Reconstructing patterns of migration and translocation of different animal taxa across the Indian Ocean and Island South-East Asia

TRINKS, ALEXANDRA,MARIA

Reconstructing patterns of migration and translocation of different animal taxa across the Indian Ocean and Island South-East Asia

Alexandra Maria Trinks

Thesis submitted for the degree of Doctor of Philosophy

Department of Archaeology
Durham University

Department of Archaeology
University of Aberdeen

2014

Abstract

The Indian Ocean represents one of the oldest exchange networks connecting South-East-Asia with India, the Arabian peninsula, as far as Africa in the West. Since the beginning of the Common Era, extensive trade between geographically distant and culturally diverse people enabled the transmission of not only new technologies, exotic goods and food items, but also diverse plant and animal species. Although archaeological remains, particularly from the 1st millennium AD, reflect an intensification of maritime connectivity across the Indian Ocean, the exact routes of travel and trade across this vast area in early times are still subject to discussion.

This thesis presents different projects that aim to assess the potential of using commensal animals, such as the house mouse *Mus musculus*, the black rat *Rattus rattus*, and the Asian house gecko *Hemidactylus frenatus*, as proxies to infer pathways of human travel and trade. Commensal species are usually small animals, that live in close association with humans and opportunistically exploit their habitat and food sources. Utilisation of these new resources has led to a close relationship between humans and certain species, and thus favoured their global distribution due to translocations through humans.

Therefore, genetic analyses from modern and museum samples of the species in question have been employed, and embedded in a phylogeographic approach. This integrative methodology connects genealogy and geography, with the aim to reconstruct evolutionary, demographic and biogeographic processes that led to the contemporary distribution of genetic lineages of the commensal species and subsequently mirrors travel routes of the humans who carried them. The incorporation of ancient DNA analysis provides a powerful method, not only enabling the detection of source populations, but direct monitoring of their genetic change through time. Given that people have moved them around for a long time, undirected distribution pattern of populations were expected for each species. However, the results demonstrate that several unique and geographically restricted lineages have been identified, reflecting past human-mediated translocation throughout the Indian and Pacific Ocean from the 1st millennium AD onwards.

Furthermore, a methodological exercise incorporating latest high-throughput sequencing techniques has been carried out. Next-generation-sequencing has been applied in order to generate the first mitogenome from museum samples of the Asian dwarf buffalo (*Bubalus mindorensis*) and assess its position with the taxonomy of the bovid family.

Phylogeographic studies are generally limited to a specific time period by the age of the samples used. However, recent developments in the field of aDNA have widened the range of applications from single genes to whole genomes, which subsequently extended the range of questions that can be addressed. Refined extraction and sequencing techniques allow the analysis of ever decreasing amounts of DNA, enabling the use of aDNA methods to understand phylogenetic relationships over space and time. Here, results demonstrate that museum samples provide a powerful opportunity to obtain information about the genetic diversity of species that are difficult to sample in the wild, like the endangered Asian dwarf buffalo, or where archaeological remains are difficult to retrieve during the excavation process, as with small commensal animals.

Declaration

The work in this thesis is based on research carried out at the Durham Evolution and Ancient DNA (DEAD) laboratory at the Department of Archaeology, Durham University, UK. This thesis is by publication, and its results are presented in form of individual manuscripts. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is entirely my own work unless referenced to the contrary in the text.

Statement of Copyright

Acknowledgments

First and foremost I would like to thank my principal supervisor Dr. Greger Larson for providing me with the opportunity to work on such diverse and fascinating projects and for his constant support, encouragement and patience throughout the last years. Moreover, I would like to thank my co-supervisor Prof. Keith Dobney for the organisation of the data collection and his guidance during sampling trips, as well as Prof. Jeremy Searle for welcoming me in his lab for a research stay, and for his assistance and insightful comments.

Durham University and University of Aberdeen generously provided financial support, not only for my research and travel costs, but coverage of my tuition fees and living expenses.

A special mention deserve the numerous museums and institutions which donated samples for the various research projects: the Field Museum (Chicago), the American Natural History Museum (New York), the Smithsonian (Washington DC), the British Natural History Museum (London), the French National Institute for Agricultural Research (Rennes) and the Johann Wolfgang von Goethe-University (Frankfurt), further Prof. Atholl Anderson (Australian National University) as well as Dr. J. Chris Hillman and Solomon Pomerantz for collecting specimens in the field.

I thank Dr. Hannes Schröder, Dr. Maria Avila, Dr. Laurent Frantz for spending time teaching me next-generation-sequencing methods, Dr. Axel Barlow and Georgios Xenikoudakis for offering valuable statistical advise. Also, a thank you to Jeff Veitch, Kristen Hopper, Anke Pohl, and Alexandra Dörhöfer for their help when technology was failing.

Furthermore, I would like to express my gratitude to my fellow PhD students in the different projects – Heidi Eager, Ardern Hulme-Beaman and Ophélie Lebrasseur – for making this collaboration an agreeable experience.

And last, but not least, I would like to thank my family and friends – Christine, Inga, Silja, Ulrike, Allar, Geerke, Tina, Sergio, Anke, Maciek & George – for their endless support and belief in me, and who kept me going along this PhD journey.

Table of Contents

List of Abbreviations

+ Γ	Gamma distribution
+ I	Invariable sites
A	Adenine
AD	*Anno Domini*
aDNA	ancient DNA
AIC	Akaike Information Criterion
BC	Before Christ
BIC	Bayesian Information Criterion
bp	base pairs
BP	Before Present
C	Cytosine
c.	calibrated
Cyt b	Cytochrome b
ddNTP	Dideoxynucleosidetriphosphates
D-loop	Displacement-loop (= control region)
dNTP	Deoxynucleosidetriphosphates
DNA	Deoxyribonucleic acid
e.g.	*exempli gratia*
emPCR	emulsion PCR
et al.	*et alii*
g	gramme
G	Guanine
gb	gigabases
GTR	General Time Reversible
HKY	Hasegawa-Kishino-Yano
hLRT	Hierarchical likelihood-ratio test
HVR	Hypervariable Region
i.e.	*id est*
IUB	International Union of Biochemistry
K2	Kimura-2-parameter
kb	kilobases
km	kilometre
mb	megabases
MCMCMC	Metropolis-Coupled Markov Chain Monte Carlo
min	minutes
µl	microlitre
ml	millilitre
ML	Maximum Likelihood
µM	micromol
mM	millimol
mtDNA	mitochondrial DNA
NCBI	National Center for Biotechnology Information
ncDNA	nuclear DNA
ng	nanogramme
NGS	Next-generation-sequencing

NJ	Neighbour joining
PCR	Polymerase Chain Reaction
pH	*pondus Hydrogeni*
rpm	rounds per minute
RrC	*Rattus rattus* Complex
RRW	Relaxed Random Walk
sec	seconds
T92	Tamura-3-parameter
T	Thymine
UV	Ultra violet

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

1 General Introduction

1.1 Introduction

The Indian Ocean represents one of the oldest exchange routes connecting South-East-Asia with India, the Arabian peninsula, as far as Africa in the West. Since the beginning of the Common Era, long-distance trade in sailboats enabled transmission of not only new technologies, exotic goods and food items, but also biological exchanges between geographically distant and culturally diverse people. Although archaeological remains, particularly from the 1st millennium AD, reflect an intensification of maritime connectivity across the Indian Ocean (Boivin *et al*. 2014), the exact travel and trade routes across this vast area in early times are still subject to discussion. In order to gain insights into past structures of exchange, usually archaeological artefacts, or skeletal and botanical remains are used. Besides traditional archaeology, the incorporation of further disciplines like linguistics, textual sources, and genetics helps to unravel earliest contacts between distinct cultures in (pre)history, especially when archaeological finds are scarce.

Within the framework of this thesis various projects are presented, that aim to assess the potential of using commensal animals as proxies to infer pathways of human travel and trade. Commensal species are usually small animals, that live in close association with humans and opportunistically exploit their habitat and food sources. Utilisation of these new resources has led to a close relationship between humans and certain species, and thus favoured their global distribution due to translocations through humans. As the introduction of commensal species to new areas, especially over long distances or onto islands, is facilitated by humans through sea and land transport, their histories are shared and the spatiotemporal arrangement of genetic lineages of the animal population is shaped by human movement. Analysing the population history of commensal species can therefore help to unravel the complexities of evolutionary processes, invasive pathways, and subsequently contributes to the understanding of human activities throughout the study area.

Here, the suitability of three different commensal species – the black rat (*Rattus rattus*), the house mouse (*Mus musculus*), and the Asian house gecko (*Hemidactylus frenatus*) – is assessed, in order to reconstruct pathways of human movements across the Indian Ocean in (pre)history. All three species have – with the help of humans – successfully invaded and established populations in new habitats outside their natural bounds, as they all combine specific behavioural-ecological characteristics. Each species is widely distributed within the study area and due to their size transported unintentionally, thus reflecting the actual composition of translocated lineages. Further, the use of multiple species offers a broader perspective in our understanding of long-distance movements of humans (Jones *et al*. 2013).

Despite the major impact and consequences of the widespread human-aided dispersal of commensal animals throughout the region, questions regarding who transported them, where they came from, and what routes they took, remain unresolved. Especially in the case of the widely scattered islands in the Indian Ocean, material artefacts or skeletal remains are scarce due to few excavation activity. Therefore, genetic analyses from modern and museum samples of the species in question have been employed, and embedded in a phylogeographic approach. This integrative methodology connects genealogy and geography, with the aim to reconstruct evolutionary, demographic and biogeographic processes that led to the contemporary distribution of genetic lineages (Avise 2000). Especially island environments are prone to changes within the diversity of their flora and fauna, e.g. through extinctions or population turnovers. Therefore, it is possible that contemporary phylogeographic pattern might not accurately reflect past structures. The application of ancient DNA analysis provides a powerful method, enabling detection of source populations and direct monitoring of their genetic change through time. How far the history of a certain population can be traced back, depends on the age of the samples. Here, the analysis of museum samples is a suitable alternative, when archaeological remains are not available or, in case of micro-mammals and –reptiles, are difficult to analyse. Furthermore, human genetic approaches have been proven to be inadequate to establish clear migratory pathways of early coloniser, due to low levels of genetic variation (Denny & Matisoo-Smith 2010).

All genetic analyses in the individual studies of this thesis use markers of the mitochondrial genome. The marker choice was based on the ecology of the study species. Due to a hierarchical structure, populations are resistant to secondary invasion by females. Therefore, newly arriving females coming into an established population are generally unable to survive or gain mates, and in consequence do not contribute to the gene pool of the population. This means the maternally inherited mitochondrial DNA may reflect signals of the first coloniser in the regions and subsequently of the humans who transported them (Boursot *et al*. 1993, Gabriel *et al*. 2010).

Through the application of phylogeographic analyses of modern and museum samples from different commensal species, the following research questions are addressed in the individual case studies of this thesis:

1. As archaeological records and historical data is scarce across the Indian Ocean area: is it possible to reconstruct patterns of past human movements by tracking the spatiotemporal arrangement of genetic lineages of commensal species that are unwittingly transported by humans?
2. Given that people have moved small commensal animals around for a long time, undirected distribution pattern of contemporary populations are expected for each species. However, through ancient DNA analyses direct monitoring of the population history through time is possible, that is otherwise rapidly erased over time. As archaeological remains of the species in question are scarce, are museum samples old enough to trace back routes of exchange in (pre)history?
3. Here, multiple commensal species are used to understand long-distance movements of humans. Due to their different ecology and dependence on humans: are certain species more suitable to answer specific questions regarding early human contacts than others?

Furthermore, in an additional chapter, a methodological exercise incorporating latest high-throughput sequencing techniques has been carried out. Through the innovative methodology, the first mitogenome from museum samples of the Asian dwarf buffalo (*Bubalus mindorensis*) has been generated, in order to assess its position within the taxonomy of the bovid family. Recent developments in the field of ancient DNA have widened the range of applications from single genes to whole genomes, which subsequently extended the range of questions that can be addressed. Refined extraction and sequencing techniques allow the analysis of ever decreasing amounts of DNA, enabling the use of aDNA methods to understand phylogenetic relationships over space and time.

Thesis structure

In the following introductory chapters I will summarise the current evidence of earliest contacts and emerging trade activities across the Indian Ocean. Different sources of evidence are taken into account, including archaeological finds, such as zooarchaeological and archaeobotanical remains or pottery, as well as linguistics, historical texts, and genetics. Furthermore, the concept of commensalism will be explained. I will give a definition of the term 'commensalism', summarise how and where it originated and outline the application of the commensal model in this study. Additionally, I will explain characteristics of commensal animals with focus on their invasion success. On that basis, I provide brief background information about the ecology and behaviour of the different study species, as well as about their relationship to humans. Thereafter, I will give an overview about palaeogenetic research – the general fields of application and latest technological developments, as well as the structure and properties of ancient DNA and the related application of anti-contamination measurements in order to ensure the generation of authentic and meaningful data. In addition, characteristics of the mitochondrial DNA, the genetic marker used in all studies here, are described. The last chapter of the general introduction provides basic information about the methodological approaches used within this thesis.

After the general introduction, three case studies are presented (chapter 2, 3, 4), where phylogeographic analyses of commensal animals, such as the house mouse (*Mus musculus*), the black rat (*Rattus rattus*), and the Asian house gecko (*Hemidactylus frenatus*) are employed in order to unravel invasive pathways and relate them to human trade and travel routes. In a further methodological exercise (chapter 5), latest next-generation-sequencing techniques have been incorporated, in order to generate the first mitogenome of the Asian dwarf buffalo (*Bubalus mindorensis*) and resolve its position of within the bovid family. These case studies are written in paper manuscript format, addressing specific but closely related questions. Conclusions and future work are discussed in an overall summary in chapter 6. All methodological protocols used during labwork, as well as sample lists of all specimens used in each study are found in the Appendix (chapter 7).

1.2 Overview of the Indian Ocean history

The Indian Ocean represents one of the oldest maritime trade networks connecting South-East Asia and Australia in the east, with Asia in the north, and Africa in the west. Since the 4th millennium BC, earliest civilisations arose throughout this vast area, e.g. in Mesopotamia, Egypt or the Indus Valley (Beaujard 2005). Despite its ecological and cultural diversity, the region has experienced remarkable interaction, trade and exchange between theses different cultures. Thereby, not only new technologies, religious beliefs, and exotic goods have been spread, but also various animal and plant species (Boivin *et al*. 2013). Based on geographic factors, the Indian Ocean can be divided into several regional subsystems: the China Sea, the Bay of Bengal, the Persian Gulf, and the Red Sea. From these core regions, interaction with the neighbouring trading spheres developed over time (Beaujard 2005).

Figure 1.1: Map of the study area.

In the following paragraphs, earliest evidence of contact and emerging trade activities of the prehistoric Indian Ocean will be summarised, before intensification of the connections between South-East Asia, Arabia and East Africa occurred during the 1st millennium AD. Different sources of evidence are taken into account, including archaeological finds, such as zooarchaeological and archaeobotanical remains or pottery, as well as linguistics, historical texts, and genetics.

Prehistoric evidence

Direct evidence of maritime subsistence before the 1st millennium AD is scarce. A few sites of human occupation from the Later Stone Age have been identified along the East African coast, e.g. Kilwa, and on offshore islands such as Mafia or Zanzibar, indicating that local hunter-gatherer communities had sufficient maritime knowledge and were able to navigate basic canoes through coastal waters (Breen & Lane 2003). Based on these small-scale coastal societies, cross-cultural contacts throughout the Indian Ocean have been established and intensified, leading to the vast trade network connecting Asia and Africa by the 1st millennium AD (Boivin & Fuller 2009).

Although archaeological remains of prehistoric periods are still insufficient, there is growing evidence that from the Bronze Age onwards the previously separate regional trading networks became interlinked. By the end of the 3rd millennium BC, a trans-oceanic connection between Africa and India has led to a gradually increasing trade of goods across the region (Fuller *et al*. 2011, Boivin *et al*. 2013). Thereby, several African crops, such as sorghum (*Sorghum bicolor*), pearl millet (*Pennisetum glaucum*), and finger millet (*Eleusine coracana*), as well as cowpea (*Vigna unguiculata*) and hyacinth bean (*Lablab purpureus*) reached Asia. Finds of these crops increased significantly at western Indian sites around 2,000 BC and probably arrived with Harappan seafarers (Madella & Fuller 2006). By the 1st millennium BC all five crops are widespread throughout India (Fuller & Boivin 2009). The exact routes from Africa to the East remain uncertain, but a lack of archaeobotanical remains on the Arabian peninsula points to a more direct sea dispersal. Although active trade between the Arabian peninsula and western India has been documented

during that time, there is no evidence that Indian ships reached the East African coast (Boivin *et al*. 2014).

In contrast, translocations from Asia to the west are less well understood. Asian broomcorn millet (*Panicum miliaceum*) spread from its chinese origin in westward direction, reaching Central Asia by 2,000 BC, Pakistan by 1,900 BC and the Sudan by 1,700 BC. The absence of broomcorn millet remains in adjacent regions, like Mesopotamia, indicates a seaborne transport to East Africa rather through overland trade (Fuller & Boivin 2009). At the same time two Asian domesticates, the Indian zebu cattle (*Bos indicus*) and the chicken (*Gallus gallus*), have been first introduced to East Africa, most likely via the Arabian Sea. Their early presence is based on depictions and few skeletal remains from Egyptian sites dating to the 2nd millennium BC indicating the import of chicken as exotic goods, e.g. for cockfighting, rather than as sources of food (Houlihan & Goodman 1986). Chicken bone assemblages increase by the mid 1st millennium BC, suggesting the incorporation in the African cuisine (MacDonald & Edwards 1993). The zebu cattle was present in East Africa probably from the 1st millennium BC, as humped clay figurines from Egypt and Nigeria suggest. There, hybridisation with local African taurine cattle took place during later times (Boivin & Fuller 2009).

The early contacts between India and Africa have not only led to the exchange of different crop plant and animal species, but opened up what later became the famous spice trade route. First evidence of spice trade comes from black peppercorn (*Piper nigrum*) remains at the grave of Pharaoh Ramses II., dating to 1,200 BC (Plu 1985). Black pepper is native to the tropics in southern India and has probably been moved along the Arabian Sea by hunter-gatherers and small fisher communities (Fuller *et al*. 2011). Further Asian spices such as sesame (*Sesamum indicum*) reached North Africa during the 2nd millennium BC, as residues in Tutankhamun's tomb in Egypt show. Here, additional finds of these spices in Mesopotamia suggest transport via a land route (Weisskopf & Fuller 2014). Sesame has been used in the African cuisine as leafy vegetables and its seeds have been ground to flour or oil (Blench 1997). Early evidence of cinnamon (*Cinnamomum sp*.) has been detected in Israel in the

course of the 1st millennium BC, before it became a common flavour in the African, the Middle Eastern and the Mediterranean cuisine in classical times (Namdar *et al*. 2013).

While an Arabian trade system has been established in the western Indian Ocean, evidence for regular contacts between India and the Malay peninsula throughout the Bay of Bengal emerged during the 1st millennium BC. Indian style stone beads and black polished ceramics have been extensively traded in eastward direction, as finds from sites in southern Thailand document (Bellina & Glover 2004, Bellina-Pryce & Silapanth 2008). Additionally, archaeobotanical evidence of Indian crops such as mungbean (*Vigna radiata*) and horsegram (*Macrotyloma uniflorum*) appear in the South-East Asian record at least from 400 BC (Castillo & Fuller 2010). In contrast, a variety of plants have spread in the opposite direction. For instance, the *Areca*-nut palm (*Areca catechu*), leaves of the *Piper betel*, as well as citron (*Citrus medica*), mango (*Mangifera indica*), and sandalwood (*Santalum album*) have their origin in South-East Asia, but have been moved to the Indian peninsula by humans around 1,400 to 1,300 BC (Asouti & Fuller 2008).

Although evidence of contacts across the Indian Ocean in early prehistory is limited, gradually increasing interaction between small-scale agricultural and fishing societies across the Arabian Sea in westward and the Bay of Bengal in eastward direction is confirmed at least during the 2nd and 1st millennium BC, respectively. Based on regional networks, systematic trading activities developed over time between the growing urban civilisations. Despite the evidence of translocation of diverse animal and plant species, as well as archaeological remains, the driving motives of early trans-oceanic voyages remain opaque. Factors that led to the crossing of dangerous open waters with only basic maritime technology initially may have been the recognition of the value of luxury and exotic goods as social status and power, or religious prestige (Fuller *et al*. 2011).

With the intensification of the trade across the Indian Ocean during the 1st millennium AD further motives of exchange came into focus. Progressive urbanisation, political stabilisation, and development of technological knowledge led

to improvements of the infrastructure and consequently to the increase in regular long-distance travel and trade (Boivin *et al*. 2014). Due to the requirement of continuous food supply, domestic animals, grain, and spices, as well as processed food items such as oil, fish, and wine have been transported to different areas where insufficiency of the local food production have led to the need of trans-regional imports. Furthermore, various plants have been used as fibre, dye, and perfume, medicine or fodder (Boivin *et al*. 2012). In some cases, rare or exotic items, live animals but also certain foods like sugar and rice have been acquired due to symbolic reasons, as gifts or tribute to express an elite status (Foster 1998). However, in general exchange happened due to everyday processes of travel (Boivin *et al*. 2014).

In summary, the current prehistoric evidence reveals active maritime connections in the western Indian Ocean throughout the Arabian Sea and in the eastern Indian Ocean throughout the Bay of Bengal. Thereby, different plant translocations occurred between Africa and India, as well as India and South-East Asia. However, these contacts have been restricted to the particular regions and interaction over long distances occurred only at later times, probably from the latter half of the 1st millennium BC onwards. The extension of trans-oceanic networks and the development of regular trade during the 1st millennium AD will be discussed in the following chapter.

Intensification of contacts during the 1st millennium AD

In the course of the late 1st millennium BC and the 1st millennium AD, the regional trade networks between small hunter-gatherer and fishing communities developed into trans-regional exchanges between rising port cities. The emergence of the Persian empire in the 6th century BC linked the Mediterranean Sea with the Indian Ocean, controlling trade routes as far as India. Further, the unification of the great empires in China and India during the 3rd century BC has led to the opening of overland and maritime trade routes, like the Silk Road, connecting East Asia through the Indian peninsula with the Mediterranean and East Africa. By the 4th century AD,

the Indian Gupta empire maintained frequent trading activities primarily with South East Asia via the Bay of Bengal, lesser in westward direction (Beaujard 2005). Around the 7th and 8th century AD, the trade routes throughout the Arabian Sea and along the East African coast became dominated by Arab merchants, before European explorers sailed throughout the entire Indian Ocean area during the Age of Discovery (Love 2006).

Increasing urbanisation due to constant population growth, advanced maritime technologies, and infrastructural improvements during the 1st millennium AD have led to the regular trade of a range of goods, including the translocation of biological species. Several sources of evidence demonstrate how the prospering contacts between different cultures slowly began to influence local traditions, agricultural systems, and culinary practices (Boivin *et al*. 2014).

Zooarchaeological evidence. The dynamics of exchange are visible particularly in the archaeological records. Skeletal remains of different Asian domesticates such as zebu cattle, goat, and sheep increase in urban settlements along the East African mainland, as well as on Zanzibar, Madagascar or the Comoros from the mid 1st millennium AD (Juma 2004). Although occasional hybridisation of the zebu with the African taurine ('Sanga breed') may have taken place already by the end of the 1st millennium BC, intensification of this process occurred probably only during the later half of the 1st millennium AD, with a subsequent southward spread of the Sanga cattle with Bantu-speaking pastoralists (Gifford-Gonzalez & Hanotte 2011).

Further zooarchaeological evidence of domesticates stem from bone assemblages of pig, cat, and dog. Remains of the pig (*Sus scrofa*) have been identified on Zanzibar and Madagascar, dating to the 6th to 8th century AD and the 11th to 13th century AD, respectively. The absence of this domesticate at sites along the East African coast may be explained due to the spread of the Islam and the related exclusion of pork consumption (Juma 2004, Boivin *et al*. 2013). The cat (*Felis catus*) is an South-West Asian domesticate and reached African islands probably during the mid 1st millennium AD. Finds on Zanzibar, the Comoros, and Madagascar suggest a seaborne dispersal from Asia and are dating to the 6th to 13th century AD (Horton & Mudida 1996, Juma 2004). In contrast, the domestic dog

(*Canis familiaris*) appears less frequently. While skeletal remains of dogs are absent on the Comoros, they could be recovered from coastal sites on Zanzibar (6th to 7th century AD) and on Madagascar (13th century AD) and may have been introduced by early farming groups (Juma 2004).

The East African mainland and the islands off the coast mark an important trading point and demonstrate the development of African seafaring technologies. Around the mid 1st millennium AD urban settlements and ports have been established on Madagascar, Zanzibar, and the Comoros. Since then, frequent contacts between the East African coast and nearby islands are shown by the human-mediated translocation of various wild animal species (Boivin *et al*. 2013). For example, finds of the wild bush pig (*Potamochoerus larvatus*) reveal introductions from the mainland to Zanzibar between the 6th and 8th century AD and a few centuries later to Madagascar (Juma 2004). Furthermore, the helmeted guinea fowl (*Numida meleagris*) arrived on Madagascar during the 11th to 13th century AD and becomes common in later strata. Surprisingly it is absent on other islands, such as Pemba or Zanzibar, although the helmeted guinea fowl appears at sites on the Kenyan coast (Helm 2000). The introduction of further wild animal species to Madagascar is reported during the 13th to 14th century AD. Few skeletal remains of the hinged-back tortoise (*Kinixys belliana*) and the rock dove (*Columba livia*) are currently restricted to the northwestern part of Madagascar (Radimilahy 1998).

Further evidence reveals a connection not only between the East African coast and mainly Madagascar, but a close trade and travel network between the offshore islands. In course of the 9th century AD several species, indigenous to Madagascar, have been translocated to the neighbouring Comoros islands. Among them are, e.g. the common tenrec (*Tenrec ecaudatus*), two lemur species (*Eulemur fulvus, Eulemur mongoz*), and two tortoise species (*Geochelone yniphora, Erymnochelys madagascariensis*) (Walsh 2007). Although additional species translocations appear between the Comoros and islands further north such as Mafia, Zanzibar, and Pemba, the exact time periods of their introductions are not confirmed and may be more recent (Cheke 2011).

A further source of evidence to unravel early human contacts are commensal animals. Because of their close association with humans, commensals are mostly translocated unintentionally and therefore provide direct insights into human trade activities and travel routes. Non of the three commensal study species in this thesis – the black rat (*Rattus rattus*), the house mouse (*Mus musculus*), and the Asian house gecko (*Hemidactylus frenatus*) – are native to the African continent, but have their origin in Asia. Therefore, the worldwide spread of these species was facilitated by humans over sea and land transport.

The first secured evidence of black rats in Africa stem from Egypt and date to the later half of the 1st millennium BC. Remains of the rodent have been found in stomachs of mummified large birds and cats at Quseir el-Qadim, an important port city at the Red Sea coast that has been involved in the Indian spice trade (Armitage *et al*. 1984). During the 1st millennium AD, the black rat becomes widespread along the East African coast. Finds from Zanzibar date to the 6th to 8th century AD, whereas the black rat occurs in Kenyan sites around the 8th to 10th century AD (Helm 2000). On the neighbouring island of Pemba, rat remains only date to the 13th to 14th century AD, possibly due to the lack of published material from earlier sites (Fleisher 2003). Furthermore, the black rat has arrived on Madagascar and the near by island of Mayotte possibly with the first urban settlers from the 9th century AD onwards, while it only occurs on the Comoros during the 11th century AD (Hingston *et al*. 2005). In the course of the later 1st millennium AD, rats spread southwards along the East African coast and reach South Africa between the 8th and 12th century AD, reflecting the connection of trading communities in that area (Plug 2000). In order to figure out the source populations of the translocated rats to East Africa, genetic studies have been carried out. The results suggest an introduction of Indian rats to Madagascar via the Arabian peninsula. Genetically different rats on other islands throughout the Indian Ocean and especially the neighbouring Comoros, indicate several independent colonisation events (Hingston *et al*. 2005, Tollenaere *et al*. 2009). However, sample sizes and geographical coverage of the current studies are insufficient. Therefore, this project tries to fill in the gaps of knowledge by extending the study area in order to get a broader picture of early rat translocations.

Remains of the house mouse in Africa are generally scarce. Although possible finds date as early as the 2nd millennium BC, these have to be treated with uncertainty. Due to burrowing and stratigraphic mixing, the age of these samples is in doubt (Blench 2007). First secured evidence is reported from sites on the islands along the East coast. Thereby, earliest finds from Zanzibar date to the 6th to 8th century AD. Further, on Grande Comore remains are restricted to a single tooth from a strata dating to the 9th to 10th century AD, on Madagascar skeletal remains dating to the same time period (Radimilahy 1998, Juma 2004). So far, there is no evidence of the house mouse from the African mainland, supporting a direct introduction from South-East Asia to the offshore islands. However, the lack of finds may be a reflection of a general low population size due to the competition with the dominant black rat, which is more abundant at African sites (Harper & Cabrera 2010). Current genetic analyses of the house mouse revealed, that Malagasy mice belong to the *gentilulus* lineage, which is also found on the Arabian peninsula. Whether these rodents reached the islands directly with Arab ships or took a route along the East African coast is in question (Duplantier *et al*. 2002, Boivin *et al*. 2013). Further sampling across the area needs to be carried out, in order to unravel introduction pathways of the house mouse to Africa.

A third commensal species that has been moved with humans across the Indian Ocean is the Asian house gecko *Hemidactylus frenatus*. It is native to the tropical and subtropical regions of Asia, but reached islands throughout the Indian Ocean via natural and human-mediated translocations. So far, only limited genetic research has been carried out, pointing to related haplotypes from Sri Lanka to islands in the west, such as the Mascarenes and Madagascar (Vences *et al*. 2004). Within the framework of this thesis, we extend the sampling area in eastward and westward direction from their Indian origin and try to distinguish between patterns of natural colonisation events and the introduction of *Hemidactylus frenatus* by humans to new habitats.

Archaeobotanical evidence. Besides the zooarchaeological evidence, further archaeobotanical remains not only show early translocations of plants in both directions, but indicate a significant change in crop consumption due to intensification of contacts between Asia and important trading points in East Africa during the 1st millennium AD. Throughout the Iron Age, native western African cereals such as sorghum, pearl and finger millet have been introduced to East Africa, reaching the offshore islands Pemba and Zanzibar probably by the 7th century AD (Helm *et al*. 2012). From prehistoric times onwards, Asian introductions started to replace African crops and ultimately dominated the archaeobotanical records in the late 1st millennium AD (Walshaw 2010, Boivin *et al*. 2013).

The banana (*Mus paradisiaca*), greater yam (*Dioscorea alata*), and taro (*Colocasia esculenta*) represent important sources of evidence for trans-oceanic connections between South-East Asia and Africa. Earliest archaeobotanical finds of banana come from Cameroon and date to the 5th century BC (Mbida *et al*. 2006). Unfortunately, there is no further evidence from neighbouring sites that supports an introduction to the East African mainland prior to the colonisations of the offshore islands. Unraveling the introductory pathways of this plant is difficult even with molecular genetic studies, due to its complex genetic makeup through a hybridisation event. Hybridisation of the cultivar could have taken place either in South-East Asia or India, where both types occur. Therefore, it is possible that the banana did not reach Africa on a direct route from South-East Asia, but through India (Kennedy 2009, Boivin *et al*. 2013). Currently, different scenarios are postulated: bananas may have been first cultivated in West Africa after an introduction via a trans-African direction from the East African coast (De Langhe 2007), rather than through a direct sea route (Blench 2009), as greater morphological variation in West African bananas shows. Further, linguistic evidence hypothesises a dispersal of bananas to Madagascar more likely from the East African coast, than a direct import from South-East Asia (Beaujard 2011). However, wild bananas found on Pemba share similarities with the wild-type from Java or the Moluccas (Simmonds 1962). For taro and greater yam, archaeobotanical and genetic data are lacking, wherefore the time and pathways of their introduction to Africa are still unknown. Historical and linguistic evidence point

to an early arrival with South-East Asian seafarers, due to the utilisation of taro and greater yams in ritual purposes on Madagascar as still common in current Melanesia (Beaujard 2011).

Another Asian crop that reached Africa initially in small quantities, but became abundant in the archaeological record in later times is rice (*Oryza sativa*). Both domesticated forms – *japonica*, which has been domesticated in China and subsequently spread to South-East Asia, as well as *indica*, which has been domesticated in northern India – occur on Madagascar (Fuller *et al*. 2010). While *japonica* probably came to Madagascar on a direct route from South-East Asia through early contacts, *indica* arrived later from India with Arab seafarers (Beaujard 2011). Further finds of the *japonica* rice variant from deposits on Pemba dating to the 11th to 15th century AD, suggests a dispersal throughout the East African islands from Madagascar to the Comoros and Pemba. Evidence of rice cultivation on the mainland is almost absent so far. From the 11th century AD onwards, rice became more and more important dominating the archaeobotanical record along the East African coast, where it was also grown locally by then (Walshaw 2010, Boivin *et al*. 2013). In turn, finds of African crops appear less frequent at sites on Madagascar and the Comoros during that time. The change towards a subsistence based on rice cultivation among Swahili communities on the East African offshore islands relates to an increase in population density. Furthermore, rice gained importance in ritual uses and social meanings, emphasising the need of these communities to differentiate themselves from the coastal societies towards eastern Indian Ocean trade contacts (Walshaw 2010).

A further plant that demonstrates contacts among the Indian Ocean trade network is the coconut (*Cocos nucifera*). Due to its nature of being a portable source of food and water, the coconut represents a basis of oversea travels and has been translocated to various places by humans (Gunn *et al*. 2011). Earliest evidence in assemblages at sites on Pemba, the Comoros and mainland Tanzania date to the 6th to 8th century AD, while linguistics indicate an earlier introduction to Madagascar (Beaujard 2011). Genetic studies revealed two distinct groups of coconuts, that are distributed

throughout the Pacific and the Indian Ocean, respectively. Admixture of these two groups occurs along the trade route between South-East Asia, Madagascar and East Africa, reflecting the South-East Asian and African contacts. A second introduction via the Arabian peninsula to East Africa is supported by linguistic evidence (Allibert 2008, Gunn *et al*. 2011).

While Asian plants and cultivars got established in East African communities in the course of the 1st millennium AD, similar patterns can be observed vice versa. Although several African crops, such as sorghum, pearl millet, and finger millet reached India already in early times and from there got transported to South-East Asia, they have been implemented in Asian agriculture and cuisine only centuries later. Thereby, the African imports occur alongside native crops and do not dominate the archaeobotanical records (Boivin *et al*. 2014). For example, the African baobab tree (*Adansonia digitata*) is distributed throughout Africa, Madagascar, and also found in Arabia and India. Although the exact routes of its spreading are not clear, transport by Arab seafarers is suggested, who traversed frequently between India, the Arabian peninsula and along the African coast (Pock Tsy *et al*. 2009). By the end of the 1st millennium AD another important crop plant – coffee (*Coffea arabica*) – reached first the Arabian peninsula and from there spread with European explorers during Medieval times to Europe and their colonies across the Indian Ocean (Wild 2004). Further determination of African introductions in the Asian plant record difficult, due to a lack of archaeobotanical remains. Plants of African origin that appear in India are different kinds of legumes, such as winged beans (*Psophocarpus tetragonolobus*), guar beans (*Cyamopsis tetragonoloba*), and velvet beans (*Mucuna puriens*), but their dispersal histories are still uncovered. Cultivars like okra (*Abelmoschus esculentus*), brown mustard (*Brassica juncea*), or the tamarind tree (*Tamarindus indicus*) are supposed to be translocated to South Asia, but their areas of origin are in debate (Boivin *et al*. 2014).

Pottery. In addition to zooarchaeological and archaeobotanical finds, excavations of early settlements reveal further finds, such as pottery and glassware, indicating contacts between different regions and cultures around the Indian Ocean. An important site is Hafun West located at the Horn of Africa in Somalia. Here, occupation is documented from the early 1st century BC onwards. Assemblages contained finds from the Red Sea, Mediterranean, as well as Mesopotamian and Indian pottery. The composition of finds changes in the neighbouring site Hafun Main, which has been occupied during the first half of the 1st millennium AD. While imports from the Gulf region dominate with additional finds from India, remains indicating contacts with the Mediterranean or Red Sea are scarce (Horton & Middleton 2000).

One of the earliest sites on the offshore islands is Unguja Ukuu on Zanzibar. Here, Mediterranean pottery and blue-green glazed Sassanian-Islamic ceramics dating to the 5th and 6th century AD have been found. Similar remains from the same time period also occur in the northern part of the island at Fukuchani. On the East African mainland, 5th to 6th century AD Graeco-Roman glassware is recovered in Egypt and further south in Tanzania, together with blue-green glazed pottery mentioned above. Further evidence of contacts between the Arabian peninsula and East Africa is shown by finds of green-glazed pottery in Mozambique, as well as on the Omani coast, dating to the mid 1st millennium AD (Horton & Middleton 2000). Generally, archaeological remains are lacking before the 7th century AD, when urbanisation and advanced technological knowledge led to the rise of the Swahili culture along the East African coast. If this is based on an insufficient archaeological record due to a lack of excavations or if early sites have not been occupied over long periods of time is still in debate. However, the current evidence suggests frequent trade activity with Arab merchants, who dominated the sea routes to East Africa and explains the paucity of Mediterranean finds (Boivin *et al*. 2013). Doubts continue to exist about material from early settlements on Madagascar and the Comoros from the 8th century AD. Although the decorations on the pottery may be linked to South-East Asian contacts, further archaeological evidence of an arrival of Austronesians is lacking (Allibert 2008, Boivin *et al*. 2013).

Genetics. While archaeological remains are insufficient to provide a clear picture of early contacts between South-East Asia and Madagascar, human genetics revealed an equal contribution of African and Indonesian lineages to the Malagasy population. Thereby, results place haplogroups of the Malagasy Y-chromosomal marker in Borneo, which is consistent with linguistic studies (see next paragraph) (Hurles *et al*. 2005). Furthermore, evidence of an Indian source population on Madagascar is apparent. Whether this reflects an independent colonisation event or a layover on the route from South-East Asia to Africa can not be determined (Dubut *et al*. 2009).

Genetic studies of other offshore islands along the East African coast reveal a similar pattern of mixed origins. Although African lineages – with a distinct coastal East African signature and shared with Madagascar – are dominant among Comorians, further South-East Asian and Arabian contribution has been detected, reflecting continuous interaction with the Middle East since the 8th century AD (Msaidie *et al*. 2011).

Linguistics. Furthermore, linguistic studies prove contacts between South-East Asia and East Africa. Austronesian influence is most apparent on Madagascar, were the linguistic origin of indigenous people points to a mixed African and Indonesian ancestry (Dewar & Wright 1993, Hurles *et al*. 2005). Malagasy language shares almost 90% of its basic vocabulary with an Austronesian language spoken in the Barito Valley of southern Borneo (Dahl 1951, Dewar & Wright 1993, Adelaar 1995). Furthermore it contains influences of Bantu languages, spoken in East Africa particularly in Mozambique (Dahl 1988, Dewar & Wright 1993).

Additionally, several African terms regarding maritime technology or introduced plants and animals have been derived from Malayan languages. For example, the Swahili word for outrigger canoes, which are most likely of Austronesian origin, links to the Malayan term (Blench 1996). Further, the name for the Malagasy wild pig, *lambo*, is based on the Malayan word *lembo* for bovine. Similar patterns are observed for South-East Asian plants introduced to Africa. Thereby, the Malagasy word *vary* refers to rice and is related to Malayan word *bari*, meaning 'cooked rice'. The Malagasy name for greater yam '*ovy*' corresponds to '*ubi*' in Malayan languages (Beaujard 2011).

Textual sources. Besides archaeological remains and genetic studies, a further source of evidence provides insights into early contacts among the Indian Ocean trade network: textual sources. Here, especially a text from the beginning 1st millennium AD, the *Periplus of the Erythraean Sea*, describes the trade activity between African ports and the Arabian peninsula (Casson 1989). The *Periplus* thereby focuses on the East African coast ('Azania'), which at that time has been controlled by Arab seafarers. Based on the information in the text, African offshore islands have been inhabited at the end of the 1st millennium BC, with a subsistence built on fishing and probably some form of plant cultivation. Furthermore, the text mentions the presence of canoes and sewn boats, wherewith goods or biological species may have been transported throughout the area. Certainly, frequent trade between local communities and Arabian merchants has led to the exchange of iron, glass beads or wine in return for African ivory, rhinoceros bone, and turtle shell. A further source of textual evidence is provided by Ptolemy's *Geography*, dating to the 2nd century AD, which also mentions the region Azania, even though in less detail (Horton & Middleton 2000, Boivin *et al*. 2013). In the following centuries textual evidence is scarce.

The Age of Discovery

Between the 15th and 19th century AD, European seafarers started to extensively explore sea routes across the Atlantic, Indian and Pacific Ocean. At the beginning of the 'Age of Discovery', especially Portuguese and Spanish explorers were not only driven by the spirit of adventure, but were looking for new direct trading routes to the Asian markets in order to circumvent Italian and Muslim middlemen, who controlled the overland trade routes across the Mediterranean at that time. By having direct access to ports in the East, a constant import of luxury goods, such as silk from China, gemstones and spices from India, as well as gold and ivory from Africa, would have been enabled. While Portuguese mariners sailed along the West African coast to reach the Indian Ocean, the Spanish took a route across the Atlantic and discovered – by coincidence – the Americas instead. In the following centuries,

England, France and the Netherlands equally took part in the worldwide trading networks and established seaborne relations with distant regions (Love 2006).

Initial ventures of Portuguese sailors reached out into the Atlantic, discovering various islands like the Azores, the Canaries or Madeira. Subsequent colonisation around the mid 15th century AD made these islands important stopping points for further explorations. The first voyages along the African coast were launched by Prince Henry between 1415 and 1458, sponsored by the Portuguese Crown. Main goals of the early expeditions have been the access to African gold from the Niger Delta and the search for Christian allies against the overpowering Islam. By 1448, Portugal has entered the African slave trade, which was until then dominated by Arab merchants (Russell 2001). Several trading posts have been established and by the end of the 15th century AD, Portugal had authority over all trading pathways along the West African coast and connections to the caravan trade across the Sahara. After Prince Henry's death, additional expeditions led further south and reached as far as Namibia and South Africa, opening up relations with local rulers and trade routes to Asian markets. Thereby, Portuguese merchants slowly started to replace Arabian traders in this region (Love 2006).

In the beginning of the 16th century AD, improvement of navigational techniques has led to the circumvention of Africa by Bartolomeu Dias and initial voyages across the Indian Ocean by Vasco da Gama. Thereby, Chinas sudden retreat from their international maritime supremacy ultimately led the way for Portugal's naval power over Asian markets. Due to the commercial unity of the Indian Ocean area and the construction of strategic points along the coast from Africa to China, Portugal started to defeat Arabian fleets and to dominate major waterways across the Indian Ocean. Further, Portuguese seafarers reached various islands in the course of the 16th century AD, e.g. Socotra in 1507, Sri Lanka in 1515 or the Moluccas between 1516 and 1519. However, although Portugal became the main exporter of eastern goods, such as spices from India, silk from China or carpets from Persia, their relations with African and Asian rulers were primarily based on commerce, not conquest, due to a lack of military power. As a result Arab, Indian, Chinese, and Malay traders

remained prominent in areas were Portugal was not able to gain absolute control. By the end of the 16th century AD, Portugal's position began to weaken when not only Spanish, but Dutch, English, and French ships entered the competition of establishing trade routes to the East (Love 2006).

Besides the different sources of evidence, there are still major gaps in the understanding of how early contact between the different cultures across the Indian Ocean evolved. Most regions, especially the scattered islands within the area, are currently lacking a systematic archaeological exploration. A more cohesive investigation of the regions bordering the Indian Ocean and the incorporation of advanced techniques such as ancient DNA analysis may help to overcome the present limitations and bias in sampling. So far, archaeological evidence points to a gradually developing trade and exchange of predominantly luxury goods in early times. Only during the 1st millennium AD, a change towards a more specialised commerce occurred due to growing urbanisation and the subsequent transformation in food consumption and agriculture. However, the exact pathways of oversea translocations of various goods, plant and animal species remain uncertain and are still subject to debate.

1.3 Commensalism

In the following subchapters I give a definition of the term 'commensalism', summarise how and where it originated and outline the application of the commensal model in this study. Furthermore, I explain characteristics of commensal animals with focus on their invasion success. On that basis, I provide brief background information about the ecology and behaviour of the different study species, as well as about their relationship to humans.

1.3.1 The definition and origin of commensalism

Definition of commensalism

The term 'commensal' has various usages. Tchernov (1984) defined commensalism basically as "a one-sided symbiosis, in which members of one species are benefited (the active commensal partner), while those of the other species are neither benefited nor harmed (the passive host)". Within the framework of this thesis, commensal animals represent a variety of species that opportunistically exploit human habitat and food sources. The impact of this interaction thereby varies from having no apparent negative effect on humans (e.g. the Asian house gecko) to commensals becoming pests to home and agriculture (e.g. the black rat or the house mouse) (Jones *et al*. 2013).

The transition from hunter-gatherers to sedentary societies with a lifestyle based on agriculture and husbandry is commonly accepted as a key factor in the evolving commensal relationship (Cucchi *et al*. 2012). Population growth and the related increase in settlement size and density is linked to several factors that enhanced attractiveness for small mammals or reptiles. The shift towards sedentism created new ecological niches, providing a consistent source of food, as well as artificial shelter protecting against predation or interspecific competition. Furthermore, climatic pressures are reduced (Tchernov 1968, Tchernov 1984, Boursot *et al*. 1993, Cucchi *et al*. 2005, Cucchi *et al*. 2012). Exploitation of these new resources has led

to a close relationship between humans and certain species, hence favoured their global distribution.

Origin of commensalism

Localisation of a beginning synanthropy of the commensal species used in this study is difficult, due to a lack in the archaeological documentation. Evidence of a closer interaction is found in the zooarchaeological record of early farming sites from the beginning of the Holocene. However, these remains are often difficult to retrieve. Due to their fragile nature, bones of small mammals and reptiles, such as rats, mice or geckos are prone to decay – leading to a poor preservation. Often missed during the excavation process, the fragmented condition of most of the recovered bones further creates a problem in studying them. For example, differentiation between related species is almost impossible. Moreover, the poor preservation conditions in some areas can lead to a geographical bias in the record, leaving room for misinterpretations.

Determination of the range distribution of mice and rats before their connection with humans is difficult (Ervynck 2002). The first evidence of mammalian commensalism stem from the Levant about 12,000 years ago in association with the Natufian culture, which is characterised by a semi-sedentary lifestyle including the cultivation and storage of grain to some degree. Archaeological remains of *M. musculus* are present throughout deposits at Hayonim Cave and show a massive increase in Natufian times, *R. rattus* remains have been found in Natufian layers at Kebara Cave. Both sites are located in northern Israel (Tchernov 1984, Auffray *et al*. 1990, Cucchi *et al*. 2012). In the course of the following millennia, a transition from hunter-gatherers to sedentary societies took place independently in several regions of the world. Humans started to domesticate animals and to cultivate plants. By 10,000 BP, husbandry and agriculture were the main sources of food production. These practices then spread from the Near East to other parts of the world, such as eastwards to the

Iranian plateau, southwards to Arabia and North East Africa, and to Europe in the West (Cucchi *et al*. 2012).

This process represents the basis for the development of rat and mouse synanthropy – due to the factors mentioned earlier – and their subsequent invasion throughout the world. From the Neolithic onward, an exponential increase of small mammal bone remains from Levantine human settlements has been reported. Furthermore, evidence of granaries with complex constructions to avoid rodent parasitism have been excavated, pointing to a considerably closer association of small mammals and humans in this region (Cucchi *et al*. 2012).

Human-mediated transport during the Late Bronze and Early Iron Age facilitated a westward spread of the house mouse through Eurasia via two pathways. While *M. m. musculus* took a northern route to central and northern Europe, *M. m. domesticus* colonised the Mediterranean via a southern route. This is demonstrated by *Mus* fossils recorded from various sites around the western Mediterranean Sea and in central and northern Europe dated to the Bronze Age, and from sites in northwestern Europe dating to the Iron Age (Boursot *et al*. 1993). With the beginning of extensive seafaring activities of European explorers, the house mouse progressed onto other continents, such as North Africa and America, as well as to oceanic islands in the Atlantic and Pacific (Boursot *et al*. 1993, Bonhomme & Searle 2012, Jones *et al*. 2013).

In contrast, reconstruction of early dispersal routes of commensal rats is controversial. Finds of the black rat in the Levant, besides those from Kebara Cave, seem to be absent during the Neolithic. In this region, rat bones only appear again in younger time periods, such as in Roman and medieval layers. During the Metal Ages, evidence of the black rat as a commensal animal is only present in urban settlements inside the Tigris and Euphrates basin (Tell Selenkahiye/Syria, Susa/Iran, Uruk/Iraq) (Ervynck 2002). The time gap of about four millennia between the first remains of the black rat in the Levant and Neolithic finds from the Mesopotamian area, may be explained by diverse preservation conditions.

Furthermore, commensal black rats were present on the Indian subcontinent during the Neolithic. Therefore, it is unclear whether commensal rats first appeared in India along with the rise of the Indus valley civilisations and were then transported to Mesopotamia or vice versa. Additionally, information about the subsequent spread of the black rat north, west and south from the Near East is incomplete and difficult to determine. Archaeological remains from classical Egyptian sites and south along the East African coast, suggest a further translocation of the black rat through Arab traders. These remains date between the 8th and 12th century AD (Ervynck 2002).

Although evidence is scarce, there is reason to presume that commensalism evolved in different regions for different species, as the development of sedentary societies with a subsistence based on agriculture took place independently in several regions of the world. Furthermore, it seems – at least for *Mus* – that multiple independent commensalism events may have occurred within a species. Here, the change towards commensalism through human sedentism occurred after the divergence of each subspecies, which are now occupying distinct geographical areas (Boursot *et al*. 1993, Cucchi *et al*. 2012).

Compared to rats and mice, the history of *H. frenatus* as a pest overseas is considerably more recent. Although the Asian house gecko may have been taken to Pacific islands by Polynesians and Melanesians approximately 4,000 years ago (Case *et al*. 1994), this species has started to extensively invade habitats outside its natural range over the last 65 years, with a massive range expansion especially over the last two decades. Reason for the translocations is frequent passive transport by humans through the increased international shipping and cargo movement (Newbery & Jones 2007, Hoskin 2011). As the majority of introductions is facilitated by hiding on boats, especially in shipping containers, *H. frenatus* is mainly first detected in port areas (Lever 2003). The consequences of its uncontrolled spreading is particularly apparent in Australia, where the Asian house gecko is currently found not only in urban areas, but also started infiltrating natural habitats. There, it causes a decline in endemic gecko species by outcompeting them through predation and the introduction of new parasites (Hoskin 2011).

1.3.2 Characteristics of invading species

The dispersal and colonisation success of an invading species is dependent on different factors: its transportability, i.e. the facility with which a species is moved by people and its invasibility, i.e. the receptivity of the invaded environment (Anderson 2009). Commensal species are able to establish populations in new areas, due to certain advantageous characteristics that have been identified and will be further explained in the following paragraph.

First, commensals are generally small and easily transported, especially when inactive. They have remarkably broad physiological tolerances, such as low metabolic rates, and are able to adapt to severe conditions, like food scarcity. Moreover, they are generalists who exploit a wider range of resources for foraging, nesting or shelter than their wild counterparts, and are mainly nocturnal (Braithwaite 1980, Cole 2005). The successful establishment of invasive populations is enhanced by a constant environment, which is – in case of commensal species – provided by humans. Stable environmental conditions, such as a consistent climate, regular food resources and protection is the basis for year-round breeding and consequently a higher number of offspring (Pocock *et al*. 2004, Cole 2005). The offspring of invasive species usually reaches sexual maturity in early stages of their otherwise short lifespan. Due to high reproductive rates and the resultant population density, commensals show more complex social structures (Boursot *et al*. 1993). Once introduced, the invader rapidly expands their range due to the lack of predators (e.g. the Asian house gecko in Australia, see Hoskin 2011). In contrast, the commensal rodents used in this study have lower dispersal rates and mostly stay near the territory, where they were born (see following subchapter).

Within the framework of this study, three different commensal species – the house mouse, the black rat and the Asian house gecko – have been used to address questions regarding the reconstruction of early trade and travel routes of humans. All of them successfully colonised habitats outside their natural bounds, due to specific features. In the following I will provide background information about the ecology and behaviour of the three different study species, as well as their relationship to humans in order to explain their current worldwide distribution.

Characteristics of the commensal study species

Mus musculus

Common names: House mouse, common mouse (http:// www.issg.org/ database/ species/ ecology.asp?si=97&fr=1&sts=&lang=EN)

Geographical range: The house mouse is native to the Indian subcontinent, but it has now been spread throughout the almost entire world. Its current geographic abundance is facilitated by its commensal relationship with humans (http:// www.issg.org/ database/ species/ ecology.asp?si=97&fr=1&sts=&lang=EN).

Appearance: Head to body length of house mice ranges between 65-95 mm with an equal tail length of 60-105 mm. Weight of adults averages between 12-30 g. Their fur varies in colour from light brown to black with lighter coloured undersides in white or grey. Mice living in close relationship with humans tend to have longer tails and darker fur compared to their non-commensal counterparts (http:// animaldiversity.org/ accounts/Mus_musculus/, http:// www.issg.org/ database/ species/ ecology.asp?si=97&fr=1&sts=&lang=EN).

Habitat: Due to their close association with humans, house mice are able to occupy a variety of niches – houses, barns or granaries, but are also found in fields or woods and even in inhospitable environments such as tundra or deserts. Furthermore, observations of individuals seasonally changing from indoor to outdoor living have been reported (http:// animaldiversity.org/ accounts/Mus_musculus/, http:// www.issg.org/ database/ species/ ecology.asp?si=97&fr=1&sts=&lang=EN).

Ecology / Behaviour: House mice are skilled climbers and jumpers, quick runners and can swim well. They usually nest in hidden places like woodpiles and behind rafters or build underground burrows. Materials for the nests are generally soft and can vary from textiles to paper. House mice are generally nocturnal, although they also forage during daytime when in human dwellings. Mice are omnivores – while

favouring all kind of seeds, nuts and plants, as well as insects and carrion in a natural habitat, commensal mice consume any accessible human food, as well as household materials like glue and soap. Furthermore, most mice tend to store up their food for later use. The required water is mainly met by the moisture content of the food. Due to the ability to concentrate their urine, mice are able to survive without free water, which enables the colonisation of arid environments. House mice have a polygynous mating system and breed throughout the year. Depending on the conditions, females can have between 5-10 litters per year, usually consisting of 5-6 offspring. Sexual maturity is reached at 5-7 weeks, the average lifespan is 2 years although individuals in captivity can live up to 5 years. About 3 weeks after birth the young mice leave their mother's territory, though females more likely to remain in vicinity. Social groups and territories are complex and more pronounced in commensal mice populations. Here, groups consist of a dominant male and several females with their young. Occasionally, subordinate males occupy a territory or males share, although normally territories are defended. Females establish a loose hierarchy among them, but are less aggressive than males. House mice have excellent senses, especially vision and hearing. They have a precise sense of smell and use their whiskers to feel air movement or surface textures. Communication about social dominance or reproductive readiness is ensued through squeaking and pheromones (http:// animaldiversity.org/ accounts/Mus_musculus/, http:// www.issg.org/ database/ species/ ecology.asp?si=97&fr=1&sts=&lang=EN).

Introduction pathways: House mice are moved around by humans – accidently or deliberately – through cargo transport via road vehicles, aircraft, and sea freight. Mostly transported as stowaways, e.g. hidden in bags of crops or hay in earlier times. Additionally, they have extensively broaden their range through escapes from captivity or natural dispersal (http:// www.issg.org/ database/ species/ ecology.asp?si=97&fr=1&sts=&lang=EN).

Impact on humans: Domesticated forms are used as laboratory animals for biomedical research and as pets. On the other hand, mice are pests to home and agriculture by consuming and contaminating human food through urination and

defecation, as well as destruction of furniture or clothing. In addition, mice are vectors of diseases and parasites infectious to humans and responsible for extirpations of indigenous species in ecosystems outside their natural range. The level of commensalism varies among the different subspecies of *M. musculus*. While *castaneus* is exclusively commensal, *musculus* and *molossinus* are able to occupy several niches from human-related to wild environments; *domesticus* is mainly found near human dwellings, but establishes feral populations where competitors are absent (http:// animaldiversity.org/ accounts/Mus_musculus/).

Rattus rattus

Common names: Black rat, house rat, ship rat, roof rat (http:// www.issg.org/ database/ species/ ecology.asp?si=19&fr=1&sts=sss&lang=EN)

Geographical range: The black rat is presumably native to India and other Indo-Malayan countries, but is nowadays found on almost all continents through introduction by humans. This rodent species is particularly common in costal areas, as it thrives well in areas inhabited by humans and on large ships. It favours tropical regions and has been outcompeted by the Norway rat (*Rattus norvegicus*) in colder and harsher climates (http:// www.issg.org/ database/ species/ ecology.asp?si=19&fr=1&sts=sss&lang=EN).

Appearance: Head to body length of black rats ranges between 160-220 mm with a tail longer than the body (190 mm or more). Weight of adults averages between 70-300 g, whereas males are larger and heavier than females. Although usually black, their fur sometimes varies in colour from dark to light brown with a lighter coloured ventral belly. Furthermore, it has relatively large ears and eyes (http:// www.issg.org/ database/ species/ ecology.asp?si=19&fr=1&sts=sss&lang=EN, http:// animaldiversity.org/ accounts/ Rattus_rattus/).

Habitat: Black rats are able to utilise various habitat types from grasslands and forests to urban and suburban areas, but show a preference for drier environments. This species is often found in large numbers around harbours and costal areas, due to its spread through human seafaring activities and prefers to live and nest in high places such as trees or roofs. Studies have shown that rats can reach elevations up to 250 m above sea level. Although it was formerly common in urban and rural areas with temperate climate conditions, it has now been driven out in these regions by the more aggressive Norway rat (http:// www.issg.org/ database/ species/ ecology.asp?si=19&fr=1&sts=sss&lang=EN, http:// animaldiversity.org/ accounts/ Rattus_rattus/).

Ecology / Behaviour: Black rats are avid and agile climbers, using their long tail to keep the balance. Therefore, they usually prefer to nest in upper parts of buildings or trees, though sometimes ground level nests and burrows have been located. Materials such as sticks and leaves are used for the nest building. Rats are nocturnal and omnivores. Although they prefer fruit, grain, cereals, seeds, and other vegetation, rats also consume insects, invertebrates or any accessible human food and rubbish. Thereby, rats can cause vast damage to farms and livestock by gnawing through various materials and excreting on forage stores. Furthermore, rats need about 15ml of free water per day. Black rats have a polygynous mating system and breed throughout the year, if environmental conditions allow it. Peak breeding seasons are summer and autumn. Females can produce up to 5 litters per year, usually consisting of 6-12 offspring. Sexual maturity is reached after 3-5 months, the average lifespan is 1 year although individuals in captivity can live up to 4 years. Social groups consist of multiple males and females, with a dominant male being the most successful breeder. Hierarchies are formed among the males and females, whereas 2-3 females are often dominant to all other group members, except the dominant male. Females generally show a more aggressive behaviour than males, but are less mobile than males. Subadult rats are usually forced out of their natal home range. Territories of a range maximum of about 100m^2, resources of food and mates are defended. Black rats communicate through squeaks when threatened or socialising. Oil smears mark territorial boundaries and hierarchy in groups is determined through aggressive

postures and physical contact. Rats have excellent senses and use vision, hearing, touching and smelling to explore the environment (http:// www.issg.org/ database/ species/ ecology.asp?si=19&fr=1&sts=sss&lang=EN, http:// animaldiversity.org/ accounts/ Rattus_rattus/).

Introduction pathways: Natural dispersal to islands have been reported, that involve swimming of approximately 500 m through calm waters. Furthermore, rats are transported over long distances by humans in road vehicles or sea freight (http:// animaldiversity.org/ accounts/ Rattus_rattus/).

Impact on humans: While its closest relative, the Norway rat, is often used for biomedical research and kept as a pet, there are no known benefits of *R. rattus* for humans. In contrast, the black rat is a pest to home and agriculture. The rodent is not only destructive to crops, farms, and fruit trees, but also spoils food resources that it is unable to consume through urination and defecation. Furthermore, rats are vectors of diseases that not only affect livestock and other animals but also humans, e.g. the bubonic plague (*Yersinia pestis*). When the black rat is introduced to a new environment, it is often directly responsible or contributes to the extinction through predation of indigenous species such as birds, small mammals, reptiles, invertebrates, and plants, especially on islands (http:// animaldiversity.org/ accounts/ Rattus_rattus/).

Hemidactylus frenatus

Common names: Asian house gecko, Pacific house gecko, spiny-tailed house gecko, bridled house gecko, house lizard (http:// www.issg.org/ database/ species/ ecology.asp?si=1344&fr=1&sts=sss&lang=EN)

Geographical range: The Asian house gecko is native to south and southeast Asia and the Indo-Australian archipelago, but has quickly become a cosmopolitan species through introductions by humans to almost all tropical and subtropical regions

around the world. Introduced populations are now found in Central and South America, parts of North America and Africa, as well as on many small islands across the Indian and Pacific Ocean (http:// www.issg.org/ database/ species/ ecology.asp?si=1344&fr=1&sts=sss&lang=EN, http: www.iucnredlist.org/ details/ 176130/0).

Appearance: Head to body length of the house gecko ranges between 75-150 mm, with males generally being larger and heavier than females. Their scalation is uniform with distinctive spines along the back, arranged in bands around the tail. Colouration varies from grey or light brown and beige to greenish iridescence with white undersides. Further distinct characteristics of *H. frenatus* are their vertical pupils, subdigital lamellae, and enlarged medial subcaudals, as well as a second pair of chin shields and a very individual 'chuck chuck chuck' call, which is commonly emitted at dusk and dawn (http:// www.issg.org/ database/ species/ ecology.asp?si=1344&fr=1&sts=sss&lang=EN, http: www.iucnredlist.org/ details/ 176130/0).

Habitat: Geckos mainly occur in tropical and subtropical environments, but are also sometimes found in temperate areas. Although it is known to inhabit natural environments such as woodlands, forests, fields or rocky grounds, the Asian house gecko is highly synanthropic and most abundant in urban and developed locations. There, it often dwells on building walls near artificial lighting (http:// www.issg.org/ database/ species/ ecology.asp?si=1344&fr=1&sts=sss&lang=EN, http: www.iucnredlist.org/ details/ 176130/0).

Ecology / Behaviour: *H. frenatus* is a medium-sized arboreal gecko, living in close association with humans. Predominantly nocturnal, there is diurnal activity to some degree. Although its activity range is mainly influenced by light, temperature, and human activity, *H. frenatus* has remarkable broad physiological tolerances. Usually it hides under rocks and crevices during the day and only emerges at dusk to forage all night. The Asian house gecko is a generalist and active predator, which prefers open hunting surfaces such as building walls or vertical rocks. It is perfectly adapted to

forage on concentrations of insects around artificial lighting. Due to the simplicity of flat walls, *H. frenatus* recognises intruders earlier which enables their exclusion from the food source. Besides insects and spiders, the Asian house gecko also consumes juveniles of other gecko species and even their own progeny. A further advantage is its form of locomotion: adhesive toe pads on sub-digital lamellae enable increased climbing activity on a variety of surfaces, especially very smooth ones. This ability affords deposition of eggs in high crevices and refuge from predators. In addition, the eggs are adhesive and hard-shelled, making them resistant to moisture loss and therefore they most likely survive long distance travels. Functional sperm retention of up to a year facilitates the successful establishment to new areas, without the necessity of mating. In tropical habitat breeding occurs all year, but is seasonal in cooler climates. After an average of 50 days, two eggs per clutch are laid, partially fixed to a solid surface. Sexual maturity is reached between 6 months and a year, the general life span is approximately 5 years. The Asian house gecko is territorial with a social hierarchy. Territories are established and maintained through the distinctive 'chuck chuck chuck' calls. When in contact with other geckos, *H. frenatus* is most aggressive, attacking and biting the competitor (http:// www.issg.org/ database/ species/ ecology.asp?si=1344&fr=1&sts=sss&lang=EN, http: www.iucnredlist.org/ details/ 176130/0, Csurhes & Markula 2009).

Introduction pathways: The extensive range expansion of the Asian house gecko is the result of the constant increase of international travel, trade, and transportation over the last decades. Its superior ability to cling to various surfaces, allows this gecko species to access high crevice spaces for refuge and egg deposition, e.g. in cargo and shipping containers, and therefore goes undetected to new locations. Although the majority of translocations happens via sea freight, the Asian house gecko is also commonly found on road vehicles or planes. Natural introductions are difficult to prove, but distinct characteristics of this species (adhesive hard-shelled eggs, functional sperm retention) increases the potential and facilitates the successful establishment of populations in new environments (http:// www.issg.org/ database/ species/ ecology.asp?si=1344&fr=1&sts=sss&lang=EN, http: www.iucnredlist.org/ details/ 176130/0).

Impact on humans: *H. frenatus* does not present a threat to humans and is kept as a pet in some locations. In contrast, the Asian house gecko has a great impact on endemic gecko species when introduced to new areas, especially on islands. Once introduced, it behaves territorial and aggressive towards smaller or same-sized geckos and manifests its superiority by excluding them from food sources and retreat sites. This in turn, makes native geckos more vulnerable to predation. Attacks from *H. frenatus* often lead to the loss of toes or tails, which reduces movement and gripping ability and results in a higher mortality rate for their competitors. Successful measures have to be developed in order to stop the relentless spreading of the Asian house gecko, making it a potential threat for ecosystems (http:// www.issg.org/ database/ species/ ecology.asp?si=1344&fr=1&sts=sss&lang=EN, http: www.iucnredlist.org/ details/ 176130/0).

1.3.3 Application of the commensal model

The understanding of the temporal and geographical extent of human movement, contact, and exchange networks in (pre)history is primarily based on the distribution of archaeological artefacts or historical records. However, in some areas archaeological evidence is insufficient, especially in the case of the widely scattered islands in the Indian and Pacific Ocean. Furthermore, archaeological remains can provide an indication of arrival times for certain species, but without genetic analyses it is impossible to track the source population (Boivin *et al*. 2013). Therefore, "biological data can contribute to this debate through analyses of genetic variation in domestic and commensal animals" as they are "significant components of human dispersal" (Larson *et al*. 2007).

As outlined in the previous chapter, commensals and humans share a mutual history due to their close ecological relationship. Humans facilitate the introduction of commensal species to new areas through sea and land transport, especially over long distances or onto islands, thereby shaping the distribution of the animal population and their particular genetic lineages. Conversely, the human-mediated organisms can be used not only as proxies to infer human history, but evolutionary processes and invasive pathways (Jones *et al*. 2013).

Here, analyses of several commensal species will be performed in order to reconstruct human contacts within the Indian Ocean area. Thereby, the species in question will be embedded in a phylogeographical framework. The term 'phylogeography' was first used by Avise *et al*. in 1987. This integrative approach connects genealogy and geography, with the aim to reconstruct evolutionary, demographic and biogeographic processes that led to the contemporary distribution of genetic lineages. Results are then visualised in a phylogenetic tree (Avise *et al*. 1987, Avise 2000, Avise *et al*. 2009). By adding a spatiotemporal dimension to the genetic data, not only the colonisation history of an organism can be tracked, but also the geographical, geological and climatological phenomena shaping the distribution patterns, e.g. sea-level changes that result in the formation of land-bridges (Hickerson *et al*. 2010, Jobling *et al*. 2004, Jones *et al*. 2013).

A recent extension of the methodology described above, is the comparison of phylogeographic structures of multiple widespread taxa. In regard to the context of this thesis, 'comparative phylogeography' is performed to investigate the origin and dispersal of co-distributed organisms. All three species used here have successfully invaded and established populations in new habitats, as they all combine specific behavioural-ecological characteristics. Each species is widely distributed within the study area and due to their size transported unintentionally, thus reflecting the actual composition of translocated lineages. The use of multiple species offers a broader perspective and therefore helps to fill gaps in our understanding of long-distance movements of humans, especially where the archaeological record is missing or the genetic signal of early settlers is covered by subsequent immigration (Hickerson *et al*. 2010, Jones *et al*. 2013).

So far, several studies have been carried out mainly focusing on marker in the mitochondrial genome. Recent analyses of the phylogeography of the western house mouse (*Mus musculus domesticus*) demonstrated the advantage of bioproxies to provide insights in human settlement history. The current distribution of *M. m. domesticus* is directly caused by human movements, as the observed pattern of genetic lineage dispersal can be linked to maritime activities of humans within

different time periods from the Iron Age to the 15th century (Jones *et al*. 2011, Bonhomme & Searle 2012).

Matisoo-Smith & Robins (2004, 2009) demonstrated the significance of bioproxies by revealing insights into the spread of the Lapita culture throughout the Pacific on the basis of the Pacific rat (*Rattus exulans*). The Lapita culture is defined by a distinctive patterned pottery first found in Lapita, New Caledonia. Remnants of this characteristic pottery have been found throughout Remote Oceania, suggesting that Lapita people were the first settlers in this area and therefore represent the ancestors of modern Pacific people. But questions remain about whether there have been contributions of other populations from Asia and Micronesia (Denny & Matisoo-Smith 2010). The genetic diversity among human populations in Polynesia is very low and therefore not sufficiently informative. Nevertheless, phylogenetic analysis of *Rattus exulans*, which was deliberately transported in canoes as a source of food, unraveled multiple introductions of colonists and their particular paths of migration to various Pacific islands (Addison & Matisoo-Smith 2010).

A further example of tracking human colonisation routes through the dispersal of commensal species is shown by Omar *et al*. (2011). Their phylogenetic study on the Malagasy pygmy shrew confirmed that *Suncus madagascariensis* is not an island endemic, but was rather introduced from South-East Asia or the Indian subcontinent. This adds to the current genetic (Hurles *et al*. 2005) and linguistic (Adelaar 2009) evidence supporting a colonisation of Madagascar by Austronesian people from ISEA.

As these examples demonstrate, analysing bioproxies is a valuable approach and circumvents problems arising from studying scarce material artefacts and humans directly. Wide-spread commensal species may still carry genetic signals of early settlements, even when the original human founders and their descendants have left, died out or admixture by subsequent immigration is covering initial lineages (Jones *et al*. 2013).

Comparative use of ancient and modern material

Samples for the phylogeographical approach fundamental to chapters 2, 3, and 4 stem either from modern tissue (dried or in solution) collected during field trips or dried tissue from museum samples (up to 100 years old). Unfortunately, no archaeological remains have been available for any species in this particular study. Although the museum samples have only been obtained within the last century, they are considered sources of ancient DNA.

The use of ancient DNA has the ability to give insights into complex population processes over time that otherwise remain undetected when only studying the genetic composition of extant populations in a region, e.g. sequential events of colonisation, extinction, and replacement by modern populations (Cucchi *et al*. 2005). How far the history of a population can be traced back, is dependent on the age of the samples. Furthermore, ancient DNA helps to genetically identify fragmented zooarchaeological remains, where species determination is difficult. Genetic analysis of archaeological or museum samples therefore offers a more powerful approach to reconstruct population histories, than looking at fossils, archaeological finds or historical resources solely.

1.4 Genetics

In the following subchapters I give an overview about palaeogenetic research – the general fields of application and latest technological developments – as well as explain basics of genetics.

1.4.1 Palaeogenetics

Palaeogenetic research focuses on the reconstruction of archaeological or historical questions, using natural scientific methods (Herrmann 1994). The analyses of ancient DNA (aDNA), mostly from humans and animals, less often from plants, viruses or bacteria, adds a temporal aspect to evolutionary studies and therefore provides a direct view of past genetic structures. The spectrum of resources from which aDNA can be obtained includes skeletal material, mummified tissue and hair, as well as ice cores, preserved plant remains or coprolites (Hummel 1994).

History of research and fields of application

This discipline had its breakthrough during the 1980s, when Higuchi *et al*. (1984) isolated DNA for the first time from a 140 years old tissue sample of the now extinct *Quagga* (*Equus quagga*). A few years later, the invention of the polymerase chain reaction (PCR) (Saiki *et al*. 1985, Mullis *et al*. 1987) enabled large-scale analyses of ancient material, by amplifying even minimal amounts of DNA. After that, the number of publications about extracting DNA from various materials, e.g. in amber incased insects (DeSalle *et al*. 1992, Cano *et al*. 1993), plant remains from the Miocene (Golenberg *et al*. 1990) or dinosaur bones (Woodward *et al*. 1994), increased dramatically. The initial enthusiasm about the rapid development in this field was soon replaced by doubts about the authenticity of the DNA from samples that are millions of years old. These doubts and a better understanding about taphonomic processes of organisms (Lindahl 1993) have led to a more critical revision (Pääbo and Wilson 1991, Zischler 1995) and finally to a revocation of some

papers. After scientists became aware of possible misinterpretations of their results through contamination with modern DNA, they developed a set of measures to ensure the authenticity of ancient DNA data (Hofreiter *et al*. 2001, Burger *et al*. 2004, Pääbo *et al*. 2004).

Latest applications of aDNA techniques focus on a variety of evolutionary questions, including the exploration of extinct species and their relationship to extant taxa (e.g. Krings *et al*. 1997, Loreille *et al*. 2001, Hofreiter *et al*. 2002, Burger *et al*. 2004, Miller *et al*. 2009). Furthermore, palaeogenetic research has provided insights into the location and timing of domestication events of various animal species, like cattle (Troy *et al*. 2001, Bollongino *et al*. 2006), pigs (Larson *et al*. 2005, Larson *et al*. 2007), dogs (Larson *et al*. 2012) or goats (Fernández *et al*. 2006, Naderi *et al*. 2008). Domestication enabled transportation and exchange of animals, which in turn can be related to human migration and trade. Additionally, analyses of sediments, ice cores (Willerslev *et al*. 2007) or coprolites (Poinar *et al*. 1998) have been applied to reconstruct ecosystems and dietary habits of extinct animals. Ancient DNA approaches have been proved to be highly effective in assessing population histories and their varied responses to environmental changes, that would otherwise have gone undetected (Barnes *et al*. 2002, Shapiro *et al*. 2004).

Future research will emphasise especially on the development and enhancement of new ancient DNA examination methods. The focus already shifted from studying mitochondrial DNA to incorporating nuclear DNA, and the field of application has been extended by the development of next-generation-sequencing (NGS) (Hofreiter 2009). The new high-throughput sequencing techniques have revolutionised the field of molecular genetics, especially palaeogenetic studies, by increasing the amount of DNA sequence data available from archaeological and palaeontological remains by a multiple of times (Knapp & Hofreiter 2010, Rizzi *et al*. 2012). Varies projects have recently revealed the true potential of ancient DNA samples to study the process of evolution: shortly after the introduction of NGS, Poinar *et al*. (2006) published 13 million bp of the nuclear genome of the extinct woolly mammoth. Other studies followed, peaking in the presentation of a 1-fold coverage draft nuclear genome of

the Neanderthal (Green *et al*. 2010) and a high-quality 20-fold coverage nuclear genome of a 4,500 years old Palaeo-Eskimo (Rasmussen *et al*. 2010). Moreover, the change in technology starts to replace the PCR as a method to enrich the desired target sequence, using hybridisation capture instead (see chapter 1.4.5) (Hodges *et al*. 2007; Briggs *et al*. 2009; Gnirke *et al*. 2009; Krause *et al*. 2010; Burbano *et al*. 2010).

The technology and analytical methods are constantly evolving in an unprecedented pace, trying to compensate the limitations of next-generation-sequencing methods. Currently, a 'third generation' of sequencing techniques enables an even faster and more affordable generation of DNA data by sequencing single DNA molecules without prior amplification. Thereby, a bias in coverage depth and ultimately in polymorphism identification is avoided (see chapter 1.4.5). Furthermore, third-generation-platforms have the potential to increase the read length through full exploitation of the DNA polymerase activity, leading to more accurate *de novo* assemblies (Schadt *et al*. 2010).

Over the last decade palaeogenetic research has developed in a remarkable way from analysing only a few hundred base pairs of uniparental markers to generating complete genomes. So far, high-throughput sequencing focused on producing mitochondrial genomes of individuals, but will soon reach a genomic scale. Studying as many individuals as possible from ancient populations will give new insights in past population dynamics and will provide a more complete understanding of living systems. However, the innovative technologies go together with bioinformatics challenges and will demand new approaches to look at ancient DNA.

1.4.2 Structure and properties of ancient DNA (aDNA)

The term "ancient DNA" (aDNA) describes the retrieval of genetic material from archaeological or palaeontological finds, museum specimens and fossil remains (Pääbo *et al*. 2004). Deoxyribonucleic acid (DNA), located in the cell nucleus, is the carrier of the genetic information. The double stranded, helical polymer consists of a sugar-phosphate backbone and the attached nucleobases adenine (A), guanine (G), thymine (T) and cytosine (C), whereas A strictly pairs with T and C strictly pairs with G. The distinctive order of the bases defines the sequence of the DNA (Stachan & Read 1996).

The DNA constantly underlies chemical and physical processes, which can lead to a structural change or damage of the genome. Most frequent causes include hydrolysis, oxidation, methylation and alkylation. In living cells, the integrity of DNA molecules is constantly maintained by enzymatic repair and protection mechanisms (Lindahl 1993). After the death of an organism these mechanisms break down, which leads to degradation of the DNA through endogenous enzymes such as lysosomal nucleases. In addition bacteria, fungi, and insects contribute to the decomposition of the biomass (Eglinton & Logan 1991, Hummel 1994). *Post mortem* DNA damage physically destroys the genetic material, reducing the length and number of amplifiable DNA molecules. Therefore, amplification of the target sequence in several overlapping fragments with an average length of 100 to 200 base pairs (bp) is necessary to ensure authenticity of the results (Lindahl 1993, Pääbo *et al*. 2004).

The taphonomic process is not linear in time, but highly dependent on the environment and subsequent storage conditions. Certain factors like temperature, humidity, pH-value of the soil, as well as the number of microorganisms or the exposure to electromagnetic radiation are influencing the degradation of DNA (Burger *et al*. 1999). Constant cold climates, especially permafrost deposits from arctic and subarctic regions, facilitate long-term survival of DNA (Willerslev & Cooper 2005), whereas a warm environment promotes enzymatic and chemical activity. A dry preservation (e.g. mummification) reduces oxidative and hydrolytic processes, while humidity and moist soils have a negative effect on DNA conservation. An acidic environment on the one hand leads to the reduction of

hydroxylapatite, the protective mineral of the bone, on the other hand prevents infestation of microorganisms (Kaiser 2005).

Next to the quantitative reduction of the genetic material, chemical reactions can lead to DNA modifications that can either inhibit the extension of polymerase enzymes during the PCR or lead to erroneous base incorporation and therefore to misinterpretation of the generated data (Pääbo 1989). Two types of spontaneous chemical reactions commonly affect DNA *post mortem*: oxidative and hydrolytic damage. Oxidation causes base fragmentation of cytosine and thymine. The structural change results in complete strand breakage and in a consequence in an inhibition of the PCR reaction (Hofreiter *et al*. 2001, Pääbo *et al*. 2004). Structural modification through hydrolysis leads to deamination and depurination of the bases and therefore to incorrect base pairing. Most commonly observed miscoding lesions in ancient DNA are transitions, point mutations that change a pyrimidine to another pyrimidine (cytosine to thymine C→T and *vice versa*) or a purine to another purine (adenine to guanine A→G and *vice versa*). Transversions, substitutions of a purine for a pyrimidine and *vice versa*, occur less often. Furthermore, changes in the molecular sequence can involve insertion and deletion events (indels) (Lindahl 1993, Gilbert *et al*. 2003, Gilbert *et al*. 2007a). All these influences limit the verifiability and analysis of ancient DNA. Therefore working with aDNA requires precautionary measures to ensure the validity of the results. These measures will be discussed in the following chapter.

1.4.3 Anti-contamination measurements

Analysis of ancient DNA requires a number of methodological measures and specific behaviour to ensure the generation of authentic and meaningful data. Because of its degraded state, ancient DNA is prone to contamination with modern DNA, which is predominant in quality and quantity. The hazards especially involved with ancient human DNA are apparent, as the researcher himself represents a source of contamination within each step of analysis: in the phase of sample preparation, while extraction and PCR set-up, or during the amplification step. The following paragraph

describes strategies to avoid laboratory contamination and demonstrates criteria of authenticity in aDNA work (Pääbo *et al*. 2004, Gilbert *et al*. 2005).

Of importance is the physical separation of pre and post PCR laboratories with a strict policy of daily personnel movement only from pre to post PCR areas. All work steps with ancient DNA, including the PCR set-up, are carried out in the pre PCR facilities. The amplification reaction and sequencing preparation, as well as all work with modern material are conducted in the post PCR laboratory. Strict separation of these work areas is necessary to avoid carry-over contamination of already amplified target sequences. PCR reactions produce large concentrations of PCR products that can remain stable over long periods of time and once aerosolised, can easily spread throughout the laboratory (Burger 2007). Transportation of PCR products can be via an individual or any object that has been in post PCR vicinity (Willerslev & Cooper 2005).

Therefore strict hygiene guidelines are applied at the facilities at the Department of Archaeology at Durham University: prior to and after DNA work, extensive cleaning of the work areas and equipment with dilute sodium hypochlorite is performed. Furthermore exposure of equipment and reagents to ultra violet (UV) irradiation is carried out. Laboratory coats, overshoes and gloves are worn constantly and latter are changed regularly. Additionally, all surfaces and floors in every DNA laboratory are cleaned and disinfected once a week.

Sample preparation, extraction and PCR set-up are carried out in dedicated rooms within the ancient DNA facilities, partly undertaken under a fume hood. The hoods are equipped with a separate UV light and are radiated for four hours after every use. Next to the disinfection of facilities and equipment, blank controls are used in every work step on a standard basis in order to check the purity of the chemicals and to detect potential sources of contamination. Furthermore, primer design is a crucial step to avoid amplification of contaminant DNA. In the Department of Archaeology at Durham University we are solely working with ancient animal material. Primer pairs are designed to specifically amplify the desired species and especially not match human DNA. In order to increase the likelihood of

obtaining non-damaged DNA, primer pairs typically target overlapping fragments of an average size of 150 to 250 bp.

Generally, contamination can arise not only within the laboratory work, but also from the equipment, the reagents or the sample itself. Regarding the work on rodents in this thesis, it is important to note that mice and rats are a common presence in museum and storage facilities and collected samples, equipment or reagents may have been subject to contamination with exogenous DNA from the environment. Therefore particular attention has to be paid to the correctness of the results and whether they make phylogenetic sense. Thus far, experience has shown that the problem of contamination cannot be ruled out completely, as contamination sources are diverse. But the measures described in this chapter help to significantly reduce the risk of contamination and to detect irregularities within the data generation.

1.4.4 Structure and properties of mitochondrial DNA (mtDNA)

Besides the chromosomal DNA (nuclear DNA = ncDNA) in the nucleus, eukaryotic cells comprise another independent genome: the mitochondrial DNA (mtDNA). The mtDNA is located in the mitochondria, cytoplasmic organelles that partake in energy-providing processes of the cell. The number of mitochondria in a cell varies with the cell type between a few hundred up to several thousand, each containing two to ten copies of the mitochondrial genome (Jobling *et al*. 2004). Compared to the two copies of nuclear DNA per cell, the number of mitochondrial genomes is significantly higher. Therefore the probability of mtDNA preservation is increased, which is of great importance for the analysis of degraded ancient DNA.

In mammals, the mitochondrion is a circular double-stranded DNA molecule of about 16.5 kb in length and contains 37 genes that regulate the oxidative phosphorylation (see figure 1.2). In contrast to the ncDNA, the mitochondrial genome has an exceedingly compact structure with well-conserved coding areas (Jobling *et al*. 2004). Introns or transposable elements are entirely missing (Futuyma 1998). The only non-coding segment is the control region or displacement-loop (d-loop), start point for transcription and replication of the mtDNA. Two less

conservative regions are located within the d-loop, hypervariable region I and II (HVR I and II) (Bandelt *et al*. 2006). HVR I and II represent the most polymorphic sections of the mtDNA, due to a high and constant mutation rate up to ten times greater than in the nuclear genome. Reasons for the increased variability are less efficient protection and repair mechanisms that enhance susceptibility of mtDNA to oxidative damage, as well as a higher turnover rate, requiring more replications than nuclear DNA. Furthermore the lack of selective pressure leads to the accumulation and in consequence to the manifestation of mutations (Jobling *et al*. 2004). The coding region, and especially the cytochrome b (cyt b) gene, is commonly used to determine phylogenetic relationships between and within genera. Although this area is well-conserved, it contains sufficient variation to allow inter- and intraspecific analysis. The control region evolves faster and is therefore ideal to examine processes on population level in more recent times (Knoop & Müller 2006). Further important characteristics of the mitochondrial DNA are the uniparentally inheritance through females and the absence of recombination, implying that changes occurring in mitochondrial sequences are the result of mutations only (Jobling *et al*. 2004).

All features mentioned – high copy number and mutation rate, maternal inheritance and lack of recombination – make the mtDNA a particularly useful genetic tool. The constant occurrence of new sequence variants enables the reconstruction of individual lineages over generations and therefore helps to examine population genetic questions.

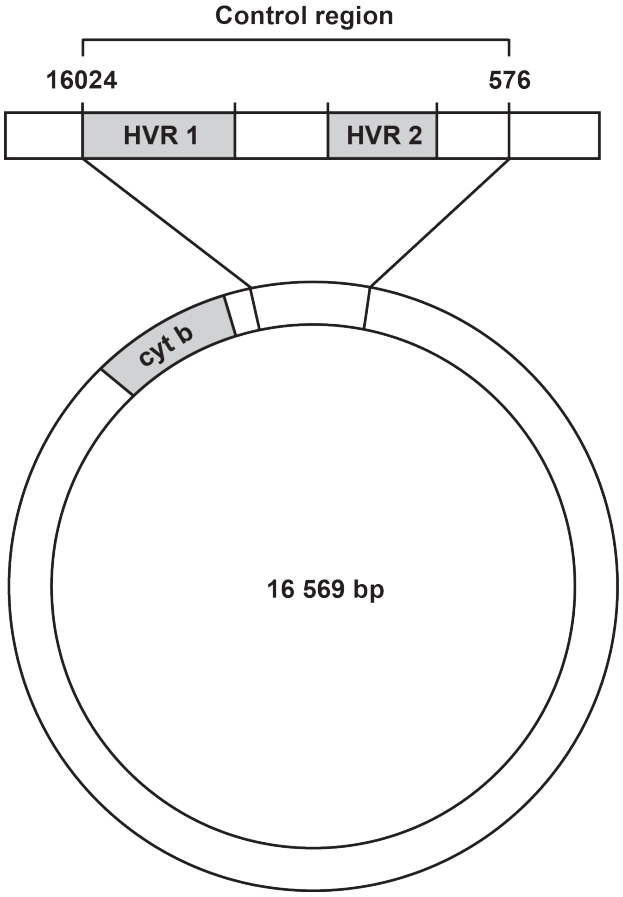

Figure 1.2: Schematic illustration of the human mitochondrial genome with focus on the marker used in this study (modified after Jobling et al. 2004).

1.4.5 Next-Generation-Sequencing (NGS)

As already mentioned in chapter 1.4.1 the development of high-throughput sequencing techniques has influenced the field of ancient DNA studies in a significant way. Since the beginning of palaeogenetic research in the 1980s, the methods to recover and analyse aDNA have been constantly improved to overcome the two main technical limitations: the poor preservation of endogenous DNA and the contamination with exogenous DNA (Rizzi *et al*. 2012).

So far, PCR and Sanger-sequencing were the main tools available to analyse ancient DNA samples, targeting particular for mtDNA regions (see chapter 1.4.4). Although NGS platforms have been commercially available since 2005, Sanger-sequencing still constitutes a large part in this thesis, wherefore I will briefly describe this method before explaining improved NGS techniques applied in chapter 5. Sanger-

sequencing was developed in 1977 (Sanger *et al*.) and is based on the selective incorporation of chain-terminating dideoxynucleotides by a DNA polymerase during DNA replication steps. After denaturation of the double stranded target fragments, the DNA polymerase, a primer (upper or lower), fluorescent dye labeled deoxynucleosidetriphosphates (dNTPs) and modified dideoxynucleosidetriphosphates (ddNTPs), which terminate DNA strand elongation by a lack of a 3'-OH group, are added to the reaction. Throughout the amplification step either a dNTP or a ddNTP is randomly incorporated in the new DNA strand, producing sequence fragments of different lengths. Assorting of the fluorescent dye labeled fragments by size and read order is ensured by automated capillary electrophoresis with a DNA-sequencing instrument (Applied Biosystems 3730 DNA Analyzer). Thereby, the fragments run through the capillary – small fragments faster than big fragments – and a laser detects the bases by their specific fluorescent dye labels at the end of each sequence (Bromham 2008). However, capillary sequencing techniques are inappropriate for large-scale sequencing – expensive and time-consuming – due to a low throughput of only 96 reads (~ 100,000 bp) at a time.

The development of next-generation-sequencing (NGS) methods has opened up new possibilities and extended the field of applications by greatly increasing the number of bases sequenced per run, while decreasing costs (Millar *et al*. 2008, Kircher *et al*. 2011, Rizzi *et al*. 2012). Most important within ancient DNA studies, the focus shifted from analysing small fragments of mitochondrial DNA over recovering full mitogenomes to generating whole nuclear genomes, allowing large-scale comparative and evolutionary studies of extant and extinct species. So far, research questions especially emphasised the relationship between Neanderthals and present-day humans. The first nuclear DNA sequence from a *Homo neanderthalensis* was reported by Green *et al*. (2006) as part of the Neanderthal Genome Project. Further papers followed, revealing that modern Eurasians share more genetic variants with Neanderthals than sub-Saharan Africans (Green 2010). In the same year another important study shed new light on the evolutionary history of the genus *Homo*. Krause *et al*. (2010) examined a phalanx of a juvenile hominin from the Denisova cave in southern Siberia, discovering a sequence distinct from both Neanderthals and

modern humans. With the possibility of reconstructing whole nuclear genomes, new research questions can be addressed, e.g. investigating ancient phenotypes or behavioural traits, providing important information about extinct hominins (Krause *et al*. 2007, Lalueza-Fox *et al*. 2007). Furthermore, NGS technology was applied to study domesticated animals (Ludwig *et al*. 2009, Lari *et al*. 2011), as well as biodiversity and extinction processes of ancient species (Allentoft *et al*. 2009, Miller *et al*. 2009, Ramirez *et al*. 2009).

At the moment there are five commonly used platforms available for massively parallel sequencing: Roche (454) GS FLX, Illumina (Solexa) Genome Analyzer, Applied Biosystems SOLiD, Ion Torrent (Life Technologies) and Helicos BioSciences HeliScope. The choice of instrument is mainly influenced by their advantages for particular applications and the quality of the DNA material (Metzker 2009; Knapp & Hofreiter 2010). Most commonly used NGS platforms for ancient DNA approaches are Roche (454) GS FLX and Illumina (Solexa) Genome Analyzer. Both technologies provide sufficient read length for fragmented aDNA, but differ in their amplification and sequencing steps due to their platform-specific sequencing chemistry, resulting in different throughputs. Whereas the Roche (454) GS FLX produces about 400 to 600 million base pairs (megabases = mb) per run with read lengths of up to 400 bp, the Illumina (Solexa) Genome Analyzer generates up to 600 gigabases (gb) of 2 x 100 bp (paired end) reads. Both platforms provide a sequencing plate with separate lanes – 16 and 8 respectively, thereby allowing for sequencing multiple libraries within in a single run (Knapp & Hofreiter 2010).

Generally, the generation of reads with either platform requires three key steps: library preparation (platform-specific modification of the DNA fragments), the actual sequencing reaction (library amplification via immobilisation on a flowcell or in-solution), followed by data analysis (Metzker 2009, Rizzi *et al*. 2012). All steps will be further explained in the following sections.

Library preparation

In order to ensure a non-biased source of template material, pre-modification of the DNA is necessary. Firstly, molecules need to be cut to ideal length between 100 to 1,000 bp, according to the different sequencing platforms. As ancient DNA is already present in a fragmented state, shearing is usually not required. Afterwards, the double-stranded DNA fragments are trimmed at both sides by enzymes and converted into blunt ends. Then universal adapters that attach to the flowcell and contain priming sites for subsequent amplification and sequencing steps are ligated to the blunt ends (Knapp & Hofreiter 2010, Rizzi *et al*. 2012). Species-specific primer used in traditional Sanger sequencing are no longer required, which allows sequencing of every single fragment present in the DNA extract (Fortes & Paijmans accepted). Usually, the content of endogenous DNA is < 1 – 2% (Green *et al*. 2010), in rare cases can reach up to 50%, depending on the preservation conditions (Poinar *et al*. 2006) or the sample type (Gilbert *et al*. 2007b). Therefore, different methods have been developed to enrich DNA-libraries for the desired target, the so called 'hybridisation capture' (Summerer 2009, Mamanova *et al*. 2010). These methods work with bait molecules, which have high sequence similarity to the actual target DNA. After hybridisation of bait and target and subsequent immobilisation of both, the residual fragments are washed away. Besides the increase of the desired target DNA, hybridisation capture methods favour particularly small fragments, thereby introducing a bias against modern contaminants (Hodges *et al*. 2007, Avila-Arcos *et al*. 2011, Burbano *et al*. 2012).

In general, library preparation protocols involve several enzymatic reactions that include purification steps inbetween. These purification steps though cause significant loss of predominantly the shortest fragments, which is disadvantageous especially in ancient DNA studies. Recently published papers have improved library preparation protocols, mainly focusing on the preservation of small fragments by replacing the previous purification steps with heat inactivation of the enzymes (see figure 1.3) (Gansauge & Meyer 2013, Fortes & Paijmans accepted). After the

libraries are built, multiple samples can be pooled in a single lane of a flowcell by inserting sample specific sequence tags (‘barcode’ or ‘index’) in the library adapters.

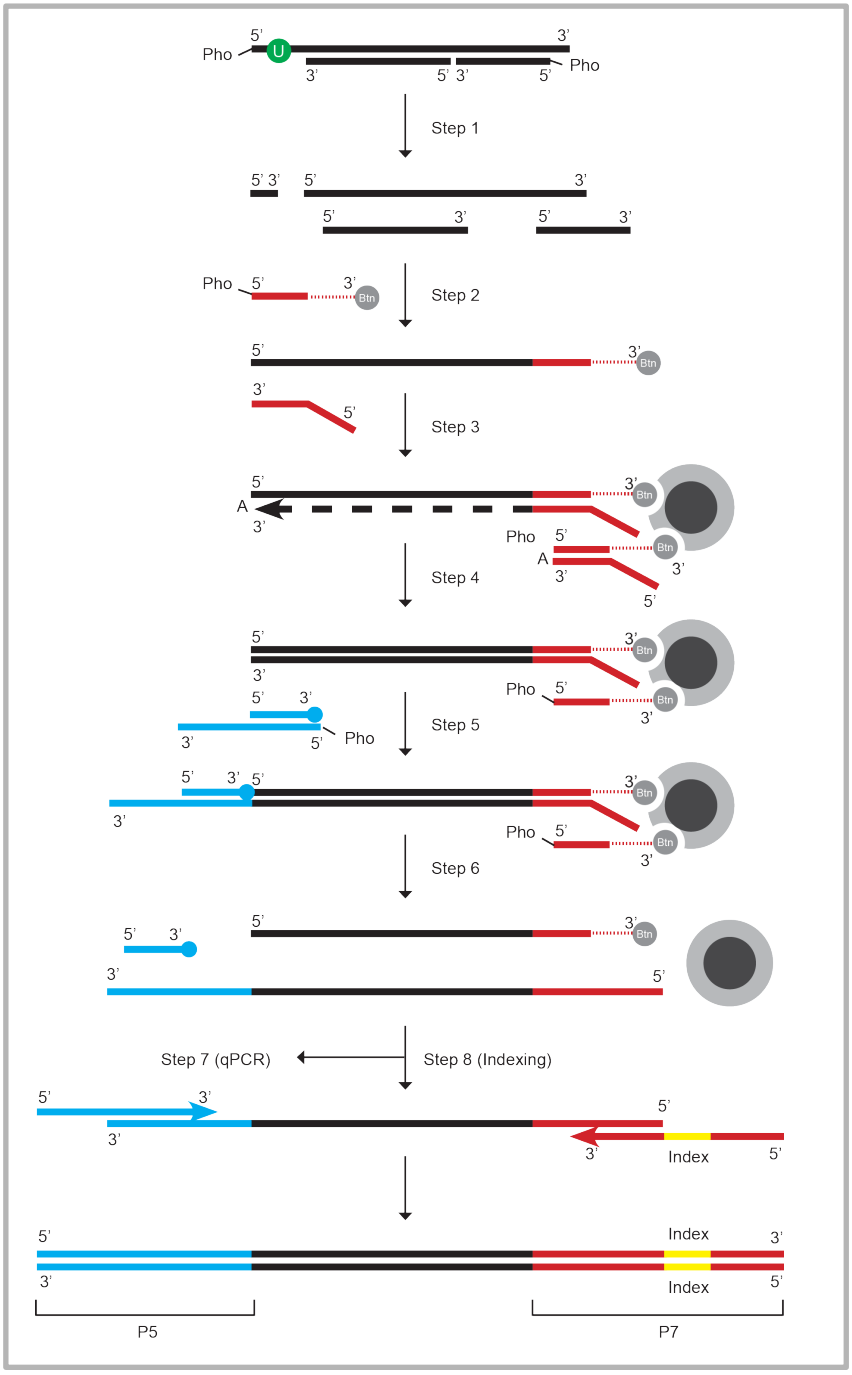

Figure 1.3: Schematic overview of single-stranded library preparation (modified after Gansauge & Meyer 2013).

(Figure 1.3 continued)
Exemplified by one strand of a double-stranded DNA molecule containing a uracil and a single-strand break. Uracil (U), Phosphate (Phos), Biotin (Btn).
Step 1: Uracil excision, dephosphorylation and heat denaturation.
Step 2-3: Ligation of first adapter, immobilisation od ligation products on beads.
Step 4: Primer annealing and extension.
Step 5-6: Blunt-end repair, ligation of second adapter and library elution.
Step 7: qPCR in order to determine cycle number in indexing PCR.
Step 8: Indexing PCR in order to incorporate sample specific barcodes.

Sequencing reaction

Most NGS imaging systems have not been designed to detect single fluorescent events. Therefore, amplification of the library is necessary in order to generate millions of DNA molecules for required signal intensity. Amplification and sequencing steps differ among platforms, depending on their sequencing chemistry (Metzker 2009, Rizzi *et al*. 2012). While Roche/454 uses an emulsion PCR (emPCR) with subsequent pyrosequencing, Illumina multiplies the target DNA via bridge PCR and detects bases through sequencing-by-synthesis.

During an emPCR, single DNA molecules are attached to beads through oligonucleotides on their surfaces that are complementary to the 454-specific adapters. Each bead is associated with a single DNA fragment, amplifying it in an aqueous microdroplet, isolated by oil. The individual water-oil micelles contain PCR reactants, generating thousands of homogeneous template molecules through thermal cycles. In order to provide a fixed location in which the sequencing reaction can be monitored, the millions of beads are loaded onto a Roche/454 picotiter plate, each well containing a single bead. Simultaneous imaging of all beads is ensured by pyrosequencing technology. Thereby, the incorporation of each nucleotide by a DNA polymerase results in the release of pyrophosphate, which initiates a series of enzymatic reactions that ultimately produces a light signal. The intensity of the light is proportional to the number of nucleotides incorporated and recorded by a CCD camera seated opposite the picotiter plate (Mardis 2008, Glenn 2011).

The Illumina system (see figure 1.4 to 1.7) uses a bridge PCR amplification to obtain multiple copies of the desired target sequences. Thus, individual sequencing templates are immobilised randomly on a solid glass surface (flowcell) and amplified into identical clusters by adding fluorescent-labelled nucleotides along with the DNA polymerase (Rizzi *et al*. 2012). The base-unique fluorescent label serves as a terminator for the polymerisation, making each incorporation a single event. After the base is identified and imaged, the terminators are enzymatically removed to allow incorporation of the next nucleotide (Mardis 2008, Glenn 2011, Rizzi *et al*. 2012). This step is important, as precise library quantification is crucial to obtain high quality and non-biased DNA sequences for further analysis (Schadt *et al*. 2010).

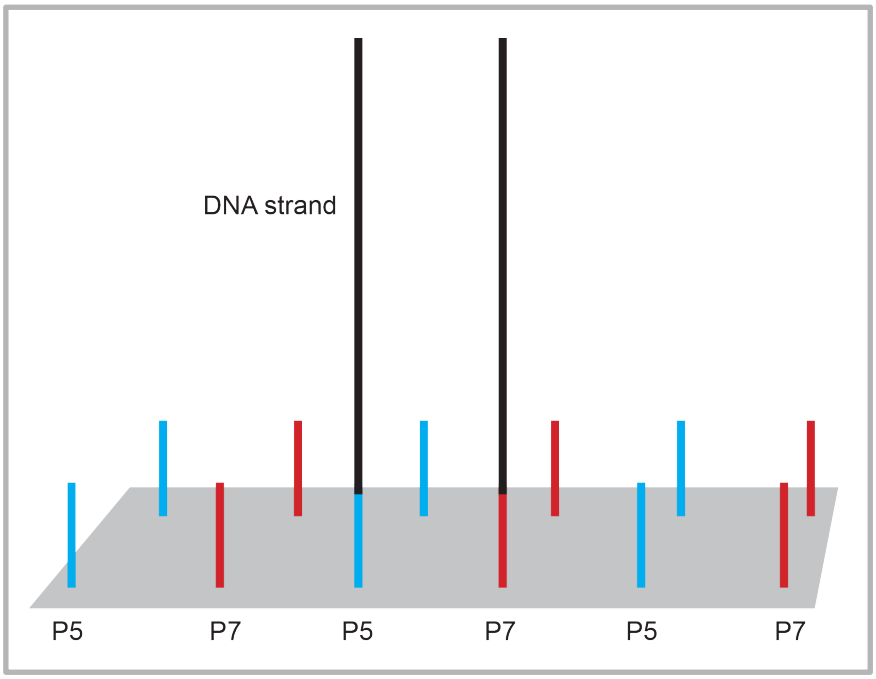

Figure 1.4: Schematic overview of next-generation-sequencing reaction: Hybridisation of the DNA to the flowcell.
Bound libraries are extended by polymerases. The double-stranded molecule is denatured and the original template is washed away. The newly synthesised strand is attached to the flowcell surface.

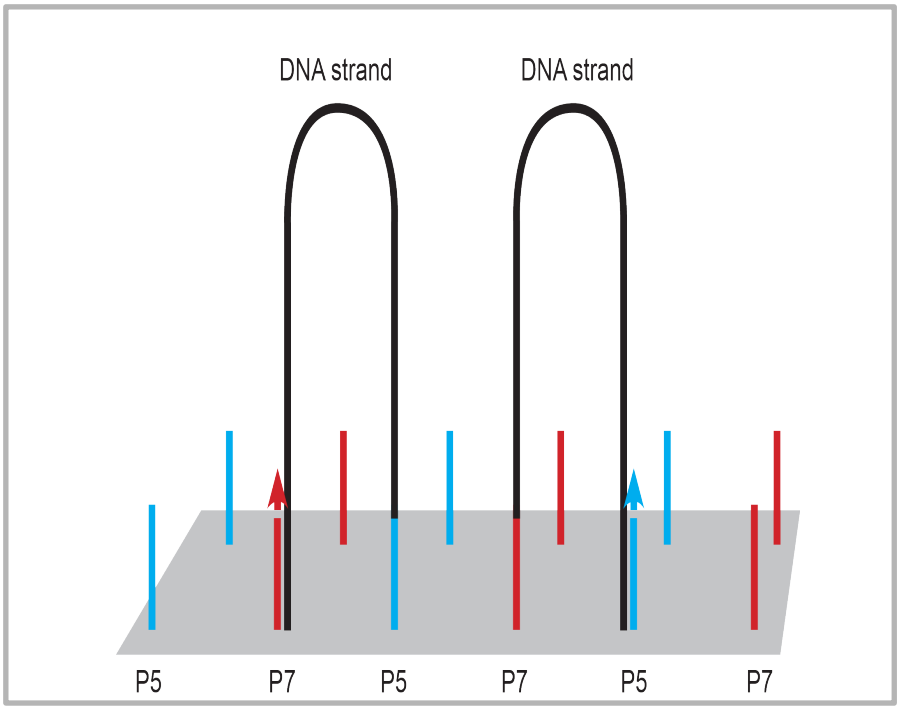

Figure 1.5: Schematic overview of next-generation-sequencing reaction: Bridge amplification.
The single-stranded molecule flips over and forms a bridge by hybridising to a complementary primer. Hybridised primer is extended by polymerases. The double-stranded bridge is denatured. Bridge amplification cycle is repeated until multiple copies of each DNA fragment are generated ('cluster').

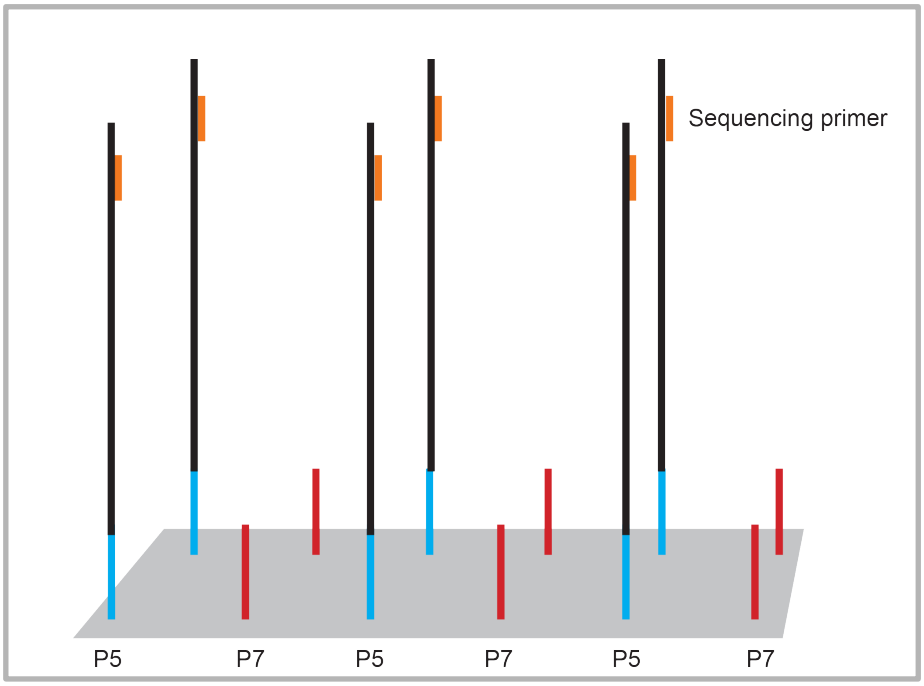

Figure 1.6: Schematic overview of next-generation-sequencing reaction: Sequencing Primer Hybridisation.
The sequencing primer is hybridised to the adapter sequence.

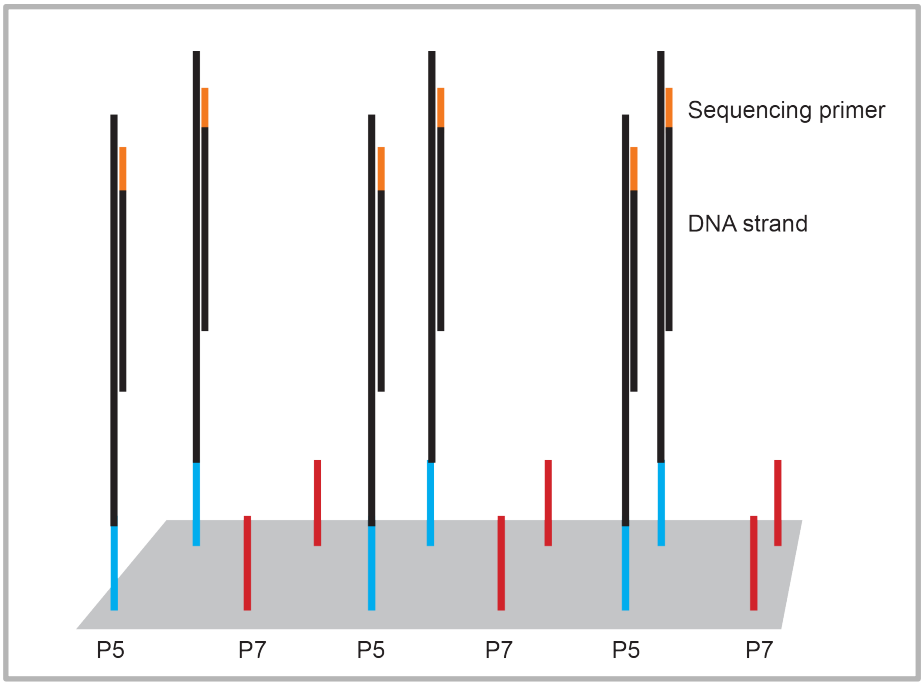

Figure 1.7: Schematic overview of next-generation-sequencing reaction: Extension and detection of incorporated bases.
Hybridised primer is extended by polymerases. Detection of the individually incorporated bases by light signals.

Data analysis

Although the advantages of the new sequencing technologies are obvious, there are specific characteristics and limitations, that must be taken into account during data analysis (Kircher *et al*. 2011). After library preparation, amplification and subsequent registration of the DNA sequence ('base calling'), the raw data needs further editing and verification.

The generated reads are displayed in a specific format (FASTQ) with an additional phred encoded output line to measure the quality of each base, i.e. the probability of the correct call of each base. A higher sequencing score thereby indicates a smaller probability of error (Cock *et al*. 2010). Prior to further analysis, sequences of a multiplex set-up are sorted into their libraries by identification of the sample-specific index or barcode. Afterwards, the specific adapters, necessary for the amplification step, are detached from the target DNA. According to the quality score, sequences of low value are removed, before the remaining sequences are mapped

and merged to a whole genome (Falgueras et al. 2010, Schmieder & Edwards 2011, Lindgreen 2012).

Depending on the sequencing strategy, the generated reads are then either aligned to a reference genome (re-sequencing) or assembled *de novo*. *De novo*-sequencing requires the preparation of several libraries with different fragment sizes and assembly through the use of algorithms that can reconstruct a previously unknown sequence. Furthermore a *de novo*-sequence demands a higher coverage compared to a re-sequencing approach to ensure informative and robust results (Rizzi *et al*. 2012). To acquire the correct coverage, clonal reads are collapsed and sequences with multiple hits to the reference genome are also removed. Limitations within the alignment process occur when reads are placed in repetitive regions or when gaps are present in the reference genome due to structural changes (Metzker 2009).

Another factor of uncertainty within the data analysis occurs through possible contamination during library preparation. Low amounts of starting material and single-stranded ligation processes make the template DNA prone to contamination with exogenous DNA. Contamination sources are diverse: impurities can originate from lab chemicals or the equipment, as well as from the experimenter or the sample itself (Kircher *et al*. 2011). Different criteria have been applied to detect the presence of modern DNA in ancient samples. Due to the degraded nature of aDNA, short fragments and a higher frequency of miscoding lesions are the main features to distinguish between ancient and modern DNA (Rizzi *et al*. 2012). Furthermore, sites of known fixed variations between species or populations help to detect contaminant molecules, especially when studying humans (Burbano *et al*. 2010, Green *et al*. 2010, Krause *et al*. 2010). If no informative sites are available, e.g. with a *de novo* approach, exploiting the sex of the sample can be useful. Y-chromosomal fragments in female samples or X-chromosomal heterozygosity in males are an indication for contamination issues (Green *et al*. 2009, Reich *et al*. 2010). Moreover, cross-contamination can appear during the preparation of the sequencing run, but can be easily identified and filtered through the sample-specific barcodes in the libraries (Mamanova *et al*. 2010, Meyer & Kircher 2010).

Sequencing approaches

Within the new high-throughput sequencing techniques different approaches have been applied to target for distinctive types of DNA. Depending on the project goal and available material we distinguish between shotgun sequencing, amplicon sequencing and sequence capture.

Shotgun sequencing represents the simplest sequencing strategy and is characterised by random targeting for DNA without a prior selection. Afterwards, the reads are identified through comparison to established sequence databases. Shotgun sequencing is commonly used in studies that aim to identify all organisms present in an isolated specimen or in areas where no DNA research has been carried out so far. Furthermore, this approach can provide valuable information about the quality and quantity of endogenous DNA. Disadvantages of this method are the commonly low amount of endogenous DNA due to the abundance of bacterial and fungal contamination. Also a high sequencing depth is required within shotgun sequencing, to ensure the generation of meaningful results. Therefore this strategy is most suitable for well-preserved samples, e.g. from permafrost preserved specimens (Knapp & Hofreiter 2010, Rizzi *et al*. 2012). Hence, less well-preserved samples, which are the majority in ancient DNA studies, require target enrichment strategies before sequencing. Target enrichment is used in amplicon sequencing or sequence capture and achieved through two methods – the widely used PCR amplification and more recently developed DNA capture via hybridisation (Knapp & Hofreiter 2010).

Amplicon sequencing is based on a PCR step prior to sequencing in order to target for a specific region within the genome. This approach is useful when aiming for SNPs or haplotype markers in a well-known DNA segment. Normally long fragments are amplified in PCR target enrichment and then sheared to suitable sizes for the NGS instrument used. Ancient DNA however is usually highly degraded, whereas numerous independent PCR amplifications of overlapping fragments are necessary. Thereby the amount of available sample extract can be a limiting factor within this approach (Knapp & Hofreiter 2010, Rizzi *et al*. 2012).

Sequence capture methods use specific probes, either immobilised on microarrays or beads, to recognise and capture target DNA. The genomic regions to be analysed are defined by target specific, biotinylated primers hybridised to the DNA fragments. Once the library is connected to the probes, non-target DNA molecules are removed and an extension step is performed to amplify the specific fragments. Subsequently this library is then sequenced with NGS. Despite the additional library amplification step, sequence capture overcomes some of the limitations of PCR enrichment in aDNA studies: the much shorter fragment length required enables the amplification of highly degraded molecules and therefore also reduces the risk of enriching modern contamination, especially when studying humans. The main disadvantage of sequence capture is the loss of template material during library preparation (Knapp & Hofreiter 2010). In order to overcome this problem, efforts have been made to improve library preparation protocols order to increase original template recovery and prevent loss of library complexity (see 'library preparation').

1.5 Analyses

In the following subchapters I explain the methodological approaches integral to the chapters 2, 3, 4, and 5 of this thesis.

1.5.1 Model of nucleotide sequence evolution

The reconstruction of evolutionary relationships between taxa, based on similarities and differences in their DNA sequence data, is defined as 'phylogenetics'. Thereby it is assumed that closely related organisms share more similarities in their genetic characteristics, than further distant related taxa. A way to visualise the evolutionary history of organisms is a phylogenetic tree, using maximum likelihood (ML), neighbour joining (NJ) or a Bayesian approach (see chapter 1.5.2). However, the correct application of these methods requires a suitable model of sequence evolution that adequately captures the underlying evolutionary process. Various methods have been developed to select models under specific criteria, which will be explained in the following paragraph.

The probability of a nucleotide substitution is dependent on two factors: the time that has passed and the rate at which mutations occur, as they vary within the genome. Furthermore, as already mentioned in chapter 1.4.2, nucleotide substitutions can be distinguished in transitions, transversions or indels, whereas transitional changes are most frequent. As genetic distances increase, the probability of multiple substitutions per site grows. Therefore, the difference between observed and effective number of changes within a molecular sequence at a particular time needs to be ascertained.

Models of nucleotide sequence evolution can be used to assess and correct the miscalculation. The several models differ in their complexity, through the adjustment of different parameters, which take the various rates of mutations into account. Further model parameters consider the variation of mutation rates within a dataset. 'Gamma distribution' (+Γ) thereby assesses site-specific variable rates, 'proportion of invariable sites' (+I) estimates the quantity of unchanging sites (Knoop & Müller 2006). Increasing the number of parameters provides a more

precise estimate of the actual number of substitutions and consequently increases the fit between the model and the dataset. Conversely, adding parameters increases the variance, which can lead to models 'over-fitting' the data. Hence, substitution models must be selected in a way to balance the bias of either being under- or overparameterised.

Different methods have been developed to evaluate the significance of adding certain parameters, e.g. the hierarchical likelihood-ratio test (hLRT), the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC). While hLRT repeats pairwise comparisons between nested models until the simplest model at a given significance level is reached, AIC and BIC measure the fit of the model based on a likelihood function (log likelihood) and add a penalty for each additional parameter (Ripplinger & Sullivan 2008).

The models of nucleotide sequence evolution used to reconstruct phylogenetic trees in chapters 2, 3, 4, and 5 of this thesis have been generated by the 'model selection' option in the programme MEGA-CC (Kumar *et al*. 2012). MEGA-CC uses the Bayesian Information Criterion (BIC) as a measure for model selection among a limited set of models. Best-fit models for the given datasets included:

K2 (Kimura-2-parameter)

Differentiates between transitions and transversions, but assumes all nucleotides occur with the same frequency.

T92 (Tamura-3-parameter)

Extension of Kimura's 2-parameter model by assuming a G+C-content bias exists.

HKY (Hasegawa-Kishino-Yano)

Assumes different nucleotide frequencies and considers variable transition and transversion rates.

GTR (General Time Reversible)

GTR is the most general model, as it assumes individual base frequencies. Except for the reversible direction of substitutions, i.e. A→G = G→A. Each pair of nucleotide substitution is given a different mutation rate.

1.5.2 Bayesian trees

As already mentioned in the previous chapter 1.5.1, the construction of phylogenetic trees is a powerful tool to reveal the evolutionary history of different taxa. In this thesis a Bayesian statistical approach is applied to address various evolutionary questions. Compared to other standard methods like maximum likelihood (ML) or neighbour joining (NJ), Bayesian allows determination of the probability of the phylogeny. Thereby, the probability of a parameter, composed of the tree topology, branch lengths, and a model of nucleotide sequence evolution, is generated by the given sequence data. As the analytical calculation of the entire range of possible tree topologies and various substitution models is computationally intense, Bayesian uses the MCMCMC-algorithm (Metropolis-coupled Markov chain Monte Carlo) to find the tree with the highest posterior probability according to the dataset.

In principle, the MCMCMC-algorithm randomly (Monte Carlo) explores a parameter space of multiple peaks – areas of trees with a high posterior probability, and in-between valleys – areas of trees with a low posterior probability. Each step involves a modification of the tree topology, the branch length or the substitution model, whereas the new tree is either accepted or rejected, according to the value of the posterior probability (Markov chain). A practical problem of this process is that the Markov chain may have difficulties in moving from one peak to another and may get stuck. Therefore multiple chains are run in parallel (Metropolis-coupled), each starting from a different and randomly selected point. The chains are partly 'heated' to move quicker and cover more parameter space. These heated chains are able to swap with the single cold chain, although only the cold chain is actually sampled.

After a sufficient number of generations, the likelihood values reach a plateau. In order to calculate the consensus tree, only trees within the plateau are incorporated. Hence, suboptimal trees with low posterior probabilities in the beginning of the chain ('burn-in') have to be discarded. Trees are sampled in fixed intervals ('sampling frequency') according to the parameters defined by the user. The consensus tree includes information of the branch lengths, as well as the clade support values, which are estimated by the proportion of sampled trees that have been recorded by the cold chain (Kelly 2005, Knoop & Müller 2006).

Bayesian inference was performed using MrBayes version 3.2.2 (Huelsenbeck & Ronquist 2001) and BEAST version 1.8.0 (Drummond & Rambaut 2007), employing the particular substitution models generated by 'model selection' in MEGA-CC (Kumar *et al*. 2012). Details about the specific parameter, generation times and sampling frequency, will be annotated in the individual chapters.

1.5.3 SPREAD

The use of analytical tools that consider the geographic context of molecular sequences provide insights into the origin and spread of species and help to reconstruct their phylogeographic history. SPREAD ('Spatial Phylogenetic Reconstruction of Evolutionary Dynamics') is a programme that incorporates geographical information into molecular phylogenetic methods in order to visualise the spatial-temporal distribution of Bayesian trees. Thereby, the generated SPREAD output file is uploaded in the virtual globe software *Google Earth* (http://earth.google.com), where the diffusion of species through space and time is animated. For the case studies in this thesis, the SPREAD template 'time slicer' has been used. Time slicer visualises the continuous diffusion by summarising "the rate and degree of geographical movement over the complete posterior distribution of trees" (Bielejec *et al*. 2011). First, the tree is divided into time intervals. Then, undetected ancestral locations along each branch that cross the different time intervals are calculated. In order to account for uncertainties among the branch-specific variation in dispersal rates, Brownian bridges have been incorporated in relaxed random walks (RRW) (Lemey *et al*. 2010). Afterwards, the distribution of all imputed locations is plotted by the use of bivariate kernel density estimates (Bielejec *et al*. 2010).

The phylogeographic history of the species in question is then visible by uploading the SPREAD output file in a virtual globe software like *Google Earth*, where the expansion in space and time is simulated. Thereby, the red branch indicates the starting point of the coalescence tree. From there, bifurcations of the internal tree nodes spread across the estimated geographical range. These nodes are

surrounded by light blue polygons, that represent their confidence intervals. The bifurcations end in the external nodes of the tree, i.e. the sampled specimens.

1.6 References

Literature

Addison DJ & Matisoo-Smith E (2010) Rethinking Polynesians origins: a West-Polynesia triple-I model. *Archaeol. Oceania* **45**, 1-12.

Adelaar A (1995) Asian roots of the Malagasy: a linguistic perspective. *Bijdragen tot de Taal-Land en Volkenkunde* **151**, 325-356.

Adelaar A (2009) Towards an integrated theory about the Indonesian migrations to Madagascar. In: P. Pergine *et al*. (eds.) Ancient Human Migrations. University of Utah Press, 149-172.

Allentoft M, Schuster SC *et al*. (2009) Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data. *Biotechniques* **46**, 195-200.

Allibert C (2008) Austronesian migration and the establishment of the Malagasy civilization: Contrasted readings in linguistics, archaeology, genetics and cultural anthropology. *Diogenes* **218**, 7-16.

Anderson A (2009) The rat and the octopus: initial human colonization and the prehistoric introduction of domestic animals to Remote Oceania. *Oceania*, 1503-1519.

Armitage PL, West B, Steedman K (1984) New evidence of black rat in Roman London. *The London Archaeologist* **4**, 375-383.

Asouti E, Fuller DQ (2008) Trees and woodlands of South India. Archaeological perspectives. Walnut Creek (CA), Left Coast Press.

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.

Auffray JC & Britton-Davidian J (2012) The house mouse and its relatives: systematics and taxonomy. In: Macholan M *et al*. (eds.) Evolution of the house mouse. Cambridge University Press, 1-34.

Avila-Arcos MC, Cappellini E, Romero-Navarro JA *et al*. (2011) Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Sci Rep* **1**, 74.

Avise JC, Arnold J *et al*. (1987) Intraspecic phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu Rev Ecol Syst* **18**, 489-522.

Avise JC (2000) Phylogeography: the History and Formation of Species. Harvard University Press: MA, USA.

Avise, JC (2009) Phylogeography: retrospect and prospect. *Journal of Biogeography*, **36**, 3-15.

Barnes I, Matheus P, Shapiro B, Jensen D, Cooper A (2002) Dynamics of Pleistocene Population Extinctions in Beringian Brown Bears. *Science* **295**, 2267-2270.

Bandelt HJ, Macaulay V, Richards M (2006) Human Mitochondrial DNA and the Evolution of Homo sapiens. Springer, Berlin, Heidelberg.

Beaujard P (2005) The Indian Ocean in Eurasian and African World-Systems before the Sixteenth Century. *Journal of World History* **16 (4)**, 411-465.

Beaujard P (2011) The first migrants to Madagascar and their introduction of plants: Linguistic and ethnological evidence. *Azania: Archaeological Research in Africa* **46**, 169-189.

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: I. Glover, P. Bellwood (eds.) Southeast Asia: from prehistory to history. Routledge Curzon, London, 68-88.

Bellina-Pryce B, Silapanth P (2008) Weaving cultural identities on trans-Asiatic networks: Upper Thai-Malay Peninsula – an early socio-political landscape. *Bullentin de l'Ecole Francaise d'Extreme-Orient* **93**, 257-293.

Bielejec F, Rambaut A, Suchard MA, Lemey P (2011) SPREAD: Spatial Phylogenetic Reconstruction of Evolutionary Dynamics. *Bioinformatics* **27**, 2910-2912. SPREAD version 1.0.6, available from www.kuleuven.ac.be/aidslab/phylogeography/SPREAD.html

Blench R (1996) The ethnographic evidence for long-distance contacts between Oceania and East Africa. In: J. Reade (ed.) The Indian Ocean in antiquity. Kegan Paul, London, 417-438.

Blench R (1997) A history of agriculture in northeastern Nigeria. In: D. Barreteau, R. Dognin, C. von Graffenried (eds.) L'homme et le milieu vegetal dans le Bassin du Lac Tchad. Orstom, Paris, 69-112.

Blench R (2007) New palaeozoogeographical evidence for the settlement of Madagascar. *Azania: Archaeological Research in Africa* **42**, 71-82.

Blench RM (2009) Bananas and plantains in Africa: re-interpreing the linguistic evidence. *Ethnobotanical Research and Applications* **7**, 363-380.

Boivin N, Fuller DQ (2009) Shell middens, ships and seeds: exploring coastal subsistence, maritime trade and the dispersal of domesticates in and around the ancient Arabian Peninsula. *Journal of World Prehistory* **22**, 113-180.

Boivin N, Fuller DQ, Crowther A (2012) Old world globalization and the Columbian exchange: Comparison and contrast. *Antiquity* **44 (3)**, 452-469.

Boivin N, Crowther A, Helm R, Fuller D (2013) East Africa and Madagascar in the Indian Ocean World. *Journal of World Prehistory* **26** (3), 213-281.

Boivin N, Crowther A, Prendergast M, Fuller DQ (2014) Indian Ocean Food Globalisation and Africa. *African Archaeological Review* **31 (4)**, 547-581.

Bonhomme F & Searle JB (2012) House mouse phylogeography. In: M. Macholán *et al*. (eds.) Evolution of the House Mouse. Cambridge University Press, 278-296.

Bollongino R, Edwards CJ, Alt KW, Burger J, Bradley DG (2006) Early history of European domestic cattle as revealed by ancient DNA. *Biology Letters* **2**, 155-159.

Boursot P, Auffray JC, Britton-Davidian J, Bonhomme F (1993) The Evolution of House Mice. *Annu. Rev. Ecol. Syst.* **24**, 119-152.

Braithwaite R (1980) The ecology of *Rattus lutreolus* III. The rise and fall of a commensal population. *Australian Wildlife Research* **7**, 199-215.

Breen C, Lane P (2003) Archaeological approaches to East Africa's changing seascapes. *World Archaeology* **35 (3)**, 469-489.

Briggs AW, Good JM, *et al*. (2009) Targeted Retrieval and Analysis of Five Neandertal mtDNA Genomes. *Science* **325**, 318-321.

Bromham L (2008) Reading the story in DNA. Oxford University Press (New York).

Burbano HA, Hodges E, *et al*. (2010) Targeted Investigation of the Neandertal Genome by Array-Based Sequence Capture. *Science* **328**, 723-725.

Burbano HA, Green RE, Maricic T *et al*. (2012) Analysis of Human Accelerated DNA Regions Using Archaic Hominin Genomes. *PloS One* **7**, 32877.

Burger J, Hummel S, Herrmann B, Henke W (1999) DNA preservation: A microsatellite-DNA study on ancient skeletal remains. *Electrophoresis* **20**, 1722-1728.

Burger J, Rosendahl W *et al*. (2004) Molecular Phylogeny of the extinct cave lion Panthera leo spelaea. *Molecular Phylogenetics and Evolution* **30**, 841-849.

Burger J (2007) Palaeogenetik. In: G. A. Wagner (ed.) Einführung in die Archäometrie. Springer, Berlin, 279-298.

Cano RJ, Poinar HN, Pieniazek NJ, Acra A, Poinar jr. GO (1993) Amplification and sequencing of DNA from a 120-135-million-year-old weevil. *Nature* **363**, 536-538.

Case TJ, Bolger DT, Petren K (1994) Invasions and Competitive Displacement Among House Geckos in the Tropical Pacific. *Ecology* **75**, 464-477.

Casson L (1989) The Perilplus Maris Erythraei: Text with introduction, translation and commentary. Princeton University Press, Princeton.

Castillo C, Fuller DQ (2010) Still too fragmentary and dependent upon chance? Advances in the study of early Southeast Asian archaeobotany. In: B. Bellina, E. Bacus, O. Pryce, J. Wisseman-Christie (eds.) Studies in honour of Ian Glover. River Books, Bangkok, 91-111.

Cheke A (2011) Was the flying fox *Pteropus comorensis* on Mafia Island (Tanzania) introduced by humans? *Journal of East African Natural History* **100**, 59-68.

Cock PJ, Fields CJ *et al*. (2010) The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res* **38**, 1767-1771.

Cole NC (2005) The ecological impact of the invasive house gecko *Hemidactylus frenatus* upon endemic Mauritian geckos. Dissertation, Bristol.

Csurhes S & Markula A (2009) Pest animal risk assessment: Asian house gecko *Hemidactylus frenatus*. Biosecurity Queensland, Queensland Primary Industries and Fisheries, Department of Employment, Economic Development and Innovation (Queensland, Australia).

Cucchi T, Vigne JD, Auffray JC (2005) First occurrence of the house mouse (Mus musculus domesticus Schwarz and Schwarz, 1943) in the Western Mediterranean: a zooarchaeological revision of subfossil occurrences. *Biological Journal of the Linnean Society* **84**, 429-454.

Cucchi T, Auffray JC, Vigne JD (2012) On the origin of the house mouse synanthropy and dispersal in the Near East and Europe: zooarchaeological review and perspectives. In: S. J. E. Baird *et al*. (eds.) Evolution of the House Mouse. Cambridge University Press, 65-93.

Dahl OC (1951) Malgache et Maanyan: Une comparaison linguistique. Egede Instituett, Oslo.

Dahl OC (1988) Bantu substratum in Malagasy. *Etudes Océan Indien* **9**, 91-132.

De Langhe E (2007) The establishment od traditional plantain cultivation in the African rain forest: a working hypothesis. In: T. Denham, J. Iriarte, L. Vrydaghs (eds.) Rethinking agriculture: archaeological and ethnoarchaeological perspectives. Walnut Creek (CA), Left Coast Press, 361-370.

Denny M & Matisoo-Smith L (2010) Rethinking Polynesian Origins: Human Settlement of the Pacific. LENScience Senior Biology Seminar Series, Liggins Institute.

DeSalle R, Gatesy J, Wheeler W, Grimaldi D (1992) DNA sequences from a fossil termite in Oligo-Miocene amber and their phylogenetic implications. *Science* **257**, 1933-1936.

Dewar RE, Wright HT (1993) The cultural history of Madagascar. *Journal of World Prehistory* **7**, 417-466.

Dubut V, Cartault F, Payet C *et al*. (2009) Complete mitochondrial sequences for haplogroups M23 and M46: Insights into the Asian ancestry of the Malagasy population. *Human Biology* **81 (4)**, 495-500.

Duplantier JM, 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.

Eglinton G, Logan GA (1991) Molecular preservation. *Philos. Trans. R. Soc. London Ser. B* **333**, 315–27.

Ervynck A (2002) Sedentism or urbanism? On the origin of the commensal black rat (Rattus rattus). *Bones and the man. Studies in honour of Don Brothwell,* 95-109.

Falgueras J, Lara AJ et al. (2010) SeqTrim: a high-throughput pipeline for pre-processing any type of sequence read. *BMC Bioinformatics* **11**, 1-12.

Fernández H, Hughes S *et al*. (2006) Divergent mtDNA lineages of goats in an Early Neolithic site, far from the initial domestication areas. *Proc. Natl. Sci. USA* **103** (42), 15375-15379.

Fleisher J (2003) Viewing stonestowns from the countryside: an archaeological approach to Swahili regional systems, AD 800-1500. Ann Arbor: University Microfilms.

Fortes GG & Paijmans JLA (accepted) Analysis of whole mitogenomes from ancient samples. In: Kroneis T (ed.) Whole genome amplification. Humana Press, USA.

Foster KP (1998) Gardens of Eden: Exotic flora and fauna in the ancient Near East. In: J. Coppock, J.A. Miller (eds.) Transformation of Middle Eastern natural environments: Legacies and lessons. Bullentin Series 102, Yale School of Forestry and Environmental Studies, 320-329.

Fuller DQ, Boivin N (2009) Crops, cattle and commensals across the Indian Ocean: current and potential archaeobotanical evidence. *Etudes Ocean Indien* **42/43**, 13-46.

Fuller DQ, Sato YI, Castillo C *et al*. (2010) Consilience of genetics and archaeobotany in the entangled history of rice. *Archaeological and Anthropological Sciences* **2**, 115-131.

Fuller DQ, Boivin N, Hoogervorst T, Allaby R (2011) Across the Indian Ocean: The prehistoric movement of plants and animals. *Antiquity* **85**, 544-558.

Futuyma DJ (1998) Evolutionary Biology (3rd ed.). Sinauer Associates Inc, Sunderland, Massachusetts.

Gansauge MT & Meyer M (2013) Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc*. **8**, 737-748.

Gifford-Gonzalez D, Hanotte O (2011) Domesticating animals in Africa: Implications of genetics and archaeological findings. *Journal of World History* **24**, 1-23.

Gilbert MTP, Hansen AJ, Willerslev E *et al*. (2003) Characterization of Genetic Miscoding Lesions Caused by Postmortem Damage. *The American Journal of Human Genetics* **72**, 48-61.

Gilbert MTP, Bandelt H, Hofreiter M, Barnes I (2005) Assessing ancient DNA studies. *Trends in Ecology & Evolution* 20, 541-544.

Gilbert MTP, Binladen J, Miller W *et al*. (2007a) Recharacterization of ancient DNA miscoding lesions: insights in the era of sequencing-by-synthesis. *Nucleic Acids Research* **35**, 1-10.

Gilbert MTP, Tomsho LP, Rendulic S *et al*. (2007b) Whole-Genome Shotgun Sequencing of Mitochondria from Ancient Hair Shafts. *Science* **317**, 1927-1930.

Glenn TC (2011) Field guide to next-generation DNA sequencers. *Molecular Ecology Resources* **11**, 759-769.

Gnirke A, Melnikov, A *et al*. (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol*. **27**, 182-189.

Google Earth (Version 5.1) Mountain View, CA: Google Inc. (2009). Available from http://www.google.com/earth/download/

Golenberg EM, Giannasi DE, Clegg MT, Smiley CJ, Durbin M, Henderson D, Zurawski G (1990) Chloroplast Sequence from a Miocene Magnolia Species. *Nature* **344**, 656-658.

Green RE, Krause J *et al*. (2006) Analysis of one million base pairs of Neanderthal DNA. *Nature* **444**, 330-336.

Green RE *et al*. (2009) The Neanderthal genome and ancient DNA authenticity. *EMBO J* **28 (17)**, 2494-2502.

Green RE, Krause J, Briggs AW *et al*. (2010) A Draft Sequence of the Neandertal Genome. *Science* **328**, 710-722.

Gunn B, Baudouin L, Olsen K (2011) Independent origin of cultivated coconut (*Cocos nucifera L.*) in the old world tropics. *PloS* **6 (6)**, 1-8.

Harper GA, Cabrera LF (2010) Response of mice (*Mus musculus*) to the removal of black rats (*Rattus rattus*) in arid forest on Santa Cruz Island, Galapagos. *Biological Invasions* **12**, 1449-1452.

Helm RM (2000) Conflicting histories: The archaeology of the iron-working, farming communities in the central and southern coast region of Kenya. Dissertation, Bristol.

Helm RM, Crowther A, Shipton C *et al*. (2012) Exploring agriculture, interaction and trade on the eastern African littoral: Preliminary results from Kenya. *Azania: Archaeological Research in Africa* **47 (1)**, 39-63.

Herrmann B (1994) Achäometrie. Naturwissenschaftliche Analyse von Sachüberresten. Berlin, Heidelberg, New York.

Hickerson MJ, Carstens BC *et al*. (2010) Phylogeography's past, present and future: 10 years after Avise, 2000. *Molecular Phylogenetics and Evolution* **54**, 291-301.

Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC (1984) DNA sequences from the quagga, an extinct member of the horse family. *Nature* **312**, 282-284.

Hingston M, Goodman S, Ganzhorn J, Sommer S (2005) Reconstruction of the colonization of southern Madagascar by the introduced *Rattus rattus*. *Journal of Biogeography* **32**, 1549-1559.

Hodges E, Xuan Z, et al. (2007) Genome-wide in situ exon capture for selective resequencing. *Nat. Genet*. **39**, 1522-1527.

Hofreiter M, Jaenicke V, Serre D, von Haeseler A, Pääbo S (2001) DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res*. **29**, 4793-4799.

Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S (2001) Ancient DNA. *Nat. Rev. Genet.* **2**, 353-359.

Hofreiter M, Capelli C *et al*. (2002) Ancient DNA analyses reveal high mitochondrial DNA sequence diversity and parallel morphological evolution of late pleistocene cave bears. *Mol. Bio. Evol.* **19**, 1244-1250.

Hofreiter M (2009) Spurensuche in alter DNA. *Biol. unserer Zeit* **39**, 176-184.

Horton M, Mudida N (1996) Subsistence at Shanga: The faunal record. In: M. Horton (ed.) Shanga: The archaeology of a Muslim trading community on the coast of East Africa. British Institute in Eastern Africa , Nairobi, 378-393.

Horton M, Middleton J (2000) The Swahili: The social landscape of a mercantile society. Blackwell, Cambridge.

Hoskin CJ (2011) The invasion and potential impact of the Asian House Gecko (*Hemidactylus frenatus*) in Australia. *Austral Ecology* **36**, 240-251.

Houlihan PF, Goodman SM (1986) The birds of ancient Egypt. Aris and Phillips, Warminster.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755. Version 3.2.2, available from http://mrbayes.sourceforge.net

Hummel S (1994) DNA aus alten Geweben. In: B. Herrmann (ed.) Archäometrie. Naturwissenschaftliche Analyse von Sachüberresten. Berlin, Heidelberg, New York, 87-100.

Hurles M, Sykes BC *et al*. (2005) The Dual Origin of the Malagasy in Island Southeast Asia and East Africa: Evidence from Maternal and Paternal Lineages. *Am*. *J. Hum*. *Genet*. **76**, 894-901.

Jobling MA, Hurles M, Tyler-Smith C (2004) Human Evolutionary Genetics. Origins, Peoples & Disease. New York.

Jones EP, Jóhannesdóttir F *et al*. (2011) The expansion of the house mouse into north-western Europe. *J. Zool.* **283**, 257-268.

Jones EP, Eager HM et al. (2013) Genetic tracking of mice and other bioproxies to infer human history. *Trends in Genetics* **29**, 298-308.

Juma A (2004) Unguja Ukuu on Zanzibar: An archaeological study of early urbanism. Uppsala University, Department of Archaeology and Ancient History, Uppsala.

Kaiser C (2005) Untersuchungen zur Degradation von DNA in Knochen und deren Anwendbarkeit für die Liegezeitbestimmung von Skelettfunden. Dissertation München.

Kelly CD (2005) Understanding mammalian evolution using Bayesian phylogenetic inference. *Mammal Rev*. **35**, 188-198.

Kennedy J (2009) Bananas: towards a revised prehistory. In: A. Fairbairn, E. Weiss (eds.) From foragers to farmers. Papers in honour of Gordon C. Hillman. Oxbow, Oxford, 190-204.

Kircher M, Heyn P, Kelso J (2011) Addressing challenges in the production and analysis of illumina sequencing data. *BMC Genomics* **12**, 382.

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

Knoop V & Müller K (2006) Gene und Stammbäume. Ein Handbuch zur molekularen Phylogenetik. Spektrum, München.

Krause J, Lalueza-Fox C *et al*. (2007) The derived FOXP2 variant of modern humans was shared with Neanderthals. *Curr Biol* **17**, 1908-1912.

Krause J, Fu QM, *et al*. (2010) The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464**, 894-897.

Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Pääbo S (1997) Neandertal DNA sequences and the origin of modern humans. *Cell* **90**, 19-30.

Kumar S, Stecher G *et al*. (2012) MEGA-CC: Computing Core of Molecular Evolutionary Genetics Analysis Program for Automated and Iterative Data Analysis. *Bioinformatics* **28**, 2685-2686. Available from http://www.megasoftware.net

Lalueza-Fox C, Römpler H *et al*. (2007) A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals. *Science* **318**, 1453-1455.

Lari M, Rizzi E *et al*. (2011) The complete mitochondrial genome of an 11450-year-old aurochen (Bos primigenius) from Central Italy. *BMC Evol Biol* **11**, 32.

Larson G, Dobney K et al. (2005) Worldwide phyleography of wild boar reveals multiple centers of pig domestication. *Science* **307**, 1618-1621.

Larson G, Cucchi T *et al*. (2007) Phylogeny and ancient DNA of Sus provides insights into neolithic expansion in Island Southeast Asia and Oceania. *PNAS* **104** (12), 4834-4839.

Larson G, Karlsson EK *et al*. (2012) Rethinking dog domestication by integrating genetics, archeology, and biogeography. *PNAS* **109** (23), 8878-8883.

Lemey P, Rambaut A, Welch JJ, Suchard MA (2010) Phylogeography takes a relaxed random walk in continuous space and time. *Molecular Biology and Evolution* **27**, 1877-1885.

Lever C (2003) Naturalized Reptiles and Amphibians of the World. Oxford University Press.

Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715.

Lindgreen S (2012) Adapter removal: Easy Cleaning of Next Generation Sequencing Reads. *BMC Research Notes* **5**, 1-7.

Loreille O, Orlando L, Patou-Mathis M, Philippe M, Taberlet P, Hänni C (2001) Ancient DNA analysis reveals divergence of the cave bear, Ursus spelaeus, and brown bear, Ursus arctos, lineages. *Curr. Bio.* **11**, 200-203.

Love RS (2006) Maritime Exploration in the Age of Discovery – 1415-1800. Greenwood Publishing Group.

Ludwig A, Pruvost M *et al*. (2009) Coat color variation at the beginning of the horse domestication. *Science* **324**, 485.

MacDonald K, Edwards D (1993) Chickens in Africa: The importance of Qasr Ibrim. *Antiquity* **67**, 548-590.

Madella M, Fuller DQ (2006) Palaeoecology and the Harappan civilisation of South Asia: a reconsideration. *Quaternary Science Reviews* **25**, 1283-1301.

Mamanova L *et al*. (2010) Target enrichment strategies for next-generation sequencing. *Nat Methods* **7 (2)**, 111-118.

Mardis ER (2008) Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics* **9**, 387-402.

Matisoo-Smith E & Robins J (2004) Origins and dispersals of Pacific peoples: Evidence from mtDNA phylogenies of the Pacific rat. *PNAS* **24**, 9167-9172.

Matisoo-Smith E & Robins J (2009) Mitochondrial DNA evidence for the spread of Pacific rats through Oceania. *Biol. Ivasions* **11**, 1521-1527.

Mbida C, De Langhe E, Vrydaghs H *et al*. (2006) Phytolith evidence for the early presence of domesticated banana (*Musa*) in Africa. In: M. Zeder, D. Bailer, E. Emshwiller, B.D. Smith (eds.) Documenting domestication: new genetic and archaeological paradigms. University of California Press, Berkeley (CA), 68-81.

Metzker ML (2009) Sequencing technologies – the next generation. *Nature Reviews Genetics* **11**, 31-46.

Meyer M & Kircher M (2010) Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harb Protoc* **6**, 5448.

Miller W, Drautz DI *et al*. (2009) The mitochondrial genome sequence of the Tasmanian tiger (Thylacinus cynocephalus). *Genome Research* **19**, 213-220.

Msaidie S, Ducournean A, Boetsch G *et al*. (2011) 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.

Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich E (1987) Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction. *Cold Spring Harbor Symposia on Quantitative Biology* **51**, 263-273.

Naderi S, Rezaei, HR *et al*. (2008) The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals, *PNAS* **105** (46), 17659-17664.

Namdar D, Gilboa A, Neumann R *et al*. (2013) Cinnamaldehyde in Early Iron Age Phoenician flasks raises the possibility of Levantine trade with Southeast Asia. *Mediterranean Archaeology and Archaeometry* **13**, 1-19.

Newbery B & Jones DN (2007) Presence of Asian House Gecko *Hemidactylus frenatus* across an urban gradient in Brisbane: influence of habitat and potential for impact on native gecko species. In: Lunney D *et al*. (eds.) Pest or Guest: the zoology of overabundance. Royal Zoological Society of New South Wales, Mosman (New South Wales, Australia).

Omar H, Adamson EAS *et al*. (2011) Phylogenetic relationships of Malayan and Malagasy pygmy shrews of the genus Suncus (Soricomorpha: Soricidae) inferred from mitochondrial cytochrome b gene sequences. *Raffles Bull. Zool*. **59**, 237-243.

Pääbo S (1989) Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *PNAS* **86**, 1939-1943.

Pääbo S, Wilson AC (1991) Miocene DNA sequences – a dream comes true? *Current Biology* **1**, 45-46.

Pääbo S, Poinar H (2004) Genetic Analyses From Ancient DNA. *Annual Review of Genetics* **38**, 645-679.

Plu A (1985) Bois et graines. In: L. Balout, C. Roubet (eds.) La momie de Ramses II. Contribution scientifique a l'egyptologie. Editions Recherches sur les Civilisations, Paris, 166-174.

Plug I (2000) Overview of Iron Age fauna from the Limpopo Valley. *Goodwin Series* **8**, 117-126.

Pock Tsy JML, Lumaret R, Mayne D *et al*. (2009) Chloroplast DNA phylogeography suggests a West African centre of origin for the baobab, *Adansonia digitat L*. (*Bombacoideae, Malvaceae*). *Molecular Ecology* **18**, 1707-1715.

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.

Poinar HN, Hofreiter M (1998) Molecular coproscopy: dung and diet of the extinct ground sloth Nothrotheriops shastensis. *Science* **281**, 402-406.

Poinar HN, Schwarz C et al. (2006) Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* **20** (311), 392-394.

Radimilahy C (1998) Mahilaka: An archaeological investigation of an early town in northwestern Madagascar. *Studies in African Archaeology* **15**, Department of Archaeology and Ancient History, Uppsala.

Ramirez O, Gigli E *et al*. (2009) Paleogenomics in a temperate environment: shotgun sequencing from an extinct Mediterranean caprine. *PloS One* **4**, 5670.

Rasmussen M, Li Y *et al*. (2010) Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762.

Reich D, Green RE *et al*. (2010) Genetic history an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053-1060.

Ripplinger J & Sullivan J (2008) Does Choice in Model Selection Affect Maximum Likelihood Analysis? *Syst. Biol*. **57**, 76-85.

Rizzi E, Lari M, Gigli E, De Bellis G, Caramelli D (2012) Ancient DNA studies: new perspectives on old samples. *Genetics Selection Evolution* **44**, 21-39.

Russell P (2001) Prince Henry 'the Navigator', a Life. New Haven, Yale University Press.

Saiki RK, Scharf S et al. (1985) Enzymatic Amplification of β-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia. *Science* **230**, 1350-1354.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences USA* **74**, 5463-5467.

Schadt EE, Turner S, Kasarkis A (2010) A window into third-generation sequencing. *Human Molecular Genetics* **19**, 227-240.

Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863-864.

Shapiro B, Drummond AJ *et al*. (2004) Rise and fall of the Beringian steppe bison. *Science* **306**, 1561-1565.

Simmonds NW (1962) The evolution of bananas. Longmans, London.

Stachan T, Read AP (1996) Molekulare Humangenetik.

Summerer D (2009) Enabling technologies of genomic-scale sequence enrichment for targeted high-throughput sequencing. *Genomics* **94**, 363-368.

Tchernov E (1968) Succession of rodent faunas during the Upper Pleistocene of Israel (Mammalia Depicta). Hamburg and Berlin.

Tchernov E (1984) Commensal animals and human sedentism in the Middle East. In: J. Clutton-Brock & G. Grigson (eds.) Animal and Archaeology: 3. Early Herders and Their Flocks. Oxford: British Archaeological Reports, 1984, 91-115.

Tollenaere C, Brouat C, Duplantier JM *et al*. (2009) 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.

Troy CS, McHugh DE et al. (2001) Genetic evidence for Near-Eastern origins of Euopean cattle. *Nature* **410**, 1088-1091.

Vences M, Wanke S, Vieites DR *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.

Walsh MT (2007) Island subsistence: Hunting, trapping and the translocation of wildlife in the western Indian Ocean. *Azania: Archaeological Research in Africa* **42**, 83-113.

Walshaw SC (2010) Converting to rice: Urbanization, Islamization and crops on Pemba Island, Tanzania, AD 700-1500. *World Archaeology* **42**, 137-154.

Weisskopf A, Fuller DQ (2014) Sesame: Origins and development. In: C. Smith (ed.) Encyclopedia of Global Archaeology. Springer, New York, 6578-6590.

Wild A (2004) Black gold: The dark history of coffee. Harper Perrenial, London.

Willerslev E, Cooper A (2005) Ancient DNA. *Proc. of the Royal Society B: Biological Sciences* **272**, 3-16.

Willerslev E, Cappellini E et al. (2007) Ancient Biomolecules from Deep Ice Cores Reveal a Forested Southern Greenland. *Science* **317**, 111-114.

Woodward SR, Weyand NJ, Bunnell M (1994) DNA sequence from Cretaceous period bone fragments. *Science* **266**, 1229-1232.

Zischler H, Höss M, Handt O, von Haeseler A, van der Kuyl AC, Goudsmit J (1995) Detecting dinosaur DNA. *Science* **268**, 1192-1193.

Electronical sources (2014)

Global Invasive Species Database:

http://www.issg.org/database/species/ecology.asp?si=97&fr=1&sts=&lang=EN
http://www.issg.org/database/species/ecology.asp?si=19&fr=1&sts=sss&lang=EN
http://www.issg.org/database/species/ecology.asp?si=1344&fr=1&sts=&lang=EN

Animal Diversity Webpage:

http://animaldiversity.org/accounts/Mus_musculus/
http://animaldiversity.org/accounts/Rattus_rattus/

ICUN Redlist:

http://www.iucnredlist.org/details/176130/0

2 *Mus musculus castaneus*

Phylogeography of *Mus musculus castaneus* reveals human-mediated dispersal routes throughout the Indian and Pacific Ocean

2.1 Abstract

Due to their remarkable adaptability, close association with humans, and ecological dependence on them, the house mouse (*Mus musculus*) is one of the most widespread commensal animals, currently occupying almost all continents around the globe. But despite their socio-economic impact, the origin and invasive pathways remain poorly understood, especially for the subspecies *M. m. castaneus*. Here, molecular genetic methods and phylogeographic diffusion analyses, based on mitochondrial d-loop data, are used to establish the geographical range and assess dispersal routes of *M. m. castaneus*. Results reveal that humans are undoubtedly associated with their range expansion, possibly reflecting emerging trade connections during the late 1st millennium AD and travel routes of European seafarers during the Age of Discovery.

2.2 Introduction

The 'house' or 'common' mouse *Mus musculus* is a widespread commensal animal, which currently inhabits almost all continents and islands (see figure 2.1) (Musser & Carleton 2005). Presently, the assumed origin of the house mouse is the Indo-Pak subcontinent, where genetic diversity is greatest (Boursot *et al*. 1993, Boursot *et al*. 1996, Prager *et al*. 1998). Several studies (Berry & Jakobson 1975, Brothwell 1981, Berry 1986) indicate that *M. musculus* and its ancestors favoured a temperate steppic climate, a circumstance that would have prevented a natural expansion across deserts, mountain ranges or tropical forests (Bonhomme & Searle 2012). But due to their remarkable adaptability, close association with humans, and ecological dependence on them, mice became one of the most successful and ubiquitous invasive mammals occupying all kinds of habitats outside their natural bounds. Due

to passive transport by humans, the house mouse is nowadays found from coastal areas over shrublands and tropical forests to mountainous regions, sub-Antarctic islands, agricultural landscapes and even in modern cities (Boursot *et al*. 1993, Lowe *et al*. 2000).

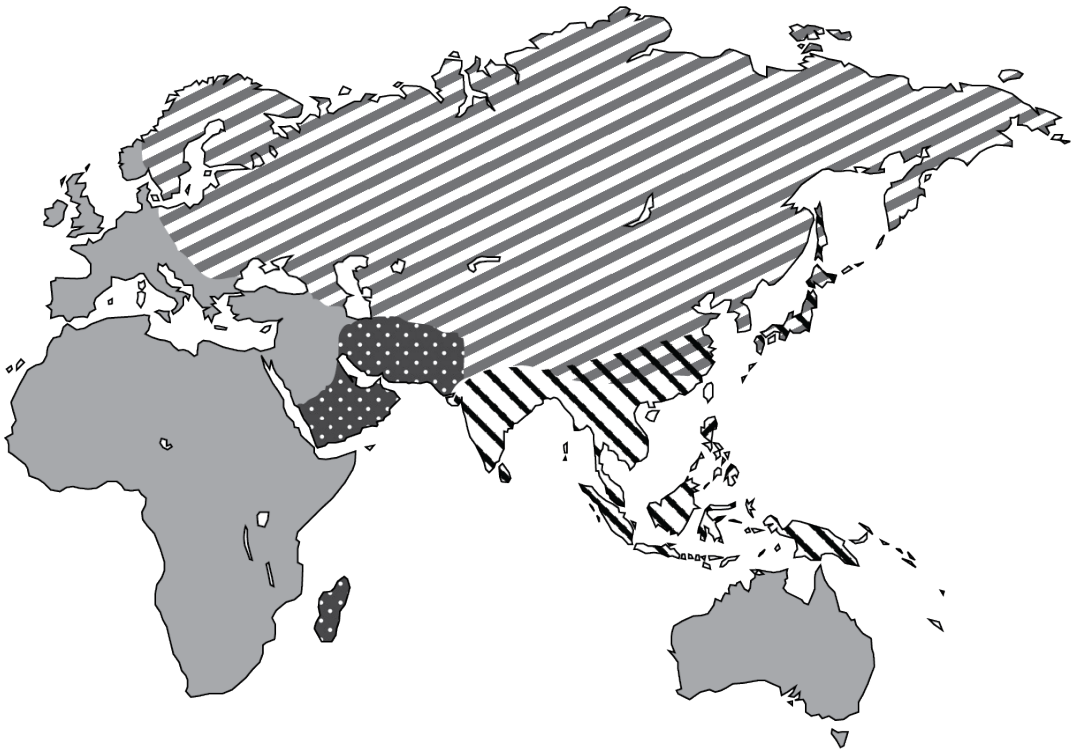

Figure 2.1: Current distribution of M. musculus within the study area.
Solid grey: M.m. domesticus; Dotted: M.m. gentilulus, Grey-striped: M.m. musculus; Black-striped: M.m. castaneus; Hybrid zones of castaneus and musculus in Japan and China: M.m. molossinus.

Within the genus *Mus* there are five currently recognised subspecies, based on distinct molecular lineages (Tucker 2007, Suzuki & Aplin 2012, Yonekawa *et al*. 2012, Bonhomme & Searle 2012): the distribution area of *M. m. musculus* ranges from Central over Eastern Europe to North-East Asia, whereas *M. m. domesticus* occurs in the Near East, North Africa, western Europe and meanwhile reached the Americas, Australia and Oceanian islands through passive transport with humans. *M. m. castaneus* is found in Central and Southeast Asia (Musser & Carleton 2005). Furthermore there are two subspecies with a more restricted range: the recently described *gentilulus* from the south-eastern coast of the Arabian peninsula (Prager *et al*. 1998) and Madagascar (Duplantier *et al*. 2002, Auffray & Britton-Davidian

2012), as well as *M. m. molossinus*, a hybrid between *musculus* and *castaneus*, is found in Japan (Yonekawa *et al*. 2012). An additional lineage *bactrianus*, supposedly in Afghanistan and Pakistan, is indicated but not well supported (Din *et al*. 1996).

The level of commensalism varies among the different subspecies. While *M. m. castaneus* is exclusively commensal, other subspecies like *musculus* or *molossinus* are able to occupy several niches from human-related to wild environments. *M. m. domesticus* is mainly found near human dwellings but establishes feral populations in areas where competitors are absent (Auffray & Britton-Davidian 2012).

The transition from hunter-gatherers to sedentary societies, with a lifestyle based on agriculture and husbandry, is commonly accepted as a key factor in the evolving commensal relationship between mice and men (Tchernov 1984, Boursot *et al*. 1993, Cucchi *et al*. 2005, Cucchi *et al*. 2012). Population growth and the related increase in settlement size and density, is linked to several factors that enhanced attractiveness for mice and therefore facilitated their dispersal. Tchernov (1968, 1984) was the first to study this biological interaction between humans and the house mouse, proposing that the shift towards sedentism created a new ecological niche for *Mus* species, as the development in production and large-scale storage of grain provided a consistent source of food. Furthermore, human dwellings do not only function as artificial shelter for the small mammals, but provide protection against predation or interspecific competition.

The earliest evidence for commensalism of house mice has been found in the Levant about 12,000 years ago in association with the Natufian culture. Archaeological remains of *M. musculus* are present throughout the stratigraphic deposits at Hayonim Cave (Israel) and show a massive increase in Natufian layers, a time when people started harvesting and storing grain to some degree (Tchernov 1984, Auffray *et al*. 1988). Although Tangri and Wyncoll (1989) bring into consideration that the change observed in rodent frequencies could be caused by variation of rodent populations in the surrounding habitat or fluctuations in the diet of birds of prey, rather than the development of a population of commensal mice, the general view supports the more parsimonious interpretation based on Tchernov's

(1984) theory (Cucchi *et al*. 2012). From there, human-mediated long-distance transport during the Late Bronze and Early Iron Age facilitated a westward spread through Eurasia via two pathways. While *M. m. musculus* took a northern route to central and northern Europe, *M. m. domesticus* colonised the Mediterranean via a southern route (Boursot *et al*. 1993, Cucchi *et al*. 2005). Furthermore, with the beginning of extensive seafaring activities of Europeans during the Age of Discovery, the dispersal of the house mouse progressed onto other continents and oceanic islands (Jones *et al*. 2011, Jones *et al*. 2013, Bonhomme & Searle 2012).

The study presented here particularly focuses on the much less understood dispersal of *M. m. castaneus*. Due to a lack of available data, the biogeographic history of this subspecies remains uncertain. Although humans are undoubtedly associated with their dispersal, precise indications are missing of where and when this commensal relationship began. Rajabi-Maham *et al*. (2012) provided first insights into the phylogeography of *M. m. castaneus*, identifying three lineages that show distinct geographic distribution. Furthermore, results indicate recent expansion, presumably through human-mediated translocation.

In order to establish the geographical range and assess dispersal routes of *M. m. castaneus* we extended the sample range of previous studies by Rajabi-Maham *et al*. (2012) and performed phylogeographic diffusion analyses, based on mitochondrial d-loop data. The marker choice was based on the assumption that commensal mouse populations are resistant to secondary invasion by females due to a hierarchical structure within the population. Therefore newly arriving females coming into an established population are generally unable to survive or gain mates, and in consequence do not contribute to the gene pool of the population. This means the maternally inherited mitochondrial DNA may reflect signals of the first coloniser in the regions considered and subsequently of the humans who transported them (Boursot *et al*. 1993, Gabriel *et al*. 2010).

2.3 Material and Methods

2.3.1 Samples

A total of 351 tissue samples of *M. musculus* were analysed, of which 239 specimens were collected from various museums (Field Museum, Chicago; American Museum of Natural History, New York; Smithsonian, Washington DC) and an additional 112 modern samples were provided by Michel Pascal (French National Institute for Agricultural Research, Rennes) and Katrin Koch (Johann Wolfgang von Goethe-University, Frankfurt). The sampling area comprises different places of the mainland and islands around the Indian Ocean, including countries from the East and West African coast, the Arabian peninsula, as well as the Indian subcontinent and South-East Asia. The project was divided geographically between the PhD students, whereas Heidi Eager (University of Oxford) processed all samples located in the western Indian Ocean (Africa, Arabian peninsula, Iran, West-India) and I focused on samples from the eastern Indian Ocean area and the Pacific (East-India, China, South-East Asia, Australia). In phylogenetic analyses, DNA sequences of specimens from India, Sri Lanka, and Pakistan were shared (see figure 2.2). A list of all 139 samples used in my part of the study is found in chapter 7.2.1, including additional information.

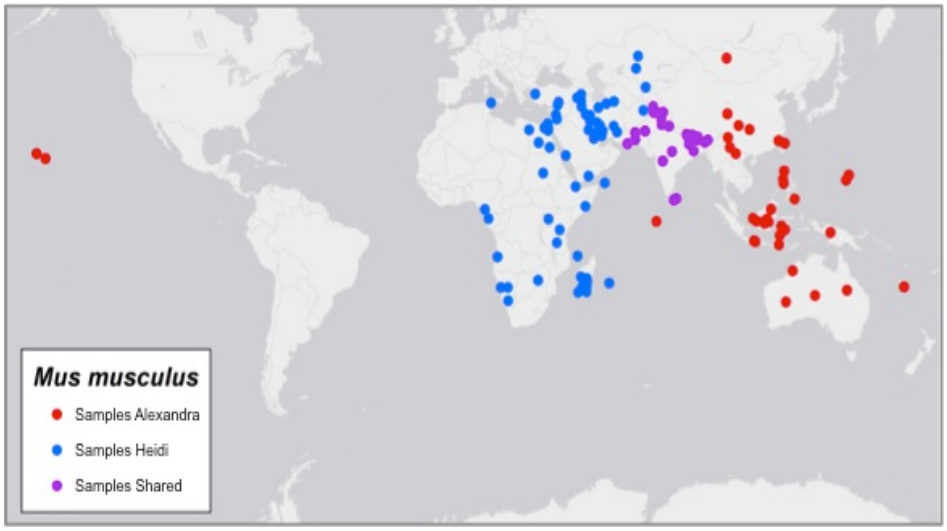

Figure 2.2: Sample distribution of M. musculus.

(Figure 2.2 continued)
Distribution of sampling locations of M. musculus specimens used in this study. Samples extracted and analysed by Heidi Eager are indicated in blue, samples extracted and analysed by Alexandra Trinks are indicated in red. Jointly extracted and analysed samples are indicated in purple.

2.3.2 Analyses

All samples were analysed in different facilities at Cornell University. Because of the age of certain samples (> 100 years old) they were treated as ancient material. Therefore all pre-PCR working steps – sample preparation, extraction and PCR set-up were conducted in a clean-room laboratory free of molecular work at the Laboratory of Ornithology (see chapter 7.1.2, extraction protocol 1 and 2). All following steps including PCR-amplification, purification and sequencing set-up were carried out in the facilities of the Searle-Lab, based in the Department of Ecology and Evolutionary Biology, while sequencing was performed by the Life Sciences Core Laboratories Center at Cornell University (see chapters 7.1.3 and 7.1.4).

In total, 870 bp of the mitochondrial control region were amplified and sequenced in 8 overlapping fragments, whereas a variety of primer combinations was used depending on the nature of the sample (see chapter 7.1.1). The sequencing chromatograms were edited manually, subsequently assembled, and a consensus sequence per individual exported (see chapter 7.1.4) using Geneious R6 version 6.0.6 (Drummond *et al*. 2011). Standard anti-contamination guidelines were followed to ensure authenticity of aDNA results (see chapters 1.4.3 and 7.1.5).

Phylogenetic analyses were performed by construction of Bayesian trees using MrBayes version 3.2.2 (Huelsenbeck & Ronquist 2001). Firstly, all sequences from the eastern Indian Ocean area and the Pacific were included in phylogenetic analyses in order to define subspecies of *Mus musculus* within the dataset. The best-fit nucleotide substitution model, ascertained in MEGA-CC (Kumar *et al*. 2012) under the Bayesian Information Criterion (BIC), was Hasegawa-Kishino-Yano +Γ. Under this model, a consensus tree with bootstrap values on the nodes was created for

5,000,000 generations, a sample frequency of 5,000 and with a burn-in period of 100 trees (see figure 2.3).

Furthermore, mitochondrial phylogeography of *M. m. castaneus* was inferred using all 65 sequences identified as this specific subspecies, including *castaneus* specimens processed by Heidi Eager (see Appendix 7.2.1; Uganda: 149873, UAE: 256649, Iran: 88788, 88912, 112270, 112280, Reunion: MP17, MP19, MP20, MP22, MP23, MP24, MP27, MP28, MP29, MP33). These were combined with 117 sequences entries from GenBank to a dataset composed of 182 individuals. A Bayesian tree of this dataset was generated, using BEAST version 1.8.0 (Drummond & Rambaut 2007). Here, the best-fit nucleotide substitution model was Hasegawa-Kishino-Yano +Γ +I. To calibrate the tree a mutation rate of 0.041 (= 4.1%) substitutions per site per million years was assumed based on different papers about rodent mutation rates (Goios *et al*. 2007, Geraldes *et al*. 2008). In order to apply the correct clock model, a relaxed clock model was used initially to test for variation in the evolutionary rate. The statistics showed that the hypothesis of a constant mutation rate could not be rejected, consequently the mutation rate is constant. Therefore, the parameter has been changed to a strict clock model in the final run. Additionally, a coalescent tree prior with exponential population growth was specified, based on the expectation that the population size of *M. m. castaneus* has not remained constant during their evolutionary history.

For the phylogeographic diffusion analysis the initially used lognormal relaxed random walk model (RRW), which allows the rate of diffusion to vary along individual branches of the topology (Lemey *et al*. 2010), was changed to the less complex homogenous Brownian model, which assumes a constant dispersal at the same rate. Complexity of the models needed to be reduced, as the statistical results of the initial run favoured three different scenarios, which prevented the tree from converging. After running several trees with combinations of different clock–, tree– and diffusion models, statistical results favoured the homogenous Brownian diffusion model, a strict clock, and exponential population growth. Coordinates of the sample locations used in diffusion analyses were recovered from Google Maps. In cases were duplicate localities confound the diffusion analysis, a random jitter was

applied, where new localities were drawn randomly within a geographic window of 0.02 x 0.02 decimal degrees (= 2 km x 2 km) from the original locality (Barlow *et al*. 2013). Two independent runs of 50,000,000 generations each were performed with a sample frequency of 5,000 and a 10% burn-in period (see figure 2.4). Afterwards, the generated BEAST tree was uploaded in SPREAD version 1.0.6 (Bielejec *et al*. 2011). SPREAD allows to visualise the full posterior distribution of the trees obtained in phylogeographic analyses and plots the results onto a map. The generated output file was uploaded in Google Earth in order to visualise the simulated spreading in space and time (see figure 2.6).

Additionally, the generation of a median-spanning network of *castaneus* specimens with PopART (http://popart.otago.ac.nz), based on the number of differences between sequences, revealed 99 different haplotypes among a total sample set of 170 sequences (see figures 2.7 and 2.8). Due to the inability of networks to deal with missing data, the original *castaneus* alignment had to be truncated to a 700 bp fragment of the mitochondrial d-loop.

2.3.3 Results

A dataset comprising 351 *M. musculus* samples has been divided regionally between Heidi Eager and myself. My part of the study focused on samples located in the eastern Indian Ocean area and the Pacific, including a total of 139 samples. Bayesian analyses of DNA sequences of a 870 bp mitochondrial d-loop fragment identified 65 samples as *M. m. castaneus* within the dataset.

Figure 2.3: Mitochondrial phylogeny of M. musculus.

(Figure 2.3 continued)
Inferred from a 870 bp fragment of the mitochondrial d-loop using MrBayes version 3.2.2 in order to identify M. m. castaneus specimens. The best-fit nucleotide substitution model under the Bayesian Information Criterion (BIC) was Hasegawa-Kishino-Yano +Γ. Under this model, a consensus tree with bootstrap values on the nodes was created for 5,000,000 generations, a sample frequency of 5,000, and with a burn-in period of 100 trees. M. m. castaneus haplogroup HG1b is indicated in red, M. m. castaneus haplogroup HG2 is indicated in green, M. m. castaneus haplogroup HG3 is indicated in blue. Other M. musculus sp. are marked in black.

In a phylogeographic diffusion analyses these 65 samples were combined with 117 sequences obtained from GenBank. Results of the analyses revealed lineages clustering with haplogroups HG1b, HG2 and HG3 as identified by Rajabi-Maham *et al*. (2012). According to this previous study, each lineage is restricted to a distinct geographic region: whereas HG1a/b occurs in northern India and Pakistan, as well as central and north-west Iran, HG3 is mostly restricted to eastern Iran, Afghanistan, and Indo-Pakistan. Furthermore, a single specimen was recorded from Taiwan. HG2 is the most abundant haplogroup, covering areas in East-Africa, India, as well as eastern China and New Zealand. The samples analysed here, fall into these defined regions. Bayesian analyses generated a well-resolved tree with posterior probability values between 0.96 and 0.99 for the three main clades apparent from the data – HG1b, HG2 and HG3. The maximum time span assumed for the divergence of these three haplogroups is 300,000 to 400,000 years ago (see figure 2.4). The extension of the sampling area in this study initially reveals congruence of the distribution range of HG1b and HG3 presented by Rajabi-Maham *et al*. (2012). Furthermore, results uncover a broader distribution area for HG2. Here, this clade is found on the East-African mainland, as well as on Reunion, further on the Indian subcontinent, in China and ISEA as far as on the Pacific island Guam (see figure 2.5).

Molecular-clock-rooted analyses place the origin of the coalescence tree of *M. m. castaneus* in the expected native range India, indicated by the initial branch in red. From there, a constant dispersal occurs in eastern direction into ISEA and through oceanic translocation to Reunion and East Africa. Furthermore, a highly branched pattern is observed in Pakistan, Afghanistan and Iran. The continuous spreading of the house mouse throughout a longer time period is demonstrated by the light blue confidence interval polygons of the internal tree nodes, that cover almost the entire area (see figure 2.6).

Figure 2.4: Molecular-clock based mitochondrial phylogeny of M. m. castaneus.

(Figure 2.4 continued)
Inferred from a 870 bp fragment of the mitochondrial d-loop using BEAST version 1.8.0. The best-fit nucleotide substitution model was Hasegawa-Kishino-Yano +Γ +I. Under this model, two independent runs of 50,000,000 generations each were performed with a sample frequency of 5,000 and a burn-in of 100 trees. To calibrate the tree a mutation rate of 0.041 (= 4.1%) substitutions per site per million years was assumed. Statistics showed that the hypothesis of a constant mutation rate could not be rejected, consequently a strict clock model was applied. Clade support is given by Bayesian posterior probabilities. Node bars show the 95% HPDs. Additionally, a coalescent tree prior with exponential population growth was specified. For the phylogeographic diffusion analysis a homogenous Brownian model was applied. Coordinates of the sample locations used in diffusion analyses were recovered from Google Maps. In cases were duplicate localities confound the diffusion analysis, a random jitter was applied, where new localities were drawn randomly within a geographic window of 0.02 x 0.02 decimal degrees (= 2 km x 2 km) from the original locality. Haplogroup HG1b is indicated in red, HG2 is indicated in green, HG3 is indicated in blue.

Figure 2.5: Haplogroup distribution of M. m. castaneus.
Samples analysed in this study are marked with a circle, samples analysed by Rajabi-Maham et al. (2012) are marked with a triangle. Haplogroup HG1b is indicated in red, HG2 is indicated in green, HG3 is indicated in blue. Samples included in the analyses, which have been processed by Heidi Eager, are black-rimmed.

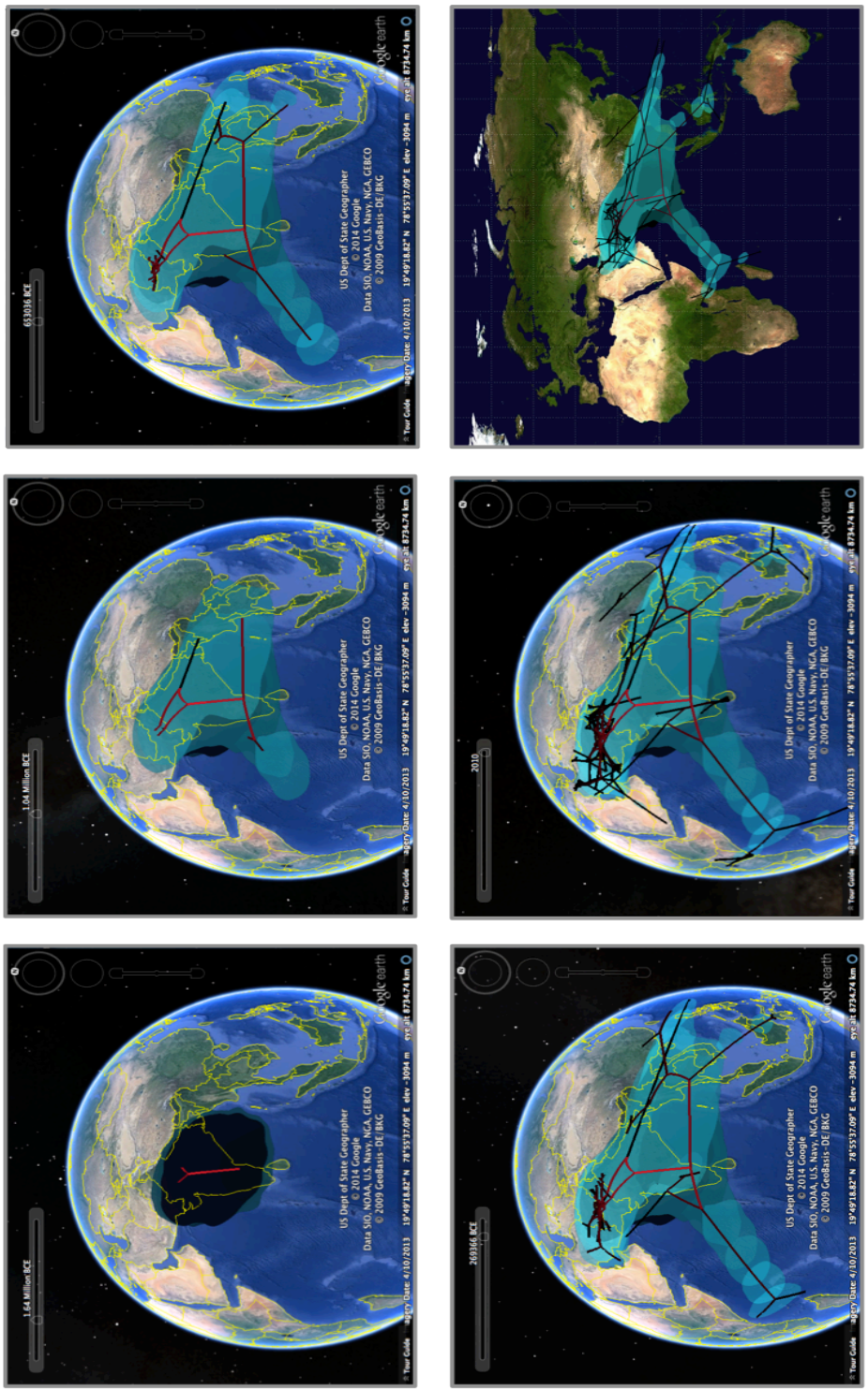

Figure 2.6: Dispersal simulation of M. m. castaneus in space and time.
Visualisation of the full posterior distribution of the trees obtained in the phylogeographic analysis in BEAST, using SPREAD version 1.0.6. Subsequently, the generated output file was uploaded in Google Earth in order to visualise the simulated spreading of M. m. castaneus in space and time.

Results of the network analysis revealed 99 haplotypes in a total of 170 sequences (see figures 2.7 and 2.8). Thereby, three groups are apparent that relate to HG1b, HG2, and HG3 identified by Rajabi-Maham *et al*. (2012). The haplotype with the highest frequency and at the same time broadest distribution is M1. Connected to the Indian haplotype M42, the widely spread M1 occurs in South-East Asian locations such as Taiwan, Thailand, Mariana Island, Java, Indonesia, and the Indian peninsula, further in China and Iran, as well as at the East African coast in Kenya and Uganda. Furthermore, results demonstrate that specimens from the areas Iran, Pakistan, and India show the highest haplotype diversity.

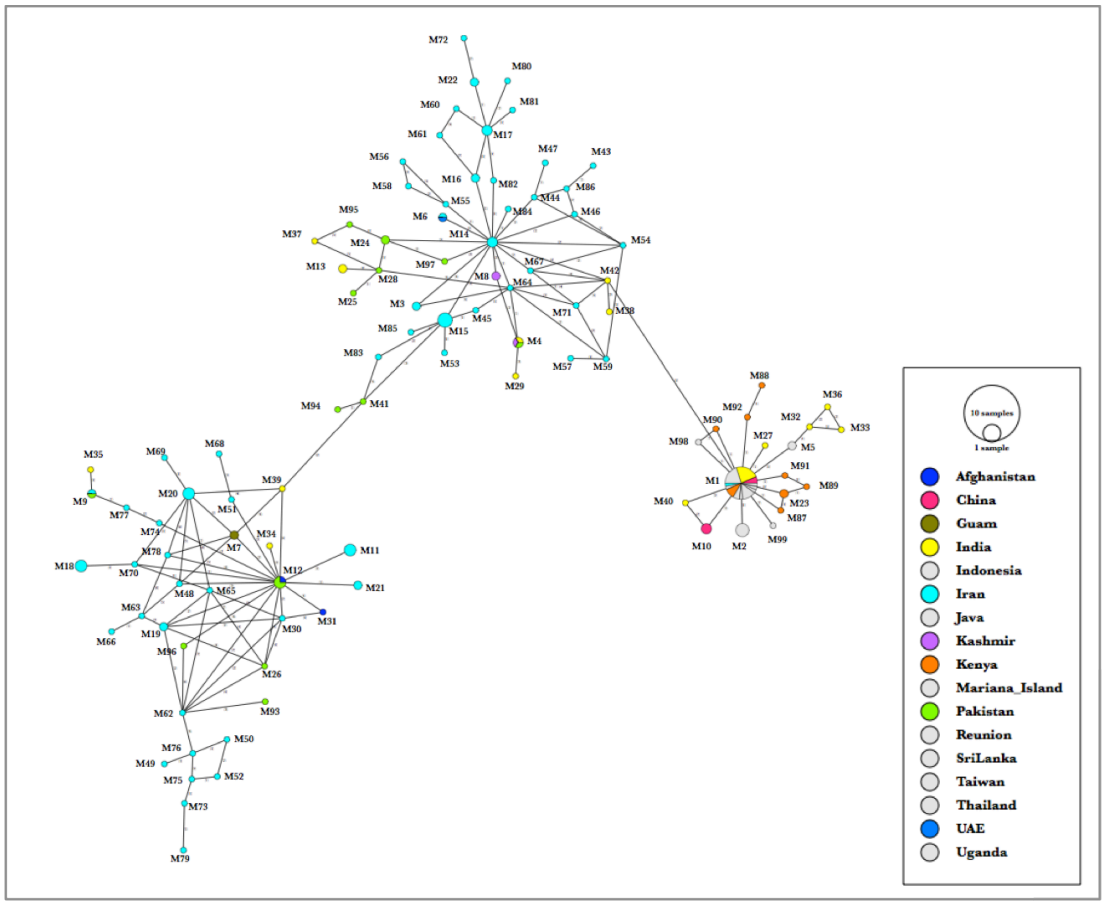

Figure 2.7: Network analysis of M. m. castaneus using PopART.
Haplotype diversity of M. m. castaneus throughout the region, assessed from a 700 bp fragment of the mitochondrial d-loop. 170 sequences have been included and revealed 99 haplotypes.

Country	Haplotype
Afghanistan	M12, M31
China	M1, M10
Guam	M7
India	M1, M4, M13, M27, M29, M32, M33, M34, M35, M36, M37, M38, M39, M40, M42,
Indonesia	M1, M2
Iran	M1, M3, M6, M9, M11, M14, M15, M16, M17, M18, M19, M20, M21, M22, M30, M43, M44, M45, M46, M47, M48, M49, M50, M51, M52, M53, M54, M55, M56, M57, M58, M59, M60, M61, M62, M64, M65, M66, M67, M68, M69, M70, M71, M72, M73, M74, M75, M76, M77, M78, M79, M80, M81, M82, M83, M84, M85, M86, M96
Java	M1
Kashmir	M4, M8
Kenya	M1, M23, M87, M88, M89, M90, M91, M92
Mariana Island	M1
Pakistan	M4, M9, M12, M24, M25, M26, M28, M41, M93, M94, M95, M96, M97
Reunion	M1, M99
Sri Lanka	M1, M5
Taiwan	M1
Thailand	M1, M98
UAE	M6
Uganda	M1

Figure 2.8: Table of the haplotype diversity of M. m. castaneus.

2.4 Discussion

The house mouse *M. musculus* is one of the most widespread commensal animals, nowadays occupying almost every corner around the globe. However, the biogeographic history – the original geographical range and the dispersal routes – of the lesser understood *M. m. castaneus* still needs to be assessed. Here, results demonstrate that all samples analysed fall into three lineages – HG1b, HG2 and HG3 as defined by Rajabi-Maham *et al*. (2012). Thereby, HG1b and HG3 are geographically restricted and only HG2 has been transferred to many other places, from the East African coast, over the Indian subcontinent, throughout ISEA to New Zealand. Different explanations can be considered for this particular distribution pattern of only one spreading haplogroup.

Firstly, the cause could lie in an incorrect taxonomy. Due to misidentifications, the three lineages could represent different subspecies, whereas one subspecies is more associated with human movement, than the others. However, more likely the three lineages represent haplogroups of *M. m. castaneus,* whereby the different populations have been geographically subdivided. Rajabi-Maham *et al*. (2012) discussed the possibility of a period of isolation linked to the last glaciation with putative refugia in Eastern Iran, Afghanistan and South Pakistan (HG3), North-East India (HG2), and the foothills of North-West India (HG1a). The picture of HG1b is more ambiguous because of its disjoint geographical distribution, which likely has been shaped by early human influence. After the period of isolation, HG2 may have occupied a specific niche that promoted commensalism with humans or coincidentally have been in the right place at the right time to get transported preferably over HG1 and HG3. Once humans started to move this particular population around and established it in various regions outside their natural habitat, it became more difficult for other specimens to invade these areas, as territories are defended by local populations (see chapter 1.3.2). In general, commensal mouse populations are resistant to secondary invasion by females due to a hierarchical structure within the population. Therefore, newly arriving females coming into an established population are unable to survive or gain mates, and in consequence do

not contribute to the gene pool of the population (Gabriel *et al*. 2010, Bonhomme & Searle 2012).

Determination of the timing of their spread is difficult. But taking current archaeological evidence and characteristics of the ecology of *M. m. castaneus* into account, may help to narrow down the phases of their range expansion. Here, results demonstrate that specimens carrying HG2 are found in South-East Asia, India, and along the East African coast. Thereby, two groups are apparent within the phylogenetic tree. While populations from Iran, China and Thailand group together with North-East-Indian specimens from Delhi, Bengal, Sikkim, Assam and further form a sister group with *M. m. castaneus* from Indonesia, East-African populations from Uganda, Kenya and Reunion cluster together with specimens from West-India (Masinagudi, Kotagiri) and Sri Lanka. As HG2 specimens of the 'western group' are absent along the land route on the Arabian peninsula, their distribution pattern in westward direction strongly indicates a human-mediated seaborne dispersal from their native range. The distribution pattern of the 'eastern group' has probably been created via transport over land from India through China and Thailand, and via maritime pathways across Indonesia.

By the 1st millennium BC, evidence of exchange between the East African coast and India is recorded through archaeobotanical remains. Several African crops are widespread across India, however, during that time there is no evidence that Indian ships reached the East African coast. Although the Persian Gulf has been controlled by an Arab trade system, there is no evidence of *castaneus* HG2 in this region (Fuller & Boivin 2009, Boivin *et al*. 2014). An early dispersal of *M. m. castaneus* across the Pacific, e.g. related to the Austronesian expansion, is possible according to the visible pattern, but unlikely due to the following reasons. Although their level of commensalism varies among the subspecies, mice are dependent on resources and shelter provided by humans. The house mouse accumulates where it can exploit food and is protected from competitors and predators. Therefore, their presence represents a relative good marker for established sedentary societies, and have unlikely been moved around by nomadic cultures. Furthermore, *M. m. castaneus* could have been transported in canoes, but would not have been preferred

over other rodents like *R. exulans*, that was deliberately transported as a source of food, throughout Oceania to various Pacific islands (Addison & Matisoo-Smith 2010).

House mice are agile climbers and occupy a variety of niches, such as houses, barns or granaries. Thereby, they usually hide in dark places and due to their small size get easily transported unseen in cargo (or in earlier times in bundles of hay or grain). Because of their remarkable adaptability and omnivorous nutrition, house mice are able to quickly adjust to different environments. In turn, the close association to humans and their constant supply with shelter and food, is the basis for year-round breeding and subsequently the reason for the successful establishment of mice populations in new areas (Cole 2005, http: //animaldiversity.org/ accounts/ Mus_musculus/). Among the house mouse subspecies, especially *M. m. castaneus* is highly synanthropic and dependent on human habitat and food sources (Auffray & Britton-Davidian 2012). Therefore, the widespread range extension of *castaneus* must have taken place in later times when settled societies have already been formed and agriculture, as well as trade and transport have been well established. Rajabi-Maham *et al*. (2012) dated the onset of the HG2 expansion into the Indo-Pacific to *c.* 1200 years ago. This correlates with the intensification of contacts between trans-regional trade networks across the Indian Ocean due to political stabilisation, growing port cities, as well as technological and infrastructural improvements (Boivin *et al*. 2014). By the end of the 1st millennium AD, the great empires in China and India got connected with East Africa and the Mediterranean, leading to frequent travel and trade within the region (Beaujard 2005). Furthermore, regular contacts in eastern direction between India and the Malay peninsula – overland and through the Bay of Bengal – have been established during the early 1st millennium AD (Beaujard 2005, Asouti & Fuller 2008, Castillo & Fuller 2010, Boivin et *al*. 2014). By that time, particularly the famous Silk Road would have facilitated the spread of not only goods, but biological specimens throughout the region. These overland and -sea corridors may have enabled the spread of *M. m. castaneus* HG2 from their native range in India in eastward direction throughout the Pacific and across the Indian Ocean as far as East Africa.

Earliest secured archaeological evidence of the house mouse have been found on the islands along the East African coast (Zanzibar, Grande Comore) and are dating to the mid to late 1st millennium AD (Juma 2004). However, these skeletal remains of the house mouse are scarce and are missing on the African mainland during the same time period. While this supports a direct seaborne introduction from South-East Asia, the lack of finds of *M. musculus* along the coastal route may be a reflection of a general low population size due to the competition with the dominant black rat, which has been more abundant at African sites (Harper & Cabrera 2010).

Generally the islands off the East African coast have been first settled by humans at different times, spanning several centuries throughout the 1st millennium AD. Reunion, where *castaneus* is found, has not been used as a regular stopping point for early seafarers. Instead, this island was discovered by Portuguese explorers around 1510 AD, and first settled by the French in 1660 AD. Mice have only been recorded around 1754 AD, although they may have been present before that (Cheke 2010). Additionally, the low haplotype diversity of *castaneus* on Reunion (M1 and M99) indicates a rather recent introduction. Between the 15th and 18th century AD, Portuguese mariners and merchants started sailing along the African West coast and further throughout the Indian Ocean in order to find a direct sea route to Asian markets. These early voyages, followed by Spanish, French, English, and Dutch explorers in later times, increased the worldwide interaction of distant cultures and the dispersal of seaborne goods (Love 2006). Hence, the new trading networks – with stopping points in India, the native range of *M. m. castaneus* HG2 – facilitated the spreading of the house mouse onto islands or remote areas throughout the Pacific as far as Guam, and on their way back to the East African coast. This is further reflected within the network analysis, where the Indian haplotype M42 connects to the currently widespread haplotype M1. The reason why *M. m. castaneus* does not travel back with European seafarers along the South and West African coast, may be due to the competition with *M. m. domesticus* populations in these regions, which already have been established while sailing from Europe to Asia.

Where and when the commensal relationship between *M. m. castaneus* and humans exactly began is still subject to discussion due to a lack of zooarchaeological analyses of early Neolithic sites (Bonhomme & Searle 2012). But combining molecular genetic data and archaeological or historical evidence can help to date the origin and subsequent dispersal routes of the house mouse. The molecular clock has been frequently used to place timescales on phylogeographic analyses. However, there is evidence that date estimates from molecular genetic data can be incorrect due to the inconsistency of mutation rates over time (Ho & Larson 2006, Herman *et al*. 2014). The uncertainty surrounding this particular parameter is illustrated by Foerster *et al*. (2009). In this study, a range of values for mutation rates in mice have been used in order to unravel the colonisation history of Madeiran house mice. Results gave a range of expansion time estimates, from prehistoric periods to Viking seafarings in the 9th century AD. In 2014, a study from Rando *et al*. used radiocarbon dating to establish an accurate chronology for the early presence of house mice on Madeira. Results confirmed one of the proposed colonisation times obtained through mitochondrial d-loop data by Foerster *et al*. (2009), pointing to an introduction during the Viking Age in the late 1st millennium AD.

This demonstrates, that radiocarbon dating is a useful tool to determine the exact age of bone samples and in combination with genetical analyses is able to reconstruct early human presence on islands. How far the history of a certain population can be traced back, depends on the age of the samples. Thereby, the analysis of museum samples is a suitable alternative, when archaeological remains are not available or, in case of micro-mammals, are difficult to analyse. Often the lack of archaeological finds and their small bone size represents a limiting factor to realise both analyses.

Therefore, it is essential to establish an appropriate mutation rate in order to assess a correct timeframe. Here, estimated timeframes within the Bayesian tree may be imprecise as the mitochondrial sequence of only 870 bp is too short for mutations to occur and therefore unable to take all changes into account. However, although an absolute assessment of dispersal times is difficult here, molecular cock based analyses rather provide a relative classification of significance. In this case, also

adding more genetic marker, i.e. whole mitogenomes or nuclear genes, is necessary to overcome the current lack of sensitivity of the data. Here, one has to bear in mind that the results are only the reflection of the evolutionary history of the mitochondrial d-loop, which may be inconsistent with the true population history. The dates assumed for the splits of haplogroups are therefore only an indicator for the maximum time span of separation. Population splits may have happened more recently, than the divergence between gene lineages. Therefore, incorporating ancient DNA or "relating the divergence of lineages directly to contemporaneous external events of known time" can be used to calibrate evolutionary events more precisely (Herman *et al*. 2014).

In summary, this study has sought to assess the phylogeographic history of *M. m. castaneus*. As only few data have been available so far, their geographical range and dispersal routes remained uncertain. Here, results of phylogeographic diffusion analyses show that humans are undoubtedly associated with their range expansion, possibly reflecting emerging trade connections during the late 1st millennium AD and travel routes of European seafarers during the Age of Discovery. In order to narrow down dispersal times, a more exact calibration of the molecular clock by adding more sequence data, ideally in combination with radiocarbon dating methods, will be necessary in future analyses.

2.5 References

Literature

Addison DJ & Matisoo-Smith E (2010) Rethinking Polynesians origins: a West-Polynesia triple-I model. *Archaeol. Oceania* **45**, 1-12.

Asouti E, Fuller DQ (2008) Trees and woodlands of South India. Archaeological perspectives. Walnut Creek (CA), Left Coast Press.

Auffray JC, Tchernov E, Nevo E (1988) Origine du commensalisme de la souris domestique (Mus musculus domesticus) vis-a-vis de l'homme. *C. R. Acad Sci. Paris* **307**, 517-522.

Auffray JC & Britton-Davidian J (2012) The house mouse and its relatives: systematics and taxonomy. In: Macholan M *et al*. (eds.) Evolution of the house mouse. Cambridge University Press, 1-34.

Barlow A, Baker K, Henry CR *et al*. (2013) Phylogeography of the widespread African puff adder (*Bitis arietans*) reveals multiple Pleistocene refugia in southern Africa. *Molecular Ecology* **22**, 1134-1157.

Beaujard P (2005) The Indian Ocean in Eurasian and African World-Systems before the Sixteenth Century. *Journal of World History* **16 (4)**, 411-465.

Berry RJ (1986) Genetical processes in wild mouse populations. Past myth and present knowledge. *Current Topics in Microbiology and Immunology* **127**, 86-94.

Berry RJ & Jakobson ME (1975) Adaptation and adaptability in wild living house mice. *Journal of Zoology* **176**, 391-402.

Bielejec F, Rambaut A, Suchard MA, Lemey P (2011) SPREAD: Spatial Phylogenetic Reconstruction of Evolutionary Dynamics. *Bioinformatics* **27**, 2910-2912. SPREAD version 1.0.6, available from www.kuleuven.ac.be/aidslab/phylogeography/SPREAD.html

Boivin N, Crowther A, Prendergast M, Fuller DQ (2014) Indian Ocean Food Globalisation and Africa. *African Archaeological Review* **31 (4)**, 547-581.

Bonhomme F & Searle JB (2012) House mouse phylogeography. In: Macholan M *et al*. (eds.) Evolution of the house mouse. Cambridge University Press, 278-296.

Boursot P, Auffray JC, Britton-Davidian J, Bonhomme F (1993) The Evolution of House Mice. *Annu. Rev. Ecol. Syst.* **24**, 119-152.

Boursot P *et al*. (1996) Origin and radiation of the house mouse: mitochondrial DNA phylogeny. *J. Evol. Biol.* **9**, 391-415.

Brothwell D (1981) The Pleistocene and Holocene archeology of the house mouse and related species. *Symposium of the Zoological Society of London. Vol. 47: The Biology of the House Mouse*, 1-13.

Castillo C, Fuller DQ (2010) Still too fragmentary and dependent upon chance? Advances in the study of early Southeast Asian archaeobotany. In: B. Bellina, E. Bacus, O. Pryce, J. Wisseman-Christie (eds.) Studies in honour of Ian Glover. River Books, Bangkok, 91-111.

Cheke A (2010) The timing of arrival of humans and their commensal animals on Western Indian oceanic islands. *Phelsuma* **18**, 38-69.

Cole NC (2005) The ecological impact of the invasive house gecko *Hemidactylus frenatus* upon endemic Mauritian geckos. Dissertation, Bristol.

Cucchi T, Vigne JD, Auffray JC (2005) The Passive transportation of the house mouse (*Mus musculus domesticus*) to Cyprus: New indirect evidence of intensive Neolithic navigation in Eastern Mediterranean. *Archaeozoology of the Near East VI. Proceedings of the sixth international symposium on the archaeozoology of southwestern Asia and adjacent areas*, 61-77.

Cucchi T, Auffray JC, Vigne JD (2012) On the origin of the house mouse synanthropy and dispersal in the Near East and Europe: zooarchaeological review and perspectives. In: Macholan M *et al*. (eds.) Evolution of the house mouse. Cambridge University Press, 65-93.

Din W, Anand R, Boursot P *et al*. (1996) Origin and radiation nuclear genes of the house mouse: clues from nuclear genes. *J. Evol. Biol* **9**, 519-539.

Drummond AJ & Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**, 214. BEAST version 1.8.0, available from http://code.google.com/p/beast-mcmc/

Drummond AJ, Ashton B *et al*. (2011) Geneious R6 version 6.0.6, available from http://www.geneious.com

Duplantier JM, 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.

Foerster DW, Guenduez I, Nunes AC *et al*. (2009) Molecular insights into the colonization and chromosomal diversification of Madeiran house mice. *Molecular Ecology* **18**, 4477-4494.

Fuller DQ, Boivin N (2009) Crops, cattle and commensals across the Indian Ocean: current and potential archaeobotanical evidence. *Etudes Ocean Indien* **42/43**, 13-46.

Gabriel S, Johannesdottir F, Jones EP, Searle JB (2010) Colonization, mouse-style. *BMC Biology* **8**, 131-133.

Geraldes A, Basset P, Gibson B *et al*. (2008) Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. *Molecular Ecology* **17**, 5349-5363.

Goios A, Pereia L, Bogue M *et al*. (2007) MtDNA phylogeny and evolution of laboratory mouse strains. *Genome Research* **17**, 293-298.

Google Earth (Version 5.1) Mountain View, CA: Google Inc. (2009). Available from http://www.google.com/earth/download/

Harper GA, Cabrera LF (2010) Response of mice (*Mus musculus*) to the removal of black rats (*Rattus rattus*) in arid forest on Santa Cruz Island, Galapagos. *Biological Invasions* **12**, 1449-1452.

Herman JS, McDevitt AD, Kawalko A et al. (2014) Land-Bridge Calibration of Molecular Clocks and the Post-Glacial Colonization of Scandinavia by the Eurasian Field Vole *Microtus agrestis*. *PloS One* **9**, 1-11.

Ho SYW, Larson G (2006) Molecular clocks: when times are a-changin'. *Trends in Genetics* **22**, 79-83.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755. Version 3.2.2, available from http://mrbayes.sourceforge.net

Jones EP, Jóhannesdóttir F *et al*. (2011) The expansion of the house mouse into north-western Europe. *J. Zool.* **283**, 257-268.

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.

Kumar S, Stecher G *et al*. (2012) MEGA-CC: Computing Core of Molecular Evolutionary Genetics Analysis Program for Automated and Iterative Data Analysis. *Bioinformatics* **28**, 2685-2686. Available from http://www.megasoftware.net

Lemey P, Rambaut A, Welch JJ, Suchard MA (2010) Phylogeography takes a relaxed random walk in continuous space and time. *Molecular Biology and Evolution* **27**, 1877-1885.

Love RS (2006) Maritime Exploration in the Age of Discovery – 1415-1800. Greenwood Publishing Group.

Lowe S, Browne M, Boudjelas S, De Poorter M (2000) 100 of the World's Worst Invasive Alien Species A selection from the Global Invasive Species Database. *The Invasive Species Specialist Group (ISSG) a specialist group of the Species Survival Commission (SSC) of the World Conservation Union (IUCN)*. School of Geography and Environmental Sciences (SGES), University of Auckland.

Musser GG & Carlton MD (1993) Family Muridae. In: Wilson DM & Reeder DM (eds.) Mammal species of the world, 2nd edition. Smithsonian Institution Press, Washington DC, 501-755.

PopART. Available from http://popart.otago.ac.nz

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.

Rajabi-Maham H, Orth A, *et al*. (2012) The south-eastern house mouse *Mus musculus castaneus* (Rodentia: Muridae) is a polytypic subspecies. *Biological Journal of the Linnean Society* **107 (2)**, 295-306.

Rando JC, Pieper H, Alcover JA (2014) Radiocarbon evidence for the presence of mice on Madeira Island (North Atlantic) one millennium ago. *Proc. R. Soc. B.* **281**, 20133126.

Suzuki H & Aplin KP (2012) Phylogeny and biogeography of the genus *Mus* in Eurasia. In: Macholan M *et al*. (eds.) Evolution of the house mouse. Cambridge University Press, 35-64.

Tangri D & Wyncoll G (1989) Of mice and men: is the presence of commensal animals in archaeological sites a positive correlate of sedentism? *Paleorient* **15**, 85-94.

Tchernov E (1968) Succession of rodent faunas during the Upper Pleistocene of Israel (Mammalia Depicta). Hamburg and Berlin.

Tchernov E (1984) Commensal Animals and Human Sedentism in the Middle East. *Animals and Archaeology: 3. Early herders and their flocks*, 91-115.

Tucker PK (2007) Systematics of the genus *Mus*. In: Fox JG, Barthold SW & Davisson MT (eds.) The Mouse in Biomedical Research, Vol. 1: Hostory, wild Mice, and Genetics. San Diego/CA, Academic Press Elsevier, 13-23.

Yonekawa H, Sato JJ *et al*. (2012) Origin and geetic status of *Mus musculus molossinus*: a typical example of reticulate evolution in the genus *Mus*. In: Macholan M *et al*. (eds.) Evolution of the house mouse. Cambridge University Press, 94-113.

Electronical sources (2014)

http://www.iucnredlist.org/details/13972/0
http://animaldiversity.org/accounts/Mus_musculus/

3 *Rattus rattus*

Reconstructing trans-oceanic dispersal routes: Phylogeographic analysis of the black rat *Rattus rattus* across the Indian Ocean

3.1 Abstract

The black rat (*Rattus rattus*) is a remarkably adaptable commensal species, nowadays occupying almost every corner of the globe. Its opportunistic lifestyle and close association with humans enabled progressive geographic expansion from their native range in Asia, leading to vast ecosystem damage. Despite its socio-economic impact, the taxonomy, evolutionary history, and invasive pathways of the black rat remain poorly understood. By combining molecular genetic, archaeological, and historical evidence it is possible to track the dispersal of *R. rattus*, which has been shaped by long-distance movements of humans, especially in the case of the widely scattered islands throughout the Indian Ocean.

Here, a 1,651 bp fragment consisting of the mitochondrial cytochrome b and the beginning of the d-loop from rat populations across the Indian Ocean was analysed to evaluate the phylogeographic structure in *R. rattus*. Results revealed Lineages I to IV of the '*Rattus rattus* Complex' (RrC), as defined by Aplin *et al*. (2011). Whereas the Lineages II, III, and IV remain in their distribution area, Lineage I is being transported to various other places outside its native range in India. Thereby, several unique geographically restricted phylogenetic clades have been identified, that may reflect early human travel and trade routes throughout the region.

3.2 Introduction

The black rat *Rattus rattus*, also known as 'house rat', 'roof rat' or 'ship rat', is a remarkably adaptable species and therefore has become the most widely distributed of all commensal animals since it started moving with humans from the early first millennium AD (Lund 1994, Aplin *et al*. 2011). Their ability to invade into different kinds of habitats around the globe – from coastlands, forests and rural environments to agricultural and urban areas, has led to vast ecosystem damage (Jones *et al*. 2008, http:// www.issg.org/ database/ species/ ecology.asp?si =19&fr =1&sts =sss&lang =EN) (see figure 3.1). Not only has the black rat become pests to human home and agriculture, but is also a transmitter of zoonotic diseases (Duplantier *et al*. 2005) and responsible for the extinction of local species through competition and predation (Aplin *et al*. 2003).

Figure 3.1: Current distribution of R. rattus.
(modified, after IUCN distribution map: http://www.iucnredlist.org/details/19360/0)

Despite its socio-economic impact, the taxonomy, evolutionary history, and invasive pathways of the black rat remain poorly understood (Aplin *et al*. 2003, Aplin *et al*. 2011). Several studies (Searle *et al*. 2009, Matisoo-Smith & Robins 2009) demonstrate that areas of origin, subsequent dispersal routes and the time frame of

colonisation events of invasive species can be inferred from molecular genetic data through phylogeographic methods (Avise 2000). Improving the understanding of pest species is necessary in order to apply conservation and health implications to control their destructive impact.

The taxonomy of *Rattus* is complex and remains highly controversial. Currently the genus contains about 65 species (Musser & Carlton 2005), but deciphering the phylogenetic history of rats is often challenging due to the difficulties in their morphological distinction (Robins *et al*. 2007). Because of their fragility, the bones of rodents are prone to decay by biotic and abiotic processes, which leads to poor preservation and as a result to misidentifications, e.g. in museum collections (Cucchi *et al*. 2005, Weissbrod *et al*. 2005). Furthermore, the abundance of synonyms for each species and the degree of taxonomic revisions reveal the complications assigning phylogenetic relationships (Musser & Carlton 2005).

The genus *Rattus* has its origin in South-East Asia, by which the black rat is native to the Indian subcontinent. Outside Asia determination of its natural distribution remains uncertain due to a general lack of fossil data and the already discussed problem of misidentifications. From that area, a dispersal in westward direction through Asia to the Mediterranean occurred at the beginning of the Holocene (Ervynck 2002). So far, several cytological (Yosida 1980, Baverstock *et al*. 1983) and morphological (Niethammer 1975) studies of the black rat indicate the distinction in two separate species: *R. rattus* for European and Indian populations (karyotype of 2n = 38-40) and *R. tanezumi*, restricted to South-East Asia (karyotype of 2n = 42).

Furthermore, Aplin *et al*. (2011) undertook a molecular genetic survey to investigate the evolutionary history of *R. rattus*, including their geographic pattern as well as direction and timing of dispersals. Comparison of mitochondrial cytochrome b sequences of 165 individuals mainly sampled from South and South-East Asia showed four well-supported lineages I-IV, designated as the '*Rattus rattus* Complex' (RrC) (see figure 3.2). By that means, Lineage I is present in western India and shows the broadest distribution outside of mainland Asia with representation in

Europe, the Americas, Africa and Madagascar, as well as Australia and various Pacific Islands. Lineage I has its maximum genetic diversity in southern India, and statistical analysis shows a stable and long term residency population structure. Outside of India haplotypes cluster in star like patterns suggesting a recent population expansion. Two groups are apparent here, comprising black rat haplotypes that were introduced to different parts of the world during the Age of Exploration, as well as haplotypes exclusively confined to Madagascar. Lineage II is found from eastern India through Myanmar, northern Laos and Vietnam as well as southern China. Linage III is common in the Himalayan foothills of Pakistan and Nepal, while Lineage IV occurs in the lower Mekong River catchment in southern Laos, Thailand and southern Vietnam. Further, Lineage II and IV overlap on mainland Asia in central Laos and Thailand.

The first three Lineages I-III form a monophyletic group, therefore labelled as Clade A. However, Lineage IV of the RrC belongs to a separate Clade B that further includes rats from Thailand and Laos identified as *R. losea* (Lineage V), as well as representatives of *R. tiomanicus* and *R. baluensis*, which form Lineage VI (Aplin *et al*. 2011).

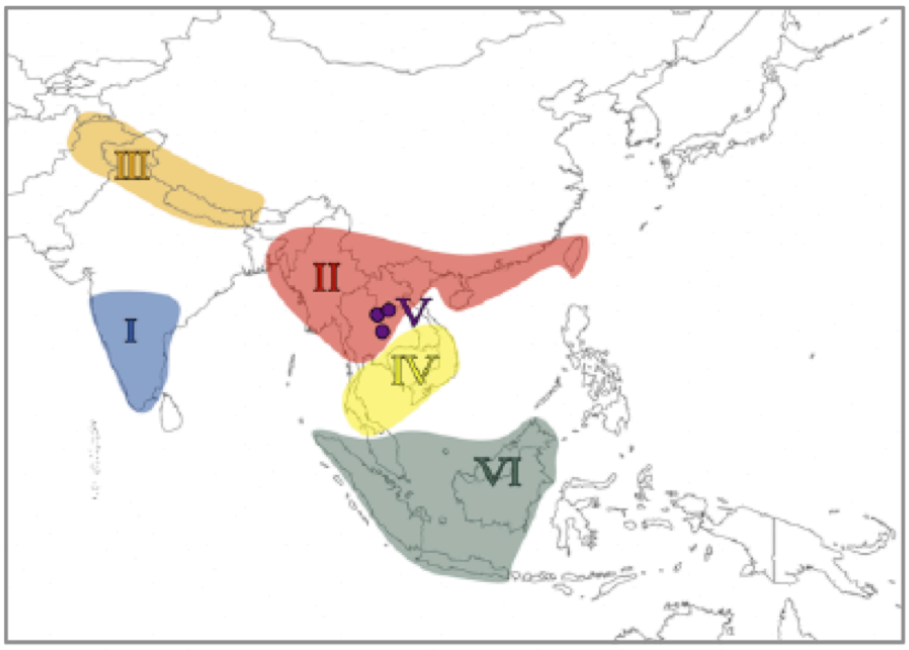

Figure 3.2: Natural ranges of Lineages I to VI, inferred by Aplin et al. (2011).
Lineage I is indicated in blue, Lineage II in red, Lineage III in orange, Lineage IV in yellow, Lineage V in purple, Lineage VI in green.

The aim of the study presented here is to fill in the gaps of the work of Aplin *et al*. (2011). While he and his colleagues focused on the rat dispersal in South-East Asia, we intended to unravel western migration routes through the Indian Ocean area. So far, the archaeological record of the widely scattered islands throughout the Indian Ocean is scarce and leaves an ambiguous picture of human discovery and settlement. Hingston *et al*. (2005) were the first to try to untangle the colonisation history of Madagascar, the largest island in the Indian Ocean, by looking at genetic signatures of *R. rattus*. Results of the analysis showed consistency with an Indian origin of southern Malagasy populations, although a lack of samples from the East African coast, the Arabian peninsula and northern Madagascar prevented a validation of this hypothesis. In 2010, Tollenaere *et al*. tried to resolve the question whether the black rat came via trans-oceanic dispersal from Asia or over the East African land route to Madagascar. Results of the phylogeographic pattern confirm that Malagasy black rats share haplotypes with specimens from either Indian or Arabian source populations and were probably introduced by a single, human-mediated colonisation event. Despite their vicinity, other islands along the East African coast seem to have a different population history. Thereby, Mayotte acquired its rats from Madagascar, while Grande Comore appears to have been colonised independently from East Africa, and rats on Reunion probably stem from Europe. Still unclear is where the small founder population on Mauritius originated. Rats from Rodrigues and the Seychelles presumably dispersed from Mauritius, but have yet to be studied in detail, because they show different colour patterns in their pelage (Tollenaere *et al*. 2010, Cheke 2010).

Currently, the taxonomy and the dispersal routes of *R. rattus* are still debated as only a few molecular genetic studies have been carried out to resolve these questions. Recent studies particular emphasise on South-East Asia (Aplin *et al*. 2011) and Madagascar in the western Indian Ocean (Hingston *et al*. 2005, Tollenaere *et al*. 2010). In order to get a more complete picture of the phylogeography and dispersal routes of the black rat within the whole Indian Ocean area, we continued the research of Aplin *et al*. (2011) by extending the sampling area from South-East Asia over the Arabian peninsula to the East African coast, including samples from various islands

throughout the Indian Ocean. Furthermore, we increased the length of sequence data by adding a further genetic marker – the mitochondrial d-loop – in order to get a higher resolution and finer substructure within the *R. rattus* topology.

3.3 Material and Methods

3.3.1 Samples

Overall, 292 tissue samples of *R. rattus* were analysed, whereas 263 specimens were obtained from various museums (Field Museum, Chicago; American Museum of Natural History, New York; British National History Museum, London) and additional 29 modern specimens were collected in the field by Dr. J. Chris Hillman. The museum material is composed of mainly dried tissue, skins or ethanol-fixed samples, modern specimens were stored in ethanol. The sampling area comprises different places of the mainland and islands around the Indian Ocean, including countries from the East and West African coast, the Arabian peninsula, as well as the Indian subcontinent and South-East Asia (see figure 3.3). A list of all samples used in this study is found in chapter 7.2.2, including additional information.

This study is a joint project between Heidi Eager (HE) and myself (AT). The first set of primers was developed by AT, the second set of primers was developed by AT and HE. Modern rats from Diego Garcia have been extracted and amplified by AT, modern material from Africa has been extracted and amplified by HE. The museum material was jointly extracted. Subsequent amplifications of the museum material were carried out individually by AT and HE on different days using different reagents, in order to obtain replicates for most of the individuals. Results of the DNA extraction and amplification success is shown in figure 3.4.

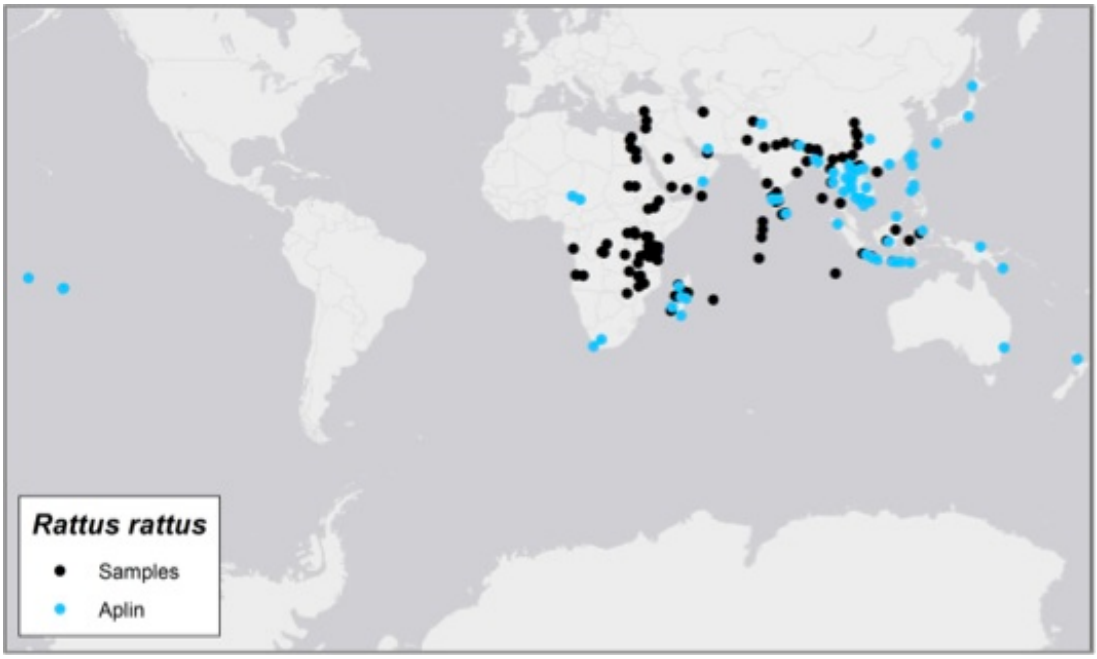

Figure 3.3: Sample distribution of R. rattus.
Distribution of sampling locations of R. rattus specimens used in this study (black) and samples locations from Aplin et al. (2011) (blue).

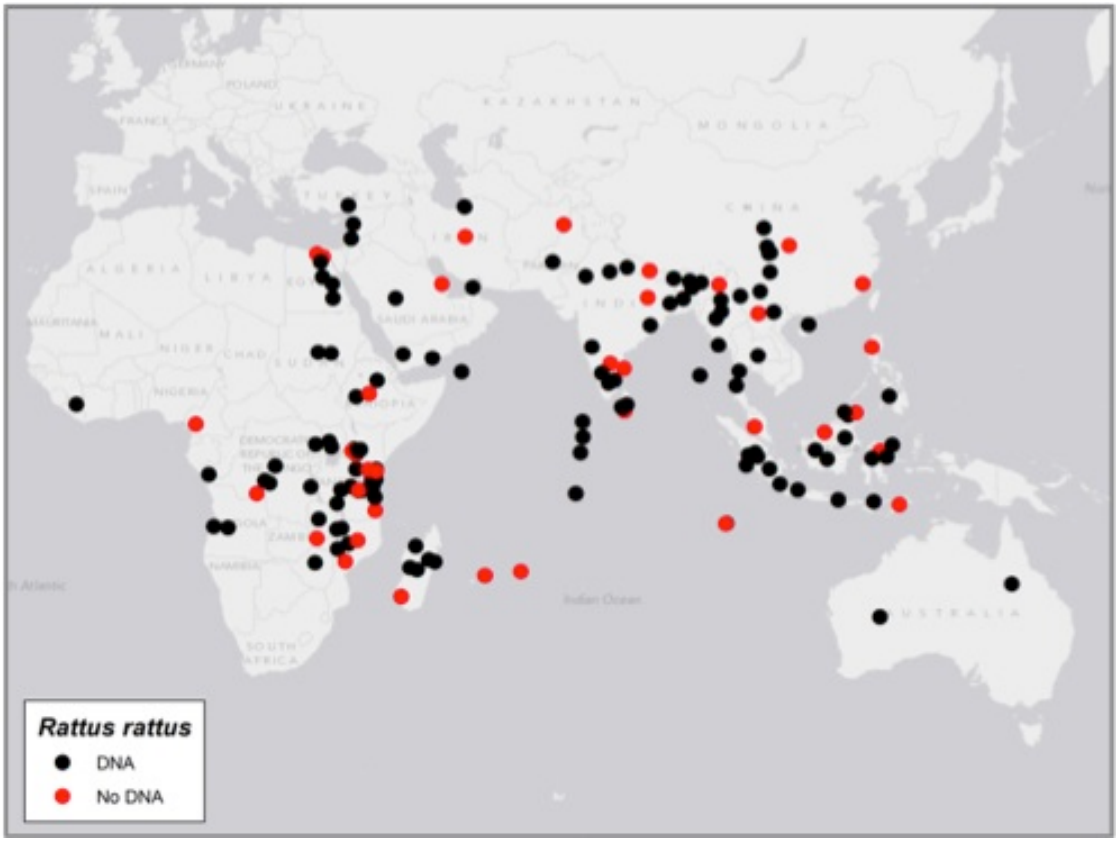

Figure 3.4: Successful DNA amplification of R. rattus samples.
Successful DNA extraction and amplification of R. rattus specimens used in this study is indicated in black, unsuccessful DNA extraction is indicated in red.

3.3.2 Analyses

All samples were analysed in the facilities at the Archaeology Department of Durham University. Because of the recent age of most of the specimens, the samples were treated as modern material. Each workstep – from DNA extraction to sequencing set-up – was conducted in the modern laboratory, following the protocols described in chapters 7.1.2 (extraction protocol 1 and 2), 7.1.3 and 7.1.4. The sequencing reaction was carried out by the DNA Sequencing Service at the School of Biological and Biomedical Sciences at Durham University. Mitochondrial DNA was amplified in 10 overlapping fragments for cytochrome b and 4 overlapping fragments for the beginning of the control region, whereas a variety of primer combinations was used depending on the nature of the sample (see chapter 7.1.1). The sequencing chromatograms were edited manually, subsequently assembled, and a consensus sequence per individual exported (see chapter 7.1.4) using Geneious R6 version 6.0.6 (Drummond *et al*. 2011). Standard anti-contamination guidelines were followed to ensure authenticity of the DNA results (see chapters 1.4.3 and 7.1.5).

Phylogenetic analyses of two different datasets were performed by the construction of Bayesian trees using BEAST version 1.8.0 (Drummond & Rambaut 2007). First, a topology of 322 sequences of the cytochrome b gene with a length of 948 bp was created. The best-fit nucleotide substitution model, selected in MEGA-CC (Kumar *et al*. 2012) under the Bayesian Information Criterion (BIC), was Kimura-2-parameter +Γ +I (see figure 3.5). Furthermore, an alignment of the combined loci, cytochrome b and d-loop, was used. The dataset consisted of 192 sequences with a length of 1,651 bp. After partitioning, the appropriate substitution model has been applied to cytochrome b (Kimura-2-parameter +Γ +I) and d-loop (Tamura-3-parameter +Γ), respectively. After initial runs, the nucleotide substitution model for cytochrome b was changed to the second-best fit and less complex Kimura-2-parameter +Γ, as the trees did not converge with the best-fit model. For each dataset, two independent runs of 50,000,000 generations each were performed with a sample frequency of 5,000. For the dataset containing cytochrome b, a 10% burn-in period of each run was applied. For the combined dataset containing cytochrome b and d-loop a 10%

burn-in period and a 20% burn-in period was applied, respectively. Subsequently, the two runs for each dataset, which converged on same values, were combined to a consensus tree each (see figure 3.7).

Additionally, the generation of a median-spanning network of *R. rattus* specimens with PopART (http://popart.otago.ac.nz), based on the number of differences between sequences, revealed 86 different haplotypes among a total sample set of 139 sequences (see figures 3.9 and 3.10). Due to the inability of networks to deal with missing data, the original *R. rattus* alignment of the combined loci cytochrome b and d-loop had to be truncated to a 1,371 bp fragment.

3.3.3 Results

Analyses of the cytochrome b gene

DNA sequences of a 948 bp fragment of the cytochrome b gene were obtained from 202 of the 292 modern and museum samples. Additionally, 120 sequence entries of *R. rattus* from Aplin *et al*. (2011) were obtained from GenBank and combined to a dataset composed of 322 individuals. Molecular-clock based Bayesian analyses generated a well-resolved tree with posterior probability values around 1 leading to the main clades (see figure 3.5). All samples analysed here fall into Lineages I to IV, the '*Rattus rattus* Complex' (RrC), as defined by Aplin *et al*. (2011). In this previous study, each lineage is restricted to a certain area: Lineage I (blue) has the broadest distribution outside the native range but is supposed to have its origin in southern India. Lineage II (red) is found from eastern India through Myanmar, northern Laos and Vietnam as well as southern China, whereas Lineage III (orange) primarily occurs in the Himalayan foothills of Pakistan and Nepal. Lineage IV (yellow) is distributed from southern Laos, over Thailand to southern Laos.

The distribution pattern of the *R. rattus* samples used in this study show congruence for the Lineages II, III, and IV, although Lineages II and IV slightly expand their range. On the contrary, Lineage I spreads from the tip of India throughout the Indian Ocean to Madagascar, along the East African coast inland and into the Arabian peninsula. Further, specimens are found in westward direction on Andaman and the Cocos Keeling Islands (see figure 3.6).

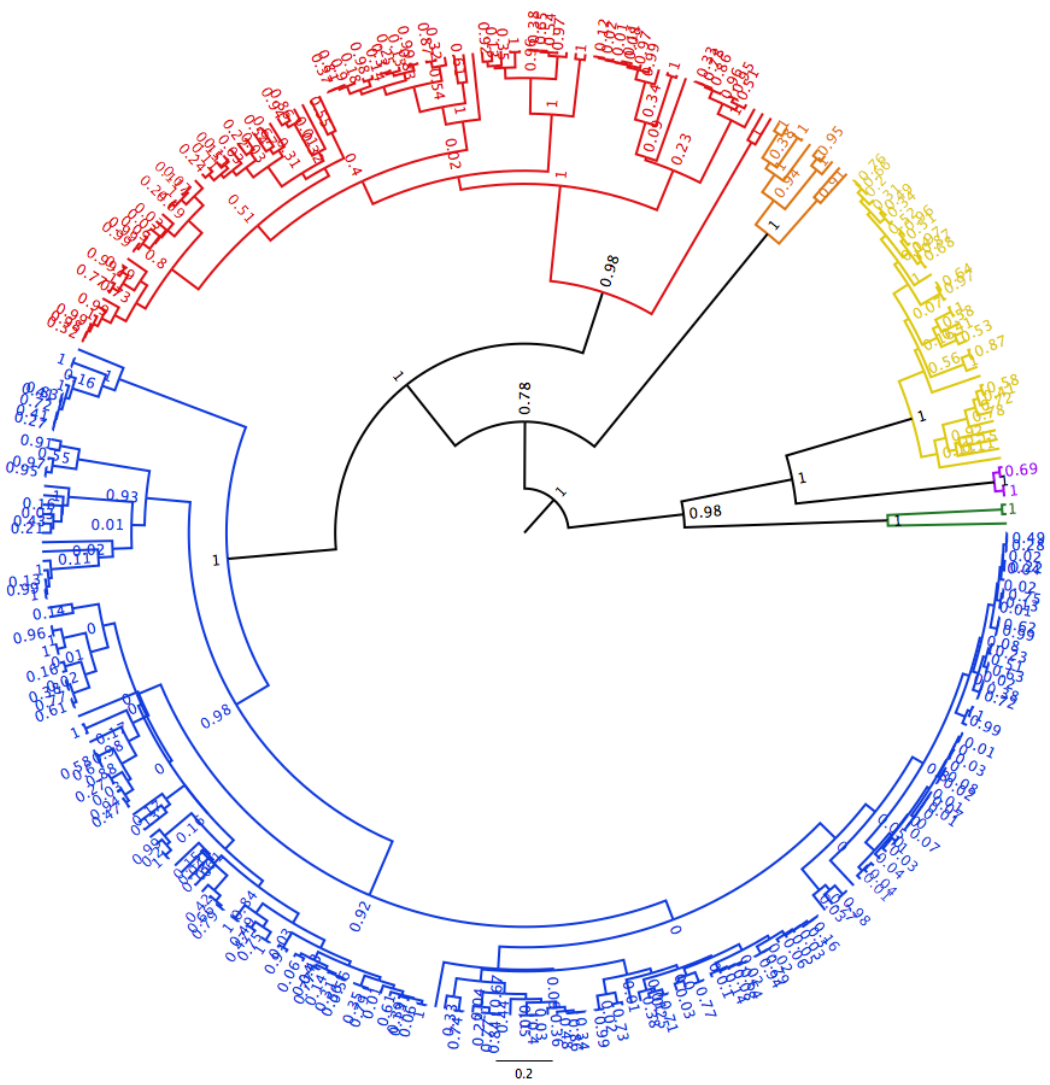

Figure 3.5: Molecular-clock based mitochondrial phylogeny of R. rattus inferred from a 948 bp fragment of cyt b.

Inferred from a 948 bp fragment of the mitochondrial cyt b using BEAST version 1.8.0. The best-fit nucleotide substitution model under the Bayesian Information Criterion (BIC), was Kimura-2-parameter +Γ. Two independent runs of 50,000,000 generations each were performed with a sample frequency of 5,000 and a 10% burn-in period and then combined to a consensus tree. Lineages I to VI according to Aplin et al. (2011) were recovered. Lineage I is indicated in blue, Lineage II in red, Lineage III in orange, Lineage IV in yellow, Lineage V in purple, Lineage VI in green.

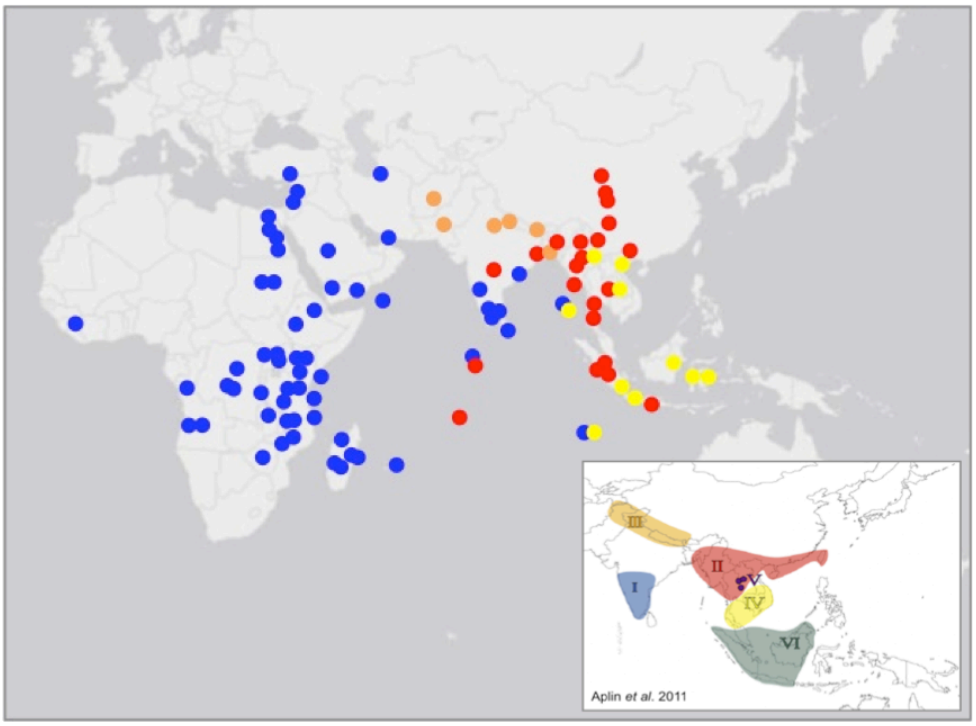

Figure 3.6: Haplogroup distribution of R. rattus inferred from a 948 bp fragment of cyt b.
The samples used in this study group with Lineages I to IV identified by Aplin et al. (2011). Lineage I is indicated in blue, Lineage II in red, Lineage III in orange, Lineage IV in yellow.

Analyses of the combined dataset of the cytochrome b gene and the d-loop

A 1,651 bp fragment containing the combined sequences of the mitochondrial cytochrome b and the d-loop of 192 specimens was used in further analyses. Molecular-clock based Bayesian analyses generated a well-resolved tree, with high posterior probability values around 1 for the main clades, but with less high values for the subclades (see figure 3.7).

In figure 3.8, the distribution of the different subclades of Lineage I is shown. Lineage II, III, and IV do not show a further subdivision in their distribution range. In contrast, Lineage I is divided in 5 subgroups (dark blue, mid blue, grey/black-rimmed, light blue/black-rimmed, turquoise/white-rimmed), which can be assigned to certain geographical areas along the East African coast and Arabian peninsula.

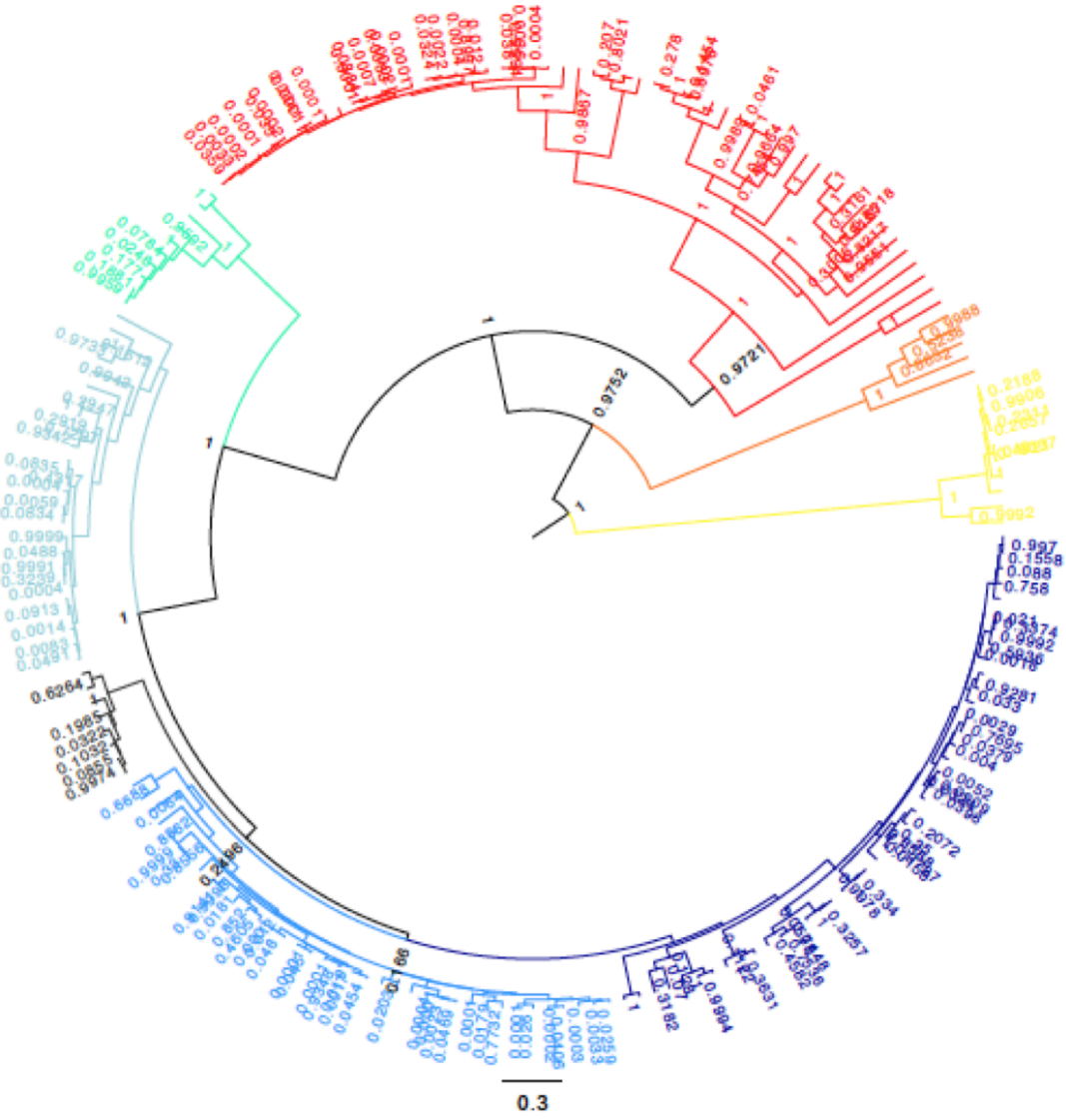

Figure 3.7: Molecular-clock based mitochondrial phylogeny of R. rattus inferred from a 1,651 bp fragment of cyt b and d-loop.

Inferred from a 1,651 bp fragment of cyt b and d-loop. After partitioning the dataset best substitution models were Kimura-2-parameter +Γ for cyt b and Tamura-3-parameter +Γ for the d-loop. Two independent runs of 50,000,000 generations each were performed with a sample frequency of 5,000 and a 10% and 20% burn-in period, respectively, and then combined to a consensus tree. The samples used in this study group with Lineages I to IV identified by Aplin et al. (2011). The dark blue, mid blue, grey, light blue, and turquoise groups are subclades within Lineage I. Lineage II is marked in red, Lineage III is marked in orange, Lineage IV is marked in yellow.

Whereas the dark blue clade spreads from India throughout the Arabian peninsula and along the eastern and western African coasts, the mid- and light blue/black-rimmed clades are more confined to Madagascar, the East African coast and African inland, respectively. Furthermore, one population only occurs in Egypt and on Socotra (grey/black-rimmed), another group is restricted to Sri Lanka and the Andaman Islands (turquoise/white-rimmed).

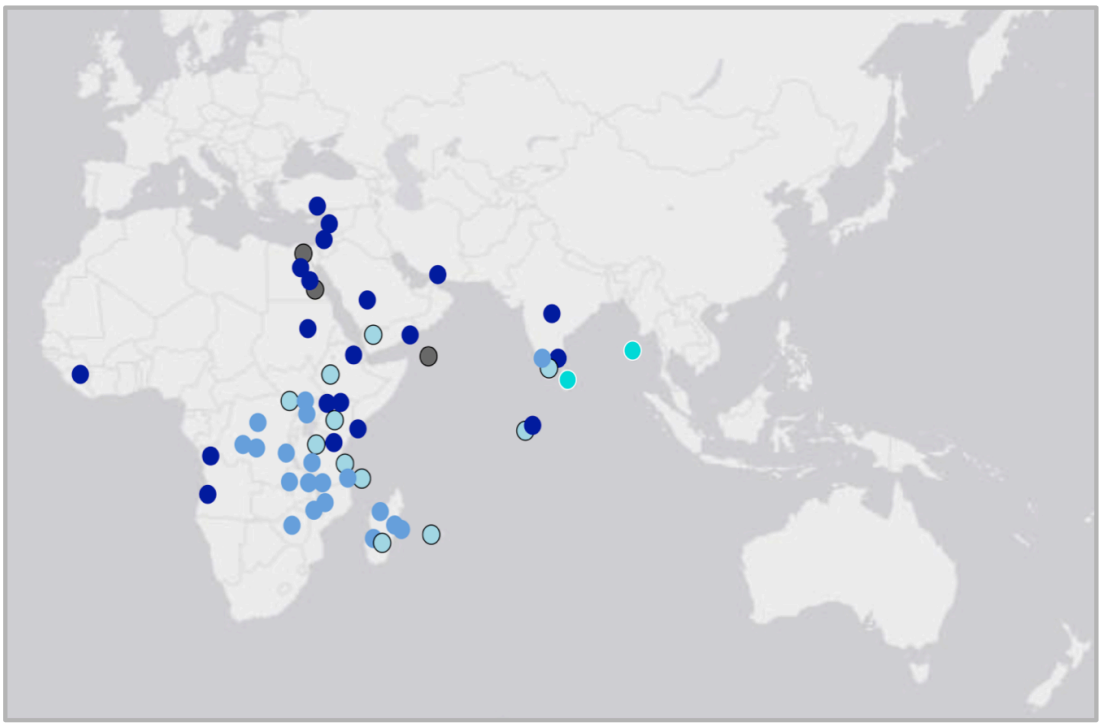

Figure 3.8: Haplogroup distribution of R. rattus Lineage I after Aplin et al. (2011) inferred from a 1,651 bp fragment of cyt b and d-loop.
Distribution of the combined dataset of cyt b and d-loop used in this study with focus on the subgroups within Lineage I identified by Aplin et al. (2011).

Results of the network analysis revealed 86 haplotypes in a total of 139 sequences (see figures 3.9 and 3.10). Thereby, haplotypes R1 and R2 show the highest frequencies. Haplotype R1 is confined to Diego Garcia, located off the Indian peninsula, and its high frequency results from the large number of samples collected from this location. Further, R2 is distributed throughout the East African coast in countries such as Tanzania, Zambia, Malawi, and Madagascar. Additionally, R4 and R9 are shared haplotypes between India and Madagascar, as well as Uganda, Tanzania, and Kenya, respectively. The haplotype with the most connections is R16

from Yemen and links to a number of haplotypes from regions such as Egypt (R7, R29), Tanzania (R3, R47), Madagascar (R20, R21), Maldives (R81), Iran (R78), India (R37), and Sri Lanka (R42). Locations with highest haplotype diversity are Indonesia, China, India and Sri Lanka, as well as Tanzania and Madagascar.

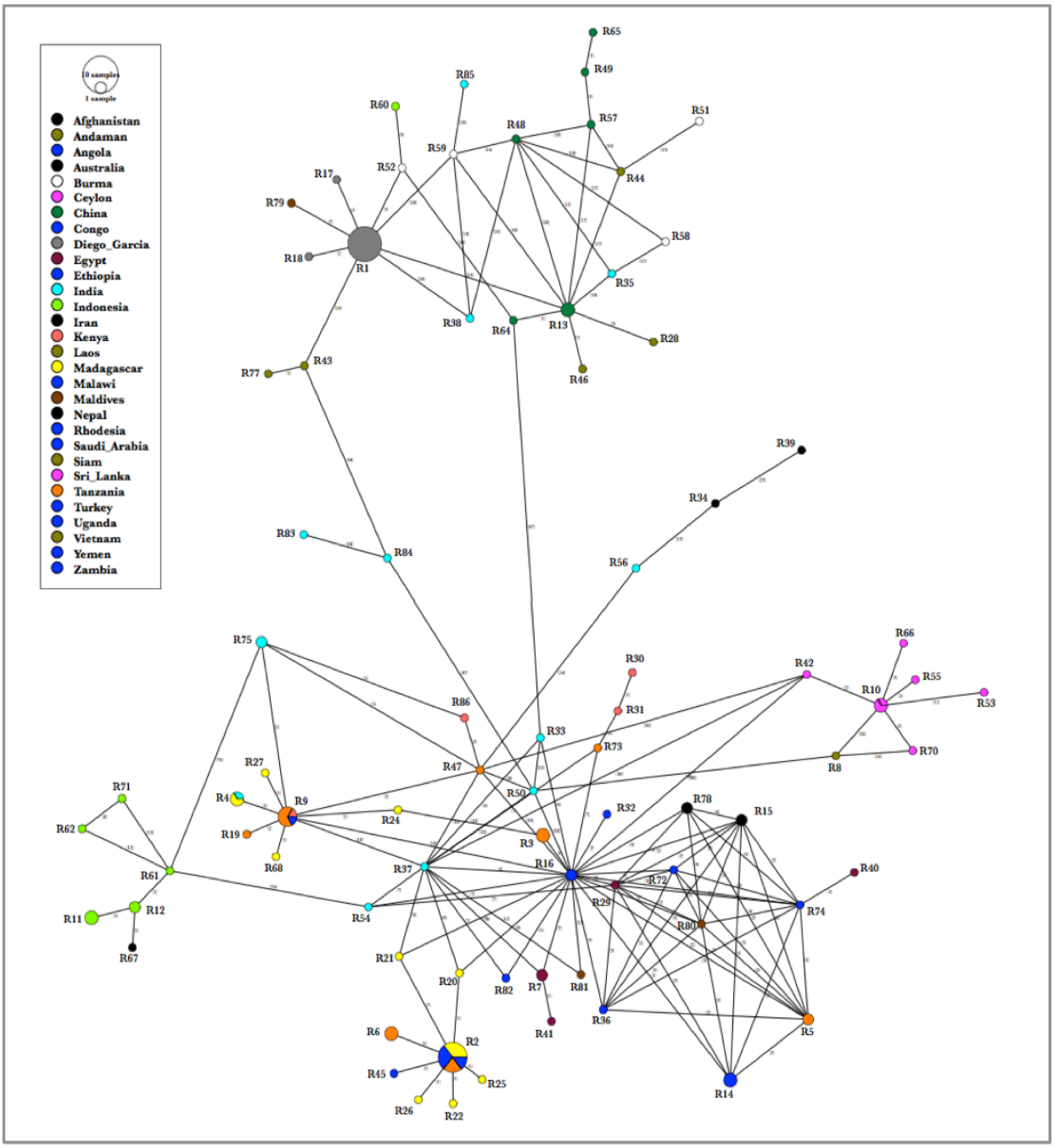

Figure 3.9: Network analysis of R. rattus using PopART.
Haplotype diversity of R. rattus throughout the region, assessed from a 1,371 bp fragment of the mitochondrial cytochrome b and d-loop. 139 sequences have been included and revealed 86 haplotypes.

Country	Haplotype
Afghanistan	R34
Andaman	R8
Angola	R14, R72
Australia	R67
Burma	R51, R52, R58, R59
China	R13, R48, R49, R57, R64, R65
Congo	R74
Diego Garcia	R1, R17, R18
Egypt	R7, R29, R40, R41
Ethiopia	R82
India	R4, R33, R35, R37, R38, R50, R54, R58, R75, R83, R84, R85
Indonesia	R11, R12, R60, R61, R62, R71
Iran	R15, R78
Kenya	R9, R30, R31, R86
Laos	R28
Madagascar	R2, R4, R20, R21, R22, R24, R25, R26, R27, R68
Malawi	R2
Maldives	R79, R80, R81
Nepal	R39
Rhodesia	R45
Saudi Arabia	R32
Siam	R44, R46
Sri Lanka/Ceylon	R10, R42, R53, R55, R66, R70
Tanzania	R2, R3, R5, R6, R9, R19, R47, R73
Turkey	R36
Uganda	R9
Vietnam	R43, R77
Yemen	R16
Zambia	R2

Figure 3.10: Table of the haplotype diversity of R. rattus.

3.4 Discussion

The black rat is an excellent example of a commensal species that progressively expanded geographically through its opportunistic lifestyle and close association with humans. By combining molecular genetic, archaeological, and historical evidence it is possible to track the dispersal of *R. rattus*, which has been shaped by long-distance movements of humans, especially in the case of the widely scattered islands throughout the Indian Ocean. The aim of this study is to continue the research of Aplin *et al*. (2011) by extending the sampling area from ISEA throughout the Indian Ocean to the East African coast. Further, more mitochondrial sequence data of the black rat was generated by adding an extra marker, in order to reveal a finer substructure of its phylogeography.

Previous studies mainly focused on the spreading of *R. rattus* out of their presumed native range in India to Europe during Roman and medieval times (Armitage 1994), on the colonisation history of Madagascar (Hingston *et al*. 2005, Tollenaere *et al*. 2010), as well as the timing and direction of historical dispersals in ISEA (Aplin *et al*. 2011). Results of the mtDNA variation pattern are generally consistent with the archaeological record, but reveal a finer substructure of the range extension.

Here, analyses show that adding an extra marker not necessarily leads to a more clear and finer substructure within the topology of *R. rattus*. In both Bayesian trees – either containing sequences of the cytochrome b gene only, or a combination of cytochrome b and d-loop – samples group in identical clades. The samples analysed here fall into Lineages I to IV of the '*Rattus rattus* Complex' (RrC), as defined by Aplin *et al*. (2011). Greatest diversity is detected on the Indian subcontinent, the expected native range. These results are consistent with observations by Aplin *et al*. (2011) of a long history of a constant and stable population in this area. Whereas the Lineages II, III, and IV remain in their distribution area, Lineage I is being transported to various other places outside its native range in India. For the combined data-set (cyt b + d-loop), five unique geographically restricted subgroups have been

identified within the globally distributed Lineage I, each reflecting different human trade and travel routes throughout the region.

In the course of the 1st millennium AD, the rise of culturally diverse, urban settlements along the East African coast from Somalia to Mozambique and on the offshore islands Zanzibar, Pemba, Mafia, Madagascar, and the Comoros has been observed. The emergence of the so-called Swahili culture is characterised by people with a subsistence based on agriculture and fishing, further a shared language (belonging to the Bantu-language-family) and religion (Islam) (Horton & Middleton 2000, Boivin *et al*. 2013). In order to differentiate themselves from their neighbours, the Swahili established trade relations with a range of different partners. Excavations of early occupation phases revealed remains of Sassanian Islamic ware from the Near East, but also Chinese pottery, as well as stone and glass beads from the Gulf and India, demonstrating frequent trade activities across the Indian Ocean. In turn, predominantly ivory, timber, and slaves have been exported from the East African coast (Horton & Middleton 2000, LaViolette 2008). By the 10th century AD, the increase in trade activities has led to growing wealth and subsequently to the rise of important urban centres and port cities. Thereby, archaeological evidence suggests a specialisation of certain port cities, favouring a connection with either Indian, Arabian or Persian merchants (Horton & Middleton 2000). Furthermore, the Swahili opened up an additional trading sphere towards the African inland, moving goods in westward directions (Horton & Middleton 2000, LaViolette 2008). The intensification of these commercial relations have facilitated the spread of small commensal animals, which will be illustrated in the following paragraphs.

Dark blue clade. Samples of the dark blue clade belong to a group that has been moved to many places outside their native range and appear in India, the Near East (Egypt, Yemen, Saudi Arabia, Jordan, Lebanon, Turkey), and mainland East Africa (Tanzania, Kenya), but also along the West African coast (Liberia, Angola). Low genetic diversity among this group observed by Aplin *et al*. (2011) and short branch lengths within the tree reported here, indicate a rather recent introduction to the areas. The observation of Aplin *et al*. (2011) is only partly confirmed by the results

of the network analysis that revealed single haplotypes for the locations Yemen (R16), Saudi Arabia (R32), Turkey (R36), as well as Zambia (R2), Uganda (R9), Rhodesia/Zimbabwe (R45), Malawi (R2) or Congo (R74). Higher haplotype diversity has been observed in Kenya, Egypt, and especially Tanzania (see figure 3.10). Furthermore, Yemen shows multiple connections to various other haplotypes, rather indicating a less recent introduction.

Archaeological records suggest a spreading of *R. rattus* to the Middle East and subsequently to Europe from the Late Pleistocene onwards (Aplin *et al*. 2011). Here, the earliest evidence of black rat remains linked to human activities has been found in Natufian layers of cave sites in the Levant probably dating around 15,000 to 11,000 BP. Due to a further range expansion, the black rat has reached the Mediterranean around 6,000 BP (Ervynck 2002). In the course of the following millennia, skeletal remains of this rodent are scarce. Within the Euphrates-Tigris-basin the black rat appears at urban sites dating from the late 3rd to the 1st millennium BC (Ervynck 2002). The long gap between finds from the Levant and the Euphrates-Tigris-basin may be explained by a bias in sampling or bone preservation, or by a change of habit in humans. As rats are highly commensal and dependent on humans, they may have disappeared when humans stopped using caves as dwelling places. The re-appearance of the black rat a few millennia later, may be connected with the increase of urban structures at that time, which provided a suitable niche for the rodents. Although rats utilise various habitat forms, they show a preference for drier environments and favour to live or nest in high places, such as trees and rocks in the wild or roofs within urban structures (see chapter 1.3.2). It is therefore possible that past distribution structures are not accurately reflected by mitochondrial marker, as ancient populations may have gone extinct or have been replaced by successful female invaders.

The current distribution of the dark blue clade has been shaped by recent trade and travel activities. Although archaeological finds are present in the Near East since the Late Pleistocene, earliest skeletal remains of the black rat along the East African coast are significantly younger (Ervynck 2002). While *R. rattus* has been reported

from classical Egyptian sites dating to the end of the 1st millennium BC, it occurs in Zambia and South Africa only between the 9th to 12th century AD (Voigt & von den Driesch 1984, Plug 2000, Helm 2000, Boivin *et al*. 2013). By that time, trade networks between small-scale hunter-gatherer and fishing communities have been developed into frequent trans-regional exchanges between growing port cities. These connections interlinked India over sea- and land-routes with East Africa and the Mediterranean (Beaujard 2005). Thereby, Arab merchants dominated trade ways through the Arabian Sea, the Persian Gulf, to India and further south along the East African coast during the 1st millennium AD. Most likely the black rat has been transported through these maritime corridors, before European explorers sailed across the Indian Ocean a few centuries later (Love 2006, Boivin *et al*. 2014). Within the network analysis, this is demonstrated by the haplotype with the most connections R16 from Yemen, that links to a number of haplotypes from regions such as Egypt (R7, R29), Tanzania (R3, R47), Madagascar (R20, R21), Maldives (R81), Iran (R78), India (R37), and Sri Lanka (R42).

In the course of the 16th century, European explorers started their search for a direct sea-route to Asian markets in order to circumvent Italian and Muslim middlemen, who controlled the overland trade routes across the Mediterranean at that time. Initially, Portuguese mariners sailed along the West African coast to reach the Indian Ocean, French, English, and Dutch ships followed later (Love 2006). Haplotypes of the dark blue clade appear along the West African coast and group with sequences from Senegal, Guinea, and France identified by Aplin *et al*. (2011), pointing to a distribution through European seafarers during the Age of Discovery. Archaeological remains of the black rat have been found in European sites between 3,100 BP (Switzerland, Spain) and 2,500 BP (Poland) (Evynck 2002). But only during the 16th to 18th century AD, the establishment of long-distance trading routes between Europe, Africa, and Asia has led to the occupation of territory along the West African coast, and therefore enabled the transmission of small commensal animals to new areas by European explorers (Love 2006, Cheke 2010). A more recent distribution of *R. rattus* is further supported by a lack of archaeological finds of *R. rattus* along the West African coast in prehistoric times. Furthermore, a direct

seaborne dispersal of this particular group is assumed due to the lack of specimens across the African inland belonging to the dark blue clade.

Light blue/black-rimmed clade. A further group identified among the Lineage I rats is the light blue/black-rimmed clade. Their distribution range partly overlaps with the dark blue clade, as they appear in India, the Arabian peninsula and further along the East African coast, as well as on islands throughout the Indian Ocean, such as Madagascar, Mafia, Zanzibar, the Maldives, and Mauritius. According to the short branches within the phylogenetic tree, the pattern of their spreading along East Africa similarly may have been shaped in recent times, probably by Arab traders.

However, the introduction to the different islands occurred through diverse pathways. Black rats probably arrived on Madagascar with the first urban settlers from the 9th century onwards (Hingston *et al*. 2005). Tollenaere *et al*. (2010) tried to resolve the question whether the black rat came via a direct sea route from Asia or over the East African land route. Results confirmed that Malagasy black rats share haplotypes with specimens from either Indian or Arabian source populations. Here, results are consistent with previous observations of multiple introductions from diverse locations. Within this study, specimens on Madagascar share haplotypes with both, Indian (haplotype R4 within the network) and mainland African populations from Uganda, Tanzania, and Kenya (haplotype R2 within the network). During the 1st millennium AD, the development in seafaring technologies and the establishment of urban settlements and ports has led to frequent contacts across the region. Thereby, the East African mainland and the islands off the coast mark important trading points, which is demonstrated by the human-mediated translocation of not only rats, but various other animal and plant species (Boivin *et al*. 2013).

A similar pattern is observed on Zanzibar, where haplotypes of *R. rattus* cluster together with rats from Tanzania. Here, earliest finds of the black rat have been excavated at archaeological sites dating to the 6th to 8th century AD (Helm 2000). Furthermore, the connection between the African mainland and Zanzibar from earliest times onwards is supported by the introduction of native western African cereals such as sorghum, pearl and finger millet first to East Africa, reaching the

offshore islands Pemba and Zanzibar probably by the 7th century AD (Helm *et al*. 2012). Additionally, Mediterranean pottery and blue-green glazed Sassanian-Islamic ceramics dating to the 5th and 6th century AD have been found at one of the earliest sites on Zanzibar, Unguja Ukuu, as well as on the East African mainland in Egypt and further south in Tanzania (Horton & Middleton 2000). The introduction of the black rat to the island Mafia, located off the East African coast, is not confirmed but may be during more recent times (Cheke 2011). Here, specimens on Mafia cluster with black rats from Tanzania.

Mauritius has been an important stopping point for European seafarers and was settled by the Dutch in 1598 AD. However, the black rat has been present on this island prior to the first recorded landing (Cheke 2010). Within the phylogenetic tree of this study, the Mauritian specimen appears as a single haplotype. These isolated haplotypes may represent single natural introductions where rats on wreckage get carried to new places by oceanic currents (Cheke 2010). A similar scenario is also assumed for the single haplotype appearing on the Maldives. Further sampling across the region may help to unravel source populations for those individual introductions.

Mid blue clade. A further group identified among the Lineage I rats is the mid blue clade, which occurs in India, but then exclusively on Madagascar and Mafia, and along the East African coast. From there, specimens of this group spread further inland in westward direction to present-day Uganda, Malawi, Zimbabwe, Zambia, and the Republic of Congo. Due to a lack of this haplogroup along the land route through the Arabian peninsula, a direct seaborne dispersal from India to the East African coast is assumed.

The rise of urban settlements along the East African coast in the course of the 1st millennium AD, has led to intensive trade relations with different partners across the Indian Ocean. Thereby, archaeological evidence revealed that certain East African port cities favoured relations with merchants from distinct areas (Horton & Middleton 2000). This may explain the pattern of geographically unique groups among the black rat Lineage I along the East Africa coast. The mid blue clade shows direct seaborne contact between India and the East African mainland. From there, further specimens reached Madagascar, as Malagasy haplotypes group together with

those from Zambia, Zimbabwe, and Malawi. Mafia shares haplotypes with Madagascar and Malawi.

The movement of the black rat in the African hinterlands may be associated with the extended Swahili trade system from the 10th century AD onwards. During that time, the frequent trade activities have led to great wealth across the region and the rise of urban centres. From there, exotic goods from India, the Arabian peninsula or the Gulf have been distributed in westward direction (Horton & Middleton 2000, LaViolette 2008).

Grey/black-rimmed clade. Furthermore, a group of black rats restricted to Egypt and the island Socotra, located off the coast of Yemen, has been identified. The first secured evidence of black rats in Africa stem from Egypt and date to the later half of the 1st millennium BC. Remains of the rodent have been found in stomachs of mummified large birds and cats at Quseir el-Qadim, an important port city at the Red Sea coast that has been involved in the Indian spice trade (Armitage *et al*. 1984). Thereby, Socotra served as an stopping point for merchants on their way along the East African coast (Casson 1989). Here, further sampling across the region may help to find the source population, from where Egyptian and Socotran black rats have been introduced.

Turquoise/white-rimmed clade. A last group among the Lineage I rats is a clade that occurs on Sri Lanka and on the Andaman islands, located in the Bay of Bengal between India and Myanmar. Due to the low genetic diversity (haplotype R8 within the network), a single introduction to the Andaman islands is assumed. Archaeological evidence has been demonstrated that contacts between India and South-East-Asia have been established since the 1st millennium BC. Not only Indian stone beads and black polished ceramics have been found in sites in Thailand, but also Indian crops such as mungbean (*Vigna radiata*) and horsegram (*Macrotyloma uniflorum*) appear in the South-East Asian record at least from 400 BC (Bellina-Pryce & Silapanth 2008, Castillo & Fuller 2010). Moreover, a variety of plants have spread in the opposite direction. For instance, the *Areca*-nut palm (*Areca catechu*), as well as citron (*Citrus medica*), and mango (*Mangifera indica*) have their origin in

South-East Asia, but have been moved to the Indian peninsula by humans around 1,400 to 1,300 BC (Asouti & Fuller 2008). That also small commensal animals travelled through these connections is shown by the additional presence of a Lineage IV black rat on the Andaman islands (figure 3.6). The distribution range of *R. rattus* IV encompasses mainland South-East Asia and Indonesia further south (Aplin *et al*. 2011), from where its range has been extended in westward direction. This is observed similarly for Lineage II, where specimens of this group have been found outside of South-East-Asia on Diego Garcia and the Maldives, where they overlap with Lineage I black rats (figure 3.6).

Analyses have shown that the phylogeography of *R. rattus* is extremely complex. Therefore, this study has sought to unravel the evolutionary history and geographic pattern of the black rat within the Indian Ocean and to reconstruct the direction of dispersals, that were mainly facilitated due to the growing long-distance trading networks over the last millennia. Due to their close association with humans, rats have become one of the most widespread animals, currently occupying almost all continents around the globe. Furthermore, this close relationship leads to a shared history of commensal populations and humans, who transported them. Given that people have moved them around for a long time, an undirected and indistinct distribution of populations was expected. However, the results demonstrate that several unique and geographically restricted lineages have been identified, which reflect past human-aided dispersal throughout the Indian Ocean. As archaeological remains are scarce and morphological distinction is challenging, this study highlights the benefits that can be obtained from molecular genetic analyses of museum material to infer human history. The incorporation of ancient DNA adds a temporal aspect to the genealogy of *R. rattus*. How far the population history is traced back, depends on the age of the samples. Thereby, the analysis of museum samples is a suitable alternative, when archaeological remains are not available or, in case of micro-mammals, are difficult to analyse. However, it is possible that ancient signals are obscured by recent introductions of populations.

Due to their small size, bones of the black rat are challenging to retrieve during the excavation process. Further, DNA analysis of archaeological remains of small mammals presents additional difficulties. Firstly, fragments may have been carried to other strata through burrowing and therefore may not reflect the correct age. Secondly, DNA recovery is challenging due to post-mortem degradation. Certain measures have to be taken into account, in order to assure the authenticity of aDNA sequences (see chapter 1.4.3). As already discussed in the previous chapter, incorporating radiocarbon dating is a valuable approach to determine the accurate age of an archaeological sample. Although the scarcity in finds limits the realisation of both, DNA and radiocarbon analyses.

In order to get a more precise picture of the transoceanic connections across the Indian Ocean, further archaeological sites have to be studied in more detail to overcome the current lack of sample numbers and subsequent inadequate coverage of archaeological and historic data across the area.

3.5 References

Literature

Aplin K, Chesser T, ten Have J (2003) Evolutionary biology of the genus *Rattus*: profile of an archetypal rodent pest. In: Singleton GR, Hinds LA, Krebs CJ, Spratt DM (eds.) Rats, Mice and people: rodent biology and management. Australian Centre for International Agricultural Research, Canberra, 487-498.

Aplin K, Suzuki H, Chinen AA *et al*. (2011) Multiple Geographic Origins of Commensalism and Complex Dispersal History of Black Rats. *PloS One* **6**, 1-20.

Armitage PL, West B, Steedman K (1984) New evidence of black rat in Roman London. *The London Archaeologist* **4**, 375-383.

Armitage PL (1994) Unwelcome companions: ancient rats reviewed. *Antiquity* **68**, 231-240.

Asouti E, Fuller DQ (2008) Trees and woodlands of South India. Archaeological perspectives. Walnut Creek (CA), Left Coast Press.

Avise JC (2000) Phylogeography: the history and formation of species. Harvard University Press, Cambride, MA, 398-410.

Baverstock PR, Adams M, Maxson LR, Yosida TH (1983) Genetic differentiation among karyotypic forms of the black rat, *Rattus rattus*. *Genetics* **105**, 969-983.

Beaujard P (2005) The Indian Ocean in Eurasian and African World-Systems before the Sixteenth Century. *Journal of World History* **16 (4)**, 411-465.

Bellina-Pryce B, Silapanth P (2008) Weaving cultural identities on trans-Asiatic networks: Upper Thai-Malay Peninsula – an early socio-political landscape. *Bullentin de l'Ecole Francaise d'Extreme-Orient* **93**, 257-293.

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: Petraglia MD, Rose JI (eds.) The Evolution of Human Populations in Arabia – Paleoenvironments, Prehistory and Genetics. Springer Science and Buisness Media B.V., 251-278.

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

Boivin N, Crowther A, Prendergast M, Fuller DQ (2014) Indian Ocean Food Globalisation and Africa. *African Archaeological Review* **31 (4)**, 547-581.

Casson L (1989) The Perilplus Maris Erythraei: Text with introduction, translation and commentary. Princeton University Press, Princeton.

Castillo C, Fuller DQ (2010) Still too fragmentary and dependent upon chance? Advances in the study of early Southeast Asian archaeobotany. In: B. Bellina, E. Bacus, O. Pryce, J. Wisseman-Christie (eds.) Studies in honour of Ian Glover. River Books, Bangkok, 91-111.

Cheke A (2010) The timing of arrival of humans and their commensal animals on Western Indian oceanic islands. *Phelsuma* **18**, 38-69.

Cheke A (2011) Was the flying fox *Pteropus comorensis* on Mafia Island (Tanzania) introduced by humans? *Journal of East African Natural History* **100**, 59-68.

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.

Cucchi T, Vigne JD & Auffray JC (2005) The Passive transportation of the house mouse (Mus musculus domesticus) to Cyprus: New indirect evidence of intensive Neolithic navigation in Eastern Mediterranean. *Archaeozoology of the Near East VI. Proceedings of the sixth international symposium on the archaeozoology of southwestern Asia and adjacent areas* 61-77.

Drummond AJ & Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**, 214. BEAST version 1.8.0, available from http://code.google.com/p/beast-mcmc/

Drummond AJ, Ashton B *et al*. (2011) Geneious R6 version 6.0.6, available from http://www.geneious.com

Duplantier JM, Catalan J, Orth A, Grolleau B, Britton-Davidian J (2003) Systematics of the black rat in Madagascar: Consequences for the transmission and distribution of plague. *Biological Journal of the Linnean Society* **78**, 335–341.

Duplantier JM, Duchemin JB, Chanteau S, Carniel E (2005) From the recent lessons of the Malagasy foci towards a global understanding of the factors involved in plague reemergence. *Veterinary Research* **36**, 437-453.

Ervynck A (2002) Sedentism or urbanism? On the origin of the commensal black rat (Rattus rattus). *Bones and the man. Studies in honour of Don Brothwell,* 95-109.

Fitzpatrick SM, Callaghan RT (2008) Seafaring Simulations and the Origin of Prehistoric Settlers to Madagascar. In: Clark G, Leach F, O'Connor S (eds.) Terra Australis 29, Islands of Inquiry: Colonization, seafaring and the archaeology of maritime landscapes. ANU, Canberra, 55-66.

Fleisher J (2003) Viewing stonetowns from the countryside: An archaeological approach to Swahili regional systems, AD 800–1500. Ann Arbor: University Microfilms.

Fuller DQ, Boivin N, Hoogervorst T, Allaby R (2011) Across the Indian Ocean: the prehistoric movement of plants and animals. *Antiquity* **85**, 544-558.

Helm RM (2000) Conflicting histories: The archaeology of the iron-working, farming communities in the central and southern coast region of Kenya. University of Bristol.

Helm RM, Crowther A, Shipton C *et al*. (2012) Exploring agriculture, interaction and trade on the eastern African littoral: Preliminary results from Kenya. *Azania: Archaeological Research in Africa* **47 (1)**, 39-63.

Hingston M, Goodman SM, Ganzhorn JU, Sommer S (2005) Reconstruction of the colonization of southern Madagascar by introduced *Rattus rattus*. *Journal of Biogeography* **32**, 1549-1559.

Horton M, Middleton J (2000) The Swahili: The social landscape of a mercantile society. Blackwell, Cambridge.

Jones HP, Tershy BR, Zavaleta ES *et al*. (2008) Severity of the effects of invasive rats on seabirds: a global review. *Conservation Biology* **22**, 16-26.

Kumar S, Stecher G *et al*. (2012) MEGA-CC: Computing Core of Molecular Evolutionary Genetics Analysis Program for Automated and Iterative Data Analysis. *Bioinformatics* **28**, 2685-2686. Available from http://www.megasoftware.net

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

Love RS (2006) Maritime Exploration in the Age of Discovery – 1415-1800. Greenwood Publishing Group.

Lund M (1994) Commensal rodents. *Rodent Pests and Their Control*, 23-43.

Matisoo-Smith L & Robins JH (2009) Mitochondrial DNA evidence for the spread of Pacific rats through Oceania. *Biological Invasions* **11**, 1521-1527.

Musser GG & Carlton MD (1993) Family Muridae. In: Wilson DM & Reeder DM (eds.) Mammal species of the world, 2nd edition. Smithsonian Institution Press, Washington DC, 501-755.

Niethammer J (1975) Zur Taxonomie und Ausbreitungsgeschichte der Hausratte (*Rattus rattus*). *Zool*. *Anz*. **194**, 405-415.

Plug I (2000) Overview of Iron Age fauna from the Limpopo Valley. *Goodwin Series* **8**, 117–126.

PopART. Available from http://popart.otago.ac.nz

Robins JH, Hingston M, Matisoo-Smith E *et al*. (2007) Identifying *Rattus* species using mitochondrial DNA. *Molecular Ecology Notes* **7**, 717-729.

Searle JB, Jones CS, Gündüz I *et al*. (2009) Of mice and (Viking?) men: phylogeography of British and Irish house mice. *Proceedings of the Royal Society B: Biological Sciences* **276**, 201-207.

Tollenaere C, Brouat C, Duplantier JM *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.

Voigt EA, von den Driesch A (1984) Preliminary report on the faunal assemblage from Ndondondwane. *Annals of the Natal Museum* **26**, 95-104.

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.

Yosida TH (1980) Cytogenetics of the Black Rat. Karyotype evolution and species differentiation. Tokyo University Press, Tokyo, 256.

Electronical sources (2014)

http://www.issg.org/database/species/ecology.asp?si=19&fr=1&sts=sss&lang=EN
http://www.iucnredlist.org/details/19360/0

4 *Hemidactylus frenatus*

Natural translocation or human-mediated introductions? Assessing dispersal routes of the commensal Asian house gecko

4.1 Abstract

The Asian house gecko *H. frenatus* is currently one of the most widespread reptiles in the world due to its fast and successful introduction to almost all subtropical and tropical environments. Because of their worldwide distribution, geckos represent an excellent model to study invasive pathways, which is fundamental in order to develop effective conservation strategies and prevent further extinctions of local species.

Here, a 463 bp fragment of the mitochondrial cytochrome b gene from gecko populations across the world was sequenced to evaluate the phylogeographic structure in *H. frenatus*. Further, different methodical approaches have been applied to unweave the pattern of old natural populations from more recent anthropogenically transported specimens. Results of phylogenetic analyses revealed three distinct clades, whereas Clade A is restricted to the Indian subcontinent. Clade B and C are spreading east- and westward from the core area, strongly indicating a recent anthropogenic introduction on several locations throughout the Indian and Pacific Ocean.

4.2 Introduction

With currently more than 80 species occupying tropical, subtropical and temperate environments around the globe, *Hemidactylus* is one of the species-rich, most diverse and widely distributed genus among reptiles. However, the majority of *Hemidactylus* gecko species is restricted to habitats in South Asia. Only a few species are responsible for most of the vast geographical area covered by the genus – one of

them being the Asian house gecko *Hemidactylus frenatus* (Rocha *et al*. 2005, Carranza & Arnold 2006).

H. frenatus is a medium-sized arboreal gecko native to large parts of Asia and the Indo-Pacific, but rapidly has become one of the most widespread species. At present, the Asian house gecko is established in various locations outside of its natural range (see figure 4.1) in parts of Africa, Australia, the Americas as well as numerous islands throughout the Indian and Pacific Ocean (Cole 2005, Newbery & Jones 2007, Csurhes & Markula 2009). The introduction of *H. frenatus* to new environments has caused significant decline in the abundance of endemic gecko species, leading to a reduction of species diversity through local extinctions (Dame & Petren 2006, Newbery & Jones 2007, Csurhes & Markula 2009).

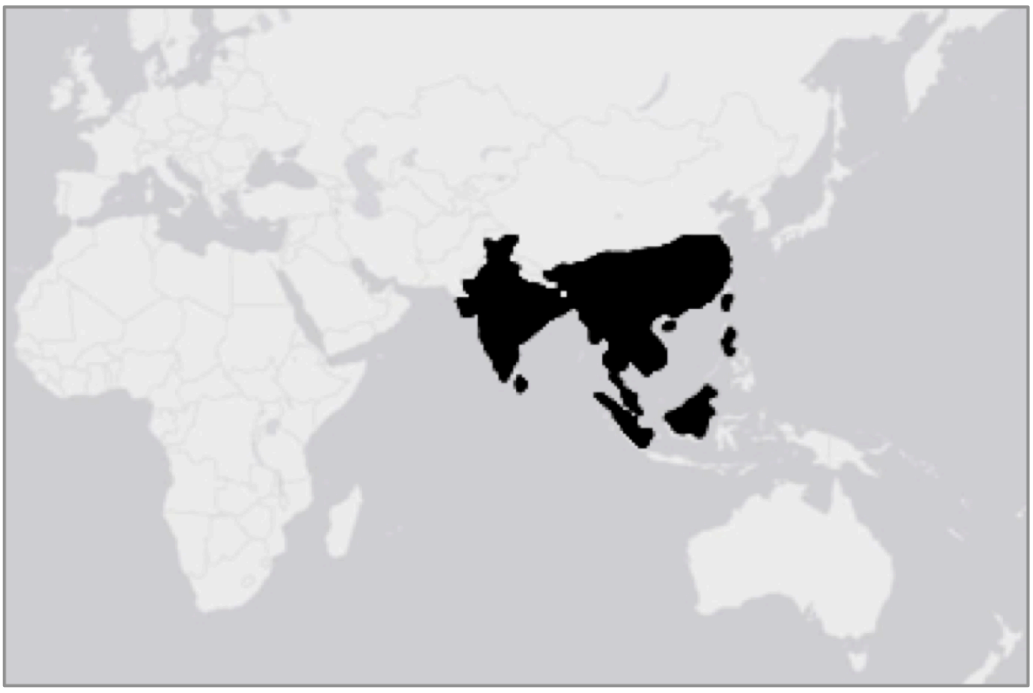

Figure 4.1: Native range of H. frenatus.
(modified, after IUCN distribution map: http://www.iucnredlist.org/details/176130/0)

H. frenatus is a highly synanthropic species, living in close association with humans. Although found in woodlands, forests as well as open fields or rocky habitat, this gecko species is most abundant in urban and suburban locations (Cole 2005, Newbery & Jones 2007, Csurhes & Markula 2009, Hoskin 2011). While there is diurnal activity to some degree, the Asian house gecko is predominantly nocturnal.

Its activity range seems to be mainly influenced by light, as well as temperature and human activity (Marcellini 1971). During the day, the animal hides under rocks, in crevices or other retreat sites and only emerges at dusk to forage all night (Csurhes & Markula 2009). Thereby, the generalist and active predator prefers open hunting surfaces, like building walls or vertical rocks. The Asian house gecko is perfectly adapted to forage on concentrations of insects that gather around artificial lights (Petren & Case 1998, Csurhes & Markula 2009). Due to the structural simplicity of flat walls – in contrast to natural environment – *H. frenatus* recognises intruders earlier, which enables the exclusion of smaller gecko species from these food resources (Case *et al*. 1994). Besides insects and spiders, *H. frenatus* consumes juveniles of other gecko species and even their own progeny (Case *et al*. 1994, Dame & Petren 2006).

A further morphological advantage of the Asian house gecko is their form of locomotion. In contrast to other gecko species with slender clawed toes, *H. frenatus* grips through adhesive toe pads on sub-digital lamellae. These enable increased climbing activity on various surfaces, especially very smooth vertical ones (Zani 2000, Pianka & Sweet 2005). The superior clinging ability affords deposition of eggs in high crevices and refuge from predators. In addition, the eggs of the Asian house gecko are adhesive and hard-shelled, unlike most reptile eggs, making them resistant to moisture loss and therefore they most likely survive long distance travels (Wilson 2006). These features in turn increase the potential to pass undetected in transit on shipping containers and cargo. Furthermore, female *H. frenatus* have the capability for functional sperm retention of up to a year. This facilitates the successful establishment to new locations with a reduced necessity of mating (Murphy-Walker & Haley 1996, Yamamoto & Ota 2006, Newbery & Jones 2007).

Once introduced, *H. frenatus* behaves territorial and aggressive towards endemic gecko species and manifests its competitive superiority by excluding them from food sources and retreat sites (Cole *et al*. 2005). The exclusion from refugia makes native geckos more vulnerable to predation and exposed to adverse climatic conditions (Schoener *et al*. 2001, Cole *et al*. 2005). Furthermore, attacks of *H. frenatus* often lead to the loss of toes or tails, which reduces movement and gripping ability as well as decreases growth, fecundity, and range size. The increased

mortality due to direct aggressive interaction with the Asian house gecko and exclusion from retreat sites have a negative impact on the abundance and ultimately on the survival of endemic gecko species (Martin & Avery 1998, Cole *et al*. 2005).

With the Age of Discovery, the frequent dispersal of organisms began and thereby the crossing of biogeographic barriers that have been isolated for millions of years. Especially over the past decades, international shipping and cargo movement via plane has constantly increased and the growing network of human travel and trade has led to a vast spread of alien species (Case *et al*. 1992, Mooney & Cleland 2001, Newbery & Jones 2007, Csurhes & Markula 2009). The number of reptiles being introduced around the world has caused a significant competitive impact on resident species, driving them to extinction. For example, Australia is currently facing a massive range expansion of the Asian house gecko, detecting high densities of this gecko species not only in urban but also in natural habitats. Although *H. frenatus* was introduced about 50 years ago, it only recently started spreading relentlessly becoming a potential threat to Australia's ecosystem (Hoskin 2011). Therefore, understanding of the ecology of commensal species as well as their invasive pathways is fundamental to develop and apply successful conservation strategies (Cole *et al*. 2005).

Because of the worldwide distribution and rapid radiation due to its close association with humans, the Asian house gecko represents an excellent model to study the underlying pattern of species translocation. Especially in recent times, *H. frenatus* has experienced multiple processes of transmarine relocation (Vences *et al*. 2004, Cole *et al*. 2005, Dame & Petren 2006, Newbery & Jones 2007, Cole 2009, Hoskin 2010) – deliberately or accidentally. In consequence of the fast and permanent movement with humans to almost all subtropical and tropical environments around the globe, I expect the recent worldwide gecko diversity to be equally distributed with no obvious phylogeographic patterns shown.

Here, in order to test that expectation I extracted and sequenced a 463 bp fragment of the mitochondrial cytochrome b gene from gecko populations across the world. Subsequently, phylogenetic trees were constructed to evaluate the strength of the phylogeographic signal in *H. frenatus*. In order to assess whether the distribution

of the Asian house gecko is shaped by natural translocation or if human-mediated introductions play a decisive role, continuous phylogeographic diffusion modelling was applied.

4.3 Material and Methods

4.3.1 Samples

Altogether, 97 tissue samples of *Hemidactylus frenatus* were analysed, whereas 51 specimens were obtained from diverse museums (Field Museum, Chicago; British National History Museum, London; Smithsonian, Washington DC), additional 46 modern specimens were collected in the field from Dr. J. Chris Hillman and Solomon Pomerantz (Sealinks Project). Museum samples were mainly stored in ethanol or formalin, the modern tissue samples were dried.

The sampling area comprises different places of the mainland and islands around the Indian Ocean, including countries from the East and West African Coast, the Indian subcontinent and South-East Asia. Furthermore, samples stem from various Pacific Islands (see figure 4.2). A list of all samples is found in chapter 7.2.3, including additional information.

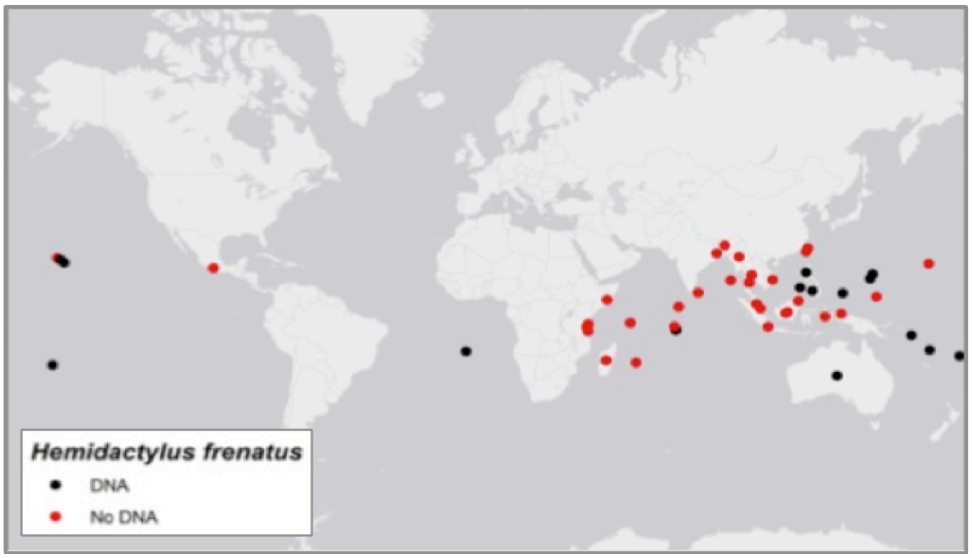

Figure 4.2: Sample distribution of H. frenatus.

(Figure 4.2 continued)
Distribution of sampling locations of H. frenatus specimens used in this study. Successful DNA extraction and amplification is indicated in black, unsuccessful DNA extraction is indicated in red.

4.3.2 Analyses

All samples were analysed in the facilities at the Archaeology Department of Durham University. Because of the recent age of most of the specimens, the samples were treated as modern material. Each workstep – from DNA extraction to sequencing set-up – was conducted in the modern laboratory, following the protocols described in chapters 7.1.2 (extraction protocol 1 to 3), 7.1.3 and 7.1.4. The sequencing reaction was carried out by the DNA Sequencing Service at the School of Biological and Biomedical Sciences at Durham University.

A 463 bp region of the mitochondrial cytochrome b was amplified and sequenced in one fragment (see chapter 7.1.1). The sequencing chromatograms were edited manually, subsequently assembled, and a consensus sequence per individual exported (see chapter 7.1.4) using Geneious R6 version 6.0.6 (Drummond *et al*. 2011). Standard anti-contamination guidelines were followed to ensure authenticity of the DNA results (see chapters 1.4.3 and 7.1.5). Phylogenetic analyses were performed by construction of Bayesian trees using MrBayes version 3.2.2 (Huelsenbeck & Ronquist 2001). The best-fit nucleotide substitution model, selected in MEGA-CC (Kumar *et al*. 2012) under the Bayesian Information Criterion (BIC), was Kimura-2-parameter +Γ. Under this model, a consensus tree with bootstrap values shown on the nodes was created for 5,000,000 generations, a sample frequency of 5,000 and with a burn-in period of 100 trees (see figure 4.3).

Furthermore, mitochondrial phylogeography was inferred using BEAST version 1.8.0 (Drummond & Rambaut 2007). To calibrate the tree a mutation rate of 0.01 (= 1%) substitutions per site per million years was assumed based on the paper by Rocha *et al*. (2010). In order to apply the correct clock model, a relaxed clock model was used initially to test for variation in the evolutionary rate. Statistics showed that the hypothesis of a constant mutation rate could not be rejected, consequently the

mutation rate is constant. Therefore, the parameter has been changed to a strict clock model in the final run. Best-fit nucleotide substitution model was Kimura-2-parameter +Γ. Additionally, a coalescent tree prior with exponential population growth was specified, based on the expectation that the population size of *H. frenatus* has not remained constant during their evolutionary history.

For the phylogeographic diffusion analysis a lognormal relaxed random walk (RRW) model was applied. The model uses the geographic sample locations of molecular sequences to reconstruct the spatiotemporal diffusion of an organism. Thereby, it allows the rate of diffusion to vary along individual branches of the topology (Lemey *et al*. 2010). The coordinates of the sample locations were recovered from Google Maps. In cases were duplicate localities confound the RRW model, a random jitter was applied, where new localities were drawn randomly within a geographic window of 0.02 x 0.02 decimal degrees (= 2 km x 2 km) from the original locality (Barlow *et al*. 2013). Two independent runs of 50,000,000 generations each were performed with a sample frequency of 5,000 and a 10% burn-in period (see figure 4.4). Afterwards, the generated BEAST tree was uploaded in SPREAD version 1.0.6 (Bielejec *et al*. 2011). SPREAD allows to visualise full posterior distribution of the trees obtained in phylogeographic analyses and plots the results onto a map. The generated output file was uploaded in Google Earth in order to visualise the simulated spreading in space and time (see figure 4.6).

Additionally, the generation of a median-spanning network of *H. frenatus* specimens with PopART (http://popart.otago.ac.nz), based on the number of differences between sequences, revealed 12 different haplotypes among a total sample set of 34 sequences (see figures 4.7 and 4.8). Due to the inability of networks to deal with missing data, the original *H. frenatus* alignment had to be truncated to a 288 bp fragment of the mitochondrial cytochrome b.

4.3.3 Results

DNA sequences of a 463 bp fragment of the cytochrome b gene were obtained from 37 of the 97 modern and museum samples. Additionally, 12 sequence entries of *H. frenatus* from GenBank were combined to a dataset composed of 49 individuals. The poor amplification success of only 37 samples is caused by the storage conditions of the museum specimens, which were either fixed in ethanol or formalin. Recovery of DNA from formalin fixed specimens is challenging. Although formalin does not physically degrade nucleic acids, it leads to the generation of DNA-protein crosslinkages. Furthermore, the nucleic acids will fragment if the fixative solution is unbuffered, due to an extreme low pH value of < 1. Crosslinking and fragmentation of the DNA complicates both the isolation of DNA and the amplification during the PCR (Gilbert *et al*. 2008).

Bayesian analyses generated a well-resolved tree (see figure 4.3). Three main clades are apparent from the data. Thereby, samples of Clade A (purple) are restricted to India and Sri Lanka, whereas Clade B (yellow) is distributed from Colombia and Hawaii throughout the Pacific to Diego Garcia Island in the West. A third Clade C (green) occurs from St. Helena Island in the West through the Indian Ocean to the Philippines and Vanuatu in the East (see figure 4.5). Results are strongly supported by bootstrap values between 96 and 100.

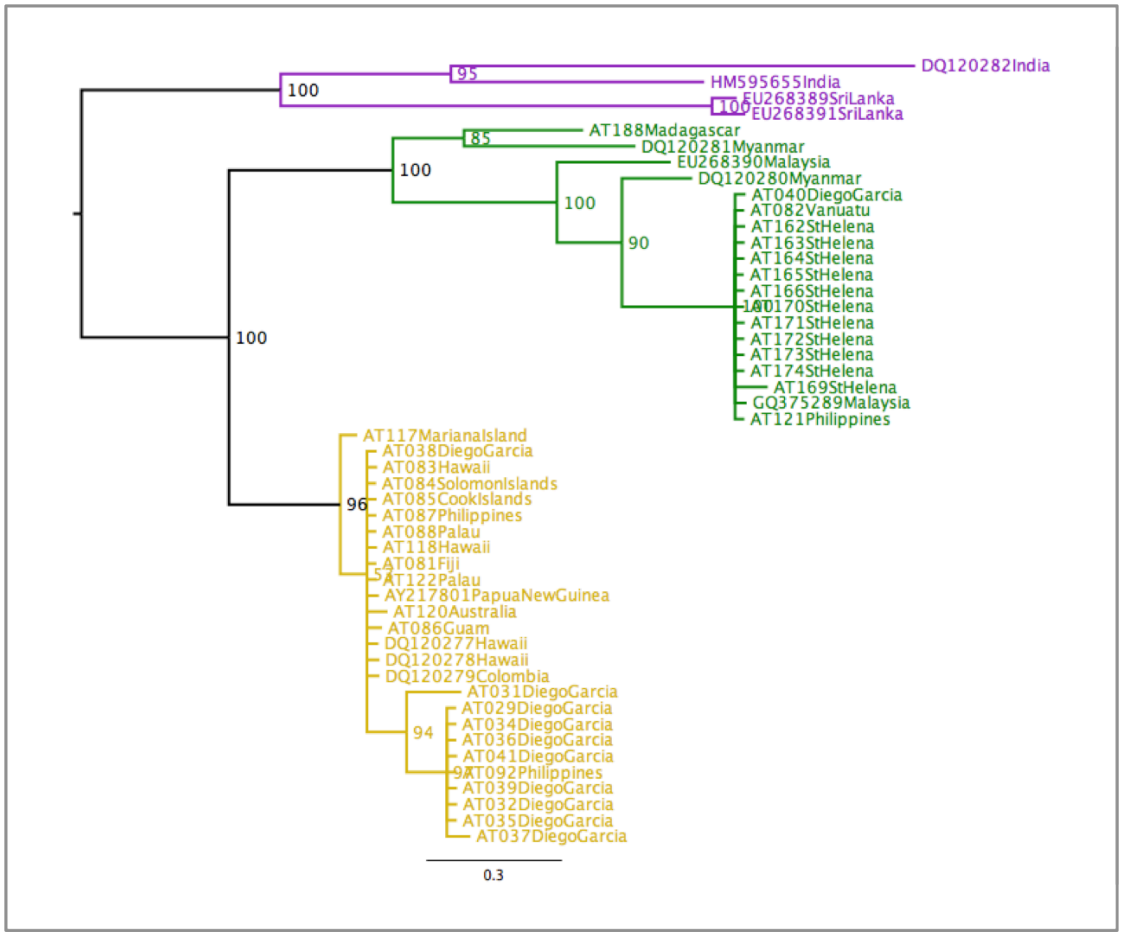

Figure 4.3: Mitochondrial phylogeny of H. frenatus.
Inferred from a 463 bp fragment of the mitochondrial cyt b. Analyses were performed using MrBayes version 3.2.2. The best-fit nucleotide substitution model under the Bayesian Information Criterion (BIC) was Kimura-2-parameter +Γ. A consensus tree with bootstrap values shown on the nodes was created for 5,000,000 generations with a sample frequency of 5,000 and a burn-in period of 100 trees. Clade A is indicated in purple, Clade B is indicated in yellow, Clade C is indicated in green.

Molecular-clock-rooted Bayesian analyses produced an identical tree topology (see figure 4.4). The calibrations applied to the mitochondrial tree indicate that divergence of the three clades occurred between 6.5 to 1 million years ago. Statistical results examined in Tracer version 1.6 (Rambaut *et al*. 2014) indicated that the parameter of the tree model (exponential growth rate) did not include zero, therefore rejecting a constant rate of diffusion and justifying the use of the RRW model. Furthermore, the parameter of the relaxed clock model (coefficient of variation) included zero, indicating no variation in the evolutionary rate. Therefore, the clock model was changed to a strict clock model, implying a constant mutation rate.

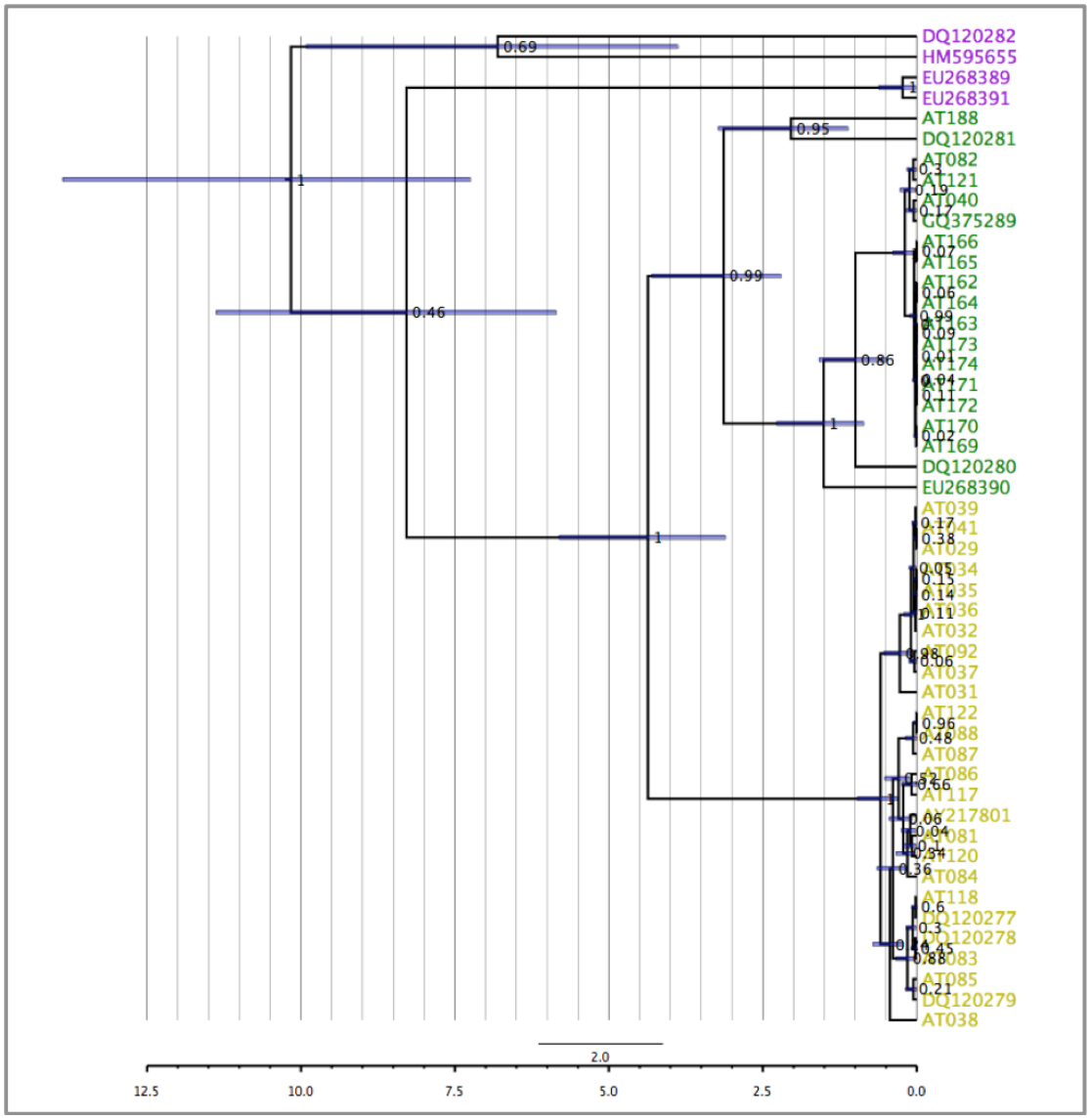

Figure 4.4: Molecular-clock based mitochondrial phylogeny of H. frenatus.
Inferred from a 463 bp fragment of the mitochondrial cyt b using BEAST version 1.8.0. Clade support is given by Bayesian posterior probabilities. Node bars show the 95% HPDs. To calibrate the tree a mutation rate of 0.01 (= 1%) substitutions per site per million years was assumed. Statistics could not reject a constant mutation rate, therefore a strict clock model was applied. Best-fit nucleotide substitution model was Kimura-2-parameter +Γ. Additionally, a coalescent tree prior with exponential population growth was specified. For the phylogeographic diffusion analysis a lognormal relaxed random walk (RRW) model was applied. The coordinates of the sample locations were recovered from Google Maps. In cases were duplicate localities confound the RRW model, a random jitter was applied, where new localities were drawn randomly within a geographic window of 0.02 x 0.02 decimal degrees (= 2 km x 2 km) from the original locality. Two independent runs of 50,000,000 generations each were performed with a sample frequency of 5,000 and a burn-in period of 1,000 trees. Clade A is indicated in purple, Clade B is indicated in yellow, Clade C is indicated in green.

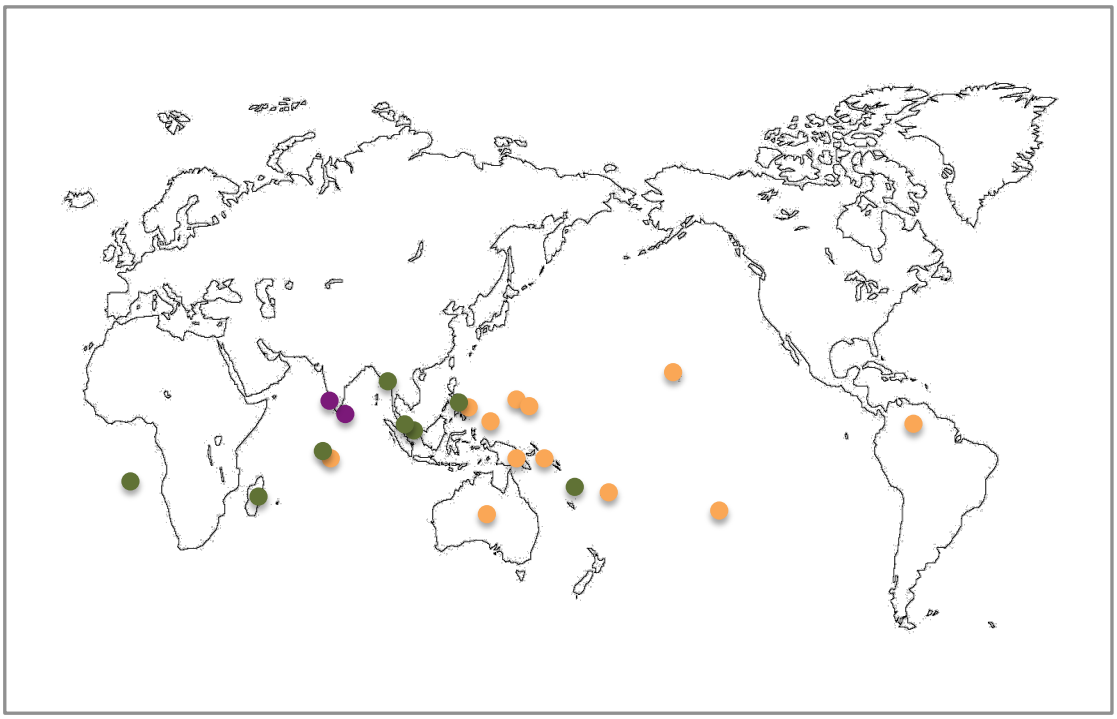

Figure 4.5: Haplogroup distribution of H. frenatus.
Clade A is indicated in purple, Clade B is indicated in yellow, Clade C is indicated in green.

Species distribution modelling (see figure 4.6) revealed that *H. frenatus* populations stay in their native range in India and South-East Asia for most of the time, before spreading out throughout the globe. The red branch thereby indicates the starting point of the coalescence tree, which is placed in the expected native range of the Asian house gecko. Furthermore, the Asian house gecko still remains in South-East Asia for a long time, represented by the bifurcations of all internal nodes within the tree as shown by their light blue confidence interval polygons. Only in recent times a rapid expansion from the native range in eastern and western directions occurred, leading to the current worldwide distribution of the Asian house gecko.

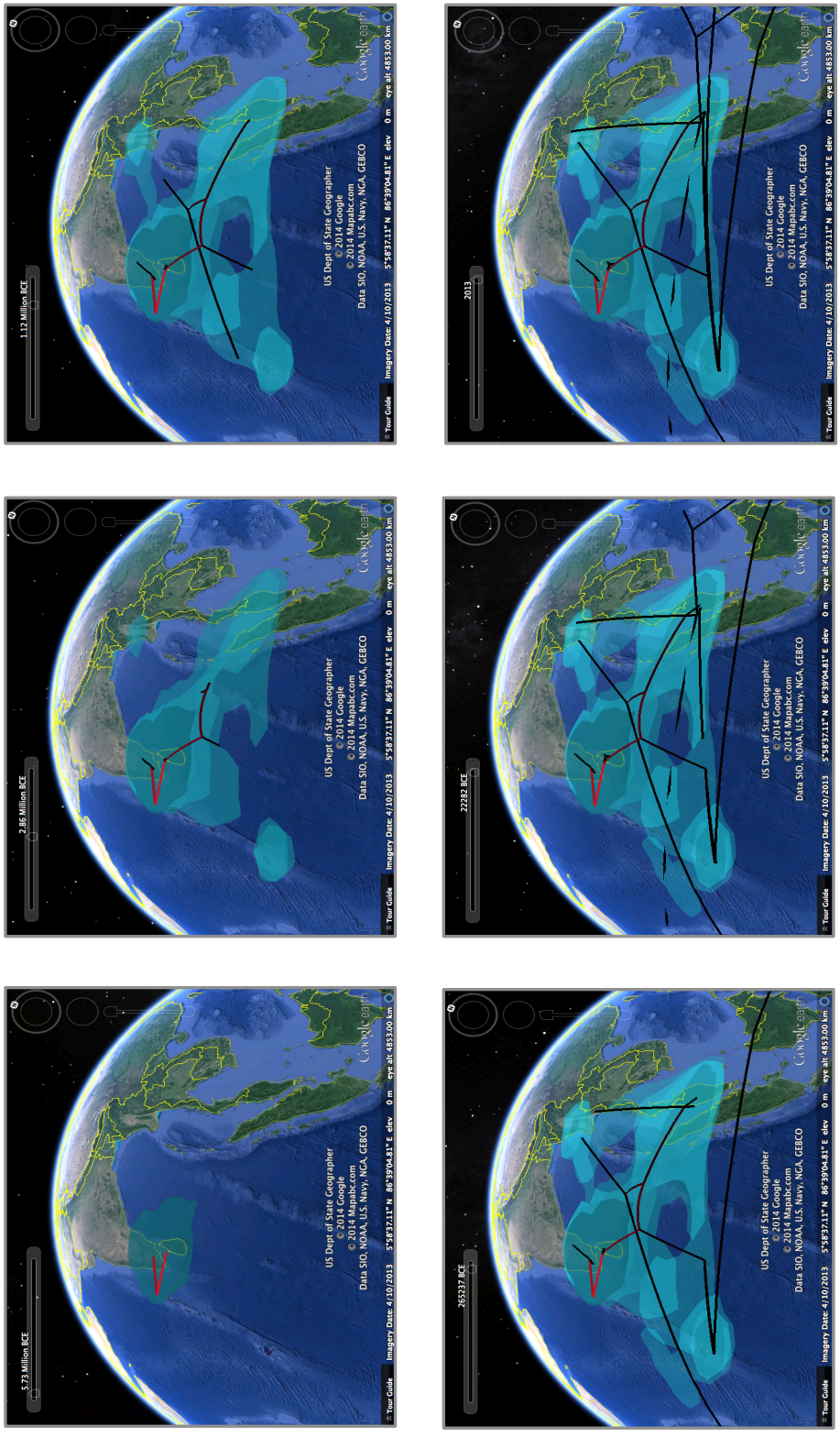

Figure 4.6: Dispersal simulation of H. frenatus in space and time.
Visualisation of the full posterior distribution of the trees obtained in the phylogeographic analysis in BEAST, using SPREAD version 1.0.6. Subsequently, the generated output file was uploaded in Google Earth in order to visualise the simulated spreading of H. frenatus in space and time.

Results of the network analysis revealed 12 haplotypes in a total of 34 sequences (see figures 4.7 and 4.8). Thereby, haplotypes G1, G2, and G3 show highest frequencies, as well as a broad geographical distribution among the collected samples. Specimens of the G1 haplotype only appear throughout the Pacific from Australia and Papua-New Guinea over several Pacific islands to Colombia. G2 is mainly found on Diego Garcia, located off the Indian peninsula, but also on the Philippines. G3 appears with a high frequency on St. Helena, located off the African West coast, but also on the Pacific island of Vanuatu. The most diverse location is Diego Garcia, where 4 different haplotypes occur, that are also found on the Philippines. However, one has to bear in mind that the number of samples per location varied and therefore different haplotypes may have been missed in certain areas.

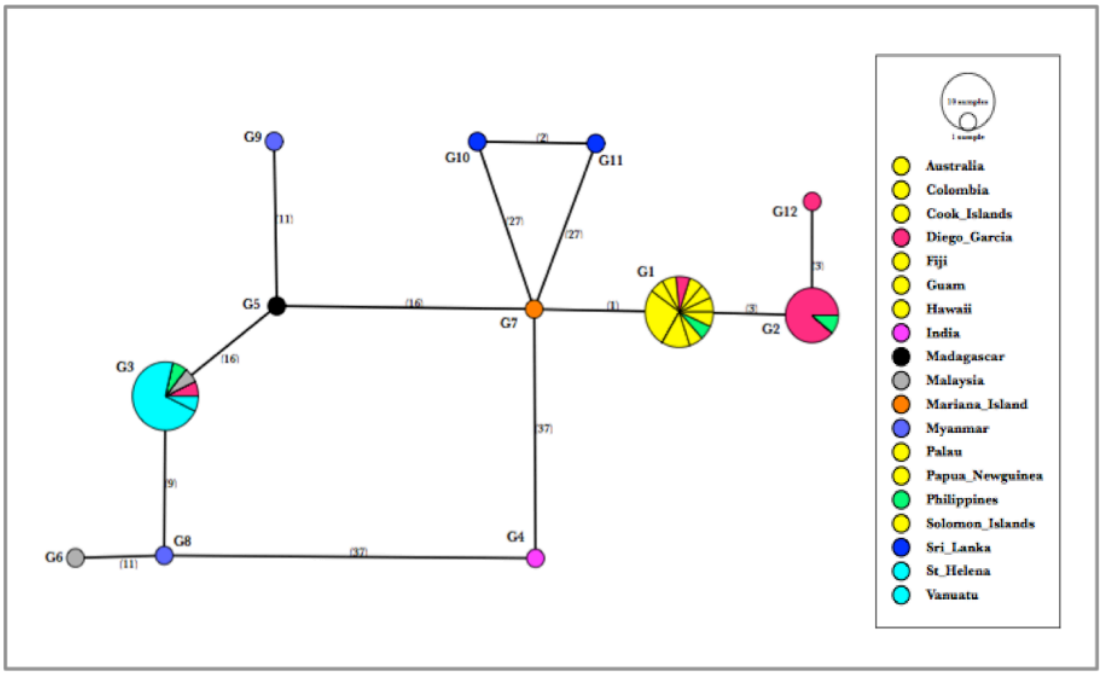

Figure 4.7: Network analysis of H. frenatus using PopART.
Haplotype diversity of H. frenatus throughout the region, assessed from a 288 bp fragment of the mitochondrial cytochrome b. 34 sequences have been included and revealed 12 haplotypes.

Country	Haplotype
Australia	G1
Colombia	G1
Cook Islands	G1
Diego Garcia	G1, G2, G3, G12
Fiji	G1
Guam	G1
Hawaii	G1
India	G4
Madagascar	G5
Malaysia	G3, G6
Mariana Island	G7
Myanmar	G8, G9
Palau	G1
Papua-New Guinea	G1
Philippines	G1, G2, G3
Solomon Islands	G1
Sri Lanka	G10, G11
St. Helena	G3
Vanuatu	G3

Figure 4.8: Table of the haplotype diversity of H. frenatus.

4.4 Discussion

The Asian house gecko *H. frenatus* is currently one of the most widespread reptiles in the world due to its fast and successful introduction to almost all subtropical and tropical environments. Because of their worldwide distribution, geckos represent an excellent model to study invasive pathways, which is fundamental in order to develop effective conservation strategies to prevent further extinctions of local species. Basically, there are two possible ways how *H. frenatus* has reached the faraway lands and small scattered islands in the Indian and Pacific Ocean – either by natural trans-oceanic colonisation or human-mediated translocation. Here, different approaches have been applied to unweave the pattern of a natural population from anthropogenically transported specimens.

Compared to other commensal animals, such as the black rat or the house mouse, archaeological evidence of the spreading of the Asian house gecko in early times is insufficient. Although *H. frenatus* is highly synanthropic and may have been initially transported to places outside their natural range by Polynesian and Melanesian seafarers approximately 4,000 years ago (Case *et al*. 1994), their history of commensalism is still uncertain. In fact, this species has only recently started to extensively invade new habitats especially over the last decades (Hoskin 2011), which is associated with the continuous urbanisation throughout the region, as well as the frequent international shipping and cargo movement at present. This development facilitated the recent dispersal of *H. frenatus* due to some distinct biological features, which appear to be advantageous especially in modern urban environments.

As mentioned earlier, their adhesive toe pads enable superior clinging ability especially on smooth, vertical surfaces such as building walls or cargo containers (Zani 2000, Pianka & Sweet 2005). Further, the Asian house gecko is perfectly adapted to hunt insects that gather near electric lights (Petren & Case 1998, Csurhes & Markula 2009), while excluding other gecko species from these sources of food. The environment shaped by humans, provides a perfect niche for *H. frenatus*, why its

mainly found in urban and suburban structures and less in natural habitats (Cole 2005, Newbery & Jones 2007, Csurhes & Markula 2009, Hoskin 2011). Besides the close association with humans and exploitation of their environment, there are several factors facilitating the current dispersal success of the Asian house gecko across the Indian and Pacific Ocean: due to their small body size and their habit of hiding in dark places during the day, detection of geckos is difficult. Thereby, they easily get transported unintentionally in cargo – mainly via sea freight as introduced specimens appear mainly in harbours first – but also in road vehicles or planes (e.g. on Diego Garcia as described later in this chapter) (Csurhes & Markula 2009). Furthermore, their advanced climbing ability enables deposition of eggs in high crevices, e.g. in the upper corners of cargo containers, where they are difficult to spot and protected from predation. Additionally, the eggs are adhesive and hard-shelled making them resistant to moisture loss and therefore they most likely survive long distance travels (Wilson 2006). Due to being a generalist and having broad physical tolerances, the Asian house gecko has been transported over long distances and successfully established new populations along the mainland and on islands across the Indian and Pacific Ocean.

Because of the rapid spreading of the Asian house gecko in recent times, the worldwide gecko diversity was expected to be equally distributed with no obvious phylogeographic pattern shown. However, results of phylogenetic analyses revealed three distinct clades, whereas Clade A is restricted to the Indian subcontinent, and Clade B and C are distributed east- and westward from the core area (see figure 4.5). In a further molecular-clock based analysis, I tested whether this phylogeographic structure is more likely explainable by natural or human-aided dispersal.

After incorporating a temporal dimension to the dataset (see figure 4.4), results demonstrate that *H. frenatus* populations remain in their native range for a long period of time (6 million to 400,000 years BP) before a change of this pattern occurs in more recent times. A rapid expansion in eastern and western directions is indicated, currently covering locations around the world. According to molecular clock estimations the dispersal occurred within the last 400,000 years. Therefore, I can not discard the hypothesis that humans were involved in the range expansion at

some point. Furthermore, the fact that *H. frenatus* is not moving for a long time and then suddenly extents its range of habitat – including back and forth movement within different islands – is more likely explainable by human-influenced dispersal. In order to have a more exact calibration of the molecular clock, adding more sequence data is necessary. Currently the timing of the dispersal might be imprecise as the mitochondrial sequence of only 463 bp is too short for mutations to occur and therefore unable to take all changes into account. However, although an absolute assessment of dispersal times is difficult here, molecular cock based analyses rather provide a relative classification of significance.

Analyses of the haplotype diversity showed, that greatest diversity is found within the native range of the Asian house gecko in South-East Asia and the Indian subcontinent. From the core area, distinct haplotypes are spreading in eastern (G1) and western directions (G2 and G3) throughout the Indian and Pacific Ocean. The lack of variation within specimens of the geographically separate islands throughout the Pacific Ocean, that mainly carry the G1 haplotype, strongly indicates recent anthropogenic introductions. Thereby, low diversity of this highly mobile species is explainable due to bottlenecks, while the Asian house gecko has been transported from island to island. In case of a natural spreading a more arbitrary phylogeographic picture is expected, as it is undirected and occurs rather infrequent, especially within the scattered islands in the wide-open Ocean.

On the other hand, the picture within the Indian Ocean area is not that clear. Although a particular *H. frenatus* population is moving from the Philippines throughout Malaysia and Diego Garcia Island as far as St. Helena Island on the west coast of Africa, distinct haplotypes likewise occur on Diego Garcia, St. Helena Island, and Madagascar (see figure 4.8). Different explanations are suggested, whereby the isolated haplotypes might represent single natural introductions where geckos on wreckage get carried by oceanic currents to new places. This may be the case for St. Helena, where only a single haplotype G3 occurs at high frequency. For other islands like Madagascar, Vanuatu or Marina Island, more samples have to be collected to assess whether the single haplotypes are the result of the low number of

samples collected in these locations. Further, current haplotype frequencies could be the result of population changes through time on these islands.

In the course of the 1st millennium AD, regional exchange networks between local communities across the Indian Ocean developed into trans-regional trade between rising port cities (Beaujard 2005). However, the permanent settlement of major islands in the Indian Ocean took place mainly between the 15th and 17th century AD when European oversea expeditions led to the establishment of vast trading routes between Asia, Africa and Europe (Cheke 2010). As a consequence the advanced maritime technologies and infrastructural improvements have led to the transfer of not only goods, but also livestock and commensal animals. This resulted in the dispersal of organisms and thereby the crossing of biogeographic barriers (Boivin *et al*. 2014). However, evidence of movements of the Asian house gecko *H. frenatus* during prehistoric times is scarce. It has been first recorded on Mauritius around 1770 AD and on Reunion around 1801 AD – at times when these islands were permanently settled by humans (Cheke 2010).

Reasons for the lack of early evidence may be the difficult recovery of gecko remains in archaeological excavations. Further, *H. frenatus* may have been dependent on humans, that it only started colonising far away islands, when these were frequently used as stopping points and subsequently settled. Moreover, it is possible that contemporary phylogeographic pattern might not accurately reflect past structures. Especially island environments are prone to changes within the diversity of their flora and fauna, e.g. through extinctions or population turnovers. Through the increased international shipping and cargo movement in modern times, the introduction of alien species to new habitats is becoming more prevalent having discernable effects on the endemics. An example of the impact of human movement is shown by the similar pattern of haplotype diversity shared by Diego Garcia Island and the Philippines. This particular structure can be explained by a constant exchange between the two places as a result of recent military activity. Since the occupation of Diego Garcia by US Americans, flights come in to the island from Guam, Singapore and the Philippines on a regular basis and subsequently introducing – deliberately or accidently – highly invasive species (Cheke 2008, pers. comm.

Anthony Cheke). In order to infer the history of a population and their complex processes over time, the incorporation of ancient DNA is necessary. However, in this case no archaeological remains of *H. frenatus* have been available.

As shown in this study, it is often difficult to distinguish natural populations from human-mediated introductions, due to the known ease with which the highly commensal Asian house gecko *H. frenatus* is transported. Nevertheless, geckos represent an excellent model to study invasive pathways and are a valuable proxy to unravel human travel routes, although preferably in modern times. Due to their different ecology, compared to other study animals within this thesis, the Asian house gecko started massively extending its range only within the last decades and therefore does not represent an ideal proxy to trace back ancient human trade routes. Furthermore, the lack of archaeological samples impedes insights into population processes of *H. frenatus* over the past millennia. In order to study beginning connections between the different cultures across the Indian and Pacific Ocean, the black rat and the house mouse represent more valuable proxies.

However, contrary to the expected undirected spreading of *H. frenatus* populations in recent times, three distinct clades with a specific distribution pattern have been identified. In order to gain further knowledge about the movement of the Asian house gecko future work should focus on a more exact calibration of the molecular clock. By adding more sequence data, the correct timeframe of the spreading is assessed. Here, results are based solely on a single gene tree and extension of the sequence data will allow a more precise timing of potential invasive pathways. These ways of migration could then be assigned to human seafaring routes as well as dates of introductions to several islands to distinguish natural populations from human-mediated translocations. Thereby, sequence variation should increase with geographic distance to the native range. Furthermore, additional samples need to be collected within the potential native range to evaluate the haplotype variation of the core area. Moreover, further sampling in several locations throughout the Indian and Pacific Ocean is necessary to verify the current phylogeographic structure of the defined Clades A, B, and C.

In summary this study has sought to examine the phylogeographic structure of one of the most widespread reptiles, the Asian house gecko *H. frenatus*. Genetic analyses revealed three distinct clades, which have been introduced to islands in the Indian and Pacific Ocean by natural colonisation as well as human-mediated translocation mainly in recent times. The spread of commensal animals has a severe impact on native species, especially on islands. Therefore, the study of the ecology and invasive pathways of *H. frenatus* is fundamental to develop effective conservation strategies and prevent further extinctions of local species.

4.5 References

Literature

Barlow A, Baker K, Henry CR *et al*. (2013) Phylogeography of the widespread African puff adder (*Bitis arietans*) reveals multiple Pleistocene refugia in southern Africa. *Molecular Ecology* **22**, 1134-1157.

Beaujard P (2005) The Indian Ocean in Eurasian and African World-Systems before the Sixteenth Century. *Journal of World History* **16 (4)**, 411-465.

Bielejec F, Rambaut A, Suchard MA, Lemey P (2011) SPREAD: Spatial Phylogenetic Reconstruction of Evolutionary Dynamics. *Bioinformatics* **27**, 2910-2912. SPREAD version 1.0.6, available from www.kuleuven.ac.be/aidslab/phylogeography/SPREAD.html

Boivin N, Crowther A, Prendergast M, Fuller DQ (2014) Indian Ocean Food Globalisation and Africa. *African Archaeological Review* **31 (4)**, 547-581.

Carranza S & Arnold EN (2006) Systematics, biogeography, and evolution of *Hemidactylus* geckos (Reptilia: Gekkonidae) elucidated using mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution* **38**, 531-545.

Case TJ, Bolger DT, Richman AD (1992) Reptilian extinctions: the last ten thousand years. In: Fiedler PL, Jain SK (eds.) Conservation Biology: The Theory and Practice of Conservation, Preservation, and Management. Chapman & Hall (New York), 91-125.

Case TJ, Bolger DT, Petren K (1994) Invasions and Competitive Displacement Among House Geckos in the Tropical Pacific. *Ecology* **75**, 464-477.

Cheke A (2008) The risk to other Indian Ocean islands of invasive species on Diego Garcia. *Phelsuma* **16**, 70-72.

Cheke A (2010) The timing of arrival of humans and their commensal animals on Western Indian oceanic islands. *Phelsuma* **18**, 38-69.

Cole NC, Jones CG, Harris S (2005) The need for enemy-free space: The impact of an invasive gecko on island endemics. *Biological Conservation* **125**, 467-474.

Cole NC (2005) The ecological impact of the invasive house gecko *Hemidactylus frenatus* upon endemic Mauritian geckos. Dissertation, Bristol.

Cole NC (2009) Herpetofaunal observations on Eagle Island, Middle Brother, North Brother and Diego Garcia, with an overview of previous records in the Chagos Archipelago. *Phelsuma* **17**, 40-48.

Csurhes S & Markula A (2009) Pest animal risk assessment: Asian house gecko *Hemidactylus frenatus*. Biosecurity Queensland, Queensland Primary Industries and Fisheries, Department of Employment, Economic Development and Innovation (Queensland, Australia).

Dame EA & Petren K (2006) Behavioural mechanisms of invasion and displacement in Pacific island geckos (*Hemidactylus*). *Animal Behaviour* **71**, 1165-1173.

Drummond AJ & Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**, 214. BEAST version 1.8.0, available from http://code.google.com/p/beast-mcmc/

Drummond AJ, Ashton B *et al*. (2011) Geneious R6 version 6.0.6, available from http://www.geneious.com

Gilbert MTP, Haselkorn T, Bunce M *et al*. (2008) The Isolation of Nucleic Acids from Fixed, Paraffin-Embedded Tissues – Which Methods Are Useful When? *PloS One* **6**, 1-12.

Google Earth (Version 5.1) Mountain View, CA: Google Inc. (2009). Available from http://www.google.com/earth/download/

Hoskin CJ (2011) The invasion and potential impact of the Asian House Gecko (*Hemidactylus frenatus*) in Australia. *Austral Ecology* **36**, 240-251.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755. Version 3.2.2, available from http://mrbayes.sourceforge.net

Kumar S, Stecher G *et al*. (2012) MEGA-CC: Computing Core of Molecular Evolutionary Genetics Analysis Program for Automated and Iterative Data Analysis. *Bioinformatics* **28**, 2685-2686. Available from http://www.megasoftware.net

Lemey P, Rambaut A, Welch JJ, Suchard MA (2010) Phylogeography takes a relaxed random walk in continuous space and time. *Molecular Biology and Evolution* **27**, 1877-1885.

Marcellini DL (1971) Activity Patterns of the Gecko *Hemidactylus frenatus*. *Copeia* **4**, 631-635.

Martin J & Avery RA (1998) Effects of tail loss on the movement patterns of the lizard, *Psammodromus algirus*. *Functional Ecology* **12**, 794-802.

Mooney HA & Cleland EE (2001) The evolutionary impact of invasive species. *PNAS* **98** (10), 5446-5451.

Murphy-Walker S & Haley SR (1996) Functional sperm storage duration in female *Hemidactylus frenatus* (Family Gekkonidae). *Herpetologica* **52**, 365-373.

Newbery B & Jones DN (2007) Presence of Asian House Gecko *Hemidactylus frenatus* across an urban gradient in Brisbane: influence of habitat and potential for impact on native gecko species. In: Lunney D *et al*. (eds.) Pest or Guest: the zoology of overabundance. Royal Zoological Society of New South Wales, Mosman (New South Wales, Australia).

Petren K & Case TJ (1998) Habitat structure determines competition intensity and invasion success in gecko lizards. *Proceedings of the National Academy of Sciences USA* **95**, 11739-11744.

Pianka ER & Sweet SS (2005) Integrative biology of sticky feet in geckos. *BioEassays* **27**, 647-652.

Rambaut A, Suchard MA, Xie D, Drummond AJ (2014) Tracer v1.6, available from http://beast.bio.ed.ac.uk/Tracer

Rocha S, Carretero MA, Harris DJ (2005) Diversity and phylogenetic relationships of *Hemidactylus* geckos from the Comoro islands. *Molecular Phylogenetics and Evolution* **35**, 292-299.

Rocha S, Carretero MA, Harris DJ (2010) On the diversity, colonization patterns and status of *Hemidactylus* spp. (Reptilia: Gekkonidae) from the Western Indian Ocean islands. *Herpetological Journal* **20**, 83-89.

Schoener TW, Spiller AW, Losos JB (2001) Predators increase the risk of catastrophic extinction of prey populations. *Nature* **412**, 183-186.

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. MEGA version 5.2.2, available from http://www.megasoftware.net

Vences M, Wanke S, Vieites DR *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.

Wilson S (2006) Asian House Geckos – Queensland Museum Fact Sheet. www.qm.qld.gov.au/inquiry/factsheets/asian_house_geckos_20080709.pdf

Yamamoto Y & Ota H (2006) Long-term Functional Sperm Storage by a Female Common House Gecko, *Hemidactylus frenatus*, from the Ryukyu Archipelago, Japan. *Current Herpetology* **25**, 39-40.

Zani PA (2000) The comparative evolution of lizard claw and toe morphology and clinging performance. *J. Evol. Biol.* **13**, 316-325.

Electronical sources (2014)

http://www.iucnredlist.org/details/176130/0

5 *Bubalus mindorensis*

Resolving the taxonomic position of the Asian dwarf buffalo: The first mitogenome of *Bubalus mindorensis*

5.1 Abstract

The phylogeny, demographic history, and origin of domestication of the Asian buffalos have been subject to a controversial discussion for a long time. In order to resolve the taxonomic status of the tamaraw within the bovid family, I employed next-generation-sequencing techniques to generate the first mitogenomes of *Bubalus mindorensis* from museum samples, collected about 80 years ago on Mindoro Island. Although tamaraws share morphological features with both Asian water buffalos and the anoas, my analyses demonstrate a closer relationship with the domestic swamp buffalo. Furthermore, the domestic river buffalo clusters out of this monophylum, revealing a paraphyletic state between the swamp and river type buffalos. The results lead to the postulation of an independent domestication of swamp and river buffalos from different wild ancestors, i.e. *B. arnee* is not the progenitor of both domestic forms, as commonly assumed.

5.2 Introduction

Currently the genus *Bubalus* comprises the wild Asian water buffalo *B. arnee*, the wild Asian dwarf buffalo *B. mindorensis* (tamaraw), as well as *B. depressicornis* (lowland anoa) and *B. quarlesi* (mountain anoa). The domestic form *B. bubalis* has been classified into two types – river (*B. bubalis bubalis*) and swamp (*B. bubalis carabenensis*) buffalo – based on their morphology, behaviour, and karyotype (Tanaka *et al*. 1996, Kumar *et al*. 2007a, Barker in press).

During the Pleistocene the genus *Bubalus* was widely distributed throughout Europe and South Asia, but due to environmental changes towards a drier climate became restricted to the Indian subcontinent and South-East Asia. In historical times,

the wild Asian water buffalo *B. arnee* ranged across South and South-East Asia, from Mesopotamia to Indochina. Presently, a dramatic decrease of its population is resulting in regional extinctions. However, if the remnant Asian water buffalos represent purebred wild animals is in dispute (Mason 1974, www.iucnredlist.org/details/3129/0).

The anoas, *B. depressicornis* and *B. quarlesi*, are wild dwarf buffalos native to the Indonesian island of Sulawesi. Although currently placed as a subgenus of *Bubalus*, uncertainties remain over their taxonomic status within the bovid family and the validity of the classification into two distinct species (Burton *et al*. 2005). The nomenclatural distinction in lowland and mountain anoa suggests altitudinal separation, with the large form (*B. depressicornis*) inhabiting low-lying coastal, swamp, and forest areas, and the smaller form (*B. quarlesi*) occupying high elevations. However, reports differ on preferred habitats and both species do not seem to be restricted to certain environments (Burton *et al*. 2005, www.arkive.org/mountain-anoa/bubalus-quarlesi/, http://www.arkive.org/lowland-anoa/bubalus-depressicornis/, http://www.iucnredlist.org/details/3126/0, http://www.iucnredlist.org/details/3128/0). According to fossil remains, both species inhabited the forests of Sulawesi since the Late Pleistocene (Groves 1969, van den Bergh *et al*. 2001), but in the last decades a significant decline of the anoa populations was recorded due to habitat loss and over-hunting (Burton *et al*. 2005).

Within the framework of this chapter, I focused especially on the wild Asian dwarf buffalo *B. mindorensis*. The tamaraw is the rarest wild buffalo, endemic to Mindoro, a Philippine island located off the coast of Luzon. Findings of several fossil teeth from surface accumulations on Luzon identified as *B. mindorensis*, suggest that the wild Asian dwarf buffalo occurred on Luzon as well as Mindoro during the Pleistocene, but has been restricted to the island of Mindoro in historical times (Beyer 1957, Kuehn 1986).

Once numerous and widespread across the entire island, the species is now classified as critically endangered. An estimated total of 250 to 300 individuals remain in three protected areas on Mindoro – Mount Iglit-Baco National Park,

Mount Calavite Preserve and Mount Aruyan Preserve. Main reasons for its population decline are relentless habitat destruction due to human population growth and the accompanying environmental degradation through farming, as well as illegal hunting (Custodio *et al*. 1996, Maala 2001, Ishihara & Kanai 2010). In order to prevent extinction of the tamaraw, effective conservation strategies are required including information about ecology, population dynamics, and genetic diversity of the Asian dwarf buffalo. However, acquisition of these information has proved to be extremely challenging due to the difficulties accessing their habitat and rare sightings of tamaraws in the field (Maala 2001, Ishihara & Kanai 2010). Tamaraws used to roam in open grasslands, swampy river valleys and upland forests (Kuehn 1986, Maala 2001). But since the reduction of their natural environment and increased hunting activity, tamaraws defensively retreat to the woodlands of Mindoro. They are fierce, nocturnal and wary animals, especially towards humans, usually living in small groups or solitary which impedes direct observation (Custodio *et al*. 1996, Maala 2001).

The position of the tamaraw within the bovid family has been subject to debate as it shares morphological features with both the Asian water buffalo and the anoas (Namikawa *et al*. 1995). So far, limited molecular genetic work has been carried out to resolve its particular demographic distribution and taxonomic status, currently placing it with the domestic swamp buffalo based on the mitochondrial cytochrome b gene (Tanaka *et al*. 1996).

Besides the unresolved phylogenetic history of the Asian buffalos, the proceedings of their domestication are still uncertain and various studies in the past gave rise to controversy. The domestic water buffalo has a great economic impact as a source of dairy, meat, and drought. But their evolutionary relationship as well as time and place of their domestication are still unresolved. *B. arnee* is commonly thought to be the progenitor of the domestic river and swamp buffalos (Lau *et al*. 1998, Kierstein *et al*. 2004, Lei *et al*. 2007, Yang *et al*. 2008). But the incomplete knowledge about the original zoogeographic distribution of the wild Asian water buffalo as well as its genetic variability and relation to the domestic forms due to the lack of DNA data,

makes it difficult to determine the proceedings of buffalo domestication (Yang *et al*. 2008, http://www.iucnredlist.org/details/3129/0).

The river buffalo is native to the Indian subcontinent, and has spread to the Middle East and South-East Europe during historical times. The swamp type is distributed in North-East India, Bangladesh, and China, further in South-East Asian countries (Cockrill 1981). The geographical ranges of the two domesticates are broadly parapatric, only overlapping in East India, Bangladesh, and possibly Myanmar, where interbreeding of the two types can occur (Kikkawa *et al*. 1997, Zhang *et al*. 2011, Barker in press). Cockrill (1981) suggested the river type has been domesticated about 4,000 to 5,000 years ago by the civilizations of the Euphrates, Tigris and Indus, whereas the centre of swamp buffalo domestication lies in the Yangtze valley. However, there is no proof that the range distribution of *B. arnee* included central China (Barker in press). Furthermore, the oldest archaeological findings of domestic buffalos come from Neolithic sites in southern China (6,500 to 8,000 years ago) and North-East Thailand (2,500 to 4,300 years ago) (Groves 2006, Barker in press).

Several genetic studies of the domestic river and swamp buffalo from various regions in Asia produced discordant conclusions about their origin of domestication. Kierstein *et al*. (2004) placed a single domestication centre in the Indian subcontinent, while Barker *et al*. (1997) and Lau *et al*. (1998) suggest domestication of the river type in India and the swamp buffalo in China from a common ancestor. More recent analyses propose an independent origin of river and swamp buffalos in West-India and South-China, respectively (Kumar *et al*. 2007a, Lei *et al*. 2007), while studies of ancient material by Yang *et al*. (2008) fail to establish direct linkages between Neolithic water buffalo remains from North-China and modern chinese water buffalos. Therefore, domestication must have taken place in regions outside of China, probably in South- or South-East-Asia. Zhang *et al*. (2011) indicate that the swamp buffalo was domesticated in the far south of China, northern Thailand, and Indochina. Subsequently, the swamp type spread south through peninsular Malaysia to Sumatra, Java and Sulawesi, as well as in northern direction through China, then Taiwan, the Philippines and Borneo.

As the current state of knowledge demonstrates, the phylogenetic relationships within the genus *Bubalus*, as well as the possible origin of domestication of the domestic river and swamp buffalo is still subject under discussion. So far, only a certain number of studies deal with the systematics of the Asian buffalos. Thereby, analyses focus on fragments of the mitochondrial d-loop and cytochrome b gene (Tanaka *et al*. 1996, Lau *et al*. 1998, Kierstein *et al*. 2004).

In order to resolve the phylogenetic status of the tamaraw within the bovid family, I generated two whole mitochondrial genomes ('mitogenomes') of *B. mindorensis* from museum samples, collected in 1935 on Mindoro Island. The efficiency of single mitochondrial genes versus complete mitochondrial genomes for inferring complex phylogenies, especially within the bovid family, has been discussed in recent papers (Rohland *et al*. 2007, Arif *et al*. 2012, Paijmans *et al*. 2013). Results show, that whole mitochondrial genomes provide a strong and reliable signal for interpreting relationships among closely related taxa, while individual gene trees can be inconsistent with the true evolutionary history of the genus, even when supported by high bootstrap values or Bayesian posterior probabilities (Rohland *et al*. 2007). As a consequence, the insufficient information provided by only partial mtDNA sequences can lead to discordant results when addressing a phylogenetic question multiple times, wherefore complete mitochondrial genomes may often be necessary to obtain correct phylogenies.

5.3 Material and Methods

Within the tamaraw project, two methodological different approaches were used to answer the question stated above. Firstly, sequences of the mitochondrial cytochrome b were generated by traditional Sanger-sequencing. Afterwards, whole mitochondrial genomes were created via a high-throughput sequencing approach. The initial analyses of a single marker served to quantify the amount of endogenous DNA within the museum samples in order to select the most suitable specimens for subsequent next-generation-sequencing. Additionally, sequences of cytochrome b were obtained in order to combine previously published DNA to a dataset with a maximum number of samples.

In a further step, whole mitochondrial genomes of *B. mindorensis* were produced via high-throughput sequencing from the two samples yielding the most endogenous DNA. Thereby, the choice of instrument and sequencing approach is mainly influenced by their advantages for particular applications, as well as the project goal, and the quality of the DNA material. Here, a shotgun sequencing approach is applied using an Illumina Hi-seq platform, as their technology provides sufficient read length for fragmented aDNA. Shotgun sequencing is characterised by random targeting for DNA without prior selection, and therefore commonly used in studies where no DNA research has been carried out so far. Furthermore, this approach can provide valuable information about the quality and quantity of endogenous DNA in museum samples, in order to test their general suitability for next-generation-sequencing.

5.3.1 Samples

A total of 9 tamaraw (*B. mindorensis*) bone or dried tissue samples were collected from the Chicago Field Museum (FMNH) and the National History Museum Gothenburg. The specimens stem all from Mindoro Island, Philippines. Furthermore, 57 modern Anoa (*B. depressicornis, B. quarlesi*) samples from Sulawesi, analysed by members of the DEAD-lab in the facilities at the Archaeology Department of Durham University as part of a different project, were incorporated in the analyses (see figure 5.1). A list of all samples is found in chapter 7.2.4, including additional information.

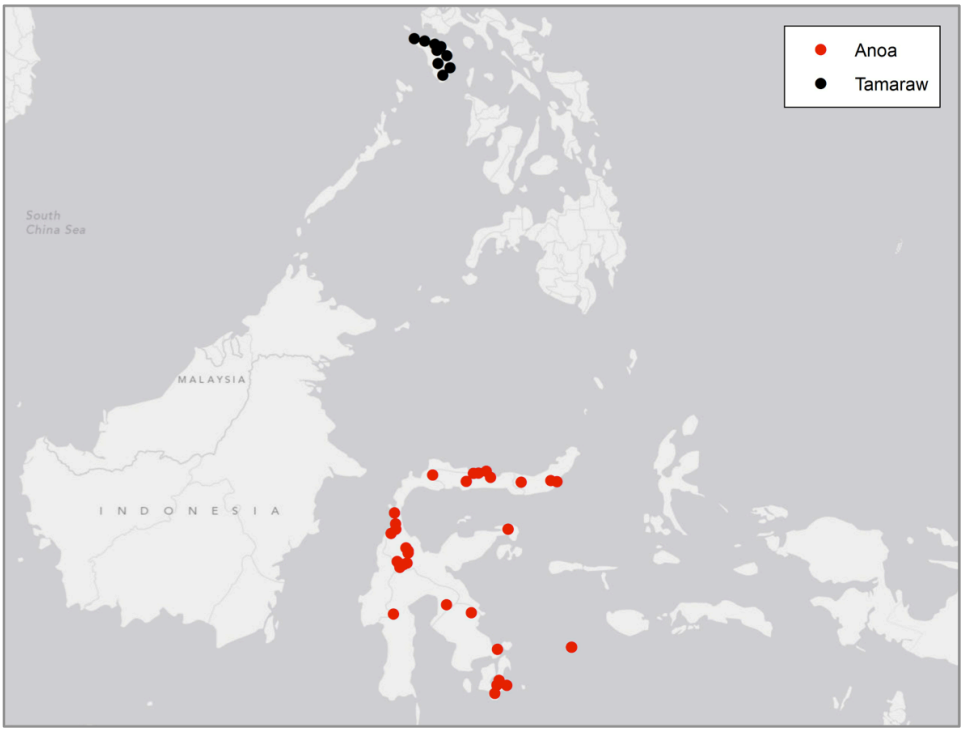

Figure 5.1: Sample distribution of the tamaraw and the anoa.
Distribution of the sampling locations of B. mindorensis (tamaraw) is indicated in black. Distribution of the sampling locations of B. depressicornis (lowland anoa) and B. quarlesi (mountain anoa) are indicated in red.

5.3.2 Analyses

Analyses of mitochondrial cytochrome b via Sanger-sequencing

The museum samples were analysed in the facilities at the Archaeology Department of Durham University. Because of the age of certain samples (~ 100 years old), all pre-PCR working steps, including sample preparation, DNA extraction and PCR set-up, were conducted in a dedicated ancient DNA laboratory. The bone samples were ground to fine powder and extracted following the protocol described in chapter 7.1.2, extraction protocol 4, the dried tissue samples were treated following the extraction protocol 1.

All subsequent steps, including PCR-amplification, purification, and sequencing set-up were carried out in a separate modern laboratory, while sequencing was performed by the DNA Sequencing Service at the School of Biological and Biomedical Sciences at Durham University (see chapters 7.1.3 and

7.1.4). Mitochondrial DNA was amplified and sequenced in 7 overlapping fragments, spanning 1142 bp of the entire cytochrome b (see chapter 7.1.1). The sequencing chromatograms were edited manually, subsequently assembled, and a consensus sequence per individual exported (see chapter 7.1.4) using Geneious R6 version 6.0.6 (Drummond *et al*. 2011). Standard anti-contamination guidelines were followed to ensure authenticity of aDNA results (see chapters 1.4.3 and 7.1.5).

Phylogenetic analyses were performed by construction of Bayesian trees using MrBayes version 3.2.2 (Huelsenbeck & Ronquist 2001). The best-fit nucleotide substitution model, ascertained in MEGA-CC (Kumar *et al*. 2012) under the Bayesian Information Criterion (BIC), was Tamura-3-parameter +Γ. Under this model, a consensus tree with bootstrap values shown on the nodes was created for 5,000,000 generations, a sample frequency of 5,000 and with a burn-in period of 100 trees.

Analyses of the whole mitochondrial genome via next-generation-sequencing

After initial analyses of the museum samples in the facilities at the Department of Archaeology at Durham University via Sanger-sequencing, extracts of the 2 samples yielding a sufficient amount of endogenous DNA were sent to the Centre for GeoGenetics at the University of Copenhagen for a shotgun sequencing approach using an Illumina Hi-seq platform. Library preparation (protocol after Schröder, see chapter 7.1.4) was carried out in the ancient DNA facilities of the Institute, considering standard anti-contamination guidelines. All following reactions – library amplification, sample quantification and the subsequent sequencing reaction were conducted in a separate modern DNA laboratory (see chapter 7.1.4).

Initial validation of the raw data was done by Dr. Maria Avila at the Centre for GeoGenetics at the University of Copenhagen. The final analysis – including editing of the reads, alignment to a reference genome, and export of a consensus file of the mitochondrial genotype was carried out by Dr. Laurent Frantz from Wageningen University (see chapter 7.1.5). Standard protocols were followed according to authenticity criteria of next-generation-sequencing of ancient DNA (see chapters 1.4.5 and 7.1.5).

Phylogenetic relationships within the bovid family were ascertained using MrBayes version 3.2.2 (Huelsenbeck & Ronquist 2001). The best-fit nucleotide substitution model, selected in MEGA-CC (Kumar *et al*. 2012) under the Bayesian Information Criterion (BIC), was General Time Reversible (GTR) +Γ +I. Under this model, a consensus tree with bootstrap values shown on the nodes was created for 5,000,000 generations, a sample frequency of 5,000 and with a burn-in period of 100 trees.

5.3.3 Results

Results of Sanger-sequencing

DNA sequences for the 1,142 bp of the cytochrome b gene were obtained from 6 of the 9 museum samples. Additionally, the sequences acquired from next-generation-sequencing were added to the alignment and furthermore combined with sequences of 57 modern Anoa (*B. depressicornis, B. quarlesi*) samples from Sulawesi, as well as 119 sequence entries of different bovine species from GenBank to generate a dataset composed of 184 individuals.

Bayesian analyses generated a well-resolved tree, separating the bovine (*Bos*, *Bison*) from the bubaline (*Bubalus*) species (see figure 5.2). Regarding *Bubalus*, the analyses showed that samples of tamaraw, swamp and river buffalo fall into separate clades, supported by high bootstrap values. Thereby, the tamaraws and swamp buffalos form sister taxa, while being paraphyletic to the river buffalos. In contrast, the anoa samples do not appear as uniform, splitting into a polytomy instead. The saola (*Pseudoryx nghetinhensis*) and African buffalo (*Syncerus caffer*) form the basis of the Asian bovids.

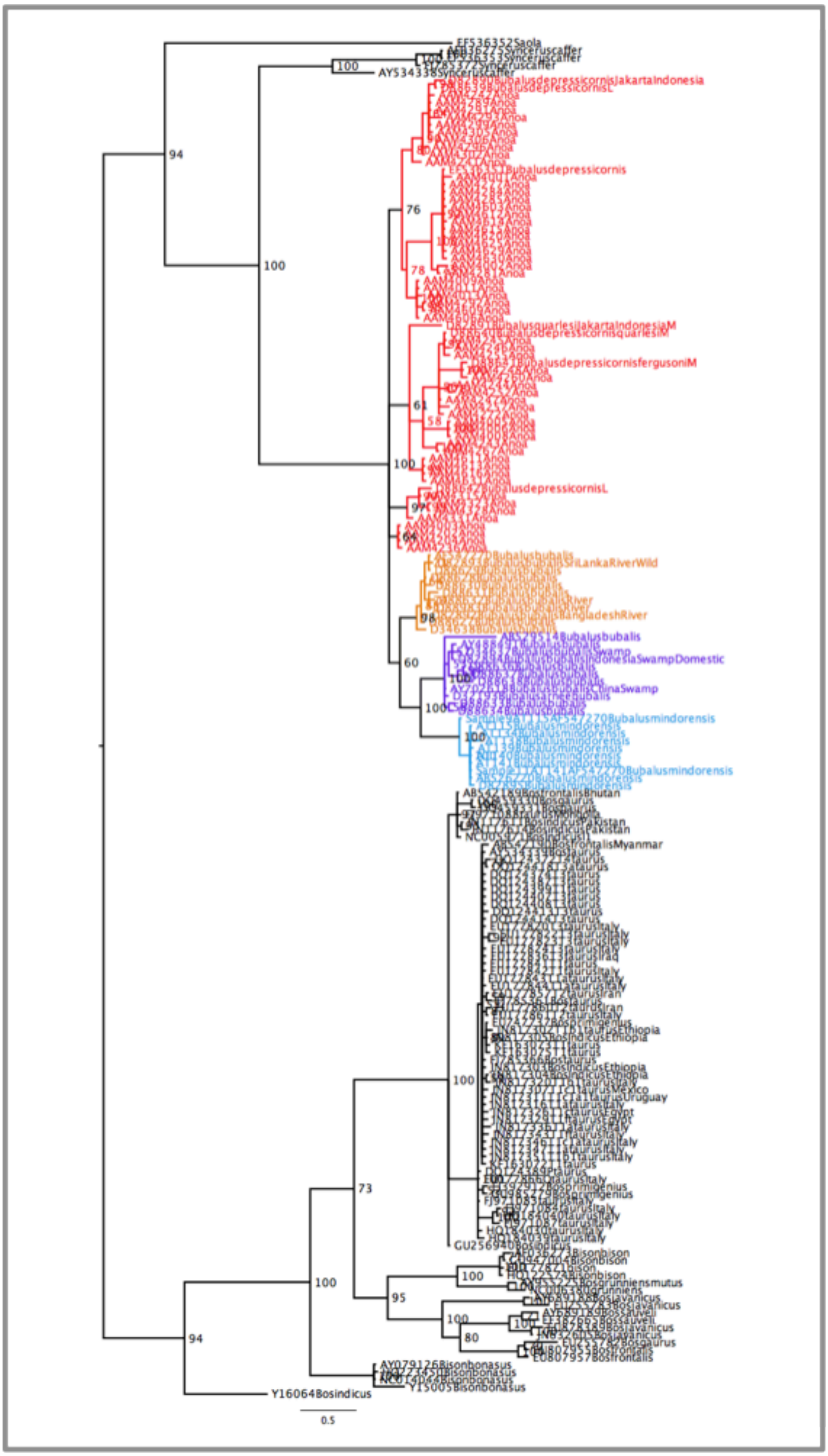

Figure 5.2: Mitochondrial phylogeny of the bovid family inferred from cyt b.

(Figure 5.2 continued)
Inferred from a 1,142 bp fragment of the mitochondrial cyt b. Phylogenetic analyses were performed using MrBayes version 3.2.2. The best-fit nucleotide substitution model under the Bayesian Information Criterion (BIC) was Tamura-3-parameter +Γ. A consensus tree with bootstrap values shown on the nodes was created for 5,000,000 generations, a sample frequency of 5,000 and with a burn-in period of 100 trees. Anoas are indicated in red, domestic river buffalos are indicated in orange, domestic swamp buffalos are indicated in purple, tamaraws are indicated in blue. Species of the genus Bos, Syncerus, and Pseudoryx are indicated in black.

Results of next-generation-sequencing

Complete mitochondrial genomes were successfully recovered from both of the museum samples via an Illumina shotgun approach. Afterwards, reads from both samples were mapped to a reference genome of *B. bubalis* (NCBI accession number: AF547270). Statistical results of the NGS analysis are shown in figure 5.3 and 5.4.

NGS results Ref AF547270	Sample 9 (AT115)	Sample 11 (AT141)
reads mapped	31,460	69,043
reads mapped uniquely	4,616	33,760
average depth	6.031	81.792
% of reads mapped	0.00047	0.00333
% of endogenous DNA	54	63
% of human DNA	4	4
% of other DNA	42	33

Figure 5.3: Statistical results of the NGS analysis.

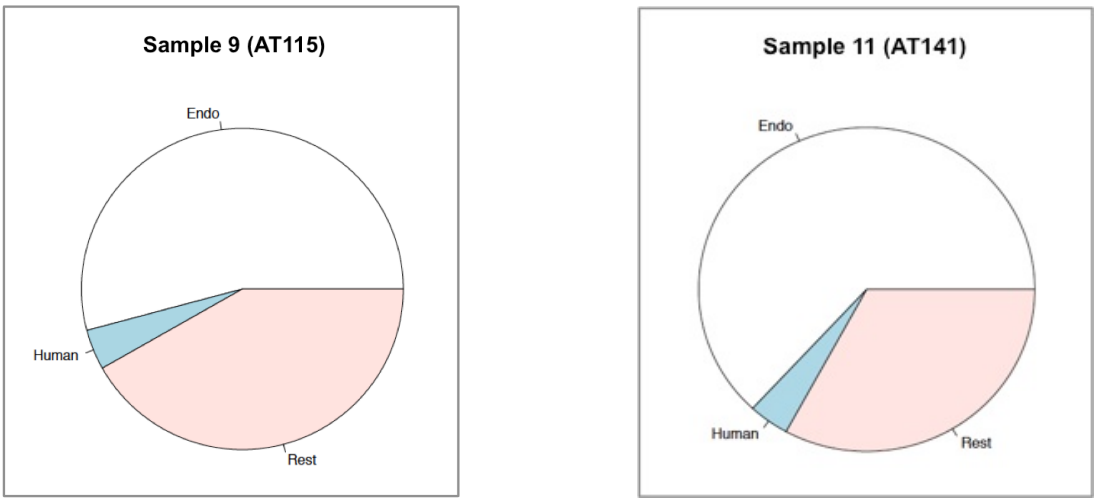

Figure 5.4: Diagram of the endogenous DNA content of the museum samples (B. mindorensis) used for NGS.

(Figure 5.4 continued)
Sample 9 (AT115): 54% endogenous DNA (Endo), 4% human DNA (Human), 42% other DNA (Rest). Sample 11 (AT141): 63% endogenous DNA (Endo), 4% human DNA (Human), 33% other DNA (Rest).

The successfully recovered sequences were combined with 38 sequence entries of different bovine species from GenBank to generate a dataset composed of 40 individuals. The phylogenetic analyses using whole mitochondrial genomes showed a well-resolved tree with maximum bootstrap values of 100 (see figure 5.5). Results appeared to be slightly different compared to the tree generated from the cytochrome b gene only (compare with figure 5.2). Again, tamaraw and swamp buffalo form a monophyletic group, but cluster more closely with *B. depressicornis* than with river buffalo, while saola and African buffalo sit on the basis of the bubaline species.

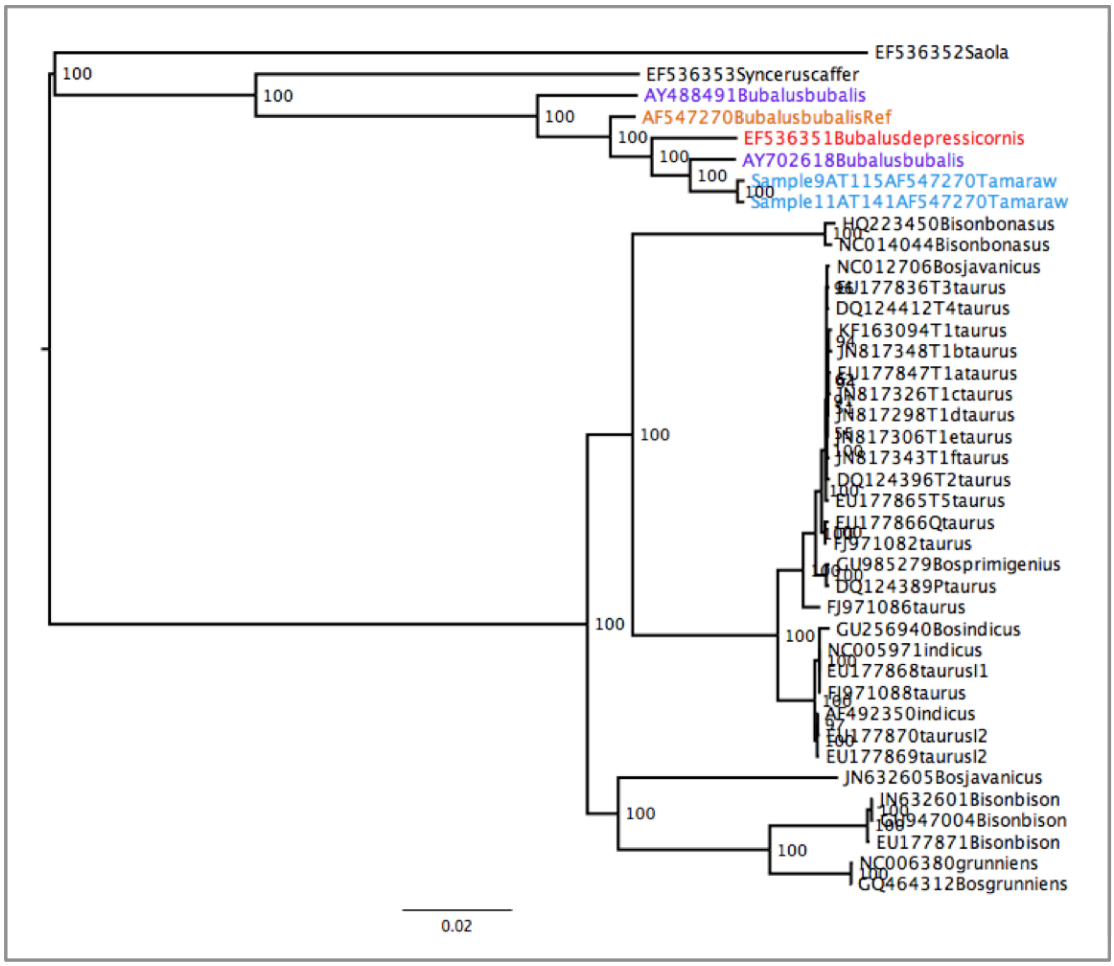

Figure 5.5: Mitochondrial phylogeny of the bovid family inferred from full mitogenomes.

(Figure 5.5 continued)
Phylogenetic analysis inferred from full mitogenomes using MrBayes version 3.2.2. The best-fit nucleotide substitution model under the Bayesian Information Criterion (BIC) was General Time Reversible (GTR) +Γ +I. A consensus tree with bootstrap values shown on the nodes was created for 5,000,000 generations, a sample frequency of 5,000 and with a burn-in period of 100 trees. Anoas are indicated in red, domestic river buffalos are indicated in orange, domestic swamp buffalos are indicated in purple, tamaraws are indicated in blue. Species of the genus Bos, Syncerus, and Pseudoryx are indicated in black.

5.4 Discussion

Taxonomic status of the tamaraw

The phylogenetic relationships recovered from this study provide new insight into both the taxonomy of the Asian bovids and the history of their domestication. So far, only limited molecular genetic work has been carried out to resolve the taxonomic status of the tamaraw within the bovid family (Tanaka *et al*. 1996).

In the study presented here, tree-topologies resulting from the cytochrome b gene and the first complete mitogenomes of *B. mindorensis*, both place the Asian dwarf buffalo monophyletically with the domestic swamp buffalo *B. bubalis carabenensis*. This is supported by high bootstrap values of 100 with each marker. Although tamaraws share morphological features with both, Asian water buffalos and the anoas (Namikawa *et al*. 1995), results demonstrate a closer relationship with the domestic swamp buffalo. Furthermore, the domestic river buffalo clusters out of this monophylum, revealing a paraphyletic state between the swamp and river type buffalos. Therefore, domestic swamp and river buffalos can be differentiated at least at the subspecies level.

Here, phylogenetic analyses of cytochrome b and the mitogenomes generally show the same results, but differ in one point. The tree-topology of the whole mitochondrial sequences show that tamaraw and swamp buffalo form a monophyletic group, but cluster more closely with *B. depressicornis* than with river buffalo. As explained earlier, the analysis of whole mitochondrial genomes is more powerful for deciphering inter- and intraspecific relationships than single genes. However, the mitogenomes of *Bubalus* species available on GenBank, that are incorporated in this study, are mostly from unpublished papers and authors provide

no further information about the provenance of the samples or methods used. Therefore, uncertainty about the authenticity of the sequences remain.

The samples used to generate whole mitochondrial genomes of *B. mindorensis* stem from museum specimens, collected about 80 years ago on Mindoro Island. Historical samples from museum collections or other archives become increasingly important for phylogenetic or phylogeographic studies as they provide significant advantages (Paijmans *et al*. 2013). Some species are difficult to sample in the wild, like the endangered tamaraw. As described earlier, sightings of the Asian dwarf buffalo are rare due to its nocturnal activity and retreat to the forests of Mindoro. Here, museum specimens offer a powerful opportunity to obtain information about the genetic diversity of a threatened species. These information are essential to develop appropriate conservation strategies in order to prevent extinction of the Asian dwarf buffalo.

Although museum samples are in general considerably younger than archaeological remains, the endogenous DNA is often of low quality due to the storage conditions. Therefore, historical samples provide similar technological challenges as samples from the fossil record used in aDNA studies. As demonstrated in this study, museum samples are a valuable source to generate mitogenomes via next-generation-sequencing. Both specimens yielded over 50% of endogenous DNA, with an average depth of 6 and 81, respectively.

Taxonomic status of the domestic water buffalos

Although several studies address the question of the origin of Asian water buffalo domestication, results come to ambiguous conclusions due to uneven sampling across Asia (Barker *et al*. 1997, Lau *et al*. 1998, Kierstein *et al*. 2004, Kumar *et al*. 2007a, Lei *et al*. 2007, Yang *et al*. 2008, Zhang *et al*. 2011). Currently, growing evidence supports a distinct genetic origin of river and swamp buffalo (Kumar *et al*. 2007a, Lei *et al*. 2007). This study reaffirms the results of latest research that the domestic swamp buffalo is more closely related to the tamaraw than to the domestic

river type. But what impact does this have on the domestication history of the Asian buffalos?

Here, in both analyses – of the cytochrome b and the whole mitochondrial genome – the domestic river buffalo clusters out of the monophylum of swamp buffalo and tamaraw, revealing a paraphyletic state between the domestic swamp and domestic river buffalos. The results lead to the postulation of an independent domestication of swamp and river buffalos from different wild ancestors, i.e. *B. arnee* is not the progenitor of both domestic forms, as commonly assumed (Lau *et al*. 1998, Kierstein *et al*. 2004, Lei *et al*. 2007, Yang *et al*. 2008). At present, a lack of DNA sequences of the wild Asian water buffalo and the scarce fossil remains of potential wild ancestors prevents resolving the particular question of ancestry.

The results of this research project have significant impact on the systematics of *Bubalus* and furthermore on the current terminology of the Asian buffalos. So far, the Asian bovids were classified into the wild Asian water buffalo *B. arnee*, the wild Asian dwarf buffalo *B. mindorensis*, as well as the anoas *B. depressicornis* and *B. quarlesi*. The domestic form *B. bubalis* has been categorised into river and swamp buffalo (Tanaka *et al*. 1996, Kumar *et al*. 2007a, Barker in press). But the paraphyletic state of the domestic forms – where the swamp type clusters more closely with the tamaraw – leads to a revision of the nomenclature of the Asian dwarf buffalo *B. mindorensis*.

Despite the results shown here, there are still open questions about the process of domestication of the wild Asian water buffalo *B. bubalis*. Future research particularly needs to focus on the current lack of DNA sequences of the wild Asian water buffalo in order to address the question of the wild progenitor of the domestic swamp and river buffalos. Furthermore, adding a temporal parameter to the analyses by molecular-clock based Bayesian methods will help to unravel the demographic history of the Asian bovids. By assessing a timeframe, the distinct geographic distribution of each species can be correlated with climatic events like sea level changes to distinguish between post-glacial and domestication-associated population expansions.

5.5 References

Literature

Arif IA, Bakir MA, Khan HA (2012) Inferring the Phylogeny of Bovidae Using Mitochondrial DNA Sequences: Resolving Power of Individual Genes Relative to Complete Genomes. *Evolutionary Bioinformatics* **8**, 139-150.

Barker JSF (in press) Water buffalo (*Bubalus bubalis*, Linnaeus 1758). In: C. Smith (ed.) Encyclopedia of Global Archaeology. Springer, New York.

Beyer HO (1957) New finds of fossil mammals from the Pleistocene strata of the Philippines. *Natl. Res. Counc. Philippines Bull.* **41**, 220-239.

Burton JA, Hedges S and Mustari AH (2005) The taxonomic status, distribution and conservation of the lowland anoa *Bubalus depressicornis* and mountain anoa *Bubalus quarlesi*. *Mammal Rev*. **35**, 25-50.

Cockrill WR (1981) The water buffalo: a review. *British Veterinary Journal* **137**, 8-16.

Custodio CC, Lepiten MV, Heaney LR (1996) Bubalus mindorensis. In: A. V. Linzey (ed.) Mammalian Species 520. American Society of Mammalogists, 1-5.

Drummond AJ, Ashton B *et al*. (2011) Geneious R6 version 6.0.6, available from http://www.geneious.com

Groves CP (1969) Systematics of the anoa (Mammalia, Bovidae). *Beaufortia* **17**, 1-12.

Groves CP (2006) Domesticated and commensal mammals of Austronesia and their histories. In: P. Bellwood *et al*. (eds.) The Austronesians Historical and Comparative Perspectives. ANU EPress, Canberra, 161-173.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755. Version 3.2.2, available from http://mrbayes.sourceforge.net

Ishihara S & Kanai Y (2010) Population Dynamics, Herd Behaviour, and Genetic Analysis of the Critically Endangered Tamaraw (Bubalus mindorensis): A Mini Review. *Journal of Developments in Sustainable Agriculture* **5**, 136-140.

Kierstein G, Vallinoto M, Silva A *et al*. (2004) Analysis of mitochondrial D-loop region casts new light on domestic water buffalo (*Bubalus bubalis*) phylogeny. *Molecular Phylogenetics and Evolution* **30**, 308-324.

Kuehn DW (1986) Population and Social Characteristics of the Tamarao (*Bubalus mindorensis*). *Biotropica* **18**, 263-266.

Kumar S, Nagarajan JS, Sandhu N *et al*. (2007a) Mitochondrial DNA analyses of Indian water buffalo support a distinct genetic origin of river and swamp buffalo. *Animal Genetics* **38**, 227-232.

Kumar S, Nagarajan M, Sandhu JS *et al*. (2007b) Phylogeography and domestication of Indian river buffalo. *BMC Evolutionary Biology* **7**, 186-191.

Kumar S, Stecher G *et al*. (2012) MEGA-CC: Computing Core of Molecular Evolutionary Genetics Analysis Program for Automated and Iterative Data Analysis. *Bioinformatics* **28**, 2685-2686. Available from http://www.megasoftware.net

Lau CH, Drinkwater RD, Yusoff K *et al*. (1998) Genetic diversity of Asian water buffalo (Bubalus bubalis): mitochondrial DNA D-loop and cytochrome b sequence variation. *Animal Genetics* **29**, 253-264.

Lei CZ, Zhang W, Chen H *et al*. (2007) Independent maternal origin of Chinese swamp buffalo (Bubalus bubalis). *Animal Genetics* **38**, 97-102.

Maala CP (2001) Endangered Philippine Wildlife Species with Special Reference to the Philippine Eagle (*Pithecophaga jefferyi*) and Tamaraw (*Bubalus mindorensis*). *Journal of International Development and Cooperation* **8**, 1-17.

Mason IL (1974) Species, types and breeds. In: W. R. Cockrill (ed.) The Husbandry and Health of the Domestic Buffalo. Food and Agriculture Organization of the UN, Rome, 1-47.

Paijmans JLA, Gilbert MTP, Hofreiter M (2013) Mitogenomic analyses from ancient DNA. *Molecular Phylogenetics and Evolution* **69**, 404-416.

Rohland N, Malaspinas AS, Pollack JL *et al*. (2007) Proboscidean Mitogenomics: Chronology and Mode of Elephant Evolution Using Mastodon as Outgroup. *PloS Biology* **5** (8), 1663-1671.

Tanaka K, Solis CD, Masangkay JS *et al*. (1996) Phylogenetic Relationship Among All Living Species of the Genus *Bubalus* Based on DNA Sequences of the Cytochrome b Gene. *Biochemical Genetics* **34**, 443-452.

van den Bergh GD, de Vos J and Sondaar PY (2001) The late quarternary palaeogeography of mammal evolution in Indonesian Archipelago. *Palaeontology, Palaeoclimatology, Palaeoecology* **171**, 385-408.

Yang DY, Liu L, Chen X, Speller CF (2008) Wild or domesticated: DNA analysis of ancient water buffalo remains from north China. *Journal of Archaeological Science* **35**, 2778-2785.

Zhang Y, Vankan D, *et al*. (2011) Genetic differentiation of water buffalo (*Bubalus bubalis*) populations in China, Nepal and south-east Asia: inferences on the region of domestication of the swamp buffalo. *Animal Genetics* **42**, 366-377.

Electronical sources (2014)

http://www.iucnredlist.org/details/3126/0
http://www.iucnredlist.org/details/3128/0
http://www.iucnredlist.org/details/3129/0
http://www.arkive.org/mountain-anoa/bubalus-quarlesi/
http://www.arkive.org/lowland-anoa/bubalus-depressicornis/

6 Summary

6.1 Conclusions and Future Directions

This thesis aimed to assess the potential of using commensal animals, like the black rat (*Rattus rattus*), the house mouse (*Mus musculus*), and the Asian house gecko (*Hemidactylus frenatus*) as a proxy to infer human history. Several studies revealed the advantage of bioproxies to provide insights in early human travel routes. For example, the current distribution of *M. m. domesticus* is directly caused by human movements, as the observed pattern of genetic lineage dispersal can be linked to maritime activities of humans within different time periods from the Iron Age to the 15th century (Jones *et al*. 2011, Bonhomme & Searle 2012). Furthermore, Matisoo-Smith & Robins (2004, 2009) showed the significance of bioproxies by revealing insights into the spread of the Lapita culture throughout the Pacific on the basis of the Pacific rat (*Rattus exulans*). The genetic diversity among human populations in Polynesia is very low and therefore not sufficiently informative. Instead, phylogenetic analysis of *R. exulans*, which was deliberately transported in canoes as a source of food, unraveled multiple introductions of colonists and their particular paths of migration to various Pacific islands (Addison & Matisoo-Smith 2010). As these examples demonstrate, analysing bioproxies is a valuable approach and circumvents problems arising from scarcity of material artefacts and studying humans directly.

Here, results of the different case studies show that wide-spread commensal species across the Indian Ocean still carry genetic signals of early human contacts, even when the original founders and their descendants have left, died out or admixture by subsequent immigration is covering initial lineages. Due to their remarkable adaptability, close association with humans, and ecological dependence on them, the house mouse, the black rat, and the Asian house gecko have become one of the most pervasive animals, currently occupying almost all continents around the globe. Moreover, this close relationship leads to a shared history of commensal populations and the humans, who transported them.

In chapter 2, assessment of the biogeographic history of *M. m. castaneus* revealed three different haplogroups after Rajabi-Maham *et al*. (2012), whereas only one (HG2) has been transferred to many places outside their natural bounds. The cause of only one widely distributed population may lie in the occupation of a specific niche, that promoted commensalism with humans or that HG2 coincidentally has been in the right place at the right time to get transported preferably over HG1 and HG3. Once humans started to move this particular population around and established it in various regions outside their natural habitat, it became more difficult for other specimens to invade these areas. Within the relocated population, two groups are apparent within the phylogenetic tree, spreading from the Indian peninsula in eastward and westward direction. Distribution patterns strongly indicate human-mediated seaborne dispersal in westward direction associated with Arab seafarers and overland transport over long distances to the East from the late 1st millennium AD onwards, when contacts between trans-regional trade networks became more frequent (Boivin *et al*. 2013). A lack of archaeological remains of the house mouse before the 1st millennium AD confirms the current view. Further, earlier translocations are less likely, due to specific behavioural-ecological characteristics of house mice. The accumulation of the highly synanthropic *M. m. castaneus* is a good indicator of established sedentary societies with a subsistence based on agriculture (Auffray & Britton-Davidian 2012). Therefore, translocation to other places must have happened in times, when settled societies have already been formed trade and transport have been well established.

In chapter 3, the phylogeographic structure in *R. rattus* has been evaluated. Again, results revealed four distinct Lineages I to IV after Aplin *et al*. (2011), of which only Lineage I has been found in various places outside its native range. Further analysis showed a subdivision of Lineage I in several unique and geographically restricted populations along the East African coast, that reflect patterns of human movement throughout the Indian Ocean associated with the developing trade activities in the 1st millennium AD and later with European seafarers during the Age of Discovery. With the emergence of the Swahili culture during the 1st millennium AD, trade relations

from the East African coast with different partners such as the Near East, India, and as far as China have been established, leading to an overall increase of seaborne connections (Horton & Middleton 2000, Boivin *et al*. 2013). These activities have not only led to the exchange of exotic goods or food items, but to translocations of commensal species. The appearance of the commensal black rat in East Africa is documented in archaeological layers from the mid 1st millennium AD, a time of urbanisation and rising port cities which provided a perfect niche for this rodent (Plug 2000, Helm 2000, Boivin *et al*. 2013). Further, archaeological evidence revealed that certain East African port cities favoured relations with merchants from distinct areas (Horton & Middleton 2000), which may explain the pattern of geographically unique groups among the black rat Lineage I along the East Africa coast.

In chapter 4, molecular genetic methods where used to explore the phylogeography of the Asian house gecko *H. frenatus*. In order to assess whether the current distribution of the Asian house gecko is shaped by natural translocation or if human-mediated introductions play a decisive role, continuous phylogeographic diffusion modelling was applied. Results revealed three distinct clades, whereas clade A is restricted to the natural habitat of *H. frenatus* in South-East Asia and the Indian subcontinent. Clade B and C are spreading east- and westward from the core area, strongly indicating a recent anthropogenic introduction on several locations throughout the Indian and Pacific Ocean, but probably in more recent times. Due to a lack of archaeological remains, the beginning of the commensal relationship between the Asian house gecko and humans is still unclear. Although *H. frenatus* may have been initially transported to places outside their natural range by Polynesian and Melanesian seafarers approximately 4,000 years ago (Case *et al*. 1994), their vast range expansion occurred only recently, particularly over the last decades (Hoskin 2011). This development may have been facilitated by some distinct biological features of the Asian house gecko, which appear to be advantageous especially in modern urban environments and current international shipping and cargo movement.

In traditional archaeology, evidence of contact and exchange between different cultures in (pre)history is based on remains of material artefacts or historical data. The distribution of populations is based on the zooarchaeological record. However, in some areas archaeological evidence is insufficient, especially in the case of the widely scattered islands in the Indian and Pacific Ocean. Furthermore, archaeological remains can provide an indication of arrival times for certain species, but without genetic analyses it is impossible to track the source population (Boivin *et al*. 2013). In the case of micro-mammals and –reptiles, the zooarchaeological record is scarce and identification of individual species is difficult. Genetic analysis of skeletal remains can therefore help to distinguish between species and unravel complex population histories, that can be related to human contacts. How far a population history can be traced back is dependent on the age of the samples.

The commensal study species used within the framework of this thesis have all been moved around by people for a long time. Therefore, an undirected and indistinct distribution of contemporal populations was expected for each species. In order to gain insights into their distribution patterns, museum and modern samples of each species have been analysed through molecular genetics. Results have demonstrated, that despite the young age of the samples of around a 100 years at maximum, several unique and geographically restricted lineages have been identified, which reflect past human-aided translocations from the 1st millennium AD onwards throughout the Indian and Pacific Ocean. In combination with the archaeological finds such as zooarchaeological and archaeobotanical remains or pottery and ornaments, a more precise picture of earliest contacts and exchange can be reconstructed.

Thereby, results have shown that the different species – due to their distinct ecological characteristics – are each suited to answer particular questions. Rats and mice are dependent on humans and exploit their habitat and food sources. Therefore, they are a good marker of settled societies with agricultural subsistence. The first evidence of mammalian commensalism stem from the Levant about 12,000 years ago in association with the Natufian culture, which is characterised by a semi-sedentary lifestyle including the cultivation and storage of grain to some degree (Cucchi *et al*.

2012). The shift from hunter-gatherers towards sedentism created new ecological niches, providing a consistent source of food, as well as artificial shelter protecting against predation or interspecific competition. Exploitation of these new resources has led to a close relationship between humans and these rodent species, hence favoured their global distribution since early times. Because of this close relationship over a long time, *R. rattus* and *M. m. castaneus* are suitable – in combination with the zooarchaeological record – to especially trace human contacts in (pre)historic times.

Contrary to that, archaeological evidence of the spreading of the Asian house gecko in early times is insufficient. Although *H. frenatus* is highly synanthropic and may have been transported already in early times (Case *et al*. 1994), their history of commensalism remains opaque. In fact, the dispersal of this species is associated with the continuous urbanisation throughout the region, as well as the frequent international sea freight at present. The modern environment shaped by humans, provides a perfect niche for *H. frenatus* to hunt successfully, why its mainly found in urban and suburban structures and cargo containers, less in natural habitats (Cole 2005, Newbery & Jones 2007, Csurhes & Markula 2009, Hoskin 2011). Therefore, the Asian house gecko does not represent an ideal proxy to trace back early human trade routes, but modern invasive pathways. Furthermore, the lack of archaeological samples impedes insights into population processes of *H. frenatus* over the past millennia.

Although this study highlights the benefits that can be obtained from molecular genetic analyses of museum material to infer human history, there are also some obstacles to overcome. The incorporation of ancient DNA adds a temporal aspect to the genealogy of the different study species. However, assessing correct mutation rates in order to date population divergence is difficult. Here, the use of single mitochondrial marker and the evaluation of changes within a recent time span, compared to the overall population history of the study species, impedes a precise timing. One has to bear in mind that the results are only the reflection of the evolutionary history of one mitochondrial marker, which may be inconsistent with the true population history. The dates assumed for the splits of haplogroups in the

individual studies are therefore only an indicator for the maximum time span of separation. Population splits may have happened more recently, than the divergence between gene lineages. In order to overcome the current lack of sensitivity of the data, incorporation of more sequence data is necessary to take all genetical changes into account. Further, adding zooarchaeological samples to modern and museum specimens gives an insight into even earlier processes of the population history. But DNA recovery, especially of small and fragile bones from warmer environments is challenging, even when the rapid technological developments in palaeogenetic research nowadays enable not only the generation of single gene sequence data, but whole mitochondrial and even nuclear genomes.

The advantages of the latest high-throughput sequencing techniques, which allow the generation of whole genomes even from old samples, is demonstrated in a further methodological exercise of this thesis. In chapter 5, the first full mitogenomes of the endangered Asian dwarf buffalo (*B. mindorensis*) have been created in order to solve its taxonomic position within the bovid family. Acquisition of information about the genetic diversity, ecology, and population dynamics of endangered species is essential to apply effective conservation strategies and save them from extinction. Results place the Asian dwarf buffalo monophyletically with the domestic swamp buffalo *B. bubalis carabenensis*, revealing a paraphyletic state between the domestic swamp and river type buffalos. This has led to the postulation of an independent domestication of swamp and river buffalos from different wild ancestors, i.e. *B. arnee* is not the progenitor of both domestic forms, as commonly assumed.

Recent developments in the field of aDNA have widened the range of applications and questions that can be addressed. Refined extraction and sequencing techniques allow the analysis of ever decreasing amounts of DNA on genomics level, enabling the use of aDNA methods to understand the spatiotemporal distribution of genetic lineages (Larson *et al*. 2007, Dabney *et al*. 2013, Gansauge & Meyer 2013). Historical samples from museum collections have become increasingly important for phylogeographic studies as they provide significant advantages (Paijmans *et al*. 2013), especially in the case of small commensal animals. Recovery of fossil remains of rodents or small reptiles is difficult due to their little size and fragile nature. Here,

museum samples provide a powerful opportunity to obtain information about the genetic diversity of species that are difficult to sample in the wild or to find during the excavation process.

6.2 References

Addison DJ & Matisoo-Smith E (2010) Rethinking Polynesians origins: a West-Polynesia triple-I model. *Archaeol. Oceania* **45**, 1-12.

Aplin KP, Suzuki H *et al*. (2011) Multiple Geographic Origins of Commensalism and Complex Dispersal History of Black Rats. *PloS One* **6**, 26357.

Auffray JC & Britton-Davidian J (2012) The house mouse and its relatives: systematics and taxonomy. In: Macholan M *et al*. (eds.) Evolution of the house mouse. Cambridge University Press, 1-34.

Boivin N, Crowther A, Helm R, Fuller D (2013) East Africa and Madagascar in the Indian Ocean World. *Journal of World Prehistory* **26** (3), 213-281.

Bonhomme F & Searle JB (2012) House mouse phylogeography. In: Macholan M *et al*. (eds.) Evolution of the house mouse. Cambridge University Press, 278-296.

Case TJ, Bolger DT, Petren K (1994) Invasions and Competitive Displacement Among House Geckos in the Tropical Pacific. *Ecology* **75**, 464-477.

Cole NC (2005) The ecological impact of the invasive house gecko *Hemidactylus frenatus* upon endemic Mauritian geckos. Dissertation, Bristol.

Csurhes S & Markula A (2009) Pest animal risk assessment: Asian house gecko *Hemidactylus frenatus*. Biosecurity Queensland, Queensland Primary Industries and Fisheries, Department of Employment, Economic Development and Innovation (Queensland, Australia).

Cucchi T, Auffray JC, Vigne JD (2012) On the origin of the house mouse synanthropy and dispersal in the Near East and Europe: zooarchaeological review and perspectives. In: S. J. E. Baird *et al*. (eds.) Evolution of the House Mouse. Cambridge University Press, 65-93.

Dabney J, Knapp M, Glocke I *et al*. (2013) Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *PNAS* **110** (39), 15758-15763.

Gansauge MT & Meyer M (2013) Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc*. **8**, 737-748.

Helm RM (2000) Conflicting histories: The archaeology of the iron-working, farming communities in the central and southern coast region of Kenya. University of Bristol.

Horton M, Middleton J (2000) The Swahili: The social landscape of a mercantile society. Blackwell, Cambridge.

Hoskin CJ (2011) The invasion and potential impact of the Asian House Gecko (*Hemidactylus frenatus*) in Australia. *Austral Ecology* **36**, 240-251.

Jones EP, Jóhannesdóttir F *et al*. (2011) The expansion of the house mouse into north-western Europe. *J. Zool.* **283**, 257-268.

Larson G, Cucchi T *et al*. (2007) Phylogeny and ancient DNA of Sus provides insights into neolithic expansion in Island Southeast Asia and Oceania. *PNAS* **104** (12), 4834-4839.

Matisoo-Smith E & Robins J (2004) Origins and dispersals of Pacific peoples: Evidence from mtDNA phylogenies of the Pacific rat. *PNAS* **24**, 9167-9172.

Matisoo-Smith E & Robins J (2009) Mitochondrial DNA evidence for the spread of Pacific rats through Oceania. *Biol. Ivasions* **11**, 1521-1527.

Newbery B & Jones DN (2007) Presence of Asian House Gecko *Hemidactylus frenatus* across an urban gradient in Brisbane: influence of habitat and potential for impact on native gecko species. In: Lunney D *et al*. (eds.) Pest or Guest: the zoology of overabundance. Royal Zoological Society of New South Wales, Mosman (New South Wales, Australia).

Paijmans JLA, Gilbert MTP, Hofreiter M (2013) Mitogenomic analyses from ancient DNA. *Molecular Phylogenetics and Evolution* **69**, 404-416.

Plug I (2000) Overview of Iron Age fauna from the Limpopo Valley. *Goodwin Series* **8**, 117-126.

Rajabi-Maham H, Orth A, *et al*. (2012) The south-eastern house mouse *Mus musculus castaneus* (Rodentia: Muridae) is a polytypic subspecies

7 Appendix

7.1 Protocols

7.1.1 Primer design

Primer design is basis for applications in PCR and DNA sequencing. The chemically synthesised oligonucleotides serve not only as a starting point for the Taq-polymerase during DNA replication, but are designed to specifically amplify the required DNA fragments of the desired species in order to avoid amplification of contaminants. Depending on the age and type of samples in this thesis, primer pairs target overlapping fragments of an average size of ~ 100 to 550 bp.

Primer pairs were designed using Geneious R6 version 6.0.6 (Drummond *et al*. 2011) and Primer3 (Untergasser *et al*. 2012). The design is based on reference sequences of each particular species obtained from Genbank NCBI (http://www.ncbi.nlm.nih.gov).

Primer pairs: Mus musculus (chapter 2)

Primer pairs for the mouse project were designed by Heidi Eager. The design is based on the sequence of a whole mitochondrial genome of *Mus musculus domesticus* (NCBI accession number FJ374660).

Fragment Ref_FJ374660.1	Primer forward 5'-3'	Primer Position	Primer reverse 5'-3'	Primer Position	Product Length	°C
D-loop 1	GCACCCAAAGCTGGTATTCT	15387 15407	TTTTATGACCTGAACCATTGAYT	15510 15533	103 bp	57°C
D-loop 2	TATGTATATCGTACATTAAAYTAT	15457 15480	GAAGGGGATAGTCATATGGAAG	15650 15672	169 bp	57°C
D-loop 2a	CCAAGCATATAAGCAAGTACAT	15486 15508	CATATGGAAGAGAAGAGTTTATG	15637 15659	129 bp	57°C
D-loop 2b	ATATCTGTGTTATCTGACATAC	15605 15627	AGAAGAGGGGCATWGGTGG	15720 15739	93 bp	57°C
D-loop 3	TTATCTGACATACACCATACAG	15614 15636	TATGGGCGATAACGCATTTGAT	15812 15833	175 bp	57°C
D-loop 3a	TCTACCATCCTCCGTGAAAC	15689 15708	TATGGGCGATAACGCATTTGAT	15812 15833	103 bp	57°C
D-loop 4	CTTTATCAGACATCTGGTTCTT	15779 15800	GCGTCTAGACTGTGTGCTGT	15968 15987	167 bp	57°C
D-loop 4a	CCTTAAATAAGACATCTCGATG	15840 15862	GCGTCTAGACTGTGTGCTGT	15968 15987	106 bp	57°C
D-loop 5	CTTTCATCAACATAGCCGTCAA	15935 15956	TGTTTTTGGGGTTTGGCATTAA	16111 16132	154 bp	57°C

D-loop 5a	CACCTACGGTGAAGAATCATT	15987 16007	GGTTTGGCATTAAGAGGAGG	16104 16123	96 bp	57°C
D-loop 6	TATTCATGCTTGTTAGACATAAA	16039 16061	TGTTTTTGGGGTTTGGCATTAA	16111 16132	49 bp	57°C
D-loop 7	CTCAATACCAAATTTTAACTCTC	16067 16089	GTCATATTTTGGGAACTACTAG	16189 16210	99 bp	57°C
D-loop 8	CTATCAAACCCTATGTCCTGA	16162 16182	CTTGTTAATGTTTATTGCGTAAT	16279 16301	96 bp	57°C

Primer pairs: Rattus rattus (chapter 3)

The first set of primer pairs U1/L1 to U4/L4 has been designed by myself. The second set of primer pairs Cyt b Rr1 to Cyt b Rr10 and RatDloop 1 to 4 have been designed in equal parts by Heidi Eager and myself. Primer pairs F1 and F2 were taken from Aplin *et al*. (2011). The design is based on the sequence of a whole mitochondrial genome of *Rattus rattus* (NCBI accession number NC_012374).

Fragment Ref_NC_012374	Primer forward 5'-3'	Primer Position	Primer reverse 5'-3'	Primer Position	Product Length	°C
Cyt b U1/L1	AATTTGTCATTATTTCTACACAGCATT	14043 14069	TAGGGTTGCTTTGTCTACTGAGAA	14628 14652	559 bp	56°C
Cyt b U2/L2	CATCTGCCGAGACGTAAACTAC	14330 14352	GTCTCCTAGTAAGTCTGGGAAGAAT	14858 14883	507 bp	56°C
Cyt b U3/L3	AGGATCAAACAACCCCACAG	14735 14755	TGTTGATGGTGGGGAGTTAGT	15353 15374	599 bp	56°C
D-loop U4/L4	TCTCAGGACATGTCAAGAAGAAG	15328 15350	GGCATCCGAAAATTAAAAAATAC	15898 15920	548 bp	56°C
Aplin Museum F1 (2011)	ATCACACCCTCTACTCAAAA	14144 14164	GGCATGTAAGTATCGRATTAG (modified)	14358 14378	194 bp	56°C
Aplin Museum F2 (2011)	TCATCAGTTACYCACATCTGC (modified)	14316 14337	CCTCAGATTCATTCGACTAGRGT (modified)	14601 14624	264 bp	56°C

Fragment Ref_NC_012374	Primer forward 5'-3'	Primer Position	Primer reverse 5'-3'	Primer Position	Product Length	°C
Cyt b Rr1	ACACAGCATTTAACTGTGACCA	14060 14082	GGGCGGGAAGGTCAATGAAGG	14176 14197	94 bp	56°C
Cyt b Rr2	TTAATCACTCCTTCATTGACCTTCC	14167 14192	AGCCGTAGTTTACGTCTCGGCAG	14333 14356	141 bp	56°C
Cyt b Rr3	TTAACAGCATTCTCATCAGTTAC	14304 14327	GTTGCTATGACTGCAAATA	14485 14504	158 bp	56°C
Cyt b Rr4	TCCTACACCTTCTTAGAAACATGAAAC	14442 14469	AGCCTCCTCAGATTCATTCGAC	14607 14629	138 bp	56°C
Cyt b Rr5	CAAACCTATTATCAGCCATTCCCTA	14566 14590	AGTTTAGTCCTGTGGGGTTGTT	14742 14764	151 bp	56°C
Cyt b Rr6	GCCCTTGCAATTGTACATCTCCT	14697 14720	TGGGTCTCCTAGTAAGTCTGGGAA	14862 14886	142 bp	56°C
Cyt b Rr7	GACTTACTTGGAGTATTCATGTTAC	14808 14833	GGGATGGAGCGTAGAATAGCG	14960 14980	127 bp	56°C
Cyt b Rr8	ACCCCACCACATATTAAGCCAGA	14916 14939	TGGGCGGAATGTTAGACTGCGT	15062 15084	123 bp	56°C
Cyt b Rr9	TTCTAATCTTAGCCTTTCTACCA	15019 15042	AACTRATGGATGCTAGTTGG	15179 15199	137 bp	56°C
Cyt b Rr10	AGCCAACCTCTTCATTTTAAC	15113 15134	GCTCTTCATTTTTGGTTTACAA	15300 15322	166 bp	56°C
RatD-loop1	AACTACTTCTTGACAGTACATAA	15396 15419	TGGTGTATGTCTAATAACACAGA	15590 15614	172 bp	56°C
RatD-loop2	ACATGAATATTCTTTCATACATT	15546 15569	TTGTTGATTTCACGGAGGAT	15680 15700	111 bp	56°C

RatD-loop3	AGACATACACCATTAAAGTCATAA	15602 15626	GCCCTGAAGTAAGAACCAGA	15776 15796	150 bp	56°C
RatD-loop4	CCCATACAACTTGGGGGTGA	15733 15753	GGCATCCGAAAATTAAAAAATAC	15898 15920	145 bp	56°C

Primer pairs: Hemidactylus frenatus (chapter 4)

Primer pairs for the gecko project were designed by myself. The design is based on the sequence of a whole mitochondrial genome of *Hemidactylus frenatus* (NCBI accession number NC_012902).

Fragment Ref_NC_012902	Primer forward 5'-3'	Primer Position	Primer reverse 5'-3'	Primer Position	Product Length	°C
Cyt b	CTAATGATCCTCCGCAAAGC	14098 14118	AATCCGCCTCAAATTCACTG	14580 14600	463 bp	58°C

Primer pairs: Bubalus mindorensis (chapter 5)

Primer pairs for the tamaraw project were designed by myself. The design is based on the sequence of a whole mitochondrial genome of *Bubalus bubalis* (NCBI accession number AF547270).

Fragment Ref_AF547270	Primer forward 5'-3'	Primer Position	Primer reverse 5'-3'	Primer Position	Product Length	°C
1 Cyt b	GAAAAACCATCGTTGTCATT	14492 14512	GAGGAGAATGCTGTTGTTGT	14703 14723	191 bp	56°C
2 Cyt b	TGATGAAACTTTGGCTCTCT	14616 14636	TGCTGTGGCTATTACTGTGA	14890 14910	254 bp	56°C
3 Cyt b	CGTAGGACGAGGCATATACT	14819 14839	GGGTTGCTTTGTCTACTGAG	15032 15052	193 bp	56°C
4 Cyt b	ATCATTCTGAGGAGCAACAG	14942 14962	GAATGGGATTTTGTCTGTGT	15169 15189	207 bp	56°C
5 Cyt b	AGGATCCAACAACCCAACAGG	15137 15158	GGAACTGATCGTAAGATTGC	15363 15383	205 bp	56°C
6 Cyt b	CCATATCAAACCTGAATGGT	15326 15346	ATATATGGGTGTTCGACTGG	15552 15572	206 bp	56°C
7 Cyt b	CAAACCTGCTAACACTCACA	15517 15537	ATAGCCTCTTCCCTGAGTCT	15737 15757	200 bp	56°C

7.1.2 DNA extraction

DNA extraction was performed in dedicated ancient and modern DNA laboratories in the Archaeology Department at Durham University (chapters 3, 4, 5) and the Laboratory of Ornithology at Cornell University (chapter 2), following strict guidelines discussed in chapter 1.3.3.

The different sample types (bone, wet and dried tissue, skin) used in this thesis were prepared prior to DNA extraction in order to minimise the risk of co-extracting exogenous contaminants and to remove preservative chemicals that can inhibit subsequent PCR, respectively. Therefore the outer surface of bone samples was removed by abrasion, before a small piece of bone (~ 0.5 g) was cut out and successively pulverised in a Micro-dismembrator. The bone powder was then transferred into a 15 ml Eppendorf tube for further processing. Dried skin samples were washed several times with phosphate buffered saline (PBS), ethanol- or formalin-fixed specimens were washed several times with purified water in order to increase the quality of DNA recovery. Afterwards, samples were placed on a petri dish and finely chopped with a sterile, disposable blade before transferral of ~ 10 mg into a 1.5 ml Eppendorf tube. The different DNA extraction protocols for each sample type are stated below. One in ten extractions were blank controls in order to detect possible contamination.

Extraction protocol 1: Dried tissue (chapters 2, 3 and 5)

DNA extraction of dried tissue samples was carried out using the Qiagen MicroKit, following the manufacturer's recommendations.

Day 1

- transfer ~ 10 mg of dried tissue in a 1.5 ml tube
- add 180 µl of ATL buffer to each sample
- add 20 µl of Proteinase K to each sample
- vortex for 15 sec
- incubate samples on a rotary shaker over night at 56°C

Day 2

- take samples out of the oven and spin down shortly (~ 3000 rpm) to dry the rim
- add 200 µl AL buffer, vortex 15 sec

- add 200 µl ethanol, vortex 15 sec
- incubate 5 min at room temperature, then spin down shortly (~ 3000 rpm) to dry the rim
- transfer lysate in filter columns, without touching the membrane or wetting the rim
- centrifuge 1 min at 8000 rpm
- transfer columns in new 2 ml tubes, discard the flow-through
- add 500 µl AW1 buffer without wetting the rim
- centrifuge 1 min at 8000 rpm
- transfer columns in new 2 ml tubes, discard the flow-through
- add 500 µl AW2 buffer without wetting the rim
- centrifuge 1 min at 8000 rpm, discard the flow-through
- centrifuge 3 min at 14.000 rpm to dry the membrane
- place columns in new 1.5 ml tubes
- add 35 µl AE buffer, incubate 5 min
- centrifuge 1 min at 14.000 rpm
- repeat adding 50µl AE buffer, incubate 5 min
- centrifuge 1 min at 14.000 rpm
- repeat adding 50µl AE buffer, incubate 5min
- centrifuge 1 min at 14.000 rpm
- end up with three elutions E1 á 35µl, E2 á 50µl, E3 á 50µl, freeze at -20°C until further use

Extraction protocol 2: Ethanol-fixed tissue (chapters 2, 3 and 4)

DNA extraction of ethanol-fixed tissue samples was carried out using the following protocol.

Day 1

- add 300 µl Extraction buffer (1M NaCl, Tris-HCl pH 8.0, 10% SDS, H_2O) to each sample
- add 3 µl Proteinase K to each sample
- vortex 15 sec
- incubate samples on a rotary shaker over night at 50°C

Day 2

- add 80 µl of saturated NaCl to remove DNA byproducts, vortex
- centrifuge 10 min at 9.000 rpm
- transfer supernatant without touching the pellet, discard pellet
- repeat step until the supernatant is clear
- add 800 µl ethanol (97-100%) to precipitate the DNA pellet
- mix by inverting the tubes several times
- centrifuge 45 min at 13.000 rpm
- pour off the supernatant of each sample and remove any fluids with a small pipette, but mind the DNA pellet

- leave the tube open to dry the pellet, wait until all fluids are dissolved
- add 200 µl 1xTE, incubate 5 min
- centrifuge 1 min at 14.000 rpm
- end up with one elution E1 á 200 µl, freeze at -20°C until further use

Extraction protocol 3: Formalin-fixed tissue (chapter 4)

DNA extraction of formalin-fixed tissue samples was carried out using the Qiagen DNEasyKit, following the manufacturer's recommendations. A preheating step was added before pipetting the Proteinase K, in order to unblock DNA-protein-crosslinkages.

Day 1

- add 180 µl ATL buffer to each sample
- incubate samples at 95°C for 10 min to unblock DNA-protein-crosslinkages
- allow samples to cool, add 20 µl Proteinase K
- vortex 15 sec
- incubate samples on a rotary shaker over night at 56°C

Day 2

- take samples out of the oven and spin down shortly (~ 3000 rpm) to dry the rim
- add 200 µl AL buffer, vortex 15 sec
- add 200 µl ethanol, vortex 15 sec
- incubate 5 min at room temperature, then spin down shortly (~ 3000 rpm) to dry the rim
- transfer lysate in filter columns, without touching the membrane or wetting the rim
- centrifuge 1 min at 8000 rpm
- transfer columns in new 2 ml tubes, discard the flow-through
- add 500 µl AW1 buffer without wetting the rim
- centrifuge 1 min at 8000 rpm
- transfer columns in new 2 ml tubes, discard the flow-through
- add 500 µl AW2 buffer without wetting the rim
- centrifuge 1 min at 8000 rpm, discard the flow-through
- centrifuge 3 min at 14.000 rpm to dry the membrane
- place columns in new 1.5 ml tubes
- add 200 µl AE buffer, incubate 5 min
- centrifuge 1 min at 14.000 rpm
- end up with one elution E1 á 200 µl, freeze at -20°C until further use

Extraction protocol 4: Bone samples (chapter 5)

The concentrated extract of the bone samples was purified using the Qiagen QIAquick PCR Purification Kit, following the manufacturer's recommendations.

Day 1

- UV-irradiate the Extraction buffer for 45 min
- add 2 ml Extraction buffer (0.5M EDTA pH 8, 1M Tris pH 8, 1% SDS) to each sample
- add 20µl Proteinase K to each sample
- vortex 15 sec
- incubate samples on a rotary shaker over night at 50°C

Day 2

- UV-irradiate PB, PE and EB buffer for 45 min
- take samples out of the oven and spin bone powder down for 2 min at 2500 rpm
- transfer solution into centricons without touching the bone powder
- centrifuge until extract is concentrated to 100 µl
- take 100 µl of extract and mix it with 5x PB buffer and transfer the lysate directly in a filter tube
- centrifuge 1 min at 13.000 rpm, discard flow through
- add 750 µl of PE buffer
- centrifuge 1 min at 13.000 rpm, discard flow through
- centrifuge 3 min at 13.000 rpm to dry the membrane
- transfer filter in a new 2 ml tube
- add 50 µl EP buffer on filter, incubate for 1 min
- centrifuge 1 min at 13.000 rpm
- add 50 µl EP buffer on filter, incubate for 1 min
- centrifuge 1 min at 13.000 rpm
- end up with one elution E1 á 100 µl, freeze at -20°C until further use

7.1.3 DNA amplification

In the different projects discussed in this thesis two main mitochondrial loci were targeted for PCR amplification: hypervariable regions of the displacement-loop (d-loop) and cytochrome b (cyt b). PCR set-up was conducted under a fume hood in dedicated ancient respectively modern PCR set-up rooms. Every PCR set-up included a negative control in order to detect possible contamination. Additionally, a positive control (modern material of each particular species) was included in each PCR to exclude that possible failure of the reaction is due to reagents or the thermal cycler. The modern positive control was stored in the dedicated post-PCR room and added to the reaction before placing the tubes in the thermal cycler.

Standard protocols of PCR set-up and thermal cycler programs are stated below. To ensure optimal PCR success, modification of the reaction conditions was repeatedly needed. Modification included altering the amount of DNA extract added or adjusting the final concentration of the other reagents used. Furthermore, cycling conditions were changed by in- or decreasing the times and numbers of cycle repetitions. PCRs were visualised on a 1.5% agarose gel, using GelRed and UV-illumination. Successfully amplified PCR products were stored at -20°C prior to sequencing.

PCR protocol: Mus musculus (chapter 2)

PCR conditions at the laboratories at Cornell University (PCR set-up final volume: 25µl).

Reagent	Start Concentration	Final Concentration	Volume per Sample
H_2O	-	-	15.4 µl
Buffer	10 x	1 x	2.5 µl
$MgCl_2$	25 mM	4 mM	4 µl
Primer forward	10 mM	0.2 mM	0.5 µl
Primer reverse	10 mM	0.2 mM	0.5 µl
dNTP-Mix	10 mM	0.2 mM	0.5 µl
Sigma Jump Start Taq	2.5 U/µl	1.5 U	0.6 µl
DNA-Extract	-	-	1 µl

Programm	Temperature	Duration
Hot Start	94˚C	1.30 min
Denaturation	94˚C	45 sec
Annealing	57˚C	45 sec
Elongation	72˚C	1.30 min
Final Extension	72˚C	4 min
Cycles	-	40 times

PCR protocol: Rattus rattus (chapter 3)

PCR conditions at the laboratories at Durham University (PCR set-up final volume: 25µl).

Reagent	Start Concentration	Final Concentration	Volume per Sample
H_2O	-	-	15.8 µl
Buffer	10x	1 x	2.5 µl
$MgCl_2$	25 mM	4 mM	4 µl
Primer forward	10 mM	0.25 mM	0.625 µl
Primer reverse	10 mM	0.25 mM	0.625 µl
dNTP-Mix	25 mM	0.25 mM	0.25 µl
AmpliTaq Gold ™	5 U/µl	1 U	0.2 µl
DNA-Extract	-	-	1 µl

Programm	Temperature	Duration
Hot Start	95˚C	7 min
Denaturation	94˚C	30 sec
Annealing	56˚C	30 sec
Elongation	72˚C	50 sec
Final Extension	72˚C	4 min
Cycles	-	30-40 times

PCR protocol: Hemidactylus frenatus (chapter 4)

PCR conditions at the laboratories at Durham University (PCR set-up final volume: 25µl).

Reagent	Start Concentration	Final Concentration	Volume per Sample
H_2O	-	-	15.8 µl
Buffer	10x	1 x	2.5 µl
$MgCl_2$	25 mM	4 mM	4 µl
Primer forward	10 mM	0.25 mM	0.625 µl
Primer reverse	10 mM	0.25 mM	0.625 µl
dNTP-Mix	25 mM	0.25 mM	0.25 µl
Sigma Jump Start Taq	5 U/µl	1 U	0.2 µl
DNA-Extract	-	-	1 µl

Programm	Temperature	Duration
Hot Start	94˚C	7 min
Denaturation	94˚C	30 sec
Annealing	58˚-59˚C	30 sec
Elongation	72˚C	50 sec
Final Extension	72˚C	4 min
Cycles	-	30-40 times

PCR protocol: Bubalus mindorensis (chapter 5)

PCR conditions at the laboratories at Durham University (PCR set-up final volume: 25µl).

Reagent	Start Concentration	Final Concentration	Volume per Sample
H_2O	-	-	15.8 µl
Buffer	10x	1 x	2.5 µl
$MgCl_2$	25 mM	4 mM	4 µl
Primer forward	10 mM	0.25 mM	0.625 µl
Primer reverse	10 mM	0.25 mM	0.625 µl
dNTP-Mix	25 mM	0.25 mM	0.25 µl
Sigma Jump Start Taq	5 U/µl	1 U	0.2 µl
DNA-Extract	-	-	1 µl

Programm	Temperature	Duration
Hot Start	94˚C	7 min
Denaturation	94˚C	30 sec
Annealing	56˚C	30 sec
Elongation	72˚C	50 sec
Final Extension	72˚C	4 min
Cycles	-	30-40 times

7.1.4 DNA sequencing

Sanger-sequencing (chapters 2, 3, 4, 5)

Prior to sequencing, successfully amplified PCR products were purified by adding 0.05 µl Thermosensitive Alkaline Phosphatase (FastAP), 0.1 µl Exonuclease I and 2 µl Exonuclease I buffer, subsequently running the samples for 10 min at 37°C followed by 15 min at 65°C. In order to obtain the appropriate concentrations needed for sequencing, 180 µl of purified water was added to each sample and primer were diluted 50:50.

Sanger-sequencing of the chapters 3, 4 and 5 was performed by the DNA Sequencing Service at the School of Biological and Biomedical Sciences at Durham University, using an Applied Biosystems Automated 3730 DNA Analyzer. Generation of DNA sequences of chapter 2 was done by the Life Sciences Core Laboratories Center at Cornell University via an Applied Biosystems Automated 3730xl DNA Analyzer.

Analysis of the sequence data was carried out using Geneious R6 version 6.0.6 (Drummond *et al*. 2011). Trace files were edited manually and uncertainties marked according to the IUB-code. Afterwards, sequences of all primer systems per sample were built into contigs using the assembly function and a consensus sequence was exported.

Next-generation-sequencing (chapter 5)

Library preparation and high-throughput sequencing was carried out in the facilities at the Centre for GeoGenetics at the University of Copenhagen. The workshop, led by Dr. Hannes Schröder, focused on a shotgun approach using an Illumina Hi-Seq sequencing platform. Different steps within this technology include library preparation (end repair, adapter ligation, adapter fill-in reaction), library amplification (‘enrichment’), sample quantification by a Bioanalyzer (Agilent) and the subsequent sequencing reaction. Library preparation was carried out in the ancient DNA facilities of the Institute, following standard anti-contamination

guidelines. All following reactions were conducted in a separate modern DNA laboratory.

An initial validation of the raw data was done by Dr. Maria Avila at the Centre for GeoGenetics at the University of Copenhagen, the final analysis was carried out by Dr. Laurent Frantz (Wageningen University). The protocol (after Schröder) for library preparation and library amplification is stated hereinafter.

Library preparation

End repair (50 µl reaction set-up per sample)

- 42.5 µl DNA extract
- 5 µl 10x End Repair Reaction buffer
- 2.5 µl End Repair Enzyme Mix

Mix gently by pipetting and spin down briefly.
Incubate for 20 min at 12°C, followed by 15 min at 37°C.

Purification (Qiagen MinElute silica spin column)

- add 5x volume of PB buffer (~ 250 µl) to each sample
- mix and transfer into a column; centrifuge for 1 min at 13.000 rpm
- discard flow through
- add 720µl of PE buffer to the column
- centrifuge for 1 min at 13.000 rpm; discard flow through
- centrifuge for 1 min at 13.000 rpm to dry the membrane
- add 15 µl EP buffer; incubate for 10 min at 37°C
- centrifuge for 1 min at 13.000 rpm
- add 15 µl EP buffer; incubate for 10 min at 37°C
- centrifuge for 1 min at 13.000 rpm
- end up with an elution á 30 µl; transfer into 8-strip tubes

Adapter Ligation (50 µl reaction set-up per sample)

- 30 µl end-repaired DNA
- 10 µl 5x Quick Ligation buffer
- 5 µl Adapter Mix (2.5 µM)
- 5 µl Quick T4 DNA Ligase (5U/µl)

Mix gently by pipetting and spin down briefly.
Incubate for 15 min at 20°C.

Purification (Qiagen MinElute silica spin column)

- add 5x volume of PB buffer (~ 250 µl) to each sample
- mix and transfer into a column; centrifuge for 1 min at 13.000 rpm

- discard flow through
- add 720µl of PE buffer to the column
- centrifuge for 1 min at 13.000 rpm; discard flow through
- centrifuge for 1 min at 13.000 rpm to dry the membrane
- add 15 µl EP buffer; incubate for 10 min at 37°C
- centrifuge for 1 min at 13.000 rpm
- add 15 µl EP buffer; incubate for 10 min at 37°C
- centrifuge for 1 min at 13.000 rpm
- end up with an elution á 30 µl; transfer into 8-strip tubes

Adapter fill-in reaction (50 µl reaction set-up per sample)

- 30 µl adapter-ligated DNA
- 13 µl purified H_2O
- 5 µl Adapter fill-in buffer
- 2 µl Bst DNA Polymerase

Mix gently by pipetting and spin down briefly.
Incubate for 30 min at 37°C, followed by 20 min at 80°C to inactivate the enzyme.

Library amplification ('enrichment')

100 µl reaction set-up per sample:

- 50 µl DNA Library template
- 26.8 µl purified H_2O
- 10 µl 10x PCR buffer
- 10 µl Mg (50 mM)
- 4 µl BSA (20 mg/ml)
- 2 µl inPE 1.0 (10 µM) = forward primer
- 2 µl Indexing Primer (10 µM) = reverse primer; barcodes each sample individually, therefore varies from 'index 1' to 'index 50'
- 0.8 µl DNTP
- 0.4 µl Gold Taq

PCR cycling conditions were 94°C for 5 min, 15 cycles of 94°C for 30 sec, 60°C for 30 sec and 72°C for 30 sec, followed by a final extension of 72°C for 7 min. Visualisation of the PCR was carried out on a 1.5% agarose gel, using GelRed.

7.1.5 Sequence authentication

In order to ensure the generation of meaningful aDNA data, strict anti-contamination guidelines have been followed (see chapter 1.3.3). For chapters 2, 3, and 4, were Sanger-sequencing was used, species-specific primer pairs were designed for overlapping fragments to facilitate identification of miscoding lesions and exogenous DNA. Additionally, multiple negative controls were included in every work step to monitor for potential contamination. Independent replication of the sequence data was carried out to verify authenticity. Thereby, particular attention has been paid to the correctness of the results in chapters 2 and 3, as the ubiquity of rats and mice increase the risk of contamination (see chapter 1.3.3). Hence, samples from same locations have not been extracted at the same time and sequences have been compared and checked, whether they make phylogenetic sense. Most of the samples used in this thesis are modern or less than 100 years old, wherefore degradation is not as progressed. This is demonstrated by the great amplification success – particularly within the mice and rat samples – of fragments partially up to 500 bp.

In chapter 5, were DNA data was generated via a high-throughput sequencing platform, editing and verification of the sequences was carried out by Dr. Laurent Frantz (Wageningen University). Different authenticity criteria have already been discussed in chapter 1.3.5. Accordingly, the sequences were first sorted by their sample specific barcodes, before trimming the adapters using the programme FASTX clipper (FASTX toolkit 2009). This step is crucial, as NGS platforms are prone to sequencing parts of the adapter when generating short fragments of degraded DNA. Afterwards, reads were aligned to a mitochondrial reference genome of *Bubalus bubalis* (NCBI accession number AF547270), using SMALT (Postingel 2010). Thereby, only uniquely aligned reads were considered. Sequences with multiple hits to the reference were discarded and duplicates collapsed. To be able to call the genotype for the mitochondrial genome, Dr. Laurent Frantz removed the first and last 4 bp of every read, where most of the damaged DNA can be found, and then computed the effective coverage of each base (number of reads covering each base position). The effective coverage was weighted by the mapping quality (phred encoded probability that the read maps on the right position in the genome) and any

sites not representing at least 70% of the overall effective coverage, were filtered out. Ultimately, a FASTA file of the consensus genotype was created.

7.1.6 References

Aplin KP, Suzuki H *et al*. (2011) Multiple Geographic Origins of Commensalism and Complex Dispersal History of Black Rats. *PloS One* **6**, 26357.

Drummond AJ, Ashton B *et al*. (2011) Geneious R6 version 6.0.6, available from http://www.geneious.com

FASTX toolkit (2009) available from http://hannonlab.cshl.edu/fastx_toolkit/

Postingel H (2010) Wellcome Trust Sanger Institute, Cambridge, UK. SMALT version 0.7.4 available from http://www.gnu.org/licenses/

Untergasser A, Cutcutache I *et al*. (2012) Primer 3 – new capabilities and interfaces. *Nucleic Acids Res.* **40**, 115-132.

7.2 Sample lists

7.2.1 Chapter 2: *Mus musculus*

Sample Nb	Lab Code	Species	Provided by	Date	Location	Material	DNA
150156	150156	*M. musculus*	AMNH	1941	Australia	skin	
150157	150157	*M. musculus*	AMNH	1941	Australia	skin	
162639	162639	*M. m. domesticus*	AMNH	1948	Australia/Townsville	dried tissue	x
162640	162640	*M. musculus*	AMNH	1948	Australia	dried tissue	
162641	162641	*M. m. domesticus*	AMNH	1948	Australia/Angelalla	dried tissue	x
162643	162643	*M. musculus*	AMNH	1948	Australia/Warrego River	dried tissue	
197468	197468	*M. musculus*	AMNH	1964	Australia/Gidley Brook	dried tissue	
32292	32292	*M. m. musculus*	AMNH	1910	China/Fengsiangfu/Shensi	dried tissue	x
32293	32293	*M. musculus*	AMNH	1910	China/Fengsiangfu/Shensi	dried tissue	
32295	32295	*M. m. musculus*	AMNH	1910	China/Fengsiangfu/Shensi	dried tissue	x
43579	43579	*M. m. castaneus*	AMNH	1916	China/Lijang	dried tissue	x
47955	47955	*M. musculus*	AMNH	1920	China	dried tissue	
47959	47959	*M. musculus*	AMNH	1920	China	dried tissue	
47961	47961	*M. musculus*	AMNH	1920	China	dried tissue	
47962	47962	*M. musculus*	AMNH	1920	China	dried tissue	
47963	47963	*M. musculus*	AMNH	1920	China	dried tissue	
55797	55797	*M. m. castaneus*	AMNH	?	China	dried tissue	x
56400	56400	*M. musculus*	AMNH	1920	China/Yenping/Fukien Prov.	dried tissue	
56416	56416	*M. musculus*	AMNH	1921	China	dried tissue	
84303	84303	*M. m. musculus*	AMNH	1925	China/Choni	dried tissue	x
84975	84975	*M. m. castaneus*	AMNH	1926	China/Yunnan	dried tissue	x
176246	176246	*M. musculus*	AMNH	1944	USA/Connecticut	dried tissue	
26857	26857	*M. m. castaneus*	AMNH	1905	India/Assam	dried tissue	x
163164	163164	*M. musculus*	AMNH	1947	India/Gokarna	dried tissue	x
171161	171161	*M. m. castaneus*	AMNH	1949	India/Umran	dried tissue	x
215634	215634	*M. musculus*	AMNH	1965	India/Howrah	dried tissue	
215638	215638	*M. musculus*	AMNH	1965	India/Howrah	dried tissue	
54858	54858	*M. musculus*	AMNH	?	Indonesia	dried tissue	
101268	101268	*M. m. castaneus*	AMNH	1931	Indonesia/Sulawesi/Roeroekan	dried tissue	x
101269	101269	*M. m. castaneus*	AMNH	1931	Indonesia/Sulawesi/Roeroekan	dried tissue	x
101528	101528	*M. musculus*	AMNH	1931	Indonesia/Cheribon	dried tissue	
102693	102693	*M. m. castaneus*	AMNH	1934	Indonesia/Cheribon	dried tissue	x
106661	106661	*M. m. castaneus*	AMNH	1936	Indonesia/Cheribon	dried tissue	x
106698	106698	*M. m. domesticus*	AMNH	1936	Indonesia/Cheribon	dried tissue	x
109164	109164	*M. m. castaneus*	AMNH	1938	Indonesia/Sulawesi/Peleng	dried tissue	x
109166	109166	*M. musculus*	AMNH	1938	Indonesia	dried tissue	
109169	109169	*M. m. castaneus*	AMNH	1938	Indonesia/Sulawesi/Peleng	dried tissue	x
109172	109172	*M. m. castaneus*	AMNH	1938	Indonesia/Sulawesi/Peleng	dried tissue	x
109173	109173	*M. m. castaneus*	AMNH	1938	Indonesia/Sulawesi/Peleng	dried tissue	x
109174	109174	*M. m. castaneus*	AMNH	1938	Indonesia/Sulawesi/Peleng	dried tissue	x
109175	109175	*M. musculus*	AMNH	1938	Indonesia	dried tissue	
153041	153041	*M. m. castaneus*	AMNH	1939	Indonesia/Bumbulan	dried tissue	x
153042	153042	*M. m. castaneus*	AMNH	1939	Indonesia/Bumbulan	dried tissue	x
242307	242307	*M. m. castaneus*	AMNH	1973	Indonesia/Pewunu	dried tissue	x
267709	267709	*M. m. castaneus*	AMNH	1993	Indonesia/Wamena	skin	x
267710	267710	*M. m. castaneus*	AMNH	1993	Indonesia/Wamena	skin	x
84095	84095	*M. m. musculus*	AMNH	1925	Mongolia/Tsang	dried tissue	x
240750	240750	*M. musculus*	AMNH	1972	Nepal	dried tissue	
240751	240751	*M. musculus*	AMNH	1972	Nepal/Lapache	skin	
240752	240752	*M. musculus*	AMNH	1972	Nepal/Lapache	skin	
240753	240753	*M. musculus*	AMNH	?	Nepal	skin	
250011	250011	*M. musculus*	AMNH	1973	Nepal	skin	
250014	250014	*M. musculus*	AMNH	1974	Nepal	skin	
250016	250016	*M. musculus*	AMNH	1974	Nepal	skin	
242213	242213	*M. musculus*	AMNH	?	Philippines	dried tissue	
242217	242217	*M. musculus*	AMNH	?	Philippines/Loa Banos	dried tissue	
242223	242223	*M. m. musculus*	AMNH	?	Philippines/Los Banos	dried tissue	x
242224	242224	*M. musculus*	AMNH	?	Philippines/Los Banos	dried tissue	

242232	242232	*M. m. domesticus*	AMNH	?	Philippines	dried tissue	**x**
20813	20813	*M. musculus*	AMNH	1901	Sri Lanka	dried tissue	
240959	240959	*M. m. castaneus*	AMNH	1971	Sri Lanka/Nanuoya	dried tissue	**x**
242697	242697	*M. m. castaneus*	AMNH	?	Sri Lanka/Boragas	skin	**x**
242702	242702	*M. m. castaneus*	AMNH	1970	Sri Lanka/Talawakele	skin	**x**
242703	242703	*M. m. castaneus*	AMNH	?	Sri Lanka/Talawakele	skin	**x**
165784	165784	*M. m. castaneus*	AMNH	1954	Taiwan	dried tissue	**x**
171127	171127	*M. musculus*	AMNH	1947	India	skin	
208173	208173	*M. m. castaneus*	AMNH	1961	India	skin	**x**
208174	208174	*M. musculus*	AMNH	1962	India	skin	
208175	208175	*M. m. castaneus*	AMNH	1962	India	skin	**x**
170225	170225	*M. musculus*	AMNH	1950	Pakistan	skin	
170247	170247	*M. musculus*	AMNH	1950	Pakistan	skin	
170248	170248	*M. musculus*	AMNH	1950	Pakistan	skin	
170249	170249	*M. musculus*	AMNH	1950	Pakistan	skin	
170273	170273	*M. musculus*	AMNH	1950	Pakistan	skin	
242326	242326	*M. musculus*	AMNH	1972	Pakistan	dried tissue	
28614	28614	*M. m. castaneus*	AMNH	1906	Kashmir	dried tissue	**x**
28623	28623	*M. m. castaneus*	AMNH	1908	Kashmir	dried tissue	**x**
28641	28641	*M. m. castaneus*	AMNH	1908	Kashmir	dried tissue	**x**
100728	100728	*M. m. gentilulus*	AMNH	1931	Madagascar/Tabiky	dried tissue	**x**
120307	120307	*M. musculus*	FMNH	1976	Western Australia	dried tissue	
120314	120314	*M. musculus*	FMNH	1976	Western Australia	dried tissue	
120315	120315	*M. musculus*	FMNH	1976	Western Australia	dried tissue	
120317	120317	*M. musculus*	FMNH	1976	Western Australia	dried tissue	
120320	120320	*M. musculus*	FMNH	1976	Western Australia	dried tissue	
120323	120323	*M. musculus*	FMNH	1976	Western Australia	dried tissue	
33157	33157	*M. m. castaneus*	FMNH	1929	China/Yunnan	dried tissue	**x**
33159	33159	*M. musculus*	FMNH	1929	China/Yunnan	dried tissue	
33160	33160	*M. musculus*	FMNH	1929	China/Yunnan	dried tissue	
40594	40594	*M. m. castaneus*	FMNH	1932	China/Szechuan	dried tissue	**x**
40645	40645	*M. musculus*	FMNH	1932	China/Szechuan	dried tissue	
34806	34806	*M. m. castaneus*	FMNH	1930	India/Haldibari	dried tissue	**x**
35554	35554	*M. musculus*	FMNH	1931	India/Bengal	dried tissue	
35556	35556	*M. musculus*	FMNH	1931	India/Bengal	dried tissue	
35565	35565	*M. m. castaneus*	FMNH	1931	India/Bengal	dried tissue	**x**
35571	35571	*M. m. castaneus*	FMNH	1931	India/Sikkim	dried tissue	**x**
35574	35574	*M. m. castaneus*	FMNH	1931	India/Sikkim	dried tissue	**x**
35577	35577	*M. musculus*	FMNH	1931	India/Sikkim	dried tissue	**x**
35578	35578	*M. musculus*	FMNH	1931	India/Sikkim	dried tissue	**x**
142196	142196	*M. musculus*	FMNH	1968	Nepal/Khumjung	dried tissue	**x**
142198	142198	*M. musculus*	FMNH	1968	Nepal/Khumjung	dried tissue	**x**
142200	142200	*M. musculus*	FMNH	1968	Nepal/Lukla	dried tissue	**x**
55232	55232	*M. musculus*	FMNH	1945	New Caledonia/Noumea	dried tissue	
55233	55233	*M. m. domesticus*	FMNH	1945	New Caledonia/Noumea	dried tissue	**x**
55235	55235	*M. musculus*	FMNH	1945	New Caledonia/Noumea	dried tissue	
56259	56259	*M. musculus*	FMNH	1946	Philippines/Mindanao	dried tissue	
99402	99402	*M. m. castaneus*	FMNH	1965	Sri Lanka/Dehiwala	dried tissue	**x**
99403	99403	*M. m. castaneus*	FMNH	1965	Sri Lanka/Dehiwala	dried tissue	**x**
32426	32426	*M. m. castaneus*	FMNH	1929	Vietnam/Tonkin	dried tissue	**x**
65424	65424	*M. m. castaneus*	FMNH	1947	India/Mussoorie	dried tissue	**x**
65425	65425	*M. m. castaneus*	FMNH	1947	India/Mussoorie	dried tissue	**x**
65426	65426	*M. m. castaneus*	FMNH	1947	India/Mussoorie	dried tissue	**x**
28953	28953	*M. musculus*	FMNH	1900	India/Kashmir	dried tissue	
140449	140449	*M. musculus*	FMNH	1990	Pakistan/Baluchistan	dried tissue	
140450	140450	*M. musculus*	FMNH	1990	Pakistan/Baluchistan	dried tissue	
140452	140452	*M. musculus*	FMNH	1990	Pakistan/Baluchistan	dried tissue	
140455	140455	*M. musculus*	FMNH	1990	Pakistan/Malakand	dried tissue	
140459	140459	*M. m. castaneus*	FMNH	1990	Pakistan/Karakar	dried tissue	**x**
140461	140461	*M. musculus*	FMNH	1990	Pakistan/Malakand	dried tissue	
140464	140464	*M. m. castaneus*	FMNH	1990	Pakistan/Yakh Tangai Swat	dried tissue	**x**
256649	256649	*M. m. castaneus*	FMNH	1984	UAE	dried tissue	**x**
88788	88788	*M. m. castaneus*	FMNH	1938	Iran/Astarabad	dried tissue	**x**
88912	88912	*M. m. castaneus*	FMNH	1938	Iran/Astarabad	dried tissue	**x**

112270	112270	*M. m. castaneus*	FMNH	1968	Iran/Isfahan	dried tissue	**x**
112280	112280	*M. m. castaneus*	FMNH	1968	Iran/Fars	dried tissue	**x**
149873	149873	*M. m. castaneus*	FMNH	1990	Uganda	dried tissue	**x**
161770	161770	*M. m. gentilulus*	FMNH	1998	Madagascar	dried tissue	**x**
1	1	*M. m. domesticus*	Frankfurt University/K. Koch	modern	Hawaii/Kaho'olawe	tissue - Ethanol	**x**
2	2	*M. m. domesticus*	Frankfurt University/K. Koch	modern	Hawaii/Kaho'olawe	tissue - Ethanol	**x**
3	3	*M. m. domesticus*	Frankfurt University/K. Koch	modern	Hawaii/Kaho'olawe	tissue - Ethanol	**x**
4	4	*M. m. domesticus*	Frankfurt University/K. Koch	modern	Hawaii/Kaho'olawe	tissue - Ethanol	**x**
54.1025	54.1025	*M. musculus*	NHM	?	Socotra	dried tissue	
57.417	57.417	*M. musculus*	NHM	?	Maldives	dried tissue	
38316	38316	*M. m. domesticus*	Smithsonian	?	Kauai	dried tissue	**x**
154878	154878	*M. m. domesticus*	Smithsonian	?	Java	dried tissue	**x**
173932	173932	*M. musculus*	Smithsonian	?	India	dried tissue	
201188	201188	*M. m. castaneus*	Smithsonian	?	India	dried tissue	**x**
277484	277484	*M. m. castaneus*	Smithsonian	?	Guam	dried tissue	**x**
277486	277486	*M. m. castaneus*	Smithsonian	?	Guam	dried tissue	**x**
279125	279125	*M. m. castaneus*	Smithsonian	?	Saipan Island	dried tissue	**x**
326614	326614	*M. m. castaneus*	Smithsonian	?	Pakistan	dried tissue	**x**
326620	326620	*M. m. castaneus*	Smithsonian	?	Pakistan	dried tissue	**x**
362205	362205	*M. m. domesticus*	Smithsonian	?	Baker Island	dried tissue	**x**
496973	496973	*M. m. castaneus*	Smithsonian	?	Java	dried tissue	**x**
538877	538877	*M. m. domesticus*	Smithsonian	?	North Pacific Ocean	dried tissue	**x**
562726	562726	*M. m. castaneus*	Smithsonian	?	Tinian Island	dried tissue	**x**
552565	552565	*M. musculus*	Smithsonian	?	Moheli Island	dried tissue	
MP-17	MP-17	*M. m. castaneus*	Michel Pascal	modern	Reunion	tissue - Ethanol	**x**
MP-19	MP-19	*M. m. castaneus*	Michel Pascal	modern	Reunion	tissue - Ethanol	**x**
MP-20	MP-20	*M. m. castaneus*	Michel Pascal	modern	Reunion	tissue - Ethanol	**x**
MP-22	MP-22	*M. m. castaneus*	Michel Pascal	modern	Reunion	tissue - Ethanol	**x**
MP-23	MP-23	*M. m. castaneus*	Michel Pascal	modern	Reunion	tissue - Ethanol	**x**
MP-24	MP-24	*M. m. castaneus*	Michel Pascal	modern	Reunion	tissue - Ethanol	**x**
MP-27	MP-27	*M. m. castaneus*	Michel Pascal	modern	Reunion	tissue - Ethanol	**x**
MP-28	MP-28	*M. m. castaneus*	Michel Pascal	modern	Reunion	tissue - Ethanol	**x**
MP-29	MP-29	*M. m. castaneus*	Michel Pascal	modern	Reunion	tissue - Ethanol	**x**
MP-33	MP-33	*M. m. castaneus*	Michel Pascal	modern	Reunion	tissue - Ethanol	**x**
T49	T49	*M. m. gentilulus*	Jean-Marc Duplantier	modern	Madagascar	tissue - Ethanol	**x**

NCBI Accession Nb	Species	Location	Author
AF074540	*M. m. gentilulus*	Yemen	Prager *et al*. 1998
AF074541	*M. m. gentilulus*	Yemen	Prager *et al*. 1998
AF074542	*M. m. gentilulus*	Yemen	Prager *et al*. 1998
AF074543	*M. m. gentilulus*	Yemen	Prager *et al*. 1998
AF074544	*M. m. gentilulus*	Yemen	Prager *et al*. 1998
AF074545	*M. m. gentilulus*	Yemen	Prager *et al*. 1998
GU384333	*M. m. domesticus*	Argentina	Jones *et al*. 2011
GU384334	*M. m. domesticus*	Argentina	Jones *et al*. 2011
GU384335	*M. m. domesticus*	Argentina	Jones *et al*. 2011
GU384336	*M. m. domesticus*	Canary Islands	Jones *et al*. 2011
GU384337	*M. m. domesticus*	Canary Islands	Jones *et al*. 2011
GU384338	*M. m. domesticus*	Canary Islands	Jones *et al*. 2011
GU384339	*M. m. domesticus*	Cyprus	Jones *et al*. 2011
GU384340	*M. m. domesticus*	Cyprus	Jones *et al*. 2011
GU384341	*M. m. domesticus*	Italy/Calabria	Jones *et al*. 2011
GU384342	*M. m. domesticus*	Italy/Calabria	Jones *et al*. 2011
GU384343	*M. m. domesticus*	Luxembourg	Jones *et al*. 2011
GU384344	*M. m. domesticus*	Niger	Jones *et al*. 2011
GU384345	*M. m. domesticus*	Spain/Barcelona	Jones *et al*. 2011
GU384346	*M. m. domesticus*	Spain/Barcelona	Jones *et al*. 2011
GU384347	*M. m. domesticus*	Spain/Barcelona	Jones *et al*. 2011
GU384348	*M. m. domesticus*	Spain/Barcelona	Jones *et al*. 2011
GU384349	*M. m. domesticus*	Spain/Barcelona	Jones *et al*. 2011
GU384350	*M. m. domesticus*	Spain/Barcelona	Jones *et al*. 2011
GU384351	*M. m. domesticus*	Spain/Barcelona	Jones *et al*. 2011
GU384352	*M. m. domesticus*	Spain/Barcelona	Jones *et al*. 2011

GU384353	*M. m. domesticus*	Spain/Barcelona	Jones *et al*. 2011
GU384354	*M. m. domesticus*	Spain/Barcelona	Jones *et al*. 2011
GU384355	*M. m. domesticus*	Spain/Barcelona	Jones *et al*. 2011
GU384356	*M. m. domesticus*	Spain/Barcelona	Jones *et al*. 2011
GU384357	*M. m. domesticus*	Spain/Barcelona	Jones *et al*. 2011
GU384358	*M. m. domesticus*	Spain/Barcelona	Jones *et al*. 2011
GU384359	*M. m. domesticus*	Spain/Barcelona	Jones *et al*. 2011
GU384360	*M. m. domesticus*	Spain/Barcelona	Jones *et al*. 2011
JF277281	*M. m. domesticus*	Australia	Gabriel *et al*. 2011
JF277282	*M. m. domesticus*	Australia	Gabriel *et al*. 2011
JF277283	*M. m. domesticus*	Australia	Gabriel *et al*. 2011
JF277284	*M. m. domesticus*	Australia	Gabriel *et al*. 2011
JF277285	*M. m. domesticus*	Australia	Gabriel *et al*. 2011
JF277286	*M. m. domesticus*	Australia	Gabriel *et al*. 2011
JF277287	*M. m. domesticus*	Australia	Gabriel *et al*. 2011
JF277288	*M. m. domesticus*	Australia	Gabriel *et al*. 2011
JF277289	*M. m. domesticus*	Australia	Gabriel *et al*. 2011
JF277290	*M. m. domesticus*	Australia	Gabriel *et al*. 2011
JF277291	*M. m. domesticus*	Australia	Gabriel *et al*. 2011
JF277292	*M. m. domesticus*	Australia	Gabriel *et al*. 2011
JF277293	*M. m. domesticus*	Australia	Gabriel *et al*. 2011
JF277294	*M. m. domesticus*	Netherlands	Gabriel *et al*. 2011
JF277297	*M. m. domesticus*	Netherlands	Gabriel *et al*. 2011
JF277298	*M. m. domesticus*	Netherlands	Gabriel *et al*. 2011
JF277299	*M. m. domesticus*	Netherlands	Gabriel *et al*. 2011
JF277300	*M. m. domesticus*	Netherlands	Gabriel *et al*. 2011
JQ692972	*M. m. domesticus*	Iceland	Jones *et al*. 2012
JQ692973	*M. m. domesticus*	Iceland	Jones *et al*. 2012
JQ692974	*M. m. domesticus*	Iceland	Jones *et al*. 2012
JQ692975	*M. m. domesticus*	Iceland	Jones *et al*. 2012
JQ692976	*M. m. domesticus*	Canada/Newfoundland	Jones *et al*. 2012
JQ692977	*M. m. domesticus*	Canada/Newfoundland	Jones *et al*. 2012
JQ692978	*M. m. domesticus*	Greenland	Jones *et al*. 2012
MMU47430	*M. m. domesticus*	UK, Germany	Prager *et al*. 1993
MMU47431	*M. m. domesticus*	UK, Germany, Denmark	Prager *et al*. 1993
MMU47432	*M. m. domesticus*	Germany	Prager *et al*. 1993
MMU47433	*M. m. domesticus*	Germany	Prager *et al*. 1993
MMU47434	*M. m. domesticus*	Scotland	Prager *et al*. 1993
MMU47435	*M. m. domesticus*	Greece	Prager *et al*. 1993
MMU47436	*M. m. domesticus*	Scotland	Prager *et al*. 1993
MMU47437	*M. m. domesticus*	Scotland	Prager *et al*. 1993
MMU47438	*M. m. domesticus*	Scotland	Prager *et al*. 1993
MMU47439	*M. m. domesticus*	Croatia	Prager *et al*. 1993
MMU47440	*M. m. domesticus*	Israel	Prager *et al*. 1993
MMU47441	*M. m. domesticus*	Germany	Prager *et al*. 1993
MMU47442	*M. m. domesticus*	Germany	Prager *et al*. 1993
MMU47443	*M. m. domesticus*	Germany	Prager *et al*. 1993
MMU47444	*M. m. domesticus*	Germany	Prager *et al*. 1993
MMU47445	*M. m. domesticus*	Switzerland	Prager *et al*. 1993
MMU47446	*M. m. domesticus*	Egypt	Prager *et al*. 1993
MMU47447	*M. m. domesticus*	Switzerland	Prager *et al*. 1993
MMU47448	*M. m. domesticus*	USA/Indiana	Prager *et al*. 1993
MMU47449	*M. m. domesticus*	USA/California	Prager *et al*. 1993
MMU47450	*M. m. domesticus*	Egypt	Prager *et al*. 1993
MMU47451	*M. m. domesticus*	Egypt	Prager *et al*. 1993
MMU47452	*M. m. domesticus*	Egypt	Prager *et al*. 1993
MMU47453	*M. m. domesticus*	Egypt	Prager *et al*. 1993
MMU47454	*M. m. domesticus*	Morocco	Prager *et al*. 1993
MMU47455	*M. m. domesticus*	Denmark, Germany	Prager *et al*. 1993
MMU47456	*M. m. domesticus*	Sweden, Finland	Prager *et al*. 1993
MMU47457	*M. m. domesticus*	Germany	Prager *et al*. 1993
MMU47458	*M. m. domesticus*	Denmark	Prager *et al*. 1993
MMU47459	*M. m. domesticus*	Denmark	Prager *et al*. 1993
MMU47460	*M. m. domesticus*	Denmark	Prager *et al*. 1993
MMU47461	*M. m. domesticus*	Denmark	Prager *et al*. 1993

MMU47462	*M. m. domesticus*	Sweden	Prager *et al*. 1993
MMU47463	*M. m. domesticus*	Denmark	Prager *et al*. 1993
MMU47464	*M. m. domesticus*	Sweden, Germany	Prager *et al*. 1993
MMU47465	*M. m. domesticus*	Switzerland, Germany	Prager *et al*. 1993
MMU47466	*M. m. domesticus*	Germany	Prager *et al*. 1993
MMU47467	*M. m. domesticus*	Germany	Prager *et al*. 1993
MMU47468	*M. m. domesticus*	Germany	Prager *et al*. 1993
MMU47469	*M. m. domesticus*	Germany	Prager *et al*. 1993
MMU47470	*M. m. domesticus*	Germany	Prager *et al*. 1993
MMU47471	*M. m. domesticus*	Italy	Prager *et al*. 1993
MMU47472	*M. m. domesticus*	Spain/Mallorca	Prager *et al*. 1993
MMU47473	*M. m. domesticus*	Germany	Prager *et al*. 1993
MMU47474	*M. m. domesticus*	Germany	Prager *et al*. 1993
MMU47475	*M. m. domesticus*	Germany	Prager *et al*. 1993
MMU47476	*M. m. domesticus*	Germany	Prager *et al*. 1993
MMU47477	*M. m. domesticus*	Italy	Prager *et al*. 1993
MMU47478	*M. m. domesticus*	Morocco	Prager *et al*. 1993
MMU47479	*M. m. domesticus*	Spain/Barcelona	Prager *et al*. 1993
MMU47480	*M. m. domesticus*	Israel/Jerusalem	Prager *et al*. 1993
MMU47481	*M. m. domesticus*	Portugal/Lisbon	Prager *et al*. 1993
MMU47482	*M. m. domesticus*	Italy/Milan	Prager *et al*. 1993
MMU47483	*M. m. domesticus*	Croatia	Prager *et al*. 1993
MMU47484	*M. m. domesticus*	Peru	Prager *et al*. 1993
MMU47485	*M. m. domesticus*	Germany	Prager *et al*. 1993
MMU47486	*M. m. domesticus*	Germany	Prager *et al*. 1993
MMU47487	*M. m. domesticus*	Austria	Prager *et al*. 1993
MMU47488	*M. m. domesticus*	Germany	Prager *et al*. 1993
MMU47489	*M. m. domesticus*	Germany	Prager *et al*. 1993
MMU47490	*M. m. domesticus*	Germany	Prager *et al*. 1993
MMU47491	*M. m. domesticus*	Germany	Prager *et al*. 1993
MMU47492	*M. m. domesticus*	Norway	Prager *et al*. 1993
MMU47493	*M. m. domesticus*	Norway	Prager *et al*. 1993
MMU47494	*M. m. domesticus*	Norway	Prager *et al*. 1993
MMU47495	*M. m. domesticus*	Croatia	Prager *et al*. 1993
MMU47496	*M. m. domesticus*	Georgia	Prager *et al*. 1993
MMU47497	*M. m. domesticus*	Georgia	Prager *et al*. 1993
MMU47498	*M. m. musculus*	Czech Republic	Prager *et al*. 1993
MMU47499	*M. m. musculus*	Germany/Bavaria	Prager *et al*. 1993
MMU47500	*M. m. musculus*	Slovakia, Germany	Prager *et al*. 1993
MMU47501	*M. m. musculus*	Austria	Prager *et al*. 1993
MMU47502	*M. m. musculus*	Germany	Prager *et al*. 1993
MMU47503	*M. m. musculus*	Germany/Bavaria	Prager *et al*. 1993
MMU47504	*M. m. musculus*	Croatia, Germany, Austria	Prager *et al*. 1993
MMU47505	*M. m. musculus*	Austria	Prager *et al*. 1993
MMU47506	*M. m. musculus*	Austria, Germany	Prager *et al*. 1993
MMU47507	*M. m. musculus*	Germany/Bavaria	Prager *et al*. 1993
MMU47508	*M. m. musculus*	Germany/Bavaria	Prager *et al*. 1993
MMU47509	*M. m. musculus*	Germany/Bavaria	Prager *et al*. 1993
MMU47510	*M. m. musculus*	Germany/Bavaria	Prager *et al*. 1993
MMU47511	*M. m. musculus*	Germany/Bavaria	Prager *et al*. 1993
MMU47512	*M. m. musculus*	Austria	Prager *et al*. 1993
MMU47513	*M. m. musculus*	Austria	Prager *et al*. 1993
MMU47514	*M. m. musculus*	Germany/Bavaria	Prager *et al*. 1993
MMU47515	*M. m. musculus*	Germany/Bavaria	Prager *et al*. 1993
MMU47516	*M. m. musculus*	Germany/Bavaria	Prager *et al*. 1993
MMU47517	*M. m. musculus*	Czech Republic	Prager *et al*. 1993
MMU47518	*M. m. musculus*	Turkmenistan	Prager *et al*. 1993
MMU47519	*M. m. musculus*	Ukraine/Crimea	Prager *et al*. 1993
MMU47520	*M. m. musculus*	Russia, Turkmenistan	Prager *et al*. 1993
MMU47521	*M. m. musculus*	Japan/Okinawa	Prager *et al*. 1993
MMU47522	*M. m. musculus*	Turkmenistan	Prager *et al*. 1993
MMU47523	*M. m. musculus*	Poland/Warsaw	Prager *et al*. 1993
MMU47524	*M. m. musculus*	Poland/Warsaw	Prager *et al*. 1993
MMU47525	*M. m. musculus*	Czech Republic, Poland	Prager *et al*. 1993
MMU47526	*M. m. musculus*	Turkmenistan	Prager *et al*. 1993

MMU47527	*M. m. musculus*	Siberia/Altai	Prager *et al*. 1993
MMU47528	*M. m. musculus*	Ukraine/Crimea	Prager *et al*. 1993
MMU47529	*M. m. musculus*	Serbia/Belgrade	Prager *et al*. 1993
MMU47530	*M. m. musculus*	Austria	Prager *et al*. 1993
MMU47531	*M. m. musculus*	Austria	Prager *et al*. 1993
MMU47532	*M. m. musculus*	Georgia	Prager *et al*. 1993
MMU47533	*M. m. musculus*	Moldova	Prager *et al*. 1993
M77119	*M. spretus*	?	She *et al*., unpublished
JN416649	*M. m. castaneus*	Afghanistan/Kaboul	Rajabi-Maham *et al*. 2012
JN416650	*M. m. castaneus*	Afghanistan/Kaboul	Rajabi-Maham *et al*. 2012
JN416651	*M. m. castaneus*	India/Masinagudi	Rajabi-Maham *et al*. 2012
JN416652	*M. m. castaneus*	India/Masinagudi	Rajabi-Maham *et al*. 2012
JN416653	*M. m. castaneus*	India/Bikaner	Rajabi-Maham *et al*. 2012
JN416654	*M. m. castaneus*	India/Bikaner	Rajabi-Maham *et al*. 2012
JN416655	*M. m. castaneus*	Iran/Ahvaz	Rajabi-Maham *et al*. 2012
JN416656	*M. m. castaneus*	Kenya/Mombasa	Rajabi-Maham *et al*. 2012
JN416657	*M. m. castaneus*	India/Punjab	Rajabi-Maham *et al*. 2012
JN416658	*M. m. castaneus*	India/Punjab	Rajabi-Maham *et al*. 2012
JN416659	*M. m. castaneus*	India/Kotagiri	Rajabi-Maham *et al*. 2012
JN416660	*M. m. castaneus*	India/Delhi	Rajabi-Maham *et al*. 2012
JN416661	*M. m. castaneus*	India/Delhi	Rajabi-Maham *et al*. 2012
JN416662	*M. m. castaneus*	India/Delhi	Rajabi-Maham *et al*. 2012
JN416663	*M. m. castaneus*	India/Delhi	Rajabi-Maham *et al*. 2012
JN416664	*M. m. castaneus*	Pakistan/Islamabad	Rajabi-Maham *et al*. 2012
JN416665	*M. m. castaneus*	India/Delhi	Rajabi-Maham *et al*. 2012
JN416666	*M. m. castaneus*	Iran/Teheran	Rajabi-Maham *et al*. 2012
JN416667	*M. m. castaneus*	Iran/Teheran	Rajabi-Maham *et al*. 2012
JN416668	*M. m. castaneus*	Iran/Teheran	Rajabi-Maham *et al*. 2012
JN416669	*M. m. castaneus*	Iran/Mashhad	Rajabi-Maham *et al*. 2012
JN416670	*M. m. castaneus*	Iran/Teheran	Rajabi-Maham *et al*. 2012
JN416671	*M. m. castaneus*	Iran/Teheran	Rajabi-Maham *et al*. 2012
JN416672	*M. m. castaneus*	Iran/Zanjan	Rajabi-Maham *et al*. 2012
JN416673	*M. m. castaneus*	Iran/Zanjan	Rajabi-Maham *et al*. 2012
JN416674	*M. m. castaneus*	Iran/Zanjan	Rajabi-Maham *et al*. 2012
JN416675	*M. m. castaneus*	Iran/Shirvan	Rajabi-Maham *et al*. 2012
JN416676	*M. m. castaneus*	Iran/Zabol	Rajabi-Maham *et al*. 2012
JN416677	*M. m. castaneus*	Iran/Zabol	Rajabi-Maham *et al*. 2012
JN416678	*M. m. castaneus*	Iran/Zabol	Rajabi-Maham *et al*. 2012
JN416679	*M. m. castaneus*	Iran/Birdjand	Rajabi-Maham *et al*. 2012
JN416680	*M. m. castaneus*	Iran/Hamadan	Rajabi-Maham *et al*. 2012
JN416681	*M. m. castaneus*	Iran/Hamadan	Rajabi-Maham *et al*. 2012
JN416682	*M. m. castaneus*	Iran/Isfahan	Rajabi-Maham *et al*. 2012
JN416683	*M. m. castaneus*	Iran/Isfahan	Rajabi-Maham *et al*. 2012
JN416684	*M. m. castaneus*	Iran/Isfahan	Rajabi-Maham *et al*. 2012
JN416685	*M. m. castaneus*	Iran/Isfahan	Rajabi-Maham *et al*. 2012
JN416686	*M. m. castaneus*	Iran/Isfahan	Rajabi-Maham *et al*. 2012
JN416687	*M. m. castaneus*	Iran/Isfahan	Rajabi-Maham *et al*. 2012
JN416688	*M. m. castaneus*	Iran/Isfahan	Rajabi-Maham *et al*. 2012
JN416689	*M. m. castaneus*	Iran/Isfahan	Rajabi-Maham *et al*. 2012
JN416690	*M. m. castaneus*	Iran/Isfahan	Rajabi-Maham *et al*. 2012
JN416691	*M. m. castaneus*	Iran/Isfahan	Rajabi-Maham *et al*. 2012
JN416692	*M. m. castaneus*	Iran/Isfahan	Rajabi-Maham *et al*. 2012
JN416693	*M. m. castaneus*	Iran/Isfahan	Rajabi-Maham *et al*. 2012
JN416694	*M. m. castaneus*	Iran/Isfahan	Rajabi-Maham *et al*. 2012
JN416695	*M. m. castaneus*	Iran/Asadabad	Rajabi-Maham *et al*. 2012
JN416696	*M. m. castaneus*	Iran/Asadabad	Rajabi-Maham *et al*. 2012
JN416697	*M. m. castaneus*	Iran/Bandian	Rajabi-Maham *et al*. 2012
JN416698	*M. m. castaneus*	Iran/Bandarabbas	Rajabi-Maham *et al*. 2012
JN416699	*M. m. castaneus*	Iran/Iranshahr	Rajabi-Maham *et al*. 2012
JN416700	*M. m. castaneus*	Iran/Iranshahr	Rajabi-Maham *et al*. 2012
JN416701	*M. m. castaneus*	Iran/Iranshahr	Rajabi-Maham *et al*. 2012
JN416702	*M. m. castaneus*	Iran/Iranshahr	Rajabi-Maham *et al*. 2012
JN416703	*M. m. castaneus*	Iran/Chabahar	Rajabi-Maham *et al*. 2012
JN416704	*M. m. castaneus*	Iran/Chabahar	Rajabi-Maham *et al*. 2012
JN416705	*M. m. castaneus*	Iran/Chabahar	Rajabi-Maham *et al*. 2012

JN416706	*M. m. castaneus*	Iran/Chabahar	Rajabi-Maham *et al*. 2012
JN416707	*M. m. castaneus*	Iran/Chabahar	Rajabi-Maham *et al*. 2012
JN416708	*M. m. castaneus*	Iran/Chabahar	Rajabi-Maham *et al*. 2012
JN416709	*M. m. castaneus*	Iran/Chabahar	Rajabi-Maham *et al*. 2012
JN416710	*M. m. castaneus*	Iran/Chabahar	Rajabi-Maham *et al*. 2012
JN416711	*M. m. castaneus*	Iran/Chabahar	Rajabi-Maham *et al*. 2012
JN416712	*M. m. castaneus*	Iran/Bandarabbas	Rajabi-Maham *et al*. 2012
JN416713	*M. m. castaneus*	Iran/Bandarabbas	Rajabi-Maham *et al*. 2012
JN416714	*M. m. castaneus*	Iran/Bandarabbas	Rajabi-Maham *et al*. 2012
JN416715	*M. m. castaneus*	Iran/Bandarabbas	Rajabi-Maham *et al*. 2012
JN416716	*M. m. castaneus*	Iran/Bandarabbas	Rajabi-Maham *et al*. 2012
JN416717	*M. m. castaneus*	Iran/Mehriz	Rajabi-Maham *et al*. 2012
JN416718	*M. m. castaneus*	Iran/Zabol	Rajabi-Maham *et al*. 2012
JN416719	*M. m. castaneus*	Iran/Zabol	Rajabi-Maham *et al*. 2012
JN416720	*M. m. castaneus*	Iran/Zabol	Rajabi-Maham *et al*. 2012
JN416721	*M. m. castaneus*	Iran/Qaene	Rajabi-Maham *et al*. 2012
JN416722	*M. m. castaneus*	Iran/Zabol	Rajabi-Maham *et al*. 2012
JN416723	*M. m. castaneus*	Iran/Qaene	Rajabi-Maham *et al*. 2012
JN416724	*M. m. castaneus*	Iran/Qaene	Rajabi-Maham *et al*. 2012
JN416725	*M. m. castaneus*	Iran/Mehriz	Rajabi-Maham *et al*. 2012
JN416726	*M. m. castaneus*	Iran/Mehriz	Rajabi-Maham *et al*. 2012
JN416727	*M. m. castaneus*	Iran/Mehriz	Rajabi-Maham *et al*. 2012
JN416728	*M. m. castaneus*	Iran/Raz Gouye Nik	Rajabi-Maham *et al*. 2012
JN416729	*M. m. castaneus*	Iran/Raz Gouye Nik	Rajabi-Maham *et al*. 2012
JN416730	*M. m. castaneus*	Iran/Taft Islamie	Rajabi-Maham *et al*. 2012
JN416731	*M. m. castaneus*	Iran/Taft Islamie	Rajabi-Maham *et al*. 2012
JN416732	*M. m. castaneus*	Iran/Mehriz	Rajabi-Maham *et al*. 2012
JN416733	*M. m. castaneus*	Iran/Taft Islamie	Rajabi-Maham *et al*. 2012
JN416734	*M. m. castaneus*	Iran/Zabol	Rajabi-Maham *et al*. 2012
JN416735	*M. m. castaneus*	Iran/Azerbaijane	Rajabi-Maham *et al*. 2012
JN416736	*M. m. castaneus*	Iran/Azerbaijane	Rajabi-Maham *et al*. 2012
JN416737	*M. m. castaneus*	Iran/Zanjan	Rajabi-Maham *et al*. 2012
JN416738	*M. m. castaneus*	Iran/Zanjan	Rajabi-Maham *et al*. 2012
JN416739	*M. m. castaneus*	Iran/Zanjan	Rajabi-Maham *et al*. 2012
JN416740	*M. m. castaneus*	Iran/Zanjan	Rajabi-Maham *et al*. 2012
JN416741	*M. m. castaneus*	Iran/Zanjan	Rajabi-Maham *et al*. 2012
JN416742	*M. m. castaneus*	Iran/Boushehr	Rajabi-Maham *et al*. 2012
JN416743	*M. m. castaneus*	Kenya/Nairobi	Rajabi-Maham *et al*. 2012
JN416744	*M. m. castaneus*	Kenya/Banburi	Rajabi-Maham *et al*. 2012
JN416745	*M. m. castaneus*	Kenya/Nairobi	Rajabi-Maham *et al*. 2012
JN416746	*M. m. castaneus*	Kenya/Mombasa	Rajabi-Maham *et al*. 2012
JN416747	*M. m. castaneus*	Kenya/Mombasa	Rajabi-Maham *et al*. 2012
JN416748	*M. m. castaneus*	Kenya/Nairobi	Rajabi-Maham *et al*. 2012
JN416749	*M. m. castaneus*	Kenya/Nairobi	Rajabi-Maham *et al*. 2012
JN416750	*M. m. castaneus*	Kenya/Kalokol	Rajabi-Maham *et al*. 2012
JN416751	*M. m. castaneus*	Kenya/Mombasa	Rajabi-Maham *et al*. 2012
JN416752	*M. m. castaneus*	Kenya/Mombasa	Rajabi-Maham *et al*. 2012
JN416753	*M. m. castaneus*	Pakistan/Rawalpindi	Rajabi-Maham *et al*. 2012
JN416754	*M. m. castaneus*	Pakistan/Rawalpindi	Rajabi-Maham *et al*. 2012
JN416755	*M. m. castaneus*	Pakistan/Rawalpindi	Rajabi-Maham *et al*. 2012
JN416756	*M. m. castaneus*	Pakistan/Rawalpindi	Rajabi-Maham *et al*. 2012
JN416757	*M. m. castaneus*	Pakistan/Rawalpindi	Rajabi-Maham *et al*. 2012
JN416758	*M. m. castaneus*	Pakistan/Rawalpindi	Rajabi-Maham *et al*. 2012
JN416759	*M. m. castaneus*	Pakistan/Tahmasapabad	Rajabi-Maham *et al*. 2012
JN416760	*M. m. castaneus*	Pakistan/Rawalpindi	Rajabi-Maham *et al*. 2012
JN416761	*M. m. castaneus*	Pakistan/Tahmasapabad	Rajabi-Maham *et al*. 2012
JN416762	*M. m. castaneus*	Pakistan/Gujarkhan	Rajabi-Maham *et al*. 2012
JN416763	*M. m. castaneus*	Pakistan/Islamabad	Rajabi-Maham *et al*. 2012
JN416764	*M. m. castaneus*	Thailand/Pathum	Rajabi-Maham *et al*. 2012
JN416765	*M. m. castaneus*	Thailand/Pathum	Rajabi-Maham *et al*. 2012
JN416766	*M. m. castaneus*	Iran/Azerbaijane	Rajabi-Maham *et al*. 2012
JN416767	*M. m. castaneus*	Iran/Azerbaijane	Rajabi-Maham *et al*. 2012
JN416768	*M. m. castaneus*	Iran/Azerbaijane	Rajabi-Maham *et al*. 2012
JN416769	*M. m. castaneus*	Madagascar	Rajabi-Maham *et al*. 2012

7.2.2 Chapter 3: *Rattus rattus*

Sample Nb	Lab Code	Species	Provided by	Date	Location	Material	DNA
1	AT001	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
2	AT002	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
3	AT003	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
4	AT004	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
5	AT005	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
6	AT006	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
7	AT007	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	
8	AT008	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	
9	AT009	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
11	AT010	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
12	AT011	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
13	AT012	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
14	AT013	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
15	AT014	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
16	AT015	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	
17	AT016	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
19	AT017	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
23	AT018	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
25	AT019	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
26	AT020	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
27	AT021	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
28	AT022	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	
29	AT023	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	
30	AT024	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	
74	AT025	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	
76	AT026	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
X12	AT027	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
X13	AT028	*R. tanezumi*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
187398	HE001	*R. rattus*	FMNH	modern	Tanzania/Mafia	tissue - Ethanol	x
187400	HE002	*R. rattus*	FMNH	modern	Tanzania/Mafia	tissue - Ethanol	x
192839	HE003	*R. rattus*	FMNH	modern	Tanzania/Pemba	tissue - Ethanol	x
192841	HE004	*R. rattus*	FMNH	modern	Tanzania/Pemba	tissue - Ethanol	x
198176	HE005	*R. rattus*	FMNH	modern	Tanzania/Zanzibar	tissue - Ethanol	x
165732	HE006	*R. r. rattus*	FMNH	modern	Madagascar	tissue - Ethanol	x
175937	HE007	*R. r. rattus*	FMNH	modern	Madagascar	tissue - Ethanol	x
194717	HE008	*R. rattus*	FMNH	modern	Madagascar	tissue - Ethanol	x
179285	HE009	*R. rattus*	FMNH	modern	Madagascar	tissue - Ethanol	x
178864	HE010	*R. rattus*	FMNH	modern	Madagascar	tissue - Ethanol	x
194574	HE011	*R. rattus*	FMNH	modern	Madagascar	tissue - Ethanol	x
204453	HE012	*R. rattus*	FMNH	modern	Madagascar	tissue - Ethanol	
178860	HE013	*R. rattus*	FMNH	modern	Madagascar	tissue - Ethanol	x
194613	HE014	*R. rattus*	FMNH	modern	Madagascar	tissue - Ethanol	x
192838	HE015	*R. rattus*	FMNH	modern	Tanzania/Pemba	tissue - Ethanol	x
192905	HE016	*R. rattus*	FMNH	modern	Tanzania/Pemba	tissue - Ethanol	x
187399	HE017	*R. rattus*	FMNH	modern	Tanzania/Mafia	tissue - Ethanol	x
187397	HE018	*R. rattus*	FMNH	modern	Tanzania/Mafia	tissue - Ethanol	x
187396	HE019	*R. rattus*	FMNH	modern	Tanzania/Mafia	tissue - Ethanol	x
166260	HE020	*R. r. rattus*	FMNH	modern	Madagascar	tissue - Ethanol	x
194586	HE021	*R. rattus*	FMNH	modern	Madagascar	tissue - Ethanol	x
178680	HE022	*R. rattus*	FMNH	modern	Madagascar	tissue - Ethanol	x
192840	HE023	*R. rattus*	FMNH	modern	Tanzania/Pemba	tissue - Ethanol	x
166083	HE024	*R. r. rattus*	FMNH	modern	Madagascar	tissue - Ethanol	x
176119	HE025	*R. r. rattus*	FMNH	modern	Madagascar	tissue - Ethanol	x
175948	HE026	*R. r. rattus*	FMNH	modern	Madagascar	tissue - Ethanol	x
166225	HE027	*R. r. rattus*	FMNH	modern	Madagascar	tissue - Ethanol	x
176023	HE028	*R. r. rattus*	FMNH	modern	Madagascar	tissue - Ethanol	x
163132	R001	*R. r. wroughtoni*	AMNH	1943	India/Madras/Kalhatti	dried tissue	x
163134	R002	*Rattus*	AMNH	1943	India/Madras/Avalance	dried tissue	x
163147	R003	*R. r. wroughtoni*	AMNH	1943	India/Madras/Kalhatti	dried tissue	x
55587	R004	*R. r. rattus*	AMNH	1916	Tanzania/Kilossa	dried tissue	x
109188	R005	*R. r. pelengensis*	AMNH	1934	Sulawesi/Peleng	dried tissue	

34045	R006	*R. r. rattus*	FMNH	1915	Madagascar	dried tissue	**x**
122589	R007	*R. r. rattus*	FMNH	1954	Turkey/Adana	dried tissue	**x**
76616	R008	*R. r. brunneusculus*	FMNH	1950	India/Assam/Manipur	dried tissue	**x**
142241	R009	*R. r. arboreus*	FMNH	1968	Nepal	dried tissue	**x**
31869	R010	*R. r. flavipectus*	FMNH	1929	Vietnam/Tonkin	dried tissue	**x**
114187	R011	*R. r. brunneus*	FMNH	1972	Nepal	dried tissue	
31953	R012	*R. r. flavipectus*	FMNH	1929	Laos/Phong Saly	dried tissue	**x**
80170	R013	*R. r. diardii*	FMNH	1951	Malaysia/Sarawak	dried tissue	**x**
108906	R014	*R. r. diardii*	FMNH	1965	Malaysia/Sabah/	dried tissue	**x**
40521	R015	*R. r. flavipectus*	FMNH	1932	China/Sezchuan/Kao Ku	dried tissue	
31912	R016	*R. r. flavipectus*	FMNH	1929	Vietnam/Tonkin	dried tissue	
82954	R017	*R. r. arboreus*	FMNH	1914	India/Daltonganj	dried tissue	
76617	R018	*R. r. brunneusculus*	FMNH	1950	India/Assam/Manipur	dried tissue	**x**
88453	R019	*R. r. diardii*	FMNH	1950	N-Borneo/Sandakan	dried tissue	
82952	R020	*R. r. kandianus*	FMNH	1913	Ceylon/Udugama	dried tissue	**x**
98814	R021	*R. r. rattus*	FMNH	1962	Egypt/Aswan	dried tissue	**x**
77730	R022	*R. r. rattus*	FMNH	1950	Egypt/Wadi Natrun	dried tissue	
100710	R023	*R. r. rattus*	FMNH	?	Egypt/Cairo	dried tissue	
85360	R024	*R. r. rattus*	FMNH	1948	Egypt/Faiyum	dried tissue	**x**
77242	R025	*R. r. rattus*	FMNH	?	Egypt/Giza	dried tissue	**x**
74632	R026	*R. r. rattus*	FMNH	1953	Egypt/Qena	dried tissue	**x**
99117	R027	*R. r. rattus*	FMNH	1929	Ethiopia/Addis Abbaba	dried tissue	
97524	R028	*R. r. rattus*	FMNH	1962	Iran/Gorgan	dried tissue	**x**
85420	R029	*R. r. rattus*	FMNH	1948	Kenya/Arabuka	dried tissue	**x**
85348	R030	*R. r. rattus*	FMNH	1948	Kenya/Kitero Village	dried tissue	**x**
85350	R031	*R. r. rattus*	FMNH	1948	Kenya/Njoro	dried tissue	**x**
85352	R032	*R. r. rattus*	FMNH	1948	Kenya/Njoro	dried tissue	**x**
42373	R033	*R. r. rattus*	FMNH	1934	Sudan/Sotuba	dried tissue	**x**
99603	R034	*R. r. rattus*	FMNH	1963	Saudi Arabia/Hillat Muhaish	dried tissue	**x**
86317	R035	*R. r. rattus*	FMNH	1956	Tanzania/Tanganyika	dried tissue	
86316	R036	*R. r. rattus*	FMNH	1956	Tanzania/Arusha	dried tissue	**x**
77983	R037	*R. r. rattus*	FMNH	1951	Yemen/San'a	dried tissue	**x**
82961	R038	*R. r. wroughtoni*	FMNH	1921	India/Benhope	dried tissue	**x**
186995	R039	*R. rattus*	FMNH	modern	Tanzania/Lindi Region	dried tissue	
187064	R040	*R. rattus*	FMNH	1995	Tanzania/Morogoro	dried tissue	**x**
81931	R041	*R. r. rattus*	FMNH	1954	Angola/Canzele	dried tissue	
103666	R042	*R. r. rattus*	FMNH	1965	Afghanistan/Jalalabad	dried tissue	**x**
82969	R043	*R. r. macmillani*	FMNH	1914	Burma/Hkmati	dried tissue	
43689	R044	*R. r. rattus*	FMNH	1931	Cameroon/Sakbayeme	dried tissue	
34835	R045	*R. r. brunneusculus*	FMNH	1930	India/Bengal/Mangpu	dried tissue	**x**
179191	R046	*R. rattus*	FMNH	2001	Jordan/Irbid/Hemmah	dried tissue	**x**
179192	R047	*R. rattus*	FMNH	2001	Jordan/Irbid/Hemmah	dried tissue	**x**
122595	R048	*R. r. rattus*	FMNH	1954	Turkey/Adana	dried tissue	**x**
122594	R049	*R. r. rattus*	FMNH	1954	Turkey/Adana	dried tissue	**x**
34046	R050	*R. r. rattus*	FMNH	1915	Madagascar	dried tissue	**x**
85366	R051	*R. rattus*	FMNH	1948	Madagascar	dried tissue	**x**
85415	R052	*R. r. rattus*	FMNH	1948	Madagascar	dried tissue	**x**
85417	R053	*R. r. rattus*	FMNH	1948	Madagascar	dried tissue	
85418	R054	*R. r. rattus*	FMNH	1948	Madagascar	dried tissue	
85353	R055	*R. r. rattus*	FMNH	1948	Kenya/Subukia	dried tissue	**x**
35581	R056	*R. r. brunneusculus*	FMNH	1961	India/Bengal/Mangpu	dried tissue	**x**
86213	R057	*R. r. rattus*	FMNH	1956	Tanzania/Tanganyika	dried tissue	**x**
85347	R058	*R. r. rattus*	FMNH	1948	Kenya/Njoro	dried tissue	
29777	R059	*R. r. narbadae*	FMNH	?	India/Alapalli	dried tissue	**x**
76726	R060	*R. r. brunneusculus*	FMNH	1951	India/Assam	dried tissue	**x**
76624	R061	*R. r. brunneusculus*	FMNH	1950	India/Assam/Manipur	dried tissue	**x**
108972	R062	*R. r. diardii*	FMNH	1965	Malaysia/Sabah/Kinabalu	dried tissue	**x**
94207	R063	*R. r. brunneus*	FMNH	1960	Nepal	dried tissue	**x**
80164	R064	*R. r. diardii*	FMNH	1951	Malaysia/Sarawak	dried tissue	
91264	R065	*R. r. rattus*	FMNH	1958	India	dried tissue	
99114	R066	*R. r. rattus*	FMNH	1929	Ethiopia/Addis Abbaba	dried tissue	**x**
93763	R067	*R. r. rattus*	FMNH	1961	Egypt/Cairo	dried tissue	**x**
98815	R068	*R. r. rattus*	FMNH	1929	Egypt/Aswan	dried tissue	**x**
92222	R069	*R. r. kandianus*	FMNH	1966	Ceylon/Dehiwala	dried tissue	**x**

82953	R070	*R. r. kandianus*	FMNH	1913	Ceylon/Udugama	dried tissue	
92223	R071	*R. r. kandianus*	FMNH	1960	Ceylon/Gonapola	dried tissue	**x**
31903	R072	R. r. flavipectus	FMNH	1929	Vietnam/Tonkin	dried tissue	**x**
31902	R073	R. r. flavipectus	FMNH	1929	Vietnam/Tonkin	dried tissue	
80167	R074	*R. r. diardii*	FMNH	1951	Malaysia/Sarawak	dried tissue	
86314	R075	*R. r. rattus*	FMNH	?	Tanzania/Arusha/Tengeru	dried tissue	**x**
66182	R076	*R. r. thai*	FMNH	1949	Siam/Kam Pang	dried tissue	**x**
95167	R077	*R. r. rattus*	FMNH	1961	Southern Rhodesia/Gatooma	dried tissue	**x**
82968	R078	*R. r. tikos*	FMNH	1914	Burma/Tenasserim	dried tissue	**x**
66183	R079	*R. r. thai*	FMNH	1949	Siam/Kam Pang	dried tissue	**x**
195591	R080	*R. rattus*	FMNH	2006	Tanzania/Tanga Region	dried tissue	**x**
32882	R081	*R. sikkimensis*	FMNH	1922	China/Hainan/Nodoa	dried tissue	**x**
39420	R082	*R. r. flavipectus*	FMNH	1917	China/Yunnan	dried tissue	**x**
32887	R083	*R. sikkimensis*	FMNH	1922	China/Hainan/Nodoa	dried tissue	**x**
40512	R084	*R. r. flavipectus*	FMNH	1932	China/Sezchuan/Kao Ku	dried tissue	**x**
26854	R085	*R. r. tistae*	AMNH	1901	India/Meghalaya	dried tissue	
245114	R086	*R. r. wroughtoni*	AMNH	1985	India/Tamil Nadu State	dried tissue	**x**
26851	R087	*R. r. tistae*	AMNH	1901	India/Meghalaya State	dried tissue	**x**
114379	R088	*R. r. rattus*	AMNH	1934	Kenya/Voi	dried tissue	**x**
114382	R089	*R. r. rattus*	AMNH	1934	Kenya/Voi	dried tissue	**x**
114378	R090	*R. r. rattus*	AMNH	1934	Kenya/Voi	dried tissue	**x**
114026	R091	*R. r. rattus*	AMNH	1932	Kenya/Aberdare Mts.	dried tissue	**x**
114027	R092	*R. r. rattus*	AMNH	1932	Kenya/Mt Kenya	dried tissue	**x**
81338	R093	*R. r. rattus*	AMNH	1925	Tanzania/Rungwe	dried tissue	**x**
161358	R094	*R. r. rattus*	AMNH	1942	Zambia/Lake Chiyaua	dried tissue	**x**
161359	R095	*R. r. rattus*	AMNH	1942	Zambia/Lake Chiyaua	dried tissue	**x**
162122	R096	*R. r. rattus*	AMNH	1942	Malawi/Kasungu	dried tissue	**x**
162121	R097	*R. r. rattus*	AMNH	1942	Malawi/Kasungu	dried tissue	**x**
54956	R098	*R. r. tikos*	AMNH	1924	Burma/Mergus	dried tissue	**x**
54955	R099	*R. r. tikos*	AMNH	1924	Burma/Mergus	dried tissue	**x**
54954	R100	*R. r. tikos*	AMNH	1924	Burma/Mergus	dried tissue	**x**
240874	R101	*Rattus*	AMNH	?	Sri Lanka	dried tissue	**x**
240878	R102	*Rattus*	AMNH	1925	Sri Lanka	dried tissue	**x**
163153	R103	*Rattus*	AMNH	1937	India/Madras	dried tissue	**x**
163142	R104	*R. r. rufescens*	AMNH	1943	India/Madras	dried tissue	**x**
152965	R105	*R. r. sapoensis*	AMNH	1935	Indonesia/Sulawesi/Togian Is.	dried tissue	**x**
240871	R106	*Rattus*	AMNH	1940	Sri Lanka	dried tissue	**x**
242124	R107	*R. r. mindanensis*	AMNH	?	Philippines/Luzon/Benguet/Baguio	dried tissue	
54541	R108	*R. r. rufescens*	AMNH	1919	India/Sonaripur/Kheri Dist.	dried tissue	**x**
163757	R109	*R. sladeni*	AMNH	1922	Burma/Mandalay	dried tissue	**x**
163760	R110	*R. sladeni*	AMNH	1925	Burma/Mandalay	dried tissue	**x**
152970	R111	*R. r. sapoensis*	AMNH	1935	Indonesia/Sulawesi/Togian Is.	dried tissue	**x**
152971	R112	*R. r. sapoensis*	AMNH	1935	Indonesia/Sulawesi/Togian Is.	dried tissue	
59872	R113	*R. r. flavipetus*	AMNH	?	China/Szechuan/Wanksien	dried tissue	**x**
118995	R114	*R. r. rattus*	AMNH	1939	Uganda/Bwamba/Bundimbale	dried tissue	**x**
44726	R115	*R. r. flavipectus*	AMNH	1913	China/Fukien Prov./Kuliang	dried tissue	
113071	R116	*R. r. flavipectus*	AMNH	1931	Burma/Gova	dried tissue	**x**
103073	R117	*R. r. diardii*	AMNH	1930	Indonesia/Mentawai Islands	dried tissue	**x**
109193	R118	*R. r. pelengensis*	AMNH	1934	Indonesia/Sulawesi/Peleng I	dried tissue	**x**
113069	R119	*R. r. flavipectus*	AMNH	1931	Burma/Tawmaw	dried tissue	**x**
107544	R120	*R. r. diardii*	AMNH	1932	Indonesia/Bali/Selat/Tseh	dried tissue	**x**
107978	R121	*R. tiomanicus*	AMNH	1945	Indonesia/Bali/Bratan	dried tissue	**x**
102397	R122	*R. r. diardii*	AMNH	1930	Indonesia/Java/Cheribon	dried tissue	**x**
107606	R123	*R. r. diardii*	AMNH	1934	Indonesia/Bali/Nusa Penida	dried tissue	**x**
102010	R124	*R. r. diardii*	AMNH	1929	Indonesia/Java/Cheribon	dried tissue	**x**
102112	R125	*R. r. diardii*	AMNH	1929	Indonesia/Java/Cheribon	dried tissue	**x**
101551	R126	*R. r. diardii*	AMNH	1928	Indonesia/Java/Cheribon	dried tissue	**x**
102005	R127	*R. r. diardii*	AMNH	1929	Indonesia/Java/Cheribon	dried tissue	**x**
55181	R128	*R. r. rattus*	AMNH	1920	Zaire/Kasai Occident/Kanaga	dried tissue	**x**
55530	R129	*R. r. rattus*	AMNH	1917	Tanzania/Kilossa	dried tissue	**x**
82669	R130	*R. r. rattus*	AMNH	1922	Zaire/Kalongi	dried tissue	**x**
84666	R131	*R. r. flavipectus*	AMNH	?	China	dried tissue	**x**
109190	R132	*R. r. pelengensis*	AMNH	1934	Indonesia/Sulawesi/Peleng I	dried tissue	**x**
85082	R133	*R. r. rattus*	AMNH	1922	Angola/Huambo/Luimbale Mt.	dried tissue	**x**

85081	R134	*R. r. rattus*	AMNH	1922	Angola/Huambo/Luimbale Mt.	dried tissue	**x**
109185	R135	*R. r. pelengensis*	AMNH	1934	Indonesia/Sulawesi/Peleng I	dried tissue	**x**
89714	R136	*R. r. rattus*	AMNH	1935	Uganda/Kampala	dried tissue	**x**
85083	R137	*R. r. rattus*	AMNH	1922	Angola/Huambo/Luimbale Mt.	dried tissue	**x**
102815	R138	*R. r. diardii*	AMNH	1930	Indonesia/Sumatra/Kalianda	dried tissue	**x**
102817	R139	*R. r. diardii*	AMNH	1930	Indonesia/Sumatra/Kalianda	dried tissue	**x**
111608	R140	*R. r. gigas*	AMNH	1930	China/Szechuan/Chengtu	dried tissue	**x**
111611	R141	*R. r. flavipectus*	AMNH	1930	China/Szechuan/Chengtu	dried tissue	**x**
111615	R142	*R. r. flavipectus*	AMNH	1930	China/Szechuan/Chengtu	dried tissue	**x**
111974	R143	*Rattus*	AMNH	1956	China/Szechuan/Tsao Po	dried tissue	**x**
163749	R144	*R. sladeni*	AMNH	1940	Burma/Popa Mt./Mandalay	dried tissue	**x**
163751	R145	*R. sladeni*	AMNH	1954	Burma/Popa Mt./Mandalay	dried tissue	**x**
167533	R146	*Rattus*	AMNH	1940	Liberia/Ganta	dried tissue	**x**
205049	R147	*R. r. rattus*	AMNH	?	Tanzania/Singida/Manyoni,	dried tissue	
214234	R148	*R. r. sumbae*	AMNH	1963	Indonesia/Timor Island	dried tissue	
217643	R149	*R. r. diardii*	AMNH	1965	Malaysia/Selangor	dried tissue	
217645	R150	*R. r. diardii*	AMNH	1965	Malaysia/Selangor	dried tissue	
239709	R151	*R. r. diardii*	AMNH	1966	Malaysia/Selangor	dried tissue	
239711	R152	*R. r. diardii*	AMNH	1966	Malaysia/Selangor	dried tissue	
240416	R153	*R. r. diardii*	AMNH	1964	Malaysia/Selangor	dried tissue	
240427	R154	*R. r. diardii*	AMNH	1964	Malaysia/Selangor	dried tissue	
240949	R155	*Rattus*	AMNH	1940	Sri Lanka/Sabaragamuwa Prov.	dried tissue	**x**
252183	R156	*R. r. sumbae*	AMNH	1971	Indonesia/Flores/Nusa Engar	dried tissue	**x**
252193	R157	*R. r. sumbae*	AMNH	1971	Indonesia/Flores/Nusa Engar	dried tissue	**x**
106385	R158	*R. r. diardii*	AMNH	1932	Indonesia/Sumatra/Boekit	dried tissue	**x**
102994	R159	*R. r. diardii*	AMNH	1930	Indonesia/Sumatra/Goemoengsoekit	dried tissue	**x**
103526	R160	*R. r. diardii*	AMNH	1931	Indonesia/Borneo/Kalimantan	dried tissue	**x**
151499	R161	*R. r. rattus*	AMNH	1935	Australia,/Cocos Keeling Island	dried tissue	**x**
103525	R162	*R. r. diardii*	AMNH	1931	Indonesia/Borneo/Kalimantan	dried tissue	**x**
103292	R163	*R. r. diardii*	AMNH	1930	Indonesia/Sumatra/Palembang,	dried tissue	**x**
252099	R164	*R. r. diardii*	AMNH	1972	Indonesia/Sumatra/Lubuk	dried tissue	**x**
102673	R165	*R. r. diardii*	AMNH	1930	Indonesia/Sumatra/Palembang	dried tissue	**x**
102580	R166	*R. r. diardii*	AMNH	1930	Indonesia/Sumatra/Palembang	dried tissue	**x**
102579	R167	*R. r. diardii*	AMNH	1930	Indonesia/Sumatra/Palembang	dried tissue	**x**
102577	R168	*R. r. diardii*	AMNH	1930	Indonesia/Sumatra/Palembang	dried tissue	**x**
102674	R169	*R. r. diardii*	AMNH	1930	Indonesia/Sumatra/Palembang	dried tissue	**x**
151492	R170	*R. r. rattus*	AMNH	1935	Australia/Cocos Keeling Island	dried tissue	**x**
100686	R171	*R. r. rattus*	AMNH	1926	Madagascar	dried tissue	**x**
100726	R172	*R. r. rattus*	AMNH	1925	Madagascar	dried tissue	**x**
150060	R173	*R. r. kandianus*	AMNH	1944	Sri Lanka	dried tissue	**x**
240955	R174	*Rattus*	AMNH	?	Sri Lanka/Uva Province/Welimada	dried tissue	**x**
106999	R175	*R. r. diardii*	AMNH	1933	Indonesia/Borneo/Kalimantan	dried tissue	**x**
107004	R176	*R. r. diardii*	AMNH	1932	Indonesia/Borneo/Kalimantan	dried tissue	**x**
107024	R177	*R. r. diardii*	AMNH	1933	Indonesia/Borneo/Kalimantan	dried tissue	**x**
107048	R178	*R. exulans*	AMNH	1983	Indonesia/Borneo/Kalimantan	dried tissue	**x**
103436	R179	*R. r. diardii*	AMNH	1931	Indonesia/Borneo/Kalimantan	dried tissue	**x**
103437	R180	*R. r. diardii*	AMNH	1931	Indonesia/Borneo/Kalimantan	dried tissue	**x**
88843	R181	*R. r. rattus*	AMNH	1934	Iran/Dar Kaleh/Astarabad	dried tissue	**x**
88844	R182	*R. r. rattus*	AMNH	1934	Iran/Dar Kaleh/Astarabad	dried tissue	**x**
101278	R183	*R. r. palelae*	AMNH	1927	Indonesia/Sulawesi/Roeroekan	dried tissue	**x**
101275	R184	*R. pesticulus*	AMNH	1978	Indonesia/Sulawesi/Roeroekan	dried tissue	**x**
152986	R185	*R. r. palelae*	AMNH	1935	Indonesia/Sulawesi/Boemboelan	dried tissue	**x**
152985	R186	*R. r. palelae*	AMNH	1935	Indonesia/Sulawesi/Boemboelan	dried tissue	**x**
162110	R187	*R. r. rattus*	AMNH	1942	Malawi/Kota Kota	dried tissue	**x**
162102	R188	*R. r. rattus*	AMNH	1942	Malawi/Zomba	dried tissue	**x**
205043	R189	*R. r. rattus*	AMNH	?	Tanzania/Singida/Manyoni	dried tissue	**x**
81971	R190	*R. r. rattus*	AMNH	1926	Zaire/Albertsville	dried tissue	**x**
85724	R191	*R. r. rattus*	AMNH	1921	Angola/Benguela/Lobito,	dried tissue	**x**
55592	R192	*R. r. rattus*	AMNH	1917	Tanzania/Kilossa	dried tissue	**x**
55591	R193	*R. r. rattus*	AMNH	1917	Tanzania/Dodoma	dried tissue	**x**
55586	R194	*R. r. rattus*	AMNH	1916	Tanzania/Kilossa	dried tissue	
120181	R195	*R. r. rattus*	AMNH	?	Congo/Dolisie	dried tissue	**x**
120189	R196	*R. r. rattus*	AMNH	?	Congo/Dolisie	dried tissue	**x**
256646	R197	*R. r. rattus*	AMNH	1979	United Arab Emirates/Mukala City	dried tissue	**x**

256645	R198	*R. r. rattus*	AMNH	1979	United Arab Emirates/Mukala City	dried tissue	**x**
242127	R199	*R. r. mindanensis*	AMNH	?	Philippines/Mindanao/Cotabato	dried tissue	**x**
43353	R200	*R. r. yunnanensis*	AMNH	?	China/Yunnan	dried tissue	**x**
43363	R201	*R. r. yunnanensis*	AMNH	?	China/Yunnan	dried tissue	**x**
163140	R202	*R. r. rufescens*	AMNH	1943	India/Mysore/Biligirirangan	dried tissue	**x**
163139	R203	*R. r. rufescens*	AMNH	1943	India/Mysore/Biligirirangan	dried tissue	**x**
97516	R204	*R. r. rattus*	FMNH	1962	Iran/Gorgan	dried tissue	**x**
57361	R205	*R. rattus*	FMNH	1948	Kenya/Tezo/Kilifi	dried tissue	**x**
77990	R206	*R. r. rattus*	FMNH	1951	Yemen/San'a	dried tissue	**x**
77977	R207	*R. r. rattus*	FMNH	1951	Yemen/San'a	dried tissue	**x**
31900	R208	*R. r. flavipectus*	FMNH	1929	Vietnam/Tonkin	dried tissue	**x**
104213	R209	*R. r. brunneus*	FMNH	1966	Nepal	dried tissue	
142243	R210	*R. r. brunneus*	FMNH	1969	Nepal	dried tissue	
76638	R211	*R. r. brunneusculus*	FMNH	1950	India/Assam/Manipur	dried tissue	**x**
142244	R212	*R. r. brunneus*	FMNH	1969	Nepal	dried tissue	
31897	R213	*R. r. flavipectus*	FMNH	1929	Vietnam/Tonkin	dried tissue	**x**
31950	R214	*R. r. flavipectus*	FMNH	1929	Laos/Phong Saly	dried tissue	
94261	R215	*R. r. rattus*	FMNH	1961	N-Lebanon/Akkar/Halba	dried tissue	**x**
179189	R216	*R. rattus*	FMNH	2001	Jordan/Irbid/Hemmah	dried tissue	**x**
122593	R217	*R. r. rattus*	FMNH	1954	Turkey/Adana/Haruniye	dried tissue	**x**
103664	R218	*R. r. rattus*	FMNH	1965	Afghanistan/Jalalabad	dried tissue	
99115	R219	*R. r. rattus*	FMNH	1929	Ethiopia/Addis Abbaba	dried tissue	**x**
68310	R220	*R. r. rattus*	FMNH	1950	Egypt/Cairo	dried tissue	
100708	R221	*R. r. rattus*	FMNH	?	Egypt/Cairo	dried tissue	
104216	R222	*R. r. brunneus*	FMNH	1966	Nepal	dried tissue	
104215	R223	*R. r. brunneus*	FMNH	1966	Nepal	dried tissue	
122590	R224	*R. r. rattus*	FMNH	1954	Turkey/Adana/Haruniye	dried tissue	**x**
85361	R225	*R. r. rattus*	FMNH	1950	Egypt/Faiyum	dried tissue	**x**
99118	R226	*R. r. rattus*	FMNH	1929	Ethiopia/Addis Abbaba	dried tissue	**x**
100733	R227	*R. r. rattus*	FMNH	?	Egypt/Asyut	dried tissue	**x**
82950	R228	*R. r. brunneus*	FMNH	1921	India/Hathiban	dried tissue	**x**
6.7.2.3	R229	*Rattus*	BNHM	?	Andaman Islands/Narcondam	dried tissue	**x**
10.7.26.2	R230	*Rattus*	BNHM	?	Andaman Islands/Brother Island	dried tissue	**x**
10.7.26.3	R231	*Rattus*	BNHM	?	Andaman Islands/Brother Island	dried tissue	**x**
1957.407	R232	*R. r. ceylonus*	BNHM	?	Maldives/Male Atoll	dried tissue	**x**
1957.406	R233	*R. r. ceylonus*	BNHM	?	Maldives/Male Atoll	dried tissue	**x**
1957.405	R234	*R. r. ceylonus*	BNHM	?	Maldives/Male Atoll	dried tissue	**x**
57,410	R235	*R. r. kandianus*	BNHM	?	Maldives/Mululay Island	dried tissue	**x**
59,549	R236	*R. r. kandianus*	BNHM	?	Maldives/Addu Atoll/Gan Island	dried tissue	**x**
59,546	R237	*R. r. kandianus*	BNHM	?	Maldives/Addu Atoll/Gan Island	dried tissue	**x**
5.5.8.42	R238	*R. r. alexandrius*	BNHM	?	Sudan/Khartoum	dried tissue	**x**
66.5810	R239	*R. r. frugivorus*	BNHM	?	Sudan/Khartoum	dried tissue	**x**
50.47	R240	*R. r. rattus*	BNHM	?	Ethiopia/Batie	dried tissue	**x**
72,371	R241	*R. rattus*	BNHM	1970	Ethiopia/Didessa	dried tissue	**x**
77.3261	R242	*R. rattus*	BNHM	?	Sudan	dried tissue	**x**
33.6.14.4	R243	*R. rattus*	BNHM	?	Tanzania/Mafia	dried tissue	**x**
55,340	R244	*R. r. frugivorus*	BNHM	?	Tanzania/Pemba	dried tissue	**x**
34.1.11.48	R245	*R. r. alexandrius*	BNHM	?	Mozambique/Lurio	dried tissue	
34.1.11.44	R246	*R. r. alexandrius*	BNHM	?	Mozambique/Tete	dried tissue	**x**
83.524	R247	*R. rattus*	BNHM	1983	Rodrigues Islands	dried tissue	
69.1243	R248	*R. rattus*	BNHM	1969	Mauritius	dried tissue	
69.1244	R249	*R. rattus*	BNHM	1969	Mauritius	dried tissue	**x**
69.1245	R250	*R. rattus*	BNHM	1969	Mauritius	dried tissue	
36.4.14.58	R251	*R. rattus*	BNHM	?	Persian Golf/Taub Island	dried tissue	**x**
36.4.14.59	R252	*R. rattus*	BNHM	?	Persian Golf/Taub Island	dried tissue	**x**
74,379	R253	*R. r. frugivorus*	BNHM	?	Bahrein	dried tissue	
99.3.14.11	R254	*R. rattus*	BNHM	?	Socotra/Adho Diemellus	dried tissue	**x**
99.3.14.6	R255	*Rattus*	BNHM	?	Socotra	dried tissue	**x**
99.3.14.5	R256	*Rattus*	BNHM	?	Socotra	dried tissue	**x**
77.3084	R257	*R. rattus*	BNHM	1974	Iran/Mirmahalent	dried tissue	
99.3.14.9	R258	*R. r. flavipectus*	BNHM	?	Socotra/Thluteed	dried tissue	**x**
99.3.14.8	R259	*R. r. rattus*	BNHM	?	Socotra/Thluteed	dried tissue	**x**
77.3084	R260 = R257	*R. rattus*	BNHM	1974	Iran/Mirmahalent	dried tissue	
83.524	R261 = R247	*R. rattus*	BNHM	1983	Rodrigues Islands	dried tissue	

34.1.11.48	R262 = R245	*R. rattus*	BNHM	?	Mozambique	dried tissue	
34045	R263 = R006	*R. r. rattus*	FMNH	1915	Madagascar/Ambantondrazaka	dried tissue	**x**
77983	R264 = R037	*R. r. rattus*	FMNH	1956	Yemen/San'a	dried tissue	**x**
99114	R265 = R066	*R. r. rattus*	FMNH	1929	Ethiopia/Addis Abbaba	dried tissue	**x**
42373	R266 = R033	*R. r. rattus*	FMNH	1934	Sudan/Sotuba	dried tissue	**x**
99117	R267 = R027	*R. r. rattus*	FMNH	1929	Ethiopia/Addis Abbaba	dried tissue	
179191	R268 = R046	*R. rattus*	FMNH	2001	Jordan/Irbid/Hemmah	dried tissue	**x**
179192	R269 = R047	*R. rattus*	FMNH	2001	Jordan/Irbid/Hemmah	dried tissue	**x**
203220	R270	*R. r. rattus*	AMNH	1957	Egypt	skin	**x**
163152	R271	*R. r. wroughtoni*	AMNH	1943	India/Madras	skin	
251685	R272	*R. rattus*	AMNH	1971	Bangladesh	skin	**x**
150166	R273	*R. r. rattus*	AMNH	1937	Australia/Cocos Keeling Island	skin	
150170	R274	*R. r. rattus*	AMNH	1937	Australia/Cocos Keeling Island	skin	
194154	R275	*R. alexandrinus*	AMNH	1963	Australia/Queensland	skin	**x**
197385	R276	*R. rattus*	AMNH	1961	Australia/Western	skin	**x**
197396	R277	*R. rattus*	AMNH	1959	Australia/Western	skin	**x**
197681	R278	*R. r. rattus*	AMNH	1963	Australia/Western	skin	**x**
215597	R279	*R. r. rufescena*	AMNH	1960	India/Mandala	skin	**x**
215598	R280	*R. r. rufescena*	AMNH	1961	India/Rajastan	skin	**x**
215599	R281	*R. r. rufescena*	AMNH	1961	India/West Bengal	skin	**x**
215604	R282	*Rattus*	AMNH	1965	India/West Bengal	skin	**x**
114376	R283	*R. r. rattus*	AMNH	1934	Kenya/Voi	skin	
27882	R284	*R. r. kijabius*	AMNH	1902	Kenya/Kijabe	skin	
187532	R285	*R. r. rattus*	AMNH	1956	Kenya/Kilifi	skin	**x**
187674	R286	*R. r. rattus*	AMNH	1956	Kenya/Kilifi	skin	
187591	R287	*R. r. rattus*	AMNH	1956	Kenya/Nyeri	skin	**x**
184680	R288	*R. r. rattus*	AMNH	1955	Pakistan	skin	**x**
170239	R289	*R. r. rattus*	AMNH	1946	Pakistan	skin	**x**
277480	R290	*R. rattus*	AMNH	1981	Uganda	skin	**x**
161360	R291	*R. r. rattus*	AMNH	1942	Zambia	skin	
HME	R292	*R. rattus*	Heidi M. Eager	modern	Zanzibar	skin	**x**

NCBI Accession Nb	Species	Location	Author
AB033702	*R. rattus*	?	Suzuki *et al*. 2000
AB096841	*R. tanezumi*	?	Suzuki *et al*. 2003
AB211039	*R. rattus*	Japan	Chinen & Suzuki 2005
AB211040	*R. tanezumi*	Japan	Chinen & Suzuki 2005
AB211041	*R. tanezumi*	Japan	Chinen & Suzuki 2005
AB211042	*R. tanezumi*	Japan	Chinen & Suzuki 2005
AB211043	*R. tanezumi*	Japan	Chinen & Suzuki 2005
AB355901	*R. tanezumi*	Vietnam/Haiphongtane	Truong *et al*. 2009
AJ428514	*R. norvegicus*	Denmark	Nilsson *et al*. 2003
DQ191286	*R. exulans*	Philippines	?
DQ191288	*R. tanezumi*	Philippines	?
DQ439816	*R. tanezumi*	South Africa	Bastos *et al*. 2011
DQ439830	*R. rattus*	South Africa	Bastos *et al*. 2011
DQ439834	*R. rattus*	South Africa	Bastos *et al*. 2011
DQ439841	*R. norvegicus*	South Africa	Bastos *et al*. 2011
DQ439842	*R. norvegicus*	South Africa	Bastos *et al*. 2011
EF186414	*R. exulans*	Cook Islands	Robins *et al*. 2007
EF186418	*R. exulans*	USA	Robins *et al*. 2007
EF186420	*R. exulans*	French Polynesia	Robins *et al*. 2007
EF186421	*R. exulans*	Indonesia	Robins *et al*. 2007
EF186423	*R. exulans*	French Polynesia	Robins *et al*. 2007
EF186426	*R. exulans*	New Zealand	Robins *et al*. 2007
EF186427	*R. exulans*	Papua New Guinea	Robins *et al*. 2007
EF186429	*R. exulans*	French Polynesia	Robins *et al*. 2007
EF186431	*R. exulans*	Samoa	Robins *et al*. 2007
EF186432	*R. exulans*	Thailand	Robins *et al*. 2007
EF186435	*R. fuscipes*	Australia	Robins *et al*. 2007
EF186436	*R. fuscipes*	Australia	Robins *et al*. 2007
EF186437	*R. fuscipes*	Australia	Robins *et al*. 2007
EF186438	*R. fuscipes*	Australia	Robins *et al*. 2007

EF186439	*R. fuscipes*	Australia	Robins *et al*. 2007
EF186440	*R. tanezumi*	China	Robins *et al*. 2007
EF186441	*R. hoffmanni*	Indonesia	Robins *et al*. 2007
EF186442	*R. hoffmanni*	Indonesia	Robins *et al*. 2007
EF186447	*R. leucopus*	Australia	Robins *et al*. 2007
EF186448	*R. leucopus*	Australia	Robins *et al*. 2007
EF186451	*R. leucopus*	Australia	Robins *et al*. 2007
EF186452	*R. leucopus*	Australia	Robins *et al*. 2007
EF186454	*R. novaeguineae*	Papua New Guinea	Robins *et al*. 2007
EF186461	*R. norvegicus*	French Polynesia	Robins *et al*. 2007
EF186462	*R. norvegicus*	French Polynesia	Robins *et al*. 2007
EF186469	*R. rattus*	French Polynesia	Robins *et al*. 2007
EF186470	*R. rattus*	New Zealand	Robins *et al*. 2007
EF186477	*R. sordius*	Australia	Robins *et al*. 2007
EF186478	*R. sordius*	Australia	Robins *et al*. 2007
EF186479	*R. sordius*	Australia	Robins *et al*. 2007
EF186491	*R. tanezumi*	Indonesia	Robins *et al*. 2007
EF186493	*R. tanezumi*	Indonesia	Robins *et al*. 2007
EF186494	*R. tanezumi*	Indonesia	Robins *et al*. 2007
EF186502	*R. tanezumi*	Indonesia	Robins *et al*. 2007
EF186508	*R. tanezumi*	Japan	Robins *et al*. 2007
EF186513	*R. tiomanicus*	Indonesia	Robins *et al*. 2007
EF186515	*R. tunneyi*	Australia	Robins *et al*. 2007
EF186516	*R. tunneyi*	Australia	Robins *et al*. 2007
EF186517	*R. tunneyi*	Australia	Robins *et al*. 2007
EF186518	*R. tunneyi*	Australia	Robins *et al*. 2007
EU273707	*R. rattus*	?	Robins *et al*. 2008
EU273708	*R. praetor*	?	Robins *et al*. 2008
EU273709	*R. exulans*	?	Robins *et al*. 2008
EU273710	*R. exulans*	?	Robins *et al*. 2008
EU273711	*R. exulans*	?	Robins *et al*. 2008
EU273712	*R. tanezumi*	?	Robins *et al*. 2008
FJ842265	*R. tanezumi*	South Africa	Bastos *et al*. 2011
FJ842268	*R. rattus*	South Africa	Bastos *et al*. 2011
FJ842274	*R. norvegicus*	South Africa	Bastos *et al*. 2011
FJ842275	*R. norvegicus*	South Africa	Bastos *et al*. 2011
FJ842277	*R. norvegicus*	South Africa	Bastos *et al*. 2011
FJ842278	*R. norvegicus*	Vietnam	Bastos *et al*. 2011
FJ842279	*R. norvegicus*	Indonesia	Bastos *et al*. 2011
FJ897498	*R. rattus*	Senegal	Bastos *et al*. 2011
FJ897499	*R. rattus*	Senegal	Bastos *et al*. 2011
FJ897500	*R. rattus*	Guadeloupe	Bastos *et al*. 2011
FJ897501	*R. rattus*	Guadeloupe	Bastos *et al*. 2011
FR775875	*R. argentiventer*	Vietnam	Balakirev & Rozhnov 2012
FR775878	*R. argentiventer*	Vietnam	Balakirev & Rozhnov 2012
FR775882	*R. argentiventer*	Vietnam	Balakirev & Rozhnov 2012
GQ891569	*R. rattus*	India/Mudumalai	Tollenaere *et al*. 2010
GQ891570	*R. rattus*	India/Mudumalai	Tollenaere *et al*. 2010
GQ891571	*R. rattus*	India/Mudumalai	Tollenaere *et al*. 2010
GQ891572	*R. rattus*	India/Avallanchi	Tollenaere *et al*. 2010
GQ891573	*R. rattus*	India/Attur	Tollenaere *et al*. 2010
GQ891574	*R. rattus*	Oman/Arazat	Tollenaere *et al*. 2010
GQ891575	*R. rattus*	Oman/Tibraq	Tollenaere *et al*. 2010
GQ891576	*R. rattus*	Oman/Tibraq	Tollenaere *et al*. 2010
GQ891577	*R. rattus*	Oman/Sahanout	Tollenaere *et al*. 2010
GQ891578	*R. rattus*	Oman/Sahanout	Tollenaere *et al*. 2010
GQ891579	*R. rattus*	Oman/Sahanout	Tollenaere *et al*. 2010
GQ891580	*R. rattus*	Oman/Sahanout	Tollenaere *et al*. 2010
GQ891581	*R. rattus*	Yemen	Tollenaere *et al*. 2010
GQ891582	*R. rattus*	Yemen	Tollenaere *et al*. 2010
GQ891583	*R. rattus*	Ethiopia	Tollenaere *et al*. 2010
GQ891584	*R. rattus*	Ethiopia	Tollenaere *et al*. 2010
GQ891585	*R. rattus*	Tanzania	Tollenaere *et al*. 2010
GQ891586	*R. rattus*	Tanzania	Tollenaere *et al*. 2010
GQ891587	*R. rattus*	Tanzania	Tollenaere *et al*. 2010

GQ891588	*R. rattus*	Mozambique	Tollenaere *et al*. 2010
GQ891589	*R. rattus*	Mozambique	Tollenaere *et al*. 2010
GQ891590	*R. rattus*	Mozambique	Tollenaere *et al*. 2010
GQ891591	*R. rattus*	Grande Comore	Tollenaere *et al*. 2010
GQ891592	*R. rattus*	Grande Comore	Tollenaere *et al*. 2010
GQ891593	*R. rattus*	Grande Comore	Tollenaere *et al*. 2010
GQ891594	*R. rattus*	Grande Comore	Tollenaere *et al*. 2010
GQ891595	*R. rattus*	Grande Comore	Tollenaere *et al*. 2010
GQ891596	*R. rattus*	Grande Comore	Tollenaere *et al*. 2010
GQ891597	*R. rattus*	Mayotte	Tollenaere *et al*. 2010
GQ891598	*R. rattus*	Mayotte	Tollenaere *et al*. 2010
GQ891599	*R. rattus*	Madagascar	Tollenaere *et al*. 2010
GQ891600	*R. rattus*	Madagascar	Tollenaere *et al*. 2010
GQ891601	*R. rattus*	Madagascar	Tollenaere *et al*. 2010
GQ891602	*R. rattus*	Madagascar	Tollenaere *et al*. 2010
GQ891603	*R. rattus*	Madagascar	Tollenaere *et al*. 2010
GQ891604	*R. rattus*	Madagascar	Tollenaere *et al*. 2010
GQ891605	*R. rattus*	Madagascar	Tollenaere *et al*. 2010
GQ891606	*R. rattus*	Madagascar	Tollenaere *et al*. 2010
GQ891607	*R. rattus*	Reunion	Tollenaere *et al*. 2010
GU570659	*R. leucopus*	Australia	Robins *et al*. 2010
GU570660	*R. leucopus*	Papua New Guinea	Robins *et al*. 2010
GU570661	*R. lutreolus*	Australia	Robins *et al*. 2010
GU570662	*R. tunneyi*	Australia	Robins *et al*. 2010
GU570663	*R. villosissimus*	Australia	Robins *et al*. 2010
GU570664	*R. fuscipes*	Australia	Robins *et al*. 2010
GU570665	*R. sordidus*	Australia	Robins *et al*. 2010
HM217362	*R. argentiventer*	Cambodia	Pages *et al*. 2010
HM217365	*R. rattus*	Tanzania	Pages *et al*. 2010
HM217366	*R. rattus*	Oman	Pages *et al*. 2010
HM217367	*R. rattus*	India	Pages *et al*. 2010
HM217368	*R. rattus*	Madagascar	Pages *et al*. 2010
HM217474	*R. nitidus*	Laos	Pages *et al*. 2010
HM217478	*R. nitidus*	Laos	Pages *et al*. 2010
HM217479	*R. nitidus*	Laos	Pages *et al*. 2010
HM217731	*R. rattus*	India/Mudumalai	Pages *et al*. 2010
HM217732	*R. rattus*	India/Mudumalai	Pages *et al*. 2010
HM217733	*R. rattus*	India/Kotagirir	Pages *et al*. 2010
HM217734	*R. rattus*	India/Kotagirir	Pages *et al*. 2010
HM217735	*R. rattus*	India/Mudumalai	Pages *et al*. 2010
HM217737	*R. rattus*	India/Avallanchi	Pages *et al*. 2010
HM217738	*R. rattus*	India/Avallanchi	Pages *et al*. 2010
HM217740	*R. rattus*	India/Attur	Pages *et al*. 2010
HM217741	*R. rattus*	India/Attur	Pages *et al*. 2010
HQ157808	*R. rattus*	South Africa	Bastos *et al*. 2011
HQ157809	*R. rattus*	South Africa	Bastos *et al*. 2011
JF718276	*R. rattus*	Juan de Nova Island	Russell *et al*. 2011
JF718277	*R. rattus*	Glorieuse Island	Russell *et al*. 2011
JF718278	*R. rattus*	Europa Island	Russell *et al*. 2011
JF718279	*R. rattus*	Europa Island	Russell *et al*. 2011
JN675476	*Bandicota savilei*	Thailand	Aplin *et al*. 2011
JN675478	*Berylmys bowersi*	Thailand	Aplin *et al*. 2011
JN675479	*R. niniventer*	Thailand	Aplin *et al*. 2011
JN675480	*R. niniventer*	Indonesia	Aplin *et al*. 2011
JN675481	*R. andamanensis*	Cambodia	Aplin *et al*. 2011
JN675482	*R. andamanensis*	Vietnam	Aplin *et al*. 2011
JN675483	*R. andamanensis*	China	Aplin *et al*. 2011
JN675484	*R. andamanensis*	China	Aplin *et al*. 2011
JN675485	*R. andamanensis*	China	Aplin *et al*. 2011
JN675486	*R. andamanensis*	China	Aplin *et al*. 2011
JN675487	*R. andamanensis*	China	Aplin *et al*. 2011
JN675488	*R. argentiventer*	Cambodia	Aplin *et al*. 2011
JN675489	*R. argentiventer*	Vietnam	Aplin *et al*. 2011
JN675491	*R. argentiventer*	Vietnam	Aplin *et al*. 2011
JN675493	*R. argentiventer*	Indonesia	Aplin *et al*. 2011

JN675494	*R. argentiventer*	Indonesia	Aplin *et al*. 2011
JN675495	*R. baluensis*	Malaysia/Sabah	Aplin *et al*. 2011
JN675496	*R. exulans*	Thailand	Aplin *et al*. 2011
JN675497	*R. exulans*	Laos	Aplin *et al*. 2011
JN675498	*R. exulans*	Vietnam	Aplin *et al*. 2011
JN675499	*R. losea*	Cambodia	Aplin *et al*. 2011
JN675500	*R. losea*	Vietnam	Aplin *et al*. 2011
JN675501	*R. losea*	Vietnam	Aplin *et al*. 2011
JN675502	*R. losea*	Vietnam	Aplin *et al*. 2011
JN675504	*R. losea*	Vietnam	Aplin *et al*. 2011
JN675505	*R. losea*	Vietnam	Aplin *et al*. 2011
JN675506	*R. losea*	Vietnam	Aplin *et al*. 2011
JN675507	*R. losea*	Vietnam	Aplin *et al*. 2011
JN675508	*R. losea*	Taiwan	Aplin *et al*. 2011
JN675509	*R. nitidus*	Laos	Aplin *et al*. 2011
JN675510	*R. norvegicus*	Vietnam	Aplin *et al*. 2011
JN675511	*R. pyctoris*	Nepal	Aplin *et al*. 2011
JN675512	*R. pyctoris*	Nepal	Aplin *et al*. 2011
JN675513	*R. satarae*	India	Aplin *et al*. 2011
JN675514	*R. tiomanicus*	Indonesia/Java	Aplin *et al*. 2011
JN675515	*R. tiomanicus*	Indonesia/Java	Aplin *et al*. 2011
JN675516	*R. tiomanicus*	Indonesia/Borneo	Aplin *et al*. 2011
JN675517	*R. rattus*	France	Aplin *et al*. 2011
JN675518	*R. rattus*	Senegal	Aplin *et al*. 2011
JN675519	*R. rattus*	Guinea	Aplin *et al*. 2011
JN675520	*R. rattus*	Madagascar	Aplin *et al*. 2011
JN675521	*R. rattus*	Madagascar	Aplin *et al*. 2011
JN675523	*R. rattus*	Oman	Aplin *et al*. 2011
JN675524	*R. rattus*	Iran	Aplin *et al*. 2011
JN675525	*R. rattus*	India	Aplin *et al*. 2011
JN675526	*R. rattus*	India	Aplin *et al*. 2011
JN675527	*R. rattus*	India	Aplin *et al*. 2011
JN675528	*R. rattus*	India	Aplin *et al*. 2011
JN675529	*R. rattus*	India	Aplin *et al*. 2011
JN675530	*R. rattus*	India	Aplin *et al*. 2011
JN675532	*R. rattus*	India	Aplin *et al*. 2011
JN675534	*R. rattus*	Japan	Aplin *et al*. 2011
JN675535	*R. rattus*	Papua New Guinea	Aplin *et al*. 2011
JN675541	*R. rattus*	USA/California	Aplin *et al*. 2011
JN675542	*R. rattus*	USA/California	Aplin *et al*. 2011
JN675545	*R. rattus*	USA/California	Aplin *et al*. 2011
JN675546	*R. rattus*	Brazil	Aplin *et al*. 2011
JN675548	*R. rattus*	Guyana	Aplin *et al*. 2011
JN675549	*R. rattus*	Guyana	Aplin *et al*. 2011
JN675550	*R. rattus*	Brazil	Aplin *et al*. 2011
JN675551	*R. rattus*	Brazil	Aplin *et al*. 2011
JN675552	*R. rattus*	Madagascar	Aplin *et al*. 2011
JN675553	*R. rattus*	Madagascar	Aplin *et al*. 2011
JN675554	*R. rattus*	Bangladesh	Aplin *et al*. 2011
JN675555	*R. rattus*	Bangladesh	Aplin *et al*. 2011
JN675556	*R. rattus*	Myanmar	Aplin *et al*. 2011
JN675557	*R. rattus*	Myanmar	Aplin *et al*. 2011
JN675558	*R. rattus*	Myanmar	Aplin *et al*. 2011
JN675559	*R. rattus*	Thailand	Aplin *et al*. 2011
JN675560	*R. rattus*	Thailand	Aplin *et al*. 2011
JN675561	*R. rattus*	Thailand	Aplin *et al*. 2011
JN675562	*R. rattus*	Thailand	Aplin *et al*. 2011
JN675563	*R. rattus*	Laos	Aplin *et al*. 2011
JN675564	*R. rattus*	Laos	Aplin *et al*. 2011
JN675565	*R. rattus*	Laos	Aplin *et al*. 2011
JN675566	*R. rattus*	Laos	Aplin *et al*. 2011
JN675567	*R. rattus*	Laos	Aplin *et al*. 2011
JN675568	*R. rattus*	Laos	Aplin *et al*. 2011
JN675569	*R. rattus*	Laos	Aplin *et al*. 2011
JN675570	*R. rattus*	Laos	Aplin *et al*. 2011

JN675571	*R. rattus*	Laos	Aplin *et al*. 2011
JN675572	*R. rattus*	Vietnam	Aplin *et al*. 2011
JN675573	*R. rattus*	Vietnam	Aplin *et al*. 2011
JN675575	*R. rattus*	China	Aplin *et al*. 2011
JN675576	*R. rattus*	China	Aplin *et al*. 2011
JN675577	*R. rattus*	China	Aplin *et al*. 2011
JN675578	*R. rattus*	China	Aplin *et al*. 2011
JN675579	*R. rattus*	China	Aplin *et al*. 2011
JN675580	*R. rattus*	Indonesia/Java	Aplin *et al*. 2011
JN675581	*R. rattus*	Indonesia/Java	Aplin *et al*. 2011
JN675582	*R. rattus*	Indonesia/Java	Aplin *et al*. 2011
JN675583	*R. rattus*	Indonesia	Aplin *et al*. 2011
JN675584	*R. rattus*	Indonesia	Aplin *et al*. 2011
JN675585	*R. rattus*	Philippines	Aplin *et al*. 2011
JN675586	*R. rattus*	Philippines	Aplin *et al*. 2011
JN675587	*R. rattus*	Philippines	Aplin *et al*. 2011
JN675588	*R. rattus*	Taiwan	Aplin *et al*. 2011
JN675589	*R. rattus*	Taiwan	Aplin *et al*. 2011
JN675590	*R. rattus*	Taiwan	Aplin *et al*. 2011
JN675591	*R. rattus*	Taiwan	Aplin *et al*. 2011
JN675592	*R. rattus*	Japan	Aplin *et al*. 2011
JN675593	*R. rattus*	Japan	Aplin *et al*. 2011
JN675595	*R. rattus*	Papua New Guinea	Aplin *et al*. 2011
JN675597	*R. rattus*	USA/California	Aplin *et al*. 2011
JN675599	*R. rattus*	Nepal	Aplin *et al*. 2011
JN675600	*R. rattus*	Pakistan	Aplin *et al*. 2011
JN675601	*R. rattus*	Pakistan	Aplin *et al*. 2011
JN675602	*R. rattus*	Pakistan	Aplin *et al*. 2011
JN675603	*R. rattus*	Sri Lanka	Aplin *et al*. 2011
JN675604	*R. rattus*	Laos	Aplin *et al*. 2011
JN675605	*R. rattus*	Cambodia	Aplin *et al*. 2011
JN675606	*R. rattus*	Cambodia	Aplin *et al*. 2011
JN675607	*R. rattus*	Cambodia	Aplin *et al*. 2011
JN675608	*R. rattus*	Cambodia	Aplin *et al*. 2011
JN675609	*R. rattus*	Cambodia	Aplin *et al*. 2011
JN675610	*R. rattus*	Cambodia	Aplin *et al*. 2011
JN675611	*R. rattus*	Vietnam	Aplin *et al*. 2011
JN675612	*R. rattus*	Vietnam	Aplin *et al*. 2011
JN675613	*R. rattus*	Vietnam	Aplin *et al*. 2011
JN675614	*R. rattus*	Vietnam	Aplin *et al*. 2011
JN675615	*R. rattus*	Indonesia/Java	Aplin *et al*. 2011
JN675616	*R. rattus*	Indonesia	Aplin *et al*. 2011
JN675617	*R. rattus*	Indonesia	Aplin *et al*. 2011
JN675618	*R. rattus*	Indonesia	Aplin *et al*. 2011
JN675619	*R. rattus*	Indonesia	Aplin *et al*. 2011
JN675621	*R. rattus*	Indonesia	Aplin *et al*. 2011
JN675622	*R. rattus*	Indonesia	Aplin *et al*. 2011
JN675624	*R. rattus*	Philippines	Aplin *et al*. 2011
JN675625	*R. rattus*	Laos	Aplin *et al*. 2011
JN675626	*R. rattus*	Laos	Aplin *et al*. 2011
JN675627	*R. rattus*	Thailand	Aplin *et al*. 2011
JN675628	*R. rattus*	Thailand	Aplin *et al*. 2011
NC001665	*R. norvegicus*	?	McLeod *et al*. 2006
X14848	*R. norvegicus*	?	Grosskopf & Feldmann 1981

7.2.3 Chapter 4: *Hemidactylus frenatus*

Sample Nb	Lab Code	Species	Provided by	Date	Location	Material	DNA
18	AT029	*H. frenatus*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
32	AT030	*H. frenatus*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	
34	AT031	*H. frenatus*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
35	AT032	*H. frenatus*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
36	AT033	*H. frenatus*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	
37	AT034	*H. frenatus*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
38	AT035	*H. frenatus*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
39	AT036	*H. frenatus*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
40	AT037	*H. frenatus*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
44	AT038	*H. frenatus*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
48	AT039	*H. frenatus*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
50	AT040	*H. frenatus*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
60	AT041	*H. frenatus*	Dr. J. Chris Hillman	2009	Diego Garcia Island	tail tip - Ethanol	x
211991	AT072	*H. frenatus*	Field Museum	1967	USA/Hawaii/Oahu	tissue - Formalin	
134893	AT073	*H. frenatus*	Field Museum	1959	Andaman Islands	tissue - Formalin	
134889	AT074	*H. frenatus*	Field Museum	1958	India/Assam/Goalpara	tissue - Formalin	
68619	AT075	*H. frenatus*	Field Museum	1952	Mexico/Guerrero	tissue - Formalin	
52054	AT076	*H. frenatus*	Field Museum	1946	Philippines/Mindanao	tissue - Formalin	
60702	AT077	*H. frenatus*	Field Museum	1948	Mariana Island/Guam	tissue - Formalin	
81931	AT078	*H. frenatus*	Field Museum	1957	Wake Island	tissue - Formalin	
122277	AT079	*H. frenatus*	Field Museum	?	Ceylon	tissue - Formalin	
142109	AT080	*H. frenatus*	Field Museum	1963	Moluccas/Ambon I	tissue - Formalin	
USNM 333169	AT081	*H. frenatus*	Smithsonian	?	Republic of Fiji	tissue - Ethanol	x
USNM 333837	AT082	*H. frenatus*	Smithsonian	?	Vanuatu	tissue - Ethanol	x
USNM 518589	AT083	*H. frenatus*	Smithsonian	?	Hawaii	tissue - Ethanol	x
USNM 533260	AT084	*H. frenatus*	Smithsonian	?	Solomon Islands	tissue - Ethanol	x
USNM 533644	AT085	*H. frenatus*	Smithsonian	?	Cook Islands	tissue - Ethanol	x
USNM 571541	AT086	*H. frenatus*	Smithsonian	?	Guam/Pacific Ocean	tissue - Ethanol	x
USNM 54656	AT087	*H. frenatus*	Smithsonian	?	Philippines	tissue - Ethanol	x
USNM 224691	AT088	*H. frenatus*	Smithsonian	?	Palau/Pacific Ocean	tissue - Ethanol	x
122243	AT089	*H. frenatus*	Field Museum	1944	Java/Batavia	tissue - Formalin	
14184	AT090	*H. frenatus*	Field Museum	1929	New Guinea/Manokwari	tissue - Formalin	
42406	AT091	*H. frenatus*	Field Museum	1943	India/Bengal	tissue - Formalin	
52137	AT092	*H. frenatus*	Field Museum	1946	Philippines/Palawan	tissue - Formalin	x
96860	AT093	*H. frenatus*	Field Museum	1958	North Taipei	tissue - Formalin	
125027	AT094	*H. frenatus*	Field Museum	1954	Thailand/Nakon Phanom	tissue - Formalin	
76242	AT095	*H. frenatus*	Field Museum	1956	North Borneo/Tawau	tissue - Formalin	
USNM 524049	AT116	*H. frenatus*	Smithsonian	1998	Burma/Sagaing	tissue - Ethanol	
USNM 536091	AT117	*H. frenatus*	Smithsonian	1999	Mariana Island/USA	tissue - Ethanol	x
USNM 536542	AT118	*H. frenatus*	Smithsonian	1999	Hawaii/Lanai City/USA	tissue - Ethanol	x
USNM 571987	AT119	*H. frenatus*	Smithsonian	2000	Caroline Islands/Micronesia	tissue - Ethanol	
USNM 36101	AT120	*H. frenatus*	Smithsonian	?	Australia	tissue - Ethanol	x
USNM 56822	AT121	*H. frenatus*	Smithsonian	?	Philippines	tissue - Ethanol	x
USNM 220439	AT122	*H. frenatus*	Smithsonian	?	Palau	tissue - Ethanol	x
1957.1.10.77	AT142	*H. frenatus*	BNHM/Tracy Heath	?	Maldives	tissue - Ethanol	
1956.1.11.31	AT143	*H. frenatus*	BNHM/Tracy Heath	?	Singapore	tissue - Ethanol	
1871.7.20.180	AT144	*H. frenatus*	BNHM/Tracy Heath	?	Indonesia	tissue - Ethanol	
1896.9.24.4	AT145	*H. frenatus*	BNHM/Tracy Heath	?	Somalia	tissue - Ethanol	
1927.5.20.16	AT146	*H. frenatus*	BNHM/Tracy Heath	?	Vietnam	tissue - Ethanol	
1860.3.19.1051/2	AT147	*H. frenatus*	BNHM/Tracy Heath	?	Malaysia	tissue - Ethanol	
1895.9.11.1	AT148	*H. frenatus*	BNHM/Tracy Heath	?	Borneo	tissue - Ethanol	
1956.1.13.40	AT149	*H. frenatus*	BNHM/Tracy Heath	?	Seychelles	tissue - Ethanol	
(18)59.7.1.28	AT150	*H. frenatus*	BNHM/Tracy Heath	?	Thailand	tissue - Ethanol	
1884.3.11.8	AT151	*H. frenatus*	BNHM/Tracy Heath	?	Taiwan	tissue - Ethanol	
1924.9.19.4	AT152	*H. frenatus*	BNHM/Tracy Heath	?	Vietnam	tissue - Ethanol	
1897.2.3.1	AT153	*H. frenatus*	BNHM/Tracy Heath	?	Malaysia	tissue - Ethanol	
1884.1.8.19	AT154	*H. frenatus*	BNHM/Tracy Heath	?	Indonesia	tissue - Ethanol	
(18)58.6.1.32	AT155	*H. frenatus*	BNHM/Tracy Heath	?	Mauritius	tissue - Ethanol	
1953.1.12.18	AT156	*H. frenatus*	BNHM/Tracy Heath	?	Seychelles	tissue - Ethanol	
1930.7.1.77	AT157	*H. frenatus*	BNHM/Tracy Heath	?	Madagascar	tissue - Ethanol	
1996.438	AT158	*H. frenatus*	BNHM/Tracy Heath	?	Chargos Archipelago	tissue - Ethanol	

1915.8.4.1	AT159	*H. frenatus*	BNHM/Tracy Heath	?	Thailand	tissue - Ethanol	
(18)70.6.7.30	AT160	*H. frenatus*	BNHM/Tracy Heath	?	Indonesia	tissue - Ethanol	
1959.1.6.23	AT161	*H. frenatus*	BNHM/Tracy Heath	?	Maldives	tissue - Ethanol	
St Helena 1	AT162	*H. frenatus*	Dr. J. Chris Hillman	2013	St Helena Island	dried tissue	**x**
St Helena 2	AT163	*H. frenatus*	Dr. J. Chris Hillman	2013	St Helena Island	dried tissue	**x**
St Helena 3	AT164	*H. frenatus*	Dr. J. Chris Hillman	2013	St Helena Island	dried tissue	**x**
St Helena 4	AT165	*H. frenatus*	Dr. J. Chris Hillman	2013	St Helena Island	dried tissue	**x**
St Helena 5	AT166	*H. frenatus*	Dr. J. Chris Hillman	2013	St Helena Island	dried tissue	**x**
St Helena 6	AT167	*H. frenatus*	Dr. J. Chris Hillman	2013	St Helena Island	dried tissue	
St Helena 7	AT168	*H. frenatus*	Dr. J. Chris Hillman	2013	St Helena Island	dried tissue	
St Helena 8	AT169	*H. frenatus*	Dr. J. Chris Hillman	2013	St Helena Island	dried tissue	**x**
St Helena 9	AT170	*H. frenatus*	Dr. J. Chris Hillman	2013	St Helena Island	dried tissue	**x**
St Helena 10	AT171	*H. frenatus*	Dr. J. Chris Hillman	2013	St Helena Island	dried tissue	**x**
St Helena 11	AT172	*H. frenatus*	Dr. J. Chris Hillman	2013	St Helena Island	dried tissue	**x**
St Helena 12	AT173	*H. frenatus*	Dr. J. Chris Hillman	2013	St Helena Island	dried tissue	**x**
St Helena 13	AT174	*H. frenatus*	Dr. J. Chris Hillman	2013	St Helena Island	dried tissue	**x**
Sealinks 1	AT175	?	Sealinks	2012	Mafia Island	dried tissue	
Sealinks 2	AT176	?	Sealinks	2012	Kilindoni, Mafia Island	dried tissue	
Sealinks 3	AT177	?	Sealinks	2012	Kilindoni, Mafia Island	dried tissue	
Sealinks 4	AT178	?	Sealinks	2012	Kilindoni, Mafia Island	dried tissue	
Sealinks 5	AT179	?	Sealinks	2012	Pemba Island	dried tissue	
Sealinks 6	AT180	?	Sealinks	2012	Pemba Island	dried tissue	
Sealinks 7	AT181	?	Sealinks	2012	Pemba Island	dried tissue	
Sealinks 8	AT182	?	Sealinks	2012	Kuumbi Cave, Zanzibar	dried tissue	
Sealinks 9	AT183	?	Sealinks	2012	Unguja Ukuu, Zanzibar	dried tissue	
Sealinks 10	AT184	?	Sealinks	2012	Unguja Ukuu, Zanzibar	dried tissue	
Sealinks 11	AT185	?	Sealinks	2012	Unguja Ukuu, Zanzibar	dried tissue	
Sealinks 12	AT186	?	Sealinks	2012	Unguja Ukuu, Zanzibar	dried tissue	
Sealinks A	AT187	?	Sealinks	2013	Madagascar	dried tissue	
Sealinks B	AT188	?	Sealinks	2013	Madagascar	dried tissue	**x**
Sealinks C	AT189	?	Sealinks	2013	Madagascar	dried tissue	
Sealinks D	AT190	?	Sealinks	2013	Madagascar	dried tissue	
Sealinks E	AT191	?	Sealinks	2013	Madagascar	dried tissue	
Sealinks F	AT192	?	Sealinks	2013	Madagascar	dried tissue	
Sealinks G	AT193	?	Sealinks	2013	Madagascar	dried tissue	
Sealinks H	AT194	?	Sealinks	2013	Madagascar	dried tissue	

NCBI Accession Nb	Species	Location	Author
AY217801	*H. frenatus*	Papua New Guinea	Whiting *et al*. 2003
DQ120277	*H. frenatus*	Hawaii	Carranza *et al*. 2006
DQ120278	*H. frenatus*	Hawaii	Carranza *et al*. 2006
DQ120279	*H. frenatus*	Colombia	Carranza *et al*. 2006
DQ120280	*H. frenatus*	Myanmar	Carranza *et al*. 2006
DQ120281	*H. frenatus*	Myanmar	Carranza *et al*. 2006
DQ120282	*H. frenatus*	India	Carranza *et al*. 2006
EU268389	*H. frenatus*	Sri Lanka	Bauer *et al*. 2008
EU268390	*H. frenatus*	Malaysia	Bauer *et al*. 2008
EU268391	*H. frenatus*	Sri Lanka	Bauer *et al*. 2008
GQ375289	*H. frenatus*	Malaysia	Bauer *et al*. 2010
HM595655	*H. frenatus*	India	Bansal & Karanth 2010

7.2.4 Chapter 5: *Bubalus mindorensis*

Sample Nb	Lab Code	Species	Provided by	Date	Location	Material	DNA
GNM4366	AT115/Sample 9	*B. mindorensis*	Goteborg NHM	?	Philippines, SW Mindoro	bone	x
216	AT134	*B. mindorensis*	Field Museum	?	Philippines, SW Mindoro	dried tissue	x
217	AT135	*B. mindorensis*	Field Museum	?	Philippines, SW Mindoro	dried tissue	
218	AT136	*B. mindorensis*	Field Museum	?	Philippines, SW Mindoro	dried tissue	
18979	AT137	*B. mindorensis*	Field Museum	?	Philippines, SW Mindoro	dried tissue	
43300	AT138	*B. mindorensis*	Field Museum	1935	Philippines, SW Mindoro	dried tissue	x
43301	AT139	*B. mindorensis*	Field Museum	1935	Philippines, SW Mindoro	dried tissue	x
43302	AT140	*B. mindorensis*	Field Museum	1935	Philippines, SW Mindoro	dried tissue	x
43303	AT141/Sample 11	*B. mindorensis*	Field Museum	1935	Philippines, SW Mindoro	dried tissue	x
AAM4001	LG373	*Anoa*	National Museum/Scotland	?	Philippines	bone	x
AAM4002	GL1252	*Anoa*	Goteborg NHM	1919	Philippines	bone	x
AAM4003	GL1253	*Anoa*	Goteborg NHM	?	Philippines	bone	x
AAM4005	GL1255	*Anoa*	Goteborg NHM	1918	Philippines	bone	x
AAM4006	GL1256	*Anoa*	Goteborg NHM	1918	Philippines	bone	x
AAM4008	GL1258	*B. depressicornis*	Goteborg NHM	1918	Philippines	bone	x
AAM4009	GL1259	*Anoa*	Goteborg NHM	1917	Philippines	bone	x
AAM4011	GL1261	*Anoa*	Goteborg NHM	1917	Philippines	bone	x
AAM4013	GL1263	*B. depressicornis*	Goteborg NHM	1917	Philippines	bone	x
AAM202	5	*Anoa*	modern	modern	Philippines	tissue	x
AAM204	7	*Anoa*	modern	modern	Philippines	tissue	x
AAM236	144	*Anoa*	modern	modern	Philippines	tissue	x
AAM241	65	*Anoa*	modern	modern	Philippines	hair	x
AAM242	66	*Anoa*	modern	modern	Philippines	tissue	x
AAM243	69	*Anoa*	modern	modern	Philippines	tissue	x
AAM244	96	*Anoa*	modern	modern	Philippines	tissue	x
AAM245	97	*Anoa*	modern	modern	Philippines	tissue	x
AAM246	98	*Anoa*	modern	modern	Philippines	tissue	x
AAM247	99	*Anoa*	modern	modern	Philippines	tissue	x
AAM248	100	*Anoa*	modern	modern	Philippines	tissue	x
AAM252	106	*Anoa*	modern	modern	Philippines	tissue	x
AAM255	120	*Anoa*	modern	modern	Philippines	tissue	x
AAM257	129	*Anoa*	modern	modern	Philippines	hair	x
AAM260	173	*Anoa*	modern	modern	Philippines	tissue	x
AAM267	184	*Anoa*	modern	modern	Philippines	tissue	x
AAM272	191	*Anoa*	modern	modern	Philippines	hair	x
AAM277	163	*Anoa*	modern	modern	Philippines	tissue	x
AAM281	167	*Anoa*	modern	modern	Philippines	tissue	x
AAM284	170	*Anoa*	modern	modern	Philippines	tissue	x
AAM285	171	*Anoa*	modern	modern	Philippines	tissue	x
AAM289	110	*Anoa*	modern	modern	Philippines	tissue	x
AAM291	112	*Anoa*	modern	modern	Philippines	tissue	x
AAM292	113	*Anoa*	modern	modern	Philippines	tissue	x
AAM293	114	*Anoa*	modern	modern	Philippines	tissue	x
AAM296	122	*Anoa*	modern	modern	Philippines	tissue	x
AAM299	125	*Anoa*	modern	modern	Philippines	tissue	x
AAM302	128	*Anoa*	modern	modern	Philippines	tissue	x
AAM305	134	*Anoa*	modern	modern	Philippines	hair	x
AAM306	135	*Anoa*	modern	modern	Philippines	hair	x
AAM315	41	*Anoa*	modern	modern	Philippines	tissue	x
AAM323	60	*Anoa*	modern	modern	Philippines	hair	x
AAM328	209	*Anoa*	modern	modern	Philippines	hair	x
AAM331	212	*Anoa*	modern	modern	Philippines	hair	x
AAM603	399	*Anoa*	Zoo	modern	Philippines	?	x
AAM606	70	*Anoa*	Zoo	modern	Philippines	?	x
AAM609	78	*Anoa*	Zoo	modern	Philippines	?	x
AAM611	287	*B. quarlesi*	Zoo	modern	Philippines	hair	x
AAM612	299	*B. depressicornis*	Zoo	modern	Philippines	swab	x
AAM613	360	*B. quarlesi*	Zoo	modern	Philippines	hair	x
AAM614	365	*B. depressicornis*	Zoo	modern	Philippines	swab	x

AAM615	389	*Anoa*	Zoo	modern	Philippines	?	**x**
AAM616	396	*B. quarlesi*	Zoo	modern	Philippines	hair	**x**
AAM620	448	*B. depressicornis*	Zoo	modern	Philippines	hair	**x**
AAM625	562	*B. depressicornis*	Zoo	modern	Philippines	blood	**x**
AAM629	429	*B. depressicornis*	Zoo	modern	Philippines	swab	**x**
AAM630	567	*B. depressicornis*	Zoo	modern	Philippines	swab	**x**
AAM631	453	*B. depressicornis*	Zoo	modern	Philippines	hair	**x**

NCBI Accession Nb	Species	Location	Author
AB526220	*B. mindorensis*	?	Ishihara *et al*. 2010
AB529514	*B. bubalis*	?	Ishihara *et al*. 2010
AB542189	*B. frontalis*	Bhutan	Tanaka *et al*. 2011
AB542190	*B. frontalis*	Myanmar	Tanaka *et al*. 2011
AF036273	*B. bison*	?	Hassanin *et al*. 1999
AF036275	*S. caffer*	?	Hassanin *et al*. 1999
AY079126	*B. bonasus*	?	Verkaar *et al*. 2004
AY488491	*B. bubalis*	?	Parma *et al*. 2004
AY534338	*S. caffer*	?	Kimwele *et al*. 2004
AY534339	*B. taurus*	?	Kimwele *et al*. 2004
AY689188	*B. javanicus*	?	Hassanin *et al*. 2004
AY689189	*B. sauveli*	?	Hassanin *et al*. 2004
AY702618	*B. bubalis* (Swamp Buffalo)	China	Qian *et al*. 2004
AY955225	*B. grunniens mutus*	?	Liu *et al*. 2005
BBUMTCB23	*B. arnee*	?	Chikuni *et al*. 1995
D32193	*B. arnee bubalis*	?	Chikuni *et al*. 1995
D34637	*B. bubalis* (Swamp Buffalo)	?	Kikkawa *et al*. 1997
D34638	*B. bubalis*	?	Kikkawa *et al*. 1997
D82890	*B. depressicornis*	Jakarta/Indonesia	Tanaka *et al*. 1996
D82891	*B. quarlesi*	Jakarta/Indonesia	Tanaka *et al*. 1996
D82892	*B. bubalis* (River Buffalo)	Bangladesh	Tanaka *et al*. 1996
D82893	*B. bubalis* (River Buffalo)	Sri Lanka	Tanaka *et al*. 1996
D82894	*B. bubalis* (Swamp Buffalo)	Indonesia	Tanaka *et al*. 1996
D82895	*B. mindorensis*	?	Tanaka *et al*. 1996
D88627	*B. bubalis*	?	Kikkawa *et al*. 1997
D88628	*B. bubalis*	?	Kikkawa *et al*. 1997
D88629	*B. bubalis*	?	Kikkawa *et al*. 1997
D88630	*B. bubalis*	?	Kikkawa *et al*. 1997
D88631	*B. bubalis*	?	Kikkawa *et al*. 1997
D88632	*B. bubalis* (River Buffalo)	?	Kikkawa *et al*. 1997
D88633	*B. bubalis*	?	Kikkawa *et al*. 1997
D88634	*B. bubalis*	?	Kikkawa *et al*. 1997
D88635	*B. bubalis*	?	Kikkawa *et al*. 1997
D88636	*B. bubalis*	?	Kikkawa *et al*. 1997
D88637	*B. bubalis*	?	Kikkawa *et al*. 1997
D88638	*B. bubalis*	?	Kikkawa *et al*. 1997
D88639	*B. depressicornis*	?	Kikkawa *et al*. 1997
D88640	*B. depressicornis quarlesi*	?	Kikkawa *et al*. 1997
D88641	*B. depressicornis fergusoni*	?	Kikkawa *et al*. 1997
D88642	*B. depressicornis*	?	Kikkawa *et al*. 1997
D88983	*B. bubalis* (River Buffalo)	?	Kikkawa *et al*. 1997
DQ124372	*B. taurus T4*	?	Shin & Kim 2005
DQ124374	*B. taurus T3*	?	Shin & Kim 2005
DQ124387	*B. taurus T3*	?	Shin & Kim 2005
DQ124389	*B. taurus P*	?	Shin & Kim 2005
DQ124399	*B. taurus T1*	?	Shin & Kim 2005
DQ124407	*B. taurus T3*	?	Shin & Kim 2005
DQ124408	*B. taurus T3*	?	Shin & Kim 2005
DQ124413	*B. taurus T3*	?	Shin & Kim 2005
DQ124414	*B. taurus T3*	?	Shin & Kim 2005
DQ124418	*B. taurus T3*	?	Shin & Kim 2005
DQ459330	*B. gaurus*	?	Chaichoune *et al*. 2006
DQ459331	*B. gaurus*	?	Chaichoune *et al*. 2006
EF382665	*B. sauveli*	?	Wajjwalku *et al*. 2007
EF536351	*B. depressicornis*	?	Hassanin *et al*. 2012

EF536353	*S. caffer*	?	Hassanin *et al*. 2012
EU177820	*B. taurus T3*	Italy	Achilli *et al*. 2008
EU177822	*B. taurus T3*	Italy	Achilli *et al*. 2008
EU177823	*B. taurus T3*	Italy	Achilli *et al*. 2008
EU177824	*B. taurus T3*	Italy	Achilli *et al*. 2008
EU177836	*B. taurus T3*	Italy	Achilli *et al*. 2008
EU177841	*B. taurus T1*	?	Achilli *et al*. 2008
EU177842	*B. taurus T1*	Italy	Achilli *et al*. 2008
EU177843	*B. taurus T1*	Italy	Achilli *et al*. 2008
EU177844	*B. taurus T1*	Italy	Achilli *et al*. 2008
EU177857	*B. taurus T2*	Iran	Achilli *et al*. 2008
EU177860	*B. taurus T2*	Iran	Achilli *et al*. 2008
EU177861	*B. taurus T2*	Italy	Achilli *et al*. 2008
EU177866	*B. taurus Q*	Italy	Achilli *et al*. 2008
EU177871	*B. bison*	?	Achilli *et al*. 2008
EU255782	*B. gaurus*	?	Riviere-Dobigny *et al*. 2007
EU255783	*B. javanicus*	?	Riviere-Dobigny *et al*. 2007
EU747737	*B. primigenius*	?	Slomski, unpublished
EU807955	*B. frontalis*	?	Geng & Chang, unpublished
EU807957	*B. frontalis*	?	Geng & Chang, unpublished
EU878389	*B. javanicus*	?	Chaichoune *et al*. 2008
FJ392912	*B. primigenius*	?	Stock *et al*. 2009
FJ785361	*B. taurus*	?	Kikkawa *et al*. 2009
FJ785366	*B. taurus*	?	Kikkawa *et al*. 2009
FJ785372	*S. caffer*	?	Kikkawa *et al*. 2009
FJ971083	*B. taurus*	Italy	Achilli *et al*. 2009
FJ971084	*B. taurus*	Italy	Achilli *et al*. 2009
FJ971087	*B. taurus*	Italy	Achilli *et al*. 2009
FJ971088	*B. taurus*	Mongolia	Achilli *et al*. 2009
GU256940	*B. indicus*	?	Wang *et al*. 2009
GU947004	*B. bison*	?	Douglas *et al*. 2011
GU985279	*B. primigenius*	?	Edwards *et al*. 2010
HQ122574	*B. bison*	?	Herrero *et al*. 2010
HQ184030	*B. taurus*	Italy	Bonfiglio *et al*. 2010
HQ184039	*B. taurus*	Italy	Bonfiglio *et al*. 2010
HQ184040	*B. taurus*	Italy	Bonfiglio *et al*. 2010
HQ223450	*B. bonasus*	?	Derr *et al*. 2010
JN117611	*B. indicus*	Pakistan	Hussain *et al*. 2011
JN117614	*B. indicus*	Pakistan	Hussain *et al*. 2011
JN632605	*B. javanicus*	?	Hassanin *et al*. 2012
JN817302	*B. taurus T1*	Ethiopia	Bonfiglio *et al*. 2012
JN817303	*B. taurus T1*	Ethiopia	Bonfiglio *et al*. 2012
JN817304	*B. taurus T1*	Ethiopia	Bonfiglio *et al*. 2012
JN817305	*B. taurus T1*	Ethiopia	Bonfiglio *et al*. 2012
JN817307	*B. taurus T1*	Mexico	Bonfiglio *et al*. 2012
JN817311	*B. taurus T1*	Uruguay	Bonfiglio *et al*. 2012
JN817316	*B. taurus T1*	Italy	Bonfiglio *et al*. 2012
JN817320	*B. taurus T1*	Italy	Bonfiglio *et al*. 2012
JN817326	*B. taurus T1*	Egypt	Bonfiglio *et al*. 2012
JN817329	*B. taurus T1*	Egypt	Bonfiglio *et al*. 2012
JN817336	*B. taurus T1*	Italy	Bonfiglio *et al*. 2012
JN817343	*B. taurus T1*	Italy	Bonfiglio *et al*. 2012
JN817346	*B. taurus T1*	Italy	Bonfiglio *et al*. 2012
JN817347	*B. taurus T1*	Italy	Bonfiglio *et al*. 2012
JN817351	*B. taurus T1*	Italy	Bonfiglio *et al*. 2012
KF163072	*B. taurus T1*	?	Horsburgh *et al*. 2013
KF163073	*B. taurus T1*	?	Horsburgh *et al*. 2013
KF163075	*B. taurus T1*	?	Horsburgh *et al*. 2013
NC005971	*B. indicus*	?	Miretti *et al*. 2004
NC006380	*B. grunniens*	?	San *et al*. 2007
NC014044	*B. bonasus*	?	Zeyland *et al*. 2012
Y15005	*B. bonasus*	?	Zimmermann *et al*. 1998
Y16064	*B. indicus*	?	Schreiber *et al*., unpublished
AF492350	*B. indicus*	?	Hiendleder *et al*. 2008

AF547270	*B. bubalis*	?	Verma *et al*. 2002
AY702618	*B. bubalis*	?	Qian *et al*. 2004
DQ124396	*B. taurus T2*	?	Shin & Kim 2005
DQ124412	*B. taurus T4*	?	Shin & Kim 2005
EU177847	*B. taurus T1*	?	Achilli *et al*. 2008
EU177865	*B. taurus T5*	?	Achilli *et al*. 2008
EU177868	*B. taurus*	?	Achilli *et al*. 2008
EU177869	*B. taurus*	?	Achilli *et al*. 2008
EU177870	*B. taurus*	?	Achilli *et al*. 2008
FJ971082	*B. taurus*	?	Achilli *et al*. 2009
FJ971086	*B. taurus*	?	Achilli *et al*. 2009
GQ464312	*B. grunniens*	?	Wang *et al*. 2010
JN632601	*B. bison*	?	Hassanin *et al*. 2012
JN817298	*B. taurus*	?	Bonfiglio *et al*. 2012
JN817306	*B. taurus*	?	Bonfiglio *et al*. 2012
JN817326	*B. taurus*	?	Bonfiglio *et al*. 2012
JN817343	*B. taurus*	?	Bonfiglio *et al*. 2012
JN817348	*B. taurus*	?	Bonfiglio *et al*. 2012
KF163094	*B. taurus*	?	Horsburgh *et al*. 2013
NC012706	*B. javanicus*	?	Lipinski *et al*. 2009
EF536352	*P. nghetinhensis (Saola)*	?	Hassanin et al. 2012