USING COMMENSALS AS PROXIES FOR HISTORICAL INFERENCE IN THE INDIAN OCEAN: GENETIC AND ZOOARCHAEOLOGICAL PERSPECTIVES

Heidi M. Eager

Linacre College

DPhil Archaeological Science, Hilary 2014

ABSTRACT

The human-abetted introduction of commensal species (i.e. those that opportunistically exploit the anthropogenic environment for food and shelter, e.g. rats, cockroaches etc.) to new areas has occurred throughout history. This has resulted in detrimental ecological changes worldwide but, from a viewpoint of human knowledge, a beneficial corollary of these translocations is that the species in question can be used as proxies to study the movement of the humans who transported them. I reconstruct colonisation histories of three widespread commensal mammalian species in the Western Indian Ocean, the black rat *Rattus rattus*, house mouse *Mus musculus* and Asian house shrew *Suncus murinus*, through phylogeographic studies (the geographic distribution of genetic lineages) of maternally-inherited mitochondrial markers, and zooarchaeological data. The DNA analyses are conducted on samples largely derived from museum specimens collected up to 110 years ago, and from archaeological bones (in the case of rats). I show considerable cryptic diversity in all three species, particularly in mice for which we find a potential major new lineage. Certain lineages within each species predominantly reveal long-distance translocations within the Indian Ocean, but high resolution geographic and genetic clustering is also evident, particularly in Asian house shrews. Phylogeographic structuring of the three species in East Africa and the southern Indian Ocean region (e.g. Madagascar, Reunion, etc.) indicate connections with Arabia, the Middle East, and India in the Islamic period from the first millennium AD, and later European connections during the Age of Exploration. Closer to the origins of the three species (the Indian subcontinent in all cases), range expansions in Eurasia and nearby islands relate to early to mid Holocene human populations, but also with signals of later secondary colonisations. Through ancient DNA studies I found genetic continuity between temporally separated populations of black rats suggesting population persistence, and high levels of diversity in Songo Mnara, a Swahili stonetown in

Tanzania. Knowledge of the colonisation history and genetic diversity of an introduced species is essential to understand their resilience in novel landscapes, and to identify pathways of invasion and, by proxy, human trade and exchange networks that facilitated their dispersal. My research contributes significantly to that end for three socially, economically and ecologically important species that are well-established in the Indian Ocean region and beyond.

PREFACE

My research forms a part of a larger multidisciplinary project spearheaded by Dr. Nicole Boivin (University of Oxford) called the Sealinks Project (www.sealinksproject.com). The project investigates early maritime connections in the Indian Ocean and seeks to understand the socio-cultural, economic and environmental impacts of these connections through the synthesis of archaeological, genetic, linguistic and palaeoecological studies. My research focuses on biological proxy lines of evidence and was guided by Dr. Jeremy Searle (Cornell University), through his interest in evolutionary biology of small mammals and their application in tracing human histories (www.sites.google.com/site/thesearlelaboratory). It quickly became a collaboration involving many archaeologists, biologists, morphometricians, historians and field researchers, and was especially influenced by members of Crossing the Green Sea project and the D.E.A.D Lab based in Durham and Aberdeen Universities. My research benefitted immensely from the ancient DNA expertise of Dr. Greger Larson (Durham University). The value of such a deeply collaborative project, involving such a diverse range of specialists is made clear by the influence it has had on the genuinely interdisciplinary nature of my research, particularly within the fields of genetics and zooarchaeology.

ACKNOWLEDGEMENTS

For my dad, in loving memory: with alacrity, equanimity, and occasionally puerility.

Acknowledging all the deserving people who have helped progress my DPhil, both academically and personally, would be a tome in itself, so instead I am inclined to keep it brief. My mum and sister deserve great recognition for always offering love, help and encouragement, especially over the last few years. My deepest thanks go to my supervisors, Nicky and Jeremy, who offered me a great opportunity and let me follow my interests in a supportive, unconstrained manner. It is only through the academic, technical and personal support from fellow PhD students, group members, researchers, field biologists, archaeologists, to name but a few, that I have been able to complete my thesis, and I would like to thank the Searle Lab, Sealinks crew, Steve Goodman and Martin Walsh in particular for their help. I am very grateful to the Cornell University community, who welcomed me from across the pond, and to Greger Larson and the DEAD Lab, who welcomed me up North. I appreciate that I could not have undertaken my research if it were not for the museums, archaeologists and biologists who provided me with samples. To my friends - thank you for keeping me sane.

DECLARATION OF OWN WORK

This research has been a collaboration involving many researchers and institutions, in particular two other research students, Alex Trinks (Durham University) and Ardern Hulme-Beaman (AHB) (University of Aberdeen). AHB contributed to molecular work for the Asian house shrew (Chapter 3) by co-designing Primer Set 1, co-extracting museum specimens, and by extracting and amplifying the bulk of modern specimens. In each chapter, I have acknowledged the work that is included in my thesis that is done by other parties (except as outlined here). All analysis and all written work are my own.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Figures

Tables

CHAPTER 1

INTRODUCTION

OVERVIEW

The Indian Ocean and the landmasses that surround it form a remarkable historical arena. The region is the home of some of the earliest 'civilisations', including those of the Indus Valley, Mesopotamia and Egypt and the remains of the earliest planked wooden boats (*ca*. 3000 BC) and commercial seagoing ships (late 3rd millennium BC) were found on the Nile and Red Sea coast (Ward 2012). The Indian Ocean was the first maritime area to witness long-distance contact and exchange between diverse cultures and the area as a whole is considered by many as the 'cradle of globalisation' (Mitchell 2005; Jamal & Moorthy 2012). The Indian Ocean is bounded by Africa in the west, Southeast Asia and Australia in the east, and Asia in the north. By mere mention of the surrounding continents we can plainly see the region is not only environmentally and climatically varied but also culturally and biologically diverse.

With such a long pedigree it is somewhat surprising that there are still major lacunae in our knowledge of the prehistory of the Indian Ocean. What is clear is that since the time of the earliest civilisations maritime exchanges were significant in the spread of flora and fauna, luxury and essential goods, raw materials, and people, and these introductions have had significant environmental, cultural, socio-political, religious and economic implications. Evidence for translocations is often gleaned from genetic, linguistic, or palaeoecological data and other proxy lines of evidence which complement, and sometimes challenge, the hypotheses described from traditional approaches in archaeology and history (Fuller *et al.* 2011).

My thesis explores trade and contact networks in the Indian Ocean using biological proxies, and evaluates the utility of such proxies for reconstructing human histories in the region. The research methodology employs maternally-inherited genetic markers to reconstruct genealogies of species and embed them in a geographical context. I characterise the relationship between the genetic diversity and geographic distributions of the proxies and seek to understand the evolutionary implications of their translocations. The rationale for my thesis stems from the mutual history that humans share with various other species brought about by intimate relationships, whether intentional or unwitting. Human-animal relationships occur along a spectrum from the process of domestication (such as those of herd animals, crops or pets), through passive symbioses (from the human point of view) such as parasitism (including zoonoses and long-lasting parasitic and pathogenic associations) and commensalism (involving species that are not direct parasites but which still opportunistically exploit humans and are often considered pests). Ultimately, many human-related drivers (e.g. artificial selection, migration, disease-resistance, among an almost infinite list of others) will affect the dynamics of populations of species with which they are intimately associated. The species in question, called a bioproxy, can then be considered analogous to material artefacts in archaeology in that it contains information relevant to the humans it was associated with (Jones *et al.* 2013). Empirical data relating to the species, such as fossils or historical documents, can be used to reconstruct the historical relationship, however a more powerful approach resides in unlocking the secrets held in an individual or population's DNA through genetic analyses.

Although the field of human genetics is increasingly powerful with substantially more complex models based on whole-genome data, and ever-increasing numbers of individuals studied (e.g. The International HapMap Project, The African Genome Variation Project (International HapMap Project 2003; Gurdasani *et al.* 2014)), the role of bioproxies is to both complement and act as an alternative for human genetic information (Jones *et al.* 2013). The colonisation of the bioproxy species may co-occur with the colonisation of the humans who took them, or the bioproxy species may colonise after human contact (e.g. through trade, pilgrimages, voyages of discovery, etc.) where humans have left no genetic trace. Where humans have colonised an area, bioproxies can be particularly informative in cases where human populations have either moved on or died out, or where multiple immigration events obscure genetic histories of populations.

One such arena where bioproxies have been used to trace the prehistoric movement of people is in the Pacific Ocean. The model most successfully used in the region is the Pacific rat *Rattus exulans,* which is hypothesised to have been carried by ancient seafarers as a food item, and is now widely distributed throughout the region (Matisoo-Smith & Robins 2004). Through the synthesis of archaeological data and maternally-inherited gene genealogies, including ancient DNA, researchers reconstructed the ancient movements of Pacific rat populations and, by proxy, the humans who carried them (Matisoo-Smith & Robins 2004; Barnes *et al.* 2006; Robins *et al.* 2007). Significant findings of the colonisation model show that one genetic clade of the Pacific rat was associated with the Lapita Cultural Complex, a prehistoric culture of Near Oceania (*ca.* 1500 BC) and, furthermore, it was introduced by the Lapita to Remote Oceania in a complex diffusion associated with the initial peopling of the region (Matisoo-smith & Robins 2004). The complexity of interactions between islands over time shown from the rat genetic data refuted simplistic models of human colonisation proposed for the region, including the “express train” theory that postulates a rapid migration out of Taiwan, through Melanesia, to Remote Oceania.

The role of the house mouse as a proxy for human mobility has been primarily exploited within the western house mouse subspecies *Mus musculus domesticus* (Bonhomme & Searle 2012; Jones *et al.* 2013). Using zooarchaeological data and mitochondrial DNA sequence variation several studies have built a model of house mouse translocation related to human movement patterns (Cucchi *et al.* 2005b; Rajabi-Maham *et al.* 2008;

Jones *et al.* 2010b; Bonhomme *et al.* 2011). Here, I focus on one hypothesised network but for an in-depth review of using bioproxies for historical inference with a focus on house mice see Jones *et al.* (2013), and Chapter 6. Based on mitochondrial genealogies, researchers hypothesise that the house mice was introduced to mainland Scandinavia, Iceland, Greenland, Ireland and the British Isles with Vikings *ca*. 1000 year BP (Searle *et al.* 2009; Jones *et al.* 2010a, 2012). What is perhaps more surprising is that individuals from the Madeiran archipelago off the coast of northwestern Africa in the Atlantic Ocean also share genetic affinity with 'Viking mice' (Förster *et al.* 2009; Searle *et al.* 2009). Madeira was discovered by Portuguese explorers in 1419 AD and has remained closely connected with Portugal ever since. Despite this close connection no genetic affinity (of the gene studied) is shared between mice originating from Portugal and Madeira. It is suggested the introduction of house mice to Madeira could be due to 9th century AD Vikings making an unexpected detour while sailing to the Mediterranean, although as yet the archaeological and historical evidence do not support such a visitation (Jones *et al.* 2013).

A VERY BRIEF ACCOUNT OF INDIAN OCEAN HISTORY

Parts of the Indian Ocean were traversed in ancient times by relatively basic sea-going vessels (such as reed boats or outriggers) and, over time, much of the littoral has been linked by economic and social maritime networks. Various inlets and subdivisions of the ocean such as the Red Sea, Persian Gulf, Arabian Sea and Bay of Bengal are integral domains within networks that facilitated the development of 'interaction spheres' (Gupta 2004; Beaujard 2005). Networks also influenced, and were influenced by, communities of the interior, which were reached through road and importantly riverine routes such as the Zambezi, Shatt al-Arab, and Indus.

Although trade between various coastal communities and states had occurred prior, the late first millennium BC/early first millennium AD saw the intensification of trade and cross-cultural interactions in the western Indian Ocean as maritime technologies advanced. From the 6th century BC the Persian Empire linked the Mediterranean Sea to the Indian Ocean via trading networks that extended to India, and the Persian Empire would later be controlled by Alexander the Great (Beaujard 2005). By the 3rd century BC unification of both India and China as well as the Greco-Roman control of western Asia allowed both terrestrial and maritime networks to open up so that by the first century AD the Silk Route and other major trade routes in the Indian Ocean would link East Asia, through the northern part of the Indian Ocean to the Mediterranean, and south to sub-Saharan East Africa (Beaujard 2005).

The collapse of the classical era's 'globalised' economy saw a transformation in the players controlling and contributing to international trade. By the 4th - 5th centuries AD India, flourishing under the Gupta Empire, became the principal nexus in the ocean network exchanging goods and ideas predominantly with South East Asia but also to the west (this period is supposed by some scholars to be a period of "Indianisation" of Southeast Asia) (Gupta 2004; Beaujard 2005). Indian trade with Southeast Asia and China via the Bay of Bengal was maintained throughout the first millennium AD (Wade 2009) . As the west recovered from recession and demographic declines there was a concomitant increase in the volume and frequency of trade which was set amidst a far-reaching network that extended from Africa in the west to South East Asia in the east and beyond to China. From around the 7th/8th century AD until the mid second millennium AD, Arabian merchants dominated trade in the Arabian Sea, extending far south below the Horn of Africa and reaching previously isolated islands such as Madagascar (Horton & Middleton 2000). From the Age of Exploration in the 15th century under the supremacy of Europeans and onwards, all parts of the Indian Ocean were traversed by explorers, merchants or whalers.

While civilizations, empires and states have all played a central role in the maritime networks that spanned the Indian Ocean, more marginal societies have also been actively involved but have either left an ephemeral archaeological or historical trace, or have not been well-studied. I use the term ‘marginal’ here not in terms of importance but rather in terms of their geography, or the degree to which they have been studied. From the brief account I have given clearly sub-Saharan African cultures are not identified as central agents in maritime networks. As these regions become more intensively studied and evidence for prehistoric links with the wider Indian Ocean world start to accumulate, it is clear that these cultures played a more crucial role in maritime networks than previously assumed (Glover & Bellwood 2004; Mitchell 2005; Helm *et al.* 2012). There is also mounting evidence that small-scale maritime societies actively contributed to cultural exchanges in the pre-Roman era, and in many cases were the pioneers of such exchanges (Fuller *et al.* 2011).

THE COMMENSAL MODEL

Although various categories of bioproxies such as domesticates or pathogens could be used for this study I use commensals because they are widespread in the Indian Ocean, are relatively easily collected, and are usually unintentionally transported and thus evolving under natural scenarios thereby avoiding the biases of artificially selected lineages and extinctions.

There are many biological definitions of the term ‘species’, which each attempt to delimit criteria – some theoretical, and some in order to make the definition operational. Various definitions encompass the concept in both a broad or narrow sense. For my thesis, I acknowledge that the theoretical biological species concept proposed by Mayr (1970), which recognises that interbreeding/reproductive isolation demarcate species, has influenced my concept of species greatly yet has serious limitations in that gene flow occurs in natural populations and hybrids of the species I study are known to occur. Therefore, for the purposes of being able to functionally delimit species, I utilise the genotypic cluster concept as outlined in Mallett (1995), which defines a species as a "distinguishable groups of individuals which have few or no intermediates when in contact". It incorporates genetic and morphological information and thus 'species' will be quantifiable as a group and will exhibit a unimodal clustering in these traits. To extend the definition to allopatric populations, tests where groups come into secondary contact should be made and species will be defined by those genotypic clusters that do not maintain fusing with its sibling. In sympatric populations species are maintained not only by low levels of gene flow (potentially due to reproductive traits or ecological adaptations) but also by the strength of selection, mutation and drift – species under this concept can therefore hybridise.

What is commensalism?

In its broad ecological sense, commensalism is a relationship between two species where their interaction has a positive effect on one species and a neutral effect on the other species. In the narrower sense (the definition my dissertation follows), commensalism is a relationship characterised by a species living in close association with humans, where the species opportunistically exploits humans without acting as a parasite or pathogen (Jones *et al.* 2013). It comes from the Latin *cum mensa* meaning "at the same table". The organism derives a benefit from the human environment, either through resource provision (food, shelter), protection from predators, or a reduction in competition. The definition is restricted to a small subset of species that have exploited this association and, as such, have become cosmopolitan in their distribution and often economically significant. Mammalian commensals include the globally distributed black rat *Rattus rattus*, Norway rat *Rattus norvegicus* and house mouse *Mus musculus* and a handful of small mammals with regional distributions, for example, the Pacific rat *Rattus exulans*, the Himalayan field rat *Rattus nitidus*, the

Turkestan rat *Rattus pyctoris* and the house shrew *Suncus murinus* (Aplin 2003). There are also non-mammalian commensals such as the house gecko *Hemidactylus frenatus*, house sparrow *Passer domesticus* and the rock pigeon *Columba livia*. In the Norway and Pacific rats the commensal relationship has developed into an obligate affiliation to the extent that it is uncertain whether these species still exist in their wild state (Aplin 2003).

How and where did commensalism evolve?

The first mammalian commensals that have been identified in the archaeological record date to the Natufian period (ca. 12500 to 9500 BC, based on calibrated radiocarbon dates) of the Levant. The house mouse first appears in Natufian deposits at Hayonim Cave and the black rat first reliably appears in Natufian deposits at Kebara Cave (Tchernov 1984; Auffray *et al.* 1988). Both sites are in the coastal hinterlands of Northern Israel. It is not possible to ascertain the distributions of either species before their association with humans, however, the Israeli deposits of the black rat are the most westerly deposits of the species by the Natufian period and the house mouse first appears in the Middle East around 10,000BC (Auffray *et al.* 1988).

The Natufian culture of the Levant is characterised by the emergence of a sedentary or semi-sedentary hunter-gatherer lifestyle and the cultivation of cereals (Byrd 1994). The proto-Neolithic culture has implications for the evolution of commensalism in that as people became more sedentary, built dwellings and started to cultivate and store grain, the conditions favoured wild species that were able to adapt to the new environment. The anthropogenic environment provided resources of food and shelter, and a reduction in interspecific competition, as wild species not able to adapt to the new conditions were excluded (Auffray *et al.* 1990). From their associative beginnings with humans, many species evolved to become facultative or obligate commensals, which led to morphological, behavioural and genetic changes (Auffray *et al.* 1988).

Based on their distributions, it is clear that commensalism evolved in different areas for different taxa. Human settlement patterns and agricultural development in the Fertile Crescent and the Mesolithic and Neolithic of Asia and Africa created the anthropogenic habitats required for the commensal association to evolve. Whether commensalism evolved only once in each species is less clear. The archaeological record is often scant with respect to small mammal assemblages due to factors associated with deposition and retrieval. Further, there are geographical biases in the record: small mammals appear frequently in Levantine sites but not in Arabian sites, which is due, in part, to preservation conditions. However, house mice genetics and zooarchaeology suggest that, at least in this species, multiple, independent commensalisations of a species can occur. The house mouse has four distinct genetic lineages (classified as subspecies) that independently spread from South Asia well before the beginning of the Neolithic transition, and now occupy largely distinct geographical ranges from each other (except for narrow hybrid zones). The earliest commensal mice (from the Neolithic of the Levant as described above) are of the western house mouse subspecies *Mus musculus domesticus,* identified through characteristic molar tooth morphotypes. A second, independent process began with the eastern house mouse *M. m. musculus* in Europe a few thousand years later (Cucchi *et al.* 2012). The other two major lineages remain unstudied. Therefore, as each lineage had genetically evolved before the first evidence of commensalism, and as early allopatric populations are associated with human cultures in different regions and periods (at least for *domesticus* and *musculus*) we can deduce that commensalism evolved independently in each lineage.

Why are commensals so successful?

The ability of an invader to colonise a new area (its 'invasiveness'), the ways in which it is moved by people (its 'transportability') and the receptiveness of the invaded community (its 'invasibility') all contribute to the success or failure of an invading individual or population (Anderson 2009). Commensal species are successful invaders because of a number of characteristics and adaptations that are advantageous for these traits.

Habitat use in commensals is less restricted and they are therefore able to utilise a wider set of biological and physical resources than their wild counterparts for foraging, nesting, protection, etc. (Braithwaite 1980). The anthropogenic environment is less seasonal than the natural environment, environmental conditions and food supply are more stable and resources such as shelter and food are often abundant, although humans or livestock can cause disturbances for commensal populations (Pocock *et al.* 2004). True commensal populations depend on humans for shelter and food and are regularly found inside human dwellings, food stores, waste dumps, farm buildings, etc. and consume grain and other agricultural produce, or household food and scraps. Feral populations, on the other hand, revert to living in natural or semi-natural environments and are able to utilise wild resources. Animals returning to the feral state tend to do so under conditions where resources are stable and competition from other species is weak (Pocock *et al.* 2004). Some species benefit from an association with humans but are not truly commensal. These associations are by convenience and do not persist and thus are deemed anthropophilic (e.g. the Australian swamp rat *Rattus lutreolus* (Braithwaite 1980)).

Commensal species owe much of their success to their plasticity and adaptability. They are able to adapt to severe conditions at the population and individual level and are able to use a wide variety of resources. Under the null expectation, founder effects and isolation in introduced populations will lead to a reduction in genetic diversity and an increase in differentiation between populations. Commensal populations tend to start founding colonies with few individuals and therefore we expect commensal populations to exhibit the above genetic traits (e.g. Brouat *et al.* (2007)). However, recruitment and migration can prevent isolation and contribute towards higher genetic diversity, as was found to be the case in a population of Norway rats (Gardner-Santana *et al.* 2009).

Adaptations to the human environment in commensal rodent populations include behavioural and biological traits that affect population structure and dynamics.

Compared to their feral or wild counterparts, commensals may exhibit (at least in some species): higher density populations (Gardner-Santana *et al.* 2009); more complex social structures (i.e. demes with more individuals to a unit) (Boursot *et al.* 1993; Pocock *et al.* 2005); higher reproductive rates (Boursot *et al.* 1993; Pocock *et al.* 2005); higher mortality (Pocock *et al.* 2004); lower dispersal rates and dispersal at shorter distances (Pocock *et al.* 2005); female-biased sex ratios (Duplantier *et al.* 2002); year-round breeding (Braithwaite 1980; Pocock *et al.* 2005); more rapid growth of juveniles and earlier sexual maturity (Braithwaite 1980); larger body size (Braithwaite 1980); less severe population depressions (Chernousova 2001); and higher population turnover (Pocock *et al.* 2004).

The economic and social cost of commensalism

Biological translocations have undoubtedly changed the face of the planet and the role of commensals in altering ecology and human society, economy and health is severe. Commensals are often pests to home and agriculture, transmitters of disease, or affect local habitats and wildlife. Space does not permit me to fully review these aspects but I will briefly give some examples.

It is their reliance on human food resources and shelter, and their invasibility and invasiveness that make commensal species considerable pests and economically important species for humans. Consider, for example, the house mouse in Australia, which exploits outdoor crops. Under certain climatic conditions, populations can reach plague proportions with mice numbers in excess of 800 individuals per hectare (and up to 2700 per hectare) with consequent severe economic losses (Singleton *et al.* 2005). From their introduction with Australian settlers from the British Isles, mice quickly invaded the country (Gabriel *et al.* 2011). A 1917 newspaper details the huge haul of mice trapped during a plague with a headline that read *Mice plague: Millions slaughtered, millions coming* (Unknown 1917) (see also Figure 1.1):

> As a result of the campaign for the protection of wheat stacks mice are being slaughtered in millions. Every wheat stack is swarming with the creatures and at places where they have been counted the captures are almost incredible. At Minyip station 90,000 represented the bag for three nights. Two drays were used to cart them to the depot, one load turning the scales at 16cwt 2qrs 14lbs.

PLAGUE OF MICE IN THE WHEAT DISTRICTS OF AUSTRALIA: A CARTLOAD OF THE RODENTS CAUGHT IN TWO NIGHTS AT MINYIP, IN THE GIPPSLAND DISTRICT, VICTORIA.

Figure 1.1 A 1917 newspaper shows part of a haul of mice from Victoria (courtesy of Sir George Grey Special Collections, Auckland Libraries, AWNS-19170517-40-6).

Although commensal species are relatively few, they are some of the best known to humans, in part because of their role in the spread of disease. While endemic fauna can also host and transmit zoonotic diseases, it is their close association with humans that makes commensals particularly important vectors for disease transmission in both humans and livestock. Transmission can occur through contamination of food or living quarters by urine or faeces, or by direct transmission through bites or arthropod vectors. There are dozens of zoonoses associated with commensal rodents. Human pathogens transmitted by rodents include hantaviruses (Bunyaviridae)*,* plague *Yersinia pestis,* rat bite fever *Streptobacillus moniliformis* and *Spirillum minus***,** Leptospirosis

Leptospira interrogans , Toxoplasmosis *Toxoplasma gondii,* Rickettsial pox *Rickettsia akari* and Salmonellosis *Salmonella* spp. to name a few.

Commensal species are, by definition, introduced species when outside their native range. The effect of introduced species on non-native environments and biota can be severe. Predation of eggs and chicks by Norway rats was responsible for massive declines of seabirds on the uninhabited, subantarctic Campbell Island that destroyed local populations of a rare flightless duck (Campbell teal *Anas nesiotis*), and other local wildlife. The rats were likely introduced by whaling ships in the 19th century but in 2001 the New Zealand government exterminated the entire population of rats by dropping poison on the island. The successful eradication has since seen the Campbell teal and other birds recolonise the island ("Campbell Island: New Zealand's subantarctic islands").

THE COMMENSAL BIOXPROXY FRAMEWORK AND MY RESEARCH

Intrinsic in the commensal relationship is a virtually obligate reliance on humans by the commensal population. Consequently, the history of a commensal organism is intricately tied to that of the human population it is, or was, associated with. Accordingly, commensal histories can be used to infer a great deal about human history, providing insights into, for example, the early biogeography and structure of human populations, the spread of agriculture, the intensity and timing of human colonisation or cultural exchange/trade, and human-mediated impacts on native environments and biota (Wyncoll & Tangri 1989; Cucchi *et al.* 2005b; Storey *et al.* 2007; Anderson 2009).

Functionally, the ability to use commensals in historical inference lies with the embedding of the species within a phylogeographical framework. Phylogeography uses evolutionary relationships and geographical distributions of taxa or populations to infer patterns of diffusion (Avise *et al.* 1987). The inference depends on estimation of phylogenetic trees (branching networks that describe genealogical relationships) and so the majority of studies have, until relatively recently, used mitochondrial DNA sequence data, often single genes, as recombination does not contribute to gene variability. With advances in next-generation sequencing technologies and the development of coalescent, quantitative and model-based methods the field has progressed to allow for multi-locus frameworks and more complex demographic scenarios (Sites & Morando 2009). Until very recently, these technologies have required high quality DNA, thus precluding the analysis of ancient or degraded remains. My research utilises predominantly ancient DNA and traditional sequencing methods for the basis of single gene phylogeographic analyses and thus complex demographic models are precluded. Phylogeographic studies of commensal organisms from gut bacteria to plants and animals have all successfully been used to reconstruct prehistoric and early historic human histories (Matisoo-smith *et al.*; Linz *et al.* 2007; Hunt & Jones 2008; Jones *et al.* 2012).

In this study I draw on analyses of three commensal species to reconstruct human movements, trade and contacts in the Indian Ocean. Using phylogeographic studies based on maternally-inherited genetic markers of mostly extant populations I reconstruct the colonisation histories of the species and, by proxy, the humans who carried them. To strengthen historical inference I incorporate, where possible, data from ancient bone samples collected from archaeological excavations. I anchor my interpretation of the genetic information in the context of zooarchaeological finds and current knowledge from archaeological and historical sources.

I use three commensal models that are now established invasives, the black rat (*Rattus rattus*), the house mouse (*Mus musculus*) and the house shrew (*Suncus murinus*) (see Figure 1.2 and 1.3). A description of each species, its ecology, life history, behaviour and relationship to humans is given in Table 1.1. Although a number of commensal species are available, the three used here are good candidates for phylogeographic studies within the Indian Ocean region as they are widely distributed, particularly in coastal areas, and are mammalian systems that benefit from r-selected life history strategies (e.g. high fecundity and short generation times), thus genetically diverging over the relatively short time scale studied here. Furthermore, they are highly unlikely to be able to naturally cross long distances over water (e.g. swimming or rafting) and therefore colonisations of islands and distant landmasses are almost certainly human-aided.

Figure 1.2: (From left to right) Black rat *Rattus rattus,* house mouse *Mus musculus,* house shrew *Suncus murinus.* Images by Mike Read/naturepl.com (rat) and Ernie Janes/photoshot.com (mouse) courtesy of Arkive.org, and mekshat .com (shrew).

All three species are highly adaptable in their habitat requirements and, depending on geographic location, can be found in agricultural and urban domains, primary and secondary forests, deserts and almost any other environment barring polar conditions. One only need look at the number of synonyms listed - Musser and Carlton (2005) list 83 synonyms for the black rat, 59 for the house shrew and 123 for the house mouse - to

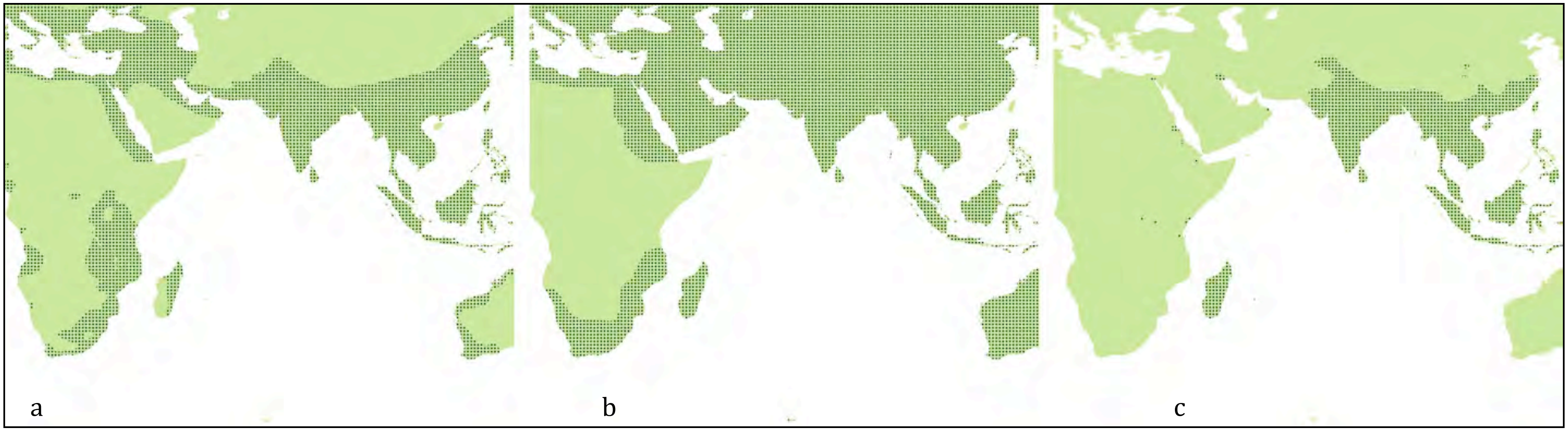


Figure 1.3 Indian Ocean distribution maps (shown as stipple) of a) black rat *Rattus rattus* b) house mouse *Mus musculus* and c) house shrew *Suncus murinus.* Data from the IUCN (IUCN 2011).

	RATTUS RATTUS (MURIDAE)	*MUS MUSCULUS* (MURIDAE)	*SUNCUS MURINUS* (SORICIDAE)
Common names	Black rat, ship rat, roof rat, house rat	House mouse, common mouse	(Asian) house shrew, grey musk shrew, (Asian) musk shrew, money shrew
Description	Head and body length (160-220mm), tail length (190mm+, longer than the body), weight (70-300g). Males are larger than females. Although usually black it can also range from dark to light brown and usually has lighter coloured underparts. It looks similar to the Norway rat *Rattus norvegicus* but is smaller with bigger eyes, ears and a relatively longer tail.	Head and body length (65-95mm), tail length (60-105mm), weight (12-30g). Morphologically house mice are similar to non-commensal mice (e.g. *Mus spretus*) but can be distinguished by the length of their tails, non-commensal species having shorter tails. House mice range in colour from light brown to almost black and usually have lighter coloured underparts in white, grey or light brown.	Head and body length (100-150mm), tail length (short), weight (23.5 g to 82.0 g (f), 33.2 g to 147.3 g (m). The house shrew is a small mammal similar in appearance and often mistaken for a rat, although it is not a rodent. It is has some distinguishing features such as a long, pointed snout, pointed teeth, small eyes and a short tail that is thicker at the base. It is a highly morphologically variable species. Pelage colour and body size varies between localities. They range

			in colour from a very light to dark grey. Males are generally larger than females. It has well-developed musk glands (variable in females) that produce a strong odour, which is probably used for marking territories and for defense.
Ecology, life history and behaviour	Black rats are agile climbers (aided by their long tail) and prefer to nest in the upper parts of structures such as roofs, or in trees, although under certain environmental conditions they may occupy burrows or nest at ground level. Nests are built from leaves and sticks. The species is usually nocturnal. It is an omnivore and although it has a preference for fruit, seeds, grain, leaves and a	House mice occupy a variety of structures such as houses, barns and granaries and can also live outside in fields, when conditions allow. In some cases they can seasonally switch between indoor and outdoor living. House mice are agile climbers, jumpers and can swim well. They generally nest in available hidden niches but can also burrow. Nests are built from a variety of materials,	House shrews are largely terrestrial and nest at ground level. Nests are built from leaves and sticks when outside, or from a variety of materials (including manmade materials) when indoors. The species is nocturnal. It is mainly insectivorous but is also an opportunistic feeder so that they also consume grain and human food resources on occasions. They have a high metabolic rate and need

	variety of other vegetation it will also consume invertebrates, rubbish and human food resources. It generally requires a free water need of approximately 15ml per day but in some cases the high water content of their food may be adequate. Under suitable conditions, black rats can breed throughout the year - typically summer and autumn are the peak breeding seasons. Up to five litters per year are produced with sexually maturity reached at age 3-5 months. Black rats form social groups with a dominant male, and a hierarchy of both females and males. It is a polygynous species with the dominant males usually securing	including man-made materials such as cloth and paper. The species is usually nocturnal but can forage during the day when inside. Mice are omnivores and in the natural environment eat leaves, seeds roots and other plant material as well as invertebrates and sometimes carrion. Commensal mice take human food including grain and can store food for later use. Mice breed throughout the year when conditions allow. Up to fourteen litters per year (although generally five to ten) can be produced with three to twelve individuals per litter. Sexual maturity is reached at 5-7 weeks. The maximum life span of a free-living mouse can be 12-18 months but is usually much less.	multiple feeds throughout the night. House shrews reproduce year round but with peaks in spring and summer and have litters of four to eight. Sexual maturity is reached at around age 35 days. The wild lifespan can be 1-2 years but is usually much less. Little is known about the social organisation or behaviour of house shrews. They are thought to be solitary and not tolerant of other individuals.

	most mates. Territories (encompassing food resources) are defended. It has an estimated home range of a maximum of 100m^2 and females are reportedly less mobile than males.	House mice are polygynous. Social groups and territories can be complex, with multiple individuals forming a deme consisting of a dominant male, several females and their young. Subordinate males can sometimes occupy the territory of a dominant male, although territories are usually defended.	
Human relationship	Pests to crops (particularly fruit trees); spoil food resources through urination and defecation; vector of diseases; implicated in the spread of the black death *Yersinia pestis* in the Middle Ages	Used as laboratory models; pets; pests to home and agriculture through destruction of buildings, fittings, and consumption of food; contaminate food through urination and defecation; vectors of disease.	Used as laboratory models; pest control (eat insects from around human environments); pests to certain crops (through consumption and nesting); pests in dwellings (mainly through nesting).

Table 1.1 Description of Black rat *Rattus rattus,* house mouse *Mus musculus,* and Asian house shrew *Suncus murinus.* Information from Animal Diversity Web (2012) and references therein, and John Long (2003).

realise that taxonomy in these species is unresolved and largely based on geography or pelage colour. Considering their ubiquity and socio-economic importance remarkably little research has been done on these species.

Each species has its benefits relative to the others in terms of their use in this study: black rats are a relatively large species that were translocated westwards from their Asian native range relatively early in prehistory and are therefore relatively common in archaeological contexts; house mice are the most widespread, and are a comparatively well-studied system; and shrews are entrenched in some environments where major pre-historical lacunae still exist, such as Madagascar. Different species offer different perspectives and using multiple bioproxies gleans a more complete model of human activities.

SAMPLE COLLECTION AND ANCIENT DNA

The rodent and shrew samples for genetic analysis collected in my study stem from three main sources. 1) modern tissue collected in the field by various field researchers and preserved in solution 2) osteocrust (remaining tissue on the bone left over from the cleaning process) and dry skin from specimens held in museum collections and 3) bone finds recovered from archaeological deposits. The latter two sources fall under the category of ancient DNA (aDNA).

Significant effort was made to source material for the commensal species studied here from throughout the Indian Ocean region, and in particular the African continent and near islands, where current data is lacking. I contacted over 200 field researchers, archaeologists, and museum curators to source material, as well as personally visiting various museum and other institutional collections. Collection information is detailed in the relevant analytical chapters. In some cases, after inspecting collections, there proved to be no available material for study either due to the lack of the species in the collection, or in the case of museum samples, the lack of non-invasive sample material. Some archaeological material from Fort Jesus Museum, Kenya, was sourced but I was unable to arrange export of the material within the timeframe of my visit. I was not able to retrieve material for analysis from the following institutions: British Museum (London); Fort Jesus Museum (Mombasa); National Museum of Kenya (Nairobi); British Institute in East Africa (Nairobi); Zanzibar Museums (Zanzibar); Yale University (New Haven); and Harvard University (Massachusetts). In some cases, contacts made collections available for study, however I was unable to visit the collections due to cost and time restrictions (e.g. in Africa). Volunteers in the institutions were not available to study the material, identify specimens, nor collect material.

Ancient DNA is the recovery of genetic information from biological samples originating from fossilised remains, archaeological finds, museum specimens, or other non-contemporary resources. It was originally made practicable by the development of enzymatic amplification methods in molecular genetics (i.e. PCR sequence amplification and subsequently whole genome amplification, etc.). Even with the advanced sequencing technologies and methods available today, aDNA is still limited by factors that affect the retrieval and accuracy of genetic data linked to the degradation of the DNA molecule *post mortem.* Enzymatic degradation of the DNA molecule from endogenous nucleases soon after death is followed by infestation by bacteria, fungi and invertebrates (Pääbo *et al.* 2004). Hydrolytic cleavage and oxidative base damage of the molecule then occurs causing nicking or base modification to DNA (Pääbo *et al.* 2004). The resulting DNA profile then characteristically contains low copy numbers, short fragments (>500 bp), and miscoding lesions, which is compiled idiosyncratically on a sample-by-sample basis (Pääbo *et al.* 2004).

Methodologically, degraded DNA is difficult to amplify due to the inherent properties of low copy numbers and short fragments. This leads to the preferential amplification of high copy number and high molecular weight DNA introduced through exogenous contamination. Contamination can be introduced *in situ*, in the field, at the identification stage, in the environment, in the laboratory, in reagents and almost any other scenario. Contamination of the sample itself is typically more difficult to detect than contamination introduced in the laboratory (Willerslev & Cooper 2005). Due to these issues, minimising contamination and validating the authenticity of genetic information is essential for aDNA studies and a number of criteria have been proposed in the laboratory. These include isolation of aDNA labs from modern ones to prevent carry-over contamination; UV radiation and bleaching to destroy contaminant DNA; extraction and amplification controls, independent amplifications and extractions, cloning, and independent reproduction to validate sequences and diagnose contamination; and critically assessing the appropriate molecular behaviour (an inverse correlation between amplicon length and amplification efficiency) (Willerslev & Cooper 2005).

In some cases where mitochondria are the target sequence, preferential amplification of nuclear insertions of the mitochondrial genome (numts) occurs and these can be mistaken for the mitochondrial sequence (Pääbo *et al.* 2004). To account for this aDNA studies are often developed using overlapping primer sequences as it is unlikely that the numt will be generated in both, and cloning is also encouraged (Willerslev & Cooper 2005). A number of data inspection protocols can help detect numts such as the inspection of chromatograms for dual peaks (the co-amplification of numt and template), inspection of translation sequences for stop codons, and identifying deviation from expected synonymous vs non-synonymous and transition/transversion ratios (Triant & Dewoody 2007).

Miscoding lesions in the DNA template caused by hydrolytic processes are sometimes recovered in genetic sequence data. For aDNA the most frequently identified process is the hydrolytic loss of amino groups leading to deamination and G>A and T>C base changes (Willerslev & Cooper 2005). Inspection of chromatograms, cloning and identifying deviation from expected synonymous vs non-synonymous and transition/transversion ratios help identify deaminated sites.

CHAPTER OUTLINE

Chapter two summarises our current knowledge relating to the distribution and dispersal of house mice and black rats from the onset of commensalism in the Neolithic of the Near East to their early first millennium AD distributions in the Indian Ocean region. Geographically, it focuses on the Indian subcontinent, Southwest Asia, the Arabian Peninsula and North Africa. It critically reviews the archaeological and genetic literature and proposes hypotheses for the natural dispersal and/or human-abetted translocation of rats and mice in the past. Chapters three to six present the results and interpretations of genetic analyses undertaken for the species examined in this study. A brief introduction, the materials and methods, results and a discussion are described as a unit. In this way each species is clearly defined in terms of the hypothesis they test and the methodology of the research. Although molecular methodologies and data analysis were broadly similar across species, their application to each taxon differed and the methods were most logically presented within the chapter for each species. Chapter three examines the genetic diversity and phylogeography of the Asian house shrew *Suncus murinus* in the Indian Ocean region and extending to Asia Pacific, inferred from the cytochrome *b* gene. It is based on a collection of modern and museum specimens and presents some of the methodological considerations that affect phylogeographic studies, especially when incorporating museum specimens in the analysis. It describes the geographic clustering of genetic lineages in terms of the human networks that may be responsible for introducing them. Chapter four explores the phylogeography of the

Rattus rattus species complex (the black rat and a collection of closely-related taxa) in the Indian Ocean focusing on Africa and Arabia. It reconstructs genealogies inferred from cytochrome *b* and D-loop data derived from modern and museum samples. Chapter five incorporates a temporal aspect to the phylogeography of the black rat through ancient DNA studies. It identifies continuity of genetic lineages and novel diversity over time, inferred from the D-loop gene. It also attempts to address the problematic zooarchaeological identification of black rats by developing a genetic identification assay based on the cytochrome *b* gene that is applicable to ancient DNA. Chapter six examines the phylogeography of the house mouse *Mus musculus* with a focus on the Western Indian Ocean. It builds on the substantial D-loop framework for the species in other regions of the world and draws the Western Indian Ocean into global and regional networks in terms of house mouse translocations and dispersals. The evolutionary history of the house mouse is considered in light of the description of novel genetic diversity. Chapter seven summarises the findings of the commensal model approach, and briefly assesses the utility of each model for inferring human history in the Indian Ocean region. It highlights the need for incorporating both genetic and zooarchaeological data in the interpretation of ancient genetic diversity and geographic distribution. It describes some of the broader impacts the research has and suggests avenues for future research.

CHAPTER 2

DISPERSAL AND TRANSLOCATION OF HOUSE MICE (*MUS MUSCULUS*) AND BLACK RATS (*RATTUS RATTUS SENSU LATO*) IN THE INDIAN OCEAN: ZOOARCHAEOLOGICAL AND GENETIC PERSPECTIVES

Trade networks in the Indian Ocean have functioned not only as channels for commerce but also as conduits for the transmission of technological, cultural and religious ideas, and biological exchanges. But these interactions also unwittingly introduced invasive pests and pathogens that quickly became important socio-economical forces. In the case of many commensal species, such as mice and rats, the unwelcome guests became pests to home and agriculture, transmitters of disease, and affected local species through competition and predation.

Commensal species have close ecological relationships with humans and are a “significant component of human dispersal and exchange networks” (Larson *et al.* 2007) to the extent that their introduction to new areas, particularly over long distances, is almost obligately facilitated by humans through sea and land transport. This relationship means that the histories of the commensal population and the humans who transported them are shared and both the gene flow and distribution of the animal population is shaped by human movement (Jones *et al.* 2013).

This review examines biological translocations of two socio-economically important small mammals that are widespread in the Indian Ocean; the black rat *Rattus rattus* and the house mouse *Mus musculus*. It considers the current zooarchaeological and genetic evidence for routes of invasion and colonization histories of the two species that lead to their establishment as invasive populations in the region and links them to patterns of human culture, trade and contact. The early translocation of rats and mice is linked to exchanges between the earliest civilizations of the Indus Valley, Mesopotamia and Egypt and this review focuses on the regions connecting these cultures. Geographically, it includes countries bordering the Arabian Sea, Persian Gulf and Red Sea beginning with the native range of the two species in South Asia (India and Pakistan) then moving west along the coast to Southwest Asia (Iran and Iraq), the Arabian Peninsula and Northeastern Africa (predominantly Egypt) (Figure 2.1).

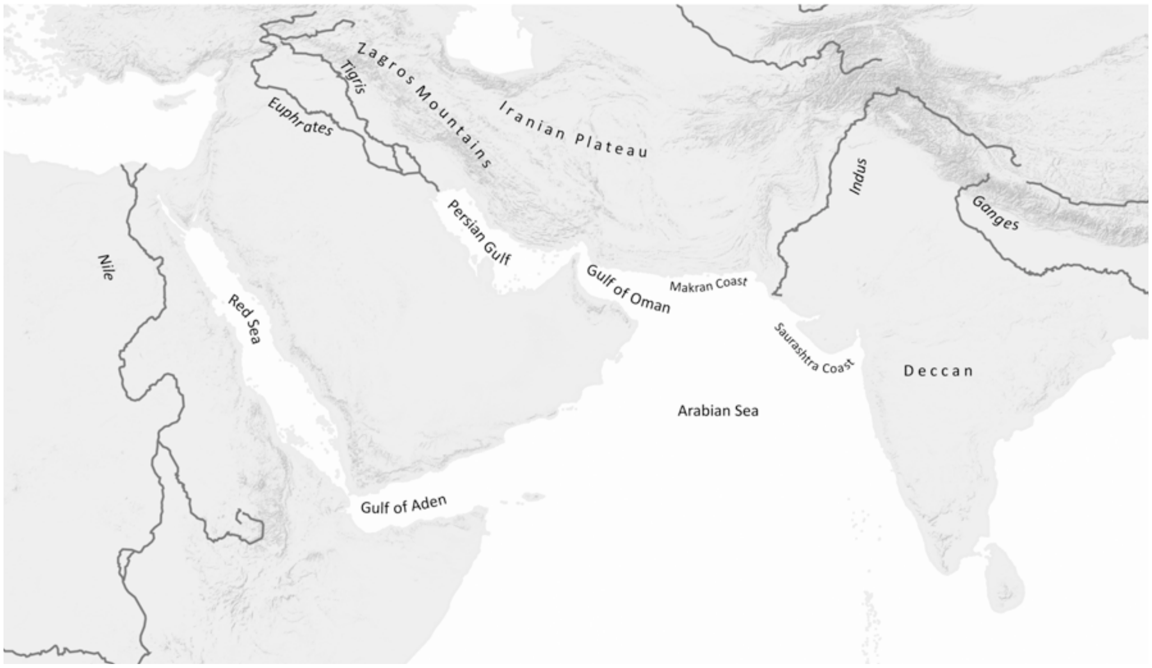


Figure 2.1 Map of the region with features discussed in the text.

BIASES IN SMALL MAMMAL ARCHAEOLOGY AND TAXONOMY

There are a number of biases that mar the interpretation of the archaeological and distributional data reviewed here and it is worthwhile introducing these before starting.

The relatively small and fragile nature of rodent and other small mammal bones leaves them prone to decay by biotic and abiotic processes and consequently they may not preserved. In the excavation process small bones may be missed by the naked eye and when sieving using larger mesh fractions (>1-2mm) (Weissbrod *et al.* 2005). Due to poor preservation and recovery, the majority of small mammal bones recovered from excavations are the larger limb bones and mandibles and they are often in fragmented or poor condition. Often these bones do not have morphological characteristics that distinguish them from other closely related species, especially if only bones of the post-cranial skeleton are present (Cucchi *et al.* 2005b). For example, the sympatric and morphologically similar species *M. musculus* and *M. spretus* can be differentiated based on skull and dental characteristics but these are often missing in archaeological contexts (Valenzuela-Lamas *et al.* 2011). Further, rodents and other small mammals are often seen as of secondary importance to economic species such as domesticates, and with limited time and/or money to analyse them, they are often ignored (Toškan & Kryštufek 2006). The burrowing ability of many commensals means they are easily labelled as 'intrusive' and while this may be true for some species (or populations) it is not true for others.

In particular regards to the study of commensalism is the issue that individuals found in deposits may represent wild-living, and not commensal, populations. There are scenarios whereby wild-living populations of typically commensal species may be introduced to archaeological deposits, for example, through raptor pellets that are frequently found in cave and rockshelters. In these cases, taphonomic information in regards to the bone is of great importance as fractures and corrosion indicate degradation due to the passing of the bone through the digestive tract (Weissbrod *et al.* 2005).

One only need look at the number of synonyms listed (Musser and Carlton 2005) and the degree of taxonomic revision for the species reviewed here to recognise that assigning phylogenetic relationships based on traits of morphology, geography and ecology has proven difficult. In part, this is due to the phenotypic variability shown by these commensals e.g. coat colour, tail length or body size, which makes identification difficult. In my own experience skin specimens in museums sometimes mistake *R. rattus* for *R. tanezumi* (only recently classified a species distinct from *R. rattus*) or other *Rattus* species. *Mus macedonicus, Mus spretus* or *Mus spicilegus* are sometimes labelled *M. musculus* (differing mainly in the length of their tails), as are juvenile rats and even shrews. Other authors have similarly noted the misidentifications in voucher specimens and the literature (Musser & Carleton 2005; Pagès *et al.* 2010).

ORIGINS, DISTRIBUTION AND MOLECULAR BIOLOGY

The black rat is also eponymously known as the 'ship rat', 'roof rat' or 'house rat' for its propensity to stowaway on ships and nest in roofs and houses. The *Rattus* genus has its origins in Southeast Asia and *R. rattus* is native to the Indian Peninsula (Musser & Carleton 2005). It is now a cosmopolitan species having spread to all continents (except Antarctica). In the 19th century it was the most widely transported rat, being carried with European traders and colonisers throughout much of the globe (Atkinson 1985). Ervynck provides an excellent review of pre-Holocene *Rattus rattus* fossils and determines that, due to misidentifications and lack of data, a natural distribution outside Asia cannot be determined with certainty (Ervynck 2002). In many areas the black rat has been displaced by the invading Norway rat *Rattus norvegicus* (Lund 1994) and thus its current distribution is likely to be a fraction of its former range.

Phylogenetic classifications of closely related *Rattus* species are not fully resolved (see Aplin *et al.* (2011) for a discussion). That aside, two distinct chromosomal forms of *Rattus rattus sensu lato* were identified by Yosida et al; the 2n=38 "Oceanian" form and the 2n=42 “Asian” form, which have since been classified *Rattus rattus* (the black rat) and *Rattus tanezumi* (the Asian house rat), respectively (Yosida *et al.* 1974). The Oceanian karyotypic form is suggested to have developed from Robertsonian fusion of four acrocentric pairs in the Asian form (Yosida *et al.* 1971). It is the more recent Oceanian/*rattus* form that is widespread having dispersed from the Indian subcontinent west through Asia to the Mediterranean by at least the beginning of the Holocene and some time later accompanying humans on migrations and with trade (Ervynck 2002). The Asian/*tanezumi* form inhabits much of continental and peninsular Southeast Asia (Aplin *et al.* 2011).

House mice are perhaps the best-known commensal species, in part due to their seemingly ubiquitous presence as pests to home and agriculture, but also because they are familiar as pets and as model systems in biomedical science. Their use in science has led them to be one of the best-characterized commensals in regards to their genetics.

House mice are native to Eurasia with the presumed cradle of the species in the northern Indian subcontinent, where genetic diversity is highest (Boursot *et al.* 1996; Prager *et al.* 1998). Their remarkable adaptability and close association with humans has enabled them to disperse so that they are now a cosmopolitan species occurring from sub-Antarctic islands to tropical forests and modern cities. Within the *Mus musculus* complex there are three major subspecies (regarded by some as species): *domesticus* is found in the Near East, North Africa, western Europe and is the form widely introduced to the Americas, Australia, parts of Africa and elsewhere in historical times; *musculus* is found in North Asia and eastern Europe; and *castaneus* is found in central and Southeast Asia (Musser & Carleton 2005). Two other lineages are proposed but not widely accepted: *gentilulus* from Yemen and possibly other parts of the Arabian Peninsula, which is a relatively new find (Prager *et al.* 1998); and *bactrianus* from Afghanistan (Din *et al.* 1996); this form is considered to be *castaneus*-like by many authors and is not well-characterized genetically.

The origins of commensalism

Commensalism apparently initiated in response to the sedentarisation of human cultures in the late Pleistocene/early Holocene of the Near East (Tchernov 1984). Whether sedentism *per se* or factors associated with it such as food storage and rubbish disposal are the contributing factors is debatable (Wyncoll & Tangri 1989). The origin of commensalism in house mice is proposed to have occurred with the *domesticus* lineage in the Natufian period of the Levant 12,000 years ago (Auffray *et al.* 1988). The Natufian find from Hayonim Cave, Israel is associated with semi-sedentary populations who harvested and stored grain to some degree. The first black rats found in association with humans come from Kebara Cave in Israel dating to the late Pleistocene (Tchernov 1984) (but see Ervynck (2002) for a discussion of the contextual surety of these finds) and by the early Holocene it is a confirmed commensal found in a number of Natufian period Levantine cave and rockshelter sites in Israel, including Hayonim Cave (Ervynck 2002). Although other commensal subfossils are found in the Levant I will not discuss them further here.

ZOOARCHAEOLOGICAL AND GENETIC EVIDENCE FOR COMMENSAL TRANSLOCATIONS

South Asia

Given that the Indian subcontinent is the native home of the black rat and house mouse it would not be surprising to find early subfossils in archaeological Pleistocene deposits in the region to rival those found in the Near East (Figure 2.2). A subfossil identified only as '*Rattus*' species was found in association with worked bone in the late Pleistocene at Devakachhar in the Narmada Valley (Madhya Pradesh, India) (Misra 2003). This could be the earliest evidence of black rat in its early commensal stages in India unless, as is possible, the finds are of wild-living black rats or other *Rattus* species (e.g. ancestral ranges of *R. tanezumi* or *R. nitidus* that no longer inhabit the area). The earliest black rat (identified to species) and house mouse in archaeological contexts do not come from the continent but from Sri Lanka. The paucity of early finds on the mainland may be due to biases in South Asian zooarchaeology, which favours analysis of subsistence and economic species. I will not go in to detail regarding commensals in Sri Lanka but will focus on the prehistoric material. The rat, mouse and a number of other rodents were found in the Batadomba-lena rockshelter in the southwest of the island (Perera *et al.* 2011). The rockshelter is a prehistoric site dated to the Pleistocene *ca.* 10, 000 -34, 000 BC, where microliths and other evidence of human habitation occur. The rodents were grouped together and the contexts not distinguished for each species so we cannot extract species-specific information relevant to the horizons they appear in. However, as only one rodent individual was identified as having come from the disturbed Holocene layers we can presume that either the black rat or the house mouse, if not both, were found in Pleistocene contexts. For the house mouse this is an unusually rare occurrence. Batadomba-lena aside, there is a complete absence of subfossil evidence for house mice in South Asia.

If the rat and mouse are dated to Pleistocene levels they would predate those found in Hayonim Cave in the Neolithic Levant where commensalism is widely suggested to have originated. While there is not sufficient evidence to indisputably claim a commensal relationship was established at Batadomba-lena - the shelter was not permanent but the occupants were complex foragers and, in a suitably rich environment could conceivably be seasonally sedentary - it certainly merits consideration as commensalism *in process*. The house mouse, black rat and a number of other species found in the rockshelter such as the field mouse *Mus booduga*, lesserbandicoot rat *Bandicota bengalensis* and greater bandicoot rat *Bandicota indica* are all modern-day pests of agriculture found in a site that was mainly occupied by foragers. Nuts *Canarium* sp., bananas *Musa* sp. and jackfruit *Artocarpus* sp. are common at the site and were a significant component of the diet of the Late Pleistocene foragers but nuts, seeds and fruits are also the preferred diet of many pest species such as the black rat which may have benefited from the plant refuse (Lund 1994; Perera *et al.* 2011). The above animal species are perhaps the ancestors of modern day commensals and syanthropes that are best-suited to the anthropogenic niche and eventually outcompete other wild-living species.

A local chromosomal form of *Rattus* is found in the highlands of Sri Lanka (a 2n=40 form) and labelled *Rattus kandianus* (it is considered by some as a subspecies), whereas the coastal lowlands are populated by the common Oceanian/*rattus* variety (although hybrids occur in both areas) (Yosida *et al.* 1974). It is believed that *kandianus* is a relict population and should represent the older of the species on the island and *rattus*, due to its similarity with other Oceanian forms, is a recent immigrant brought by humans (Yosida *et al.* 1974). We do not know when the black rat was first introduced to Sri Lanka and if securely identified, the Batadomba-lena fossils could represent very early introductions for populations that subsequently went extinct, or they could represent more ancient forms of *Rattus* such as *kandianus.*

The majority of small mammals recovered from archaeological sites in India and Pakistan are identified as *Rattus rattus* or *Rattus* sp. Whether or not these identification are accurate is difficult to determine as the criteria by which they were identified is rarely reported in the literature. Synanthropic or commensal rodents of a similar size as the black rat, or larger, such as the Indian gerbil *Tatera indica* or bandicoot rats *Bandicota* sp. are occasionally recovered, however smaller rodents such as mice *Mus* sp. are not. The most common house mouse subspecies found in India, *castaneus*, is perhaps the most obligately commensal of the major lineages being found almost exclusively in buildings which suggests a long history of commensalism with humans and therefore we should expect to see them in the archaeological record (Musser & Carleton 2005). The lack of evidence is likely a mix of biases owing to poor recovery and interpretation as well as due to the ecology of *Mus musculus.* Recovery of fauna and in particular small bones in Indian excavations is both uneven and unsystematic and identification of species may be compromised by incomplete reference collections (Joglekar 2006). House mice do not easily establish colonies in the presence of competitors and rats and gerbils are perhaps better able to exploit early human settlements. A similar scenario whereby the house mouse is outcompeted, in this case by a the native species (the wood mouse *Apodemus* spp.) is suggested for the house mouse's relatively late colonisation of Western Europe (Cucchi *et al.* 2005a).

By the Mesolithic on the subcontinent there is sufficient evidence to confirm at least an emergent commensalism between the black rat and hunter-gatherers in India. Perhaps as early as the sixth millennium BC black rats are found at Damdama in the Ganga Valley of North India (Thomas *et al.* 1995b). Damdama is a sedentary mesolithic site and key for tracing the onset of sedentism in India (Thomas *et al.* 1995b). The nearby Mesolithic site of Mahadaha also contains remains of black rats but, unlike Damdama, was thought to be occupied only seasonally by transient hunter-gatherers (Alur 1980).

One of the best-described black rats in an early context comes from Kurnool Caves in South India. Patnaik *et al.* (2008), using characteristics of dental morphology, identify "cf. *Rattus rattus*" from deposits dated around 3200 BC. The Kurnool Caves, or Muchchatla Chintamanu Gavi, saw long-term human occupation from the Palaeolithic to more recent times and have produced numerous lithic and bone artefacts. For the study of fauna it is an equally important site as it has produced a comparatively large and well-preserved faunal assemblage compared to other sites of similar age. Among the at least eleven other rodent species found in the cave are the greater bandicoot rat, lesser bandicoot rat*,* Indian bush rat *Golunda ellioti,* Miss Ryley's soft-furred rat *Millardia kathleenae* and the soft-furred field rat *Millardia meltada*. These species all inhabit disturbed and human-modified habitats today and are in many ways similar to the black rat ecologically but they appear in Pleistocene contexts before the appearance of *Rattus*

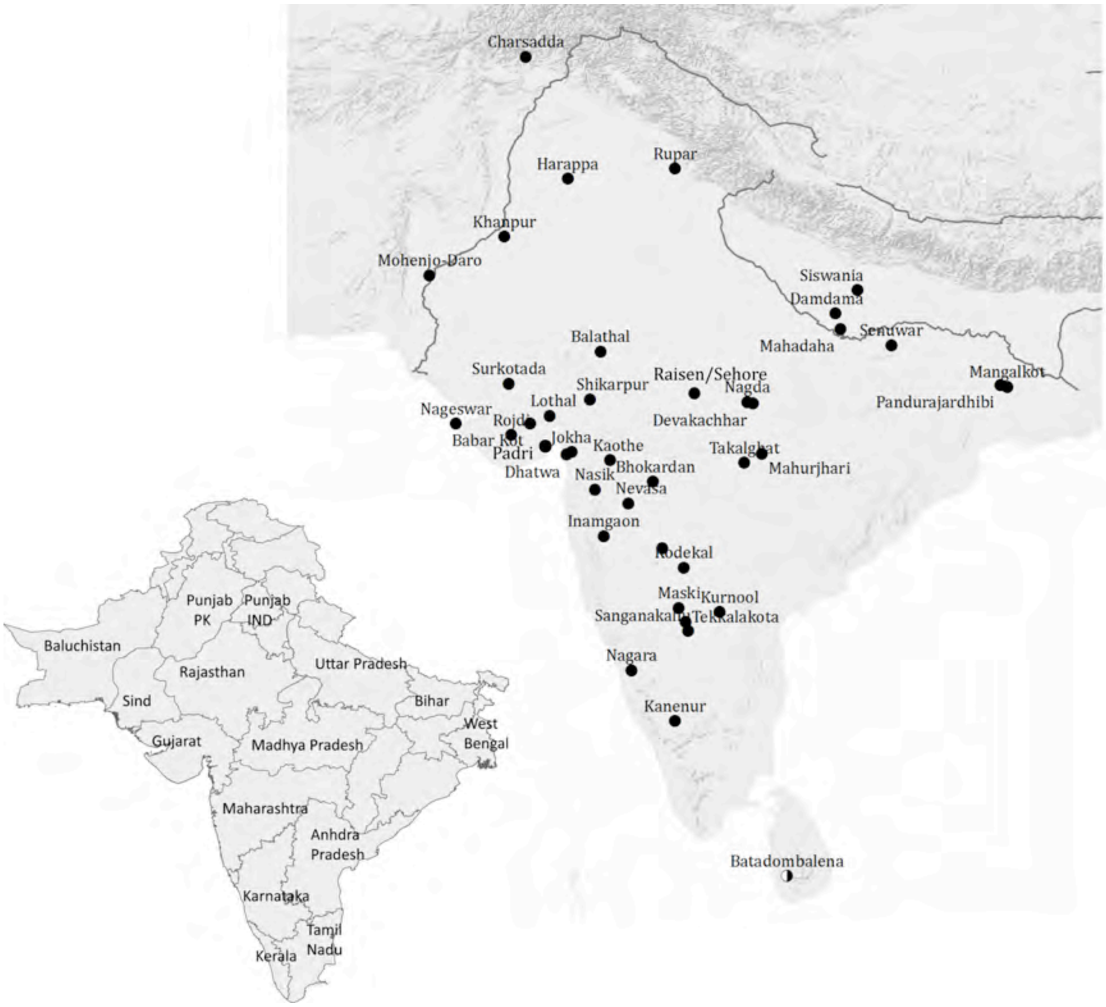


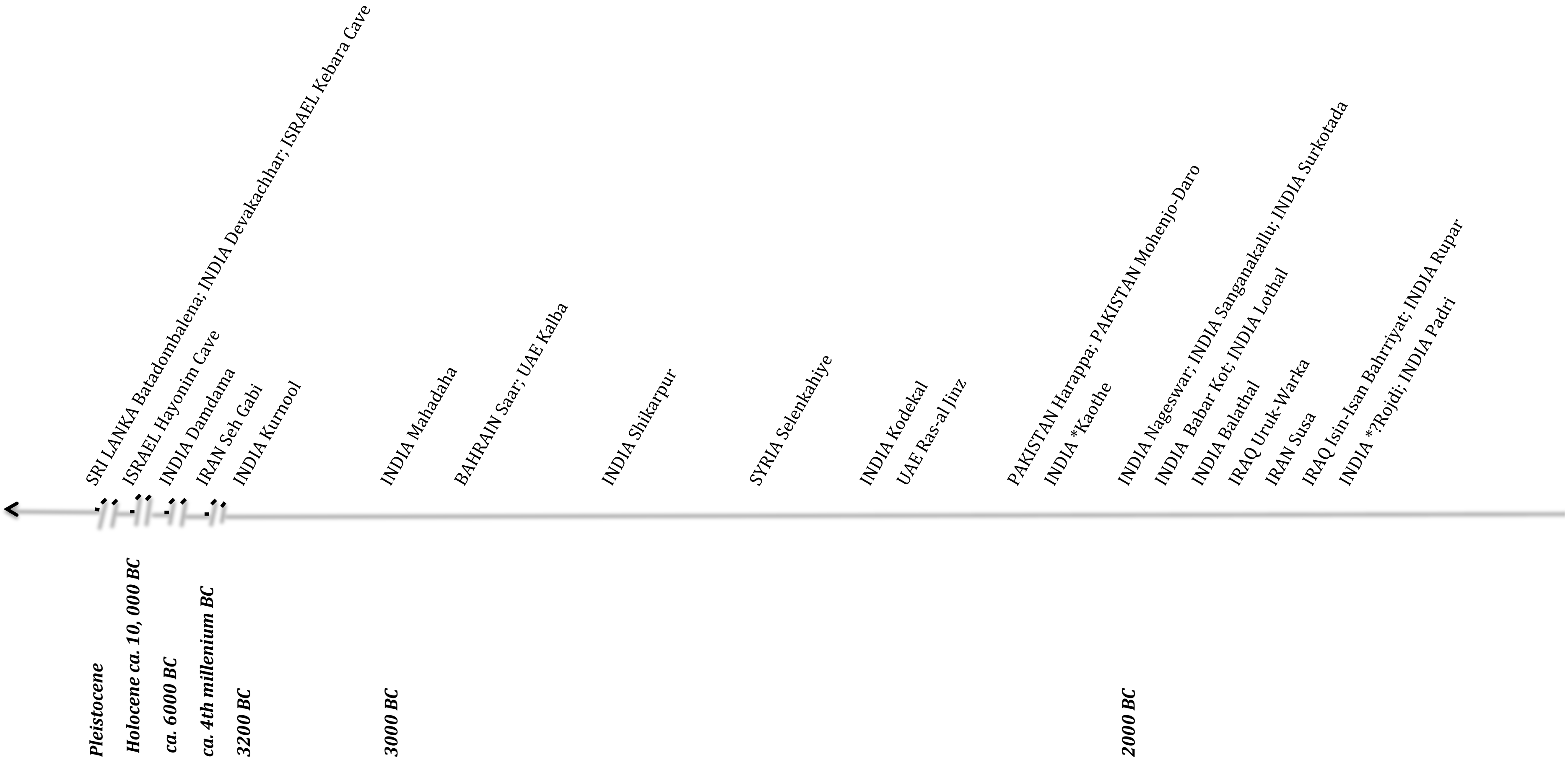
Figure 2.2 Map of archaeological sites in South Asia and provinces of India and Pakistan (inset) discussed in the text. Black circles are sites where rats are recorded and black/white circle on Sri Lanka where both rats and mice are recorded.

and continue through the Holocene. *Rattus* appears to be a late addition to the rodent fauna of the area.

As the black rat is archaeologically present in northwestern India by the Mesolithic we might expect that the organised communities of the Early Harappans of the Indus Valley, where there is also evidence of widespread food production, would have the species as commensals. While the black rat is indeed found in several Harappan culture sites in India and Pakistan, its presence is not secure until the Mature Harappan phase (*ca.*

2600-1900 BC), except for a single bone in Shikarpur in an Early Harappan context (Thomas *et al.* 1995a). By the Mature Harappan phase the black rat is an established commensal of humans. In Mature Harappan India and Pakistan the black rat was found in Surkotada, Lothal, Babar Kot, Nageswar, Shikarpur, Padri, Mohenjo-Daro, Harappa and possibly Rojdi (unidentified to species and possibly a later intrusion) (Prashad 1939; Rao 1979; Possehl & Raval 1989; Sharma 1990; Hedge 1992; Mahapatra 1995; Thomas *et al.* 1995a; Ervynck 2002). Harappa, Mohenjo-Daro and Surkodata are major urban centres with evidence for food cultivation and storage (e.g. granaries). In Surkotada the black rat accounts for a relatively high proportion of the fauna constituting 6% of fauna in the period *ca*. 2055-1970 BC, 6% in the period 1940-1790 BC and 3.24% in the period 1790-1660 BC (Joshi 1990). The connection between humans and black rats in Harappa is indisputable, if the identification of the bones is valid, and it is perhaps the earliest evidence of the rat as a domestic pest in the region - eight bones of the rat are found inside a jar and the authors speculate that the rat got stuck inside the jar and died(Prashad 1939).The Mature Harappan is a period of intensified urbanization that led to the large-scale cities of the Indus Valley Civilization. It is most likely the growth of urban centres and the various resources implemented to sustain them (e.g. dwellings, foodstores and granaries) that enabled commensal black rats to thrive. The period is also characterised by a marked increase in trade and the integration of towns and cities and this would have facilitated the spread of rats in the period.

At around the same time the black rat was occupying coastal and inland towns and cities of the Indus Valley Civilization it was simultaneously inhabiting (albeit to a lesser degree) the less-urbanised, early farming communities of the Southern Neolithic. The culture flourished around 2500-1000 BC and the black rat is found in its earliest levels at Kodekal Ashmound, Karnataka around 2335 BC (± 105 years, radiocarbon dated) (Shah 1973). The Neolithic culture of South India extended across Karnataka, Andhra Pradesh and Tamil Nadu, but rat fossils from this period are found only in the inland districts of



Timeline of *Rattus rattus* subfossils.

? Unidentifed to species; *possibly intrusive or not securely dated; date is averaged where there is a range

Figure 2.3 Continued next page

Figure 2.3 Continued.

Bellary and Gulbarga in Karnataka, i.e. in Kodekal, Sanganakallu (a Neolithic house but contexts not given for bone finds) (Alur 1969) and possibly in nearby Tekkalakota (identified by Reddy in his review but the original report only identifies "rat bones") (Rami Reddy 1985)′(Nagarajarao & Malhotra 1965). Sanganakallu is a site with a stone tool manufacturing industry and, as with Tekkalakota, it is a village complex with evidence of early agriculture (e.g. cattle, sheep, pulses and millets). In Maski (Karnataka), the black rat was found in the Neolithic/Chalcolithic boundary around the early first millennium BC, in association with microliths, small amounts of copper and beads of carnelian, agate, coral, shell and glass(ASI 1957). The inhabitants of Maski in the Chalcolithic were largely pastoralists but by the late first millennium BC they had developed an agricultural economy and an urbanised core (ASI 1957) thereby providing suitable conditions for occupation by commensals. It is surprising then that in this later period (*ca.* 200 BC) rats are absent. Once rats are established it is rare for localised extinctions to occur and there is no evidence of environmental change or interspecific competition that accounts for the decline of rats at the site. Such absences are a feature of rodent zooarchaeology. This may be a sampling artefact (bones are rare and therefore may be missed), or due to subtle shifts in habitat use by some rodents that sees them moving from peridomestic to feral lifestyles.

By the end of the Harappan period there is an established commensal relationship and this denotes a reliance on humans by black rats to some degree. With the de-urbanization of the Indus Valley around the beginning of the second millennium BC (*ca.* 1900-*ca.* 1700 BC), which signalled the decline of the Harappan civilisation, we observe a decrease in rat fossils in the Indus Valley and an increase in rat fossils to the south and east. This may have been linked with what has suggested to be a Harappan diaspora, with political instability and environmental change encouraging eastward population shifts to Punjab, Rajasthan and surrounding areas possibly as far as Bihar, or south to southern Gujarat and the Deccan. The migration out of the Indus Valley is variously attributed to climate change, a shift in economy from urban to agricultural, a concomitant decline in trade, and in some areas the shift from farming to herding (Sonawane 2000; Agrawal 2007). *R. rattus* is found in the Chalcolithic layers of the post-urban Indus site of Balathal (Rajasthan) (Thomas 2008). Rats are also found in Khanpur, Punjab (*ca.* 2000-1000 BC) although this is only identified to genus and is possibly intrusive but a more secure find of *Rattus rattus* from Rupar, Punjab (*ca.* 2000-1400 BC) exists (Dutta 1984; Thomas 1984). In southern Gujarat, black rats are found in Jokha (*ca.* 1500-1000 BC) where a shift from farming to herding occurred in the late Harappan (Mehta & Chowdry 1971; Sonawane 2000). In Senuwar, Bihar in the northeast of India the black rat is found in deposits dated between 1700-1400 BC and even further east it is found in the Chalcolithic of West Bengal in Pandurajardhibi and Mangalkot (Banerjee 1981; Sathe & Badam 1996)

The commensal status of rats is questionable for early farming cultures of South India such as the Savaldas of Kaothe (*ca.* 4200 BP), and Malwa/Jorwe cultures of Inamgaon (site dated from *ca.* 1600 BC and rats found in the earliest levels), Tuljapur Garhi (site dated *ca.* 14th century BC and rats found in the earliest levels) and Nevasa (only identified to genus) (Thomas 1984, 1988, 1996; Thomas & Joglekar 1990). The zooarchaeological evidence does not clarify the situation as all finds from these sites (identified by the same faunal analyst) are suspected to be non-contemporaneous intrusives and at least in Inamgaon, burrows were found in the excavation. Black rats are scansorial, that is they are adapted for climbing and prefer to nest in trees and the upper portions of dwellings (e.g. roofs) and it is likely that the remains, if intrusive, were not black rats.

The following finds complete the zooarchaeological dataset for black rats in South Asia from the Iron Age to Medieval periods: Nagda (Madhya Pradesh) in levels dated 500-200 BC; Takalghat (Maharashtra) in levels dated 555 BC; Siswania (Uttar Pradesh), Jokha

(Gujarat) and Kanenur (Karnataka) (possibly a later intrusion but not explained why) in the Early Historic; Nagara (Gujarat) in the Early Historic and Medieval; Dhatwa (Gujarat) and Nasik (Maharashtra) (earlier excavations at the site assumed these were intrusive) in the Iron Age/Early Historic; and Mahurjhari (Maharashtra), Bhokardan (Maharashtra) and Charsadda (Pakistan) in the early Medieval(Sankalia & Deo 1955; Mehta 1968; Deo 1970; Mehta *et al.* 1971, 1975; Deo & Shastri 1974; Banerjee 1986; *Indian Archaeology a review 1996-97.* 2002; Thomas *et al.* 2006; Young 2007; Deshpande-Mukherjee *et al.* 2010). Finally, black rats were found in rockshelters in the Raisen and Sehore districts of Madhya Pradesh in unidentified contexts (the site stretches from Late Stone Age to Medieval) (Alur 1990). There is no taphonomic information as to whether these may be introduced by raptors, or are a natural introduction.

The Makran Coast is a historically important navigational pathway incorporating both overland routes through the narrow coastal plain and maritime routes via ancient ports. There is a lack of zooarchaeological evidence for exchanges via the Makran Coast, however modern-day house mice and black rat populations from the coast share a curious molecular signal: house mice have profiles typical of the *castaneus* lineage from Thailand; and a distinct isolate of a *tanezumi*/*rattus* hybrid population occurs which suggests a translocation of the 42 chromosome form of the Asian rat in to a pre-existing 38 chromosome population of black rats (Yosida & Sagai 1975; Prager *et al.* 1998). These are almost certainly sea-borne introductions due to human-based trade or migration, although a relict Asian rodent population cannot currently be ruled out. The route would have relatively recently linked Pakistan with Southeast Asia, possibly via an Indian hub as Pakistani mice are closely related to individuals from Southwest India (Prager *et al.* 1998).

Southwest Asia

Both genetic and zooarchaeological evidence suggests house mice occupied Iran since the Middle to Upper Palaeolithic. The geographic distribution of the genetic lineages of

Mus musculus subspecies suggest that house mice were widely dispersed in ancient times having spread from northern India to western Asia prior to sedentary occupation by humans. Two Palaeolithic cave sites in Iran, Qaleh Bozi at the foothills of the Zagros Mountains (dated *ca.* 30,000 BC) and Kani Mikaeil in the central Zagros (dated *ca.* 20,000 BC – mid-Holocene) contain house mouse subfossils (Hashemi *et al.* 2006). However, these remains are not directly dated and owing to the difficulty ascribing burrowing animals such as house mice to discrete depositional environments they are not securely dated. Furthermore, they probably represent wild-living forms as the authors contend that fractures and corrosion found on the rodent bones is evidence of introduction by raptors (Hashemi *et al.* 2006).

Mus musculus lineages began to diverge approximately 250,000 years ago and it is clear that independent lineages had colonized much of Eurasia in the Pleistocene as wild forms (Bonhomme & Searle 2012). As lineages had genetically and geographically differentiated prior to the first advent of commensalism in the Levant we can deduce that *domesticus*, *castaneus* and *musculus* all developed commensalism independently in connection with their human counterparts (Din *et al.* 1996). Geometric morphometric studies of subfossil house mice molars (including the Palaeolithic material mentioned above) identified that the ancestors of the Iranian house mouse were similar to the *castaneus* lineage (Shabani *et al.* 2010). Conversely, early Levantine commensals were thought be of the *domesticus* lineage (Tchernov 1991) and it is therefore likely that mice developed a commensalism in association with the human environment in the Neolithic of Iran, much like in the Levant. Although some modern-day house mice populations in the northwest of Iran are *domesticus* they share only one common haplotype with Israeli (Levantine) populations indicating that these proximal populations independently colonized the region (Bonhomme *et al.* 2011).

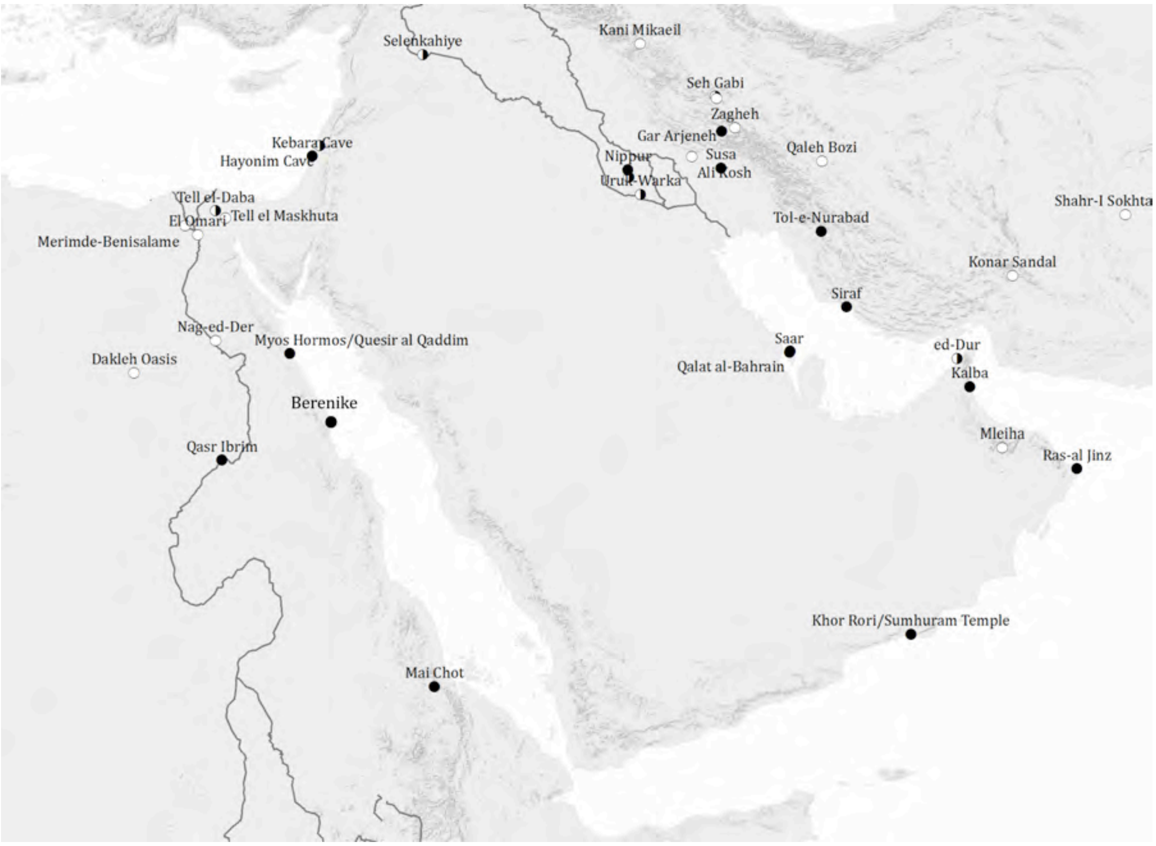


Figure 2.4 Map of archaeological sites in Southwest Asia, the Arabian Peninsula and Northeast Africa discussed in the text. Black circles are sites where rats are recorded, white circles are where mice are recorded and white/black are where both species are recorded.

The site of Ganj Dareh in the central Zagros (site dated 8400-7000 BC) is suggested to be one of the earliest sites for the initial domestication of goats *Capra hircus* some 10,000 years ago (Zeder 2000). The transition from hunting to herding could have facilitated the process of commensalism in mice and indeed mice are found at the site (Potts 1999) (Figure 2.4). Some 500-1000 years later, domesticated goats were introduced to the lowlands in Tepe Ali Kosh on the Deh Luran plain (site dated 7500-5600 BC) and mice are found once again (Zeder 2000). By the Late Neolithic (*ca*. seventh/sixth millennium BC) mice are found at Zagheh, a settlement on the Central Iranian Plateau connecting the Northern Zagros to the Eastern Iranian Plateau where there is evidence of trade with the east in the form of turquoise beads (Shabani *et al.* 2010). Mice are found slightly later in the Chalcolithic of Seh Gabi in the Central Zagros (Cuyler Young Jr. &

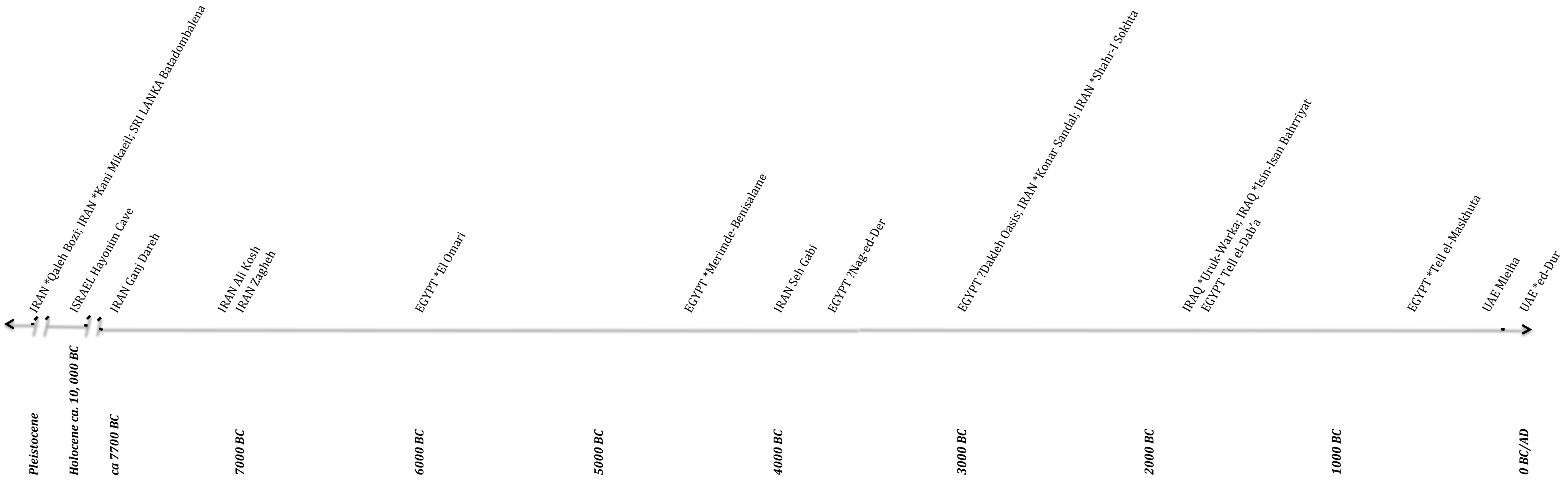
Levine 1974). The Zagros mountains would have acted as a formidable geographic barrier to colonisation in mice and the mountain chain is suggested to be preventing contact between modern day *domesticus* populations found in the Zagros and south central Asian *castaneus* populations found further east (Prager *et al.* 1998). It is therefore likely that, although wild populations would probably have existed in Iran prior to the beginnings of commensalism, mice were also being transported with early Neolithic dispersal and exchange networks.

It is not until the Bronze Age that house mice are found in eastern Iran at Konar Sandal and Shahr-I-Sokhta, although these finds are not securely dated (Caloi & Compagnoni 1977; Shabani *et al.* 2010). The Konar Sandal subfossils are of a type consistent with those found in the Neolithic of the Zagros (i.e. *castaneus*-like). Genetically, modern populations from these areas are also of the *castaneus*-like form that is common in North India and Afghanistan (Darvish *et al.* 2006). At present we are unable to identify a source for these populations, which would require genetic sampling and zooarchaeological identification of subspecies. In the early third millennium BC, Elam (Southwest Iran) was connected to eastern Iran and specifically Shahr-I-Sokhta via overland trade routes across the Iranian Plateau, which then continued to the Indus Valley and Afghanistan and it could be these early exchanges that has left the zooarchaeological and genetic signature we see.

Perhaps surprisingly mice are not found in the Tigris and Euphrates Basin until Babylonian times *ca.* 1800 BC in Uruk-Warka and Isin (Boessneck 1977; Boessneck *et al.* 1984). By this time Mesopotamia is connected to the Middle Asian Interaction Sphere, which I discuss later in reference to rats, and it is possible that mice were introduced through these networks.

The black rat appears relatively late in the archaeological record of Iran and Iraq given that it had extended its native range to western Asia by the end of the Pleistocene and was present in archaeological contexts in the Neolithic of the Levant. Although potentially wild-living black rats are found in the Pleistocene cave site of Gar Arjeneh (Central Zagros) lack of patination on the bones leads the authors to conclude that they are a recent deposit and so they will be disregarded here (Hashemi *et al.* 2006). Ervynck emphasizes the Levantine subfossils are restricted to cave sites or rockshelters, which he thinks could indicate a wild-living rather than a commensal population of Neolithic rats and the first true commensals might be Bronze Age finds from Mesopotamia (Ervynck 2002). However, black rats occur in the Chalcolithic of Seh Gabi (Central Zagros), a village site with evidence of herding, agriculture and of other commensals (i.e. house mice) (Cuyler Young Jr. & Levine 1974). Although the contexts of the finds are not given, the bones are thought to be contemporaneous with the site, which was occupied until the late fourth millennium BC. This find supports a pre-Bronze Age advent of commensalism in the region. The Zagros Mountains could have maintained isolated relict native populations from the Pleistocene expansion, or even immigrant commensal populations from the Levant, which may then have developed commensalism in response to increasing urbanisation and farming. However, a caveat is that no wild-living black rat sub-fossils have been found in the early Holocene of the Zagros, or even the corridor connecting South Asia with the Levant.

It is within a broadly Mesopotamian context that the next finds of black rats occur, firstly in the Bronze Age of Selenkahiye (Syria) *ca.* 2500 BC and then in the Old Babylonian period of Isin (*ca.* 2000-1500 BC), Uruk-Warka (*ca*. 1850 BC) and Susa (*ca.* 1800 BC) (Boessneck 1977; Boessneck *et al.* 1984; Uerpmann 2001; Ervynck 2002). For Ervynck "…it remains remarkable that the black rat is absent from early (proto)urban sites from outside the Tigris and Euphrates basin, and that the animal also remains absent within the Tigris and Euphrates before the second half of the third millennium BC" (Ervynck 2002). The Seh Gabi material allows us to reconsider the first clause but



Timeline of *Mus Musculus* subfossils

?unidentified to species; *possibly intrusive or not securely dated; dates are averaged where there is a range

Figure 2.5 Timeline of *Mus musculus* subfossils are they linked to Mesopotamian rats? Seh Gabi was abandoned around 3200 BC and rats are not present in other contemporaneous Central Zagros sites such as Godin Tepe whose occupation continued throughout the third and second millenniums BC. Seh Gabi and Godin Tepe lie on a highway connecting Mesopotamia with the Iranian plateau and there is evidence for links with Mesopotamia at Godin since *ca.* 3200-3000 BC (Schmandt-Besserat 1977). Rats could have maintained commensal populations in the Central Zagros after the Seh Gabi phase and then been transported to Mesopotamia through overland networks in the third/second millennium BC.

Alternatively, Mesopotamian rats could have been introduced from further afield. Armitage was the first to suggest that black rats spread from commensal populations in the Indian subcontinent westwards via ancient trade routes to reach Mesopotamia by the second millennium BC (Armitage 1994). In the early second millennium BC Dilmun (Bahrain) became the main entrepôt of trade in a Persian Gulf network that connected Mesopotamia with India. The zooarchaeological distribution of rats supports the Dilmun connection as they are found in Bahrain by the second millennium BC (discussed later) and, conspicuous by their absence, there are no rats found along the northern coast of the Persian Gulf during the same period (although this could be an artefact of recovery and identification). The development of sea-going vessels meant large quantities of goods could be traded directly over long distances where once pack animals or carts were used to transport smaller quantities of goods, and the size of the vessels meant that bulky consumables such as grain could be traded with relative ease. It is the quantity and type of the goods being transported and the volume and frequency of trade that would likely have facilitated the spread of rats in the period.

There is continuing evidence of occupation by black rats in the Tigris and Euphrates Basin in to the first millennium BC where they are found in Assyrian Nippur *ca.* 700-600 BC (Gibson *et al.* 1978). Finally, black rats are found in Tol-e-Nurabad in a period described as transitional between NeoElamite and Achaemenid, which places the finds at *ca.* 600 – 500 BC (Hashemi 2009). Tol-e-Nurabad is located on the Iranian Plateau and rats are rather a late arrival to the Plateau given they were present close by in the Tigris and Euphrates Basin a millennium earlier.

Although rats are found on the southern shores of the Persian Gulf since the second millennium BC, they do not appear on the northern shores until the Sasanian period (*ca.* fourth to the end of the sixth/seventh century AD) at Siraf (von den Driesch & Dockner 2002). The bones were identified as *Rattus* sp. Three species of *Rattus* are known from Iran: *R. norvegicus* the Norway rat, which is a second millennium AD introduction; *R. rattus* the black rat, which is an early introduction; and *R. pyctoris* the native Himalayan rat, which is not found in urban areas and occupies only mountainous regions of northeast and central Iran (Rad *et al.* 2009). Given the bones from Siraf were found in an urbanised context in the first millennium AD they are likely to be black rat. Modern genetics has thus far not recovered the ancient signal of rats along the coast as only a single rat identified from Bandar–e Hamiran on the Persian Gulf has been typed. It belongs to the 'ship rat' cluster, which also includes individuals from coastal west Africa, USA, New Zealand, Australia and Brazil and that was suspected to have been translocated during or after the Age of Exploration (Aplin *et al.* 2011).

A total of eight *Rattus* sp. bones were found at the Great Mosque in Siraf during the peak of its prosperity as an entrepôt of long distance trade (the Sasanian and early Islamic periods until the mid ninth century AD). As the importance and wealth of the port city declined (*ca.* 1050-1275/1325) the number of bones vastly increases to 48 by which time the rat is a prominent commensal. However, by *ca.* 13th/early 14th-16th century AD rat bones are absent from the mosque. It would be unusual for the black rat as an established commensal to go extinct in Siraf and it is more likely that the pattern is one of scale - in 1262-3 AD the mosque was rebuilt at a fraction of its former size and rats probably moved to other preferred habitats in Siraf. The absence of rats in the fourteenth century and beyond is fortuitous as in 1347 the bubonic plague *Yersinia pestis* swept through Europe and the Middle East causing catastrophic mortality and massive societal and demographic upheaval among the living as inhabitants fled ports and cities *en masse* and farmers left their fields untended. The disease, which came to be known as the Black Death, was spread by commensal black rats and sporadic outbreaks continued until the eighteenth century. The Black Death is not recorded in coastal Iran although it was present in the north of the country (Azizi & Azizi 2010).

Arabian Peninsula

Commensal rodents and even native synanthropic rodents are practically unknown in the Arabian Gulf before the Bronze Age (see Beech (2000) for a review of thirteen sites dated to the fourth and fifth millennia BC, only one of which lists unidentified rodents in the fauna). By the late third millennium BC black rats are found at the site of RJ2 in R'as Al-Jinz, which marks the beginning of occupation by commensal rats and some syanthropes (e.g. *Mastomys* sp.) in Arabia (Cleuzio & Tosi 2000). These Bronze Age finds are considerably earlier than on the northern coast of the Gulf of Oman and Persian Gulf. The development of agriculture, in particular date cultivation (a preferred nesting and food source for black rats) and the lack of inter-specific competition from local species would have enabled sea-borne invaders to successfully colonise the region.

R'as Al-Jinz lies just south of the Gulf of Oman on the littoral of the Arabian Sea and is thought to be a seasonal fishing settlement with over 95% of the faunal assemblage being marine-based. Whether R'as Al-Jinz was directly trading with the Indus or not (it is the first landfall if crossing the Arabian Sea from the Indus Valley and Harappan ceramics and seals have been found at the site), it was certainly part of a coastal network trading regionally in resources that maintained the agricultural areas of the interior and dealt in goods such as pyrolusite that entered the trading system(Cleuzio & Tosi 2000; Potts 2008). Genetically, Omani black rats cluster with Indian rats (Tollenaere *et al.* 2010) and, as the maritime connections between the Indus Valley and Arabia are well attested in the Bronze Age (Boivin *et al.* 2009), it is likely that rats were imported through these maritime connections, though subsequent Iron Age re-introductions are likely.

The importance of maritime trade networks for the translocation of rats is again evidenced by the appearance of rats at sites in the Gulf of Oman and the Persian Gulf in the early second millennium BC. At Kalba, a site strategically located at the gateway to the Persian Gulf near the Straits of Hormuz a relatively large number of rat bones were found (47 and all but 2 reliably identified as *R. rattus*) (Phillips & Mosseri-Marlio 2002). In Saar, Bahrain (the entrepôt of trade in the Persian Gulf mentioned in reference to Mesopotamian-Indus connections) 29 *R. rattus* bones were found (Uerpmann & Uerpmann 2005). The majority of imported pottery at Saar comes from Saurashtra, Gujarat, India and some of the important ancient ports and maritime centres along the Saurashtran Coast include the aforementioned sites of Lothal, Padri and Nageswar. There is also typical Mesopotamian style pottery at Saar and Kalba, and Iranian and Harappan pottery in the latter. Dilmunite pottery typical of Saar is found in Kalba suggesting direct associations between the two sites and finally Dilmunite seals are found in the aforementioned Susa, Iran. The maritime connections between distant cultures where contemporaneous rat bones are found are evident in material culture but as yet the routes of biological translocations are unknown.

Rats are unexpectedly absent in the contemporaneous settlement layers of Qala'at al Bahrain with respect to Saar (i.e. second millennium BC) Qala'at al Bahrain was the capital of Bahrain in the Bronze Age and a place where we should expect to find rats as it had an urbanised core and a farming fringe. However, it is not until the late Dilmun period (800-300 BC) that rats (unidentified but thought to be *R. rattus*) are found in in the city (Uerpmann & Uerpmann 1999).

In the Bronze Age of Arabia, occupation sites predominantly flanked the Oman Peninsula and the northern coast of the Gulf of Oman. Outside this settlement zone, Bronze Age rats are absent. It is not until the late first millennium BC that rats appear on the Oman littoral at Khor Rori/Sumhuram (Carenti & Wilkens 2008). Not only were bones found in a building dating to 360 BC -130 AD but rodent gnaw marks were apparent on almost every mammal bone from the temple and almost every stratigraphical context in the occupation area. The port of Khor Rori is identified as Moscha Limen from the *Periplus of the Erythrean Sea* (a first century AD seafaring account of trade in the Indian Ocean) and traded in frankincense imported and exported through both overland and maritime routes (Casson 1989). Specific ties between the port of Kane (southern Yemen) and Moscha Limen are described in the *Periplus* and archaeological evidence of ties between India and Khor Rori are demonstrated by the find of a bronze statue from India in the Omani port (Casson 1989). Again, as Omani rats cluster genetically with Indian rats they are probably imported from India, possibly via an intermediate port, in cargoes unloaded on the frankincense route.

House mice are widespread in the Arabian Peninsula today but their zooarchaeology suggests they were a relatively recent introduction, at least in commensal form. Only two instances of archaeological house mice are found in the region. The first, labelled '*Mus domesticus*' is at the Persian Gulf site of Mleiha *ca.* 300 BC – 1 BC and the second follows shortly after at the nearby site of ed-Dur *ca.* 1 – 200 AD, where they are found in tandem with black rats and described as "penecontemporaneous or a later intrusive" (Potts 2006). The zooarchaeological and genetic evidence are at odds with one another in the case of Arabia as modern house mice in Yemen have an ancient origin. Mice in Yemen are of a distinct, major clade tentatively called *gentilulus* whose native range is restricted to South Arabia (further sampling may find them in other parts of the Arabian Peninsula) (Prager *et al.* 1998). The migration of mice out of their putative Indian cradle occurred in the Pleistocene and we can assume that mice colonised Yemen at this time.

Incidentally, *gentilulus* is the form found in Madagascar (Duplantier *et al.* 2002) that may have been introduced through Arab trade in the late first millennium AD, or perhaps less plausibly through Yemeni-East African contacts in the classical period as described in the *Periplus of the Erythrean Sea* (Casson 1989). The absence of an early mouse zooarchaeological signal may be due to preservation bias as bone preservation in South Arabia is poor, or the Yemeni lineage may be a late-forming commensal and thus opaque in archaeological contexts. The Mleiha/ed-Dur mice could be of the *gentilulus* form introduced through regional exchange networks with South Arabia and the interior, or could be another lineage introduced through Persian Gulf or Indian Ocean trade networks. The origins of house mice in the Arabian Peninsula will remain unknown until wider genetic sampling is undertaken in the region.

Northeast Africa

The long and intimate association between mice and humans is most apparent in Egypt. Papyri from Ancient Egypt detail the ancient origins of the mouse in medicine (Brothwell 1981) and the mummified remains of a child from Naga-ed-Der on the Nile in Upper Egypt (*ca.* 4000-3500 BC) demonstrate how they may have been used (Neolitzky 1911). The mouse was found in the intestines of the child after being skinned and eaten. While seemingly an odd use of the mouse in medicine to me now it has, until relatively recently, been used in such a fashion. Warren R. Dawson writing in 1924 recounts "To-day the mouse is chiefly used for children. I have found instances all over England and Wales. It is usually flayed, fried or boiled or made into a pie, and is given to children for incontinence of the bladder, for dribbling, and for whooping-cough" (Dawson 1924). The ingested mouse from Naga-ed-Der is identified as either the house mouse *Mus musculus* or the morphologically similar spiny mouse *Acomys cahirinus* (a native African species), the former being preferred by the authors.

Ascertaining the Neolithic occupation of Egypt by house mice is not straightforward. Another archaeological instance of mice in the Neolithic occurs in the Dakleh Oasis during a human habitation period that lasted from 6000-2000 BC (Churcher 1999). Although they are certainly contemporaneous (the mice were found under house floors), once again their identification is uncertain and they could be either the house mouse or the spiny mouse. Conversely, numerous positively identified house mice were found in the Neolithic of El Omari in the Nile Delta but the quantity of the finds (30 bones in total) led the authors to propose they are intrusive (Boessneck & von den Driesch 1990). In this case, house mice were found with a number of other native pest species including Sundevall's jird *Meriones crassus*, the African grass rat *Arvicanthis niloticus* and the short-tailed bandicoot rat *Nesokia indica* suggesting pest species were a significant problem at the site. A review of fauna from the Neolithic site of Merimde-Benisalame (Nile Delta) dated 4750-4250 BC, lists the house mouse but it is not discussed further (Boessneck 1988). Remembering that house mice are zooarchaeologically in the Levant by the terminal Pleistocene and in Iran by the eighth millennium BC, and genetics shows they are also native to Yemen, it is possible that they colonised Egypt in the Neolithic from either a source in Southwest Asia or the Arabian Peninsula. In fact, *domesticus* is the common subspecies in northern Africa and Egyptian mice are genetically similar to those from the eastern Mediterranean (i.e. Israel) which could be the source for ancient populations (Bonhomme *et al.* 2011).

The final occurrences of house mice in ancient Egypt are from the Eastern Delta. House mice are found in the Late Hyksos period at Tell el-Dab'a in 1750 BC (Boessneck & von den Driesch). There is perhaps a question mark over this material as identification of rats found at the site in the same period were later rescinded by the authors (discussed later in relation to rats) and the same might apply to mice. The house mouse and other synanthropic and pest species such as the spiny mouse, Sundevall's jird and the greater musk shrew *Crocidura flavescens* briefly occupy Tell el-Dab'a from 1750 but by 1680 - 1620 BC all but the shrew disappeared. Soon after this period black rats occupied Tell el-Dab'a (but see below for a discussion relating to the certainty of the species at the site), perhaps aided by the growing size of the settlement and taking advantage of the commensal niche vacated by earlier commensal and synanthropic species. Finally, "*Mus musculus praetextus"* was listed in faunal remains from Tell el-Maskhuta although no contextual information was described (Boessneck 1988).

Egyptian house mice also cluster genetically with those from the western Mediterranean and Tunisia where subfossil house mice are found in Roman times in Carthage, a significant port of the Roman Empire that supplied it with grain in bulk, and mice may have been stowaways in this cargo (von den Driesch & Baumgartner 1997; Bonhomme *et al.* 2011).

Black rats reported from El Khattara *ca.* 4000 BC proved to be false, which makes the earliest archaeological rats date from the late Hyksos period (1600-1550 BC) at Tell el-Dab'a (Boessneck 1976; Ervynck 2002). In a recent review of the material the authors no longer wished to maintain the claim of the identification as the material is no longer available for study (Ervynck 2002). Admittedly, the next black rat subfossils do not appear in Egypt until the late first millennium BC, which makes the Tell el-Dab'a finds a unique early incidence of black rats in Egypt. However, black rats are in archaeological deposits in Selenkahiye, Syria dated to the third millennium BC and in the Tigris and Euphrates Basin by the Old Babylonian period and by the early second millennium BC Tell el-Dab'a already had substantial Syro-Palestinian immigration that could have been an avenue for commensal translocations.

The only securely dated and identified archaeological rats in Egypt come from ports in the Red Sea. The earliest is from Ptolemaic Berenike where they are found in all stratigraphic layers, and they continue into the Roman period (Van Neer & Ervynck 1998, 1999). Other Ptolemaic rats found in bird mummies identified in the early 20th century are suspected to be misidentified (Lortet & Gaillard 1903; Ervynck 2002). In Quesir al-Qaddim the remains of black rats were found in the stomach and dung balls of a mummified cat in the Roman period (Driesch & Boessneck 1983). Quesir al-Qaddim is the ancient Red Sea port of Myos Hormos described in the *Periplus of the Erythrean Sea* and, along with Berenike, was trading with Rome and the Mediterranean in the west as well as the Indian Ocean network in the east. The black rat was present in the western Mediterranean by the beginning of the Roman period if not before (remains have been found on Corsica dated 151 - 393 BC) and they are potentially imported from this source (Ruffino & Vidal 2010). Or, they could be introduced through ties with the east, namely India (by the late first century BC there was regular traffic between Myos Hormos and north and south India) a destination where rats were firmly established.

Lastly, at Qasr Ibrim "numerous well-preserved rats" were found in excavations (pers. comm. P. Rowley-Conwy). The city was an important inland trade corridor with a long history of occupation (*ca.* 1000 BC – 1811 AD) but as yet the material is unpublished and the stratigraphical contexts of the material are unknown.

In other areas of Northeast Africa (outside Lower Egypt and the Red Sea coast of Egypt) both *Rattus* and *Mus* are exceedingly rare. Only a single example of commensal rats (not identified to species) outside of Egypt occurs at Mai Chot, Eritrea in the period 800-400 BC (Shoshani *et al.* 2008). This may be an artefact of zooarchaeological recovery and identification, but if it is a real signal either individuals are not being transported to Upper Egypt or inland Northeastern Africa or they are unable to colonise it. It is likely to be the latter given the connections between Upper and Lower Egypt via the conduit of the Nile in the ancient period, and between the Mediterranean/Red Sea coasts and inland associated with the Kingdom of Kush in the first millennium BC and the Aksum Empire in the first millennium AD. Lower Egypt has a milder, more stable climate than Upper Egypt and in their early phases of establishment (when human settlements were generally smaller and lower density than they are today) commensals would have relied more heavily on outside habitats than domestic ones. The anthropogenic environment is less seasonal than the natural environment for commensals in that both environmental conditions and food supply are more stable and resources such as shelter and food are often abundant and distributed locally. Although they are able to tolerate short periods of severe conditions, long-term uncertainty can prevent commensals from colonising new territories. House mice were the first commensals to occupy Lower Egypt in the Neolithic but are strikingly absent from the late first millennium BC when rats emerge.

OVERVIEW

Although modern-day house mice and black rats are cosmopolitan, inhabiting almost all the globe, their ancient counterparts had a far more restricted distribution. Zooarchaeologically, house mice tend to be found in inland and riverine sites rather than the coast, whereas, outside their native range black rats tend to inhabit coastal sites. House mice prefer dry, open habitats and black rats prefer humid, agriculturally productive environments and this partly explains their distributions. Mice are also more amenable than black rats to transport by pack animals and carts that would have taken them through the Zagros and the Iranian Plateau, and small boats and canoes that would have unwittingly transported them along the Nile, Tigris and Euphrates rivers. In sites where both mice and rats are found it is usually mice that colonise first, quickly followed by rats. Mice are poor competitors and would not easily be able to colonise new territories where other commensals are already established.

The translocation of commensals by humans across naturally impenetrable geographic barriers such as oceans, mountains and deserts has mediated gene-flow in the species and left an archaeological signature that can be used to trace the history of the humans who carried them. Using well-identified and dated zooarchaeological remains to establish a chronology for the introduction of species to an area combined with analyses of genetic variation in a phylogeographic framework, we can reconstruct histories of the commensal population and by proxy the humans who unknowingly transported them. There are obvious deficiencies in the database upon which this paper is based, mainly due to biases discussed in the introduction. The two major tasks that need to be implemented to make this potentially highly informative proxy more robust are 1. dedicated recovery of commensals in archaeological excavations, employing routine sieving using a fine mesh followed by identification and 2. intensive and widespread sampling of modern specimens for genetic analysis.

CHAPTER 3

THE TAMING OF THE SHREW: MITOCHONDRIAL DIVERSITY AND PHYLOGEOGRAPHY OF THE ASIAN HOUSE SHREW *SUNCUS MURINUS*

INTRODUCTION

Understanding the geographic distribution, genetic diversity, and ecology of an invasive species is key to determining pathways by which they are translocated, and factors associated with their successful colonization. In some species that expand their distribution via human agency, there is the advantage that the species can act as a proxy for inferring human contact networks (Jones *et al.* 2013). The Asian house shrew *Suncus murinus* (Linnaeus 1766) is a widespread commensal species that is native to the Indian subcontinent and has expanded its distribution out of its native India and Sri Lanka in tandem with humans (Hutterer 2005). Its dispersal into Island Southeast Asia (ISEA) as well as the coast and islands of the western Indian Ocean indicates that it is particularly amenable to maritime transportation.

Although some wild populations occur, most Asian house shrews live in close proximity to humans, often inside or underneath dwellings. Such a close relationship suggests a long history of association between humans and the shrew and their distribution and commensal habit make them an ideal proxy for studying nodes of maritime networks in history. Cheke (2010) surveyed historical documentation relating to the introduction of the shrew to oceanic islands in the western Indian Ocean but no review of its archaeological distributions across its entire range has previously been undertaken.

Two main population groups, the ancestral South Asian or “continental” type and the Southeast Asian or “island” type have so far been identified through genetic analyses and a further two contact areas are recognised: the Malay group identified through mitochondrial DNA; and the “Burma” group identified through nuclear DNA (Yamagata *et al.* 1995; Kurachi *et al.* 2007a; b). Hybrid or introgressed populations occur in Sri Lanka that are phenotypically Asian house shrews both morphologically and in their nuclear DNA but are a wild sister species, the Sri Lankan Highland shrew *Suncus montanus*, in their mitochondrial DNA (Meegaskumbura *et al.* 2010). Knowledge of the taxonomy and diversity of the Asian house shrew, and its potential to hybridize with wild species throughout its range, is important for understanding routes of invasion and the ecological and human health implications associated with a species that can adapt to both urban and sylvatic lifestyles.

Early studies on the species identified mainland populations that had diversified within the subcontinent and then were passively spread to ISEA with humans (Yamagata *et al.* 1995; Ruedi *et al.* 1996; Kurachi *et al.* 2007a). Asian house shrews have also been introduced to the Indian Ocean throughout coastal East Africa and the Arabian Peninsula and many islands in the region. The house shrew remains confined to coastal ports and cities in most of its westerly range and based on morphological data it is argued that the species was probably introduced along early maritime pathways, probably with Arabic seafarers in the Islamic period (Hutterer & Tranier 1990). The Mauritian population is thought to be related to an expansion out of Asia in recent times, as shown by mitochondrial restriction fragment length polymorphism (RFLP) analysis (Yamagata *et al.* 1995).

The species is of particular significance as it is used as a laboratory model, transmits emerging infectious diseases such as the plague and hantavirus, and affects local biota through competition and predation (Velimirovic 1972; Varnham *et al.* 2004; Kang *et al.* 2011). Little research regarding genetic diversity or geographic distribution has been conducted on this important animal, and species diversity, introduction pathways and its historical relationship with humans are largely unknown.

This study aims to 1. assess the phylogenetic relationship of *S. murinus* populations through mitochondrial sequence data 2. test whether patterns of translocation to the Indian and Pacific Oceans are related to human movement networks and 3. describe the historical commensal ecology between humans and house shrews. The cytochrome *b* gene is used to infer phylogenies. Cytochrome *b* is commonly used for phylogeographic studies and has been previously used for *Suncus* spp. (Dubey *et al.* 2007; Meegaskumbura *et al.* 2010). Although the rapidly evolving D-loop tends to offer higher resolution in relation to recent evolutionary scenarios that are perhaps more relevant to the hypothesised translocation scenarios of Asian house shrews, its use was precluded in my study as Asian house shrews have been shown to be heteroplasmic in this gene (Yamagata & Namikawa 1999). I sampled heavily in the understudied native range of India and I focus on the islands of Madagascar and Réunion in the Indian Ocean, where introduced house shrews are entrenched in the environment. I surveyed archaeological reports to determine the historical and prehistorical commensal status of the Asian house shrew in the Indian and Pacific Ocean regions.

MATERIALS & METHODS

Taxon Sampling

Morphologically identified *S. murinus* were sampled from localities within its native distribution of India and Sri Lanka as well as colonised ranges in the Pacific and Indian Oceans mostly from museum specimens. I also searched museum databases for *Suncus montanus* (incorporating the Nilgiri Highland shrew *Suncus niger*, a newly classified taxon previously *S. montanus niger*) outside of their native range of Sri Lanka and India in case of hybrids that are known to occur but found none (Meegaskumbura & Schneider 2008).

Material was collected from the American Museum of Natural History, New York (AMNH) in June 2011 by Heidi Eager, Rodrigo Vega, Michelle Mundee and Jeremy Searle; The Field Museum of Natural History, Chicago (FMNH) in October 2011 by Heidi Eager, Alex Trinks, Ardern Hulme-Beaman, Ophelie Lebrasseur, Anna Linderholm and Keith Dobney; and the Smithsonian Institution National Museum of Natural History, Washington D.C. (NMNH) in December 2012 by Heidi Eager and Soraia Barbosa. The Natural History Museum, London (NHM) collected and sent samples after a visit by Heidi Eager and Ardern Hulme-Beaman to identify suitable samples that would yield material.

Osteocrusts (brain material remaining in the cranium after the bone cleaning and preservation process) were picked with either a surgical curette or needle onto clean aluminium foil and poured into a sterile 1.5ml tube. Curettes and needles were sterilised between samples by washing in 20% bleach, ddH_20, and finally 99% ethanol. This is a minimally invasive and minimally destructive technique that preserves the prepared specimen. Skins samples (approx. $5mm^2$) were sliced from either the suture area or a previously damaged area with a sterile, disposable razorblade and placed in a sterile 1.5ml tube.

Previously collected tissue stored in either ethanol or DMSO was supplied by Michel Pascal for Réunion (60 samples, collected in 2006-2007); Jean-Marc Duplantier for Madagascar and Comoros (12 samples, collected in 1999); and FMNH for Madagascar and Comoros (19 samples, collected between 1995-2006 by AJ Tianarifidy, SM Goodman and Z Rakotomalala). A total of 217 tissue, dry-preserved skin and osteocrust samples were available for analysis. Not all samples yielded enough sequence to be included in the analysis, and some sequences were not reliable and so were excluded from analysis. This is discussed in more detail later. A subset of modern Réunion shrews showed very little sequence variation (many individuals shared the same haplotype) and so the collection was only partially sequenced to reduce cost. Appendix 3.I lists sample information and Figure 3.1 shows a map with samples new to this study.

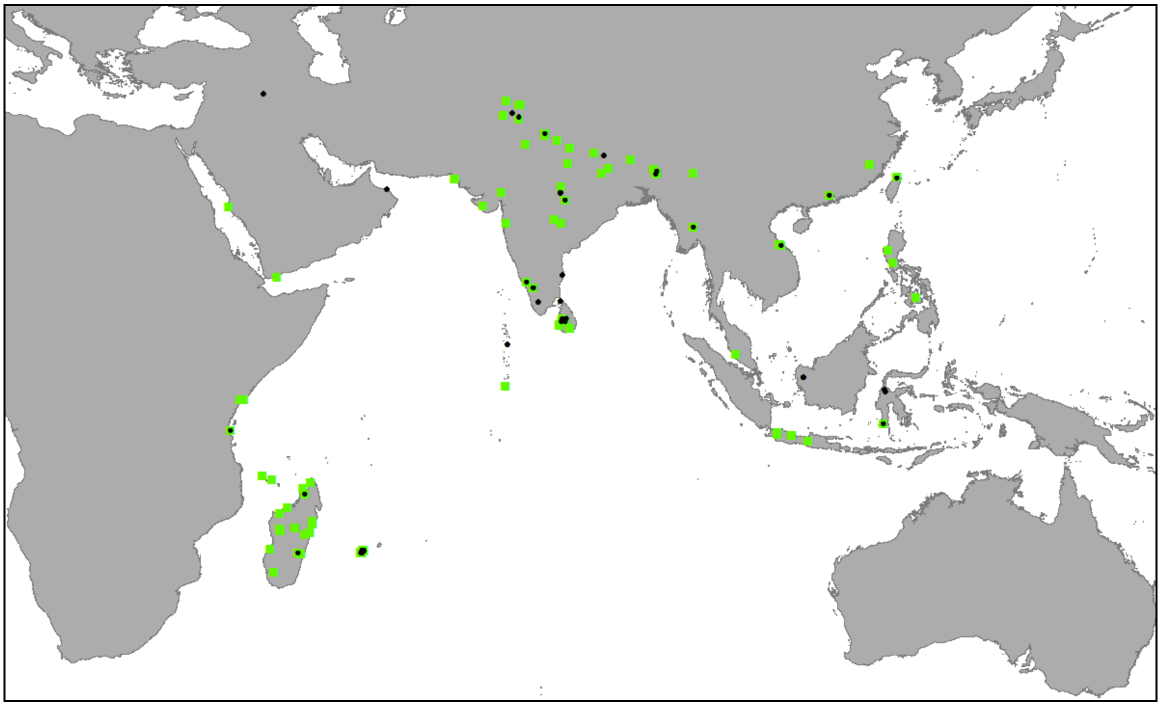


Figure 3.1 Location of *Suncus murinus* collected. Green squares=sequences used in phylogenetic analysis; black circles=sequences not used (see 3.1 for a description of why sequences were not used).

A comprehensive search for archaeological material available for ancient DNA (aDNA) analysis yielded only one *S.murinus* element, a well-preserved rostrum and teeth from Vallée Secrète, in Cilaos, Réunion, provided by Jean-Denis Vigne/Anne-Laure Dijoux (Fig. 3.2). As only one element was available for analysis, the bone was probably from a modern deposit, and due to the expense and time associated with aDNA extraction and amplification, I did not analyse the bone.

Figure 3.2 Dorsal, lateral and ventral views of a *Suncus murinus* rostrum and teeth from Vallée Secrète, Cilaos, Réunion.

Molecular methods

Samples were divided into “modern” tissue material, and “museum” osteocrusts and skins, and were processed separately at the extraction and polymerase chain reaction (PCR) stages. Much of the material sampled (osteocrusts and skins) was collected at tropical latitudes as much as 110 years ago and has since been stored at room temperature with no special DNA preservation techniques. As such, samples will feature many of the signatures of degraded DNA such as low copy numbers, short fragments, and post-mortem mutations that make them more susceptible to contamination and amplification of nuclear encoded copies. I adopted a number of specialized protocols to prevent contamination and non-specific amplification.

Storage, extraction and PCR setup of osteocrusts and skins were conducted in a dedicated degraded pre-PCR facility that has a dedicated isolated air flow system on a different site to modern tissue and post-PCR work. Work was conducted under a laminar flow hood and all non-organic materials and work spaces were UV-irradiated for 30 min prior to use and 4 h after use; gloves were changed between handling each specimen tube at the extraction stage; negative controls were interspersed 1:11 at the extraction stage and one per primer pair at the PCR setup stage; samples from the same locations were extracted on different days; and amplicons were generated replicatively on different days and with different primer pairs.

Osteocrust and skin samples were extracted using Qiagen QIAmp DNA Micro Kits (Qiagen, Hilden, Germany) following the manufacturer's protocol. If a specimen did not digest overnight, 10 µl Proteinase K was added and the sample left to digest for an extra 4 h. Skins were washed in 1 x Phosphate Buffered Saline two times, followed by three washes in ddH$_2$O. DNA was eluted twice (50 µl each) and the second elution used in PCR reactions.

Modern tissue was extracted using Qiagen DNeasy Blood and Tissue Kits (Qiagen, Hilden, Germany) following the manufacturer's protocol. Tissue in DMSO was washed in 500ml ddH$_2$O two times prior to extraction. Tissue in ethanol was evaporated prior to extraction. DNA was eluted twice (50 µl each) and the second elution used in PCR reactions.

The whole of the cytochrome *b* gene and portions of the flanking tRNAs were amplified in overlapping fragments. Table 3.1 lists the primers used in this study. Modern samples were amplified with primers MVZ05/SmF2R and SmF2F/MVZ14, except in cases where degradation or inhibition prevented successful amplification. In these cases various combinations of other primers were used to complete the sequence. Historical samples were amplified in four fragments using Primer Set 1: MVZ05/SmF1R, SmF2F/SmF2R, SmF3F/SmF3R, and SmF4F/MVZ14. Due to degradation through age or preservation, museum samples sometimes generated hairpin loops with the above primer sets, or did not generate some or all amplicons. In these cases amplicons were generated using a second set of primers with shorter fragment length (Primer Set 2): Smv21F/Smv21R, Smv22F/Smv22R, Smv23F/Smv23R, Smv24F/Smv24R, Smv25F/Smv25R. In cases where sequences were obtained for museum samples using Primer Set 1, I also sequenced individuals using Primer Set 2 in order to act as an independent control to identify any potential numts amplified by a particular primer pair.

Primer Name	5' - 3'	Reference
MVZ05	CGAAGCTTGATATGAAAAACCATCGTTG	Meegaskambura & Schneider 2008
SmF1R	GCCATAATAAAATCCTCGGCCC	This study
SmF2F	GCAAATGGCGCYTCC	This study
SmF2R	TTGTTTGAGCCAGTTTCGTG	This study
SmF3F	TTTATCGTTACTGCCCTTG	This study
SmF3R	TATGCTTCGTTGTTTGGCTG	This study
SmF4F	AATTAGGCGGAGTTCTGGCA	This study
MVZ14	GGTCTTCATCTYHGGYTTACAAGAC	Meegaskambura & Schneider 2008
Smv21F	CGTTGTCTTTCAACTACAGAAACC	This study
Smv21R	ACATGGAGAAAAAGGCAAATAAAA	This study
Smv22F	CGAGACGTAAATTATGGCTGACT	This study
Smv22R	GAGAAGCCGCCTCAGATTCA	This study
Smv23F	TCTCTTATCAGCTATTCCCTATATC	This study
Smv23R	CTGGAGAAAATAGGACTAAGGAGG	This study
Smv24F	AAAGATATCCTTGGAGCACT	This study
Smv24R	TTCAGAATATACATTGGCTTATTG	This study
Smv25F	CTTCACACAGCCAAACAACG	This study
Smv25R	CCGGTTTACAAGACCAGAGT	This study

Table 3.1. *Suncus murinus* cytochrome *b* primers used in this study.

As the quality of my DNA samples varied due to age and preservation I did not follow a uniform PCR protocol. Twenty-five µl PCR reactions contained ddH$_2$0, 1 µl of DNA template and final concentrations of 1 x buffer, 2 mM - 4 mM MgCl$_2$, 2-2.5 mM dNTPs, 2 - 2.5 mM each primer, and 0.6-1.5 units JumpStart *Taq* (Sigma-Aldrich, St. Louis, USA), Platinum Taq or Platinum Taq High Fidelity (Life Technologies, Carlsbad, USA). Products were thermocycled 30 times for modern material and 35-40 times for historical material. Annealing temperatures (T_a) were 55°C for all primer pairs, except SmF2F/SmF2R and Smv25F/Smv25R (T_a=51°C).

PCR reactions were cleaned with ExoSAP-IT (Affymetrix, Santa Clara, USA) following the manufacturer's protocol. Products were sequenced in either one (for some individuals from modern material only) or both directions on an Applied Biosystems ABI 3730. One µl of modern and 2 µl of historical template were used in sequencing reactions.

Sequence alignment and data generation

Contigs were edited and assembled in Geneious v 6.1.6. Missing nucleotides in partial sequences were replaced with N. The amplification of nuclear encoded copies of mitochondrial genes (numts) is a serious concern when the starting DNA template is degraded (Ho & Gilbert 2010). Inspection of chromatograms, inspection of translation sequences, analysis of transition/transversion ratios and analysis of synonymous/non-synonymous ratios did not show any evidence of numts within the sequences generated. Furthermore, I had molecular controls of overlapping fragments and alternative primers with which I could detect numts.

The final dataset did not include all collected individuals (Appendix 3.I). Twenty individuals did not amplify at all. Thirteen individuals had large amounts of missing data - only two were included in the final dataset (from coastal Kenya) as the remaining individuals had genetically similar counterparts from very close locations. Sequences of three individuals showed probable deamination as heterozygous C/T and G/A peaks (but no other combinations of states) were evident, even after reamplification. Five individuals were potentially cross-contaminated with exogenous DNA as chromatograms showed background amplification and portions of the sequence were unreadable. Reamplification and in some cases re-extraction (where material was available) did not improve the sequence quality. DNA from multiple individuals (conspecific or not) might be expected from samples stored in dry museum collections. It is therefore necessary to check the validity of sequences and phylogenies produced from such data. I found no evidence of cross-contamination in the final dataset. Time and money did not permit me to clone potentially deaminated or contaminated sequences. In most cases where sequences with missing, deaminated or contaminated bases were excluded, a closely related individual within a proximal location was sequenced, except in Borneo where no sequence was obtained.

The taxonomy of *S. murinus* is unresolved with respect to other closely related species. Historically, species within the *Suncus* genus were identified based on morphological traits and geographic locations but morphological variability and a wide distribution within *S. murinus* has resulted in unclear taxonomic boundaries for the species. Hybridisation between *S. murinus* and closely related species has further complicated systematic identification (Meegaskumbura & Schneider 2008). To check for misidentified or hybrid/introgressed individuals in my dataset an initial phylogenetic tree was generated with my samples and with representatives of each taxon with the genus name *"Suncus"* available on GenBank (data not shown). Then, new sequences obtained in my study were aligned with previously published sequences available on GenBank that clustered within the closely related *S. murinus*, *S. niger* and *S. montanus* (sister group to *S. murinus*) groups (Table 3.2). I omitted some published sequences that were relatively short and some sequences from Dubey *et al.* (2007) that had unreliable readings at the beginning of the sequence. For phylogenetic analysis, alignments were truncated to the cytochrome *b* gene only, excluding the tRNAs (1,140bp). The final dataset consisted of 179 sequences, 140 new to this study and 39 published sequences.

As some of the sequences contained missing nucleotides, an abridged dataset was produced in order to calculate genetic diversity/neutrality statistics, haplotype networks and mismatch distributions. The five *S. niger* sequences were not included as most of the sequences contained stretches of missing nucleotides. The start codon and terminal three codons (including the stop codon) were eliminated as these contained missing data in many sequences but were not variable (containing only one singleton variation within the *S. murinus* group). As there was no uniform pattern to the position of the missing nucleotides within the cytochrome *b* gene, sequences (rather than sites) were removed. Of a possible 174 sequences, 143 were included in the dataset. The majority of sequences that were eliminated had counterparts that were similar in both their location and sequence. By chance, sequences were eliminated roughly proportionally throughout the dataset (clade by clade).

GenBank Accession	Length	Country	Reference
AB033610	1,140	?Japan	1
JF784169	1,140	China	2
JF784170	1,140	China	2
JF784171	1,140	China	2
FJ716833	1,140	Sri Lanka	3
FJ716834	1,140	Sri Lanka	3
FJ716835	1,140	Sri Lanka	3
FJ716837	1,140	Sri Lanka	3
EF524756	395	Philippines	4
EF524669, EF524777	375, 391	India	4
EF524668, EF524776	375, 393	?	4
EU122224	1,140	Sri Lanka	5
FJ813943	685	Philippines	6
FJ813949	1,074	Philippines	6
FJ813963	1,140	Philippines	6
DQ630390	1,122	?	7
AB175074	1,140	Japan	8
AB175075	1,140	Taiwan	8
DQ975251	985	?	9
AB066247	402	Japan	10
D85370	402	Japan	11
GQ290372	1,140	Sri Lanka	12
GQ290379	1,140	Sri Lanka	12
GQ290380	1,140	Sri Lanka	12
GQ290365	1,140	Sri Lanka	12
GQ290366	1,140	?	12
GQ290367	1,140	Sri Lanka	12
GQ290368	1,140	Sri Lanka	12
GQ290369	1,140	Sri Lanka	12
GQ290370	1,140	Sri Lanka	12
GQ290371	1,140	Sri Lanka	12
GQ290373	1,140	Sri Lanka	12
GQ290374	1,140	Sri Lanka	12
GQ290375	1,140	Sri Lanka	12
GQ290376	1,140	Sri Lanka	12
GQ290377	1,140	Sri Lanka	12
GQ290378	1,140	Sri Lanka	12

Table 3.2 Published *Suncus* spp. sequences used in this study. 1.(Onuma *et al.* 2000) 2. (Guo *et al.* 2011) 3. (Meegaskumbura & Schneider 2008) 4. (Dubey *et al.* 2008b) 5. (Meegaskumbura *et al.* 2007) 6. (Esselstyn *et al.* 2009) 7. (Dubey *et al.* 2007) 8. (Ohdachi *et al.* 2006) 9. (Dubey *et al.* 2008a) 10. (Motokawa *et al.*) 11. (Ohdachi *et al.* 1997). 12. (Meegaskumbura *et al.* 2010).

Data analysis

Phylogenetic analyses were conducted using Bayesian Markov Chain Monte Carlo (MCMC) simulations in MrBayes v3.2.1, and Maximum Likelihood (ML) with bootstrapping in PhyML3.0 (Guindon *et al.* 2010; Tamura *et al.* 2011; Ronquist *et al.* 2012). Data was not reduced to haplotypes given that missing nucleotides were a feature of the dataset. *Suncus dayi* was used as an outgroup as it is within the *Suncus* genus and closely related but external to the *S. murinus/S. niger/S. montanus* group (Meegaskumbura & Schneider 2008). The best-fit substitution models for Bayesian inference as chosen by Bayesian Information Criterion (BIC) and ML as chosen by Akaike Information Criterion (AIC) were estimated in jModelTest v2.1.4, using 11 substitution schemes (Posada 2008). A model with a HKY+I+Γ substitution scheme was implemented. For Bayesian analysis, two independent analyses were run from random starting trees with four chains for 20 million generations sampled every 1000 generations. A 25% burn-in was used. To check for convergence and stationarity of MrBayes runs I examined the standard deviation of split frequencies statistics of the two independent runs and used the slide and compare functions in Are We There Yet (AWTY) (Nylander *et al.* 2008). FigTree v 1.4.0 was used to visualise the tree (Rambaut 2014a). For ML analysis, gamma rate categories was set to 4, and all other substitution parameters were estimated by the program. Tree searching was done by subtree pruning and regrafting (SPR) and nearest neighbour interchange (NNI). To test the phylogeny, 500 bootstrap replicates were executed.

Post-mortem DNA damage is a characteristic of the osteocrust and skin samples I collected. Degradation of the DNA molecule causes deamination resulting in predominantly type II transitions (G > A and C > T). While my molecular approach of amplifying overlapping fragments and replicating sequences minimised error in the final dataset, I implemented a Bayesian MCMC sequence damage model (for transitions only) to test the robustness of my phylogeny in order to account for uncertainty in my data (Ho *et al.* 2007). Briefly, the model accounts for DNA damage by adding delta (Δ) additional substitutions per site to each external branch of the tree in an age-independent process. The model was run on a dataset consisting only of *S. murinus* individuals. XML files were compiled in BEAUTI v1.8.0 and MCMC run in BEAST v1.8.0 (Drummond *et al.* 2012). Bayesian MCMC sampling was implemented with and without the Δ damage parameter. A HKY+I+Γ substitution model was implemented. A strict clock (estimated rate) was used and a constant size prior placed on the tree (Jeffrey's prior on the parameter). A uniform prior (0, 0.1) was placed on the Δ parameter. The upper bound was <1 as molecular controls limited error in the sequences. Two independent analyses for each model were run from random starting trees with four chains for 20 million generations, sampled every 1000 generations. Convergence and appropriate mixing were checked by visually inspecting the trace and model parameters in Tracer v1.3, and in AWTY using the slide and compare functions (Rambaut 2014b). A 10% burn-in was used. Runs were combined with LogCombiner v 1.7.5 and a median tree generated with TreeAnnotator v 1.7.5. FigTree v 1.4.0 was used to visualise the consensus tree.

Even though sequences in my dataset were temporally spaced, differing by as much as 100 years in their collection date, I did not implement a phylogenetic inference model to account for uncertainty in the genealogy of non-contemporaneous sequences (Drummond *et al.* 2002). The temporal distance between the samples is relatively shallow in evolutionary terms. House shrews are polymorphic for breeding behaviour with females in some populations producing litters only once a year, while others breed year round (Chang *et al.* 1999). Sexual maturity is reached at about 35 days. Although my oldest sample could theoretically differ from my modern samples by several hundred generations, it is likely to be far fewer, and all other museum samples will be younger than this in evolutionary terms. I did not expect to see divergence in cytochrome *b* sequences purely based on non-contemporaneous collection, and in fact where I had temporally spaced data from Madagascar, Taiwan, China (my sequences and published data) I recovered identical haplotypes.

Phylogeography uses genetic markers to examine population history. A theory that supposes a clock-like rate of molecular evolution is often practically used to infer the timing of species or population differentiation in phylogeography, and is termed the 'molecular clock'. Typically, the molecular clock is calibrated at a species boundary using the date of a fossil from paleontological deposits as the date of divergence from the most recent common ancestor (tMRCA). However, a decay in nucleotide substitution rates with time has been noted and therefore the inconstant rate of molecular evolution across the time-scale from the present to the fossil era has brought criticism to current molecular clock dating methodologies (Ho *et al.* 2005; Ho *et al.* 2007). At both long and short term scales the molecular clock is inefficient at determining the tMRCA due to nucleotide substitutions that are either saturated (more than one change at a site) or not fixed (mutations that are not common between all descendants of a most recent common ancestor), respectively (Ho *et al.* 2005; Ho *et al.* 2007).

In a study examining divergence of the Eurasian field vole *Microtus agrestis*, Herman *et al.* (2011) calibrated the molecular clock directly from the data and recent geophysical events (introduction to islands that could only have occurred during glacial periods) and calculated times to the most recent common ancestor using coalescent simulations. They found a nucleotide substitution rate that was four times higher than the intrageneric substitution rate commonly used for the *Microtus* genus, and twenty times higher than the frequently employed rodent mitochondrial DNA rate.

Appropriate calibration of the molecular clock is critical for examining the timing of population divergence. For my study this is particularly pertinent as the human-aided introduction of commensals functions within very recent history with regards to the biological history of the species. An accurate mutation rate can be determined by incorporating known sequence data from ancient DNA, or by relating lineage divergence with contemporaneous external (e.g. geologic) events (Herman *et al.* 2011).

I attempted to estimate the tMRCAs of the major clades and subclades by calibrating the molecular clock of *S. murinus* using prior distributions on the tMRCA of two introductions to islands that presumably could only have been human-mediated (Réunion or Madagascar) using Bayesian MCMC in BEAST v1.7.5. Simulations were run using only one calibration point, or both. The tMRCA prior of Réunion was assigned a gamma distribution with upper and lower limits [486, 0], respectively. The tMRCA prior of Madagascar was assigned a gamma distribution with upper and lower limits [1514, 0], respectively. Upper limits were truncated to the presumed earliest possible introduction date, currently the first evidence of human occupation of the islands. A HKY+I+Γ substitution model was implemented with both strict and uncorrelated lognormal relaxed molecular clocks tested, under either expansion growth or Bayesian Skyline models. Two independent analyses for each model were run from random starting trees with four chains for 50 million generations, sampled every 1000 generations. A 20% burn-in was used. The data analysis was unsuccessful as evidenced by lack of suitable parameter statistics and erroneous tMRCA estimation and therefore data is not shown in this thesis. Shallow calibrations have been experimentally shown to underestimate clade ages by up to three orders of magnitude and, from all models assessed here, the resulting tMRCAs for clades were severely underestimated, perceivably by up to two orders of magnitude (Duchêne *et al.* 2014). With no further external evidence with which to calibrate medium-scale clades I was unable to reliably calibrate the *S. murinus* tree and therefore I refer only to relative timing of events between lineages.

Neighbour-net networks were generated in SplitsTree v 4.13.1 using an equal angle algorithm and uncorrected p-distance. Median-joining haplotype networks (ε=0) were generated in Network v 4.6.1.2 for Clades A and C (Bandelt *et al.* 1999). To reduce cube dimensionality between median vectors in Clade C, multiple substitutions at a character were downweighted to 8.

Genetic distances between populations (uncorrected p-distance) were calculated in MEGA5. Population genetic and neutrality statistics were calculated in dnasp5 for clades (except where number of sequences was too low) (Rozas *et al.* 2003). Statistics were calculated for populations within the native range of the Indian subcontinent and Sri Lanka, as well as supposedly introduced populations from Madagascar & Comoros (treated as a single unit) and Réunion. Finally, in one case a repeat dataset was calculated for the C_{IV} lineage with two anomalous published sequences removed. The sequences (AB175074 andAB033610, Table 3.2) comprised unusually high levels of nucleotide substitutions and are shown as long branches in Figure 3.3. Both are from papers where research was conducted in Japanese institutions and one details the collection location as Okinawa, Japan (AB175074) (Onuma *et al.* 2000; Ohdachi *et al.* 2006). Neither paper lists the method of collection. It is possible these individuals are from inbred laboratory lines, whose origin of collection was Japan, as the house shrew is a common laboratory model in Japan.

I report haplotype diversity, nucleotide diversity (π) (uncorrected for multiple substitutions), population mutation rate θ (Watterson's) per site from S (segregating sites). Tajima's D (sites), Fu's F_S (sites) and Ramos-Onsins & Rozas' R_2 (pairwise differences under a constant population size model). F_S is more sensitive to population expansion and genetic hitchhiking than D, and R_2 is able to detect recent population growth in small sample sizes (Fu 1997; Ramírez-Soriano *et al.* 2008). $D<0$ and $F_S <0$ indicates population expansion or purifying selection/genetic hitchiking, $D>0$ and $F_S >0$ indicates a bottleneck or overdominant selection, and R_2 close to zero (positive) indicates recent population growth. P-values for F_S and R_2 were calculated by coalescent simulation based on segregating sites (S) and for D based on θ with no recombination. Ten thousand iterations and a 95% confidence interval were employed. $P<0.05$ for D and $R_{2,}$ and $P<0.02$ for F_S were used to determine significance.

Mismatch distribution plots, the expansion parameter tau (τ) and theta values (theta initial θ_0 and theta final θ_1) for clades, subclades (excluding subclade C_I due to the low number of sequences), for geographical populations of Madagascar/Comoros and Réunion, were calculated in dnasp5 under a neutral infinite sites population growth and decline model. A known mutation rate μ can be used to estimate the time of expansion, however as no reliable μ estimate exists for *S. murinus* I do not perform this analysis. The central tenet of the mismatch distribution is that the trajectory of population size makes waves in the distribution of pairwise differences between individuals in a non-recombining population (Rogers & Harpending 1992). It relies on three parameters, the population mutation rate before population size changes (θ_0), the population mutation rate after population size changes (θ_1), and the time of the population size change measured in generations (τ, τ=2μt, where μ is the sequence mutation rate per generation, and t is generations before present). Given that $\theta = 2N_e\mu$ in a haploid population (where N_e is the effective population size), it is evident that mismatch distribution models depend heavily on estimations of μ and N_e and therefore sequence diversity and number of sequences included in the analysis will influence the model. To tease apart the effect that sampling biases have on mismatch plots, I reproduced these tests with modified datasets: the full sequence set incorporating all individuals; a haplotype dataset; a downsampled dataset (in Madagascar & Comoros and Réunion, where there were relatively many individuals sequenced, only five of each sequence were included); and with anomalous sequences removed in subclade C_{IV}.

Ranges of clades A and C within the subcontinent were interpolated in ArcGIS v10 using a natural neighbour algorithm.

RESULTS

Taxonomy and origin

According to the cytochrome *b* gene phylogeny (Figure 3.3, Appendix 3.II), three of my individuals are the Nilgiri Highland shrew *S. niger*. All other shrews new to this study cluster within a monophyletic group to which *S. montanus* is sister, tentatively labelled *S. murinus.* All branches leading to species groups are well supported.

S. murinus has high haplotype and nucleotide diversity (H_d= 0.924 ± 0.000; π= 0.0241 ± 0.0018) compared to, for example, the common shrew *Sorex araneus* from Russia, Poland and Scotland (H_d= 0.912 ± 0.014; π=0.0036 ± 0.0023), or another widely distributed species such as *Rattus norvegicus* the Norway rat (H_d=0.86 π= 0.0047) (Table 3.3) (Raspopova & Shchipanov 2011; Song *et al.* 2014).

Due to their widespread distributions it is not a practical approach to describe clades with regional names as is often used in naturally dispersed species. Here, I follow the naming schemes that have been used for humans and house mice that use letters for clades and numerals for subclades. *S. murinus* individuals cluster into a further three monophyletic groups (A, B and C) , except for one singleton from Burma. All named clades are well supported. Clade A is new to this study for the cytochrome *b* gene but likely represents the “continental” group identified through earlier mitochondrial RFLP analysis and Clade C contains individuals belonging to the previously identified “island” group (Yamagata *et al.* 1995). Clade B grouped with published sequences of the subspecies *Suncus murinus murinus* that were identified by Meegaskambura *et al.* (2010).

To test their relatedness of clades within *S.murinus,* I calculated pairwise differences between clades, and to the sister taxon *S. montanus*. Overall, the topology of the tree, topology of the Neighbour-net network (Figure 3.4) and pairwise distances between

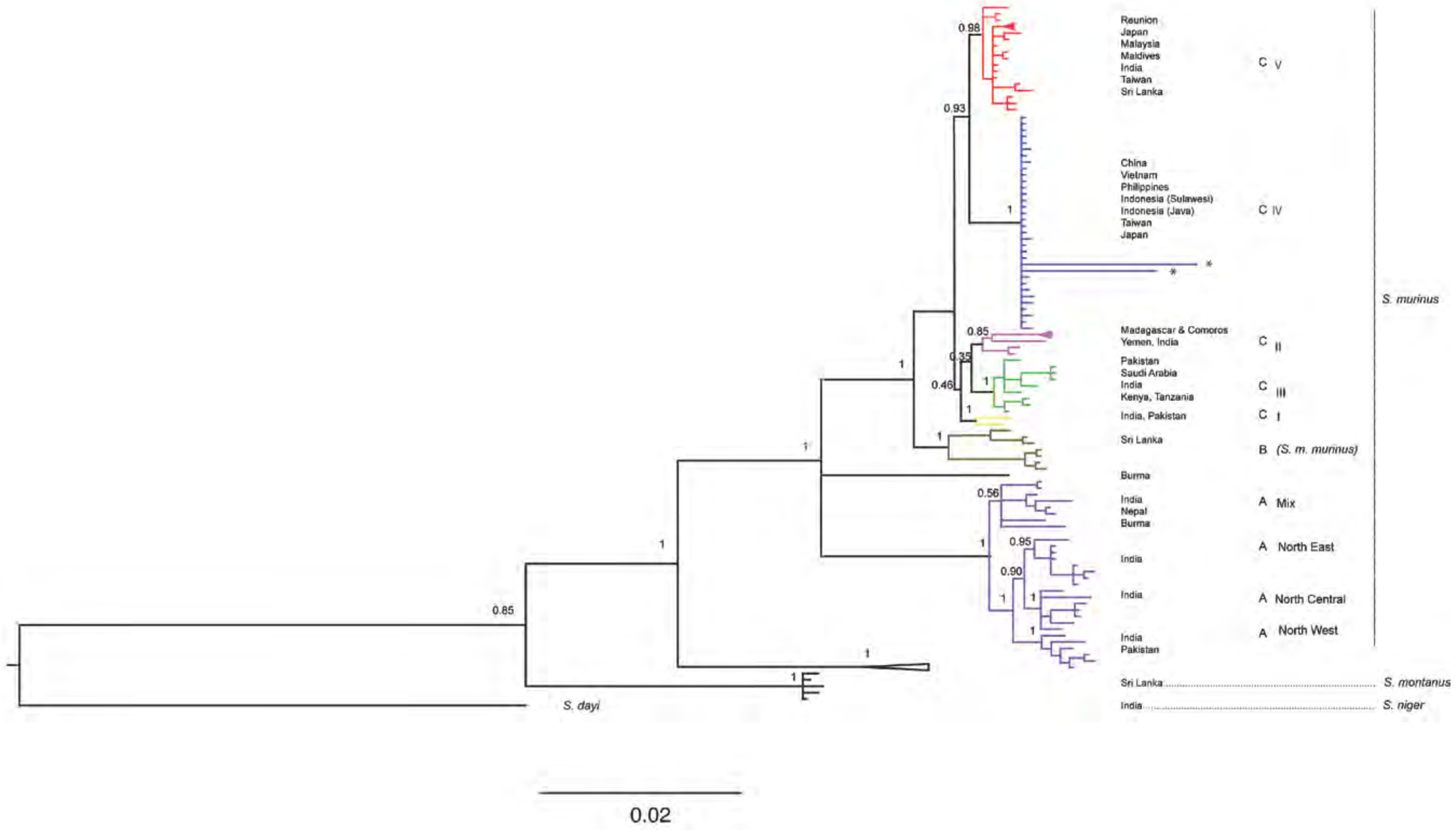


Figure 3.3 Bayesian MCMC tree **(TOP)** and Maximum Likelihood (ML) tree **(BOTTOM)** for the cytochrome *b* gene of *Suncus murinus*. Posterior probabilities or bootstrap values are shown on branches. Clades and subclades are listed to the right, with major phylogeographic units within *S. murinus* colour-coded. Trees are rooted with *Suncus dayi.* The anomalous Japanese sequences are marked with an asterisk (*) at the end of the branch.

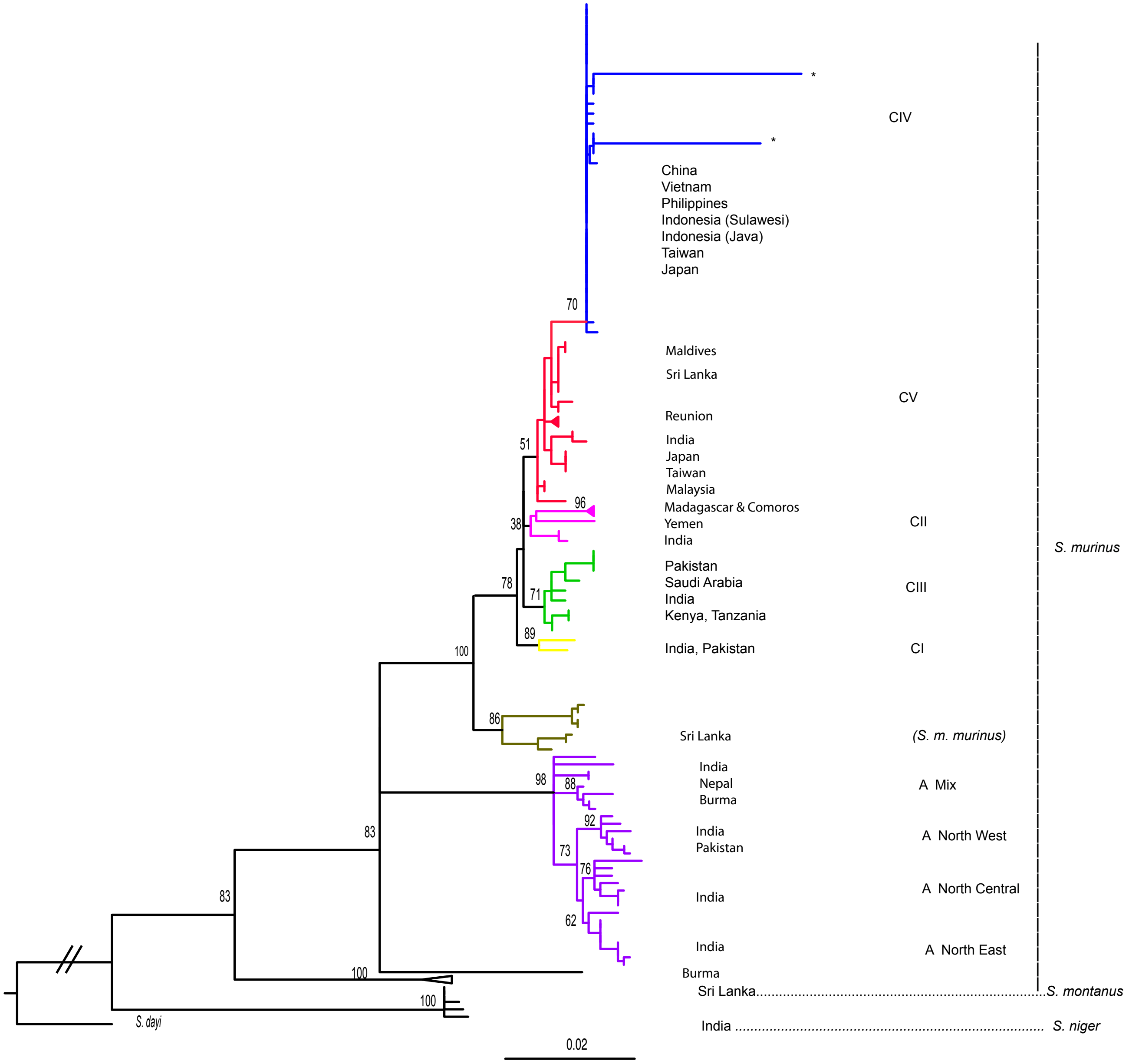


Figure 3.3 continued

Clade/ Population	N	h	H_d (±SD)	π (±SD)	θ (±SD)	D (*P*)	F_S (*P*)	R_2 (*P*)
S. murinus (total)	136	62	0.924 (0.000)	0.02407 (0.00176)	0.03393 (0.00804)			
A (total)	27	23	0.989 (0.013)	0.01176 (0.00071)	0.01817 (0.00600)	-1.4562 (0.051)	-7.36* (0.019)	0.0696* (0.014)
B (total)	7	6						
C (total)	102	33	0.866 (0.020)	0.00954 (0.00070)	0.01996 (0.00509)	-1.7374* (0.014)	-3.035 (0.304)	0.0477* (0.015)
C_I	2	2						
C_{II}	30	6	0.31 (0.109)	0.00174 (0.00098)	0.00515 (0.00187)	-2.3356* (0.001)	0.347 (0.618)	0.0751* (0.05)
C_{III}	7	5						
C_{IV}	28	9	0.497 (0.117)	0.00366 (0.00206)	0.0123 (0.00413)	-2.6827* (0.000)	0.265 (0.590)	0.1144 (0.447)
C_{IV} (no anomalous)	26	7	0.415 (0.121)	0.00041 (0.00014)	0.00139 (0.0007)	-2.0912* (0.000)	-6.233* (0.002)	0.0702* (0.021)
C_V	35	11	0.667 (0.087)	0.00237 (0.00042)	0.00431 (0.00157)	-1.52038 (0.052)	-2.339 (0.194)	0.0643* (0.028)
Indian subcontinent	34	28	0.988 (0.011)	0.02675 (0.00327)	0.02840 (0.00876)			
Sri Lanka	14	9	0.912 (0.059)	0.0148 (0.00158)	0.01171 (0.00461)			
Madagascar/ Comoros	25	4	0.230 (0.110)	0.00021 (0.00011)	0.00070 (0.00045)	-1.7333* (0.001)	-3.084* (0.020)	0.1083 (0.326)
Réunion	21	2	0.095 (0.084)	0.00008 (0.00007)	0.00025 (0.00025)	-1.16356 (0.096)	-0.9189 (0.319)	0.2130 (0.718)

Table 3.3. Genetic diversity and neutrality statistics for *Suncus murinus,* within each clade, within C subclades (and a modified dataset with anomalous Japanese sequences removed), and within geographic populations of the Indian subcontinent, Sri Lanka, Madagascar/Comoros (C_{II}) and Réunion (C_{IV}). N number of sequences. h number of haplotypes. H_d haplotype diversity. π nucleotide diversity. θ from sites, population mutation rate. D Tajima's D. F_S Fu's F_S. R_2 Ramos-Onsins & Rozas' R_2. *indicates statistically significant from coalescent simulations ($P<0.05$ for D and R_2, $P<0.02$ for F_S). Statistics not calculated for B, C_I and C_{III} due to lack of sequences.

clades and an outgroup (Table 3.4) indicates relatively ancient diversification of Clade A and Clades B/C into separate evolutionary lineages, followed by diversification within *B*/C into separate lineages.

There is further genetic structure within Clades A and C as the MJ haplotype networks also show (Figures 3.5 and 3.6), and possibly also B. Substructure of B is not investigated further given the limited number of sequences within it. The trees identify three well-supported clades in the C group (C_I, C_{III}, and C_{IV}) and a further clade with low support (C_{II}). Clade C_{IV} is nested within Clade C_V in the ML tree but is distinct in the Bayesian tree, where it has relatively high branch support. The unstable topology of these two clades is also reflected in the MJ network, which shows that C_{IV} is connected to C_V by a median vector but it is not connected to any other clade, and they are in either a sister, or ancestral-derived relationship. The unstable topology of these two clades is likely due to sampling bias in the core C_V clade, which is more closely related to every other sublade than any other subclade is to each other (Table 3.6). As Clade C_{IV} is monophyletic and well-supported in both Bayesian and ML trees, I refer to C_{IV} and C_V as distinct subunits in my analysis, and refer hereafter to the Bayesian tree when statistics are reported, unless otherwise stated. Affinity of sequences within clades is congruent between both Bayesian and ML trees, except that some sequences that cluster within the poorly-supported "Mix" group of the A clade in the Bayesian MCMC tree are basal in the ML tree. MJ networks within clades show a number of median vectors in their output, which represent either unsampled or extinct ancestral haplotypes (Figures 3.5 and 3.6) and there is low support for relationships between taxa within clade C in the tree (PP=0.35 - 0.93) (Fig 3.3).

The subcontinental population has higher haplotype and nucleotide diversity than Sri Lanka, although both populations show relatively high diversity at the intraspecific level (Table 3.3).

	S. murinus A	*S. murinus* B	*S. murinus* C
S. murinus B	5.1		
S. murinus C	5.1	2.5	
S. montanus	7.2	6.3	6.4

Table 3.4 Percent pairwise uncorrected distance between *Suncus murinus* clades, and between *Suncus montanus.*

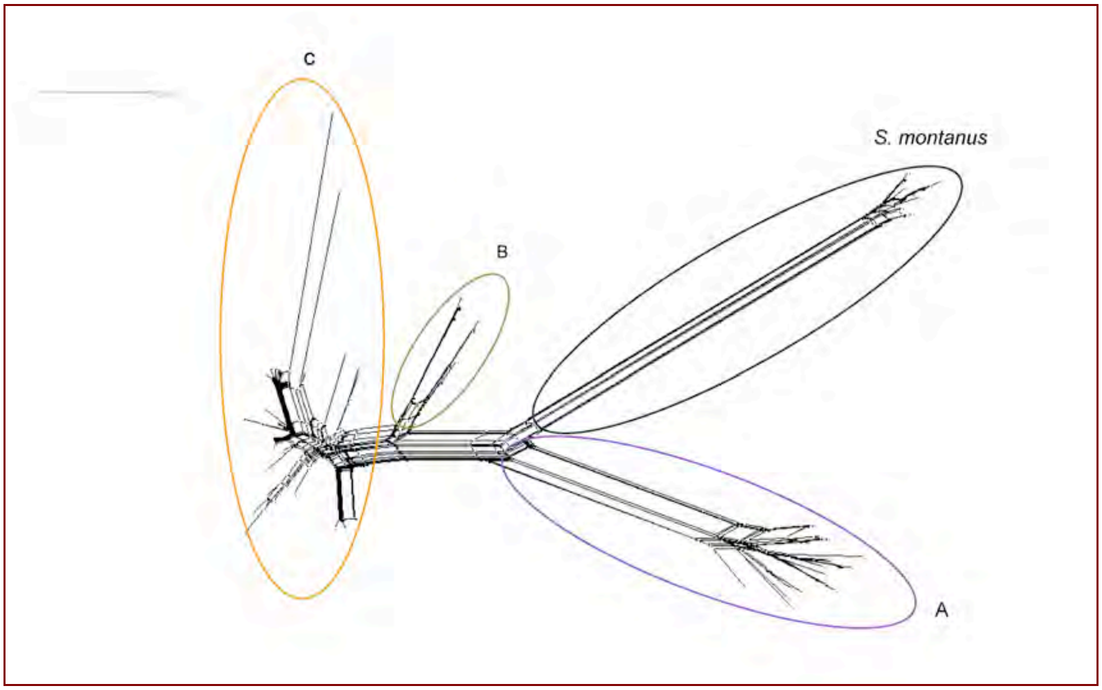


Figure 3.4. Neighbour-net network (uncorrected p-distance) for clades A (purple), B (brown), and C (orange), and *Suncus montanus.*

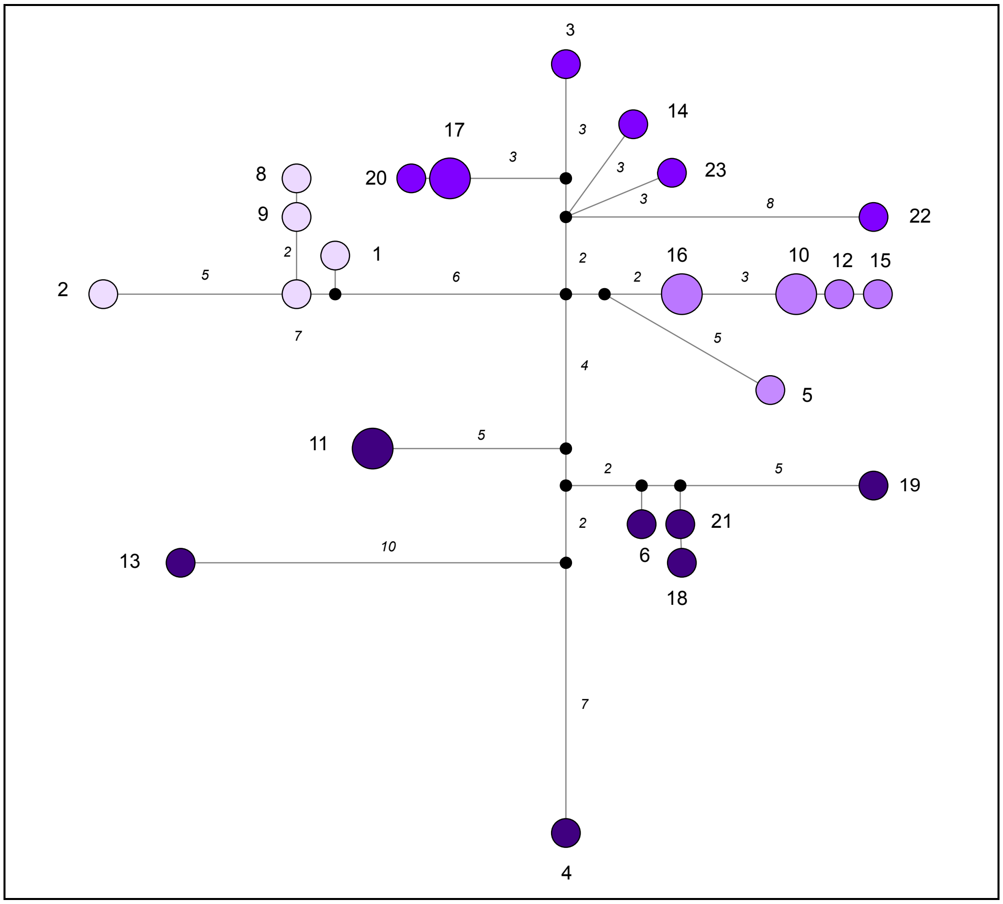


Figure 3.5. Median- joining haplotype network for Clade A. Subclades are coloured shades of purple and are (clockwise from top) North Central, North East, Mix, and North West. Circles represent haplotypes and area is proportional to number of sequences, with haplotype numbers given outside circles. Black dots are median vectors. Number of substitutions (> 1) are given in italics along the branch.

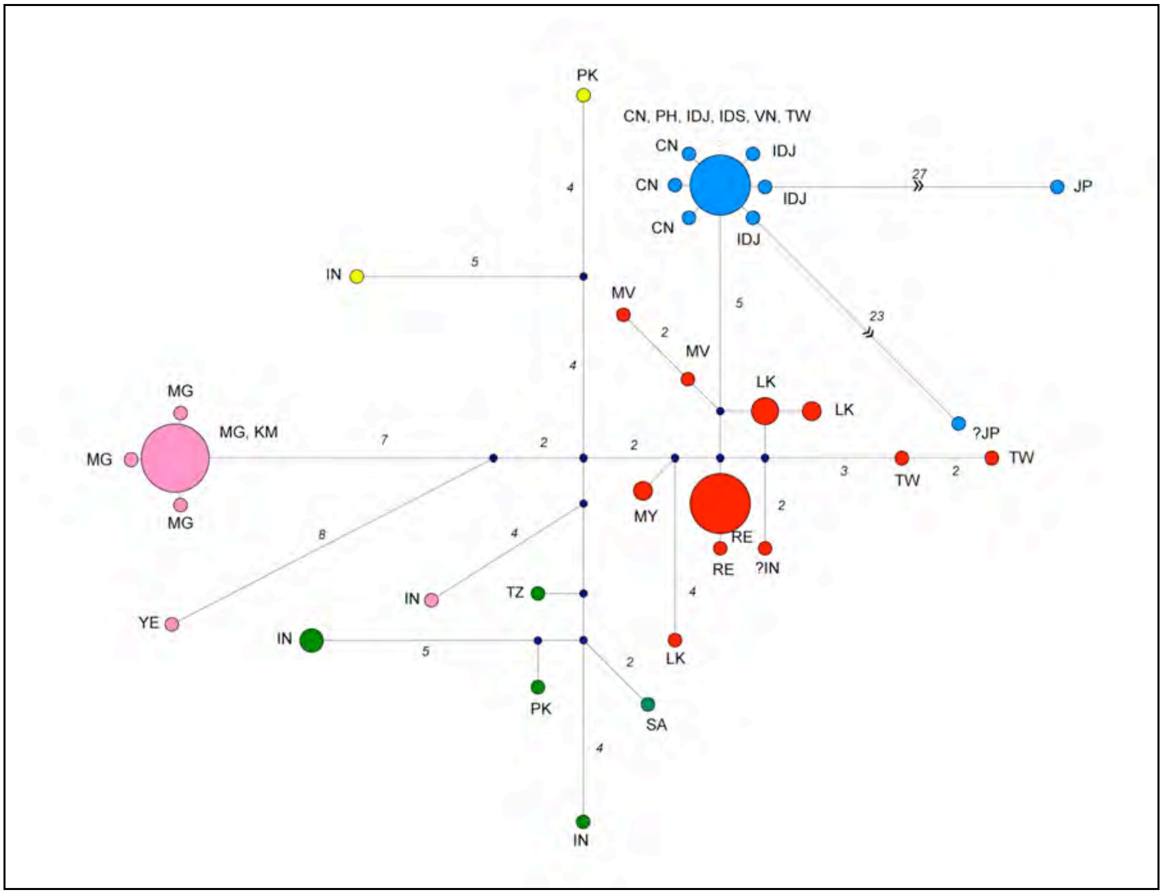


Figure 3.6. Median joining haplotype network for Clade C. Subclades are coloured (as in Figure 3.3): I yellow; II pink; III green; IV blue; V red. Circles represent haplotypes and area is proportional to number of sequences. Black dots are median vectors. Number of substitutions (>1) are given along the branch (branch lengths shortened for Clade C_{IV} in Japan as indicated by double arrow). ? are published sequences with inferred collection locations due to lack of collection information. Pakistan (PK), India (IN), Tanzania (TZ), Saudi Arabia (SA), Yemen (YE), Madagascar (MG), Comoros (KM), China (CN), Philippines (PH), Indonesia Java (IDJ), Indonesia Sulawesi (IDS), Vietnam (VN), Taiwan (TW), Japan (JP), Maldives (MV), Sri Lanka (LK), Malaysia (MY), Réunion (RE).

Accounting for damage in the phylogeny

Potential unidentified deamination in sequence data did not significantly affect the phylogeny (Figure 3.7). The tree accounting for damage in transitions (Δ damage model) differed slightly from the equivalent tree without added damage (0 damage model) but only in relation to bipartitions with low support (PP<0.50). In the Δ damage tree, C_{IV} is an outgroup to C_{II} /C_{III} and A North West is an outgroup to A Central/A North East, whereas in the 0 damage tree C_{III} is an outgroup to C_{II} /C_{IV} and A North East is an outgroup to A Central/A North West. The Δ damage tree has marginally reduced prior probabilities for two clades compared to the 0 damage tree: C_{II} $PP_{0\ damage}$=0.87 and $PP_{\Delta\ damage}$=0.85; A North East $PP_{0\ damage}$=0.97 and $PP_{\Delta\ damage}$=0.96. Between the two trees all sequences cluster within the same subclades, except for two individuals within Clade A ("171137 India" and "113033 Burma"). The two sequences are outgroups to sublineages within Clade A in the 0 damage tree but cluster within the "Mix" lineage (PP=0.49) in the Δ damage tree.

Comparison of the Δ damage/0 damage trees to the MrBayes tree shows that all clades and subclades are recovered and all sequences show the same clade affinity (except the two sequences previously mentioned in the 0 damage model, which are similar between the Δ damage tree and the MrBayes tree). However, once again subclade topology within Clades A and C differs between trees, which further emphasizes the lack of a robust relationship between subclades.

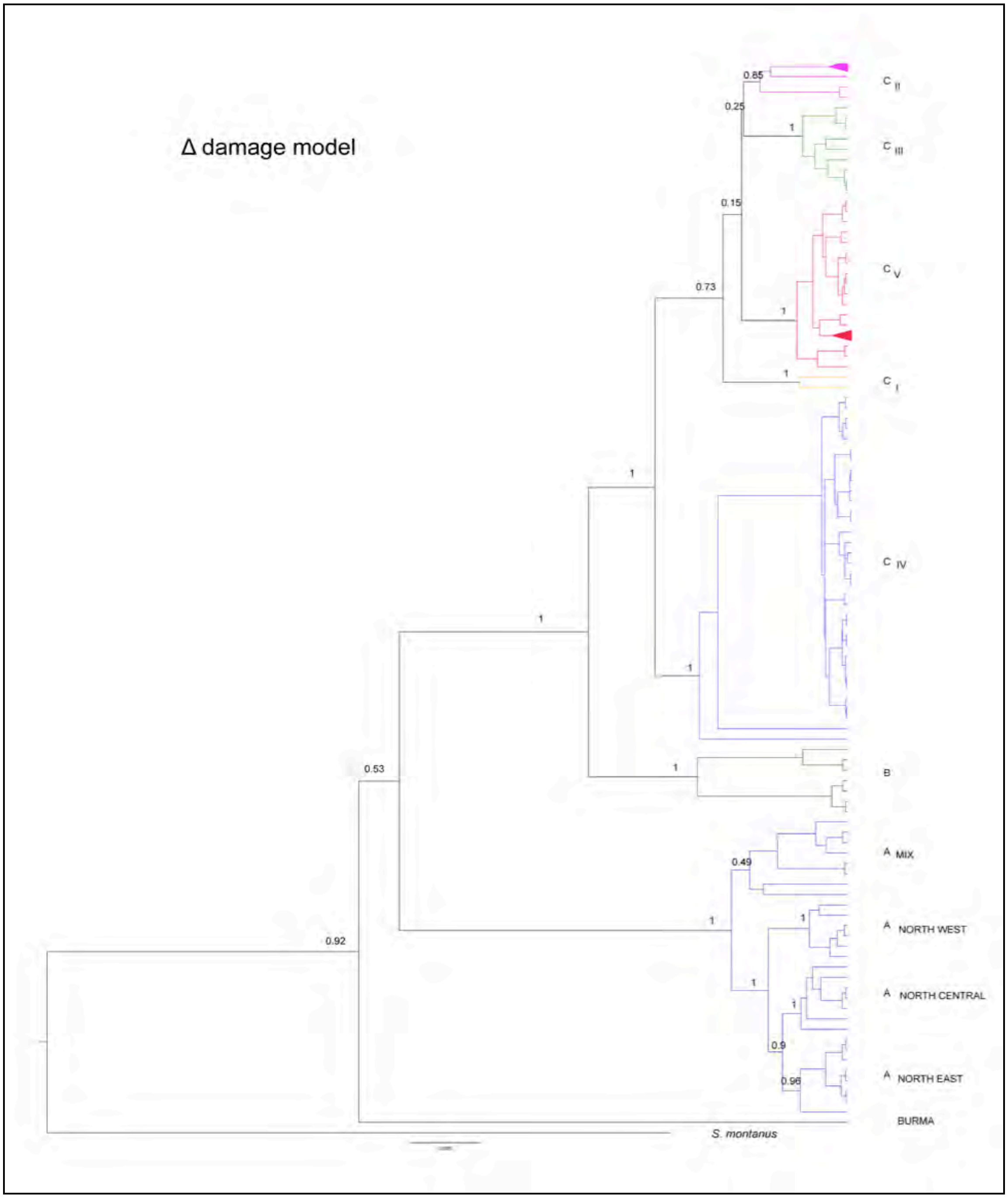


Figure 3.7 Bayesian MCMC analysis of *Suncus murinus* accounting for deamination (“Δ damage”) (above) and the same model without the Δ damage parameter (“0 damage”) for comparison (next page).

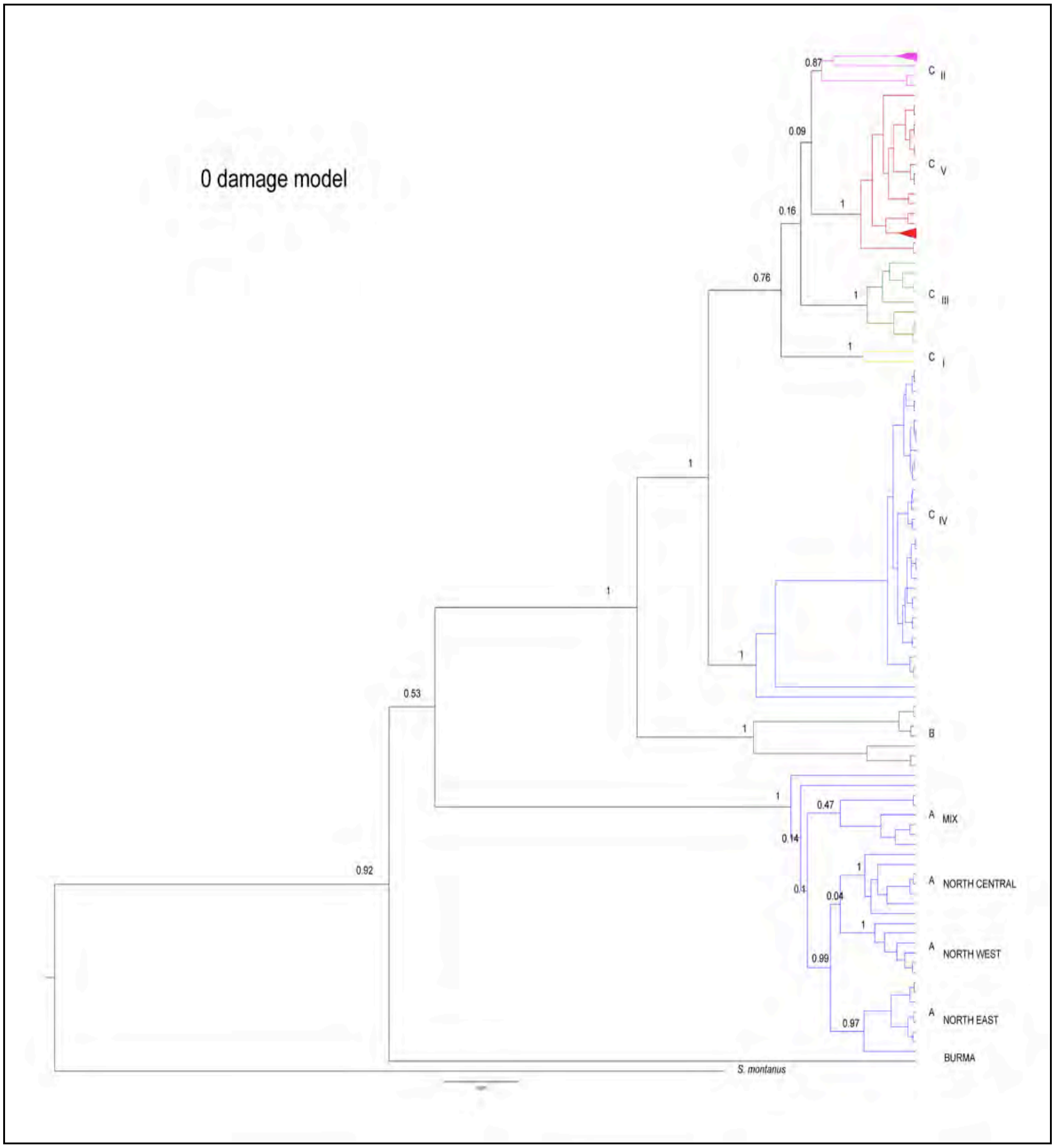


Figure 3.7 continued.

Geographic distribution

Within the Indian subcontinent clades A and C do not appear to have significantly overlapping populations and are probably parapatric (Figure 3.8), although further sampling in central India may reveal areas of sympatry. Clade A is restricted to the

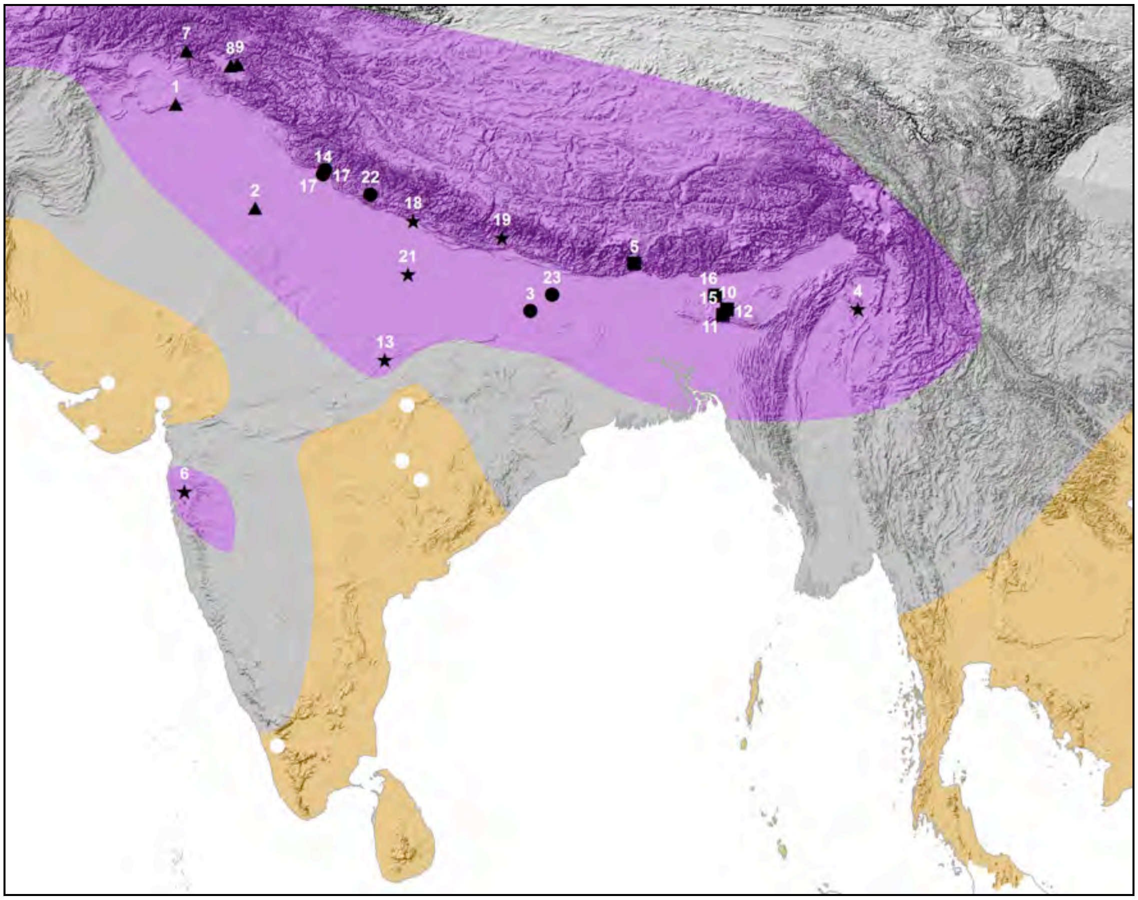


Figure 3.8. Interpolated ranges (nearest neighbour) for clade A (purple) and C (orange) within the subcontinent, and distribution of Clade A sequences by haplotypes (see also Fig. 5). Grey is where either lineage is equally likely. (Black) Triangle= North West, circle= North Central, square= North East , star= Mix. Numbers refer to haplotypes and correspond to sample IDs in Figure 3.5 and Appendix 3.I. 1. 3021186, 2. 3021153, 3. 150087, 4. 113033, 5. 35402, 6. 29767, 7. 326328, 8. 28632, 9. 27643, 10. 76178 and 76173, 11. 76177 and 76174, 12. 76176, 13. 171137, 14. 171136, 15. 76171, 16. 76166 and 76159, 17. 65399 and 65398, 18. 94133, 19. 94132, 20. 65397, 21. 54539, 22. 82521, 23. 82516. White dots are locations of haplotypes of C clade.

northern Indian subcontinent, except for one sequence from southern India. Sequence distribution within Clade A is shown in Figure 3.8 and is labelled by haplotype in reference to the MJ network (Figure 3.5). Subclades are geographically patterned in the subcontinent and three are highly supported, and are labelled according to their geographic distribution 'North East', 'North Central' and 'North West'. A further lineage 'Mix' with lower support is found in the central, eastern and southern subcontinent; some sequences within this clade are found basally within the A clade in the ML phylogeny.

Clade C is widespread and consists of five phylogeographic subunits (labelled C_I, C_{II}, C_{III}, C_{IV} and C_V) that are highly supported in the Bayesian phylogeny (PP=0.98-1.0), except C_{II} (PP=0.85) (Figure 3.3 and 3.9). Clade C_I contains 2 (2) individuals from India and Pakistan (the number in brackets represents the number of individuals in the dataset with no missing nucleotides); C_{II} 37 (30) individuals from Madagascar, Comoros, Yemen and India; C_{III} 9 (7) individuals from Kenya, Tanzania, Saudi Arabia, Pakistan and India; C_{IV} 33 (28) individuals from China, Vietnam, Indonesia (Java and Sulawesi), Philippines, Japan, Taiwan; and C_V 38 (35) individuals from Sri Lanka, Taiwan, Japan, India, Réunion, Malaysia, Maldives. All groups are found in India except C_V. The groups can roughly be described by their introduced distributions (Table 3.5): C_I Western South Asian; C_{II} Madagascar-Yemen; C_{III} African-Arabian; C_{IV} East and Southeast Asian; and C_V Indian Ocean and northern Southeast Asia (Table 3.5).

Clade	Distribution	Countries
C_I	Western South Asian	India, Pakistan
C_{II}	Madagascar-Yemen	Madagascar, Comoros, Yemen, India
C_{III}	African-Arabian	Kenya, Tanzania, Saudi Arabia, Pakistan, India
C_{IV}	East & Southeast Asian	China, Vietnam, Indonesia (Java and Sulawesi), Philippines, Japan, Taiwan
C_V	Indian Ocean & northern Southeast Asia	Sri Lanka, Taiwan, Japan, India, Réunion, Malaysia, Maldives

Table 3.5 Summary of distributions of clade C subgroups.

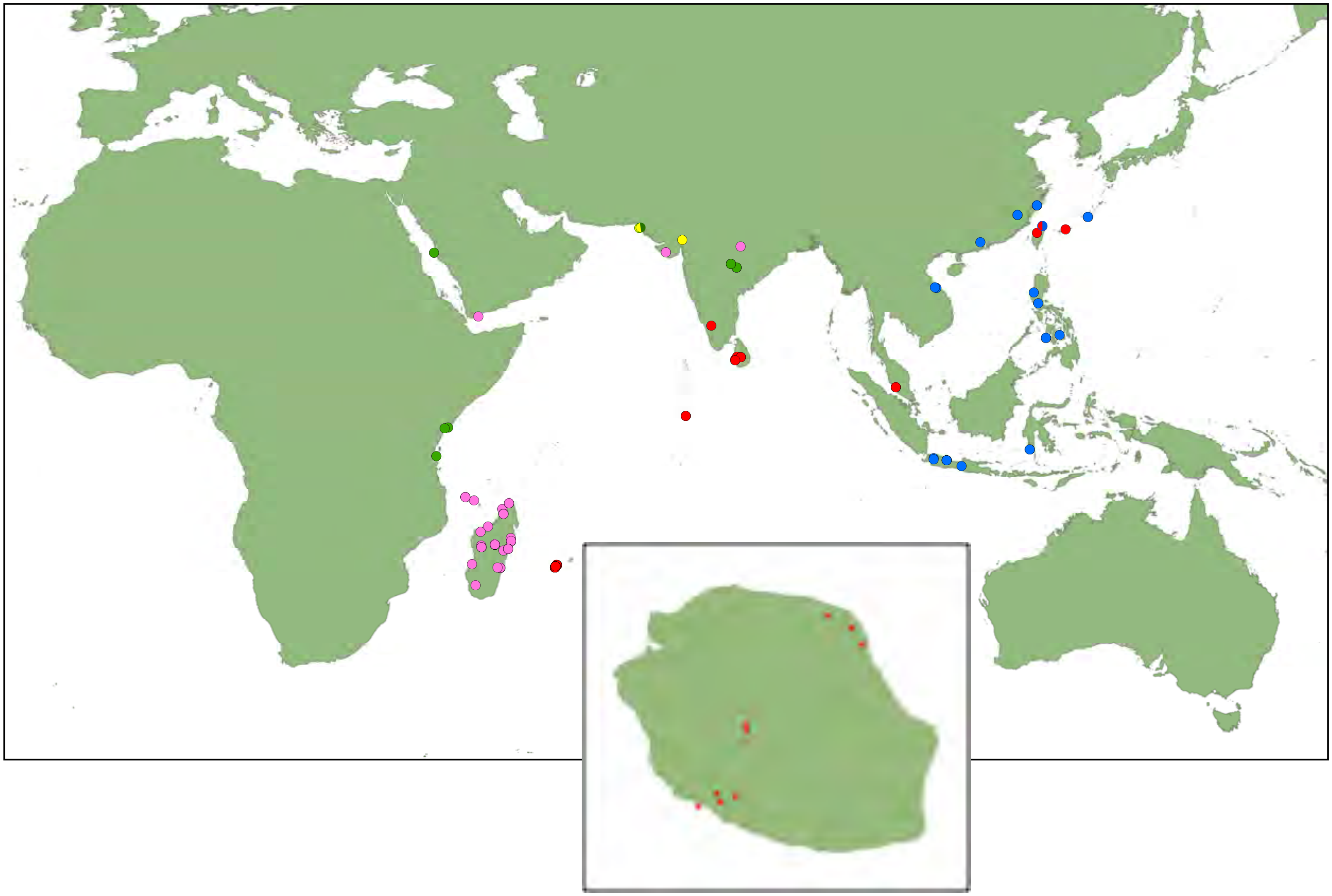


Figure 3.9. Distribution map of sequences in subclades within Clade C. C_I =yellow, C_{II} =pink, C_{III} =green, C_{IV} =blue, and C_V =red. Réunion (inset). Clade C_I India and Pakistan; C_{II} Madagascar, Comoros, Yemen and India; C_{III} Kenya, Tanzania, Saudi Arabia, Pakistan and India; C_{IV} China, Vietnam, Indonesia (Java and Sulawesi), Philippines, Japan, Taiwan; and C_V Sri Lanka, Taiwan, Japan, India, Réunion, Malaysia, Maldives.

Population genetic diversity and demography

Genetic diversity and neutrality statistics for all clades (where there were enough sequences) and populations are given in Table 3.3.

Clades

Clade A has higher nucleotide and haplotype diversity than Clade C. Neutrality statistics tend to indicate that both clades have undergone recent expansion (although in the case of Clade A, Tajima's D is not significant and in Clade C Fu's F_S is not significant). Although not a perfect fit, the mismatch distribution curve for Clade A supports an expansion event at some time in the deeper past (Figure 3.10). Mismatch distributions for B and C do not show evidence of expansion. Rather, the multimodal distribution for B indicates substructure in this subclade, which is further evidenced by the high support for bipartitions of sequences in Bayesian MCMC trees (Appendix 3.II).

Subclades

According to pairwise distances, subclade C_V is more closely related to every other subclade (0.8-1.2%) than any other subclade is to another (1.4-1.7%) (Table 3.6).

	V	IV	III	II
IV	0.8			
III	1.1	1.6		
II	1.2	1.6	1.5	
I	1.1	1.7	1.4	1.6

Table 3.6. Percent pairwise uncorrected distance between subclades within Clade C.

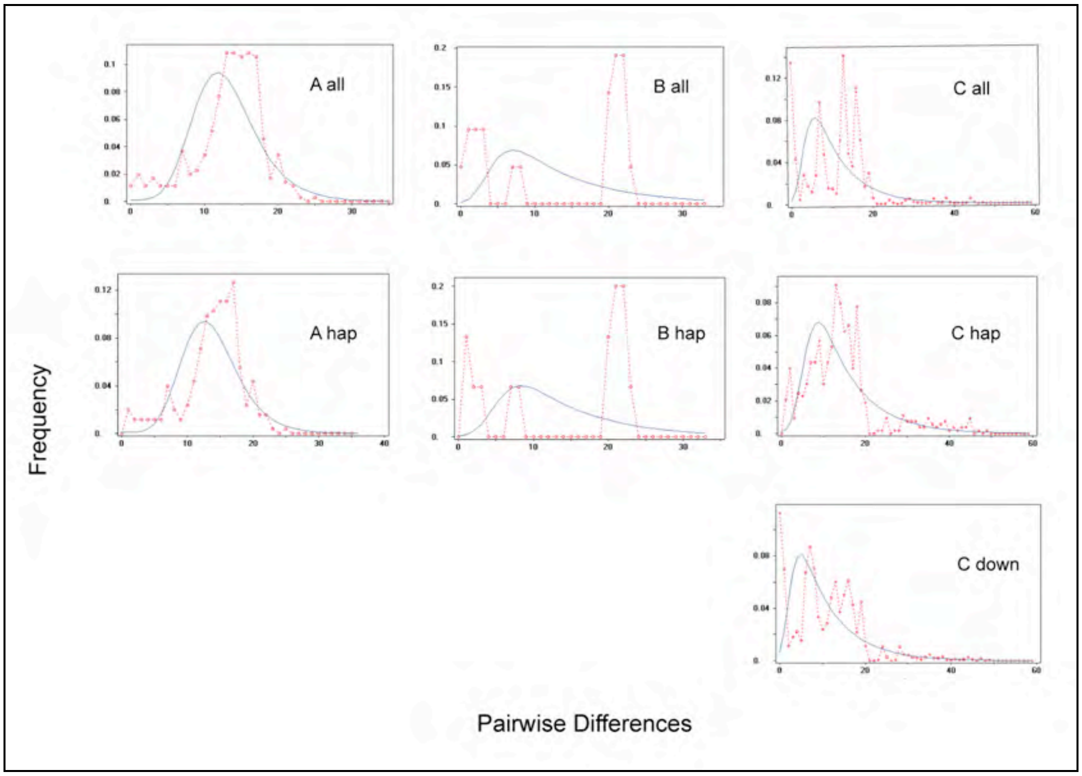


Figure 3.10. Mismatch distribution plots for Clade A, B and C (left to right) for all sequences (top row), haplotypes (second row), and Clade C downsampled (bottom right). The blue line is the expected distribution under a growth and decline model and the red line the observed distribution. Top row: A all (N=27, θ_0=2.9, θ_1=1000, τ=10.30), B all (N=7, θ_0=8.7, θ_1=1000, τ=4.73), C all (N=102, θ_0=6.95, θ_1=1000, τ=3.81). Second row: A hap (N=23, θ_0=2.65, θ_1=1000, τ=11.21), B hap (N=6, θ_0=8.36, θ_1=1000, τ=5.90), C hap (N=33, θ_0=8.27, θ_1=1000, τ=6.24). Bottom right: C down (N=62, θ_0=7.58, θ_1=1000, τ=3.0).

The MJ network shows the two anomalous Japanese sequences are related to two haplotypes, found in wild Javanese individuals, although this could be due to homoplasy in the sequences (Figure 3.6). Inbred lines from wild caught stock populations on Java are also common in the laboratory. Unsurprisingly, removal of the sequences reduces haplotype diversity, nucleotide diversity and the population mutation rate (Table 3.3). The reduced dataset also shows an increase in the strength and significance of population expansion/neutrality tests and the effect on mismatch distribution is particularly striking (Figure 3.11). Removal of the aberrant sequences (which show as a double peak followed by a lower frequency single peak at high values of pairwise distances in the complete dataset) causes the plot to conform well to a sudden expansion model, where previously it showed substructure.

The patterns of population genetic diversity within C subclades is varied (Table 3.3, Figure 3.11). In relation to Clade C_{IV}, the dataset without the anomalous sequences will be discussed. All subclades show moderate levels of haplotype diversity and moderate to low levels of nucleotide diversity, with C_{V} showing the greatest diversity overall. The mismatch distribution for C_{II} shows a multimodal distribution (clearly three peaks in the haplotype plot) and indicates substructure within the subclade consistent with the MJ haplotype network (Figure 3.6). Conversely, significant D and R_2 statistics indicate a recent population expansion in this subclade. C_{III} also shows structure in mismatch distributions with three, possibly four groups, although this is less than we might expect from MJ haplotype networks. C_{IV} shows a clear signal of a recent population bottleneck followed by rapid expansion (or selective sweep) in L-shaped mismatch distributions and this is supported by all neutrality and expansion tests, which are significant and, in the case of D and Fu's F_S, highly negative. The L-shaped mismatch distribution also attests to the low genetic diversity in this clade. Mismatch distributions calculated for the haplotype and downsampled datasets in C_{V} show clear signals of sudden expansion, particularly in the downsampled plot, although the plot calculated using all sequences does not conform. There is a shift in the crest of the wave to the right as the dataset is reduced, with the implications being a comparatively older expansion event in the downsampled/haplotypes datasets than the dataset compiled using all sequences. In this case, all datasets indicate a recent population expansion that probably occurred a significant length of time before that of C_{IV} but the age of the expansion is feasibly more akin to that suggested by the downsampled dataset (given it reduces bias introduced through oversampling but allows for more accurate measures of N_e than haplotype

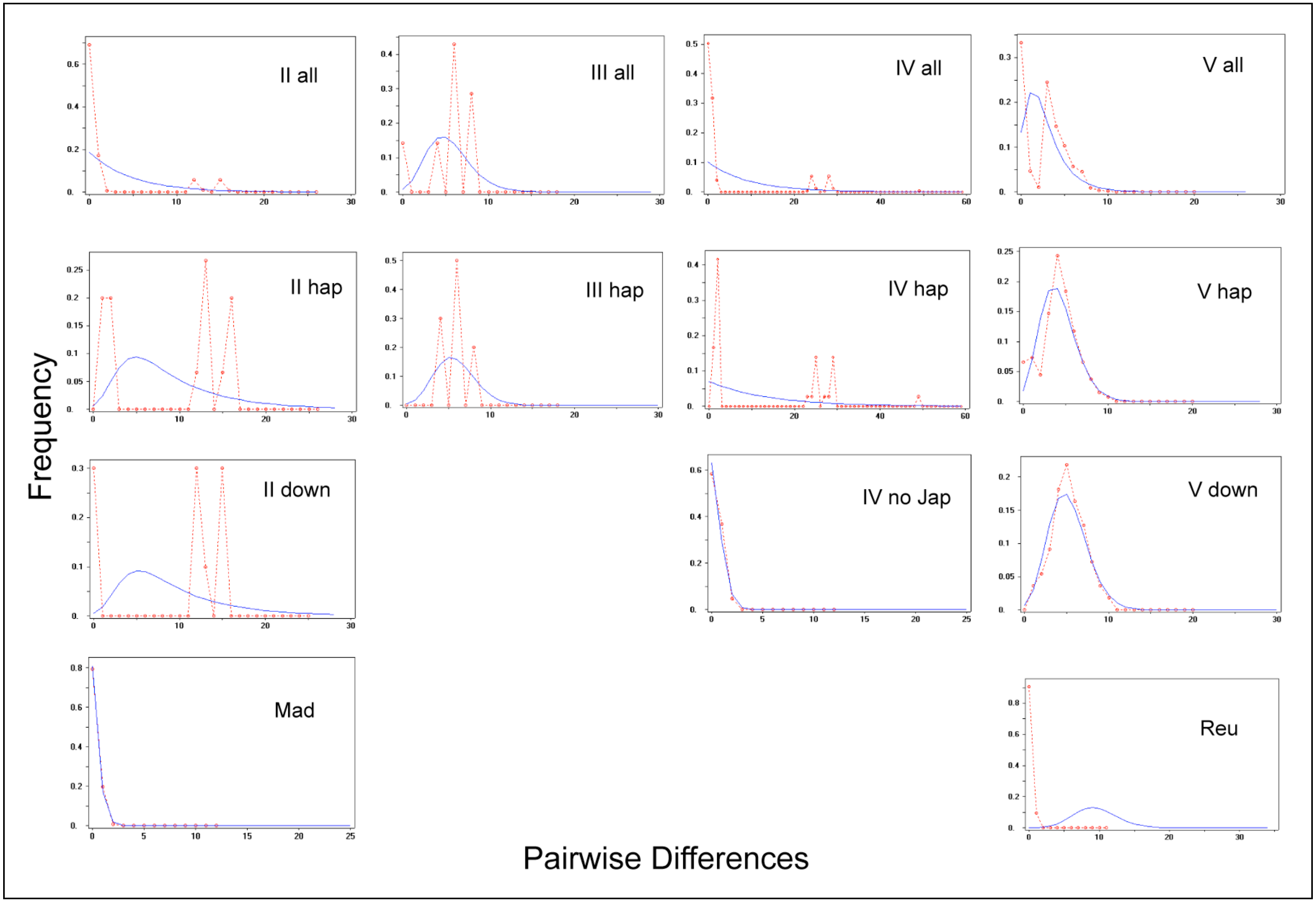

Frequency
Pairwise Differences
II all
III all
IV all
V all
II hap
III hap
IV hap
V hap
II down
IV no Jap
V down
Mad
Reu

Figure 3.11. Mismatch distribution plots for C subclades, Madagascar & Comoros and Réunion. The blue line is the expected distribution under a growth and decline model and the red line the observed distribution. Left to right C_{II} (Madagascar/Comoros bottom), C_{III}, C_{IV}, C_{V} (Réunion bottom). Top row: all sequences; second row: haplotypes; third row: C_{II} downsampled, C_{IV} with anomalous Japanese sequences removed, and C_{V} downsampled; and bottom row: Madagascar & Comoros and Réunion (left to right). Top row: C_{II} all (N=30, θ_0=4.351, θ_1=1000, τ=0), C_{III} all (N=7, θ_0=1.195, θ_1=1000, τ=4.233), C_{IV} all (N=28, θ_0=8.978 θ_1=1000, τ=0), C_{V} all (N=35, θ_0=1.622, θ_1=1000, τ=1.053). Second row: C_{II} hap (N=6, θ_0=5.78, θ_1=1000, τ=3.28), C_{III} hap (N=5, θ_0=0, θ_1=1000, τ=5.8), C_{IV} hap (N=9, θ_0=13.28, θ_1=1000, τ=0) ,C_{V} hap (N=11, θ_0=0, θ_1=1000, τ=5.2). Third row: C_{II} down (N=5, θ_0=5.849, θ_1=1000, τ=3.522), C_{IV} no Jap (N=26, θ_0=0, θ_1=1000, τ=0.462), C_{V} down (N=17, θ_0=0.6, θ_1=1000, τ=3.562). Bottom row: Madagascar & Comoros (N=28, θ_0=0, θ_1=1000, τ=0.214), Réunion (N=21, θ_0=0, θ_1=1000, τ=9.523).

data). While not significant, D and F_S for this subclade are negative and a significant R_2 value supports the idea of a recent population expansion.

The observed mismatch distributions in some populations in my data showed near identity to expansion curves, while in others there was no fit, rather multimodal peaks were present. Clades B and C do not conform to expected curves under a population expansion model no matter which dataset is used. I did not investigate B further due to the small number of sequences, however, analysing Clade C at a finer scale teases apart the unique population expansion distributions for each subclade. By comparing mismatch distributions for my data at different levels (major clades versus subclades) and compiled under different data approaches (all sequences, haplotypes and, where appropriate, downsampled datasets) I can identify where there is true substructure in the data, or even which sequences are influencing the pattern.

Geographic populations

A population bottleneck followed by expansion in recent history in the Madagascar/ Comoron population is evident from L-shaped mismatch distributions, low genetic diversity, and significantly negative D and Fu's F_S values. In this case a non-significant R_2 value does not conform to the overall pattern, although it is a relatively low value, indicative of recent expansion. There is one common haplotype on Madagascar, which is distributed throughout the island and three unique haplotypes that are located on the northwest and central west part of the island (Figure 3.12).

The observed pattern for the mismatch distribution on Réunion is indicative of a recent population bottleneck but the population shows no significant expansion trend. Although negative, D and Fu's F_S were not significant, and neither was R_2.

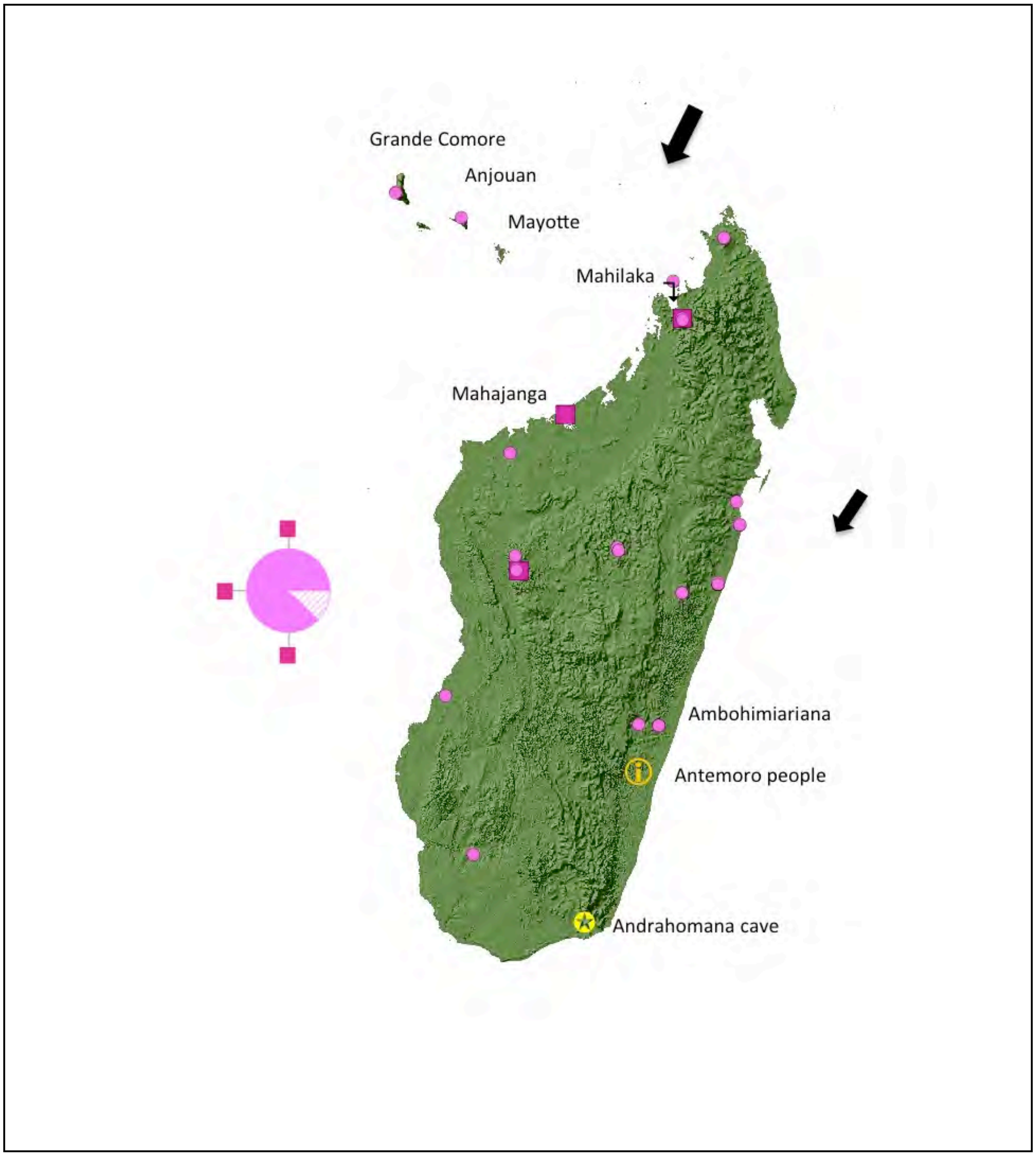


Figure 3.12. Map of Madagascar and the Comoro Islands with sites discussed in the text, and location and relationship of haplotypes (proportion of sequences within the central haplotype for Comoros is indicated by hatching and for Madagascar by solid colour). The squares on the median-joining network and map indicate unique haplotypes. Large arrows are suggested introduction routes.

Sri Lanka hosts multiple species and/or clades (Figure 3.13). Both *S. montanus* (a Sri Lankan endemic) and *S. murinus* are found on the island, although *S. montanus* is mostly restricted to the highlands. Two clades within *S. murinus*, clades B and C, also co-occur on the island. Under current taxonomic classifications, clade B is described as the subspecies *Suncus murinus murinus* and clade C as the subspecies *Suncus murinus caerulescens* (Meegaskambura *et al.* 2010).

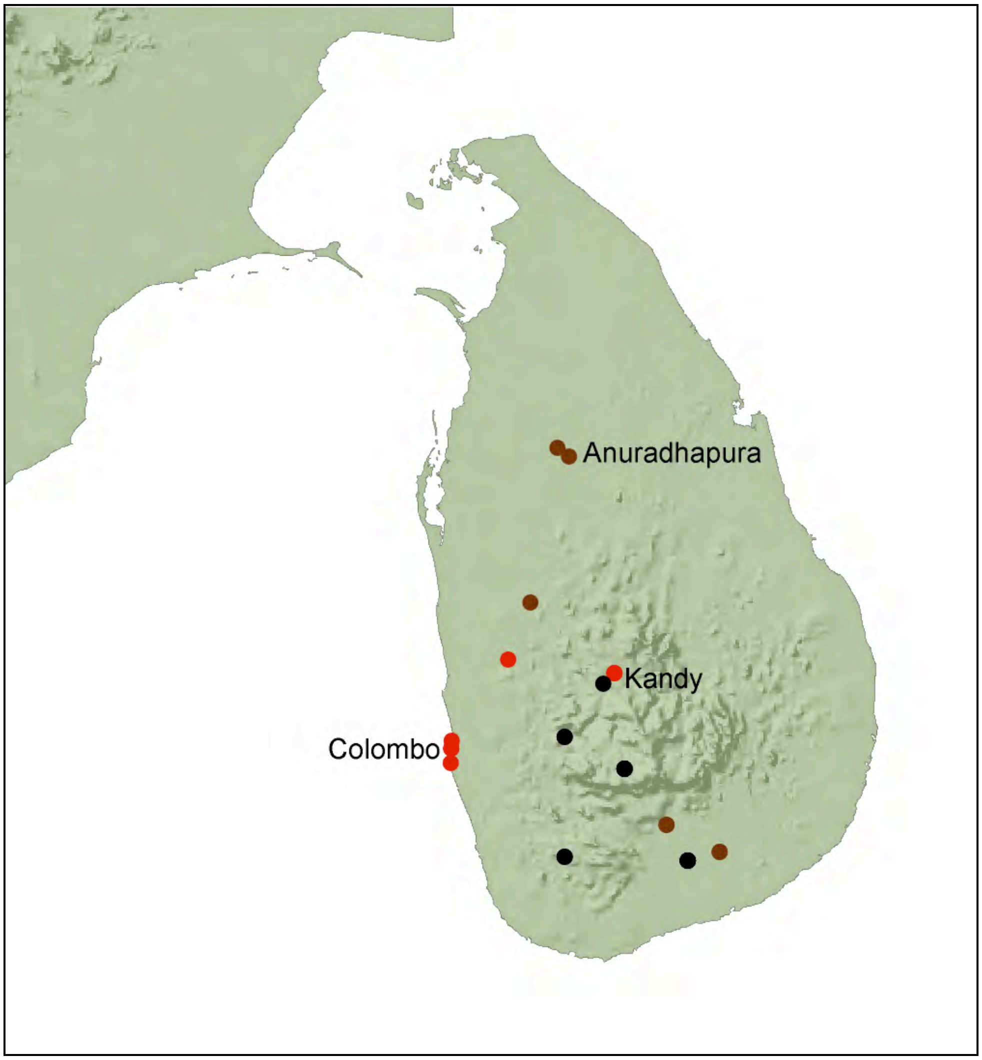


Figure 3.13. Map of Sri Lanka with location of sequences used here, coloured according to their identity clade B (brown), clade C (red) and *Suncus montanus* (black).

DISCUSSION

Phylogenetic structure

The monophyly of major lineages is robust, as is the affinity of the majority of individuals within mostly well-supported groups. Phylogenetic trees and MJ networks clearly show evolutionarily distinct lineages at multiple levels (Clade A, B and C, and within clade). However, the relationship between groups at the subclade level is not well-supported, and thus variable. This is apparent in the unstable topology of phylogenetic trees around bipartitions with low support (compare MrBayes, ML, the Δ damage, and 0 damage trees), and the connection of subclades or haplogroups through median vectors in MJ networks. The choice of included taxa or outgroups within a dataset can cause a change in tree topology due to long branch attraction. This is of particular relevance to rapidly evolving lineages where high mutation rates increase the probability of shared homoplasies between taxa or between a taxon and an outgroup. The result is an inferred relationship that is not based on a true synapomorphy and thus the correct relationship may not be recovered. The house shrew potentially has high mutation rates for the mitochondrial genome compared to other mammalian species as do other species of shrew; the greater white-toothed shrew *Crocidura russula* for example, has a rate of 1.59×10^{-8} substitutions/site/year compared to the mammalian average of $1.106 \pm 0.189 \times 10^{-8}$ substitutions/site/year (Lawson Handley *et al.* 2006). It is therefore not prudent to discuss the relationship between subclades or subpopulations in *S. murinus* to a great degree.

A similar substructure is found at the intra-subclade level. The relationship between sub-populations within some groups is variable for the same reasons described above (compare C_V between MrBayes, ML, Δ damage and 0 damage trees). The topology and bipartition support of terminal branches will also be influenced by the proportion of missing data in the sequences but the substructure remained true when MJ networks were calculated on the dataset that did not contain missing nucleotides (as shown by median vectors between sub-populations within subclades).

Is this substructure an effect of sampling strategy, i.e. unsampled haplotypes? As a general strategy, I favoured a shallow, broad approach to sampling in order to maximise geographic coverage in the Indian and Pacific Ocean regions. However, I also sampled heavily in, or near, the putative origin of the species - the northern Indian subcontinent. Sequences from the northern subcontinent proved to be a monophyletic clade (Clade A) within which were divergent subclades. The same pattern was seen for the rest of the distribution (Clade C). While adding taxa may resolve some relationships to some degree, particularly in more recent lineages, the overall pattern of multiple, highly divergent lineages is probably an authentic evolutionary signal.

The house shrew is a highly adaptable, rapidly evolving species with a high reproductive rate - factors that contribute to its success as an invasive species. Like any species, house shrew populations are subject to factors such as disease, habitat and climate change, resource scarcity and competition from other species that cause demographic changes in an established population. Although a successful invader, wild house shrew populations are known to have gone extinct in recent history (Jogahara *et al.* 2008). Further, as they are considered a pest by humans they are often actively extirpated (Varnham *et al.* 2004). The scenario by which the phylogenetic structure in house shrews evolved can best be described as a recurrent pattern of introduction of an individual/population from a source with founder effects limiting diversity in the colonising population, followed by rapid expansion in some cases, and population bottlenecks and extinctions in both source and introduced populations. Consistent with my data, the trace left by the pattern of founder effects and bottlenecks/extinctions will be the accumulation of differences between two populations and the loss of deeper evolutionary connections.

Apart from a slight reduction in the PP of some clades there were no substantial changes caused by the addition of the Δ damage parameter to the tree model, *per se*. I realistically set the upper bound on the model prior as <1 as I did not expect large amounts of error in my sequences due to molecular controls. However, even within these bounds I can expect that the addition of substitutions (for transitions) along the terminal branches might weaken support of bipartitions. The recovery of the same clades, subclades and affinity of most individuals within subclades attests to the robust nature of the tree overall.

Mismatch distribution and data sampling

Mismatch distributions (distributions of pairwise differences) are used to estimate demographic parameters of population expansion, such as the size of the starting population, the rate of population expansion, and the timing of the expansion event (Rogers & Harpending 1992). In order to understand the effect that sampling bias and data preparation has on my data I analysed populations under various sampling conditions that account for frequency of sequence similarity in different ways (all sequences, haplotypes, and downsampled). Estimates of θ, a factor used to model mismatch distributions, will be affected by the frequency of sequences in the dataset and their identity to one another. Density of sampling in my study was higher in Madagascar/Comoros (N=25) and Réunion (N=21) (sampling density given for the dataset without missing nucleotides) than other locations, and low genetic diversity was a feature of these islands.

A number of trends can be identified in the observed and expected distribution curves of mismatch distributions as the data is reduced from the full dataset using all sequences, to downsampling oversampled areas, and then reducing to haplotypes (where no two sequences will be identical).

Observed distribution

As all sequences are downsampled the frequency of identity of oversampled sequences also reduces, and in haplotype datasets identity=0. The matrix of pairwise differences between each individual and all others is also affected as samples become more dissimilar to each other, especially in the haplotype dataset. The observed frequencies of pairwise differences are thus affected in two ways:

1. Identical sequences cause a high frequency peak at 0 pairwise differences, as sequences are downsampled the height of the peak reduces, and for haplotypes there is no peak as frequency of identity=0.

2. The proportion of pairwise distances >1 will increase overall as sequence diversity increases.

Expected distribution

When there is a high frequency of low pairwise differences in the observed data (many identical or near identical sequences), the reduction of the dataset influences the expected distribution curve in two ways:

1. There is a reduction in the vertical intercept (where initially, intercept>0), implying a comparatively larger increase in population size.

2. The crest of the wave moves to the right along the x-axis. The crest of the wave is determined by τ (which is affected by pairwise diversity of sequences). In the case of C_{IV}, the anomalous highly diverse sequences (extremely high μ relative to other sequences) swamp this effect, and consideration of outliers should therefore be taken into account.

Overall, where substructure exists in the data, a poor fit of the observed data to the expected curve will follow no matter which dataset is used. As the scale of analysis becomes finer (i.e. at the subclade or population level) the choice of data compilation has substantial effects on the observed and expected distributions. These two conclusions are of particular importance for invasive species. Recurrent population bottleneck and expansion events seen in invasive species (at deeper levels) will follow multiple trajectories that are obscured in a single trajectory model (such as a mismatch distribution model), and the implications drawn from mismatch distributions (at shallower levels) are affected by data compilation and may not necessarily reflect the true evolutionary signal.

Taxonomy

The Asian house shrew shares overlapping ranges with two closely related species that are smaller and darker in form, the Sri Lankan Highland shrew *S. montanus* in Sri Lanka and the Nilgiri Highland shrew *S. niger* in India. A recent study has shown partial introgression through hybridization of *S. montanus* DNA into *S. murinus* individuals in the Sri Lankan Highlands (Meegaskumbura *et al.* 2010). Here, I identify potential hybrids (or introgression from an ancient hybridization event) of *S. niger* and *S. murinus* in southern India. The three individuals mitochondrially clustering with *S. niger* in my data were collected from two different museums and are probably correctly identified as *S. murinus* based on morphology - *S. murinus* being larger and paler than *S. niger*. Asian house shrews are highly variable in size ranging between 11.3 - 14.6 cm in females and 13.5 - 16 cm in males, with males always being larger within a population (Kurachi *et al.* 2007a). Mating incompatibilities between small males and large females have experimentally been shown in a laboratory population of *S. murinus,* yet mating the reverse size forms produced no such incompatibility (Ishikawa *et al.* 1991). As with crosses in *S. murinus*, there appears to be no physical reproductive isolation barrier between small *S. montanus* or *S. niger* females and large *S. murinus* males. Analysis of nuclear DNA is needed to confirm my hypothesis.

My data suggest a mitochondrial contribution from *S. niger* in *S. murinus* individuals from southern India (Nilgiri Hills and Karnataka). In contrast, no nuclear gene flow in sympatric populations of *S. niger* (under their former name of *S. montanus niger*) and

S.murinus in the Nilgiri Hills was found in a study by Ruedi *et al.* (1996). Sri Lankan *S. montanus* x *S. murinus* were *S. montanus* in their mitochondrial DNA (mtDNA) and *S. murinus* in their nuclear DNA (Meegaskumbura *et al.* 2010). Although it is rather speculative given the limited number of sequences and lack of genetic studies, gene flow in *S. montanus* x *S. murinus* and *S. niger* x *S. murinus* crosses may be unidirectional with *S. montanus* and *S. niger* contributing mtDNA to hybrid offspring or backcrossed lines.

MtDNA protein electrophoresis, RFLP and D-loop sequence data identified two main groups of *S. murinus* - the “continental” and “islands” types, with a further contact group in Malaysia (Yamagata *et al.* 1995; Yamagata & Namikawa 1999). These probably correlate to my Clades A and C_{IV}, respectively, with my Malay individuals clustering in Clade C_V. Pairwise genetic distances between groups show B and C are more closely related to each other than Clade A, which is equally distant from both. Further, *S. montanus* is more closely related to B and Clade C than A.

The high genetic distance between Clade A and both *S. montanus* and clades B and C (Table 3.4, Figure 3.4), along with its relatively high haplotype and nucleotide diversity (clade A $H_d = 0.989 \pm 0.013$; $\pi = 0.0118 \pm 0.0007$) suggest that the clade has been evolutionarily isolated for a substantial length of time. Interpolated ranges within the subcontinent for Clades A and C are non-overlapping (Clade A is restricted to the north, except for one individual in the south) and the mode of divergence was probably parapatric between these lineages. Given the wide distribution of the species, there are presumably contact zones, or even hybrid zones, in central and southern India (although there is no data at present) (Figure 8). Clade A has further phylogeographic structure with ‘North Western’, ‘North Central’, ‘North Eastern’, and a possible fourth ‘Mixed’ population, which probably evolved from a rapid population expansion sometime in the more distant past, as indicated by a right shift in the mismatch distribution curve (Table 3.3, Figure 3.10). Given their distribution in the Himalayas, which are a barrier to natural colonization, it is feasible that the house shrew's rapid colonization of the northern subcontinent is linked to migration with humans in its commensal form. Today, *S. murinus* is the commonest shrew species in the Himalayas (of fifteen species in total) inhabiting most regions of the mountain chain and covering a broad altitudinal range (61-2286 m) (Jenkins 2013). Median vectors in the MJ haplotype network for Clade A can be mostly interpreted as extinct ancestral haplotypes, and therefore the relationship among these subclades is uncertain.

The house shrew is a highly polymorphic species with size and colour being the two most obvious differences. Species are generally described based on morphological characteristics, however, such classifications are difficult for taxa such as the house shrew that show high levels of polymorphism which are not correlated genetically (Kitchener *et al.* 1994). Similarly, highly monomorphic species can exhibit high levels of genetic variation and thus be a 'cryptic' taxon consisting of more than species. Using a multi-locus, coalescent-based approach, Pauperio *et al.* (2012) found that three morphologically similar European populations of the field vole *Microtus agrestis* showed genetic divergence at levels considered to be at the cusp of speciation. In their study, the cytochrome *b* gene showed a particularly strong signal with Dxy (average raw DNA divergence) ranging between 3.49 – 6.59 % for pairwise distances between the three populations. Pairwise distances based on the cytochrome *b* gene have been used to certify species level divergences in other mammal studies, with most sister species differing by between 4-11% (Bradley & Baker 2001). Divergence between clade A and each of clades B and C for Asian house shrews is 5.1%, which places clade A in the spectrum of species level divergence. Further genetic, morphological, and breeding studies are warranted in Clade A populations to establish their status.

Clade B is restricted to Sri Lanka. The mismatch distribution plot shows substructure consistent with the phylogenetic tree, although sample size is limited. Clade B (*Suncus murinus murinus*) co-occur with a larger lighter form, clade C_V (*Suncus murinus* *caerulescens*) in Sri Lankan towns and sea ports and they readily interbreed. Based on their low genetic distance (2.5%) Meegaskambura *et al.* (2010) suggest they are a single taxon.

Clade C is a relatively diverse, widespread clade ($H_d = 0.866 \pm 0.020$; $\pi = 0.00954 \pm 0.00509$). Expansion throughout the Indian and Pacific Oceans probably occurred in recent history, as neutrality statistics tend to indicate (Table 3.3), possibly in the last few millennia. Substructure within the clade obscures any signal of expansion in mismatch distributions (Figure 3.10) but these signals are recovered for some subclades when analysed separately (Figure 3.11). Clade C is discussed in more detail in relation to colonisation histories.

Based on mitochondrial DNA RFLP and autosomal blood protein/enzyme variation, Kurachi *et al.* (2007a; b) identify two distinct population groups in Burma, plus a probable hybrid of the two. They identify a South Asian type in the north and west of the country (the “continental” type) and a Southeast Asian type similar to ISEA populations in the south and east of the country (the “island” type), and an individual from central Burma having mtDNA of the South Asian variety and a nuclear profile of the Southeast Asian variety. My south-central Burmese singleton is interesting in that its mitochondrial DNA relationship to either clade A or B/C is not well-defined - in the MrBayes tree it falls between Clade A and B/C but in the damage model trees it falls as an outgroup to both major lineages. It is possible that the singleton is an intermediate form (between A and B/C), or may represent as yet unidentified diversity within the mitochondrial genome.

Origin and dispersal

Population genetic indices were calculated for *S. murinus* individuals within India and Sri Lanka to identify an origin. Haplotype and nucleotide diversity was higher in the Indian population (Table 3.3). My results support previous hypotheses that the Indian subcontinent is the origin of the species (Yosida 1982; Yamagata *et al.* 1995; Kurachi *et al.* 2007b).

Yosida (1982) proposed that the standard 2n=40 karyotype originated in central India and migrated east to Bangladesh and south to southern India, where it underwent a chromosomal reduction event and thence migrated to Sri Lanka. As previous researchers have suggested, the genetic diversity accumulated in the house shrew is probably a result of isolation and dispersal during periods of glaciation and warming (Yamagata *et al.* 1995; Kurachi *et al.* 2007a). Much of the extant diversity patterns are probably a result of range contractions around the Last Glacial Maximum (LGM) *ca.* 26, 000 - 13, 000 years BP. During this period the central subcontinent aridified and it is hypothesised that house shrews were maintained in isolated refugia in the south of the continent during this time (Kurachi *et al.* 2007a).

Colonisation histories

Judging by their coastal and island distributions, the introduction of the house shrew to areas outside its native range is largely attributable to passive transportation with humans in relatively recent history. To test this, I compare phylogeographic patterning in a mitochondrial gene in shrews with known human networks.

From this study and others we can conclude that in most of their introduced ranges, house shrews are remarkably homogenous for mtDNA type at the population level (Yamagata *et al.* 1995; Kurachi *et al.* 2007a). As mitochondria are maternally inherited, this suggests that female recolonisation of previously colonised areas is limited (although see clade C_V for an exception) and a similar pattern is found in house mice (Bonhomme & Searle 2012). Mitochondrial DNA phylogeographies of extant house shrews therefore probably reflect ancient (or historical) human-mediated introduction events and in this way the species acts as an excellent bioproxy for human movements.

Clade C type *S. murinus* has successfully colonised many islands and coastal areas in the Indian and Pacific Oceans. The house shrew's current distribution outside India was probably achieved mostly through passive transportation with humans and they can be considered successful ship-borne invasives in much the same way as the ship rat lineage of the black rat *Rattus rattus* (Aplin *et al.* 2011). Previous chromosomal, mitochondrial and nuclear analysis has so far focused on South and Southeast Asia (and Mauritius). Here I provide the first evidence of molecular differentiation in house shrews in the western Indian Ocean (outside Mauritius).

From the phylogeographic distribution of subclades, it is evident that there are a number of interaction spheres within the Indian and Pacific Oceans, at least as they relate to maternally-inherited markers describing shrew dispersal, and each is discussed below.

C_I Western South Asian

Two individuals from Sind (Pakistan) and Gujarat (India) cluster within this subclade. Clearly, further sampling is required to identify the extent of its distribution. Based on similarities between morphotypes, Hutterer & Harrison (1988) proposed that large, grey-brown shrews from Bahrain and Iraq were introduced from India (possibly the north of India). Although I had samples from Iraq and Oman in the Persian Gulf I was not able to amplify DNA from them.

C_{II} Madagascar-Yemen

The two Indian individuals from Gujarat and central India that are part of this, albeit poorly supported, subclade are closely related (213 bp of the central Indian sequence are missing, but where there is dual coverage the two sequences differ by only 1 bp). Their status in the phylogeny is weak: based on the MJ network (Figure 3.6) and pairwise differences between the Gujarati sequence (omitting the Central Indian sequence due to missing nucleotides) and representatives within subclades C_{II} and C_{III}

(uncorrected p-distance: Gujarat & C_{II} =1.1%; Gujarat & C_{III} =0.9%) the Indian sequences appear more related to C_{III}, however they are more closely related to C_{II} when a likelihood model that incorporates a model of evolution is used (see the MrBayes tree, Figure 3.3). It is likely this discrepancy is due to the different treatment of multiple mutations at a single site (the non-parsimonious state) between the network algorithm and the tree search.

Madagascar population (C_{II})

Madagascar and Comorian individuals are closely related, and these in turn are more distantly related to the individual from Yemen, although there is low support for this latter relationship - the source of these populations is unknown. High sequence diversity between Yemeni and Madagascar/Comorian haplogroups indicates the split from the ancestral population occurred in the relatively distant past.

In much of its introduced range, the house shrew is limited to urbanised areas around ports, cities and towns. In Madagascar, however, it is found throughout the island and is embedded in urbanised, disturbed, and natural environments, including forests and mountains up to about 1500 m (Soarimalala & Goodman 2011). The first securely dated house shrew described on the island was documented in 1858, however its wild establishment in natural habitats suggest a more ancient origin (Hutterer & Tranier 1990).

Mismatch distributions and tests of neutrality clearly indicate that Madagascar house shrews underwent a bottleneck upon introduction, followed by rapid expansion on the island. Madagascar shrews are strictly a monophyletic group with a common central haplotype found throughout the island and multiple haplotypes differing by a single base pair found on the northwest of the island (Figure 3.12). The core haplotype is identical between a museum sample collected 87 years ago and extant samples, and the extent of its distribution on the island suggest that the haplotype represents the first wave of colonisation, although an earlier introduction followed by replacement by the core haplotype cannot be ruled out. The common haplotype is the only haplotype found on Grande Comore and Anjouan. Higher diversity on Madagascar suggests it was the first island to be colonised in the region, although it is noted that sampling is significantly less dense on the Comoro Islands (N=3 Comoros; N=31 Madagascar).

Based on the presence of pale grey and dark grey morphotypes on Madagascar, Hutterer and Trainier (1990) proposed that house shrews were genetically diverse and imported from multiple sources. I found no evidence of multiple introductions via a maternally inherited pathway, although I had no information on either pelage colour or autosomal/Y-linked pathways with which to validate my hypothesis. A number of genetic studies have been conducted since Hutterer and Trainier published their morphological study on house shrews, and no significant correlation between geography or ecology and colour morphology has been found thus far (Ruedi *et al.* 1996).

Mitochondrial studies of both house mice and black rats *Rattus rattus* show kinship between Madagascar and Arabian populations (Duplantier *et al.* 2002; Tollenaere *et al.* 2010). Madagascar house mice show a signature of a single, recent introduction and a kinship with Yemeni mice that is very similar to that of the house shrew, and this is confirmed by Y-chromosomal data in mice. Similarly, rats are thought to be a recent introduction from a single source population, possibly in Oman. Rats from Mayotte are closely related to those from Madagascar and are expected to have colonised the island contemporaneously. Rats from Grande Comore, on the other hand, are distinct from populations east of it and are probably a separate introduction from a source population in Oman.

House shrews probably invaded Madagascar through a gateway in the northwest of the island, where genetic diversity is highest and the oldest populations are expected to occur. Unique house shrew haplotypes are found on the northwest coast near two historically important trading ports and these are the most plausible ports of entry for the species. The first, Mahilaka, was a large town, which rose to prominence around the 12th century AD and marked the start of high volume trade within the Indian Ocean network from where it garnered its Arab-Islamic influences. It was partially abandoned around 1400 AD. With the onset of Mahilaka's demise, other ports along the northeast and northwest coast replaced it, one of the largest being Mahajanga, a town founded by Indian traders in the 18th century AD, near where the second unique haplotype is found. The common name for the house shrew in Malagasy is "Arab rat" *voalavo arabo*, and along with the kinship occurring between Yemeni, Madagascar and Comorian populations (although they are both from an unknown source population that could equally be India, or elsewhere) I hypothesise an introduction with Arab traders in the early to mid second millennium AD through the north of the island. Shrew migrations to the south and east would have rapidly taken place after initial colonisation in vehicles and cargoes moved by humans (a mechanism central to the rapid dispersal of shrews on other islands) as the interior was settled by humans particularly in the central highlands and the arid south (Fritts *et al.* 1998; Dewar & Richard 2012). With very few competitors or predators the shrew was able to establish populations throughout the island. Madagascar is very heterogeneous with regards to habitat, topography and climate in all corners of the island and the shrew's rapid colonisation attests to its plasticity in novel environments (Dewar & Richard 2012).

In black rats, the highest diversity, oldest expansion signal and closest kinship with Omani rats was found in individuals in Ambohimiariana in the southeast of the country, and this is the proposed port of entry for the species (Tollenaere *et al.* 2010). The southern corridor is also a possible gateway for house shrews and cannot be ruled out, although genetically there is less evidence for such a route. Settlement along the southeast coast by human population groups show Arab-Islamic influences in the traditions and cultures of the people that indicate ties with Arabia since the early second millennium (the Onjatsy, the ZafiRaminia (ca. 1200 AD), and the Antemoro (ca. 1475 AD)) (Capredon *et al.* 2013). Although Comorian populations show a Middle Eastern genetic contribution in human populations, genetic evidence of shrew introductions (lower diversity in the Comoros) indicates the Comoro Islands were not the gateway by which shrews colonised this part of the Indian Ocean (Msaidie *et al.* 2010).

C_{III} *African-Arabian*

Although the small number of sequences within this subclade prevents calculation of diversity indices, from the MJ network I note that there seems to be high haplotype diversity and moderate diversity within it.

House shrews are not documented until the 19th century AD in the Red Sea region and other parts of East Africa but genetically, the accumulation of moderate nucleotide differences suggests older colonisation histories (Hutterer & Harrison 1988; Cheke 2010). Indian and Pakistani sequences are distinct from one another as expected from individuals within their native range that have diversified over a long period. The Saudi Arabian sequence is most closely related to that of Pakistan but with some degree of diversification. Based on morphotypes, Hutterer and Harrisson (1988) propose separate introduction routes to the Red Sea for individuals reaching Jeddah (Saudi Arabia) and Aden (Yemen), and this is confirmed through mitochondrial genotypes. Considering their proximity, the genetic distance between Yemen and Saudi Arabia is high (uncorrected p-distance=1.2%), probably resulting from the introduction of propagules from different sources that were already diversified, followed by isolation leading to further distance. Although there is a substantial proportion of missing nucleotides in the two individuals from the Kenyan coast, they are identical to each other and differ by only one nucleotide from the individual from Zanzibar, Tanzania (442 bp of sequence co-occur between all three). Importation from either the Red Sea or Arabian Gulf to East

Africa is likely to have been through a single, recent pathway, at least for the mitochondria.

In their introduced ranges in mainland Africa and Arabia, house shrews are locally abundant but have not expanded their distributions and in almost all cases are found in close association with humans. House shrews can adapt to a wide range of environmental conditions and it is likely that their restricted distribution is due to competition from other small mammals. The cosmopolitan black rat is a competitor of the house shrew and could be responsible for limiting the spread of house shrews, as could other similarly sized shrew species on the African mainland, e.g. the African giant shrew *Crocidura (flavescens) olivieri* (Hutterer & Tranier 1990; Varnham *et al.* 2004).

C_{IV} East and Southeast Asian

Individuals of this subclade are geographically restricted to continental and island East and Southeast Asia. A clear star-like population expansion and population bottleneck are evident from the MJ network and mismatch distribution (when anomalous sequences are removed). Introduction from a single core source to areas along the coast and to islands must have been very rapid as evidenced by the low genetic diversity but large geographic distance between pairs of populations. Although genetically homogeneous in the mitochondrial aspect, molecular characterization of autosomal markers reveals higher diversity in shrews from Southeast Asia than in South Asia (Kurachi *et al.* 2007b). Kurachi *et al* (2007b) propose that original populations in East and Southeast Asia were of a genotype more similar to South Asian populations than extant populations and recent immigration of the type commonly found today rapidly spread and displaced the original maternally inherited genotype. Thus, East and Southeast Asian populations may be significantly older than their mitochondria advocate. House shrews in the region are remarkably phenotypically diverse in size and colour and this may be due to multiple introduction events that are not witnessed in their mitochondrial inheritance (Kitchener *et al.* 1994).

This subclade is the only group not to be found on mainland South Asia. C_{IV} and C_V (Indian Ocean and northern Southeast Asia) are more closely related to each other than any other subclades are to one another (uncorrected p-distance=0.8%). The ML phylogeny places clade C_{IV} as a subclade of C_V. The MJ network indicates the C_{IV} subclade is related to C_V via an unknown ancestor (an unsampled or extinct haplotype). The feasible candidates identified by the MJ network as proximal source populations for Clade C_{IV} (those near the native range) are Malaysia and Sri Lanka. Forty-chromosome races of house shrews are found in Vietnam, Indonesia, Taiwan and Japan (and perhaps other areas of East and Southeast Asia that have not been karyotyped), and the same race is found in Malaysia but not Sri Lanka (where 2n=30-32 races are found) (Yosida 1982; Rogatcheva *et al.* 2000). It is therefore plausible that the source of East and Southeast Asian introductions is around the Malay Peninsula. The reduced chromosomal form is also found in Malaysia but this form is not widespread and it may represent a restricted, secondary colonisation of Malaysia after the standard karyotype was introduced, and therefore may have missed its window of opportunity for dispersal to the east.

C_V Indian Ocean and northern Southeast Asia

Subclade C_V is more closely related to all other subclades than any other subclade is to another (uncorrected p-distance = 1.4-1.7%) (Table 3.6). It is also the most diverse of all the clades (when anomalous Japanese sequences are removed from C_{IV}), and mismatch distributions (based on the downsampled dataset) show clear evidence of a rapid population expansion sometime in the more distant past.

Populations of C_v are sympatric with wild counterparts in the Nilgiri Hills in South India (*S. niger*) and the Sri Lankan Highlands (*S. montanus*) and, as discussed earlier, they can probably hybridize. Clade C_V in Sri Lanka is ascribed a subspecies taxon identification *Suncus murinus caerulescens* and it is distinct from Clade B, *S. m. murinus,* both genetically and morphologically although the two subspecies can interbreed freely (Meegaskumbura *et al.* 2010). Figure 3.13 shows a map of Sri Lanka with locations of individuals used in this study. *S. montanus* are restricted to the highlands and the rainforest in the south and southwest of the island and are not associated with humans but co-occur in the highlands with clade C_V (e.g. Kandy). *S. murinus* (clades B and C) is found throughout the island, except in rainforests of the south and southwest and is predominantly in commensal form (Meegaskumbura & Schneider 2008). A further subspecies *S. m. kandianus* (the Kandyan shrew) occurs in the central provinces, preferring less commensal, grassland habitats. Its taxonomic affinity is uncertain and I had no examples of this subspecies (Meegaskumbura *et al.* 2010).

Clade B is probably the older of the two *S. murinus* lineages and was probably associated with protohistoric people on Sri Lanka and is found in some of the oldest cities on the island (e.g. Anuradhapura). It is widespread throughout the island and is highly commensal. Clade C_V could have been a secondary invasion from the mainland some time after that of clade B and, although similarly highly commensal, it is found mainly in the bigger ports, towns and cities (Meegaskumbura & Schneider 2008). Even though clades B and C_V are both common, highly commensal and co-occur in major sea-ports (e.g. Colombo), there is evidence for only one taxon, clade C_V, being translocated outside of Sri Lanka. Clade C_V might be the more adaptable of the two subspecies and thus more capable of surviving boat journeys and colonising new territories.

Reduced chromosomal forms, 2n=35-40, 2n=30-32, and 2n=30-32, 37 and 40 are found in the Malay Peninsula, Sri Lanka, and southern India respectively (Rogatcheva *et al.* 2000). From a South Indian origin, house shrews with reduced forms were probably translocated to Sri Lanka. Sri Lankan shrews probably migrated from Sri Lanka to the Malay Peninsula, where they met an already established population of the standard karyotype and hybridized producing intermediate karyotypic forms. Although I only have two individuals from Malaysia that both cluster within C_V, previous studies identified both continental and island types of mtDNA on the peninsula (Kurachi *et al.* 2007a).

Réunion population (C_V)

Shrews are found over most of Réunion, in urban, ruderal and natural settings. The population has very low genetic diversity ($H_d = 0.095 \pm 0.084$; $\pi = 0.00008 \pm 0.00007$), which is evident in the tall peak around 0 in the mismatch distribution plot. Shrew colonisation of Réunion probably occurred in very recent history given the low diversity on the island and lack of a signal of population expansion. Réunion is an isolated island in the Mascarenes east of Madagascar. Although it was perhaps known to Arab traders it was not colonised until the 17th century by the French. House shrews were introduced to the island around the 1730s, around the same time as the Norway rat *Rattus norvegicus,* but died out; a subsequent recolonisation took place in the 1860s (Cheke 2010). My genetic data supports an East Indies introduction, possibly with the French from the eastern side of South India or elsewhere from the Bay of Bengal region where other merchant polities were located.

Although it doesn't show in the haplotype network (as it contained missing nucleotides in the sequence), an individual from Taramajima Island (Japan) clusters within the common haplotype from Réunion and is an anomaly. The individual is a laboratory shrew from a closed colony (AB066247, Table 3.2) that originated from wild caught animals on the southern Japanese island. Other Japanese populations are of the C_{IV} East and Southeast Asian variety and although Taiwan has instances of the C_V that are also somewhat anomalous they are of a distinctly different haplotype. Taramajima may have been very recently colonised from Réunion through modern transport and trade, or it may have been colonised from the same source as Réunion in the 19th or early 20th centuries.

Origins of commensalism

I conducted a comprehensive search of archaeological reports to determine the presence of the Asian house shrew in archaeological sites throughout their current distribution from East Africa throughout Southwest Asia and South Asia to ISEA.

The earliest fossil *S. murinus* (identified as “Suncus cf. murinus) is from Pliocene deposits in the putative origin region of northern India, dated to 2.3 to 5.6 million years ago (Patnaik 1997). The earliest archaeological finds of the Asian house shrew date to the Pleistocene and were found in Middle Palaeolithic (Mousterian) deposits in Qafzeh, Israel, a region the species no longer inhabits (Rabinnovich & Tchernov 1996). Shrew specimens found in late Lower Palaeolithic deposits in Qesem Cave, Israel could predate the Qafzeh finds but are yet to be securely identified (Maul *et al.* 2011). Today, the natural distribution of the species extends as far west as Pakistan and modern distributions further west are thought to be “probably the result of casual or intentional imports by man” (Hutterer & Tranier 1990). The Qafzeh finds represent the northernmost and westernmost range limits of the house shrew in the Pleistocene and it is assumed to be a natural range expansion, possibly due to the climatic warming in the Middle Palaeolithic (Rabinnovich & Tchernov 1996). This extinct and somewhat anomalous population indicates a much earlier expansion out of its native range than previously thought.

Outside of the Levantine finds, the Asian house shrew is mostly absent from archaeological deposits as far as I have determined, except for an indeterminate identification of an ancient Egyptian shrew mummy (David 1997), and a well-preserved rostrum and teeth from a recent deposit (USFR1) in Vallée Secrète, Cilaos, Réunion, shown earlier. In the case of the Egyptian mummy, an alternative identification is given - the African giant shrew (*Crocidura flavescens olivieri*). The species is synanthropic and its native range covers Egypt and most of Africa, and it is therefore a more plausible candidate. However, a positive identification as *S. murinus* cannot be ruled out as the

North African status of house shrew populations is unknown but they could represent a relict population from a colonisation in prehistory.

The shrew’s absence in archaeological contexts is probably for a number of reasons (beyond preservation and recovery issues as discussed in Chapter 2, though note that house shrew mandibles are sturdy and are expected to be relatively well preserved). In some areas where they are introduced, they may be a late addition to the faunal assemblage. In both their introduced and native ranges, their absence is perhaps best explained by their ecology: around human habitations house shrews, being insectivorous, forage in fields and other non-structural areas and, although they build dens under dwellings and other structures, their low density populations (usually a single large male and a few subadult males and females) and large ranges (1,490–3,773 m^2) compared to mice, for example, means that the frequency and density of bone deposits will be lower (Pocock *et al.* 2005; Nakamoto & Nakanishi 2013).

The lack of sub-fossils makes it difficult to determine the conditions by which commensalism evolved in house shrews and the origin of this association, however their genetics indicates an ancient origin for commensalism in the species. The house shrew is thought to have genetically diverged before the arrival of humans to the Indian subcontinent and the already diverse populations were transported by humans along migration routes to other parts of South and Southeast Asia in prehistory. Wild and commensal forms form a monophyletic genetic clade suggesting either a relatively recent commensalisation or the ability to go from a wild/feral to a commensal ecology. Given their sympatric ranges and ability to hybridize with wild species such as *S. montanus* I think the latter is more likely.

CHAPTER 4

PHYLOGEOGRAPHY OF THE *RATTUS RATTUS* SPECIES COMPLEX IN THE INDIAN OCEAN WITH A FOCUS ON AFRICA AND ARABIA

INTRODUCTION

Rattus is a speciose genus with an unresolved taxonomy but at least 66 valid species currently described (Musser & Carleton 2005). Of a multitude of species, four are widely distributed, the Pacific rat *R. exulans,* the Norway rat *R. norvegicus,* the black rat *R. rattus,* and the Asian black rat *R. tanezumi,* and this is largely attributable to their commensalism with humans with whom they have travelled throughout much of the world (Aplin 2003). The species *R. rattus* is part of a polytypic group of closely related species that are native to South and Southeast Asia, termed the *R. rattus* complex. There is enormous difficulty in separating the species morphologically, to the extent that even more distantly related taxa such as *R. argentiventer* and *R. tiomanicus* are difficult to distinguish and are often confused for members of the *R. rattus* complex (Aplin 2003). Coat colour has historically been used to identify forms but it is now realised that this is a very variable and unreliable trait.

Aplin *et al.* (2011) identified four main lineages within the *R. rattus* species complex (LI-IV) that occupy mostly distinct ranges from one another but with some areas of overlap in Southeast Asia. LI corresponds to the black rat *R. rattus sensu stricto,* which is sister to LII, the Asian black rat *R. tanezumi.* LIII is an uncategorised variant they place within a clade that also contains LI and LII. LIV is described as part of the Asian black rat form and is thus currently incorporated into the paraphyletic *R. tanezumi* group. Here, I distinguish *R. rattus* I*, R. tanezumi, R. rattus* III and *R. rattus* IV, together forming the *R. rattus* species complex. These categories follow Lack *et al.* (2012) and correspond to Aplin *et al.*'s LI-LIV. "*R. rattus"* is used to describe historical or archaeological records, where the lineage is uncertain.

R. rattus I is essentially cosmopolitan and is the lineage most prolifically translocated by humans. Tollenaere *et al.* (2009) studied phylogeographic patterns of *R. rattus* I in the Western Indian Ocean region and particularly the islands of Madagascar, the Comoro Islands and Réunion. Even though the islands are geographically close, the authors found evidence for multiple colonisation events, with Madagascar and Grande Comore populations originating in either India or the Arabian Peninsula, Mayotte being colonised from Madagascar, and a distinct invasion of Réunion from Europe. The addition of more samples to the phylogeographic framework allowed Aplin *et al.* (2011) to identify distinct translocation networks that separated the Madagascar variety from the "ship rat". The "ship rat" was linked to Age of Exploration voyages that reached both the Atlantic and Indian Oceans but that had source areas in both the Middle East and Europe.

Here, I extend the phylogeographic study of the black rat to encompass more of the Indian Ocean region, and specifically Africa and Arabia. Archaeological data suggest an ancient occurrence for black rats in the Bronze Age of Arabia and a first millennium AD introduction to East Africa, and I test whether such a temporal discrepancy is evident in the mitochondrial genome. Mitochondrial genes have been shown to provide useful phylogeographic structure in many taxa, including those associated with human translocations (Jones *et al*. 2013). Following previously published work (Tollenaere *et al.* 2010) I chose cytochrome *b* gene and the hypervariable region 1 (HVR1) of the D-loop for this study of rats.

MATERIALS AND METHODS

Taxon sampling

For this study, 375 morphologically identified "*Rattus rattus*" were sampled from museum collections, focusing on previously under-sampled locations within their native distribution as well as colonised areas on the continental landmasses surrounding the Indian Ocean, her islands, and the islands of Southeast Asia (Figure 4.1, Appendix 4.I). Due to morphological similarities between congeners within the *R. rattus* species complex, I expected "*R. rattus*" specimens to include many that were not *R. rattus* I. We did not sample "*R. tanezumi*" from the Indian Ocean region as the museums we visited did not contain specimens with this name outside their native ranges aside from the Cocos (Keeling) Islands and the Andaman and Nicobar Islands.

Collection of osteocrusts and skins was conducted as in Chapter 3. Previously collected tissue stored in dimethyl sulphoxide saline solution (DMSO) was provided by the FMNH (in collaboration with Jean-Marc Duplantier and Steve Goodman and collected by S.M. Goodman, W.T. Stanley, A.J. Tianarifidy, Z. Rakotomalala, B. Rasolonandrasana, and V.L. Soarimalala). J.C. Hillman provided tissue stored in ethanol or isopropyl alcohol. Material was sent by the sample providers for analysis. I also collected one individual from Zanzibar. The specimen was found dead at the House of Wonders Museum, Old Town, probably through a cat attack. A portion of the tail was removed and air-dried.

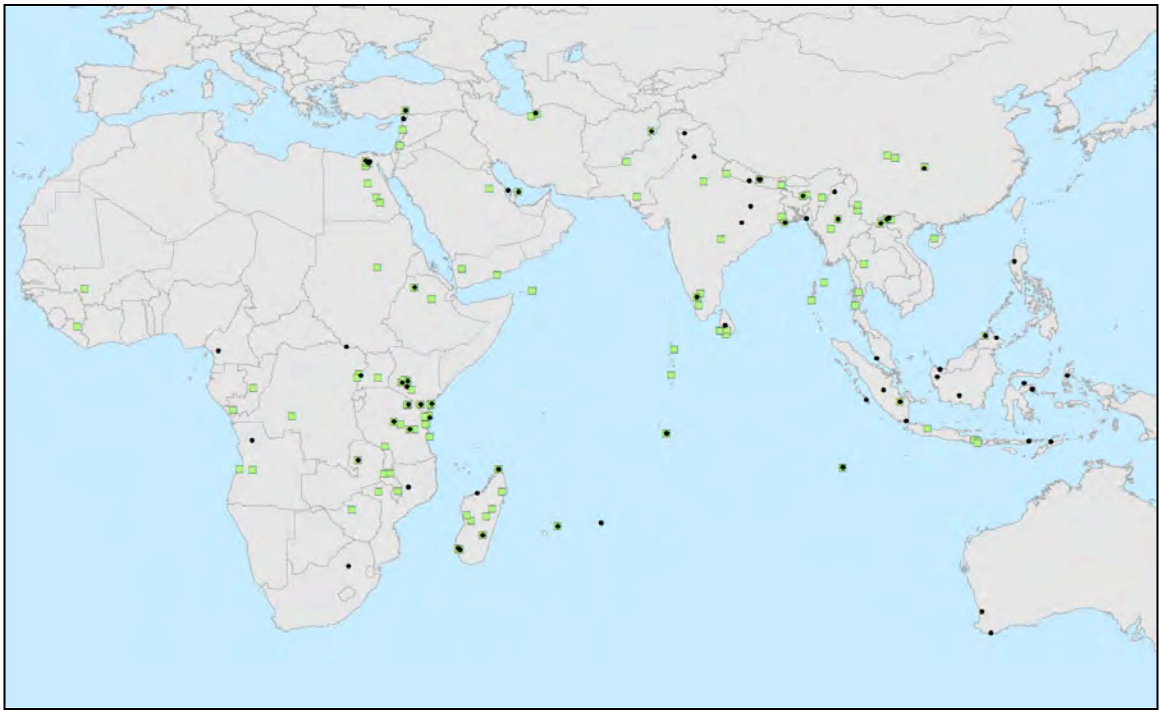


Figure 4.1. Sampling locations for specimens used in this study. Green squares indicate localities providing sequences incorporated in the study, black circles indicates where they were not (see Appendix 4.I for more information).

Molecular methods

Museum specimens were jointly extracted, amplified, and sequenced by Alex Trinks (AT) and myself (HME) in the DEAD Lab, Durham University. AT extracted and amplified modern tissue from Diego Garcia, and I extracted and amplified modern tissue from East Africa and Madagascar. The data are used by AT and myself for independent analyses.

Extraction and polymerase chain reaction (PCR) setup were conducted in a dedicated pre-PCR room. A number of specialized protocols to limit contamination were implemented: PCR was setup in a containment hood where all equipment was UV-irradiated for 30 min prior to use and 4 h after use; gloves were changed between handling each specimen tube at the extraction stage; negative controls were interspersed with samples 1:11 at the extraction stage (1:14 for modern tissue) and one per primer pair at the PCR setup stage; sequences were replicated; random replicates were independently amplified by AT and myself on different days, using different reagents.

Samples were extracted as in Chapter 3, except that museum skins were washed in 5% Tween 20 (a nonionic detergent) instead of phosphate buffered saline .

The cytochrome *b* gene and 490 bp of the D-loop incorporating hypervariable region 1 (HVR1) and flanking tRNAs were amplified in overlapping fragments. All positions described herein relate to the published whole mitochondrial genome sequence EU27307 from *R. rattus* isolate RNZRrTit01 (Robins *et al.* 2008). See Table 4.1 for a description of all primers. Modern tissue samples were amplified using primer sets U1/L2, U2/L2, U3/L3 and U4/L4 to produce fragments of 507-599bp. The same primer sets were used with osteocrusts and skins, however degraded DNA is typically fragmented and in many cases we did not achieve amplification products of this size. We designed primers that produced fourteen overlapping fragments, ranging in size from 94-172 bp, that incorporated variation in both *R. rattus* I and *R. tanezumi*. The primer nomenclature indicates the forward and reverse primer pairs used in this study (e.g. RrF1/RrR1 or RrDloopF1/RrDloopR1: Table 4.1), however primer pairs RrF1/RrR2 or RrF7/RrR8 were also used. Primers Aplin F1/Aplin R1 were used for R001-R006 as a trial.

As the quality of the DNA samples varied, I likewise varied my polymerase chain reaction (PCR) protocol. Twenty-five µl PCR reactions contained ddH_20, 1 µl of DNA template and final concentrations of 1 x buffer, 2mM-4mM $MgCl_2$, 2-2.5mM dNTPs, 2-2.5mM each primer, and 0.6-1.5 units Taq (Naxo Smart Taq Hot (Tartu, Estonia) or AmpliTaq Gold DNA Polymerase (Roche, USA)). Products were thermocycled 30 times for modern material and 35-40 times for museum material.

Locus	Primer Name	Sequence 5' to 3'	Position (EU273707)	Reference	Product (bp)
tRNA	U1	AATTTGTCATTATTTCTACACAGCATT	14043-14069	this study (AT)	559
Cyt *b*	L1	TAGGGTTGCTTTGTCTACTGAGAA	14628-14651	this study (AT)	
Cyt *b*	U2	CATCTGCCGAGACGTAAACTAC	14330-14351	this study (AT)	507
Cyt *b*	L2	GTCTCCTAGTAAGTCTGGGAAGAAT	14858-14882	this study (AT)	
Cyt *b*	U3	AGGATCAAACAACCCCACAG	14735-14754	this study (AT)	599
tRNA	L3	TGTTGATGGTGGGGAGTTAGT	15353-15373	this study (AT)	
tRNA	U4	TCTCAGGACATGTCAAGAAGAAG	15328-15350	this study (AT)	548
D-loop	L4	GGCATCCGAAAATTAAAAAATAC	15898-15920	this study (AT and HME)	
tRNA	RrF1	ACACAGCATTTAACTGTGACCA	14060-14081	this study (AT and HME)	94
Cyt *b*	RrR1	GGGCGGGAAGGTCAATGAAGG	14176-14196	this study (AT and HME)	
Cyt *b*	RrF2	TTAATCACTCCTTCATTGACCTTCC	14167-14193	this study (AT and HME)	139
Cyt *b*	RrR2	AGCCGTAGTTTACGTCTCGGCAG	14333-14355	this study (AT and HME)	
Cyt *b*	RrF3	TTAACAGCATTCTCATCAGTTAC	14302-14346	this study (AT and HME)	138
Cyt *b*	RrR3	GTTGCTATGACTGCAAATA	14485-14503	this study (AT and HME)	
Cyt *b*	RrF4	TCCTACACCTTCTTAGAAACATGAAAC	14442-14468	this study (AT and HME)	138
Cyt *b*	RrR4	AGCCTCCTCAGATTCATTCGAC	14607-14628	this study (AT and HME)	
Cyt *b*	RrF5	CAAACCTATTATCAGCCATTCCCTA	14566-14590	this study (AT and HME)	151
Cyt *b*	RrR5	AGTTTAGTCCTGTGGGGTTGTT	14742-14763	this study (AT and HME)	
Cyt *b*	RrF6	GCCCTTGCAATTGTACATCTCCT	14697-14719	this study (AT and HME)	142
Cyt *b*	RrR6	TGGGTCTCCTAGTAAGTCTGGGAA	14862-14885	this study (AT and HME)	
Cyt *b*	RrF7	GACTTACTTGGAGTATTCATGTTAC	14808-14832	this study (AT and HME)	127

Cyt *b*	RrR7	GGGATGGAGCGTAGAATAGCG	14960-14980	this study (AT and HME)	
Cyt *b*	RrF8	ACCCCACCACATATTAAGCCAGA	14916-14938	this study (AT and HME)	123
Cyt *b*	RrR8	TGGGCGGAATGTTAGACTGCGT	15062-15083	this study (AT and HME)	
Cyt *b*	RrF9	TTCTAATCTTAGCCTTTCTACCA	15019-15041	this study (AT and HME)	137
Cyt *b*	RrR9	AACTRATGGATGCTAGTTGG	15179-15198	this study (AT and HME)	
Cyt *b*	RrF10	AGCCAACCTCTTCATTTTAAC	15113-15133	this study(AT and HME)	166
tRNA	RrR10	GCTCTTCATTTTTGGTTTACAA	15300-15321	this study (AT and HME)	
tRNA/D-loop	RrDloopF1	AACTACTTCTTGACAGTACATAA	15394-15418	this study (AT and HME)	172
D-loop	RrDloopR1	TGGTGTATGTCTAATAACACAGA	15591-15613	this study (AT and HME)	
D-loop	RrDloopF2	ACATGAATATTCTTTCATACATT	15546-15568	this study (AT and HME)	111
D-loop	RrDloopR2	TTGTTGATTTCACGGAGGAT	15680-15699	this study (AT and HME)	
D-loop	RrDloopF3	AGACATACACCATTAAAGTCATAA	15602-15625	this study (AT and HME)	150
D-loop	RrDloopR3	GCCCTGAAGTAAGAACCAGA	15776-15795	this study (AT and HME)	
D-loop	RrDloopF4	CCCATACAACTTGGGGGTGA	15733-15752	this study (AT and HME)	146
D-loop	RrDloopR4	same as L4		this study (AT and HME)	
Cyt *b*	Aplin F1	ATCACACCCTCTACTCAAAA	14144-14163	Aplin *et al.* 2011	194
Cyt *b*	Aplin R1	GGCATGTAAGTATCGRATTAG	14358-14378	Aplin *et al.* 2011	

Table 4.1. Primers used in this study. Their locus position relative to GenBank accession number EU273707 (Robins *et al.* 2008) and their product length in base pairs is given.

PCR reactions were cleaned with FastAP Thermosensitive Alkaline Phosphatase and Exonuclease I (ThermoScientific, Pittsburgh, USA) following the manufacturer's protocol. Products were sequenced in either one (modern only for some individuals) or both directions on an Applied Biosystems ABI 3730. One µl of modern and 2 µl of museum template were used in sequencing reactions.

Sequence alignment and data generation

Not all specimens yielded usable sequence data (Appendix 4.I). Amplification failure was most likely due to severe degradation of DNA or misidentification of specimens (as the primers were developed for *R. rattus* I and the closely related *R. tanezumi* and would have failed on more distantly related forms). In four individuals, highly divergent sequence data and a nucleotide BLAST (National Institute of Health 2014) showed they were clearly not the expected species. Individuals with <50% sequence coverage across either the cytochrome *b* gene (cyt*b*), or the cytochrome *b* + D-loop genes (cyt*b* + D-loop), were removed from analysis as a high proportion of missing data reduces the likelihood of resolving the true tree in phylogenetic inference. Limitations of time and money meant that a small proportion of samples were not extracted. Some sequences were potentially cross-contaminated because they showed inconsistency between amplicons generated on different days, by different researchers, or using different primer sets. Although we re-amplified dubious fragments to detect the valid bases, I chose to exclude these sequences as in most cases we had others from the same location within the same haplogroup.

Sequence data were inspected, assembled, and checked for numts as in Chapter 3. Chromatograms showed irregularities (messy and/or dual peaks) in some museum sample sequences between positions 14,913-14,936 of the cyt*b* gene and I therefore excluded these positions from analysis for all sequences. The region was one of low polymorphism for *R. rattus* I and *R. tanezumi* and so it is not expected that the removal of these sites will influence phylogenetic inference to any great degree. The affected sequences were scattered among lineages and geographically and showed no evidence of heteroplasmy. I examined synonymous/non-synonymous and transition/transversion ratios and translation sequences but did not find any evidence of numts in this stretch of the sequence. Rather, I believe the irregularities were caused by suboptimal primers but because we did not clone these individuals I cannot be certain.

The mitochondrion is a non-recombining organelle that can be considered a single locus in a meta-analysis and therefore the two mitochondrial genes employed as markers here were concatenated to form a single dataset. However, an extensive published dataset is available for *R. rattus* and *R. tanezumi* for the cyt*b* gene, and fewer sequences are available for the cyt*b* + D-loop genes. Therefore I aligned my individuals with published sequences of the above taxa to create two datasets (cyt*b* and cyt*b* + D-loop) in order to maximise both individual sequence coverage (cyt*b*) and breadth of coverage in the mitochondrial genome (cyt*b* + D-loop loci). The alignments spanned positions 14,127 - 15,169 (1143bp) for cyt*b*, and 15,408 -15,887 (480bp) for D-loop. The cyt*b* + D-loop concatenated dataset did not include the tRNAs. The sequence alignments (cyt*b* n=1087; cyt*b* + D-loop n=250) were reduced to haplotypes in FaBox yielding 605 and 210 haplotypes for the cyt*b* and cyt*b* + D-loop datasets, respectively (Villesen 2007).

Data analysis

Phylogenetic trees were constructed using the cyt*b* and cyt*b* + D-loop datasets applying Bayesian Markov Chain Monte Carlo (MCMC) simulations in MrBayes v 3.2.1 (Ronquist *et al.* 2012). *Niviventer excelsior* (GenBank JQ927552) was used as an outgroup (following Aplin *et al.* 2011). *N. excelsior* is a member of the Rattini tribe and is closely related to the genus *Rattus* but is sufficiently distant to the *R. rattus* species complex to avoid unresolved taxonomic issues. The best-fit substitution models as chosen by the Bayesian Information Criterion (BIC) were estimated separately for the cyt*b* and D-loop genes in jModelTest v2.1.4, and were HKY+Γ and HKY+I+Γ, respectively (Posada 2008). A partitioned analysis was conducted for the concatenated dataset (cyt*b* + D-loop) with a separate substitution model for each locus. Two independent analyses were run from random starting trees with four chains for 15 million generations for the cyt*b* dataset and 5 million generations for cyt*b*+D-loop dataset, sampled every 1000 generations. To check for convergence and stationarity, I examined the standard deviation of split frequencies statistics of the two independent runs, examined the trace in Tracer v 1.5, and used the Compare function in Are We There Yet (AWTY) (Nylander *et al.* 2008). A 25% burn-in was used.

Maximum Likelihood (ML) trees for the cyt*b* dataset were computed in PhyML3.0 with the same substitution model as for the Bayesian analysis (Guindon *et al.* 2010). Gamma rate categories was set to 4, and all other substitution parameters were estimated by the program. Tree searching was done by subtree pruning and regrafting (SPR) and nearest neighbour interchange (NNI). To test the phylogeny, 1000 bootstrap replicates were executed, as well as an approximate likelihood ratio test (aLRT).

Aplin *et al.* (2011) used divergence time based on the *Mus*/*Rattus* split in the fossil record to calibrate a *Rattus* species tree, and calculated the tMRCA for the species clades. However, at the intraspecific or sister-species level (the level at which this study is targeted), a single, deep calibration point will overestimate tMRCAs giving erroneous estimates. At the shallow timescale at which black rat migrations associated with humans are expected to have occurred, this will tend towards unreliable inferences relating to human networks. For the black rat species complex no well-dated fossil records are available. Although we have aDNA sequence (see Chapter 5), the bones have not been radiocarbon dated, and the sequences are from shallow evolutionary scales (that will tend to underestimate tMRCAs of clades) and are therefore not used in this study. Calibration points at the medium-scale (either through aDNA with radiocarbon dates or known external events) are required to help resolve the calibration of the *Rattus* tree.

A Neighbour-net network for all individuals within the *R. rattus* I, *R. tanezumi* and *R. rattus* III and IV groups for the cyt*b* + D-loop concatenated dataset was generated in SplitsTree v 4.13.1 using an equal angle algorithm and uncorrected p-distance.

RESULTS

According to the phylogenetic analyses of the cyt*b* and cyt*b* +D-loop datasets (Figure 4.2, Figure 4.3a), the majority of my sequences reside within the major lineages *R. rattus* I and *R. tanezumi*, corresponding to lineages I and II of the *R. rattus* species complex after Aplin *et al*. (2011) (see also Appendix 4.I). There was high support for all major lineages in both datasets, except for the bifurcation leading to the sister taxon *R. rattus* III and *R. rattus* IV in the Bayesian cyt*b* tree and this relationship is not maintained in the ML tree. In the cyt*b* + D-loop tree this relationship changes so that *R. rattus* III becomes sister to the *R. rattus* I */R. tanezumi* cluster, and *R. rattus* IV becomes sister to the *R. rattus* I/*R. tanezumi/ R. rattus* III cluster. Bootstrap support for the *R. rattus* I taxon was lower than for other test statistics.

Within the *R. rattus* I and *R. tanezumi* lineages there are further clades. The various tests of branch support were congruent between Bayesian PP and and ML with an aLRT of branch support, yet ML and bootstrapping values were often below the standard support threshold of 70-80%. Shallow divergence levels and missing data for some individuals at informative sites in the *R. rattus* I clade may lead to low support for branches in phylogenetic tests. As such, I combine information from Bayesian PP, and ML trees with both bootstrap and aLRT tests of the phylogeny (the latter is expected to perform well in cases where branch length is short) to describe clades within the *R. rattus* I taxon (Guindon *et al.* 2010). Clades are determined to be units where both PP and aLRT values are above the support thresholds (0.95 – 0.98, and 0.8 – 0.9, respectively). Unless otherwise stated, I will refer to the Bayesian trees hereafter. Clades are labelled R1 - R4 (cyt*b* tree) and A1-A3, B, C (cyt*b* + D-loop tree) in the *R. rattus* I group, and are unlabelled in the *R. tanezumi* group as they are not discussed further (Figure 4.2 and 4.3). Overall, tree topology and sequence affinity within lineages and clades was the same between Bayesian and ML methods, except that for *R. rattus* I there was more phylogenetic structure in the ML tree (Figure 4.2 inset). The structure was not supported under a bootstrapping test of the phylogeny, however aLRT values show that two clades have moderate support. Clades R1 and R3 are nested within clades on the ML tree that contain sequences from India, which are ungrouped in the Bayesian tree. *R. rattus* I clade R3 has lower support in the cyt*b* tree but is described here for comparison with the concatenated cyt*b* + D-loop dataset (Clade B, Figure 4.3a). The corresponding names of clades between the cyt*b* and cyt*b* + D-loop trees are R1=C; R2=A2; R3=B (one individual in R3 crosses into group A); R4=A3; and A1 is an extra clade in the concatenated dataset. Clades A1-A3 are nested within clade A.

A polytomy is an unresolved dichotomous branch that may or may not reflect a true biological signal in the gene or the species in question (i.e. multiple divergences from a common ancestor at the same time). There is a polytomy in the Bayesian tree within the main branch of *R. rattus* I for the cyt*b* gene and within the internal branches of the three main lineages for which we have sequences (*R. rattus* I, *R. tanezumi,* and *R. rattus* IV). With the addition of the HVR1 of the D-loop gene to the data, bifurcations within the phylogenetic tree are slightly better resolved (Figure 4.3a).

The affinity of sequences within lineages is maintained in the distance-based Neighbour-Net network (Figure 4.4). A recent, rapid expansion within the *R. rattus* I lineage can be seen by the bush of short terminal branches. On the basis of this analysis *R. tanezumi* has three main groups within it: 1) the core type, 2) two basal sequences (sample R279/280 from India), and 3) another separate basal sequence (sample R044 from

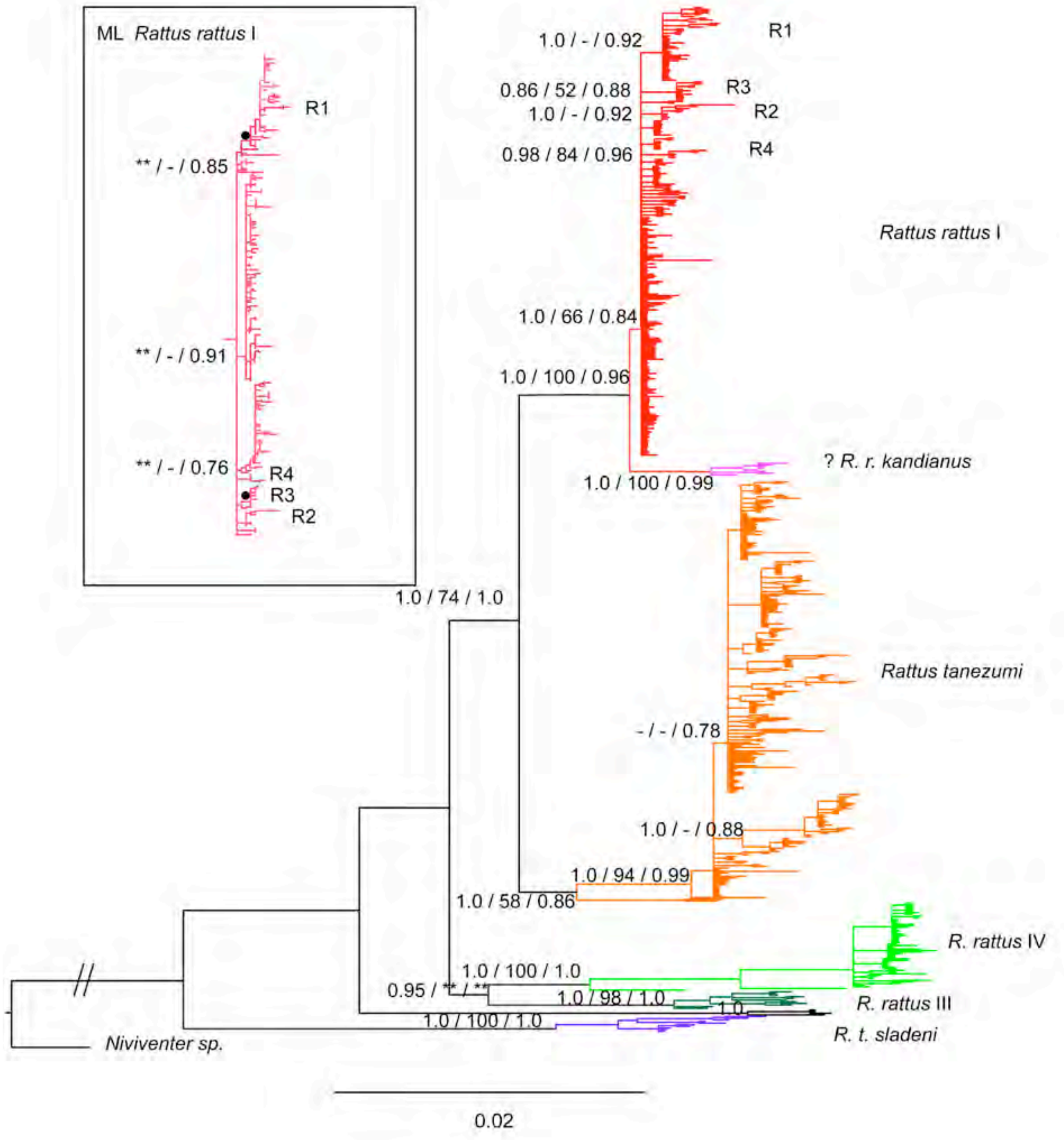


Figure 4.2. Bayesian MCMC trees for the cytochrome *b* gene for the major lineages and clades discussed in the text, rooted with *Niviventer excelsior*. Inset: Maximum Likelihood (ML) tree for *Rattus rattus* I. Bayesian posterior probabilities / ML bootstrap values / ML approximate likelihood ratio test values > 50% are placed at the node (- indicates <50 % support). ** indicates the node was not resolved in the analysis. Black dots in the ML tree are the node for clade recovered in the Bayesian tree. The unnamed black lineage contains one of my individuals plus published sequences but an unresolved taxonomy.

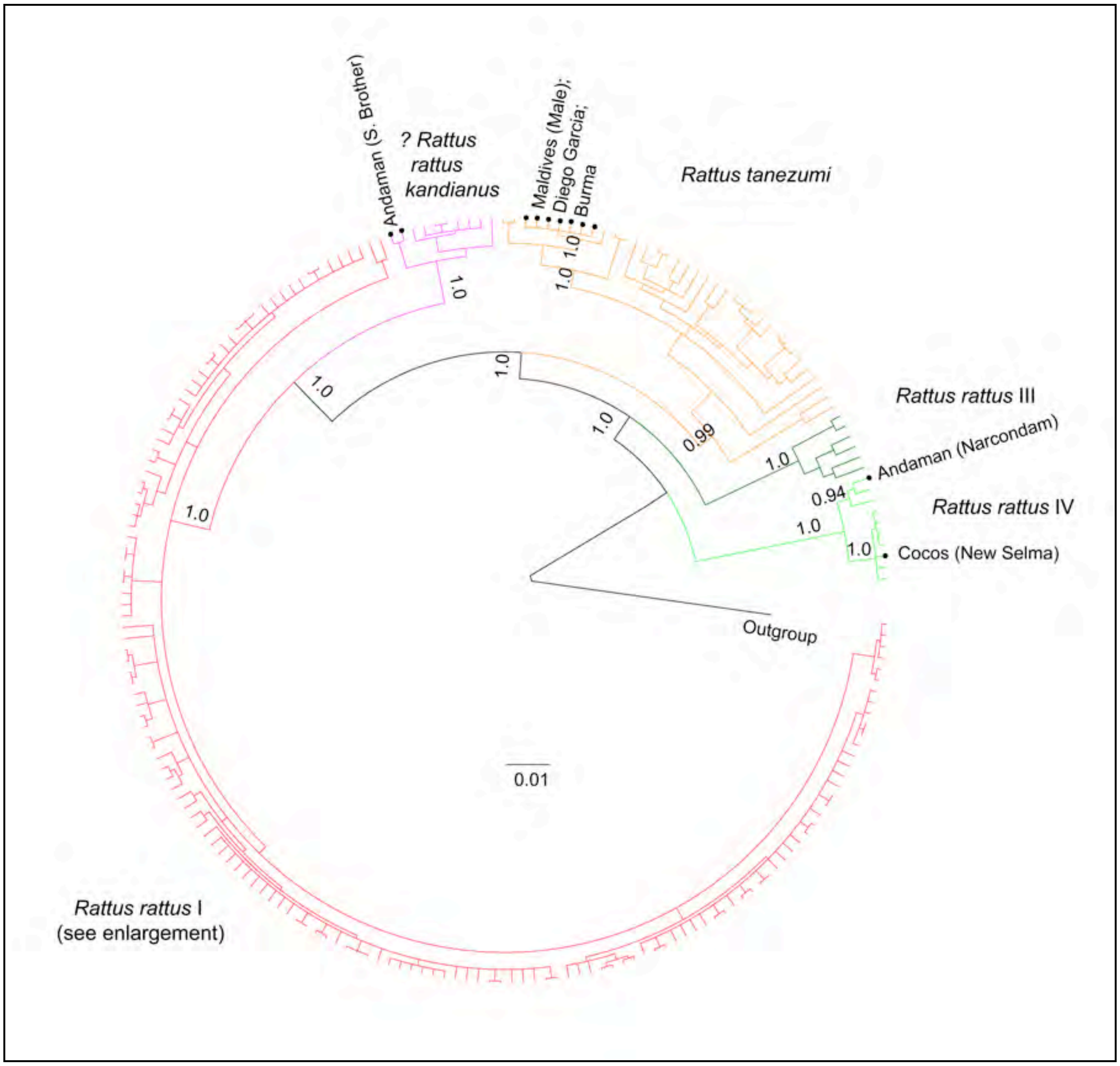


Figure 4.3a. Bayesian MCMC consensus tree of the cytochrome *b* and HVR 1 of the D-loop genes. The tree is rooted with *Niviventer excelsior.* Posterior probabilities for major branches and those discussed in the text are given. Locations for sequences (indicated by a black dot on the terminal branch) relating to human translocations are given for all taxa except *Rattus rattus* I*,* which is enlarged in Figure 3b.

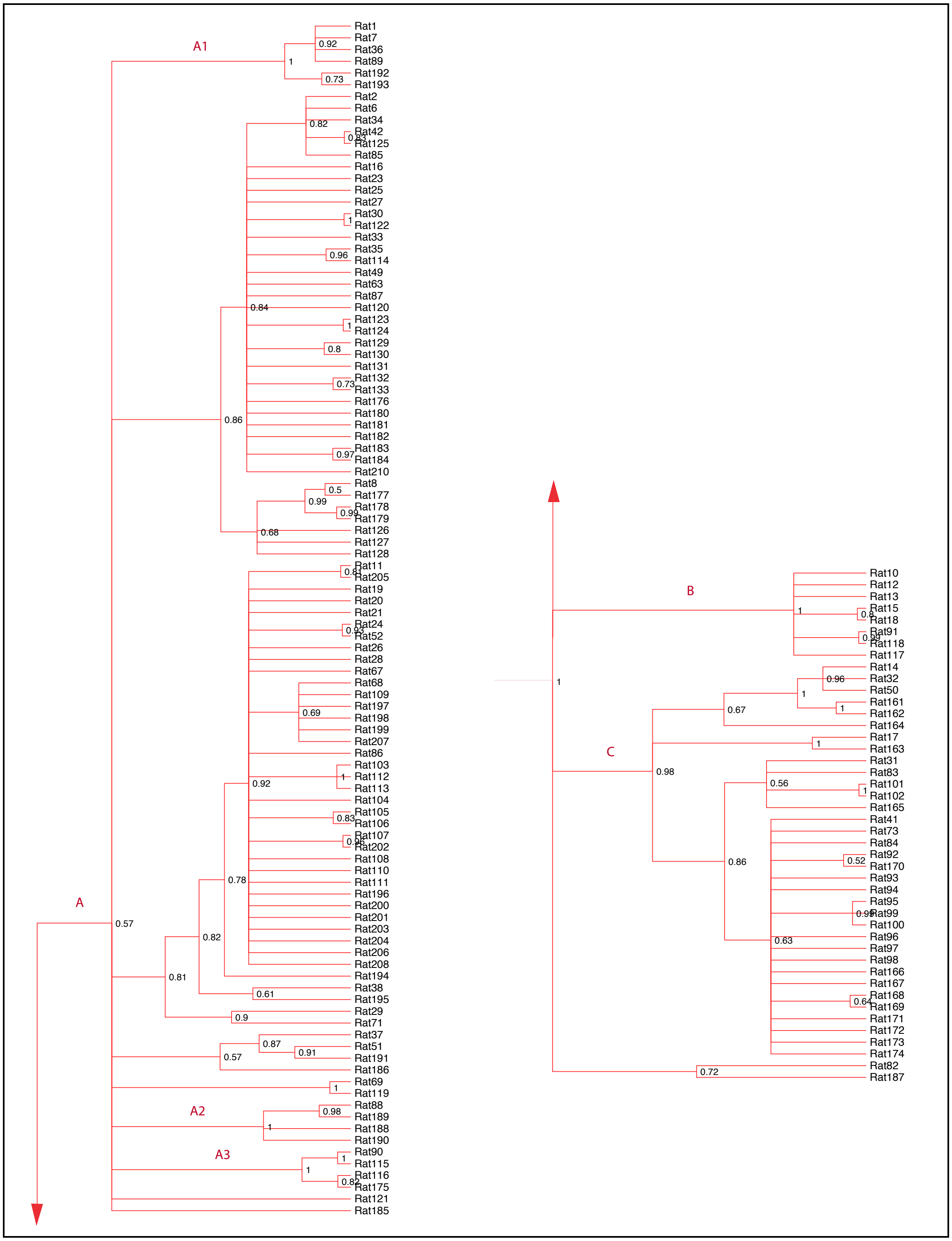


Figure 4.3b. Enlargement of the Bayesian MCMC tree in Figure 3a for the *Rattus rattus* I lineage. The tree is divided in two to fit within a page and is joined at the arrows. Haplotype identifications are listed in Appendix 4.II.

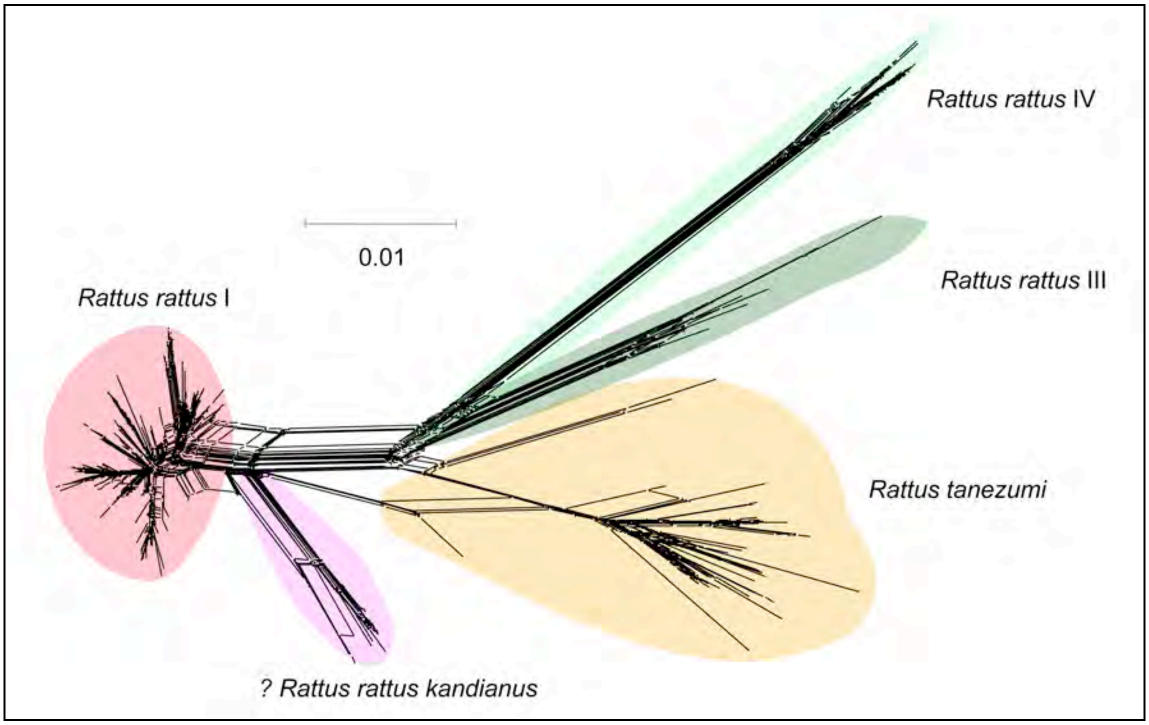


Figure 4.4. Neighbour-Net network (uncorrected p-distance) for the cytochrome *b* + D-loop concatenated dataset.

India), whose position is ambiguous as indicated by the box like signal between *R. tanezumi* and *R. rattus* I / *?R. kandianus.* Therefore, groups 2) and 3) may represent undescribed taxa within the native range of *Rattus*, although this is uncertain and requires further work as they are not distinctive in phylogenetic analysis (Figure 4.3a).

Although *R. rattus* III and IV are distinctive, they appear closely related in this analysis indicating that they could potentially be a single species. Aplin *et al.* (2011) calculate nucleotide diversity based on cytochrome *b* gene sequences for *R. rattus* I and *R. tanezumi* as 0.5% and 0.8%, respectively. This is relatively low diversity at the species level (for *R. rattus* I especially) compared to, for example, the Asian house shrew, which is 2.4% for the same gene (N.b. the Asian house shrew intraspecies diversity is 1.1% for the northern subcontinental clade and 0.9% for the widely distributed clade).

Within the unnamed black lineage in Figure 4.2, we contributed one specimen (R081 from Hainan, China). The published sequences of this black lineage were identified as *R. rattus* IV from Vietnam (GenBank JQ823537) and *R. tanezumi/ R. losea* from China (GenBank HM031707 and HM031708), and it is clearly an unresolved taxon. Our individual is described as *R. sikkimensis* (a synonym for *R. andamenensis*) in the museum database from which it was collected. Its collection location falls within the known ranges for *R. andamenensis* and *R. losea*.

The geographic distributions of the different lineages are shown in Figure 4.5. There are several aspects of these distributions that are worthy of note:

1. Ten specimens (eight haplotypes) formed a distinctive subgroup of *R. rattus* I in the mitochondrial phylogenies (Figures 4.2, 4.3a and 4.4), and most individuals of these were morphologically identified as Ceylonese black rats *R. rattus kandianus* (or *R. kandianus)* in the AMNH and FMNH databases. As expected, these were collected from Sri Lanka as well as two individuals from Brother Island, in the Andaman Islands group (Figure 4.5). I label these "? *R. rattus kandianus*".

2. One individual from Narcondam Island, also in the Andaman Islands group, and one from Selma Island in the Cocos Islands group, clustered within the *R. rattus* IV lineage.

3. I expand the likely native ranges of *R. rattus* III westward to western and southwestern Pakistan and eastern Afghanistan.

4. Two individuals from Male, in the Maldives, and all the rats from Diego Garcia, a remote atoll south of India (n=21), clustered within a group with an individual from Burma (within the native range) that formed a well-supported clade within *R. tanezumi* in the cyt*b* + D-loop tree (PP=1.0) (Figure 4.3a). These are clearly introductions and *R.*

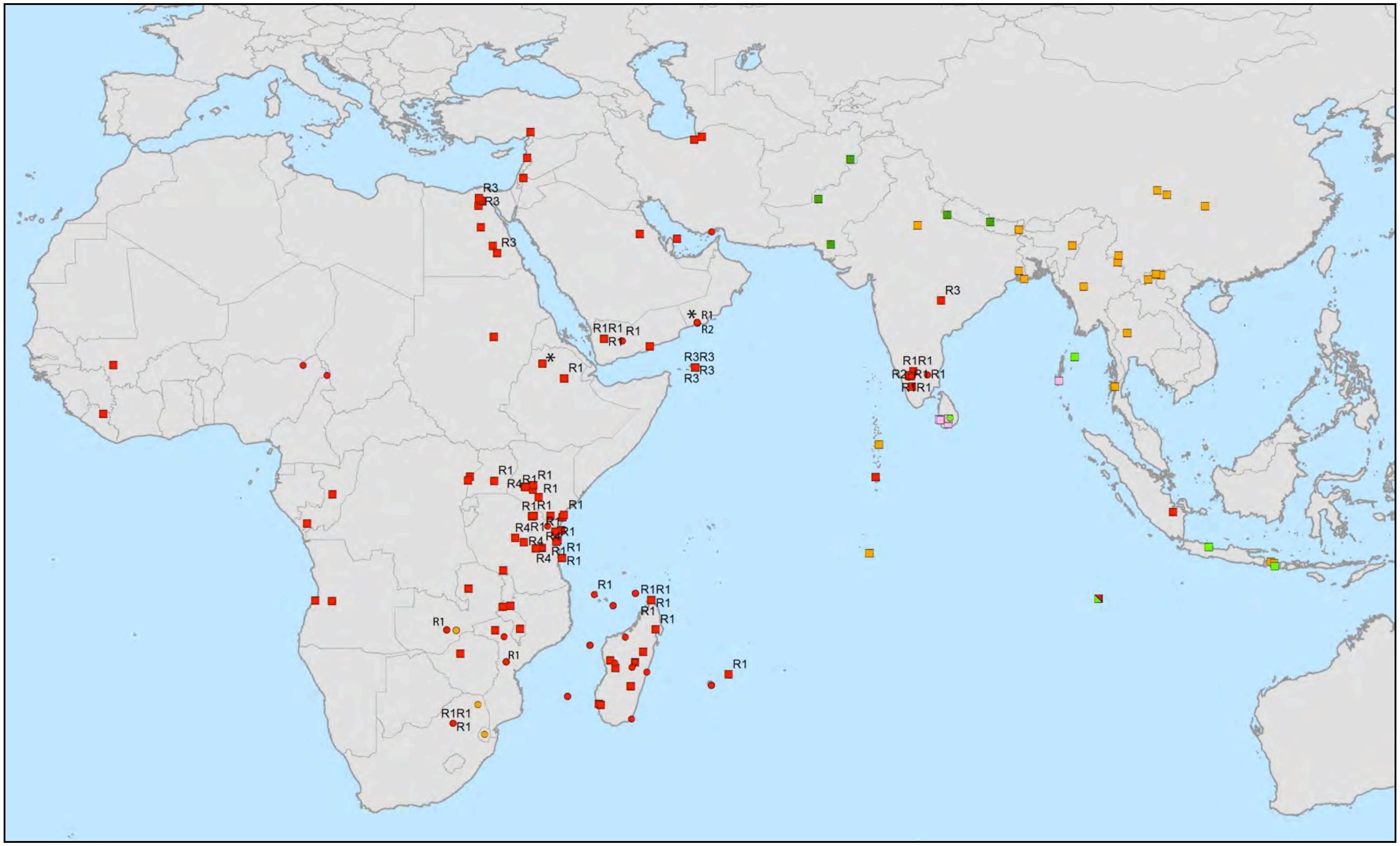


Figure 4.5. Distribution of cytochrome *b* lineages based on my data (squares) and published sequences for the western Indian Ocean region only (circles). More data are available for cyt*b* than for cyt*b* + D-loop (see Table 4.2 for a description of congruent clades between the cyt*b* and cyt*b* + D-loop trees). Red= *R. rattus* I*;* orange=*R. tanezumi;* dark green= *R. rattus* III; light green= *R. rattus* IV; pink= *?R. rattus kandianus*. *R. rattus* I subclades are marked (R1 to R4) and those unmarked are the major group. * indicates the extra clade in the cyt*b* + D-loop tree.

Clade	Location
A [R]	India, Egypt, Sudan, mainland Tanzania, mainland Kenya, Mozambique, Zimbabwe, Zambia, Uganda, Malawi, Mali, Congo, Angola, Liberia, Senegal*, [Nigeria*], [Cameroon*], Mafia Island, Pemba Island, [South Africa*], Madagascar, Îles Éparses*, Mayotte*, Réunion*, Maldives, Taub Island (Persian Gulf), Yemen, Saudi Arabia, Iran, Jordan, Turkey, Lebanon, Guadeloupe*, New Zealand*, [New Selma Island (Cocos)], [Papua New Guinea*], [French Polynesia*], [Samoa*], [France*], [Guinea*], [Australia*], [USA*], [Brazil*], [Venezuela*], [Guyana*]
A1 [R]	Ethiopia, Oman*
A2 [R2]	India, Oman*
A3 [R4]	Mainland Kenya and Tanzania
B [R3]	Socotra, Egypt, [India]
C [R1]	India, Ethiopia, mainland Kenya, mainland Tanzania, Uganda, Mozambique*, [Zambia*], [South Africa*], Mafia Island, Zanzibar Island, Madagascar, Mauritius, Grande Comore*, Yemen, Oman*

Table 4.2. Geographic distribution of sequences belonging to the clades and subclades (A1-A3, B, C) of the *Rattus rattus* I lineage in the Bayesian MCMC cyt*b* + D-loop tree and for the ungrouped sequences within clade A. The cyt*b* tree equivalent clade names are in square brackets (R1-4, ungrouped R) and locations of sequences that are not present in the concatenated dataset are added in square brackets. * additional countries/islands from the literature. See Appendix 4.I for data on individual specimens. Published sequences are predominantly from Robins *et al.* (2008), Tollenaere *et al.* (2010) and Russell *et al.* (2011b).

Location	Clade/group affinities
Yemen, South Africa, Zambia, Mozambique, Madagascar, Mafia Island	A†, C
Oman	A1, A2, C
Egypt	A†, B
Ethiopia	A1, C
Kenya, Tanzania	A†, A3, C

Table 4.3. *Rattus rattus* I clade affinities (A1-A3, B, C) and ungrouped sequences (A†) for countries/islands where more than one clade (or subclade) occurs.

tanezumi has also been translocated to South Africa and Zambia (Figure 4.5) (Bastos *et al.* 2011; Nakamura I *et al.* 2013).

5. *R. rattus* I is the type found most widely distributed in the Western Indian Ocean including its putative native range of India, plus Africa, Arabia and the Middle East and islands of the Indian Ocean. It is also found in Europe, the Americas, and the Pacific Ocean region (from published sequences, not shown in Figure 4.5). The geographic locations of haplotypes for supported clades within *R. rattus* I are shown in Table 4.2. In some countries or islands, multiple clades of *R. rattus* I co-occur or, in some cases, clades and ungrouped sequences co-occur (Table 4.3).

DISCUSSION

Taxonomy

The *Rattus rattus* species complex is widely distributed throughout the Indian Ocean region and subdivided into multiple distinct lineages which have largely parapatric distributions but with some areas of overlap in Indonesia, the Indian subcontinent, Sri Lanka and southern Africa (this study and Aplin *et al.* 2011). The distribution pattern of the species complex shows it has undoubtedly benefited from its association with humans in terms of expanding its range, and this is especially true of *R. rattus* I. The benefit is derived not only from the movement of founders from one place to another in ships and other land and sea vessels, but also from the environment that they meet when they arrive, and in which they thrive as commensals (e.g. disturbed, agricultural, and urban environments).

It is not clear whether *R. rattus* III and IV are valid species. They appear as distinctive monophyletic groups in all analyses. They are sister taxa, most evident from the

Neighbour-Net network, but whether they are highly distinct forms within one species, or separate species is unclear, and would benefit from further study with nuclear markers and morphometrics.

Aplin *et al.* (2011) and Lack *et al.* (2012) recognise four taxa within the *R. rattus* species complex. Based on this study there is potentially a fifth taxon: *? R. rattus kandianus*. This monophyletic group is clearly closely related to *R. rattus* I, but it is undoubtedly genetically distinct on the basis of mitochondrial DNA sequence. However there are some doubts as to how this fifth taxon relates to the form that has been defined morphologically either as *R. rattus kandianus* or *R. kandianus* (Musser & Carleton 2005). Previous studies have sequenced the cytochrome *b* gene of individuals described as *R. (rattus) kandianus* and based on this have placed the taxon within the *R. rattus* IV lineage (Robins *et al.* 2007; Pagès *et al.* 2010). The published sequences also cluster within that lineage in my phylogeny. Pagès *et al* (2010) state that the lineage contained a number of cryptic variants of the *Rattus* complex and that there is a mismatch between the species identification of *R. (rattus) kandianus* and its genetic identity.*? R. rattus kandianus* individuals here are clearly of a different lineage from those previously sequenced by Robins *et al.* (2007) and Pagès *et al.* (2010). The attribution of taxonomic status based on morphological or geographical characters is clearly problematic in the *R. rattus* complex and we had individuals that were identified as *R. rattus kandianus* but phylogenetically clustered with both *R. rattus sensu stricto* and *R. tanezumi.*

A complicating factor in the identification of individuals and to the validation of taxonomic units within the *R. rattus* complex is that at least some of the lineages are known to have overlapping distributions and hybridise in natural settings (Aplin 2003; Bastos *et al.* 2011; Lack *et al.* 2012). The extent of hybridisation in both geographic and genetic terms is unknown. This is relevant for my study as hybridisation may lead to mitochondrial introgression, which could lead to a mismatch of morphology and genetics. With the addition of samples in this study I have extended the known range of mitochondrial lineages *R. tanezumi* and *R. rattus* III to a point where there may be a natural contact zone in eastern Nepal, and this could provide a natural laboratory for testing the strength of reproductive isolation and hybridization between the lineages.

Polytomies in the phylogenetic tree

The cyt*b* tree has multiple polytomous nodes within the main branch of the *R. rattus* I lineage and within the internal branches of *R. tanezumi,* and *R. rattus* III and IV. Many of the polytomous nodes are resolved when the HVR1 of the D-loop is added to the sequence dataset.

Typically, a true polytomy occurs when there is simultaneous radiation from a source population and so multiple (not just two) descendants are equally related to a common ancestor. While this may be feasible for an invasive species where multiple progenitors can be translocated to a new and isolated range at the same time and found a population, it is unlikely that this is the signal we are seeing in the data. Rather, the signal is probably due to a rapid (but not simultaneous) radiation from one or more source areas that are unsampled or have gone extinct and is thus more likely a molecular polytomy rather than a species polytomy.

Molecular polytomies typically occur when sampling is incomplete, when branch length is small (in this case a rapid recent radiation), or due to homoplasy, all of which are considerations within this dataset. Pages *et al.* (2010) found that rapidly evolving mitochondrial markers helped resolve terminal nodes of the phylogeny for the Rattini tribe, and here I find that the highly mutable D-loop gene resolves terminal nodes when concatenated with the more conserved cytochrome *b* gene.

Phylogeographic patterns within lineages

Rattus rattus I

The *R. rattus* I lineage is the most widespread of the *R. rattus* species complex occurring on all continents, except Antarctica, although it has colonised as far south as the subantarctic Macquarie Island, where it was introduced by sealers in the 19th century AD (Pye *et al.* 1999). It is found widely distributed on islands throughout the world.

According to the mitochondrial markers used here, within the *R. rattus* I lineage are a number of clades that are geographically distributed, and a number of ungrouped haplotypes that form, in many cases, a polytomy within clade A. The units discussed here are the clades A (plus subclades A1-A3), B and C. Given that many of the *R. rattus* I nodes are polytomies, it is hard to distinguish relative timings between clades or haplotypes in most cases.

Subfossil *R. rattus* are known from the Levant by at least the beginning of the Holocene and were probably in Western Europe, as far west as Switzerland, in the pre-Roman period *ca.* 1000 BC (Toškan & Kryštufek 2006). Sequences that derive from the Near East, Middle East or Western Europe (Iran, Jordan, Lebanon, Turkey, France) all form polytomies within an undefined group on the A branch of the Bayesian cyt*b* + D-loop tree. Haplotypes of *R. rattus* I in the Americas are similarly part of the ungrouped clade A but their introduction to the region is much later, they spread quickly and were established in coastal North, Central and South America by the late 16th century AD (Musser & Carleton 2005). Aplin *et al.* (2011) found the most widespread genetic form of rats, that here make up part of the A clade, had low diversity and is probably the result of lineage restriction due to sequential founder events and I find a similar pattern. Extinctions of local populations, even over a large area, are possible. For example, *R. rattus* is found relatively frequently in cave and rockshelter sites in the early Holocene of the Levant (e.g. Hayonim Cave) but subfossil remains disappear by the end of the Natufian period, suggesting an ancient extinction event possibly linked to the cooling of the climate during the Younger Dryas. From my data, it is therefore likely that ancient populations have either gone extinct, or have been successively replaced by secondary colonisers, or there has been replacement of the mitochondrial genome by a successful female invader, or group of invaders from a single or few closely related populations. For my study, this means the mitochondrial locus may not always reflect the ancient distribution of populations with respect to their genetic identity within the *R. rattus* group. Black rats are implicated in the spread of the Black Death (the plague *Yersinia pestis*) in the 14th century AD (McCormick 2003). Whether or not black rats are the agent of the plague pandemic is under debate but their significant co-distribution with the disease is undeniable, and massive population declines of black rats, caused by high mortality in response to infection with the plague bacterium, are archaeologically and historically documented (McCormick 2003; Panagiotakopulu 2004). With significant declines covering much of the ancient black rat range it is possible that the modern phylogeographic signal we see is due to extirpation of lineages in the mid second millennium AD brought about by mortality in response to plague.

Multiple genetic signatures, where clades or clades and ungrouped sequences co-occur, are most likely due to multiple colonisations by *R. rattus* I (at least the females) that have been introduced from different source locations via different transport routes. However, an ancient introduction followed by divergence of populations within the colonised range cannot be ruled out as polytomous nodes prevent ancestral relationships being inferred. Although contemporaneous populations may consist of descendants of multiple lineages that have secondarily colonised an area, it is also possible that secondary colonisers have filled vacant niches or replaced their predecessors; the scale of my analysis is too coarse to be able to discern population structure at the local level. In Madagascar however, the disjunct distributions of the C (R1) clade in the very north of the island and the A (R) clade elsewhere on the island suggests neighbouring ranges that may reflect niche-filling or replacement events sometime after initial colonisation of the island. This is quite feasible as black rats form social groups with a dominant male, and a hierarchy of both males and females, and defend their territories. As such, even a female interloper would have difficulty securing mates in an established population.

The broad phylogeographic pattern within Africa divides Egypt, East/Northeast Africa, and West Africa into different zones of interaction. Phylogeographic inference from current (albeit limited) sampling distributions, and the lack of zooarchaeological remains prior to the early 19th century AD, suggest West Africa (occupied by the globally-distributed A clade) is part of a European incursion at the end of the second millennium AD. The genetic pattern fits with that suggested for house mice – an introduction to West Africa with European traders and colonisers in the same period (Bonhomme *et al.* 2011). However, the lack of systematic retrieval and identification of faunal remains (particularly micromammals) in West Africa limits the inferences I can make about the arrival of rats to the region based on zooarchaeological data.

Clade A is the lineage on Madagascar reported by Tolleneare *et al.* (2009), which they attribute to introduction through Arab trade in the early second millennium AD mainly based on zooarchaeological evidence (assuming population/haplotype persistence). With the addition of new samples, I rethink the introduction of rats to Madagascar. In their analysis Tolleneare *et al.* (2009) find that a haplotype from Mahajanga in the west of the island (haplotype 31) clusters with those from Réunion and South Africa, which would place it in my clade C, and they attribute this variant to a post-colonisation immigration event. However, in my analysis, this haplotype clusters in clade A. Therefore, in my analysis, all previously published sequences on Madagascar are within the globally distributed clade A. With the addition of samples in the north of the island I recover a second clade (clade C) on the island, restricted to the very northern tip.

There is some structure in clade A as it pertains to Africa although limited sampling and resolution of the phylogenetic tree make it difficult to describe these clades beyond simple inferences. Clade A1 is found in Ethiopia and Oman only. Relative to all other clades, this appears to be a later intrusion. Clade A3 is currently only found in Kenya and Tanzania and could represent *in situ* diversification from a clade A ancestor, or rapid dispersal after introduction from an unknown source.

Egypt is predominantly the B clade, which is currently also found only on the island of Socotra (off the coast of Yemen in the Arabian Sea). The low genetic diversity of rats in Egypt, which indicates a relatively recent introduction, is at odds with the zooarchaeological data for the country. The earliest recorded black rats in Egypt are in the Hyksos period *ca.* 1750 BC, although whether these are reliably identified and stratified is questionable, but they are certainly on the Red Sea Coast by the Ptolemaic/Roman period (see Chapter 2 for a more thorough review).

The distribution of clade C is restricted to the Western Indian Ocean region only. It includes southern India, the Arabian Sea coast of the Arabian Peninsula, East Africa from Kenya to South Africa and including Mafia and Zanzibar islands, Ethiopia, and more distant islands including Madagascar, Grande Comore and Réunion. In Arabia, clade C rats are found in Oman and Yemen (from the capital, Sana'a, and unknown locations). The distribution of clade C shows Gulf of Aden ties with external regions. The fact that under current sampling regimes, Persian Gulf and Red Sea locations are excluded from this network suggests direct connections between the Southern Arabian Peninsula coast and cultures east and south of it in a strictly Indian Ocean network, although this may be an artefact of sampling. The pattern is indicative of Arab trade networks that flourished in the region and extended via the East African coast to Madagascar (Chami 2002; Boivin *et al.* 2013a). The phylogeographic pattern probably also reflects the later Omani Empire dominance of the coastal flanks of the Arabian Peninsula, East Africa and the Western Indian Ocean in the 18th and 19th centuries AD. By the 19th century AD, Omani and Swahili caravan routes from coastal East Africa to the interior were well established, trading grain and other commodities, raw materials and luxury goods (Mitchell 2005). These caravan routes and the types of cargoes moved along them would have facilitated the spread of rats inland during this period. In Uganda, black rats are documented from around 1920 near Lake Victoria and within 50 years had expanded westward to occupy most of the country (Kingdom 1974).

How well does the phylogeographic pattern fit with zooarchaeological data for East Africa? There are no published ancient DNA studies, nor any systematic revisions of the zooarchaeological data for the black rat in East Africa. The difficulties in distinguishing introduced black rats from native rats in Africa (e.g. species of *Mastomys*, *Tatera* or *Arvicanthis* genera), along with the commensal niche being occupied by native species (e.g. African grass rat *Arvicanthis niloticus*), probably results in cases of misidentification of black rat in archaeological excavations and I discuss this more in the next chapter. That being said, black rats are first described from Unguja Ukuu on Zanzibar in the 6th-8th centuries AD and from then on throughout coastal Kenya by the 10th century AD (e.g. Lamu, Shanga, Chombo) (Boivin *et al.* 2013a). In southern Africa the rat appears in Bosutswe in Botswana and Ndondonwane in KwaZulu-Natal, South Africa in the 8th century AD, and Pont Drift in the Limpopo Valley (South Africa) between the 8th-12th centuries AD (Boivin *et al.* 2013a). In Isamu Pati, Zambia it arrives by at least the 14th century AD and possibly as early as the 10th century AD (Fagan 1963). Southern Africa and Zambia are linked to Indian Ocean coastal networks in the Iron Age (the late first millennium AD) (Mitchell 2005). The translocation of rats along trade routes that were in operation between East Africa and the interior in the Iron Age is not clear: the lack of integrading rat remains along the routes and the very short period between the purported onset of colonisations in the regions (at least for South Africa and East Africa) suggests there is little possibility the populations are related. The absence of zooarchaeological specimens could be an artefact of preservation, recovery, and/or patterns of zooarchaeological study, which is almost certainly obscuring ancient distributional patterns. The ancient geographic distribution of black rats appears to be reflected in modern phylogeographic patterns, especially the range of clade C, and this could indicate the introduction of rats in the earliest phase of the East Africa-Arabian network *ca.* 8th century AD.

With regards to the Arabian Peninsula, there were probably two phases of colonisation as it relates to the mitochondria (clades A and C). It is, however, difficult to infer the relative timing of these introductions. The widespread A clade is found throughout the Arabian peninsula. The ungrouped haplotypes are found in the Persian Gulf region of Saudi Arabia and Taub Island, and the Gulf of Aden in Yemen. Oman hosts two subclades within clade A (A1 and A2). A1 is linked to haplotypes found in Ethiopia only, as previously discussed for the African region. A2 is linked to a single sequence from Tamil Nadu, in southern India.

? Rattus rattus kandianus

Outside of its probable origin in Sri Lanka, I also identified two individuals of this lineage from South Brother Island in the Andaman Islands group that were presumably translocated there. The two Andaman Islands haplotypes cluster with each other and therefore likely reflect a single introduction.

Rattus tanezumi

Although its proximity to mainland South Asia, where *R. tanezumi* is indigenous, means it could represent a native taxon, *R. tanezumi* on the Maldives clusters with counterparts from Diego Garcia, a remote atoll in the middle of the Indian Ocean, along with an individual from Burma in a clade (Bayesian cyt*b* + D-loop tree, PP=1.0) with low genetic diversity, and thus probably represent a recent invasion. A source population from Southeast Asia is likely but whether these lineages represent separate colonisations from a source or a stepping stone model (e.g. introduced to Maldives then rapidly from Maldives to Diego Garcia) cannot be ascertained with the limited number of sequences. The black rat (at the time of documentation the distinction between species was not made) was reportedly introduced to the Maldives prior to 1886 and to

Diego Garcia prior to 1884 (Atkinson 1985; Long 2003). Note, the distinction as to which islands in the Maldives were colonised in the 19th century is not made and the *R. rattus* I clade found on the island probably predates its documented arrival. Assuming *R. tanezumi* represents the earliest coloniser, I hypothesise the black rat colonised Diego Garcia before it is first documented and prior to the early 19th century AD as, around this time, the island began to be administered from Mauritius, where the black rat was well established, and if a stable *Rattus* population was not already present on the island it would have been rife for invasion by the *R. rattus* I form.

Bastos *et al.* (2011) made a comprehensive study of *Rattus* in South Africa and found cryptic *R. tanezumi* in sympatry with *R. rattus* I*,* the known occupant of the region. At the time of their publication only a few published sequences were available to incorporate into a phylogeographic framework and they were unable to deduce relationships between populations. Since then, a number of *R. tanezumi* haplotypes have been published and I find that South African *R. tanezumi* forms a clade with individuals from Lusaka, Zambia (cyt*b* tree, PP=1.0) (Nakamura I *et al.* 2013). The clade has low diversity and the same haplotype is found in both countries.

R. tanezumi hybridizes with both *R. rattus* I and *R. rattus* IV in an apparently unidirectional manner as regards genetic introgression, from *R. rattus* I and *R. rattus* IV to *R. tanezumi*, based on nuclear markers, and the *R. tanezumi* form is identified through mitochondrial DNA only (Lack *et al.* 2012). For the purposes of this study, where my phylogeographic inference is based on mitochondrial markers, the *R. tanezumi* lineage I recover is therefore likely to be the older of the signals in relation to rat colonisations.

Rattus rattus III and IV

Rattus rattus III is not discussed here as I did not identify new sequences belonging to this lineage.

R. rattus IV probably originated in mainland Southeast Asia and its translocation to Sri Lanka from the Sunda Shelf/Cambodia region across the Bay of Bengal suggests an ancient maritime connection between the regions (Aplin *et al.* 2011). The haplotype from the Andaman Islands is also probably an ancient introduction and its grouping in a clade distinct from Sri Lanka (cyt*b* tree, both clades PP=1.0) suggests it was imported from the Sunda Shelf region via a separate introduction pathway. Black rats (*rattus/tanezumi*) are known from the Cocos Keeling Islands since 1878 and my haplotype from New Selma Island, which clusters in a clade of low diversity, support a relatively late colonisation of the island (Atkinson 1985). The occurrence of both *R. rattus* and *R.rattus* IV on the island is an extension of the geographic overlap that occurs between various lineages in southwestern Indonesia.

CHAPTER 5

ANCIENT DNA SHOWS GENETIC CONTINUITY AND NOVEL DIVERSITY IN INDIAN OCEAN BLACK RATS (*RATTUS RATTUS* SPECIES COMPLEX)

INTRODUCTION

Colonisation histories of naturally dispersed and introduced species are often reconstructed from modern genetic data using a phylogeographic approach (Avise *et al.* 1987). However, phylogeographic studies are usually limited by the period in which the samples were collected (usually relatively recently), and thus only give a snapshot of genetic and geographic relationships at a slice in time. In some cases, historical documents or the documentation of subfossil remains are available for the species and region in question, which can help elucidate ancient or at least pre-modern distributions (Cucchi *et al.* 2005b; Bonhomme *et al.* 2011). Recent advances in genetic and genomic technologies have added ancient DNA (aDNA) to the tools used to infer phylogenetic relationships and genetic diversity over space and time (Larson *et al.* 2007; Haile *et al.* 2010).

Modern or contemporary phylogeographic patterns may not accurately characterise earlier ones. In many cases we may expect change in the distribution of haplotypes, for example, where extinctions or population turnovers have occurred. Conversely, the continuity of haplotypes through time is also expected, for example, where populations have been isolated, or are stable. To examine the geographical distribution, phylogenetic relationships, and genetic diversity of a population, species, or group of species, in space and time requires information contributed through aDNA (Hadly *et al.*

2004). Chronological studies of rodents using aDNA and modern DNA incorporated into phylogeographic frameworks have variously identified continuity of populations and the extinction of previous diversity. For example, Viking Age house mouse (*Mus musculus*) bones from Iceland are of the same clade that is found on there today (Jones *et al.* 2012). Similarly, ancient Pacific rats (*Rattus exulans*) from Rapa Nui are of the dominant haplotype that has been widely introduced throughout East Polynesia (Barnes *et al.* 2006a). In contrast, aDNA from Pacific rats on Norfolk Island uncovers haplotypes that are unknown from modern populations and may represent extinct diversity (Matisoo-Smith *et al.* 2001).

Identifying the distribution of micromammals in pre-historical contexts has so far largely relied on zooarchaeological identifications of bone retrieved from archaeological or paleontological sites. Recovery issues aside, this process has been marred by the difficulty in positively identifying bone elements to the species level. Biotic and abiotic processes cause taphonomic changes to bones, often rendering them fragmented and unidentifiable, and the small size of micromammal bones makes them more prone to decay than the bones of larger animals. Even intact micromammal bones often lack morphological features that distinguish them from similar species and morphological variation within species or even within geographical populations can further complicate diagnostic tests. Furthermore, the bias towards economically important species in archaeology (e.g. domesticates or prestige species), until recently at least, has meant a lack of expertise for and/or study of those groups traditionally considered less important (including micromammals). The difficulty in accurately identifying micrommamal remains to the species level has resulted in patchy or, in some cases, erroneous information relating to pre-historical distributions. To address such issues in relation to rats of the Pacific region, Robins *et al.* (2007) developed a molecular assay that distinguishes between morphologically similar species in the *Rattus* genus. They were able to make their study applicable to aDNA studies by using short lengths of sequence data that encompassed diagnostic genetic variation in the mitochondrial genome (Robins *et al.* 2007).

Here, I extend my phylogeographic study of the *Rattus rattus* species complex in Chapter 4 to include a temporal aspect by incorporating aDNA. Specifically, I compare the ancient phylogeographic pattern in a region of the Indian Ocean with the modern one. I evaluate zooarchaeological identifications by positively identifying zooarchaeological remains of the black rat *Rattus rattus*. Ancient DNA, as it is defined here, is DNA retrieved from any remains recovered from archaeological excavations and not necessarily determined by the age of the bone. Profiles of aDNA are characteristically highly degraded and fragmented, especially in tropical environments, where ancient DNA is notoriously difficult to amplify. Moreover, in micromammals we expect smaller quantities of DNA to remain in the bone substrate compared to larger animals simply due to the small size of the bones being more amenable to decay. The methodological implications are that only very short fragments of DNA are likely to be amplified and I therefore designed my study around this limitation. For my phylogeographic study I targeted the rapidly evolving hypervariable region 1 (HVR 1) of the D-loop gene and obtained gene products with Sanger sequencing. To identify *R. rattus* I targeted a diagnostic portion of the cytochrome *b* gene and obtained the gene products with *de novo* Pyrosequencing.

MATERIALS & METHODS

Samples

In this aDNA study I aimed to examine the relationship between ancient and modern haplotypes, and to test the zooarchaeological identification of bones against genetic identities, within the *Rattus rattus* species complex in the Indian Ocean region. An extensive search for rat bones from archaeological sites throughout the Indian Ocean region, and particularly from the Indian subcontinent westward, resulted in approximately 246 bones made available for aDNA study.

Material was supplied by Anne-Laure Dijoux/Jean-Denis Vigne from Vallée Secrète, Réunion (unpubl.); Carl Phillips from Kalba, Oman (Phillips & Mosseri-Marlio 2002); Mark Horton from Unguja Ukuu and Fukuchani, Zanzibar (unpubl.); Richard Helm from Mbuyni, Chombo, Mtsengo, Panga Ya Saidi, Kenya (Helm 2000), Stephanie Wynne-Jones and Jeff Fleisher from Songo Mnara and Vumba Kuu, Tanzania and Kenya (Wynne-Jones, 2009; Wynne-Jones & Fleisher, 2010); Wim Van Neer (faunal analyst) from Fort Frederik Hendrik, ed-Dur, Shenhur, Berenike, Aqaba, Abu Sha'ar, and Dembeni (Van Neer & Gautier 1993; Van Neer & Lentacker 1996; Van Neer & Ervynck 1998; Peters *et al.* 2009); Wijerathne Bohingamuwa/Sealinks Project from Kantharodai and Manthai, Sri Lanka (unpubl.); and the Sealinks Project excavations from Fukuchani, Unguja Ukuu, Panga Ya Saidi, and Sinseme Cave (unpubl). I also visited museum and research collections to identify material in flots, sediments and faunal assemblages as outlined in Chapter 1.

The time-consuming and costly nature of aDNA studies meant that only a subset (n=63) of the collected bones were analysed (Table 5.1). Samples analysed were from: Paithan and Sanganakallu in India; Manthai and Kantharodai in Sri Lanka; ed- Dur in United Arab Emirates; Kalba in Oman; Aqaba in Jordan; Shenhur and Berenike in Egypt; Sinseme Cave, Panga Ya Saidi, Mbuyuni, Chombo and Vumba Kuu in Kenya; Songo Mnara in Tanzania; Fukuchani and Unguja Ukuu on Zanzibar; Fort Frederik Hendrik on Mauritius; and Vallée Secrète on Réunion (Figure 5.1). The East African and South Asian material is from locations within the tropics, and the North African, Near East and Arabian material is from within the subtropics.

Sample ID	Taxon ID	ID by	Principal Investigator	Site	Excavation Code/Context	GMM
¯ RAT_01	rat	H. Eager*	R. Helm	Panga Ya Saidi	PYS 10/2 F	
RAT_02	rat	E. Quintana Morales	S. Wynne-Jones	Songo Mnara	SM 10020	
RAT_03	rat	E. Quintana Morales	S. Wynne-Jones	Songo Mnara	SM 10035	
¯ RAT_04	rat	H. Eager*	D. Fuller	Manthai	SL MA 09 F1.22	
RAT_05†	rat	H. Eager*	R. Helm	Sinseme Cave	SC 10/2 203 SPIT B	
¯ RAT_06†	rat	H. Eager*	R. Helm	Sinseme Cave	SC 10/2 203 SPIT B	
RAT_07†	rat	E. Quintana Morales	S. Wynne-Jones	Vumba Kuu	VMB 15 Spit 2 Con 2 (0101-4)	
RAT_08	*R. rattus*	N. Mudida	R .Helm	Mbuyuni	MBU 01-TPO1 (11)	
¯ RAT_09	*R. rattus*	N. Mudida	R .Helm	Chombo	CHO 01-TP04 (14)	
RAT_10	*R. rattus*	E. Quintana Morales	S. Wynne-Jones	Songo Mnara	SM 10035 (S0116-22)	
¯ RAT_11	*R. rattus*	E. Quintana Morales	S. Wynne-Jones	Songo Mnara	SM 15009 (S0126-7)	
RAT_12†	*R. rattus*	J. Denis-Vigne	A.-L. Dijoux	Vallée Secrète	VS? STR2? US1 ensemble 3	
RAT_13†	*R. rattus*	J. Denis-Vigne	A.-L. Dijoux	Vallée Secrète	VS? STR2? US2 F72	
RAT_14	*R. rattus*	N. Mudida	M. Horton	Unguja Ukuu	UU1 07	
RAT_15	*R. rattus*	N. Mudida	R .Helm	Chombo	CHO 01-TP04 (13)	
¯ RAT_16	*R. rattus*	N. Mudida	R .Helm	Mbuyuni	MBU 01-TPO1 (11)	
¯ RAT_17	Rat cf. *Rattus rattus*	C. Mosseri	C. Phillips	Kalba K4	K4 2000 34.014 (229)	
RAT_18	Rat cf. *Rattus rattus*	C. Mosseri	C. Phillips	Kalba K4	K4 2000 34.014 (229)	
¯ RAT_19	Rat cf. *Rattus rattus*	C. Mosseri	C. Phillips	Kalba K4	K2A 007 upper bag 32 6 5 m	
¯ RAT_20	Rat cf. *Rattus rattus*	C. Mosseri	C. Phillips	Kalba K4	K4 32 031	
¯ RAT_21	Rat cf. *Rattus rattus*	C. Mosseri	C. Phillips	Kalba K4	K4 2000 34-009 (227)	

RAT_22	*R. rattus*	W. Van Neer	?	Fort Frederik Hendrik	FH04.151.03	
RAT_23	*R. rattus*	W. Van Neer	?	Fort Frederik Hendrik	FH04.165.13	
¯RAT_24	*R. rattus*	W. Van Neer	?	ed-Dur	AE UF2249	
¯RAT_25	*R. rattus*	W. Van Neer	?	ed-Dur	AP3 UF4163	Y
¯RAT_26	*R. rattus*	W. Van Neer	?	ed-Dur	SN305 UF2431 tomb 3846	
¯RAT_27	*R. rattus*	W. Van Neer	?	ed-Dur	SN305 UF2431 tomb 3846	Y
¯RAT_28	*R. rattus*	W. Van Neer	?	Shenhur	CH01/S1/8/sw17/4mm	Y
¯RAT_29	*R. rattus*	W. Van Neer	?	Shenhur	CH01/S1/8/sw16/4mm	Y
¯RAT_30	*R. rattus*	W. Van Neer	?	Shenhur	CH01/S1/8/F32/hand	Y
¯RAT_31	*R. rattus*	W. Van Neer	S. Sidebotham	Berenike	BE97/13-002	Y
¯RAT_32	*R. rattus*	W. Van Neer	S. Sidebotham	Berenike	BE97/13-002	Y
¯RAT_33	*R. rattus*	W. Van Neer	S. Sidebotham	Berenike	BE95/6-008	Y
RAT_34	*R. rattus*	W. Van Neer	S. Sidebotham	Berenike	BE95/5-012/PB74	
¯RAT_35	*R. rattus*	W. Van Neer	S. Sidebotham	Berenike	BE95/4/3123	Y
RAT_36	*R. rattus*	W. Van Neer	?	Aqaba	AQ08 U50 loc4 upper layer	
¯RAT_37	Rat cf. *Rattus rattus*	C. Mosseri	C. Phillips	Kalba K4	K2A 007 upper bag 32 6 5 m	
¯RAT_38	Rat cf. *Rattus rattus*	C. Mosseri	C. Phillips	Kalba K4	K2A 0007, 9 Sm mammal bones	
¯RAT_39	rodent	H. Eager*	D. Fuller	Sanganakallu	SGK 98A1 (13.2.98)	
¯RAT_40	Rat	H. Eager*	D. Fuller	Sanganakallu	SAN04 tr10(1170)<1132>	
¯RAT_41	Rat	H. Eager*	D. Fuller	Sanganakallu	SAN04 tr10(1053)<1068>	
¯RAT_42	Rat	H. Eager*	D. Kennet	Paithan	PTN 97 13.3D Ringwell Sth	
¯RAT_43	Rat	H. Eager*	D. Kennet	Paithan	PTN 97 11.3D	
¯RAT_44	Rat	E. Quintana Morales	S. Wynne-Jones	Songo Mnara	SM 10020 (S0103-7, S0110-13)	

RAT_45	*R. rattus*	E. Quintana Morales	S. Wynne-Jones	Songo Mnara	SM 15010 C (S0128-9)
RAT_46	Rat	E. Quintana Morales	S. Wynne-Jones	Songo Mnara	SM 100026 (S0115)
¯RAT_47	*Rattus* sp.	J. Perera	Sealinks	Kantharodai	KTD 2011 Pit 2 49
¯RAT_48	*Rattus* sp.	J. Perera	Sealinks	Kantharodai	KTD 2011 Pit 2 32
¯RAT_49	*R. rattus*	J. Perera	Sealinks	Kantharodai	KTD 2011 Pit 2 13
¯RAT_50	*R. rattus*	J. Perera	Sealinks	Manthai	MA 2009 phase 1
¯RAT_51	*R. rattus*	J. Perera	Sealinks	Manthai	MA 2009 Phase 4, cont 50
¯RAT_52	*R. rattus*	J. Perera	Sealinks	Manthai	MA 2009 Phase 5, Cont 37
¯RAT_53	*Rattus* sp.	J. Perera	Sealinks	Manthai	MA 2009 Phase 7, cont 59
RAT_54	*R. rattus*	J. Perera	Sealinks	Manthai	MA 2009 Phase 8 , Cont 11
¯RAT_55	*R. rattus*	J. Perera	Sealinks	Manthai	MA 2009 Phase 8, cont 26
¯RAT_56	*R. rattus*	J. Perera	Sealinks	Manthai	MA 2009 Phase 10, cont 24
RAT_57	Rat	E. Quintana Morales	S. Wynne-Jones	Songo Mnara	SM 10020 (S0103-7, S0110-13)
RAT_58	Rat	E. Quintana Morales	S. Wynne-Jones	Songo Mnara	SM 13007 (S0123)
RAT_59	Rat	E. Quintana Morales	S. Wynne-Jones	Songo Mnara	SM 100024 (S0114)
RAT_60	*R. rattus*	E. Quintana Morales	S. Wynne-Jones	Songo Mnara	SM 10002C (S0101)
RAT_61	matches *R. rattus*	M. Prendergast	Sealinks	Fukuchani	FK2011 12 003
RAT_62	Rodent Cf *R. rattus*	M. Prendergast	Sealinks	Panga Ya saidi	PYS2011 03 301 A
RAT_63	Rodent Cf *R. rattus*	M. Prendergast	Sealinks	Panga Ya saidi	PYS2011 03 305 E

Table 5.1. Archaeological bones analysed in this study, their identification and provenance. ¯ indicates sample did not amplify. † denotes three extractions from the one bone. * indicates a reference collection was not used to identify bones. Y indicates molars were sent to Ardern Hulme-Beaman for geometric morphometric (GMM) analysis. Site locations are given in the map in Figure 5.1.

Figure 5.1. Location of bone samples from archaeological excavations in the Indian Ocean region used in this study (see Table 5.1).

Zooarchaeological identifications of the analysed material ranged from ‘rat’ to species-level ‘*Rattus rattus*’. The difficulty in positively identifying *Rattus* sp. bone elements, and the partial or damaged character of many of the bones, meant that it is expected that a proportion of the bones collected would not be of the *R. rattus* species complex. Usually the material provided would have been identified through the use of a specimen reference collection, and unless known otherwise, this was assumed to be the case. This can be problematic for aDNA studies in that it increases the probability of cross-contamination. The alternative is not to use a reference collection, which makes identification more difficult but reduces the risk of cross-contamination. Material I personally collected from museums and research institutions was not identified using a reference collection (i.e. for bones analysed here this concerns Dorian Fuller’s material from Manthai and Sanganakallu archaeobotanical flots; Richard Helm’s collection from Sinseme Cave and Panga Ya Saidi; and Derek Kennet’s collection from Paithan).

Archaeological material was photographed prior to destruction, except where bones were fragmented. Where molars were intact they were sent to Ardern Hulme-Beaman for geometric morphometric analysis (Figure 5.2).

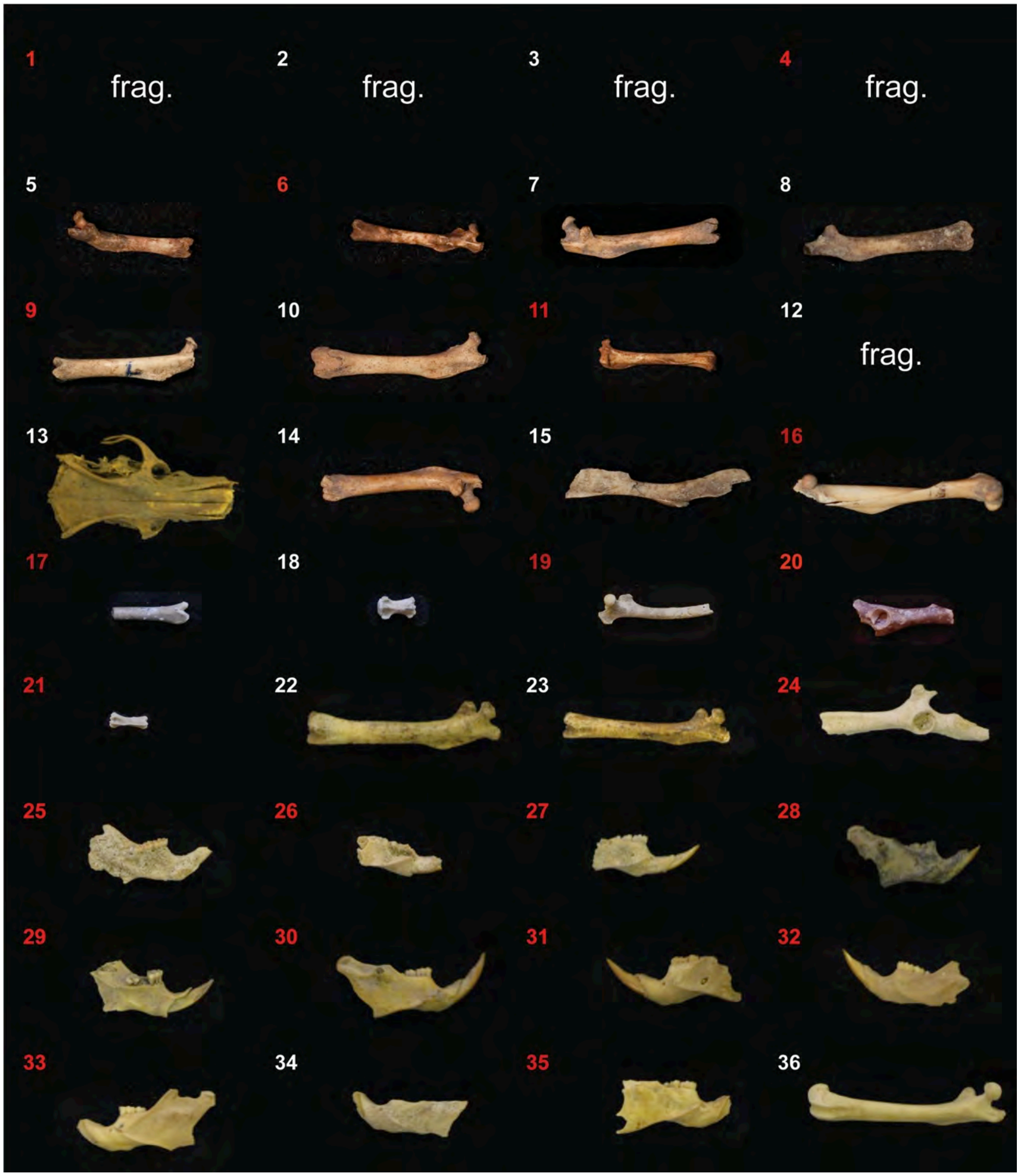


Figure 5.2. Archaeological bones used in this study (approximately to scale). Numbers refer to Sample ID in Table 1. DNA did not amplify in samples with numbers coloured red. Fragmented bones were not photographed ("Frag."). Image continued on the next page.

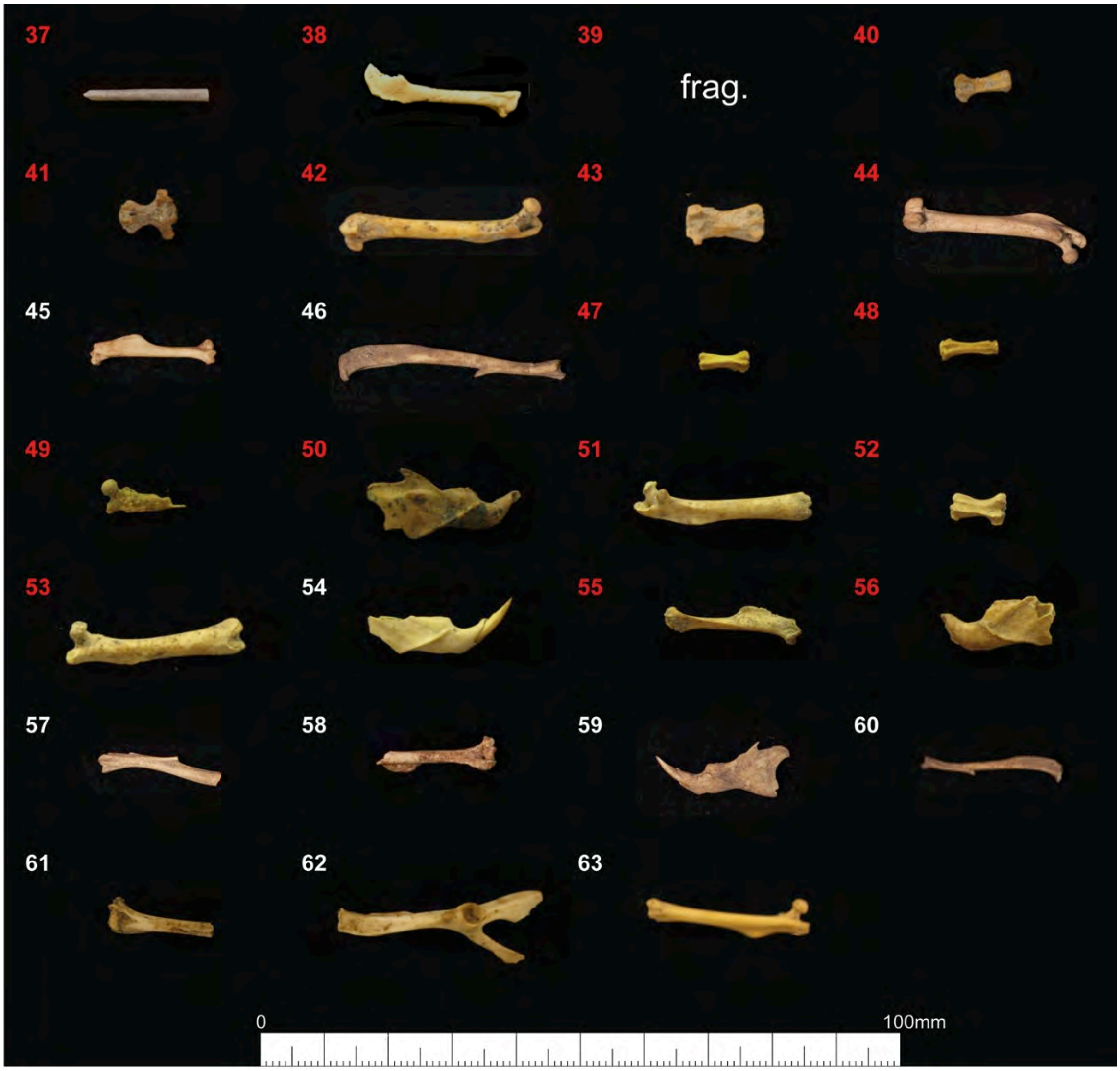


Figure 5.2 continued

DNA extraction and polymerase chain reaction (PCR)

All molecular work was conducted in the DEAD Lab, Durham University.

Bones were prepared for molecular analysis in an isolated, dedicated ancient DNA sample preparation room. All surfaces and instruments were bleached prior to and after use. Gloves were changed between handling each sample. Samples were either processed whole or a Dremel drill, with a clean, disposable sanding disc was used to cut a portion of the bone. Samples were prepared under an extractor fan on clean foil. The bone sample was placed in a canister and microdismembrated (5-10 s at 3000 rpm) until a fine powder was achieved. The powder was transferred to a 15 ml falcon tube.

Where the size of the bone allowed, I conducted multiple extractions of a single bone. Extractions were conducted in an isolated aDNA lab, separate from sample preparation. All inorganic buffers were UV irradiated for 30 min prior to use. A control per five samples or fewer was used. Two ml of buffer (85% EDTA 0.5 M, pH8; 10% Tris-HCl 1 M, pH 8; and 5% SDS 1% w/v), 20µl Proteinase K (20 mg/ml), and a small quantity (approx. 50 - 100 mg) of Dithiothreitol (DTT) was added to the bone powder. Samples were left overnight in a rotator at 56-60 °C. Samples were briefly centrifuged to sediment undissolved bone powder and the supernatant transferrred to an Amicon Ultra 30K tube. Samples were centrifuged at 4000*g* (RCF) until 100-200 µl liquid remained. The sample was transferred to a 1.5 ml Eppendorf tube and a QiaQuick PCR purification protocol was undertaken, following the manufacturer's protocol, except that the phosphate buffer eluate was poured back into the column and centrifuged again to recover lost DNA. Final elutions were 100 µl.

Generally, bone preparation for aDNA analyses involves cleaning the bone to remove contaminants by either grinding away the surface of the bone, or by bleaching the bone surface and recovering internal bone material. However, their small size prevents cleaning in small mammals as grinding removes a substantial portion of the bone and bleach penetrates the bone's interior. Therefore, small mammal bone extractions will contain contaminants, the most common of which will be soil-associated microbes and chemical substances. An additional problem is that PCR inhibitors will be carried over to the DNA extract, in many causes resulting in the inhibition of the PCR reaction and unsuccessful amplification of DNA template. Fulvic and humic acids common in soil are inhibitors of PCR probably acting through the enzymatic inhibition of DNA polymerases (Matheson *et al.* 2010). DNA extracts proved to be substantially inhibited during PCR and it is expected this was due to soil contaminants.

There are a number of molecular methodologies that can help overcome inhibition. For example, common strategies involve adding reagents to the PCR reaction that decouple the interaction between inhibitors and DNA or DNA polymerase, including bovine serum albumin (BSA), Tween 20, betaine, *Taq* polymerase, and modified DNA polymerases (Bartlett & Stirling 2003). I used all the above reagents, and in some cases a combination of the reagents, to overcome inhibition, where necessary, on a sample-by-sample basis. Shorter amplicons and higher melting temperatures can reduce inhibition in some cases (McCord *et al.* 2011). I designed primers for Pyrosequencing to amplify the smallest length of sequence and the highest melting temperature possible. However, this approach was somewhat restricted as those variable regions in mitochondrial genes that are the most useful to genotype are also the most difficult for the development of primers as they require invariant stretches of DNA.

KAPA2G (Kapa Biosystems, Boston, USA) is a modified *Taq* enzyme (with proprietary buffers) that enhances processivity and specificity of the PCR reaction thus increasing yield and specificity as well as overcoming inhibitors. A PCR trial on difficult aDNA templates using KAPA2G Robust DNA polymerase showed that this polymerase generated amplicons of appropriate size as diagnosed by a 2% agarose gel, however products could not be sequenced by either Sanger or Pyrosequencing methods.

PCR reactions were 25 µl for Sanger sequencing and 50 µl for Pyrosequencing reactions. PCR reactions contained ddH_20, 1 x buffer, 2.5-4 mM Mg, 0.2-0.3 mM dNTPs, 0.2-0.3 µM each primer, 1-2 U *Taq* DNA polymerase (AmpliTaq Gold), 0.5 µg/µl BSA and in some cases where inhibition was present 0.5% Tween 20 and 0.7M betaine. Multiple sequences from the same extracts and different extracts from the same individual (where available) were attempted. A PCR control per primer pair per reaction was employed. In some cases a modern positive control was used but the addition of modern DNA occurred in an isolated lab after processing ancient samples.

The primers used in PCR reactions for Pyro- and Sanger sequencing methods are described under their respective sections below.

Pyrosequencing

Pyrosequencing is a DNA sequencing technology that uses the sequencing-by-synthesis principle to detect the incorporation of nucleic acids during DNA synthesis (Ronaghi *et al.* 1998). It is an alternative to Sanger sequencing and the main advantage (as it relates to this study) is that it can be used for *de novo* DNA sequencing of very short (1 - ~100 bp) sequences. It relies on the quantitative detection by fluorescence of pyrophosphate (that is converted to ATP and ultimately oxyluciferin) that is released when each nucleotide is incorporated into the DNA sequence under synthesis (Chowdhury 2012).

Based on a combined dataset of published sequences and sequences generated for the modern phylogeographic study (hereafter called the *Rattus* species dataset) I identified a short length of sequence in the cytochrome *b* (cyt*b*) gene that could distinguish *Rattus rattus* I (nomenclature from Aplin *et al.* (2011)) from all other taxa within the *Rattus rattus* species complex plus other common commensals of the *Rattus* genus (the Norway rat *Rattus norvegicus* and the Pacific rat *Rattus exulans*). The assay could not distinguish between *Rattus rattus* I and the ?*kandianus* variety described in the previous chapter. Partial cytochrome *b* sequences from position 14,250 – 14,273 relative to

EU27307 from *R. rattus* isolate RNZRrTit01 (all positions with respect to it hereafter) were amplified in a PCR reaction and *de novo* sequenced using Pyrosequencing technology (Robins *et al.* 2008). The fragment length of the PCR product was 88 bp and the target length for sequencing was 19 bp. The cascade release of deoxynucleotide triphosphates (dNTPs), followed by incorporation (or not), has implications for the design of assays using Pyrosequencing technology. I designed the sequencing primer and associated release of dNTPs to strictly match a length of sequence upstream (3'-5') of the target area. Therefore, if genetic variation within the length of primer sequence occurs the released nucleotide will not be incorporated and the sequencing reaction will terminate. For my study this adds confidence to the positive identification of *Rattus rattus* based on short stretches of sequence. However, as mitochondrial DNA is evolving rapidly, mutations may also occur in *Rattus rattus* gene sequences, and this could be another reason for failure of the sequencing reaction.

At the time the assay was developed a single nucleotide polymorphism (SNP) within the region chosen for sequencing, at position 14,261, distinguished *R. rattus* I from the Sladen's rat *Rattus tanezumi sladeni.* However, since the molecular work was undertaken two sequences of *R. t. sladeni* from locations in China have proved to be identical to *R. rattus* I for the fragment generated (Conroy *et al.* 2012). Considering these sequences are rare and are from a lineage that is not widely dispersed (and not found in the area where I have ancient material) I do not consider the lineage to be of concern for my study in terms of misidentification. Furthermore, genetic variation in *R. t. sladeni* across the portion of the gene where the sequencing primer anneals would likely cause the sequencing reaction to fail.

A single-stranded DNA (ssDNA) template is required in Pyrosequencing in order to synthesize DNA by incorporation of dNTPS. Here, I use a method that attaches a biotin tag to one of the DNA strands by use of a biotinylated primer in the PCR reaction. The biotinylated double-stranded (dsDNA) PCR product is denatured in solution and the biotinylated ssDNA strand captured by streptavidin-coated sepharose beads. The biotinylated ssDNA is then used as a template for DNA synthesis in the sequencing reaction.

Primer design is one of the key features for successful Pyrosequencing. There needs to minimal background or non-specific amplification of the PCR product (i.e. avoidance of primer dimers, hairpin loops, duplexes, mispriming). I used PyroMark Assay Design software (Qiagen, USA) to develop primers with high melting temperature (T_m) inputs (61 – 66 °C) in order to counteract inhibition in the PCR reaction. The forward primer was biotinylated. Primers used in the PCR reaction were: *LinF* 5’-TCTTCTAGGAGTATGCCTTATAGT-biotinylated-3’ (T_m=61.7 °C) and *LinR* 5’-TGAGAATGCTGTTAAAGTGTC-3’ (T_m=62 °C). Primers corresponded to positions 14,231 – 14,254 and 14,298 – 14,318, respectively. The sequencing primer was *LinS* 5’-ATAGTACAAATTATCACAGG-3’, corresponding to position 14,274 – 14,293. See Figure 5.3 for a schematic of primers relative to the template sequence.

Figure 5.3 Position of primers used in Pyrosequencing relative to *Rattus rattus* EU27307 position 14,231 – 14,318 (5’-3’). Template sequence is in grey with the region *de novo* sequenced in black, polymerase chain reaction primers (green), and sequencing primer (orange) with the direction of extension shown by arrows.

PCR reactions were carried out under the following conditions: 95 °C x 5 min, then 50 cycles of 95 °C x 35 s, 58 °C x 35 s, 68 or 72 °C (depending on *Taq* requirements) x 45 s.

Products were sequenced on a PyroMark Q24 following the manufacturer's guidelines as outlined for Q24 sequencing and are briefly described here. To bind the biotinylated PCR product to the streptavidin-coated sepharose beads 10 µl of PCR product, 40 µl binding buffer, 2 µl sepharose beads and 28 µl milli-Q H_20 were agitated for 15 min at 1400 rpm. Denaturation of the dsDNA was carried out in a vacuum preparation tool apparatus. Under vacuum, the bead-bound PCR product was washed in 70% ethanol for 5 sec, followed by denaturation in NaOH for 5 sec, then washed in buffer for 10 sec. With the vacuum off, bead-bound ssDNA product was released into 25 µl primer solution (0.3 µM sequencing primer in annealing buffer) and heated on a heat block 80 °C for 5 min.

DNA template was *de novo* sequenced. In practical terms this requires the cascade release of all four dNTPs for a number of nucleotide sites determined by the user (n=19 here). For the sequencing reaction, a Q24 cartridge was filled with appropriate volumes of substrate, enzyme and dNTPs as determined by the PyroMark Q24 Analysis software (Qiagen, USA), the solution containing ssDNA and sequencing primer was introduced to the machine, and the Q24 sequencing program was run under the AQ/SQA (*de novo*) protocol. Inspection of pyrograms (analagous to chromatograms) and interpretation of results was conducted in PyroMark Q24 Analysis software. The passed peak height threshold was set to 20 and the checked peak height threshold was set to 10. The resulting sequence was aligned with the *Rattus* species dataset and their identification determined by comparison of SNPs between target sequence and known sequences.

Sanger Sequencing

In order to identify species I sequenced a portion of the cytochrome *b* gene that covered the same positions as the Pyrosequencing assay, using primers described in Chapter 4:

RrF2 5’ – TTAATCACTCCTTCATTGACCTTCC -3’ and *RrR2* 5’- AGCCGTAGTTTACGTCTCGGCAG -3’; *RrF3* 5’ – TTAACAGCATTCTCATCAGTTAC -3’ and *RrR3* 5’- GTTGCTATGACTGCAAATA -3’. Results are reported for non-*Rattus* species but are not reported for species that were positively identified as *Rattus rattus* species complex as they were subsequently used in phylogeographic analyses.

For phylogeographic inference, I sequenced the HVR1 and part of the conserved region of the D-loop to incorporate with modern and museum material in a phylogenetic tree. Primer combinations were: *RrDloopF1* 5’ -AACTACTTCTTGACAGTACATAA -3’ and *RrDloopR1* 5’ - TGGTGTATGTCTAATAACACAGA -3’; *RrDloopF2* 5’ - ACATGAATATTCTTTCATACATT -3’ and *RrDloopR2* 5’ - TTGTTGATTTCACGGAGGAT -3’; *RrDloopF3* 5’ - AGACATACACCATTAAAGTCATAA -3’ and *RrDloopR3* 5’ - GCCCTGAAGTAAGAACCAGA -3’; *RrDloopF4* 5’ -CCCATACAACTTGGGGGTGA -3’ and *RrDloopR4* 5’ - GGCATCCGAAAATTAAAAAATAC -3’.

PCR reactions were carried out as for the Pyrosequencing PCR and cleaned of unincorporated dNTPs with FastAP Thermosensitive Alkaline Phosphatase and Exonuclease I (ThermoScientific, Pittsburgh, USA) following the manufacturer’s protocol. Products were sequenced in both directions on an Applied Biosystems ABI 3730 using 3µl of template.

Sequence alignment and data analysis

DNA sequence fragments were aligned to a *R. rattus* reference (EU27307) sequence in Geneious v6.1.6. Where sequences deviated from expected variation (on visual inspection), I performed a sequence search in the GenBank nucleotide database using a BLASTn algorithm to search for highly similar sequences (National Institute of Health 2014). In some cases where sequences did not overlap within an individual, they were subjected to BLAST in multiple fragments. The sequence(s) with the highest Bit-scores are reported. *R. rattus* species complex sequence data were inspected, contigs assembled, and checked for numts as in Chapter 4.

Published D-loop sequences for where there was geographical information and those generated for my study in Chapter 4 (n=348) were aligned in Geneious v 6.1.6 using a FFT-NS-I x2 algorithm (MAFTT v7.017) and checked by eye. I reduced sequences to haplotypes in FaBox yielding 198 haplotypes (Villesen 2007). Published sequences were predominantly from Hingston *et al.* (2005); Robins *et al.* (2007); Tollenaere *et al.* (2010); Russell *et al.* (2011); and Lu *et al.* (2012). The alignment spanned positions 15,408 - 15,887 of the D-loop. Ancient DNA sequences were added after processing the modern and museum data into a haplotype alignment so that aDNA sequences could be individually parameterised in Bayesian MCMC phylogenetic inference analysis.

To test the placement of aDNA sequences, phylogenetic inference was conducted with MrBayes v3.2.1 and stationarity and convergence checking was conducted in Tracer v1.5 and AWTY, as described in Chapter 4. Briefly, a HKY+I+Γ substitution model was implemented and two independent simulations with four chains were run for 5 million generations. The tree was rooted with *Niviventer* sp. A Maximum Likelihood tree under a HKY+I+Γ model, using 1000 bootstrap replicates to test the phylogeny was generated in PhyML3.0 under the condition described in Chapter 4. All sites were used in the analysis as many of the sequences contained missing data, especially ancient sequences. Appendix 5.I shows a schematic of the input file for genetic analysis, with missing nucleotides (N) and gaps highlighted.

As sequences in my dataset were from specimens separated in time I implemented an additional phylogenetic inference model to account for uncertainty in the genealogy of non-contemporaneous sequences hereafter called the ageBEAST model (Drummond *et al.* 2002). The model scales the mutation parameters μ and θ applied to the terminal branches of tree according to the variation in the ages of the samples using Bayesian inference. XML files were compiled in BEAUTI v1.8.0 and the MCMC run conducted in BEAST v1.8.0 (Drummond *et al.* 2012). Bayesian MCMC sampling of a Bayesian skyline coalescent model was implemented with HKY+I+Γ substitutions under a relaxed lognormal clock. The temporal spacing of sequence data was parameterized with modern and museum material treated as contemporaneous, and ancient sequences individually assigned an age (Table 5.2). For archaeological bones, an age was assigned according to contextual information relating to their position *in situ.* Where a date range was all that was available, the mean was taken. There is uncertainty relating to the ages of all bones analysed here as none of the material has been directly dated. In three cases the temporal context from where the rat bone was discovered has not been described and I estimate ages according to the following: 1. the period of the site is taken as the age range of RAT_07 from Vumba Kuu 2. RAT_12 from Vallée Secrète was from a disturbed surface layer and probably modern and therefore assigned an age of 1 year before present (1 BP) 3. RAT_12 from Vallée Secrète was from the first layer below the surface (which contains anthropological remains) and I designated this pre-modern and an age of 99 BP. Modern/museum material was dated 1 BP. I assigned a lognormal prior (mean=1; standard deviation=1.25) for all age parameters. Two simulations were run from random starting trees with four chains for 50 million generations, sampled every 1000 generations. Convergence and appropriate mixing were checked by visually inspecting the trace and model parameters in Tracer v1.3, and in AWTY using the compare functions (Rambaut 2014b). A 20% burn-in was used, as determined by visual inspection of the trace file. Runs were combined with LogCombiner v1.7.5 and a median tree generated with TreeAnnotator v1.7.5.

Sample ID	Site	Period	Years (BP)
RAT_02	Songo Mnara (Tanzania)	15th -16th c. AD	562
RAT_03	Songo Mnara (Tanzania)	15th -16th c. AD	562
RAT_07	Vumba Kuu (Kenya)*	14th- 15th	662
RAT_10	Songo Mnara (Tanzania)	15th -16th c. AD	562
RAT_12	Vallée Secrète (Réunion)*	<50 BP	1
RAT_13	Vallée Secrète (Réunion)*	>50 BP	99
RAT_22	Fort Frederik Hendrik (Mauritius)	Late 17th c. AD	322
RAT_23	Fort Frederik Hendrik (Mauritius)	Late 17th c. AD	322
RAT_36	Aqaba (Jordan)	End 13th – 15th c. AD	667
RAT_45	Songo Mnara (Tanzania)	15th -16th c. AD	562
RAT_46	Songo Mnara (Tanzania)	15th -16th c. AD	562
RAT_58	Songo Mnara (Tanzania)	15th -16th c. AD	562
RAT_59	Songo Mnara (Tanzania)	15th -16th c. AD	562

Table 5.2 Dates for samples used in the MCMC BEAST analysis. * are from layers without contextual dating information.

RESULTS

Thirty-eight of the 63 bones failed to generate sequence. The negative results are probably due to a combination of low quantities of endogenous DNA and assay design. In some cases PCR amplification followed by Sanger sequencing did generate sequences that were not *R. rattus* and were probably of native rodents. Degraded DNA in the bone samples prevented successful amplification of some fragments, even in individuals where other fragments were successfully generated.

Identification

Based on Sanger sequenced cyt *b* and D-loop gene sequences, six individuals were determined to be rodents but probably not of *Rattus* species (Table 5.3). The assay was designed to positively identify *R. rattus* complex species but not other rodent species and therefore the possible species described in Table 5.3 are listed only to show that the sequence identity recovered from the BLASTn algorithm does not reveal a member of the *R. rattus* species complex. In many cases the BLASTn sequence identity for the bone sample is highly improbable, e.g. *Pseudomys oralis* the Hastings river mouse, an Australian endemic. This is probably due to a lack of publicly available gene sequences for many native taxa within the geographic region concerned, and the short sequence length analysed, in some cases.

Sample ID	**Site (country)**	**Locus (bp)**	**BLASTn ID**
RAT_08	Mbuyuni (Kenya)	Cyt *b* (44 including 10 Ns)	Sequence short and degraded. Possibly *R. rattus* but most likely another taxon.
RAT_15	Chombo (Kenya)	D-loop (149)	*Apodemus draco; Mus macedonicus; Praomys taitae*
RAT_54	Manthai (Sri Lanka)	Cyt *b* (91 including 15 Ns)	*Maxomys surifer; Taeromys celebensis*
RAT_60	Songo Mnara (Tanzania)	Cyt *b* (301); D-loop (464)	*Otomys irroratus; Pseudomys orlais*
RAT_62	Panga Ya Saidi (Kenya)	Cyt *b* (227)); D-loop (162); D-loop (149)	*Arvicanthis nairobae; Arvicanthis* sp.; *Peromyscus leucopus; Pseudomys oralis; Praomys taitae; Apodemus draco; Mastomys natalensis*
RAT_63	Panga Ya Saidi (Kenya)	D-loop (147)	*Mastomys natalensis; Praomys taitae*

Table 5.3 BLASTn identification of archaeological bone samples that were not attributable to the *Rattus rattus* species complex. The locus and number of base pairs (bp) used in the BLAST is given.

Positive identifications of *R. rattus* I (incorporating the ?*kandianus* variety described in Chapter 4) from *de novo* sequence data (Pyrosequencing assay) are shown in Table 5.4. The sequences are invariant, except for SNPs at positions 14,255 and 14,270. Only ten individuals produced sequence from the Pyrosequencing assay and this is probably due to the robust nature of the methodology; PCR product generation is stringent and sequencing by incorporation of nucleotides fails when unexpected variation is introduced into the sequence. Five of the samples identified by the Pyrosequencing assay did not produce amplicons that could be used for the phylogeographic study. This is likely due to degradation of the DNA molecule fragmenting the DNA smaller than the products required by Sanger sequencing technology.

Phylogeographic inference

Phylogenetic inference based on the HVR1 of the D-loop resulted in the polyphyly of the *Rattus tanezumi* taxon (Figure 5.4) when compared to the monophyly exhibited in Chapter 4. The clustering of sequences within more derived clades remained largely consistent between phylogenetic trees inferred from different loci for the Bayesian trees and the topology of the ML tree is consistent, except for further structure in the *R. rattus* I taxon that has low bootstrap support. The *R. rattus* I taxon was monophyletic and highly supported in the Bayes phylogeny (PP=0.99) but ML bootstrap support was low (Prob=26), probably due to the effect of low divergence levels and missing nucleotides on branch length and topology. Although branch support differs between Bayesian and ML trees, the Bayesian tree makes phylogenetic sense in that sequence affinity is the same as in Chapter 4, and ancient sequences cluster with modern sequences from the same location (where both occur). I therefore refer to the Bayesian tree hereafter when describing the relationship of ancient sequences with other known sequences.

The ageBEAST model forces a bifurcation where the MrBayes model would infer a polytomous node and therefore the tree topologies differ between the two models

(Figure 5.5). Most of the branches in the ageBEAST model are poorly supported (PP<0.5 and in many cases PP=0) and this reflects the polytomy inferred by the MrBayes model. Overall, the average to well-supported clades in the MrBayes model are retained in the ageBEAST model e.g. the clade containing the two ancient Réunion sequences (REU_12 and REU_13) contain the same taxa in both models and have PP=0.9 for the branch leading to the clade. All reports are based on the MrBayes tree unless otherwise stated.

Although *R. tanezumi* is found in present-day Zambia and South Africa it has not yet been found on the coast of East Africa. I found a mitochondrial lineage of *R. tanezumi* from Songo Mnara, a 15th - 16th century AD town on the coast of Tanzania. Three sequences (RAT_02, RAT_46, RAT_59) were attributable to *R. tanezumi* and formed a monophyletic clade (PP=1.0). There is only one polymorphism between the three sequences (calculated from where there is corresponding nucleotides i.e. no Ns). The Songo Mnara *R. tanezumi* cluster in a clade with individuals from Java, Sulawesi, Malaysia and the Cocos (Keeling) Islands. All four remaining rats from Songo Mnara (RAT_03, RAT_10, RAT_45, RAT_58) were *R. rattus* I and were unresolved (i.e. polytomous nodes) within the clade. There was low sequence variation between all four sequences; each had a singleton variation.

RAT_07 from Vumba Kuu on the Kenyan coast clusters with a haplotype found on the East African coast in the modern period in Tanga, Tanzania and Kilifi, Kenya (PP=0.92). The group is nested within a clade (PP=0.74) with two ancient sequences from Fort Frederik Hendrik on Mauritius (RAT_22 and RAT_23) plus modern haplotypes from Mauritius, Tanzania, Kenya, Mozambique, Madagascar, Grande Comore and India. This clade is a constituent part of the C clade described in Chapter 4. The two Fort Frederik Hendrik sequences differ by a singleton polymorphism (based on where sequence coverage corresponds i.e. no Ns).

The two sequences from Vallée Secrète on Réunion (RAT_12 and RAT_13) cluster with Haplotype 39 from Tolleneare *et al.* (2010), also from Réunion, and all three sequences are identical where there is parallel sequence coverage (PP=0.67). They are nested within a clade (PP=0.9) that is of the A type rats described in Chapter 4.

The sequence from the Aqaba site on the Red Sea coast of Jordan clusters with modern sequences solely from Yemen (pp=0.88). In Chapter 4 these haplotypes were within the C clade, however here the C clade is polyphyletic.

Sample ID	Site	Taxon ID	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
RAT_03*	Songo Mnara	*Rr* I	C	C	A	A	A	T	T	A	T	C	A	C	A	G	G	C	T	T	A
RAT_05*	Sinseme Cave	*Rr* I	T	C	A	A	A	T	T	A	T	C	A	C	A	G	G	C	T	T	A
RAT_07ex1	Vumba Kuu	*Rr* I	T	C	A	A	A	T	T	A	T	C	A	C	A	G	G	C	T	T	A
RAT_07ex2	Vumba Kuu	*Rr* I	T	C	A	A	A	T	T	A	T	C	A	C	A	G	G	C	T	T	A
RAT_07ex3	Vumba Kuu	*Rr* I	T	C	A	A	A	T	T	A	T	C	A	C	A	G	G	C	T	T	A
RAT_10	Songo Mnara	*Rr* I	C	C	A	A	A	T	T	A	T	C	A	C	A	G	G	C	T	T	A
RAT_13ex1	Vallée Secrète	*Rr* I	T	C	A	A	A	T	T	A	T	C	A	C	A	G	G	A	T	T	A
RAT_13ex2	Vallée Secrète	*Rr* I	T	C	A	A	A	T	T	A	T	C	A	C	A	G	G	A	T	T	A
RAT_13ex3	Vallée Secrète	*Rr* I	T	C	A	A	A	T	T	A	T	C	A	C	A	G	G	A	T	T	A
RAT_34*	Berenike	*Rr* I	T	C	A	A	A	T	T	A	T	C	A	C	A	G	G	C	T	T	A
RAT_45	Songo Mnara	*Rr* I	C	C	A	A	A	T	T	A	T	C	A	C	A	G	G	C	T	T	A
RAT_57*	Songo Mnara	*Rr* I	C	C	A	A	A	T	T	A	T	C	A	C	A	G	G	C	T	T	A
RAT_58	Songo Mnara	*Rr* I	C	C	A	A	A	T	T	A	T	C	A	C	A	G	G	C	T	T	A
RAT_61*	Fukuchani	*Rr* I	T	C	A	A	A	T	T	A	T	C	A	C	A	G	G	C	T	T	A

Table 5.4 Sequence data and taxon identification of ancient rat material from the *de novo* Pyrosequencing assay. Positions 14255-14273 (53-71) of the cytochrome *b* gene relative to EU27307. Polymorphisms between sequences are coloured. * indicates the sample was not used in parallel phylogeographic analyses. Three DNA extractions of each of RAT_07 and RAT_13 were analysed.

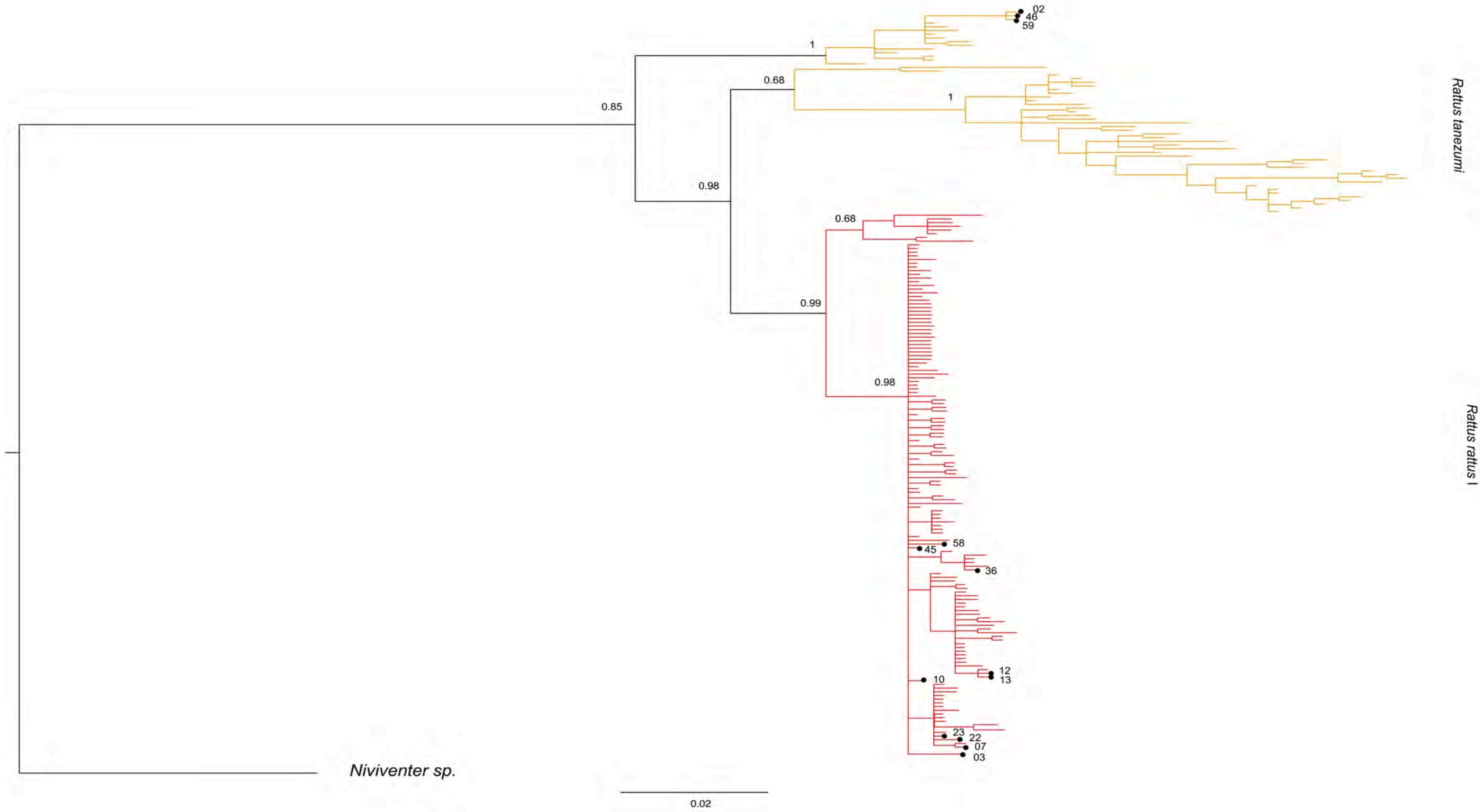


Figure 5.4. MrBayes tree **(TOP)** and Maximum Likelihood tree **(BOTTOM)** of HVR1 of the D-loop rooted with *Niviventer* sp. for *Rattus rattus* I (red) and *Rattus tanezumi* (orange) with ancient DNA samples highlighted by black dots on the terminal branches. Numbers correspond to Sample "RAT" IDs (Table 5.1). Posterior probabilities or bootstrap values are shown along the branch.

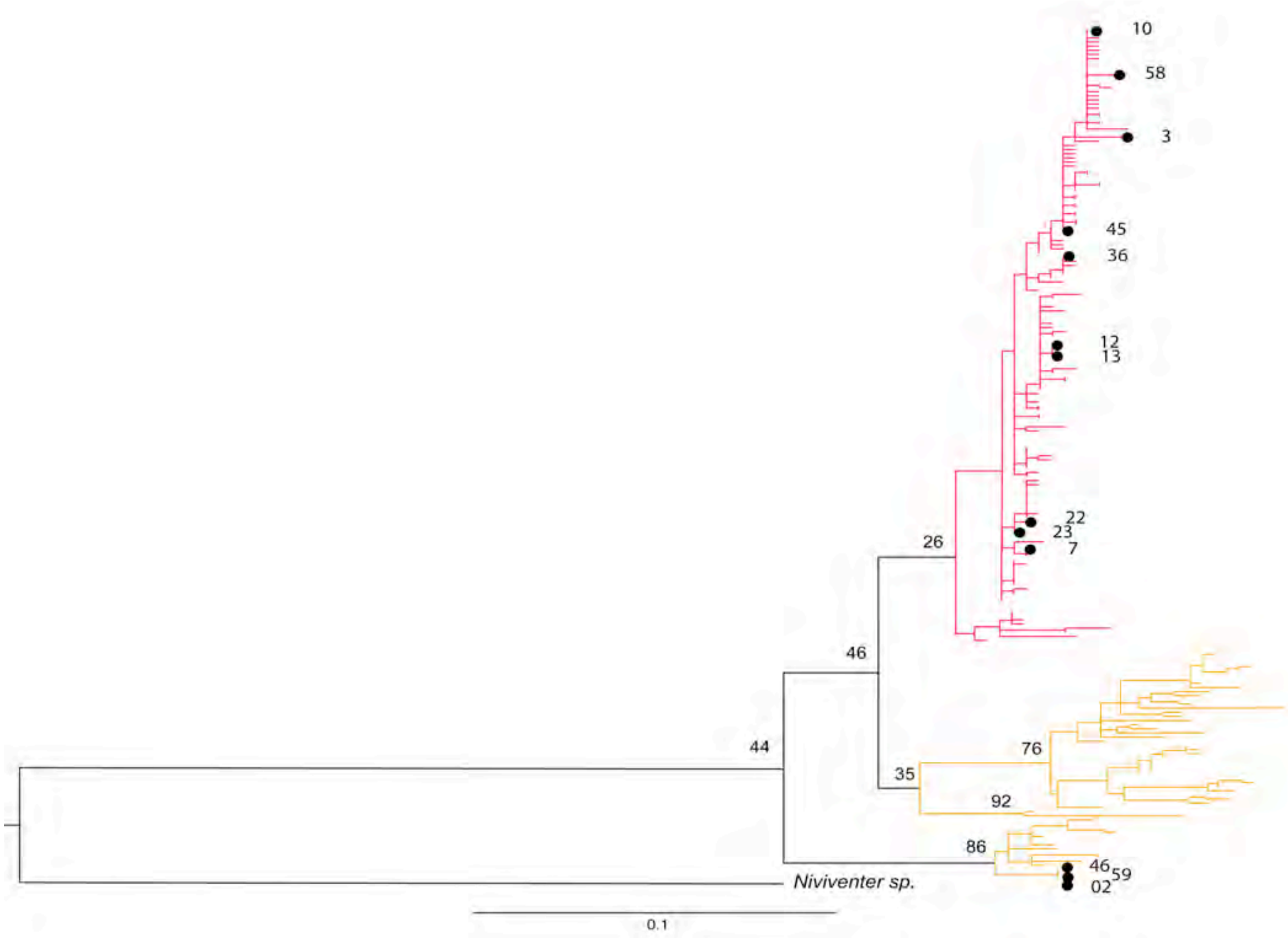


Figure 5.4 continued

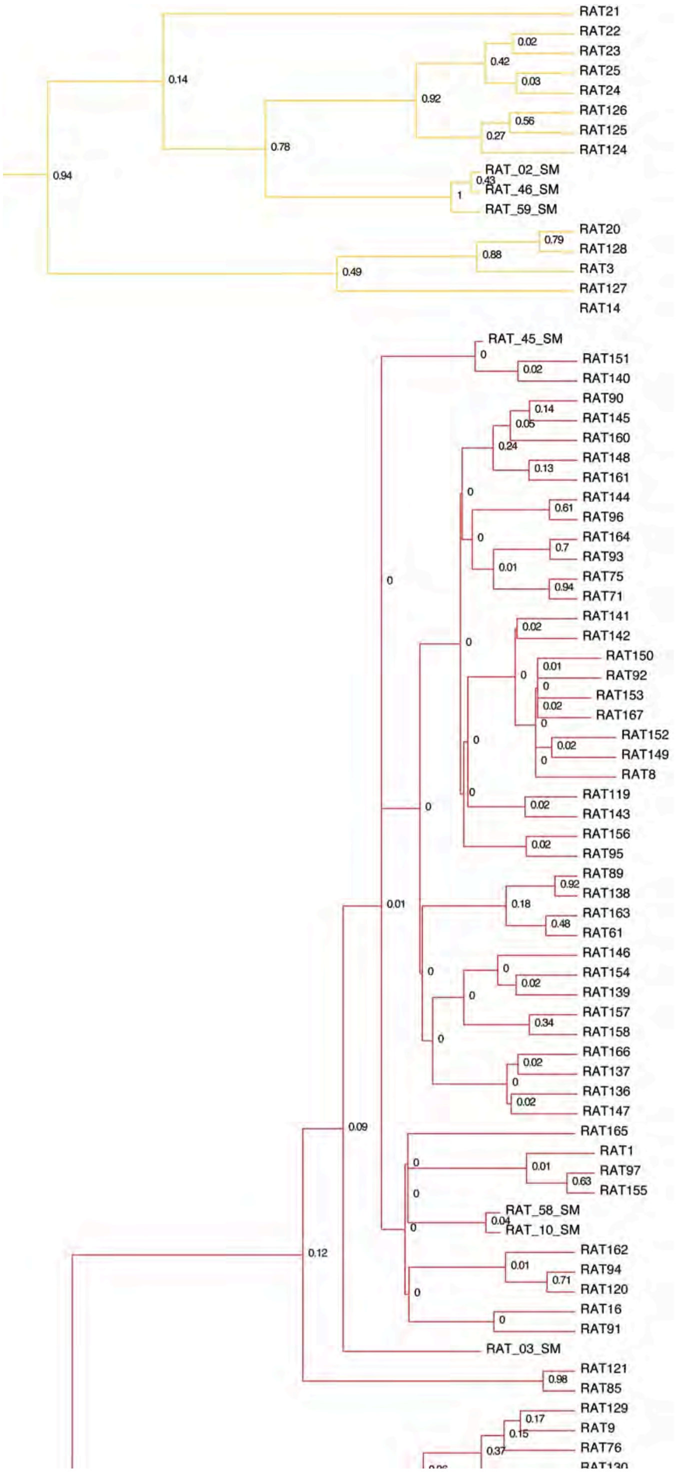


Figure 5.5. Bayesian MCMC phylogenetic inference with age of samples parameter (BEAST). Only *Rattus rattus* I (red) and *Rattus tanezumi* (orange) clades containing ancient rats are shown. Note the low posterior probabilities of many branches indicative of polytomous nodes. Image continued on the next page.

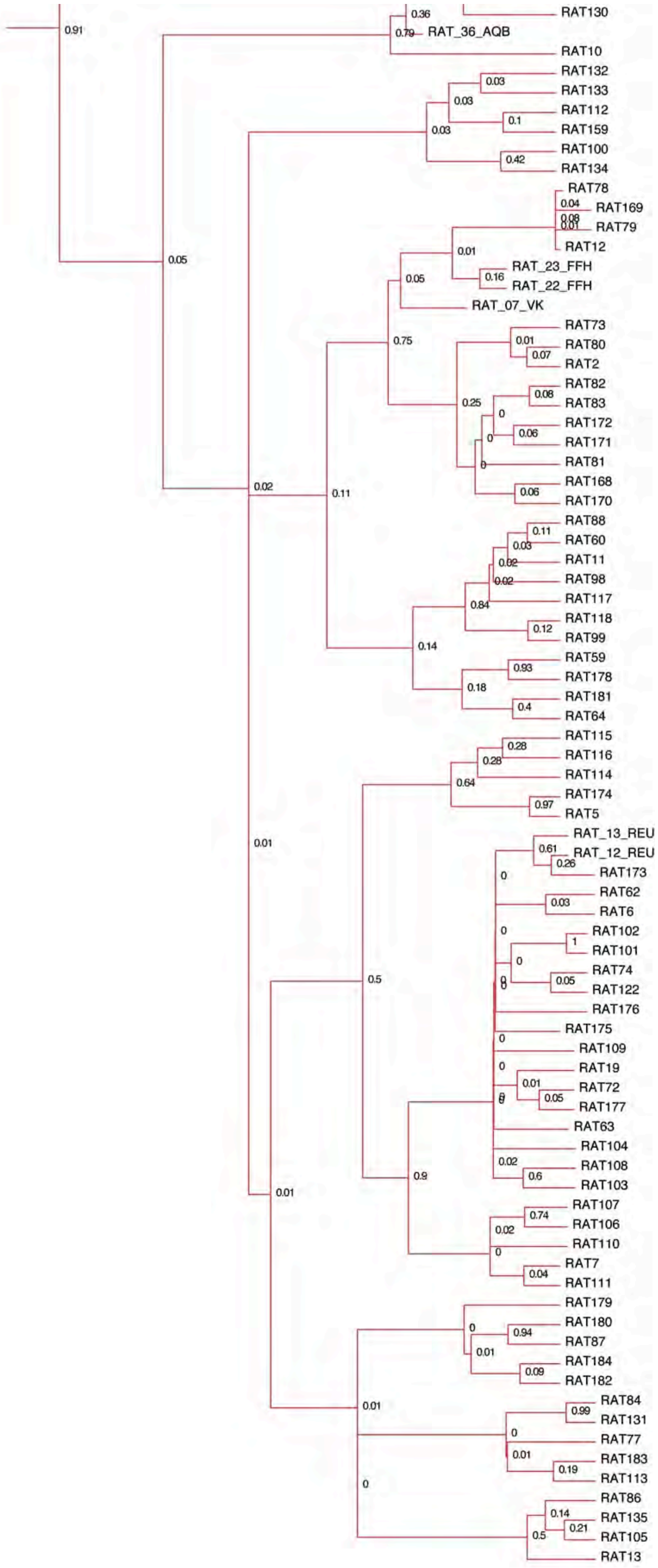


Figure 5.5. continued

DISCUSSION

The dearth of positively identified micromammals at the species level (or even the genus level), and the sometimes erroneous taxon identification of bone (or other preserved) elements in the literature, contributes to the uncertainty surrounding the earlier distributions of micromammals (Robins *et al.* 2007). This has significant implications for the interpretation of the colonisation and population genetic histories of a species that may have been transported by humans. The lack of diagnostic features on many micromammal bone elements, the often fragmented or decayed nature of the bone, the lack of expertise or interest in micromammal identifications, plus many other factors contribute to the problem. Modern techniques using ancient DNA recovered from subfossil bone or other preserved elements can help overcome this historical impasse.

Here, I positively identify bone from several archaeological excavations to the *Rattus rattus* I taxon. Three specimens, RAT_05 from Sinseme Cave (Kenya), RAT_34 from Berenike (Egypt) and RAT_61 from Fukuchani (Zanzibar) produced data suitable for species identification but not for phylogeographic inference (probably due to DNA degradation). These were identified as *R. rattus* I and are discussed below. The presence of *R. rattus* I on several other sites will be discussed later in relation to phylogeographic inference.

Sinseme Cave (the site of Panga ya Mwandzumari) is a rockshelter situated on the coast of Kenya. The Sealinks Project undertook excavations at the cave site in 2010 and the bone analysed here was recovered from sediments taken from Trench 2, which was situated outside an entrance to a small open chambered cave (Helm *et al.* 2012). There were no clear occupation layers within the trench, but pottery sherds found in layer 203A date the context to the late first/early second millennium AD and the bone was located below this layer (203B) (Helm *et al.* 2012). The DNA profile of the sample was highly degraded, as indicated by the difficulty in amplifying small DNA fragments, and this is consistent with what is expected of an older sample. Elsewhere on the eastern African coast, the black rat is first described from Unguja Ukuu on Zanzibar *ca.* 6th- mid 8th century AD and is first documented on the mainland in Chombo, Kenya, *ca.* late 8th-10th century AD (Helm 2000; Boivin *et al.* 2013b). However, the identification of *Rattus rattus* at Chombo is probably invalid as I analysed the assemblage in this study and found at least one of the bones not to be a *Rattus* form. The Sinseme Cave result places the black rat on the mainland coast of East Kenya in the earliest period of its presumed occupation of East Africa. However, the specimen from Sinseme Cave that yielded a DNA sequence has not been directly dated and there could be timing errors due to stratigraphic uncertainties. The genetic identity of the bone is not likely to be an issue as it was identified without the use of a reference collection and therefore the risk of cross-contamination is low.

Fukuchani is an archaeological site on Zanzibar that was excavated by Mark Horton in 1992 and again by the Sealinks Project in 2011 (Crowther *et al.* 2013). The chronology of the site has not been confirmed by direct dating methods, however the presence of Tana tradition pottery indicates a Middle Iron Age occupation in the latter half of the first millennium AD (Crowther *et al.* 2013). The rat bone analysed here was recovered from a shell midden layer (FK12 003) attributed to Phase 3 (the middle occupation period) in the 2011 excavations, and identified by Mary Prendergast as “matches *Rattus rattus*”. The degraded DNA profile of the bone suggests it is contemporaneous with the site, or at least not a modern incursion. Current evidence (the genetic identification of the bone along with the documentation of *R. rattus* at Unguja Ukuu by the late 8th century AD) supports a Middle Iron Age introduction of rats to Zanzibar that is probably through connections with the Indian Ocean world, perhaps via mainland East Africa (see Chapter 4).

Berenike, a port on the Red Sea coast of Egypt near Sudan, was founded by the Ptolemies and in the Roman period was a trade emporium linking the Red Sea with Alexandria and Rome. Thirteen rat bones from Berenike were provided by Wim Van Neer from excavations conducted in 1995 and 1997 that were dated from contextual information from the Ptolemaic to the early 6th century AD (Van Neer & Lentacker 1996; Van Neer & Ervynck 1999). I extracted DNA from five but was only able to amplify a small fragment in one of them. This is probably due to poor DNA preservation and was somewhat unexpected as bone remains appeared relatively well preserved (see Figure 5.2, RAT_31 - RAT_35). The bone analysed here is from a structure that was probably a warehouse near the ancient harbour and is dated to the 4th/early 5th century AD (Van Neer & Lentacker 1996). Black rats frequently occur in Berenike from the Ptolemaic period on, and from another Red Sea port, Myos Hormos, in the Roman period, however no genetic validation of zooarchaeological remains had, thus far, been made (Driesch & Boessneck 1983; Van Neer & Ervynck 1998, 1999).

In some cases, sequence data from species that were variously identified by zooarchaeologists as“rat”, “cf. *Rattus rattus”* or “*Rattus rattus”* did not match with *Rattus rattus* I (Table 5.3). In most instances there was enough sequence data to be confident that the specimen was not from the *Rattus rattus* species complex. I will not go in to detail here regarding species identification as the assay was not designed to identify other than the *Rattus rattus* species complex. DNA of RAT_08 (Mbuyuni) was severely degraded. Sequence data for the individual was shallow and unreliable (poor sequence quality, multiple peaks indicative of possible deamination) across the portion of the gene containing identifying single nucleotide polymorphisms and therefore I reserve my judgment as to whether this may or may not have been *Rattus*. A point to note is that where clean sequence was recovered from RAT_08, it contained nucleotide variants different from those described in the *Rattus rattus* species complex. Most of the bones for which I recovered sequence that did not belong to the *Rattus rattus* species complex came from East Africa. Given the limited number of reported black rat finds from East Africa, each identification contributes significantly to the interpretation of ancient distributions and equally, misidentifications will substantially mar interpretations. This dilemma exists for many regions where the black rat is introduced.

Although *R. tanezumi* is found in present-day Zambia and South Africa, where it was presumed to be a recent introduction, it has not yet been described from the coast of East Africa (Bastos *et al.* 2011). The genetic identification of three individuals of *R. tanezumi* from Songo Mnara, a 15th – 16th century town on the Kilwa coast of Tanzania is therefore unexpected. All three bones are from a trench located in the back room of House 44 in the deepest layers (but within different contexts) and therefore are probably contemporaneous with its earliest occupation (Wynne-Jones & Fleisher 2010). There is contextual reliability of the finds as an internal wall from the house structure had fallen, sealing the archaeological deposits beneath it. *R. rattus* I identified from the pyrosequencing assay is found in the same context as *R. tanezumi* (RAT_57 and RAT_02 in context 10020) which indicates the lineages are both synchronous and syntopic. It is possible this population consists of hybrids as hybridisation is known to occur in extant populations (Lack *et al.* 2012). *R. rattus* I (RAT_03 and RAT_10, although these bones may belong to the same individual) is also found in midden deposits beneath the floor of the house in association with shell and other fauna, ceramics and beads.

The majority of *Rattus* species bones excavated at Songo Mnara come from House 44 as detailed above but less frequently they were found in trenches described as the steps of a courtyard, the base of "monumental steps" leading to House 23, and in the central room of House 23 (Wynne-Jones & Fleisher 2010). Two bones were analysed from these contexts: RAT_58 from the base of the monumental steps in the active cultural layer; and RAT_45 from the central room in a midden layer composed mostly of charcoal and ash that was situated below the earth floor levels. Both bones were identified as *R. rattus* I.

Mitochondrial genetic diversity is low in the *R. rattus* I lineage (Aplin *et al.* 2011). The presence of at least three *R. rattus* I haplotypes at Songo Mnara (albeit only distinguished by singleton nucleotide variants) therefore suggests moderate diversity in ancient East Africa due to either *in situ* diversification after some period of colonization, or introduction of multiple propagules from one or a few localised source populations. Variable missing nucleotides in the ancient rat sequence data prevented calculation of diversity indices for the Songo Mnara population. With the addition of *R. tanezumi,* the ancient rat population of Songo Mnara appears genetically diverse and, based on current knowledge, more so than modern populations from the East African coast, perhaps because of an intervening population bottleneck.

Songo Mnara was a Swahili town operating within the Kilwa trade network of the early to mid second millennium AD that reached as far afield as Arabia, Southwest Asia, India and China (Chittick 1974; Wynne-Jones 2007). With the multiplicity of international connections it is perhaps not surprising to see diversity within the *Rattus* population. *R. tanezumi* from Songo Mnara cluster with modern mitochondrial haplotypes found in Indonesia and Malaysia, and the Cocos Islands (which are probably derived from Indonesia).

Vumba Kuu, Kenya, was a Swahili settlement on the East African coast slightly earlier than Songo Mnara, dated to the 14th-15th centuries AD and abandoned in the 16th century AD (Wynne-Jones 2009). There is no contextual information for the bone analysed here and it is possible it is a recent assemblage or a later intrusion. It clustered with modern haplotypes from the same region in phylogenetic analysis and, assuming it is from a contemporaneous deposit (14th-15th century AD), it shows continuity of the haplotype through time and thus population stability. Interestingly, Vumba Kuu and Songo Mnara were occupied around the same time, yet with regards the indication they give of the genetic variation of black rats in East Africa they suggest independent introductions from outside sources rather than a coastal migration within East Africa.

Rattus rattus I identified through genetic analyses from Vallée Secrète on Réunion and Fort Frederik Hendrik on Mauritius also show continuity of haplotypes through time. The ancient Réunion rats are identical to the modern haplotype reported by Tollenaere *et al.* (2010), and may in fact be relatively recent assemblages. RAT_12 was from a disturbed layer of topsoil and is probably modern and, although RAT_13 is from a layer with historical artefacts, its remarkable preservation suggests it too is not considerably old (see Figure 5.2) (Anne-Laure Dijoux pers. comm.). The fauna described from Fort Frederik Hendrik is from a refuse layer rich in animal remains, particularly those of introduced species, and is dated to the end of the 17th century AD during the second period of Dutch occupation (Peters *et al.* 2009). Black rats were present on the island prior to the Dutch first landing on the island in 1598 AD and may have been introduced by Arab or Portuguese mariners either through direct contact with the island or due to a shipwreck (Peters *et al.* 2009; Cheke 2010). There are no recorded extinctions of rats on the island in the Dutch period and the haplotype recovered in both the ancient Fort Frederik Hendrik assemblage and in modern populations probably represents the early colonising form. The Mauritian haplotype clusters with Clade C rats described in Chapter 4 - it is the strictly Indian Ocean clade that mainly incorporates India, East Africa and the Arabian Peninsula.

Aqaba is an ancient port city that lies on the Red Sea coast of Jordan. Rats analysed here are dated to *ca.* 13th – 15th century AD, during the Islamic period of the town's history (Wim Van Neer, unpubl.). The ancient Aqaba sequences cluster with modern haplotypes solely from the Red Sea region of Yemen (San'a and surrounds). Phylogenetic inference incorporating the cytochrome *b* gene places the Yemen clade within clade C (Chapter 4), however this lineage becomes polyphyletic in the aDNA analysis. The modern phylogeography shows that rats from the Mediterranean coast of Jordan are grouped in the widespread A clade but I had no examples from the Red Sea coast of the country. Moreover, I do not have ancient samples from the Mediterranean coast. Therefore, it cannot be determine whether replacement of the ancient haplogroup may have occurred (the modern A clade replacing the ancestral C clade), or whether haplotypes are maintained and the signal reflects ancient diversity within Jordan. The ancient Aqaba sample groups within an Indian Ocean network (clade C) that probably reflect its ties to the Arab world in the Islamic period. The emerging picture for clade C is one of a network involving the Red Sea and Arabian Peninsula, India, East Africa and islands of the Indian Ocean that was responsible for translocating rats in the early second millennium AD. This network may also reflect earlier trading systems of the pre-Islamic period.

Temperature, and to a lesser extent pH and availability of water, contribute to hydrolitic depurination of the DNA molecule, which is a major factor in its degradation (Smith *et al.* 2003). Therefore, the expectation is that hotter, wetter climates will present less success of DNA amplification and, in practice, successfully amplifying aDNA from tropical or subtropical environments has proven notoriously difficult. Very few studies have reported amplified aDNA from tropical environments. Two studies report aDNA on subfossil lemurs from Madagascar, although it is noted that several of the samples came from outside the tropical zone (Karanth *et al.* 2005; Orlando *et al.* 2008).
A further study amplified aDNA from rodent jawbones from Loltún, a humid tropical cave in the Yucatan peninsula (Gutiérrez-García *et al.* 2014). Here, I present the results of successfully amplified aDNA from a number of tropical and subtropical sites throughout the Indian Ocean region, in particular East Africa, which form an important contribution to understanding ancient diversity in the tropics. Ancient DNA was retrieved from both cave and open-air sites, including building structures.

In the study reported in this chapter I have incorporated a temporal aspect to the genealogical history of black rats in the Indian Ocean through the analysis of aDNA. A couple of considerations regarding the accuracy of information gleaned from genetic analysis of ancient remains should be considered here. Firstly, the context in which the bone is deposited is not necessarily contemporaneous with the bone itself. Secondly, the DNA sequence analysed may not necessarily accurately reflect the original DNA profile due to degradation of the DNA molecule *post mortem* (see Chapter 1). Black rats are scansorial and nest above ground-level. They are therefore not a serious concern in terms of intrusion into archaeological deposits through burrowing, as many other micromammals are. However, the small size of rat bones means they can filter through sediment in to deposits due to bioturbation or other physical processes. Radiocarbon dating is necessary to determine with confidence the age of rat bones analysed here. . For rats (and other micromammals) this poses a serious limitation as the size of some bone elements are often too small to conduct both analyses on. However, in many cases there is material remaining allowing me to both genetically characterise and date a single element. Ancient DNA material for which bone remains (RAT_02; RAT_05; RAT_07; RAT_10; RAT_45; RAT_46) was sent for accelerator mass spectrometry radiocarbon dating, however, the results were not received before the completion of my thesis. Molecular controls limited the error associated with *post mortem* degradation of DNA. I confirmed identical coverage across sequences that were variously a) extracted more than once b) generated from multiple PCRs on different days and c) generated from overlapping fragments. Where unresolved dual peaks did occur in sequence data (indicating deamination) I coded them accordingly so they did not unduly influence phylogenetic inference.

CHAPTER 6

PHYLOGEOGRAPHY OF THE HOUSE MOUSE *MUS MUSCULUS* WITH A FOCUS ON THE WESTERN INDIAN OCEAN, AND REPORT OF A NEW LINEAGE FROM THE INDIAN SUBCONTINENT

INTRODUCTION

The house mouse *Mus musculus* first showed evidence of commensalism with humans in the Near East around 12,000 years ago (Auffray *et al.* 1988). It has since become one of the most widely distributed commensal species on earth, in large due to its ability to exploit, and form large populations, in anthropogenic environments, and because it can stowaway in transport and cargoes unnoticed. A useful corollary of this relationship with people is that the house mouse can be used as a proxy to reconstruct human movement and migration patterns (Bonhomme & Searle 2012; Jones *et al.* 2013). Although a well-studied system in many respects, particularly within the biomedical sciences, research relating to the geographical distribution and natural genetic variation of the species is at a relatively early stage.

The house mouse is a morphologically variable species and is globally distributed. Taxonomists have historically attributed taxonomic status to populations of the house mouse based largely on morphological and geographical differences resulting in over 123 subspecies or alternative species names, which are now mostly viewed as dubious (Musser & Carleton 2005). The house mouse is currently considered to be a single species *Mus musculus* (Auffray et al. 1990; Auffray and Britton-Davidian 2012). Three main, wide-ranging, subspecies are now recognised: the eastern house mouse *M. m.*

musculus; the south eastern house mouse *M. m. castaneus*; and the western house mouse *M. m. domesticus*. A further subspecies, the south Arabian house mouse *M. m. gentilulus* is known from Yemen and has been introduced to Madagascar (Auffray and Britton-Davidian 2012). Genetics has been influential in the consideration of other populations, for example, the Japanese house mouse *M. m. molossinus* has recently been shown to be a complex hybrid of the three main subspecies based on a variety of genetic markers (Nunome *et al.* 2010), and based on mitochondrial data, a *M. musculus* form from Afghanistan, once labelled *M .m. bactrianus,* is now subsumed under the *M. m. castaneus* taxon (Yonekawa *et al*. 2012; Rajabi-Maham *et al.* 2012). As genetic research relating to unstudied or understudied geographical populations accumulates, particularly in the native range, I expect that genetic variation akin to the differences seen between subspecies will be seen, and this may already be evident in a potential taxon from Nepal identified through mitochondrial DNA (Suzuki *et al.* 2013).

The role of the house mouse as a biological proxy for long-distance movements of humans has been chiefly exploited within the western house mouse subspecies *M. m. domesticus* (Bonhomme & Searle 2012; Jones *et al.* 2013). Mitochondrial D-loop data has identified six main clades that are associated with various human cultures over different time periods, including Iron Age links between the Mediterranean and Central Europe, Viking Age connections between the British Isles, Scandinavia (and Iceland), and more surprisingly the Madeiran archipelago in the Atlantic Ocean, and Age of Discovery translocations that extended as far as Australia (Forster & Al. 2009; Searle *et al.* 2009; Bonhomme *et al.* 2011; Jones *et al.* 2012). Other subspecies are also useful proxies and of particular interest to this study are the Madagascar connections with the Arabian Peninsula in the Islamic period inferred from the presence of the *M. m. gentilulus* lineage on the island (Duplantier *et al.* 2002).

In this study, I further investigate the distribution and genetic diversity of house mice with foci on the Indian subcontinent as the native area of the species, and introductions to Africa, Arabia and the Western Indian Ocean, including the islands of Madagascar, Réunion, Socotra and the Maldives. My study is based on samples mainly collected as museum osteocrusts (dried flesh left on the bone after skull cleaning) and small pieces of tissue from skins over 100 years old and as such gives an indication of genetic variation prior to the emergence of the much more intensive networks that have built up over the last 50 years of globalisation. While house mouse populations may be subject to repeated population turnover and replacement at the local level, at the regional level they are remarkably persistent. Therefore, using genetic data from relatively recent specimens is most likely able to reflect ancient and historical signatures of the species at a regional level. Based on numerous previous studies and building on available databases, I sequence the mitochondrial D-loop gene.

MATERIALS AND METHODS

Specimens

I sampled morphologically identified house mice *M. musculus* osteocrusts and skin pieces from museum collections described in Chapter 3. Additional material was supplied by various researchers and institutions: 48 samples (in ethanol) from Réunion were supplied by Michel Pascal; nine samples from Madagascar and Tanzania (in dimethyl sulphoxide saline solution (DMSO)) were supplied by the FMNH in collaboration with Jean-Marc Duplantier and Steven Goodman; and Jean-Marc Duplantier also supplied a tissue sample (in ethanol) from Madagascar. A total of 288 mice from localities within Eurasia, and the Indian and Pacific Ocean regions were collected (Appendix 6.I).

The specimens were divided geographically into Western Indian Ocean (Africa, Arabia, South West Asia, and island Indian Ocean) and Eastern Indian Ocean/Pacific (China, mainland and island South East Asia, Australia, and the Pacific). This division was employed in order to support two separate phylogeographic analyses by different researchers (AT and myself), and based on the hypothesis that there were separate eastern and western networks (as they relate to human history) in the ancient and historical periods that were responsible for the translocation of mice from one place to another. All molecular work and data analysis was conducted independently by AT and myself, except for an area of overlap in and around the Indian subcontinent (the putative origin) where samples were processed by both researchers and used in both analyses as appropriate. Samples used in this study are identified in Appendix 6.I and Figure 6.1.

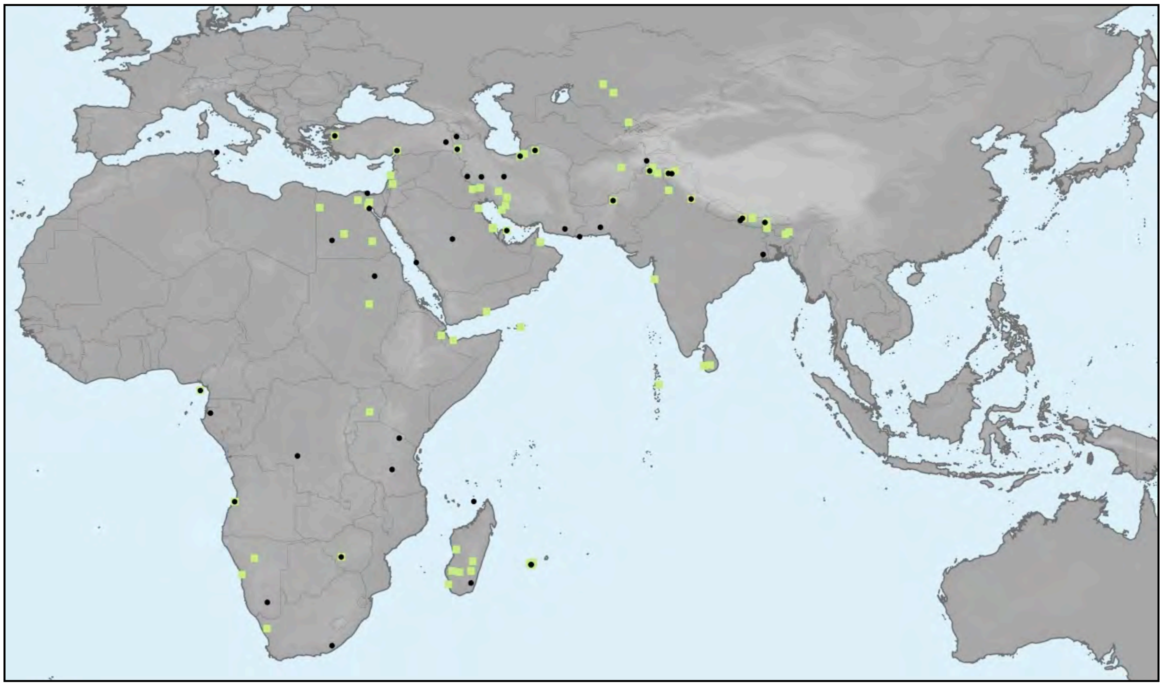


Figure 6.1. Locations of samples used in this study (excluding those analysed by AT). Green squares=sequence generated and used; black dots= samples not used (see Appendix 6.I for a description of why sequences were not used).

I collected bones from archaeological deposits with the intention of conducting ancient DNA (aDNA) analysis on them (Appendix 6.II). House mice remains are rarely recovered from archaeological excavations, and the small size of mouse bones means that not only is the entire bone required for DNA extraction but also the probability of extracting DNA is low (see Chapters 1 and 5, and Figure 6.2). Therefore, these valuable bones were left intact for non-destructive analysis (e.g. geometric morphometric). Where molars were present in mandible elements, they were sent to Thomas Cucchi for geometric morphometric analysis. I intend to conduct aDNA analysis on the bones in the future, once I have an established protocol for maximising DNA extraction from minute elements.

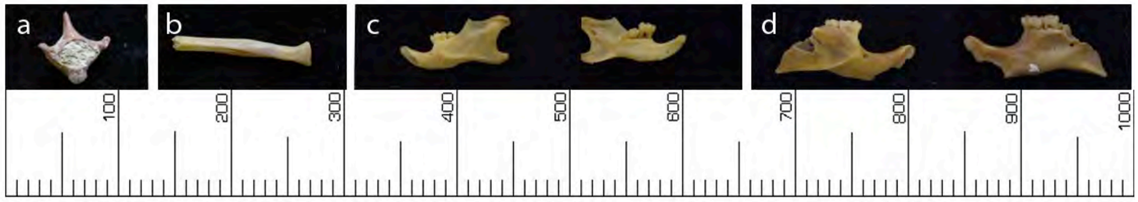


Figure 6.2. Examples of archaeological mouse elements collected and a scale bar (mm) to indicate their small size (N.b. Image is enlarged). **a** Vertebra of *Mus* sp. (with substrate embedded) from Kantharodai, Sri Lanka (Mouse07); **b** radius of *Mus* sp. from Kantharodai, Sri Lanka (Mouse04); **c** lateral and medial views of the left mandible of the western house mouse *Mus musculus domesticus* from 4th c. AD deposits in Abu Sha'ar, Egypt (Mouse02); **d** lateral and medial views of the right mandible of the spiny mouse *Acomys cahirinus* from 5th -7th c. AD deposits in Shenhur, Egypt (Mouse19). Note the similarity between house mouse and spiny mouse mandibles. Sample ID in brackets refers to Appendix 6.II.

Molecular methods

Specimens were extracted and amplified in the Searle Lab, Cornell University, following the same work flow (aDNA and modern) and extraction techniques as outlined in Chapter 3.

The sequence V00711 (Bibb *et al.* 1981) available through GenBank is used as a reference and all positions and fragment lengths discussed hereafter are with respect to it. I amplified 945 bp of the D-loop in overlapping fragments (not accounting for insertions and deletions) using the primers from Jones *et al.* (2012), which are detailed in Table 6.1. Modern tissue samples (tissues stored in ethanol or DMSO after removal from a recently killed animal) were amplified in two fragments using primers L15380/H15989 and L15780/H16325. Museum samples were amplified in multiple fragments using various combinations of primers to generate sequences approximately 147-350 bp in length. Amplicons generated by primer pairs that created longer fragments were attempted first (e.g. L15380/H15671) and, when amplification was not successful, a combination of primers that generated shorter amplicons were used (e.g. L15380/15528 and L15450/H15671). For museum material (osteocrusts and small pieces of tissue from dried skins), in most cases amplicons were generated using various combinations of primers resulting in multiple fragments covering any single nucleotide site, and sometimes replicated. Multiple fragments covering a single site ensured accurate calling of base states from the chromatograms as well as the ability to detect the amplification of possible numts, or contamination, if overlapping fragments were not identical.

Due to the varying quality of my DNA template I did not follow a uniform PCR protocol. Twenty-five µl PCR reactions contained ddH_20, 1 µl of DNA template and final concentrations of 1 x buffer, 2 mM - 4 mM $MgCl_2$, 2-2.5 mM dNTPs, 2-2.5 mM each primer, and 0.6-1.5 units JumpStart Taq (Sigma-Aldrich, St. Louis, USA), Platinum Taq or Platinum Taq High Fidelity (Life Technologies, Carlsbad, USA).

Fragment name	Primer position	Sequence 5' to 3'	Fragment length
Frag 1	L15380	GCACCCAAAGCTGGTATTCT	149
	H15528	TTTTATGACCTGAACCATTGAYT	
Frag 2	L15450	TATGTATATCGTACATTAAAYTAT	219
	H15671	GAAGGGGATAGTCATATGGAAG	
Frag 2a	*L15480*	*CCAAGCATATAAGCAAGTACAT*	*179*
	H15659	*CATATGGAAGAGAAGAGTTTATG*	
Frag 2b	*L15602*	*ATATCTGTGTTATCTGACATAC*	*138*
	H15740	*AGAAGAGGGGCATWGGTGG*	
Frag 3	L15614	TTATCTGACATACACCATACAG	222
	H15835	TATGGGCGATAACGCATTTGAT	
or	*L15689*	*TCTACCATCCTCCGTGAAAC*	*147*
Frag 4	L15780	CTTTATCAGACATCTGGTTCTT	209
	H15989	GCGTCTAGACTGTGTGCTGT	
or	*L15842*	*CCTTAAATAAGACATCTCGATG*	*147*
Frag 5	L15937	CTTTCATCAACATAGCCGTCAA	215
	H16195	TGTTTTTGGGGTTTGGCATTAA	
Frag 5a	*L15990*	*CACCTACGGTGAAGAATCATT*	*150*
	H16139	*GGTTTGGCATTAAGAGGAGG*	
Frag 6	L16041	TATTCATGCTTGTTAGACATAAA	112
	H16195	TGTTTTTGGGGTTTGGCATTAA	
Frag 7	L16069	CTCAATACCAAATTTTAACTCTC	160
	H16229	GTCATATTTTGGGAACTACTAG	
Frag 8	L16181	CTATCAAACCCTATGTCCTGA	144
	H16325	CTTGTTAATGTTTATTGCGTAAT	

Table 6.1. Primers used in this study after Jones *et al.* (2012), their position relative to V00711 (Bibb *et al.* 1981), and the length of the fragment (not accounting for insertions and deletions).

PCR products were cleaned prior to sequencing using ExoSAP (Affymetrix, Santa Clara, USA) following the manufacturer’s protocol and sequenced in one direction for modern samples and both directions for museum samples on an Applied Biosystems ABI 3730.

Sequence alignment and data generation

Contigs were edited and assembled in Geneious v 6.1.6. Missing nucleotides in partial sequences were replaced with 'N'. Macholan e*t al.* (2012) identified nuclear encoded copies of mitochondrial gene fragments (numts) for the D-loop in house mice of the *domesticus* lineage from Socotra. These are the first numts identified for the D-loop gene in house mice that I am aware of. I checked sequences for contamination or numts by examining the chromatogram for dubious peaks, and concordance between overlapping/alternative fragments and, for numt identification, by examining the frequency of transitions and transversions in the fragments and examining the phylogenetic trees for dubiously positioned taxa. I found no evidence of either cross-contamination or numts in the fragments generated.

Amplicons and assembled sequences were compared to the GenBank nucleotide database using a BLAST search algorithm to identify species that were not *M. musculus* (National Institute of Health 2014). Individuals that were not *M. musculus* were not analysed further. Some samples did not produce any sequence and some *M. musculus* sequences contained too many missing nucleotides to be reliably informative (particularly if they were missing nucleotides in the hypervariable regions 1 and 2) (Appendix 6.II). A total of 145 sequences new to this study were used in the analysis.

I aligned my samples with published sequences retrieved from GenBank and with published alignments from Suzuki *et al.* (2013). The details of particular alignments are described in the data analysis section. Alignments were made in Geneious v 6.1.6 using MAFFT v 7.107 and then manually realigned in MEGA 5.2. This involved removing a 76 bp insertion from some *M. m. castaneus* sequences (nine of my individuals) and a 10 bp insertion from some *M. m. domesticus* sequences (two of my individuals) (Table 6.2). The alignment was truncated to approx. 850 bp from the start codon to allow comparison with previously published sequences. In some analyses sequences were collapsed to haplotypes in FaBox v1.41 (specified in the data analysis section for each dataset) (Villesen 2007). As the sequence was truncated, the loss of some polymorphisms in some sequences meant that some published haplotypes were assimilated within others.

Sample ID	HG ID	Country	Insertion (bp)
85876	DOM	Angola	10
85877	DOM	Angola	10
88788	CAS	Iran	76
88825	CAS	Iran	76
171127	CAS	India	76
88912	CAS	Iran	76
88916	CAS	Iran	76
103679	CAS	Afghanistan	76
88911	CAS	Iran	76
140464	CAS	Pakistan	76

Table 6.2. Samples with insertions removed for analysis in the *M. m. castaneus* (CAS) and *M. m. domesticus* (DOM) haplogroups (HG).

Data analysis

To assign all my individuals to haplogroups (HGs) I implemented a Bayesian Markov Chain Monte Carlo (MCMC) phylogenetic analysis in MrBayes v3.2.1 (Ronquist *et al.* 2012). I used the alignment *Appendix_4_Concatenated_30tax.nex.txt* published by Suzuki *et al.* (2013), truncated, and with some modifications. Three sequences of each of *M. m. castaneus* (CAS), *M. m. musculus* (MUS) , *M. m. domesticus* (DOM), *M. m. gentilulus* (GEN) and two sequences of the NEP lineage (as yet taxonomically undefined) were retained along with *Mus macedonicus.* I added an undefined sequence, that I name as part of the SIK haplogroup, from Prager *et al.* (1998) (GenBank ID AF074526), plus *Mus cypriacus*, *Mus spicilegus,* and *Mus spretus*, and all my sequences*. M. spretus* was used as an outgroup following the methods of (Suzuki *et al.* 2013). This was the ALL (all new individuals) dataset.

The best-fit substitution model for Bayesian inference as chosen by Bayesian Information Criterion (BIC) and Aikake's Information Criterion corrected (AICc) was estimated in jModelTest v2.1.4, using 11 substitution schemes, and similarly for all subsequent Bayesian MCMC analyses (Posada 2008) (Table 6.3). In the case of the ALL and DOM datasets, the best-fit model (TIM2 +I+Γ) is not implemented in MrBayes, and so I applied the closest available model, that of GTR+I +Γ (Zakharov *et al.* 2009).

Dataset	**AICc**	**BIC**	**Implemented**
ALL	TIM2 +I+Γ	TIM2 +I+Γ	GTR+I +Γ
NEP	HKY+I	HKY+I	HKY+I
SIK	HKY+Γ	HKY+Γ	HKY+Γ
GEN	HKY+I	HKY+I	HKY+I
CAS	HKY+I+Γ	HKY+I+Γ	HKY+I+Γ
DOM	TIM2 +I+Γ	TIM2 +I+Γ	GTR+I +Γ

Table 6.3. Best-fit substitution models under Bayesian Information Criterion (BIC) and Aikake's Information Criterion corrected (AICc) as estimated by jModelTest v2.1.4 for each haplogroup dataset, and the model implemented.

Two independent runs were implemented with four chains for 5 million generations sampled every 1000 generations. A 25% burn-in was used. To check for convergence and stationarity of MrBayes runs I examined the standard deviation of split frequencies statistics of the two independent runs, examined the log files in Tracer v 1.4 and used the "Compare" function in Are We There Yet (AWTY) (Nylander *et al.* 2008).

After assigning all individuals to HGs, independent analyses were run on each HG. Bayesian MCMC analysis were run for the CAS, DOM, NEP, GEN, and a new HG (SIK, described in more detail in the results section) based on haplotype data. The same run and sampling specifications as for the ALL dataset was used on all independent HGs, except that 10 and 15 million generations were used for CAS and DOM, respectively. Only ten new individuals, all within the known range, were sequenced for MUS and a NN network was generated to assign individuals to groups within it.

Two of the HGs I identified contained only small numbers of haplotypes (NEP and SIK). For these HGs, haplotypes were aligned with one example of every other HG plus *M. macedonicus*, used as an outgroup following the methods of Suzuki *et al.* (2013). I removed three sequences (from Madagascar) from the GEN dataset (samples 156238, 175935 and 165722) as they contained a large number of missing nucleotides but in every other respect were identical to sample 172728. For the DOM dataset, my sequences were aligned with an existing alignment of all published sequences (excluding the recent paper by Suzuki *et al.* (2013)) created by Sofia Gabriel, with the addition of those Suzuki *et al.* sequences, and aligned as shown in the supplementary information *Appendix_6_Cr_DOM.nex.txt.* The dataset consisted of 626 haplotypes. Two of each published CAS and MUS sequences were used as outgroups, following the methods of Jones *et al.* (2010b). CAS sequences were aligned with all published sequences available on GenBank for which there was location information (some sequences with a high proportion of missing nucleotides were removed) following the alignment methodology of Suzuki *et al.* (2013), resulting in 192 haplotypes.

Maximum Likelihood (ML) trees were constructed in PhyML3.0 (Guindon *et al.* 2010). In most datasets, all sequences were included, even where they consisted of a large number of missing nucleotides, as many of the museum samples I generated would have been excluded from the analysis if sequences with high proportions of missing nucleotides were excluded. However, 20 sequences from the DOM dataset that contained high proportions of missing nucleotides were removed. Similarly, all sites were included as removing sites with missing data would have removed informative sites in the dataset. Substitution models were implemented following the AICc in Table 6.3. Number of gamma rate categories was set to 4, and all other substitution parameters were estimated by the program. Tree searching was done by nearest SPR and NNI. Bootstrapping to test the phylogeny was conducted with 1000 replicates for all datasets, except CAS and DOM, which had 500 replicates. For DOM, node support was also estimated with approximate likelihood ratio test (aLRT) (Anisimova & Gascuel 2006).

Although house mice are a comparatively genetically well-studied system, calibrating the molecular clock is not without issue. Suzuki *et al.* (2013) conducted genetic analysis on cytochrome *b* and D-loop genes, yet only calibrated the molecular clock for cytochrome *b* estimation. They identify that high inequality in D-loop branch lengths (compared to cytochrome *b*) suggests irregular substitution fixation rates over evolutionary time. Rate heterogeneity between lineages, especially at the subspecies level, will lead to either under- or over- estimation of the molecular clock along different branches. There are potentially two calibration points with which to date clades within the *M. musculus* group: 1.7 million years ago for the root of the *M. musculus* group estimated from nuclear gene sequences and the fossil record of the genus *Mus* and other Murinae; and approximately 1000 years ago for a shallow branch within clade F of the *domesticus* lineage estimated from aDNA from subfossil Iceland/Greenland mice (Jones *et al.* 2012; Suzuki *et al.* 2013). With no further calibration points along the evolutionary timescale, nor calibration points for each subspecies, the molecular clock will inaccurately estimate divergence times for clades and historical inference, especially at shallow timescales will be unduly affected. Therefore, I did not calibrate the molecular clock of *M. musculus.*

To calculate genetic distances between the HGs I aligned my new individuals with multiple representatives from each HG taken from *Appendix_1_Cr_761taxa.nex.txt* resulting in 904 sequences (Suzuki *et al.* 2013). Genetic distances between populations (uncorrected p-distance) were calculated in MEGA5 (Tamura *et al.* 2011) with a 95% partial deletion parameter, which removed sites with Ns and indels. Due to the limits of computational power, I reduced this dataset, removing some highly similar (within a HG) CAS, MUS and DOM sequences to create a dataset of 559 sequences with which I produced a Neighbour-Net network (NN). NN networks for the ALL and MUS datasets were generated in SplitsTree v 4.13.1 using an equal angle algorithm and uncorrected p-distance.

Median-joining (MJ) haplotype networks (ε=0) were generated for NEP and GEN datasets in Network v 4.6.1.2 (Bandelt *et al.* 1999). For the NEP dataset, networks were based on all haplotypes (with Ns) resulting in 598 bp of sequence that remained after ambiguous sites were removed. For GEN, networks were based on 269 bp of the HVR1 region plus some of the conserved region, as this was the stretch of the D-loop gene for where there was site coverage in all haplotypes. The SIK dataset did not have enough haplotypes to generate a network and the MUS, CAS and DOM datasets contained missing nucleotides that variously covered the whole D-loop gene, thereby removing all sites in the analysis, and thus offering no extra information than previous studies (e.g. Suzuki *et al.* (2013)).

Population genetic and neutrality statistics, and mismatch distribution plots were calculated in dnasp5 as in Chapter 3, for geographical populations of Madagascar and Réunion (Rozas *et al.* 2003). To allow calculation of these statistics I removed missing nucleotides from the dataset by removing both sequences and nucleotide sites. Statistics and plots were calculated on 530 nucleotide sites and 24 sequences for Madagascar, and 767 nucleotide sites and 41 sequences for Réunion.

RESULTS

Identifying taxa and the relationship of HGs

I identified six sequences that were species other than *M. musculus* (Table 6.4). The remaining individuals (n=145) clustered within six HGs, including the four major groups MUS (n=10), CAS (n=49), DOM (n=70), and GEN (n=10) (Figure 6.3 and Appendix 6.I). Additionally, two more HGs, NEP (n=3) (a newly identified lineage after Suzuki *et al.* 2013), and SIK (n=3) were identified. The SIK lineage is new to this study, except for one individual published by Prager *et al.* (1988). I named the lineage (SIK) after its most easterly range in Sikkim, India, in order to discriminate it from the NEP variety in Nepal, with which it probably has an overlapping range. All HGs were highly supported, except for CAS, which showed lower support as is known from other studies (e.g. Suzuki *et al.* (2013)). There is low support for groupings between HGs.

According to genetic distances, GEN and DOM are the most distinct taxa, and CAS and MUS are the most closely related (Table 6.5). The NN network (Figure 6.4) shows DOM as an outsider to the other lineages with a relatively late expansion signal as shown by the cluster of sequences at the tip of the arm. Clear genetic substructure can be seen within the CAS HG and to a lesser extent within the MUS and DOM HGs.

All three methods (phylogenetic tree, genetic distances, network) suggest that MUS and CAS are more closely related than either is to DOM.

Sample ID	Supplier	Type	Taxon ID	Date	Country	Seq. (bp)	BLASTN ID
86360	FMNH	Osteocrust	*Mus musculus*	1956	Tanzania	250 & 224	*Mastomys natalensis* Natal multimammate mouse
100175	FMNH	Osteocrust	*Mus musculus praetextus*	?	Egypt	285	Indeterminate rodent (*Uromys, Leggadina, Praomys, Mus caroli*)
112251	FMNH	Osteocrust	*Mus musculus praetextus*	1968	Iran	730	*Mus macedonicus* Macedonian mouse
112255	FMNH	Osteocrust	*Mus musculus praetextus*	1968	Iran	870	*Mus macedonicus* Macedonian mouse
122319	FMNH	Osteocrust	*Mus musculus praetextus*	1953	Turkey	104	*Mus macedonicus* Macedonian mouse
168098	FMNH	Tissue in DMSO	*Mus musculus*	2000	Tanzania	872 (165 Ns)	*Mus callewaerti* Callewaert's mouse, or *Mus bufo* toad mouse, or *Mus macedonicus* Macedonian mouse

Table 6.4. Individuals that were defined as *M. musculus* in museum collections but were actually different species according to D-loop sequence data. The length of the sequence and the best match hit from a nucleotide BLAST (NCBI) are given.

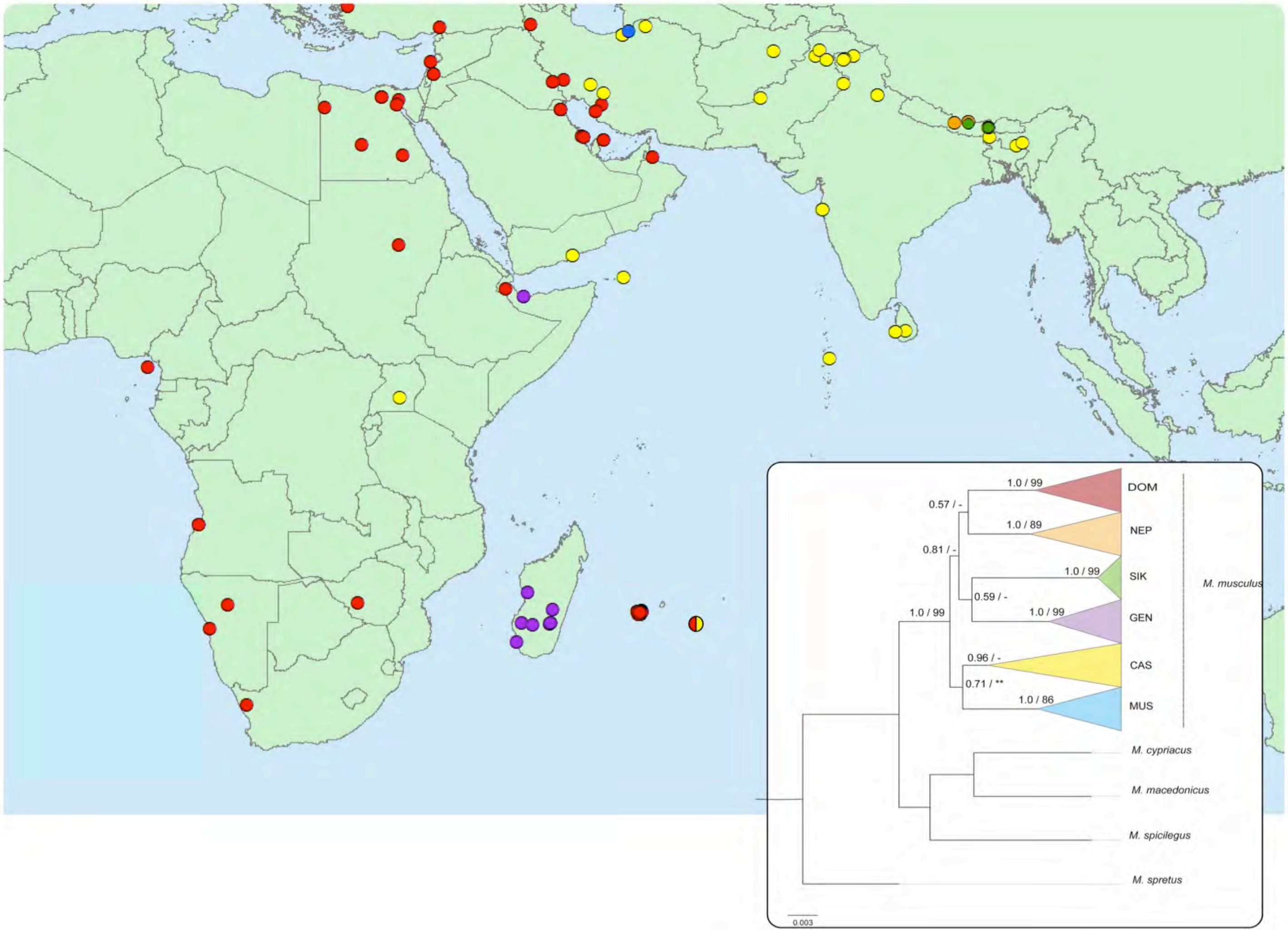


Figure 6.3. Phylogeny and location of sequences new to this study to determine haplogroup (HG) identity. The tree includes published examples of all HGs, plus aboriginal forms, and is rooted with *M. spretus.* Bayesian posterior probabilities / Maximum Likelihood (ML) bootstrap values > 50% for *Mus musculus* are placed at the node (- indicates <50 % support). ** indicates the node was not recovered in the ML tree.

	GEN	NEP	SIK	CAS	MUS
GEN (16)					
NEP (5)	3.66				
SIK (4)	3.81	2.76			
CAS (136)	3.51	2.32	2.69		
MUS (142)	3.22	2.45	2.85	1.93	
DOM (601)	4.67	3.03	4.07	3.52	3.45

Table 6.5. Percent pairwise uncorrected p-distance (95% partial deletion) between HGs based on 904 sequences. The number of sequences analysed within each HG is given in brackets.

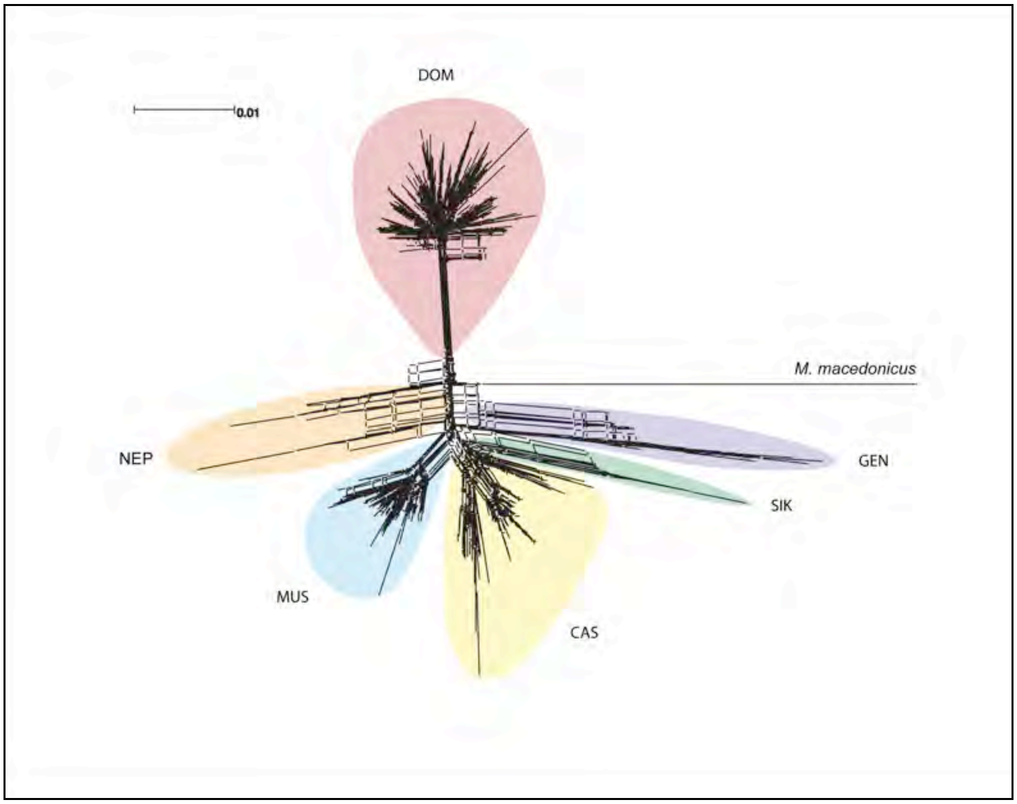


Figure 6.4. NN network (uncorrected p-distance) based on 559 sequences, rooted with *M. macedonicus.*

Characteristics of HGs

Currently undetermined taxa (NEP and SIK)

On the basis of the analysis of all new sequences (ALL), NEP and SIK are genetically distinct clades and this is confirmed in separate analyses (Figure 6.5). Although geographically close, they are only moderately genetically related (uncorrected p-distance = 2.76%) when compared to the relationships between other HGs (smallest distance = CAS vs MUS (1.93%) and largest distance = GEN vs DOM (4.67%)). Sequences currently known (mine and published) for NEP and SIK are distributed in the foothills of the Himalayas in Nepal, and for SIK only, in the Sikkim region of India (Figure 6.5, Table 6.6). The MJ network for the NEP lineage shows two main population groups, one consisting of NEP6 and NEP7 from the Kathmandu area, and a further group (NEP1-NEP5) that is also found in Kathmandu but is also more widely distributed.

Hap ID	**Sample ID**	**Location**	**Study**
NEP1	AB649584	Tukuche	Suzuki *et al.* (2013)
NEP2	142198	Khumjung	This study
NEP3	AF074524	Thamel	Prager *et al.* (1998)
NEP4	142196	Khumjung	This study
NEP5	163164	Gokarana	This study
NEP6	AB649585	Kathmandu	Suzuki *et al.* (2013)
NEP7	AF074525	Nuwakot	Prager *et al.* (1998)
SIK1	AF074526	Jamnagaon	Prager *et al.* (1998)
SIK2	142200	Lukla	This study
SIK3	35577	Lingtam	This study
SIK3	35578	Lingtam	This study

Table 6.6. Haplotype and sample identifications for NEP and SIK HGs, their collection location and the source of the data (see Figure 6.5).

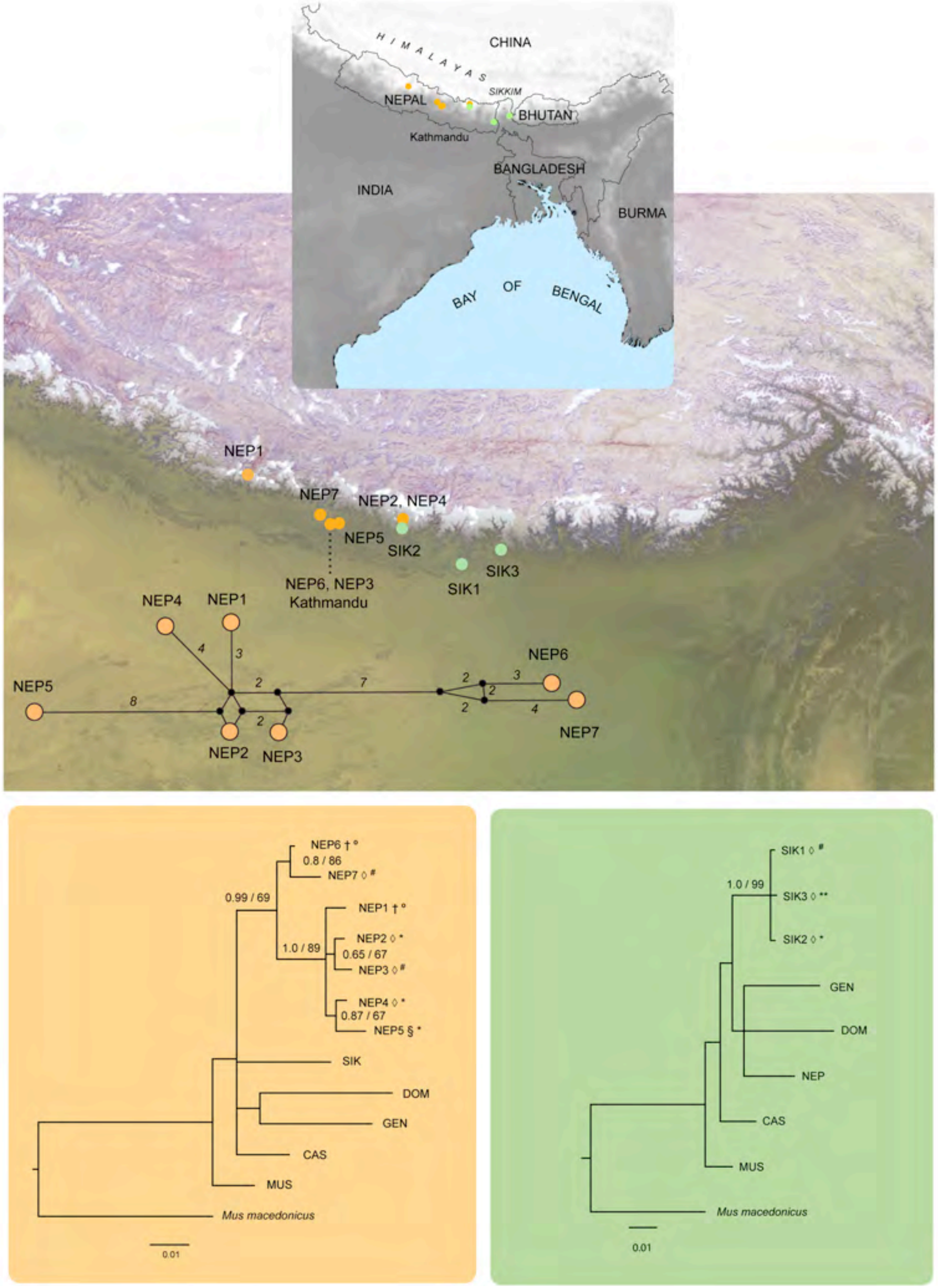


Figure 6.5. Geographic distribution and genetic relatedness of haplotypes within NEP (orange) and SIK (green) haplogroups (HGs) (refer to Table 6.6). An overview map and their distribution in the Himalayas are shown with haplotype identifications corresponding to sequences in the trees and the MJ network for NEP. Median vectors in the MJ network are shown with black dots and number of substitutions (>1) are shown along the branches. Bayesian MCMC trees are rooted with *M. macedonicus* and Bayesian posterior probabilities / Maximum Likelihood bootstrap values for the NEP and SIK HGs are placed at the node. Sample suppliers are identified as FMNH (◊), AMNH (§), and Hokkaido University (†). Sequences from this study (*), Suzuki et al 2013 (°), and Prager et al 1998 (#).

The Arabian house mouse M. m. gentilulus (GEN)

GEN has a previously recorded distribution in Yemen and Madagascar (Prager *et al.* 1998; Duplantier *et al.* 2002). All my Madagascar sequences were of the GEN HG. The MJ network clearly indicate that Madagascar mice result from a star-like expansion from a central haplotype. Large distances and high haplotype diversity in the MJ network indicate the Yemen populations are relatively old, compared to Madagascar populations. The T2 haplotype from Hodeida, on the Red Sea coast is of interest because it falls within the (albeit poorly supported) Madagascar clade, indicating derivation from a recent common ancestor with the Madagascar mice, of particular interest in terms of movements between south Arabia and Madagascar. I identified a GEN haplotype along the Red Sea coast of Somalia that fits within the Yemen diversification.

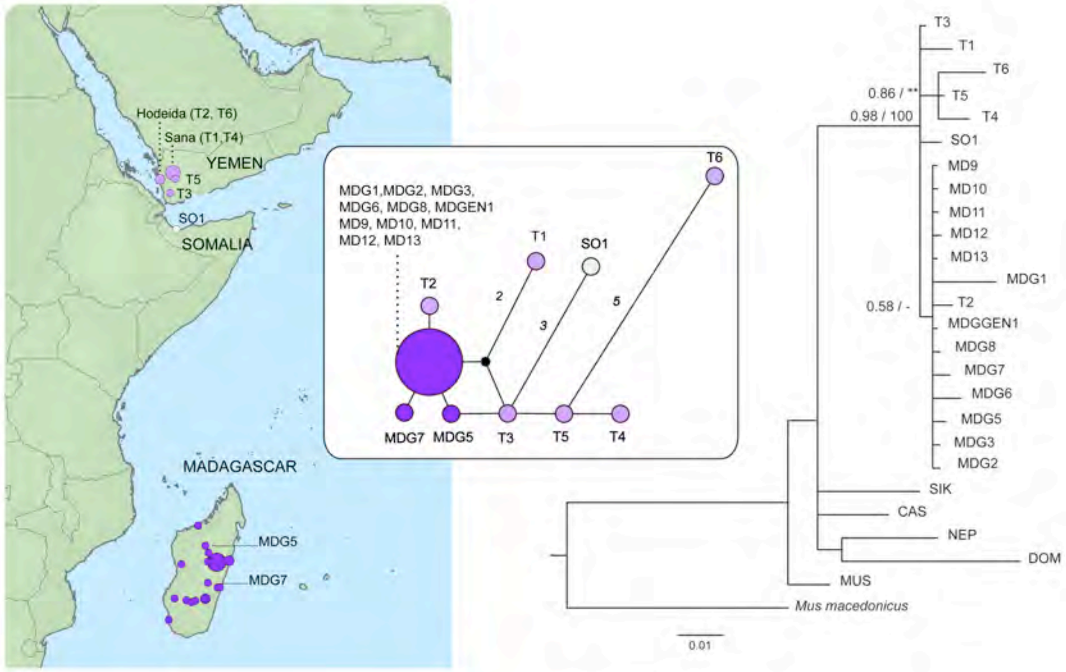


Figure 6.6. Geographic distribution and genetic relatedness of haplotypes within the GEN haplogroup (HG) (see Table 6.7). Bayesian MCMC trees are rooted with *M. macedonicus* and Bayesian posterior probabilities / Maximum Likelihood bootstrap values > 50% for the GEN HG are placed at the node (- indicates <50 % support). ** indicates the node was not recovered in the ML tree. Haplotypes are labelled according to the original studies and mine are prefixed 'MD'. The area of the circles in the map are proportional to the number of haplotypes at a location (haplotypes used in the Bayesian MCMC tree). Geographic regions are colour-coded Madagascar (dark purple), Yemen (medium purple) and Somalia (light purple). Only non-core haplotypes are labelled on Madagascar. The MJ network is based on 269 bp incorporating the HVR1 region, and area of the circles is proportional to the number of sequences within a haplotype (labelled by haplotype name). Geographical regions are colour-coded as in the map. Median vectors in the MJ network are shown with black dots and number of substitutions (>1) are shown along the branches.

Sample ID	Country	Location	Reference	Hap. ID	No.
AY091517	Madagascar	Mahajanga	Duplantier *et al.* (2002)	MDG6	1
AY091518	Madagascar	Ambodifaho	Duplantier *et al.* (2002)	MDG3	1
AY091520	Madagascar	Antananarivo	Duplantier *et al.* (2002)	MDG1	1
AY091521	Madagascar	Antafotenona	Duplantier *et al.* (2002)	MDG7	1
AY091522	Madagascar	Kamolandy	Duplantier *et al.* (2002)	MDG5	1
AY091523	Madagascar	Firarazana	Duplantier *et al.* (2002)	MDG2	1
AY091523	Madagascar	Ambodifaho	Duplantier *et al.* (2002)	MDG2	1
AY091523	Madagascar	Manakasina	Duplantier *et al.* (2002)	MDG2	2
AY091524	Madagascar	Nanarena	Duplantier *et al.* (2002)	MDG8	1
AY091524	Madagascar	Ampadratokana	Duplantier *et al.* (2002)	MDG8	1
AY091524	Madagascar	Ambalatenona	Duplantier *et al.* (2002)	MDG8	1
AY091524	Madagascar	Manakasina	Duplantier *et al.* (2002)	MDG8	2
JN416769	Madagascar	Manakasina	Rajabi-Maham *et al.* (2012)	MDGEN1	1
AF074540	Yemen	Sana	Prager *et al.* (1998)	Type 1	3
AF074541	Yemen	Hodeida	Prager *et al.* (1998)	Type 2	1
AF074542	Yemen	Wadi Maleh	Prager *et al.* (1998)	Type 3	1
AF074543	Yemen	Sana	Prager *et al.* (1998)	Type 4	1
AF074544	Yemen	Ma'bar	Prager *et al.* (1998)	Type 5	1
AF074545	Yemen	Hodeida	Prager *et al.* (1998)	Type 6	1
1010320	Somalia	Bulhar	This study	SOM1	1
156238	Madagascar	R. N. Integrale d'Andringitra	This study	MD11	1
161770	Madagascar	Foret d'Ankazomivady	This study	MD9	1
165722	Madagascar	P.N. d'Andringitra	This study	MD10	1
165723	Madagascar	P.N. d'Andringitra	This study	MD9	1
172728	Madagascar	P.N. de Bemaraha	This study	MD9	1
173206	Madagascar	P.N. de Tsimanampetsotsa	This study	MD12	1
175935	Madagascar	P.N. de l'Isalo	This study	MD13	1
T49	Madagascar	Antananarivo	This study	MD11	1
100728	Madagascar	Tabiky	This study	MD9	1

Table 6.7. Haplotype and sample identifications for the GEN HG, their collection location, number of individuals, and the source of the data (see Figure 6.6).

The eastern house mouse M. m. musculus (MUS)

My MUS samples are all from known *M. m. musculus* ranges within western Eurasia and cluster with published sequences from the same countries (Figure 6.7). Individuals 98115 and 98116 collected from the AMNH were recorded as topotypes of *Mus musculus wagneri* based on morphological data. In my genetic analysis they are divergent within the MUS HG clustering within separate groups (pairwise uncorrected-p distance between the two sequences = 0.6%).

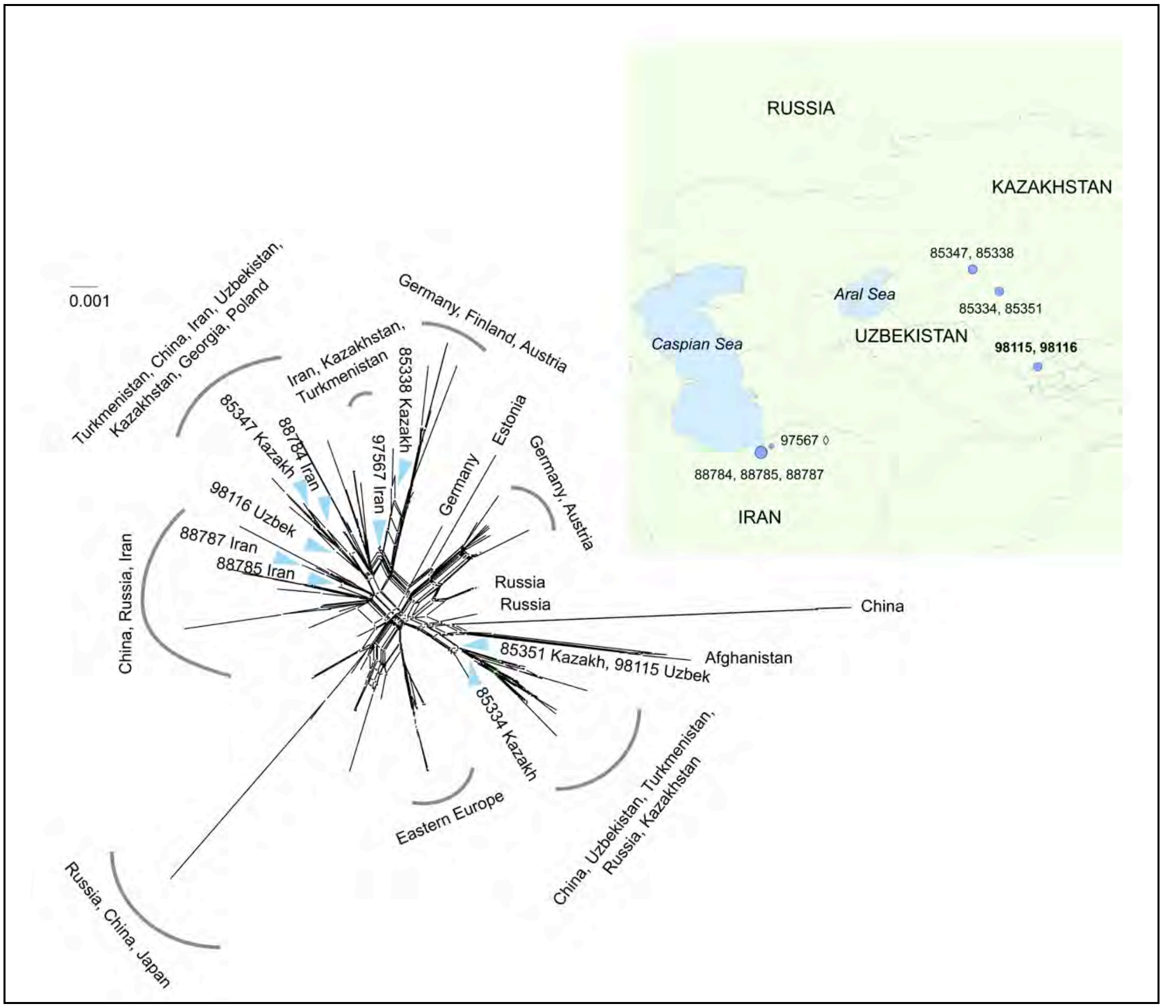


Figure 6.7. Position of my MUS samples in a NN network indicated by blue triangles and their location on a map. Size of the circles is proportional to number of sequences. All individuals are from AMNH, except ◊=FMNH. Topotypes of *Mus musculus wagneri* are in bold.

The southeastern house mouse M. m. castaneus (CAS)

My sequences clustered within three of the four known CAS clades (CAS1, CAS2, and CAS3) as described by Suzuki *et al.* (2013) (Figure 6.8). Topology between Bayesian and ML phylogenetic methods was similar except that three individuals that were in the CAS2 clade in the Bayesian tree were basal in the ML tree (88915 from Iran; 171127 from India; and a published sequence from Iran GenBank accession number JN416693). All eight of my individuals and published sequences that contained the 76 bp tandem repeat (removed for analysis) clustered within the CAS2 clade. The support for CAS clades varies with CAS1 and CAS4 being well-supported under Bayesian MCMC estimation (but not under ML estimation) and CAS2 and CAS3 having generally low support. As with other published studies of the D-loop gene, support is generally low for CAS clades. For example, in a ML analysis and using aLRT to test the support of nodes (the selection threshold for aLRT values is typically around 0.8 - 0.9), Rajabi-Maham *et al.* (2012) showed support values of 0.76 for CAS1 (their HG2), 0.90 for CAS2 (their HG3), and 0.73 for CAS3 (their HG1) (Guidon *et al.* 2010).

Here, I extend the known range of *M. m. castaneus* to incorporate Maldives, Réunion, Yemen and Uganda. These areas together with Socotra are highlighted in the phylogenetic tree in Figure 6.8 as locations where humans must have transported this subspecies within the western Indian Ocean. CAS1 is the most widely distributed clade with a range from western Iran throughout southern India and Sri Lanka, and throughout East and Southeast Asia. It was introduced to the Maldives, Réunion and mainland Africa (Figure 6.8). CAS 3 is distributed in western Iran and northern India/Pakistan and has been introduced to the Arabian Peninsula and Socotra (Figure 6.8).

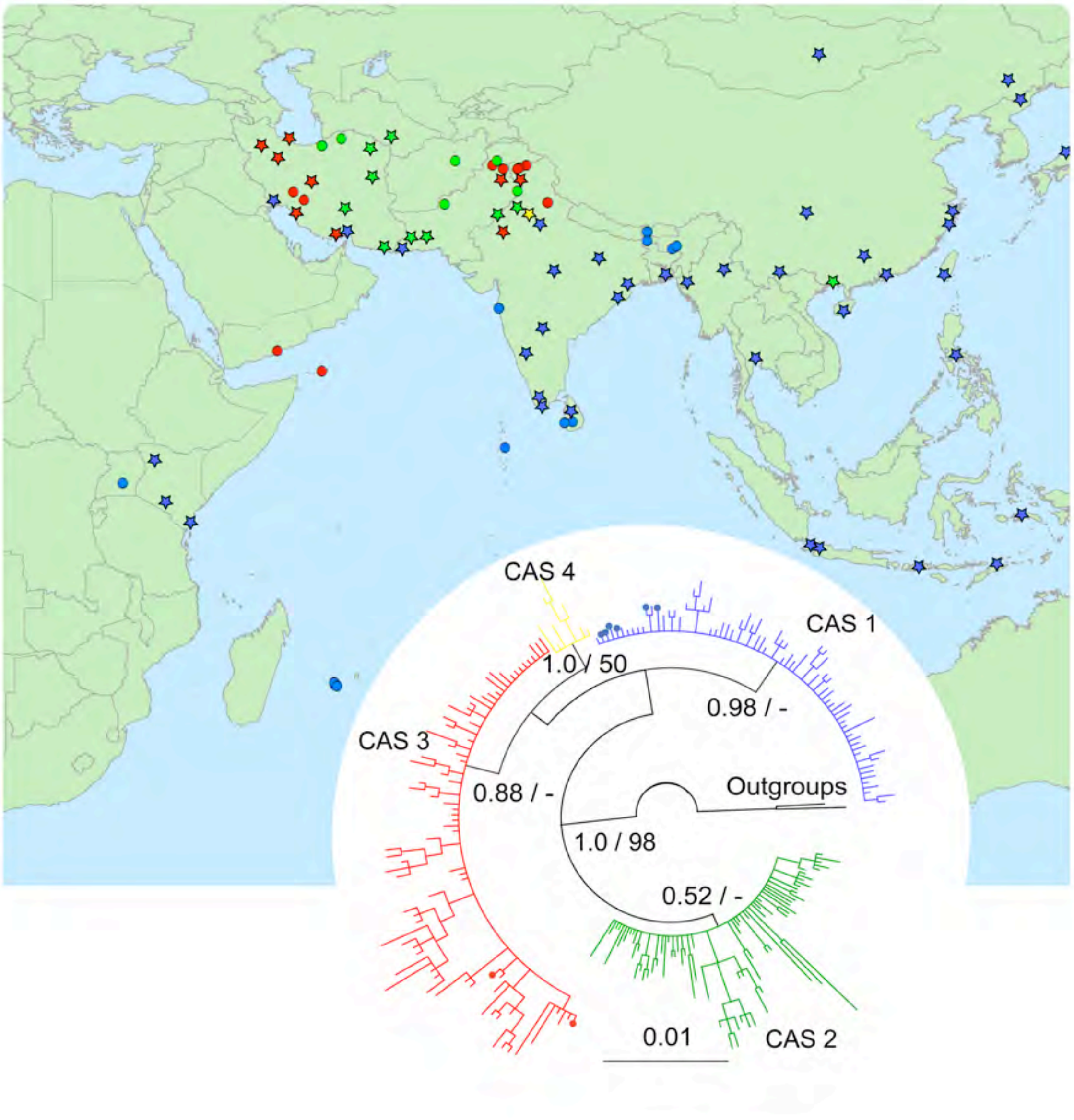


Figure 6.8. Geographic distribution and genetic relatedness of haplotypes within the CAS haplogroup (HG). Bayesian MCMC trees are rooted with GEN and Bayesian posterior probabilities / Maximum Likelihood bootstrap values > 50% are placed at the node (- indicates <50 % support). Dots at the terminal end of branches indicate haplotypes from (clockwise from top left) Réunion x3, Uganda, Réunion, Maldives, Yemen, and Socotra. Circles on the map indicate locations of clades from my study, and stars are the approximate location from Suzuki *et al.* (2013) and Rajabi-Maham *et al.* (2012). See Appendix 6.I for clade affinity within CAS for individual samples. N.b. colours are not consistent with previous figures.

The western house mouse M .m. domesticus (DOM)

Clades described here are labelled 'A-F' and follow the nomenclature of Jones *et al.* (2010b) plus a clade labelled '6' after Bonhomme *et al.* (2011). Clade affinity of sequences between Bayesian MCMC and ML phylogenetic analysis was identical. My sequences were placed in all the previously described clades, plus some were ungrouped, and two sequences were sister to clade F proper (see Figure 6.9 and Appendix 6.I). Based on the 945 bp sequence analysed here, one individual from Oman (551476) was identical to the Iran.1 type from Gündüz *et al.* (2000) (ungrouped in my tree) and one individual from Saudi Arabia (166948) was identical to Itl-dom8 from Rajabi-Maham (2008) (part of the sister group to clade F). Three sequences from Suzuki *et al.* (2013) (AB649594 from Java; AB649608 from Senegal; AB649609 from Somalia) and possibly a fourth (AB649595 from Java, not clear from the text where it is placed) cluster within my Clade E but are described as Clade F by the authors. Some haplotypes described as belonging to clades in publications fell ungrouped in my tree (e.g. some sequences belonging to haplogroups 1, 9, 10 and 11 from Bonhomme *et al.* (2011)).

Support for the clades within the DOM lineage varies; low support for some DOM clades is a noted feature of the lineage due to its shallow genealogy (Rajabi-Maham *et al.* 2008). Interestingly, the various tests of branch support (Bayesian, ML and bootstrapping, and ML and aLRT) showed very different support values, yet the tree topology remained the same and there is clearly a strong phylogenetic signal. Support for all clades using ML and bootstrapping are well below the standard threshold, support for all clades using ML and aLRT are high, and support for all clades using Bayesian MCMC varies with some clades well supported and others moderately to well below the standard threshold. Here, the aLRT seems liberal, bootstrapping conservative, and Bayesian moderate. As parametric tests, Bayesian and aLRT tests of branch support may be unduly influenced by model violations in my data as the correct model (TIM2 +I+Γ) is not implemented, and so I implemented an alternative model (GTR+I +Γ). However, both tests are resistant to moderate model oversimplifications, and here the

GTR+I +Γ model is probably suitably complex. There are shallow divergence levels in the DOM lineage as evidenced by the short branch terminal radiation in the NN network (Figure 6.4) and very short branches can lead to low support for branches in phylogenetic tests; the aLRT is expected to perform well in cases where branch length is short and where phylogenetic signal is strong (Guindon *et al.* 2010). With a clear phylogenetic signal and the support of published studies, I refer to clades within the DOM lineage as described above.

I extend the known range of the DOM HG to parts of the Arabian Peninsula, southern Africa and Réunion. Geographic distributions with respect to clade identity for sequences new to this study are: A) Persian Gulf and Eastern Mediterranean; B) Northwest Iran; C) Egypt and, for C1, Réunion; D) Eastern Mediterranean and, for D1, Southwest Africa; E) Southern Africa and Réunion; F) Persian Gulf, Egypt and Southwest Africa; Ungrouped) Eastern Mediterranean, Northeastern Mediterranean, Persian Gulf, Gulf of Aden, Northeastern Africa and Western Africa.

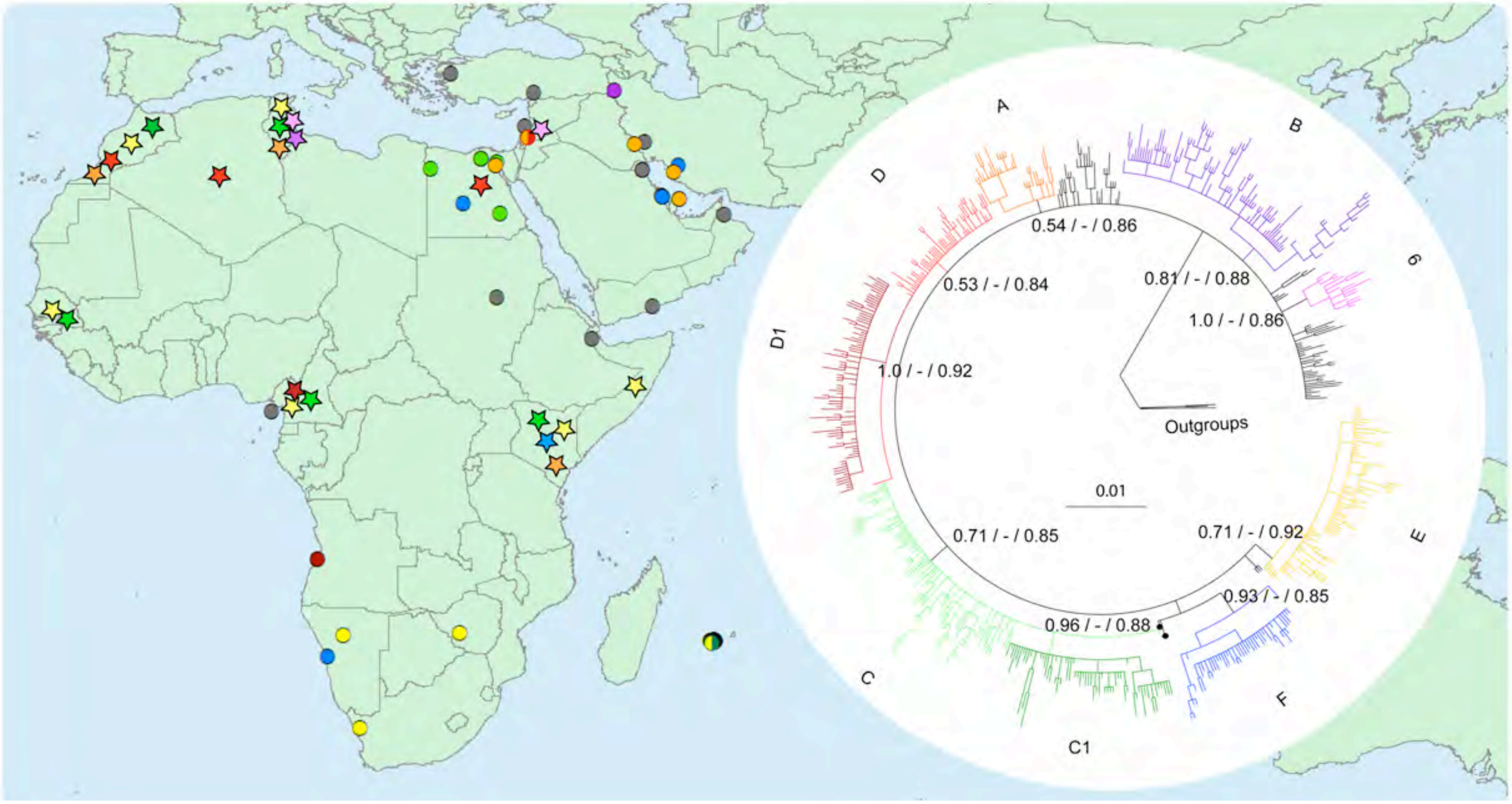


Figure 6.9. Geographic distribution and genetic relatedness of haplotypes within the DOM haplogroup (HG). Bayesian MCMC trees are rooted with MUS/CAS and Bayesian posterior probabilities / Maximum Likelihood (ML) bootstrap values / ML approximate likelihood ratio test values > 50% are placed at the node (- indicates <50 % support). Circles on the map indicate locations of clades from my study, and stars are approximate locations from published studies for locations from the Arabian peninsula, Africa and the Indian Ocean (taken from Bonhomme *et al.* (2011) and Suzuki *et al.* (2013) Supplementary Information). See Appendix 6.I for clade affinity within DOM for individual samples. Ungrouped haplotypes are grey, and two haplotypes that are outgroups to Clade F are marked with black dots on the tree. N.b. colours are not consistent with previous figures.

Geographical populations: Madagascar and Réunion

To put the genetic diversity data for Madagascar and Réunion in context I report some diversity statistics for other populations. In Nepal, close to the putative origin of the species, Suzuki *et al.* (2013) report high haplotype and nucleotide diversity (π=0.02703 and Hd=1). In Norwegian populations that are descendants of introduced *M. m. domesticus* diversity is much lower (π= 0.00818 ± 0.0042 and Hd= 0.894 ± 0.0002) and introduced *domesticus* in Ireland and France are even lower (Ireland π= 0.00503 ± 0.0002 and Hd= 0.9049; France π= 0.00231 ± 0.00003 and Hd= 0.8253) (Jones *et al.* 2010a; b). Madagascar populations of *M. m. gentilulus* show even lower haplotype and nucleotide diversity and the presence of two HGs (DOM and CAS) on Réunion elevate nucleotide diversity slightly, but haplotype diversity remains low (Table 6.8).

Although neutrality statistics do not uniformly signal that the Madagascar population (GEN) underwent an expansion, a statistically significant negative Tajima's D, along with a negative Fu's F_S and small Ramos-Onsins & Rozas' R_2 all tend to indicate a recent expansion event that is partly supported by a relatively good fit of the observed to expected distributions in the mismatch distribution (MMD) plot (Table 6.8, Figure 6.10). The MMD plot shows an L-shaped pattern in the observed frequency distribution of pairwise difference consistent with a population bottleneck and some substructure in the data where the peak at five pairwise differences is seen.

The Réunion population does not show a signal of expansion in neutrality statistics or MMD plots. The MMD shows an L-shaped pattern similar to the Madagascar population, and significant substructure because of the presence of two diverse lineages on the island, is also present. HGs are geographically distributed on Réunion (Figure 6.11): clade E (DOM) are the type found in the central and northern ranges of the island and are the predominant lineage in the west. The westerly range is co-occupied by Clade 1 CAS, which also predominantly occupies the southern end of the island. One individual in the south is an outlier – it is of the DOM haplogroup but is the only individual of the C1 clade so far identified on the island.

Population	**bp**	**N**	**h**	**H_d (±SD)**	**π (±SD)**	**θ (±SD)**	**D (*P*)**	**F_S (*P*)**	**R_2 (*P*)**
Madagascar	530	24	5	0.312	0.0014	0.0046	-2.258*	-1.192	0.1192
				(0.121)	(0.0008)	(0.0021)	(0.000)	(0.089)	(0.298)
Réunion	767	41	5	0.382	0.0099	0.0101	-0.171	10.484	0.1142
				(0.091)	(0.0025)	(0.0033)	(0.486)	(0.995)	(0.605)

Table 6.8. Genetic diversity and neutrality statistics within geographical populations of Madagascar (GEN) and Réunion (DOM & CAS). bp is the number of base pairs analysed for each dataset (indels removed), N number of sequences, h number of haplotypes, H_d haplotype diversity, π nucleotide diversity, θ population mutation rate from sites, D Tajima's D, F_S Fu's F_S, and R_2 Ramos-Onsins & Rozas' R_2. *indicates statistically significant from coalescent simulations ($P<0.05$ for TD and R_2, $P<0.02$ for F_S).

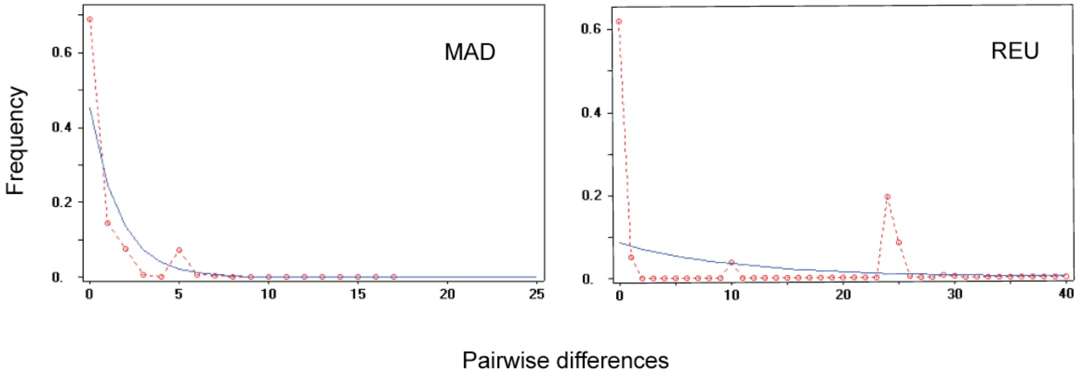


Figure 6.10. Mismatch distribution plots for Madagascar (MAD) and Réunion (REU). The blue line is the expected distribution under a growth and decline model and the red line the observed distribution. MAD (N=24, θ_0=1.2083, θ_1=1000, τ=0), REU (N=41, θ_0=10.66, θ_1=1000, τ=0).

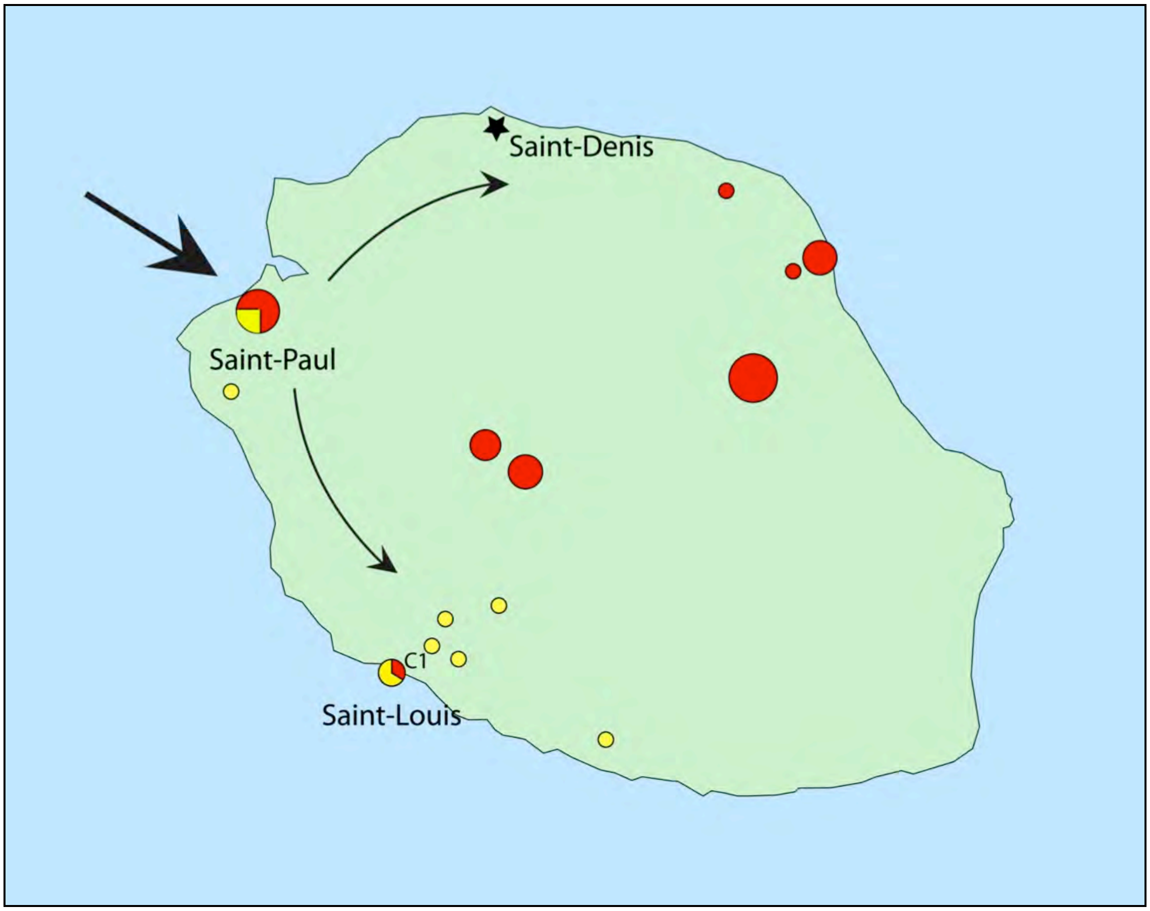


Figure 6.11. Location of CAS (yellow) and DOM (red) individuals on Réunion. Size of the circle is proportional to the number of haplotypes, and area within the circles identifies the proportion of sequences belonging to a haplogroup at that location. DOM individuals are Clade E, except for 1 individual from C1 marked on the map, and CAS are Clade 1. Hypothesised port of entry (thick black arrow) and routes of dispersal (thin black arrows) are shown.

DISCUSSION

Erroneous identification

The subgenus *Mus* contains 11 currently recognised species and several unvalidated ones (Wilson & Reeder 2005). Among the specimens that I sampled as *Mus musculus* there were 6 individuals that are likely misidentified in museum databases (Table 6.4). This is probably not due to introgression of the non-*M. musculus* mitochondrial genome into *M. musculus*; there is no evidence that the mislabelled species can hybridise with *M. musculus*. The species assignments made here for these six individuals are based on very short fragments of the D-loop gene in some cases, and should not be considered definite taxon identifications.

Here I attempt to explain the identification errors.

Sample 86360 (Tanzania): External morphology of the Natal multimammate mouse is very similar to the house mouse, except in females, which have multiple, prominent mammary glands that distinguish them from other species. It is a common commensal species distributed in sub-Saharan Africa. The consistent morphology, ecology, and distribution between the house mouse and the Natal multimammate mouse means that non-specialists could easily confuse the two species, as probably occurred in this case.

Sample 100175 (Egypt): The short fragment length and lack of similar sequences in the BLAST database meant that this individual does not have a solid determination (a number of rodent species are equally likely) and here I label it ‘indeterminate rodent’.

Samples 112251, 112255, 122319 (Iran and Turkey): Three samples were identified as *M. m. praetextus* (a synonym for the *domesticus* form) in the FMNH museum database but their genetic identity was that of an aboriginal form *Mus macedonicus* the Macedonian mouse. The Macedonian mouse is sympatric, but not syntopic, with the house mouse in parts of southwest Asia (extending to western Iran) and Eastern Europe (Boursot *et al.* 1993). Again, they are morphologically very similar to the house mouse, differing mainly in the length of their tails (a longer tail in the house mouse being adaptive to vertical movements and thus hypothetically a commensal adaptation to the built environment) (Slabova & Frynta 2007). *M. macedonicus* was originally described as a subspecies of *Mus hortulans* and assigned valid species status in 1986 (Musser & Carleton 2005). The misidentified individuals were collected prior to the taxonomic revision and are thus likely to be subsumed under an invalid taxon identification.

Given their collection locations fall within the range of *M. macedonicus* they are likely to belong to this species.

Sample 168098 (Tanzania): Three possible identifications based on sequence similarity were identified through BLAST - *Mus callewaerti* Callewaert's mouse, *Mus bufo* toad pygmy mouse, and *M. macedonicus*. Callewaert's mouse is recorded from Angola and Congo, and the toad pygmy mouse from Uganda, Rwanda, Burundi and Congo, however the extent of both species' ranges are unknown (Kingdon 2013). The two African endemics fall into the body size range of *M. musculus*. Published D-loop sequences for the endemics are only approx. 280 bp, whereas published *M. macedonicus* sequences cover the whole fragment for which I have sequence data. The heightened coverage probably leads to a hit in the BLAST database for *M. macedonicus*, whose geographic range does not cover Tanzania. My misidentified sample could belong to an uncharacterised species.

The diversity and relationship of HGs in the house mouse

Four established HGs were represented by individuals in my dataset (DOM, CAS, MUS, and GEN), plus sequences that clustered within a recently identified taxon (NEP), and sequences that cluster within a potential new HG with a single sequence from Prager *et al.* (1998) (SIK). All HGs were well supported and monophyletic in an analysis using all new sequences. However, the relationships between the HGs based on the mitochondrial sequence was less clear, as noted by other authors (e.g. Suzuki *et al.* 2013).

Based on current evidence, the house mouse is thought to have originated in the northern Indian subcontinent where genetic diversity is the highest (Bonhomme & Searle 2012). From this area, extant eastern house mouse MUS, south-eastern house mouse CAS, and western house mouse DOM populations are known, and recently a fourth (as yet taxonomically indeterminate) lineage (NEP) was identified through mitochondrial DNA (Suzuki *et al.* 2013). Here, I add a potential fifth lineage (SIK) to the region.

With respect to the haplogroup identity of *M. musculus* in museums, some individuals were erroneously recorded in the database but correctly labelled on specimen tags, e.g. *Tapecomys primus* (a south American endemic). Others were labelled with alternative subspecies names (e.g. *M. m. praetextus* or *M. m. wagneri*). For some individuals, their genetic identity, as ascertained through mitochondrial DNA, did not match with their museum designation, presumably identified through morphological and geographical criteria. This could be either misidentification, or mismatch between the mitochondrial and dominant genetic identity of the individual due to introgression of the mitochondrial genome after hybridization (Darvish *et al.* 2006).

Characteristics of HGs

Currently undetermined taxa (NEP and SIK)

Here, I describe a new major mitochondrial lineage from the foothills of the Himalayas in the border region of Nepal and India that I label SIK. Only a single published sequence that clusters within the SIK clade was known previously and hypothesised from phylogenetic inference to be part of a new fifth lineage of CAS (Prager *et al.* 1998; Suzuki *et al.* 2013). With the addition of more sequences I show that the SIK lineage is potentially a previously unrecognised, major new lineage of *M. musculus.* Genetic diversity and neutrality statistics could not be calculated for the SIK HG as there were too few sequences. However, shallow branches on the SIK tree indicate a relatively recent divergence within the group compared to other HGs. Sampling bias may affect the phylogenetic inference for SIK and I await the inclusion of more sequences in future studies to further evaluate this group.

Two of the three sequences identified as SIK from my mitochondrial DNA study were identified as *M. m. castaneus* in the FMNH museum database (the CAS HG is also found in the Sikkim region of India). The third sequence (also FMNH) was identified as *M. m. homourus. M. m. homourus* (Hodgson, 1845) is also used to describe two NEP and one CAS individuals in FMNH and AMNH databases, respectively. *M. m. homourus* is purportedly distributed in the foothills of the Himalayas from Nepal to China, and possibly extends to South and Southeast Asia

(Schwarz & Schwarz 1943). It is subsumed within *M. m. domesticus* under current taxonomy (Wilson & Reeder 2005). The disjuncture between the known range of *domesticus* and that of *homourus* suggest the latter taxon needs to be reexamined. In any case, *M. m. homourus* does not closely represent my SIK individuals, which are geographically restricted to the Himalayan foothills of Nepal and India.

I added sequences to the NEP HG, a recently recognised group that still needs further clarification (Suzuki *et al.* 2013). The NEP HG is geographically restricted to Nepal. NEP and SIK are probably make contact in the far east of Nepal but whether they occur syntopically, or hybridise, is unknown. The MJ haplotype network for the NEP HG shows two main population groups, one consisting of NEP6 and NEP7 from the Kathmandu area and a further group (NEP1-NEP5) that is also found in Kathmandu but is also distributed to the east and west of it.

There is considerable cryptic diversity in house mice of the northern Indian subcontinent as evidenced by the addition of the NEP and SIK lineages. With the addition of more samples, relationships and timings of lineage diversification can be estimated, which will inform hypotheses related to the origin and dispersal of mice and lineages within the species. Furthermore, there are potentially more lineages awaiting discovery in this topographically complex region.

The Arabian house mouse M. m. gentilulus (GEN)

The Arabian house mouse has, so far, been recorded from Yemen - its putative origin - and Madagascar, where it was presumably introduced (Prager *et al.* 1998; Duplantier *et al.* 2002). Yemen developed as an early trade nexus due to its importance in the incense trade. The significance of the commercial ports of southwestern Arabia is described in the first century AD *Periplus of the Erythrean Sea* (Casson 1989), and this continued through to the Arab trade of the Islamic period from the late first millennium AD. Given the importance of Yemen in ancient trade networks, it is surprising that the GEN HG has not been translocated more extensively. The GEN form's ability to stowaway and survive long journeys on board ships is not in doubt as it successfully reached Madagascar, a 4000 km journey from Yemen. A likely explanation is that ancient GEN stowaways were met with already established house mouse populations or native species that successfully outcompeted the invaders. The successful colonisation of Madagascar can easily be understood in these terms, as all native rodents on Madagascar are endemic nesomyids that were isolated for millions of years and thus were no competition for invasive, commensal rodents, such as mice.

A recent, single origin of introduction for the house mouse on Madagascar was inferred by Duplantier *et al.* (2002) who found populations on the island to be narrowly monophyletic within the *M. m. gentilulus* lineage. The same pattern is seen with the addition of my data. A clear star-like pattern with a core haplotype and few derived haplotypes can be seen in the MJ network and there is low nucleotide and haplotype diversity in the population. Neutrality statistics and MMD plots are consistent with a population bottleneck during colonisation of Madagascar, followed by a subsequent population expansion. It is quite remarkable that from few invading individuals the Madagascar population has boomed, and successfully colonised almost the entire island, including urban, disturbed, and natural forest environments.

Duplantier *et al.* (2002) suggest an introduction to Madagascar with Arab trade via the East African coast and the Comoro Islands between the 8th and 11th century AD. There are no sequences from the islands of Pemba, Zanzibar, or the Comoro Islands with which to authenticate an East African pathway. A recent biodiversity report from Pemba Island documents the presence of the white-bellied house mouse *M. m. gentilis* (Brants 1827) on the island but the observation is made without descriptive data on individuals (Nahonyo *et al.* 2005). *M. m. gentilis* is considered as subsumed within *M. m. domesticus* and, disregarding highly polymorphic belly colour, its size should presumably distinguish it from the smaller *M. m. gentilulus* (Harrisson 1970). The emerging pattern for mainland East Africa is one of multiple colonisations but overwhelmingly of the *M. m. domesticus* and *M. m. castaneus* lineages, neither of which are found on Madagascar. This distribution suggests a direct route between the Arabian Peninsula and Madagascar. House mouse bones from deposits that date to the 9th to 10th centuries AD in Mahilaka, an Islamic port in northwestern Madagascar, are contemporaneous with bones found at M'Bachile on Grande Comore and thus link the islands in terms of house mouse distribution in the Islamic period (Wright 1984; Radimilahy 1998). An introduction to Grande Comore from Madagascar, or vice versa, from source populations that originated ultimately from Yemen could be possible. With regards the linkage between Yemen and Madagascar, the individual with the T2 haplotype in Yemen is of particular interest. A more detailed genetic comparison between this individual and those on Madagascar would be very interesting, and may help date the colonization of Madagascar. It is also possible that this individual is a descendant of a back-migrant from Madagascar to Yemen.

Madagascar's connections with Iran's ancient ports (e.g. Siraf) through maritime trade networks in the later Swahili period are conspicuously absent with the lack of a genetic signature of either *M. m. castaneus* or *M. m. domesticus*, which occupy present day ports and harbours of Iran, and Egyptian (e.g. Fatimid) links are also missing in the absence of the *M. m. domesticus* lineage (Boivin *et al.* 2013b). Moreover, it follows that a *M. m. castaneus* contribution from Southeast Asia should be evident to correspond with the significant Southeast contribution in the language, culture and genetic identity of the people of Madagascar. With further sampling, or characterization of genetic markers other than mitochondrial DNA, a signature of Southeast Asian and/or East African mice may appear.

I identify a new occurrence of the GEN HG in Bulhar, near Berbera, in Somalia. Bulhar lies across the Gulf of Aden opposite from Aden, Yemen. Interestingly, another house mouse typed from Somalia (unknown location) and a sample from neighbouring Ali Sabieh in Djibouti (near to the Gulf of Aden) are of the DOM HG. The Somali DOM type is of the Clade E variety (discussed in more detail in the *M. m. domesticus* section), which is widely distributed and is presumably a relatively recent introduction (the same type is found in West Africa in Senegal).

The eastern house mouse M. m. musculus (MUS)

The origin of the eastern house mouse is suspected to be near Transcaucasia, or the Caspian Sea, with subsequent range expansions east and west so that it now occupies most of northern Eurasia up to Western Europe, and extends in to Japan (Boursot *et al.* 1993; Prager *et al.* 1996). Specimens sequenced in my study that belong to the MUS HG were from the southern Caspian Sea and east of the Aral Sea in Central Asia - all within the known range of the subspecies.

Suzuki *et al.* (2013) identified two subgroups within the MUS HG based on mitochondrial data, with an estimated divergence date of 150,000 ± 13 000 years bp, and an origin for both subgroups in the ancestral range in the vicinity of the Caspian sea. These two groups can be seen in the northwesterly and southeasterly groupings on my NN network. Although geographically close, individuals from two populations in Kazakhstan cluster in different groups within the MUS HG and two individuals from the same location in Uzbekistan are also in separate groups; this pattern has been recorded previously in these countries (Suzuki *et al.* 2013). Kazakhstan and Uzbekistan clearly harbour genetically diverse populations of the MUS HG, and their proximity to the putative origin of the HG suggests the Central Asian populations are an ancient, probably natural, colonisation.

Two specimens from Tashkent, Uzbekistan sampled from the AMNH are recorded as topotypes (examples from the type locality) of *Mus musculus wagneri* (Eversmann, 1848), and cluster in separate (presumably anciently divergent) groups of the MUS HG. *M. m. wagneri* is a taxon historically identified through morphological, ecological and geographical criteria with a range that covers Uzbekistan, Tajikistan, Turkmenistan and Kazakhstan (Schwarz & Schwarz 1943). It is a morphologically heterogenous taxon and based on matrilineal genetic traits described here, and biparentally-inherited markers described elsewhere, the population can probably be subsumed under *M. m. musculus,* in a trinomial classification (Milishnikov *et al.* 2004; Wilson & Reeder 2005).

The south-eastern house mouse M. m. castaneus (CAS)

From a disputed origin in either the Near East or the northern Indian subcontinent, the south-eastern house mouse dispersed to southern India and Sri Lanka, Southeast Asia and Japan, and in historical times has been transported to some parts of Africa and islands of the Indian and Pacific Oceans (Boursot *et al.* 1993, 1996; Prager *et al.* 1998). Four clades within the HG are described, each with mostly discrete distributions, although with an area of overlap in the north of India/Pakistan and the Persian Gulf coast of Iran/Pakistan. CAS1 and CAS3 are predominantly involved in long distance translocations with regards to presumably introduced populations in recent history. CAS1 is the most widely distributed clade found throughout South India, Sri Lanka, Southeast Asia, coastal Iran and it has also been introduced to the Maldives, Réunion, Uganda, Kenya, and New Zealand. Interestingly, estimates of the timing of expansion of CAS clades (based on MMD and the tau parameter) identified CAS1 as the clade that has most recently expanded, and yet it is the clade that has most successfully colonised the globe (Rajabi-Maham *et al.* 2012). While in general *M. m. castaneus* is the least commensal of all the subspecies - in the tropical environments of Southeast Asia (i.e. CAS1) they are practically obligate commensals, unable to live as free-living form, instead relying on the built environment for resources and protection (Bonhomme & Searle 2012). CAS1 is perhaps the most commensal of the CAS clades and its recent expansion can be linked to the growing urbanisation of areas where it successfully colonises.

Haplotypes from Uganda, Kenya and the Maldives do not show definitive relationships to previously described haplotypes within the CAS1 clade and thus more detailed links involving these populations cannot be drawn. There are no archaeological finds of house mouse in Uganda known to me (this may be an artefact of zooarchaeological study), and the species is reportedly not known in the country until the 1960s (Long 2003). The house mouse may have been a stowaway from Asia in rice that was introduced to Uganda in the early 20th century, again during World War II, or again in the 1960s (Haneishi *et al.* 2013). Invasion of Réunion by house mice is discussed in more detail later.

As well as CAS1, CAS3 has been transported within the western Indian Ocean area. Both Yemen (PP=0.96) and Socotra (PP=0.57) cluster with Iranian sequences in a wider network that is mostly within the clades native range of Iran, and northwestern India and Pakistan (Figure 6.8). The sequences do not appear to be in a derived-ancestral relationship, as may be expected from the close proximity and long history between the two locations.

The western house mouse M. m. domesticus (DOM)

From a homeland in the Near East and Fertile Crescent the western house mouse expanded in tandem with humans *ca.* 10, 000 BP along two main routes to Europe, one via the Mediterranean and the other via the Bosphoros/Black Sea and Asia Minor (Bonhomme & Searle 2012). The Mediterranean route encompassed two house mouse progressions that are supported by both genetic and zooarchaeological data. First, an emergence into the eastern Mediterranean with early farmers and, second, a much later expansion to the western Mediterranean/Western Europe (*ca*. first millennium BC and later) with maritime trade (Bonhomme *et al.* 2011, Cucchi *et al.* 2005). In the last millennia the western house mouse has been widely transported so that it is now the most widely globally distributed of all the major subspecies, having colonised parts of Eurasia, Southeast Asia, Africa, Australia and the Americas, as well as many islands (Bonhomme & Searle 2012).

The lack of broad secondary structure shown in the NN network (compared to, for example, the four groups of CAS or the two groups of MUS), and the explosive terminal radiation at the end of the DOM branch (also shown on the Neigbour-Net network) support Bonhomme *et al.*'s (2011) hypothesis that the expansion of *M. m. domesticus*, or at least the maternal lineage, is a relatively recent phenomenon of regionalised dispersals from a centralised core (i.e. Near East and Fertile Crescent). Here, I discuss the geographic distribution of the western house mouse primarily as it relates to Africa and Arabia.

I did not have any new sequences clustering with clade 6 but these are known from Lebanon, Syria, Tunisia, Cyprus and Italy. The signature of Clade 6 dispersal, at least to Tunisia, is very much that of a Phoenician or Roman maritime trade expansion in the late first/early second millennium BC, as suggested by Bonhomme *et al.* (2011).

Clade A is also found in the Near East but is distinguished from clade 6 by a conspicuous absence of central Mediterranean locations (e.g. Cyprus and Italy, which are relatively well-sampled) and has a more widespread north African distribution. I add samples from Jordan, Egypt, Iraq, and the Persian Gulf, and the clade is also known from Turkey, Bulgaria, Greece, Israel, Syria, Lebanon, Iran, Morocco, Tunisia, Kenya and an outlier in the Philippines. It probably followed a number of routes to expand its distribution into Southwest Asia, the Arabian Peninsula and North Africa, and these are discussed below.

Clade A type mice probably followed a terrestrial route from the Fertile Crescent through the Sinai Peninsula into Egypt. Determining the legitimate occupation of ancient Egypt by house mice is difficult due to morphological bone similarities between commensal and wild species, and their proclivity for burrowing, which makes them potentially intrusive into archaeological deposits (see Chapter 2). However, the collection of zooarchaeological finds in Neolithic Egypt and into the late Hyksos Period point to an ancient occupation of the region by house mice. The genetic relatedness and geographical proximity between Fertile Crescent and Egyptian populations, and the level and type of human occupation in Egypt (which would be able to support house mice), in association with zooarchaeological finds, shows an accumulation of evidence for the ancient occupation of Egypt by the species. From genetic evidence I hypothesise that from Egypt the house mouse dispersed to northwestern Africa reaching Tunisia and Morocco either by terrestrial North African routes that probably represent a Neolithic expansion, or via the Mediterranean in the Iron Age (Bonhomme *et al.* 2011). Apart from the site of Kobbat Bent el Rey (Carthage) in Tunisia, house mice deposits are absent in zooarchaeological assemblages in North Africa, at least prior to the modern period. A lack of genetic or zooarchaeological data from intervening Libya makes a terrestrial versus a maritime route difficult to establish, but a stepwise colonisation across the north of Africa from Egypt to

Tunisia following the progression of agriculture in the Neolithic is feasible given the association of house mice with agriculturalists in the Neolithic Levant.

A second, later route from the eastern Mediterranean limits of the Fertile Crescent to ancient Mesopotamia probably took place in clade A. Given the complexity of urbanised human occupation in ancient Mesopotamia and its proximity to the Near East and Iran (where multiple lineages are present by this time) we might expect house mice to be present prior to their first zooarchaeological occurrence in Babylonian Uruk-Warka *ca.* 1800 BC. However, current data suggests a relatively late colonisation of ancient Mesopotamia by house mice, which may reflect a later replacement process.

Clade A populations also occur on the northern and southern coasts of the Persian Gulf. The first zooarchaeological occurrence of the species in this region occurs in the Mleiha period (*ca.* 300 BC – 1 BC) on the southeastern fringe of the Gulf (Potts 2006). The Mleiha period is characterised by trade with Greece (where clade A mice also occur) and a Greco-Arabian trade network may be responsible for the introduction of mice along the northern coast of the Arabian Peninsula (Potts 2006). Clade F mice, also found in the Persian Gulf region, are mostly associated with European haplotypes. Strictly speaking, the Iranian and Saudi Arabian individual are not within clade F (they are the black dots on the phylogenetic tree in Figure 6.9) and the Egyptian haplotype occupies a basal position within the clade. Their basal positioning suggests an ancient colonisation of the Persian Gulf and Egypt, from where the European haplotypes subsequently derived (probably via an intermediary in the Mediterranean, e.g. Cyprus).

Numerous clades are distributed throughout the African continent, the most common of which are C, D, D1, E and F. For the area under study here, Clade C is distributed in Africa above the equator, more specifically in Morocco, Tunisia, Egypt, Kenya, Cameroon and Senegal. Other locations containing this clade are almost exclusively from locations within the Mediterranean including Spain, Portugal, Turkey, Israel and Cyprus (excluding some long distance migrants and not considering the C1 subclade). Clade D is found in the North of Africa, and D1 in West Africa (and elsewhere, but not discussed here). Clade E is the most widely distributed clade on the African continent and is found in Namibia, South Africa, Zimbabwe, Kenya, Somalia, Senegal, Morocco and Tunisia. It is also found throughout the globe e.g. British Isles, France, Germany, Netherlands, Greece, Canada, New Zealand, Australia, Indonesia, China, and as far south as Kerguelen and the Falkland Islands. Clade F is the variety that was widely transported by the Vikings (Searle *et al.* 2009). In Africa, mice of this clade are found in Kenya and Namibia, and elsewhere they are also found in Scandinavia, the British Isles, France, Cyprus, Australia, New Zealand and the Falkland Islands.

There are clearly multiple, complex colonisations of Africa but, overall, the observed pattern suggests introductions through European networks, probably in the Age of Exploration and later. Clade C suggests an interaction sphere that very much links the Mediterranean with northern and central Africa probably via Portugal and Spain that have variously explored, colonised and traded with the region. Clades E and F are related to more Northern European and Western European haplotypes and probably reflect British, French and Dutch exploration, colonisation, and trade passageways. Zooarchaeological evidence from both West and East Africa is noticeably lacking and while this may well be a false signal due to the lack of preservation, recovery, or study of mouse bones, it may also reflect a true signal of the late introduction of the species to the sub-Saharan continent. In sub-Saharan Africa, the A clade found in Kenya stands out – it is related to haplotypes that are part of an early colonization of Northern Africa and Arabia. Transportation to Kenya is most likely from either Red Sea Egyptian (e.g. Fatimid) connections, or Persian Gulf Iranian-Arabian connections earlier than the European connections described elsewhere. Kenya also hosts CAS1 variants that are found on the coast of Iran. The accumulation of multiple lineages in Kenya that also occur in the Persian Gulf region could reflect the intensity of maritime trade between East Africa and the Gulf from ports such as Siraf in Iran, in the Islamic period.

House mice of Réunion

Mice are documented on Réunion prior to 1754 and presumably colonised the island around this time (documentary evidence of invasion by black rats a century earlier, but no mention of mice, suggests an invasion window in the early to mid 18th century) (Cheke 2010). Haplotype and nucleotide diversity are relatively low in the Réunion population (0.382 ± 0.091 and 0.0099 ± 0.0025, respectively) and the MMD plot shows a evidence of a population bottleneck consistent with colonisation by a few individuals, but no signature of a recent expansion – all of which supports a relatively recent introduction of the house mouse to the island.

Réunion hosts multiple clades of house mice; two *M. m. domesticus* clades (E and C1) and a third *M. m. castaneus* (CAS1) type. Under an island invasion paradigm, Réunion, being a remote island with no native competitors or predators, should be readily invaded by house mice (even if rats were already on the island). The occurrence of multiple lineages on the island therefore suggests either replacement of an existing population (or genetic locus) by a secondary coloniser, or colonisation by multiple lineages almost simultaneously. Female house mice tend not to integrate into established populations and thus gene flow (i.e. maternally-inherited mitochondrial DNA) upon secondary contact is unlikely (Bonhomme & Searle 2012). Ignoring missing nucleotides in sequence data, each of both CAS1 and DOM clade E populations on the island have three haplotypes, all differing by one bp, and a two bp insertion in the case of DOM. Although substitution rates between house mouse lineages may differ, and assuming only one haplotype colonised the island, this indicates an approximately equivalent time of introduction to the island for both lineages.

Current sampling shows CAS1 is distributed in the south central and western parts of the island, and DOM (clade E) is distributed in the central and northeast region, with an area of overlap in Saint-Paul on the west coast. Saint-Paul is the oldest town on the island and was the capital until 1738, when Saint-Denis usurped the role as administrative centre. Given the time frame, Saint-Paul is likely to be the port of introduction for both clades. From there, CAS moved south and west, and DOM moved north and east. However, an alternative scenario, whereby CAS was introduced in the south and DOM in the north and they meet in a zone of secondary contact in Saint-Paul, or another alternative scenario, cannot be ruled out.

Clade C1 (DOM) is currently only represented by one specimen from Saint-Louis in the south, where it overlaps with the distribution of mice in the CAS1 clade. It is difficult to determine whether this represents a modern or historical invasion given the limited data. Its proximity to Pierrefonds Airport and the Port of Saint-Pierre (<10km), two major international ports of entry to the island, may indicate a recent introduction. The haplotype clusters with France and Germany (PP=0.96) and it is presumably an introduction from Europe through trade or tourism. However, the town has been settled by the French since 1719 and an introduction in historical times is not out of the question.

CHAPTER 7

CONCLUSION

Trade in necessities is hard to document, for such staples as foodstuffs and cloth leave no archaeological trace

(Pearson 2007)

The statement above illustrates one area of research where the use of proxy lines of evidence for inferring human trade and contact is invaluable. It is also particularly germane for my research as the biological proxies I employ are all living in and off agricultural products and foodstuffs to some degree, and are transported in cargoes of them, making them good indicators of such exchanges. Furthermore, many ancient commensal populations opportunistically exploited early human agricultural cultures, making them excellent indicators of the diffusion (and intensity) of agricultural subsistence strategies in prehistory. For example, house mice, whose origin of commensalism can be traced to the onset of agriculture in the Levant, can be traced to the emergence of agriculture in other areas (Bonhomme & Searle 2012; Cucchi *et al.* 2012). By extension of their transportability, commensals can be used as indicators of long-distance trade in general.

Primarily, commensal models have successfully been used to reconstruct prehistoric and historic human population movements and exchanges in the Pacific and North Atlantic/Europe, where they have both validated and challenged standing theories of human expansion and contact (see Storey *et al.* 2013 and Jones *et al.* 2013 for an overview). However, using commensals as proxies for inferring human mobility and contact in the Indian Ocean has not been substantially explored. To address this, I examined the phylogeographic patterns and zooarchaeological distributions of three widespread, important commensal species, focusing mainly on the Western Indian Ocean region.

The value of the commensal approach, as it is used in my research, is reliant upon phylogeographic inference and zooarchaeological data. Previously, genetic and geographic relationships of the Asian house shrew had only been studied in South and Southeast Asia (e.g.Yamagata *et al.* 1990; Meegaskumbura *et al.* 2010). By adding data for the Western Indian Ocean and understudied parts of South Asia, clear geographical groupings of genetically related sequences have emerged and they have proven to be a useful indicator of human-abetted translocation. However, zooarchaeological data for the species is practically absent and this limits inferences that can be made regarding the absolute dating of a species' occurrence in a location. In contrast, the black rat shows low genetic variation at the level of genetic resolution used in this study and thus only shallow geographic and genetic clustering is apparent, but there is a relatively abundant subfossil record for the species that has permitted ancient DNA analysis. The house mouse has clear phylogeographic clustering and an ever-growing database of geographically distributed haplotypes that has allowed a more complete framework to evolve, and thus stronger historical inferences to be drawn. Their zooarchaeological presence falls between the other two species but their proclivity for burrowing means they are often considered intrusive in archaeological deposits. Direct-dating for potentially intrusive mice remains (and for micromammals generally) would significantly enhance my understanding of the geographic distribution of the species through time.

The overall pattern that emerges when synthesising mitochondrial phylogeographies for the three species studied here is that the eastern Indian Ocean (the eastern side of Bay of Bengal and island Southeast Asia extending into the Pacific) and the western Indian Ocean (Africa, islands of the western Indian Ocean and the Arabian peninsula) operate within different translocation spheres with respect to commensal dispersals. Eastern forms are predominantly *Rattus tanezumi, R. rattus III and R. rattus IV, M. m. castaneus* and Clade C_{IV} Asian house shrews, whereas the western forms are predominantly *R. rattus I, M. m. domesticus/gentilulus* and Clade C_{II} and C_{III} Asian shrews. An area of overlap is seen in the South Asia, the putative origin of all three species. This does not necessarily follow what we would expect from human historical and archaeological accounts as the two areas are far more connected throughout history than is evidenced by commensal phylogeographies. From around 500 BC there is archaeological evidence of the interaction between Northern India and countries in the Bay of Bengal in the form of ceramics. Northern Black Polished Ware is a luxury ceramic associated with the urbanised states of Northern India in the first millennium BC. In the latter half of the first millennium BC it is found on coastal sites in Eastern India and Sri Lanka (Gupta 2007). Rouletted Ware, another form of ceramic, was manufactured in the late BC/early AD era, probably in Tamil Nadu in Southern India, and is found contemporaneously in sites in Bali, Java and Vietnam (Bellina & Glover 2004).

Prehistoric evidence for Southeast Asia's connection with Africa comes in the form of biological translocations. Murdoch (1959) first proposed the import of a 'tropical food kit' from Southeast Asia via a coastal route to Africa that consisted of banana (*Musa* x *paradisiaca*), greater yam (*Dioscorea alata*) and taro (*Colocasia esculenta*). These species are all Southeast Asian in origin but are now widespread throughout Africa suggesting an ancient introduction. The absence of these species in intervening coastal areas such as Northern India and Arabia suggest a coastal route is not the most parsimonious hypothesis, and these species may in fact be trans-oceanic introductions. Zooarchaeological and genetic evidence for the commensal species studied here suggest that they were introduced in a much later period than the tropical crops listed above. Although yet to be AMS dated, three bones of *Rattus tanezumi* from Songo Mnara, a 15th – 16th century town on the Kilwa coast of Tanzania are the first evidence of introductions of these species from a putative Southeast Asian origin. Contact between Southeast Asia and the East African coast is known in this period. Admiral Zheng He visited the Kilwa coast on a voyage of expedition in the early 15th century AD passing through Java and peninsular Malaysia to reach the Indian Ocean (Hsu 2008). Secondly, the Portuguese economic and missionary incursion of the Indian Ocean (in particular India and to some extent Southeast Asia) targeted Kilwa in the early 16th century AD, attacking and seizing the trading nexus, and for a short time maintaining control over the region (Alpers 1975). Prior to these historical accounts, Chinese celadon ceramics found at Mapungubwe Hill, an Iron Age site in the Limpopo valley, South Africa, show connections with Southeast Asia in the 13th-14th Century AD (Prinsloo *et al.* 2005). Chinese ceramics (e.g. Changsha) are found even earlier on coastal sites in East Africa probably reaching the coast in the early 9th century AD (Crowther *et al.* 2013).

Why are we lacking signals for Southeast Asian introductions of commensals to the Western Indian Ocean? This is of particular interest in more recent history as with an increase in technology, trade and connectivity over time, we expect a correlated increase in the introductions of non-native species as propagules are transported more frequently and in higher numbers in larger vessels over longer distances. Clearly, the broad-scale geographic level at which many phylogeographic studies of commensals operate simplify models of translocation that are otherwise complex. However, even with increased local density of sampling (e.g. Madagascar or Réunion in my study) populations can remain narrowly monophyletic, or exhibit low diversity. Boivin *et al.* (2013) note the apparent monophyly of Madagascar mice, which seems at odds with multitude of networks that Madagascar was involved in prior to, during, and after the house mouse was purportedly introduced to the island. It is entirely possible that introduced mice in the northern region of Madagascar (where there is a lack of samples) are distinct from the *M. m. gentilulus* lineage found elsewhere on the island, as seen in the phylogeography of the black rat, and a separate invasion pathway linking Madagascar with East Africa is therefore plausible.

An increasingly complex (although not necessarily clarifying) picture may emerge with the addition of multiple genetic markers for historical inference. For example, Nunome *et al.* (2010) found diversity, which was obscured in the mitochondrial DNA, in the nuclear genome of some Japanese wild mice *Mus musculus* due to introgression of *M. m. castaneus,* and to a lesser extent *M. m. domesticus,* into a predominantly *M. m. musculus* haplotype. From their data they were able to reconstruct an ancient hybridisation event that occurred in Japan between introduced populations from South China and the Korean peninsula probably in parallel with agricultural developments in East Asia. Methodological issues aside, there are also biological parameters that contribute significantly to the secondary colonisation of sink areas by commensals and thus influence genetic patterns. The species studied here tend to be territorial, although each species, and even populations within species, may exhibit different behaviours in response to interlopers depending on their behaviours (e.g. hierarchies in rats and mice, or solitary lifestyles in shrews). Therefore, the likelihood of successful secondary invasion by an interloper is low, and it follows then that they do not contribute significantly (or at all) to the local gene pool. Conversely, low diversity may also be a signal of population (or gene) replacement, or dominance, through recolonisation that has occurred following a population bottleneck or extinction event. Such a scenario was suggested by Kurachi *et al.* (2007) for the Asian house shrew, whereby a rapid replacement of the mitochondrial genome occurred following a secondary invasion of female house shrews in Southeast Asia.

When a species is invasive (as are commensals) it is very often a very clear marker of human cultural contact as introduction to a new area is often obligately facilitated by humans through transport. Commensal species therefore, are a particularly useful marker of contact for island archaeology, which seeks to understand island-island and island-mainland connections, as well as the impact of humans on island ecosystems. One arena where a picture of the human-abetted introduction to islands of multiple synanthropic, commensal and introduced species is emerging is Madagascar (and nearby Comoro Islands). Madagascar is a unique natural laboratory for understanding the colonisation history of invasive species as it is a remote island that can only be reached through long-distance sea-borne (and now air-borne) transport, for terrestrial species at least. Further, the vast majority of native fauna on the island is endemic, and they offer little competition or threat to introduced species. Phylogeographies of the black rat and house mouse on Madagascar have both identified invasion pathways that are probably linked to Arab trade in the late first millennium AD to mid second millennium AD and here I add Asian house shrews to the list of potential introductions (Duplantier *et al.* 2002; Tollenaere *et al.* 2010).

Linguistic, cultural and human genetic affinities suggest an Austronesian element to the peopling of Madagascar (Alibert 2008; Cox *et al.* 2012). The stick zither and leaf-funnel clarinet are instruments found on Madagascar that closely resemble Austronesian ones, and various cultural practices are also shared between the two regions e.g. the use of canoes as coffins, and secondary burial (Alibert 2008, Blench 1996). The linguistic evidence shows that Malagasy, the most common language on the island, is of the Austronesian language family and is related to Barito that is spoken in Southeast Borneo (Adelaar 2009). Post-migration development of the language shows secondary contact between the first migrants and Bantu speakers in the form of nominal prefixes *ki*- and *tsi*- and the future tense markers *hu* and *h*-, among other lexical borrowings (Adelaar 2010). Human genetic studies (Y chromosomal polymorphisms and mitochondrial sequence diversity) show that the Malagasy are of approximately equal African and Indonesian descent, however some groups on the southeast coast also have genetic contributions from Arab origins (e.g. the Onjatsy, the ZafiRaminia, and the Antemoro) (Hurles et al. 2005; Capredon *et al.* 2013).

It is perhaps surprising then that we do not recover such a signal in commensal introductions that are abundant and native to Southeast Asia. One such line of evidence comes from the wild Malagasy pygmy shrew *Suncus madagascariensis*. Once thought to be endemic it has recently been shown to be a recent sea-borne introduction, probably from Asia, that is part of the Etruscan shrew *Suncus etruscus* complex (Omar *et al.* 2011). Subfossil *S. madagascariensis* is found in Andrahomana cave in the southeast of the island in conjunction with house mice *Mus musculus* and rats *Rattus* sp in a post-human setting (Burney *et al.* 2008). Unfortunately, stratigraphic mixing between the cave deposits means the bones cannot be securely dated. Socotran *S. etruscus* may be more closely related to Malagasy *S. madagascariensis* (part of the *S. etruscus* complex) than more geographically proximate *S. etruscus* populations (although the author does not elucidate by what characteristics this assumption is made) (Hutterer 2005). Could this be evidence of an ancient Indian Ocean network that connected Asia, Arabia and Madagascar? The *Periplus of the Erythrean Sea* identifies an Indian trading post on Socotra (“Dioscurides”) that was linked to the wider classical network incorporating Arabia, specifically Yemen, down the coast of East Africa, possibly as far as Madagascar (Casson 1989). Socotra maintained its key role as a node in trading networks throughout the Classical period, and beyond, and this could be the corridor by which *S. etruscus* has travelled.

Asian zebu cattle *Bos (primigenius) indicus* are found in archaeological deposits on both the Comoro Islands and northwest Madagascar dated to approximately the 9th-10th centuries AD (Boivin *et al.* 2013a). A small number of propagules of zebu cattle were probably first introduced to the continent through the Horn of Africa in the second millennium BC; a second incursion, probably via a maritime pathway in the Swahili period, contributed considerably to the genetic make-up of contemporary populations (Gifford-Gonzales and Hanotte 2011). Hanotte *et al.* (2002) conducted an autosomal microsatellite study of cattle and found substantial introgression of Asian zebu alleles into Malagasy cattle, more so than in mainland populations suggesting importation from a source other than East Africa. As with mainland Africa, biological translocations of cultivated plants to Madagascar such as the greater yam, bananas, taro, rice (*Oryza sativa*), and the coconut tree (*Cocos nucifera*) are of South or Southeast Asian origin (Jones et al. 2013).

Trade networks (and other networks such as migration, slave or pilgrimage routes) in the period of Arab trade (*ca.* late first millennium AD - mid second millennium AD) are probably responsible for the introduction of house mice, black rats, and Asian house shrews to the Western Indian Ocean region. From India or Pakistan, the species may have travelled to Arabia with merchant vessels (in some cases probably prior to the Arab period), followed by transport to East Africa and islands in the southwestern Indian Ocean, and interconnective translocations between the latter two regions. The Swahili coast was a central part of early maritime networks that operated in the Indian Ocean from the first millennium AD. Although trade along the coast is indicated from classical sources in Roman times, it is not until the 7th century AD that regular trade between African communities and Islamic traders began. Within the Swahili coastal network that extended from the East African coast, near islands and as far as Madagascar, the movement of both goods and people was unfettered. The Arabic influence on the Swahili people is evident in the language, which is of Bantu origin but contains loan words from Arabic, and this is considered to be a relatively late development probably during the period when Islamic influence was great (Spear 2000). A number of sites along the East African coast are associated with the emergence of the urbanised Swahili, including Shanga, Unguja Ukuu, Kilwa, and Chibuene. Evidence of Arabic trade from these sites comes in the form of pottery of Sassanian Islamic ware, and stone and glass beads originating in the Gulf and India (LaViolette 2008). The Swahili exported various products, including luxury goods, some of those commonly traded were tortoise shell, ambergris, ivory and slaves (Horton and Middleton 2000).

A European incursion in the Age of Exploration and colonial period is evident in rats of clade A, *M. m. domesticus* (particularly the C/C1, D/D1, E and F clades), and probably shrews of clade C_V - for rats and mice at least this expansion was on a global scale. French settlement of Réunion, a remote island in the Mascarenes, in the 17th century marks a relatively late human colonisation compared to coastal and near islands in the Indian Ocean region. In the main, we would there expect a correspondingly late colonisation of the island by commensals and other introduced biota. Historical documents identify black rat, house mice and Asian house shrew introductions in the 17th and 18th centuries, after human settlement (Cheke 2010). My genetic data supports a relatively late colonisation of the island by the above species, and the pathway could have been via European East Indies connections. Since settlement, Réunion has seen an influx of immigrants of Mozambican, Malagasy, Comorian, Indian and Chinese origin, and these interactions have undoubtedly influenced the biota of the island. House geckos *Hemidactylus spp.* have been introduced to the island and one species, *H. brooki*, not only shows genetic affinities with neighbouring Mascarene Islands (Mauritius and Rodrigues), but also Moheli, in the Comoros (Vences *et al.* 2004).

The ancient distributions and population histories of micromammals are often based on zooarchaeological data, which can be problematic for many reasons, and can result in patchy or erroneous data. A study evaluating the zooarchaeological identifications of lagomorphs (differentiated largely on size) through parallel ancient DNA sequencing analysis found that

10% of identifications were incorrect; mostly in bones of intermediate size (Yang *et al.* 2005). Furthermore, the researchers discovered a previously unidentified (within their assemblage) taxon. Size variation is also a complicating factor for differentiating bones of *Rattus exulans*, a popular commensal proxy for human migrations in the Pacific, from rodent congeners, and researchers advocate identification through ancient DNA analysis (Matisoo-Smith & Allen 2001). There is no standard convention for reporting faunal remains in archaeological literature and most of the time no outline of the criteria used to identify micromammals is given. Genetic identification of ancient DNA using short fragments of mitochondrial DNA goes someway to overcoming these obstacles and I have positively identified several remains from East Africa, Egypt and the western Indian Ocean. Consequently, I was also able to determine forms that had been identified as black rats (in the broad sense) but were not. Clearly, the zooarchaeological identifications of micromammals in the current literature needs to be re-examined, especially considering the utility of micromammals for inferring not only human mobility, but also paleoenvironments and paleoecologies is being realised (D. Avery 2007; Weissbrod *et al.* 2014). Cucchi *et al.* (2005) have critically examined the published data for house mice in the Western Mediterranean using a number of metrics and this goes some way to addressing the material for one of species I have discussed here (Cucchi *et al.* 2005a).

The benefit of ancient DNA analysis for phylogeographic studies resides in the direct correlation between genetic and zooarchaeological information, which introduces a robust temporal element to phylogeographic frameworks. Ancient DNA studies can uncover cryptic population histories, such as where populations may have gone extinct, suffered a population bottleneck, or population turnover has occurred, that are obscured by modern (or at least subsequent) replacements. Although this study does not have high resolution modern data for the region (and therefore there may be sampling biases), I recovered a previously unknown (for the area) *Rattus* lineage on the Tanzanian coast from around the mid second millennium AD. The lineage is predominantly South/Southeast Asian and its presence in East Africa during a period when the region was dominated by Arab trade suggests direct contacts with the eastern Indian Ocean region. Ancient DNA also allows me to examine colonisation-extinction dynamics and, conversely, population persistence through time. Many of the ancient samples (bone) I examined were of identical haplogroups (and even haplotypes in some cases) to modern (tissue from freshly killed animals) and historical (museum skins and osteocrusts) ones. This attests to the black rat's stability over time at the population level and confirms the lack of secondary invasion in many locations for the species. Moreover, it indicates that distributional patterns inferred from phylogeographies generated from modern (or historical) data, for the mitochondrial genome at least, reflect ancient patterns.

All three species studied here communicate a multitude of diseases to both humans and livestock and are implicated in epidemics and pandemics that have had serious health, social and economic ramifications. Here, I highlight one disease - the plague and its etiologic agent the bacterium *Yersinia pestis* – that illustrates the role of commensals in the spread of disease. One of the best-known pandemics is the spread of the bubonic plague in the 14th century AD from the steppes of Asia to Europe along the Silk Road, purportedly by the black rat *R. rattus*. The disease, which came to be known as the Black Death, devastated Europe killing around one third of the population and caused massive societal and demographic upheaval among the living as inhabitants fled ports and cities *en masse*, farmers left their fields untended and people lost faith in religion (Kelly 2005). The modern pandemic, which began in China in 1855, resulted in 12 million deaths and is arguably ongoing, with current plague foci in South and South East Asia, parts of Africa, the USA and South America (Morelli *et al.* 2010). Plague is maintained in rodent and other small mammal reservoirs and the black rat, house mouse and Asian house shrew all act as vertebrate amplifying hosts of the bacterium (Velimirovic 1972). Translocation of commensal hosts will therefore also readily translocate the bacterium. In its new environment the bacterium can be maintained in the invading host population, or may be transferred to other established small mammals. The commensal's role in plague transmission is multipart. Due to their unique ecology which sees them navigate a spectrum of domains (e.g. switching between feral/urban lifestyles or foraging in wild/domestic environments) they act as a vector between the "sylvatic" (or wild) and "urban" plague cycles (Chamberlain 2004). Simplified, in areas of plague foci the plague bacterium is maintained in wild rodents or other small mammals. A commensal that comes in to contact with a wild reservoir and is infected by a flea bite can then transmit the disease to humans via the flea vector as it enters the urban or rural environment.

Future directions

This study, and other bioproxy studies would benefit from deepening the genetic and temporal aspects of the study, to get higher resolution, long-term data. The majority of mammalian commensal model phylogeographic studies have thus far employed single gene mitochondrial markers. With the rapid development of sequencing technologies, whole genome, or high resolution single nucleotide polymorphism (SNP), data can be collected relatively quickly and cheaply. Sequencing ancient DNA as part of these high resolution studies is also feasible. To this end, I have trialed DNA from museum specimens (skins and osteocrusts) on a high density SNP bead array (Appendix 7.I). With deeper coverage of the genome across time, researchers will be able to investigate demographic histories of populations in more detail, for example, using the Approximate Bayesian Computation approach (Beaumont 2010).

REFERENCES

Adelaar, A. (2009) Towards an Integrated Theory About the Indonesian Migrations to Madagascar. In: *Ancient Human Migrations* (ed Pergine P, *et al.*), pp. 149–172. University of Utah Press, Utah.

Adelaar, A (2010) The Amalgamation of Malagasy. In: *A Journey Through Austronesian and Papuan Linguistic and Cultural Space. Papers in Honour of Andrew K. Pawley* (ed Bowden J, Himmelmann N, Ross M) Pacific Linguistics, Canberra.

Agrawal DP (2007) *The Indus Civilization: An Interdisciplinary Perspective*. Aryan Books International, Delhi.

Alibert C (2008) Austronesian Migration and the Establishment of the Malagasy Civilization: Contrasted Readings in Linguistics, Archaeology, Genetics and Cultural Anthropology. *Diogenes*, **55**, 7–16.

Alpers EA (1975) *Ivory and Slaves: Changing Pattern of International Trade in East Central Africa to the Later Nineteenth Century.* University of California Press, Berkeley and Los Angeles.

Alur KR (1969) Animal Remains from Sanganakallu. In: *Excavation at Sangankallu, 1964-65; Early Neolithic House at Bellary* (eds Ansari Z., Nagarajarao MS), pp29-38. Deccan College Pune.

Alur KR (1980) Faunal Remains from Vindhyas and the Ganga Valley. In: *Beginnings of Agriculture* (eds Sharma G, Misra VD, Mandal D, B.B. M, Pal JN), pp. 220–226. Abinash Prakashan, Allahabad.

Alur KR (1990) *Studies in Indian Archaeology andPpalaeontology*. Shrihari Prakashana on behalf of Alur Venkatarao Pratishthana, Dharwad.

Alpers EA (1975) *Ivory and Slaves: Changing Pattern of International Trade in East Central Africa to the Later Nineteenth Century.* University of California Press, Berkeley and Los Angeles.

Anderson A (2009) The Rat And The Octopus: Initial Human Colonization And The Prehistoric Introduction Of Domestic Animals To Remote Oceania. *Biological Invasions* **11**, 1503–1519.

Animal Diversity Web (2012) University of Michigan. www.animaldiversity.ummz.umich.edu

Anisimova M, Gascuel O (2006) Approximate Likelihood-Ratio Test for Branches: A Fast, Accurate, and Powerful Alternative. *Systematic Biology*, **55**, 539-552.

Aplin K (2003) Evolutionary Biology of the Genus Rattus: Profile of an Archetypal Rodent Pest. In: *Rats , Mice And People: Rodent Biology and Management* (eds Singleton EGR, Hinds LA, Krebs CJ, Spratt DM), pp. 487–498. Australian Centre for International Agricultural Research, Canberra.

Aplin KP, Suzuki H, Chinen AA *et al.* (2011) Multiple Geographic Origins of Commensalism and Complex Dispersal History of Black Rats. *PLoS ONE* **6**: e26357.

Armitage L (1994) Unwelcome Companions: Ancient Rats Reviewed. *Antiquity* **68**, 231-240.

ASI (1957) Maski 1954: A Chalcolithic Site of the Southern Deccan. *Ancient India: Bulletin of the Archaeological Survey of India Volume 13.*

Atkinson IAE (1985) The Spread Of Commensal Species Of Rattus To Oceanic Islands And Their Effects On Island Avifaunas. In: *Conservation of Island Birds* (ed Moors PJ), pp. 35–73. ICBP Technical Publication, Cambridge, UK.

Auffray J-C, Tchernov E, Nevo E (1988) Origine Du Commensalisme De La Souris Domestique (*Mus musculus domesticus*) Vis-A-Vis de l'Homme. *Compte Rendus de l'Académie des Sciences Paris*, **307**, 517–522.

Auffray J, Vanlerberghe F, Britton-Davidian J (1990) The House Mouse Progression In Eurasia: A Palaeontological And Archaeozoological Approach. *Biological Journal of the Linnean Society*, **41**, 13–25.

Avise JC, Arnold J, Ball RM *et al.* (1987) Intraspecific Phylogeography: The Mitochondrial DNA Bridge Between Population Genetics and Systematics. *Ecology*, **18**, 489–522.

Azizi MH, Azizi F (2010) A history of the human plague in Iran. *Archives of Iranian medicine*, **13**, 563–9.

Bandelt HJ, Forster P, Röhl A (1999) Median-Joining Networks For Inferring Intraspecific Phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Banerjee NR (1981) Animal Remains from Bharatpur, District Burdwan, West-Bengal, India. *Records of the Zoological Survey of India*, **97**, 193–201.

Banerjee NR (1986) *Nagda (1955-57). Memoirs of the Archaeological Survey of India 85*. Archaeological Survey of India, New Delhi.

Barnes SS, Matisoo-Smith E, Hunt TL (2006) Ancient DNA of the Pacific Rat (*Rattus exulans*) from Rapa Nui (Easter Island). *Journal of Archaeological Science*, **33**, 1536–1540.

Bartlett JMS, Stirling D (Eds.) (2003) *Methods in Molecular Biology (Vol 226): PCR Protocols*. Humana Press, New Jersey.

Bastos AD, Nair D, Taylor PJ *et al.* (2011) Genetic Monitoring Detects an Overlooked Cryptic Species and Reveals the Diversity and Distribution of Three Invasive *Rattus* Congeners in South Africa. *BMC Genetics*, **12**, 26.

Bellina B, Glover I (2004) The Archaeology of Early Contact with India and The Mediterranean World, from the Fourth Century BC to the Fourth Century AD. In: *Southeast Asia from Prehistory to History* (ed Glover I, Bellwood P), pp. 68-88. Routledge/Curzon, London.

Beaujard P (2005) The Indian Ocean in Eurasian and African World-Systems Before the Sixteenth Century. *Journal of World History*, **16**, 411-465.

Bibb MJ, Van Etten RA, Wright CT, Walberg MW, Clayton DA (1981) Sequence and Gene Organization of Mouse Mitochondrial DNA. *Cell*, **26**, 167–180.

Blench R (1996) The Ethnographic Evidence for Long-Distance Contacts Between Oceania and East Africa. In: *The Indian Ocean in Antiquity* (ed Reade J), pp 417–438 Kegan Paul, London

Boessneck J (1976) *Tell El-Dab'a III: Die Tierknochenfunde 1966-1969*. Austrian Academy of Sciences, Wien.

Boessneck J (1977) Tierknochenfunde aus Isan Bahriyat (Isin). In: *Isin Isan Bahriyat I* (ed Hrouda B), pp. 111–133. Verlag der Bayerischen Akademie der Wissenshaften, Munich.

Boessneck J (1988) *Die Tierwelt des Alten Ägypten Untersucht Anhand Kulturgeschichtlicher und Zoologischer Quellen*. C H Beck, Munich.

Boessneck J, von den Driesch A *Tell el-Dab'a VII, Untersuchungen der Zweigstelle Kairo 10*. Austrian Academy of Sciences, Wien.

Boessneck J, von den Driesch A (1990) Tierreste aus der Vorgeschichtlichen Siedlung von El-Omari bei Heluan/ UnterAgypten. In: *El-Omari: A Neolithic Settlement and Other Sites in the Vicinity of Wadi Hof, Helwan* (eds Debono F, Mortensen B), pp. 99–107. Phillip von Zabern, Mainz am Rhein.

Boessneck J, von den Driesch A, Von Steger U (1984) Tierknochenfunde in Uruk-Warka. *Baghdader Mitteilungen*, **15**, 150–189.

Boivin N, Blench R, Fuller DQ (2009) Archaeological, Linguistic and Historical Sources on Ancient Seafaring: A Multidisciplinary Approach to the Study of Early Maritime Contact and Exchange in the Arabian Peninsula. In: *The Evolution of Human Populations in Arabia, paleoenvironments, prehistory and genetics* (eds Petraglia M, Rose J), pp 251-278. Springer, Dordecht

Boivin N, Crowther A, Helm R, Fuller DQ (2013a) East Africa and Madagascar in the Indian Ocean World. *Journal of World Prehistory*, **26**, 213–281.

Bonhomme F, Orth A, Cucchi T *et al.* (2011) Genetic Differentiation of the House Mouse Around the Mediterranean Basin: Matrilineal Footprints of Early and Late Colonization. *Proceedings of the Biological Sciences/The Royal Society*, **278**, 1034–43.

Bonhomme F, Searle JB (2012) House Mouse Phylogeography. In: *Evolution of the House Mouse* (ed Macholan M), pp. 278–296. Cambridge University Press, Cambridge, UK.

Boursot J, Auffray P, Britton-Davidian J, Bonhomme F (1993) The Evolution of House Mice. *Ecology*, **24**, 119–152.

Boursot P, Din W, Anand R *et al.* (1996) Origin and Radiation of the House Mouse: Mitochondrial DNA Phylogeny. *J. Evol. Biol.*, **9**, 391–415.

Bradley R, Baker R (2001) A test of the Genetic Species Concept: Cytochrome-b Sequences and Mammals. *Journal of Mammology*, **82**, 960–973.

Braithwaite RW (1980) The Ecology of *Rattus lutreolus III*. The Rise and Fall of a Commensal Population. *Australian Wildlife Research*, **7**, 199–215.

Brothwell D (1981) The Pleistocene and Holocene Archeology of the House Mouse and Related Species. In: *Symposium of the Zoological Society of London. Vol. 47: The Biology of the House Mouse* (ed Berry RJ), pp. 1–13. Academic Press, London.

Burney DA, Vasey N, Godfrey LR *et al.* (2008) New Findings at Andrahomana Cave, Southeastern Madagascar. *Journal of Cave and Karst Studies*, **70**, 13–24.

Byrd B (1994) From Early Humans to Farmers and Herders Recent Progress on Key Transitions in Southwest Asia. *Journal of Archaeological Research*, **2**, 221–253.

Caloi L, Compagnoni B (1977) I Mammiferi. In: *La Citta Bruciata del Deserto Salato* (ed Basaglia AO), pp. 183–213. Maniago, Venice.

Campbell Island: New Zealand's Subantarctic Islands (n.d.). Retrieved from http://www.doc.govt.nz/conservation/land-and-freshwater/offshore-islands/new-zealands-subantarctic-islands/campbell-island/

Capredon M, Brucato N, Tonasso L *et al.* (2013) Tracing Arab-Islamic Inheritance in Madagascar: Study of the Y-chromosome and Mitochondrial DNA in the Antemoro. *PLoS ONE*, **8**, 1–15.

Carenti G, Wilkens B (2008) Terrestrial Fauna and Marine Produce in Sumhuram. In: *A Port in ArabiaBbetween Rome and the Indian Ocean (3rd C.BC. - 5th C.AD) Khor Rori Report 2* (ed Avanzani A), pp. 477–548. Erma di Bretschneider, Rome.

Casson L (1989) *The Periplus Maris Erythraei*. Princeton University Press, Princeton (NJ).

Chami F (2002) East Africa and the Middle East relationship from the first millennium BC to about 1500 AD. *Journal des Africanistes*, **72**, 21–37.

Chang AC, Lin J, Lin L, Yu JY, Chang C (1999) Annual Reproductive Patterns of Female House Shrew, *Suncus murinus*, in Taiwan. *Zoological Science,* **16**, 819–826.

Cheke A (2010) The Timing Of Arrival Of Humans And Their Commensal Animals On Western Indian Ocean Oceanic Islands. *Phelsuma*, **18**, 38–69.

Chernousova NF (2001) Specific Features of the Dynamics of Murine Rodent Communities under the Effects of Urbanization: 1. Dynamics of Species Composition and Abundance. *Russian Journal of Ecology*, **32**, 122–125.

Chittick N (1974) *Kilwa: An Islamic Trading City on the East African Coast*. British Institute in East Africa, Nairobi.

Chowdhury F, Chowdhury A (2012) Pyrosequencing-An Alternative to Traditional Sanger Sequencing. *American Journal of Biochemistry and Biotechnology*, **8**, 14–20.

Churcher CS (1999) The Neolithic Fauna From Archaeological Contexts In Dakhleh Oasis, Egypt. *Archaeozoologica*, **10**, 47–54.

Cleuzio S, Tosi M (2000) R'as al-Jinz and the Prehistoric Coastal Cultures of the Ja'alan. *Journal of Oman Studies*, **11**, 19–73.

Conroy CJ, Rowe KC, Rowe KMC *et al.* (2012) Cryptic Genetic Diversity in Rattus of the San Francisco Bay Region, California. *Biological Invasions*, **15**, 741–758.

Cox M, Nelson M, Tumonggor M, Ricaut F, Sudoyo H (2012) A Small Cohort of Island Southeast Asian Women Founded Madagascar. *Proceedings of the Royal Society B*, **279**, 2761–2768.

Crowther A, Horton M, Kotarba-Morley A *et al.* (2013) *Sealinks Project 2011: Report on Fieldwork at Fukuchani and Unguja Ukuu, Zanzibar. July 18-31, 2011*. Oxford.

Cucchi T, Auffray J-C, Vigne J-D (2005a) First Occurrence of the House Mouse (*Mus musculus domesticus* Schwarz & Schwarz, 1943) in The Western Mediterranean: A Zooarchaeological Revision of Subfossil Occurrences. *Biological Journal of the Linnean Society*, **84**, 429–445.

Cucchi T, Auffray J, Vigne JD (2012) On the Origin of the House Mouse Synanthropy And Dispersal in the Near East And Europe: Zooarchaeological Review and Perspectives. In:

Evolution of the House Mouse (ed Macholan M), pp. 65-93. Cambridge University Press, Cambridge, UK.

Cucchi T, Vigne J-D, Auffray J-C (2005b) The Passive Transportation of the House Mouse (*Mus musculus domesticus*) to Cyprus: New Indirect Evidence of Intensive Neolithic Navigation in Eastern Mediterranean. In: *Archaeozoology of the Near East VI. Proceedings of the Sixth International Symposium on the Archaeozoology of Southwestern Asia and Adjacent Areas* (eds Buitenhuis H, Choyke AM, Martin L, Bartosiewicz L, Mashkour M), pp. 61–77. ARC-Publicaties 123, Groningen.

Cuyler Young Jr. T, Levine LT (1974) *Excavations of the Godin project: second progress report*. Royal Ontario Museum of Art and Archaeology, Toronto.

D. Avery (2007) Micromammals as Palaeoenvironmental Indicators of the Southern African Quaternary. *Transactions of the Royal Society of South Africa*, **62**, 17–23.

Darvish J, Orth A, Bonhomme F (2006) Genetic Transition in the House Mouse, *Mus musculus* of Eastern Iranian Plateau. *Current Topics in Microbiology and Immunology*, **55**, 349–357.

David R (Ed.) (1997) *Multidisciplinary Research on Ancient Egyptian Mummified Remains*. Manchester University Press, Manchester.

Dawson WR (1924) The Mouse in Egyptian and Later Medicine. *The Journal of Egyptian Archaeology*, **10**, 83–86.

Von den Driesch A, Baumgartner (1997) Die Dpatantiken Tierreste aus der Kobbat Bent el Rey in Karthago. *Archaeozoologica*, **9**, 155–172.

Von den Driesch A, Dockner A (2002) Animal Exploitation in Medieval Siraf, Iran, Based on The Faunal Remains From The Excavations at the Great Mosque (Seasons 1966-1973). *Bonner Zoologische Beiträge*, **50**, 227–247.

Deo SB (1970) *Excavations at Takalghat and Khapa, 1968-69*. Nagpur University, Nagpur.

Deo S., Shastri AM (1974) *Excavations at Bhokardan (Bhogavardhana), 1973*. Nagpur University, Nagpur.

Deshpande-Mukherjee A, Thomas PK, Mohanty RK (2010) Faunal Remains from the Iron Age and Early Historical Settlement at Mahurjari, District Nagpur, Maharashtra. *Man and Environment*, **35**, 87–102.

Dewar RE, Richard AF (2012) Madagascar: A History of Arrivals, What Happened, and Will Happen Next. *Annual Review of Anthropology*, **41**, 495–517.

Din W, Anand R, Boursot P, Darviche D, Dad B (1996) Origin and Radiation of the House Mouse: Clues from Nuclear Genes. *Journal of Evolutionary Biology*, **9**, 519–539.

Driesch AV Den, Boessneck J (1983) A Roman Cat Skeleton from Quseir on the Red Sea Coast. *Journal of Archaeological Science*, **10**, 205–211.

Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W (2002) Estimating Mutation Parameters, Population History and Genealogy Simultaneously From Temporally Spaced Sequence Data. *Genetics,* **161**, 1307–1320.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian Phylogenetics with BEAUti and the BEAST 1 . 7. *Molecular Biology and Evolution,* **29**, 1969-1973.

Dubey S, Koyasu K, Parapanov R, Ribi M (2008a) Molecular Phylogenetics Reveals Messinian, Pliocene, and Pleistocene Colonizations of Islands by North African Shrews. *Molecular Phylogenetics and Evolution,* **47**, 877–882.

Dubey S, Salamin N, Ohdachi SD, Barrière P, Vogel P (2007) Molecular Phylogenetics of Shrews (Mammalia: Soricidae) Reveal Timing of Transcontinental Colonizations. *Molecular phylogenetics and evolution*, **44**, 126–37.

Dubey S, Salamin N, Ruedi M *et al.* (2008b) Biogeographic Origin And Radiation of the Old World Crocidurine Shrews (Mammalia: Soricidae) Inferred From Mitochondrial and Nuclear Genes. *Molecular phylogenetics and evolution*, **48**, 953–63.

Duchêne S, Lanfear R, Ho Y (2014) The Impact of Calibration and Clock-Model Choice on Molecular Estimates of Divergence Times. *Molecular Phylogenetics and Evolution,* **78**, 277-289.

Duplantier J, Orth A, Catalan J, Bonhomme F (2002) Evidence for a Mitochondrial Lineage Originating from the Arabian Peninsula in the Madagascar House Mouse (*Mus musculus*). *Heredity*, **89**, 154–158.

Dutta BC (1984) *Rupar, Ancient Cultural Complex of India*. B.C. Dutta, Calcutta.

Ervynck A (2002) Sedentism or Urbanism? On the Origin of the Commensal Black Rat (*Rattus rattus*). In: *Bones and the Man. Studies in Honour of Don Brothwell.* (eds Dobney K, O'Connor T), pp. 95–109. Oxbow, Oxford.

Esselstyn J, Timm RM, Brown RM (2009) Do Geological or Climatic Processes Drive Speciation in Dynamic Archipelagos? The Tempo and Mode of Diversification in Southeast Asian Shrews. *Evolution: International Journal Of Organic Evolution*, **63**, 2595–610.

Fagan B (1963) The Iron Age Sequence in the Southern Province of Northern Rhodesia. *Journal of African History*, **4**, 157–177.

Forster DW, Al E (2009) Molecular Insights Into The Colonization And Chromosomal Diversification Of Madeiran House Mice. *Molecular Ecology*, **18**, 4477–4494.

Fritts TH, Frittsusgsgov T, Rodda GH (1998) The Role Of Introduced Species In Degradation Of Island Ecosysytems: A Case History Of Guam. *Annual Review of Ecology and Systematics,* **29**, 113-140.

Fu Y-X (1997) Statistical Tests of Neutrality of Mutations Against Population Growth, Hitch-Hiking, and Background Selection. *Genetics*, **147**, 915–925.

Fuller DQ, Boivin N, Hoogervorst T, Allaby R (2011) Across the Indian Ocean: The Prehistoric Movement of Plants and Animals. *Antiquity*, **85**, 544–558.

Gabriel SI, Stevens MI, Mathias L, Searle JB (2011) Of Mice and "Convicts": Origin of the Australian House Mouse , *Mus musculus*. *PLoS ONE*, **6**, 8–13.

Gardner-Santana LC, Norris D, Fornadel C *et al.* (2009) Commensal Ecology, Urban Landscapes, and Their Influence on the Genetic Characteristics of City-Dwelling Norway Rats (*Rattus norvegicus*). *Molecular Ecology*, **18**, 2766–78.

Gibson M, Franke JA, Civil M *et al.* (1978) *Excavations at Nippur: Twelfth Season.* Oriental Institute, Chicago.

Gifford-Gonzalez D, Hanotte O (2011). Domesticating Animals in Africa: Implications of Genetic and Archaeological Findings. *Journal of World Prehistory*, **24**, 1–23.

Glover I, Bellwood P (Eds.) (2004) *Southeast Asia: From Prehistory to History*. Routledge, Abingdon.

Guindon S, Dufayard J, Lefort V *et al.* (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, **59**, 307-21.

Guo W-P, Lin X-D, Wang W *et al.* (2011) A New Subtype of Thottapalayam Virus Carried by the Asian House Shrew (*Suncus murinus*) in China. *Infection, Genetics And Evolution: Journal of Molecular Epidemiology And Evolutionary Genetics In Infectious Diseases*, **11**, 1862–7.

Gupta S (2004) The Bay of Bengal Interaction Sphere (1000 BC – AD 500). *Bulletin of the Indo-Pacific Prehistory Association*, **25**, 21–30.

Gurdasani D, Carstensen T, Tekola-Ayele F *et al.* (2014) The African Genome Variation Project Shapes Medical Genetics in Africa. *Nature.* Published online 03 December 2014 doi:10.1038/nature13997.

Gutiérrez-García T, Vázquez-Domínguez E, Arroyo-Cabrales J *et al.* (2014) Ancient DNA and the Tropics: A Rodent's Tale. *Biology Letters*, **10**: 20140224.

Hadly EA, Ramakrishnan U, Chan YL *et al.* (2004) Genetic Response to Climatic Change: Insights from Ancient DNA and Phylochronology. *PLoS Biology*, **2**, 1600–1609.

Haile J, Larson G, Owens K, Dobney K, Shapiro B (2010) Ancient DNA Typing of Archaeological Pig Remains Corroborates Historical Records. *Journal of Archaeological Science*, **37**, 174–177.

Haneishi Y, Okello SE, Asea G *et al.* (2013) Exploration of Rainfed Rice Farming in Uganda Based on a Nationwide Survey: Evolution, Regionality, Farmers and Land. *African Journal of Agricultural Research*, **8**, 3318–3329.

Hanotte O, Bradley DG, Ochieng JW*, et al.* (2002) African pastoralism: Genetic imprints of origins and migrations. *Science* **296**, 336-339.

Harrisson DL (1970) Some Observations on the Status of the South Arabian House Mouse *Mus musculus gentilulus* Thomas, 1919. (Rodentia: Muridae). *Mammalia*, **34**, 244–247.

Hashemi N (2009) Appendix- Rodent remains. In: *The Mamasani Archaeological Project Stage One* (eds Potts DT, Roustaei K, Petrie CA, Weeks LR), pp. 145–146. Archaeopress, Oxford.

Hashemi N, Darvish J, Mashkour M, Biglari F (2006) Rodents and Lagomorphs Remains from Late Pleistocene and Early Holocene Caves and Rochshelter Sites in the Zagros Region, Iran. *Iranian Journal of Animal Biosystematics*, **2**, 25–33.

Hedge KTM (1992) *Excavation at Nageswar (Gujarat): A Harappan Shell Working Site on the Gulf of Kutch. Maharaja Sayajirao University Archaeology Series*. Department of Archaeology & Ancient History, Faculty of Arts, M.S. University of Baroda.

Helm RM (2000) *Conflicting Histories: The Archaeology of the Iron-Working, Farming Communities in the Central and Southern Coast Region Of Kenya.* Unpublished PhD dissertation, Department of Archaeology and Anthropology, University of Bristol.

Helm R, Crowther A, Shipton C (2012) Exploring Agriculture, Interaction and Trade on the Eastern African Littoral: Preliminary Results From Kenya. *Azania: Archaeological Research in Africa*, **41**, 39-63.

Hingston M, Goodman SM, Ganzhorn J, Sommer S (2005) Reconstruction of the Colonization of Southern Madagascar by Introduced *Rattus rattus*. *Journal of Biogeography*, **32**, 1549–1559.

Ho SYW, Gilbert MTP (2010) Ancient Mitogenomics. *Mitochondrion*, **10**, 1–11.

Ho SYW, Heupink TH, Rambaut A, Shapiro B (2007) Bayesian Estimation Of Sequence Damage In Ancient DNA. *Molecular Biology And Evolution*, **24**, 1416–22.

Horton M, Middleton J (2000) *The Swahili: The Social Landscape of a Mercantile Society*. Wiley-Blackwell, Oxford.

Hsu M-L (2008) Chinese Marine Cartography: Sea Charts Of Pre-Modern China. *Imago Mundi:The International Journal for the History of Cartography*, **40**, 96–112.

Hunt H, Jones M (2008) Pathways Across Asia: Exploring The History Of Panicum And Setaria In The Indian Subcontinent. *Pragdhara*, **18**, 53–68.

Hurles M, Sykes B, Jobling M, Forster P (2005) The Dual Origin of the Malagasy in Island Southeast Asia and East Africa: Evidence from Maternal and Paternal Lineages. *American Journal of Human Genetics*, **76**, 894–901.

Hutterer R (2005) Order Soricomorpha. In: *Mammal species of the world: A Taxonomic and Geographic Reference 3rd ed* (eds Wilson DE, Reeder D). John Hopkins University Press, Baltimore.

Hutterer R, Harrison D (1988) A New Look At The Shrews (Soricidae) of Arabia. *Bonn Zoological Bulletin*, **39**, 59–72.

Hutterer R, Tranier M (1990) The Immigration of the Asian House Shrew (*Suncus murinus*) into Africa and Madagascar. In: *Vertebrates In the Tropics: Proceedings of the International Symposium on Vertebrate Biogeography and Systematics in the Tropics, Bonn, June 5-8, 1989* (eds Peters G, Hutterer R), pp. 309–319. Alexander Koenig Zoological Research Institute and Zoological Museum, Bonn.

Indian Archaeology a review 1996-97. (2002) New Delhi.

The International HapMap Consortium (2003). The International HapMap Project.
Nature, **426**, 789-796.

Ishikawa A, Yamagata T, Namikawa T (1991) An Attempt at Reciprocal Crosses Between Laboratory Strains of Large and Small Musk Shrews (*Suncus murinus*) - Influence of Body-Weight Difference Between Sexes on Mating Success. *Experimental Animals*, **40**, 145–152.

IUCN (2011) The IUCN Red List of Threatened Species. Version 2011.2. Retrieved from www.iucn.org.

Jamal A, Moorthy S (Eds.) (2012) *Indian Ocean Studies: Cultural, Social, and Political Perspectives*. Routledge Indian Ocean Series, New York.

Jenkins P (2013) An Account of the Himalayan Mountain Soricid Community, with the Description of a New Species of Crocidura (Mammalia: Soricomorpha: Soricidae). *Raffles Bulletin of Zoology*, **Supp No. 2**, 161–175.

Jogahara T, Oda S-I, Kawai T, Hanamura H, Koyasu K (2008) Numerical Variation of Teeth in the Wild House Musk Shrew *Suncus murinus* Captured From Nagasaki, Japan. *Archives Of Oral Biology*, **53**, 617–21.

Joglekar PP (2006) A Fresh Appraisal of the Animal-Based Subsistence and Domestic Animals in The Ganga Valley. *Pragdhara,* **18** , 309-321.

Jones EP, Eager HM, Gabriel SI, Jóhannesdóttir F, Searle JB (2013) Genetic Tracking Of Mice And Other Bioproxies To Infer Human History. *Trends In Genetics,* **29**, 298-308.

Jones EP, van der Kooij J, Solheim R, Searle JB (2010a) Norwegian House Mice (*Mus musculus musculus ⁄ domesticus*): Distributions, Routes of Colonization and Patterns of Hybridization. *Molecular Ecology*, **23**, 5252–64.

Jones EP, Richards MB, Searle JB (2010b) The Expansion of the House Mouse into North-Western Europe. *Journal of Zoology,* **283**, 257-268.

Jones EP, Skirnisson K, McGovern TH *et al.* (2012) Fellow Travellers: A Concordance of Colonization Patterns Between Mice And Men in The North Atlantic Region. *BMC Evolutionary Biology*, **12**, 35.

Joshi JP (1990) *Excavations at Surkotada and Exploration in Kutch: Memoirs of the Archaeological Survey of India*. Archaeological Survey of India, Janpath, New Delhi.

Kang HJ, Kosoy MY, Shrestha SK *et al.* (2011) Short report: Genetic Diversity Of Thottapalayam Virus, A Hantavirus Harbored by the Asian House Shrew (*Suncus murinus*) in Nepal. *The American Journal of Tropical Medicine and Hygiene*, **85**, 540–5.

Karanth K, Delefosse T, Rakotosamimanana, B *et al.* (2005) Ancient DNA from Giant Extinct Lemurs Confirms Single Origin of Malagasy Primates. *Proceedings of the National Academy of Sciences*, **102**, 5090-5095.

Kelly J (2005) *The Great Mortality: An Intimate History Of The Black Death, The Most Devastating Plague Of All Time*. Harper Collins, New York.

Kingdom J (1974) *East African Mammals Vol II B*. Academic Press, London.

Kingdon J (2013) *Mammals of Africa Vol. III*. Bloomsbury, London, UK.

Kitchener D., Schmitt L., Maharadatunkamsi X (1994) Morphological And Genetic Variation in *Suncus murinus* (Soricidae: Crocidurinae) from Java, Lesser Sundra Islands, Maluku and Sulawesi, Indonesia. *Mammalia*, **48**, 433–451.

Knapp M, Hofreiter M (2010) Next Generation Sequencing of Ancient DNA: Requirements, Strategies and Perspectives. *Genes*, **1**, 227–243.

Kurachi M, Chau B, Dang V *et al.* (2007a) Population Structure of Wild Musk Shrews (*Suncus murinus)* in Asia Based on Mitochondrial DNA Variation, with Research in Cambodia and Bhutan. *Biochemical Genetics*, **45**, 165-183.

Kurachi M, Kawamoto Y, Tsubota Y, Ba-Loc C, Al E (2007b) Phylogeography of Wild Musk Shrew (Suncus Murinus) Populations in Asia Based on Blood Protein/Enzyme Variation. *Biochemical Genetics*, **45**, 543–563.

Lack JB, Greene DU, Conroy CJ *et al.* (2012) Invasion Facilitates Hybridization with Introgression in the *Rattus rattus* Species Complex. *Molecular Ecology*, **21**, 3545–61.

Larson G, Cucchi T, Fujita M *et al.* (2007) Phylogeny And Ancient DNA Of Sus Provides Insights into Neolithic Expansion in Island Southeast Asia. *Proceedings of the National Academy of Sciences,* **104**, 4834-4839.

LaViolette A (2008) Swahili Cosmopolitanism in Africa and the Indian Ocean World, AD 600–1500. *Archaeologies*, **4**, 24–49.

Lawson Handley L, Berset-Brandli L, Perrin N (2006) Disentangling Reasons for Low Y Chromosome Variation in the Greater White-Toothed Shrew (Crocidura russula). *Genetics,* **173**, 935-942.

Linz B, Moodley Y, Manica A *et al.* (2007) An African Origin for the Intimate Association Between Humans and *Helicobacter pylori*. *Nature*, **445**, 915–918.

Long J (2003) *Introduced Mammals of the World: Their History, Distribution and Influence*. CSIRO Publishing, Melbourne.

Lortet L, Gaillard C (1903) La Faune Mommifiee De L'ancienne Egypte. *Archives De Museum d'Histoire Naturelle De Lyon*, **8**, 38–40.

Lu L, Chesters D, Zhang W *et al.* (2012) Small Mammal Investigation in Spotted Fever Focus with DNA-Barcoding and Taxonomic Implications on Rodents Species from Hainan of China. *PLoS ONE*, **7**: e43479

Lund M (1994) Commensal Rodents. In: *Rodent Pests and Their Control* (eds Buckle AP, Smith RH), pp. 23–43. CAB International, Wallingford.

Mahapatra SK (1995) *Indian Archaeology 1990-91- A Review*. Archaeological Survey of India, Janpath, New Delhi.

Mallet J (1995) A Species Definition for the Modern Synthesis. *Trends in Ecology and Evolution,* **10**, 294-299.

Matheson CD, Gurney C, Esau N, Lehto R (2010) Assessing PCR Inhibition from Humic Substances. *The OPen Enzyme Inhibition Journal,* **3**, 38–45.

Matisoo-Smith E, Allen JS (2001) Name That Rat: Molecular And Morphological Identification Of Pacific Rodent Remains. *International Journal of Osteoarchaeology*, **11**, 34–42.

Matisoo-Smith E, Horsburgh KA, Robins J, Anderson A (2001) Genetic Variation in Archaeological *Rattus exulans* Remains from the Emily Bay Settlement Site, Norfolk Island. *Archaeology*, **27**, 81–84.

Matisoo-Smith E, Itwin GJ, Lambert DM (1999) Rodents of the Sunrise: Mitochondrial DNA Phylogenies of Polynesian *Rattus exulans* and the Settlement of Polynesia. *In: The Pacific from 5000 to 2000 Bp: colonization and transformations* (ed JC Galipaud), pp. 259-276. IRD, Paris.

Matisoo-Smith E, Robins J (2004) Origins And Dispersals Of Pacific Peoples: Evidence from MtDNA Phylogenies of the Pacific Rat. *Proceedings of the National Academy of sciences*, **101**, 9167–9172.

Maul LC, Smith KT, Barkai R *et al.* (2011) Microfaunal Remains At Middle Pleistocene Qesem Cave, Israel: Preliminary Results on Small Vertebrates, Environment and Biostratigraphy. *Journal of Human Evolution*, **60**, 464–480.

Mayr E (1970) *Populations, Species, and Evolution*. Harvard University Press, Cambridge, MA.

McCord B, Opel K, Funes M, Zoppis S, Jantz LM (2011) *An Investigation of the Effect of DNA Degradation and Inhibition on PCR Amplification of Single Source and Mixed Forensic Samples*. National Institute of Justice, Washington.

Mccormick M (2003) Rats , Communications, and Plague: Toward an Ecological History *Journal of Interdisciplinary History*, **34**, 1–25.

Meegaskumbura S, Meegaskumbura M, Pethiyagoda M, Manamendra-Arachchi K, Schneider C. (2007) *Crocidura hikmiya*, a New Shrew (Mammalia: Soricomorpha: Soricidae) from Sri Lanka. *Zootaxa*, **1665**, 19–30.

Meegaskumbura S, Meegaskumbura M, Schneider CJ (2010) Molecular Phylogenetics and Evolution Systematic Relationships and Taxonomy of *Suncus montanus* and *S . murinus* from Sri Lanka. *Molecular Phylogenetics and Evolution*, **55**, 473–487.

Meegaskumbura S, Schneider CJ (2008) A Taxonomic Evaluation Of The Shrew Suncus Montanus (Soricidae: Crocidurinae) Of Sri Lanka. *Ceylon Journal of Science (BioScience)*, **37**, 129–136.

Mehta RN (1968) *Excavation at Nagara; Technical Studies by D. R. Shah*. Department of Archaeology and Ancient History, Faculty of Arts, M.S. University of Baroda., Baroda.

Mehta RN, Chowdhary SN, Hedge KTM, Shah DR (1971) *Excavation At Jokha; Being The Report Of The Excavations Conducted At Jokha, Taluka Kamrej, District Surat, From 31st December 1966 To 23rd February 1967*. University Archaeology Series No. 11. Faculty of Arts. Department of Archaeology and Ancient History. The Maharaja Sayajirao University of Baroda, Baroda.

Mehta RN, Chowdhary SN, Hedge KTM, Shah DR (1975) *Excavation at Dhatwa.* M. S. University Archaeology Series No. 12. Faculty of Arts. Department of Archaeology and Ancient History. The Maharaja Sayajirao University of Baroda, Baroda.

Milishnikov AN, Lavrenchenko LA, Lebedev VS (2004) Origin of the House Mice (Superspecies Complex *Mus musculus* sensu lato) from the Transcaucasia Region: A New Look at Dispersal Routes and Evolution. *Genetika*, **40**, 1234-1250.

Misra OP (2003) *Archaeological Excavations in Central India: Madhya Pradesh and Chhattisgarh*. Mittal Publications, New Delhi.

Mitchell P (2005) *African Connections: Archaeological Perspectives on Africa and the Wider World*. Alta Mira Press, Walnut Creek.

Morelli G, Song Y, Mazzoni CJ *et al.* (2010) Yersinia Pestis Genome Sequencing Identifies Patterns of Global Phylogenetic Diversity. *Nature*, **42**, 1140–1143.

Motokawa M, Suzuki H, Harada M *et al.* Phylogenetic Relationships among East Asian Species of Crocidura (Mammalia, Insectivora) Inferred from Mitochondrial Cytochrome *b* Gene Sequences. *Zoological Science*, **17**, 497-504.

Msaidie S, Ducourneau A, Boetsch G *et al.* (2010) Genetic Diversity On The Comoros Islands Shows Early Seafaring as Major Determinant of Human Biocultural Evolution in the Western Indian Ocean. *European Journal of Human Genetics*, **19**, 89-94.

Murdoch G (1959) *Africa, Its Peoples and their Culture History.* McGraw-Hill, New York.

Musser G, Carleton M (2005) Superfamily Muroidea. In: *Mammal Species of the World: A Taxonomic and Geographic Reference 3rd ed.* John Hopkins University Press, Baltimore.

Nagarajarao MS, Malhotra KC (1965) *Stone Age Hill Dwellers of Tekkalakota: Preliminary Report of the Excavations at Tekkalakota. Deccan College Building Centenary and Silver Jubilee Series 31*. Deccan College Postgraduate and Research Institute, Poona.

Nahonyo CL, Mwasumbi LB, Msuya CA *et al.* (2005) *NGEZI – Vumawimbi Forest Reserves Biodiversity Inventory Report*. Dar Es Salaam.

Nakamoto A, 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.

Nakamura I, Hang'ombe B, Sawa H *et al.* (2013) Cross-Reactivity of Secondary Antibodies Against African Rodents and Application for Sero-Surveillance. *The Japanese Society of Veterinary Science*, **75**, 819–825.

National Institute of Health (2014) BLAST. *The National Center for Biotechnology Information*. Retrieved from www.ncbi.org.

Van Neer W, Ervynck A (1998) The Faunal Remians. In: *Berenike 1996. Report of the 1996 Excavations at Berenike (Egyptian Red Sea Coast) and the Survey of the Eastern Desert* (eds Sidebotham SE, Wendrich W.), pp. 349–388. Centre of Non-Western Studies, Leiden.

Van Neer W, Ervynck A (1999) The faunal remains. In: *Berenike 1997. Report of the 1997 excavations at Berenike and the Survey of the Eastern Desert, Including Excavtions at Shenshef* (eds Sidebotham S, Wendrich WZ), pp. 325–348. Centre of Non-Western Studies, Leiden.

Van Neer W, Gautier A (1993) Preliminary Report on the Faunal Remains from the Coastal Site of Ed-Dur, 1st-4th Century AD Umm Al-Quwain, United Arab Emirates. In: *Archaeozoology of the Near East : Proceedings of the First International Symposium on the Archaeozoology of South-Western Asia And Adjacent Areas* (eds Buitenhuis H, Uerpmann H), pp. 110–118. Backhuys, Leiden.

Van Neer W, Lentacker A (1996) The Faunal Remains. In: *Berenike '95. Preliminary Report of the Excavations at Berenike (Egyptian Red Sea Coast) and the Survey of the Eastern Desert.* (eds Sidebotham S, Wendrick W), pp. 337–355. CNWS Publications, Leiden.

Neolitzky F (1911) Nahrungs- und heilmittel der Uragypter. *Zeitschrift fur Untersuchung der Nahrungs- und Genussmittel der Uragypter*, **21**, 607–613.

Nunome M, Ishimori C, Aplin KP *et al.* (2010) Detection of Recombinant Haplotypes in Wild Mice (*Mus Musculus*) Provides New Insights Into the Origin of Japanese Mice. *Molecular Ecology*, **19**, 2474–2489.

Nylander J a a, Wilgenbusch JC, Warren DL, Swofford DL (2008) AWTY (Are We There Yet?): A System for Graphical Exploration of MCMC Convergence in Bayesian Phylogenetics. *Bioinformatics (Oxford, England)*, **24**, 581–3.

Ohdachi SD, Hasegawa M, Iwasa M a. *et al.* (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.

Ohdachi S, Masuda R, Abe H, Adachi J (1997) Phylogeny of Eurasian Soricine Shrews (Insectivora, Mammalia) Inferred from the Mitochondrial Cytochrome *b* Gene Sequences. *Zoological Science*, **14**, 527–532.

Olson DM, Dinerstein E (2002) The Global 200: Priority Ecoregions for Global Conservation. *Annals of the Missouri Botanical Garden*, **89**, 199–224.

Omar H, Adamson EAS, Goodman SM, Hashim R, Ruedi M (2011) Phylogenetic Relationships of Malayan And Malagasy Pygmy Shrews of the Genus Suncus (SORICOMORPHA: SORICIDAE) Inferred from Mitochondrial Cytochrome *b* Gene Sequences. *Raffles Bulletin of Zoology*, **59**, 237–243.

Onuma M, Cao Y, Hasegawa M, Kusakabe S (2000) A Close Relationship of Chiroptera With Eulipotyphla (Core Insectivora) Suggested by Four Mitochondrial Genes. *Zoological Science*, **17**, 1327–1332.

Orlando L, Calvignac S, Schnebelen C *et al.* (2008) DNA from Extinct Giant Lemurs Links Archaeolemurids to Extant Indriids. *BMC Evolutionary Biology*, **8**, 121.

Pääbo S, Poinar H, Serre D *et al.* (2004) Genetic Analyses from Ancient DNA. *Annual Review of Genetics*, **38**, 645–79.

Pagès M, Chaval Y, Herbreteau V *et al.* (2010) Revisiting the Taxonomy of the Rattini Tribe: A Phylogeny-Based Delimitation of Species Boundaries. *BMC Evolutionary Biology*, **10**.

Panagiotakopulu E (2004) Pharaonic Egypt and the Origins of Plague. *Journal of Biogeography*, **31**, 269–275.

Patnaik R (1997) New Murids and Gerbellids (Rodentia, Mammalia) from Pliocene Siwalik Sediments of India. *Palaeovertebrata*, **26**, 129–165.

Pauperio J, Herman JS, Melo-Ferreira J *et al.* (2012) Cryptic Speciation in the Field Vole: A Multilocus Approach Confirms Three Highly Divergent Lineages in Eurasia. *Molecular Ecology*, **21**, 6015–6032.

Pearson M (2007) *The Indian Ocean*. Routledge, New York.

Perera N, Kourampas N, Simpson IA *et al.* (2011) People of the Ancient Rainforest: Late Pleistocene Foragers at the Batadomblena Rockshelter, Sri Lanka. *Journal of Human Evolution*, **61**, 254–269.

Peters N, Van Neer W, Debruyne S, Peters S (2009) Late 17th Century AD Faunal Remains from the Dutch "Fort Frederik Hendrik" at Mauritius (Indian Ocean). *Archaeofauna*, **18**, 159–184.

Petrie WKF (1914) The Shipwrecked Sailor. In: *The World's Story: A History of the World in Story, Song, and Art Vol III Egypt, Africa and Arabia* (ed Tappan E. M). Houghton-Mifflin, Boston.

Phillips CS, Mosseri-Marlio CE (2002) Sustaining Change: The Emerging Picture of the Neolithic to Iron Age Subsistence Economy at Kalba, Sharjah Emirate, UAE. In: *Archaeozoology of the Near East V* (eds Buitenhuis H, Choyke AM, Mashkour M), pp. 195–210. ARC, Groningen.

Pocock M, Hauffe H, Searle J (2005) Dispersal in house mice. *Biological Journal of the Linnean Society*, **84**, 565–583.

Pocock M, Searle J, White P (2004) Adaptations of Animals to Commensal Habitats: Population Dynamics of House Mice *Mus musculus domesticus* on Farms. *Journal of Animal Ecology*, **73**, 878–888.

Posada D (2008) jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution*, **25**, 1253–6.

Possehl G, Raval MH (1989) *Harappan Civilization and Rojdi*. E. J. Brill, Leiden and New York.

Potts DT (1999) *The Archaeology of Elam: Formation and Transformation of an Ancient Iranian State*. Cambridge World Archaeology, Cambridge, UK.

Potts DT (2006) Before the Emirates: An Archaeological and Historical Account of Developments in the Region c. 5000 BC to 676 AD. In: *United Arab Emirates- A New Perspective* (eds Abed I Al, Hellyer P), pp. 28–69. Trident Press Ltd, London.

Potts DT (2008) Arabian Peninsula. In: *Encyclopedia of Archaeology* (ed Pearsall D), pp. 827–834. Elsevier, New York.

Prager EM, Orrego C, Sage RD (1998) Genetic Variation and Phylogeography of Central Asian and Other House Mice , Including a Major New Mitochondrial Lineage in Yemen. *Genetics,* **150**, 835-861.

Prager EM, Tichyt H, Sage RD (1996) Mitochondrial DNA sequence variation in the eastern house mouse, *Mus musculus*: comparison with other house mice and report of a 75-bp tandem repeat. *Genetics*, **143**, 427-446.

Prashad B (1939) *Animal Remains from Harappa. Memoirs of the Archaeological Survey of India no. 51*. Archaeological Survey of India, Janpath, New Delhi.

Prinsloo L, Wood N, Loubser M *et al.* (2005) Re-dating of Chinese Celadon Shards Excavated on Mapungubwe Hill, a 13th Century Iron Age Site in South Africa, Using Raman Spectroscopy, XRF and XRD. *Journal of Raman Spectroscopy,* **38**, 808-816.

Pye T, Swain R, Seppelt RD (1999) Distribution and Habitat Use of the Feral Black Rat (*Rattus rattus*) on Subantarctic Macquarie Island. *Journal of Zoology, London*, **247**, 429–438.

Rabinnovich R, Tchernov E (1996) Chronological, Paleoecological and Taphonomical Aspects of the Middle Paleolithic site of Qafzeh, Israel. In: *Archaeozoology of the Near East* (eds Buitenhuis H, Uerpmann H), pp. 5–44. Backhuys, Leiden.

Rad SAA, Jalal RA, Darvish JA, Matin MAM (2009) Identification of Three Iranian Species of the Genus *Rattus* (Rodentia, Muridae) using a PCR-RFLP Technique on Mitochondrial DNA. *Hystrix Italian Journal of Mammalogy*, **20**, 69–77.

Radimilahy C (1998) *Mahilaka: An Archaeological Investigation of an Early Town in Northwestern Madagascar*. Department of Archaeology and Ancient History, Uppsala.

Rajabi-Maham H, Orth A, Bonhomme F (2008) Phylogeography and Postglacial Expansion of *Mus musculus domesticus* Inferred from Mitochondrial DNA Coalescent, From Iran To Europe. *Molecular Ecology*, **17**, 627–41.

Rajabi-Maham H, Orth A, Siahsarvie R *et al.* (2012) The South-Eastern House Mouse *Mus musculus castaneus* (Rodentia: Muridae) is a Polytypic Subspecies. *Biological Journal of the Linnaean Society*, **107**, 295–306.

Rambaut A (2014a) Fig Tree. Retrieved from *http://tree.bio.ed.ac.uk*.

Rambaut A (2014b) Tracer. Retrieved from *http://tree.bio.ed.ac.uk*.

Rami Reddy V (1985) South Indian Neolithic Culture Seen in Retrospect. *East and West*, **35**, 43–65.

Ramírez-Soriano A, Ramos-Onsins SE, Rozas J, Calafell F, Navarro A (2008) Statistical Power Analysis of Neutrality Tests Under Demographic Expansions, Contractions and Bottlenecks with Recombination. *Genetics*, **179**, 555–67.

Rao SR (1979) *Lothal-A Harappan Town (1955-62). Memoirs of the Archaeological Survey of India No 78*. Archaeological Survey of India, Janpath, New Delhi.

Raspopova AA, Shchipanov NA (2011) Variability of a Cytochrome *b* Region In Different Chromosome Races And Populations Of The Common Shrew *Sorex araneus* L., 1758. *Russian Journal of Genetics*, **47**, 527–536.

Robins JH, Hingston M, Matisoo-Smith E, Ross HA (2007) Identifying *Rattus* Species Using Mitochondrial DNA. *Molecular Ecology Notes*, **7**, 717–729.

Robins JH, Mclenachan PA, Phillips MJ *et al.* (2008) Dating of Divergences Within the *Rattus* Genus Phylogeny Using Whole Mitochondrial Genomes. *Molecular Phylogenetics and Evolution*, **49**, 460–466.

Rogatcheva MB, Ono T, Sonta S, Oda S, Borodin PM (2000) Robertsonian Metacentrics of the House Musk Shrew (*Suncus murinus*, Insectivora, Soricidae) Lose the Telomeric Sequences in the Centromeric Area. *Genes and Genetic Systems*, **75**, 155–8.

Rogers R, Harpending H (1992) Population Growth Makes Waves in the Distribution of Pairwise Genetic Differences. *Molecular Biology and Evolution*, **9**, 552–69.

Ronaghi M, Uhlen M, Nyren P (1998) A Sequencing Method Based On Real-Time Pyrophosphate. *Science*, **281**, 363– 365.

Ronquist F, Teslenko M, van der Mark P *et al.* (2012) MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, **61**, 539–542.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA Polymorphism Analyses by the Coalescent and Other Methods. *Bioinformatics*, **19**, 2496–2497.

Ruedi M, Courvoisier C, Vogel P (1996) Genetic Differentiation and Zoogeography of Asian *Suncus murinus* (Mammalia: Soricidae). *Biological Journal of the Linnaean Society*, **57**, 307–316.

Ruffino L, Vidal E (2010) Early Colonization Of Mediterranean Islands by *Rattus rattus*: A Review Of Zooarcheological Data. *Biological Invasions,* **12**, 2389–2394.

Russell JC., Faulquier L., Tonione MA (2011a) Rat Invasion of Tetiaroa Atoll, French Polynesia. In: *Island Invasives: Eradication and Management.* (eds Veitch CR., Clout MN., Towns DR), pp. 118–123. IUCN (International Union for Conservation of Nature), Gland, Switzerland.

Russell JC, Gleeson DM, Le Corre M (2011b) The Origin of *Rattus rattus* on the îles Eparses, Western Indian Ocean. *Journal of Biogeography*, **38**, 1834–1836.

Sankalia HD, Deo SB (1955) *Report on the Excavations at Nasik and Jorwe 1950-51*. Deccan College Postgraduate and Research Institute, Poona.

Sathe V, Badam GL (1996) Animal Remains from the Neolithic and Chalcolithic Periods at Senuwar, District Rohtas, Bihar. *Man and Environment*, **21**, 43–48.

Schmandt-Besserat D (1977) Review: Excavations of the Godin Project: Second Progress Report by T . Cuyler Young , Louis D . Levine. *Journal of the American Oriental Society*, **97**, 61–64.

Schwarz E, Schwarz HK (1943) The Wild and Commensal Stocks of the House Mouse , *Mus musculus* Linnaeus. *Journal of Mammology*, **24**, 59–72.

Searle J, Jones C, Gündüz I *et al.* (2009) Of Mice and (Viking?) Men: Phylogeography of British and Irish House Mice. *Proceeding of Biological Sciences,* **276,** 201-207.

Shabani M, Darvish J, Mashkour M, Ghasemzadeh F, Mirshamsi O (2010) Contemporary and Sub-Fossil House Mice (*Mus musculus* Linnaeus, 1758)(Rodentia:Muridae) from Iran. *Iranian Journal of Animal Biosystematics*, **6**, 45–54.

Shah DR (1973) Animal Remains from the Kodekal Excavation. In: *Investigations into the Neolithic Culture of the Shorapur Doab, South India* (ed Paddaya K). Brill, Leiden.

Sharma AK (1990) Animal Bone Remains. In: *Excavation at Surkotada 1971-72 and Exploration in Kutch* (ed Joshi JP). Archaeological Survey of India, New Delhi.

Shoshani J, Ghebregiorgis MW, Peter R. Schmidt (2008) Interpreations of Faunal Remains from Archaelogical Sites on the Asmara Plateau of Eritrea. In: *The Archaeology of Ancient Eritrea* (eds Schmidt PR, Curtis MC, Teka Z). Red Sea Press, Trenton, NJ.

Singleton GR, Brown PR, Pech RP *et al.* (2005) The Genus *Mus* as a Model for Evolutionary Studies: One Hundred Years of Eruptions of House Mice in Australia – A Natural Biological Curio. *Biological Journal of the Linnean Society*, **84**, 617–627.

Sites JW, Morando M (2009) Phylogeography. In: *Encyclopedia of Life Sciences* . John Wiley & Sons.

Slabova M, Frynta D (2007) Morphometric Variation in Nearly Unstudied Populations of the Most Studied Mammal: The Non-Commensal House Mouse (*Mus musculus domesticus*) in the Near East and Northern Africa. *Zoologischer Anzeiger*, **246**, 91–101.

Smith CI, Chamberlain AT, Riley MS *et al.* (2003) The Thermal History of Human Fossils and the Likelihood of Successful DNA Amplification. .*Journal of Human Evolution*, **45**, 203-217.

Soarimalala V, Goodman S (2011) *Les Petits Mammiferes de Madagascar*. Association Vahatra, Antananarivo, Madagascar.

Sonawane VH (2000) Early farming Communities of Gujarat, India. *The Melaka Papers Vol III. Bulletin of the Indo-Pacific Prehistory Association Vol 19*, **3**, 137–146.

Song Y, Lan Z, Kohn M (2014) Mitochondrial DNA Phylogeography of the Norway Rat. *PLoS ONE,* **9**: e88425.

Spear T (2000) Early Swahili Reconsidered. *International Journal of African Historical Studies*, **33**, 257–290.

Storey AA, Clarke AC, Ladefoged T, Judith R, Matisoo- Smith E (2013) DNA And Pacific Commensal Models : Applications, Construction, Limitations, and Future Prospects. *The Journal of Island and Coastal Archaeology*, **8**, 37–65.

Storey AA, Ramirez J, Quiroz D *et al.* (2007) Radiocarbon and DNA Evidence for a Pre-Columbian Introduction of Polynesian Chickens to Chile. *Proceedings of the National Academy of Sciences,* **104**, 10335-10339.

Suzuki H, Nunome M, Kinoshita G *et al.* (2013) Evolutionary and Dispersal History of Eurasian House Mice *Mus musculus* Clarified by More Extensive Geographic Sampling of Mitochondrial DNA. *Heredity*, **111**, 375–90.

Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Tchernov E (1984) Commensal Animals and Human Sedentism in the Middle East. In: *Animals and Archaeology: 3. Early Herders and Their Flocks* (eds Clutton-Brock J, Grigson C), pp. 91–115. BAR International Series.

Tchernov E (1991) Of Mice and Men: Biological Markers for Long-Term Sedentism : A Reply. *Paléorient*, **17**, 153–160.

Thomas PK (1984) The Faunal Background of the Chalcolithic Culture of Western India. In: *Animals and Archaeology: Early Herders and Their Flocks* (eds Clutton-Brock J, Grigson C), pp. 355–361. BAR International Series, London.

Thomas PK (1988) Faunal Assemblage. In: *Excavations at Inamgaon, Vol 1, Part 2* (eds Dhavalikar MK, Sankalia H, Ansari ZD). Deccan College Postgraduate and Research Institute, Pune.

Thomas PK (1996) Animal Remains from Tuljaphur Garhi. In: *Excavations at Tuljaphur Garhi, 1984-85 (Vidarbha, Maharashtra)* (ed Bopardikar BP). Archaeological Survey of India, New Delhi.

Thomas PK (2008) Animal subsistence in the Chalcolithic Culture of Western India (with special reference to Balathal). *Bulletin of the Indo-Pacific Prehistory Association. [Online] 19:0*.

Thomas PK, Deshpande-Mukherjee A, Shobha V (2006) Faunal Assemblages from Kanenur, Kaveripura and Kannikere: Iron Age/Early Historic Sites in Karnataka. *Man and Environment*, **31**, 75–80.

Thomas PK, Joglekar P (1990) Faunal Remains. In: *Excavations at Kaothe* (eds Dhavalikar MK, Shinde V, Atre S). Deccan College Postgraduate and Research Institute, Pune.

Thomas PK, Joglekar PP, Deshpande-Mukherjee A, Pawankar SJ (1995a) Harappan Subsistence Patterns with Special Refernce to Shikarpur, a Harappan Site in Gujarat. *Man and Environment*, **20**, 33–41.

Thomas PK, Joglekar P, Mishra VD, Pandey JN, Pal JN (1995b) A Preliminary Report on the Faunal Remains from Damdam. *Man and Environment*, **20**, 29–36.

Tollenaere C, Brouat C, Duplantier J *et al.* (2010) Phylogeography of the Introduced Species *Rattus rattus* in the Western Indian Ocean, with Special Emphasis on the Colonization History of Madagascar. *Journal of Biogeography*, **37**, 398–410.

Toškan B, Kryštufek B (2006) Noteworthy Rodent Records from the Upper Pleistocene and Holocene of Slovenia. *Mammalia*, **70**, 98–105.

Triant DA, Dewoody JA (2007) The Occurrence, Detection, and Avoidance of Mitochondrial DNA Translocations in Mammalian Systematics and Phylogeography. *Journal of Mammalogy*, **88**, 908–920.

Uerpmann H (2001) The Fauna. In: *Selenkahiye: Final Report on the University of Chicago and University of Amsterdam Excavations in the Tabqa Reservoir, Northern Syria 1967-1975* (ed Van Loon MN). Nederlands Historisch- Archaeologisch Instituut, Istanbul.

Uerpmann M, Uerpmann H-P (1999) The Animal Economy of Ancient Dilmun in the Light of Faunal Remains from Excavations at Saar And Qala'at Al-Bahrain. *ISIMU Revista Sobre Oriente Proximo Y Egipto en la Antiguedad*, **2**, 635–646.

Uerpmann M, Uerpmann H (2005) Animal Bone Finds And Their Relevance To The Ecology And Economy Of Saar. In: *The Early Dilmun Settlement at Saar* (eds Killick R, Moon J). Archaeology International, Ludlow.

Unknown (1917) Millions slaughtered, millions coming. *Ararat Adviser*.

Valenzuela-Lamas S, Baylac M, Cucchi T, Vigne J-D (2011) House mouse dispersal in Iron Age Spain: A Geometric Morphometrics Appraisal. *Biological Journal of the Linnean Society*, **102**, 483–497.

Varnham KJ, Roy SS, Seymour A *et al.* (2004) Eradicating Indian Musk Shrews (*Suncus murinus*, Soricidae) from Mauritian Offshore Islands. In: *Turning the Tide: The Eradication of Invasive Species* (eds Veith C, Clout M), pp. 342–349. IUCN, Gland, Switzerland and Cambridge, UK.

Velimirovic B (1972) Plague in South East Asia. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **66**, 479–504.

Vences M, Wanke S, Vieites D *et al.* (2004) Natural Colonization or Introduction? Phylogeographical Relationships and Morphological Differentiation of House Geckos (*Hemidactylus*) from Madagascar. *Biological Journal of the Linnean Society*, **83**, 115-130.

Villesen P (2007) FaBox: an online fasta sequence toolbox. Retrieved from http://users-birc.au.dk/biopv/php/fabox/

Voigt EA (1984) Iron Age Herders of the Northern Transvaal South Africa, in the First Millennium AD. In: *Animals and Archaeology* (ed Juliet Clutton-Brock)

Wade G (2009) An Early Age of Commerce in Southeast Asia, 900–1300 CE. *Journal of Southeast Asian Studies*, **40**, 221.

Ward C (2012) Building Pharaoh's Ships: Cedar, Incense and Sailing the Great Green. *British Museum Studies in Ancient Egypt and Sudan*, **18**, 217–232.

Weissbrod L, Dayan T, Kaufman D, Weinsteinevron M (2005) Micromammal Taphonomy of El-Wad Terrace, Mount Carmel, Israel: Distinguishing Cultural from Natural Depositional Agents in The Late Natufian. *Journal of Archaeological Science*, **32**, 1–17.

Weissbrod L, Malkinson D, Cucchi T *et al.* (2014) Ancient Urban Ecology Reconstructed from Archaeozoological Remains of Small Mammals in The Near East. *PLoS ONE,* **9**: e91795.

Willerslev E, Cooper A (2005) Ancient DNA. *Proceedings of the Royal Society B*, **272**, 3–16.

Wilson DE, Reeder DAM (Eds.) (2005) *Mammal Species of the World. A Taxonomic and Geographic Reference (3rd ed)*. Johns Hopkins University Press, Baltimore, USA.

Wright HT (1984) Early Seafarers of the Comoro Islands: The Dembeni Phase of the Ninth–Tenth Centuries AD. *Azania Archaeological Research in Africa*, **19**, 13–59.

Wyncoll G, Tangri D (1989) Of Mice and Men: Is the Presence of Commensal Animals in Archaeological Sites a Positive Correlate of Sedentism? *Paléorient*, **15**, 85–94.

Wynne-Jones S (2007) Creating Urban Communities at Kilwa Kisiwani, Tanzania, AD 800-1300. *Antiquity*, **81**, 368–380.

Wynne-Jones S(2009) Excavations at Vumba Kuu, Southern Kenya Coast: Report On Fieldwork 2007- 2009. University of Bristol.

Wynne-jones S, Fleisher J (2010) Archaeological Investigations at Songo Mnara, Tanzania, 2009. *Nyame Akuma*, **73**, 2–9.

Yamagata T, Namikawa T (1999) Sequence Variation and Evolution of the Mitochondrial DNA Control Region in the Musk Shrew, *Suncus murinus*. *Genes and Genetic Systems*, **74**, 257–266.

Yamagata T, Ohishi K, Faruque M (1995) Genetic Variation and Geographic Distribution on the Mitochondrial DNA in Local Populations of the Musk Shrew, *Suncus murinus*. *Japan Journal of Genetics*, **70**, 321–337.

Yamagata T, Tamaka Y, Ishikawa A, Namikawa T, Tomita T (1990) Genetic Relationship Among the Musk Shrews, *Suncus murinus* Insectivora, Inhabiting Islands and the Continent Based on Mitochondrial DNA Types. *Biochemical Genetics*, **28**, 185–195.

Yang D, Woiderski J, Driver J (2005) DNA Analysis of Archaeological Rabbit Remains from the American Southwest. *Journal of Archaeological Science*, **32**, 567–578.

Yonekawa H, Moriwaki K, Gotoh O (1981) Evolutionary Relationships Among Five Subspecies of *Mus musculus* Based on Restriction Enzyme Cleavage Patterns of Mitochondrial DNA. *Genetics*, **98**, 801–816.

Yosida TH, Tsuchiya K, Moriwaki K (1971) Karyotypic Differences of Black Rats, *Rattus rattus*, Collected in Various Localities of East and Southeast Asia and Oceania. *Chromosoma,* **33**, 252-267.

Yosida TH, Kato H, Tsuchiya K (1974) Cytogenetical Survey of Black Rats, *Rattus rattus* , in Southwest and Central Asia, with Special Regard to the Evolutional Relationship between Three Geographical Types. *Chromosoma*, **109**, 99–109.

Yosida TH, Sagai T (1975) Variation of C-bands in the Chromosomes of Several Subspecies of *Rattus rattus*. *Chromosoma*, **50**, 283–300.

Yosida TH (1982) Cytogenetic studies on Insectivora. 2. Geographical variation of chromosomes in the house shrew, Suncus murinus (Soricidae), in east, southeast, and southwest Asia, with a note on the karyotype evolution and distribution. *Japan Journal of Genetics*, **57**, 101–112.

Young R (2007) The Archaeozoological Remains. In: *Excavations at Charsadda, NWFP, Pakistan* (ed Conningham RAE), pp. 247–257. Society for South Asian Studies, London.

Zakharov E V, Lobo NF, Nowak C, Hellmann JJ (2009) Introgression as a Likely Cause of mtDNA Paraphyly in Two Allopatric Skippers (Lepidoptera: Hesperiidae). *Heredity*, **102**, 590–9.

Zeder M a. (2000) The Initial Domestication of Goats (*Capra hircus*) in the Zagros Mountains 10,000 Years Ago. *Science*, **287**, 2254–2257.

APPENDICES

Appendix 3.I *Suncus murinus* collection information. * denotes used in this study. AMNH=American Museum of Natural History; FMNH=Field Museum of Natural History; NHM=British Natural History Museum; NMNH=Smithsonian Institution. DATE=collection year. ACC=accuracy of geographic coordinates (1=given by sample provider; 2=known precise location (e.g. city); 3=known district; 4=known region or country; 5=no information). REASON=if sample not used, reason why (None=no amplification; Miss=missing sequence data; Cont= possible contamination of sequence; Deam= possible deamination of sequence; No Loc).

SAMPLE ID	PROVIDER	TYPE		DATE	TAXON ID	COUNTRY
57.398	NHM	Osteocrust			Suncus murinus	Maldives
59.539	NHM	Osteocrust	*	1958	Suncus murinus	Maldives
59.540	NHM	Osteocrust	*	1958	Suncus murinus	Maldives
27642	AMNH	Osteocrust		1901	Suncus murinus	India
27643	AMNH	Skin	*	1903	Suncus murinus	India
28632	AMNH	Skin	*	1904	Suncus murinus	Kashmir
29760	FMNH	Osteocrust	*	1925	Suncus murinus caerulescens	India
29761	FMNH	Osteocrust	*	1925	Suncus murinus caerulescens	India
29767	FMNH	Osteocrust	*	1926	Suncus murinus caerulescens	India
30966	FMNH	Osteocrust		1929	Suncus murinus murinus	Vietnam
30969	FMNH	Osteocrust	*	1929	Suncus murinus murinus	Vietnam
35402	FMNH	Osteocrust	*	1931	Suncus murinus soccatus	India
47969	AMNH	Osteocrust	*	1916	Suncus murinus murinus	China
47970	AMNH	Osteocrust	*	1916	Suncus murinus murinus	China
54539	AMNH	Osteocrust	*	1919	Suncus murinus	India
57605	FMNH	Osteocrust		1967	Suncus murinus soccatus	Nepal
62300	AMNH	Osteocrust			Crocidura sp.	?
65396	FMNH	Osteocrust		1947	Suncus murinus caerulescens	India
65397	FMNH	Osteocrust	*	1947	Suncus murinus caerulescens	India
65398	FMNH	Osteocrust	*	1947	Suncus murinus caerulescens	India
65399	FMNH	Osteocrust	*	1947	Suncus murinus caerulescens	India
76159	FMNH	Osteocrust	*	1949	Suncus murinus griffithi	India
76161	FMNH	Osteocrust	*	1949	Suncus murinus griffithi	India
76166	FMNH	Osteocrust	*	1949	Suncus murinus griffithi	India
76169	FMNH	Osteocrust		1952	Suncus murinus soccatus	India
76171	FMNH	Osteocrust	*	1952	Suncus murinus soccatus	India

76173	FMNH	Osteocrust	*	1951	Suncus murinus soccatus	India
76174	FMNH	Osteocrust	*	1951	Suncus murinus soccatus	India
76176	FMNH	Osteocrust	*	1951	Suncus murinus soccatus	India
76177	FMNH	Osteocrust	*	1951	Suncus murinus soccatus	India
76178	FMNH	Osteocrust	*	1951	Suncus murinus soccatus	India
82510	FMNH	Osteocrust	*	1912	Suncus murinus caerulescens	India
82511	FMNH	Osteocrust	*	1912	Suncus murinus caerulescens	India
82516	FMNH	Osteocrust	*	1922	Suncus murinus caerulescens	India
82521	FMNH	Osteocrust	*	1913	Suncus murinus tytleri	India
82522	FMNH	Osteocrust	*	1922	Suncus murinus tytleri	India
82525	FMNH	Osteocrust	*	1912	Suncus murinus sindensis	India
82533	FMNH	Osteocrust		1913	Suncus murinus malabaricus	India
82534	FMNH	Osteocrust	*	1913	Suncus murinus malabaricus	India
82536	FMNH	Osteocrust		1917	Suncus murinus viridescens	India
82538	FMNH	Osteocrust		1927	Suncus murinus murinus	Burma
82539	FMNH	Osteocrust	*	1927	Suncus murinus murinus	Burma
82544	FMNH	Osteocrust		1920	Suncus murinus soccatus	India
85857	FMNH	Osteocrust		1955	Suncus murinus murinus	Taiwan
92215	FMNH	Osteocrust		1960	Suncus murinus zeylanicus	Sri Lanka
92216	FMNH	Osteocrust	*	1960	Suncus murinus zeylanicus	Sri Lanka
94132	FMNH	Osteocrust	*	1960	Suncus murinus soccatus	Nepal
94133	FMNH	Osteocrust	*	1961	Suncus murinus soccatus	Nepal
98449	FMNH	Osteocrust	*	1963	Suncus murinus murinus	Malaysia
98450	FMNH	Osteocrust	*	1963	Suncus murinus murinus	Malaysia
99690	AMNH	Osteocrust	*	1928	Suncus murinus	Sri Lanka
99691	AMNH	Osteocrust	*	1928	Suncus murinus	Sri Lanka
99778	FMNH	Osteocrust	*	1968	Suncus murinus	Madagascar
100713	AMNH	Osteocrust	*	1927	Suncus murinus	Madagascar
100715	AMNH	Osteocrust	*	1926	Suncus murinus	Madagascar
100977	AMNH	Osteocrust	*	1927	Suncus murinus murinus	Indonesia

100979	AMNH	Osteocrust	*	1927	Suncus murinus murinus	Indonesia
100980	AMNH	Osteocrust		1927	Suncus murinus murinus	Indonesia
101908	AMNH	Osteocrust	*	1929	Suncus murinus murinus	Indonesia
101910	AMNH	Osteocrust	*	1929	Suncus murinus murinus	Indonesia
101918	AMNH	Osteocrust	*	1929	Suncus murinus murinus	Indonesia
101919	AMNH	Osteocrust	*	1929	Suncus murinus murinus	Indonesia
101920	AMNH	Osteocrust	*	1929	Suncus murinus murinus	Indonesia
106739	AMNH	Osteocrust		1933	Suncus murinus murinus	Indonesia
106740	AMNH	Osteocrust		1933	Suncus murinus murinus	Indonesia
106743	AMNH	Osteocrust		1933	Suncus murinus murinus	Indonesia
106744	AMNH	Osteocrust		1933	Suncus murinus murinus	Indonesia
106893	AMNH	Osteocrust		1933	Suncus murinus murinus	Indonesia
106894	AMNH	Osteocrust		1933	Suncus murinus murinus	Indonesia
113033	AMNH	Osteocrust	*	1931	Suncus murinus	Burma
143866	AMNH	Osteocrust	*		Suncus sp.	Philippines
146753	FMNH	Osteocrust	*	1990	Suncus murinus	Madagascar
150075	AMNH	Osteocrust		1944	Suncus murinus	Sri Lanka
150076	AMNH	Osteocrust		1944	Suncus murinus	Sri Lanka
150087	AMNH	Osteocrust	*	1943	Suncus murinus	India
154598	FMNH	Tissue in DMSO	*	1995	Suncus murinus	Madagascar
155232	NMNH	Osteocrust	*	2005	Suncus murinus	Indonesia
156338	FMNH	Tissue in DMSO	*	1996	Suncus murinus	Madagascar
163037	AMNH	Osteocrust	*	1943	Suncus murinus	India
163038	AMNH	Osteocrust	*	1943	Suncus murinus	India
163040	AMNH	Osteocrust		1943	Suncus murinus	India
163042	AMNH	Osteocrust		1943	Suncus murinus	India
163043	AMNH	Osteocrust		1943	Suncus murinus	India
165448	FMNH	Tissue in DMSO	*	1997	Suncus murinus	Madagascar

165498	FMNH	Tissue in DMSO	*	1997	Suncus murinus	Madagascar
165499	FMNH	Tissue in DMSO	*	1997	Suncus murinus	Madagascar
165501	FMNH	Tissue in DMSO	*	1997	Suncus murinus	Madagascar
165570	FMNH	Tissue in DMSO	*	1997	Suncus murinus	Madagascar
165571	FMNH	Tissue in DMSO	*	1997	Suncus murinus	Madagascar
165776	AMNH	Osteocrust	*	1950	Suncus murinus	Taiwan
165779	AMNH	Osteocrust	*	1950	Suncus murinus	Taiwan
165781	AMNH	Osteocrust	*	1950	Suncus murinus	Taiwan
165790	AMNH	Osteocrust	*	1950	Suncus murinus	Taiwan
165791	AMNH	Osteocrust	*	1950	Suncus murinus	Taiwan
166242	FMNH	Tissue in DMSO		1999	Suncus murinus	Madagascar
166243	FMNH	Tissue in DMSO	*	1999	Suncus murinus	Madagascar
166244	FMNH	Tissue in DMSO	*	1999	Suncus murinus	Madagascar
171135	AMNH	Osteocrust		1943	Suncus murinus	India
171136	AMNH	Skin	*	1944	Suncus murinus	India
171137	AMNH	Skin	*	1943	Suncus murinus	India
171138	AMNH	Osteocrust		1943	Suncus murinus	
171141	AMNH	Skin		1943	Suncus murinus	India
176115	FMNH	Tissue in DMSO	*	2002	Suncus murinus	Madagascar
178588	FMNH	Tissue in DMSO	*	2003	Suncus murinus	Madagascar
183968	FMNH	Tissue in	*	2005	Suncus murinus	Madagascar

		DMSO				
183969	FMNH	Tissue in DMSO	*	2005	Suncus murinus	Madagascar
184587	AMNH	Osteocrust	*	1941	Suncus luzoniensis	Philippines
184666	AMNH	Skin	*	1955	Suncus murinus tytleri	Pakistan
194437	FMNH	Tissue in DMSO	*	2002	Suncus murinus	Comoros Islands
194579	FMNH	Tissue in DMSO	*	2006	Suncus murinus	Madagascar
194596	FMNH	Tissue in DMSO	*	2006	Suncus murinus	Madagascar
194597	FMNH	Tissue in DMSO	*	2006	Suncus murinus	Madagascar
207468	AMNH	Osteocrust	*	1958	Suncus occultidens	Philippines
215539	AMNH	Osteocrust	*	1960	Suncus murinus	India
215541	AMNH	Osteocrust		1960	Suncus stoliczkanus stoliczkanus	India
217270	AMNH	Skin	*	1960	Suncus murinus tytleri	Pakistan
226976	AMNH	Osteocrust		1969	Suncus murinus	Indonesia
226984	AMNH	Osteocrust		1969	Suncus murinus	Indonesia
240757	AMNH	Skin	*	1965	Suncus murinus	Vietnam
240805	AMNH	Osteocrust		1966	Suncus murinus kandianus	Sri Lanka
240810	AMNH	Osteocrust	*	1966	Suncus murinus kandianus	Sri Lanka
240811	AMNH	Osteocrust		1966	Suncus murinus kandianus	Sri Lanka
240813	AMNH	Osteocrust	*	1966	Suncus murinus kandianus	Sri Lanka
240814	AMNH	Osteocrust	*	1966	Suncus murinus kandianus	Sri Lanka
240819	AMNH	Osteocrust		1966	Suncus murinus kandianus	Sri Lanka
240824	AMNH	Osteocrust	*	1966	Suncus murinus kandianus	Sri Lanka
240825	AMNH	Osteocrust	*	1966	Suncus murinus kandianus	Sri Lanka
240927	AMNH	Osteocrust	*	1966	Suncus murinus giganteus	Sri Lanka
244951	AMNH	Osteocrust		1978	Suncus murinus murinus	China
244952	AMNH	Osteocrust	*	1978	Suncus murinus murinus	China

257640	NMNH	Osteocrust	*	1928	Suncus murinus	Indonesia
326323	NMNH	Osteocrust		1958	Suncus murinus	Pakistan
326328	NMNH	Osteocrust	*	1958	Suncus murinus	Pakistan
359624	NMNH	Osteocrust	*	1964	Suncus murinus	Indonesia
1-IN-010606-SM-1	Pascal	Tissue in EtOH	*	2006	Suncus murinus	Réunion
1-IN-010606-SM-2	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
1-IN-010606-SM-3	Pascal	Tissue in EtOH	*	2006	Suncus murinus	Réunion
1-IN-010606-SM-4	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
1-IN-190107-SM-5	Pascal	Tissue in EtOH		2007	Suncus murinus	Réunion
1-IN-190107-SM-6	Pascal	Tissue in EtOH		2007	Suncus murinus	Réunion
1-IN-190107-SM-7	Pascal	Tissue in EtOH		2007	Suncus murinus	Réunion
1-IN-230107-SM-21	Pascal	Tissue in EtOH		2007	Suncus murinus	Réunion
1-IN-230107-SM-7	Pascal	Tissue in EtOH	*	2007	Suncus murinus	Réunion
1-IN-230107-SM-8	Pascal	Tissue in EtOH	*	2007	Suncus murinus	Réunion
1-IN-230107-SM-9	Pascal	Tissue in EtOH	*	2007	Suncus murinus	Réunion
1-IN-230506-SM-6	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
1-IN-230506-SM-7	Pascal	Tissue in EtOH	*	2006	Suncus murinus	Réunion
1-IN-250506-SM-2	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
1-IN-260506-SM-1	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
1-IN-260506-SM-2	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
1-IN-300506-SM-2	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
1-IN-300506-SM-4	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
1-IN-300506-SM-5	Pascal	Tissue in EtOH	*	2006	Suncus murinus	Réunion
1-IN-310506-SM-1	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
18.8.5.1	NHM	Osteocrust			Suncus murinus	Iraq
19.6.9.12	NHM	Osteocrust	*		Suncus murinus	Tanzania
2-IN-130606-SM-1	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
2-IN-130606-SM-3	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
2-IN-130606-SM-4	Pascal	Tissue in EtOH	*	2006	Suncus murinus	Réunion
2-IN-130606-SM-5	Pascal	Tissue in EtOH	*	2006	Suncus murinus	Réunion

2-IN-130606-SM-6	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
2-IN-140606-SM-1	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
2-IN-140606-SM-2	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
2-IN-140606-SM-3	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
2-IN-140606-SM-4	Pascal	Tissue in EtOH	*	2006	Suncus murinus	Réunion
2-IN-150606-SM-1	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
2-IN-150606-SM-10	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
2-IN-150606-SM-2	Pascal	Tissue in EtOH	*	2006	Suncus murinus	Réunion
2-IN-150606-SM-3	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
2-IN-150606-SM-4	Pascal	Tissue in EtOH	*	2006	Suncus murinus	Réunion
2-IN-150606-SM-5	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
2-IN-150606-SM-7	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
2-IN-150606-SM-8	Pascal	Tissue in EtOH	*	2006	Suncus murinus	Réunion
2-IN-150606-SM-9	Pascal	Tissue in EtOH	*	2006	Suncus murinus	Réunion
2-IN-160606-SM-2	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
2-IN-160606-SM-4	Pascal	Tissue in EtOH	*	2006	Suncus murinus	Réunion
24.3.8.10	NHM	Osteocrust			Suncus murinus	Tanzania
3-IN-270606-SM-10	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
3-IN-270606-SM-11	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
3-IN-270606-SM-12	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
3-IN-270606-SM-13	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
3-IN-270606-SM-15	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
3-IN-270606-SM-16	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
3-IN-270606-SM-17	Pascal	Tissue in EtOH	*	2006	Suncus murinus	Réunion
3-IN-270606-SM-18	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
3-IN-270606-SM-19	Pascal	Tissue in EtOH	*	2006	Suncus murinus	Réunion
3-IN-270606-SM-20	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
3-IN-270606-SM-5	Pascal	Tissue in EtOH	*	2006	Suncus murinus	Réunion
3-IN-270606-SM-7	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
3-IN-280606-SM-1	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion

3-IN-280606-SM-10	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
3-IN-280606-SM-3	Pascal	Tissue in EtOH	*	2006	Suncus murinus	Réunion
3-IN-280606-SM-4	Pascal	Tissue in EtOH	*	2006	Suncus murinus	Réunion
3-IN-280606-SM-5	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
3-IN-280606-SM-6	Pascal	Tissue in EtOH		2006	Suncus murinus	Réunion
3-IN-280606-SM-7	Pascal	Tissue in EtOH	*	2006	Suncus murinus	Réunion
3-IN-280606-SM-9	Pascal	Tissue in EtOH	*	2006	Suncus murinus	Réunion
30.2.11.105	NHM	Osteocrust	*		Suncus murinus	India
30.2.11.53	NHM	Osteocrust	*		Suncus murinus	India
30.2.11.86	NHM	Osteocrust	*		Suncus murinus	India
40.1.65	NHM	Osteocrust	*		Suncus murinus	Saudi Arabia
87.11.15.1	NHM	Osteocrust	*		Suncus murinus	Kenya
94.1.21.2	NHM	Osteocrust	*		Suncus murinus	Kenya
94.3.9.23	NHM	Osteocrust			Suncus murinus	Oman
98.6.9.22	NHM	Osteocrust	*		Suncus murinus	Yemen
CO 122	Duplantier	Tissue in EtOH	*	1999	Suncus murinus	Cormoros Islands
CO 20	Duplantier	Tissue in EtOH	*	1999	Suncus murinus	Cormoros Islands
DA 229	Duplantier	Tissue in EtOH	*	1999	Suncus murinus	Madagascar
DA 240	Duplantier	Tissue in EtOH	*	1999	Suncus murinus	Madagascar
DA 241	Duplantier	Tissue in EtOH	*	1999	Suncus murinus	Madagascar
DA 248	Duplantier	Tissue in EtOH	*	1999	Suncus murinus	Madagascar
DF 501	Duplantier	Tissue in EtOH	*	1999	Suncus murinus	Madagascar
DF 503	Duplantier	Tissue in EtOH	*	1999	Suncus murinus	Madagascar
DF 504	Duplantier	Tissue in EtOH	*	1999	Suncus murinus	Madagascar
DF 517	Duplantier	Tissue in EtOH	*	1999	Suncus murinus	Madagascar
DF 740	Duplantier	Tissue in EtOH		1999	Suncus murinus	Madagascar
DF 741	Duplantier	Tissue in EtOH	*	1999	Suncus murinus	Madagascar

SAMPLE ID	LOCALITY	LAT	LONG	ACC	REASON
57.398	Male Atoll	4.42	73.50	2	None
59.539		-0.69	73.16	3	
59.540	Addu atoll, Gan Island	-0.69	73.16	2	
27642	Mahl Valley	33.00	74.00	3	Cont
27643	Arapal	34.10	75.10	2	
28632	Pandrettan	34.056	74.860	2	
29760	Chanda Dist, Allapalli Forest	19.46	80.06	2	
29761	Chanda Dist, Allapalli Forest	19.46	80.06	2	
29767	Ghatgain Kasra	19.38	73.29	3	
30966	Phouc Mon	16.70	107.18	2	Miss
30969	Phouc Mon	16.70	107.18	2	
35402	Lingtam	27.23	88.72	1	
47969	Yenping	26.64	118.18	2	
47970	Yenping	26.64	118.18	2	
54539	Luchnow	26.86	80.95	2	
57605	Kathmandu, Swiss Dairy	27.72	85.32	1	Miss
62300		?	?	5	No Location
65396	Mussoorie, Landour	30.47	78.10	1	Miss
65397	Mussoorie, Landour	30.47	78.10	1	
65398	Mussoorie, Landour	30.47	78.10	1	
65399	Mussoorie, Landour	30.47	78.10	1	
76159	Palasbari	26.13	91.50	1	
76161	Palasbari	26.13	91.50	1	
76166	Palasbari	26.13	91.50	1	
76169	Khasi Hills, Mawphlang	25.47	91.77	1	Miss
76171	Khasi Hills, Mawphlang	25.47	91.77	1	
76173	Khasi Hills, Mawryngkueng	25.67	91.90	1	
76174	Khasi Hills, Mawryngkueng	25.67	91.90	1	
76176	Khasi Hills, Mawryngkueng	25.67	91.90	1	

76177	Khasi Hills, Mawryngkueng	25.67	91.90	1	
76178	Khasi Hills, Mawryngkueng	25.67	91.90	1	
82510	Chanda	19.95	79.30	1	
82511	Chanda	19.95	79.30	1	
82516	Bahgownie, Darbhanga	26.17	85.90	2	
82521	Almora, Kumaon	29.62	79.67	1	
82522	Kangra Valley, Dhamtal	32.27	74.90	1	
82525	Kathiawar, Junagadh, Kudra	21.56	70.47	2	
82533	Coorg, Virajpet	12.20	75.80	1	Miss
82534	Coorg, Virajpet	12.20	75.80	1	
82536	Madura, Cumbum	9.74	77.29	2	Deam
82538	Toungoo	18.93	96.43	1	Miss
82539	Toungoo	18.93	96.43	1	
82544	Khasi Hills, Nangboh	25.90	91.88	1	None
85857	Taipei	25.05	121.53	1	Cont
92215	Makaragama, seaboard	7.35	80.13	1	None
92216	Makaragama, seaboard	7.35	80.13	1	
94132	Deorali	28.12	84.17	1	
94133	Kanchanpur, Sati, 32 mi E Dhangadi	28.68	81.12	1	
98449	Kuala Lumpur	3.20	101.68	1	
98450	Kuala Lumpur	3.20	101.68	1	
99690	Colombo Town	6.93	79.86	2	
99691	Colombo Town	6.93	79.86	2	
99778	Tamatave, Moramanga, Perinet	-18.93	48.42	1	
100713	Lac Kin Kony	-15.71	46.32	2	
100715	Hellville.	-13.31	48.26	2	
100977	Lombosang	-5.26	119.85	2	
100979	Lombosang	-5.26	119.85	2	
100980	Lombosang	-5.26	119.85	2	Miss
101908	Cheribon	-6.71	108.57	2	

101910	Cheribon	-6.71	108.57	2	
101918	Cheribon	-6.71	108.57	2	
101919	Cheribon	-6.71	108.57	2	
101920	Cheribon	-6.71	108.57	2	
106739	Ngabang, Landak River	0.38	109.95	2	None
106740	Ngabang, Landak River	0.38	109.95	2	None
106743	Ngabang, Landak River	0.38	109.95	2	Deam
106744	Ngabang, Landak River	0.38	109.95	2	None
106893	Ngabang, Landak River	0.38	109.95	2	Deam
106894	Ngabang, Landak River	0.38	109.95	2	None
113033	Lonkin	25.66	96.38	2	
143866	Rice paddy	16.10	120.40	2	
146753	Beza Mahafaly Special Reserve	-23.65	44.65	1	
150075		7.70	80.70	4	None
150076		7.70	80.70	4	None
150087	Patna	25.61	85.14	2	
154598	Parc National de la Montagne d'Ambre	-12.53	49.17	1	
155232	Depok	-6.42	106.81	2	
156338	Parc National de la Montagne d'Ambre	-12.52	49.17	1	
163037	Nilgiri Hills	11.49	76.69	2	
163038	Nilgiri Hills	11.49	76.69	2	
163040	Nilgiri Hills	11.49	76.69	2	None
163042	Nilgiri Hills	11.49	76.69	2	None
163043	Avalanch	13.04	80.25	2	None
165448	Station Forestiere de Tampolo	-17.29	49.41	1	
165498	RS d'Ambohitantely, 24 km NE Ankazobe	-18.17	47.28	1	
165499	RS d'Ambohitantely, 24 km NE Ankazobe	-18.17	47.28	1	
165501	RS d'Ambohitantely, 24 km NE Ankazobe	-18.11	47.25	1	
165570	RS d'Ambohitantely, 24 km NE Ankazobe	-18.17	47.28	1	
165571	RS d'Ambohitantely, 24 km NE Ankazobe	-18.17	47.28	1	

165776	Taipei	25.10	121.55	2	
165779	Taipei	25.10	121.55	2	
165781	Taipei	25.10	121.55	2	
165790	Taipei	25.10	121.55	2	
165791	Taipei	25.10	121.55	2	
166242	RS de Manongarivo	-13.98	48.42	1	Miss
166243	RS de Manongarivo, 12.8 km SW Antanambao	-13.98	48.42	1	
166244	RS de Manongarivo, 14.5 km SW Antanambao	-14.00	48.43	1	
171135	Jabbulpore	23.19	79.95	2	Cont
171136	Dehra Dun	30.32	78.03	2	
171137	Rithi	23.91	80.15	2	
171138		?	?	5	No Location
171141	Jubbulpore	23.17	79.95	2	None
176115	PN de Kirindy-Mite	-20.79	44.15	1	
178588	RNI de Namoroka	-16.41	45.31	1	
183968	Foulpointe, Foret d'Analalava	-17.70	49.46	1	
183969	Foulpointe, Foret d'Analalava	-17.69	49.46	1	
184587	Manila	14.60	120.98	2	
184666	Karachi City	24.92	67.04	2	
194437	Patsi, Universite d'Anjouan	-12.16	44.43	1	
194579	RS Ambohijanahary	-18.27	45.41	1	
194596	Ambohijanahary Mountain	-18.53	45.47	1	
194597	Ambohijanahary Mountain	-18.52	45.43	1	
207468	Cebu City	10.30	123.89	2	
215539	Kanda National Park	22.31	80.60	2	
215541	Kanda National Park	22.31	80.60	2	Miss
217270	Las Bela; Hawke's Bay	24.86	66.86	2	
226976	Gumbasa Proj.,Kalawara.	-1.18	119.94	2	None
226984	Gumbasa Proj.,Kalora.	-1.34	120.04	2	None
240757	Bong Song	16.78	106.97	2	

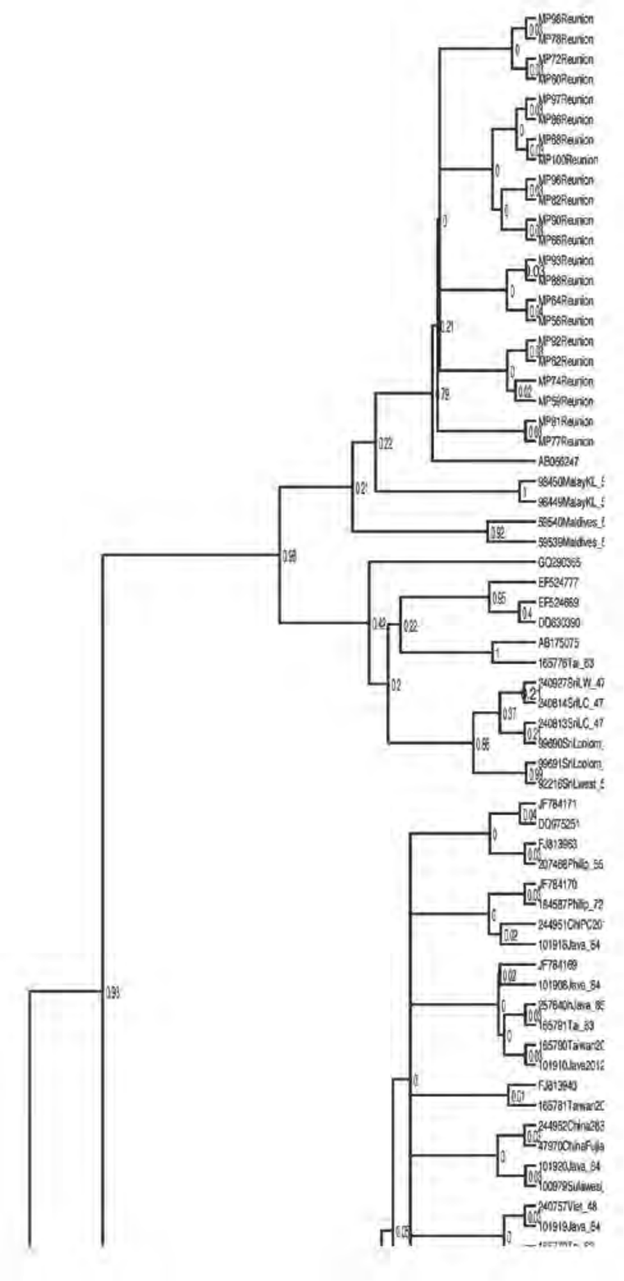
240805	Wariyapola	7.62	80.24	2	Cont
240810	Wariyapola	7.62	80.24	2	
240811	Kandy	7.287	80.64	2	None
240813	Kandy	7.29	80.64	2	
240814	Kandy	7.29	80.64	2	
240819	Kankesanturai	9.81	80.04	2	None
240824	Tanamalwila	6.44	81.13	2	
240825	Tanamalwila	6.44	81.13	2	
240927	Dehiwala (seaboard)	6.86	79.86	2	
244951	Xigiao	22.92	113.16	3	Miss
244952	Xigiao	22.92	113.16	3	
257640	Buitenzorg	-6.59	106.80	2	
326323	Sialkot District, Charwa Village	32.48	74.86	2	Cont
326328	Hazara District, Balakot	34.55	73.35	2	
359624	Bojolali Regency, Paras; Panderedjo	-7.51	110.60	3	
1-IN-010606-SM-1	Saint-Louis	-21.26	55.46	1	
1-IN-010606-SM-2	Saint-Louis	-21.26	55.46	1	Low diversity
1-IN-010606-SM-3	Saint-Louis	-21.26	55.46	1	
1-IN-010606-SM-4	Saint-Louis	-21.26	55.46	1	Low diversity
1-IN-190107-SM-5	Saint-Louis	-21.22	55.46	1	Low diversity
1-IN-190107-SM-6	Saint-Louis	-21.22	55.46	1	Low diversity
1-IN-190107-SM-7	Saint-Louis	-21.22	55.46	1	Low diversity
1-IN-230107-SM-21	Saint-Louis	-21.28	55.38	1	Low diversity
1-IN-230107-SM-7	Saint-Louis	-21.28	55.38	1	
1-IN-230107-SM-8	Saint-Louis	-21.28	55.38	1	
1-IN-230107-SM-9	Saint-Louis	-21.28	55.38	1	
1-IN-230506-SM-6	Saint-Louis	-21.26	55.42	1	Low diversity
1-IN-230506-SM-7	Saint-Louis	-21.26	55.42	1	
1-IN-250506-SM-2	Saint-Louis	-21.25	55.43	1	Low diversity
1-IN-260506-SM-1	Saint-Louis	-21.24	55.43	1	Low diversity

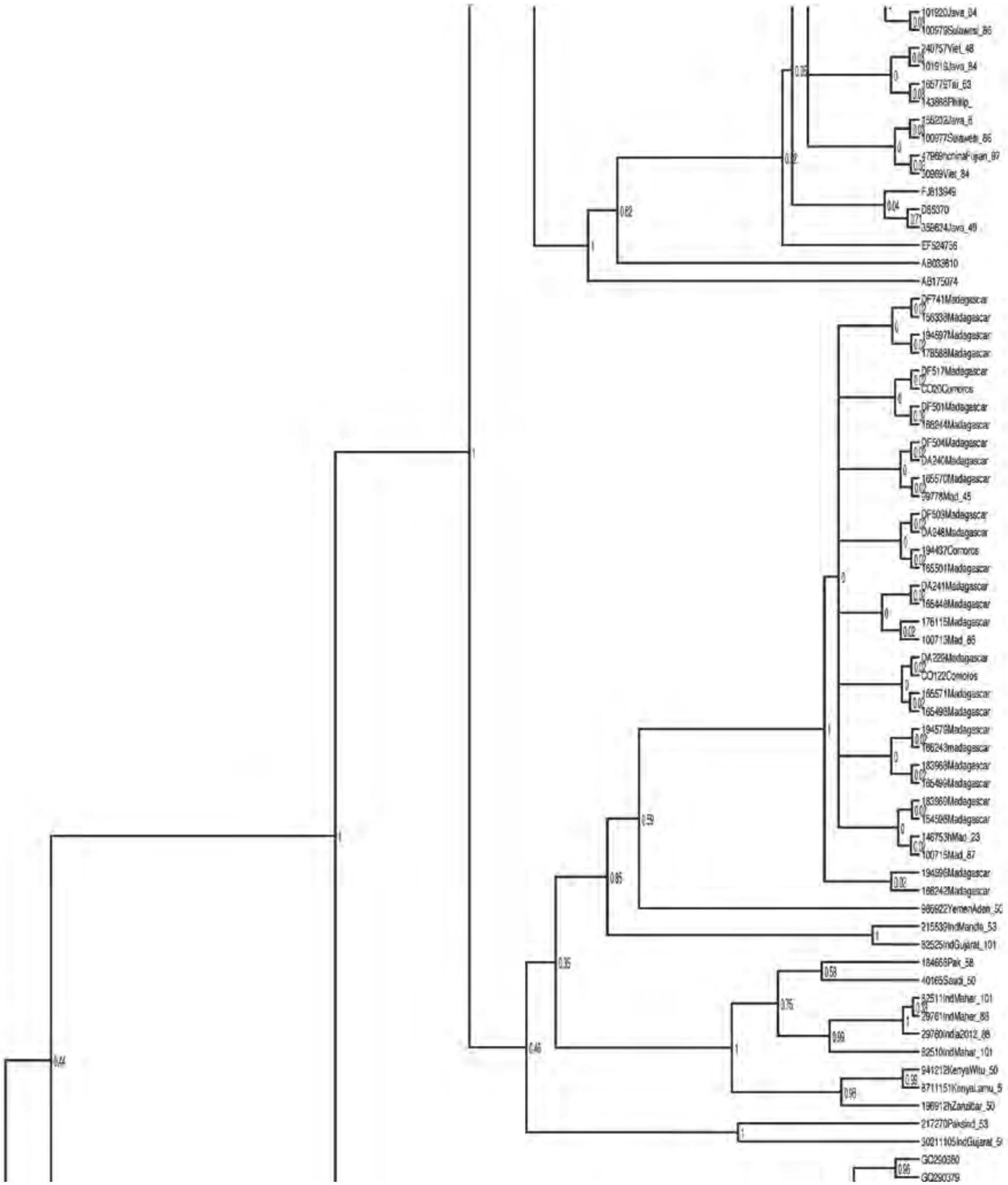
1-IN-260506-SM-2	Saint-Louis	-21.24	55.43	1	Low diversity
1-IN-300506-SM-2	Saint-Louis	-21.27	55.43	1	Low diversity
1-IN-300506-SM-4	Saint-Louis	-21.28	55.43	1	Low diversity
1-IN-300506-SM-5	Saint-Louis	-21.28	55.43	1	
1-IN-310506-SM-1	Saint-Louis	-21.27	55.43	1	Low diversity
18.8.5.1	Tigris River	35.33	43.35	1	None
19.6.9.12	Zanzibar	-6.13	39.32	2	
2-IN-130606-SM-1	Saint André	-20.97	55.70	1	Low diversity
2-IN-130606-SM-3	Saint André	-20.97	55.70	1	Low diversity
2-IN-130606-SM-4	Saint André	-20.97	55.70	1	
2-IN-130606-SM-5	Saint André	-20.97	55.70	1	
2-IN-130606-SM-6	Saint André	-20.97	55.70	1	Low diversity
2-IN-140606-SM-1	Saint André	-20.92	55.63	1	Low diversity
2-IN-140606-SM-2	Saint André	-20.97	55.70	1	Low diversity
2-IN-140606-SM-3	Saint André	-20.97	55.70	1	Low diversity
2-IN-140606-SM-4	Saint André	-20.97	55.70	1	
2-IN-150606-SM-1	Saint André	-20.92	55.63	1	Low diversity
2-IN-150606-SM-10	Saint André	-20.92	55.63	1	Low diversity
2-IN-150606-SM-2	Saint André	-20.92	55.63	1	
2-IN-150606-SM-3	Saint André	-20.92	55.63	1	Low diversity
2-IN-150606-SM-4	Saint André	-20.92	55.63	1	
2-IN-150606-SM-5	Saint André	-20.92	55.63	1	Low diversity
2-IN-150606-SM-7	Saint André	-20.92	55.63	1	Low diversity
2-IN-150606-SM-8	Saint André	-20.92	55.63	1	
2-IN-150606-SM-9	Saint André	-20.92	55.63	1	
2-IN-160606-SM-2	Saint André	-20.94	55.68	1	Low diversity
2-IN-160606-SM-4	Saint André	-20.94	55.68	1	
24.3.8.10	Zanzibar	-6.13	39.32	3	None
3-IN-270606-SM-10	Cilaos	-21.13	55.47	1	Low diversity
3-IN-270606-SM-11	Cilaos	-21.12	55.48	1	Low diversity

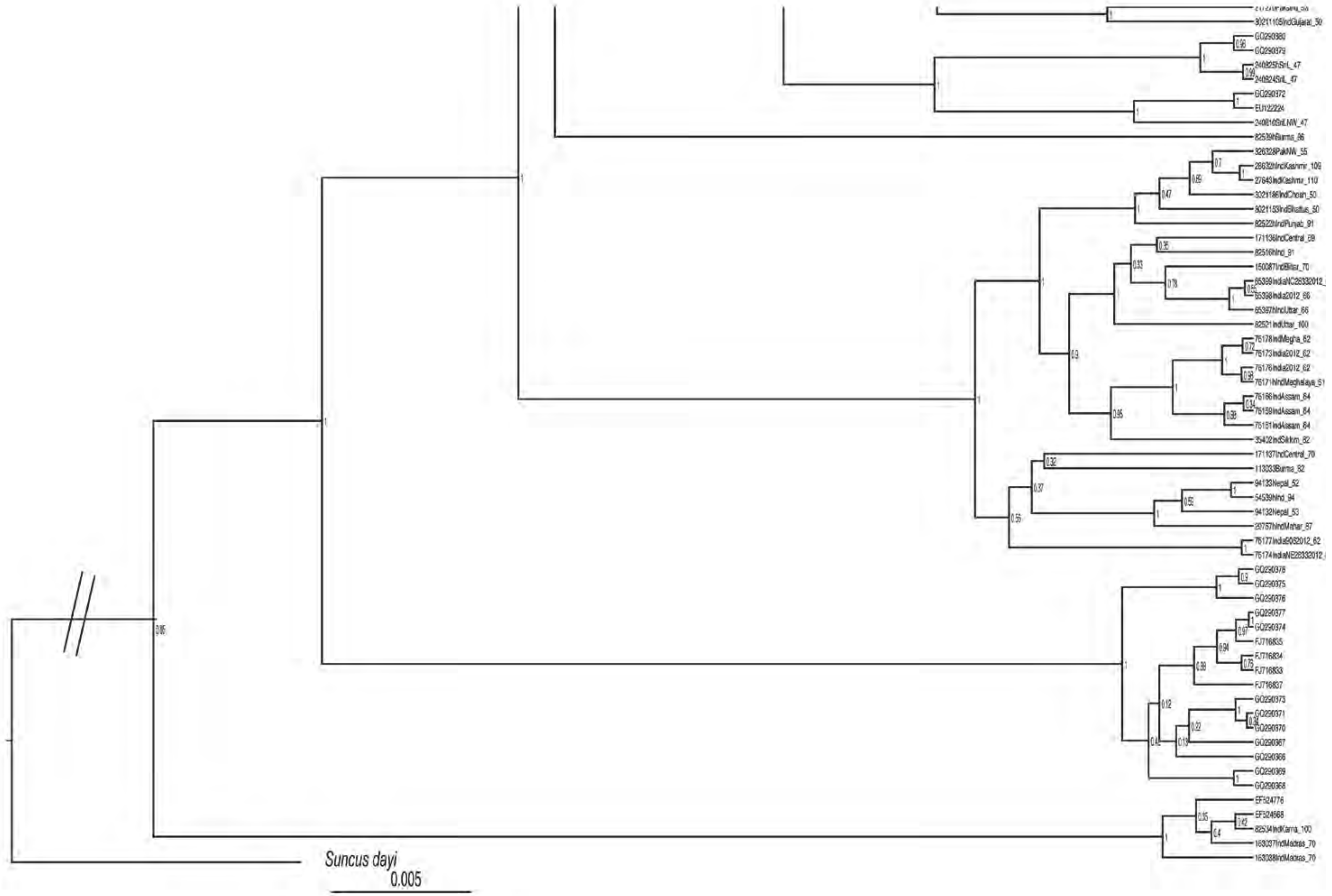
3-IN-270606-SM-12	Cilaos	-21.13	55.47	1	Low diversity
3-IN-270606-SM-13	Cilaos	-21.12	55.48	1	Low diversity
3-IN-270606-SM-15	Cilaos	-21.13	55.47	1	Low diversity
3-IN-270606-SM-16	Cilaos	-21.13	55.47	1	Low diversity
3-IN-270606-SM-17	Cilaos	-21.14	55.48	1	
3-IN-270606-SM-18	Cilaos	-21.13	55.47	1	Low diversity
3-IN-270606-SM-19	Cilaos	-21.14	55.48	1	
3-IN-270606-SM-20	Cilaos	-21.14	55.48	1	Low diversity
3-IN-270606-SM-5	Cilaos	-21.13	55.47	1	
3-IN-270606-SM-7	Cilaos	-21.13	55.47	1	Low diversity
3-IN-280606-SM-1	Cilaos	-21.14	55.46	1	Low diversity
3-IN-280606-SM-10	Cilaos	-21.13	55.48	1	Low diversity
3-IN-280606-SM-3	Cilaos	-21.14	55.48	1	
3-IN-280606-SM-4	Cilaos	-21.13	55.48	1	
3-IN-280606-SM-5	Cilaos	-21.14	55.48	1	Low diversity
3-IN-280606-SM-6	Cilaos	-21.13	55.48	1	Low diversity
3-IN-280606-SM-7	Cilaos	-21.14	55.48	1	
3-IN-280606-SM-9	Cilaos	-21.13	55.48	1	
30.2.11.105	Gujarat, Cutch	23.22	72.68	3	
30.2.11.53	Bhattus, Hissar	29.15	75.70	3	
30.2.11.86	Choah salt range	32.73	72.98	2	
40.1.65	Jeddah	21.50	39.00	1	
87.11.15.1	Lamu	-2.29	40.91	2	
94.1.21.2	Witu	-2.39	40.44	2	
94.3.9.23	Muscat	23.61	58.54	2	None
98.6.9.22	Aden	12.80	45.03	2	
CO 122	Moroni city	-11.70	43.25	1	
CO 20	Moroni city	-11.71	43.25	1	
DA 229	Ambodifaho	-18.77	49.07	1	
DA 240	Ambodifaho	-18.77	49.07	1	

DA 241	Namahoaka	-18.77	49.07	1	
DA 248	Ambodifaho	-18.77	49.07	1	
DF 501	Ambohimiarina II	-21.32	48.00	1	
DF 503	Ambohimiarina II	-21.32	48.00	1	
DF 504	Ambohimiarina II	-21.32	48.00	1	
DF 517	Ambohimiarina II	-21.32	48.00	1	
DF 740	Antafotenona	-21.30	47.63	1	Miss
DF 741	Antafotenona	-21.30	47.63	1	

Appendix 3.II Bayesian MCMC tree for the whole dataset with full sample identifications (see Figure 3.3)


MP98Reunion
MP78Reunion
MP72Reunion
MP80Reunion
MP97Reunion
MP86Reunion
MP58Reunion
MP100Reunion
MP96Reunion
MP82Reunion
MP90Reunion
MP66Reunion
MP93Reunion
MP88Reunion
MP64Reunion
MP56Reunion
MP92Reunion
MP62Reunion
MP74Reunion
MP59Reunion
MP81Reunion
MP77Reunion
AB066247
98450MalayKL_
96449MalayKL_
59540Maldives_
59539Maldives_
GQ290365
EF524777
EF524669
DQ630390
AB175075
165776Tai_63
240927SriLW_47
240814SriLC_47
240813SriLC_47
99690SriLcolom_
99691SriLcolom_
92216SriLwest_
JF784171
DQ975251
FJ813963
207488Philip_55
JF784170
184587Philip_72
244951ChPC20
101918Java_64
JF784169
101906Java_64
257640hJava_6
165781Tai_63
165790Taiwan20
101910Java2012
FJ813940
165781Taiwan20
244952China283
47970ChinaFujia
101920Java_64
100979Sulawesi_
240757Viet_48
101919Java_64
0.03
0
0.21
0.79
0.22
0.21
0.98
0.92
1
0.95
0.4
0.22
0.49
0.2
0.37
0.56
0.93
0.04
0.02
0.01
0.05


101920Java_84
100979Sulawesi_86
240757Viet_48
101919Java_84
165779Tai_63
143868Philip_
155202Java_6
100977Sulawesi_86
47968hchinaFujian_97
30969Viet_84
FJ813949
D65370
358624Java_40
EF524736
AB033610
AB175074
DF741Madagascar
156338Madagascar
194597Madagascar
178588Madagascar
DF517Madagascar
CO20Comoros
DF501Madagascar
186244Madagascar
DF504Madagascar
DA240Madagascar
165570Madagascar
99778Mad_45
DF509Madagascar
DA248Madagascar
194437Comoros
165501Madagascar
DA241Madagascar
165448Madagascar
176115Madagascar
100713Mad_85
DA228Madagascar
CO122Comoros
165571Madagascar
165498Madagascar
194579Madagascar
186243madagascar
183968Madagascar
165499Madagascar
183969Madagascar
154586Madagascar
146753hMad_23
100715Mad_87
194596Madagascar
186242Madagascar
985922YemenAden_50
215539IndMandle_53
82525IndGujarat_101
184666Pak_58
40165Saudi_50
82511IndMahar_101
29761IndMahar_88
29760India2012_88
82510IndMahar_101
941212KenyaWitu_50
8711151KenyaLamu_5
196912hZanzibar_50
217270Paksind_53
30211105IndGujarat_5
GQ290380
GQ290379
0.82
0.46
0.35
0.85
0.59
0.44


30211105IndGujarat_50
GQ290380
GQ290379
240825hSriL_47
240824SriL_47
GQ290372
EU122224
240810SriLNW_47
82539hBurma_86
326328PakNW_55
28632hIndKashmir_109
27643IndKashmir_110
3021186IndChoah_50
3021153IndBhattus_50
82522hIndPunjab_91
171136IndCentral_69
82516hInd_91
150087IndBihar_70
65399IndiaNC28332012_
65398India2012_66
65397hIndUttar_66
82521IndUttar_100
76178IndMegha_62
76173India2012_62
76176India2012_62
76171hIndMeghalaya_61
76166IndAssam_64
76159IndAssam_64
76161IndAssam_64
35402IndSikkim_82
171137IndCentral_70
113033Burma_82
94133Nepal_52
54539hInd_94
94132Nepal_53
29757hIndMahar_87
76177India9082012_62
76174IndiaNE28332012_
GQ290378
GQ290375
GQ290376
GQ290377
GQ290374
FJ716835
FJ716834
FJ716833
FJ716837
GQ290373
GQ290371
GQ290370
GQ290367
GQ290366
GQ290369
GQ290368
EF524776
EF524868
82534IndKarna_100
163037IndMadras_70
163038IndMadras_70
Suncus dayi
0.005

Appendix 4.I Nominal *Rattus rattus* collection information. * denotes used in this study. AMNH=American Museum of Natural History; FMNH=Field Museum of Natural History; NHM=British Natural History Museum. REASON=if sample not used, reason why (NO=no amplification; MISS=missing sequence data; CONT= possible contamination of sequence; NONE=not extracted; ID?=Not *Rattus*). Cyt*b*+Dloop and Cyt*b* refer to the taxon affinity within clades of the Bayesian MCMC analysis for the Cyt*b* + D-loop and Cyt*b* trees respectively (A, A1, A2, A3, B, C are clades within *R. rattus* Cyt*b* + D-loop tree; R1, R2, R3, R4 are clades within *R. rattus* Cyt*b* tree; K= *R. rattus* kandianus (?); T= *R. tanezumi;* III=*R. rattus* III; IV= *R. rattus* IV, SIK= unknown *Rattus* sp.)

	Supplier	Sample ID	Lab Code	Reason	Cytb+Dloop	Cytb	Type	Taxon ID	Location
*	Hillman	Hillman001	AT001		T	T	Tissue	Rattus rattus	Diego Garcia
*	Hillman	Hillman002	AT002		T	T	Tissue	Rattus rattus	Diego Garcia
*	Hillman	Hillman003	AT003		T	T	Tissue	Rattus rattus	Diego Garcia
*	Hillman	Hillman004	AT004		T	T	Tissue	Rattus rattus	Diego Garcia
*	Hillman	Hillman005	AT005		T	T	Tissue	Rattus rattus	Diego Garcia
*	Hillman	Hillman006	AT006		T	T	Tissue	Rattus rattus	Diego Garcia
	Hillman	Hillman007	AT007	NO			Tissue	Rattus rattus	Diego Garcia
	Hillman	Hillman008	AT008	NO			Tissue	Rattus rattus	Diego Garcia
*	Hillman	Hillman009	AT009		T	T	Tissue	Rattus rattus	Diego Garcia
*	Hillman	Hillman010	AT010		T	T	Tissue	Rattus rattus	Diego Garcia
*	Hillman	Hillman011	AT011		T	T	Tissue	Rattus rattus	Diego Garcia
*	Hillman	Hillman012	AT012		T	T	Tissue	Rattus rattus	Diego Garcia
*	Hillman	Hillman013	AT013		T	T	Tissue	Rattus rattus	Diego Garcia
*	Hillman	Hillman014	AT014		T	T	Tissue	Rattus rattus	Diego Garcia
	Hillman	Hillman015	AT015	NO			Tissue	Rattus rattus	Diego Garcia
*	Hillman	Hillman016	AT016		T	T	Tissue	Rattus rattus	Diego Garcia
*	Hillman	Hillman017	AT017		T	T	Tissue	Rattus rattus	Diego Garcia
*	Hillman	Hillman018	AT018		T	T	Tissue	Rattus rattus	Diego Garcia
*	Hillman	Hillman019	AT019		T	T	Tissue	Rattus rattus	Diego Garcia
*	Hillman	Hillman020	AT020		T	T	Tissue	Rattus rattus	Diego Garcia
*	Hillman	Hillman021	AT021		T	T	Tissue	Rattus rattus	Diego Garcia
	Hillman	Hillman022	AT022	NO			Tissue	Rattus rattus	Diego Garcia
	Hillman	Hillman023	AT023	NO			Tissue	Rattus rattus	Diego Garcia
	Hillman	Hillman024	AT024	NO			Tissue	Rattus rattus	Diego Garcia
	Hillman	Hillman025	AT025	NO			Tissue	Rattus rattus	Diego Garcia
*	Hillman	Hillman026	AT026		T	T	Tissue	Rattus rattus	Diego Garcia
*	Hillman	Hillman027	AT027		T	T	Tissue	Rattus rattus	Diego Garcia
*	Hillman	Hillman028	AT028		T	T	Tissue	Rattus rattus	Diego Garcia
*	FMNH	187398	HE001		A	R	Tissue	Rattus rattus	Tanzania, Mafia
*	FMNH	187400	HE002		A	R	Tissue	Rattus rattus	Tanzania, Mafia
*	FMNH	192839	HE003		A	R	Tissue	Rattus rattus	Tanzania, Pemba
*	FMNH	192841	HE004		A	R	Tissue	Rattus rattus	Tanzania, Pemba

*	FMNH	198176	HE005		C	R1	Tissue	Rattus rattus	Tanzania, Zanzibar
*	FMNH	165732	HE006		A	R	Tissue	Rattus rattus rattus	Madagascar
*	FMNH	175937	HE007		A	R	Tissue	Rattus rattus rattus	Madagascar
*	FMNH	194717	HE008		A	R	Tissue	Rattus rattus	Madagascar
*	FMNH	179285	HE009			R	Tissue	Rattus rattus	Madagascar
*	FMNH	178864	HE010		C	R1	Tissue	Rattus rattus	Madagascar
*	FMNH	194574	HE011		A	R	Tissue	Rattus rattus	Madagascar
	FMNH	204453	HE012	CONT			Tissue	Rattus rattus	Madagascar
*	FMNH	178860	HE013		C	R1	Tissue	Rattus rattus	Madagascar
*	FMNH	194613	HE014		A	R	Tissue	Rattus rattus	Madagascar
*	FMNH	192838	HE015		A	R	Tissue	Rattus rattus	Tanzania, Pemba
*	FMNH	192905	HE016		A	R	Tissue	Rattus rattus	Tanzania, Pemba
	FMNH	187399	HE017	MISS			Tissue	Rattus rattus	Tanzania, Mafia
*	FMNH	187397	HE018		C	R1	Tissue	Rattus rattus	Tanzania, Mafia
*	FMNH	187396	HE019			R1	Tissue	Rattus rattus	Tanzania, Mafia
*	FMNH	166260	HE020		C	R1	Tissue	Rattus rattus rattus	Madagascar
*	FMNH	194586	HE021			R	Tissue	Rattus rattus	Madagascar
*	FMNH	178680	HE022		A	R	Tissue	Rattus rattus	Madagascar
*	FMNH	192840	HE023		A	R	Tissue	Rattus rattus	Tanzania, Pemba
*	FMNH	166083	HE024		A	R	Tissue	Rattus rattus rattus	Madagascar
*	FMNH	176119	HE025		A	R	Tissue	Rattus rattus rattus	Madagascar
*	FMNH	175948	HE026		A	R	Tissue	Rattus rattus rattus	Madagascar
*	FMNH	166225	HE027		C	R1	Tissue	Rattus rattus rattus	Madagascar
*	FMNH	176023	HE028		A	R	Tissue	Rattus rattus rattus	Madagascar
	AMNH	28639	NONE	NONE			Skin	Mus musculus bactrianus	Kashmir
	AMNH	54731	NONE	NONE			Skin	Rattus rattus thai	Thailand
	AMNH	54834	NONE	NONE			Skin	Rattus rattus rufescens	Thailand
	AMNH	83840	NONE	NONE			Skin	Rattus rattus rattus	Tanzania
	AMNH	83841	NONE	NONE			Skin	Rattus rattus rattus	Tanzania
	AMNH	114380	NONE	NONE			Skin	Rattus rattus rattus	Kenya
	AMNH	150165	NONE	NONE			Skin	Rattus rattus rattus	Australia
	AMNH	150169	NONE	NONE			Skin	Rattus rattus rattus	Australia
	AMNH	151493	NONE	NONE			Skin	Rattus rattus rattus	Australia

	FMNH	142241	R009	MISS			Osteocrust	Rattus rattus arboreus	Nepal
	FMNH	31869	R010	MISS			Osteocrust	Rattus rattus flavipectus	Vietnam, Bac Tan Trai
	FMNH	114187	R011	CONT			Osteocrust	Rattus rattus brunneus	Nepal
*	FMNH	31953	R012		T	T	Osteocrust	Rattus rattus flavipectus	Laos, Phong Saly
	FMNH	80170	R013	MISS			Osteocrust	Rattus rattus diardii	Malaysia, Sarawak
	FMNH	108906	R014	MISS			Osteocrust	Rattus rattus diardii	Malaysia, Sabah, Kinabalu
	FMNH	40521	R015	CONT			Osteocrust	Rattus rattus flavipectus	China, Sezchuan, Kao Ku
	FMNH	31912	R016	CONT			Osteocrust	Rattus rattus flavipectus	Vietnam, Tonkin
	FMNH	82954	R017	CONT			Osteocrust	Rattus rattus arboreus	India, Daltonganj
*	FMNH	76617	R018		T	T	Osteocrust	Rattus rattus brunneusculus	India, Assam, Manipur
	FMNH	88453	R019	CONT			Osteocrust	Rattus rattus diardii	Borneo, Sandakan
	FMNH	82952	R020	MISS			Osteocrust	Rattus ratttus kandianus	Sri Lanka, Udugama
*	FMNH	98814	R021		A	R	Osteocrust	Rattus rattus rattus	Egypt, Aswan
	FMNH	77730	R022	CONT			Osteocrust	Rattus rattus rattus	Egypt, Wadi Natrun
	FMNH	100710	R023	NO			Osteocrust	Rattus rattus rattus	Egypt, Cairo
*	FMNH	85360	R024		A	R	Osteocrust	Rattus rattus rattus	Egypt, Faiyum
*	FMNH	77242	R025		B	R3	Osteocrust	Rattus rattus rattus	Egypt, Giza
*	FMNH	74632	R026		A	R	Osteocrust	Rattus rattus rattus	Egypt, Qena
	FMNH	99117	R027 (R267)	ID?			Osteocrust	Rattus rattus rattus	Ethiopia, Addis Abbaba
*	FMNH	97524	R028		A	R	Osteocrust	Rattus rattus rattus	Iran, Gorgan
	FMNH	85420	R029	MISS			Osteocrust	Rattus rattus rattus	Kenya, Arabuka
*	FMNH	85348	R030		A	R	Osteocrust	Rattus rattus rattus	Kenya, Kitero Village
*	FMNH	85350	R031		A3	R4	Osteocrust	Rattus rattus rattus	Kenya, Njoro
*	FMNH	85352	R032		A3	R4	Osteocrust	Rattus rattus rattus	Kenya, Njoro
*	FMNH	42373	R033 (R266)		A	R	Osteocrust	Rattus rattus rattus	Mali, Sotuba
*	FMNH	99603	R034		A	R	Osteocrust	Rattus rattus rattus	Saudi Arabia, Hillat Muhaish
	FMNH	86317	R035	NO			Osteocrust	Rattus rattus rattus	Tanzania, Tanganyika, Tengeru
*	FMNH	86316	R036		C	R1	Osteocrust	Rattus rattus rattus	Tanzania, Arusha, Tengeru
*	FMNH	77983	R037 (R264)		C	R1	Osteocrust	Rattus rattus rattus	Yemen, San'a
*	FMNH	82961	R038		A2	R2	Osteocrust	Rattus rattus wroughtoni	India, Benhope
	FMNH	186995	R039	ID?			Osteocrust	Rattus rattus	Tanzania, Lindi Region
*	FMNH	187064	R040		A	R	Osteocrust	Rattus rattus	Tanzania, Morogoro Region
	FMNH	81931	R041	NO			Osteocrust	Rattus rattus rattus	Angola, Canzele, Quai Sai

	AMNH	163133	NONE	NONE			Skin	Rattus rattus wroughtoni	India
	AMNH	167919	NONE	NONE			Skin	Rattus rattus thai	Thailand
	AMNH	168564	NONE	NONE			Skin	Rattus rattus rattus	South Africa
	FMNH	170887	NONE	NONE			Tissue	Rattus rattus rattus	Madagascar
	AMNH	171147	NONE	NONE			Skin	Rattus rattus rufescens	India
	FMNH	172622	NONE	NONE			Tissue	Rattus rattus rattus	Madagascar
	FMNH	172676	NONE	NONE			Tissue	Rattus rattus rattus	Madagascar
	FMNH	173155	NONE	NONE			Tissue	Rattus rattus rattus	Madagascar
	FMNH	175914	NONE	NONE			Tissue	Rattus rattus rattus	Madagascar
	FMNH	175938	NONE	NONE			Tissue	Rattus rattus rattus	Madagascar
	FMNH	175939	NONE	NONE			Tissue	Rattus rattus rattus	Madagascar
	AMNH	187533	NONE	NONE			Skin	Rattus rattus rattus	Kenya
	AMNH	187599	NONE	NONE			Skin	Rattus rattus rattus	Kenya
	AMNH	187667	NONE	NONE			Skin	Rattus rattus rattus	Kenya
	AMNH	203218	NONE	NONE			Skin	Rattus rattus rattus	Egypt
	AMNH	203219	NONE	NONE			Skin	Rattus rattus rattus	Egypt
	AMNH	205086	NONE	NONE			Skin	Rattus rattus rattus	Tanzania
	AMNH	205135	NONE	NONE			Skin	Rattus rattus rattus	Tanzania
	AMNH	215602	NONE	NONE			Skin	Rattus sp.	India
	AMNH	215603	NONE	NONE			Skin	Rattus sp.	India
	AMNH	215605	NONE	NONE			Skin	Rattus sp.	India
	AMNH	251684	NONE	NONE			Skin	Rattus rattus	Bangladesh
	AMNH	267661	NONE	NONE			Skin	Rattus rattus	Indonesia
	FMNH	35586	NONE	NONE				Rattus rattus brunneusculus	India, Bengal, Mangpu
	FMNH	85340/6	NONE	NONE				Rattus rattus rattus	Kenya, Rift Valley
*	AMNH	163132	R001		A	R	Osteocrust	Rattus rattus wroughtoni	India
*	AMNH	163134	R002		A	R	Osteocrust	Rattus sp.	India
*	AMNH	163147	R003		A	R	Osteocrust	Rattus rattus wroughtoni	India
*	AMNH	55587	R004		A	R	Osteocrust	Rattus rattus rattus	Tanzania
	FMNH	109188	R005	NO			Osteocrust	Rattus rattus pelengensis	Indonesia
*	AMNH	34045	R006 (R263)		A	R	Osteocrust	Rattus rattus rattus	Madagascar, Ambantondrazaka
	FMNH	122589	R007	MISS			Osteocrust	Rattus rattus rattus	Turkey, Adana, Haruniye
*	FMNH	76616	R008		T	T	Osteocrust	Rattus rattus brunneusculus	India, Assam, Manipur

*	FMNH	103666	R042		III	III	Osteocrust	Rattus rattus rattus	Afghanistan, Jalalabad
	FMNH	82969	R043	NO			Osteocrust	Rattus rattus macmillani	Burma, Hkmati
	FMNH	43689	R044	NO			Osteocrust	Rattus rattus rattus	Cameroon, Sakbayeme
*	FMNH	34835	R045		T	T	Osteocrust	Rattus rattus brunneusculus	India, Bengal, Mangpu
*	FMNH	179191	R046 (R268)		A	R	Osteocrust	Rattus rattus	Jordan, Irbid, Hemmah
*	FMNH	179192	R047 (R269)		A	R	Osteocrust	Rattus rattus	Jordan, Irbid, Hemmah
*	FMNH	122595	R048		A	R	Osteocrust	Rattus rattus rattus	Turkey, Adana, Sabun Cayl
*	FMNH	122594	R049		A	R	Osteocrust	Rattus rattus rattus	Turkey, Adana, Haruniye
*	FMNH	34046	R050		A	R	Osteocrust	Rattus rattus rattus	Madagascar, Miandrivaza
*	FMNH	85366	R051		A	R	Osteocrust	Rattus rattus	Madagascar, Ambohimahavelona
*	FMNH	85415	R052		A	R	Osteocrust	Rattus rattus rattus	Madagascar, Ambohimahavelona
	FMNH	85417	R053	NO			Osteocrust	Rattus rattus rattus	Madagascar, Ambohimahavelona
	FMNH	85418	R054	NO			Osteocrust	Rattus rattus rattus	Madagascar, Ambohimahavelona
	FMNH	85353	R055	MISS			Osteocrust	Rattus rattus rattus	Kenya, Subukia
*	FMNH	35581	R056		T	T	Osteocrust	Rattus rattus brunneusculus	India, Bengal, Mangpu
*	FMNH	86213	R057		C	R1	Osteocrust	Rattus rattus rattus	Tanzania, Tanganyika, Loljoro
	FMNH	85347	R058	NO			Osteocrust	Rattus rattus rattus	Kenya, Njoro
*	FMNH	29777	R059		A	R3	Osteocrust	Rattus rattus narbadae	India, Alapalli
	FMNH	76726	R060	MISS			Osteocrust	Rattus rattus brunneusculus	India, Assam
*	FMNH	76624	R061		T	T	Osteocrust	Rattus rattus brunneusculus	India, Assam, Manipur
	FMNH	108972	R062	MISS			Osteocrust	Rattus rattus diardii	Malaysia, Sabah, Kinabalu
*	FMNH	94207	R063		III	III	Osteocrust	Rattus rattus brunneus	Nepal
	FMNH	80164	R064	CONT			Osteocrust	Rattus rattus diardii	Malaysia, Sarawak
	FMNH	91264	R065	NO			Osteocrust	Rattus rattus rattus	India
*	FMNH	99114	R066 (R265)		A1	R	Osteocrust	Rattus rattus rattus	Ethiopia, Addis Abbaba
*	FMNH	93763	R067		A	R	Osteocrust	Rattus rattus rattus	Egypt, Cairo
*	FMNH	98815	R068		B	R3	Osteocrust	Rattus rattus rattus	Egypt, Aswan
*	FMNH	92222	R069		K	K	Osteocrust	Rattus ratttus kandianus	Sri Lanka, Dehiwala
	FMNH	82953	R070	NO			Osteocrust	Rattus ratttus kandianus	Ceylon, Udugama
*	FMNH	92223	R071		K	K	Osteocrust	Rattus ratttus kandianus	Sri Lanka, Gonapola
*	FMNH	31903	R072		T	T	Osteocrust	Rattus rattus flavipectus	Vietnam, Tonkin
	FMNH	31902	R073	NO			Osteocrust	Rattus rattus flavipectus	Vietnam, Tonkin
	FMNH	80167	R074	NO			Osteocrust	Rattus rattus diardii	Malaysia, Sarawak

*	AMNH	54541	R108		III	III	Osteocrust	Rattus rattus rufescens	India
	AMNH	163757	R109	MISS			Osteocrust	Rattus sladeni	Burma
	AMNH	163760	R110	MISS			Osteocrust	Rattus sladeni	Burma
	AMNH	152970	R111	MISS			Osteocrust	Rattus rattus sapoensis	Indonesia
	AMNH	152971	R112	NO			Osteocrust	Rattus rattus sapoensis	Indonesia
*	AMNH	59872	R113		T	T	Osteocrust	Rattus rattus flavipetus	China
*	AMNH	118995	R114		A	R	Osteocrust	Rattus rattus rattus	Uganda
	AMNH	44726	R115	NO			Osteocrust	Rattus rattus flavipectus	China
*	AMNH	113071	R116		T	T	Osteocrust	Rattus rattus flavipectus	Burma
	AMNH	103073	R117	MISS			Osteocrust	Rattus rattus diardii	Indonesia
	AMNH	109193	R118	MISS			Osteocrust	Rattus rattus pelengensis	Indonesia
*	AMNH	113069	R119		T	T	Osteocrust	Rattus rattus flavipectus	Burma
*	AMNH	107544	R120		T	T	Osteocrust	Rattus rattus diardii	Indonesia
*	AMNH	107978	R121			T	Osteocrust	Rattus tiomanicus roquei	Indonesia
*	AMNH	102397	R122		IV	IV	Osteocrust	Rattus rattus diardii	Indonesia
*	AMNH	107606	R123		IV	IV	Osteocrust	Rattus rattus diardii	Indonesia
*	AMNH	102010	R124		IV	IV	Osteocrust	Rattus rattus diardii	Indonesia
*	AMNH	102112	R125		IV	IV	Osteocrust	Rattus rattus diardii	Indonesia
*	AMNH	101551	R126		IV	IV	Osteocrust	Rattus rattus diardii	Indonesia
*	AMNH	102005	R127			T4	Osteocrust	Rattus rattus diardii	Indonesia
	AMNH	55181	R128	MISS			Osteocrust	Rattus rattus rattus	Zaire
*	AMNH	55530	R129		A	R	Osteocrust	Rattus rattus rattus	Tanzania
*	AMNH	82669	R130		A	R	Osteocrust	Rattus rattus rattus	Congo
*	AMNH	84666	R131		T	T	Osteocrust	Rattus rattus flavipectus	China
	AMNH	109190	R132	MISS			Osteocrust	Rattus rattus pelengensis	Indonesia
*	AMNH	85082	R133		A	R	Osteocrust	Rattus rattus rattus	Angola
*	AMNH	85081	R134		A	R	Osteocrust	Rattus rattus rattus	Angola
	AMNH	109185	R135	MISS			Osteocrust	Rattus rattus pelengensis	Indonesia
*	AMNH	89714	R136		C	R1	Osteocrust	Rattus rattus rattus	Uganda
*	AMNH	85083	R137		A	R	Osteocrust	Rattus rattus rattus	Angola
	AMNH	102815	R138	MISS			Osteocrust	Rattus rattus diardii	Indonesia
	AMNH	102817	R139	MISS			Osteocrust	Rattus rattus diardii	Indonesia
*	AMNH	111608	R140		T	T	Osteocrust	Rattus rattus gigas	China

*	AMNH	111611	R141		T	T	Osteocrust	Rattus rattus flavipectus	China
*	AMNH	111615	R142		T	T	Osteocrust	Rattus rattus flavipectus	China
*	AMNH	111974	R143		T	T	Osteocrust	Apodemus sp.	China
*	AMNH	163749	R144			T	Osteocrust	Rattus sladeni	Burma
	AMNH	163751	R145	MISS			Osteocrust	Rattus sladeni	Burma
*	AMNH	167533	R146		A	R	Osteocrust	Mus sp.	Liberia
	AMNH	205049	R147	NO			Osteocrust	Rattus rattus rattus	Tanzania
	AMNH	214234	R148	NO			Osteocrust	Rattus rattus sumbae	Indonesia
	AMNH	217643	R149	NO			Osteocrust	Rattus rattus diardii	Malaysia
	AMNH	217645	R150	NO			Osteocrust	Rattus rattus diardii	Malaysia
	AMNH	239709	R151	NO			Osteocrust	Rattus rattus diardii	Malaysia
	AMNH	239711	R152	NO			Osteocrust	Rattus rattus diardii	Malaysia
	AMNH	240416	R153	NO			Osteocrust	Rattus rattus diardii	Malaysia
	AMNH	240427	R154	NO			Osteocrust	Rattus rattus diardii	Malaysia
*	AMNH	240949	R155		K	K	Osteocrust	Rattus sp.	Sri Lanka
	AMNH	252183	R156	MISS			Osteocrust	Rattus rattus sumbae	Indonesia
	AMNH	252193	R157	MISS			Osteocrust	Rattus rattus sumbae	Indonesia
	AMNH	106385	R158	MISS			Osteocrust	Rattus rattus diardii	Indonesia
	AMNH	102994	R159	MISS			Osteocrust	Rattus rattus diardii	Indonesia
	AMNH	103526	R160	MISS			Osteocrust	Rattus rattus diardii	Indonesia
*	AMNH	151499	R161			R	Osteocrust	Rattus rattus rattus	Australia
	AMNH	103525	R162	MISS			Osteocrust	Rattus rattus diardii	Indonesia
	AMNH	103292	R163	MISS			Osteocrust	Rattus rattus diardii	Indonesia
	AMNH	252099	R164	MISS			Osteocrust	Rattus rattus diardii	Indonesia
	AMNH	102673	R165	MISS			Osteocrust	Rattus rattus diardii	Indonesia
	AMNH	102580	R166	MISS			Osteocrust	Rattus rattus diardii	Indonesia
*	AMNH	102579	R167			R	Osteocrust	Rattus rattus diardii	Indonesia
	AMNH	102577	R168	MISS			Osteocrust	Rattus rattus diardii	Indonesia
	AMNH	102674	R169	MISS			Osteocrust	Rattus rattus diardii	Indonesia
*	AMNH	151492	R170		IV	IV	Osteocrust	Rattus rattus rattus	Australia
*	AMNH	100686	R171		C	R1	Osteocrust	Rattus rattus rattus	Madagascar
*	AMNH	100726	R172		A	R	Osteocrust	Rattus rattus rattus	Madagascar
*	AMNH	150060	R173		K	K	Osteocrust	Rattus rattus kandianus	Sri Lanka

* AMNH	240955	R174		K	K	Osteocrust	Rattus sp.	Sri Lanka
AMNH	106999	R175	MISS			Osteocrust	Rattus rattus diardii	Indonesia
AMNH	107004	R176	MISS			Osteocrust	Rattus rattus diardii	Indonesia
AMNH	107024	R177	MISS			Osteocrust	Rattus rattus diardii	Indonesia
AMNH	107048	R178	MISS			Osteocrust	Rattus exulans ephippium	Indonesia
* AMNH	103436	R179			T4	Osteocrust	Rattus rattus diardii	Indonesia
* AMNH	103437	R180		IV	IV	Osteocrust	Rattus rattus diardii	Indonesia
* AMNH	88843	R181		A	R	Osteocrust	Rattus rattus rattus	Iran
* AMNH	88844	R182		A	R	Osteocrust	Rattus rattus rattus	Iran
* AMNH	101278	R183		IV	IV	Osteocrust	Rattus rattus palelae	Indonesia
* AMNH	101275	R184		IV	IV	Osteocrust	Rattus pesticulus eurous	Indonesia
AMNH	152986	R185	MISS			Osteocrust	Rattus rattus palelae	Indonesia
AMNH	152985	R186	MISS			Osteocrust	Rattus rattus palelae	Indonesia
* AMNH	162110	R187		A	R	Osteocrust	Rattus rattus rattus	Malawi
* AMNH	162102	R188		A	R	Osteocrust	Rattus rattus rattus	Malawi
* AMNH	205043	R189			R4	Osteocrust	Rattus rattus rattus	Tanzania
* AMNH	81971	R190		A	R	Osteocrust	Rattus rattus rattus	Congo
* AMNH	85724	R191		A	R	Osteocrust	Rattus rattus rattus	Angola
* AMNH	55592	R192		A	R	Osteocrust	Rattus rattus rattus	Tanzania
* AMNH	55591	R193		A3	R4	Osteocrust	Rattus rattus rattus	Tanzania
AMNH	55586	R194	MISS			Osteocrust	Rattus rattus rattus	Tanzania
AMNH	120181	R195	MISS			Osteocrust	Rattus rattus rattus	Congo
* AMNH	120189	R196		A	R	Osteocrust	Rattus rattus rattus	Congo
* AMNH	256646	R197		A	R	Osteocrust	Rattus rattus rattus	Yemen
* AMNH	256645	R198		A	R	Osteocrust	Rattus rattus rattus	Yemen
AMNH	242127	R199	MISS			Osteocrust	Rattus rattus mindanensis	Philippines
* AMNH	43353	R200			T	Osteocrust	Rattus rattus yunnanensis	China
* AMNH	43363	R201		T	T	Osteocrust	Rattus rattus yunnanensis	China
* AMNH	163140	R202		C	R1	Osteocrust	Rattus rattus rufescens	India
* AMNH	163139	R203		C	R1	Osteocrust	Rattus rattus rufescens	India
* FMNH	97516	R204		A	R	Osteocrust	Rattus rattus rattus	Iran, Gorgan
* FMNH	57361	R205			R	Osteocrust	Rattus rattus	Kenya, Tezo, Kilifi
* FMNH	77990	R206		C	R1	Osteocrust	Rattus rattus rattus	Yemen, San'a

*	FMNH	77977	R207		C	R1	Osteocrust	Rattus rattus rattus	Yemen, San'a
*	FMNH	31900	R208		T	T	Osteocrust	Rattus rattus flavipectus	Vietnam, Tonkin
	FMNH	104213	R209	NO			Osteocrust	Rattus rattus brunneus	Nepal
	FMNH	142243	R210	NO			Osteocrust	Rattus rattus brunneus	Nepal
	FMNH	76638	R211	ID?			Osteocrust	Rattus rattus brunneusculus	India, Assam, Manipur
	FMNH	142244	R212	NO			Osteocrust	Rattus rattus brunneus	Nepal
*	FMNH	31897	R213			T	Osteocrust	Rattus rattus flavipectus	Vietnam, Tonkin
	FMNH	31950	R214	NO			Osteocrust	Rattus rattus flavipectus	Laos, Phong Saly
*	FMNH	94261	R215		A	R	Osteocrust	Rattus rattus rattus	Lebanon, Akkar, Halba
*	FMNH	179189	R216			R	Osteocrust	Rattus rattus	Jordan, Irbid, Hemmah
	FMNH	122593	R217	MISS			Osteocrust	Rattus rattus rattus	Turkey, Adana, Haruniye
	FMNH	103664	R218	NO			Osteocrust	Rattus rattus rattus	Afghanistan, Jalalabad
*	FMNH	99115	R219		A1	R	Osteocrust	Rattus rattus rattus	Ethiopia, Addis Abbaba
	FMNH	68310	R220	NO			Osteocrust	Rattus rattus rattus	Egypt, Cairo
	FMNH	100708	R221	NO			Osteocrust	Rattus rattus rattus	Egypt, Cairo
	FMNH	104216	R222	NO			Osteocrust	Rattus rattus brunneus	Nepal
	FMNH	104215	R223	NO			Osteocrust	Rattus rattus brunneus	Nepal
	FMNH	122590	R224	MISS			Osteocrust	Rattus rattus rattus	Turkey, Adana, Haruniye
*	FMNH	85361	R225		A	R	Osteocrust	Rattus rattus rattus	Egypt, Faiyum
*	FMNH	99118	R226		A1	R	Osteocrust	Rattus rattus rattus	Ethiopia, Addis Abbaba
*	FMNH	100733	R227		A	R	Osteocrust	Rattus rattus rattus	Egypt, Asyut
*	FMNH	82950	R228		III	III	Osteocrust	Rattus rattus brunneus	India, Hathiban
*	NHM	6.7.2.30	R229		IV	IV	Osteocrust	Rattus	Andaman Islands
*	NHM	10.7.26.2	R230		K	K	Osteocrust	Rattus	Andaman Islands
*	NHM	10.7.26.3	R231		K	K	Osteocrust	Rattus	Andaman Islands
*	NHM	57.407	R232		T	T	Osteocrust	Rattus rattus ceylonus	Maldives
*	NHM	57.406	R233		T	T	Osteocrust	Rattus rattus ceylonus	Maldives
*	NHM	57.405	R234		T	T	Osteocrust	Rattus rattus ceylonus	Maldives
*	NHM	57.41	R235		T	T	Osteocrust	Rattus rattus kandianus	Maldives
*	NHM	59.549	R236		A	R	Osteocrust	Rattus rattus kandianus	Maldives
*	NHM	59.546	R237		A	R	Osteocrust	Rattus rattus kandianus	Maldives
*	NHM	5.5.8.42	R238			R	Osteocrust	Rattus rattus alexandrinus	Sudan
*	NHM	66.5810	R239		A	R	Osteocrust	Rattus rattus frugivorius	Sudan

*	FMNH	86313/86314	R075		C	R1	Osteocrust	Rattus rattus rattus	Tanzania, Arusha, Tengeru
*	FMNH	66182	R076		T	T	Osteocrust	Rattus rattus thai	Thailand, Kam Pang
*	FMNH	95167	R077		A	R	Osteocrust	Rattus rattus rattus	Zimbabwe, Gatooma
	FMNH	82968	R078	MISS			Osteocrust	Rattus rattus tikos	Burma, Tenasserim
*	FMNH	66183	R079		T	T	Osteocrust	Rattus rattus thai	Thailand, Kam Pang
*	FMNH	195591	R080		C	R1	Osteocrust	Rattus rattus	Tanzania, Tanga Region
*	FMNH	32882	R081		SIK	SIK	Osteocrust	Rattus sikkimensis	China, Hainan, Nodoa
*	FMNH	39420	R082		T	T	Osteocrust	Rattus rattus flavipectus	China, Yunnan
	FMNH	32887	R083	ID?			Osteocrust	Rattus sikkimensis	China, Hainan, Nodoa
*	FMNH	40512	R084		T	T	Osteocrust	Rattus rattus flavipectus	China, Sezchuan, Kao Ku
	AMNH	26854	R085	NO			Osteocrust	Rattus rattus tistae	India
*	AMNH	245114	R086		C	R1	Osteocrust	Rattus rattus wroughtoni	India
	AMNH	26851	R087	MISS			Osteocrust	Rattus rattus tistae	India
*	AMNH	114379	R088		A	R	Osteocrust	Rattus rattus rattus	Kenya
	AMNH	114382	R089	MISS			Osteocrust	Rattus rattus rattus	Kenya
	AMNH	114378	R090	MISS			Osteocrust	Rattus rattus rattus	Kenya
*	AMNH	114026	R091			R1	Osteocrust	Rattus rattus rattus	Kenya
*	AMNH	114027	R092			R1	Osteocrust	Rattus rattus rattus	Kenya
*	AMNH	81338	R093		A	R	Osteocrust	Rattus rattus rattus	Tanzania
*	AMNH	161358	R094		A	R	Osteocrust	Rattus rattus rattus	Zambia
*	AMNH	161359	R095		A	R	Osteocrust	Rattus rattus rattus	Zambia
*	AMNH	162122	R096		A	R	Osteocrust	Rattus rattus rattus	Malawi
*	AMNH	162121	R097		A	R	Osteocrust	Rattus rattus rattus	Malawi
*	AMNH	54956	R098		T	T	Osteocrust	Rattus rattus tikos	Burma
*	AMNH	54955	R099		T	T	Osteocrust	Rattus rattus tikos	Burma
*	AMNH	54954	R100		T	T	Osteocrust	Rattus rattus tikos	Burma
*	AMNH	240874	R101		K	K	Osteocrust	Rattus sp.	Sri Lanka
*	AMNH	240878	R102		K	K	Osteocrust	Rattus sp.	Sri Lanka
*	AMNH	163153	R103		A	R	Osteocrust	Rattus sp.	India
*	AMNH	163142	R104		C	R1	Osteocrust	Rattus rattus rufescens	India
	AMNH	152965	R105	MISS			Osteocrust	Rattus rattus sapoensis	Indonesia
*	AMNH	240871	R106		K	K	Osteocrust	Rattus sp.	Sri Lanka
	AMNH	242124	R107	NO			Osteocrust	Rattus rattus mindanensis	Philippines

*	NHM	50.47	R240		C	R1	Osteocrust	Rattus rattus rattus	Ethiopia
*	NHM	72.371	R241		A1	R	Osteocrust	Rattus rattus	Ethiopia
	NHM	77.3261	R242	MISS			Osteocrust	Rattus rattus	Sudan
*	NHM	33.6.14.4	R243		A	R	Osteocrust	Rattus rattus	Tanzania
	NHM	55.34	R244	MISS			Osteocrust	Rattus rattus frugivorius	Tanzania
	NHM	34.1.11.48	R245	MISS			Osteocrust	Rattus rattus alexandrinus	Mozambique
*	NHM	34.1.11.44	R246		A	R	Osteocrust	Rattus rattus alexandrinus	Mozambique
	NHM	83.524	R247	NO			Osteocrust	Rattus rattus	Rodrigues Islands
	NHM	69.1243	R248	NO			Osteocrust	Rattus rattus	Mauritius
*	NHM	69.1244	R249		C	R1	Osteocrust	Rattus rattus	Mauritius
	NHM	69.1245	R250	NO			Osteocrust	Rattus rattus	Mauritius
*	NHM	36.4.14.58	R251		A	R	Osteocrust	Rattus rattus	Persian Gulf
	NHM	36.4.14.59	R252	MISS			Osteocrust	Rattus rattus	Persian Gulf
	NHM	74.379	R253	NO			Osteocrust	Rattus rattus frugivorius	Bahrain
*	NHM	99.3.14.11	R254		B	R3	Osteocrust	Rattus rattus	Socotra
*	NHM	99.3.14.6	R255		B	R3	Osteocrust	Rattus	Socotra
*	NHM	99.3.14.5	R256		B	R3	Osteocrust	Rattus	Socotra
	NHM	77.3084	R257	NO			Osteocrust	Rattus rattus	Iran
*	NHM	99.3.14.9	R258		B	R3	Osteocrust	Rattus rattus flavipectus	Socotra
*	NHM	99.3.14.8	R259		B	R3	Osteocrust	Rattus rattus rattus	Socotra
	AMNH	77.3084	R260	NO			Osteocrust	Rattus rattus	Iran
	NHM	83.524	R261	NO			Osteocrust	Rattus rattus	Rodrigues Islands
	NHM	34.1.11.48	R262	NO			Osteocrust	Rattus rattus alexandrinus	Mozambique
*	AMNH	203220	R270		B	R3	Skin	Rattus rattus rattus	Egypt
	AMNH	163152	R271	NO			Skin	Rattus rattus wroughtoni	India
	AMNH	251685	R272	MISS			Skin	Rattus rattus	Bangladesh
	AMNH	150166	R273	NO			Skin	Rattus rattus rattus	Australia
	AMNH	150170	R274	NO			Skin	Rattus rattus rattus	Australia
	AMNH	194154	R275	MISS			Skin	Rattus alexandrinus	Australia
	AMNH	197385	R276	MISS			Skin	Rattus rattus	Australia
	AMNH	197396	R277	MISS			Skin	Rattus rattus	Australia
	AMNH	197681	R278	MISS			Skin	Rattus rattus rattus	Australia
*	AMNH	215597	R279		T	T	Skin	Rattus rattus rufescena	India

*	AMNH	215598	R280		T	T	Skin	Rattus rattus rufescena	India
*	AMNH	215599	R281		T	T	Skin	Rattus rattus rufescena	India
*	AMNH	215604	R282		T	T	Skin	Rattus sp.	India
	AMNH	114376	R283	NO			Skin	Rattus rattus rattus	Kenya
	AMNH	27882	R284	NO			Skin	Rattus rattus kijabius	Kenya
*	AMNH	187532	R285		C	R1	Skin	Rattus rattus rattus	Kenya
	AMNH	187674	R286	NO			Skin	Rattus rattus rattus	Kenya
*	AMNH	187591	R287		C	R1	Skin	Rattus rattus rattus	Kenya
*	AMNH	184680	R288		III	III	Skin	Rattus rattus rattus	Pakistan
*	AMNH	170239	R289		III	III	Skin	Rattus rattus rattus	Pakistan
	AMNH	277480	R290	MISS			Skin	Rattus rattus	Uganda
	AMNH	161360	R291	NO			Skin	Rattus rattus rattus	Zambia
*	HME	Sealinks01	R292		C	R1	Tail	Rattus	Zanzibar

Appendix 4.II Haplotype identifications for the Cyt*b* + D-loop tree (Figure 4.3b). n= number of sequences in each haplotype but for our sequences only. Our sequence identifications are three numbers prefixed by R, HE or AT and correspond to Appendix 4.I. GenBank accession numbers followed by information from the source publications are given. Published sequences are: GQ and FJ= (Tollenaere *et al.* 2010); JF= (Russell *et al.* 2011b); HQ= (Russell *et al.* 2011a); EU=(Robins *et al.* 2008).

Hap ID	n	Sequence ID
Rat1	1	R241EthiopiaDidessa
Rat2	1	R266R033MaliSotuba
Rat3	1	R231AndamanBrotherIsland
Rat4	1	R229AndamanNarcondam
Rat5	1	R100BurmaMergus
Rat6	1	R239Sudan
Rat7	1	R219EthiopiaAddisAbbaba
Rat8	1	R251PersianGulfTaubIsland
Rat9	1	R228IndiaHathiban
Rat10	1	R256Socotra
Rat11	1	R243Mafia
Rat12	1	R258Socotra
Rat13	1	R259Socotra
Rat14	1	R206YemenSana
Rat15	1	R254Socotra
Rat16	1	R225EgyptFaiyum
Rat17	1	R240EthiopiaBatie
Rat18	1	R255Socotra
Rat19	1	R190CongoAlbertsville
Rat20	1	HE014Madagascar
Rat21	1	R246MozambiqueTete
Rat22	1	R201ChinaYunnan
Rat23	1	R026EgyptQena
Rat24	1	R114Uganda
Rat25	1	R269R047JordanHemmah
Rat26	1	R050MadagascarMiandrivaza
Rat27	1	R048TurkeySabunCayl
Rat28	1	R052MadagascarAmbohimahavelona
Rat29	1	R088KenyaVoi
Rat30	1	R030KenyaKitero
Rat31	1	R249Mauritius

Rat32	1	R207YemenSana
Rat33	1	R215LebanonAkkar
Rat34	1	R227EgyptAsyut
Rat35	1	R040TanzaniaMorogoro
Rat36	1	R226EthiopiaAddisAbbaba
Rat37	1	R002IndiaNilgiriHills
Rat38	1	R024EgyptFaiyum
Rat39	1	R127IndonesiaJava
Rat40	1	R234MaldivesMale
Rat41	1	R292TanzaniaZanzibar
Rat42	1	R146LiberiaGanta
Rat43	1	R083ChinaHainan
Rat44	1	R018IndiaManipur
Rat45	1	R289Pakistan
Rat46	1	R288Pakistan
Rat47	1	R281India
Rat48	1	R056IndiaBengalMangpu
Rat49	1	R268R046JordanHemmah
Rat50	1	R264R037YemenSana
Rat51	1	R001IndiaNilgiriHills
Rat52	1	R130CongoKalongi
Rat53	1	R232MaldivesMale
Rat54	1	R081ChinaHainan
Rat55	1	R170CocosSelmaIsland
Rat56	2	R124IndonesiaJava
		R125IndonesiaJava
Rat57	1	R184IndonesiaSulawesi
Rat58	1	R122IndonesiaJava
Rat59	1	R123IndonesiaBali
Rat60	1	R183IndonesiaSulawesi]
Rat61	1	R126IndonesiaJava]
Rat62	1	R180IndonesiaBorneoKalimantan

Rat63	1	NC_012374
Rat64	1	R042AfghanistanJalalabad
Rat65	1	R108IndiaSonaripur
Rat66	1	R063Nepal
Rat67	1	R051MadagascarAmbohimahavelona
Rat68	1	R263R006MadagascarAmbantondrazaka
Rat69	1	R003IndiaNilgiriHills
Rat70	1	R008IndiaAssamManipur
Rat71	1	R086IndiaTamilNadu
Rat72	1	R174SriLankaUva
Rat73	1	R287Kenya
Rat74	1	R155SriLankaSabaragamuwa
Rat75	1	R230AndamanBrotherIsland
Rat76	1	R102SriLankaCentralProvince
Rat77	1	R173SriLanka
Rat78	1	R071SriLankaGonapola
Rat79	1	R101SriLankaCentralProvinceHortonPlains
Rat80	1	R069CeylonDehiwala
Rat81	1	R106SriLankaCentralProvince
Rat82	1	R237MaldivesAdduAtollGanIsland
Rat83	1	R285Kenya
Rat84	1	HE018TanzaniaMafia
Rat85	1	R049TurkeyAdanaHaruniye
Rat86	1	HE024Madagascar
Rat87	1	HE016TanzaniaPemba
Rat88	1	R038IndiaBenhope
Rat89	1	R265R066EthiopiaAddisAbbaba
Rat90	1	R032KenyaNjoro
Rat91	1	R270Egypt
Rat92	1	R080TanzaniaTanga
Rat93	1	R171Madagascar
Rat94	1	HE020Madagascar

Rat95	1	HE013Madagascar
Rat96	1	HE005TanzaniaZanzibar
Rat97	1	HE027Madagascar
Rat98	4	R075TanzaniaArushaTengeru
		R057TanzaniaTanganyikaLoljoro
		R136UgandaKampala
		R036TanzaniaArushaTengeru
Rat99	1	R104IndiaMadrasNilgiriHillsKalhatti
Rat100	1	HE010Madagascar
Rat101	1	R202IndiaMysore
Rat102	1	R203IndiaMysore
Rat103	1	R129TanzaniaKilossa
Rat104	1	R095ZambiaLakeChiyaua
Rat105	1	HE022Madagascar
Rat106	1	HE028Madagascar
Rat107	1	R077Zimbabwe
Rat108	1	HE026Madagascar
Rat109	1	HE007Madagascar
Rat110	1	HE011Madagascar
Rat111	12	R187MalawiKotaKota
		R096MalawiKasungu
		R094ZambiaLakeChiyaua
		R093TanzaniaRungwe
		R172MadagascarMahajanga
		R097MalawiKasungu
		HE006Madagascar
		R188MalawiZomba
		HE001TanzaniaMafia
		HE008Madagascar
		HE002TanzaniaMafia
		HE025Madagascar
Rat112	1	R004TanzaniaKilossa

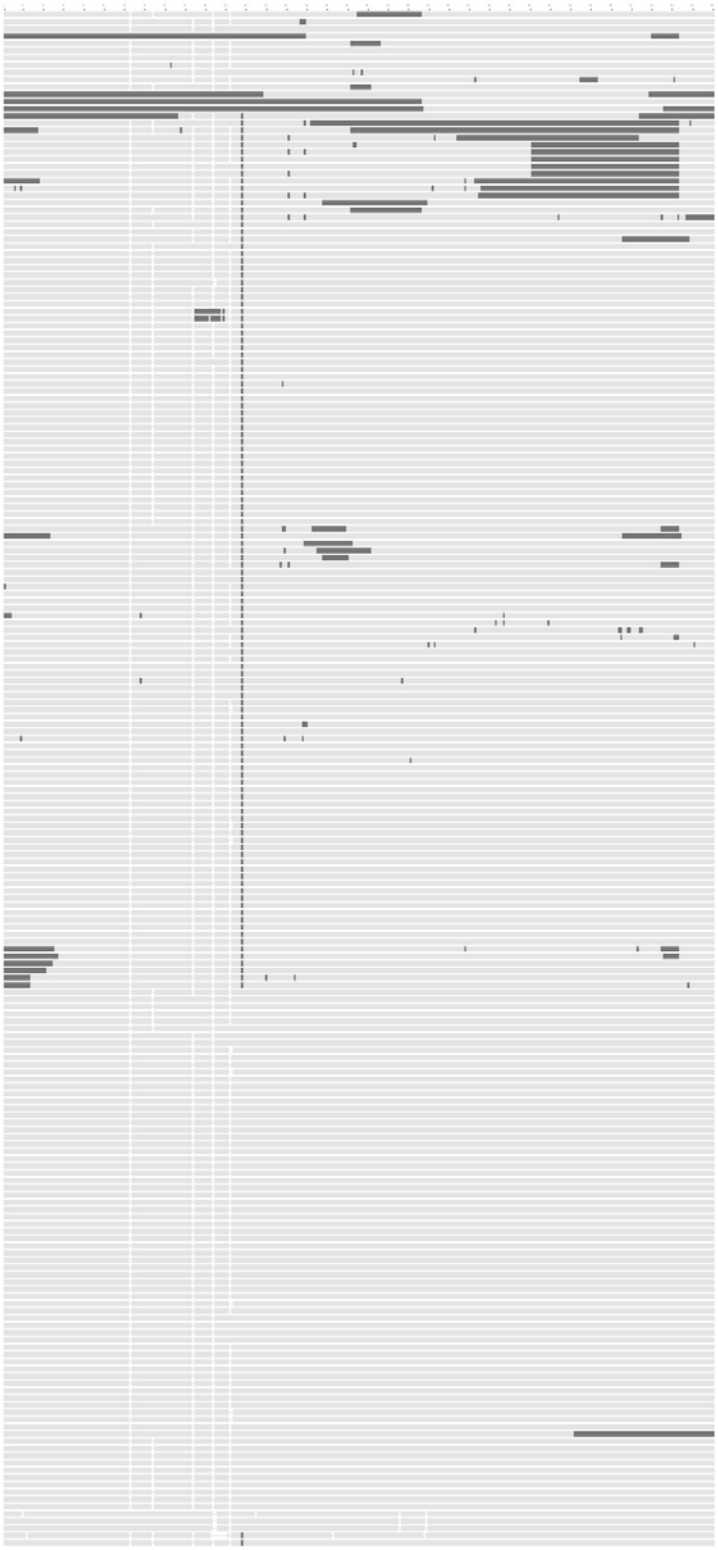
Rat113	1	R192TanzaniaKilossa
Rat114	1	R021EgyptAswan
Rat115	1	R031KenyaNjoro
Rat116	1	R193TanzaniaKilossa
Rat117	1	R068EgyptAswan
Rat118	1	R025EgyptGiza
Rat119	1	R103IndiaNilgiriHills
Rat120	1	R236MaldivesAdduAtollGanIsland
Rat121	1	R059IndiaAlapalli
Rat122	2	R028IranGorgan
		R204IranGorgan
Rat123	2	R133AngolaHuambo
		R134Angola
Rat124	1	R137AngolaHuambo
Rat125	1	R191AngolaLobitoBay
Rat126	1	R034SaudiArabiaHillatMuhaish
Rat127	3	HE004TanzaniaPemba
		HE003TanzaniaPemba
		HE023TanzaniaPemba
Rat128	2	R198UAEYemenMukala
		R197YemenMukala
Rat129	1	R067EgyptCairo
Rat130	1	R196CongoDolisie
Rat131	1	HE015TanzaniaPemba
Rat132	1	R181IranDarKaleh
Rat133	1	R182IranDarKaleh
Rat134	1	R280India
Rat135	1	R279India
Rat136	1	R235MaldivesMululayIsland
Rat137	1	R282India
Rat138	1	R045IndiaBengal
Rat139	1	R142ChinaSzechuan

Rat140	1	R079SiamKamPang
Rat141	1	R012LaosPhongSaly
Rat142	3	R131China
		R140ChinaSzechuan
		R141ChinaSzechuan
Rat143	1	R084ChinaSezchuan
Rat144	1	R113ChinaSzechuan
Rat145	1	R143ChinaSzechuan
Rat146	1	R076SiamKamPang
Rat147	1	R098BurmaMergus
Rat148	1	R208VietnamTonkin
Rat149	1	R072VietnamTonkin
Rat150	1	R120IndonesiaBali
Rat151	1	R233MaldivesMale
Rat152	1	R099BurmaMergus
Rat153	1	AT006DiegoGarcia
Rat154	1	AT002DiegoGarcia
Rat155	1	AT021DiegoGarcia
Rat156	18	AT026DiegoGarcia
		AT027DiegoGarcia
		AT003DiegoGarcia
		AT028DiegoGarcia
		AT004DiegoGarcia
		AT005DiegoGarcia
		AT010DiegoGarcia
		AT011DiegoGarcia
		AT009DiegoGarcia
		AT012DiegoGarcia
		AT013DiegoGarcia
		AT016DiegoGarcia
		AT014DiegoGarcia
		AT017DiegoGarcia

		AT018DiegoGarcia
		AT019DiegoGarcia
		AT020DiegoGarcia
		AT001DiegoGarcia
Rat157	1	R082ChinaYunnan
Rat158	1	R116BurmaGova
Rat159	1	R061IndiaManipur
Rat160	1	R119BurmaTawmaw
Rat161	1	GQ891581 Rattus rattus haplotype 13 Yemen
Rat162	1	GQ891582 Rattus rattus haplotype 14 Yemen
Rat163	1	GQ891574 Rattus rattus haplotype 6 Oman Arazat
Rat164	1	GQ891573 Rattus rattus haplotype 5 India Attur
Rat165	1	GQ891572 Rattus rattus haplotype 4 India Avallanchi
Rat166	1	GQ891594 Rattus rattus haplotype 26 Grand Comore
Rat167	1	GQ891591 Rattus rattus haplotype 23 Grand Comore
Rat168	1	GQ891596 Rattus rattus haplotype 28 Grand Comore
Rat169	1	GQ891592 Rattus rattus haplotype 24 Grand Comore
Rat170	1	GQ891587 Rattus rattus haplotype 19 Tanzania Lushoto
Rat171	1	GQ891590 Rattus rattus haplotype 22 Mozambique Maputo
Rat172	1	GQ891595 Rattus rattus haplotype 27 Grand Comore
Rat173	1	GQ891586 Rattus rattus haplotype 18 Tanzania, Mozambique, Grand
Rat174	1	GQ891593 Rattus rattus haplotype 25 Grand Comore
Rat175	1	GQ891585 Rattus rattus haplotype 17 Tanzania Lushoto & Morogoro
Rat176	1	GQ891607 Rattus rattus haplotype 39 Reunion
Rat177	1	GQ891599 Rattus rattus haplotype 31 Madagascar Madiomiangana
Rat178	1	JF718278 Rattus rattus haplotype Hap43 Europa Isl
Rat179	1	JF718279 Rattus rattus haplotype Hap44 Europa Isl
Rat180	1	FJ897498 Rattus rattus Oussouye Senegal
Rat181	1	HQ588111 Rattus rattus JRussell-73
Rat182	2	GQ891608 Rattus rattus haplotype 40 South Africa Cape Town
		FJ897499 Rattus rattus Keur Daouda Sarr Senegal
Rat183	1	FJ897501 Rattus rattus isolate BLP Guadalupe

Rat184	1	FJ897500 Rattus rattus isolate GEF Guadalupe
Rat185	1	GQ891575 Rattus rattus haplotype 7 Oman Tibraq
Rat186	1	GQ891580 Rattus rattus haplotype 12 Oman Sahanout
Rat187	1	GQ891571 Rattus rattus haplotype 3 India Mudumalai
Rat188	1	GQ891570 Rattus rattus haplotype 2 India Mudumalai
Rat189	1	GQ891569 Rattus rattus haplotype 1 India Mudumalai
Rat190	1	GQ891576 Rattus rattus haplotype 8 Oman Tibraq
Rat191	1	GQ891583 Rattus rattus haplotype 15 Ethiopia
Rat192	1	GQ891577 Rattus rattus haplotype 9 Oman Sahanout
Rat193	1	GQ891578 Rattus rattus haplotype 10 Oman Sahanout
Rat194	1	GQ891603 Rattus rattus haplotype 35 Madagascar Ambalatenona
Rat195	1	GQ891579 Rattus rattus haplotype 11 Oman Sahanout
Rat196	1	JF718277 Rattus rattus haplotype Hap42 Glorieuse Isl
Rat197	1	GQ891604 Rattus rattus haplotype 36 Madagascar Ambohimiariana
Rat198	1	GQ891601 Rattus rattus haplotype 33 Madagascar Brickaville
Rat199	1	GQ891605 Rattus rattus haplotype 37 Madagascar Ambohimiariana
Rat200	1	JF718276 Rattus rattus haplotype Hap41 Juan de Nova Isl
Rat201	1	GQ891584 Rattus rattus haplotype 16 Ethiopia
Rat202	1	GQ891589 Rattus rattus haplotype 21 Mozambique Zambezi
Rat203	1	GQ891597 Rattus rattus haplotype 29 Mayotte
Rat204	1	GQ891598 Rattus rattus haplotype 30 Mayotte
Rat205	1	GQ891600 Rattus rattus haplotype 32 Madagascar Brickaville
Rat206	1	GQ891602 Rattus rattus haplotype 34 Madagascar Ambalatenona
Rat207	1	GQ891606 Rattus rattus haplotype 38 Madagascar Ambohimiariana
Rat208	1	GQ891588 Rattus rattus haplotype 20 Mozambique, Mayotte, Madagascar
Rat209	1	JQ927552 Niviventer excelsior Outgroup
Rat210	1	EU273707 Rattus rattus RNZRrTit01 New Zealand
Rat211	1	EU273712 Rattus tanezumi RJPNAna02 9Oct06

Appendix 5.I Schematic of the FASTA alignment of the D-loop for *Rattus* showing missing nucleotides in dark grey, and gaps in white.



Appendix 6.I Samples collected for this study. Supplier= American Museum of Natural History, New York (AMNH); The Field Museum of Natural History, Chicago (FMNH); Smithsonian Institution National Museum of Natural History, Washington D.C. (NMNH); Natural History Museum, London (NHM); Michel Pascal (Pascal); and Jean-Marc Duplantier (Duplantier). Reason= the reason why data was not used in this study: analysed by AT (AT); majority of missing nucleotides (MISS); no amplification of DNA (NO AMP); and species not *M. musculus* (NOT MUS). Acc= accuracy of the geographic co-ordinates (as in Appendix 3.I).

Sample ID	Supplier	Type	Date	Reason	Taxon ID	mtDNA HG	CAS clade	DOM clade
85873	AMNH	osteocrust	3-May-25	MISS	Tapecomys primus Anderson & Yates, 2000	NA		
85874	AMNH	osteocrust	3-May-25	MISS	Tapecomys primus Anderson & Yates, 2000	NA		
* 85876	AMNH	osteocrust	6-May-24		Tapecomys primus Anderson & Yates, 2000	DOM		D1
* 85877	AMNH	osteocrust	6-May-24		Tapecomys primus Anderson & Yates, 2000	DOM		
162641	AMNH	osteocrust	8-Jan-48	AT	Mus musculus	NA		
162639	AMNH	osteocrust	20-Feb-48	AT	Mus musculus	NA		
162640	AMNH	osteocrust	20-Feb-48	AT	Mus musculus	NA		
197468	AMNH	osteocrust	11-Mar-64	AT	Mus musculus	NA		
150156	AMNH	skin	6-Feb-41	AT	Mus musculus	NA		
150157	AMNH	skin	7-Feb-41	AT	Mus musculus	NA		
162643	AMNH	osteocrust	13-Jan-48	AT	Mus musculus	NA		
55797	AMNH	osteocrust		AT	Mus musculus homourus	NA		
56416	AMNH	osteocrust	14-Nov-21	AT	Mus musculus homourus	NA		
43579	AMNH	osteocrust	27-Oct-16	AT	Mus musculus	NA		
32292	AMNH	osteocrust	28-Nov-10	AT	Mus musculus	NA		
32295	AMNH	osteocrust	7-Dec-10	AT	Mus musculus	NA		
32293	AMNH	osteocrust	22-Nov-10	AT	Mus musculus	NA		
84303	AMNH	osteocrust	14-Sep-25	AT	Mus musculus	NA		
56400	AMNH	osteocrust	11-Dec-20	AT	Mus musculus	NA		
47955	AMNH	osteocrust	16-Jan-20	AT	Mus musculus	NA		
47959	AMNH	osteocrust	7-Feb-20	AT	Mus musculus	NA		
47961	AMNH	osteocrust	9-Feb-20	AT	Mus musculus	NA		
47962	AMNH	osteocrust	19-Feb-20	AT	Mus musculus	NA		
47963	AMNH	osteocrust	20-Feb-20	AT	Mus musculus	NA		
84975	AMNH	osteocrust	7-Dec-26	AT	Mus musculus bactrianus	NA		
* 203205	AMNH	skin	15-Jan-62		Tapecomys primus Anderson & Yates, 2000	DOM		C
* 203206	AMNH	skin	16-Jan-62		Tapecomys primus Anderson & Yates, 2000	DOM		C
120250	AMNH	skin		MISS	Potamogale velox	NA		
* 171127	AMNH	skin	25-Dec-47		Chimarrogale platycephala himalayica	CAS	2	
* 26857	AMNH	osteocrust	2-Aug-05		Mus musculus	CAS	1	
* 163164	AMNH	osteocrust	23-Apr-47		Mus sp.	NEP		
215634	AMNH	osteocrust	8-Mar-65	NO AMP	Mus musculus	NA		
215638	AMNH	osteocrust	10-Mar-65	NO AMP	Mus musculus	NA		
* 208173	AMNH	skin	21-Jan-61		Mus musculus	CAS	1	

*	208174	AMNH	skin	1-Jan-62		Mus musculus	CAS	1
*	208175	AMNH	skin	21-Jan-62		Mus musculus	CAS	1
*	171161	AMNH	osteocrust	14-Apr-49		Rattus nitidus nitidus	CAS	1
	106661	AMNH	osteocrust	27-Jun-36	AT	Mus musculus castaneus	NA	
	153041	AMNH	osteocrust	24-Sep-39	AT	Mus sp.	NA	
	153042	AMNH	osteocrust	25-Sep-39	AT	Mus sp.	NA	
	101528	AMNH	osteocrust	6-May-31	AT	Mus musculus castaneus	NA	
	102693	AMNH	osteocrust	30-Oct-34	AT	Mus musculus	NA	
	106698	AMNH	osteocrust	18-Mar-36	AT	Mus musculus castaneus	NA	
	109164	AMNH	osteocrust	10-Jul-38	AT	Rattus sp.	NA	
	109166	AMNH	osteocrust	20-Jul-38	AT	Rattus sp.	NA	
	109169	AMNH	osteocrust	21-Jul-38	AT	Rattus sp.	NA	
	109172	AMNH	osteocrust	22-Jul-38	AT	Rattus sp.	NA	
	109173	AMNH	osteocrust	13-Jul-38	AT	Rattus exulans	NA	
	109174	AMNH	osteocrust	13-Jul-38	AT	Rattus exulans	NA	
	109175	AMNH	osteocrust	14-Jul-38	AT	Rattus exulans	NA	
	242307	AMNH	osteocrust	29-Jan-73	AT	Mus musculus	NA	
	101268	AMNH	osteocrust	21-Jan-31	AT	Mus sp.	NA	
	101269	AMNH	osteocrust	20-Jan-31	AT	Mus sp.	NA	
	54858	AMNH	osteocrust		AT	Mus musculus	NA	
	267709	AMNH	skin	4-Aug-93	AT	Mus musculus	NA	
	267710	AMNH	skin	4-Aug-93	AT	Mus musculus	NA	
	88895	AMNH	osteocrust	23-Nov-38	MISS	Mus musculus wagneri	NA	
	171206	AMNH	skin	25-Nov-42	MISS	Crocidura russula gueldenstaedti	NA	
	171209	AMNH	skin	28-Nov-42	NO AMP	Crocidura russula gueldenstaedti	NA	
*	88825	AMNH	osteocrust	20-Nov-38		Mus musculus wagneri	CAS	2
	88959	AMNH	osteocrust		NO AMP	Mus musculus wagneri	NA	
*	88784	AMNH	osteocrust	24-Oct-38		Mus musculus wagneri	MUS	
*	88785	AMNH	osteocrust	24-Oct-38		Mus musculus wagneri	MUS	
*	88787	AMNH	osteocrust	29-Oct-38		Mus musculus wagneri	MUS	
*	88911	AMNH	osteocrust	16-Oct-38		Mus musculus kaleh-peninsularis	CAS	2
*	88912	AMNH	osteocrust	16-Oct-38		Mus musculus kaleh-peninsularis	CAS	2
	88913	AMNH	osteocrust	16-Oct-38	NO AMP	Mus musculus kaleh-peninsularis	NA	

	88914	AMNH	osteocrust	16-Oct-38	NO AMP	Mus musculus wagneri	NA		
*	88915	AMNH	osteocrust	16-Oct-38		Mus musculus kaleh-peninsularis	CAS	2	
*	88916	AMNH	osteocrust	16-Oct-38		Mus musculus kaleh-peninsularis	CAS	2	
*	88788	AMNH	osteocrust	30-Oct-38		Mus musculus wagneri	CAS	2	
*	28623	AMNH	osteocrust	7-Apr-08		Mus musculus bactrianus	CAS	3	
*	28641	AMNH	osteocrust	25-Mar-08		Mus musculus homourus	CAS	3	
	28644	AMNH	osteocrust	21-Mar-08	MISS	Mus musculus homourus	NA		
*	28614	AMNH	osteocrust	21-May-06		Apodemus pallipes	CAS	3	
*	85338	AMNH	osteocrust	16-Oct-29		Tapecomys primus Anderson & Yates, 2000	MUS		
*	85347	AMNH	osteocrust	27-Oct-29		Tapecomys primus Anderson & Yates, 2000	MUS		
*	85334	AMNH	osteocrust	11-Oct-29		Tapecomys primus Anderson & Yates, 2000	MUS		
*	85351	AMNH	osteocrust	1-Nov-29		Tapecomys primus Anderson & Yates, 2000	MUS		
*	100728	AMNH	osteocrust	3-May-31		Mus musculus	GEN		
	84095	AMNH	osteocrust	14-Jun-25	AT	Mus musculus wagneri	NA		
*	165341	AMNH	osteocrust	13-Jan-53		Mus sp.	DOM		F
*	168584	AMNH	osteocrust	21-May-53		Micaelamys namaquensis	DOM		E
	uncat	AMNH	skin		NO AMP	uncat	NA		
	250011	AMNH	skin	8-Dec-73	NO AMP	Mus musculus castaneous	NA		
	250014	AMNH	skin	31-Aug-74	NO AMP	Mus musculus urbanus	NA		
	250016	AMNH	skin	18-Sep-74	NO AMP	Mus musculus urbanus	NA		
	240750	AMNH	osteocrust/skin	24-Feb-72	NO AMP	Mus musculus homourus	NA		
	240753	AMNH	skin		NO AMP	Mus musculus homourus	NA		
	240751	AMNH	skin	11-Mar-72	NO AMP	Mus musculus homourus	NA		
	240752	AMNH	skin	11-Mar-72	NO AMP	Mus musculus homourus	NA		
	242326	AMNH	osteocrust	3-Dec-72	NO AMP	Mus musculus	NA		
	170225	AMNH	skin	9-Sep-50	NO AMP	Rodentia	NA		
	170247	AMNH	skin	2-Sep-50	NO AMP	Rattus rattus rattus	NA		
*	170248	AMNH	skin	3-Sep-50		Rattus rattus rattus	CAS	2	
	170249	AMNH	skin	4-Sep-50	NO AMP	Rattus rattus rattus	NA		
	170273	AMNH	skin	21-Oct-50	MISS	Tapecomys primus Anderson & Yates, 2000	NA		
	242217	AMNH	osteocrust		AT	Mus musculus castaneus	NA		
	242223	AMNH	osteocrust		AT	Mus musculus castaneus	NA		
	242224	AMNH	osteocrust		AT	Mus musculus castaneus	NA		

	242213	AMNH	osteocrust		AT	Mus musculus castaneus	NA		
	242232	AMNH	osteocrust		AT	Mus musculus castaneus	NA		
	166950	AMNH	skin	31-May-54	MISS	Mus musculus bactrianus	DOM?		
*	166947	AMNH	osteocrust	1-Jan-54		Mus musculus bactrianus	DOM		
*	166948	AMNH	osteocrust	13-Dec-53		Mus musculus bactrianus	DOM		F^
*	166949	AMNH	osteocrust/skin	13-Dec-53		Mus musculus bactrianus	DOM		
	118204	AMNH	skin	24-Mar-38	NO AMP	Mus musculus	NA		
*	168708	AMNH	osteocrust	22-Jan-01		Mus sp.	DOM		E
	20813	AMNH	osteocrust	1-Aug-01	AT	Mus sp.	NA		
	242702	AMNH	skin	12-Aug-70	AT	Mus musculus castaneus	NA		
	242697	AMNH	skin		AT	Mus musculus castaneus	NA		
*	240959	AMNH	osteocrust/skin	15-Oct-71		Mus musculus castaneus	CAS	1	
	242703	AMNH	skin		AT	Mus musculus castaneus	NA		
	165784	AMNH	osteocrust	15-Oct-54	AT	Mus sp.	NA		
	176404	AMNH	osteocrust	18-Mar-57	NO AMP	Mus musculus brevirostris	NA		
*	256648	AMNH	skin	1-Jan-84		Mus domesticus praetextus	DOM		
*	256649	AMNH	skin	1-Jan-84		Mus domesticus praetextus	CAS	3	
	176246	AMNH	osteocrust	16-Sep-44	AT	Mus musculus wagneri	NA		
*	98115	AMNH	osteocrust			Mus musculus wagneri	MUS		
*	98116	AMNH	osteocrust			Mus musculus wagneri	MUS		
	42053	AMNH	skin	6-Feb-08	MISS	Mus musculus	NA		
*	42054	AMNH	skin	6-Feb-08		Mus musculus	DOM		E
*	T_49	Duplantie	tissue in alc				GEN		
*	103679	FMNH	osteocrust	1955		Mus musculus	CAS	2	
	120307	FMNH	osteocrust	1976	AT	Mus musculus domesticus	NA		
	120314	FMNH	osteocrust	1976	AT	Mus musculus domesticus	NA		
	120315	FMNH	osteocrust	1976	AT	Mus musculus domesticus	NA		
	120317	FMNH	osteocrust	1976	AT	Mus musculus domesticus	NA		
	120320	FMNH	osteocrust	1976	AT	Mus musculus domesticus	NA		
	120323	FMNH	osteocrust	1976	AT	Mus musculus domesticus	NA		
	40594	FMNH	osteocrust	1932	AT	Mus musculus tantillus	NA		
	33157	FMNH	osteocrust	1929	AT	Mus musculus castaneus	NA		
	33159	FMNH	osteocrust	1929	AT	Mus musculus castaneus	NA		

	33160	FMNH	osteocrust	1929	AT	Mus musculus castaneus	NA		
	40645	FMNH	osteocrust	1932	AT	Mus musculus tantillus	NA		
*	68258	FMNH	osteocrust	1950		Mus musculus	DOM		
*	98818	FMNH	osteocrust	1962		Mus musculus praetextus	DOM		C
*	100127	FMNH	osteocrust			Mus musculus	DOM		C
*	100173	FMNH	osteocrust			Mus musculus	DOM		C
	100633	FMNH	osteocrust		MISS	Mus musculus	NA		
	107030	FMNH	osteocrust	1972	NO AMP	Mus musculus praetextus	NA		
*	101261	FMNH	osteocrust			Mus musculus praetextus	DOM		F^
	100175	FMNH	osteocrust		NOT MUS(	Mus musculus praetextus	NA		
*	100623	FMNH	osteocrust			Mus musculus praetextus	DOM		A
*	202873	FMNH	osteocrust	2007		Mus musculus	DOM		
	203039	FMNH	osteocrust	2007	NO AMP	Mus musculus	NA		
	55232	FMNH	osteocrust	1945	AT	Muscicapa striata striata	NA		
*	65424	FMNH	osteocrust	1947		?	CAS	3	
	65425	FMNH	osteocrust	1947	NO AMP	?	NA		
*	65426	FMNH	osteocrust	1947		?	CAS	3	
*	34806	FMNH	osteocrust	1930		Mus musculus castaneus	CAS	1	
*	35571	FMNH	osteocrust	1931		Mus musculus castaneus	CAS	1	
*	35574	FMNH	osteocrust	1931		Mus musculus castaneus	CAS	1	
*	35577	FMNH	osteocrust	1931		Mus musculus castaneus	SIK		
*	35578	FMNH	osteocrust	1931		Mus musculus castaneus	SIK		
	28953	FMNH	osteocrust	1900	NO AMP	Mus musculus homourus	NA		
	35554	FMNH	osteocrust	1931	NO AMP	Mus musculus castaneus	NA		
	35556	FMNH	osteocrust	1931	NO AMP	Mus musculus castaneus	NA		
*	35565	FMNH	osteocrust	1931		Mus musculus castaneus	CAS	1	
*	97541	FMNH	osteocrust	1962		Mus musculus	DOM		B
*	97567	FMNH	osteocrust	1962		Mus musculus	MUS		
	97584	FMNH	osteocrust	1962	NO AMP	Mus musculus	NA		
	97589	FMNH	osteocrust	1962	NO AMP	Mus musculus	NA		
	112240	FMNH	osteocrust	1968	NOT MUS(	Mus	NA		
*	97593	FMNH	osteocrust	1963		Mus musculus praetextus	DOM		A
*	112270	FMNH	osteocrust	1968		Mus musculus praetextus	CAS	3	

	112251	FMNH	osteocrust	1968	NOT MUS(	Mus musculus praetextus	NA		
*	97587	FMNH	osteocrust	1962		Mus musculus praetextus	DOM		F^
	112255	FMNH	osteocrust	1968	NOT MUS(	Mus musculus praetextus	NA		
*	97595	FMNH	osteocrust	1963		Mus musculus praetextus	DOM		
*	112280	FMNH	osteocrust	1968		Mus musculus praetextus	CAS	3	
*	179178	FMNH	osteocrust	2001		Mus musculus	DOM		D
*	179181	FMNH	osteocrust	2001		Mus musculus	DOM		A
*	84459	FMNH	osteocrust	1955		Mus musculus	DOM		
*	161770	FMNH	tissue in alc	1998		Mus musculus	GEN		
*	165722	FMNH	tissue in alc	1998		Mus musculus	GEN		
*	165723	FMNH	tissue in alc	1998		Mus musculus	GEN		
*	175935	FMNH	tissue in alc	2002		Mus musculus	GEN		
*	173206	FMNH	tissue in alc	2002		Mus musculus	GEN		
*	172728	FMNH	tissue in alc	2001		Mus musculus	GEN		
	178678	FMNH	tissue in alc	2003	NO AMP	Mus musculus	NA		
*	156238	FMNH	tissue in alc	1995		Mus musculus	GEN		
	82044	FMNH	osteocrust	1954	NO AMP	Mus musculus	NA		
*	142196	FMNH	osteocrust	1968		Mus musculus homourus	NEP		
*	142198	FMNH	osteocrust	1968		Mus musculus homourus	NEP		
*	142200	FMNH	osteocrust	1968		Mus musculus homourus	SIK		
	55232	FMNH	osteocrust	1945	AT	Mus musculus	NA		
	55233	FMNH	osteocrust	1945	AT	Mus musculus	NA		
	55235	FMNH	osteocrust	1945	AT	Mus musculus	NA		
	140449	FMNH	osteocrust	1990	MISS	Mus musculus praetextus	DOM?		
	140450	FMNH	osteocrust	1990	NO AMP	Mus musculus praetextus	NA		
	140452	FMNH	osteocrust	1990	NO AMP	Mus musculus praetextus	NA		
*	140459	FMNH	osteocrust	1990		Mus musculus praetextus	CAS	3	
	140461	FMNH	osteocrust	1990	MISS	Mus musculus homourus	NA		
	140455	FMNH	osteocrust	1990	NO AMP	Mus musculus praetextus	NA		
*	140464	FMNH	osteocrust	1990		Mus musculus praetextus	CAS	2	
	56259	FMNH	osteocrust	1946	AT	Mus musculus castaneus	NA		
	125129	FMNH	osteocrust	1983	NO AMP	Mus musculus praetextus	NA		
*	99402	FMNH	osteocrust	1965		Mus musculus castaneus	CAS	1	
*	99403	FMNH	osteocrust	1965		Mus musculus castaneus	CAS	1	

	168098	FMNH	tissue in alc	2000	NOT MUS(	Mus musculus	NA		
	86360	FMNH	osteocrust	1956	NOT MUS(	Mus musculus	NA		
	82209	FMNH	osteocrust	1954	MISS	Mus musculus musculus	NA		
	82211	FMNH	osteocrust	1954	NO AMP	Mus musculus praetextus	NA		
	82214	FMNH	osteocrust	1954	MISS	Mus musculus musculus	NA		
	122319	FMNH	osteocrust	1953	NOT MUS(	Mus musculus praetextus	NA		
*	122324	FMNH	osteocrust	1953		Mus musculus praetextus	DOM		
	122328	FMNH	osteocrust	1953	NO AMP	Mus musculus praetextus	NA		
*	122442	FMNH	osteocrust	1955		Mus musculus domesticus	DOM		
	122460	FMNH	osteocrust	1955	MISS	Mus musculus domesticus	DOM?		
*	149873	FMNH	osteocrust	1990		Mus musculus	CAS	1	
	32426	FMNH	osteocrust	1929	AT	Mus musculus castaneus	NA		
*	57.417	NHM	osteocrust			Mus musculus castaneus	CAS	1	
*	18.7.4.23	NHM	osteocrust			Mus musculus gentilis	DOM		A
	47.1128	NHM	osteocrust		NO AMP	Mus musculus bactrianus	NA		
*	27.6.7.4	NHM	osteocrust			Mus	DOM		A
*	36.14.62	NHM	osteocrust			Mus musculus gentlis	DOM		u
*	54.1025	NHM	osteocrust			Mus musculus	CAS	3	
*	10.10.3.20	NHM	osteocrust			Mus musculus	GEN		
*	66.5837	NHM	osteocrust	27-Jan-64		Mus musculus	DOM		
	26.11.1.91	NHM	osteocrust		NO AMP	Mus musculus	NA		
	551475	NMNH	osteocrust	6/19/58	NO AMP	Mus musculus domesticus	NA		
	362205	NMNH	osteocrust	2/6/66	AT	Mus musculus castaneus	NA		
	552565	NMNH	osteocrust	11/11/79	NO AMP	Mus musculus castaneus	NA		
	277484	NMNH	osteocrust	3/10/45	AT	Mus musculus castaneus	NA		
	277486	NMNH	osteocrust	3/10/45	AT	Mus musculus castaneus	NA		
*	173932	NMNH	osteocrust	5/14/11		Mus musculus domesticus	CAS	4	
*	201188	NMNH	osteocrust	10/14/15		Mus musculus domesticus	CAS	3	
*	496973	NMNH	osteocrust			Mus musculus castaneus	CAS	1	
*	154878	NMNH	osteocrust	3/29/09		Mus musculus castaneus	DOM		E
	38316	NMNH	osteocrust		AT	Mus musculus castaneus	NA		
*	279703	NMNH	osteocrust	11/10/45		Mus musculus domesticus	DOM		
*	279702	NMNH	osteocrust	11/8/45		Mus musculus domesticus	DOM		
	538877	NMNH	osteocrust	11/17/60	AT	Mus musculus castaneus	NA		
*	551476	NMNH	osteocrust	3/21/67		Mus musculus domesticus	DOM		
*	326614	NMNH	osteocrust	9/22/62		Mus musculus domesticus	CAS	2	

*	326620	NMNH	osteocrust	9/22/62		Mus musculus domesticus	CAS	3	
	279125	NMNH	osteocrust	4/20/45	AT	Mus musculus castaneus	NA		
	318062	NMNH	osteocrust	3/11/48	MISS	Mus musculus domesticus	DOM?		
	562726	NMNH	osteocrust	1/7/81	AT	Mus musculus castaneus	NA		
*	MP10	Pascal	tissue	2006		Mus musculus	DOM		E
*	MP1	Pascal	tissue	2006		Mus musculus	DOM		E
*	MP2	Pascal	tissue	2006		Mus musculus	DOM		E
*	MP3	Pascal	tissue	2006		Mus musculus	DOM		E
*	MP4	Pascal	tissue	2006		Mus musculus	DOM		E
*	MP5	Pascal	tissue	2006		Mus musculus	DOM		E
*	MP6	Pascal	tissue	2006		Mus musculus	DOM		E
*	MP7	Pascal	tissue	2006		Mus musculus	DOM		E
*	MP8	Pascal	tissue	2006		Mus musculus	DOM		E
*	MP9	Pascal	tissue	2006		Mus musculus	DOM		E
*	MP11	Pascal	tissue	2006		Mus musculus	DOM		E
*	MP12	Pascal	tissue	2006		Mus musculus	DOM		E
*	MP13	Pascal	tissue	2006		Mus musculus	DOM		E
*	MP14	Pascal	tissue	2006		Mus musculus	DOM		E
*	MP15	Pascal	tissue	2006		Mus musculus	DOM		E
*	MP16	Pascal	tissue	2006		Mus musculus	DOM		E
*	MP35	Pascal	tissue	2009		Mus musculus castaneus?	DOM		E
*	MP36	Pascal	tissue	2009		Mus musculus castaneus?	DOM		E
*	MP37	Pascal	tissue	2009		Mus musculus castaneus?	DOM		E
*	MP38	Pascal	tissue	2009		Mus musculus castaneus?	DOM		E
*	MP39	Pascal	tissue	2009		Mus musculus castaneus?	DOM		E
*	MP40	Pascal	tissue	2009		Mus musculus castaneus?	DOM		E
*	MP41	Pascal	tissue	2009		Mus musculus castaneus?	DOM		E
*	MP42	Pascal	tissue	2009		Mus musculus castaneus?	DOM		E
*	MP43	Pascal	tissue	2009		Mus musculus castaneus?	DOM		E
*	MP44	Pascal	tissue	2009		Mus musculus castaneus?	DOM		E
*	MP17	Pascal	tissue	2007		Mus musculus	CAS	1	
	MP18	Pascal	tissue	2006	MISS	Mus musculus	NA		
*	MP19	Pascal	tissue	2006		Mus musculus	CAS	1	
*	MP20	Pascal	tissue	2006		Mus musculus	CAS	1	
	MP21	Pascal	tissue	2006	MISS	Mus musculus	NA		

*	MP22	Pascal	tissue	2006		Mus musculus	CAS	1	
	MP25	Pascal	tissue	2006	MISS	Mus musculus	NA		
	MP26	Pascal	tissue	2006	MISS	Mus musculus	NA		
*	MP27	Pascal	tissue	2006		Mus musculus	CAS	1	
*	MP23	Pascal	tissue	2007		Mus musculus	CAS	1	
*	MP24	Pascal	tissue	2007		Mus musculus	CAS	1	
*	MP45	Pascal	tissue	2009		Mus musculus castaneus?	DOM		C1
*	MP28	Pascal	tissue	2007		Mus musculus	CAS	1	
*	MP29	Pascal	tissue	2007		Mus musculus	CAS	1	
*	MP30	Pascal	tissue	2007		Mus musculus	DOM		E
*	MP31	Pascal	tissue	2007		Mus musculus	DOM		E
*	MP32	Pascal	tissue	2007		Mus musculus	DOM		E
*	MP46	Pascal	tissue	2009		Mus musculus castaneus?	DOM		E
*	MP47	Pascal	tissue	2009		Mus musculus castaneus?	DOM		E
*	MP48	Pascal	tissue	2009		Mus musculus castaneus?	DOM		E
*	MP33	Pascal	tissue	2007		Mus musculus	CAS	1	
*	MP34	Pascal	tissue	2007		Mus musculus	DOM		E

Sample ID	Country	Locality	Lat	Long	Acc
85873	Angola	Lobito Bay, Lobito, Benguela	-12.37	13.53	2
85874	Angola	Lobito Bay, Lobito, Benguela	-12.37	13.53	2
85876	Angola	Lobito Bay, Lobito, Benguela	-12.37	13.53	2
85877	Angola	Lobito Bay, Lobito, Benguela	-12.37	13.53	2
162641	Australia	(Western Line), Angelalla, Queensland	-26.40	147.00	2
162639	Australia	Cromarty, south foot of Mt. Elliot, Townsville, Queensland	-19.30	146.80	2
162640	Australia	Cromarty, south foot of Mt. Elliot, Townsville, Queensland	-19.30	146.80	2
197468	Australia	Gidley Brook Western Australia			
150156	Australia	Home Island, Cocos (Keeling) Isls	-12.09	96.91	2
150157	Australia	Home Island, Cocos (Keeling) Isls	14.79	96.91	2
162643	Australia	Warrego River, Goolburra Hole Queensland	-34.60	149.90	2
55797	China				
56416	China				
43579	China	Lijiang, Yunnan Province	26.88	100.23	2
32292	China	45 mi So. of Fengsiangfu, Shensi			
32295	China	45 mi So. of Fengsiangfu, Shensi			
32293	China	45 mi So. of Fengsiangfu, Shensi			
84303	China	Mountains 10 miles S.W. of Choni			
56400	China	Mts. near Yenping, Fukien Prov Yenping, Fukien Prov.	26.64	118.18	2
47955	China	Yenping, Fujian Province	26.64	118.18	2
47959	China	Yenping, Fujian Province	26.64	118.18	2
47961	China	Yenping, Fujian Province	26.64	118.18	2
47962	China	Yenping, Fujian Province	26.64	118.18	2
47963	China	Yenping, Fujian Province	26.64	118.18	2
84975	China	Yunnan Kiang, Yunnan			
203205	Egypt	Abu Ghalib, Imbaba, Giza	30.28	30.94	2
203206	Egypt	Abu Ghalib, Imbaba, Giza	30.28	30.94	2
120250	Gabon	Kango Estuaire Reg.	0.17	10.11	2
171127	India	Amritsan, Punjab	31.64	74.87	2
26857	India	Assam, Tura, Meghalaya State	25.47	91.37	2
163164	India	Gokarna Nepal	27.73	85.40	2
215634	India	Howrah (# 6 Mullick Ghat Rd.), Calcutta, West Bengal	22.60	88.26	2
215638	India	Howrah (# 6 Mullick Ghat Rd.), Calcutta, West Bengal	22.60	88.26	2
208173	India	Nr, Bombay City, Maharashtra	19.09	72.88	2

208174	India	Nr, Bombay City, Maharashtra	19.09	72.88	2
208175	India	Nr, Bombay City, Maharashtra	19.09	72.88	2
171161	India	Umran, Khasia Hills, Assam	25.77	91.88	2
106661	Indonesia	Cheribon, Java	-6.72	108.57	2
153041	Indonesia	Bumbulan, Sulawesi	0.49	122.07	2
153042	Indonesia	Bumbulan, Sulawesi	0.49	122.07	2
101528	Indonesia	Cheribon Java			
102693	Indonesia	Cheribon Java			
106698	Indonesia	Cheribon, Java	-6.72	108.57	2
109164	Indonesia	Peleng I Sulawesi	-1.31	123.03	2
109166	Indonesia	Peleng I Sulawesi	-1.31	123.03	2
109169	Indonesia	Peleng I Sulawesi	-1.31	123.03	2
109172	Indonesia	Peleng I Sulawesi	-1.31	123.03	2
109173	Indonesia	Peleng I Sulawesi	-1.31	123.03	2
109174	Indonesia	Peleng I Sulawesi	-1.31	123.03	2
109175	Indonesia	Peleng I Sulawesi	-1.31	123.03	2
242307	Indonesia	Pewunu, Gumbasa Valley, Sulawesi			
101268	Indonesia	Roeroekan Sulawesi			
101269	Indonesia	Roeroekan Sulawesi			
54858	Indonesia	Sea level Komodo Islands	-8.58	119.49	2
267709	Indonesia	Wamena, Irian Jaya	-4.00	139.00	2
267710	Indonesia	Wamena, Irian Jaya	-4.00	139.00	2
88895	Iran	Dasht, Bujnurd Dist.	37.31	56.01	2
171206	Iran	Bruyino, Luristan	33.59	48.40	3
171209	Iran	Bruyino, Luristan	33.59	48.40	3
88825	Iran	Dasht, Bujnurd Dist.	37.31	56.01	2
88959	Iran	Dasht, Bujnurd Dist.	37.31	56.01	2
88784	Iran	Mazandaran, Astarabad, Dar Kaleh	36.51	53.87	3
88785	Iran	Mazandaran, Astarabad, Dar Kaleh	36.51	53.87	3
88787	Iran	Mazandaran, Astarabad, Dar Kaleh	36.51	53.87	3
88911	Iran	Mazandaran, Astarabad, Dar Kaleh	36.51	53.87	3
88912	Iran	Mazandaran, Astarabad, Dar Kaleh	36.51	53.87	3
88913	Iran	Mazandaran, Astarabad, Dar Kaleh	36.51	53.87	3

88914	Iran	Mazandaran, Astarabad, Dar Kaleh	36.51	53.87	3
88915	Iran	Mazandaran, Astarabad, Dar Kaleh	36.51	53.87	3
88916	Iran	Mazandaran, Astarabad, Dar Kaleh	36.51	53.87	3
88788	Iran	Turkman Desert, Astarabad, Dar Kaleh	36.51	53.87	3
28623	Kashmir	Dachgam	34.13	75.02	2
28641	Kashmir	Pandrittan	34.06	74.86	2
28644	Kashmir	Pandrittan	34.06	74.86	2
28614	Kashmir	Sadak, Dras	34.44	75.80	2
85338	Kazakhstan	150 mi. N. of Kizil Orda, Perovsk Tuz Bulak	46.70	65.60	2
85347	Kazakhstan	160 mi. N. of Kizil Orda, Perovsk Tuz Bulak	46.65	65.60	2
85334	Kazakhstan	75 mi. N.E. of Kizil Orda, Perovsk Telekul	45.46	67.08	2
85351	Kazakhstan	75 mi. N.E. of Perovsk Telekul	45.46	67.08	2
100728	Madagascar	Tabiky, west of Ank	-22.15	44.26	2
84095	Mongolia	Tsagan Nor			
165341	Namibia	Henties Bay Swakopmund	-22.67	14.53	2
168584	Namibia	Paresis Mts.	-20.37	16.27	2
uncat	Nepal				
250011	Nepal	Hetaura	27.42	85.04	2
250014	Nepal	Kathmandu	27.70	85.33	2
250016	Nepal	Kathmandu	27.70	85.33	2
240750	Nepal	Zom Khola			
240753	Nepal	Zom Khola			
240751	Nepal	Lapche Zom Khola			
240752	Nepal	Lapche Zom Khola			
242326	Pakistan	None, Chitral Gol, Chitral	35.84	71.78	2
170225	Pakistan	Quetta, Balochistan	30.21	67.02	2
170247	Pakistan	Quetta, Balochistan	30.21	67.02	2
170248	Pakistan	Quetta, Balochistan	30.21	67.02	2
170249	Pakistan	Quetta, Balochistan	30.21	67.02	2
170273	Pakistan	Quetta, Balochistan	30.21	67.02	2
242217	Philippines	Los Banos, Rizal Prov Luzon	14.17	121.24	2
242223	Philippines	Los Banos, Rizal Prov Luzon	14.17	121.24	2
242224	Philippines	Los Banos, Rizal Prov Luzon	14.17	121.24	2

242213	Philippines	None, Manila, Luzon	14.60	120.98	2
242232	Philippines	None, None, None			
166950	Saudi Arabia		24.80	44.30	4
166947	Saudi Arabia	Darwish Bldg. Dammam	26.39	49.98	2
166948	Saudi Arabia	Dhahran	26.29	50.12	2
166949	Saudi Arabia	Dhahran	26.29	50.12	2
118204	South Africa	King Williams Town, Pirie Forest,	-32.69	27.29	2
168708	South Africa	Lilly Fontein, Namaqualand, Northern Cape	-30.29	18.08	2
20813	Sri Lanka				
242702	Sri Lanka	Talawakele, Central Prov	6.94	80.66	2
242697	Sri Lanka	Boragas, Central Prov			
240959	Sri Lanka	Nanuoya Central Province	6.94	80.74	2
242703	Sri Lanka	Talawakele, Central Prov	6.94	80.66	2
165784	Taiwan	Taipei			
176404	Tunisia	Haouaria	37.06	11.01	2
256648	UAE	Mukala City, South Yemen	14.54	49.13	2
256649	UAE	Mukala City, South Yemen	14.54	49.13	2
176246	USA	Hartford			
98115	Uzbekistan	Tashkent	41.27	69.22	2
98116	Uzbekistan	Tashkent	41.27	69.22	2
42053	Zimbabwe	Bulawayo	-20.16	28.58	2
42054	Zimbabwe	Bulawayo	-20.16	28.58	2
T_49	Madagascar				
103679	Afghan		34.89	68.17	2
120307	Australia	Kimberly 154 km NW Brooking Springs Homestead Western Australia	-17.13	124.60	1
120314	Australia	Kimberly 154 km NW Brooking Springs Homestead Western Australia	-17.13	124.60	1
120315	Australia	Kimberly 154 km NW Brooking Springs Homestead Western Australia	-17.13	124.60	1
120317	Australia	Kimberly 154 km NW Brooking Springs Homestead Western Australia	-17.13	124.60	1
120320	Australia	Kimberly 154 km NW Brooking Springs Homestead Western Australia	-17.13	124.60	1
120323	Australia	Kimberly 154 km NW Brooking Springs Homestead Western Australia	-17.13	124.60	1
40594	China	Lu Chang Pu 14 mi E Kao Ku Wu R, Fuling Pref, Sichuan	29.42	108.30	1
33157	China	Nguluko, Lijiang Pref, Yunnan	27.05	100.20	1
33159	China	Nguluko, Lijiang Pref, Yunnan	27.05	100.20	1

33160	China	Nguluko, Lijiang Pref, Yunnan	27.05	100.20	1
40645	China	Pien Ngai Sichuan			
68258	Djibouti		11.15	42.71	2
98818	Egypt		24.47	32.94	2
100127	Egypt		29.21	25.52	2
100173	Egypt		29.97	32.55	2
100633	Egypt		31.27	32.30	2
107030	Egypt		24.60	27.30	2
101261	Egypt	Dakhla Oasis Mut 3 mi N Wadi Gedeed	25.50	28.98	1
100175	Egypt	Ras Zafarana Wadi Araba Bir Zafarana Red Sea	29.12	32.55	1
100623	Egypt	Wadi Abu Seyala Suez	29.53	32.37	1
202873	Eq Guinea		3.36	8.66	2
203039	Eq Guinea		3.36	8.66	2
55232	Germany	Augsburg Bavaria			
65424	India		30.45	78.07	2
65425	India		30.45	78.07	2
65426	India		30.45	78.07	2
34806	India	Haldibari, Bhutan Duars, West Bengal	26.33	88.77	1
35571	India	Lingtam Sikkim	27.23	88.72	1
35574	India	Lingtam Sikkim	27.23	88.72	1
35577	India	Lingtam Sikkim	27.23	88.72	1
35578	India	Lingtam Sikkim	27.23	88.72	1
28953	India	Pahlgam, Ladakh Dist, Kashmir	34.03	75.33	1
35554	India	Sangsir, Bhutan Duars, West Bengal	27.12	88.53	1
35556	India	Tista Valley Tarkhola Sikkim			
35565	India	Tista Valley Tarkhola Sikkim	?	?	?
97541	Iran		37.56	45.07	2
97567	Iran		36.84	54.44	2
97584	Iran		26.23	60.22	2
97589	Iran		33.60	51.60	4
112240	Iran		33.65	46.44	2
97593	Iran	Ahram Bushehr	28.90	51.29	1
112270	Iran	Lordegan nr; Sar Dasht Chahar Mahal and Bakhtiari	31.51	50.83	1

112251	Iran	Maku 21 mi E West Azarbaijan	39.28	44.91	1
97587	Iran	Pol-i-Abgineh 5 km ESE Fars	29.53	51.81	1
112255	Iran	Rezaiyeh 5.8 mi SW West Azarbaijan	37.49	44.99	1
97595	Iran	Shush 12 mi S Khuzistan	32.01	48.25	1
112280	Iran	Yasuj 25 mi E Fars	30.67	52.00	1
179178	Jordan	Birket al Ara'is Irbid	32.55	35.86	3
179181	Jordan	Hemmah Irbid	32.55	35.86	3
84459	Lebanon		33.78	35.55	2
161770	Madagascar	28 km SSW Ambositra, 5 km SW Ambalamanakana, Foret d'Ankazomivady Prov de Fianarantsoa	-20.78	47.17	1
165722	Madagascar	Parc National d'Andringitra, 8.5 km SE Antanifotsy Province de Fianarantsoa	-22.17	46.95	1
165723	Madagascar	Parc National d'Andringitra, 8.5 km SE Antanifotsy Province de Fianarantsoa	-22.17	46.95	1
175935	Madagascar	Parc National de l'Isalo, along Sahanafa River, 28 km SE Berenty-Betsileo Province de Fianarantsoa	-22.32	45.29	1
173206	Madagascar	Parc National de Tsimanampetsotsa, 6.5 km NE Efoetse, near Mitoho Cave Province de Toliara	-24.05	43.75	1
172728	Madagascar	PN de Bemaraha, 3.5 km E Bekopaka Province de Toliara	-19.14	44.83	1
178678	Madagascar	PN de Midongy-Sud, NE slope Mt Papango, 3.5 km SW Befotaka Province de Fianarantsoa	-23.84	46.96	1
156238	Madagascar	Reserve Naturelle Integrale d'Andringitra, Plateau d'Andohariana Province de Fianarantsoa	-22.19	46.90	1
82044	Namibia	Kariros, Keetmanshoop, Karas	-26.57	18.13	2
142196	Nepal	Khumjung Solukhumbu	27.82	86.73	1
142198	Nepal	Khumjung Solukhumbu	27.82	86.73	1
142200	Nepal	Lukla Airport Solukhumbu	27.68	86.73	1
55232	New Caledonia	Noumea	-22.27	166.45	1
55233	New Caledonia	Noumea	-22.27	166.45	1
55235	New Caledonia	Noumea	-22.27	166.45	1
140449	Pakistan	Awaran, Khuzdar Dist, Baluchistan Prov	26.45	65.23	1
140450	Pakistan	Awaran, Khuzdar Dist, Baluchistan Prov	26.45	65.23	1
140452	Pakistan	Gwadar, Makran Dist, Baluchistan Prov	25.12	62.32	1
140459	Pakistan	Karakar Pass 1 km SE; Jaba, Swat Dist, Malakand Div	34.43	72.22	1
140461	Pakistan	Karakar Pass 1 km SE; Jaba, Swat Dist, Malakand Div	34.43	72.22	1
140455	Pakistan	Karakar Pass, Swat Dist, Malakand Div	34.43	72.22	1
140464	Pakistan	Yakh Tangai, Swat Dist, Malakand Div	34.92	72.63	1
56259	Philippines	Mt McKinley east slope, Davao City Prov, Mindanao I	7.10	125.30	1
125129	Saudi Arabia	Jeddah King Abdelaziz Univ Herbarium	21.48	39.19	1
99402	Sri Lanka	Dehiwala sea level Western Prov	6.85	79.87	1
99403	Sri Lanka	Dehiwala sea level Western Prov	6.85	79.87	1

168098	Tanzania	Kibebe Farms, approx 6 km ESE Iringa, Iringa District, Iringa Region,	35.77	-7.80	1
86360	Tanzania	Tengeru, Arusha District, Arusha Region	-3.38	36.80	1
82209	Turkey		36.51	53.87	2
82211	Turkey		36.51	53.87	2
82214	Turkey		38.50	43.41	2
122319	Turkey	Haruniye Subdist, Bahce Dist, Adana Prov	37.28	36.45	1
122324	Turkey	Haruniye Subdist, Bahce Dist, Adana Prov	37.28	36.45	1
122328	Turkey	Haruniye Subdist, Bahce Dist, Adana Prov	37.28	36.45	1
122442	Turkey	Savastepe Dist, Balikesir Prov	39.37	27.67	1
122460	Turkey	Savastepe Dist, Balikesir Prov	39.37	27.67	1
149873	Uganda	Kamwokya, Kampala Dist, Central Buganda	0.33	32.58	1
32426	Viet Nam	Lai Chau, Cac ba I, Tonkin	22.07	103.17	1
57.417	Maldives	Male island, North male	4.18	73.50	2
18.7.4.23	Mesopotamia	Amara	31.83	47.15	2
47.1128	Persian Gulf	Gach Qara Goli	26.00	52.00	4
27.6.7.4	Persian Gulf	Mishum	26.00	52.00	4
36.14.62	Persian Gulf	Naibu:Taub Island	26.00	52.00	4
54.1025	Socotra	Nankad Plain	12.35	53.94	2
10.10.3.20	Somalia	Bulhar, Somaliland	10.42	44.42	2
66.5837	Sudan	Khartoum	15.57	32.53	2
26.11.1.91	Sudan	Luluabourg	-5.89	22.40	2
551475	?		?	?	
362205	Baker Island				
552565	Comoros	Wanani Comoros	-12.33	47.33	2
277484	Guam	Sinajanja			
277486	Guam	Sinajanja			
173932	India	Dachin, Khistwar Jammu And Kashmir			
201188	India	Dandwar, Pir Panjal Region Jammu And Kashmir			
496973	Java	Djakarta, Indonesia			
154878	Java	Buitenzorg, Indonesia			
38316	Kauai	Kauai Island, Kauai County, Hawaiian Islands			
279703	Kuwait	Al Kuwait, 8 Mi S,	29.11	47.99	2
279702	Kuwait	Khaitan, 8 Mi S Al Kuwait,	29.11	47.99	2
538877	North Pacific Oce	Johnston Atoll, Sand Island,			
551476	Oman	Sohar, Batinah,	24.35	56.72	2
326614	Pakistan	Dungagali, Hazara Division, North-West Frontier	34.05	73.30	2

326620	Pakistan	Dungagali, Hazara Division, North-West Frontier	34.05	73.30	2
279125	Saipan Island	Mariana Islands			
318062	Sudan	Abu Hamed, Berber, An-Nil Province, Ash-Shamaliyah	19.54	33.32	2
562726	Tinian Island	Mariana Islands			
MP10	Réunion	Cilaos	-21.13	55.48	1
MP1	Réunion	Cilaos	-21.11	55.45	1
MP2	Réunion	Cilaos	-21.11	55.45	1
MP3	Réunion	Cilaos	-21.11	55.45	1
MP4	Réunion	Cilaos	-21.11	55.45	1
MP5	Réunion	Cilaos	-21.11	55.45	1
MP6	Réunion	Cilaos	-21.13	55.47	1
MP7	Réunion	Cilaos	-21.13	55.47	1
MP8	Réunion	Cilaos	-21.13	55.48	1
MP9	Réunion	Cilaos	-21.13	55.48	1
MP11	Réunion	Saint André	-20.92	55.63	1
MP12	Réunion	Saint André	-20.97	55.70	1
MP13	Réunion	Saint André	-20.97	55.70	1
MP14	Réunion	Saint André	-20.97	55.70	1
MP15	Réunion	Saint André	-20.97	55.70	1
MP16	Réunion	Saint André	-20.97	55.70	1
MP35	Réunion	Saint Benoit, Takamaka (Mr Dennemont)	-21.06	55.65	1
MP36	Réunion	Saint Benoit, Takamaka (Mr Dennemont)	-21.06	55.65	1
MP37	Réunion	Saint Benoit, Takamaka (Mr Dennemont)	-21.06	55.65	1
MP38	Réunion	Saint Benoit, Takamaka (Mr Dennemont)	-21.06	55.65	1
MP39	Réunion	Saint Benoit, Takamaka (Mr Dennemont)	-21.06	55.65	1
MP40	Réunion	Saint Benoit, Takamaka (Mr Dennemont)	-21.06	55.65	1
MP41	Réunion	Saint Benoit, Takamaka (Mr Dennemont)	-21.06	55.65	1
MP42	Réunion	Saint Benoit, Takamaka (Mr Dennemont)	-21.06	55.65	1
MP43	Réunion	Saint Benoit, Takamaka (Mr Dennemont)	-21.06	55.65	1
MP44	Réunion	Saint Benoit, Takamaka (Mr Dennemont)	-21.06	55.65	1
MP17	Réunion	Saint-Gilles les Hts	-21.07	55.26	1
MP18	Réunion	Saint-Louis	-21.23	55.46	1
MP19	Réunion	Saint-Louis	-21.23	55.46	1
MP20	Réunion	Saint-Louis	-21.24	55.42	1
MP21	Réunion	Saint-Louis	-21.24	55.42	1

MP22	Réunion	Saint-Louis	-21.26	55.41	1
MP25	Réunion	Saint-Louis	-21.27	55.43	1
MP26	Réunion	Saint-Louis	-21.27	55.43	1
MP27	Réunion	Saint-Louis	-21.27	55.43	1
MP23	Réunion	Saint-Louis, Etang du Gol	-21.28	55.38	1
MP24	Réunion	Saint-Louis, Etang du Gol	-21.28	55.38	1
MP45	Réunion	Saint-Louis, Etang du Gol	-21.28	55.38	1
MP28	Réunion	Saint-Paul, Etang	-21.01	55.28	1
MP29	Réunion	Saint-Paul, Etang	-21.01	55.28	1
MP30	Réunion	Saint-Paul, Etang	-21.01	55.28	1
MP31	Réunion	Saint-Paul, Etang	-21.01	55.28	1
MP32	Réunion	Saint-Paul, Etang	-21.01	55.28	1
MP46	Réunion	Saint-Paul, Etang Saint-Paul	-21.01	55.28	1
MP47	Réunion	Saint-Paul, Etang Saint-Paul	-21.01	55.28	1
MP48	Réunion	Saint-Paul, Etang Saint-Paul	-21.01	55.28	1
MP33	Réunion	Saint-Pierre, Montvert	-21.33	55.54	1
MP34	Réunion	Sainte-Suzanne, Rivière du Mât	-20.98	55.68	1

Appendix 6.II The provenance of mouse bones collected from archaeological excavations.

Bones were not analysed for aDNA analyses.

Sample ID	Site	Country	Date	PI	Taxon ID	Element	Analyst
Mouse01	Abu Sha'ar	Egypt, Red Sea	4th c. AD	W. Van Neer	*Mus musculus domesticus*	mandible left	W. Van Neer/ T. Cucchi
Mouse02	Abu Sha'ar	Egypt, Red Sea	4th c. AD	W. Van Neer	*Mus musculus domesticus*	mandible left	W. Van Neer/ T. Cucchi
Mouse03	Kantharodai	Sri Lanka		Sealinks	*Mus* sp.	cervical vertebra	J. Perera
Mouse04	Kantharodai	Sri Lanka		Sealinks	*Mus* sp.	radius	J. Perera
Mouse05	Manthai	Sri Lanka		D. Fuller	?*Mus* sp.	mandible	H. Eager
Mouse06	Manthai	Sri Lanka		D. Fuller	?*Mus* sp.	scapula	H. Eager
Mouse07	Manthai	Sri Lanka		Sealinks	*Mus* sp.	lumbar vertebra	J. Perera
Mouse08	Paithan	India		D. Kennet	?*Mus* sp.	femur	H. Eager
Mouse09	Paithan	India		D. Kennet	?*Mus* sp.	mandible	H. Eager
Mouse10	Paithan	India		D. Kennet	?*Mus* sp.	incisor	H. Eager
Mouse11	Paithan	India		D. Kennet	?*Mus* sp.	caudal vertebra	H. Eager
Mouse12	Paithan	India		D. Kennet	?*Mus* sp.	caudal vertebra	H. Eager
Mouse13	Paithan	India		D. Kennet	?*Mus* sp.	fragment	H. Eager
Mouse14	Panga Ya Saidi	Kenya		R. Helm	?*Mus* sp.	mandible	H. Eager

Mouse15	Sanganakallu	India		D. Fuller	?*Mus* sp.	caudal vertebra	H. Eager
Mouse16	Sanganakallu	India		D. Fuller	?*Mus* sp.	caudal vertebra	H. Eager
Mouse17	Sanganakallu	India		D. Fuller	?*Mus* sp.	fragment	H. Eager
Mouse18	Shenhur	Egypt, Nile	5th-7th c. AD	W. Van Neer	*Mus musculus domesticus/Acomys cahirinus*	mandible right	W. Van Neer/ T. Cucchi
Mouse19	Shenhur	Egypt, Nile	5th-7th c. AD	W. Van Neer	*Mus musculus domesticus/Acomys cahirinus*	mandible right	W. Van Neer/ T. Cucchi
Mouse20	Shenhur	Egypt, Nile	5th-7th c. AD	W. Van Neer	*Mus musculus domesticus/Acomys cahirinus*	mandible left	W. Van Neer/ T. Cucchi
Mouse21	Shenhur	Egypt, Nile	5th-7th c. AD	W. Van Neer	*Mus musculus domesticus/Acomys cahirinus*	mandible left	W. Van Neer/ T. Cucchi
Mouse22	Shenhur	Egypt, Nile	5th-7th c. AD	W. Van Neer	*Mus musculus domesticus*	mandible left	W. Van Neer/ T. Cucchi
Mouse23	Shenhur	Egypt, Nile	5th-7th c. AD	W. Van Neer	*Mus musculus domesticus*	mandible left	W. Van Neer/ T. Cucchi
Mouse24	Shenhur	Egypt, Nile	5th-7th c. AD	W. Van Neer	*Mus musculus domesticus*	mandible right	W. Van Neer/ T. Cucchi
Mouse25	Shenhur	Egypt, Nile	5th-7th c. AD	W. Van Neer	*Mus musculus domesticus*	mandible left	W. Van Neer/ T. Cucchi
Mouse26	Shenhur	Egypt, Nile	5th-7th c. AD	W. Van Neer	*Mus musculus domesticus*	mandible right	W. Van Neer/ T. Cucchi
Mouse27	Shenhur	Egypt, Nile	5th-7th c. AD	W. Van Neer	*Mus musculus domesticus*	mandible right	W. Van Neer/ T. Cucchi
Mouse28	Shenhur	Egypt, Nile	5th-7th c. AD	W. Van Neer	*Mus musculus domesticus*	mandible left	W. Van Neer/ T. Cucchi
Mouse29	Shenhur	Egypt, Nile	5th-7th c. AD	W. Van Neer	*Mus musculus domesticus*	mandible right	W. Van Neer/ T. Cucchi

Mouse30	Shenhur	Egypt, Nile	5th-7th c. AD	W. Van Neer	*Mus musculus domesticus*	mandible left	W. Van Neer/ T. Cucchi
Mouse31	Shenhur	Egypt, Nile	5th-7th c. AD	W. Van Neer	*Mus musculus domesticus*	mandible left	W. Van Neer/ T. Cucchi
Mouse32	Shenhur	Egypt, Nile	5th-7th c. AD	W. Van Neer	*Mus musculus domesticus*	mandible left	W. Van Neer/ T. Cucchi
Mouse33	Shenhur	Egypt, Nile	5th-7th c. AD	W. Van Neer	*Mus musculus domesticus*	mandible left	W. Van Neer/ T. Cucchi
Mouse34	Shenhur	Egypt, Nile	5th-7th c. AD	W. Van Neer	*Mus musculus domesticus*	mandible right	W. Van Neer/ T. Cucchi
Mouse35	Shenhur	Egypt, Nile	5th-7th c. AD	W. Van Neer	*Mus musculus domesticus*	mandible right	W. Van Neer/ T. Cucchi

Appendix 7.I Genotyping specimens using MUGA/MegaMUGA

We aimed to genotype house mice sampled from museum skins and osteocrusts using genome-wide, high coverage, SNP data generated through a commercially available bead array called the Mouse Universal Genotyping Array (MUGA), implemented commercially by Geneseek (www.neogen.com) (Yang *et al.* 2012). The first generation of the array had probes for 7,851 single nucleotide polymorphisms (SNPs), which then increased ten-fold with the second generation of the array (called MegaMUGA) to over 77,800 SNPs. On MegaMUGA, SNPs are assessed at an average of 33 Kb apart across the whole genome, including mitochondria and the Y chromosome. MUGA was initially designed around markers developed for the Collaborative Cross Project, which took eight founder strains of the three major lineages (*musculus*, *castaneus* and *domesticus*) to create recombinant-inbred lines (Collaborative Cross Consortium 2012). With the success of the array, markers were extended to include wild-type mice. In addition, MegaMUGA includes probes that detect structural variants such as insertions, deletions, duplications, etc.

In collaboration with Fernando Pardo Manuel de Villena at University of North Carolina, we sent two samples (on two occasions) for genotyping on the MUGA and MegaMUGA platforms. Initial trials of museum mice DNA extracts on the array failed or were borderline (high no-call rates; N:H ratios deviating from expected; samples did not cluster with known others) (see Figures 1 and 2; data supplied for the MUGA trial by John Didion at UNC and analysed by HME). The trialed samples had relatively high molecular weight DNA, adequate quantities of DNA (albeit partially non-endogenous), and a 600 bp fragment of mitochondrial DNA was amplifiable, all of which indicated good quality museum samples and thus good candidates for the trial. The array requires high quality “intact” DNA over 5 kb - degraded historical samples inherently violate this assumption. Therefore, a DNA enhancement technique that randomly ligates small fragments to mimic intact DNA was undertaken, following the protocol of Richard Croojimans (pers. comm.) Briefly, DNA was blunt-end repaired using T4 DNA polymerase and Klenow, followed by a kinase/ligase treatment. 2 µl of treated DNA was run on a 1.5% agarose gel in parallel with untreated DNA. If a shift in size was detected, 15 µl of sample DNA was sent for genotyping to Geneseek (three samples). There is a waiting period of several months for the array to be run and for results to be returned, and therefore the timeframe in which this analysis is expected to completed does not fit in the scope of my PhD. If the DNA enhancement technique successfully produces valid SNP array data, we intend to genotype 96 museum samples and continue this research in the future.

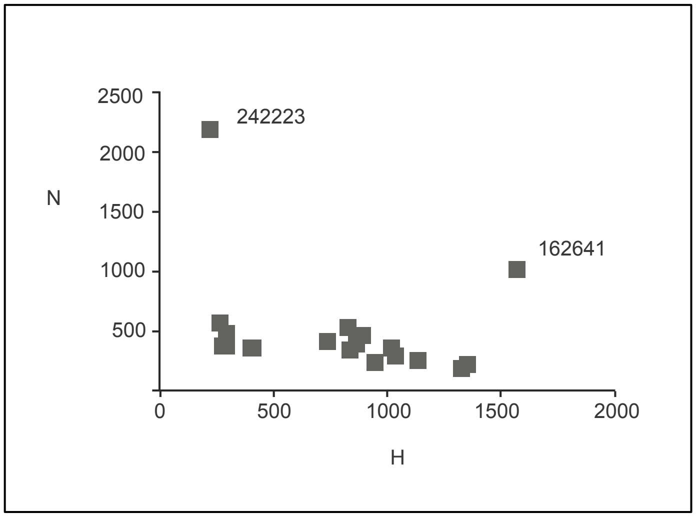


Figure 7.I Plot of number of No-calls (N) to number of heterozygotes (H) for MUGA (7,810 SNPs), showing our two samples (242223 and 162641) as clear outliers.

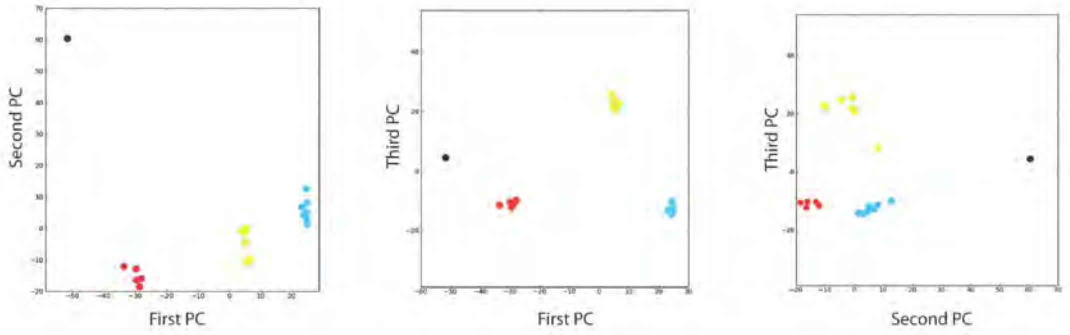


Figure 7.II. Principal component analysis (PCA) of our sample 162641 (black), and Collaborative Cross *domesticus* (red), *castaneus* (yellow), and *musculus* (blue). Our sample was a borderline pass in terms of data quality controls but is a clear outlier within the PCA, and therefore we expect inaccurate base calling has occurred.

REFERENCES

Collaborative Cross Consortium, Iraqi F, Mahajne M, Salaymah A, Sandovsky H *et al.* (2012) The genome architecture of the Collaborative Cross mouse genetic reference population. *Genetics,* **190**, 389–402.

Yang H, Ding Y, Hutchins L, Szatkiewicz J *et al.* (2009) A customized and versatile high-density genotyping array for the mouse. *Nature Methods,* ***6***, 663–666.