		N. fulvescens	GU457003	Vietnam	0.090	
			KY068838	China		
			MG748246	China		
		N. niviventer	AB973113	India	0.095	
			GU456998	Vietnam		
	Rattus	*R. andamanensis*	JQ814208	China	0.040	0.130
		R. nitidus	AB973108	India		
			KU214575	Nepal		
		R. pyctoris	KY587429	Nepal		
		R. rattus	HM217741	India	0.060	
			KY002805	Nepal		
		R. tanezumi	AB096841	China		
			KP343797	Afghanistan		
Soricidae	*Episoriculus*	*E. caudatus*	AB175112	Nepal, Gosain Kund	0.080	0.120
			AB175115	Nepal, Gosain Kund		
		E. leucops	AB175111	Nepal, Syng Gomba		
		E. macrurus	GU981290	China		
	Sorex	*S. araneus*	GQ374412		0.090	0.150
		S. bedfordiae	KC192655	China		
		S. cylindricauda	AB175121	Nepal, Kyanjin Gomba		
			KJ547369	China		
		S. minutus	DQ065610			
	Soriculus	*S. nigrescens*	AB175101	Nepal, Syng Gomba	0.050	(0.160)
			GU981297	China		

For both species, all the conspecific p-distances were much smaller than the congeneric p-distances (Figure 5.1). From this it is possible to identify a conspecific threshold of anywhere between 0.04 and 0.10. Consequently, a p-distance between a reference sequence and an unidentified sample sequence of, for example 0.03, would have a much high probability of belonging to that species, compared to another species within the genus. Conversely, a p-distance of 0.12, would indicate with

a high probability that the unidentified sequence did not belong to that species. These p-distance thresholds obtained from the reference sequence analysis have been used to provide a quantitative assessment of the identity of the sample sequences in the following sub-sections.

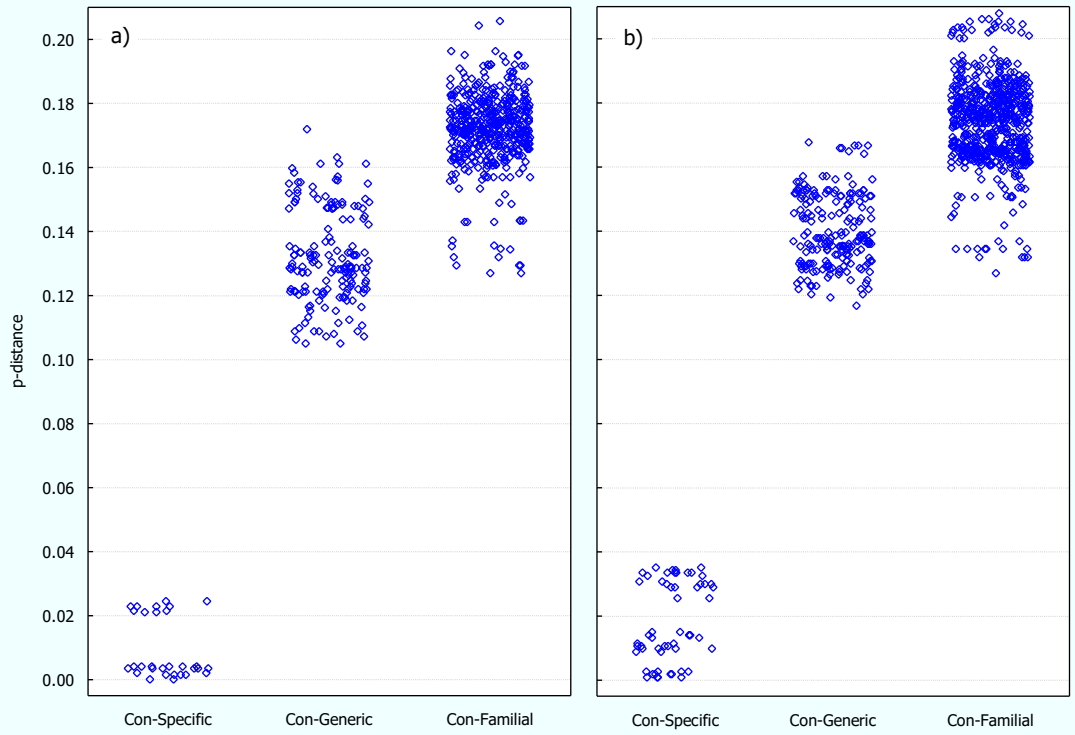


Figure 5.1 P-distances between sequences of a) *Mus booduga* and b) *M. musculus*. For each species, the p-distances have been categorised into three taxonomic levels; con-specific, con-generic and con-familial. Individual points have been randomly jittered on the x-axis for clarity, but not on the y-axis.

5.3.3 Analysis of Sample Sequences of Cricetidae with GenBank Reference Sequences

The six sample sequences in this family all had field identifications of *N sikimensis*. They included one pair of duplicate sequences from animal 14-1-6-5-05 with two different PCR runs (17 and 20). The square matrix of these sequences (Table 5.3a) showed that the duplicates were not identical and had a p-distance of 0.046. The con-specific threshold for this species was 0.09, which clearly discriminated two branches with a mean p-distance between them of 0.122 (Figure 5.2). Two of the animals in the upper branch came from the Pipar transect (13-1-4-5-02 & 14-1-6-5-05) and animal 14-3-1-6-02 came from Langtang, whilst the two animals in the other branch came from Sagarmatha. Two of the pair-wise distances between these two branches were 0.134, which was greater than the con-generic threshold of 0.124.

The rectangular matrix (Table 5.3b) showed that the mean p-distance of the single out-group reference sequence (*A. stoliczkanus*) from all sample sequences was 0.139, confirming their cross-generic status. However, it was difficult to assign the six sample sequences to any of the other references.

Two samples were just under the con-specific threshold with *N. leucurus*, but one of them was also below the threshold for *N. sikimensis*. In contrast, one sample was above the threshold with *N. irene* which put it beyond the con-generic threshold. This ambiguity is supported by the low bootstrap values for three of the higher level nodes.

Table 5.3 Pair-wise p-distance matrices for sequences for Cricetidae; a) The square matrix between six sample sequences and b) the rectangular matrix between them and the six reference sequences. Distances highlighted in red are less than the con-specific threshold (0.090) and those highlighted in blue are greater than the con-generic threshold (0.124). Sequences marked # are duplicate sequences from the same animal.

a)

	Neod siki 13-1-4-5-02 (01/16-15)	Neod siki 14-1-6-5-05 (07/14-17) #	Neod siki 14-1-6-5-05 (07/14-20) #	Neod siki 14-3-1-6-02 (07/03-16)	Neod siki 15-4-6-1-06 (05/12-05)
Neod siki 14-1-6-5-05 (07/14-17) #	0.048				
Neod siki 14-1-6-5-05 (07/14-20) #	0.060	0.046			
Neod siki 14-3-1-6-02 (07/03-16)	0.063	0.071	0.084		
Neod siki 15-4-6-1-06 (05/12-05)	0.115	0.134	0.121	0.107	
Neod siki 15-4-6-2-07 (05/15-05)	0.115	0.134	0.121	0.110	0.007

b)

	Alti stol MF040961 China	Neod iren KU214684 China	Neod iren GU908336 China	Neod leuc KP190222 China	Neod siki JF906124 China	Neod siki HQ123606 China
Neod siki 13-1-4-5-02 (01/16-15)	0.138	0.106	0.101	0.089	0.103	0.103
Neod siki 14-1-6-5-05 (07/14-17) #	0.141	0.115	0.127	0.113	0.120	0.113
Neod siki 14-1-6-5-05 (07/14-20) #	0.137	0.098	0.114	0.105	0.110	0.105
Neod siki 14-3-1-6-02 (07/03-16)	0.136	0.098	0.096	0.082	0.093	0.089
Neod siki 15-4-6-1-06 (05/12-05)	0.142	0.103	0.107	0.105	0.114	0.105
Neod siki 15-4-6-2-07 (05/15-05)	0.144	0.105	0.105	0.103	0.116	0.107

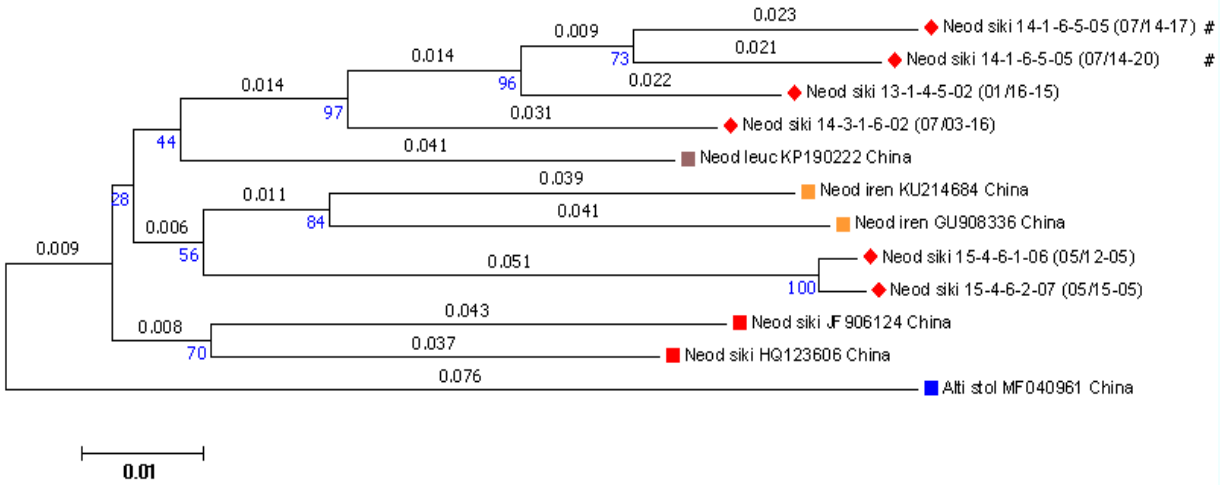


Figure 5.2 The phylogeny of six sample sequences (colour-coded diamond symbols) and six reference sequences (colour-coded with square symbols) of Cricetidae. The dendrogram was generated using the neighbour-joining algorithm and branch lengths are proportional to p-distances (displayed as black figures above the lines when >0.005). Blue figures below the nodes represent percentage of 1000 bootstrap simulations. The two sequences marked # were obtained from the same tissue sample.

5.3.4 Analysis of Sample Sequences of Muridae with GenBank Reference Sequences

There was a total of 53 sample sequences belonging to the family Muridae, plus 28 selected reference sequences. All 81 sequences were included in a common alignment, a single distance matrix and a

single phylogeny. However, to clarify the presentation of these results, they have been divided into three groups; *Apodemus/Mus*, *Niviventer* and *Rattus*.

Table 5.4 Pair-wise p-distance matrices for sequences for *Apodemus* and *Mus*; a) The square matrix between six sample sequences and b) the rectangular matrix between them and the six reference sequences. Distances highlighted in red are less than the con-specific threshold (0.070) and those highlighted in blue are greater than the con-generic threshold (0.160).

a)

	Apod gurk 13-1-4-3-01 (01/14-15)	Apod gurk 13-1-4-3-06 (10/01-14)	Apod gurk 13-1-4-4-01 (01/15-15)	Apod gurk 14-1-6-5-09 (07/15-20)	Mus bood 15-4-1-1-01 (03/01-03)	Mus bood 15-4-3-1-01 (03/14-04)	Mus bood 15-4-5-1-01 (05/06-05)
Apod gurk 13-1-4-3-06 (10/01-14)	0.149						
Apod gurk 13-1-4-4-01 (01/15-15)	0.002	0.152					
Apod gurk 14-1-6-5-09 (07/15-20)	0.005	0.149	0.007				
Mus bood 15-4-1-1-01 (03/01-03)	0.184	0.213	0.187	0.181			
Mus bood 15-4-3-1-01 (03/14-04)	0.191	0.208	0.193	0.187	0.023		
Mus bood 15-4-5-1-01 (05/06-05)	0.185	0.207	0.188	0.182	0.018	0.005	
Mus bood 15-4-6-1-02 (05/11-05)	0.192	0.221	0.194	0.191	0.021	0.035	0.030

b)

	Apod gurk AB032852 Nepal	Apod pall MG748222 China	Apod sylv AB033695 Japan	Apod sylv AF159395	Mus bood KY587424 Nepal	Mus cerv AB125766 Cambodia	Mus cook KX790791 Pakistan	Mus musc AB649506 Nepal Tukuche	Mus paha AY057814	Mus paha EU349767	Mus plat AB125782 India	Mus terr AB125777 Nepal
Apod gurk 13-1-4-3-01 (01/14-15)	0.030	0.169	0.176	0.174	0.178	0.188	0.197	0.183	0.201	0.197	0.188	0.165
Apod gurk 13-1-4-3-06 (10/01-14)	0.159	0.231	0.214	0.222	0.227	0.236	0.224	0.205	0.236	0.236	0.219	0.205
Apod gurk 13-1-4-4-01 (01/15-15)	0.032	0.172	0.178	0.176	0.181	0.190	0.199	0.185	0.204	0.199	0.190	0.167
Apod gurk 14-1-6-5-09 (07/15-20)	0.028	0.166	0.175	0.170	0.182	0.184	0.193	0.179	0.198	0.193	0.186	0.166
Mus bood 15-4-1-1-01 (03/01-03)	0.183	0.179	0.179	0.204	0.147	0.138	0.167	0.021	0.167	0.183	0.144	0.156
Mus bood 15-4-3-1-01 (03/14-04)	0.188	0.181	0.178	0.215	0.151	0.149	0.172	0.007	0.169	0.181	0.142	0.151
Mus bood 15-4-5-1-01 (05/06-05)	0.187	0.180	0.178	0.210	0.148	0.146	0.169	0.002	0.169	0.182	0.144	0.153
Mus bood 15-4-6-1-02 (05/11-05)	0.196	0.187	0.187	0.212	0.157	0.152	0.177	0.032	0.180	0.189	0.161	0.166

There were eight sample sequences in the first group, four *Apodemus* and four *Mus*. The square matrix shows that all the pair-wise p-distances between the two genera (mean = 0.194) were greater than the con-generic threshold of 0.160 (Table 5.4a). The four *M. booduga* samples all had very small p-distances (mean = 0.022), as did three of the *A. gurkha* sequences. However, one *A. gurkha* sequence (13-1-4-3-06) had all three p-distances greater than the con-specific threshold of 0.070, which effect

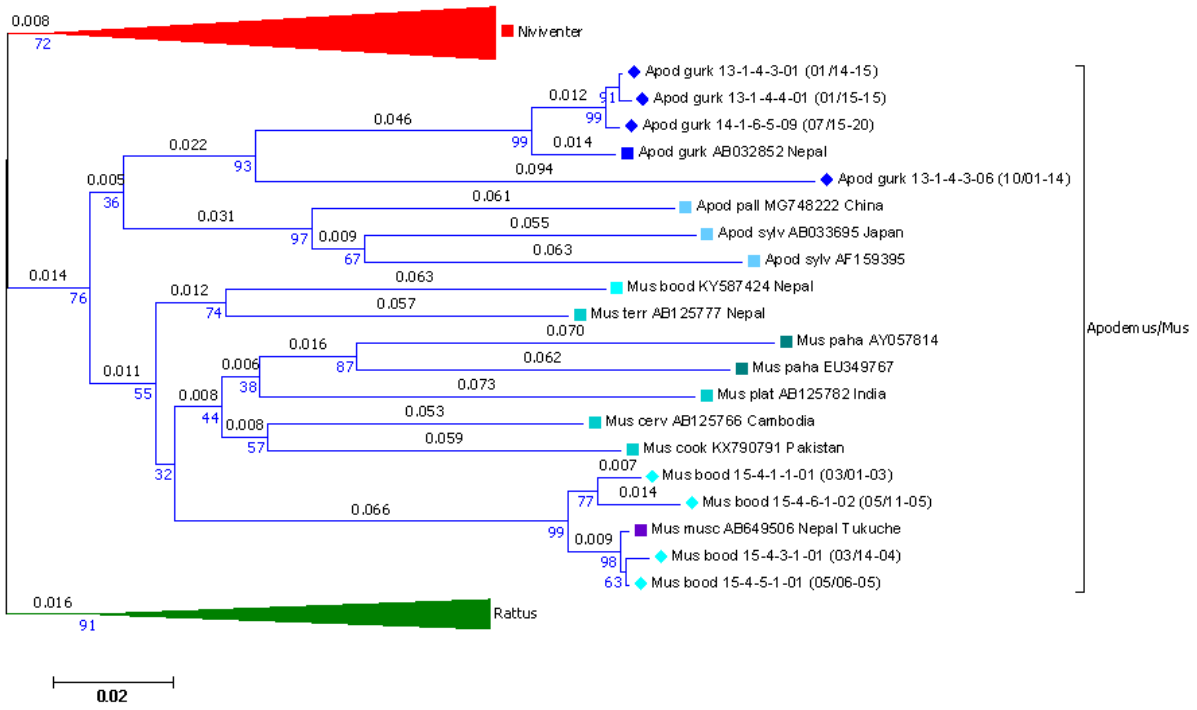


Figure 5.3 The phylogeny of eight sample sequences (colour-coded diamond symbols) and 12 reference sequences (colour-coded with square symbols) for *Apodemus/Mus*. Figures and symbols are explained in Figure 5.2. The sub-trees of the other two genera within this family are collapsed.

is clearly seen in Figure 5.3. However, this sequence had a very low phred score (27.7) so the p-distance was probably an artefact.

Twelve reference sequences were selected from these two genera. For the *A. gurkha* samples, the rectangular matrix (Table 5.3b and Figure 5.3) showed three very small p-distances with the reference sequence from Nepal, but not for the aberrant sample described above. However, all *A. gurkha* samples had p-distances greater than the con-generic threshold of 0.160 to the other three *Apodemus* reference sequences. The square matrix for the *Mus* sequences showed that all four *M. booduga* samples had extremely small p-distances from the *M. musculus* reference sequence (Table 5.4b). In contrast, the p-distances to the *M. booduga* reference sequence were all considerably greater than the con-specific threshold; indeed, they were near the con-generic threshold of 0.160. The p-distances to the other five references for *Mus* species were also much greater than the con-specific threshold.

By far the largest number of good quality sample sequences were identified in the field as genus *Niviventer*; nine *N. eha*, 22 *N. fulvescens* and one *N. niviventer*. This large number of samples precludes the display of the square matrix, but reference to the phylogeny (Figure 5.4) will allow between-sample p-distances to be inferred and exact values are given below where relevant.

Duplicate sequences were obtained for three specimens of *N. fulvescens*. All three sets clustered very closely together, with p-distances of 0.012, 0.011 and a mean of three equal to 0.006, indicating a

high degree of consistency in the evaluation of these sequences. The sample sequences partition into two very clear branches; a very tightly clustered *N. fulvescens* group with a mean p-distance of only 0.010, and a more dispersed group of *N. eha*, with a mean p-distance within the group of 0.046.

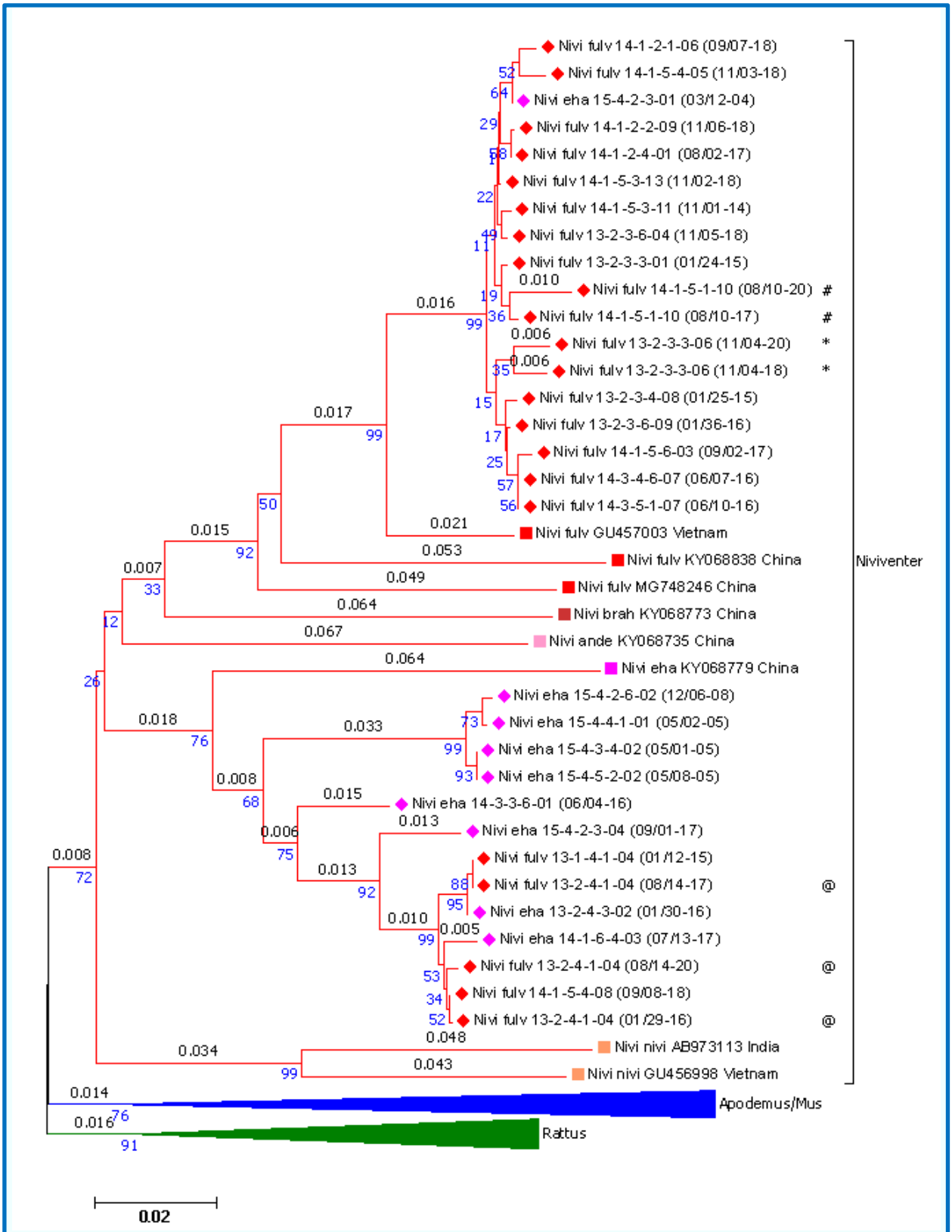


Figure 5.4 The phylogeny of 32 sample sequences (colour-coded diamond symbols) and eight reference sequences (colour-coded with square symbols) for *Niviventer*. Figures and symbols are explained in Figure 5.2. The sub-trees of the other three genera within this family are collapsed. Sequences marked with #, * or @ symbols indicate duplicate sequences from three individual animals.

Both of these values were well below the con-specific thresholds for their taxa (0.090 and 0.095 respectively). The mean p-distance between the two groups was 0.122, indicating a clear species distinction. However, there was a degree of overlap between these two groups, with one *N. eha* specimen (15-4-2-3-01) clearly mis-identified in the field, with three putative specimens of *N. fulvescens* (13-1-4-1-04, 13-2-4-1-04 & 14-1-5-4-08) found in the *N. eha* group. Note that the single sample sequence of *N. niviventer* does not appear in this phylogeny (see below).

Table 5.5 Pair-wise rectangular distance matrices for sample sequences of a) nine *N. eha*, b) 22 *N. fulvescens* and c) one *N. niviventer*, with eight reference sequences. Distances highlighted in red are less than the con-specific threshold (0.095, 0.090 & 0.095 respectively) and those highlighted in blue are greater than the con-generic threshold (0.150). The three animals marked with #, * and @ symbols indicate the mean p-distances from duplicate sequences.

a)	Nivi ande KY068735 China	Nivi brah KY068773 China	Nivi eha KY068779 China	Nivi fulv GU457003 Vietnam	Nivi fulv KY068838 China	Nivi fulv MG748246 China	Nivi nivi AB973113 India	Nivi nivi GU456998 Vietnam
Nivi eha 13-2-4-3-02 (01/30-16)	0.151	0.117	0.106	0.144	0.161	0.128	0.161	0.154
Nivi eha 14-1-6-4-03 (07/13-17)	0.152	0.125	0.109	0.143	0.162	0.129	0.162	0.155
Nivi eha 14-3-3-6-01 (06/04-16)	0.128	0.110	0.112	0.121	0.140	0.126	0.162	0.156
Nivi eha 15-4-2-3-01 (03/12-04)	0.128	0.119	0.152	0.040	0.091	0.089	0.154	0.149
Nivi eha 15-4-2-3-04 (09/01-17)	0.144	0.112	0.110	0.137	0.160	0.137	0.169	0.158
Nivi eha 15-4-2-6-02 (12/06-08)	0.132	0.123	0.096	0.132	0.148	0.123	0.157	0.150
Nivi eha 15-4-3-4-02 (05/01-05)	0.130	0.121	0.098	0.130	0.150	0.121	0.155	0.153
Nivi eha 15-4-4-1-01 (05/02-05)	0.128	0.123	0.096	0.132	0.148	0.123	0.158	0.153
Nivi eha 15-4-5-2-02 (05/08-05)	0.130	0.121	0.098	0.130	0.150	0.121	0.155	0.153
b)								
Nivi fulv 13-1-4-1-04 (01/12-15)	0.156	0.119	0.110	0.149	0.162	0.130	0.165	0.156
Nivi fulv 13-2-3-3-01 (01/24-15)	0.133	0.121	0.156	0.043	0.094	0.092	0.156	0.151
Nivi fulv 13-2-3-3-06 (11/04-20) *	0.134	0.126	0.158	0.045	0.096	0.098	0.157	0.153
Nivi fulv 13-2-3-4-08 (01/25-15)	0.135	0.124	0.158	0.046	0.096	0.094	0.158	0.153
Nivi fulv 13-2-3-6-04 (11/05-18)	0.130	0.121	0.155	0.041	0.091	0.089	0.153	0.148
Nivi fulv 13-2-3-6-09 (01/36-16)	0.132	0.123	0.158	0.043	0.094	0.091	0.155	0.151
Nivi fulv 13-2-4-1-04 (08/14-20) @	0.151	0.115	0.105	0.144	0.158	0.126	0.160	0.151
Nivi fulv 14-1-2-1-06 (09/07-18)	0.137	0.123	0.158	0.048	0.098	0.096	0.160	0.155
Nivi fulv 14-1-2-2-09 (11/06-18)	0.128	0.119	0.158	0.043	0.094	0.091	0.155	0.151
Nivi fulv 14-1-2-4-01 (08/02-17)	0.125	0.113	0.152	0.039	0.090	0.088	0.152	0.148
Nivi fulv 14-1-5-1-10 (08/10-20) #	0.133	0.125	0.157	0.048	0.095	0.093	0.156	0.152
Nivi fulv 14-1-5-3-11 (11/01-14)	0.129	0.120	0.154	0.039	0.090	0.088	0.152	0.147
Nivi fulv 14-1-5-3-13 (11/02-18)	0.130	0.121	0.155	0.041	0.091	0.089	0.153	0.148
Nivi fulv 14-1-5-4-05 (11/03-18)	0.131	0.120	0.154	0.044	0.090	0.097	0.156	0.152
Nivi fulv 14-1-5-4-08 (09/08-18)	0.149	0.113	0.103	0.143	0.156	0.124	0.159	0.149
Nivi fulv 14-1-5-6-03 (09/02-17)	0.132	0.119	0.158	0.039	0.094	0.096	0.155	0.146
Nivi fulv 14-3-4-6-07 (06/07-16)	0.130	0.121	0.155	0.041	0.091	0.094	0.158	0.148
Nivi fulv 14-3-5-1-07 (06/10-16)	0.130	0.121	0.155	0.041	0.091	0.094	0.158	0.148
c)								
Nivi nivi 14-1-5-6-07 (07/08-17)	0.170	0.149	0.165	0.177	0.183	0.163	0.172	0.170

The rectangular matrix included eight reference sequences, one of *N eha*, three *N. fulvescens* and two *N. niviventer*, plus two out-group species (Table 5.5). The single *N. eha* reference sequence was most closely linked to the *N. eha* group, but none of the p-distances to the sample sequences were within the con-specific threshold (0.095), although four distances were less than 0.10. Furthermore, these sequences (including the three putative *N. fulvescens*,) were all more distant to the *N. fulvescens* reference sequences and they were generally above the con-generic threshold distance to the two *N. niviventer* reference sequences. The sample sequences in the *N. fulvescens* group were all within the con-specific threshold distance to at least one of the reference sequences (GU457003) and many to all three reference sequences. The p-distances to the outgroup references were much greater, and the *N. niviventer* reference sequences were often further than the con-generic distance threshold. The single sample sequence of *N. niviventer* in Figure 5.4 was extremely distant from all eight reference sequences; with p-distances generally greater than the con-generic threshold (see below).

Seven animals identified in the field as *Rattus nitidus* generated eight good quality sequences. In addition, two animals were recorded as unknown *Rattus sp.* and three animals were identified as two different unknown murid species. The square matrix (Table 5.6a) shows that the duplicate sequences from two specimens had very close p-distances (0.007 and 0.005). The pattern of p-distances was complex, with all sequences except one having values of less than the con-specific thresholds with at least one other sequence. The exception (murid sp2. 15-4-1-1-11) had a reasonably good quality sequence with a phred score of 42.5, so it is difficult to explain its outlying status (Figure 5.5). Note also that the single *N. niviventer* sequence had very small p-distances with five other sequences in this group, which indicated a high probability of an incorrect field identification.

The rectangular matrix (Table 5.6b) shows that only one *R. nitidus* sample sequence (14-3-5-4-03) had p-distances less than the con-specific threshold with the two reference sequences. All other sample sequences had p-distances greater than 0.11 to the *R. nitidus* references and three of them were greater than the con-generic threshold. Two of the *R. nitidus* samples, the two *Rattus* sp. samples and the two murid sp2 samples all fell within a group that included the *R. pyctoris* reference sequence – even the aberrant murid sp2 mentioned above only had a p-distance of 0.067, although this was greater than the con-specific threshold. The final grouping included three *R. nitidus*, the murid sp1 and the *N. niviventer* sample sequences, with the *R. andamanensis* reference sequence. All five samples had p-distances greater than the con-specific threshold to the reference sequences of *R. rattus* and *R. tanezumi*.

Table 5.6 Pair-wise p-distance matrices for sequences for *Rattus* and *N. niviventer*; a) The square matrix between 13 sample sequences and b) the rectangular matrix between them and the eight reference sequences. Distances highlighted in red are less than the con-specific thresholds (0.050 & 0.060) and those highlighted in blue are greater than the con-generic threshold (0.130). The two animals marked with # and * symbols indicate duplicate sequences, with mean values given in b).

a)	Muri Sp1 14-1-1-6-06 (04/12-16)	Muri Sp2 15-4-1-1-11 (03/04-03)	Muri Sp2 15-4-1-2-05 (03/05-04)	Ratt niti 13-1-2-1-06 (01/05-14)	Ratt niti 14-1-2-1-11 (04/13-20) #	Ratt niti 14-1-2-1-11 (04/13-16) #	Ratt niti 14-1-5-1-04 (08/09-17)	Ratt niti 14-1-5-4-07 (08/12-20) *	Ratt niti 14-1-5-4-07 (08/12-17) *	Ratt niti 14-1-5-5-04 (07/07-16)	Ratt niti 14-3-5-4-03 (06/12-20)	Ratt sp 15-4-1-4-03 (03/06-04)	Ratt sp 15-4-1-4-14 (03/08-04)
Muri Sp2 15-4-1-1-11 (03/04-03)	0.148												
Muri Sp2 15-4-1-2-05 (03/05-04)	0.094	0.065											
Ratt niti 13-1-2-1-06 (01/05-14)	0.092	0.074	0.021										
Ratt niti 14-1-2-1-11 (04/13-20) #	0.094	0.090	0.039	0.041									
Ratt niti 14-1-2-1-11 (04/13-16) #	0.094	0.088	0.039	0.044	0.007								
Ratt niti 14-1-5-1-04 (08/09-17)	0.007	0.152	0.099	0.096	0.096	0.099							
Ratt niti 14-1-5-4-07 (08/12-20) *	0.007	0.152	0.094	0.100	0.089	0.094	0.014						
Ratt niti 14-1-5-4-07 (08/12-17) *	0.002	0.150	0.094	0.096	0.092	0.094	0.009	0.005					
Ratt niti 14-1-5-5-04 (07/07-16)	0.002	0.152	0.096	0.096	0.094	0.099	0.009	0.005	0.005				
Ratt niti 14-3-5-4-03 (06/12-20)	0.119	0.180	0.124	0.126	0.138	0.140	0.121	0.112	0.117	0.117			
Ratt sp 15-4-1-4-03 (03/06-04)	0.113	0.076	0.023	0.044	0.062	0.062	0.117	0.112	0.112	0.115	0.142		
Ratt sp 15-4-1-4-14 (03/08-04)	0.096	0.067	0.000	0.023	0.041	0.041	0.100	0.096	0.096	0.098	0.126	0.027	
Nivi nivi 14-1-5-6-07 (07/08-17)	0.002	0.150	0.097	0.094	0.097	0.097	0.005	0.009	0.005	0.005	0.122	0.115	0.098

b)	Ratt anda JQ814208 China	Ratt niti AB973108 India	Ratt niti KU214575 Nepal	Ratt pyct KY587429 Nepal	Ratt ratt HM217741 India	Ratt ratt KY002805 Nepal	Ratt tane AB096841 China	Ratt tane KP343797 Afghanistan
Muri Sp1 14-1-1-6-06 (04/12-16)	0.016	0.114	0.121	0.096	0.085	0.076	0.082	0.080
Muri Sp2 15-4-1-1-11 (03/04-03)	0.156	0.175	0.182	0.067	0.149	0.143	0.154	0.140
Muri Sp2 15-4-1-2-05 (03/05-04)	0.089	0.114	0.121	0.000	0.085	0.085	0.089	0.076
Ratt niti 13-1-2-1-06 (01/05-14)	0.105	0.126	0.128	0.023	0.103	0.098	0.105	0.091
Ratt niti 14-1-2-1-11 (04/13-20) #	0.100	0.135	0.142	0.041	0.101	0.092	0.100	0.082
Ratt niti 14-1-5-1-04 (08/09-17)	0.018	0.116	0.123	0.100	0.087	0.078	0.080	0.084
Ratt niti 14-1-5-4-07 (08/12-20) *	0.014	0.112	0.119	0.096	0.082	0.073	0.080	0.080
Ratt niti 14-1-5-5-04 (07/07-16)	0.016	0.114	0.121	0.098	0.084	0.075	0.082	0.082
Ratt niti 14-3-5-4-03 (06/12-20)	0.123	0.025	0.011	0.126	0.128	0.119	0.128	0.121
Ratt sp 15-4-1-4-03 (03/06-04)	0.110	0.133	0.140	0.025	0.110	0.108	0.110	0.101
Ratt sp 15-4-1-4-14 (03/08-04)	0.089	0.116	0.123	0.000	0.084	0.087	0.089	0.075
Nivi nivi 14-1-5-6-07 (07/08-17)	0.018	0.117	0.124	0.099	0.087	0.078	0.085	0.083

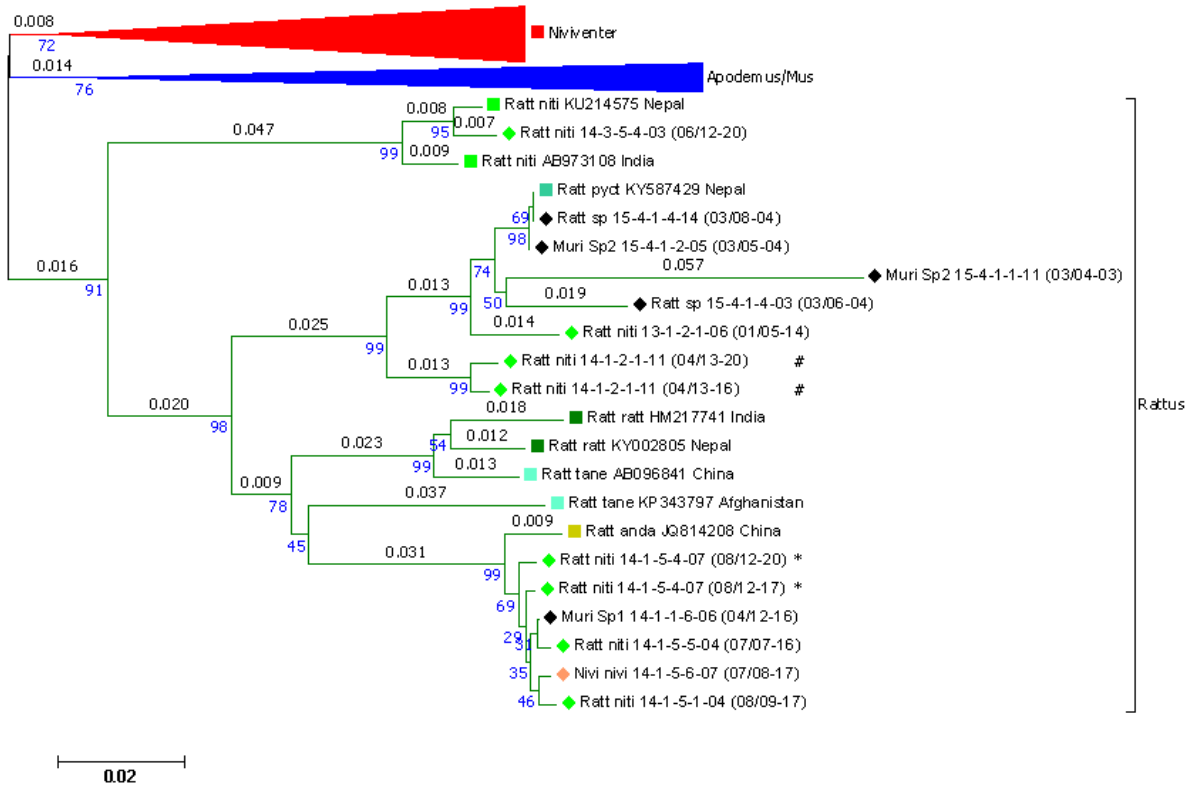


Figure 5.5 The phylogeny of 13 sample sequences (colour-coded diamond symbols) and eight reference sequences (colour-coded with square symbols) for *Rattus*. Figures and symbols are explained in Figure 5.2. The sub-trees of the other three genera within this family are collapsed. Sequences marked with # or * symbols indicate duplicate sequences from two individual animals.

5.3.5 Analysis of Sample Sequences of Soricidae with GenBank Reference Sequences

There were 31 good sample sequences from 28 soricid specimens. Three sequences from a single animal (14-3-2-5-01) had extremely small p-distances, averaging 0.002. However, the other two duplicates (from animal 14-3-1-2-01) had a p-distance of 0.181 (see Section 5.4 for a discussion of this issue). The square matrix of the 13 *E. leucops* sample sequences had p-distances less than the con-specific threshold of 0.070. Furthermore, they showed a clear geographic split into two branches (Figure 5.6), one with seven specimens from Annapurna, with a mean p-distance within the group of 0.007, and one of five from Sagarmatha plus one from Langtang with a mean intra-group distance of 0.014. The mean distance between groups was 0.039, which was still well below the con-specific

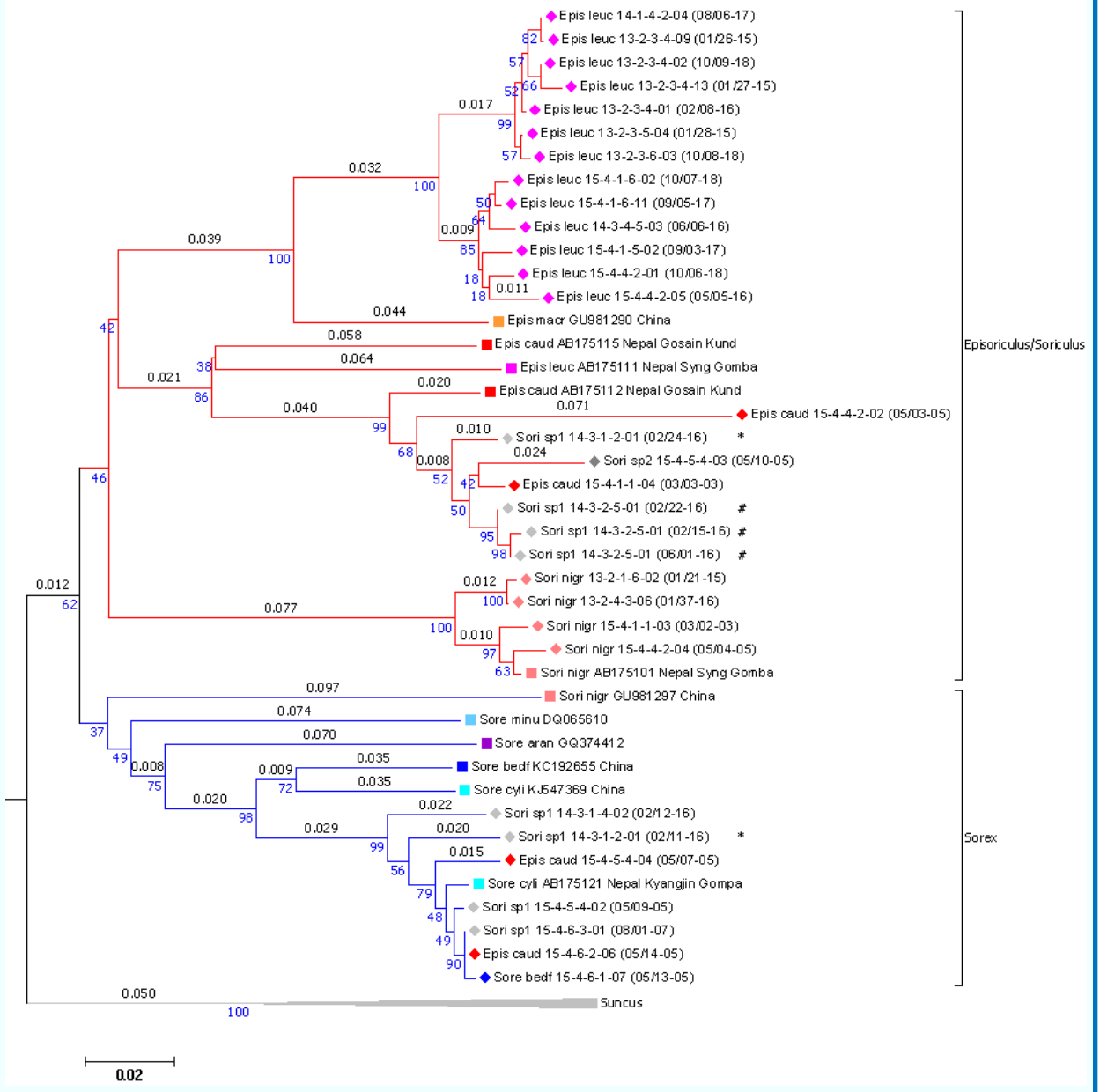


Figure 5.6 The phylogeny of 31 sample sequences (colour-coded diamond symbols) and 11 reference sequences (colour-coded with square symbols) for *Soricidae*. Figures and symbols are explained in Figure 5.2. The sub-tree of the genus *Suncus* has been collapsed. Sequences marked with # or * symbols indicate duplicate sequences from two individual animals.

threshold of 0.080. The four sequences of *S. nigrescens* also clustered very tightly together, with a mean p-distance of 0.022, and a mean p-distance to all other sequences in the *Episoriculus/Soriculus* group of 0.184. The remaining sample sequences were four *E. caudatus*, one *Sorex bedfordiae* and nine soricid sequences. These split equally between two totally separate groups, with no taxonomic or geographic pattern.

The rectangular matrix (Table 5.7) confirmed the identity of the four *S. nigrescens* samples with the *S. nigrescens* reference sequence from Nepal, with a mean p-distance of 0.021. However, they were very distant from the *S. nigrescens* reference from China; approx. 0.194, although the bootstrap

Table 5.7 Pair-wise rectangular p-distance matrix for 29 sample sequences and 11 reference sequences of Soricidae. Distances highlighted in red are less than the con-specific thresholds (0.08, 0.09 & 0.05) and those highlighted in blue are greater than the con-generic threshold (0.12, 0.15 & 0.16) for *Episoriculus*, *Sorex* and *Soriculus* respectively. The two sequences marked with * symbols indicate duplicate sequences with very different p-distances and the single animal marked # represent the mean p-distances of three duplicate sequences (see text for explanation).

	Epis caud AB175112 Nepal Gosain Kund	Epis caud AB175115 Nepal Gosain Kund	Epis leuc AB175111 Nepal Syng Gomba	Epis macr GU981290 China	Sore aran GQ374412	Sore bedf KC192655 China	Sore cyli AB175121 Nepal Kyangjin Gompa	Sore cyli KJ547369 China	Sore minu DQ065610	Sori nigr AB175101 Nepal Syng Gomba	Sori nigr GU981297 China
Sori sp1 14-3-1-2-01 (02/24-16) *	0.042	0.119	0.133	0.173	0.173	0.194	0.180	0.166	0.183	0.175	0.182
Sori sp1 14-3-1-2-01 (02/11-16) *	0.190	0.197	0.209	0.188	0.149	0.098	0.026	0.095	0.167	0.169	0.188
Sori sp1 14-3-1-4-02 (02/12-16)	0.188	0.181	0.204	0.193	0.153	0.105	0.042	0.102	0.151	0.179	0.193
Sori sp1 14-3-2-5-01 (02/15-16) #	0.033	0.126	0.129	0.172	0.180	0.201	0.189	0.178	0.185	0.181	0.180
Sori sp1 15-4-5-4-02 (05/09-05)	0.184	0.173	0.186	0.177	0.129	0.070	0.009	0.077	0.161	0.164	0.193
Sori sp1 15-4-6-3-01 (08/01-07)	0.180	0.173	0.186	0.177	0.134	0.073	0.009	0.073	0.156	0.161	0.193
Sori sp2 15-4-5-4-03 (05/10-05)	0.054	0.141	0.150	0.188	0.197	0.218	0.206	0.190	0.206	0.200	0.204
Epis caud 15-4-1-1-04 (03/03-03)	0.032	0.118	0.125	0.168	0.177	0.198	0.180	0.166	0.182	0.179	0.181
Epis caud 15-4-4-2-02 (05/03-05)	0.104	0.172	0.179	0.226	0.238	0.259	0.236	0.229	0.238	0.233	0.235
Epis caud 15-4-5-4-04 (05/07-05)	0.195	0.184	0.200	0.188	0.138	0.079	0.025	0.091	0.172	0.175	0.204
Epis caud 15-4-6-2-06 (05/14-05)	0.180	0.173	0.186	0.177	0.134	0.073	0.009	0.073	0.156	0.161	0.193
Epis leuc 14-1-4-2-04 (08/06-17)	0.174	0.174	0.183	0.099	0.193	0.182	0.193	0.195	0.184	0.188	0.197
Epis leuc 13-2-3-4-01 (02/08-16)	0.169	0.169	0.176	0.092	0.188	0.178	0.185	0.188	0.178	0.189	0.196
Epis leuc 13-2-3-4-02 (10/09-18)	0.175	0.175	0.182	0.099	0.194	0.185	0.192	0.194	0.185	0.194	0.203
Epis leuc 13-2-3-4-09 (01/26-15)	0.176	0.176	0.185	0.100	0.197	0.185	0.197	0.197	0.185	0.189	0.201
Epis leuc 13-2-3-4-13 (01/27-15)	0.185	0.185	0.194	0.109	0.203	0.196	0.203	0.205	0.194	0.200	0.210
Epis leuc 13-2-3-5-04 (01/28-15)	0.168	0.168	0.175	0.092	0.187	0.178	0.185	0.187	0.178	0.191	0.196
Epis leuc 13-2-3-6-03 (10/08-18)	0.172	0.166	0.172	0.096	0.182	0.173	0.182	0.182	0.173	0.189	0.196
Epis leuc 14-3-4-5-03 (06/06-16)	0.171	0.167	0.169	0.096	0.190	0.197	0.199	0.195	0.188	0.192	0.210
Epis leuc 15-4-1-5-02 (09/03-17)	0.166	0.168	0.175	0.097	0.187	0.189	0.189	0.187	0.187	0.184	0.209
Epis leuc 15-4-1-6-02 (10/07-18)	0.167	0.167	0.169	0.092	0.183	0.188	0.190	0.188	0.188	0.188	0.211
Epis leuc 15-4-1-6-11 (09/05-17)	0.171	0.171	0.174	0.096	0.188	0.190	0.192	0.190	0.192	0.189	0.215
Epis leuc 15-4-4-2-01 (10/06-18)	0.160	0.156	0.162	0.080	0.177	0.179	0.186	0.183	0.177	0.183	0.204
Epis leuc 15-4-4-2-05 (05/05-16)	0.171	0.166	0.162	0.094	0.176	0.176	0.183	0.176	0.178	0.183	0.206
Sore bedf 15-4-6-1-07 (05/13-05)	0.182	0.175	0.189	0.180	0.136	0.075	0.011	0.075	0.159	0.164	0.196
Sori nigr 13-2-1-6-02 (01/21-15)	0.172	0.181	0.192	0.167	0.193	0.165	0.170	0.170	0.209	0.030	0.201
Sori nigr 13-2-4-3-06 (01/37-16)	0.169	0.178	0.189	0.164	0.190	0.162	0.167	0.167	0.206	0.030	0.199
Sori nigr 15-4-1-1-03 (03/02-03)	0.175	0.175	0.172	0.170	0.184	0.157	0.166	0.164	0.198	0.014	0.188
Sori nigr 15-4-4-2-04 (05/04-05)	0.175	0.184	0.181	0.175	0.193	0.161	0.170	0.168	0.202	0.009	0.188

support for this distinction was not strong. The 13 *E. leucops* samples formed a very distinct cluster, but had no con-specific link to any reference sequence. The nearest was *E. macrurus* (mean 0.095), and the mean distance to the *E. leucops* reference was 0.175 – greater even, than the cross-generic threshold of approx. 0.165.

One of the *E. caudatus* and three of the unknown soricid specimens clustered with one of the *E. caudatus* reference sequences, with p-distances less than the con-specific threshold. Two of the unknown specimens were from sites in Langtang which were only a few kilometres from the location of the reference sequence. However, the *E. caudatus* sequence and the single unknown sp2 came from Sagarmatha, which was 200km away. The other cluster shown in Figure 5.6, containing seven sample sequences, all had p-distances to a *Sorex cylindricauda* reference sequence from Nepal well below the con-specific threshold. Five of them also had distances below the threshold to the reference sequence of *S. bedfordiae*.

5.4 Discussion

This study is the first to include large-scale collection of DNA samples from small mammals in Nepal. I collected tissue samples from 720 animals. These belonged to at least 15 species covering a geographic zone of approx. 400km in the centre of the country and over an altitudinal range from 1300m to 4400m. In contrast, the few contemporary studies of small mammals in Nepal that utilise genetic techniques are based on as few as five (Adhikari *et al.*, 2018a) or 28 specimens (Adhikari *et al.*, 2018b). Even larger epidemiological studies, such as Gundi *et al.* (2010), only used tissue samples from 108 animals.

The main aim of this chapter, the use of genetic techniques to aid identification, has been invaluable to this study of small mammals. There is a dearth of good field identification guides to small mammals in this remote terrain and on-line information is sparse. Even creating my own field handbook from museum specimens (see Chapter 2) did not ensure confident identification of all specimens in the field.

In summary, the phylogenetic analyses have confirmed the identity of nine species, representing over two thirds of all the animals caught (Table 5.8). One species was reassigned to a sister species. Two previously unrecorded species were identified within the sample, comprising eight specimens. Furthermore, 11 unidentified specimens were assigned to a species with a high degree of probability, although I have retained an unidentified status for eight specimens.

More intriguingly, there was clear evidence within my own sample sequences that three species; the vole (*N. sikimensis*), the white-bellied rat (*N. eha*) and the shrew (*E. leucops*) had sufficient geographic variation to constitute at least sub-species distinctions. Furthermore, these three taxa had no clear alignments with reference sequences, with p-distances that were greater than conspecific thresholds to the nearest reference sequences. These issues are discussed further in sub-section 5.4.2.

There appeared to be a number of anomalies in the identities of the GenBank sequences that I used as references. In particular, there was clear confusion in two groups; the Niviventers and the *S. bedfordiae / cylindricauda* group. Furthermore, at least two species appeared to contain cryptic species, with taxonomic groupings that had p-distances well above the general conspecific threshold. These issues are discussed further in sub-section 5.4.1

Table 5.8 Summary of alignments of sample sequences with reference sequences obtained from Gen-bank. Decimal values are mean p-distances between sample and reference sequences. Figures in brackets are the number of specimens when sample sequences are aligned with more than one refer-ence species. Cells highlighted in red have p-distances that are less than the con-specific distances, and blue highlighted cells have p-distances greater than the con-generic distances.

Sample Sequences		N	Reference Sequences						
a)			*Alticola stoliczkanus*	*N. irene*	*N. leucurus*	*N. sikkimensis*			
Neodon	*N. sikkimensis*	5	0.141	0.106	0.101	0.107			
b)			*A. gurkha*	*A. pallipes*	*A. sylvaticus*				
Apodemus	*A. gurkha*	4	0.030	0.169	0.176				
c)			*M. booduga*	*M. cervicolor*	*M. cookii*	*M. musculus*	*M. pahari*	*M. platythrix*	*M. terricolor*
Mus	*M. booduga*	4	0.151	0.146	0.171	0.016	0.178	0.148	0.157
d)			*N. andersoni*	*N. brahma*	*N. eha*	*N. fulvescens*	*N. niviventer*	*R. andamanensis*	
	N. eha?	9	0.136	0.119	0.103 (8)	0.040 (1)	0.157	0.169	
Niviventer	*N. fulvescens*?	18	0.135	0.120	0.152 (3)	0.076 (15)	0.153	0.174	
	N. niviventer?	1	0.170	0.149	0.165	0.173	0.171	0.018	
e)			*R. andamanensis*	*R. nitidus*	*R. pyctoris*	*R. rattus*	*R. tanezumi*		
	Murid sp1	1	0.016	0.118	0.096	0.081	0.081		
	Murid sp2	2	0.123	0.148	0.034	0.116	0.115		
Rattus	*R. nitidus*	6	0.016 (3)	0.018 (1)	0.032 (2)	0.093	0.093		
	Rattus sp	2	0.100	0.128	0.013	0.097	0.094		
f)			*E. caudatus*	*E. leucops*	*E. macrurus*	*S. bedfordiae/ cylindriccauda*	*S. nigrescens*		
	Soricid sp1	6	0.038 (2)	0.204 (1)	0.193 (1)	0.050 (3)	0.180		
	Soricid sp2	1	0.054	0.150	0.188	0.205	0.202		
Episoriculus	*E. caudatus*	4	0.032 (1)	0.179	0.226	0.058 (2)	0.195		
	E. leucops	13	0.171	0.175	0.096	0.188	0.197		
Sorex	*S. bedfordiae*	1	0.179	0.189	0.180	0.054	0.180		
Soriculus	*S. nigrescens*	4	0.176	0.184	0.169	0.166	0.021		

5.4.1 The use of GenBank as a Reference Source for DNA Barcoding

Many studies have utilised the enormous resources of the National Center for Biotechnology Infor-mation (NCBI). One of their services is the Basic Logical Alignment Search Tool (BLAST) which facilitates the procedure of DNA barcoding (Abdo and Golding, 2007; Cristescu, 2014; Hebert *et al.*, 2003; Kress and Erickson, 2008). Several candidates for barcoding genes have been proposed, which are highly conserved within species, but exhibit sufficient variation between species to allow discrimination. In particular, mitochondrial cytochrome c oxidase sub-unit 1 (CO1) (Batovska *et al.*, 2016; Hebert *et al.*, 2003; Robins *et al.*, 2007) and cytochrome-b (Lack *et al.*, 2012; Parson *et al.*, 2000) have been widely used. In this study I followed the procedures of many previous researchers on small mammals in south, south-east and east Asia (Jing *et al.*, 2007; Ohdachi *et al.*, 2004; Suzuki *et al.*, 2013) by utilising the cytochrome-b gene.

In Supplementary Information 5.C, I analysed 200 sequences selected from the NCBI GenBank database. This preliminary analysis showed that there was a high degree of variation in the mean p-distances between taxonomic groupings. For example, the four sequences of *Mus cookii*, two from

Laos, one from Pakistan and one other (Figure 5.30), had virtually identical sequences, with mean pair-wise p-distances of < 0.001. In contrast, I selected seven sequences of the shrew *Soriculus nigrescens*, five from China and two from Nepal (Figure 5.39). Within their geographical groups, they had mean p-distances of < 0.030, but the mean distance between these two groups was 0.160 (Figure 5.38d). A third scenario existed for the genus *Niviventer* (Figure 5.35), from which I selected 36 sequences from six species. The three species of interest in this study were scattered throughout the phylogeny. One specimen of *N. niviventer* from Vietnam was virtually identical to two *N. fulvescens* from China and Thailand. One sequence of *N. fulvescens* fell into a group of *N. niviventer* and the *N. eha* sequences fell into two groups, one of which was closely linked to the *N. niviventer* group.

These examples highlight one of the major problems with uncritical use of BLAST and GenBank. I have provided very clear evidence from the *Niviventer* sequences that basic taxonomic misidentification has occurred and is now lodged, presumably permanently, in GenBank. Collins and Cruickshank (2013) eloquently describe this as their "Second Deadly Sin" (of DNA barcoding). They cite the example of Barcode Index Number clusters of cyprinid fish in the Barcode of Life (Ratnasingham and Hebert, 2007). Although these BIN clusters were supposed to contain unique species, 37 out of 70 clusters (53%) contained more than one species name. Applying the BLAST algorithm to an unknown sequence, simply matches it to a large number of GenBank sequences ordered by probability value and/or putative genetic distance. It says nothing about the relationship between the sequences to which it has been matched, for example whether they come from the same "pop-set", which would imply that identifications were likely to be consistent.

My approach of undertaking a preliminary phylogenetic analysis of reference sequences allowed some of these issues to be exposed. Thus, I was able to select single reference sequences that were representative of tight species clusters, but two or more sequences from species where there was clear discrepancy in identification or large variation in p-distances. It also allowed me to quantify three taxonomic levels of genetic distance; conspecific, congeneric and confamilial. After separating the 200 GenBank sequences into family groups, I was able to calculate 17,184 pair-wise p-distances. Excluding the obvious misidentifications discussed above, 70.0% of the variation in p-distance could be explained by the taxonomic level. The median distances were; conspecific, 0.023; congeneric, 0.127 and confamilial, 0.172. However, a further 7.5% of variation could be explained by cross-tabulating with the three families (Figure 5.7). This shows that the increase in p-distance with increasing taxonomic distance was different in the three families ($F_{(4,13980)} = 228$, $p = 0$). But, more interesting than the actual values are the zones between the taxonomic levels. So, for example, there is a clear distinction at around p-distance = 0.09 for Cricetidae between conspecific and congeneric distances. Similarly, between congeneric and confamilial, there

is a clear zone at 0.125. The distinctions are not so clear for the other two families but, overall, there appears to be a conspecific threshold of around 0.08.

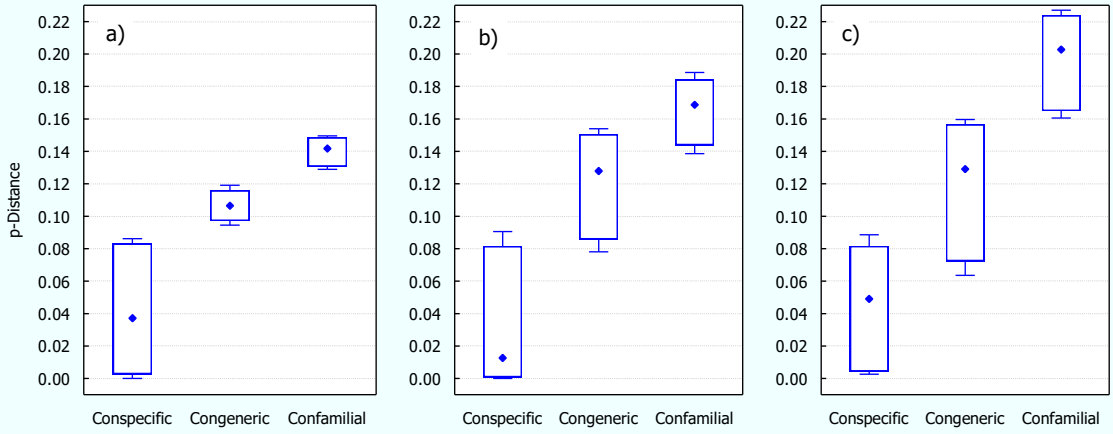


Figure 5.7 Box & whisker plot of pair-wise p-distances for three taxonomic levels for a) Cricetidae, b) Muridae and c) Soricidae. Diamonds show medians, boxes show the 10th and 90th percentiles and whiskers show 5th and 95th percentiles.

The concept of thresholds has been researched extensively, both for specimen identification and species discovery, as distinguished by (Collins and Cruickshank, 2013). For example, Meyer and Paulay (2005) indicated that for specimen identification, a conspecific threshold of 0.034 in cowries would result in only 17% error rate, and for turbinid gastropods the threshold was as low as 0.014. These values are considerably lower than those I have indicated here, although (Meyer and Paulay, 2005) indicated that 0.08 was a suitable threshold for species discovery. Bradley and Baker (2001) reported mean con-specific p-distances of 0.021 for four genera of rodents, although the range was wide (<0.001 to > 0.080). Their con-generic p-distances averaged 0.132, again with a wide range. By applying the appropriate conspecific threshold for each family to the data I have presented above, there was only a 9.6% error rate in 3731 pair-wise comparisons. So, although I concur with (Meyer and Paulay, 2005) that there was no true "barcoding gap" in these data, the error rate in applying such a threshold would only be in the order of 5% in each direction.

A few studies of small mammals in south Asia have attempted to quantify genetic distances in this way. Adhikari *et al.* (2018a) reported p-distances for cytochrome-b between *R. rattus* and *R. tanezumi* (a con-generic distance) in Nepal of only 0.043. (Shimada *et al.*, 2010) gave conspecific distances for *M. booduga* from Nepal and India of 0.022 and congeneric distances between *M. booduga* and *M. musculus* of 0.150. Suzuki *et al.* (2004) reported conspecific distances in four species of *Mus* (including *M. booduga* and *M. musculus*) of up 0.034. In a detailed study of subspecies of *Mus musculus*, (Adhikari *et al.*, 2018b) found maximum conspecific distances 0.040.

By taking this approach, I have uncovered a number of examples in GenBank where species misi-dentifications appear to have occurred in the taxa of interest to this study;

- One specimen of *Apodemus wardi* fell into a group of four sequences of *A. pallipes*, with a mean p-distance of 0.025 (Figure 5.28).
- One specimen of *Mus saxicola* was identical to a sequence of *M. platythrix* (Figure 5.30).
- There was a general confusion in the identification of three species of *Niviventer* (Figure 5.33 to Figure 5.35). 66 out of 202 (32.7%) conspecific p-distances were > 0.085 and 110 out of 1058 (10.4%) congeneric p-distances were < 0.085.
- One sequence of *R. tanezumi* fell into a group of *R. andamanensis* with a mean p-distance of 0.005 (Figure 5.36 & Figure 5.37).
- There was also general confusion in the identification of *Sorex bedfordiae* and *S. cylindricauda* (Figure 5.41). All 98 conspecific p-distances were < 0.078, so there was no clear evidence that these were two distinct species.

In addition to misidentifications, I have evidence for the presence of cryptic species within the GenBank sequences:

- One specimen of *Neodon irene* from Nepal had a mean p-distance of 0.102 from the main *N. irene* group of eight specimens (Figure 5.27). This could have been a misidentification as it was slightly closer to two sequences of *N. leucurus*, but the mean p-distance to them was still 0.095.
- One specimen of *Mus Pahari* was very remote from three other sequences of this species (Figure 5.32), with a mean p-distance of 0.112.
- Finally, the strongest evidence for cryptic species comes from the seven sequences of *Soriculus nigrescens*.(Figure 5.39). In India, Nepal and China this is a mono-specific genus (Baral and Shah, 2008; Smith and Xie, 2008). However, these sequences clearly fell into two branches; two specimens from Nepal and five from China, with a mean p-distance between them of 0.122.

5.4.2 A Comparison of Field Identifications with Phylogenetic Analysis

In Sub-sections 5.3.3 to 5.3.5, I presented three phylogenetic analyses, one for each family. These combined the selected reference sequences with the 94 sample sequences for 16 of the 18 putative species identified in the field. Overall, these were assigned to 13 species groups (Table 5.8), although eight samples remained as two unidentified taxa. Although the main objective of these analyses was to confirm or reassign species identification made in the field, they also revealed a number of important taxonomic characteristics of my samples.

Firstly, the five specimens of *Neodon sikimensis* that I collected clearly fell into two distinct branches (Figure 5.2). One branch had two specimens from Annapurna in the west and one from Langtang

in the centre of the country. The two specimens in the other branch came from Sagarmatha, at least 200km distant. The mean p-distance between these two branches was 0.120, which was significantly greater than the conspecific threshold derived from the reference sequences (Figure 5.7). Furthermore, neither branch showed a close link to any of the reference sequences, although the p-distance between the Langtang sample and the *N. leucurus* reference was 0.082. This suggests strongly, that I have two distinct, cryptic species that may not belong to any described species on GenBank. This merits further analysis of other mitochondrial or nuclear genes, but for now I will continue to refer to these specimens as *N. sikimensis*.

Three of the four sample sequences of *Apodemus gurkha* align very closely with the reference sequence (Figure 5.3). The one sequence that did not align well had a low Phred score, so was probably an artefact. This provides very strong evidence that the field identifications for this species were correct. However, given the importance of this species as the only endemic non-volant mammal species in Nepal, it merits further analysis of the 33 tissue samples that I collected from this species.

I caught 209 species of small *Mus* that I identified in the field as *M. booduga*. From these I collected 197 tissue samples. As the imperative was to sequence the rarer species, only four good sample sequences of this species were obtained, by chance all from Sagarmatha. There is no doubt from the phylogeny that these sequences belonged to *M. musculus* (Figure 5.3). I am confident that the specimens that I collected in the field belonged to the same species so, unless there were allopatric cryptic species in the other locations, I must assume that all specimens were *M. musculus*.

As I discussed in the previous sub-section, the *Niviventer* group showed considerable variation among the reference sequences, which was reflected to some degree by the sample sequences. I obtained 22 good quality sequences of *N. fulvescens*, 19 of which clustered very strongly with one of the three reference sequences (Figure 5.4), so I consider the field identifications for this species to be good. Eight of the nine sequences of the putative *N. eha* clustered well, but the average p-distance to the reference sequence was 0.103. This is strong evidence that another cryptic species may be present here. Furthermore, there is a clear geographic segregation between four sequences from Sagarmatha and five sequences from Annapurna; 400km apart.

The phylogenetic analysis of the remaining 12 murid sequences assigned five unknown samples and reassigned six putative *Rattus nitidus* into three species; *R.andamanensis*, *R. nitidus* and *R. pyctoris*. In addition, the one good sequence of *N. niviventer*, was assigned to *R. andamanensis*, so I have reassigned the other four specimens to this species.

I obtained 13 good sequences from the 20 specimens of *Episoriculus leucops*. These clustered extremely tightly together, into two sub-groups, one of five sequences from the Sagarmatha transect (plus one from Langtang) and seven from the Annapurna region (Figure 5.6). The mean p-distance between these two sub-groups of 0.038 suggested sub-specific status. However, this group had a mean p-distance to the *E. leucops* reference sequence of 0.175, which is greater even than the congeneric threshold. This reference sequence came from Syng Gompa in Nepal, which was about 10km from my Langtang transect. Even if the identification of the reference or sample sequencess was incorrect, one would expect the phylogenies to be closely aligned. The closest alignment was to the sister species *E. macrurus*, although the mean p-distance was still 0.096. In the absence of other evidence, I have retained my field identification for these specimens.

The four sample sequences of *Soriculus nigrescens* aligned closely together, although they also split into two geographical sub-groups (Annapurna and Sagarmatha) with a mean p-distance between them of 0.032. Both clusters aligned with the reference sequence from Syng Gompa, so the field identifications were confirmed.

Six of the twelve unidentified, high altitude soricids provided good quality sequences. Two of these aligned with *E. caudatus* and three with the *S. bedfordiae/cylindricauda* reference sequences. I have reassigned these specimens, but in the absence of confirming evidence I have retained the unidentified status of the other six. The four *E. caudatus* samples also showed ambiguous alignment. Given that this was the most frequently caught species (241 animals), which was clearly distinct from all others, I have retained the field identification of the 237 specimens that were not sequenced. Given the dominance of this species in the small mammal fauna (see Chapter 6), there is merit in further analysis of the 212 tissue samples that I collected from this species.

5.4.3 The Logistics of Conservation Genetics in Nepal

As explained in Chapter 2, all fieldwork had to be undertaken on foot as there was no road transport in any of the areas visited. This meant that tissue samples had to be carried at ambient temperature, sometimes for up to 20 days. The high altitude meant that ambient temperatures could fluctuate between –6° and +25° during a single transect. Tissue samples were stored in 100% ethanol which, when obtained locally, could not always be assured of highest quality. These problems were mitigated in the field by wrapping tube boxes in insulating material and utilising refrigeration whenever available. I also checked and refreshed ethanol at frequent intervals during the fieldwork.

Permission to undertake the collection of tissue samples was granted by the Ministry of Forests and Soil Conservation, on the condition that all DNA analysis was undertaken at CMDN in

Kathmandu. This is a commercial laboratory which, in 2013 at the start of my study, was in the early stages of development as a conservation genetics facility. Resources and expertise were limited, but I was not able to obtain exemption to bring tissue samples or extracted DNA back to UEA to utilise the vast range of expertise at the university and the wider Norwich Research Park. Instead, I collaborated closely with CMDN and, to ensure quality of DNA extractions and PCR, I supplied laboratory consumables, equipment and PCR primers from the UK.

In Supplementary Information 5.B I analysed whether these issues had any systematic effects on the final success and quality of DNA sequences. Tissue samples were grouped into batches for DNA extraction based on different combinations of field identity and location. Then extractions from different batches were selected for specific PCR runs so that these factors were not confounded. Overall success rates varied randomly across these factors, although there was an indication that two of the PCR runs had significantly lower success rates (Table 5.12). These corresponded with lower Phred scores which came from PCR runs using the full-length cytochrome-b gene, undertaken by one lab technician using CMDN primers. There was evidence that samples from one transect had lower Phred scores, but this was slightly confounded with species identification. Finally, the analysis of duplicate samples gave a strong indication that the DNA extractions were not the source of sequence variation, whilst variation due to the PCR runs occurred in 62% of cases. In summary, over a two year period these procedures generated 94 good DNA sequences that could be used in taxonomic analysis to confirm species identity.

5.5 References

Abdo , Z. and Golding, G.B. (2007) A Step Toward Barcoding Life: A Model-Based, Decision-Theoretic Method to Assign Genes to Preexisting Species Groups. *Systematic Biology* **56**: 44-56.

Adhikari, P., Han, S.-H., Kim, Y.-K., Kim, T.-W., Thapa, T.B., Subedi, N., Kunwar, A., Banjade, M., and Oh, H.-S. (2018a) New record of the Oriental house rat, *Rattus tanezumi*, in Nepal inferred from mitochondrial Cytochrome B gene sequences. *Mitochondrial DNA Part B* **3**: 386-390.

Adhikari, P., Han, S.H., Kim, Y.K., Kim, T.W., Thapa, T.B., Subedi, N., Adhikari, P., and Oh, H.S. (2018b) First molecular evidence of *Mus musculus bactrianus* in Nepal inferred from the mitochondrial DNA cytochrome B gene sequences. *Mitochondrial DNA A DNA Mapp Seq Anal* **29**: 561-566.

Anon. (2006) *DNeasy Blood & Tissue Handbook*. Qiagen.

Anon. (2017) Statistica (data analysis software system) v.13. TIBCO Software Inc. http://statistica.io.

Anon. (2018) Chromas 2.6.5. Technelysium Pty Ltd.

Baral, H.S. and Shah, K.B. (2008) *Wild Mammals of Nepal*. Himalayan Nature, Kathmandu. pp 188.

Batovska, J., Blacket, M.J., Brown, K., and Lynch, S.E. (2016) Molecular identification of mosquitoes (Diptera: Culicidae) in southeastern Australia. *Ecol Evol* **6**: 3001-3011.

Bradley, R.D. and Baker, R.J. (2001) A test of the genetic species concept: cytochrome-b sequences and mammals. *Journal of Mammalogy* **82**: 960-963.

Chen, W., Li, Y., Liu, Y., Liu, S., and Yue, B. (2011) Complex topographic configuration in the Hengduan Mountains shaped the phylogeographic structure of Chinese white-bellied rats. *Journal of Zoology* **284**: 215-223.

Collins, R.A. and Cruickshank, R.H. (2013) The seven deadly sins of DNA barcoding. *Mol Ecol Resour* **13**: 969-975.

Corbet, G.B. and Hill, J.E. (1992) *The Mammals of the Indomalayan Region: A systematic Review*. Oxford University Press. pp 488.

Cristescu, M.E. (2014) From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. *Trends in Ecology & Evolution* **29**: 566-571.

Dubey, S., Salamin, N., Ohdachi, S.D., Barriere, P., and Vogel, P. (2007) Molecular phylogenetics of shrews (Mammalia: Soricidae) reveal timing of transcontinental colonizations. *Mol Phylogenet Evol* **44**: 126-137.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accurace and high throughput. *Nucleic Acids Research* **32**: 1792-1797.

Felsenstein, J. (1985) Confidence Limits on Phylogenies: an Approach Using the Bootstrap. *Evolution* **39**: 783-791.

Ge, D., Lu, L., Cheng, J., Xia, L., Chang, Y., Wen, Z., Lv, X., Du, Y., Liu, Q., and Yang, Q. (2017) An endemic rat species complex is evidence of moderate environmental changes in the terrestrial biodiversity centre of China through the late Quaternary. *Sci Rep* **7**: 1-14.

Gundi, V.A., Kosoy, M.Y., Myint, K.S., Shrestha, S.K., Shrestha, M.P., Pavlin, J.A., and Gibbons, R.V. (2010) Prevalence and genetic diversity of bartonella species detected in different tissues of small mammals in Nepal. *Applied and Environmental Microbiology* **76**: 8247-8254.

Hebert, P.D., Cywinska, A., Ball, S.L., and deWaard, J.R. (2003) Biological identifications through DNA barcodes. *Proc Biol Sci* **270**: 313-321.

Jing, M., Yu, H.T., Wu, S.H., Wang, W., and Zheng, X. (2007) Phylogenetic relationships in genus *Niviventer* (Rodentia: Muridae) in China inferred from complete mitochondrial cytochrome b gene. *Mol Phylogenet Evol* **44**: 521-529.

Karmacharya, D.B., Thapa, K., Shrestha, R., Dhakal, M., and Janecka, J.E. (2011) Noninvasive genetic population survey of snow leopards (*Panthera uncia*) in Kangchenjunga conservation area, Shey Phoksundo National Park and surrounding buffer zones of Nepal. *BMC Res Notes* **4**: 516.

Kress, W.J. and Erickson, D.L. (2008) DNA barcodes: genes, genomics, and bioinformatics. *Proc Natl Acad Sci U S A* **105**: 2761-2762.

Kroken, S. and Taylor, I.R. (2001) A gene genealogical approach to recognize phylogenetic species boundaries in the lichenized fungus *Letharia*. *Mycologia* **93**: 38-53.

Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33**: 1870-1874.

Lack, J.B., Greene, D.U., Conroy, C.J., Hamilton, M.J., Braun, J.K., Mares, M.A., and Van Den Bussche, R.A. (2012) Invasion facilitates hybridization with introgression in the Rattus rattus species complex. *Molecular Ecology* **21**: 3545-3561.

Lissovsky, A.A., Petrova, T.V., Yatsentyuk, S.P., Golenishchev, F.N., Putincev, N.I., Kartavtseva, I.V., Sheremetyeva, I.N., and Abramson, N.I. (2018) Multilocus phylogeny and taxonomy of East Asian voles Alexandromys (Rodentia, Arvicolinae). *Zoologica Scripta* **47**: 9-20.

Liu, S.Y., He, K., Chen, S.D., Jin, W., Murphy, R.W., Tang, M.K., Liao, R., and Li, F.J. (2018) How many species of *Apodemus* and *Rattus* occur in China? A survey based on mitochondrial cyt b and morphological analyses. *Zoological Research* **39**: 309-320.

Liu, S.Y., Sun, Z.Y., Liu, Y., Wang, H., Guo, P., and Murphy, R.W. (2012) A new vole from Xizang, China and the molecular phylogeny of the genus *Neodon* (Cricetidae: Arvicolinae). *Zootaxa* **3235**: 1-12.

Matubara, K., Ishibashi, Y., Ohdachi, O., and Matsuda, Y. (2001) A new primer set for sex identification in the genus *Sorex* (Soricidae, Insectivora). *Molecular Ecology Notes* **1**: 241-242.

Meyer, C.P. and Paulay, G. (2005) DNA Barcoding: Error Rates Based on Comprehensive Sampling. *PLoS Biology* **3**: 2229-2238.

Molur, S., Srinivasulu, C., Srinivasulu, B., Walker, S., Nameer, P., and Ravikumar, L. (2005) *Status of South Asian Non-volant Small Mammals: Conservation Assessment and Management Plan (C.A.M.P.) Workshop Report.* Zoo Outreach Organisation / CBSG-South Asia, Coimbatore, India, . pp 618.

Motokawa, M., Harada, M., Mekada, K., and Shrestha, C. (2008) Karyotypes of three shrew species (*Soriculus nigrescens*, *Episoriculus caudatus* and *Episoriculus sacratus*) from Nepal. *Integrative Zoology* **3**: 180-185.

Naidu, A., Fitak, R.R., Munguia-Vega, A., and Culver, M. (2012) Novel primers for complete mitochondrial cytochrome b gene sequencing in mammals. *Mol Ecol Resour* **12**: 191-196.

Nei, M. and Kumar, S. (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press, New York. pp 352.

Ohdachi, S.D., Iwasa, M.A., Nesterenko, V.A., Abe, H., Masuda, R., and Haberl, W. (2004) Molecular phylogenetics of *Crocidura* shrews (Insectivora) in east and central Asia. *Journal of Mammalogy* **85**: 396-403.

Parson, W., Pegoraro, K., Neiederstatter, H., and Foger, M. (2000) Species Identification by means of the cytochrome b gene. *International Journal of Legal Medicine* **114**: 23-28.

Pearch, M.J. (2011) A review of the biological diversity and distribution of small mammal taxa in the terrestrial ecoregions and protected areas of Nepal. *Zootaxa* **3072** 1-286.

Ratnasingham, S. and Hebert, P.D. (2007) BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes* **7**: 355-364.

Robins, J.H., Hingston, M., Matisoo-Smith, E., and Ross, H.A. (2007) Identifying *Rattus* species using mitochondrial DNA. *Molecular Ecology Notes* **7**: 717-729.

Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**: 406-425.

Shimada, T., Aplin, K.P., and Suzuki, H. (2010) *Mus lepidoides* (Muridae, Rodentia) of central Burma is a distinct species of potentially great evolutionary and biogeographic significance. *Zoolog Sci* **27**: 449-459.

A.T. Smith and Xie, Y. *Eds*. (2008) *A Guide to the Mammals of China*. pp 544. Princeton University Press.

Suzuki, H., Numone, M., Kinoshita, G., Aplin, K., Vogel, P., Kryukov, A.P., Jin, M.-L., Han, S.-H., Maryanto, I., Tsuchiya, K., Ikeda, H., Shiroishi, T., Yonekawa, H., and Moriwaki, K. (2013) Evolutionary and dispersal history of Eurasian house mice *Mus musculus* clarified by more extensive geographic sampling of mitochondrial DNA. *Heredity* **111**: 375-390.

Suzuki, H., Sato, J., Tsuchiya, K., Luo, J., Zhang, Y.-P., Wang, Y.-X., and Jiang, X.-E. (2003) Molecular phylogeny of wood mice (*Apodemus*, Muridae) in East Asia. *Biological Journal of the Linnean Society* **80**: 469-481.

Suzuki, H., Shimada, T., Terashima, M., Tsuchiya, K., and Aplin, K. (2004) Temporal, spatial, and ecological modes of evolution of Eurasian *Mus* based on mitochondrial and nuclear gene sequences. *Mol Phylogenet Evol* **33**: 626-646.

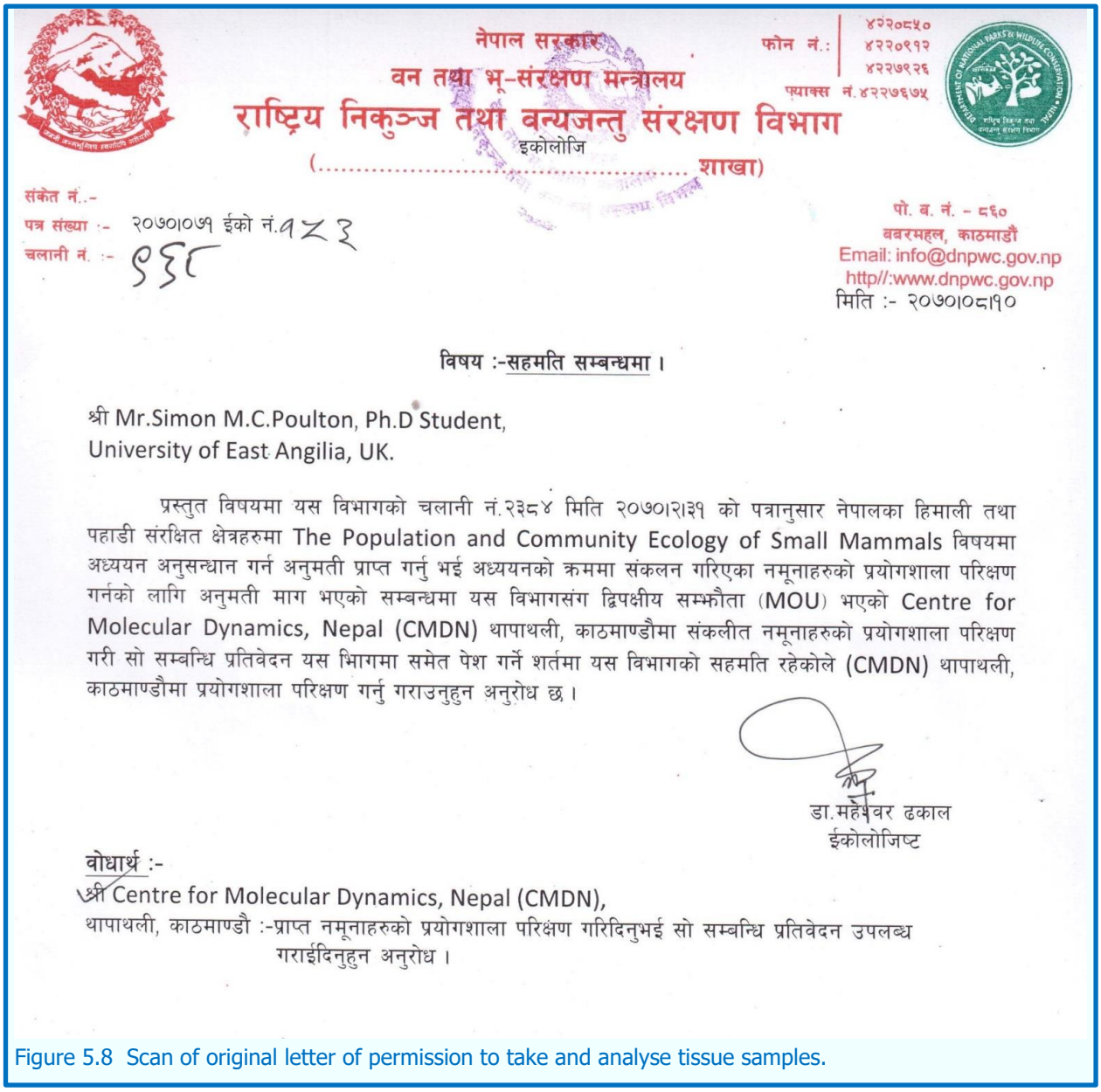
Thapa, K., Manandhar, S., Bista, M., Shakya, J., Sah, G., Dhakal, M., Sharma, N., Llewellyn, B., Wultsch, C., Waits, L.P., Kelly, M.J., Hero, J.M., Hughes, J., and Karmacharya, D. (2018) Assessment of genetic diversity, population structure, and gene flow of tigers (*Panthera tigris tigris*) across Nepal's Terai Arc Landscape. *PLoS One* **13**: e0193495.

Werhahn, G., Senn, H., Kaden, J., Joshi, J., Bhattarai, S., Kusi, N., Sillero-Zubiri, C., and Macdonald, D.W. (2017) Phylogenetic evidence for the ancient Himalayan wolf: towards a clarification of its taxonomic status based on genetic sampling from western Nepal. *Royal Society Open Science* **4**: 170186.

Zhang, B., He, K., Wan, T., Chen, P., Sun, G., Liu, S., Nguyen, T.S., Lin, L., and Jiang, X. (2016) Multi-locus phylogeny using topotype specimens sheds light on the systematics of *Niviventer* (Rodentia, Muridae) in China. *BMC Evolutionary Biology* **16**: 261.

5.6 Appendices

5.6.1 Appendix I. Letters of Permission to Undertake Tissue Sample Collection and Laboratory Analysis



नेपाल सरकार
वन तथा भू-संरक्षण मन्त्रालय
राष्ट्रिय निकुञ्ज तथा वन्यजन्तु संरक्षण विभाग
इकोलोजि
(..... शाखा)

फोन नं.: ४२२०८५०, ४२२०९१२, ४२२७९२६
फ्याक्स नं. ४२२७६७५

संकेत नं..-
पत्र संख्या :- २०७०।०७१ ईको नं.१५३
चलानी नं. :- ९६८

पो. ब. नं. - ८६०
बबरमहल, काठमाडौं
Email: info@dnpwc.gov.np
http//:www.dnpwc.gov.np
मिति :- २०७०।०८।१०

विषय :-सहमति सम्बन्धमा ।

श्री Mr.Simon M.C.Poulton, Ph.D Student,
University of East Angilia, UK.

प्रस्तुत विषयमा यस विभागको चलानी नं.२३८४ मिति २०७०।२।३१ को पत्रानुसार नेपालका हिमाली तथा पहाडी संरक्षित क्षेत्रहरुमा The Population and Community Ecology of Small Mammals विषयमा अध्ययन अनुसन्धान गर्न अनुमती प्राप्त गर्नु भई अध्ययनको क्रममा संकलन गरिएका नमूनाहरुको प्रयोगशाला परिक्षण गर्नको लागि अनुमती माग भएको सम्बन्धमा यस विभागसंग द्विपक्षीय सम्झौता (MOU) भएको Centre for Molecular Dynamics, Nepal (CMDN) थापाथली, काठमाण्डौमा संकलीत नमूनाहरुको प्रयोगशाला परिक्षण गरी सो सम्बन्धि प्रतिवेदन यस भागमा समेत पेश गर्ने शर्तमा यस विभागको सहमति रहेकोले (CMDN) थापाथली, काठमाण्डौमा प्रयोगशाला परिक्षण गर्नु गराउनुहुन अनुरोध छ ।

डा.महेश्वर ढकाल
ईकोलोजिष्ट

बोधार्थ :-
श्री Centre for Molecular Dynamics, Nepal (CMDN),
थापाथली, काठमाण्डौ :-प्राप्त नमूनाहरुको प्रयोगशाला परिक्षण गरिदिनुभई सो सम्बन्धि प्रतिवेदन उपलब्ध गराईदिनुहुन अनुरोध ।

Figure 5.8 Scan of original letter of permission to take and analyse tissue samples.

Ministry of forest and Soil Conservation
Department of National Parks and Wildlife Conservation

Letter no. 2070/071 eco. No. 153
Ref. No. 968

Date: 2070/8/10

Sub: Agreement paper

Mr. Simon M.C Poulton , Ph.D Student
University of East Anglia, UK

This permission is issued in reference to the requested submitted for laboratory analysis of samples collected during the study "The Population and community Ecology of Small Mammals" in high altitude protected areas of Nepal, reference letter No. 2384 on 2070/2/31. All the laboratory analysis on samples should be performed at the recommended laboratory having bilateral agreement (MoU) with the Department, The Center for Molecular Dynamics Nepal (CMDN), Thapathali, Kathmandu and documents regarding the analysis should be shared with the DNPWC.

Dr. Maheshwor Dhakal
Ecologist

CC:
Center for Molecular Dynamics Nepal
Thapathali, Kathmandu:- Requested to provide required documents regarding and perform the laboratory analysis on received samples.

Figure 5.9 Translation of letter of permission to take and analyse tissue samples.

SUPPLEMENTARY INFORMATION

5.A The comparison of short and full-length primers

Before primers were selected for the final PCR stage at CMDN, I analysed the ability of the short PREDICT primers, compared to full-length cytochrome-b primers, to distinguish species. 44 full-length cytochrome-b sequences, belonging to 12 species in seven genera of Muridae and Soricidae, were down-loaded from GenBank. These were aligned in Mega 7 and a distance matrix based on p-distances was calculated. Primer-Blast was then used to extract the 479bp sequence (inclusive) defined by the PREDICT primers, and the exercise repeated. This gave two distance matrices, with identical structure, from which were extracted four classes of p-distances; con-specific, con-generic, con-familial and cross-family. These allowed a pair-wise comparison of the 946 p-distances between the two sequence lengths (Figure 5.10). The overall R^2 was 95.3%, with the con-specific r2 =98.2%. This graph also shows that there was a distinct threshold of approx. 10% which clearly demarcated the difference between con-specific and con-generic p-values.

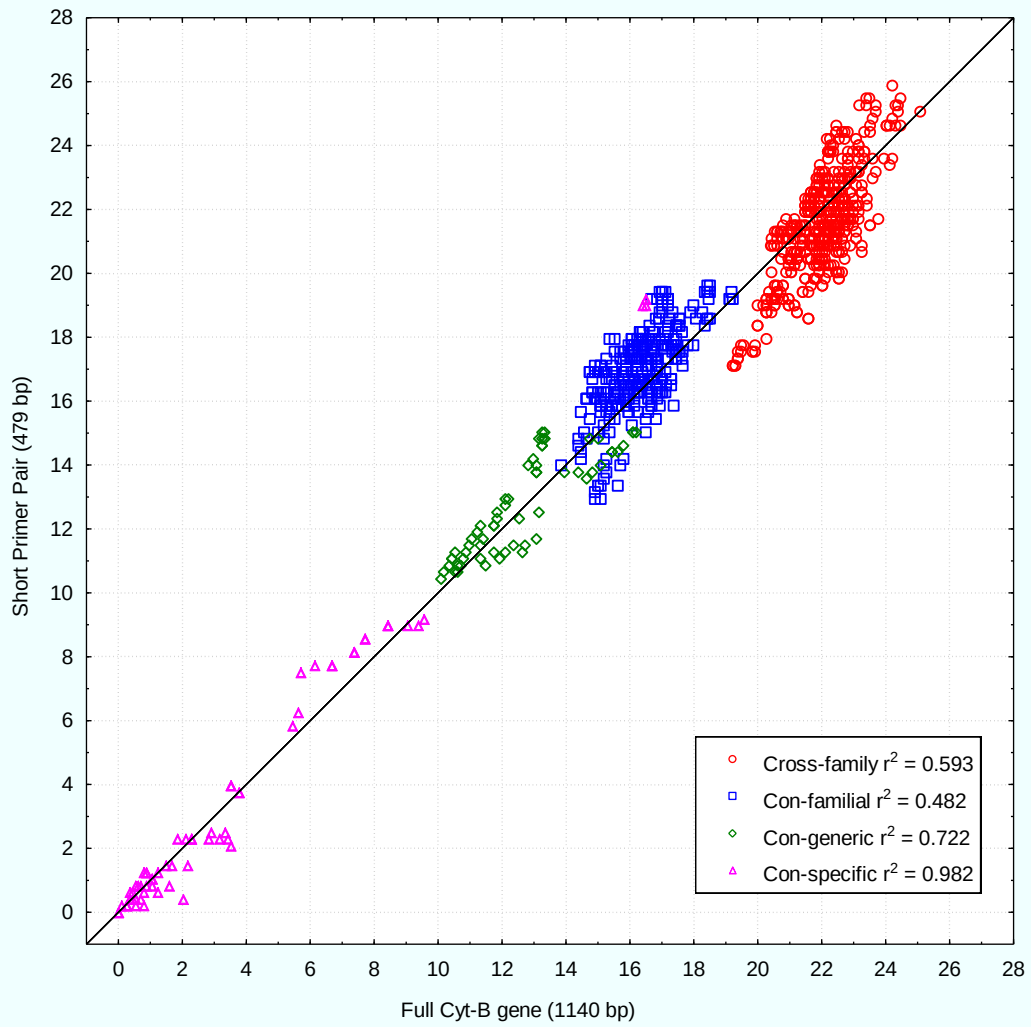


Figure 5.10 Pair-wise comparison of p-distances (expressed as percentages) from short and full-length cytochrome-b sequences, colour-coded by category of comparison. The black diagonal line represents unity. Note the three con-specific comparisons near the con-familial cluster, which is discussed further in the text.

To explore these effects further, a crossed and repeated-measure ANOVA model was constructed with species and taxonomic separation (plus their interaction) as between-subject factors, and sequence length as the within-subject factor. All factors except the main effect of sequence length were very highly significant (Table 5.9). This indicated that there was no difference overall in the p-distances between the full-length and the short sequences. The most important effect was the three-way interaction (Figure 5.11) which confirmed that the short-length primers were able to give a very clear distinction between con-specific and con-generic p-distances. However, note that the two species of Niviventer and three species of shrews had considerably higher least-squares means than the other species. This effect is explored in more detail in Supplementary Information 5.C: Analysis of Cytochrome-b Sequences Obtained From GenBank.

Table 5.9 Repeated-measures ANOVA of p-distances, categorised by species, degree of taxonomic separation and sequence length.

Term	SS	DF	MS	F	*p*
Intercept	242973.5	1	242973.5	76168.48	0.000
Species	1655.1	11	150.5	47.17	0.000
Taxonomic separation	59949.9	3	19983.3	6264.46	0.000
Species × Taxonomy	2790.3	31	90.0	28.22	0.000
Between-subject Error	5888.6	1846	3.2		
Sequence length	0.9	1	0.9	2.79	0.095
Length × Species	67.3	11	6.1	18.03	0.000
Length × Taxonomy	207.5	3	69.2	203.98	0.000
Length × Species × Taxonomy	179.6	31	5.8	17.08	0.000
Within-subject Error	626.1	1846	0.3		

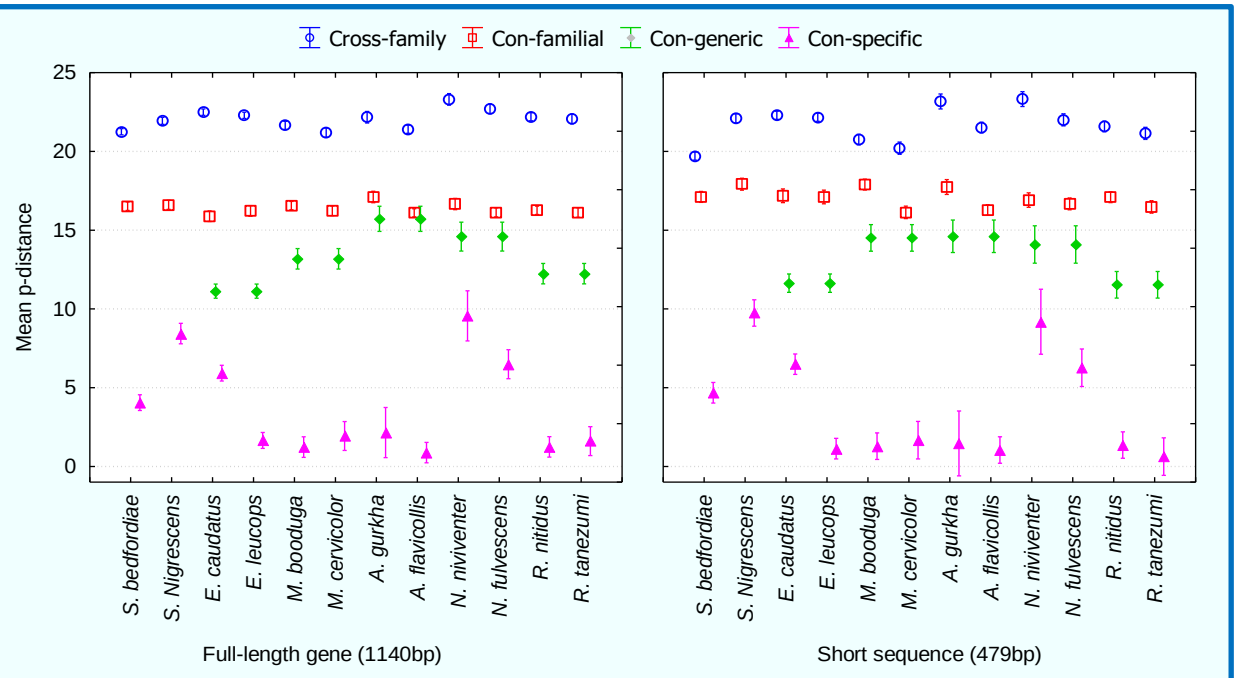


Figure 5.11 The three-way interaction of the repeated-measures ANOVA of p-distances, categorised by species, degree of taxonomic separation and sequence length. Least-squares means and 95% C. I. are shown.

5.B Analysis of the Process of Tissue Sample Collection, DNA Extraction, PCR and Sequencing

Permission to collect tissue samples and undertake their analysis was granted by the Department of National Parks and Wildlife Conservation (DNPWC) through their MOU with CMDN (see 5.6.1 Appendix I. Letters of Permission to Undertake Tissue Sample Collection and Laboratory Analysis).

The majority of the DNA-based work was carried out by the Centre for Molecular Dynamics Nepal (CMDN) in Kathmandu. The main collaborator for this project was the Managing Director, Dibesh Karmacharya (DK). Three members of staff were directly involved in laboratory procedures; Ajay Sharma (AjS), Adarsh Serchan (AdS) and Hemanta Choudhary (HC). The PCR products were sent to UEA for sequencing and a small amount of additional PCR was carried out by Michael Strinden (MS).

5.B.a Tissue Samples

From the total of 792 animals caught (see Chapter 2), tissue samples were taken from 720 animals using a 1.5mm ear punch (Harvard Apparatus Ltd, UK). The majority of tissue samples were collected in 2014 and 2015, whereas in 2013 tissue samples were collected from 71% of the animals caught (Table 5.10a). Samples were obtained for all putative taxa and almost every specimen of the rarer species contributed a tissue sample. Most of the specimens not providing tissue samples belonged to *Episoriculus caudatus* and *Soriculus nigrescens*, although eight specimens of the putative *Apodemus gurkha* did not provide tissue samples (Table 5.10b).

Table 5.10 Total number of specimens and number of tissue samples broken down by a) year and transect and b) field identification.

a) Year	Transect	Animals	Tissue Samples	
2013	1	108	68	63%
	2	120	94	78%
2014	1	245	241	98%
	3	118	118	100%
2015	4	201	199	99%
Total		792	720	91%

b) Family	Taxon	Animals	Tissue Samples	
Ochotonidae	*Ochotona tibetanus*	1	1	100%
Cricetidae	*Neodon sikimensis*	5	5	100%
Muridae	*Murid sp1.*	3	3	100%
	Murid sp2	2	2	100%
	Apodemus gurkha	41	33	80%
	Mus booduga	209	197	94%
	Niviventer eha	10	10	100%
	Niviventer fulvescens	26	25	96%
	Niviventer niviventer	5	3	60%
	Rattus nitidus	6	7	117%
	Rattus sp.	2	2	100%
Soricidae	*Soricid sp1.*	11	11	100%
	Soricid sp2.	1	1	100%
	Episoriculus caudatus	240	212	88%
	Episoriculus leucops	20	18	90%
	Sorex bedfordiae	1	1	100%
	Soriculus nigrescens	187	168	90%
	Suncus murinus	21	20	95%

5.B.b DNA Extractions

Almost all tissue samples were large enough to be divided into two sub-sets, to provide backup samples in the event of DNA extraction failure. This generated 1405 sub-samples, which included an additional division of the samples used in the first batch of DNA extractions. From these, 509 DNA extractions were completed from one sub-set. Eight extractions were duplicates of extractions made in batches 1 and 2, meaning that only 501 unique tissue samples were actually extracted. Thirty extraction batches, including 436 individual extractions, had NanoDrop readings taken. Of these, 387 (89%) had DNA concentrations >= 5ng/µl. Twenty-five batches had mean concentrations above 10ng/µl, with the highest mean concentration being 165ng/µl. Twenty batches had mean A260:A280 ratios greater than the generally accepted minimum value of 1.8, and a further two batches had ratios of 1.79. There were no obvious patterns in the mean concentrations by year/transect nor by taxon. The first two batches of DNA extractions were carried out by AjS and AdS respectively. I undertook all subsequent extractions.

5.B.c PCR

All PCR was carried out by the staff of CMDN. The first run of 37 samples, carried out by AjS, for the CO1 gene yielded poor results and was not utilised further. Four samples from this batch of DNA extractions were re-run in November 2014, also by AjS, for a short sequence of the cytochrome-b gene. These all yielded good results and sequenced successfully (Table 5.11). Six extraction

Table 5.11 Success rates for PCR runs at CMDN over four periods.

Period	PCR Run	Total	Failed	Faint	Strong	Success Rate
Nov 2014	2	4			4	100%
Feb 2017	3	4		2	2	100%
	4	26	16	9	1	38%
	5	15	1	10	4	93%
	6	8	8			0%
	7	29	25	3	1	14%
	8	31	29	2		6%
	Total	113	79	26	8	30%
Aug 2017 (Optimisation)	9	6			6	100%
	10	10		2	8	100%
	11	6	5		1	17%
	12	6	1	1	4	83%
	Total	28	6	3	19	79%
Sep 2017	13	20	20			0%
	14	6	1	1	4	83%
	15	10			10	100%
	16	25		3	22	100%
	17	22	9		13	59%
	18	22	8	5	9	64%
	Total	105	38	9	58	64%

batches, totalling 113 reactions, included eight NECs, were run in February 2017 by AdS for short length cytochrome-b, using CMDN primers. In addition an NTC for each run and seven PTCs were included. Success rates from these six batches, as revealed by electrophoretic gels, was very variable. Firstly, all eight NECs showed no bands, indicating that the DNA extractions were not contaminated. The overall success rate from the 113 specimen reactions was only 30% and of the 34 reactions considered successful, only 8 showed a strong band in the gel plate. One batch of eight reactions failed entirely, including the NTC and PTC. The first run (#3) was actually very successful with no band for the NTC, a strong band for the PTC no bands for the four NECs and two each of faint and strong bands for the four samples (Figure 5.13 in Appendix 5.B.h). However, even run 5, which had 93% success rate had a majority of faint bands. The other runs had very poor success rates, including no band for the two PTCs in run 7, which throws doubt on the four positive bands in this run (Appendix 5.B.h).

Following the four optimisation runs carried out in August 2017 by HC, the seven runs in September 2017 were generally more successful (64% of 105 reactions) and the proportion of strong bands was much higher in this period. However, run 13 was still a total failure, including one NTC and two PTCs (Figure 5.18 in Appendix 5.B.h). All other runs had NTCs with no bands and PTCs with strong bands. The difference between the success rates in the two main periods was very highly significant ($\chi^2_{(1)} = 27.34$. $p < 10^{-6}$).

Success rates for PCRs were not significantly different between putative taxa ($\chi^2_{(16)} = 20.34$. $p > 0.20$). Even when amalgamated into the three common families there was no significant difference ($\chi^2_{(2)} = 4.08$. $p > 0.10$). However, there was a highly significant difference in the success rates of PCRs from different Year/Transect combinations ($\chi^2_{(4)} = 20.33$. $p < 0.0005$). The contingency table showed that this was entirely due to the low success rate in samples from Transect 1 in 2014 (37%) and the very high success rate from Transect 4 in 2015 (82%). A two-way ANOVA with binary response variable and logit link function confirmed that both this effect and the PCR period were independently highly significant (Figure 5.12).

PCR products from 71 failed reactions carried out by CMDN in Feb 2017 were rerun at UEA. Of these, 21 achieved Qubit measurements of >=10.

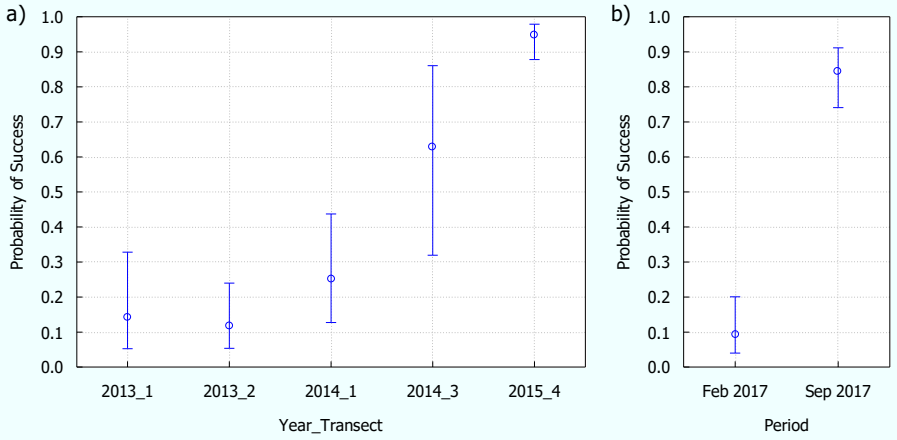


Figure 5.12 Means & 95% C. I. from two-way logistic ANOVA of PCR success by Year_Transect and laboratory analysis period.

5.B.d Sequences

A total of 126 PCR products were sequenced; four full-length cytochrome-b PCR products from November 2014, 34 from the February 2017 batch comprising full-length sequences, 67 from the September 2017 batch comprising short sequences and 21 from the reruns at UEA also based on the PREDICT short sequences.

All sequences except the first four had trace files which included quality scores for each nucleotide. The trimming algorithm worked well in the majority of cases, especially for the PREDICT primer sequences where it usually removed between 10 and 15 nucleotides from the start of the sequence, and usually only one or two trailing nucleotides (e.g. Figure 5.24 in Appendix II. Example Sequence Traces). Eleven sequences had such poor traces that they were rejected entirely (e.g. Figure 5.25); six from the full-length reactions and five from the PREDICT reactions. This left 115 sequences of sufficient quality for subsequent analysis.

The overall quality of the sequences was assessed in two ways; a) the proportion of sequences that were classified as good and b) the mean Phred score. The predictor variables of interest fell into two broad categories;

- Field variables; Taxonomic family, Year & Transect
- Lab variables; Extraction batch, Extractor, PCR run, PCR personnel, Primer length

Due to the unbalanced, nested nature of these predictors and the binary or severely skewed nature of the response variables, it was not possible to run generalized linear models. Instead, for the binary success rates, separate 2 × n cross-tabulations were analysed by contingency χ^2, and for the Phred scores, separate Kruskal-Wallis non-parametric ANOVAs were run. There was little significant variation in the success rates across the predictor variables (Table 5.12a), although the success rate was

significantly lower in PCR Runs 7 and 8 than others. In contrast, there was significant variation in Phred scores. In particular, there were lower scores for sequences obtained from animals in Year 2015 / Transect 5 (year and transect were confounded). Both the extractions and PCR runs showed significant variation, but most importantly, the personnel who carried out the PCR and the primer length (which were confounded) were very highly significant. AdS had a significantly lower median Phred score – approximately 41, compared to approx. 50 for the other two people running PCRs. However, he only ran full length primers, which also had significantly lower Phred scores.

Table 5.12 Analysis of a) binary success rates and b) Phred scores for eight predictor variables. p-values < 0.01 are marked in red.

	a) Binary Success			b) Phred Score		
Predictor	χ^2	DF	*p*	H	DF	*p*
Family	1.22	2	0.542	1.38	2	0.503
Year	1.20	2	0.548	12.15	2	0.002
Transect	2.27	3	0.518	11.48	3	0.009
Extraction Batch	17.40	11	0.096	29.89	11	0.002
Extractor	4.07	2	0.131	7.63	2	0.022
PCR Run	25.26	11	0.008	24.49	10	0.006
PCR Personnel	2.90	3	0.408	16.10	2	<0.001
Primer Length	4.65	1	0.031	15.89	1	<0.001

5.B.e Analysis of Multiple Sequences per Animal

A small degree of duplication was included at each stage of DNA analysis. Seven animals had DNA extractions from different tissue sample series. A number of the 222 PCRs run in CMDN were repeat reactions due to initial failure, although a small number of successes were also re-run. 85 animals had two or more PCRs, of which 6 had two successful reactions. In addition, the rerun of 71 failed reactions at UEA, yielded a further 21 success of which 15 already had one other successful sequence. In total 21 animals had two or three successful PCR runs.

This duplication allowed an initial appraisal of the quality of the PCR and, to some extent, the DNA extractions. The 44 sequences (19 pairs and 2 triples) were imported into the Sequence Alignment module of Mega 7 and aligned separately within their animals. Pair-wise distances were then calculated within animals, using all nucleotides with pair-wise deletion of gaps and ambiguous sites.

A total of 25 pair-wise distances was calculated, belonging to 21 animals from ten field identifications (Table 5.13). The results were highly variable and require detailed interpretation. Firstly, most animals had two PCR runs from the same DNA extraction. However, one animal (Soricid sp. 1; 14-3-1-2-01) had two different extractions that were then included in the same PCR run. The other unidentified soricid had three different extractions (from two batches), all amplified in the same

Table 5.13 Pairwise distances for 21 animals with duplicate or triplicate sequences, derived either from multiple DNA extractions and/or multiple PCR runs. Columns 1 v 2 to 2 v 3 show number of variable sites, based on approximately 435 sites per comparison.

Family	Field Identity	Animal Code	DNA Extraction	PCR Run	1 v 2	1 v 3	2 v 3
Cricetidae	*Neodon sikkimensis*	14-1-6-5-05	07/14	17 20	13		
		14-3-1-6-02	07/03	16 20	107		
Muridae	Murid sp. 1	14-1-1-6-06	04/12	16 20	102		
	Apodemus gurkha	13-1-4-4-05	10/03	18 20	4		
	Niviventer fulvescens	14-1-5-1-10	08/10	17 20	3		
		13-2-4-1-04	01/29 08/14	16 17 20	4	1	4
		13-2-3-3-06	11/04	18 20	5		
		14-1-2-1-06	08/07	18 20	121		
		14-1-2-2-09	11/06	18 20	119		
	Niviventer niviventer	14-1-5-6-07	07/08	17 20	100		
	Rattus nitidus	13-1-2-1-06	01/05	2 14	210		
		14-1-2-1-11	04/13	16 20	3		
		14-1-5-4-07	08/12	17 20	2		
Soricidae	Soricid sp. 1	14-3-1-2-01	02/11 02/24	16	77		
		14-3-2-5-01	02/15 02/22 06/01	16	2	0	0
	Episoriculus leucops	14-1-4-2-04	08/06	17 20	83		
		15-4-1-5-02	09/03	17 20	107		
		15-4-1-6-02	10/07	18 20	109		
	Soriculus nigrescens	13-2-1-6-02	01/21	2 15	223		
		13-2-4-3-06	01/37	2 16	203		

PCR run. Finally, the *N. fulvescens* (13-2-4-1-04) had two DNA extractions amplified in three different PCR runs.

Eight animals had very low p-distances, six of which had less than 1% variation in sites, based on approximately 435 sites. The two animals which had triplicate sequences had all six pair-wise distances less than 1%. In contrast, four animals had extremely high p-distances (> 55% based on less than 390 sites). These were the only four samples that comprised the first cytochrome-b PCR run (#2) at CMDN in 2014. Careful inspection of the sequences showed that they were not the same part of the cytochrome-b gene as the other runs, which used the PREDICT primers, so should not have been compared in this way. The remaining nine pair-wise comparisons had poor p-values, mostly around 25%. All of these animals except one (Soricid sp. 1; 14-3-1-2-01) included one sequence from PCR run 20, which was the rerun of PCR products from the original failed runs at CMDN. However, it should be pointed out that eight good pair-wise comparisons also included sequences from PCR run 20. These are discussed further in the next section.

5.B.f Analysis of the sequences from the UEA Re-run of failed PCR products

PCRs were carried out in UEA on the 71 failed PCR products that were sent from CMDN. Of these, 21 were of sufficient quality to be sequenced. One sequence failed entirely, but the other 20 were generally good quality; the median Phred score was 49, with a minimum and maximum of 29.5 and 53.4 respectively. 15 of these sequences have already be compared to other sequences that were derived from the same animal in the previous section, but there were five more that were not matched in this way.

These 20 sequences were run through NCBI using the BlastN algorithm. Of these, 11 were matched to human sequences, all with E-values of $<10^{-178}$ (Table 5.14a). The lowest identity match was 97.1% and seven had identity matches of >99%. In contrast, nine specimens had Blast identities that were highly probable (Table 5.14b), although one was clearly a sample handling error. The final outcome of the reruns was two more useful sequences and six confirmations of identities from original duplicates.

Table 5.14 Reruns of 20 failed PCRs from CMDN. The "Re-runs" block shows the Phred scores from the sequences of the re-run PCRs and the outcome of Blast. The table is divided into two blocks; a) the 11 samples where the Blast result was *Homo sapiens* and b) the nine samples which had good Blast results. The sample in red indicates a good blast result which was considered to be an error in sample handling. The right hand columns indicate the PCR runs from the original duplications, with their 15 Blast results that correspond to the 15 duplicates with PCR run 20 shown in Table 5.13 and discussed in the text.

Original Failed PCR Runs				Re-runs				Duplications	
Family	Field Identity	Extraction Tube	PCR Run	Phred Score	Matched Length	Identities	Blast Identity	PCR Run	Blast Identity
a)									
Cricetidae	*Neodon sikimensis*	07/03	7	29.59	440	97.05%	*H. sapiens*	16	*N. leucurus*
	Apodemus gurkha	07/08	7	44.56	438	98.86%	*H. sapiens*	17	*R. andamanensis*
	Apodemus gurkha	10/03	8	48.52	439	99.54%	*H. sapiens*	18	*H. sapiens*
	Apodemus gurkha	12/02	8	53.27	436	99.54%	*H. sapiens*		
Muridae	*Apodemus gurkha*	12/04	8	52.57	440	99.32%	*H. sapiens*	18	Bad sequence
	Mus booduga	04/15	4	46.90	437	97.94%	*H. sapiens*		
	Niviventer eha	04/12	4	44.30	436	99.08%	*H. sapiens*	16	*R. andamanensis*
	Niviventer niviventer	09/07	8	49.63	439	99.54%	*H. sapiens*	18	*N. fulvescens*
	Niviventer niviventer	11/06	8	48.22	437	99.54%	*H. sapiens*	18	*Niviventer sp.*
Soricidae	*Episoriculus leucops*	09/03	8	42.39	437	98.86%	*H. sapiens*	17	*E. macrurus*
	Episoriculus leucops	10/07	8	52.81	437	99.54%	*H. sapiens*	18	*E. macrurus*
b)									
Cricetidae	*Neodon sikimensis*	07/14	7	52.45	436	97.94%	*N. irene*	17	*N. Irene*
	Apodemus gurkha	08/10	7	46.02	437	97.25%	*Niviventer sp.*	17	*Niviventer sp.*
	Apodemus sp.	07/15	7	49.31	436	98.85%	*A. gurkha*		
	Apodemus sp.	08/14	7	50.57	423	94.80%	*Niviventer sp.*	17	*Niviventer sp.*
Muridae	*Niviventer niviventer*	11/04	8	39.25	439	97.95%	*Niviventer sp.*	18	*Niviventer sp.*
	Rattus nitidus	04/13	4	53.50	430	96.74%	*R. pyctoris*	16	*R. pyctoris*
	Rattus nitidus	06/12	7	51.91	439	99.09%	*R. nitidus*		
	Rattus nitidus	08/12	7	52.14	439	98.63%	*R. andamanensis*	17	*R. andamanensis*
Soricidae	*Episoriculus leucops*	08/06	7	52.23	438	98.63%	*R. andamanensis*	17	*E. macrurus*

5.B.g Exploration of four samples from PCR Run 2

The four samples from PCR Run 2 were identified in the within-animal alignments as not encompassing the same region of the cytochrome-b gene as the PREDICT primer products (see Sub-section 5.B.e above). Consequently, the only way to compare the original sequences with the sequences from the PCR reruns (Table 5.13) was to run Blast on each sequence independently and compare the matches. This confirmed that the PCR in Run 2 isolated a sequence approximately 380bp long between sites 40 and 420. In contrast, the PREDICT primers isolated 457bp between sites 446 and 873. This total lack of overlap explained the inability to align the pair-wise matches (Table 5.13).

The sequences from the first PCR runs all matched species which were identified in the field (Table 5.15). One of the second runs had a very poor Blast match, with only 201 sites aligned, with an African rabbit species. However, the other three sequences matched over 98% of sites in NCBI sequences and the best matches in each case were that same species as the field identities. Although

these four samples cannot be used in the taxonomic analyses based on phylogenies, their field identities have been confirmed with duplicate PCR and/or successful Blast identification.

Table 5.15 Comparison of Blast results from four sequences obtained from PCR Run 2 and subsequent PCR runs.

		First PCR Run (#2)			Second PCR Run (#14, 15 or 16)		
Animal Code	Field Identity	Blast Identity	Matched length	Identities	Blast Identity	Matched length	Identities
13-1-1-4-01	*Suncus murinus*	*S. murinus*	368	99.2%	*Sylvisorex lunaris*	201	81.4%
13-1-2-1-06	*Rattus nitidus*	*R. pyctoris*	379	99.7%	*R. pyctoris*	421	99.3%
13-2-1-6-02	*Soriculus nigrescens*	*S. nigrescens*	383	98.7%	*S. nigrescens*	426	98.1%
13-2-4-3-06	*Soriculus nigrescens*	*S. nigrescens*	329	97.1%	*S. nigrescens*	419	98.4%

5.B.h Appendix I. Electrophoretic Gel Plate Photographs

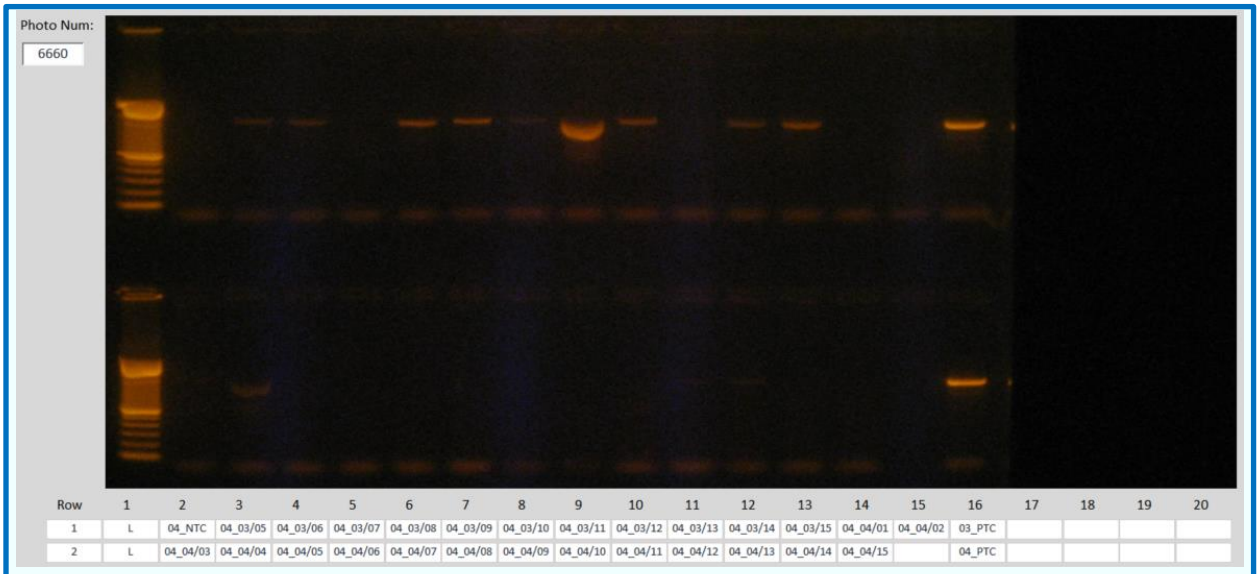


Figure 5.13 Gel plate for PCR run 4

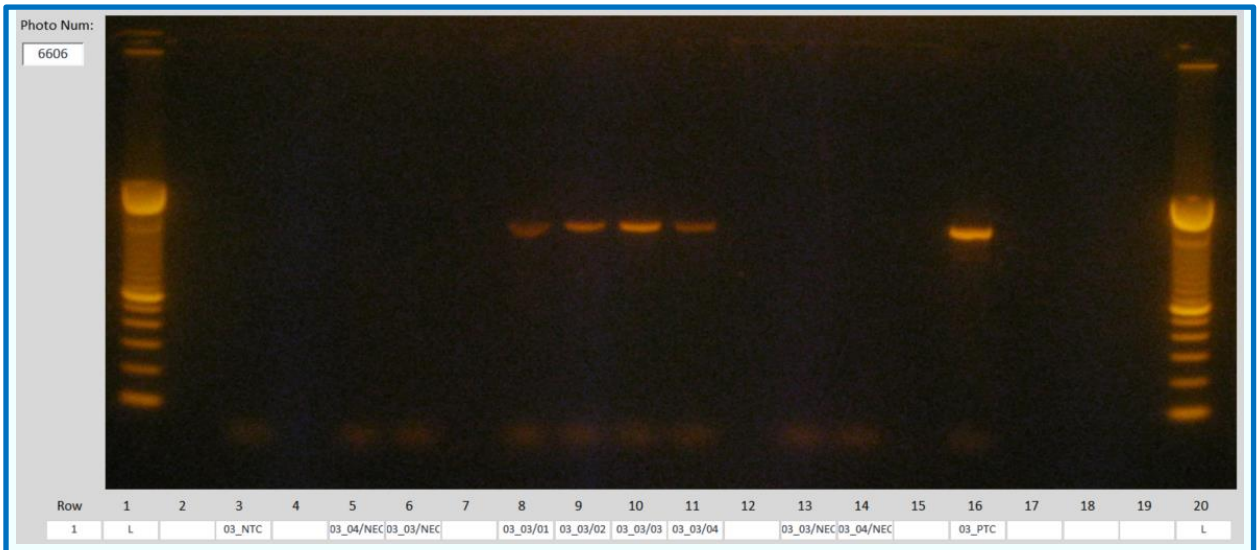
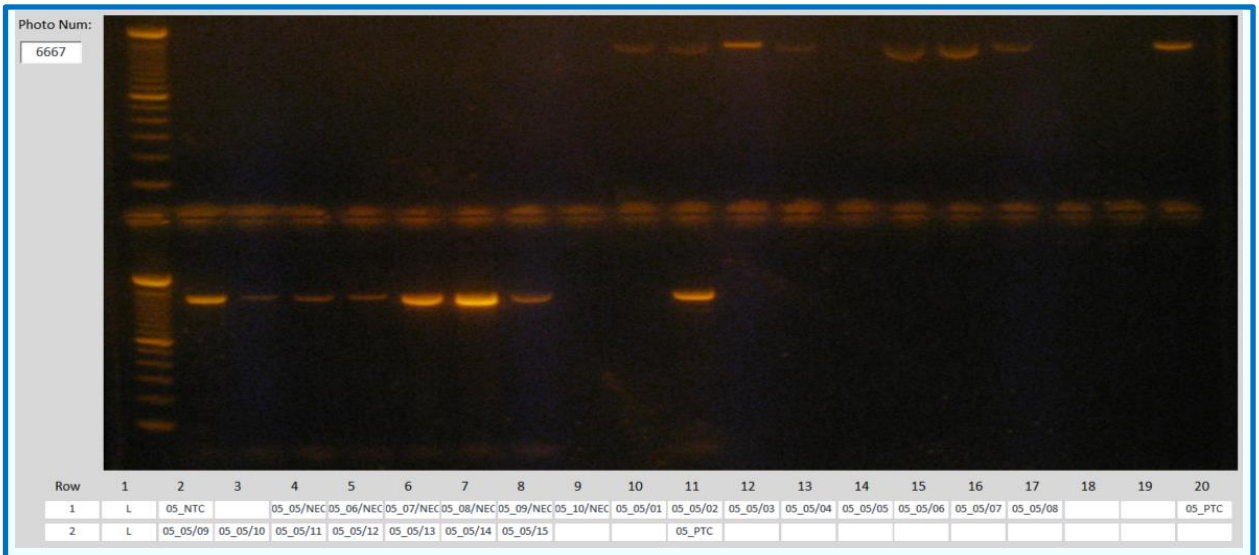
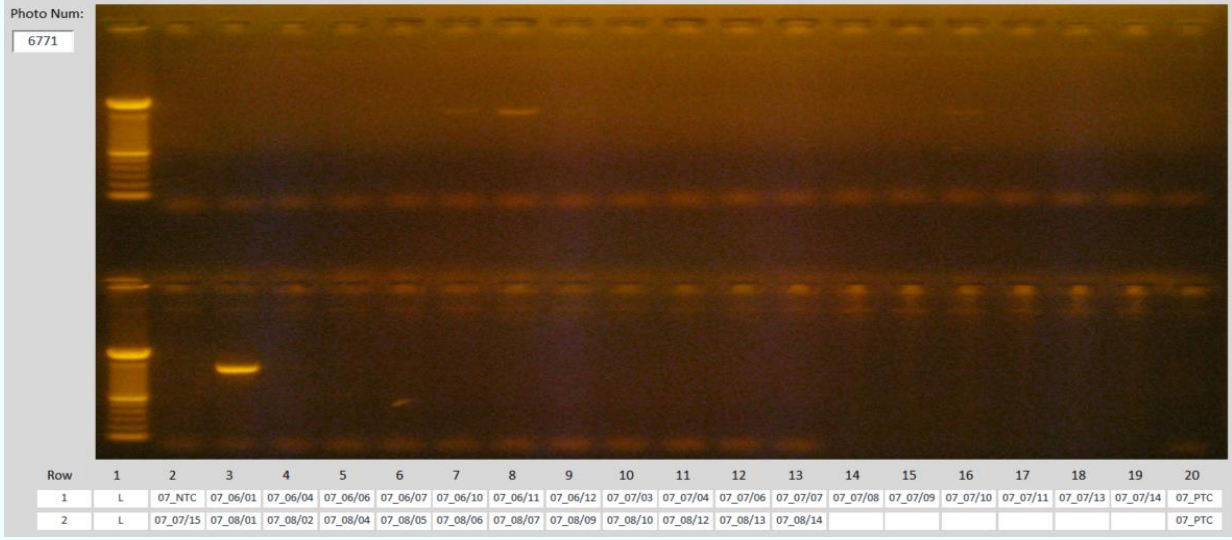


Figure 5.14 Gel plate for PCR run 3.



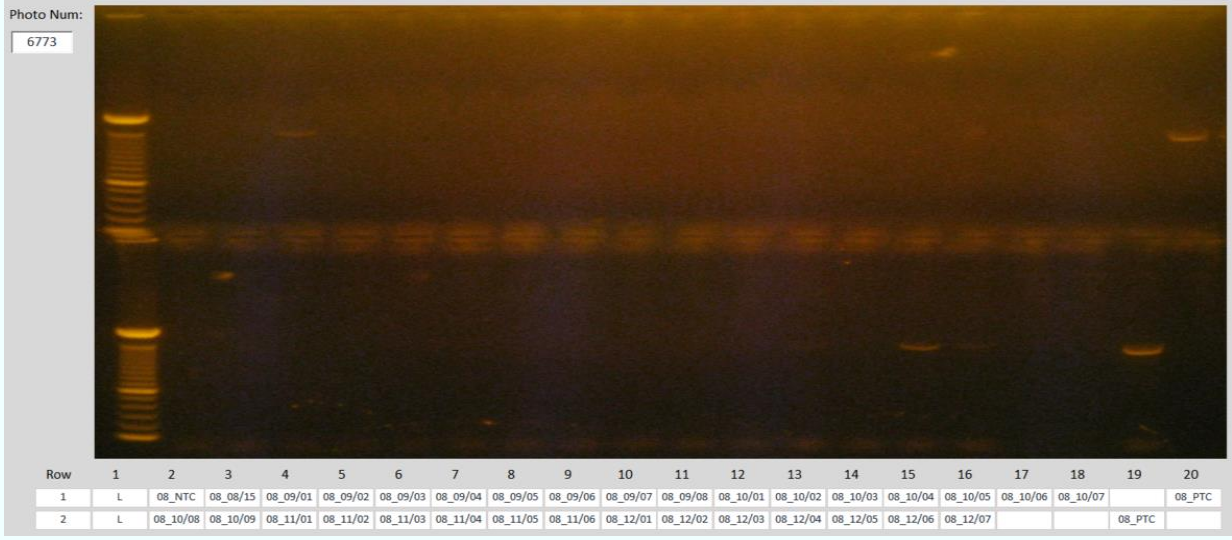
Row	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	L	05_NTC		05_05/NEC	05_06/NEC	05_07/NEC	05_08/NEC	05_09/NEC	05_10/NEC	05_05/01	05_05/02	05_05/03	05_05/04	05_05/05	05_05/06	05_05/07	05_05/08			05_PTC
2	L	05_05/09	05_05/10	05_05/11	05_05/12	05_05/13	05_05/14	05_05/15			05_PTC									

Figure 5.15 Gel plate for PCR run 5.



Row	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	L	07_NTC	07_06/01	07_06/04	07_06/06	07_06/07	07_06/10	07_06/11	07_06/12	07_07/03	07_07/04	07_07/06	07_07/07	07_07/08	07_07/09	07_07/10	07_07/11	07_07/13	07_07/14	07_PTC
2	L	07_07/15	07_08/01	07_08/02	07_08/04	07_08/05	07_08/06	07_08/07	07_08/09	07_08/10	07_08/12	07_08/13	07_08/14							07_PTC

Figure 5.16 gel plate for PCR run 7.



Row	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	L	08_NTC	08_08/15	08_09/01	08_09/02	08_09/03	08_09/04	08_09/05	08_09/06	08_09/07	08_09/08	08_10/01	08_10/02	08_10/03	08_10/04	08_10/05	08_10/06	08_10/07		08_PTC
2	L	08_10/08	08_10/09	08_11/01	08_11/02	08_11/03	08_11/04	08_11/05	08_11/06	08_12/01	08_12/02	08_12/03	08_12/04	08_12/05	08_12/06	08_12/07			08_PTC	

Figure 5.17 Gel plate for PCR run 8

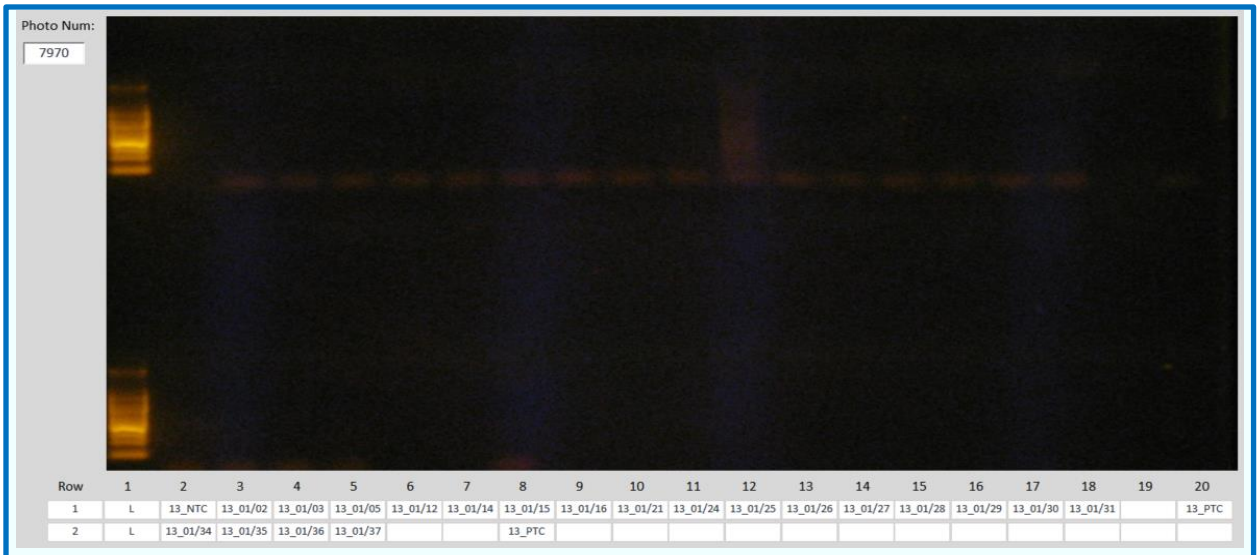


Figure 5.18 Gel plate for PCR run 13

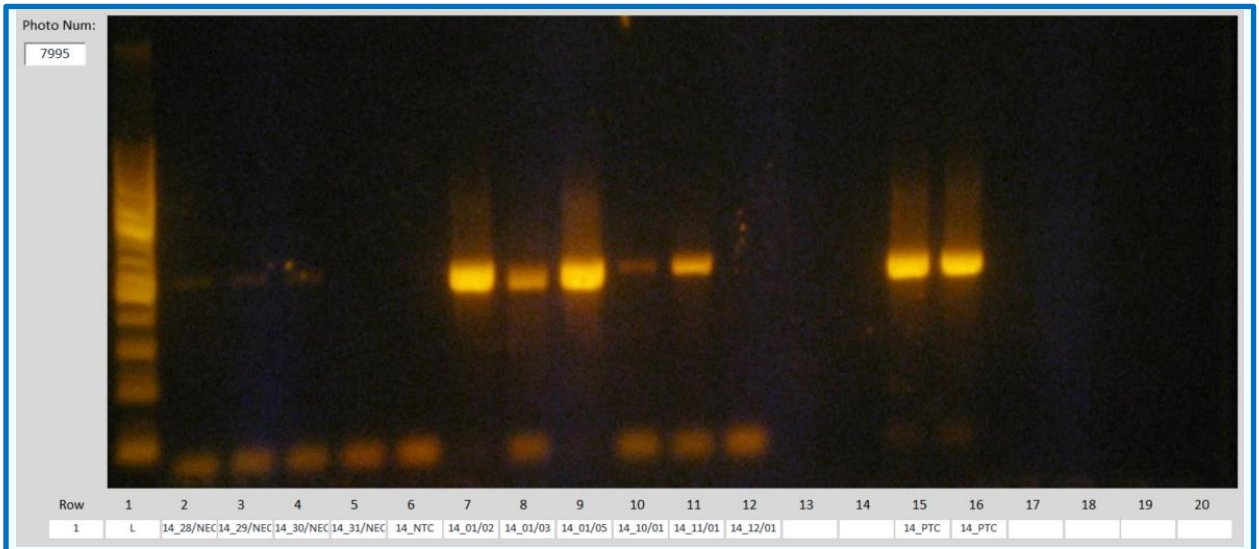


Figure 5.19 Gel plate for PCR run 14

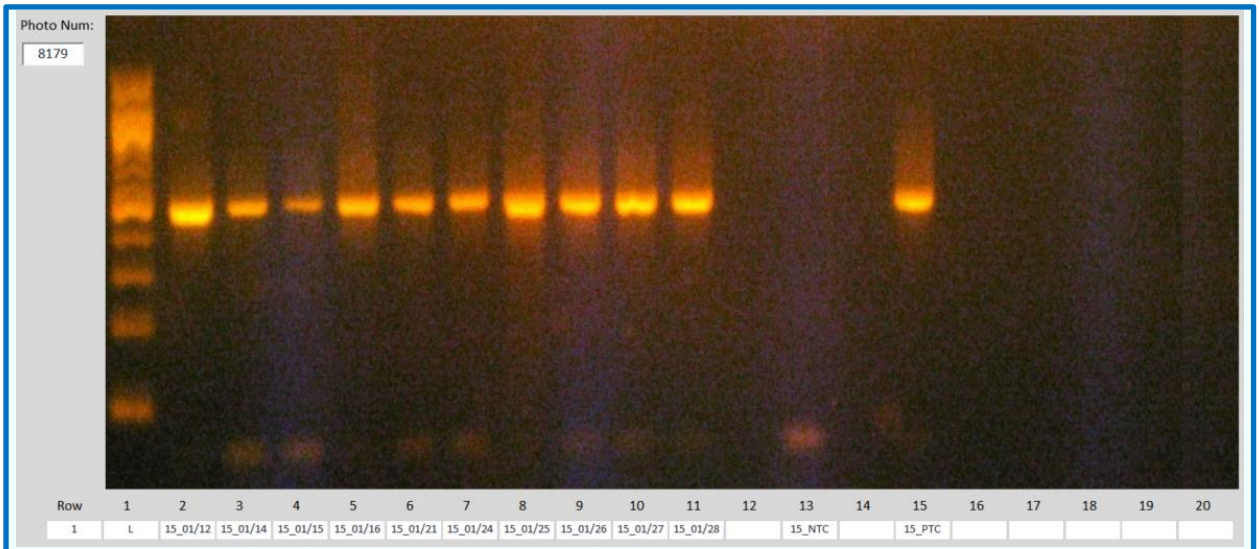


Figure 5.20 Gel plate for PCR run 15

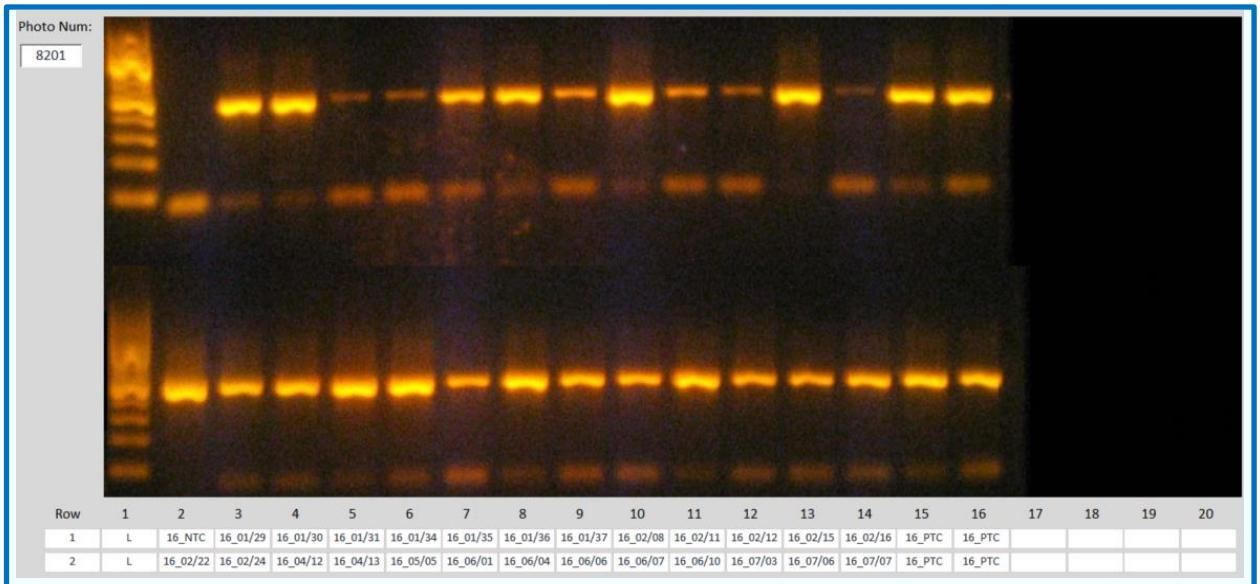


Figure 5.21 Gel plate for PCR run 16

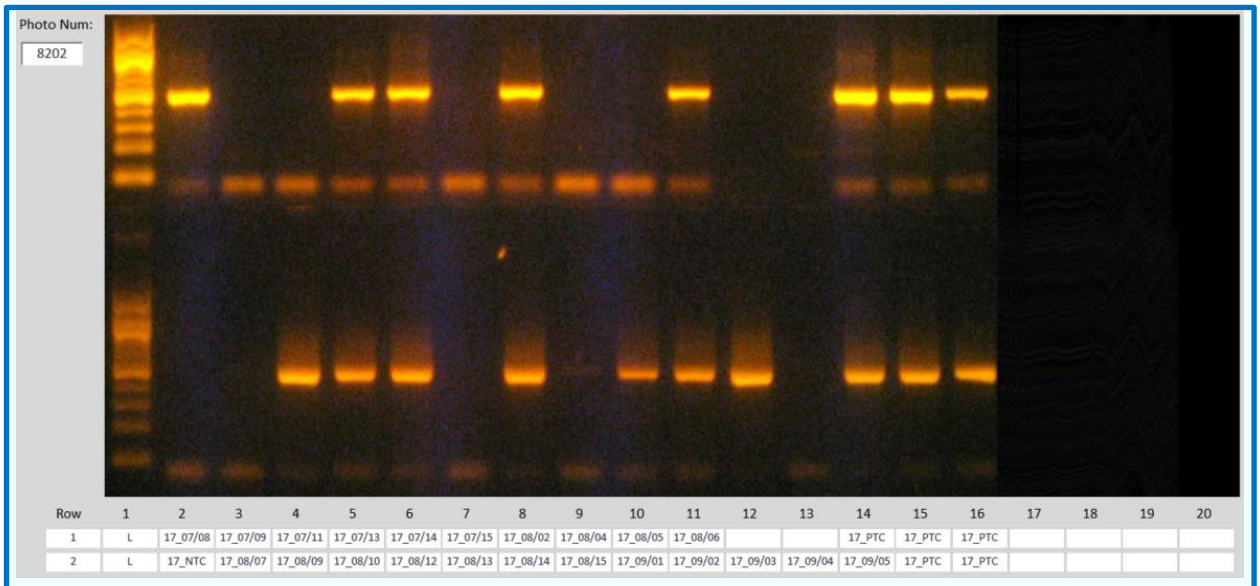


Figure 5.22 Gel plate for PCR run 17

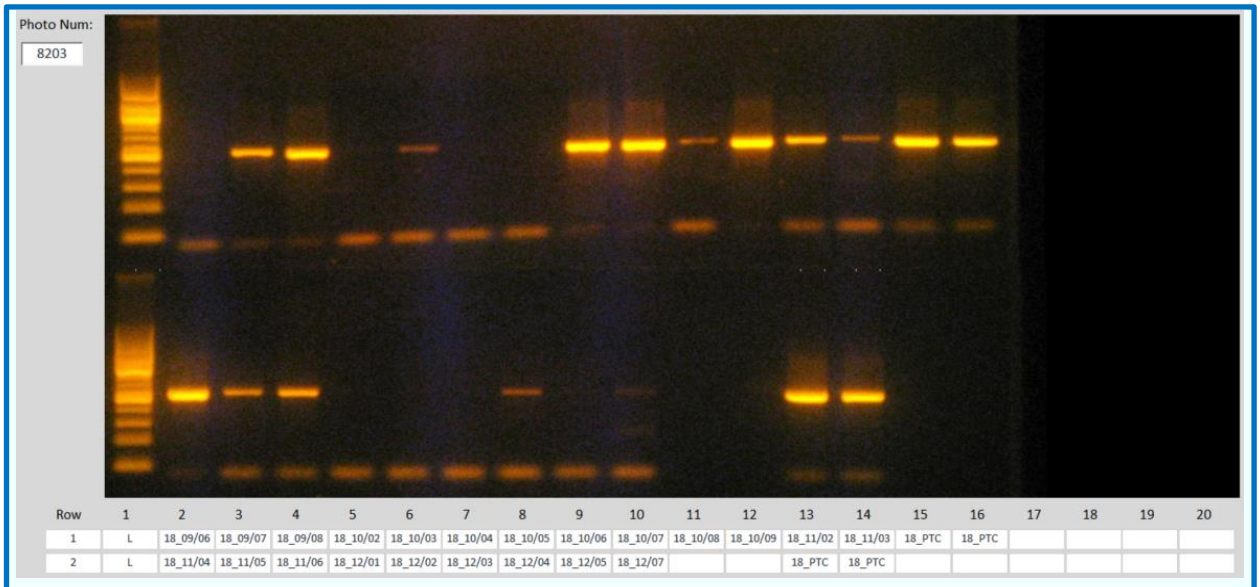


Figure 5.23 Gel plate for PCR run 18

5.B.i Appendix II. Example Sequence Traces

445356101_01_28_Predict_F_F01.pdf

Figure 5.24 Embedded PDF object showing the trace of a successful sequence from a PREDICT extraction. The trimming algorithm removed the first nine and no trailing nucleotides. This object has been appended to the main thesis in Supplementary Information Fig 5-24.pdf.

445356101_03_11_L14724_C09.pdf

Figure 5.25 Embedded PDF object showing the trace of an unsuccessful sequence from a full-length cytochrome-b extraction. The trimming algorithm removed the first 201 nucleotides and all nucleotides after 417. However, the quality of the sequence was so poor that it was rejected. This object has been appended to the main thesis in Supplementary Information Fig 5-24.pdf.

5.C Analysis of Cytochrome-b Sequences Obtained From GenBank

5.C.a Methods

This analysis is based solely on sequences of the cytochrome-b gene available from the NCBI GenBank database. Overall, I selected 200 sequences from the three families of small mammals relevant to this study; 22 cricetid, 112 murid and 66 soricid sequences. They were selected subjectively; either all available sequences from each species or, if they were very common, I selected a variable number reflecting their provenance and the number of pop-sets (as indicated by the first two letters of their accession number) up to a maximum of ten. Where possible I selected full-length sequences, but for species with few specimens, I accepted shorter sequences if available. These sequences were labelled with the abbreviated taxon, the GenBank accession number and the location from where the specimen was collected, if recorded on GenBank or in referenced papers. They were downloaded to the MS Access database used to store the field data (see Chapter 2 for a full description of this system) and are hereafter known as the "reference sequences".

The sequences were then aligned in Mega 7 (Kumar *et al.*, 2016), separately for each family to give a reference phylogeny. These were annotated with colour-coded square symbols to indicate species and genus identities, plus branch lengths and bootstrap percentages for each node. They were assessed in four stages;

- Firstly, the identification of anomalies in the phylogenies, where members of one species clustered separately with other species or formed outliers.
- To investigate the similarity of con-specific sequences, based on the p-distance matrices.
- The selection of suitable con-specific and con-generic p-distance thresholds, either for the family as a whole or for individual genera or species where necessary.
- Using these thresholds, one sequence was selected subjectively to represent phylogenetic clusters in subsequent analyses. Occasionally more than one sequence was selected to represent different pop-sets or to include sequences of particular relevance from Nepal.

5.C.b Cricetidae

22 nucleotide sequences were obtained belonging to four species in two genera (Table 5.16a), which were all the cytochrome-b sequences available for these species on GenBank. A full pair-wise matrix of p-distances was calculated yielding 462 comparisons (Table 5.16b). The *Alticola stoliczkanus* sequences clustered tightly together with a maximum con-specific p-distance of 0.047 (Figure 5.26a

Table 5.16 The number of a) sequences and b)pair-wise comparisons of four species of Cricetidae obtained from GenBank.

Taxon		a) Number of Sequences	b) Pair-wise comparisons			
			Con-specific	Con-generic	Con-familial	Total
Alticola	*stoliczkanus*	5	20	0	85	105
	irene	9	72	72	45	189
Neodon	*leucurus*	2	2	30	10	42
	sikimensis	6	30	66	30	126
	Total	22	124	168	170	462

and Figure 5.27). As it was the only member of the genus in the phylogeny, there were no con-generic distances, but the minimum p-distance to sequences of the other genera (con-familial distance) was 0.128. *Neodon irene*, formed a distinct series of groups, which were manifest as the different clusters of con-specific distances in Figure 5.26b. The cluster between 0 and 0.02 represent the distances within the cluster of five similar sequences (Figure 5.27), while the cluster between 0.05 and 0.06 represents the distances of that cluster with the pair beneath it in the phylogeny. The p-distances greater than 0.08 involve the two outlying sequences of this species. One of these (GU908336) had p-distances between 0.09 and 0.10, and is probably a good con-specific sequence. However, the single specimen from Nepal (KU214577) had p-distances greater than 0.10 and its status is discussed further in Section 5.4. The two sequences of *N. leucurus* had a distinct cluster of con-generic distances, indicating that its relationship to the other two species was equivalent; the low bootstrap values for these nodes indicating some ambiguity in the exact phylogenies. Finally, the sequences of *N. sikimensis* fell into two branches (Figure 5.26d), which still formed a discrete

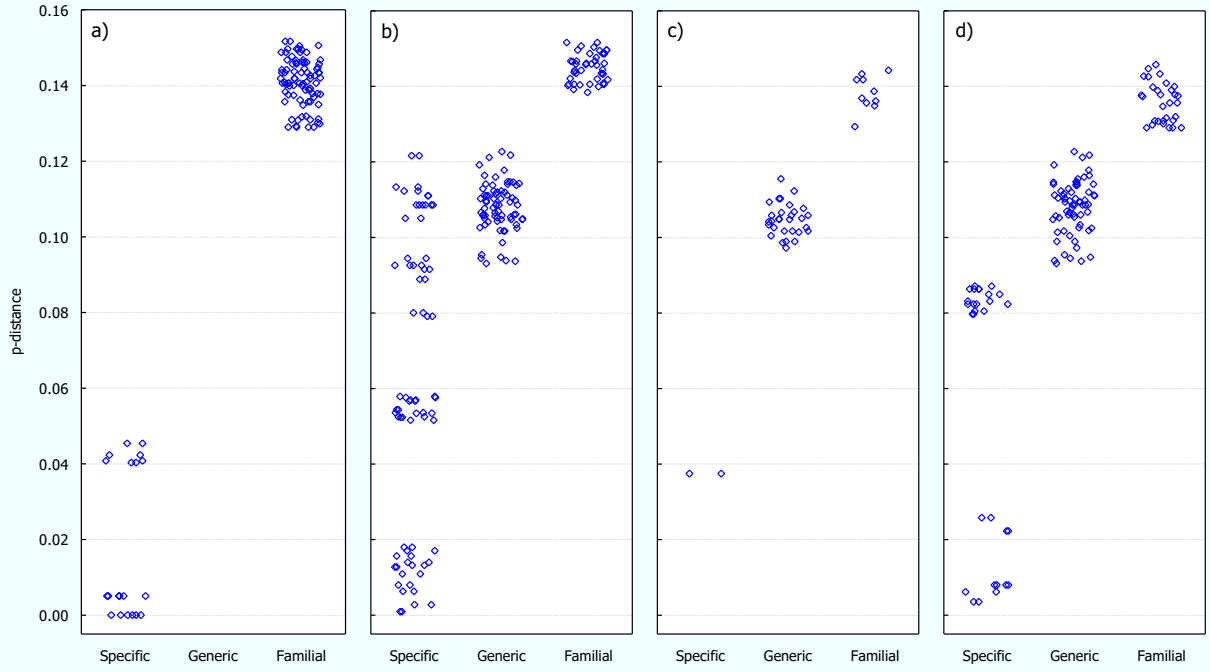


Figure 5.26 P-distances between sequences of four species of Cricetidae; a) *Alticola stoliczkanus*, b) *Neodon irene*, c) *Neodon leucurus* and d) *Neodon sikimensis*. For each species, the p-distances have been categorised into three taxonomic levels; con-specific, con-generic and con-familial. Individual points have been randomly jittered on the x-axis for clarity, but not on the y-axis.

group separate from the other two species. This resulted in a clear con-generic distance threshold of 0.09, which was also applicable to the other species if the two outlying *N. irene* were discounted. There was also a con-familial (cross-generic) threshold of approx. 0.125.

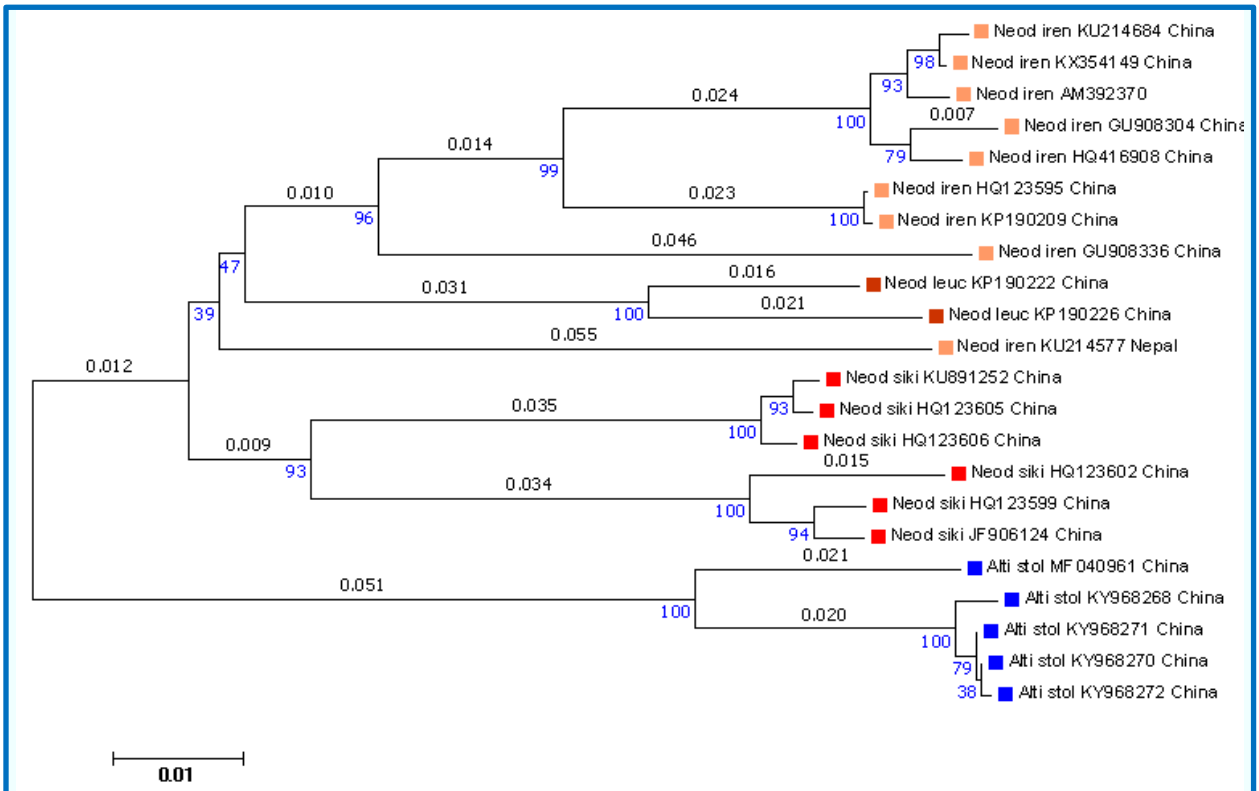


Figure 5.27 The phylogeny of 22 sequences of Cricetidae belonging to four (colour-coded) species. Figures above the lines are p-distances calculated from the distance matrix and are represented by line lengths. Blue figures below and left of the nodes are bootstrap percentages (based on 1000 replicates)

5.C.c Muridae

The 112 murid sequences were obtained from four genera and 23 species (Table 5.17). The full pair-wise matrix contained 12,432 cells, which were categorised into 526 con-specific and 2816 con-generic comparisons. Due to the large number of comparisons, the analysis of this family has been divided into the four genera and presented separately.

Table 5.17 The number of a) sequences and b)pair-wise comparisons of 23 species of Muridae obtained from GenBank.

Taxon		a) Number of Sequences	b) Pair-wise comparisons			
			Con-specific	Con-generic	Con-familial	Total
Apodemus	*gurkha*	4	12	36	396	444
	pallipes	4	12	36	396	444
	sylvaticus	4	12	36	396	444
	wardi	1	0	12	99	111
Mus	*booduga*	6	30	162	474	666
	cervicolor	4	12	116	316	444
	cookii	4	12	116	316	444
	musculus	8	56	200	632	888
	pahari	4	12	116	316	444
	platythrix	2	2	62	158	222
	saxicola	1	0	32	79	111
	terricolor	4	12	116	316	444
Niviventer	*andersoni*	8	56	224	608	888
	brahma	3	6	99	228	333
	eha	6	30	180	456	666
	excelsior	4	12	128	304	444
	fulvescens	8	56	224	608	888
	niviventer	7	42	203	532	777
Rattus	*andamanensis*	6	30	144	492	666
	nitidus	7	42	161	574	777
	pyctoris	5	20	125	410	555
	rattus	6	30	144	492	666
	tanezumi	6	30	144	492	666
Total		112	526	2816	9090	12432

5.C.c.i The Genus Apodemus

The four species of Apodemus fell into two distinct groups; *A. gurkha*, with a mean con-specific p-distance of approx. 0.012 and the other three species which had a mean con-specific p-distance of approx. 0.06 (Figure 5.28). The mean p-distance between members of these two groups was 0.16, which constitutes a difference at the generic level – greater than the distance between *Niviventer* and *Rattus* – indicating that this species probably requires revision. The phylogeny for this genus is confused by two sequences. Firstly, *A. sylvaticus* (AF159395) had a mean con-specific distance 0.105, which is shown as the cluster of six points in Figure 5.29c. Unfortunately, the provenance of this sequence was unknown, but given the wide distribution of the other three, there is strong evidence that it was more than just geographic variation. Secondly, the single sequence of *A. wardi* (AB096838) fell within the *A. pallipes* cluster, which looked like a clear mis-identification.

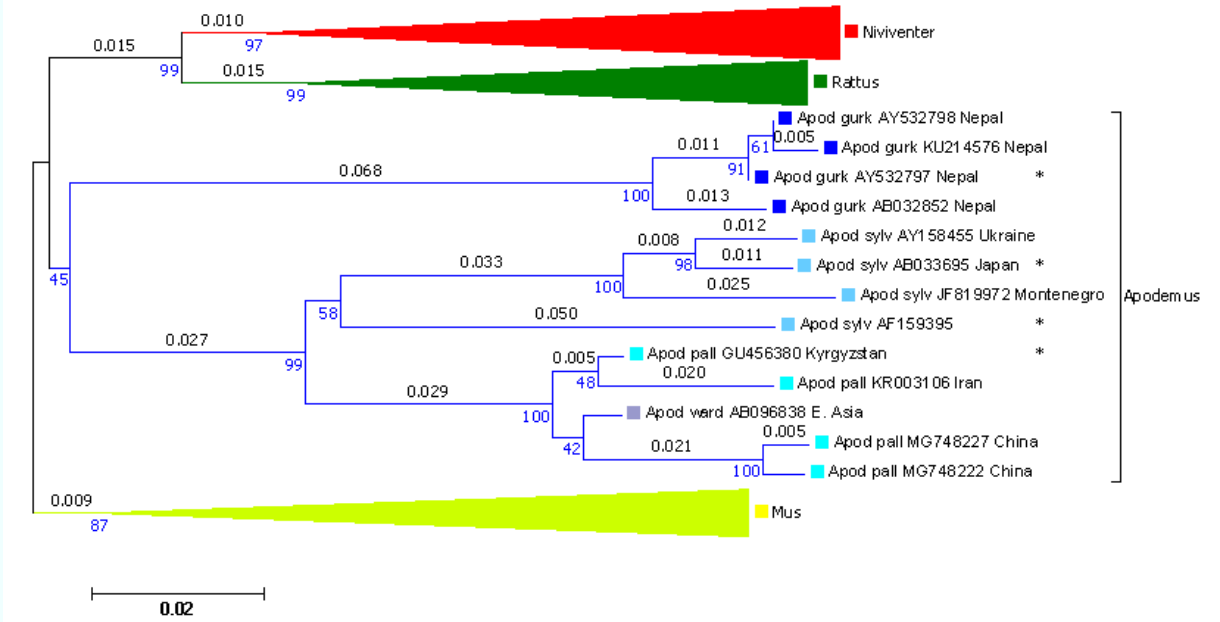


Figure 5.28 The phylogeny of 13 sequences of *Apodemus* belonging to four (colour-coded) species. The sub-trees of the other three genera within this family are collapsed. Sequences selected for subsequent analysis are indicated with *.

If the *pallipes/wardi* group was considered to be a single species, and the single A. *sylvaticus* is discounted, then the maximum con-specific distance within this genus was 0.055, and the minimum con-generic distance was 0.097. This gave a clear p-distance threshold of around 0.07 between con-specific and con-generic assigment of individuals. The four specimens selected for subsequent analysis are indicated with an * in Figure 5.28.

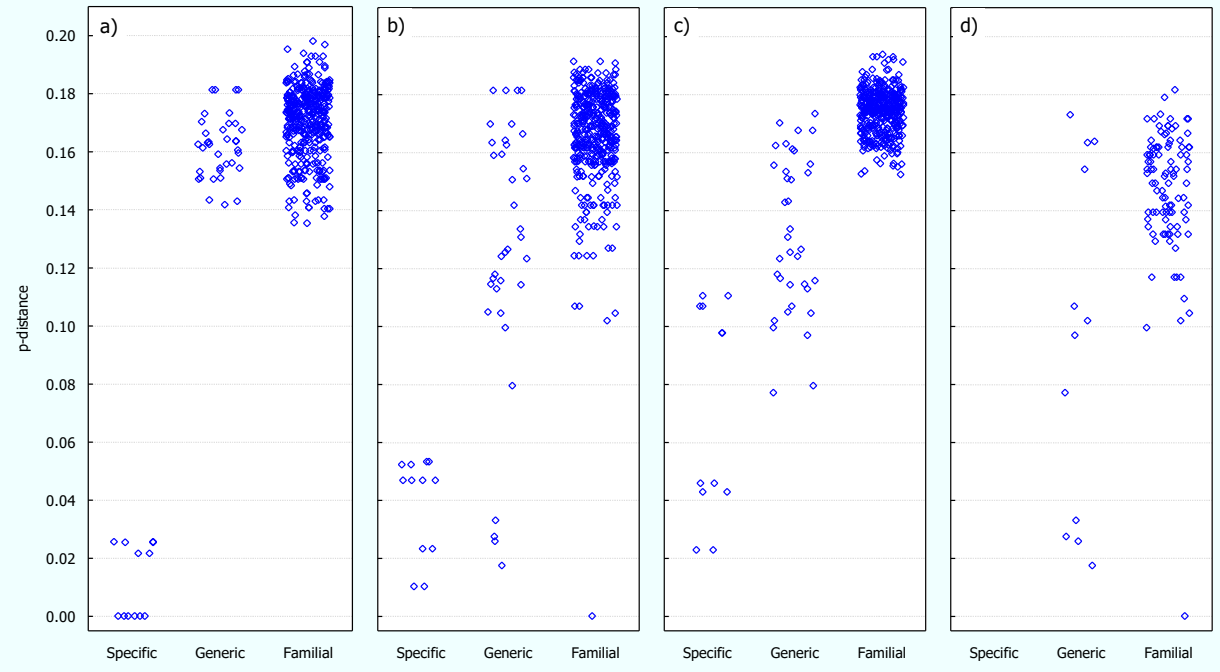


Figure 5.29 P-distances between sequences of four species of *Apodemus*; a) *A. gurkha*, b) *A. pallipes*, c) *A. sylvaticus* and d) *A. wardi*. Other details are explained in Figure 5.26.

5.C.c.ii The Genus Mus

Overall, the 32 sequences from this genus segregated very clearly into species groups, with only two ambiguous sequences (Figure 5.30). The first was the single sequence of M. *Saxicola* (AY057815) which was identical to one of the two M *platythrix* sequences. This was clearly a mis-identification and the three sequences have been clustered into the M.*platythrix* group for further analysis. The second was M. *Pahari* (EU349767), which had a mean con-specific p-distance of 0.113 (Figure 5.32a). Excluding this specimen, the mean con-specific p-distance was 0.015 (Figure 5.31 & Figure 5.32), which was the smallest of the four genera. The maximum con-specific distance was 0.05, and the minimum con-generic distance was slightly greater than 0.10, which gave a huge taxonomic gap. A threshold of 0.075 was selected subjectively for subsequent analysis. This allowed the selection of

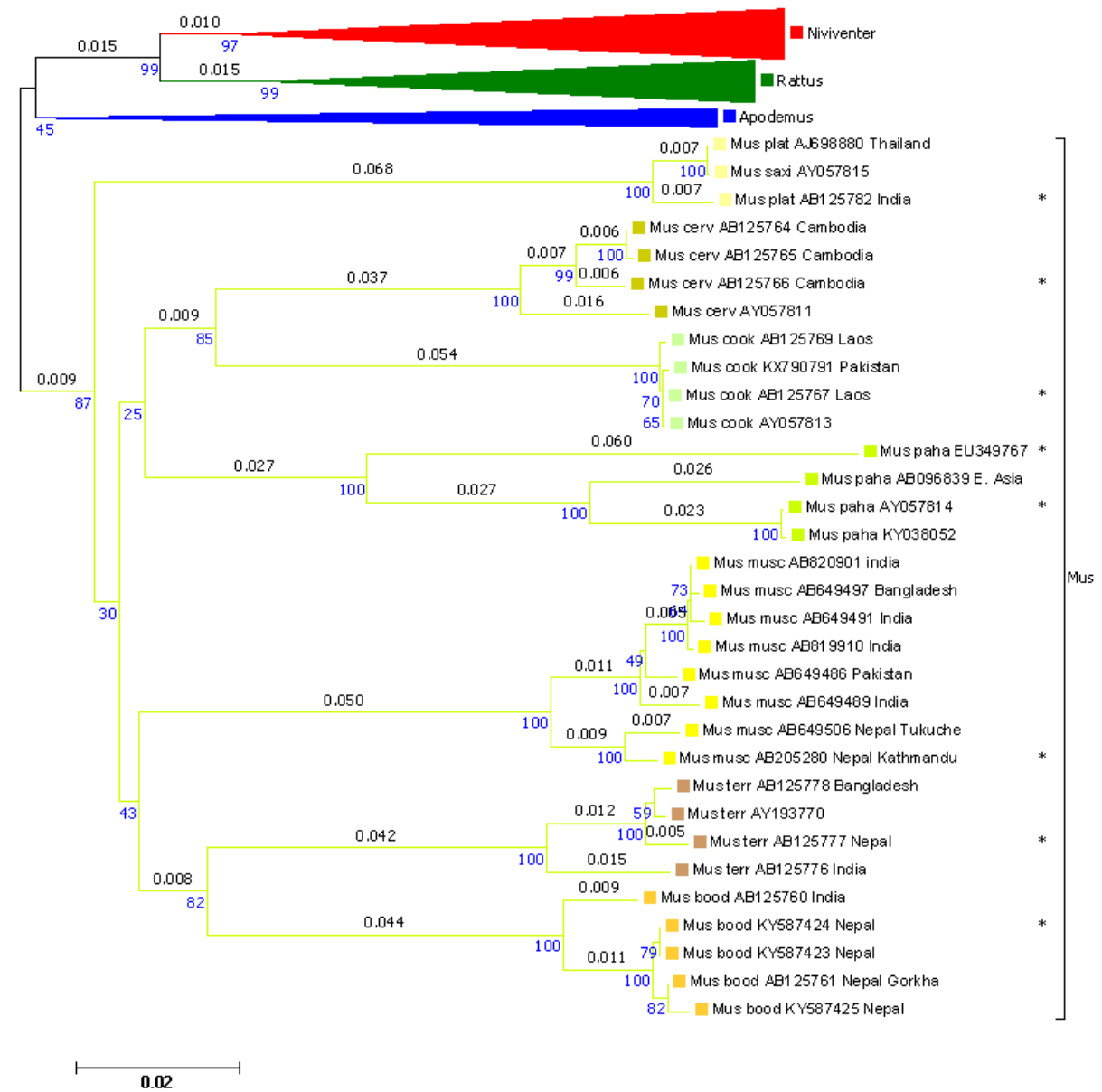


Figure 5.30 The phylogeny of 32 sequences of *Mus* belonging to eight (colour-coded) species. Other details are explained in Figure 5.28

one sequence from each species (group), plus the additional specimen of M. *pahari*, for subsequent analysis.

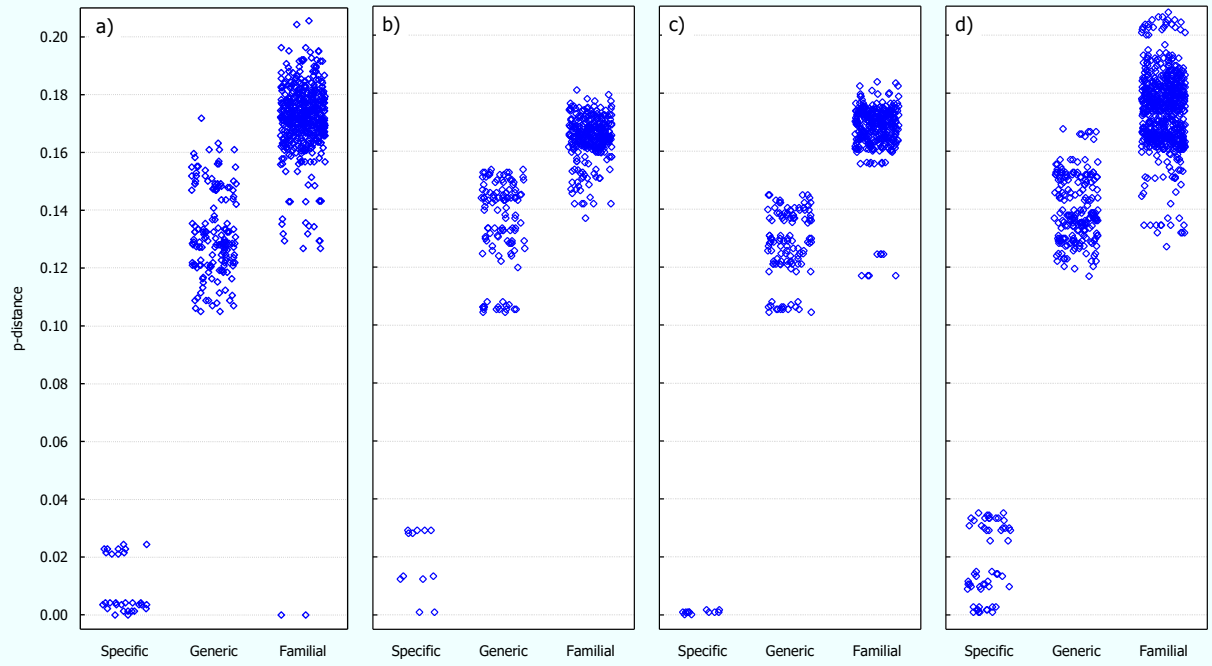


Figure 5.31 P-distances between sequences of four species of *Mus*; a) *M. booduga*, b) *M. cervicolor*, c) *M.cookii* and d) *M. musculus*. Other details are explained in Figure 5.26.

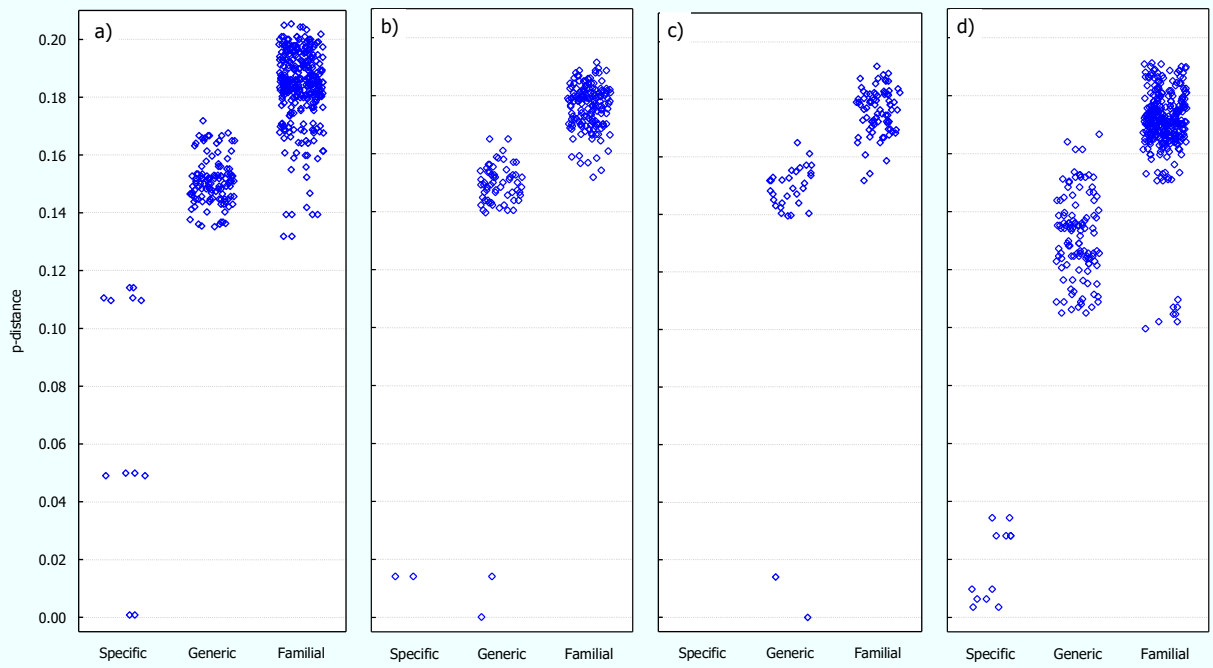


Figure 5.32 P-distances between sequences of four species of *Mus*; a) *M. pahari*, b) *M. platythrix*, c) *M.saxicolor* and d) *M. terricolor*. Other details are explained in Figure 5.26.

5.C.c.iii The Genus Niviventer

There was considerable ambiguity in the species identities of the 36 sequences from this genus. This is most clearly seen in the scatterplots of p-distances (Figure 5.33 and Figure 5.34). Only one of the six species, *N. brahma* had all con-specific and con-generic sequences in tight clusters (Figure 5.33b). All other species had considerable overlap between con-specific and con-generic p-distances. The

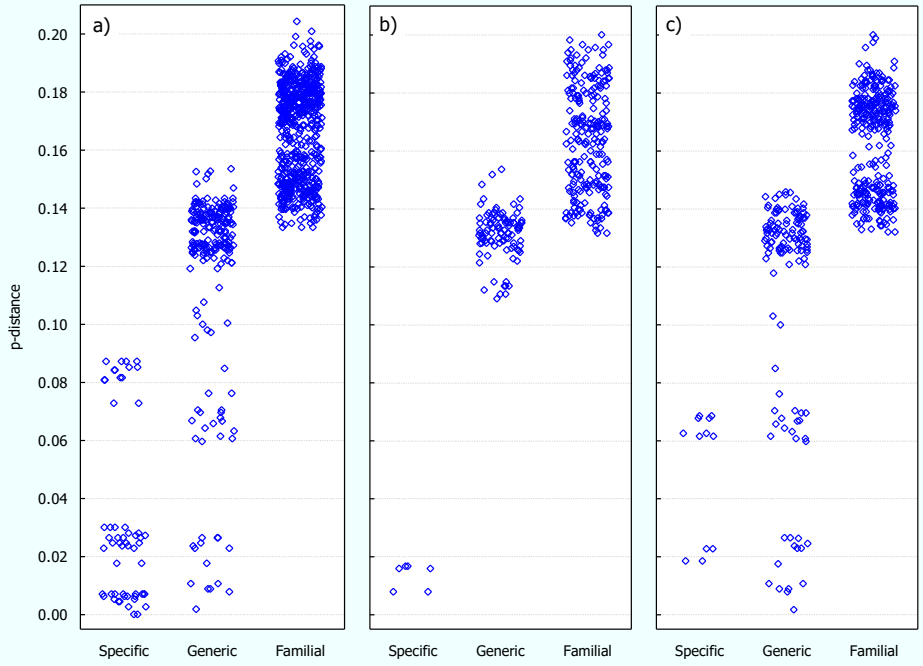


Figure 5.33 P-distances between sequences of three species of *Niviventer*; a) *N. anderson*i, b) *N. brahma* and c) *N. excelsior*. Other details are explained in Figure 5.26.

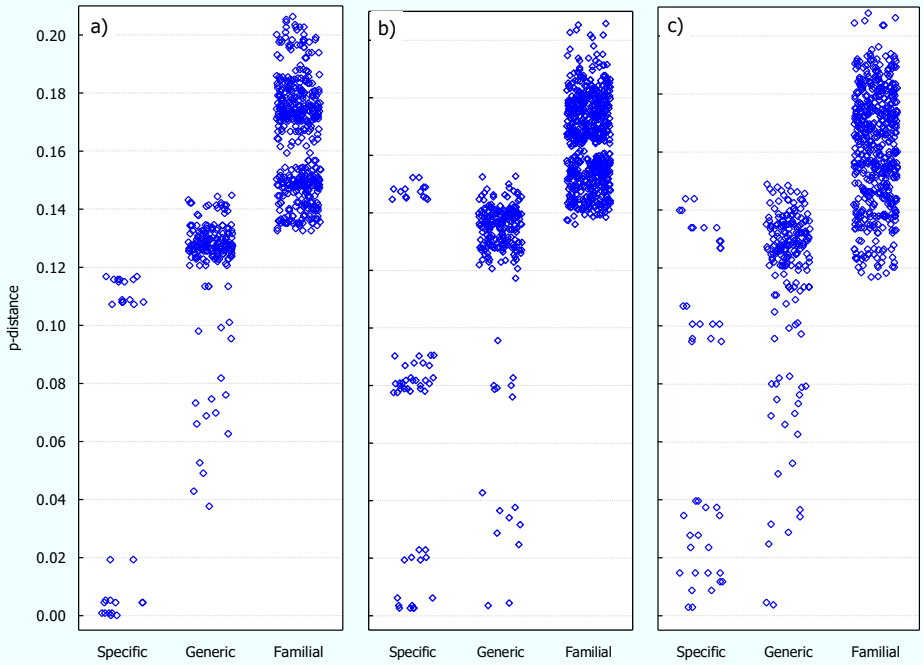


Figure 5.34 P-distances between sequences of three species of *Niviventer*; a) *N. eha*, b) *N. fulvescens* and c) *N. niviventer*. Other details are explained in Figure 5.26.

two other species (besides *N. brahma*) that were included as out-groups (*N. andersoni* and *N. excelsior*; Figure 5.33 & Figure 5.35) formed a distinct, but mixed, grouping with a maximum con-specific p-distance of approx. 0.08. However, this was caused entirely by one sequence (*N. andersoni*, EF053001), which generated the cluster of points in Figure 5.33a. Sequences of *N. eha* formed two distinct clusters, one of four specimens with extremely small p-distances (<= 0.02) and one pair of sequences that were approx. 0.11 distant (Figure 5.34a). The latter fell near a cluster of five *N. niviventer* from which they were between 0.04 and 0.08 distant. This cluster also included one sequence of *N. fulvescens*, which was represented by the cluster of con-specific points around 0.15 and

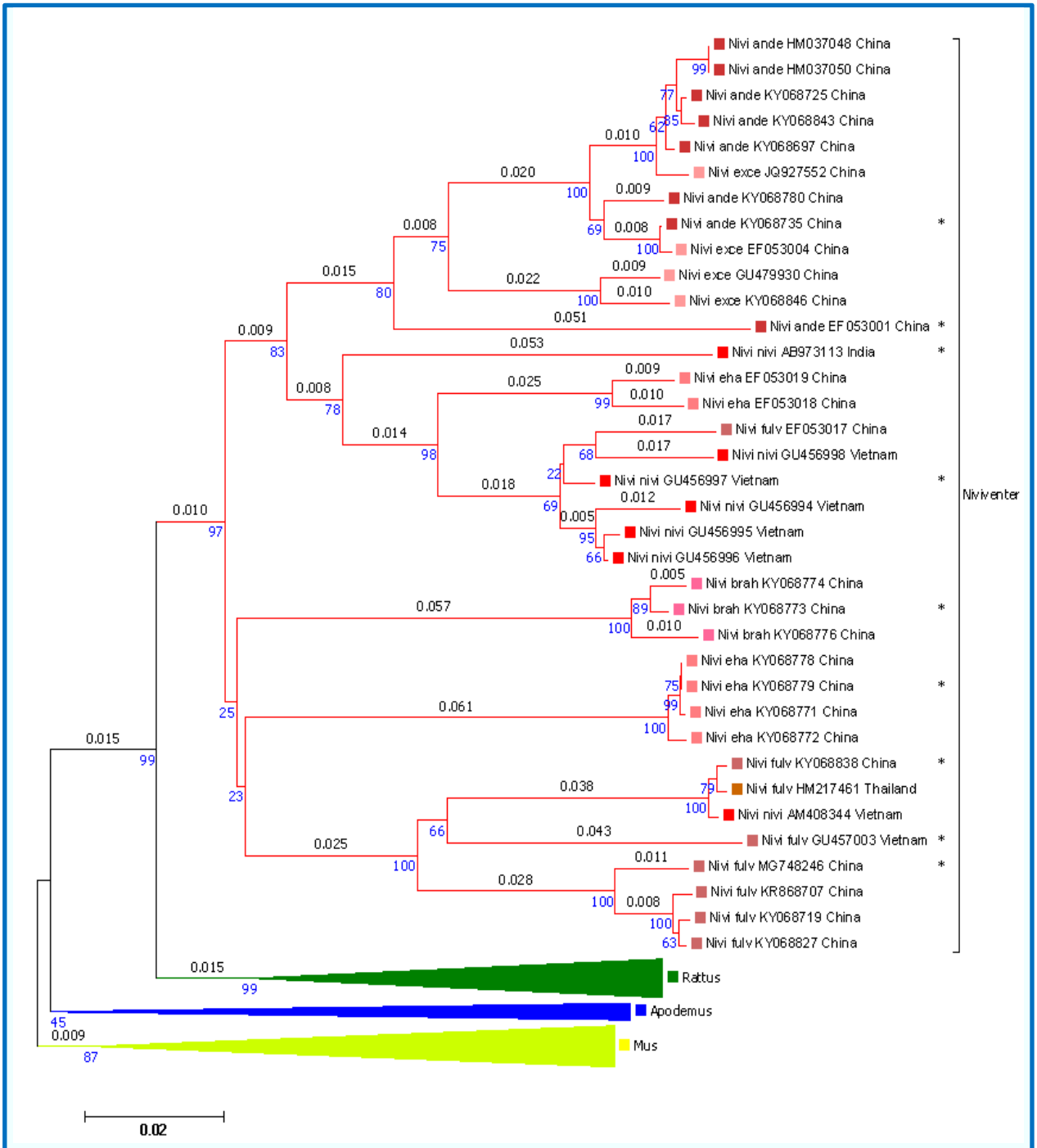


Figure 5.35 The phylogeny of 36 sequences of *Niviventer* belonging to six (colour-coded) species. Other details are explained in Figure 5.28.

con-generic points between 0.02 and 0.04 in Figure 5.34b. This was almost certainly a mis-identification. In addition, one sequence of *N. niviventer* formed an outlier to the main cluster, with a mean p-distance of approx.. 0.01 (Figure 5.34c). The *N. fulvescens* branch was quite disparate with three groups with p-distances of between 0.08 and 0.09 between them (Figure 5.34b). This branch also included one sequence of N. niviventer which was extremely close to two specimens, but which had a mean con-specific distance of about 0.135 (Figure 5.34c). This is almost certainly a mis-identification.

Considerable judgement was required to select suitable representative sequences for subsequent analysis. Firstly, if the obvious mis-identification (AM408344) was excluded, then the *N. fulvescens* grouping could be distinguished with a threshold of 0.095. However, three sequences were selected to represent the diversity of this group. Excluding the other obvious mis-identification (EF053017), the *N. niviventer* cluster could be distinguished with a threshold of about 0.05. However, this still left the outlying *N. niviventer* (AB973113), which had a mean p-distance with the rest of the group of approx. 0.10. For the purposes of this analysis it was considered to be a separate, putative species. Furthermore, this group was complicated by the presence of the two *N. eha* sequences; ignoring them removed the con-generic distances between 0.04 and 0.08. Finally, with a threshold of 0.09, the three out-group species could be distinguished as a group of *N. brahma*, one distinct specimen of *N. andersoni* (EF053001) and a mixed cluster of *N. andersoni/excelsior*. Generally, there was an upper con-generic threshold for all these species of approx. 0.15.

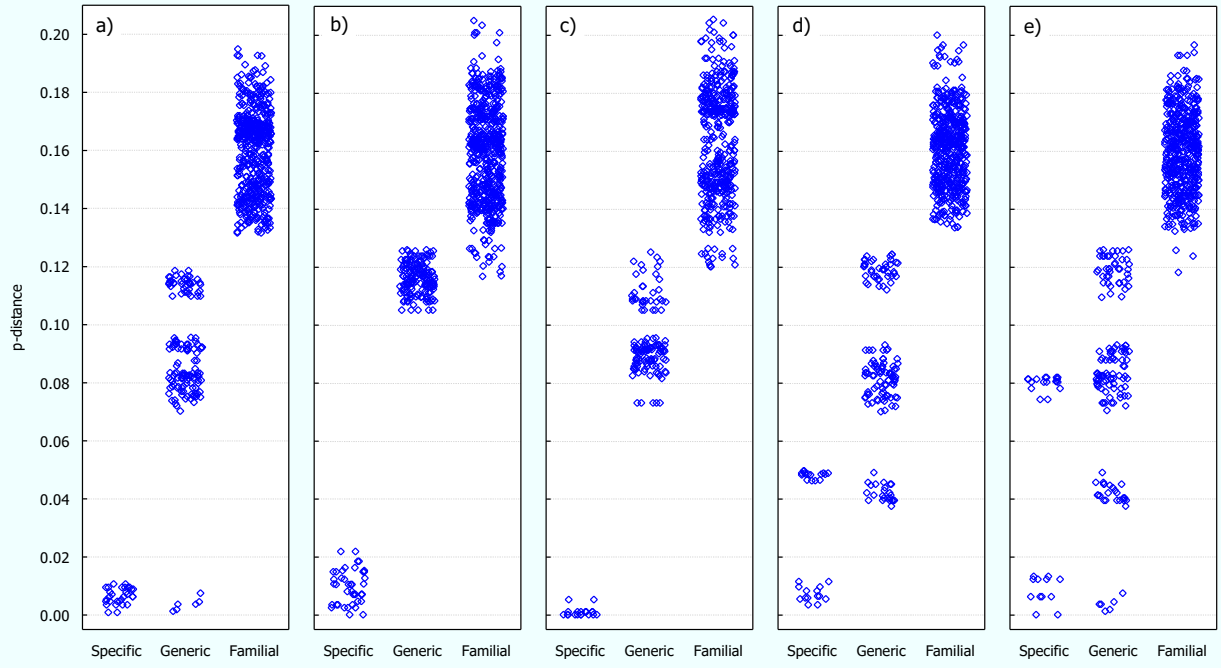


Figure 5.36 P-distances between sequences of five species of *Rattus*; a) *R. andamanensis*, b) *R. nitidus*, c) *R. pyctoris*, d) *R. rattus* and e) *R. tanezumi*. Other details are explained in Figure 5.26.

5.C.c.iv The Genus Rattus

The 30 sequences of this genus were selected from five species and formed at least three very distinct branches. The most distinct branch was *R. nittidus* (Figure 5.36b) which had no con-specific p-distances greater than 0.025 and no con-generic distances less than 0.10. *R. pyctoris* also had a very tight cluster of five sequences (p-distances <= 0.005) with no con-generic distances less than 0.07 (Figure 5.36c). Thirdly, the six sequences of *R. andamanensis* were very tightly clustered with all con-specific distances <= 0.012. However this cluster contained one sequence of *R. tanezumi*, which was clearly a mis-identification. Excluding this sequence meant that *R. andamanensis* had no con-generic distances less than 0.07 (Figure 5.36a). The other two species displayed some ambiguity. *R. rattus* fell into two groups with a mean p-distance of 0.05, but the second of these, from Nepal, had a slightly closer link to the *R. tanezumi* group, which was supported by 100% bootstrap simulations. Finally, there was one outlying *R. tanezumi* from Afghanistan (KP343797) which was about 0.08 distant from its con-specifics. An overall threshold p-distance of 0.07 distinguished five branches, although one was the singleton *R. tanezumi* and one was a compound *R. rattus/tanezumi*. The con-generic threshold for all species in this genus was very consistent at approx. 0.13.

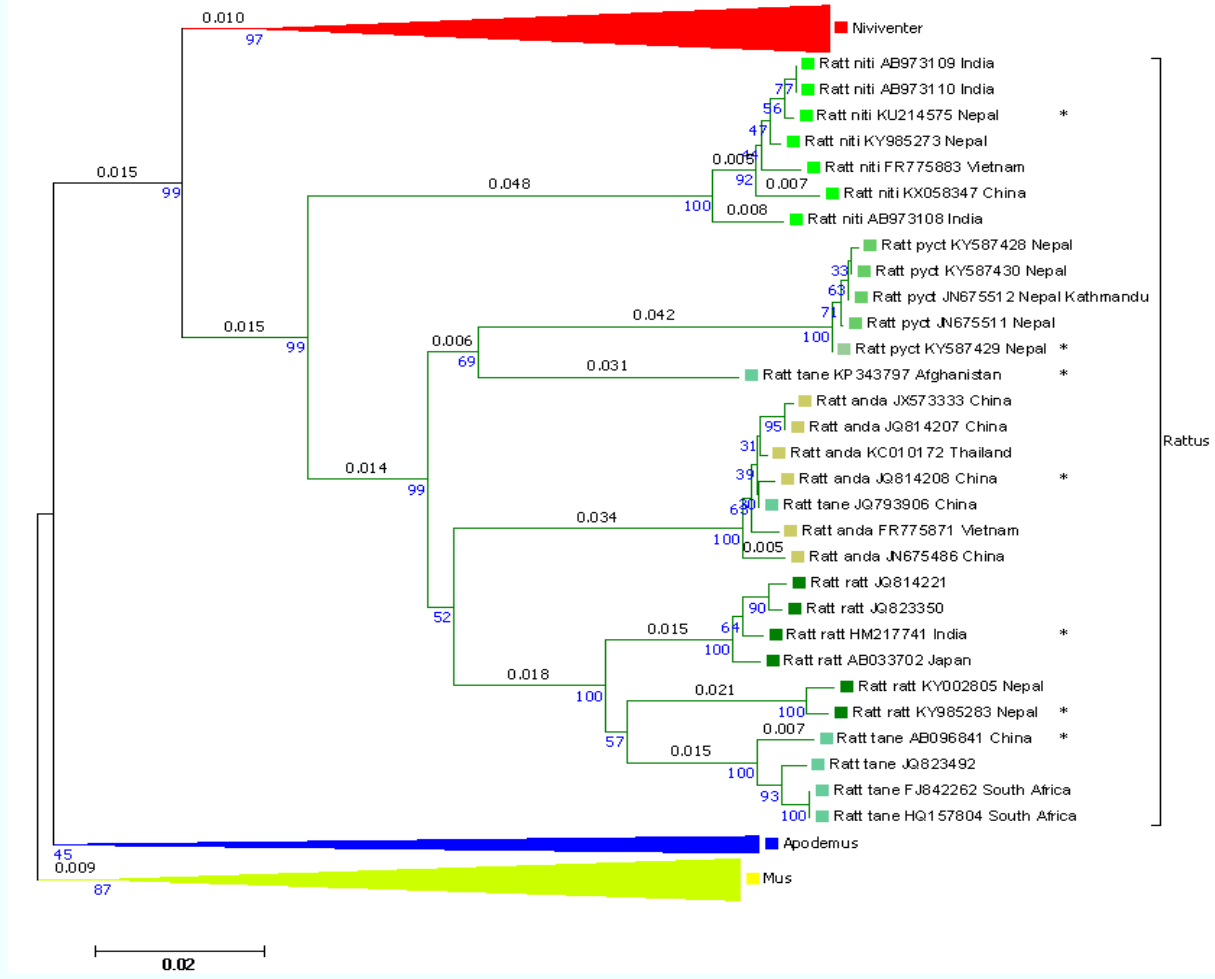


Figure 5.37 The phylogeny of 30 sequences of *Rattus* belonging to five (colour-coded) species. Other details are explained in Figure 5.28.

5.C.d Soricidae

The 66 sequences of Soricidae belonged to five genera and 14 species (Table 5.18). They partitioned into three distinct groups; a) *Episoriculus/Soriculus*, b) *Sorex* and c) *Crocidura/Suncus*.

Table 5.18 The number of a) sequences and b)pair-wise comparisons of 14 species of Soricidae obtained from GenBank.

Taxon		a) Number of Sequences	b) Pair-wise comparisons			
			Con-specific	Con-generic	Con-familial	Total
Crocidura	*attenuata*	4	12	8	240	260
	horsfieldii	2	2	8	120	130
Episoriculus	*caudatus*	7	42	84	329	455
	leucops	5	20	70	235	325
	macrurus	7	42	84	329	455
Sorex	*araneus*	3	6	54	135	195
	bedfordiae	8	56	104	360	520
	cylindricauda	7	42	98	315	455
	minutus	3	6	54	135	195
Soriculus	*nigrescens*	7	42	0	413	455
Suncus	*etruscus*	4	12	36	212	260
	montanus	3	6	30	159	195
	murinus	4	12	36	212	260
	stoliczkanus	2	2	22	106	130
Total		66	302	688	3300	4290

5.C.d.i Group A: Episoriculus/Soriculus

This group contained four species, two of which clustered into distinct branches; *E.leucops* (Figure 5.38b) and *E. macrurus* (Figure 5.38c). These showed clear con-specific thresholds of 0.070 and 0.120 respectively. The seven sequences of *E. caudatus* fell into two separate groups; four from China

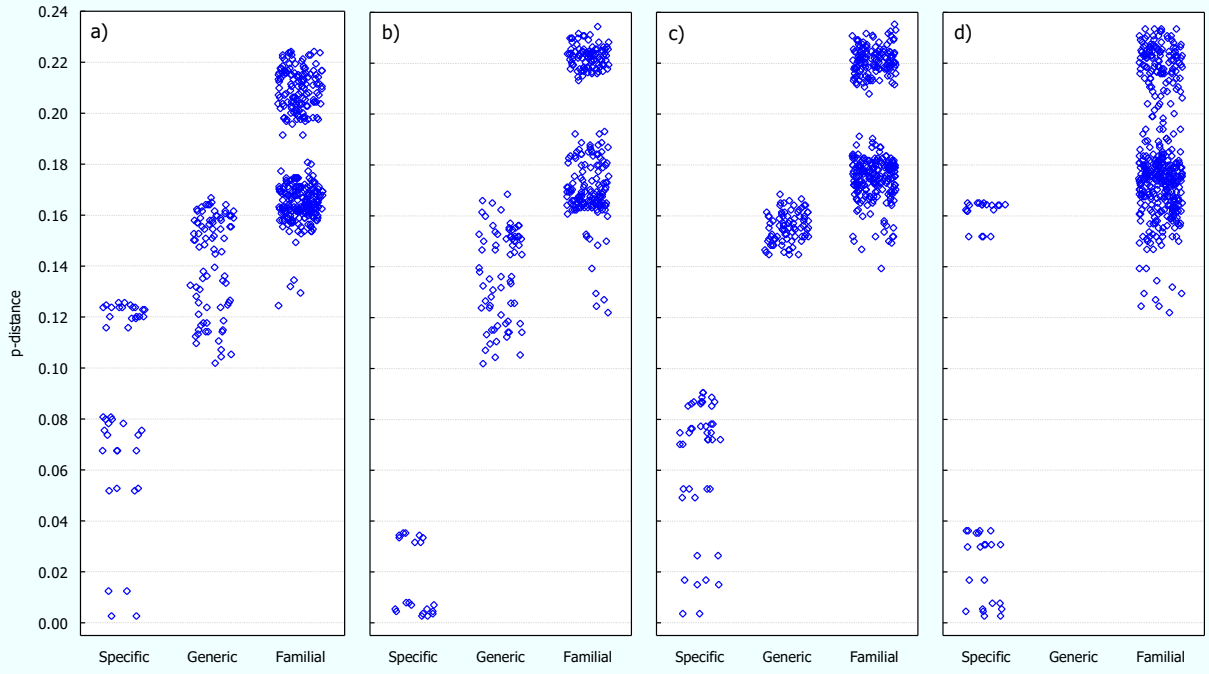


Figure 5.38 P-distances between sequences of three species of *Episoriculus*; a) *E. caudatus*, b) *E. leucops*, c) *E. macrurus* plus one species of *Soriculus*; d) *S. nigrescens*. Other details are as explained in Figure 5.26.

plus one from Nepal, which formed a group with maximum con-specific p-distances of 0.08, and a separate pair from Nepal which were at least 0.12 distant from the other group (Figure 5.38a and Figure 5.39). This was difficult to explain as the three Nepal sequences came from the same geographic location, were part of the same popset on GenBank and had virtually sequential accession numbers. Finally, the sequences of *S. nigrescens* divided into two quite distinct branches with a mean p-distance between them of 0.16 (Figure 5.38d and Figure 5.39). The cluster of two sequences came from Nepal and the other five from a single popset from China. The mean genetic distance between them approximately 0.16. Overall, using a con-specific threshold of 0.095, these sequences fell into six groupings, two *E. caudatus*, two *S. nigrescens* and one each of the other two species.

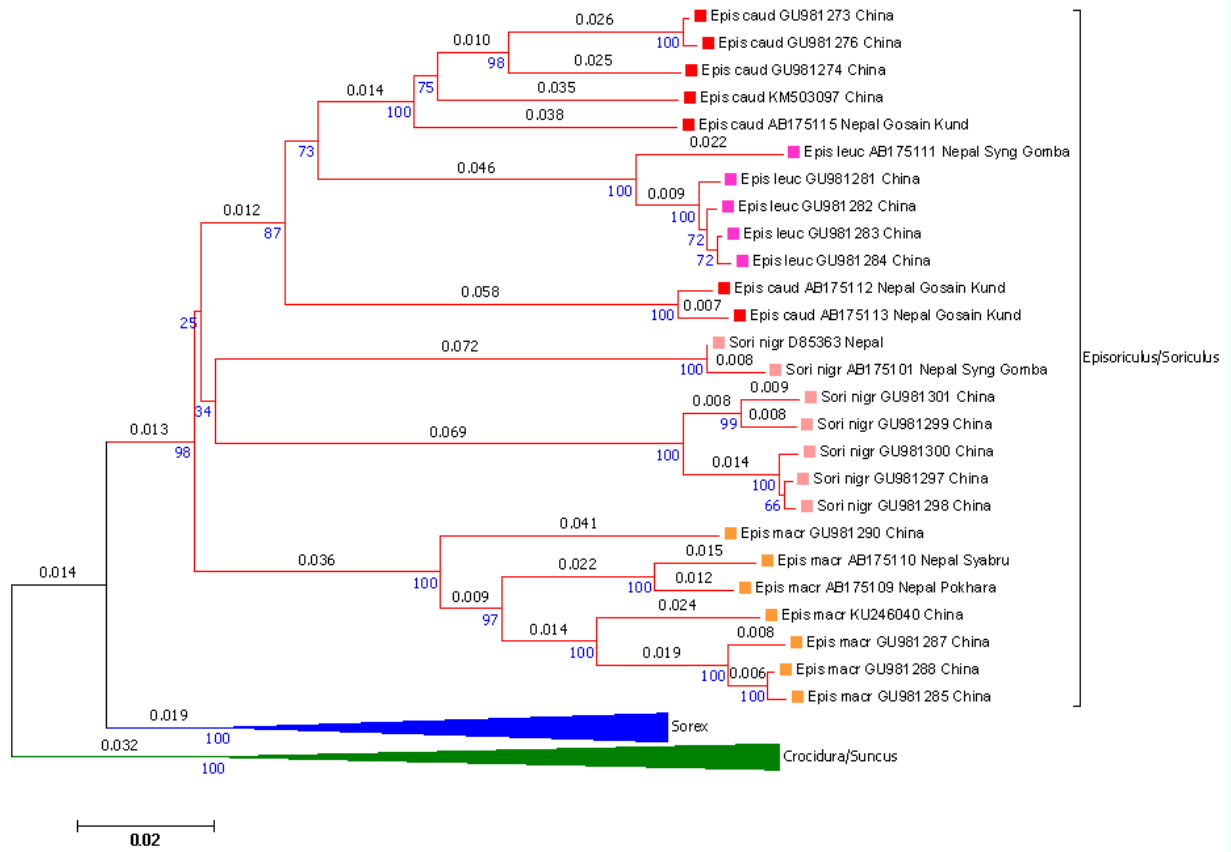


Figure 5.39 The phylogeny of 19 sequences of *Episoriculus* belonging to three (colour-coded) species and one species of *Soriculus*. Other details are explained in Figure 5.28.

5.C.d.ii Group B: Sorex

This genus comprised 21 sequences from four species. Two Eurasian species (*S. araneus* and *S. minutus*) segregated distinctly (Figure 5.40a & d respectively). The other two species showed a very confused pattern of con-specific and con-generic p-distances (Figure 5.40 b & c) which is reflected in the bootstrap values from their phylogenetic analysis (Figure 5.41). Two of the sequences clustered tightly together, came from Nepal, and both were identified as *S. cylindricauda*. However, they fell within a cluster of *S. bedfordiae* from China and, given that Smith and Xie (2008) declare that *S.*

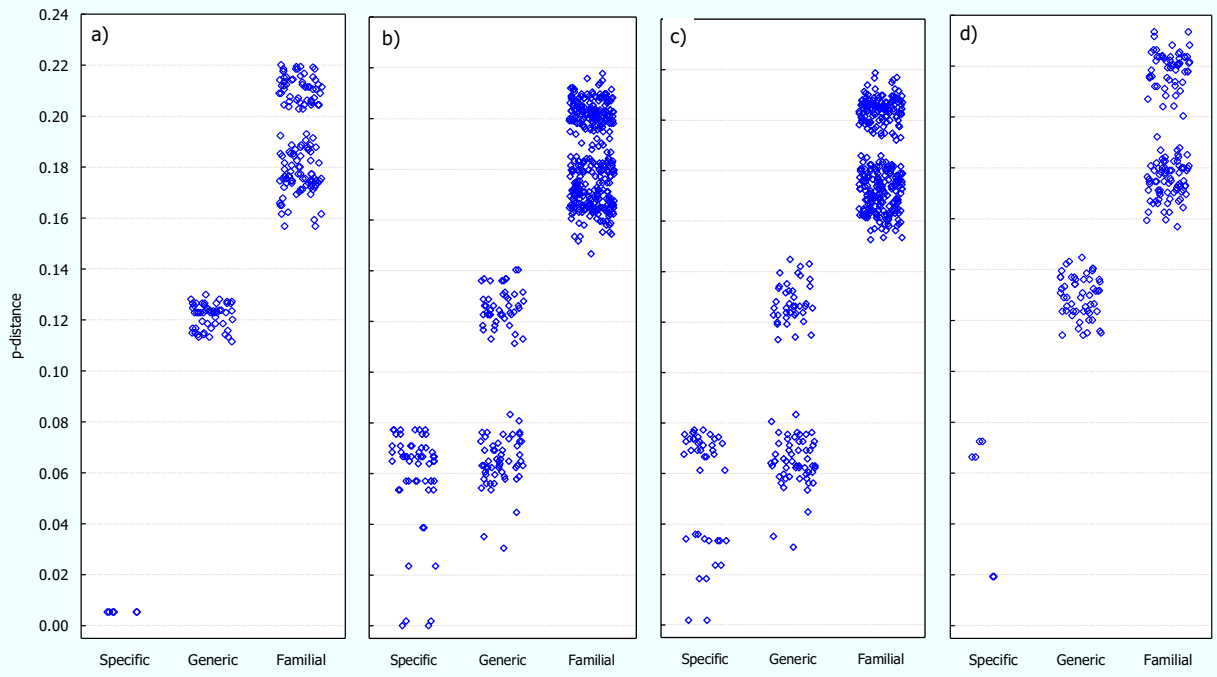


Figure 5.40 P-distances between sequences of four species of *Sorex*; a) *S. araneus*, b) *S. bedfordiae*, c) *S. cylindricauda* and d) *S. minutus*. Other details are as explained in Figure 5.26.

cylindricauda is endemic to China, there is a clear case for treating these two sequences as mis-identifications. The converse situation occurred with two sequences identified as *S. bedfordiae* which fell into a distinct cluster of sequences of *S. cylindricauda*, all from China.

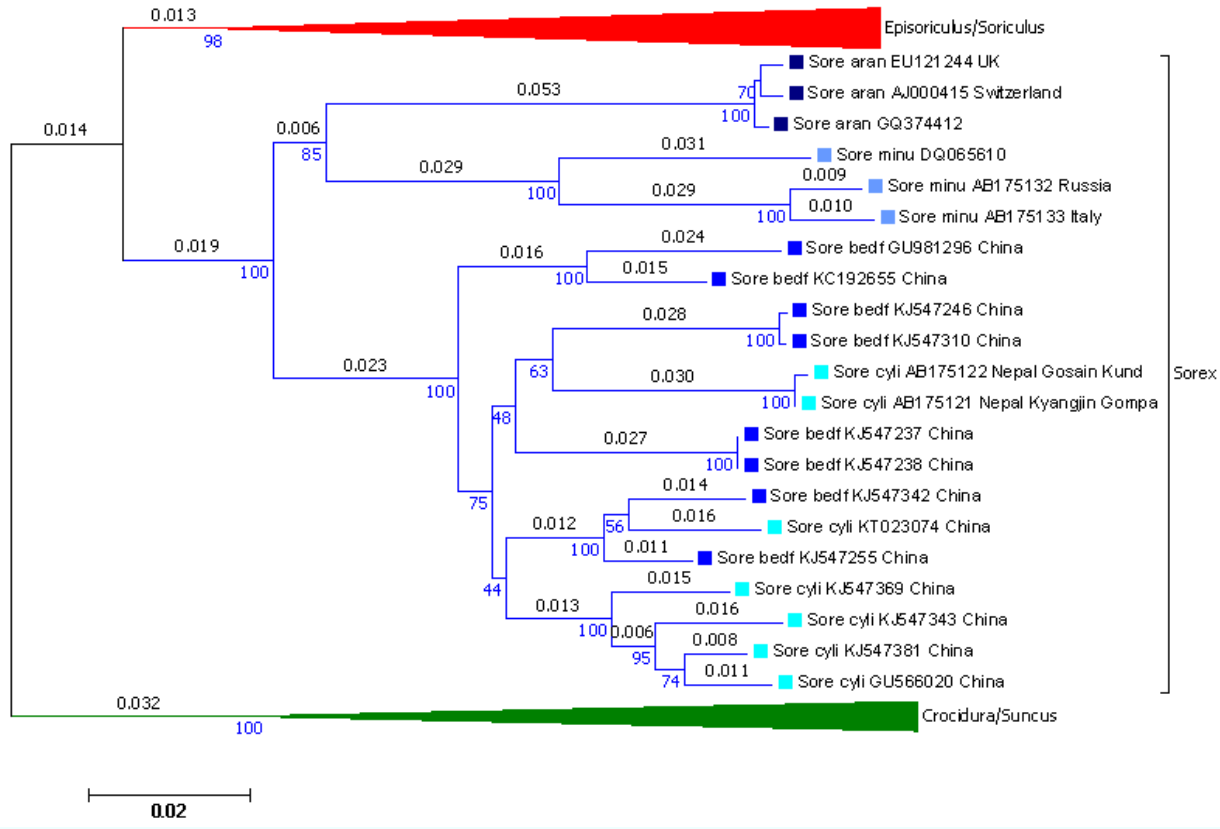


Figure 5.41 The phylogeny of 21 sequences of *Sorex* belonging to four (colour-coded) species. Other details are explained in Figure 5.28

Finally, two further sequences of *S. bedfordiae* from China fell outside the whole of this cluster, with a p-distance of approx. 0.08. I made a subjective decision to take three sequences from this cluster for subsequent analysis, including one from Nepal (AB175122).

5.C.d.iii Group C: Crocidura/Suncus

These two genera comprised 19 sequences belonging to six species. The two genera segregated distinctly and, given that the only species of interest from the current study was *Suncus murinus*, *Crocidura* was not included in the subsequent analysis. *S. etruscus* segregated distinctly (Figure 5.42a), with a minimum con-generic p-distance of 0.14. The remaining three species were not clearly distinguished (Figure 5.43), with minimum con-generic distances of approx. 0.65 and a maximum con-specific distance of 0.08 for *S. montanus*. The *S. murinus* cluster is quite distinct, although one sequence from Myanmar had a con-specific distance of approx. 0.06, which I have taken as the p-distance threshold.

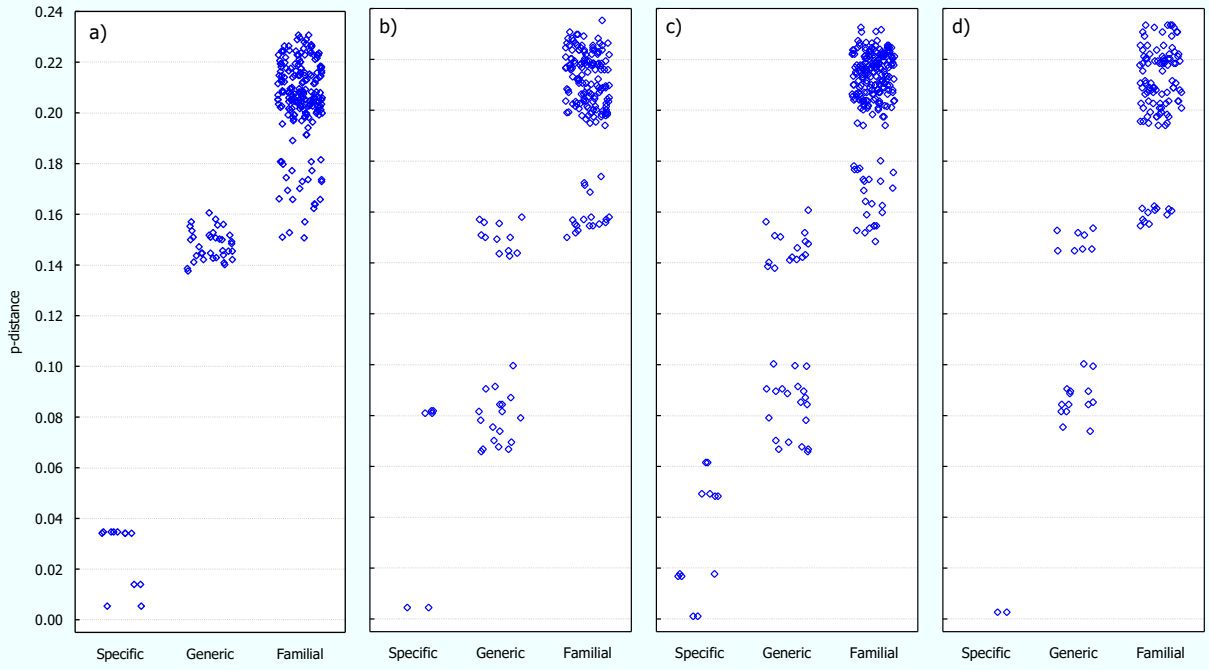


Figure 5.42 P-distances between sequences of four species of *Suncus*; a) *S. etruscus*, b) *S. montanus*, c) *S. murinus* and d) *S. stoliczkanus*. Other details are as explained in Figure 5.26

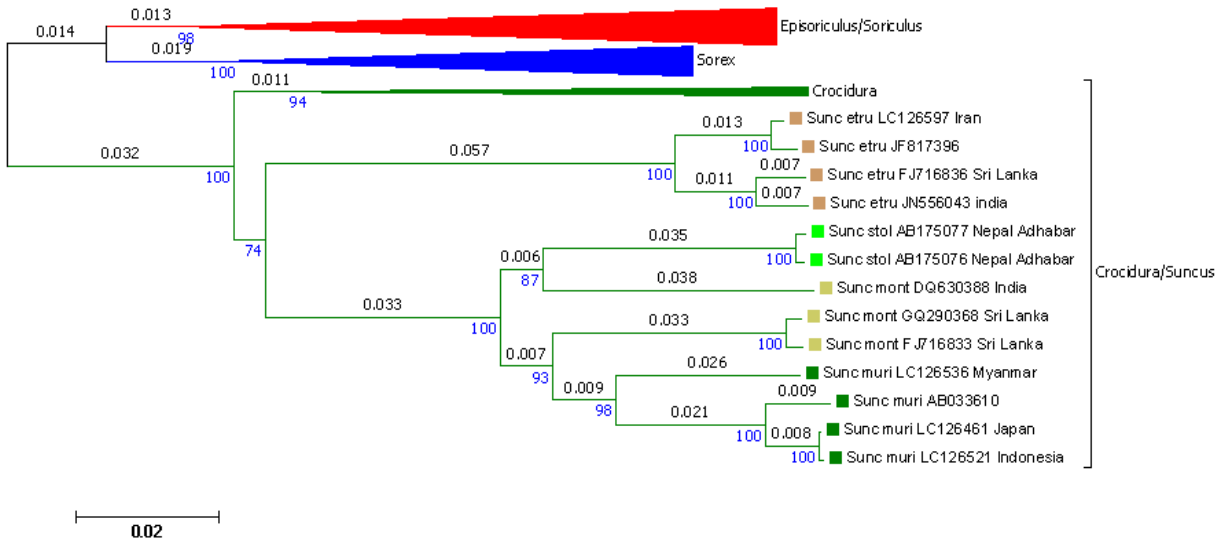


Figure 5.43 The phylogeny of 13 sequences of *Suncus* belonging to four (colour-coded) species. Other details are explained in Figure 5.28

Chapter 6 The Distribution, Abundance and Community Structure of Small Mammal Species in the Nepali Himalaya

Abstract

Altitudinal variation in small mammal communities has been studied in many parts of the world, such as the Rocky Mountains in USA, the Andes of South America, the highest isolated peaks in Africa, a number of locations in China and several mountains in Malaysia and the Philippines. However, no specific elevational studies have been carried out in the Himalaya of Nepal – the region with the largest altitudinal gradients in the world. In this final chapter, I bring together the results of the previous four data chapters to address this shortfall. The original field identifications of 18 putative taxa from Chapter 2 are combined with the probabilistic distributions from the predictive modelling in Chapter 4 and the phylogenetic analyses from Chapter 5. Together, these are used to consolidate the species identities into 17 taxa, including two possible cryptic species of vole not previously described. I then analyse the geographical and altitudinal variation in overall abundance of small mammals and show a significant decline with increasing elevation. Over 99% of the animals caught during the study were either murids or soricids and I show that there was a significant difference in the proportions of these two families found in the four different transects. The analysis of individual species shows highly significant altitudinal zonation and there may be evidence for competitive exclusion between two species of mouse; Apodemus gurkha *and* Mus musculus. *By incorporating the habitat data derived in Chapter 3, I show that several species have significant associations with the habitat clusters, especially proximity to certain features and botanical composition. Finally, I derive two measures of species diversity at different resolutions. Species richness at the trapping grid level (α–diversity) shows a significant decline with altitude, although the pattern is not consistent across all transects. No evidence for the frequently described mid-elevational "hump" was found. Species turnover between sites (β–diversity) shows a very strong association with altitudinal distance measures, indicating that the species composition of low altitude and high altitude sites is very different. There is also a significant association with geographic distance, reflecting different small mammal communities in the central and eastern regions of Nepal.*

Contents

Tables

Figures

6.1 Introduction

Over the last 50 years, the majority of field work on small mammals in Nepal has focused on collection expeditions (Abe, 1971, 1977; Martens and Niethammer, 1972; Mekada *et al.*, 2001; Mitchell and Punzo, 1976; Mitchell, 1977; Niethammer and Martens, 1975). Furthermore, none of these employed a systematic survey design with standardised field techniques, allowing robust statistical analysis. In the previous four chapters I have shown that the geographical and temporal scope of this study makes it the largest ecological small mammal survey ever undertaken in Nepal. Its systematic survey design, with spatial and temporal replication, utilising altitudinal transects with standardised trapping grids makes this study unique. The field work was undertaken over three years on four transects at altitudes ranging from 1300m to 4400m. A total of 147 visits were made to 126 trapping grids, comprising 8046 trap events. This yielded 890 captures (11.1% capture rate) of 792 individual small mammals. A detailed analysis of these trapping data, plus many methodological issues, such as capture locations and recapture rates, formed the basis for Chapter 2.

There have been more extensive surveys of the flora of Nepal, mainly over the last 10 years (Bhagat and Shrestha, 2010; Chaudhary *et al.*, 2016). Several studies have been made of specific regions of Nepal (Paudel *et al.*, 2010; Sapkota *et al.*, 2017), some of which coincided with my own geographic areas, such as Sagarmatha National Park (Bhattarai and Upadhyay, 2013). During my fieldwork, I recorded a total of three topographical, 13 habitat and 81 binary botanical variables for all trapping grids. In Chapter 3, I used novel computer intensive techniques to generate and describe derived variables which defined the major axes of environmental variation. These were then used to characterise the grids and analyse the geographical and elevational distribution of different habitats.

Museum collections of small mammals have been recognised as an extremely valuable source of data to explore ecological themes (Holmes *et al.*, 2016; Pergams and Lawler, 2009) and evolutionary changes (Moritz *et al.*, 2008). As part of my collaboration with organisations in Nepal, India and the UK, I built a photographic reference guide to many of the species likely to be encountered during the survey. In addition, the Bombay Natural History Society (BNHS) allowed me access to their database of small mammal specimens (Bajaru, *pers. comm.*). From this, and the eloquent review of small mammals in Nepal by Pearch (2011), I collated and geo-referenced over 8000 records of small mammals collected throughout the Indian sub-continent during most of the 20th century (Chapter 4). I used these data to develop predictive models of the likely distribution of 54 species within the 1529 0.1° grid squares in Nepal, and the probability of occurrence in the 10 squares covered during the current study.

Very few studies have explored the conservation genetics of small mammals in Nepal. Those that have, tended to focus on single species or sub-species identification (Adhikari *et al.*, 2018a; Adhikari *et al.*, 2018b) and there have been some recent studies on karyotypes of three shrew species in Nepal (Motokawa *et al.*, 2008). There have been more genetic studies of small mammals in neighbouring countries, such as China (Chen *et al.*, 2011; Ge *et al.*, 2017; Zhang *et al.*, 2016) and Myanmar (Shimada *et al.*, 2010). From the 792 individual small mammals in my study, I collected tissue samples from 720 animals (90.9%). In Kathmandu, I carried out DNA extraction from 501 of these, which yielded 115 successful PCR products, from which 94 good quality sequences of the cytochrome-b gene were obtained. In Chapter 5, I carried out a phylogenetic analysis on these sequences in combination with 200 sequences from GenBank. This provided strong evidence to confirm or amend the identities of 784 specimens.

In this final chapter, I combine these four sources of information. Firstly, this allows the consolidation of species identities, using both the predictive modelling of the BNHS and Pearch datasets and the phylogenetic analyses. Secondly, using these final identifications, I analyse the overall abundance and the abundance of selected taxa against geographical and topographical variables. In particular, I introduce the concept of distribution along altitudinal gradients. There are surprisingly few examples where overall abundance of small mammals has been explored along elevational gradients, as most studies have focused on community structure and composition. However, studies in China (Li *et al.*, 2003; Wen *et al.*, 2018), the Philippines (Heaney, 2001) Australia (Bateman *et al.*, 2010) and Tanzania (Mulungu *et al.*, 2008) have found various relationships between abundance and altitude, including monotonic increases and decreases, and non-linear responses. Thirdly, I analyse the effects of habitat variables on the distribution and abundance of selected species. Finally, I compile the species identities into measures of alpha and beta species diversity. In contrast to abundance, there is a vast literature on the relationship between altitude and communities of, *e.g.*, plants, several of them in Nepal (Bhattarai *et al.*, 2004; Carpenter, 2005; Vetaas and Grytnes, 2002). Similarly, in small mammals, α and β–diversity has been explored extensively, both with empirical studies (Liu *et al.*, 2008; Lomolino, 2001; McCain, 2005; Rowe *et al.*, 2015; Shuai *et al.*, 2017) and metanalyses including museum records at a continental or global scale (McCain and Beck, 2016; Rickart, 2001; Rowe, 2009). Central to this field of study is the concept of the "Mid-domain Effect" (MDE) (Colwell and Hurtt, 1994; Colwell and Lees, 2000; Colwell *et al.*, 2004; Rahbek, 1995) and its application to small mammal communities (McCain, 2004; Rowe *et al.*, 2015; Wen *et al.*, 2018; Wu *et al.*, 2013). I explore this phenomenon with my own data and compare and contrast the results with existing studies.

6.2 Methods

6.2.1 Analytical Methods for Overall Abundance and Abundance by Taxa

Data for the analyses in this chapter were drawn from previous chapters, where details of their storage, manipulation and analysis will be found. For the statistical analyses in this chapter, data were extracted from the MS Access and SQL Server databases using SQL queries. Interim data manipulation was carried out in MS Excel.

I carried out two preliminary analyses to account for methodological effects;

- Variation in trap exposure, due to void traps and incomplete or additional trapping sessions,
- Annual variation in numbers of animals caught in the trapping grids that were visited twice.

Full details of these analysis are given in Supplementary Information 6.A and 6.B. In summary, I calculated Corrected Trap Hours (CTH) to take account of variation in trap exposure and used the tallies of numbers of animals across the two years as the response variable for the subsequent analyses.

All statistical analyses in this section were based on generalised linear mixed models (GLMM), using the GENLINMIXED procedure in SPSS v25 Anon. (2017a). All models used the 126 trapping grids as cases, with the abundance of animals in each grid (N) as the response variable. Models used either a Poisson or negative binomial error term with a log link function. Due to unbalanced designs, I used Satterthwaite approximation to calculate degrees-of-freedom and robust covariances to test effects and coefficients. All models used CTH (see above and Sub-section 6.A) as an offset to account for variable trap exposure. All models included transect as a fixed, four-level categorical predictor and altitude as a covariate (Table 6.1). Supplementary models included topographical variables and habitat-based clusters as fixed factors. Models which included taxa treated them as repeated-measures factors, with a diagonal covariance matrix.

Table 6.1 Independent predictor variables used in the analysis of abundance.

Name	Description	Type	Values
Transect	Four altitudinal transects selected as part of the survey design	Categorical	Pipar, Annapurna, Langtang & Sagarmatha
Altitude	Altitude in metres of each trapping grid	Continuous	1297m – 4388m
Slope	Slope in degrees of each trapping grid	Continuous	0° – 39°
Aspect	Aspect in octants of each trapping grid	Categorical	N NE E SE S SW W NW flat
Structural	Clusters of structural variables, such as canopy height or ground cover.	Categorical	10 clusters
Proximity	Clusters based on proximity variables such as distance to human habitation or woodland	Categorical	8 clusters
Botanical	Clusters based on presence of 82 botanical taxa.	Categorical	7 clusters

6.2.2 Analytical Methods for Habitat Cluster Variables

To analyse the influence of structural, proximity and botanical clusters (from Chapter 3) on individual species I used cross-tabulations of the numbers of grids in which a species was recorded by the number of grids in each cluster. The rationale for excluding the design variables (transect and altitude) was that, although they had already been shown to have a significant effect on abundance, analysing the habitat variables without them would reveal a functional effect on small mammal numbers. I attempted to build both logistic and log-linear models to analyse these predictor variables together, but none of them would resolve due to multi-collinearity or insufficient degrees-of-freedom. Instead, I used contingency χ^2 analyses to test for associations in the 2 × n tables. To allow for multiple testing, I set a conservative *a priori* α-value of 0.005, to reduce the risk of Type I errors. These analyses and all graphic production were undertaken in Statistica v13 Anon. (2017b).

6.2.3 Analytical Methods for Species Diversity

Alpha-diversity was calculated at the trapping grid level, defined as the number of species recorded in each grid. As described for animal abundance (Supplementary Information 6.A), trap exposure theoretically has an effect on α–diversity, which is monotonically asymptotic (Stanley and Hutterer, 2007; Wu *et al.*, 2013). In the current study, however, variation in trap exposure resulting from the three factors discussed in S. I. 6.A had little effect on α–diversity, which only ranged from zero to seven. Firstly, the amount of variation caused by void-traps was very small (2.9%). By chance, this would have resulted in the under-recording of three species records. Secondly, the missing sessions (in Annapurna, Sites 4 & 5), may also have resulted in missing species records. To assess the size of this effect, I calculated species richness for each grid after one session and also after three sessions for the 138 trapping grids that had all three. By applying these ratios to the nine grids that only had one dawn session, the most likely number of species records that were missed in the truncated grids was four. Finally, the additional trapping sessions in Pipar Site 4, resulted in a single extra species being recorded in one trapping grid. In summary, I contend that, overall, the variation in trap exposure resulted in only six species records being missed, out of 331 (1.8%). This level of noise in the data was small compared to the stochastic effects in live trapping, so I used uncorrected α–diversity measures in all subsequent analyses.

To assess inter-annual effects, I calculated the α–diversity obtained in the 21 trapping grids in Pipar visited in both 2013 and 2014. In 16 grids α–diversity was higher in 2014 or equal to 2013, five of which recorded two additional species. Selecting the highest α–diversity from each grid, or calculating α–diversity across both years, would have doubled the degree of trap exposure used to obtain

these values. Instead, I chose to drop the 2013 cohort in its entirety, which still provided α–diversity values for all 126 trapping grids.

I analysed α–diversity in two stages using generalized linear models with a log link function and Poisson error terms (Bateman *et al.*, 2010). Firstly, I used the best-subsets procedure of the GLZ module in Statistica v13 (Anon., 2017b) on models including the two main effects (altitude and transect) and their interaction, plus 2nd and 3rd order polynomials of altitude and their interactions with transect, following the approaches of Novillo and Ojeda (2012) and Rowe and Lidgard (2009). I used the Akaike Information Criterion (AIC) for selection of the best model following the approach of, *e.g.*, Chen *et al.* (2017). Secondly, I used the best-subsets procedure on augmented models using slope as a continuous predictor, aspect and the three cluster variables derived from Chapter 3 (structural, proximity and botanical) as categorical predictors, with no interactions.

Beta diversity was used as a measure of species turn-over and was calculated at the site level. I used the formula of Whittaker (1960) as re-expressed by Koleff *et al.* (2003);

$$\beta_w = \frac{a+b+c}{(2a+b+c)/2} - 1 \qquad \text{Equation 6.1}$$

where a = the number of species recorded in both sites, b = the number of species recorded in the comparative site but not in the focal site and c = the number of species recorded in the focal site but not the comparative site. This index ranges from a value of zero when all species are recorded in both sites, i.e. they have identical species components, to one when no species are common to both sites. This makes β_w an index of diversity or dissimilarity and was recommended on the basis of four "good" criteria by Wilson and Shmida (1984).

I explored two other β indices; the Sorenson Index β_{Sor} and the Simpson Index β_S, as formulated in Koleff *et al.* (2003). The former is algebraically the same as β_w but the scale is reversed, making it an index of similarity. (So $\beta_{Sor} = 1 - \beta_w$.) In contrast, β_S is a dissimilarity index, also scoring from 0 to 1, but it has a fundamentally different algebraic formulation. It is less sensitive than β_w, always scoring zero when either b or c are zero, whereas β_w will be greater than zero in that situation. I used β–diversity in two ways.

- Firstly, to investigate the relationship between diversity and altitude, I created two distance matrices, one for β–diversity and one for altitudinal distances. I used all pair-wise comparisons of sites – but only within transects, thereby ignoring any β–diversity between transects. Using all combinations of Site A / Site B, this resulted in a triangular matrix of 50 cells; 2 transects ×

Combin(6, 2) + 2 transects × *Combin*(5, 2). I used a Mantel test (Manley, 1986) of association between the two distance matrices (as employed by Krystufek *et al.* (2010) and (Wen *et al.*, 2014)), using 99,999 permutations, calculating the sums-of-cross-products to evaluate the significance. This test was implemented in VBA using the derived data held in the MS Access database.

- Secondly, to explore β–diversity between transects, I coded the transects with approximate values representing their linear distances in kilometres from west to east; Annapurna 0, Pipar 20, Langtang 200, Sagarmatha 400. So the geographic distance matrix took discrete values; 0, 20, 180, 200, 380 & 400. To reduce the effect of β–diversity caused by altitudinal distance, I only included pair-wise comparisons between sites with an altitudinal distance < 1000m. It was impossible to remove the altitudinal effect entirely, because no two sites had the same altitude. This resulted in triangular matrices of 130 pair-wise distances. I used the same Mantel test for evaluating the significance of this association.

6.3 Results

6.3.1 Species Identities

This section brings together results from Chapter 2 (field identifications), Chapter 4 (predictions of distributions from the BNHS collections and the Pearch (2011) monograph) and Chapter 5 (genetic phylogenies). Here I make the final assignment of species identities. The decision-making process is presented schematically in Figure 6.1 and the final species identities shown in Table 6.2.

I consider that 11 of the 13 species identified in the field were correctly identified, being strongly supported by the predictive modelling and the genetic phylogenies. In particular, *Apodemus gurkha*, *Episoriculus caudatus* and *Soriculus nigrescens*, were confidently predicted from all three sources. Seven of the other eight species had strong support from one or both sources.

Two species were recorded in the field whose identities could not be confirmed; *Mus booduga* and *Niviventer niviventer*. In the case of *M. booduga*, there was unequivocal evidence from the predictive distributions and phylogenies that these specimens were *M. musculus*. However only four specimens were successfully sequenced, all from one transect (Sagarmatha), so there is still a possibility that *M. booduga* was present, especially in the Pipar and Annapurna transects which had low altitude sites.

Two other species that I have included in the final list were not recorded in the field; *Rattus andamanensis* and *R. pyctoris*. There was good phylogenetic evidence for the former, although there were no specimens in the BNHS dataset or recorded in the Pearch monograph. The identification of *R. pyctoris* was strongly supported by the predictive distributions and, especially, the genetic phylogenies.

The final list of putative species includes the two cryptic species of *Neodon* which the genetic analysis clearly revealed. I have also included two specimens of currently unidentified murid and six unidentified soricids. This resulted in a minimum of 17 putative species (Table 6.2), with confident identifications for 779 (98.4%) specimens.

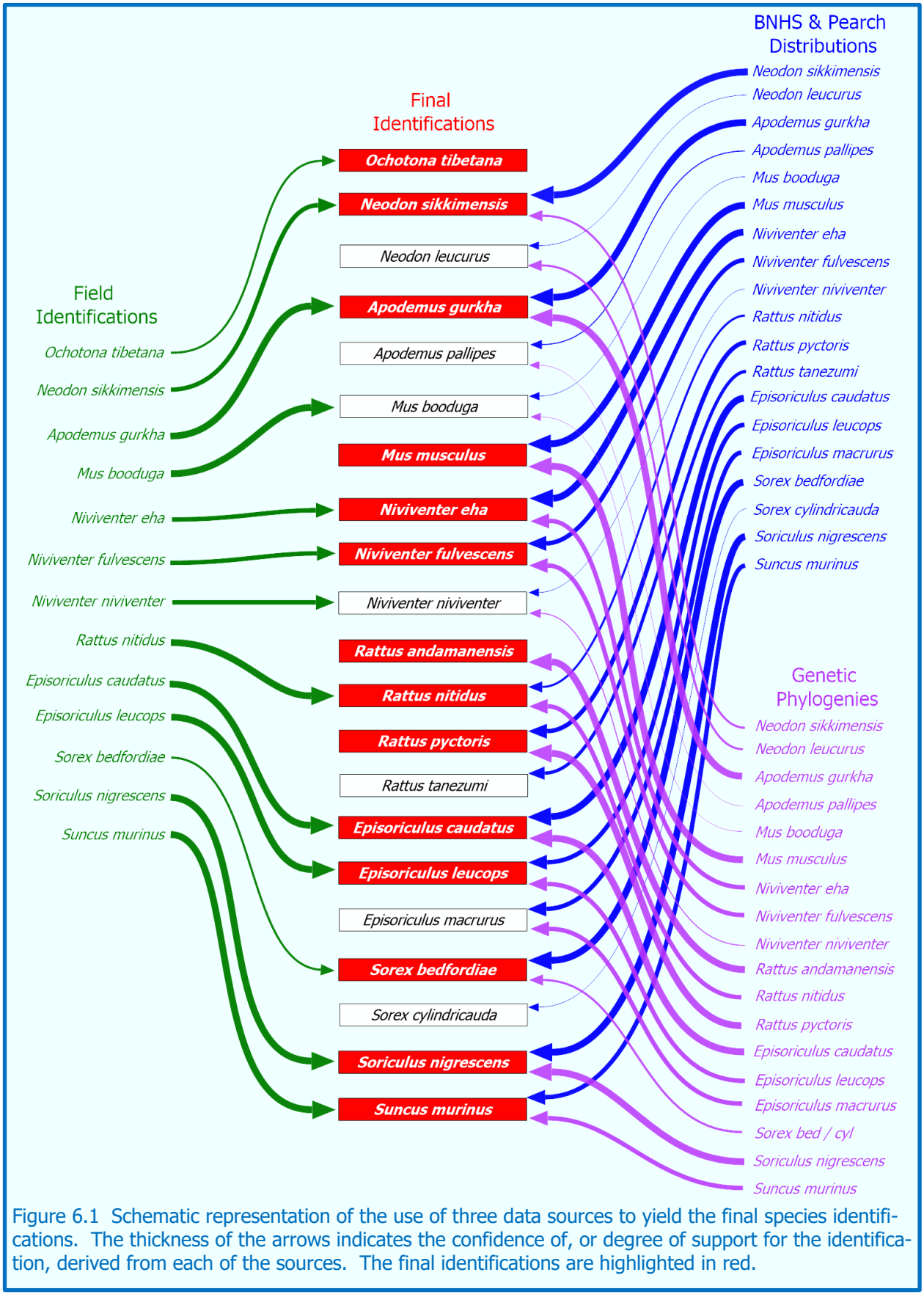


Figure 6.1 Schematic representation of the use of three data sources to yield the final species identifications. The thickness of the arrows indicates the confidence of, or degree of support for the identification, derived from each of the sources. The final identifications are highlighted in red.

Table 6.2 Final identities of the 17 putative species, with numbers of specimens.

Order	Family	Genus	Species	N	
Lagomorpha	Ochotonidae	*Ochotona*	*O tibetanus*	1	1
Rodentia	Cricetidae	*Neodon*	*N sikimensis A*	3	5
			N sikimensis B	2	
	Muridae	Murid	Murid sp.	2	305
		Apodemus	*A gurkha*	41	
		Mus	*M musculus*	209	
		Niviventer	*N eha*	13	
			N fulvescens	23	
		Rattus	*R andamanensis*	9	
			R nitidus	2	
			R pyctoris	6	
Soricomorpha	Soricidae	Soricid	Soricid sp.	6	481
		Episoriculus	*E caudatus*	241	
			E leucops	20	
		Sorex	*S bedfordiae*	6	
		Soriculus	*S nigrescens*	187	
		Suncus	*S murinus*	21	

6.3.2 The Analysis of Overall Abundance

To analyse the overall abundance of small mammals I used a GLMM with a negative binomial error term and log link function, as described in Sub-section 6.2.1. I began with a simple model incorporating transect as a fixed factor, altitude as a covariate and the interaction between these two main effects. This overall model was highly significant ($F_{(7,118)} = 11.99$, $p < 10^{-10}$), with an AIC of 682. I compared this with a model using a Poisson error term, which yielded an AIC of 770. This indicated strong over-dispersion so all subsequent models used a negative binomial error term.

All three effects in the model were highly significant (Table 6.3). The overall mean abundance was 5.79 animals per grid. Predicted means in Pipar and Sagarmatha were significantly greater than those in Annapurna and Langtang (Figure 6.2a). The effect of the altitude covariate showed a significant decline in abundance with increasing altitude. At the lowest altitude in which trapping took place (approx. 1300m), the predicted abundance was 12 animals per grid (95% C. I. approx. 8.8 – 17) declining to approx. 3 per grid (2 – 3.5) at 4400m (Figure 6.2b).

Table 6.3 Results of negative binominal GLMM of N against topographical variables. Significant effects are highlighted in red.

Effect	F	Num. DF	Denom. DF	*p*
Transect	13.35	3	118	$< 10^{-6}$
Altitude	28.76	1	118	$< 10^{-6}$
Transect × Altitude	15.21	3	118	$< 10^{-7}$
Corrected model	11.99	7	118	$< 10^{-10}$

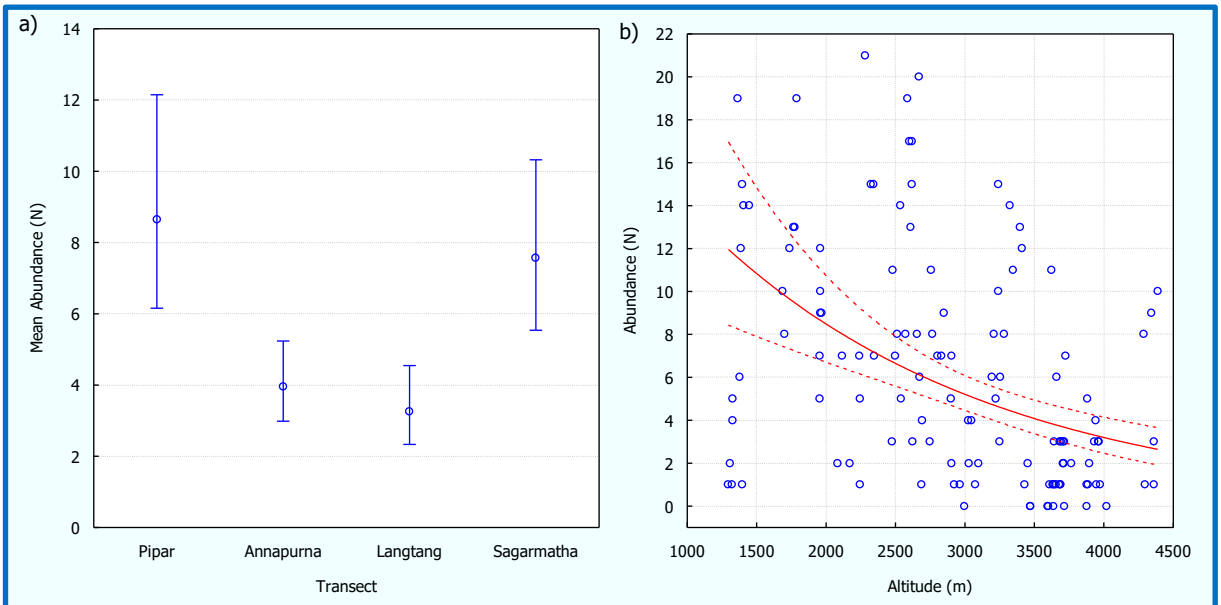


Figure 6.2 The relationships between the abundance of all small mammals per grid with a) transect and b) altitude.

However, the interaction between these two effects was also very highly significant (Figure 6.3). This showed that the negative relationship was strongest in Langtang as its interaction coefficient was negative with respect to Sagarmatha (the reference category), whereas the other two were positive. Pipar and Annapurna had virtually linear responses that were not significantly negative.

I included the main effects for slope and aspect, plus their two-way interactions with transect in a subsequent model. The AIC was considerably greater (737), and none of the additional effects were significant.

Finally, to test for the effects of structural, proximity and botanical predictor variables, I created a model including all three variables and removed the least significant in turn, testing for a decrease in AIC. This resulted in a model where the structural clusters were highly significant ($F_{(9,103)} = 2.679$, $p < 0.01$) and the botanical clusters were marginally significant ($F_{(6,103)} = 2.809$, $p < 0.02$). Two structural clusters (4 & 10) showed marginally significant deviation contrasts and Cluster 9 had a highly significant negative deviation ($t_{(102)} = -2.721$, $p < 0.01$). This cluster described woodland with a dense canopy and a moderately dense shrub layer approx. 4m in height (see Chapter 3) and had a predicted mean abundance of 4.00 animals per grid. Two botanical clusters (1 & 3) showed significant negative contrasts, with predicted mean abundances of 3.99 and 3.87 respectively. Cluster 1 dominated in the second site in Sagarmatha and described grids with scattered *Pinus* in pasture with *Artemisia* or *Pieris* present. Cluster 3 was dominant in four sites in Langtang and Sagarmatha between 3500m and 4000m, which was described as *R. campanulatum* present with bracken and *Scabious* absent.

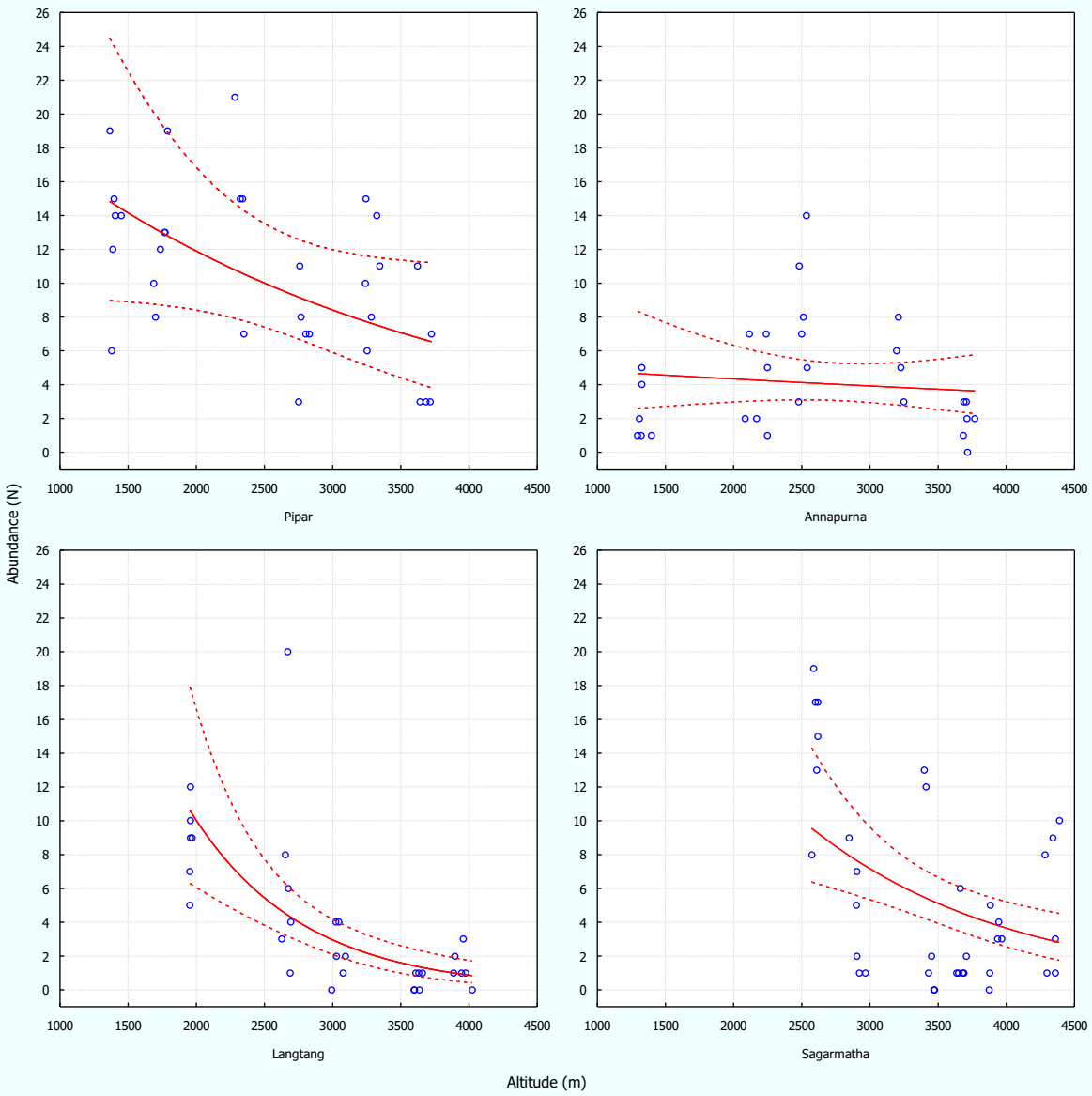


Figure 6.3 The interaction between transect and altitude on small mammal abundance. Solid lines show best fit negative binomial functions calculated separately for each transect. Dotted lines show 95% C. I. of the line.

6.3.3 The Analysis of Abundance by Taxa

6.3.3.1 Families

Four families were recorded during the survey, although two of these, Ochotonidae (pikas) and Cricetidae (voles) only had one and five individuals respectively. Voles were recorded in three of the four transects and, with the exception of one vole recorded in the fifth site in Pipar, they were only found in the highest sites in the transects (Figure 6.4). All vole records were above 3250m altitude.

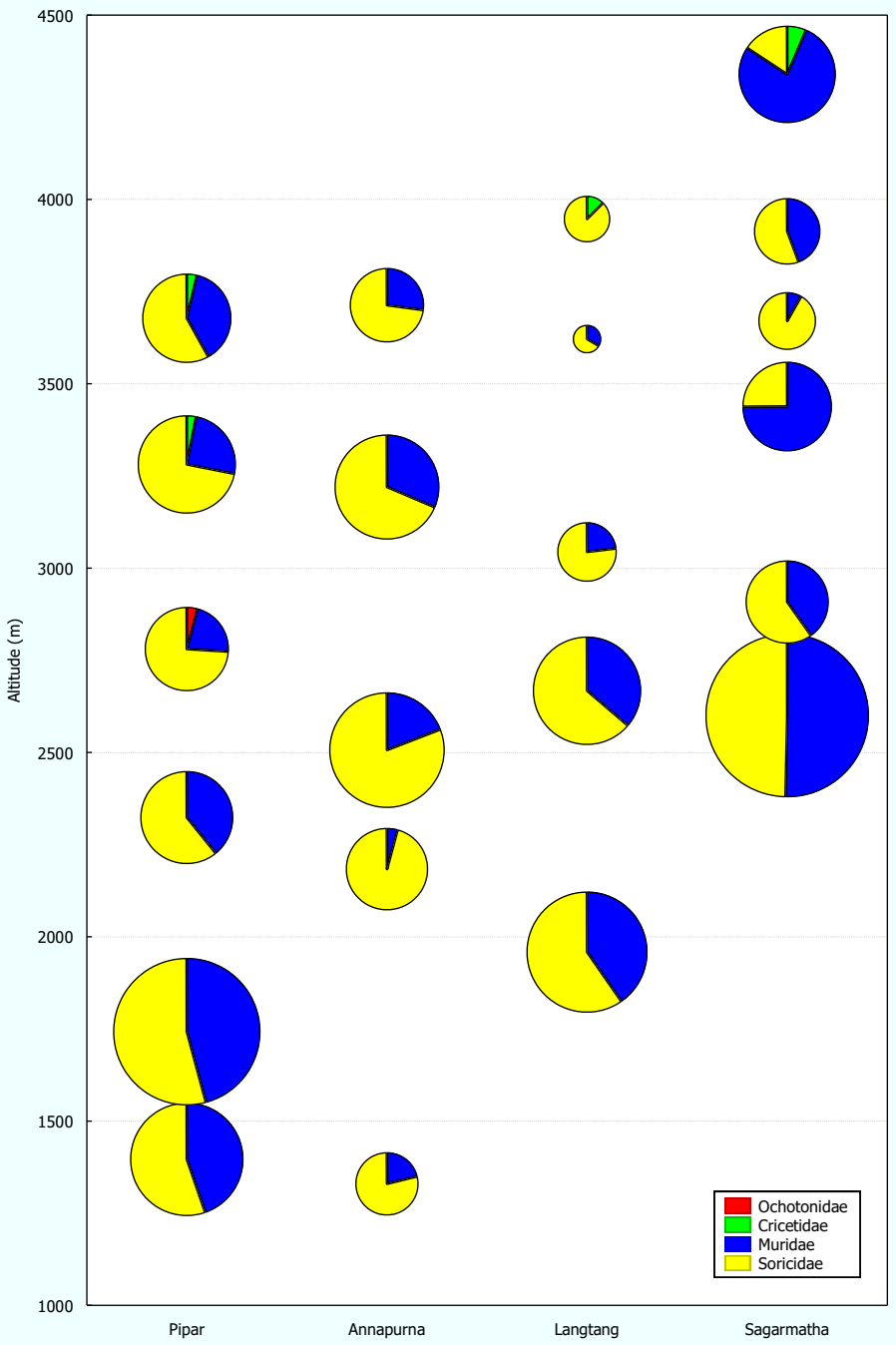


Figure 6.4 Corrected Abundance of animals for the four families in 22 sites plotted by transect and altitude. The areas of the pies are proportional to the total corrected abundance of animals within the site.

There were 305 murid records (38.5% of all animals caught). They were found in every site except one (the highest site in Langtang). Shrews constituted 60.7% (481) of all animals caught and they were ubiquitous. Furthermore, they were the dominant family in every site except two in Sagarmatha plus one other where the corrected abundance in each family was almost identical (Site 1 in Sagarmatha).

To analyse the abundances shown in Figure 6.4, I incorporated these two families as a repeated-measures factor into the GLMM used in Sub-section 6.3.2. I included all interactions between family

and the three terms already in the model, which generated two more 2-way and one 3-way interaction. A negative binomial error term and log link function were used, with robust estimators and Satterthwaite degrees-of-freedom.

The overall model was very highly significant, with all three topographical effects remaining significant (Table 6.4). The main effect of family was not significant, indicating that the observed difference in abundance between the two families was explained by higher order interactions of family with topographical effects. Firstly, the highly significant family × transect interaction indicated that there were significantly more shrews in Pipar and Annapurna but there was no difference in numbers of murids and shrews in Langtang and Sagarmatha. Secondly, the relationship between abundance and altitude varied significantly between the eight combinations of family and transect, represented by the 3-way interaction (Figure 6.5). This showed that the rates of decline in abundance with altitude in Annapurna and Langtang did not differ significantly between families. However, in Pipar there was a significant decline of murids with increasing altitude, which did not occur in shrews, and in Sagarmatha there was the opposite effect, with no decline in murids, but a highly significant decline of shrews with altitude. The third interaction (family × altitude) was not significant, indicating that there was no overall difference in the elevational distribution of the two families.

Table 6.4 Results of negative binominal GLMM of N against topographical variables and Family treated as a repeated-measure.

Effect	F	Num. DF	Denom. DF	*p*
Transect	26.81	3	236	$< 10^{-10}$
Alt	37.30	1	162	$< 10^{-8}$
Transect × Alt	22.97	3	201	$< 10^{-10}$
Family	2.09	1	178	0.150
Family × Transect	5.77	3	236	< 0.001
Family × Altitude	0.01	1	151	0.924
Family × Transect × Altitude	4.12	3	196	< 0.01
Corrected Model	11.54	15	236	0

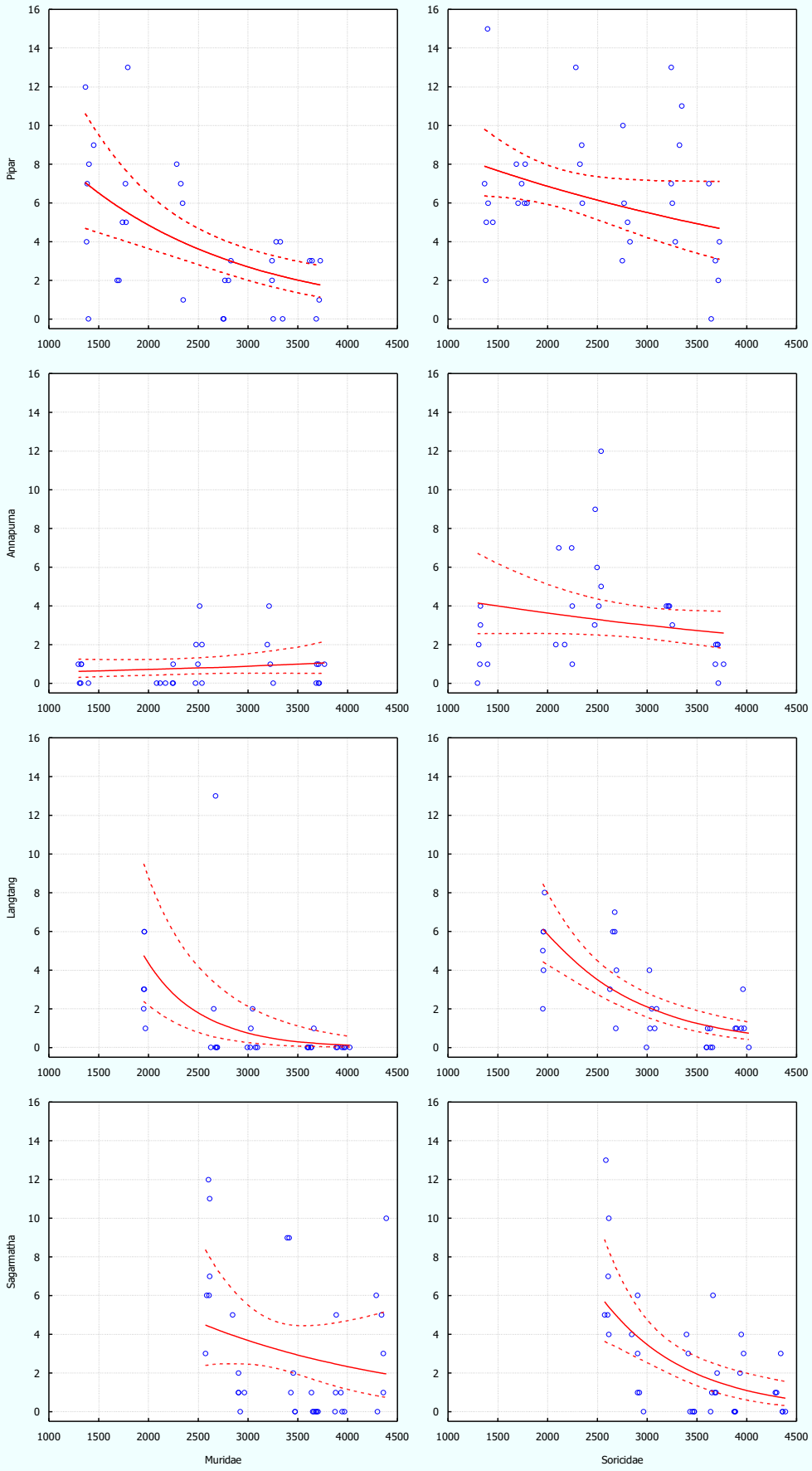


Figure 6.5 The three-way interaction between altitude (x-axis), transect and family on abundance (y-axis). Solid lines show the predicted negative binomial functions calculated separately for each transect × family combination. Dotted lines show 95% C. I. of the predicted function.

6.3.3.2 *Species of Muridae*

The geographical and altitudinal distribution of the eight murid species showed a very distinct pattern (Figure 6.6). *Mus musculus* comprised 68.5% of murid records (209 out of 305). It was fairly ubiquitous, being found in 15 of the 22 sites (68.2%), and was the dominant murid in 14 of them. It was the only murid found above 4000m, but showed an apparent preference for lower altitudes, with a median location of 2602m. It was also more dominant in Langtang and Sagarmatha. In contrast, the 41 specimens of *Apodemus gurkha* were only found in the two transects in the Annapurna region, and at altitudes above 2500m. Here, it was the dominant murid above 3000m and had a median location of 3284m.

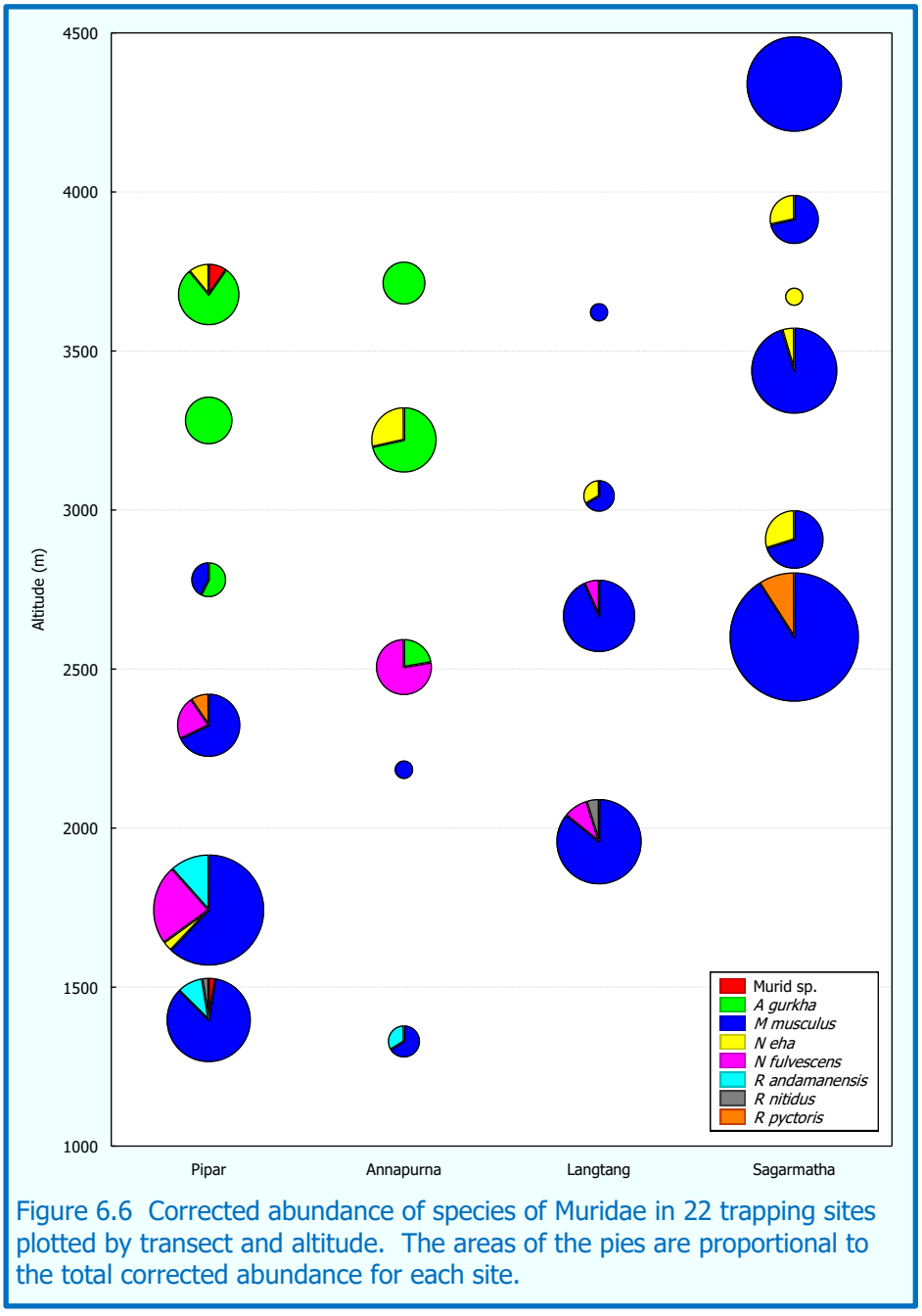


Figure 6.6 Corrected abundance of species of Muridae in 22 trapping sites plotted by transect and altitude. The areas of the pies are proportional to the total corrected abundance for each site.

I tested the significance of these patterns with the negative binomial model presented above, substituting species (only *A. gurkha* and *M. musculus*) for family. All terms that included species were very highly significant (Table 6.5). Overall, the predicted mean abundance for *A. gurkha* was approx. 1.68 animals per trapping grid , compared to 4.86 for *M. musculus*. However, the two-way interaction between species and altitude was also very highly significant (Figure 6.7). This showed a highly significant increase in predicted abundance with altitude for *A. gurkha*, whereas there was no clear altitudinal effect for *M. musculus*.

Table 6.5 Results of negative binominal GLMM of N against topographical variables and species (*A. gurkha* v *M. musculus*) treated as a repeated-measure. Effects including species are marked in red.

Effect	F	Num DF	Denom DF	*p*
Transect	12.13	3	40	$< 10^{-5}$
Alt	7.22	1	30	< 0.01
Transect × Alt	14.67	3	43	$< 10^{-5}$
Species	28.93	1	28	$< 10^{-4}$
Species × Transect	11.70	3	40	$< 10^{-4}$
Species × Altitude	27.49	1	30	$< 10^{-4}$
Species × Transect × Altitude	9.19	3	41	$< 10^{-4}$
Corrected Model	12.57	13	47	$< 10^{-10}$

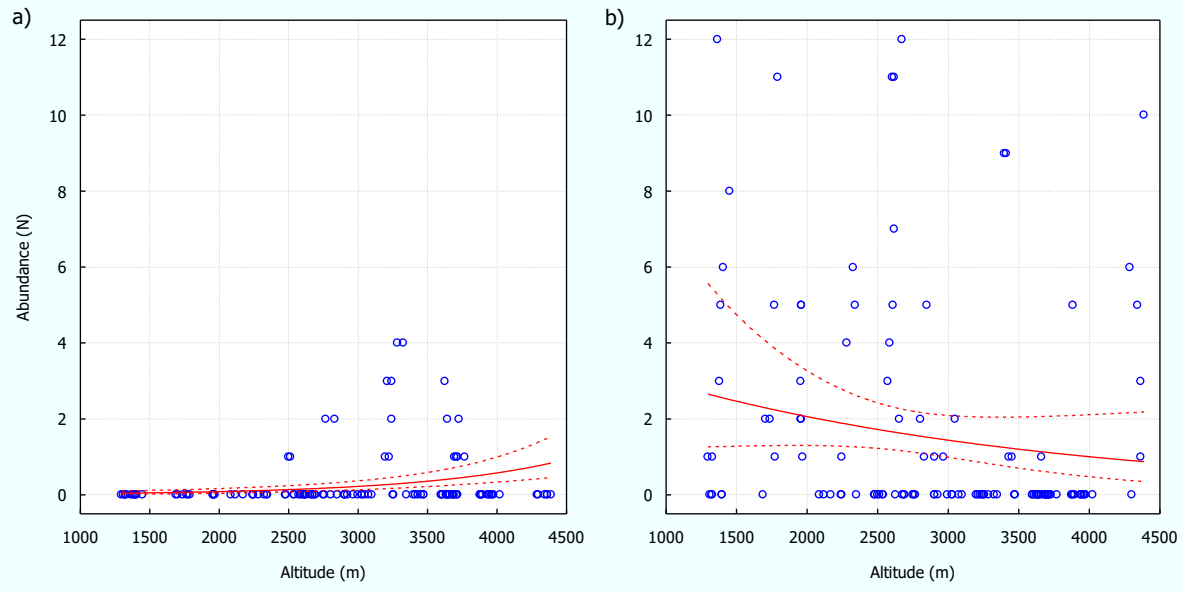


Figure 6.7 The two-way interaction between altitude and a) *A. gurkha* and b) *M. musculus* on abundance. Solid lines show the predicted negative binomial functions calculated separately for each species. Dotted lines show 95% C. I. of the lines.

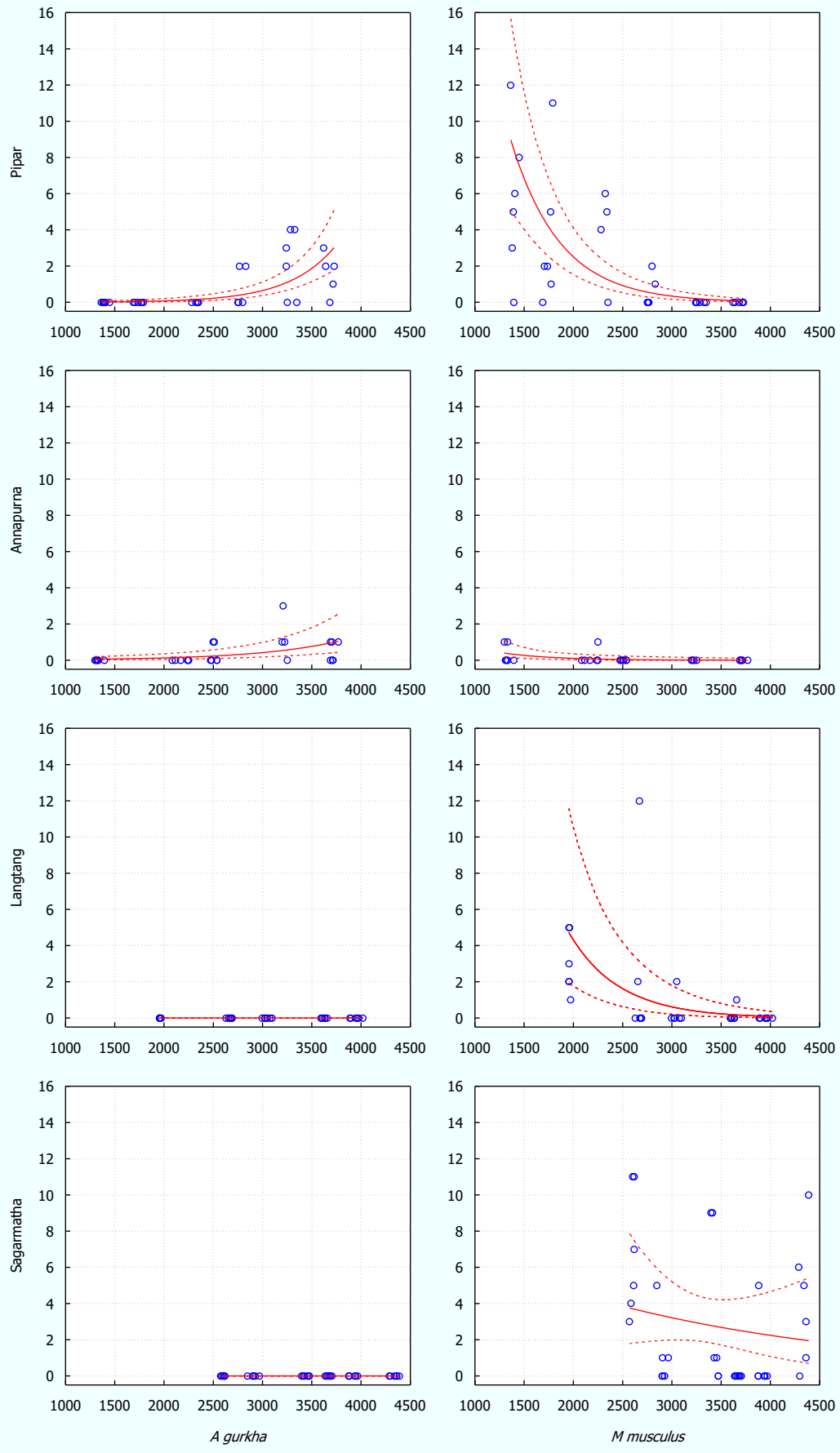


Figure 6.8 The three-way interaction between altitude (x-axis), transect and two murid species on abundance (y-axis). Solid lines show the predicted negative binomial functions calculated separately for each transect × species combination with 95% C. I. of the lines.

Most importantly, the three-way interaction (Figure 6.8) showed that *A. gurkha* had significant increases with altitude in the two transects in which it was present; Pipar and Annapurna. *M. musculus* was present in all transects at very different abundances, but only declined significantly in three of them (Pipar and Langtang, but Annapurna was marginal due to intrinsically low abundance).

The other six species of murid were captured too infrequently to warrant detailed statistical analysis, but two clear patterns emerged;

- Firstly, the two *Niviventer* species were partitioned perfectly by altitude. The 13 specimens of *N. eha* were recorded in 11 grids in all four transects, but only above 2750m. In contrast, 23 specimens of *N. fulvescens* were found in 15 grids, in three transects, but only below 2750m.
- Three species of *Rattus* were recorded from all four transects and showed strong altitudinal partitioning. Of these, nine *R. andamanensis* were found in eight grids, all below 1800m. Six *R. pyctoris* were found in four grids, all between 2400m and 2700m. The two records of *R. nitidus* were recorded in two different transects at their lowest sites.

6.3.3.3 Species of Soricidae

The 481 shrews recorded were dominated by two species (Figure 6.9); *Episoriculus caudatus* and *Soriculus nigrescens*, which together comprised 428 specimens (90%). Both species were found in all four transects. *E. caudatus* was recorded in every one of the 22 sites and was the dominant soricid in 12 of them. *S. nigrescens* was absent from three sites (13.6%) and was the dominant soricid in only four sites. They appeared to have similar elevational locations as the median altitude at which *E. caudatus* was recorded was 2654m, whereas for *S. nigrescens* it was 2497m.

To test these effects, I ran the negative binomial model with these two species as the repeated-measures factor, which gave a very highly significant model overall (Table 6.6). The two main topographical effects and their interaction were highly significant. There was an overall difference in abundance between transects, with significantly greater predicted abundance in Pipar (13.2 per grid) and significantly fewer caught in Langtang (4.15 per grid). There was a significant negative relationship with altitude, but the significant interaction with transect indicated that this was only apparent in Langtang and Sagarmatha, and marginally so in Pipar, roughly following the pattern shown the four right-hand graphs in Figure 6.5.

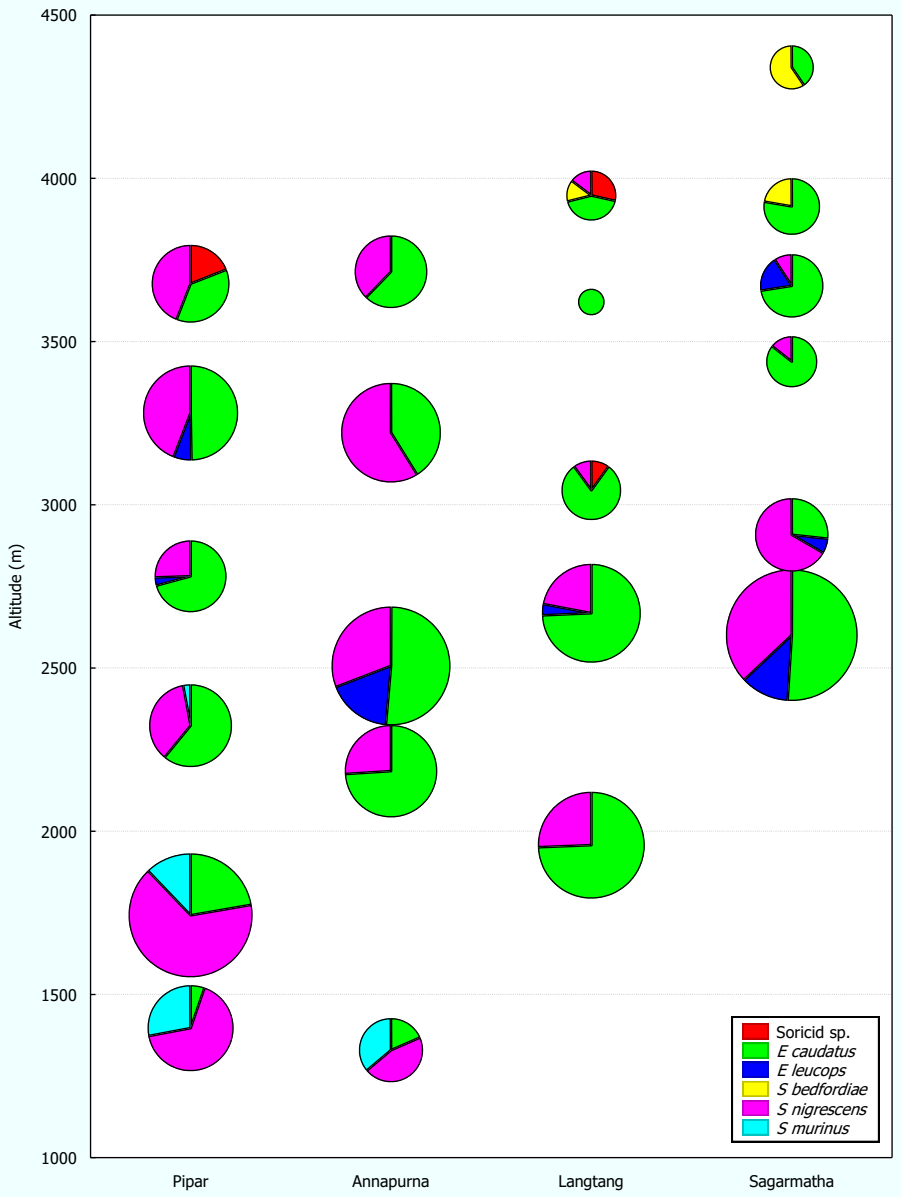


Figure 6.9 Corrected abundance of six species of Soricidae in 22 trapping sites plotted by transect and altitude. The areas of the pies are proportional to the total corrected abundance for each site.

Table 6.6 Results of negative binominal GLMM of N against topographical variables and species (*E. caudatus* v *S. nigrescens*) treated as a repeated-measure. Significant effects that included species are highlighted in red.

Effect	F	Num. DF	Denom. DF	*P*
Transect	9.28	3	60	$< 10^{-4}$
Altitude	17.51	1	72	$< 10^{-4}$
Transect × Altitude	13.74	3	72	$< 10^{-6}$
Species	0.01	1	98	0.940
Species × Transect	5.71	3	86	< 0.002
Species × Altitude	0.80	1	88	0.374
Species × Transect × Altitude	4.75	3	103	< 0.005
Corrected Model	12.56	15	143	0

The species main effect was not significant, but there was a significant interaction with transect. This indicated that, although there was no overall difference in abundance of the two species, there were significantly more *E. caudatus* recorded in Langtang, but not in the other transects. Furthermore, the significant three-way interaction (Figure 6.10) explained the pattern shown by the pie charts in Figure 6.9. There was no altitudinal relationship with abundance of *E. caudatus* in Pipar or Annapurna, but there was a strong decline in Langtang and a moderate decline in Sagarmatha. In contrast, there was a marginal decline in abundance with increasing altitude for *S. nigrescens* in Pipar and Langtang, but a very strong decline in Sagarmatha.

The remaining four shrew species revealed three very clear distributional patterns;

- The six specimens of *Sorex bedfordiae* and five of the unidentified soricid, were all found above 3600m. There was clear partitioning between transects, as the soricid was found in Pipar and Langtang, while five of the six *S. bedfordiae* were recorded in Sagarmatha.
- The 20 records of *Episoriculus leucops* were found in seven sites and in every transect, but were restricted to mid elevations between 2500m and 3700m. It was even more strongly concentrated in the zone between 2500m and 2900m, with 13 specimens (65%) being found across eight of the 23 (35%) trapping grids.
- Finally, all but one of the 21 *Suncus murinus* were found below 1750m, and all but four specimens were recorded in the Pipar transect.

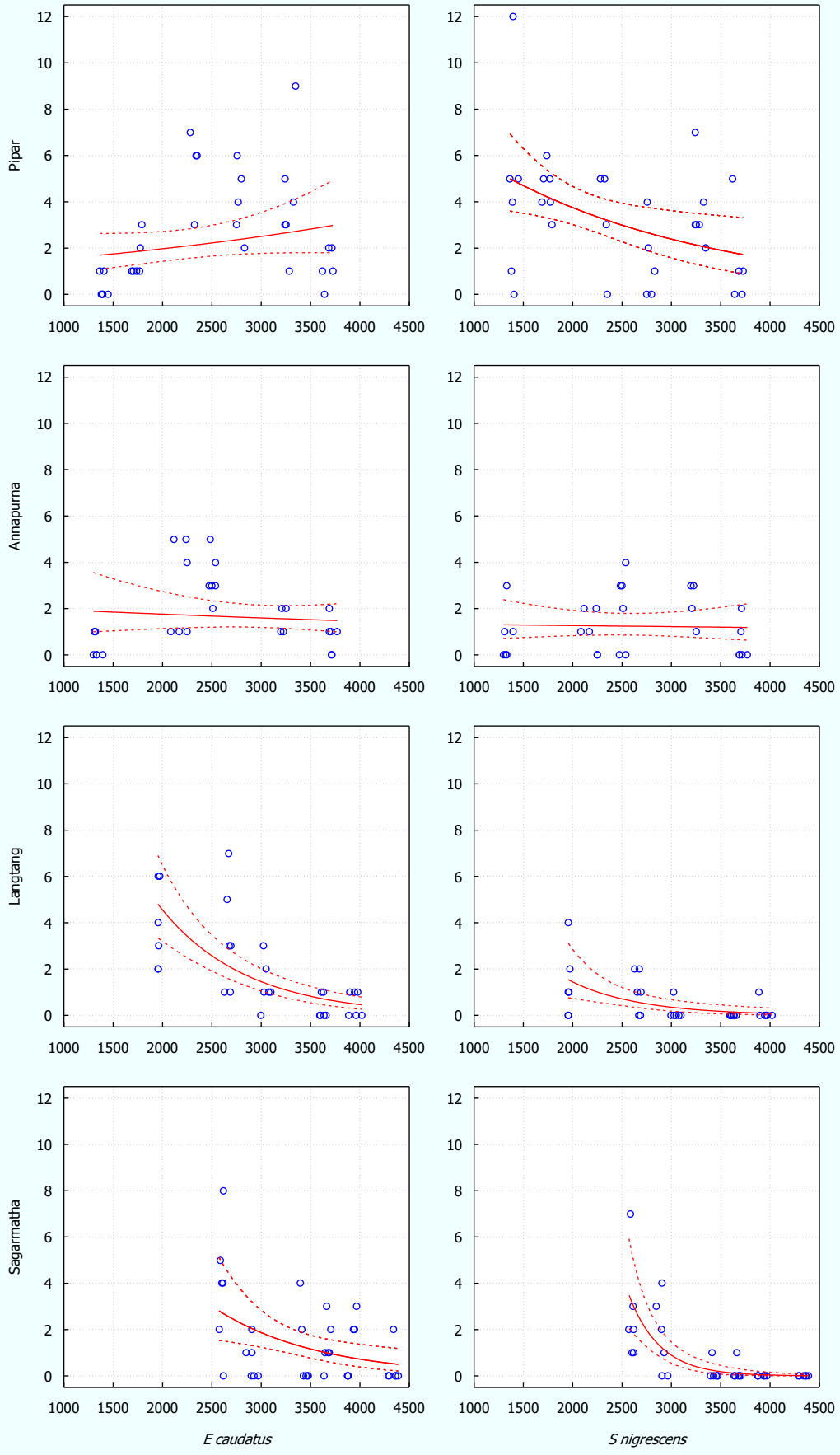


Figure 6.10 The three-way interaction between altitude (x-axis), transect and two species of shrew on abundance (y-axis). Solid lines show the predicted negative binomial functions calculated separately for each transect × species combination. Dotted lines show 95% C. I. of the lines.

6.3.4 The Influence of Habitat Structure, Proximity to Features and Botanical Characteristics on Species Distributions

Eight species were caught in sufficient numbers to allow individual analysis of habitat preferences. The three categorical variables containing cluster codes for habitat structure, feature proximity and botanical composition presented in Table 6.1 were used for these analyses. Given that the strong influences of altitude and transect have already been shown for many of these species, I have simply constructed separate cross-tabulations for each species/habitat combination and used contingency χ^2 analysis to test for associations.

The most useful predictor was botanical composition, which had signification associations with four species, while habitat structure only had an association with one species (Table 6.7). Five of the eight species had significant associations with at least one of predictor variables.

Table 6.7 Results of contingency χ^2 analysis of the abundance of eight species of small mammal against three different categorical predictor variables. P-values < 0.005 are highlighted in red, and marginal p-values (<0.05) are highlighted in blue.

Family	Species	Predictor Variables (Clusters)					
		Habitat Structure (10)		Feature Proximity (8)		Botanical Composition (7)	
		χ^2	p	χ^2	p	χ^2	p
Muridae	*A. gurkha*	13.74	0.1319	20.53	0.0045	26.15	0.0002
	M. musculus	13.66	0.1349	16.32	0.0224	29.01	< 0.0001
	N. eha	10.50	0.3158	11.78	0.1080	4.37	0.6267
	N. fulvescens	19.43	0.0218	28.42	0.0002	20.28	0.0025
Soricidae	*E. caudatus*	13.02	0.1616	6.09	0.5290	12.82	0.0461
	E. leucops	6.73	0.6657	5.90	0.5515	15.14	0.0191
	S. nigrescens	18.31	0.0318	16.27	0.0227	21.98	0.0012
	S. murinus	33.31	0.0001	22.27	0.0023	16.86	0.0098

A. gurkha showed a significant positive association with proximity cluster 7 and botanical cluster 4 (Figure 6.11). These described, respectively, trapping grids that had "grassland very close by and no crops, or woodland remote" and "*Betula* absent and grassland herbs present". There was also a negative association with proximity clusters 6 and 8 which indicated crops or human habitation nearby.

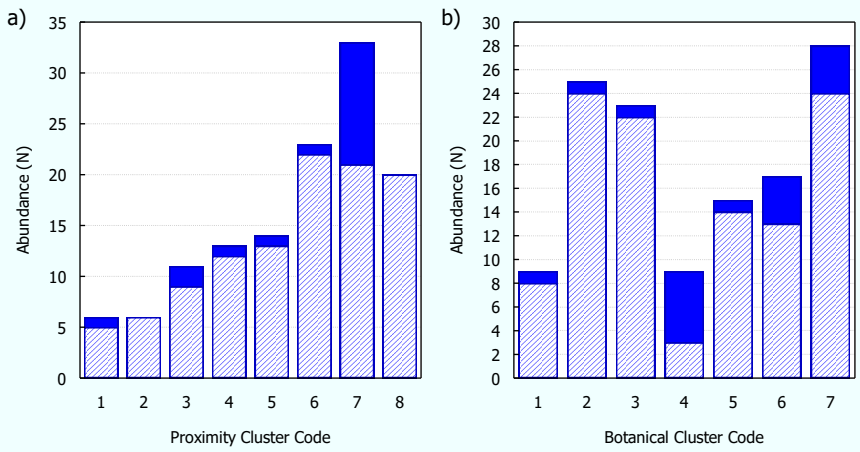


Figure 6.11 The associations between *Apodemus gurkha* and a) proximity and b) botanical clusters. The solid bars show the number of grids where the species was present, while the hatched bars show number of grids where it was absent.

M. *musculus* showed a very strong affinity with botanical clusters 2 and 7 (Figure 6.12), which described the presence of crops such as millet or fodder, or species such as *Artemisia* or nettles which are present in pasture. This species also had a significant aversion to clusters 3 and5, which described *Rhododendron campanulatum* or *Quercus* woodland with nettles absent.

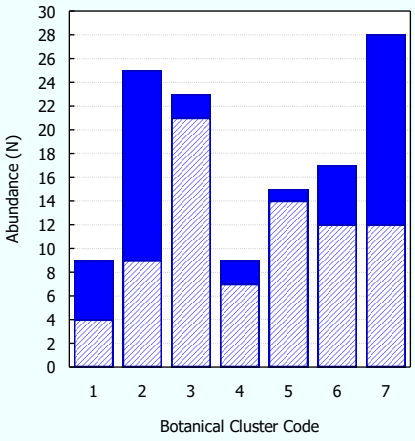


Figure 6.12 The association between *Mus musculus* and botanical clusters.

Niviventer eha showed no significant associations, but its congenor, *N. fulvescens* had very significant associations with proximity and botanical clusters (Figure 6.13). It showed a strong positive association with grids that had crops more than 200m distant, but not remote (cluster 1) and a negative association with scrub and grassland. This species was only found (with one exception) in botanical clusters 2 and 7, which is somewhat contradictory as these were the crop and pasture clusters favoured by M. *musculus*. More significantly, it had a negative aversion to botanical cluster 3, which was dominated by *R. campanulatum*.

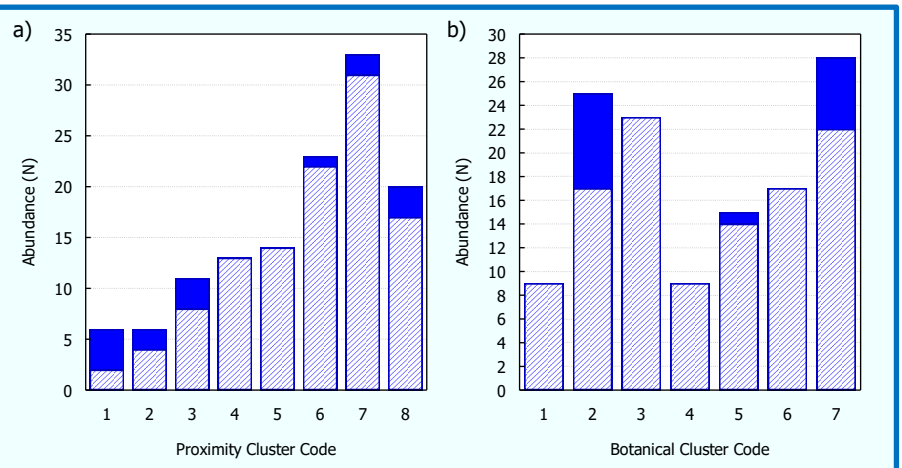


Figure 6.13 The associations between *Niviventer fulvescens* and a) proximity and b) botanical clusters.

Of the four shrew species, neither species of *Episoriculus* showed any significant associations with the habitat variables. The other common shrew, *S. nigrescens*, only had an association with the botanical clusters (Figure 6.14). In this case, due to its ubiquity, it was more an aversion to cluster 3 and, to a lesser extent to cluster 6, which described low, scrubby woodland and dwarf shrubs such as *Azalea* and *Salix*.

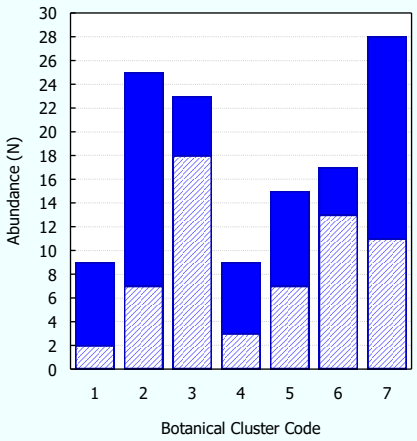


Figure 6.14 The association between *Soriculus nigrescens* and botanical clusters.

The final species of shrew was *S. murinus*, which had a strong positive association with structural cluster 4, and to a lesser degree cluster 3 (Figure 6.15). These described a medium height canopy (8 – 10m) and no leaf litter. This species also showed a strong affinity to crops and woodland or grassland within 50m (proximity cluster 2) and an aversion to grids with crops and woodland remote (proximity cluster 7).

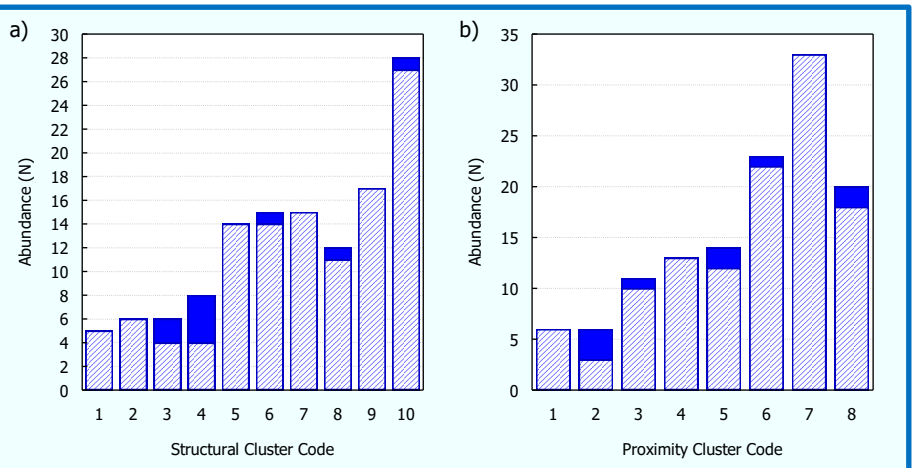


Figure 6.15 The association between *Suncus murinus* and a) structural and b) proximity clusters.

6.3.5 Species Diversity

6.3.5.1 Alpha Diversity

α–diversity ranged from zero to seven and conformed well to a Poisson distribution ($\chi^2_{(4)}$ goodness-of-fit = 3.77; $p \approx 0.44$; Figure 6.16). Both the mean empirical α–diversity and λ from the theoretical distribution were 2.26.

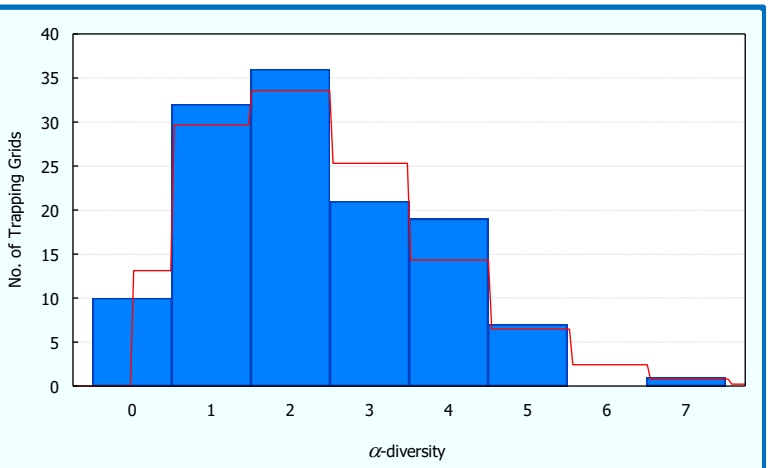


Figure 6.16 Frequency distribution of α–diversity for 126 trapping grids. The red line indicates expected Poisson distribution.

The first stage best-subsets analysis generated 127 models. AIC ranged from 407.05 to 425.43 for the worst fitting model. 34 models had AIC values < AIC_{min} + 2, the criterion suggested by Burnham and Anderson (2004) representing insufficient difference in AIC to merit discrimination. The best model included five terms; transect, transect × altitude, altitude, altitude^2 and altitude^3 and was an extremely good fit, with highly normalised residuals. However, the basic model with only three terms; transect, altitude and transect × altitude was also a very good fit with an AIC of 410.48. Not only was this a more parsimonious model, but the Type III tests were much more significant.

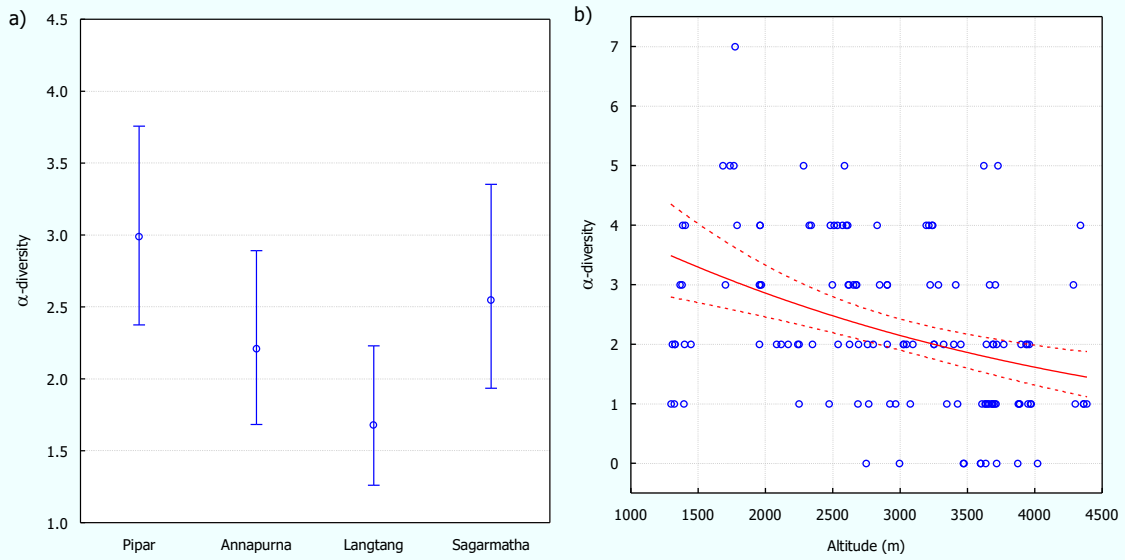


Figure 6.17 The main effects of a) transect (mean + 95% C. I.), b) altitude on α–diversity. The solid line in (b) shows the linear best-fit from a Poisson model with 95% C. I.

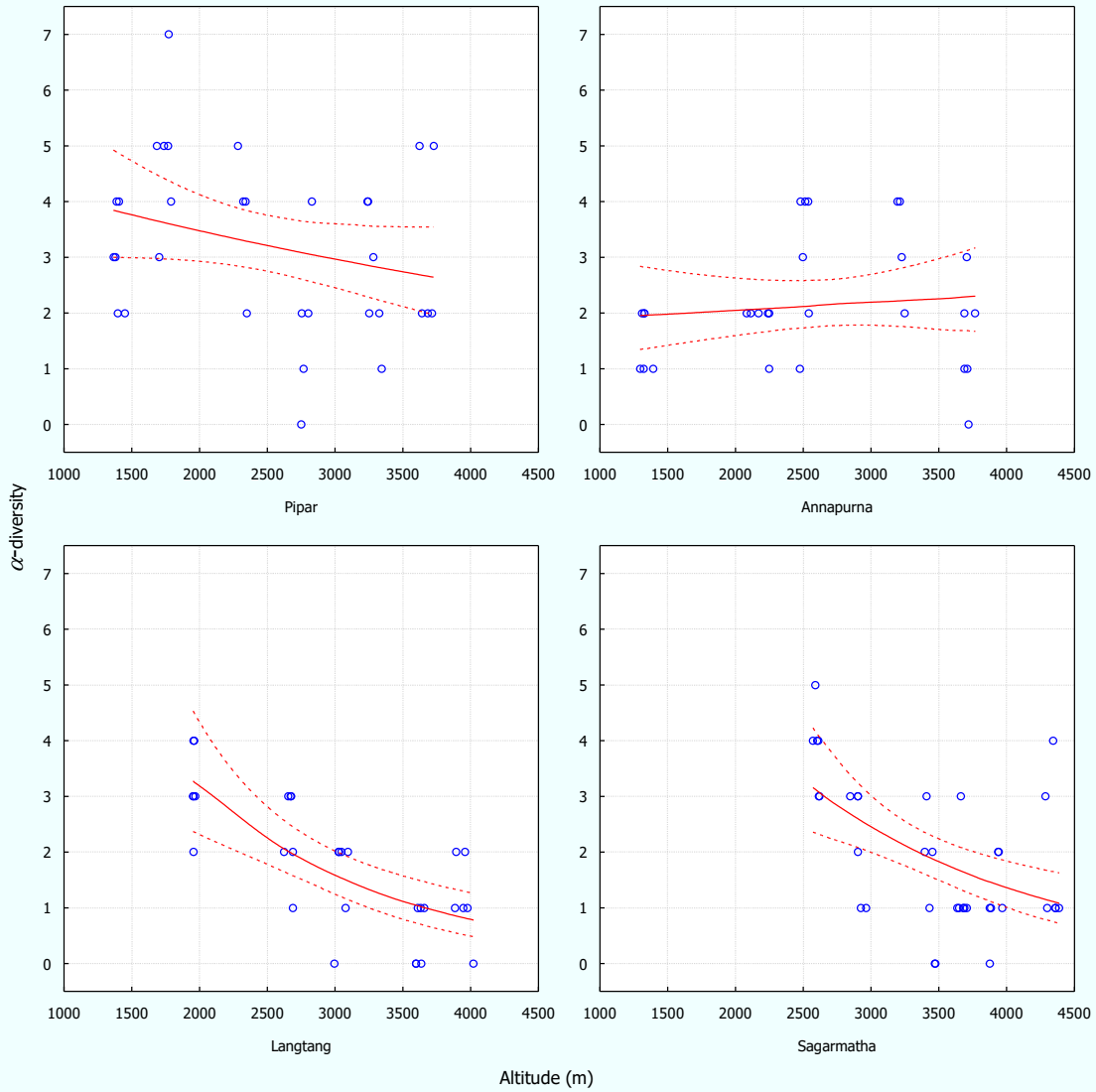


Figure 6.18 The 2-way interaction between transect and altitude on α–diversity. The solid line shows the linear best-fit from a Poisson distribution with 95% C. I.

The main effect of transect ($p < 0.01$) showed that α–diversity was significantly higher in Pipar (Figure 6.17a). There was also a highly significant decline in α–diversity ($p < 0.001$) with altitude (Figure 6.17b), so that predicted species richness in the lowest grids was approx. 3.5, declining to approx. 1.5 at 4400m. Finally, their interaction was also highly significant ($p < 0.005$), indicating that this relationship between α–diversity and altitude varied across transects (Figure 6.18). Pipar and Annapurna showed no clear decline with increasing altitude, whereas the effect was very strong in Langtang and Sagarmatha.

I also investigated whether a non-linear relationship might exist between α–diversity and altitude. A second-order polynomial model was a slightly better fit than a linear model (AIC = 419.86 & 421.60 respectively) but neither of the two terms in the polynomial model were significant, compared to the highly significant coefficient ($p < 10^{-4}$) for the linear decline illustrated in Figure 6.17b.

6.3.5.2 Beta Diversity

The Mantel test of β–diversity against altitudinal distance was very highly significant, with 20 of the statistics from 99,999 randomisations exceeding the test statistic, giving a Monte Carlo probability = 2×10^{-4} (Figure 6.19a). This result showed that as the difference in altitude between sites increased, the community composition diverged. In other words, there was significant species turnover with altitude. A similar result, although less significant, was obtained for the Mantel test of β–diversity against geographic distance ($p = 1.59 \times 10^{-3}$; Figure 6.19b). This showed that, on average, sites had

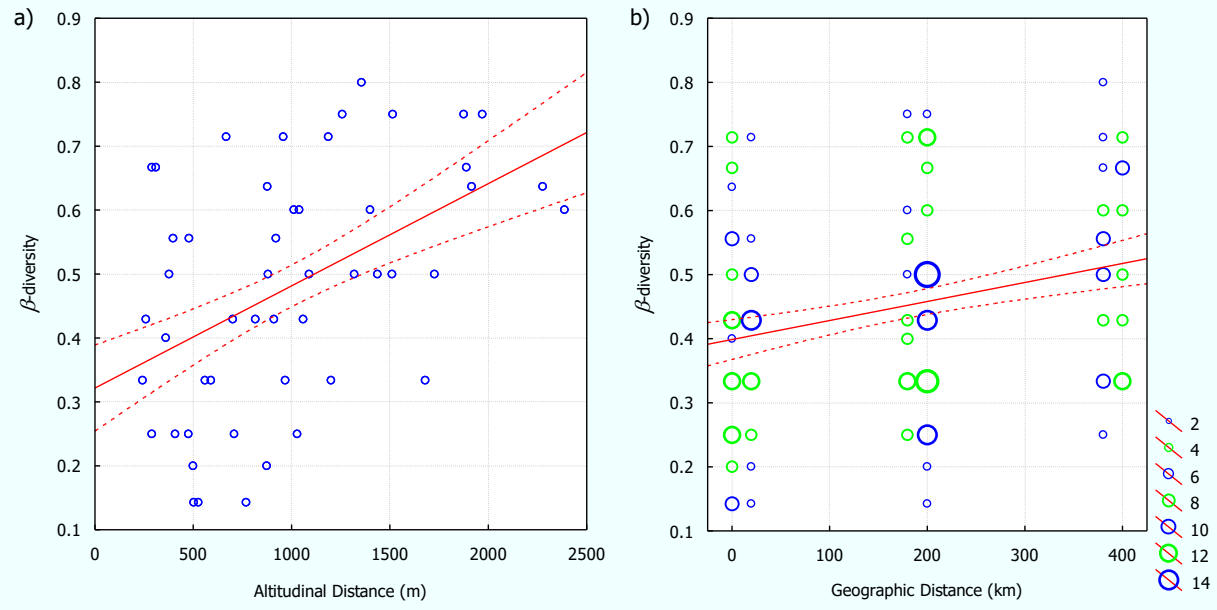


Figure 6.19 The effects of a) altitudinal and b) geographical distance on β–diversity. The points in (a) represent all pair-wise comparisons of sites within transects, thereby excluding any between transect diversity. The multiple points in (b) are size and colour-coded and represent all pair-wise comparisons of sites within 1000m altitudinal zones. The red lines show the best-fit linear relationships, although significance of the associations were tested with Monte Carlo tests (see text for details).

communities that were more similar to sites within their own transect (or between Pipar and Annapurna which were only 20km apart), than when compared with sites in other transects. In particular, β–diversity was greatest when comparing sites from Sagarmatha with Pipar and Annapurna, which were approximately 400km apart.

6.4 Discussion

In previous chapters I have shown that the geographical and temporal scope of this study makes it the largest ecological small mammal survey ever undertaken in Nepal. The field work was undertaken over three years on four transects at altitudes ranging from 1300m to 4400m in 126 trapping grids. A total of 792 individual small mammals were captured. At the same time, I collected habitat and botanical data for all trapping grids. I was also able to collect tissue samples from 720 animals (90.9%), some of which were analysed in Kathmandu. These yielded 94 good quality sequences of the cytochrome-b gene which were used to create phylogenies to aid identification. Finally, I used data from the BNHS collection of small mammals and those collated by Pearch (2011), to create predictive models of species distributions in Nepal. In this chapter I have brought together these various sources of information. This has allowed me to explore four themes;

- The consolidation of species identities,
- The analysis of overall abundance and abundance of selected taxa against geographical and topographical variables,
- The analysis of the effects of habitat variables on species distribution and abundance and
- The analysis of alpha and beta species diversity.

6.4.1 The Consolidation of Species Identities

One of the major features of the fieldwork design for this project was the use of live-catch traps. Historically, many researchers have resorted to the use of snap-traps because they are cheap, lightweight and easy to carry and use (Abe, 1971, 1977; Daniel, 2015; Mitchell, 1977). I decided to employ live-catch traps for two main reasons; ethics and efficacy, which I have discussed in Chapter 2. However, one of the consequences of live-trapping is that it requires identification of animals in the hand. This is not always easy, as most animals make vigorous attempts to escape (or bite), the fur may be wet or dirty and diagnostic characteristics, such as ears and feet, are often difficult to see. Furthermore, time pressures and humaneness issues are often paramount, so identifications may be cursory. Finally, the dearth of good quality identification material, limited to my own photographic compilation and a few sources such as Baral and Shah (2008) and Smith and Xie (2008), added to the difficulties of making confident identifications. I needed to rely considerably on my 35 years experience of small mammal trapping in the UK and overseas.

This is the main reason for my concerted efforts to confirm the field identifications. The use of two independent datasets to predict likely presence of species in Nepal and, more specifically, in the

current study locations, was explored in detail in Chapter 4. This provided strong evidence for the presence of 12 species. The phylogenetic study based on cytochrome-b, described in Chapter 5, confirmed that 13 species were recorded during the study, three of which had not been recorded during the field work (Figure 6.1).

Most importantly, 41 specimens of *Apodemus gurkha*, the only endemic non-volant mammal in Nepal (Molur *et al.*, 2005), were identified in the field, strongly supported by the distributional and phylogenetic analyses. The Bombay Natural History Society (BNHS) collection only has four specimens and the Zoological Survey Museum, Kolkata (ZSI) appears to have only two. In addition, Pearch (2011) collated 129 records from museum specimens and there has been a recently completed, unpublished MSc study (Karmacharya, *pers. comm.*). So, the 41 specimens from my study increase the number of the known records for this endangered species (Jnawali *et al.*, 2011) by nearly one third. Furthermore, I collected 33 tissue samples which represent a unique genetic resource, given that there are only 11 DNA sequences of this species on GenBank, four or which are incomplete cytochrome-b sequences.

Both the distributional and phylogenetic analyses were important in reassigning my 209 specimens of *Mus booduga* to *M. musculus*. This is a difficult genus with some dispute over specific identities (Shimada *et al.*, 2010; Suzuki *et al.*, 2013; Suzuki *et al.*, 2004). The analysis of collector effects for three species of *Mus* that I presented in Chapter 4, made a strong case for the misidentification of museum specimens. The analysis of GenBank sequences undertaken in Chapter 5 also showed unequivocal evidence of taxonomic confusion, which is of considerable concern if these sequences propagate new, incorrectly identified sequences through the uncritical use of the programme BLAST. Although it is disconcerting to find that the second most abundant species recorded in the current survey was probably misidentified, I am reassured that my efforts have resulted in the correct final identification, despite the fact that Marshall (1977) stated that there was a void in the outdoor distribution of *Mus musculus* in low and mid elevations of Nepal. Given that I obtained 197 tissue samples (94% of specimens) from every site in which the species was recorded, there is clear scope for an enormous amount of valuable genomic work on this species.

Two other species, *Rattus andamanensis* and *R. pyctoris*, were identified from the DNA sequences, although the latter was also predicted, from the distributional analysis, to be the most likely species of *Rattus* in my study areas. This is another problematic genus (Adhikari *et al.*, 2018a; Aplin *et al.*, 2011; Lack *et al.*, 2012; Robins *et al.*, 2007), for which I found evidence of misidentification in both the BNHS specimens and the GenBank sequences.

The only species of vole recorded in the current study, *Neodon sikimensis*, was very strongly predicted by the distributional analysis. However, the phylogenetic support was ambiguous, with roughly equal indication of *N. sikimensis* or *N. leucurus*. But far more interesting was the possibility that my five specimens represented two cryptic species, as the p-distances between the two clades exceeded even the mean con-generic p-distance calculated from GenBank sequences. Furthermore, the two clades were separated geographically by a distance of 400km, with the specimen found in the middle of this range (Langtang) having the smallest p-distances to all other specimens. Recently, Liu *et al.* (2012) described a new species of *Neodon* from Tibet, which was collected in close proximity to specimens of *N. sikimensis*. They recorded two other species of this genus from neighbouring Sichuan province, and a fifth species which they proposed should be removed from the genus. This is clear evidence of a fluid taxonomy, which would merit further genomic investigation of my specimens.

Five species of shrews were identified in the field, all of which had strong predictions from the distributional analysis. The two dominant species, *Episoriculus caudatus* and *Soriculus nigrescens*, plus *Suncus murinus* were also afforded strong support from the phylogenetic analysis. In particular, I may have uncovered evidence for another pair of cryptic species within *S. nigrescens*, although in this case it was in the GenBank sequences rather than my own specimens. Two sequences from Nepal had p-distances to five sequences from China that were nearly twice the con-specific distance for this group. This is currently considered to be a mono-specific genus (Motokawa *et al.*, 2008; Ohdachi *et al.*, 2006), so unless this was another case of misidentification, there is clearly scope for further taxonomic investigation. There are 51 sequences of this species lodged on GenBank, although they may not all have come from different individuals. I have tissue samples from 168 animals collected from all 19 sites in which they were recorded.

In addition to *Episoriculus caudatus*, I also recorded 20 specimens of *E. leucops*. The distributional and phylogenetic analyses both supported this identification, but also pointed equally to a con-generic species, *E. macrurus*. Although both are recorded as present in Nepal (Baral and Shah, 2008; Pearch, 2011; Smith and Xie, 2008), there are no specimens of *E. macrurus* in the BNHS or ZSI collections. Abe (1971) recorded one specimen of *E. leucops*, but neither were recorded by Daniel (2015) or Rands *et al.* (1979). My final decision was to retain the identification of *E. leucops* in the absence of contrary evidence. Furthermore, with reference to future genomic studies, I collected 18 tissue samples from these specimens and 212 tissue samples from *E. caudatus*, from every trapping site in the study. There are only 78 sequences of this species listed on Genbank.

The final identifications (Table 6.2) resulted in 17 putative species, two of which I have separated as cryptic species based on their very high phylogenetic distance. In addition, eight specimens remained unidentified as two potential species, one murid and one soricid. Furthermore, the analysis of altitudinal distribution indicated that the two specimens of murid were recorded from the extreme ends of the Pipar transect with an altitudinal separation of 2300m, so may also represent different species. The six unidentified soricids remain a puzzle as, morphologically, they were very distinct from *Episoriculus caudatus* (Figures 2.13 & 2.14 in Chapter 2), although their measurements were generally indistinguishable. In the field, they were generally found in grids above 3500m and I am confident that these were at least a different sub-species, so further analysis of their DNA sequences would be highly profitable.

6.4.2 Small Mammal Abundance and the Relationship with Altitude

The 792 individual small mammals caught during this survey gave an individual capture-rate of 9.8%. This was very similar to the 9.2% recorded by McCain (2004) in Costa Rica, the 8.2% of Bateman *et al.* (2010) in Queensland, Australia and Liu *et al.* (2008), who recorded 7.8% in China. Some studies have achieved higher rates than this, such as Nor (2001), who achieved 21.8% in Sabah, Malaysia and 21.1% obtained by Mulungu *et al.* (2008) on Mt. Kilimanjaro. Other studies have been based on a much lower absolute number of captures or capture rates (Li *et al.*, 2003; Wen *et al.*, 2018; Wu *et al.*, 2013). Higher capture rates and greater species richness may be achieved by using a number of different types and size of trap, such as the snap-trap, pitfall, cage and box traps employed by Rowe *et al.* (2015). The use of a single trap in the current study was driven partly by logistics, but also by a desire to ensure a completely standardised trapping grid, which allowed for more consistent trap exposure and analytical methods (Lomolino, 2001).

The overall abundance of small mammals showed a clear geographic effect, as two transects (Pipar and Sagarmatha) had significantly greater numbers than Annapurna and Langtang (Figure 6.2). A similar effect was shown by Li *et al.* (2003) in three transects on Mt. Qilian in China. More importantly there was a highly significant decline in abundance with altitude, again supported by Li *et al.* (2003), working over a similar elevational range from 1600m to 3700m. However, they used simple linear regressions on the transects individually, and took no account of interactions. Mulungu *et al.* (2008) also showed a decline in trap success with elevation on two transects on Mt. Kilimanjaro, although they provided no statistical support. In contrast, Bateman *et al.* (2010) showed a significant increase with altitude, although over a much smaller elevational range (650m), as did Heaney (2001) in two out of three gradients in the Philippines. In their study of small

mammal abundance on three mountains in China, Wen *et al.* (2018) showed highly significant positive relationships with species elevational ranges, rather than elevation *per se*. The highly significant two-way interaction between transect and altitude that I demonstrated (Figure 6.3) indicated that the negative relationship between abundance and elevation was not universal, which might explain some of the variation between the studies described above.

Although I caught many more shrews (481) than mice and rats (305), the GLMM (Table 6.4) showed that, if an adequate statistical model was constructed, this overall distinction was not significant. The two and three-way interaction terms explained this variance in a way that, for example, multiple χ^2 tests (Mulungu *et al.*, 2008) or separate linear regressions (Li *et al.*, 2003) could not. I did catch more shrews than rodents, but not on every transect, and the elevational effect also differed between species. The GLMM very accurately described the pattern of abundances shown by the multiple pie-charts (Figure 6.4) and gave a robust statistical basis. This ratio of shrews to rodents is very unusual, with most studies showing the opposite effect . For example on an elevational gradient in Tanzania, Stanley and Kihaule (2016) collected 276 shrews and 475 rodents (37% shrews).

These distinctions were explored further with two inter-species comparisons, one between two mouse species and one between two shrews. The first of these indicated a very strong distributional aversion between *Apodemus gurkha* and *Mus musculus* (Figure 6.6). *M. musculus* was found in all transects and at every altitude, from the lowest sites in Pipar and Annapurna (1300m) to the highest grids in Sagarmatha (4400m). In contrast, *A. gurkha* was only found in the two adjacent, western transects (Pipar and Annapurna), only above 2500m and dominant above 3200m. The GLMM confirmed that there was a significant difference between predicted abundance of the two species (1.68 & 4.86 per grid respectively) but, taken together, these two species showed no altitudinal decline. However, the three interaction terms that included species fully explained the pattern (Figure 6.8). In summary, a) *M. musculus* had the capacity to exploit all habitats, b) *A. gurkha* had a distribution limited to the western transects only and greater than 2500m altitude, so that c) when *A. gurkha* was present, *M. musculus* was absent. Although this appears to be an example of competitive exclusion (Grant, 1972), without removal experiments it would be impossible to confirm this. Unrecorded, underlying environmental factors may be the cause of this allopatric distribution, rather than direct competition between the species. Nevertheless, this is a direct contradiction to the findings of Martens and Niethammer (1972), who claimed that the two species occurred together. I can find little recent evidence of this effect in small mammals in the wild and most of it appears to be based on small-scale experiments (Abramsky *et al.*, 1979; Cameron, 1964; Dueser and Porter, 1986). Kelt *et al.* (1995) explored the effect of three types of competition on the composition of small

mammal communities across an ecotone in southern Chile and Suzuki *et al.* (2013) quote Marshall (1977), who attributed the absence of outdoor *Mus musculus* in India to the presence of other species of *Mus*.

The second inter-specific comparison was between *Episoriculus caudatus* and *Soriculus nigrescens* (Figure 6.9 & Table 6.6). There appeared to be no spatial exclusion between these species, which had similar overall abundance and both were recorded in 19 of the 22 sites (86%). This might be explained by the nearly 3:1 ratio in their mean body masses (*S. nigrescens* = 13.98 ± 0.19 SE, *E. caudatus* = 5.06 ± 0.05), which could result in different microhabitat or food preferences (Dueser and Shuggart, 1979; Vieira and Port, 2007). Taken together, these two species showed a significant decline in abundance with altitude, although both species were still recorded in small numbers at 4000m. In the Philippines, Rickart *et al.* (2011) showed the opposite effect in a single species of *Sorex*, which was caught in larger numbers at higher elevations, although their gradient only extended from 925m to 2150m. This pattern was also shown by Stanley and Hutterer (2007) in Tanzania, where total shrew numbers increased nearly four-fold between 600m and 2000m. However, on two higher mountains, the lowest capture rates of shrew species were found at 3600m (Stanley and Kihaule, 2016) and 4000m (Stanley *et al.*, 2014). This evidence indicates that shrew numbers decline only at the highest altitudes, which may indicate a metabolic and/or trophic limitation at low temperatures (Churchfield *et al.*, 2012; Genoud, 1988; Merritt, 1986).

6.4.3 The Influence of Habitat Variables on Small Mammal Abundance

The analysis of habitat preferences for eight species of small mammal showed significant associations for five of them (Table 6.7). In particular *Apodemus gurkha* and *Niviventer fulvescens* were associated with both feature proximity and botanical composition, but with different clusters. *A. gurkha* showed a preference for high altitude grassland with herbs or dwarf shrubs (Pearch, 2011) and a strong aversion to human habitation and crops. In contrast, although the preferences of *N. fulvescens* were significant, they were difficult to interpret. It appeared to have an aversion to grids with crops nearby, but was positively associated with grids in crops and pasture. *Mus musculus* only had significant associations with botanical composition, showing a positive association with human-influenced habitats, such as crops and pasture, and a strong aversion to *Quercus* woodland and to *Rhododendron campanulatum* scrub woodland found at higher altitudes.

Episoriculus caudatus was the most abundant species, but showed no evidence of habitat preferences. Despite its ubiquity, there is very little published material on its habitat preferences. Although it was also generally ubiquitous, *Soriculus nigrescens* showed a strong aversion to *Rhododendron*

campanulatum scrub woodland and dwarf shrubs, such as *Azalea* and *Salix* found above the tree-line. *Suncus murinus* showed a strong affinity for grids with crops present or nearby, and an aversion to high altitude grassland. This species is known to be commensal (Chenchittikul *et al.*, 1983; Olsen, 1984; Tung *et al.*, 2013). However, Nakomoto and Nakanishi (2013) suggested that it favoured grassland as a nocturnal foraging habitat, and its association with human habitation was more likely to be for favourable den sites. Anecdotally, although most of the specimens were in close proximity to fields and crops, I caught a number in woodland at least one hundred metres from human habitation.

6.4.4 The Analysis of Alpha and Beta Diversity Along Altitudinal Gradients

In the analysis of species diversity, I have focused on the concepts of α and β–diversity (Akhtar and Bergmeier, 2015; Jost, 2007; Koleff *et al.*, 2003; McCain, 2005; Wilson and Shmida, 1984). I have defined α–diversity as species richness at the trapping grid resolution. In this sense it conforms to the definitions of Bhatta *et al.* (2018) for alpine vegetation in Langtang, Nepal, and McCain (2004) for small mammals in Costa Rica. My analysis showed that α–diversity of small mammals declined significantly with increasing altitude, although the pattern was not consistent across transects, with two showing no effect and two having strong declines. Furthermore, this was essentially a log-linear response, unlike many studies which have shown second and third-order polynomial relationships with altitude; see *e.g.*, (Chen *et al.*, 2017; Tello *et al.*, 2015). I tested these effects with the Akaike information criterion (AIC), but the improvement in the models were so small that I considered the risk of over-fitting too great and chose the more parsimonious model.

Surprisingly few studies of small mammals on altitudinal gradients explicitly examine α–diversity. On Mt. Kilimanjaro there was a significant negative relationship between numbers of rodent and shrew species in trapping sites between 1800m and the summit (3590m) (Mulungu *et al.*, 2008), which was supported for shrews by Stanley *et al.* (2014), but not for all species. In contrast, on nearby Mt. Meru, α–diversity of all small mammals species showed no significant altitudinal effect (Stanley and Kihaule, 2016). On two transects in southern Mexico, Sánchez-Cordero (2001) showed no significant elevational effects for rodents, but there did appear to be a modest decline in bat species richness. Between 800m and 3200m on Mt Kinabalu in Malaysia, Nor (2001) showed an apparent hump-shaped distribution, with a peak around 1800m, although the data were too sparse to show this statistically. In the Philippines Rickart *et al.* (2011) showed a significant positive relationship between native species richness (but not total species richness) and altitude.

A more common approach is to analyse field data in terms of gamma-diversity (γ–diversity), which seems to have multiple definitions. For example McCain (2005) defined γ–diversity as "*...species richness patterns compiled from trapping records, specimen records and field notes for an entire mountain or mountainous region regardless of slope, area...*". In contrast, Lomolino (2001) restricted his definition to the total richness of an entire elevational zone. Predating this explicit definition, Hunter and Yonzon (1993) used 500m elevational zones to collate bird and mammal species richness for the whole of Nepal. In my study, I could have calculated γ–diversity as 17, being simply the number of putative species I recorded throughout the whole study (Table 6.2). Alternatively, I could have calculated this statistic for each of my "mountains" (transects) or for elevational zones. However, given that I have made detailed analyses of both α and β–diversity by transect and altitude, γ–diversity seems to add little.

Nevertheless, several authors have taken the γ–diversity approach of collating species into altitudinal zones, interpolating their presence in any intermediate zones where they were absent, and then tallying total species richness for the zone. This has been done with field data (Heaney, 2001; Lomolino, 2001) and with historical data from museum specimens (Rickart, 2001; Rowe, 2009). Many of these studies show a "mid-elevation hump" in species richness, a phenomenon which has received a great deal of attention for small mammal communities (McCain, 2004; Rowe *et al.*, 2015; Wen *et al.*, 2018; Wu *et al.*, 2013) and for other taxa (Cardelús *et al.*, 2006; Carpenter, 2005; García-López *et al.*, 2012). One explanation for this is the "Mid-domain Effect" (MDE) (Colwell and Hurtt, 1994; Colwell and Lees, 2000; Colwell *et al.*, 2004). This asserts that species richness will follow a modal pattern with elevation (and latitude), whenever there are physical constraints to the distribution of species, such as the bottom and top of mountains, although its importance has been disputed (Currie and Kerr, 2008). My data did not show this effect for α–diversity, possibly because the altitudinal range in which I worked did not encompass the full range available and, therefore, relaxed the constraining effects (Koleff and Gaston, 2001; McCain and Grytnes, 2010). Two of my four transects had inaccessible, v-shaped valleys below my lowest sites, and all transects had some vegetated habitat above my top sites which were unavailable due to adverse weather conditions and logistical constraints.

I have used β–diversity to represent a measure of species turnover. Like γ–diversity, it seems to have multiple definitions. Jankowski *et al.* (2009) expressed it (after Whittaker) as "*...the ratio of alpha species richness and regional ('gamma') diversity.*". This multiplicative relationship gives a "static" assessment for any single site within a region. If I had calculated γ–diversity across my whole study (17) or by 500m elevational zones (8, 8, 8, 8, 9, 8 & 4 respectively), it would have resulted, more or less,

in a constant value, so would simply have reflected the magnitude of α–diversity. Of more use is the definition given by Nguyen and Gomez-Zurita (2016); "*Beta-diversity describes the changes in species composition across local samples in a particular region…*" (also after Whittaker). There are many diversity indices which can be used to calculate β–diversity values between two sites, which are reviewed comprehensively by Koleff *et al.* (2003). The most commonly used appears to be that of Sørensen (1948), which is a similarity index scaling from zero to one. In contrast, I have used the index of Whittaker (1960), which is algebraically equivalent, but is a dissimilarity index, scaling from zero to one as diversity increases. This seems more appropriate as a measure of diversity and is the method employed by (Mena and Vázquez Domínguez, 2005).

I showed that there was a highly significant association between pair-wise β–diversity and altitudinal distance within transects (Figure 6.19a). In other words, this gave a statistical basis to the turnover in species composition in sites along the elevational gradients shown in Figure 6.6 and Figure 6.9. Sites that were within 500m altitude of each other generally had a lower β–diversity index, with an average predicted dissimilarity of 0.4. In contrast, pairs of sites that were at the extremes of their transects (approx. 2000m apart) had predicted dissimilarity of 0.7. The same pattern was shown clearly on three transects in China over very similar altitudinal ranges (Wen *et al.*, 2014), and on north and south-facing slopes of the same mountain (Shuai *et al.*, 2017). In terms of individual species, I have already shown (Sub-section 6.3.3) that there was a significant altitudinal exclusion between *Apodemus gurkha* and *Mus musculus*, but I did not analyse the less abundant species. The clear altitudinal zonation shown by the rodents *Niviventer eha*, *N. fulvescens*, *Rattus andamanensis* and *R. pyctoris* (Figure 6.6), and the shrews *Episoriculus leucops*, *Sorex bedfordiae* and *Suncus murinus* (Figure 6.9), all contributed to the significance of these results. Furthermore, the analysis of geographic distance also showed that communities diverged in their composition with the distance between transects (Figure 6.19b). This reflects the distributions of *A. gurkha*, *R. andamanensis* and *S. murinus* restricted to Pipar and Annapurna, and *S. bedfordiae* to Langtang and Sagarmatha.

In summary, I derived small mammal α and β–diversity from 126 trapping grids, with a total elevational range of 3100m. The survey design conformed with the criteria laid down by Rahbek (1995) that sites be sufficient in number and equally distributed over the gradient. These statistics have not been obtained before from small mammal studies in Nepal and rarely from empirical studies elsewhere with such large sample sizes. Most others relied on one or two transects with usually around 10, at most 20, trapping localities, over altitudinal ranges between 1000m and 2000m. I have shown that both α and β–diversity in small mammal communities in the Himalaya of central Nepal have significant altitudinal and geographic variation.

6.5 Concluding Remarks and Suggestions for Future Research

I contend that this has been a ground-breaking study. As is often the case, it has revealed many more questions regarding the ecology, taxonomy and conservation of small mammals in the Himalayan region of Nepal. I would like to conclude by exploring some options for future research in this area.

6.5.1 Further Genomic Analysis

The phylogenetic analysis presented in Chapter 5 was based on 94 partial sequences of one gene (cytochrome-b). This represents a minute fraction of the genomic information available in the 720 tissue samples that I collected. Given that 501 DNA extractions have already been undertaken and are stored in Kathmandu, this gives a vast resource for further phylogenetic analysis. I propose that additional genomic research be undertaken in three stages;

- Complete the PCRs for all current DNA extractions. This will ensure adequate sequences for all specimens of the less common species, especially the endemic *Apodemus gurkha*, as well as the three common species; *Mus musculus*, *Episoriculus caudatus* and *Soriculus nigrescens*.
- For specimens where sequences are inadequate or alignments are ambiguous, undertake PCR for the CO1 gene, or others suitable for DNA barcoding.
- Additional studies based on micro-satellites or other nuclear DNA loci.

The three common species would provide excellent subjects for research into population structure and evolutionary processes, as they were recorded in all transects and at all altitudes. As recommended by Emerson *et al.* (2017), I have grids that are only 100m apart and grids that are separated in altitude by 3000m and spatially by 400km. Furthermore, the degree of physical isolation imparted by the topography means that two sites, which may have been only 5km apart on the map, could have been isolated to small mammals for thousands of years. The hierarchical design of the field survey means that population structure could be investigated at four spatial scales; within the trapping grid, within the trapping site, within and between transects.

Given the restrictions on removal of tissue samples or extracted DNA from Nepal, I propose that modern sequencing techniques be employed, including portable nanopore technology such as the MinION (Oxford Nanopore Technologies, 2018; Srivathsan *et al.*, 2018) or loop-mediated iso-thermal amplification (LAMP) (Lee, 2017). Finally, this would give ample opportunity to obtain the sequences necessary to identify new species or sub-species and quantify geographical variation.

6.5.2 Additional Analysis of Historical Datasets

The reference specimens from the BNHS collection described in Chapter 4 have now been fully geo-referenced. Three specific areas come to mind for additional analyses of these data:

- Over 90% of the 7285 small mammal specimens had morphometric data. These would provide an invaluable resource for additional taxonomic analyses, with important conservation consequences.
- I used two spatial modelling methods, logistic regression and Maxent. I suggested that injudicious used of these models, especially Maxent can provide spurious predictions. Maxent is often used as a "black-box", with little consideration of the vast number of combinations of input parameters (Merow *et al.*, 2013). This would merit considerable further investigation, as would the use of other modelling techniques such as general additive models (Pearce and Ferrier, 2000) and genetic algorithms (Elith *et al.*, 2006). Furthermore, I only used a digital elevation model as the source data, so the inclusion of human demographic datasets and satellite-based datasets of vegetation cover, human habitation and disturbance would be extremely profitable.
- Permission has been granted by the officers of the BNHS for tissue samples to be taken for DNA extraction and sequencing (Koht, *pers. comm.*). In concert with the proposals made in the previous sub-section, this would provide a unique historical perspective to small mammal taxonomy, ecology and evolution in the Indian sub-continent.

6.5.3 Progression of Fieldwork

The field study described in Chapters 2 and 3 could provide a framework for long-term monitoring, with particular reference to species conservation, human-mediated habitat modification and climate change. Every trapping grid was geo-referenced and could be relocated to within 10m accuracy. Whilst not promoting complete repetitions of this study, I would propose three additional aspects;

- Longer-term trapping in selected sites. I made the decision to increase the spatial coverage of this study by only trapping for two nights in each grid, which precluded the effective estimation of population sizes. By extending the number of trapping sessions, it would be possible to obtain valuable capture-mark-recapture estimates of population sizes and use cumulative species-abundance curves to ensure complete recording of all species. Furthermore, the use of additional trap types would expand the suite of species recorded, especially larger-bodied rodents.
- All fieldwork carried out in this study took place in the post-monsoon period (September to November), which is the post-breeding season. It would be very valuable from an ecological

perspective to trap in the period between winter and the onset of the monsoon (March to June), which would sample the over-wintered, breeding populations of small mammals.

- Collection of additional habitat data. The rapid assessment method that I used to record habitat data in each trapping grid, was essential given the time constraints in the current study. However, a fully quantitative study, using precise structural measurements of the habitat and detailed, quantitative botanical survey should be completed. This would provide intrinsically valuable data, which could also be used for powerful retrospective analysis of the small mammal trapping results from this and future surveys.

6.5.4 Building Local Expertise

The acknowledgements at the beginning of this thesis indicate the huge amount of support and assistance that I have had over six years from the people of Nepal. In particular, there is now a cohort of enthusiastic young scientists, including field ecologists, conservationists, taxonomists, geneticists and software engineers. I believe it is imperative that the proposals outlined above be implemented to encourage, guide and further the capacity of the next generation of Nepali research scientists and technicians.

6.6 References

Abe, H. (1971) Small Mammals of Central Nepal. *Journal of the Faculty of Agriculture, Hokkaido University* **56**: 367-423.

Abe, H. (1977) Variation and taxonomy of some small mammals from central Nepal. *J. Mamm. Soc. Japan* **7**: 63-73.

Abramsky, Z., Dyer, M.I., and Harrison, P.D. (1979) Competition among Small Mammals in Experimentally Perturbed Areas of the Shortgrass Prairie. *Ecology* **60**: 530-536.

Adhikari, P., Han, S.-H., Kim, Y.-K., Kim, T.-W., Thapa, T.B., Subedi, N., Kunwar, A., Banjade, M., and Oh, H.-S. (2018a) New record of the Oriental house rat, *Rattus tanezumi*, in Nepal inferred from mitochondrial Cytochrome B gene sequences. *Mitochondrial DNA Part B* **3**: 386-390.

Adhikari, P., Han, S.H., Kim, Y.K., Kim, T.W., Thapa, T.B., Subedi, N., Adhikari, P., and Oh, H.S. (2018b) First molecular evidence of *Mus musculus bactrianus* in Nepal inferred from the mitochondrial DNA cytochrome B gene sequences. *Mitochondrial DNA A DNA Mapp Seq Anal* **29**: 561-566.

Akhtar, N. and Bergmeier, E. (2015) Species richness, alpha and beta diversity of trees, shrubs and herbaceous plants in the woodlands of Swat, Pakistan. *Pakistan Journal of Botany* **47**: 2107-2113.

Anon. (2017a) SPSS Statistics v. 25. IBM Corporation. https://www.ibm.com/support/knowledgecenter/en/SSLVMB_25.0.0/

Anon. (2017b) Statistica (data analysis software system) v.13. TIBCO Software Inc. http://statistica.io.

Aplin, K.P., Suzuki, H., Chinen, A.A., Chesser, R.T., Ten Have, J., Donnellan, S.C., Austin, J., Frost, A., Gonzalez, J.P., Herbreteau, V., Catzeflis, F., Soubrier, J., Fang, Y.P., Robins, J., Matisoo-Smith, E., Bastos, A.D., Maryanto, I., Sinaga, M.H., Denys, C., Van Den Bussche, R.A., Conroy, C., Rowe, K., and Cooper, A. (2011) Multiple geographic origins of commensalism and complex dispersal history of Black Rats. *PLoS One* **6**: e26357.

Baral, H.S. and Shah, K.B. (2008) *Wild Mammals of Nepal*. Himalayan Nature, Kathmandu. pp 188.

Bateman, B.L., Kutt, A.S., Vanderduys, E.P., and Kemp, J.E. (2010) Small-mammal species richness and abundance along a tropical altitudinal gradient: an Australian example. *Journal of Tropical Ecology* **26**: 139-149.

Bhagat, I.M. and Shrestha, S. (2010) Ferns and Fern-allies of Eastern Terai, Nepal. *Our Nature* **8**.

Bhatta, K.P., Grytnes, J.A., and Vetaas, O.R. (2018) Scale sensitivity of the relationship between alpha and gamma diversity along an alpine elevation gradient in central Nepal. *Journal of Biogeography* **45**: 804.

Bhattarai, K.R. and Upadhyay, T.P. (2013) Rangeland Management in Sagarmatha (Mount Everest) National Park and Buffer Zone, Nepal: An Ecological Perspective. *Mountain Research and Development* **33**: 19-28.

Bhattarai, K.R., Vetaas, O.R., and Grutnes, J.A. (2004) Fern species richness along a central Himalayan elevational gradient, Nepal. *Journal of Biogeography* **31**: 389-400.

Burnham, K.P. and Anderson, D.R. (2004) Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods Research* **33**: 261-304.

Cameron, A.W. (1964) Competitive exclusion betweent the rodent genera *Microtus* and *Clethrionomys*. *Evolution* **18**: 630-634.

Cardelús, C.L., Colwell, R.K., and Watkins Jr, J.E. (2006) Epiphyte Distribution Patterns: Explaining the Mid-Elevation Richness Peak. *Journal of Ecology* **94**: 144-156.

Carpenter, C. (2005) The environmental control of plant species density on a Himalayan elevation gradient. *Journal of Biogeography* **32**: 999-1018.

Chaudhary, R.P., Uprety, Y., and Joshi, S.P. (2016) *Plant Biodiversity in Nepal: Conservation and Legal Status*. Pages 224-268 *in* P.K. Jha, Siwakoti, M., and Rajbhandary, S., editors. *Frontiers of Botany*. Central Department of Botany, Tribhuvan University, Kathmandu, Nepal.

Chen, W., Li, Y., Liu, Y., Liu, S., and Yue, B. (2011) Complex topographic configuration in the Hengduan Mountains shaped the phylogeographic structure of Chinese white-bellied rats. *Journal of Zoology* **284**: 215-223.

Chen, Z., He, K., Cheng, F., Khanal, L., and Jiang, X. (2017) Patterns and underlying mechanisms of non-volant small mammal richness along two contrasting mountain slopes in southwestern China. *Sci Rep* **7**: 13277.

Chenchittikul, M., Daengpium, S., Hasegawa, M., Itoh, T., and Phanthumachinda, B. (1983) A study of commensal rodents and shrews with reference to the parasites of medical importance in Chanthaburi Province, Thailand. *Southeast Asian Journal of Tropical Medicine & Public Health* **14**: 255-259.

Churchfield, S., Rychlik, L., and Taylor, J.R.E. (2012) Food resources and foraging habits of the common shrew, *Sorex araneus*: does winter food shortage explain Dehnel's phenomenon? *Oikos* **121**: 1593-1602.

Colwell, R.K. and Hurtt, G.C. (1994) Nonbiological gradients in species richness and a spurious Rapoport efffct. *The American Naturalist* **144**: 570-595.

Colwell, R.K. and Lees, D.C. (2000) The mid-domain effect: geometric constraints on the geography of species richness. *Trends in Ecology & Evolution* **15**: 70-76.

Colwell, R.K., Rahbek, C., and Gotelli, N.J. (2004) The mid-domain effect and species richness patterns:what have we learned so far? *The American Naturalist* **163**: E1-E23.

Currie, D.J. and Kerr, J.T. (2008) Tests of the mid-domain hypothesis: a review of the evidence. *Ecological Monographs* **78**: 3-18.

Daniel, M. (2015) Expansion of small terrestrial mammals and their parasites into the Barun Valley (Makalu Mt. Region, Nepal Himalaya) linked with changes in glaciation and human activities. *Journal of Mountain Science* **12**: 14-29.

Dueser, R.D. and Porter, J.H. (1986) Habitat Use by Insular Small Mammals: Relative Effects of Competition and Habitat Structure. *Ecology* **67**: 195-201.

Dueser, R.D. and Shuggart, H.H. (1979) Niche Pattern in a Forest-Floor Small-Mammal Fauna. *Ecology* **60**: 108-118.

Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M., and E. Zimmermann, N. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**: 129-151.

Emerson, B.C., Casquet, J., Lopez, H., Cardoso, P., Borges, P.A.V., Mollaret, N., Oromi, P., Strasberg, D., and Thebaud, C. (2017) A combined field survey and molecular identification protocol for comparing forest arthropod biodiversity across spatial scales. *Mol Ecol Resour* **17**: 694-707.

García-López, A., Micó, E., and Galante, E. (2012) From lowlands to highlands: searching for elevational patterns of species richness and distribution of scarab beetles in Costa Rica. *Diversity and Distributions* **18**: 543-553.

Ge, D., Lu, L., Cheng, J., Xia, L., Chang, Y., Wen, Z., Lv, X., Du, Y., Liu, Q., and Yang, Q. (2017) An endemic rat species complex is evidence of moderate environmental changes in the terrestrial biodiversity centre of China through the late Quaternary. *Sci Rep* **7**: 1-14.

Genoud, M. (1988) Energetic strategies of shrews: ecological constraints and evolutionary implications. *Mammal Review* **18**: 173-193.

Grant, P.R. (1972) Interspecific Competition Among Rodents. *Annual Review of Ecology and Systematics* **3**: 79-106.

Heaney, L.R. (2001) Small Mammal Diversity along Elevational Gradients in the Philippines: An Assessment of Patterns and Hypotheses. *Global Ecology & Biogeography* **10**: 15-39.

Holmes, M.W., Hammond, T.T., Wogan, G.O.U., Walsh, R.E., LaBarbera, K., Wommack, E.A., Martins, F.M., Crawford, J.C., Mack, K.L., Bloch, L.M., and Nachman, M.W. (2016) Natural history collections as windows on evolutionary processes. *Molecular Ecology* **25**: 864-881.

Hunter, M.L.J. and Yonzon, P. (1993) Altitudinal Distributions of Birds, Mammals, People, Forests, and Parks in Nepal. *Conservation Biology* **7**: 420-423.

Jankowski, J.E., Ciecka, A.L., Meyer, N.Y., and Rabenold, K.N. (2009) Beta diversity along environmental gradients: implications of habitat specialization in tropical montane landscapes. *Journal of Animal Ecology* **78**: 315-327.

Jnawali, S.R., Baral, H.S., Lee, S., Acharya, K.P., Upadhyay, G.P., Pandey, M., Shrestha, R., Joshi, D., Laminchhane, B.R., Griffths, J., Khatiwada, A.P., Subedi, N., and Amin, R. (2011) *The Status of Nepal's Mammals: The National Red List Series.* IUCN, Kathmandu, Nepal. pp 266.

Jost, L. (2007) Partitioning diversity into independent alpha and beta components. *Ecology* **88**: 2427-2439.

Kelt, D.A., Taper, M.L., and Meserve, P.L. (1995) Assessing the Impact of Competition on Community Assembly: A Case Study using Small Mammals. *Ecology* **76**: 1283-1296.

Koleff, P. and Gaston, K.J. (2001) Latitudinal gradients in diversity: real patterns and random models. *Ecography* **24**: 341-351.

Koleff, P., Gaston, K.J., and Lennon, J.J. (2003) Measuring beta diversity for presence–absence data. *Journal of Animal Ecology* **72**.

Krystufek, B., Donev, N.R., and Skok, J. (2010) Species richness and distribution of non-volant small mammals along an elevational gradient on a Mediterranean mountain. *Mammalia* **75**: 3-11.

Lack, J.B., Greene, D.U., Conroy, C.J., Hamilton, M.J., Braun, J.K., Mares, M.A., and Van Den Bussche, R.A. (2012) Invasion facilitates hybridization with introgression in the Rattus rattus species complex. *Molecular Ecology* **21**: 3545-3561.

Lee, P.L. (2017) DNA amplification in the field: move over PCR, here comes LAMP. *Mol Ecol Resour* **17**: 138-141.

Li, J.S., Song, Y.L., and Zeng, Z.G. (2003) Elevational gradients of small mammal diversity on the northern slopes of Mt. Qilian, China. *Global Ecology & Biogeography* **12**: 449-460.

Liu, J.-y., Du, H., Tian, G.-b., Yu, P.-h., Wang, S.-w., and Peng, H. (2008) Community Structure and Diversity Distributions of Small Mammals in Different Sample Plots in the Eastern Part of Wuling Mountains *Zoological Research* **29**: 637-645.

Liu, S.Y., Sun, Z.Y., Liu, Y., Wang, H., Guo, P., and Murphy, R.W. (2012) A new vole from Xizang, China and the molecular phylogeny of the genus *Neodon* (Cricetidae: Arvicolinae). *Zootaxa* **3235**: 1-12.

Lomolino, M.V. (2001) Elevation gradients of species-density: historical and prospective views. *Global Ecology & Biogeography* **10**: 3-13.

Manley, B.F.J. (1986) *Multivariate Statistical Methods*. Chapman & Hall, London. pp 159.

Marshall, J.T. (1977) A synopsis of Asian species of *Mus* (Rodentia, Muridae). *Bulletin of the American Museum of Natural History, New York* **158**: 193-220.

Martens, J. and Niethammer, J. (1972) Die Waldmäuse (*Apodemus*) Nepals. *Zeitschrift für Säugetierkunde* **37**: 144-154.

McCain, C.M. (2004) The Mid-Domain Effect Applied to Elevational Gradients: Species Richness of Small Mammals in Costa Rica. *Journal of Biogeography* **31**: 19-31.

McCain, C.M. (2005) Elevational Gradients in Diversity of Small Mammals. *Ecology* **86**: 366-372.

McCain, C.M. and Beck, J. (2016) Species turnover in vertebrate communities along elevational gradients is idiosyncratic and unrelated to species richness. *Global Ecology and Biogeography* **25**: 299-310.

McCain, C.M. and Grytnes, J.-A. (2010) Elevational Gradients in Species Richness.

Mekada, K., Koyasu, K., Harada, M., Narita, Y., Shrestha, K.C., and Oda, S.-I. (2001) Faunal survey of small mammals in central Nepal, with reference to the distribution of the genus *Soriculus* (Insectivora, Mammalia). *Biogeography* **3**: 33-40.

Mena, J.L. and Vázquez Domínguez, E. (2005) Species turnover on elevational gradients in small rodents. *Global Ecology & Biogeography* **14**: 539-547.

Merow, C., Smith, M.J., and Silander, J.A. (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* **36**: 1058-1069.

Merritt, J.F. (1986) Winter Survival Adaptations of the Short-Tailed Shrew (*Blarina brevicauda*) in an Appalachian Montane Forest. *Journal of Mammalogy* **67**: 450-464.

Mitchell, R. and Punzo, F. (1976) New Mammal Records from Nepal. *Journal of the Bombay Natural History Society* **73**: 54-58.

Mitchell, R.M. (1977) *Accounts of Nepalese mammals and analysis of the host-ectoparasite data by computer techniques*. PhD Thesis. Iowa State University, Ames, Iowa, USA.

Molur, S., Srinivasulu, C., Srinivasulu, B., Walker, S., Nameer, P., and Ravikumar, L. (2005) *Status of South Asian Non-volant Small Mammals: Conservation Assessment and Management Plan (C.A.M.P.) Workshop Report.* Zoo Outreach Organisation / CBSG-South Asia, Coimbatore, India, . pp 618.

Moritz, C., Patton, J., Conroy, C., Parra, J., White, G., and Beissinger, S. (2008) Impact of a Century of Climate Change on Small-Mammal Communities in Yosemite National Park, USA. *Science* **322**: 261-264.

Motokawa, M., Harada, M., Mekada, K., and Shrestha, C. (2008) Karyotypes of three shrew species (*Soriculus nigrescens*, *Episoriculus caudatus* and *Episoriculus sacratus*) from Nepal. *Integrative Zoology* **3**: 180-185.

Mulungu, L.S., Makundi, R.H., Massawe, A.W., Machang'u, R.S., and Mbije, N.E. (2008) Diversity and distribution of rodent and shrew species associated with variations in altitude on Mount Kilimanjaro, Tanzania. *Mammalia* **72**.

Nakomoto, A. and Nakanishi, N. (2013) Home Range, Habitat Selection, and Activity of Male Asian House Shrews, *Suncus murinus*, on Okinawa-Jima Island. *Mammal Study* **38**: 147-153.

Nguyen, D.T. and Gomez-Zurita, J. (2016) Subtle Ecological Gradient in the Tropics Triggers High Species-Turnover in a Local Geographical Scale. *PLoS One* **11**: e0156840.

Niethammer, J. and Martens, J. (1975) Die Gattungen *Rattus* and *Maxomys* in Afghanistan und Nepal. *Zeitschrift für Säugetierkunde* **40**: 325-355.

Nor, S.M. (2001) Elevational Diversity Patterns of Small Mammals on Mount Kinabalu, Sabah, Malaysia. *Global Ecology and Biogeography* **10**: 41-62.

Novillo, A. and Ojeda, R.A. (2012) Diversity and distribution of small mammals in the South American Dry Andes. *Austral Ecology* **37**: 758-766.

Ohdachi, S.D., Hasegawa, M., Iwasa, M.A., Vogel, P., Oshida, T., Lin, L.K., and Abe, H. (2006) Molecular phylogenetics of soricid shrews (Mammalia) based on mitochondrial cytochrome b gene sequences: with special reference to the Soricinae. *Journal of Zoology* **270**: 177-191.

Olsen, A.R. (1984) Fecal pellets from a commensal shrew (*Suncus murinus*) and a house gecko (Gekkonidae). *Journal of the Association of Analytical Chemists* **67**: 1035-1036.

Oxford Nanopore Technologies. (2018) MinION. Oxford Nanopore Technologies. https://nanoporetech.com/products/minion

Paudel, E.N., Shrestha, K.K., and Bhuja, D. (2010) *Enumeration of Herbaceous Plants in Imja Valley, Sagarmatha National Park, Nepal*. Pages 173-188 *in* P.K. Jha and Khanal, I.P., editors. *Contemporary Research in Sagarmatha (Mt. Everest) Region, Nepal*. Nepal Academy of Science and Technology, Lalitpur, Kathmandu, Nepal.

Pearce, J. and Ferrier, S. (2000) An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling* **128**: 127-147.

Pearch, M.J. (2011) A review of the biological diversity and distribution of small mammal taxa in the terrestrial ecoregions and protected areas of Nepal. *Zootaxa* **3072** 1-286.

Pergams, O.R.W. and Lawler, J.J. (2009) Recent and widespread rapid morphological change in rodents. *PLoS One* **4**: e6452-e6452.

Rahbek, C. (1995) The elevational gradient of species richness: a uniform pattern? *Ecography* **18**: 200-205.

Rands, M., Bowden, C., Brown, P., and Newton, P. (1979) *UEA Nepal Expedition 1978/79*. UEA. pp 92.

Rickart, E.A. (2001) Elevational diversity gradients, biogeography and the structure of montane mammal communities in the intermountain region of North America. *Global Ecology and Biogeography* **10**: 77-100.

Rickart, E.A., Heaney, L.R., Balete, D.S., and Tabaranza, B.R. (2011) Small mammal diversity along an elevational gradient in northern Luzon, Philippines. *Mammalian Biology* **76**: 12-21.

Robins, J.H., Hingston, M., Matisoo-Smith, E., and Ross, H.A. (2007) Identifying *Rattus* species using mitochondrial DNA. *Molecular Ecology Notes* **7**: 717-729.

Rowe, R.J. (2009) Environmental and geometric drivers of small mammal diversity along elevational gradients in Utah. *Ecography* **32**: 411-422.

Rowe, R.J., Heaney, L.R., and Rickart, E.A. (2015) Scale effects on the pattern and predictors of small mammal diversity along a local elevational gradient in the Great Basin. *Journal of Biogeography* **42**: 1964-1974.

Rowe, R.J. and Lidgard, S. (2009) Elevational gradients and species richness: do methods change pattern perception? *Global Ecology and Biogeography* **18**: 163-177.

Sánchez-Cordero, V. (2001) Elevation gradients of diversity for rodents and bats in Oaxaca, Mexico. *Global Ecology and Biogeography* **10**: 63-76.

Sapkota, S., Pandey, B., and Shrestha, K.K. (2017) Diversity of Flowering Plants in Nubri Valley, Manaslu Conservation Area, Central Nepal. *American Journal of Plant Sciences* **8**: 1484-1498.

Shimada, T., Aplin, K.P., and Suzuki, H. (2010) *Mus lepidoides* (Muridae, Rodentia) of central Burma is a distinct species of potentially great evolutionary and biogeographic significance. *Zoolog Sci* **27**: 449-459.

Shuai, L.Y., Ren, C.L., Yan, W.B., Song, Y.L., and Zeng, Z.G. (2017) Different elevational patterns of rodent species richness between the southern and northern slopes of a mountain. *Sci Rep* **7**: 8743.

A.T. Smith and Xie, Y. *Eds*. (2008) *A Guide to the Mammals of China*. pp 544. Princeton University Press.

Sørensen, T.A. (1948) A method of establishing groups of equal amplitude in plant sociology based on similarity of species content, and its application to analyses of the vegetation on Danish commons. . *Kongelige Danske Videnskabernes Selskabs Biologiske Skrifter* **5**: 1-34.

Srivathsan, A., Baloglu, B., Wang, W., Tan, W.X., Bertrand, D., Ng, A.H.Q., Boey, E.J.H., Koh, J.J.Y., Nagarajan, N., and Meier, R. (2018) A MinION-based pipeline for fast and cost-effective DNA barcoding. *Mol Ecol Resour*.

Stanley, W.T. and Hutterer, R. (2007) Differences in abundance and species richness between shrews and rodents along an elevational gradient in the Udzungwa Mountains, Tanzania. *Acta Theriologica* **52**: 261-275.

Stanley, W.T. and Kihaule, P.M. (2016) Elevational Distribution and Ecology of Small Mammals on Tanzania's Second Highest Mountain. *PLoS One* **11**: e0162009-e0162009.

Stanley, W.T., Rogers, M.A., Kihaule, P.M., and Munissi, M.J. (2014) Elevational distribution and ecology of small mammals on Africa's highest mountain. *PLoS One* **9**: e109904-e109904.

Suzuki, H., Numone, M., Kinoshita, G., Aplin, K., Vogel, P., Kryukov, A.P., Jin, M.-L., Han, S.-H., Maryanto, I., Tsuchiya, K., Ikeda, H., Shiroishi, T., Yonekawa, H., and Moriwaki, K. (2013) Evolutionary and dispersal history of Eurasian house mice *Mus musculus* clarified by more extensive geographic sampling of mitochondrial DNA. *Heredity* **111**: 375-390.

Suzuki, H., Shimada, T., Terashima, M., Tsuchiya, K., and Aplin, K. (2004) Temporal, spatial, and ecological modes of evolution of Eurasian *Mus* based on mitochondrial and nuclear gene sequences. *Mol Phylogenet Evol* **33**: 626-646.

Tello, J.S., Myers, J.A., Macía, M.J., Fuentes, A.F., Cayola, L., Arellano, G., Loza, M.I., Torrez, V., Cornejo, M., Miranda, T.B., and Jørgensen, P.M. (2015) Elevational Gradients in β-Diversity Reflect Variation in the Strength of Local Community Assembly Mechanisms across Spatial Scales. *PLoS One* **10**: e0121458.

Tung, K.-C., Hsiao, F.-C., Wang, K.-S., Yang, C.-H., and Lai, C.-H. (2013) Study of the endoparasitic fauna of commensal rats and shrews caught in traditional wet markets in Taichung City, Taiwan. *Journal of Microbiology, Immunology and Infection* **46**: 85-88.

Vetaas, O.R. and Grytnes, J.A. (2002) Distribution of Vascular Plant Species Richness and Endemic Richness along the Himalayan Elevation Gradient in Nepal. *Global Ecology & Biogeography* **11**: 291-301.

Vieira, E.M. and Port, D. (2007) Niche overlap and resource partitioning between two sympatric fox species in southern Brazil. *Journal of Zoology* **272**: 57-63.

Wen, Z., Wu, Y., Cheng, J., Cai, T., Du, Y., Ge, D., Xia, L., and Yang, Q. (2018) Abundance of small mammals correlates with their elevational range sizes and elevational distributions in the subtropics. *Ecography* **41**: 1888-1898.

Wen, Z., Wu, Y., Du, Y., Xia, L., Ge, D., Yang, Q., and Chen, L. (2014) Seasonal Change of Species Diversity Patterns of Non-volant Small Mammals along Three Subtropical Elevational Gradients. *Biotropica* **46**: 479-488.

Whittaker, R.H. (1960) Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs* **30**: 279-338.

Wilson, M.V. and Shmida, A. (1984) Measuring beta diversity with presence-absence data. *Journal of Ecology* **72**: 1055-1062.

Wu, Y., Yang, Q., Wen, Z., Xia, L., Zhang, Q., and Zhou, H. (2013) What drives the species richness patterns of non-volant small mammals along a subtropical elevational gradient? *Ecography* **36**: 185-196.

Zhang, B., He, K., Wan, T., Chen, P., Sun, G., Liu, S., Nguyen, T.S., Lin, L., and Jiang, X. (2016) Multi-locus phylogeny using topotype specimens sheds light on the systematics of *Niviventer* (Rodentia, Muridae) in China. *BMC Evolutionary Biology* **16**: 261.

SUPPLEMENTARY INFORMATION

6.A Preliminary Analysis of Trap Exposure

The standard trapping grid comprised 18 traps laid for two nights, with two dawn and one dusk trapping sessions. On average, each dawn session lasted 16 hours and each dusk session lasted eight hours. This gave an "exposure" to small mammals of 40 hours per trap or 720 trap hours (TH) in a standard grid. However, this rate of exposure could have been affected by three factors;

- **Void Traps**. In Chapter 2 (Section 2.D.b) I analysed the rate and possible cause of void traps. In five cases this was due to "real" missing traps which had been stolen or deliberately damaged. However, traps that failed to trip, or were closed through outside agency could also be considered as "missing". Overall void trap rate was low (2.9%), but still resulted in reduced trap exposure.
- **Incomplete Trapping Sessions**. On three occasions, severe weather conditions prevented sessions being undertaken (see Chapter 2, Section 2.D.a). This can be considered an extreme case of missing traps, resulting in reduced trap exposure.
- **Additional Trapping Sessions**. In one site in Pipar in 2013 up to four additional trapping sessions were carried out (see Chapter 2, Section 2.D.a), which resulted in increased exposure.

The outcome of these factors was that the number of small mammals recorded in trapping grids could not be compared directly, without taking account of differential exposure. To do this, I calculated corrected trap hours (CTH) for each trapping grid. However, this could not be just a simple tally of traps available × session hours, because capture rates were intrinsically different in different sessions. For example, a missing trap in the first dawn session would have a different effect compared to a missing trap in the second dawn session, and especially in the dusk session, when capture rates were intrinsically lower.

To analyse this effect, I tallied the numbers of animals caught in each of the three sessions, excluding grids where three sessions were not carried out. Based on the assumption that void traps were distributed randomly, this indicated that numbers of animals caught in the three sessions were very significantly different from expected ($\chi^2_{(2)} = 120$, $p < 10^{-26}$; Figure 6.20a). This showed, that Session 1 (dawn) had a disproportionately greater number of animals captured and Session 2 (dusk) had far fewer. Even partitioning these data to compare the two dawn sessions only, showed a highly significant difference ($\chi^2_{(1)} = 50.7$, $p < 10^{-13}$). To allow for this effect, I calculated the effective hours (EH)

in a session to reflect these proportions, whilst keeping the total of the three sessions equal to 40. In this way, Session 1 had an effective time of 21.50 hours, Session 2 was 1.98 and Session 3 was effectively 16.52 hours long. I carried out a similar analysis using the grids from Pipar in 2013 (Figure 6.20b).

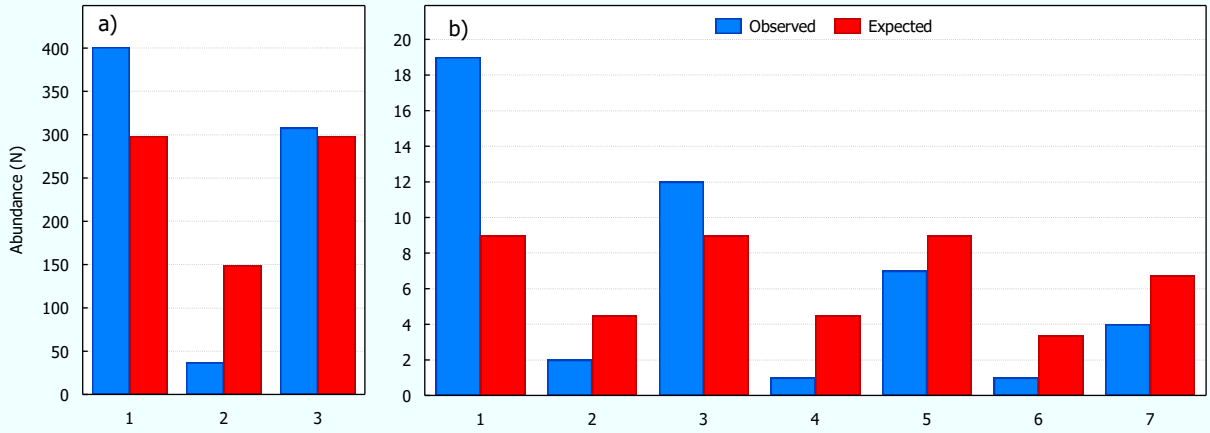


Figure 6.20 Total abundance by trap session for a) the all grids where three sessions were completed and b) the six grids where five or seven sessions were completed. Bars show observed counts and expected counts based on proportions of trap hours for each session.

The values for EH for each trapping grid session were multiplied by the number of effective traps in the session to give corrected trap hours (CTH). These were summed across sessions to give total CTH for the trapping grid.

6.B Preliminary Analysis of Annual Variation in Numbers of Animals Caught

The Pipar transect was visited twice (2013 and 2014) to obtain an indication of the variation in numbers of animals across years. However, the two visits differed in the number of sites and the number of sessions carried out in one of the sites (see Chapter 2). By removing these discrepancies, there were four sites and 21 grids visited in both years. After removing the additional sessions carried out in Site 4 in 2013 (see above) this left a total of 84 dawn sessions and 42 dusk sessions.

During these sessions, a total of 238 animals were captured; 82 murids and 154 shrews. I analysed these data using a two-way repeated-measures ANOVA, with year as the within-subjects factor and site as the between subjects factor. There was no significant difference between numbers of animals caught in the four sites ($F_{(3, 17)} = 3.053$, $p > 0.05$). However, there was a significant difference in the number of animals caught between years (mean of 2013 = 4.53, mean of 2014 = 6.89; $F_{(1, 17)} = 8.438$, $p < 0.01$). However, there was a marginally significant interaction between site and year ($F_{(3, 17)} = 4.096$, $p < 0.025$; Figure 6.21). This showed that the annual difference was only significant in Site 1. To simplify the subsequent analyses, I tallied the number of animals per grid (by species) across

the two years in these four sites. I also calculated the sum of the CTH per grid across the two years as the offset value.

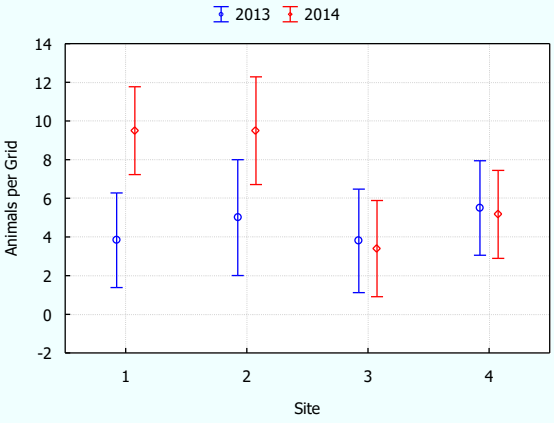


Figure 6.21 Number of animals caught on the Pipar transect in four sites over two years.