Phylogeography of two Sulawesi rodents: Testing the Effect of Climatic Variation on Population Structure

RALPH FRANCIS ROBERT ADRI WAINER

THE UNIVERSITY OF DURHAM

Phylogeography of two Sulawesi rodents: Testing the Effect of Climatic Variation on Population Structure

Ralph Wainer

Thesis for MSc by Research

Abstract

In order to support effective conservation and sustainable land management, it is important to understand the population structure and rates of population differentiation in lesser-known species. Biogeographers, evolutionary biologists and ecologist share an interest in learning about patterns of species diversity, and the identification of such a pattern (e.g. the effects of Pleistocene glacial cycles) can help predict levels of diversity according to the most likely evolutionary history of a species. A phylogenetic study focusing on the population genetics and phylogeography of two sympatric Indonesian murids (*Maxomys hellwaldi* & *Bunomys andrewsi*) was performed on different populations on the islands of Sulawesi and Buton. Indicators for population differentiation due to genetic structure were calculated using RAD-tagged Illumina sequencing data. Phylogenetic consensus trees were created using BEAST and MrBayes and dated on the basis of published references. Two hypotheses were tested, firstly that the putative populations are cryptic species, secondly that a defined geological event equally affected the evolutionary biology of both species of interest. A significant amount of differentiation was found between populations of each species on the different islands over a small geographic range. It was concluded that island colonisations during the early Pleistocene associated with the Pliocene to Pleistocene transition and its impact on sea level had resulted in incipient speciation in each genus in parallel.

Index

Chapter 1 - Introduction

1.1 Biota of Sulawesi

With over 20,000 islands, the Indo-Pacific is widely considered one of the most biotically diverse and complex areas on Earth (Lohman *et al*., 2011). These islands have varying degrees of isolation from each other and the from mainland of South-East Asia, which has made this region of particular interest to biogeographers who have studied the multiple processes that have formed patterns of biodiversity in this region (Gorog *et al*., 2003). Among these islands, Indonesia has the world's second highest level of biodiversity (Brown, 1997) and second highest count of total endemic species (Lambertini, 2000). Being the 4th most populated country (United Nations, 2013) the nation faces serious environmental issues, mainly due to pollution, rapid industrialization and illegal timber extraction (Curran, 2004). Large sections of tropical rainforest are in danger of being cleared (Sodhi *et al*., 2004) including lowland tropical forests which are considered exceptionally diverse (Lambert & Collar, 2002). Although local support for the conservation of protected areas in Indonesia has been documented with optimistic results (Walpole & Goodwin, 2001), there is an urgent need to conserve and study the country's biota. Woodruff (2010) provides a good case emphasizing the practical need to conserve tropical rainforests in South-East Asia. As 60% of Indonesia's population is concentrated on the island of Java (Monk *et al*., 1996), conservation projects are more common in the less populated outlying islands such as Sulawesi and Buton (Figure 1; adapted from Martin & Blackburn, 2010).

With a surface area of 189,000 km^2, Sulawesi (formerly Celebes) is the fourth largest island in the Indo-Australian archipelago and the largest island in the Indonesian biodiversity hotspot region termed Wallacea. Situated between the larger islands of Borneo and New Guinea, Sulawesi has a unique fauna composed of a mixture of taxa shared by neighbouring land masses, endemic forms and some wide-ranging faunal elements (Groves, 2001; Merker & Groves, 2006). As Sulawesi borders Wallace's faunal divide in the centre of the Indonesian archipelago, the island's

biogeography is of great interest due to its high level of species endemism (Moss & Wilson, 1998). The island's biodiversity is represented by nearly half of all known extant orders of non-marine placental mammals (Groves, 1976). Typical of islands found in Wallacea, there is a distinct biotic overlap of endemic Sulawesi mammals, which have evolved by vicariance, and continental dispersers, which originate from Asia (e.g. murines & squirrels) and Sahul (e.g. endemic phalangers) (Fabre *et al*., 2013).

Most studies on the biodiversity in the Wallacea region are focused on terrestrial mammals (den Tex *et al*., 2010). 30% of Sulawesi's mammalian fauna (Musser, 1991) and 52% of its endemic species are from the subfamily Murinae of the family Muridae (as defined by Carleton & Musser, 1984). The Murinae are composed of 368 extant species (Aplin & Helgen, 2010) and is the most diverse mammalian group in the Indo-Pacific. This group of mammals has long been the topic of extensive research and more genera are being described each year (e.g. *Halmaheramys*, Fabre *et al*., 2013; *Paucidentomys*, Esselstyn *et al*., 2012). As well as being the largest, Muridae is arguably the most complex and least understood mammalian sub-family (Alpin, 2003), which includes some of the most commonly used laboratory species such as *Mus musculis* and *Rattus norvegicus* (Musser & Carleton, 2005). Rodent systematists such as Musser (1987) and Medway & Yong (1976) have found the study of different phenotypes to be challenging, notably within the genus *Rattus*. The uncertainty surrounding the taxonomic classification is chiefly due to many species displaying plesiomorphies (characters which are present in both ancestors and outgroups) and mosaic convergences (the independent yet similar evolution of specific features in different species) (Medway & Yong, 1976). Previous systematic reviews (Musser & Carleton, 2005; Corbet & Hill, 1992) have also indicated large gaps in our general knowledge of this area's mammalian biodiversity. One of the most thoroughly sampled murid phylogenies to date can be found in Schenk *et al*. (2013).

Rodents in general have the highest diversification rates among mammalian taxonomic groups according to Heaney (1986). With over 50 species in Sulawesi, and 156 species in Indonesia,

the Rattini tribe encompasses the most diverse assemblage of murids in this region (Fabre *et al*., 2013). There are over 40 species of murid rodents that are native to Sulawesi (Musser, 1990). The island is regarded as 'a centre of an extraordinary radiation of Murid rodents' (Musser, 1987), yet there are substantial deficiencies in our understanding of rodent assemblages across the complex topography of the Indo-Pacific (Fabre *et al*., 2013). Moreover, many endemic rodent genera in Sulawesi are not represented by published DNA databases (Fabre *et al*., 2013) which has hindered extensive research of the area's biota. Sulawesi has only recently been colonized by murine rodents in comparison to other South-East Asian islands, such as the Philippine archipelago (Fabre *et al*., 2013). Sulawesi has lost approx. 20% of its forest between 1985 and 1997, with its remaining woodland being highly fragmented (Seymour, 2006). A number of invasive species (e.g. *R. exulans*, *R. norvegicus*) have also been introduced to Sulawesi, threatening the island's natural biota. Deforestation on this particular island has been shown to dramatically affect its endemic wildlife (Whitten *et al*., 2002; Rosenbaum *et al*., 1998; O'Brien & Kinnaird, 2000) yet little is known of its effect on rodents (Schulze *et al.*, 2004). It is certain however, in comparative terms, that the order Rodentia has lost the most species in the modern era (MacPhee & Fleming, 1999) with at least 80 insular rodent extinctions occurring on oceanic islands such as Sulawesi (Flannery, 1995). It is important to further study these Sulawesi rodents, partly because they are key members of a complex ecosystem in which they are essential as seed dispersers and/or seed predators (Vander Wall *et al*., 2005).

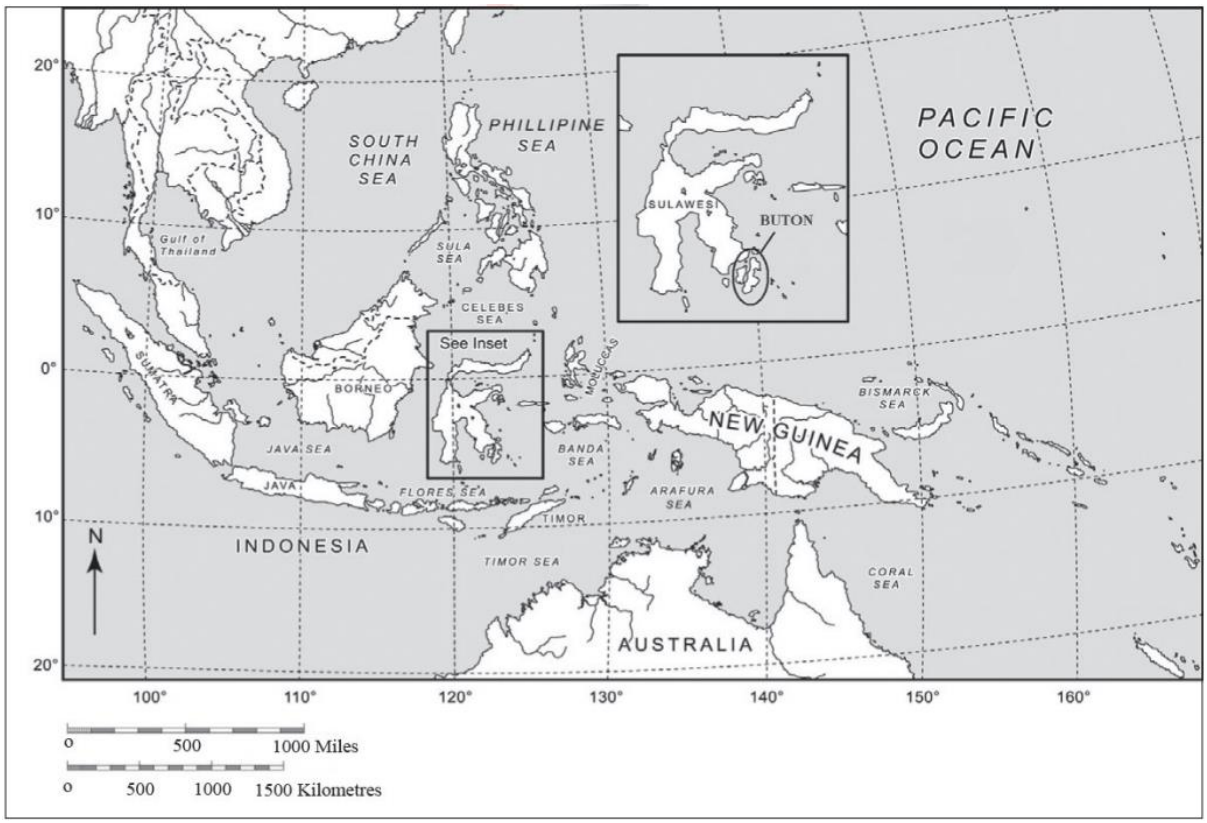


Figure 1. The Indonesian archipelago. Inset displays location of Sulawesi and Buton Island. Adapted from Martin & Blackburn (2010).

1.2. Geology of Sulawesi

45 million years ago, the northward moving Australo-Papuan tectonic plate collided with both the Pacific and Asian plates resulting in a prolonged interval of constant tectonic activity (Hall, 2002 & 2009). This activity caused the repeated submerging and emerging of multiple islands (Hall *et al.*, 2011; Hall, 1996). Beginning 5 million years ago during the Plio-Pleistocene, climate oscillations caused constant fluctuations in sea level which would have affected landmass configurations in South-East Asia. During the course of the last 3 Million years, repeated glacial and interglacial cycles have caused dramatic changes to the landform of Sulawesi, and indeed most of the Earth (Hewitt, 2004). Knowing when these sea-level fluctuations occurred is essential in studying the distribution of terrestrial biota (Darlington, 1957). Land area would have been significantly greater during periods

of low sea-level, which would facilitate the dispersal of terrestrial biota between islands (Lohman *et al*. 2011). A prime example of an increase in land area can be seen in the case of the Sunda Shelf close to Sulawesi, which had twice as much land area during the Pleistocene as today (Woodruff, 2010). Climate oscillations during the Pleistocene would have led to either the conjoining of separate islands or to a reduction in the distance between islands (Voris, 2000). The resulting 'land-bridges' formed between neighbouring islands would have facilitated the migration of terrestrial mammals such as rodents. Examples of studies which suggested that rodents specifically dispersed via land bridges can be found in Rowe *et al*. (2011) and Voris (2000). The island of Sulawesi in particular would have undergone considerable alterations in its habitable area, according to depth contour maps by Voris (2000). It has been documented how such changes in sea level has influenced the colonization of terrestrial mammals in the Indo-Pacific (Hall, 2009; Mercer & Roth, 2003). Voris (2000) demonstrated how the global changes in sea level would have predominantly affected the shallow ocean floors of the Indo-Pacific archipelago, where many islands, which are now separated by sea, were almost certainly formerly conjoined . Although this study focuses on Quaternary geological changes, geological reconstructions of the Indo-Australian Archipelago during the Neogene and Paleogene can be found in Lohman *et al.* (2011).

It is estimated that sea levels fell as much as 116m during the apex of the Last Glacial Maximum 22'000 years ago (Hanebuth *et al*., 2000). Other studies suggest that sea levels were as low as 120m below present levels during the Pleistocene (Gaither *et al*., 2010). The exposed shelf, termed Sunda Land, would have enabled a temporary connection between the Greater Sunda Islands (e.g. Sumatra) with continental Asia (Sathiamurthy & Voris, 2006). According to the historical sea-level transgression maps by Sathiamurthy & Voris (2006), which date from 4.2 Mya to 21 thousand years before present, the mainland of Sulawesi would not have been connected to the mainland of Asia. It is however possible that the mainland of Sulawesi would have connected to the small outlying islands such as Buton. Buton Island is located off the south-east coast of Sulawesi and is approximately 100km long with a width of 42 km at its widest point (Jennings *et al*., 2005). Around

Buton's coast, the altitude ranges from 0 to 200m. In the centre of Buton, there are peaks that rise to over 1000m. Buton shares the same tropical monsoon climate as Sulawesi (Whitten *et al*., 2002) with a mean annual rainfall between 1500-2000 mm (Whitten *et al*., 2002). At present, Buton is separated from mainland Sulawesi by approximately 5km of ocean at its closest point (Figure 2) with channels that are never more than 75m deep (Voris 2000). On the other side of Sulawesi, it is impossible that there would have been a connection to Borneo during the Last Glacial Maximum, as the Makassar Strait (between Sulawesi and Borneo) is in some sections too deep (Voris 2000, Groves, 2001). A sketched map showing varying degrees of current water depth in the Indonesian region can be found in Figure 3 (adapted from van den Bergh *et al.,* 2001). Voris (2000) estimated the current depth between Sulawesi and Buton to be between 50 and 70 meters .

Biogeography is a study which focuses on explaining observed patterns of species distribution within a temporal and spatial framework (Whittaker *et al*., 2005). The complicated biogeography of the Indonesian Islands has induced intricate patterns of variance and spatio-temporal dispersal among lineages. There are various opinions as to the relative importance of vicariance and dispersal in Sulawesi (Stelbrink *et al*., 2012; Lohman *et al*., 2011). Stelbrink *et al*. (2012) concluded that dispersal seemed to the most important mechanism of bringing taxa to Sulawesi and refuted the idea that vicariance caused diversification in some Sulawesi taxa, which had Continental Asian origins. Lohman *et al.* (2011) found that the present distribution patterns of Sulawesi's species were primarily shaped by pre-Pleistocene dispersal and vicariance events. Much of the vicariance/dispersal debate in the Indo-Australian Archipelago is centred on the origin of Sulawesi's taxa in particular (Lohman *et al*., 2011). Molecular analyses have disclosed that Sulawesi's taxa are often basal groups which cluster with Continental Asian taxa (Randi *et al*. 1996) or Australian taxa (Sparks & Smith, 2004). According to Fabre *et al*. (2013), no Murinae colonisations took place in the Indo-Pacific during a period of high sea level which ensued throughout the first half of the Pliocene. Although few studies have focused on the dispersal of Sulawesi rodents, research on South-East Asian phylogeny has been performed on the dispersal pattern of reptiles (Williams *et al.*,

2009) and terrestrial mammals such as marsupials (Macqueen *et al.*, 2009). The patterns of dispersal observed in those studies have been linked to changes in sea level, which would of course affect rodent colonisations as well.

Unlike many Wallacean and Philippine islands, which are usually regarded as oceanic landmasses as there was never any terrestrial connection to any surrounding land since their emergence (van Oosterzee, 1997), the biota of Sulawesi may possibly have vicariant origins from Sundaland or New Guinea (Wilson & Moss, 1999). Whereas the biota on an oceanic island would arise predominantly via dispersal, Sulawesi's biota is formed of a mixture of both vicariance and dispersal.

Phylogenetic approaches, which involve reconstructing historical relationships among populations or species, are commonplace in biogeographical publications. Through these means, researchers are able to understand more about the relative importance of dispersal events, extinction and speciation as the main factors behind current observed patterns. Of the three main factors mentioned, extinction stands out as the least understood factor due to a scarcity in fossil information (Raup, 1986). Insights from biogeographical studies can be put to good use in identifying areas of unique biological interest (Whittaker *et al*., 2005) and a more thorough understanding of the murid phylogeny would have useful implications for tracing phylogeographic relationships worldwide. Studying and revising the phylogeny of murids will help in understanding the evolutionary history of such a key component of mammalian evolution (Mercer & Roth, 2003). The key papers published on South-East Asia's rodent phylogenies are by Bryant *et al.* (2011), Rowe *et al.* (2011) and Mercer & Roth (2003). The most interesting point in Bryant *et al*. (2011) is a relatively recent split between the Papuan *Melomys* clade and *Rattus,* which occurred slightly more than 5 Mya. This split occurred somewhat recently when compared to other older nodes between a South-East Asian murid and *Rattus*, such as in Fabre *et al.* (2013) which will be discussed further on. Rowe *et al*. (2011) studied that the split between the murid *Uromys* and its closest relative *Pseudomys*

occurred during the middle to late Pliocene although their study focused on Australo-Papuan rodents, as opposed to the species of interest in this paper.

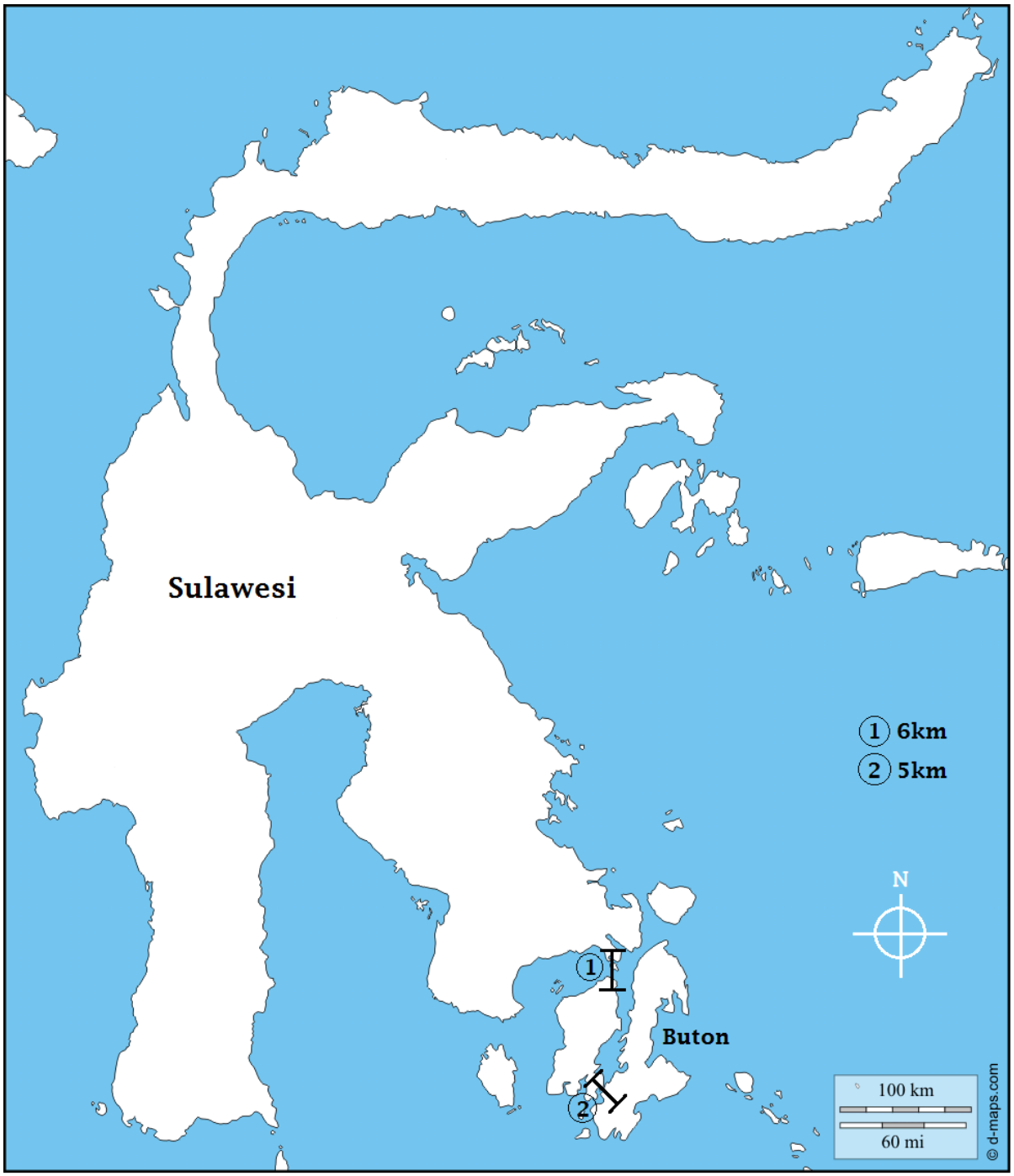


Figure 2. Map of present Buton Island (*Pulau Buton*) and mainland Sulawesi. Current distances between islands measured via Google Earth (Imagery ©2015 DigitalGlobe, TerraMetrics, CNES). Approximate minimal distances between islands shown as ① & ②. Adapted from *dmaps.com*.

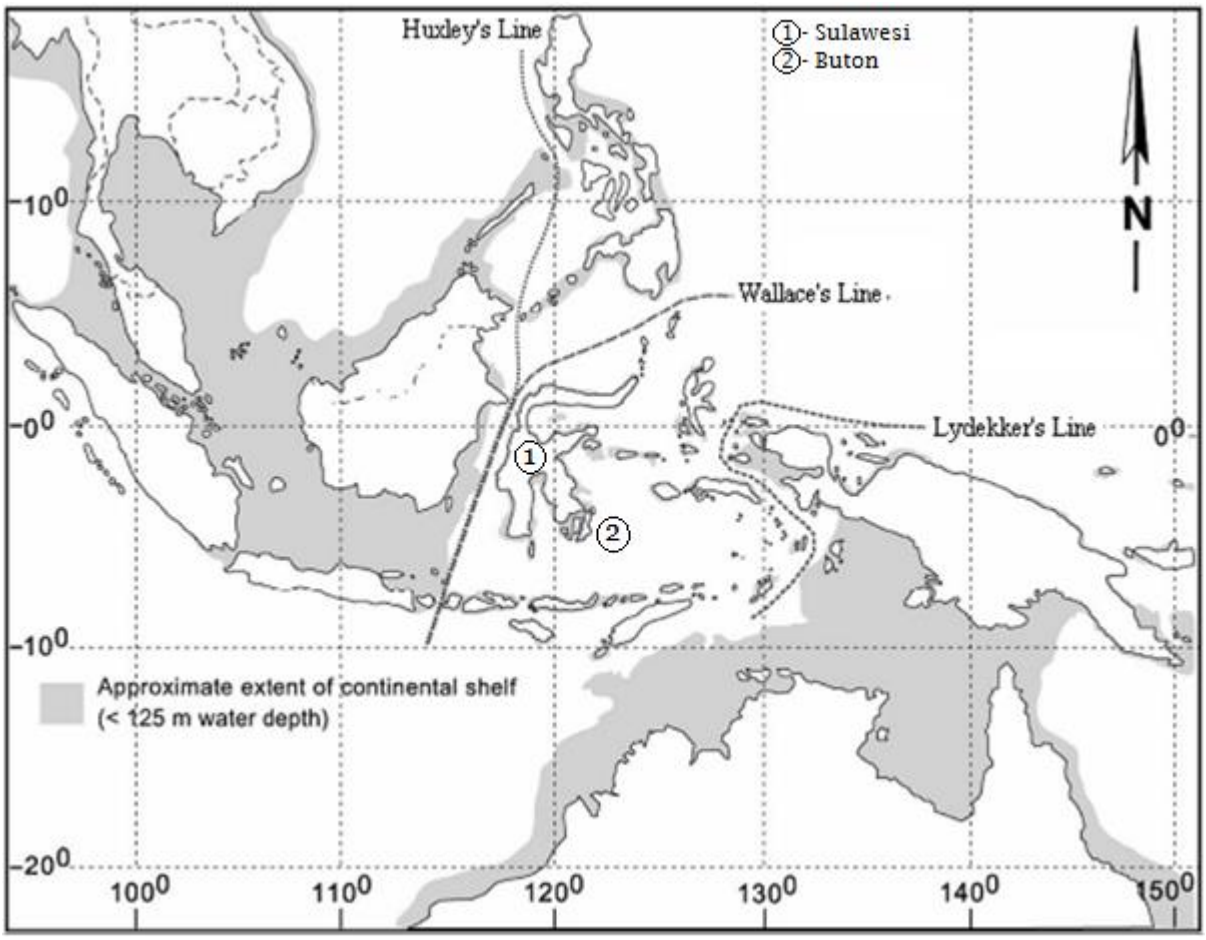


Figure 3. Map of the Indonesian region, showing 125m isobath and the most significant zoogeographic boundaries (Huxley's, Wallace's and Lydekker's Lines). Adapted from van den Bergh *et al.* (2001).

1.3. *Bunomys* & *Maxomys*

This study centres on 2 rodent species, namely *Bunomys andrewsi* (Allen, 1911) and *Maxomys hellwaldi* (Jentink, 1878). Although six genera are native to Sulawesi (Musser, 1987), *Maxomys* occurs throughout South-East Asia. The *Bunomys* genus is entirely confined to the forests of Sulawesi (Musser & Newcomb, 1983) with the possible exception of *Paulamys nasu*, a rodent found on the island of Flores, which is considered by some to be a member of the *Bunomys* genus (Kitchener & Yani, 1998; Corbet & Hill, 1992; Kitchener *et al.*, 1991). Some research such as that by Groves (2001) has already been carried out on the assumption that Flores is home to populations of

Bunomys. *Bunomys andrewsi* (BA) is found throughout the southwestern and southeastern peninsulas of Sulawesi as well as in its central mainland. At the present moment, there is no account as to what extent either of these focal rodent species inhabit the island of Buton. The most extensive paper which gives a systematic review of the *Bunomys* genus (Musser, 2014) only deals with specimen collected on mainland Sulawesi. Currently, there are no papers in the literature which specifically deal with the distribution of rodents on the island of Buton. Heanley *et al.* (2005) pointed out that rodent species in South-East Asia tend to be more widely distributed when their habitats were disturbed. Fortunately, rodent species from this region seem to be comparatively less affected by disturbances in their habitat when compared to other, usually larger species (Kitchener & Maryanto, 1994).

The habitat of BA overlaps with that of *Maxomys hellwaldi* (MH) (Ruedas, 2008). This overlap has been reconfirmed during recent fieldwork by Musser (2014). Both species also share the same ecological attributes (Musser, 2014): both are terrestrial, nocturnal, found in lowland evergreen tropical rain forest and feed on fruit, arthropods, snails, earthworms and vertebrates. An identification description for *Maxomys* can be found in Francis (2008). Approx. 39% of the *Maxomys* species are either classified by IUCN under 'Vulnerable' or 'Endangered' and approx. 42% of *Bunomys* are threatened, one being 'Critically Endangered' (Musser & Carleton, 2005). Several phylogenetic trees have been proposed for both species (Fabre *et al*., 2013; Verneau *et al.*, 1998; Rowe *et al*., 2008; Steppan *et al*., 2005; Geffen *et al*., 2011; Ruedas & Kirsch, 1997). The most recent of these phylogenies (Fabre *et al*. 2013) was chosen for node-dating in subsequent chapters. There is some debate as to the taxonomy of *Bunomys* and *Maxomys*, the taxonomy of *Maxomys* being particularly unstable (Achmadi *et al.,* 2013), although the two genera are clearly divergent. Only scarce literature exists on both BA and MH, existing literature being primarily focussed on their morphology and the parasitic species to which they are hosts (Durden, 1991 & 1986; Dewi & Hasegawa, 2010).

The taxonomic status of *Rattini* genera has been a consistent topic of debate and several murid lineages (which have previously been regarded as part of *Rattus*) have been reassigned to separate genera (Musser & Carleton, 2005; Musser & Newcomb, 1983). The debate surrounding the classification of the Rattini genera is due to complexities in delimiting species which have heterogeneity in rates of molecular evolution and can also display traits of morphological homoplasy (Jansa *et al*., 2006). Until relatively recently, the actual genus of MH was unknown. Ellerman (1961) and Marrison-Scott (1951) both classified *Maxomys* as one of 550 distinct forms the genus *Rattus,* until *Maxomys* was given its own genus by Misonne (1969). The closest relative to *Maxomys* is probably *Ratchaburimys ruchae*, an extinct rodent which has been described from early Pleistocene fossils found in South Thailand (Chaimanee, 1998).

Of the two genera, *Maxomys* has attracted the most scientific attention. According to Musser & Holden (1991), this genus is more diverse than any other rodent genus in the peninsula of Malaysia, the Sundra Shelf and its outlying islands. Currently, there are 17 species of *Maxomys* that are recognized, of which 14 originate in Indonesia (Musser & Charleton, 2005). MH is considered monophyletic and has been previously used as a test organism for the study of zoogeographical relationships of the landmasses in Wallacea and Sundaland (Ruedas & Kirsch, 1997). Specimens of BA and MH were collected from 3 sites (see Chapter 2, Figure 1), 1 on the large island of Sulawesi, and 2 on the smaller island of Buton, also known as Butung or Boeton (Goodall, 1943). Jennings *et al.* (2005) gives an ecological description of some of the sites in Buton and the sampling sites are further discussed in the second chapter. An image from Jennings *et al.* (2005) is also included, showing 2 population sites on the island of Buton (Figure 4). It is worth noting that most site names have minor linguistic variations and that Buton Island is also approx. 40 times smaller than Sulawesi and has approx. 39 times less inhabitants (Corbett, 1992).

Musser (1981) indicated how *Maxomys* required an in-depth taxonomical study to be able to further analyse the phylogeny of the genus. Ruedas & Kirsch (1997) in a paper on the phylogeny of

Maxomys, also emphasized the need for additional work required for this particular genus. More recently, Gorog *et al*. (2004) requested that further DNA sampling should be performed in order to perform tests on the regional monophyly of *Bunomys* and to further access the divergence depth in the specimen of *Maxomys* which he studied. Concerning South-East Asian murids in general, many authors have encouraged further studies on their phylogenetic relationships (Musser & Carleton, 2005; Achmadi *et al*., 2012). Even though *Maxomys* is treated as a distinct division (Musser & Carleton, 2005) its relationship to other murid genera is likely to be revised in future. Below are brief introductions to some of the most relevant principles and methods used in this study before stating aims and objectives.

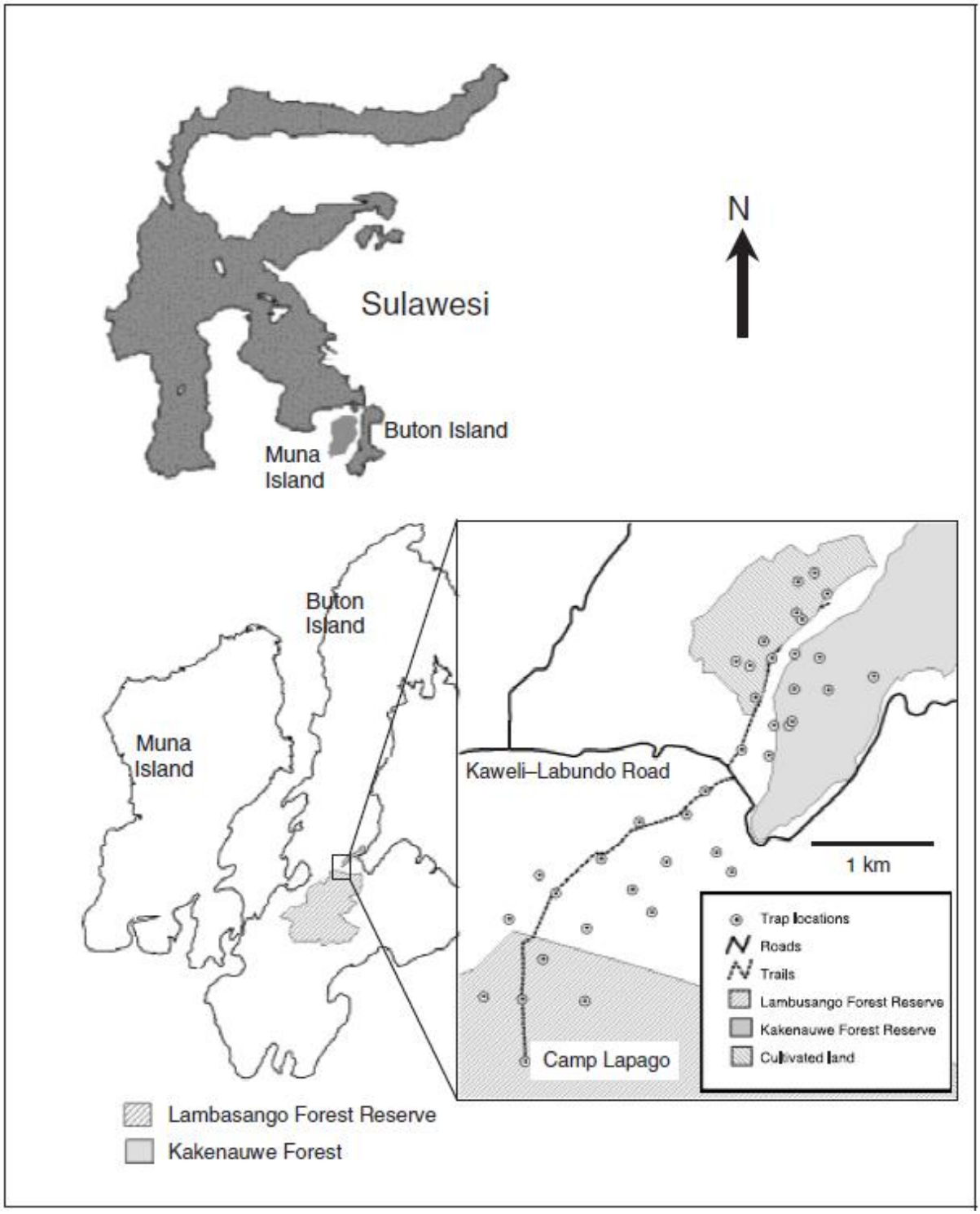


Figure 4. Map showing 2 of 3 populations (Kakenauwe and Lambusango) both on Buton Island. Created by Jennings et al. (2005). The depicted trap locations are not relevant to this study.

1.4. Population Structure

The collected specimens were derived from 3 separate sampling sites on 2 different islands. Although the specimens would have been geographically isolated, this separation does not necessarily classify individuals from a distinct sample site into a distinct 'population'. The definition of a population has long been relatively subjective, as some population structures take factors into account (e.g. language for human populations) which are not genetically related. A crucial step is to first determine a hypothetical set of populations, based solely on genetic data as opposed to the physical location of the given specimen.

Pritchard & Rosenberg (1999) discussed how genetic information may be applied to identify cryptic population structure. Generally, they tried to identify subpopulations, using genetic data only, and to probabilistically assign individuals to a subpopulation. Their program *Structure* was used to help confirm population boundaries. The method assumes a model with a *K* number of populations, in which *K* can be unknown despite the samples being derived from a fixed number of distinct sites. Each individual is assigned to a K population according to their genotypes. This method assumes Hardy-Weinberg equilibrium within populations as well as linkage equilibrium. Pritchard *et al*. (2000) presents a revised version of this method which enables individuals to be clustered into an appropriate population (*K*) by use of genomic data only. This approach varies in reliability according to the number of individuals, number of loci and allele-frequency differences among populations.

1.5. Population Migration and the Origin of Tropical Biodiversity

The migration between populations is a key evolutionary driving force and among the core concepts in population genetics (Abdo *et al*., 2004). A high rate of migration can hinder local adaptation and is one of the principle causes of uncertainty in defining population boundaries. If two populations are geographically separated by hostile terrain, one can expect increases in population differentiation over time through genetic drift, leading to clines in allele frequencies. Such isolation can both hinder

the spread of beneficial mutations and form a starting point for species differentiation (Meirmans, 2014). The measurement of the rates of gene flow is therefore and important factor in molecular evolution and conservation biology. Migration rates can be conveniently inferred by use of genetic markers, which is often preferred over practical studies which involve mark-release-recapture methods. Genetic methods detect genetic migration (movement plus successful reproduction in the new location), and this is an important distinction from mark-recapture methods. Additionally, it is possible to detect dispersal events using genetic markers which could be overlooked by mark-release-capture methods (Pritchard & Rosenberg, 1999).

A clear pattern of the earth's biodiversity is the latitudinal gradient (Mittelback *et al*., 2007), which one can see by observing how species diversity is more pronounced in tropical regions, when compared to temperate regions. Such an observation has led to researchers asking why this is the case, and what other patterns they could observe. Any species found in Sulawesi would have occurred as a result of the area's complex differentiation and extinction dynamics, and as a result of immigration of species from other areas (Heaney, 2005). The full reason behind Sulawesi's diverse biota is still a topic of much debate (den Tex *et al*., 2010). There are two main hypotheses that may explain the high levels of diversification in Sulawesi, termed by some as the 'Pleistocene Pump' and 'Museum Hypotheses' (den Tex, 2011). The Museum Hypothesis states that low extinction rates and the gradual build-up of different species over time is responsible for the observed high levels of biodiversity (den Tex, 2011). The Pleistocene Pump Hypothesis states that the diversity in Sulawesi arose during the Pleistocene, because widespread populations were broken down into smaller, more fragmented and isolated populations with impeded population migration. This is due to recurring sea-level changes brought forward by climatic fluctuations. These events would have allowed some newly formed populations to diverge in isolation, and then become sympatric when the areas eventually reconnected.

1.6. Use of DNA sequencing

The IUCN has listed the necessity to conserve genetic diversity as one of three international conservation priorities (McNeely *et al.* 1990). Genetic diversity relates to a population's evolutionary potential and is primarily necessary to ensure that populations can respond and adapt to environmental change, both natural and anthropogenic (Majerus, 1998; Kettlewell, 1973). Such genetic diversity can be reflected both in phenotypic traits and DNA sequences.

Genetic diversity can be inferred by analysing the variety of genotypes and alleles in a studied population or species (Vila *et al.*, 1997). Allelic diversity, average heterozygosity and polymorphism are most typically used to describe the extent of genetic diversity in a defined population. Historically, most data generated in the field of population genetics was collected from the study of proteins, usually via electrophoresis ('allozymes', see Frankham *et al*., 2002). As only approximately 30% of DNA base changes can be detected via allozyme electrophoresis, this method underestimates the extent of genetic diversity (Lewontin, 1974; Leberg, 1996). Regular improvements in procedures used to analyse genetic diversity, as well as decreases in equipment cost, has given rise to novel methods of DNA analysis (Hoelzel, 1998). For population studies DNA sequencing of mitochondrial genes and microsatellite DNA loci have been popular (Avise, 1994), however next generation sequencing methods provide much greater resolution and are now becoming affordable enough to supersede the earlier methods. Restriction associated DNA (RAD) sequencing data was used for this analysis, a brief description of which is given below (see Chapter 2 for further details).

1.7. Use of RAD sequencing

The study of natural and induced genomic variation is one of the key approaches to understanding the evolutionary history of particular populations (Baird *et al*., 2008). At a molecular level, genetic mapping is essential to decipher the foundations of phenotypic variation. Single nucleotide

polymorphisms (SNPs) are often used to learn about the inheritance of genomic regions. This type of genetic marker is considered favourable due to its high density and repeated occurrence (Stickney *et al*., 2002; Wicks *et a*l., 2001). Drawbacks of using SNPs are usually centred on the preliminary need to discover them in the first place (Chen *et al*., 2008), as well as the cost of genotyping numerous individuals (van Orsouw *et al*., 2007).

RAD sequencing is a relatively new procedure, introduced by Baird *et al.* (2008), which was developed as a cost-effective, high-throughput method to enable researchers to rapidly identify genomic regions which are subjected to both natural phenotypic and mutagen-induced variation (Orr & Presgraves, 2000). RAD tag libraries are here used to facilitate high-throughput genotyping of large populations. These RAD tags are genome-wide fragments of DNA which are isolated from the rest of the genome by use of particular restriction enzymes in order to subsequently facilitate rapid genetic mapping. They are used as a form of genetic marker which is densely spread throughout a given genome (Miller *et al*., 2007).

1.8. Bioinformatics

Bioinformatics has firmly established itself as an essential discipline in molecular biology and encompasses broad areas of research such as conservation ecology and genomics. Bioinformatics is essentially the application of a wide range of computational techniques in order to analyse genomic information, increasingly on a large scale. Some disciplines such as phylogeography have developed rapidly throughout the last 20 years through the improvement of molecular techniques and the software required to process the acquired data (Avise, 2009). A computational approach to analysing any acquired data has become indispensable, as a researcher would need to handle extensive amounts of information obtained by sequencing genomic DNA from a given number of species. Analysing the genetic diversity within a species is essential to appreciate the evolutionary processes which occur both at the genomic and population levels. In general, when beginning computer

analysis of any data obtained via sequencing, the basic characteristics of the data are first uncovered, followed by a more specific, in-depth analysis (Excoffier & Heckel, 2006).

1.9. Aims and Objectives

There is a wealth of discovery still to be made in Sulawesi, which was recently accentuated by the description of 2 new species of *Bunomys* (Musser, 2014). It is essential to continue studying this comparatively neglected area of biodiversity in hopes of facilitating future long-term conservation projects in Indonesia, particularly on the islands of Sulawesi and Buton. There have been accounts of increased extinction rates in rodent taxa (Waser & Ayers, 2003) and more information is required for a more complete study on the conservation status of multiple murid rodents. In terms of mammalian biodiversity, Sulawesi remains one of the least documented regions in the Indo-Pacific and further research would help answer more general questions as to the biogeographical origins and the evolution of native rodents. Taking a broad perspective, island endemic rodents tend to be a focus point in general studies of biological diversity (Ceballos & Brown, 1995), which can also be of use in planning conservation projects. In Borneo and the Sundra Shelf in general, periodical land bridges during times of temporary low sea levels during the late Quaternary have allowed a certain degree of dispersal which provides context that can help explain patterns of diversity (Amori *et al*. 2008). In brief, climate oscillations would not cause Sulawesi to join with Asia or Sundaland, however, the minor islands of Sulawesi (such as the islands of Buton and Muna) would have been connected with mainland Sulawesi during periods of low sea level (Voris 2000). Whether the entire island complex was connected via land bridges is something yet to be studied as well as the general extent of Sulawesi's historical connection with its neighbouring islands.

Using murine DNA sequences which were collected from two native Indonesian species found on two proximate islands, I test the hypothesis that both rodent species had a continuous distribution across Sulawesi and Buton before separation of the islands during the Pleistocene. In order to achieve this, I performed Illumina sequencing on both species, one which is endemic to

Sulawesi and Buton (*Bunomys andrewsi*), and one of which is distributed across South-East Asia (*Maxomys hellwaldi*). Using nuclear DNA sequences, I identify polymorphic sites (SNPs) and test hypotheses about the extent and direction of gene flow. I also use the sequence data to construct phylogenies and use coalescent methods to assess the timing of separation among populations. I aim to address the following questions: (1) what were the geographical mechanisms that led to the evolution of diversity among populations, (2) does the substructure of populations found on separate islands reinforce the idea of a geographic isolation and (3) whether both species have similar ancestral nodes, which could indicate whether their diversification occurred after a singular event, such as the Pleistocene's climatic oscillations.

Chapter 2 considers the extent of current genetic diversity between conspecific populations. In Chapter 3 I will attempt to link genetic disjunctions with historical events such as physical barriers (e.g. the formation of an ocean current or mountain range), past climates (e.g. glacial periods) and limited dispersal.

Chapter 2 – Population Genetics

2.1. Introduction

This current chapter deals with the population structures of the two focal species (*Bunomys andrewsi* & *Maxomys hellwaldi*) and the processes which could have caused differentiation among their populations. The mechanisms which lead to species differentiation have long been the subject of research since the inception of modern evolutionary biology (Simpson, 1953). One advantage of studying two different genera, as opposed to a single one, is the ability to compare the genetic structures of both species with each other, possibly revealing how specific factors (e.g. changes in sea level) caused genetic divergence between given populations (Knowles, 2009; Hickerson *et al*., 2009). A change in significant environmental variables would have the potential to impact the evolutionary trajectories and dynamics of multiple taxa. This in turn can potentially lead to evolutionary divergence among different populations as well as eventual speciation (Klicka & Zink, 1999). The study of environmental variables and population structure (as opposed to dating the occurrence of phylogeographical events which would have impacted the evolution of *Bunomys* and *Maxomys*) is the focus of this chapter.

Briefly, the occurrence of species differentiation can be summarized by asking two questions: 1) "how do new populations within a species develop?" and 2) "how do such populations become reproductively isolated from their parental populations?" (Mayr, 1947). There are therefore two essential steps involved in species differentiation, 1) the establishment of new interbreeding populations and 2) the establishment of reproductive isolation (Mayr, 1947). Much work over the last 25 years has applied a method that integrates population level and phylogenetic methods to better understand this transition between population differentiation and incipient speciation, called 'phylogeography' by Avise (2000; see review by Hickerson *et al.* 2010).

2.2. Geographic & Reproductive Isolation

Geographic isolation is a broadly-used term which refers to any set of environmental factors which effectively inhibits gene-flow between two populations. This can occur on a macro-geographical scale, where two populations are on separate islands, yet it is also possible that micro-geographic variation exists within an island, a good example being the work of Brown & Thorpe (1991) on Canary Island skinks who described phenotypic variation over a small geographic range on Grand Canaria Island associated with altitude, vegetation and climate. There are many cases of micro-geographical isolation, where populations live in close proximity to each other without formidable barriers to movement. Examples of such micro-isolation (sometimes termed topographical isolation) can be found in studies by Dehais *et al*. (2010), who worked on microsatellite loci in non-migratory French fish, and Chuquiyauri *et al.* (2013), whose research centred on the micro-geographical differences of Amazonian parasites.

Geographic isolation is frequently accompanied by varying degrees of reproductive isolation. In any given study, the degree of interbreeding and subsequent reproductive isolation can differ significantly. Long-term isolation from other populations can lead to reproductive isolation sufficient to reduce fitness in hybrid offspring; yet on the other hand a low level of reproductive isolation, which allows a moderate level of continuing gene flow, is able to support population structure within a species. A well-known example of the process causing reproductive isolation was brought forward by Schluter (2001), who studied how postglacial freshwater fish adapted to different habitats as unique ecotypes. Although reproductive isolation is generally considered a long-term process which occurs over several millennia, it has been documented that two populations can be reproductively isolated in as few as 13 generations (Hendry *et al*., 2000). A simple model where the process of incipient speciation can arise is, for example, the evolution of barriers to gene flow caused by ecologically-based divergent selection (Rundle & Nosil, 2005; Vines & Schluter, 2006). An illustration of ecological differentiation can be found in Michel *et al.* (2008), who published one of

the first studies which clearly demonstrated how intraspecific differentiation can be initiated by a loss of migration. Michel *et al*. (2008), in a study on the New Zealand endemic *Gobiomorphus*, showed how distinct ecotypes of freshwater fish found in separate rivers can be formed solely as a consequence of migration loss (all other factors remaining equal), which would cause a subsequent species diversification. Michel *et al*. (2008) gave good evidence to support the theory that if reproductive isolation persists, this process can provide a mechanism to facilitate speciation. One can expect to observe such incipient speciation in many Sulawesi species. In the literature, species differentiation is sometimes described in two stages: the process is first 'initiated' by a specific factor and then 'reinforced' by the same or other factors. Two populations which became geographically isolated by the formation of mountains or rivers would also begin differentiation by the relatively simple process of genetic drift (which is a stochastic process of differentiation).

Other factors which can reinforce incipient speciation include sexual selection. Sexual selection is one of the primary drivers of speciation, and Sawyer & Hartl (1981) demonstrated how completely arbitrary changes in mate preference could reinforce speciation. Ecological speciation can also be enforced culturally, as opposed to genetically. This is exemplified in the case of killer whales (*Orcinus orca*; Hoelzel et al. 1998, 2007, Foote *et al*., 2009), as populations of orcas exhibit taught behaviours which are not genetically heritable and which can lead to different mating choices.

2.3. Population Structure & Differentiation

Very few data exist on *M. hellwaldii* and *B. andrewsi*, so their population structure had to be inferred by studying similar species and relatable accounts of population differentiation. In a biogeographical paper by Gorog *et al.* (2004) which discusses the biogeography of two related species (*Maxomys surifer* and *Maxomys whiteheadi*), one is reminded that the Sunda islands are fundamentally different from well-studied island systems such as the Galapagos (Grant, 1999) because Sulawesi is partially surrounded by a shallow continental shelf (Voris, 2000) which would allow the vicariance of

certain populations. These populations, now endemic to different islands, would have formerly shared the same habitat during periods of low sea-level (Heaney, 2005). As discussed in Chapter 1, these periodical connections are very important to the evolution of the Sunda Shelf, yet few data exist on how it affected population structure in rodents. A broad review of the population structure of different non-rodent species inhabiting this region has allowed us to formulate hypotheses about the population structure for *M. hellwaldii* and *B. andrewsi*.

There are several relevant papers which offer insight into the population structure of mammals in this region (Schmitt *et al.*, 1995; Karns *et al*., 2000; Inger & Voris; 2001; Ruedi & Fumagalli, 1996; Heaney 1986). Of these papers, Schmitt *et al.* (1995) studied fruit bats and Karns *et al.* (2000) performed an analysis on water snakes. Both involve animals which can disperse relatively easily across water, hence were less informative concerning the population structure of land-mammals.

Many papers deal with population structure along with phylogeography, hence some references to phylogeography are here made, which will be discussed in more detail in Chapter 3. A paper by Ruedi (1996) on shrews found in the Malay Archipelago showed strong evidence of dispersal as well as vicariance having an impact on the present population structure of shrews and that neither factor is solely responsible for the current population structure. However, in the previously mentioned study by Gorog *et al.* (2003) on two similar species found on Sulawesi (*Maxomys surifer* and *Maxomys whiteheadi*), the author refuted the hypothesis of widespread migration and emphasized the importance of pre-glacial vicariance in South-East Asia. This pre-glacial vicariance occurred during a time of relatively high sea level, which would have prevented populations from breeding with rodents from different islands. Although little is known about the distribution of *M. hellwaldii* and *B. andrewsi*, one can assume that repetitive cycles of connection-isolation would be one of the primary forces in creating species differentiation, which was proposed in the case of *Hylomys* by Ruedi & Fumagalli (1996). These connection-isolation cycles would have

been caused by sea-level changes as previously discussed in Chapter 1. It is desirable to know if pre-glacial vicariance had a greater impact than island colonisations, although some colonization would have taken part in forming the present population structure. The colonization of a new environment can often lead to rapid diversification (Hendry e*t al.,* 2000) due to the development of different selection regimes to suit the new surroundings.

There are two main evolutionary routes which could have occurred on Sulawesi and Buton, both which are described in detail by Brooks and McLennan (1991). The first alternative is that geographical isolation occurred through the colonization of unoccupied islands from a source population. The second alternative is that incipient differentiation was sufficiently pronounced to account for the current population structure. Fragmentation of the geographic range of a species often grants an opportunity for genetic differentiation, partly because adaptation to differing environmental conditions would favour certain morphological changes. Examples of how adaptive processes can leave genomic signatures in organisms can be found in the work of Simonson *et al.* (2010) and McCracken *et al.* (2009), who studied humans and birds respectively. Simonson *et al*. (2010) presented genetic evidence showing high-altitude adaptation in humans native to the high-lands of Tibet while McCracken *et al*. (2009) similarly showed how pintails from the Andes have adapted to high altitude. Although this is a topic of much debate, habitat fragmentation can potentially be considered to be the most prevalent mode of species differentiation (Mayr, 1982; Bush, 1975), as opposed to rapid speciation through colonization of an uninhabited island.

My objective is to study the extent of evolutionary diversity between different populations of two rodent species naturally found on two separate islands. I test the hypothesis that my putative populations are differentiated across the narrow channel representing a current physical boundary separating the islands. I use RAD data to provide sufficiently high resolution to permit the assessment of differentiation among parapatric populations on the island of Buton.

2.4. Methods

DNA samples were extracted from 146 specimen of *Bunomys andrewsi* and 73 *Maxomys hellwaldi* during years 1997 to 1998 by Operation Wallacea. The samples used were selected from a long-term DNA archive at Durham University (United Kingdom). All samples were preserved in TE Buffer (10mM Tris, pH 8, 1 nM EDTA) at temperatures below -20 °C. Unnecessary freeze-thaw cycles were avoided to best maintain the integrity and yield of the DNA (Ross *et al*., 1990). Due to the possible detrimental effects of long-term storage on the stability of DNA (Röder *et al.*, 2010), the genomic material contained in all available samples was quantified in order to single out the best preserved specimens. The most appropriate samples were selected by first comparing gel runs and the accompanying notes taken by technician Colin Nicholson during 2001 to 2002.

To ensure that the concentration of DNA had not deteriorated since 2001, agarose gel electrophoresis tests were performed on all available samples. A 1% agarose gel was prepared using 1X Tris/Borate/EDTA buffer solution (Brody & Kern, 2004) and Molecular Biology Grade agarose (Melford Laboratories Limited). For each sample, 4 µl of undiluted DNA extraction was mixed with 2 µl of 5X GelPilot® DNA loading dye (QIAGEN Sciences). The solution was loaded on the prepared gel and run for 80 minutes at 100v.

From the 219 available samples, the 80 best were chosen after their DNA concentrations were quantified. Although all 80 samples underwent sequencing, 66 were used in the analysis (see section 2.6 for an explanation as to why some of the samples had to be omitted). The selected samples are outlined in Table 1 and come from 3 different populations (abbreviated as populations 1, 2 and 3; Figure 1). Initially, samples from four sites were analysed, two of which (the village Labundo Bundo and the forest of Kakenauwe) have been combined together as 'Point 1' in Figure 1. The proximity of those sites can be seen in Figure 2, and they were therefore merged into the single site 'Kakenauwe' . The samples were pooled because the sampling sites are close to each other and from a similar habitat. Jennings *et al.* (2006) and Martin & Blackburn (2010) both give some

description of the study sites. A description of Sulawesi's general forest area (both lower montane and upper montane) can be found in Hasnawir *et al*. (2006). The Lambusango Forest Reserve is close to the Kakenauwe Forest Reserve and together encompasses a large section of southern Buton (Figure 2). It is worth noting that Kakenauwe is adjacent to cultivated land and that the two reserves are separated by a road. Lambusango Forest Reserve (5°10'–5°24'S, 122°43'–123°07'E) is a 65,000 hectare area of lowland tropical forest which is mostly uninhabited. The forest is split into 35,000 hectare limited production forest (Singer & Purwanto, 2006) and 28,510 hectare strict forest reserve. The geology of the area is primarily formed of Quaternary karst coral limestone (O'Donovan, 2001).

		Bunomys andrewsi	*Maxomys hellwaldii*	Total
Small Island	Kakenauwe *	32	15	**47**
	Lambusango	8	2	**10**
Mainland	Lanowulu	7	2	**9**
Total		**47**	**19**	**66**

* Initially 2 adjacent sites, Kakenauwe and Labundo Bundo were combined

Table 1. Distribution of Rodent samples

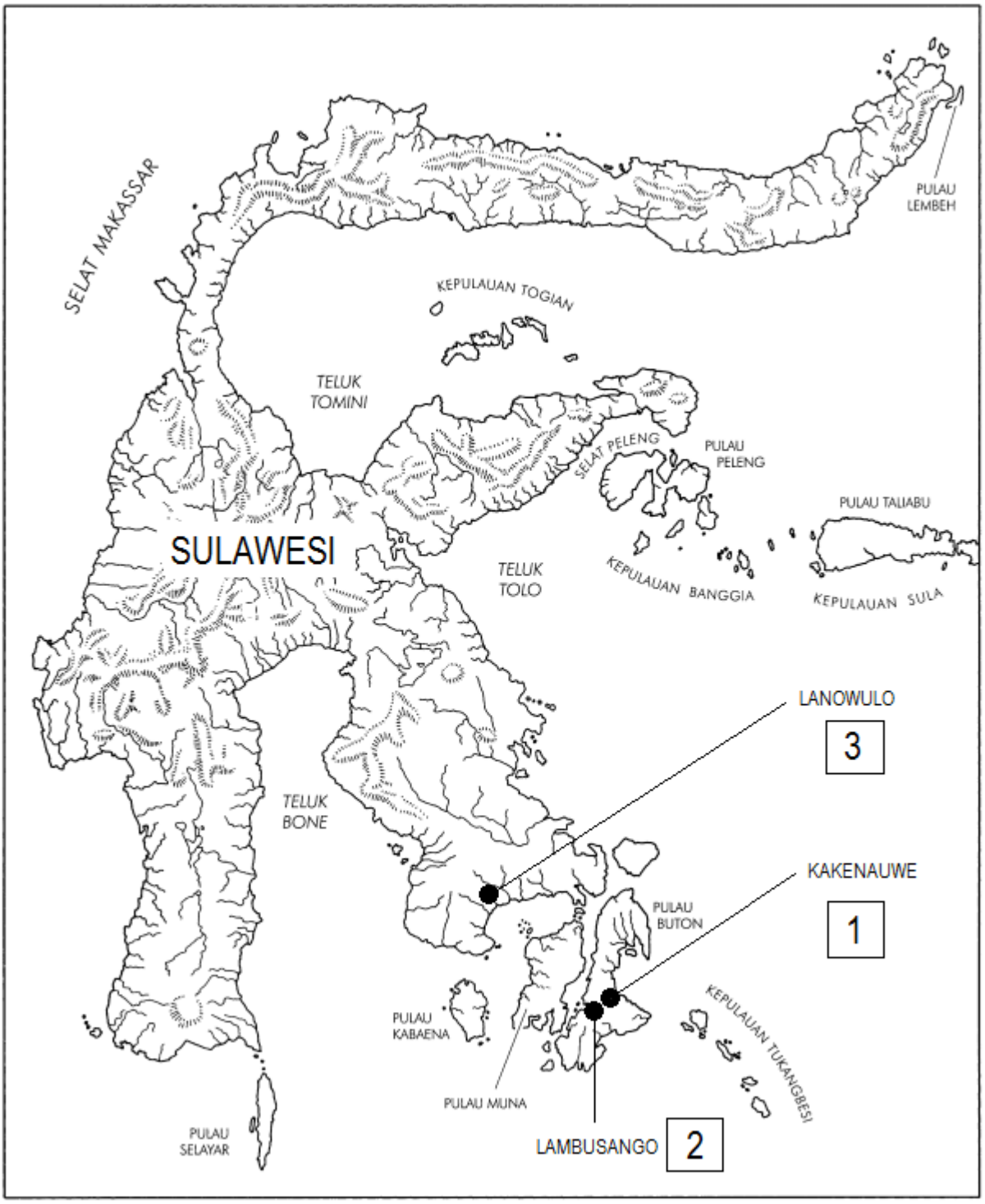


Figure 1. Sampling Sites on Sulawesi, marked as populations 1-3, where population 3 is the only population on mainland Sulawesi. Map adapted from Musser (1991).

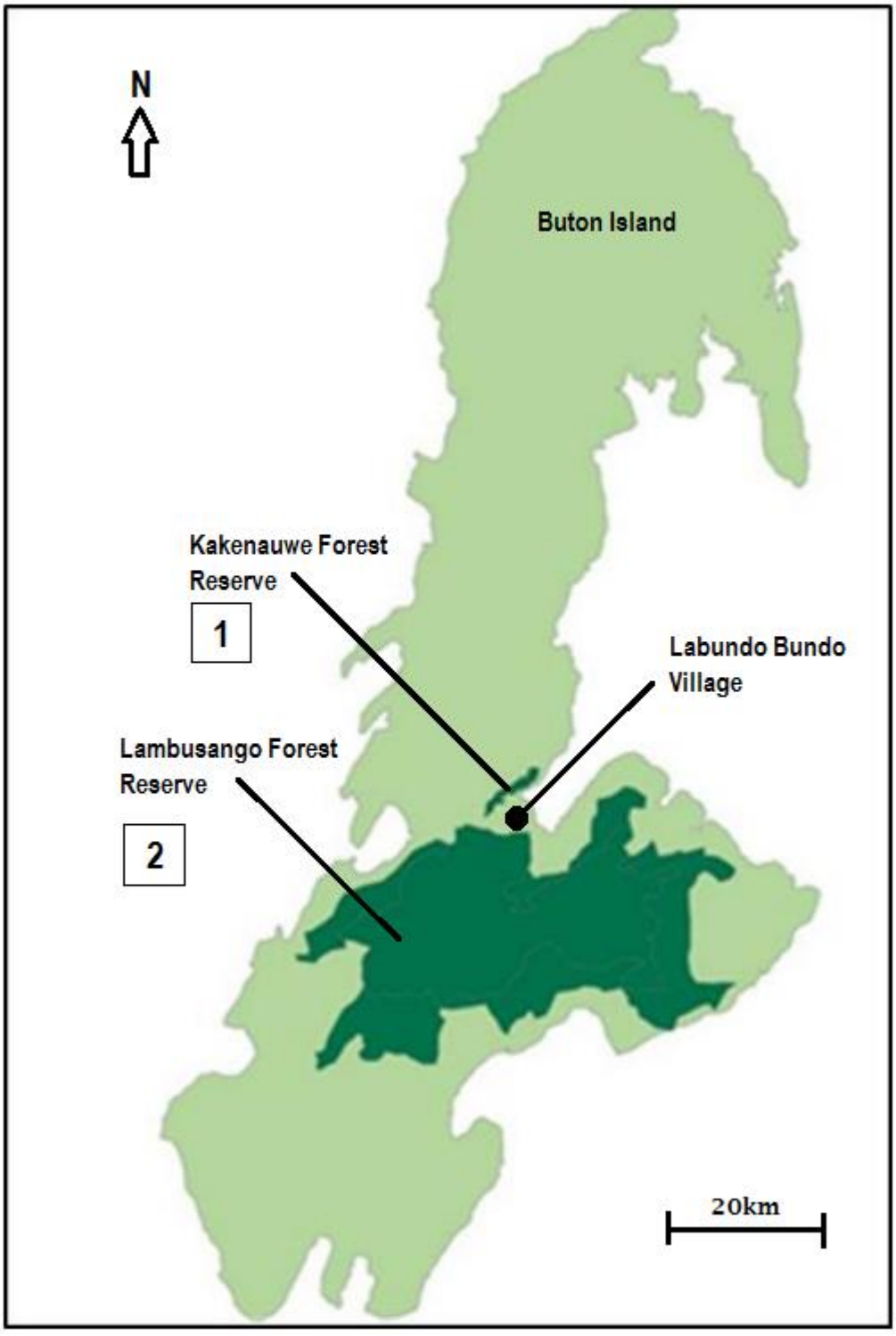


Figure 2. The Kakenauwe and Lambusango Forest Reserves (populations 1 and 2 respectively) shown within Buton Island (dark green). Sampling in Kakenauwe was undertaken in the mid-east of Buton, close to the village Labundo Bundo. Sampling in Lambusango was performed in the Midwest. Adapted from Wheeler (2011).

2.5. RAD Protocol

Peterson *et al.* (2012) provides an overview of the methods used in RAD-tagging. Baird *et al*. (2008) goes through a comprehensive protocol which form the backbone of the RAD procedures. Although his methodology is in some way similar to the following experimental procedure, several steps were appropriately altered (e.g. no Steptavidin beads were used). Additional detail on some on experimental procedures, such as PCR, can be found in Hoelzel (1998).

DNA libraries were prepared with 10 samples at a time, using 10 different barcodes and 8 indexes. The best samples from the available archive were quantified, which was achieved by AccuBlue™ High Sensitivity dsDNA and Quant-iTTM PicoGreen dsDNA quantification kits. All samples were set at the same concentration (250ng of DNA in 12.5 µl of distilled, deionized water). The appropriate quantity was measured out for each sample and then centrifuged under vacuum at 50 °C until dry. The DNA which remained in each sample tube was suspended in 12.5 µl ddH_2O and each tube was lightly vortexed to dislodge the dried DNA from the sides of the sample tube. The sample tubes were then left standing for at least 10 minutes to allow any traces of the dried DNA to diffuse into the water. A master mix for the restriction digest was prepared using MspI and HindIII restriction enzymes from New England Biolabs.

Restriction Digest Master Mix

Reagent	µl/Sample
H_2O	35
Buffer	2
BSA	0.2
Spermidine	1
MspI	0.4
HindIII	0.4
DNA	12.5

Termocycle

Time	Temperature (°C)
16 hours	37
20 mins	60
Store	< 8

After lightly vortexing and briefly centrifuging the resulting mixture, the reaction tubes were placed in a pre-warmed water-bath at 37 °C for 16 hours to enable digestion. Following this period of incubation, the tubes were centrifuged to concentrate any condensed solution towards the bottom

and then placed in a PCR machine at 65 °C for 20 minutes to heat-inactivate the digestion process. Although it could have been beneficial to purify the resulting solution, the solution was used in the next step (the ligation of adaptors) after the digestion process without extracting the restriction enzymes. This decision was made in order to minimize the amount of DNA lost after each purification step. The digestion enzymes were also heat inactivated and would not interfere with the subsequent PCR. Another master mix was created for the adaptor ligation and added directly to the sample tubes containing the product of the restriction digest.

Ligation Master Mix

Reagent	µl/Sample
H_2O	10.5
Buffer	4
Ligase	0.5
Adaptor Mix	5
DNA	20

Termocycle

Time	Temperature (°C)
2 hours	22
20 mins	65
Store	< 8

After the samples had successfully undergone the adaptor ligation, all ligation products were pooled together and then cleaned using PureLink® PCR Micro Kit (Invitrogen™) columns, which allowed up to 95% recovery of DNA fragments (Life Technologies Corporation, 2009). This purification step removed over 98.5% of the available enzymes and buffer. Three PureLink® columns were used to clean up a total of 300 µl of the ligation product (100 µl of ligation product per column, eluting in 10 µl per column). The purified DNA was eluted in a buffer consisting of 10 mM Tris–HCl (pH 8.5) and subsequently stored at -20 °C.

DNA fragments which were sized between 300bp to 400bp were obtained by running the ligation product on an agarose gel and selectively extracting the DNA fragments of the required length. A prepared 1.5% agarose gel was immersed in a Bio-Rad PowerPac 300 electrophoresis cell containing 1X TBE. 20µl of the ligation product was mixed with 6 µl of 5X GelPilot loading dye and loaded in a single sample lane on the prepared gel. In order to be as precise as possible with the subsequent gel cutting, the ligation product ran with Thermo Scientific™ GeneRuler™ 100bp DNA

Ladder on both sides of its lane. The gel ran for 100 minutes at 90V. After the run was completed, the gel was stained in a solution of distilled water and Ethidium Bromide (EtBr). The area on the gel which contained DNA fragments sized 300bp to 400bp was extracted under ultraviolet light, using clean surgical blades and the DNA ladder as reference.

A Boekel™ Dry Bath Incubator set at 50 °C was used to melt the obtained gel fragment in binding buffer for the subsequent purification. The DNA contained within the extracted gel was cleaned using a QIAGEN® QIAquick Gel Extraction Kit, eluting in 50 µl. The extraction kit removed the agarose gel and EtBr and ensured up to 80% recovery of DNA.

The next step in the procedure involved attaching an index to the purified, size-specific ligation product. As noted before, one index was used per 10 samples, so only a single index was required per library made. A Phusion® High-Fidelity PCR Kit from New England BioLabs was used to clone the ligation product and further attach an index to the DNA fragments. 8 PCR reaction tubes each with 6 µl of gel extracted library were used together with the master mix prepared from the Phusion® PCR kit. After completion of the PCR, the contents of all reaction tubes were pooled and a second agarose gel was prepared.

Phusion Master Mix

Reagent	µl/Sample
H_2O	2.76
Buffer	3.2
dNTPs	0.32
Forward Index	1.28
Reverse Index	1.28
Polymerase	0.16
DNA	6

Termocycle

Time	Temperature (°C)	
30 sec	98	
30 sec	98	x13
20 sec	62	
45 sec	72	
5 mins	72	
Store	< 8	

The second gel was set using the same reagents and methods of the first gel, except that the GeneRuler™ 100bp DNA Ladder was not placed directly adjacent to the library sample, in order to avoid accidentally gel-extracting some of the ladder. 30 µl of the Phusion PCR product was placed into a single lane, and the gel ran for 85 minutes at 90V. The gel was then stained with EtBr, using

the same methods as the first gel, and DNA fragments between 450bp and 550bp were extracted as previously described. This slight increase in selected base pair size accounted for the additional barcodes and indexes, which were now attached to the DNA fragment.

Using methods similar to the QIAGEN® QIAquick Gel Extraction Kit, the gel fragment was purified using MinElute Gel Purification Kit columns (Qiagen), eluting in 10 µl of purified DNA library. Although the Qiagen MinElute columns usually did yield positive results, I encountered difficulties with some of the libraries which were too diluted and decided to use a magnetic bead cleaning procedure to compensate. The procedure is a modified version of the Agencout® AMPure® XP PCR Purification, which filters PCR products by use of carboxylated, magnetic beads, the principle of which is described by DeAngelis *et al*. (1995). Amplicons which are above a certain number of base pairs are bound to paramagnetic beads, allowing excess nucleotides, salts and enzymes to be removed. The bead-buffer solution required for the procedure was obtained by creating a mixture of the following:

- 0.1% Sera-Mag™ SpeedBeads, carboxylate-modified micro-particles (ThermoScientific)
- 18% PEG-8000 Powder (w/v) (Promega)
- 10 nM Tris-HCl, pH 8.0
- 1 M NaCl
- 1 mM EDTA
- 0.05% Tween-20

The following steps were used to create the SpeedBeads solution:

1. Add 9 g PEG-8000 powder to a 50 ml Falcon tube.
2. Add 100 µl 0.5 M EDTA, 10 ml 5 M NaCl and 500 µl l 1 M Tris-HCl. Top up the Falcon tube with ddH_2O until the tube is approximately 49 ml full.
3. Mix the Falcon tube until the PEG-8000 powder is completely dissolved. Add 27.5 µl 0.05% Tween-20 to the mixture and then shake again vigorously.

4. Take the stock solution of Sera-Mag beads and re-suspend them by shaking. Using a 2 ml Eppendorf tube, transfer 1 ml of the Sera-Mag bead suspension and transfer the solution to a magnetic rack. Wait 10 minutes, or until the metallic beads are all pulled towards the magnet.

5. After the beads have formed a pellet, carefully remove the storage buffer and wash the beads twice with 1 ml TE. Do this while the Eppendorf tube remains on the magnetic rack. Take the tube off the magnetic rack and re-suspend the beads in 1 ml TE.

6. After re-suspending, immediately add the contents of the Eppendorf tube to the Falcon tube. Use instantly, or store the Falcon tube in a fridge at 4 °C, away from any strong sunlight (tube covered in aluminium foil).

The Agencout® AMPure® XP PCR Purification protocol was then used to purify the solution, eluting in 10 µl. Care was taken to ensure that the stored SpeedBeads solution was at room temperature and thoroughly mixed before use, so that it should appear homogenous in colour. Fresh 70% ethanol, which is used to clean the resulting bead pellet, was prepared on the day and purified using a micro-filter. The final step before sending the library off to sequencing was to dilute the library to 2nM, which was accurately quantified by both Tapestation (Agilent) and qPCR (KAPA). The library was run on 2 separate lanes, both lanes containing PhiX (Baird *et al*., 2008), which served as a control library during the sequencing run. The PhiX library provided a calibration control and a quality control for cluster generation.

2.6. Data analysis

The data from the Illumina sequencer was received in FastQ format and separated according to restriction site associated DNA markers (RAD index), forward and reverse reads and sequencing lane. The Phred quality scores provided by machines using Illumina 1.8 onwards was 33 to a base, which equates to a base call accuracy of over 99.9%. A server provided by Durham University was used to process the data, which was accessed by a Secure Shell client using a Unix-like operating system on a Microsoft Computer (SSH version 3.2.2 for Microsoft Windows). The SSH Secure Shell is the client-side application which is required as a workstation for end-users. The server at Durham University

already had several required modules at disposal (e.g. Stacks) and some compilers (e.g. Java) which were subsequently used to organize the received data. The programs used for analysing the RAD data were as follows:

The command 'process_radtags' from the Stacks pipeline was used to clean the raw sequence data and separate the samples by barcode. Having been successfully implemented in recent publications (Catchen et al., 2013; Amores et al., 2011), Stacks enables the identification of thousands of SNP markers (McCormack et al., 2013) and is a software which can be used to build loci from short-read Illumina platform sequences. It was developed in response to the rapid rise in parallel short-read sequencing technologies, which brought serious challenges in data processing and analysis (Glen, 2011; Shendure & Ji, 2008). The software allows us to assemble reads together and track genotypes using a highly statistical framework (Gompert et al., 2010). In order to run the software, text files containing the barcodes of the samples were referenced to stacks, which gave FQ (FastQ) files that were distinguishable according to barcode and index. Tracing back the barcodes and indexes, these files were separated by species and collection site. Any uncalled bases were here removed and reads with low quality scores were discarded. If the score dropped below a 90% probability of being correct (a raw phred score of 10), the read was discarded .The program was told that the number of base-pairs was 125, which was the average for the single-pairs data.

The command 'bwa index' from the BWA pipeline was used to index a reference genome. BWA is a software designed to map low-divergent sequences against a given genome (Li & Durbin, 2009), in this case the Norway Rat (*Rattus norvegicus)*. A masked version of the full *Rattus norvegicus* genome was downloaded from Ensembl (Wellcome Trust Sanger Institute) in compressed FASTA format. Indexing a genome facilitates the rapid retrieval of specific alignments which overlap a specific genomic region so future programs do not need to sort through all the alignments. With the help of RepeatMasker tools (Bergman & Quesneville, 2007), the reference genome was screened for low complexity regions and interspersed repeats, allowing these specific reads to be masked. For

the compressed indexing and local alignment of DNA, the software BWT-SW (Lam *et al*., 2008) was implemented, which is capable of finding all local alignments. The same pipeline was then used to align the sequence data to the indexed *Rattus norvegicus* genome and subsequently create SAM files (Sequence Alignment/Map format), using the commands 'bwa aln' and 'bwa sampe' respectively. The command 'bwa sampe' was specifically designed to generate alignments in the SAM format with paired-end reads, as opposed to single-end (Li, 2012). After this step was concluded, all forward and reverse reads were concatenated and the individual samples were only separated by lane.

Various commands from the Samtools pipeline were used to convert, sort, index and merge the given SAM files. Samtools is composed of a set of programs which can manipulate alignments in the SAM/BAM format. The changes in format also allowed further programs to retrieve selected reads more swiftly. The command 'samtools view' was first run to convert all SAM files into BAM files which were required by the program for the next steps. 'samtools sort' was subsequently used to sort the alignments by the leftmost coordinates and then 'samtools index' was performed to enable fast access of the data. 'samtools merge' was used to concatenate the data from both lanes, giving us 80 separate SAM files (66 of which were used in the analysis), one for each specimen, which contained all relevant data from the forward and reverse reads of both lanes. The same command line was again repeated for the merged SAM files, to ensure that no change in format occurred after merging the files.

The command 'ref_map.pl' from the Stacks pipeline was used to clean the referenced SAM files and find loci. The program was run twice, creating a reference map for each species. The minimum depth coverage to report a stack was set at 10 for both runs. The command itself can be separated into 3 steps:

- First, 'pstacks' sorts through the referenced sequences and aligns them to matching stacks. After comparing the stacks, the program then detects SNPs at each locus by use of a maximum likelihood framework (Hohenlohe *et al*., 2010).

- Secondly, ‘cstacks’ builds a catalogue using the process samples from ‘pstacks’. The program merges alleles together and computes sets of consensus loci. In the event of a genetic cross, the program would catalogue sets of possible alleles.
- Lastly, ‘sstacks’ was used to match each sample against the created catalogue.

The command ‘populations’ from the Stacks pipeline was lastly implemented to create a Genepop file which could then be used to calculate population statistics. The program compares all given populations pairwise in order to compute population genetics statistics such as F_{ST}. This command was run multiple times to evaluate which individuals should be omitted in order to obtain the highest number of loci. The percentage of individuals in a population required to process a locus for that population was set at 100%. The minimum number of populations required to process a locus was set at the maximum (all populations were included).

Of the 80 samples sequenced, 6 returned insufficient reads and were discarded from this analysis. Performing the ‘populations’ command using the remaining 49 *Bunomys* and 25 *Maxomys* specimen, 46 loci were initially found for *Bunomys* and 272 loci for *Maxomys*. The same program was repeated for both species, each time omitting a single specimen from the analysis (e.g., *Bunomys* specimen ‘A’ would be removed, and the analysis performed on the remaining 45 specimen. Specimen ‘A’ would then be added back to the analysis and specimen ‘B’ would subsequently be removed, before repeating the process). By this method, I was able to note any increase in the number of loci for a given population, upon omission of a particular specimen. I found that I received most loci when omitting 5 specific *Bunomys* and 4 *Maxomys* specimen. Analysing 44 specimens of *Bunomys*, I received 3410 loci, while 22 specimens of *Maxomys* yielded 422 loci (for a total of 66 samples as indicated in Table 1). No calculations were made on how many loci one would receive by omitting every possible combination of specimens, yet it was determined that a smaller sample size would only provide a small increase in loci. Outlier loci were further on removed using *LOSITAN* (methods to follow).

As some of the necessary programs have different degrees of interoperability (e.g. separate input files are required according to the program) the program PGDSpider (Lisher & Excoffier, 2012) was used throughout the analysis to convert files from one particular format to another. PGDSpider is a Java program which can read multiple different file formats and adapt them to the particular format required by software. The program SAMtools (Li *et al*., 2009) was also used alongside PGDSpider to convert specific files. PGDSpider forms a convenient substitute for earlier programs such as CONVERT (Glaubitz, 2004) and was used to create input files for several of the following programs.

A brief summary of the functions of the programs used in the analysis is given below:

- *LOSITAN* assesses the relationship between H_e (expected heterozygosity) in an island model (Wright, 1931) and F_{ST}. Lositan can be used as a more computationally intensive substitute for 'fdist' (Excoffier & Heckel, 2006), which would require additional independent runs in order to fine-tune optional parameters. Unlike 'fdist', *LOSITAN* is less prone to usage error, which can be caused by inappropriately estimating average F_{ST} values.

- *Genetix* (Belkhir *et al*., 2001) was used to perform a Factorial Correspondence Analysis (FCA) on each species. All loci were used for *Maxomys* when performing FCA (421 loci), yet the *Bunomys* file was shortened to 1000 loci (few enough to permit efficient running of the program, and sufficient to give high resolution). PGDSpider Version 2.0. was used to create the required input files. Hubert *et al.* (2005) provides an introduction for principle component analysis (PCA) used in *Genetix*.

- *Structure* was used to detect underlying genetic populations among a cluster of individuals from multilocus genotypes. The program was developed by Pritchard *et al.* (2000) and further expanded by Falush *et al.* (2003). *Structure* enacts a model-based clustering method which infers the structure of a population by analysing genotype data. It can be used to assign an individual to a given population, to identify discrete genetic populations and to

demonstrate the presence of population structure. *Structure* probabilistically designates each individual to a predefined hypothetical cluster number (K). The program works on the surmise that within each population, the loci are both at linkage equilibrium and Hardy-Weinberg equilibrium. Structure allows the option to use an 'Admixture Model'. Using an Admixture Model means that individuals are allowed to have a mixed ancestry as opposed to 'No-Admixture Model', where all individuals are assumed to be drawn solely from one of the K populations. . The USEPOPINFO selection flag option was not used when formatting the input data. An Admixture Model was used along with correlated allele frequencies. After finding the appropriate K values, Structure ran for 500,000 MCMC with a 10% Burnin for 4 iterations.

- *BayesAss* is used to estimate the rate of migration between populations using Markov chain Monte Carlo (MCMC) simulations (Wilson & Rannala, 2003. The program also estimates the inbreeding coefficient (*F*) and allele frequencies and uses assignment methods (Manel *et al.,* 2005; Paetkau *et al*., 1995) to single out individuals which have a migrant ancestry. The program assumes a relatively low level of migration between populations. Additionally, *BayesAss* presumes that when migrant individuals form a new population, this younger population cannot exceed 1/3 of the population total each generation, or in other words, the proportion of a population which are non-migrants cannot be lower than 2/3. It can hence be uncertain, partly due to differences in sample sizes, as to whether a population is ancestral or derived from separate one. *BayesAss* permits deviations from the Hardy-Weinberg equilibrium; however the data are assumed to be in linkage equilibrium. The accuracy of the program's results is proportional to the sample size and number of loci, so I tried to create a Genepop file in Stacks which incorporated as many individuals as possible. Generally, a decrease in sample size would yield an increased number of loci, however in this scenario there is a surplus of loci and the population size is the limiting factor. Another

element which would affect the accuracy of the *BayesAss* results is the strength of population differentiation, which was tested using Genepop (Rousset, 2008).

- *Clumpp* was run for each species individually using the population and individual result files obtained when K was set to all values between 1 and 10 and then subsequently for all values between 1 and 6 using more iterations. Structure Harvester allowed us to combine the data received from all 4 runs, creating two summary files per species, one with individual and one with population data. The 'FullSearch' algorithm in *Clumpp* was used (as opposed to 'Greedy' or 'LargeKGreedy') and the G' pairwise matrix similarity statistic. The datasets were also run using the 'Greedy' algorithm, but due to the small K value, 'FullSearch' was deemed more appropriate. 'FullSearch' finds the optimum group alignment in multiple runs and considers all possible permutation vectors.

- *Distruct* was used to graphically display the population structure of both species. In theory, Distruct could be used with the results directly obtained from Structure, yet Structure only shows results from each individual iteration (of which I ran 4) and can only calculate a histogram using the data from a specific run. Processing out data through Structure Harvester and Clumpp allowed us to consider data from all 4 iterations. After modifying the output files from Clumpp according to the program's manual (Rosenberg, 2004) and setting the appropriate parameters in Distruct, a postscript (.ps) file was created for each species, which was transcribed into a visual .pdf file via GSView (Ghostgum Software Pty Ltd.)

2.7. Results: Genetic diversity

The genepop file created by the 'populations' command in Stacks was run through the program *LOSITAN* (Antao *et al*., 2008) in order to identify outlier loci that had an F_{ST} which was either too high or too low in comparison to neutral expectations. The *LOSITAN* confidence intervals were set so that the extreme top 2.5% and bottom 2.5% of loci were located in the genepop file. The outliers were removed by manually excluding them from the 'populations' command in stacks by use of the 'whitelist' program option. By whitelisting specific markers, the program only processes predefined markers which were specified by *LOSITAN*. The resulting genepop files were used in all subsequent analyses and contained 2771 loci for *Bunomys* (43 specimen) and 421 loci for *Maxomys* (19 specimen).

Using *Geneious* 8.0.5 (Biomatters Limited) one could visualize the alignments for the FASTA and NEXUS files created for *Bunomys*, *Maxomys* and both species combined. *Bunomys* had the longest sequence length at 1,342,178 bp with 95.8% identical sites (the percentage of columns in the alignment for which all sequence is identical). *Maxomys* had a sequence length of 178,364 bp with 94.9% identical sites while both species combined had a sequence length of 102,655 bp with 94.2% identical sites. Arlequin 3.5 (Excoffier, 2010) was then used to gather population properties on both species which are summarized in Table 2. Note that Arlequin calculated population properties in Table 2 with double the actual number of loci. In summary, Arlequin used 241 loci for *Maxomys* (N=19) and 2771 loci (N=43) for *Bunomys,* which accounted for >99.5% of all available loci for both species.

F_{ST} values were calculated using *Genepop* (Table 3) and *Arlequin* (Table 4) for both species. Details on the inbreeding coefficient (F_{ST}) can be found in Wright (1943). F_{ST} values in Arlequin were calculated using standard settings (e.g. no Reynold's or Slatkin's distance) for 5000 permutations after Bonferroni correction. As previously mentioned, *Bunomys* and *Maxomys* were split into 3 populations, the details of which can be seen in Table 1 and Figure 1 (Methods Section), of which

Lanowulo (Population Number 3) is the only one derived from mainland Sulawesi, whereas the other populations are both from the island of Buton.

		Kakenauwe	Lambusango	Lanowulo	Mean (with S.D.)
M.hellwaldi	No. of Gene Copies	**15**	**2**	**2**	**6.33 (7.5)**
	No. of Loci	**844**	**844**	**844**	**844 (0)**
	No. of usable loci	**840**	**844**	**844**	**842.67 (2.3)**
B. andrewsi	No. of Gene Copies	**29**	**7**	**7**	**14.33 (12.702)**
	No. of Loci	**5542**	**5542**	**5542**	**5542 (0)**
	No. of usable loci	**5518**	**5526**	**5510**	**5518 (8)**

Table 2. Population Properties calculated by Arlequin (Excoffier 2010)

		Kakenauwe	Lambusango
Bunomys andrewsi (1)	Lambusango	**0.0184**	
	Lanowulo *	**0.9133**	**0.9516**
Maxomys hellwaldi (2)	Lambusango	**0.074**	
	Lanowulo *	**0.668**	**0.5899**

* Lanowulo is the only site located in mainlad Sulawesi.

(1) Fst calculated with 2771 loci, 43 specimen

(2) Fst calculated with 421 loci, 19 specimen

Table 3. Pairwise Fst for all 3 populations calculated by Genepop (Version 4.2)

		Kakenauwe	Lambusango
Bunomys andrewsi (1)	Lambusango	**0.13(0)**	*
	Lanowulo *	**0.71 (0)**	**0.64 (0)**
Maxomys hellwaldi (1)	Lambusango	**0.049 (0.23)**	*
	Lanowulo *	**0.93 (0)**	**0.97 (0.32)**

* Lanowulo is the only site located in mainlad Sulawesi.

Table 4. Pairwise Fst (with P values) for all 3 populations calculated by Arlequin (Excoffier 2010)

2.8. Results: Population Structure at neutral loci

Genetix software (Belkhir *et al*., 2001) was used to perform a Factorial Correspondence Analysis (FCA) on each species. GTX files were created for Genetix using PGDSpider with 422 loci and 19 specimen for *Maxomys* along with 1000 loci and 43 specimen for *Bunomys*. For *B. andrewsi*, the first factor in the PCA, which was initially calculated for 4 factors, accounted for 95.14% of the data's variability while the second factor accounted for the remaining 4.86%. *M. hellwaldii* had a principle factor of 98.9%, with a single remaining factor accounting for the other 1.12%. The FCA results drawn by *Genetix* are presented in Figures 3 and 4 for *M. hellwaldii* and *B. andrewsi* respectively. These charts only take into account the first 2 factors (the third factor was 0% and was therefore omitted, leaving us with 2-dimensional graphs). The x-axis represents the first principle component and the y-axis the second. It is worth noting that scales of the charts vary for each species and both axes do not meet at point (0,0). After calculating FCA, Fis values for each population was estimated in *Genetix*, running for 5000 permutations with 422 loci and 19 specimen for *Maxomys* and 2771 loci and 43 specimen for *Bunomys* (Table 5).

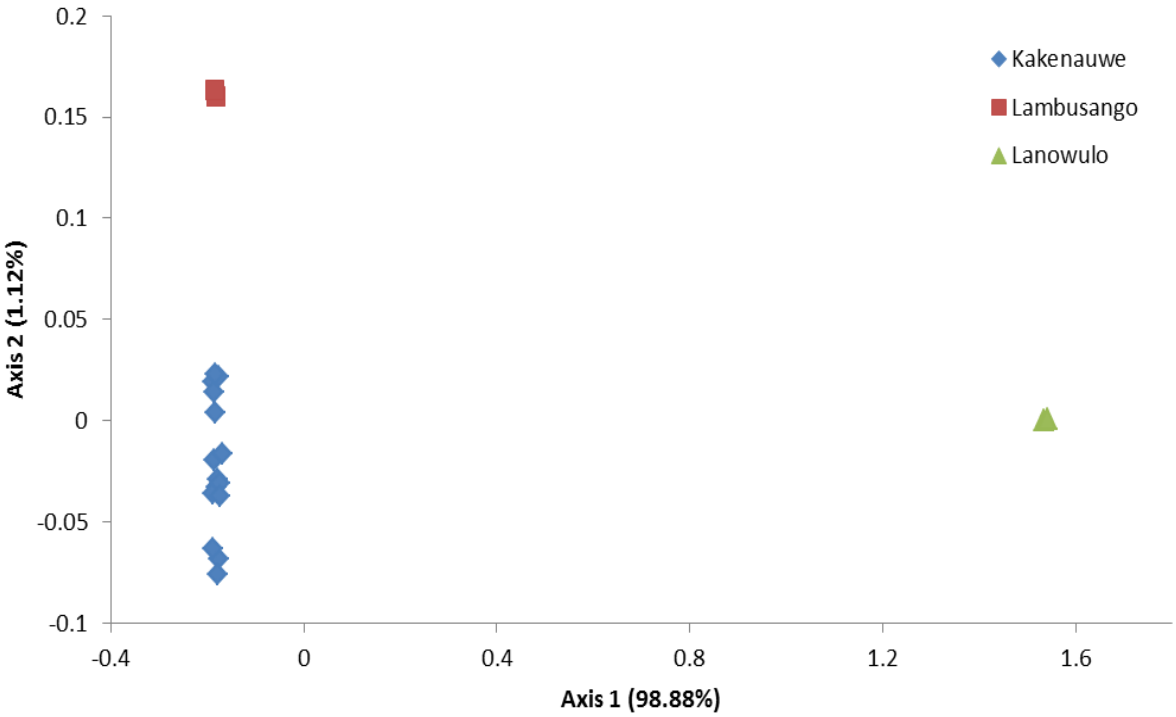


Figure 3. FCA calculated in Genetix (Belkhir *et al*., 2001) for *M. hellwaldi* with 2 principle factors.

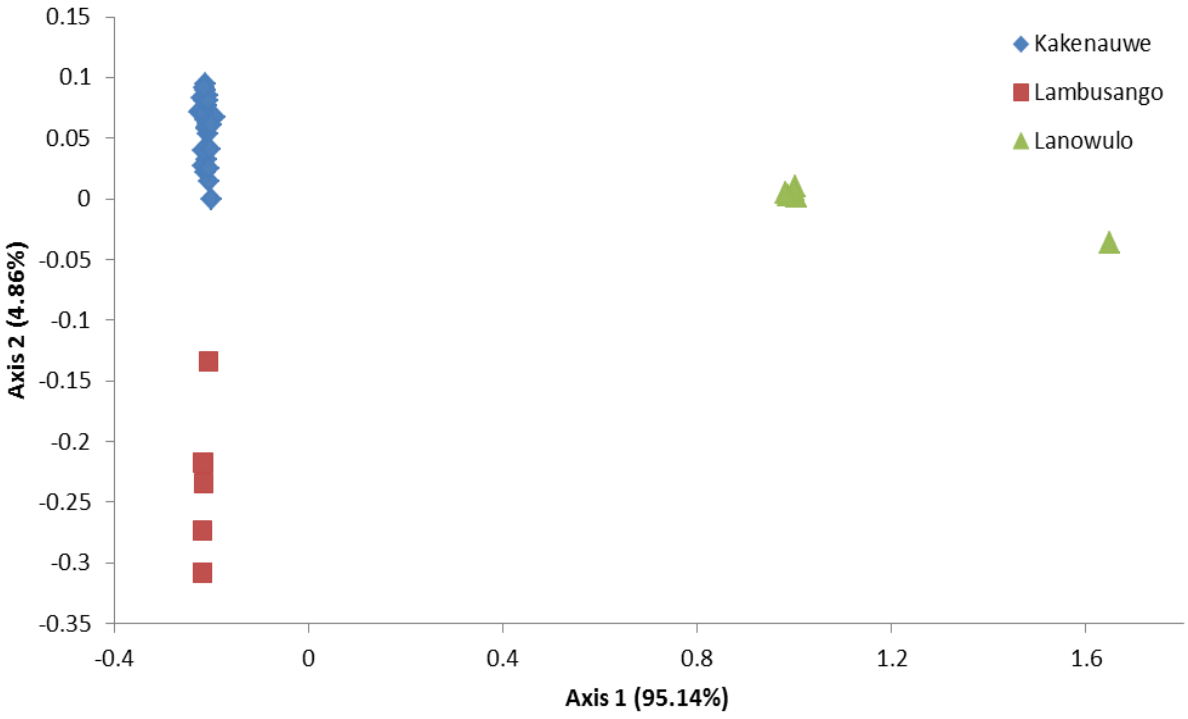


Figure 4. FCA calculated in Genetix (Belkhir *et al*., 2001) for *B. andrewsi* with 2 principle factors.

		N	Fis values
M. hellwaldi	Kakenauwe	15	-0.07392
	Lambusango	2	-0.19048
	Lanowulo	2	-0.75
B. andrewsi	Kakenauwe	32	0.00909
	Lambusango	8	-0.20336
	Lanowulo	7	0.50615

Table 5. Estimates of Fis values for each population provided by Genetix version 4.05 (Belkhir *et al.* 1996). Significance level was estimated by permutations of alleles within each population. Genetix ran for 5000 permutations, using all available loci.

Structure was used to detect underlying genetic populations among a cluster of individuals from multilocus genotypes. Structure was run separately for each species with input files which contained all individuals (diploid data) and the putative population origin for each individual*B. andrewsi* ran with 43 individuals and 2771 loci while *M. hellwaldii* ran with 19 individuals and 422 loci.

To start off with, *Structure* ran once for each separate species testing all values of K between 1 and 10 for a single iteration (5,000 steps with a Burn-in period of 50,000). This provided a broad overview of how probable each value of K was for each species and permits the user to subsequently narrow their approach. In all the *Structure* analyses, four independent runs were carried out for each value of K using an Admixture Model. After evaluating the initial data, *Structure* ran again for each species individually, testing all values of K from 1 to 6 (100,000 MCMC with a Burn-in period of 10,000) (Table 6). Using the Evanno method (Evanno *et al*., 2005), which was performed and visualized on the program Structure Harvester (Earl & vonHoldt, 2012) I concluded that *Maxomys* most probably has a K value of two, although this was uncertain as LnP(K) indicated 3. *Structure Harvester* was used to calculate Delta K, which is the mean of |Ln''(K)| divided by the standard deviation of L(K) (Figure 4). While Delta K = 2, LnP(K) indicated K = 3 for both species, so histograms were drawn for both K values using more stringent parameters (500,000 MCMC with a 10% Burnin

and 4 iterations) (Figures 5 & 6). *Clumpp* was used to combine data from all 4 runs per K value and *DISTRUCT* was used to graphically display the results.

	# K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
M. hellwaldii	1	4	-7380.775	1.3124	NA	NA	NA
	2	4	-2421.125	0.7762	4959.65	3480.525	4484.00648
	3	4	-942	1.1576	1479.125	1505.95	1300.94266
	4	4	-968.825	16.1298	-26.825	48.45	3.003764
	5	4	-1044.1	30.9472	-75.275	101.225	3.270898
	6	4	-1018.15	50.7365	25.95	NA	NA
B. andrewsi	1	4	-56567.775	4.5073	NA	NA	NA
	2	4	-34170.925	14.1085	22396.85	17020.225	1206.38274
	3	4	-28794.3	29.3575	5376.625	5667.5	193.051494
	4	4	-29085.175	191.7534	-290.875	45.4	0.236762
	5	4	-29421.45	403.7172	-336.275	462.1	1.144613
	6	4	-29295.625	223.5068	125.825	NA	NA

Table 6. Delta K and Ln'(K) values calculated by *Structure* for *M. hellwaldi* and *B. andrewsi*

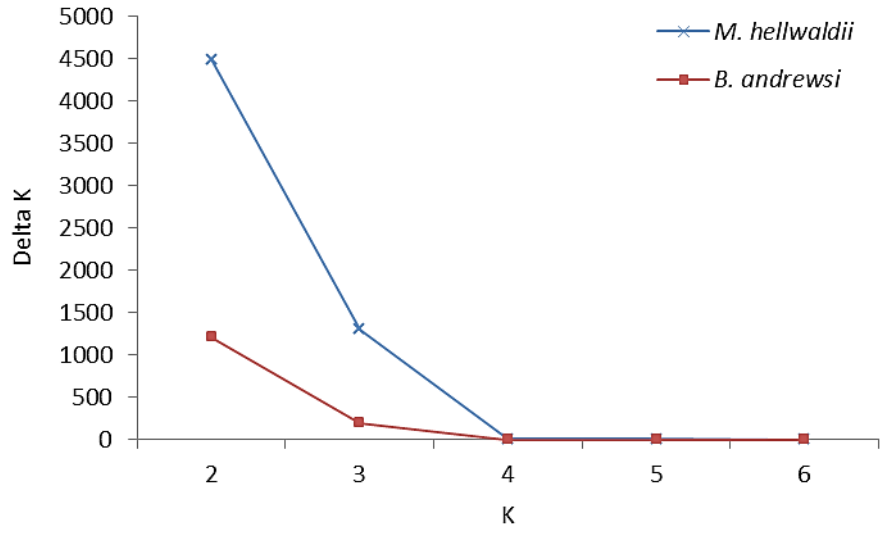


Figure 4. Delta K values against hypothetical cluster number (K) for *M. hellwaldi* and *B. andrewsi*

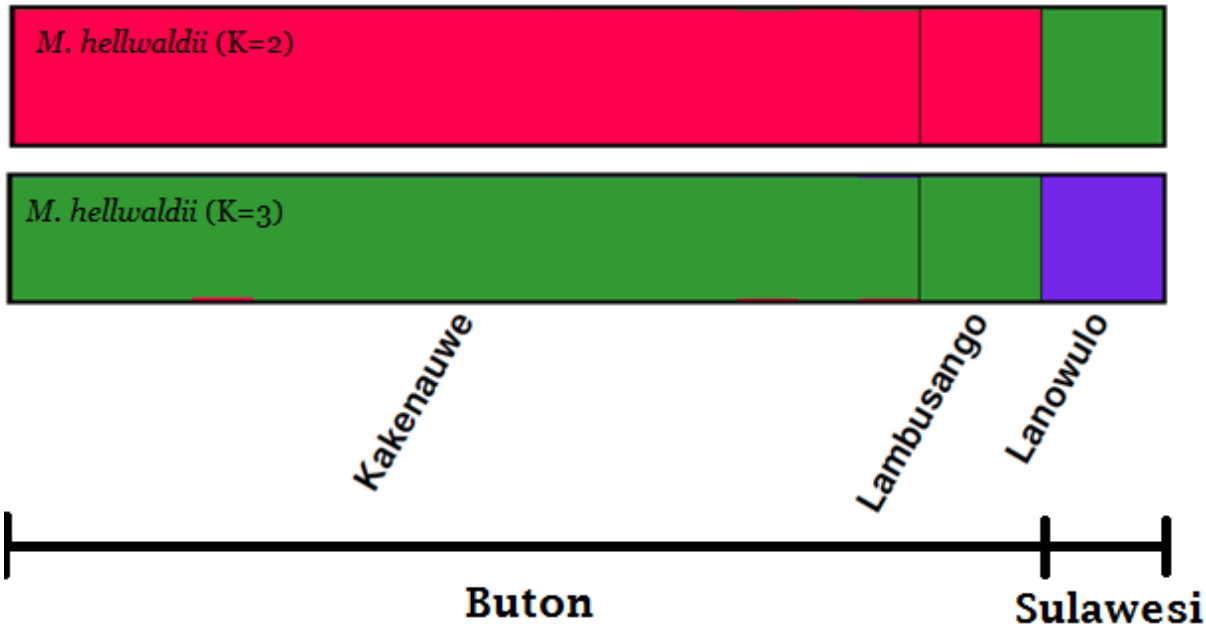


Figure 5. Structure Histograms for *M. hellwaldi* (4 iterations; 500,000 MCMC, 50,000 Burnin) for K values of 2 and 3.

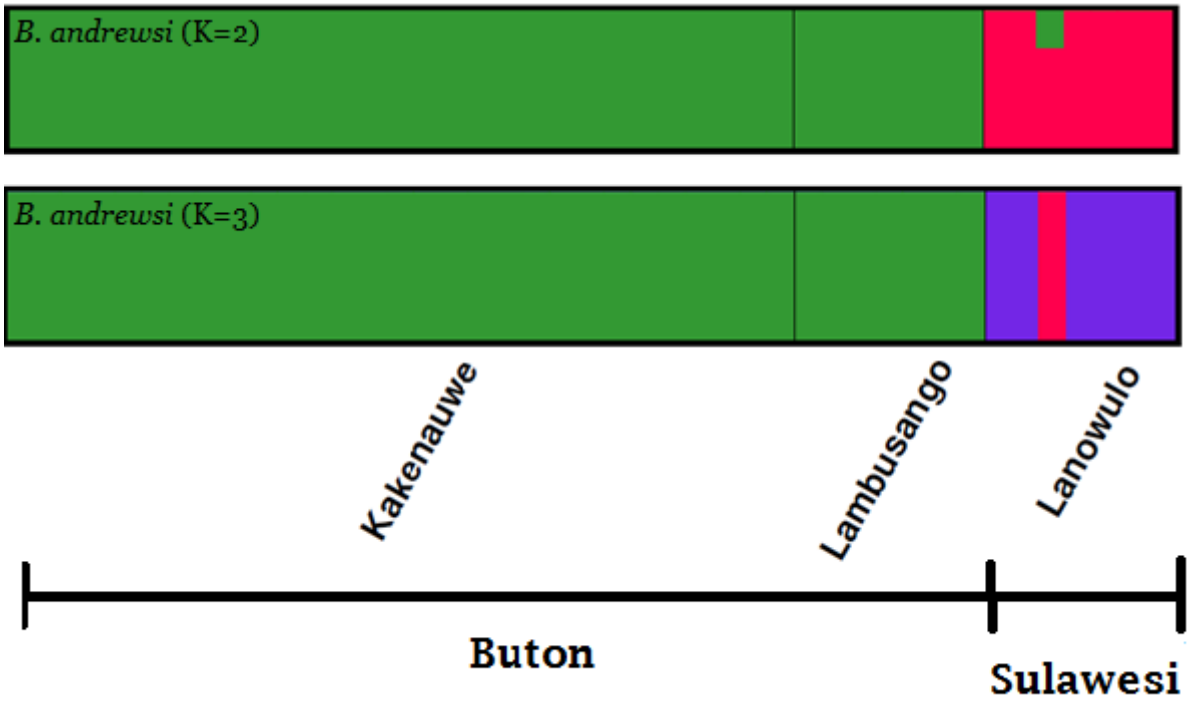


Figure 6. Structure Histograms for *B. hellwaldii* (4 iterations; 500,000 MCMC, 50,000 Burnin) for K values of 2 and 3.

Diploid SNP data for each species was used in *BayesAss (Rannala, 1997)*. Files were created in Immanc format (Rannala & Mountain, 1997) using PGDSpider. *BayesAss* was run for each species individually, first using a small dataset. After evaluating the results, two larger Immanc files were created, one per species and each containing 415 loci. A newer version of *BayesAss* (Windows Version 3.0.) was then used, which ran at a default seed of 10 for 5 Million MCMC iterations with a burn-in length of 500'000 (Table 7). The delta values for allele frequency, migration rate and level of inbreeding were all set to 0.1. The delta value (ranging from 0 to 1) defines the maximum amount a parameter can be changed after each iteration and is ideally <0.35. An increase in delta value leads to an increase in the rejected number of changes in the chain. Additionally, the program was set to compute 'Trace' files which would subsequently be read by the program Tracer (Rambout *et al*., 2014). The number of iterations was progressively increased to 5 Million, at which value Tracer recorded a sufficiently high Effective Sample Size (ESS) for Ln(P) which was 87.475 for *B. andrewsi* and 50.593 for *M. hellwaldi*.

B. andrewsi	m[1][1]:	0.9793(0.0139)	m[1][2]:	0.0104(0.0100)	m[1][3]:	0.0103(0.0100)
	m[2][1]:	0.2765(0.0362)	m[2][2]:	0.7010(0.0301)	m[2][3]:	0.0225(0.0217)
	m[3][1]:	0.0336(0.0301)	m[3][2]:	0.0333(0.0298)	m[3][3]:	0.9331(0.0403)
M. hellwaldii	m[1][1]:	0.9627(0.0239)	m[1][2]:	0.0186(0.0176)	m[1][3]:	0.0186(0.0175)
	m[2][1]:	0.2182(0.0653)	m[2][2]:	0.7396(0.0557)	m[2][3]:	0.0422(0.0400)
	m[3][1]:	0.0653(0.0530)	m[3][2]:	0.0652(0.0527)	m[3][3]:	0.8695(0.0648)

M [i] [j] is the fraction of individuals in population 'I' that are migrants derived from population 'j' per generation
Populations: 1 Kakenauwe (Buton); 2 Lambusango (Buton); 3 Lanowulo (Mainland Sulawesi)

Table 7. Mean Migration Rates with Standard Deviation calculated by BayesAss

2.9. Discussion

The results show that the population structure of *B. andrewsi* and *M. hellwaldi* are both very diverse, even within a small geographic area. There is a clear differentiation between islands for both species. Within Buton, there is a strong directional migration from East to West and from North to South. A greater lineage diversity than initially expected was found within *B. andrewsi*. Strong differentiation was observed in both species, despite of the proximity of the islands and the narrow sea which separates them. The results discourage the idea that gradual divergence over time is responsible for the observed pattern of diversity (the 'Museum Hypothesis'; see Chapter 1). The 'Pleistocene Pump Hypothesis' would most likely explain the high levels of observed diversity.

2.9.1. Validity of Methods

Studying the gene flow between different populations is essential in understanding the micro-evolutionary patterns that exist between separate populations and their geographic structure (Bossart & Prowell, 1998). This study tried to gain some insight into how given populations have diversified, a process which can ultimately lead to speciation. Bossart & Prowell (1998) described two main limitations to genetic estimates of population structure: 1) molecular markers having insufficient resolving power to distinguish patterns of gene flow and 2) limitations imposed by the software used and the mathematical models they rely on. Also, in some complex systems, there can exist different, yet equally valid conclusions drawn from the same set of results. Although these limitations still remain relevant, there has been much progress in both the creation of molecular data and its subsequent data-analysis since the publication of Bossart & Prowell (1998). The RAD data were obtained via Illumina sequencing (which results in very high resolving power) and generally well-reputed and often used software was utilized, addressing each of these issues.

What is difficult to estimate by merely studying population structure (as opposed the molecular phylogeny of a genus) is the age of population divisions. There is a risk that the population

demography of a species is interpreted on an ecological timescale (an incident occurring over the course of a few decades) when events actually happened in the distant past, such as the effects of Pleistocene glaciations and the accompanying range expansions (Larson *et al*., 1984; Merila *et al*., 1997). In this discussion, the focus is on results concerning genetic structure.

2.9.2. Factors causing observed genetic differentiation

In evolutionary history, murids have undergone a relatively recent basal radiation (Schenk *et al*., 2013) and the accentuated diversity of murids has previously been described as occurring in multiple 'bursts' of adaptive radiation (Patterson & Pascual, 1968). Recent diversification studies (which use molecular phylogenies) show varying diversification rates among lineages that are equally aged (Fabre *et al*., 2012; Schenk *et al*., 2013). The differences among populations on Buton are significant but small, while the differences between islands are very substantial.

There is some slight, yet insignificant indication of directional migration from Buton to Sulawesi. While the given data can suggest possible 'source' and 'sink' populations in *Maxomys* and *Bunomys*, Schenk *et al.* (2013) found that movement between populations and biogeographic transitions do not provide sufficient explanation for the current phenotypic diversity in murid rodents.

It is of course impossible to consider all factors, especially external ones, which would account for the observed rates of differentiation. Paradoxically, one driving force which underlies the differentiation (and eventual speciation) of both species is the rate of speciation (and extinction) of all other species with which *Maxomys* and *Bunomys* share their habitat (FitzJohn, 2009); a complex relationship that is unquantifiable. To further demonstrate the complexity of the driving forces, rates of species differentiation have been known to be affected by a multitude of ecological attributes such as diet (Rojas *et al*., 2012; study on bats) dispersal ability (Wollenberg *et al*., 2011; study on frogs) and body size (Wollenberg *et al.,* 2011).

In fact, there have been phylogenetic studies which show a correlation between an increased rate of diversification and a small body size (Cardillo *et al*., 2005). This would obviously not be a unique driving force in this study, yet simply one of many relevant factors. Briefly, the association between small body size and increased levels of species differentiation is due to the large population size and fertility of small mammals, as well as an increased capacity to finely subdivide into microenvironments (Isaac *et al*., 2005). A more well-known factor would be the correlation between elevation and species differentiation. Sulawesi and Buton are both mountainous areas with some peaks above 1000m. Species native to such 'mid-elevation' habitats are particularly subjected to increased differentiation rates (Wiens *et al*., 2007; study on salamanders).

2.9.3. The population structure of *B. andrewsi* & *M. hellwaldii*

F_{ST} calculated by Arlequin and Genepop was low between the two Buton populations for both *Maxomys* and *Bunomys*. The magnitude of F_{ST} will be affected by small sample sizes in some cases, including the comparison among the two putative populations on Buton. However, as FCA places individuals in Euclidian space based on their multi-locus genotypes, small sample sizes are less important. The most prominent principal component (x-axis) from the FCA accentuates the variation between the Sulawesi and Buton populations. The second component (y-axis) represents a very small proportion of the total variance (<2%). However, it is notable that for each species the 2 populations on Buton, separated by a small geographical distance in separate forest fragments and divided by a road, show such clearly separate clusters. At the same time, the BayesAss data shows that the poorly sampled population on Buton (Lambusango) is poorly defined (with values at 0.70 and 0.74 close to the 2/3 minimum set by the program for assignment from a population into the same population; see Table 8). Consistently, there is a strong directional signal for migration from the larger sample (Kakenauwe) into Lambusango. F_{IS} values for the best sampled population are close to zero for both species, as expected, but likely distorted be sampling effects for the populations with smaller sample sizes. More data are needed (especially based on larger sample

sizes) to assess the degree to which these populations should be identified as separate conservation units.

The *Structure* result only provided clear support for differentiation between the two islands for each species, another indication of the relatively weak signal for differentiation among the two populations on Buton. For the *Structure* histogram for *Bunomys* with a K value of 3 (Figure 6), I identified which particular specimen caused the anomaly in the Sulawesi population. The specimen (shown in red in Figure 6) was correctly labelled and should belong within the given population (Lanowulo). In the phylogenetic analysis in the next chapter, this particular specimen diverged from a node basal to the division between Buton and Sulawesi. This is likely a sampling effect, given the small sample size available from the Sulawesi population, and may reflect greater lineage diversity on Sulawesi than reflected in the given sample set.

The strong differentiation for both species between two proximate islands seems less consistent with the idea that gradual divergence over time is responsible for the pattern of diversity observed (the 'Museum Hypothesis'; see Chapter 1). The findings from Chapter 3 on phylogenetics address the question of whether these observed high levels of diversity can be explained via the 'Pleistocene Pump Hypothesis'.

Chapter 3 – Phylogenetics

3.1. Introduction

Biogeographers, evolutionary biologist and ecologists share an interest in learning about patterns of species diversity (Harmon-Threatt & Ackerly, 2013). This chapter centres on using a phylogenetic analysis to pinpoint specific historical events (e.g. periods of glacial climate) which might have caused a pattern in species diversification by simultaneously affecting a multitude of genera, including *Maxomys* and *Bunomys*.

Phylogeography is a relatively new discipline that studies the spatial distribution of gene lineages within a given population or among closely related species (Avise, 2009). It integrates aspects from phylogenetics and biogeography and compares historical hypotheses (e.g. changes in sea level) with the current distribution of gene lineages (Avise, 2009). Phylogeographers are interested in how phylogenetic components and historical events (e.g. continental drift, glacial periods) affect the distribution of present-day gene lineages. A comparative phylogeographic study of multiple species provides valuable insight into the historical and demographic nature of intraspecific evolution. The majority of phylogeographic studies fall in one of three categories: 1) the study focuses on endangered species or species which might require conservation (e.g. Bowen & Karl, 2007), 2) the study is required to answer questions in invasion biology (e.g. Rius *et al*., 2008) or 3) the study's focus is centred on investigating species complexes (e.g. Teske *et al*., 2009). In this phylogeographic assessment I compare two species in shared habitat across a potential barrier to gene flow in the context of historical environmental factors that likely affected that barrier.

The aim of this study falls primarily within category 3 (investigating species complexes) although the results can be valuable for conservation programs, as little is known about the focal species, and endangered species which share the same genus as *MH* and *BA* may share similar phylogeographic histories. Whereas the goal of Chapter 2 was to assess current population structure and differentiation, this chapter considers historical events which shaped the current-day

populations of *MH* and *BA*. Phylogenetic data will be considered in the context of historical events such as past climates (e.g. glacial periods) with the potential to generate barriers and limit dispersal. I compare species from two different genera to test the hypothesis that one single specific event (e.g. a glacial period) caused simultaneous genetic divergence in both genera (c.f. Knowles, 2009; Hickerson *et al*., 2009). A glacial period, or any other significant historical event, could have impacted the evolutionary trajectories and dynamics of a multitude of taxa which can potentially lead to evolutionary divergence among different populations as well as eventual speciation (Klicka & Zink, 1999). The mechanisms which lead to species differentiation have long been the subject of research since the inception of modern evolutionary biology (Simpson, 1953).

I am interested in knowing which event (or which series of events) ultimately caused the development of interbreeding populations and possible reproductive isolation. Quaternary climate events generated significant changes affecting migration corridors and barriers to gene flow (e.g. Avise 2000, Hoelzel 2010). Pleistocene glacial cycles would have caused changes in ocean level, which in turn would affect the physical environment inhabited by *Maxomys* and *Bunomys*. Both Avise & Walker (1998) and Avise *et al*. (1998) regard the Pleistocene species differentiation events as having near global impacts, though some areas of the world would have been more affected than others.

I wish to produce phylogenies with time points that can be interpreted in the context of known geologic phenomenon. A problem which here arises is that in order to pinpoint a geologic event, one must have a good estimate of the mutation rate (e.g. Hoelzel, 2010). In a deep phylogeny the 'molecular clock' may vary across the tree, but calibration points can facilitate node dating, and calibrations have been identified and discussed for the broader phylogeny including the study species (see Fabre et al. 2013).

3.2. Theories on Indonesian Phylogeography

The biogeographical ambivalence of Sulawesi has often been attributed to the island's geological history (Stelbrink *et al*., 2012; Lohman *et al*., 2011). High levels of endemism observed on many island complexes (including Sulawesi) should in part imply that species began to diversify after the isolation of an island. Intra-island diversification was seen in the results from Chapter 2, where two populations on the same island (Kakenauwe and Lambusango on Buton Island) showed significant levels of differentiation. Although the results in this study only cover a small number of populations on the same island, there have been recent studies on island-wide phylogenies in Sulawesi showing differing levels of diversity such as Poettinger & Schubart (2014) and Von Rintelen *et al*. (2014) who studied freshwater crabs and freshwater snails respectively.

Some authors (Shekelle *et al*., 2010; Evans *et al*., 2003) have described the presence of numerous allopatric and parapatric species on Sulawesi as being the result of an initial dispersal event, which was then followed by the isolation of different populations due to fragmentation. Studying the resulting species boundaries, some of which will be on separate islands such as Buton and mainland Sulawesi, should indicate a common set of evolutionary forces. Evans *et al.* (2003) proposed that the defining event which shaped the population structure of many Sulawesi species was the fragmentation of Sulawesi into different islands caused by alternating sea levels during the Pleistocene, a period when glacial cycles reduced sea levels to as low as 120m below present levels (Gaither *et al*., 2010). Spakman & Hall (2010) have produced a movie (*http://searg.rhul.ac.uk/pubs/spakman_hall_2010_banda*) showing a convenient overview of the geological history of Wallacea. Haq *et al.* (1987) provide a chronology of fluctuating sea levels since the Triassic.

My aim is to use phylogenetic methods to construct evolutionary trees in order to assess when *MH* and *BA* started forming distinct populations. There has been a recent growth in the field of molecular phylogenetics concerning murid rodents (Schenk *et al*., 2013; Rowe *et al*., 2011; Jansa *et*

al., 2006; Steppan *et al.,* 2005) including a recent study incorporating the species of this study into a broader phylogeny (Fabre et al. 2013).

3.3. Applications to calculate Genealogy

In the field of evolutionary genetics, the extensive availability of gene sequences and the development of the coalescent theory (Kingman, 1982) have incited the development of 'genealogy'-based methods of inference (methods of deduction based on ancestry), which in turn have modernized computational analyses. Even though the validity of a particular phylogenetic tree can be uncertain, such trees can nonetheless withhold useful information for further analysis (Yang & Rannala, 2012). Yang & Rannala (2012) provide a review of different software (e.g. BEAST, Geneious) used for phylogenetic inference using sequence data, along with an overview of computer processing methods. Next-generation sequencing techniques, such as Illumina sequencing which was used for this dataset, are producing increasingly large quantities of genetic information. The analysis of large amounts of information has come with a new set of issues unique to large data sets, which I have overcome using the methods described in the previous chapter.

3.4. Phylogenetic Trees

The key aspect of phylogenetics is to reconstruct evolutionary histories from observing physiological, morphological and molecular characteristics of given organisms. Although the idea of representing ideas in the form of evolutionary trees is a method dating back to before Darwin, the numerical calculation of phylogenetic trees is a relatively recent venture (Sneath & Sokal, 1973), along with its application to molecular data (Zuckerkandl & Pauling, 1965). The primary objective of molecular phylogenetics is to deduce the order of evolutionary events, such as allopatric speciation, and present the results in the form of a phylogeny that visually represent the relationship of species over a prolonged period of time (Baldauf, 2003). If little evolution occurred, then generating a tree can be relatively straightforward. In practice however, the real evolutionary difference between two given

sequences is rapidly obscured by multiple different mutations or 'changes on top of changes'. In such cases, merely counting the differences between two sequences is due to lead to an underestimate as to how much molecular evolution really occurred.

As phylogenies try to make sense of increasingly large amounts of data, this process can be complicated as there is no direct way to avoid all phylogenetic problems when formulating a tree. Numerous different stochastic and evolutionary models can be presented within a single dataset (Baldauf, 2003). Complications can arise as different lineages could have different mutation rates and be subjected to diverse factors which influence their evolutionary change. The best method to use for any given phylogenetic analysis is dependent on the nature of the study and the type of data collected (Ewens & Grant, 2005). Nowadays, phylogenies are being used by a diversity of branches in biology, which has led to a rapid increase in software packages available to perform phylogenetic analysis. A common method for phylogeny reconstruction involves branch lengths that correspond to the degree of evolution (the percent difference in molecular sequences) between nodes (Linder & Warnow, 2005). The relationship between branch length and real time can sometimes be erratic for any single gene, yet the lengths still give a good overview of the relative rates of change in a phylogenetic tree (Baldauf, 2003). The different types of phylogenetic trees, in particular the difference between Maximum Likelihood and Maximum Posterior Probability in Phylogenetics, are reviewed by Svennblad *et al.* (2006). Molecular clock analyses can be performed using either relaxed or strict clock conditions, which respectively assume either heterogenous or homogenous substitution rates among branches (Stelbrink *et al*., 2012).

3.5. Comparable Phylogenetic Studies

As indicated in Chapter 2, previous systematic reviews have shown that there are substantial gaps in our knowledge of Indonesia's mammalian biodiversity (Musser & Carleton, 2005; Corbet & Hill, 1992). There is little information about the phylogeny of *MH* and *BA* in the existing literature; hence papers on the phylogeny of other species found in Sulawesi and Buton were referred to in order to

make comparisons, such as a study on Sulawesi freshwater fish (Mokodongan & Yamahira, 2015). One could expect that due to the relatively small size of Buton Island, not many diverse populations would occur, however Kingston & Rossiter (2004) demonstrated how a rare species of bat found in several locations on Buton showed distinct populations according to a tree based on parsimony analysis of mtDNA haplotypes. In a non-genetic study, Burton & Nietsch (2010), in a paper about Sulawesi tarsier (*Tarsius spectrum*), showed behavioural differences between specimens observed in Kabaena (an island a few km West off the coast of Sulawesi) and Buton. One key point to emphasize is that Sulawesi was not joined to any other major island (e.g. Malaku) from the Pleistocene onwards (Hall, 2001).

As explained in Chapter 2, it is know that Sulawesi is a highpoint on a very shallow continental shelf (Heaney, 1986). These islands surrounding mainland Sulawesi, such as Buton, would have repeatedly been connected to Sulawesi during periods of low sea levels during the Pleistocene (Voris, 2000). Voris (2000) estimated that the current depth between Sulawesi and Buton is between 50 and 75 meters. As the sea level would have decreased by as much as 120m during the Pleistocene (Gaither *et al*., 2010), it is probable that both islands were connected. The island's complex evolutionary history presented researchers with the opportunity to test multiple hypotheses of diversification and population differentiation. A study on two different species found in Sulawesi by Evans *et al.* (2003) neatly puts two alternative theories in evolutionary structure into perspective: one species (*Macaca tonkeana*; Macaque monkeys) was found to have diversified into several endemic allopatric species and subpopulations, whereas the other species (*Bufo celebensis*; Celebes toad) was considered to be morphologically undifferentiated throughout Sulawesi (but see Bridle *et al*., 2003). These two species have different evolutionary structures due to the influence of differing evolutionary and ecological processes (e.g. toads could be less affected by an ocean barrier than monkeys).

Not all studies supported a role for Pleistocene climatic variation. In a study on South-East Asian gyumnures (genus *Hylomys*), Ruedi and Fumagalli (1996) suggest that adaptive radiation of their species of interest may have predated the climatic changes during the Pleistocene. In a separate study on Indo-Malayan shrews (genus *Crocidura*), Ruedi (1996) theorized that sea level variation only had a minor effect on his data on genetic differentiations. Gorog *et al.* (2004) showed how his results refuted the hypothesis of Late Pleistocene migration in the Sunda Shelf, arguing instead that pre-glacial vacariance is responsible for the evolutionary history of several species of *Maxomys* (yet not specifically *Maxomys hellwaldi,* nor specifically in Sulawesi).

3.6. Node dating for *Bunomys* and *Maxomys*

Concerning *Maxomys* and *Bunomys,* most previous phylogenetic studies have focused on the genus *Maxomys,* although not exclusively for Sulawesi. According to Musser & Charleton (2005), *Maxomys* is also found in Malaysia, Indonesia and the Philippines. Ruedas & Kirsch (1997) suggested that the *Maxomys* genus diverged with *Rattus* 7.6 Mya and that the *Maxomys* genus began to diversify 4.8 Mya. The time when *Maxomys* diverged from *Rattus* (7.6 Mya) is considered as a period of major Sundaic South-East Asian murid radiation, a statement which is backed up by fossil evidence (Ruedas & Kirsch, 1997). Watts & Baverstock (1995) had relatively similar estimates (8 ± 1 Mya) as Ruedas & Kirsch (1997) concerning the South-East Asian murid radiation. This early division is relevant to this study as I included the Norway Rat (*Rattus norvegicus*) as a reference genome. Ruedas & Kirsch (1997) use *Rattus* in a broad sense. Although I settled with 7 Mya as my divergence time between *Maxomys* and the *Bunomys*, *Rattus* lineage, I am aware that this date is based on phylogenetic inference and not a fossil calibration, and so may be inaccurate. During the Pleistocene sea level fluctuations caused mainland Sulawesi to connect with its outlying islands through land-bridges (Voris 2000, Bird *et al*., 2005), and then isolated these islands multiple times. The transient land-bridges would have lasted many thousands of years (den Tex *et al.*, 2010). Sulawesi and its outlying islands became isolated around 10,000-12,000 years ago when sea levels rose to a relatively high

level of approximately 50m below present (Voris, 2000). Voris (2000) suggested that sea levels were 50m below present for around 40% of the time within the period of the last 250,000 years.

I decided to use data given in Fabre *et al*. (2013) to date my nodes when using the program BEAST. *Maxomys hellwaldi* is itself not featured in their data, yet other members of the *Maxomys* genus are. Fabre *et al*. (2013) places the turning point between the Pliocene and the Pleistocene at 1.8 Mya, a figurewhich is no longer widely supported as Gibbard & Head (2009), who were part of a sub-commission on quaternary stratigraphy, officially placed the date at 2.58 Mya. In this study, 2.58 Mya is the figure used for reference in comparison with estimated node dates (Figure 1). Partridge (1997) provides an historical overview on the debate concerning this particular transition between geological ages, explaining the transition back to accepting the earlier date after Haq *et al.* (1977) proposed that the Quaternary base (which is defined as the end of the Pliocene and the start of the Pleistocene) should be dated at approximately 1.6 Mya.

3.7. Vicariance and Dispersal

Sulawesi has a wide range of allopatric and parapatric taxa. It is generally agreed that the presence of such numerous taxa has resulted as the product of range expansion through dispersal, which was subsequently affected by fragmentation or isolation. According to Evans *et al*. (2003), Pleistocene range fragmentation, which was triggered by rising sea levels, remains the most probable cause for the patterns in genetic diversity observed in *MH* and *BA*. It is technically possible that only one of the two species of interest dispersed across available land bridges during relevant periods of the Pleistocene, or the timing and frequency of such movements may have differed for these species. A study on Sulawesi grasshoppers by Bridle *et al.* (2004) was very particular in cautioning against the use of oversimplified evolutionary models, which primarily rely on isolation and exclude alternative explanations and mechanisms. According to revisions performed by Bridle *et al.* (2004), the effect of vicariance was overstated in the aforementioned study by Evans *et al*. (2003). Other studies in Sulawesi by Walton *et al.* (1997) and Bridle *et al.* (2001) were both suggested to have been based on

incorrect biogeographical models. In this study I test the hypothesis that both *Maxomys hellwaldii* and *Bunomys andrewsi* were affected by historical events in a similar way.

3.8. Methods

An extensive description of the laboratory work and computer editing can be found in Chapter 2. Here I present the methods used after creating FASTA and NEXUS files used for the analysis. Details of other programs used (e.g. PGDSpider) can also be found in Chapter 2 and are not repeated here. Even though the scope of phylogenetic studies can vary considerably and require different sets of computational methods, there are three main steps in phylogenetic analysis: 1) the dataset needs to be assembled and appropriately prepared in a predefined format, 2) phylogenetic trees are then subsequently inferred using various computational methods and finally 3) statistical tests are performed to grade the strength of the trees.

Concerning point 2) in particular, there has historically been considerable difficulties in developing statistical applications for phylogenetic inference. Computational biologist have tried to overcome such challenges, yet very often at the loss of some form of analytical accuracy (Ronquist & Huelsenbeck, 2003). One of the key improvements in the field of phylogenetic inference has been the application of Markov chain Monte Carlo (MCMC), which has been regarded as more efficient than the standard Maximum Likelihood bootstrapping approach (Larget & Simon, 1999). At present, the MCMC technique is the computational method most commonly used for fitting Bayesian models (Robert & Casella, 2004). Two reviews of the uses of MCMC methods and general Bayesian Variable Selection Methods are given by O'Hara & Sillanpaa (2009) as well as Cowles & Carlin (1996).

As a starting point to most of the analyses, files in Nexus format were created by using PGDSpider to convert existing Fasta files from the previous studies. Nexus files were created containing data from both species and each species individually. Details on population properties and alignment lengths that were calculated by Arlequin and Geneious can be found in Chapter 2,

Table 2. Some programs (e.g. MrBayes) would not run due to the large size of the data files. In such a case, the Nexus file was shortened, which was achieved by manually excluding random specimens from the populations which contained the most individuals.

Some discrete character data were used in the trees. According to Kumar *et al.* (2004), when using discrete-character methodology, DNA sequences are used to build an evolutionary tree by taking into account the relationships of sequence data at each nucleotide position. These discrete character methods are used to create Maximum Likelihood and Maximum Parsimony trees, which can be visualized in MEGA Version 6.06 (not shown) to help confirm the validity of the NEXUS files and in order to give one a rough overview of the data.

The two main programs used in the analyses were 1) *MrBayes* and 2) *BEAST*. Each of these is accompanied by a suit of supplementary programs (e.g. Tracer; LogCombiner) in order to diagnose, visualize and summarize the results. Brief summaries of the used programs are given below:

1. *MrBayes*, in some ways similar to BEAST, also uses Bayesian MCMC for phylogenetic inference. MrBayes performs Bayesian phylogenetic analysis by combining information obtained from anonymous nuclear data (which was used in this study), data from different partitions or subsets which have evolved under differing stochastic evolutionary models (Ronquist & Huelsenbeck, 2003).
2. *BEAST* is a Bayesian MCMC program which is used to root trees under a chosen type of model (e.g. relaxed clock model) and can produce time-measured phylogenies (Drummond & Rambout, 2007). BEAST uses MCMC to average over tree space, which implies that each of the many trees that BEAST produces will be weighted in proportion to its posterior probability. The program is complemented by additional software such as Tracer, Figtree and TreeAnnotator. BEAUti (Bayesian Evolutionary Analysis Utility) was used for the creation of BEAST XML input files, as instructed by the software's manual (Huelsenbeck & Ronquist, 2001). CIPRES Science Gateway Version 3.3 was used to run BEAST.

The methods used for creating the trees for MrBayes and then BEAST are given below:

MrBayes (Windows Version 3.2; Huelsenbeck & Ronquist, 2001) was run for 20 Million iterations until the standard deviation of clade frequencies was <0.01. The default computational parameters in the MrBayes software were changed to accommodate the relatively large Nexus file used to produce a non-linearized tree from the program. The Nexus file was created in Geneious 8.1 (Kearse *et al.,* 2012) and contained data from both species of interest together with the Norway Rat. During the final MrBayes run, the sample frequency was changed to 1000, as well as print frequency (1000) and diagonal frequency (10,000). The number of chains was increased to 5 and invgamma rate settings were used along with the 'lset' evolutionary model. During preliminary tests using MrBayes, I was unable to generate the appropriate result files, irrespective of the number of iterations used. This issue was addressed by shortening the Nexus file before rerunning the program. The Nexus file was manually shortened by randomly removing sequence data from populations with the most samples. Groups of 6 samples were systematically subtracted from the Nexus file until the shortened file was able to run on MrBayes.

After testing the number of possible samples, I settled with using 36 of the initial 66 samples from my Nexus file. Care was taken to include as many samples as possible by gradually subtracting 6 predefined individuals from the Nexus file until the program was able to run properly. Even small increases in sample size led to unavoidable problems with MrBayes software. Just an increase from 36 to 42 samples caused a malfunction, as the standard deviation did not reach <0.05, irrespective of the number of iterations. After each run, the 'sump' command was used to generate log likelihood values, which is a plot of the generation versus the log probability of the data. I checked whether the plot showed any signs of increasing or decreasing tendencies (in which case I would have discarded it). A summary posterior probability tree was created using 250 Burnin. Figtree was used to visualize the posterior probability tree, where clade credibility values provided by MrBayes were manually added (Figure 1). The Bremer value for all key nodes is 1.

BEAUti files for the program BEAST Version 1.8.2. (Drummond *et al.*, 2012) were created using an uncorrelated lognormal relaxed clock model and the Yule speciation process (Gernhard, 2008) with a random starting tree model. Although several alterations were made, the initial settings for the first few BEAST runs were based on those of Stelbrink *et al.* (2012). Unlike in MrBayes, where the input files required to be shortened, all 66 available specimen from *Maxomys* and *Bunomys* were used for BEAST. The substitution model was set at HKY and a Gamma Site Heterogeneity Model was used with 4 Gamma categories. The 'States' and 'Operators' options were kept at the default values in BEAUti and the sequence error model was kept off. Normal TMRCA priors were used which were initially set at 3.3 and 7 MYA, according to Fabre *et al.* (2013), to represent the Bunomys/Rattus and Maxomys/Bunomys nodes respectively. For my analysis, I tried both different clocks (e.g. Exponential Relaxed Clock) and different priors (e.g. Uniform TMRCA), checking the consensus trees and ESS values in Tracer. When I found that a certain option led to an average increase in Tracer ESS (number of MCMC iterations and all other options being kept equal) I tended to favour those options.

For the final BEAST consensus tree (Figure 2), BEAST was run twice, each time for 50Million iterations, logging parameters every 5000 iterations. A maximum clade credibility tree was created in TreeAnnotator Version 1.8.2. using a 10% Burnin after confirming a ESS value of >50 in Tracer for each individual file. Logcombiner Version 1.8.2. was used to combine the .tre and .log files created from multiple identical runs and the final trees were visualized in Figtree.

3.9. Results

MrBayes was run until the standard deviation of clade frequencies was <0.01.The used Nexus file contained data from both species of interest together with the Norway Rat. Sample frequency was changed to 1000, as well as print frequency (1000) and diagonal frequency (10,000). The number of chains was increased to 5 and invgamma rate settings were used along with the 'lset' evolutionary model. 36 of the initial 66 samples were used. A summary posterior probability tree was created using 250 Burnin. Figtree was used to visualize the posterior probability tree, where clade credibility values provided by MrBayes were manually added (Figure 1). The Bremer value for all key nodes is 1.

BEAUti files for the program BEAST Version 1.8.2. (Drummond *et al.*, 2012) were created using an uncorrelated lognormal relaxed clock model and the Yule speciation process (Gernhard, 2008) with a random starting tree model. All 66 available specimen from *Maxomys* and *Bunomys* were used for BEAST (Figure 2). Normal TMRCA priors were used which were initially set at 3.3 and 7 MYA, according to Fabre *et al.* (2013), to represent the Bunomys/Rattus and Maxomys/Bunomys nodes respectively. The Bremer support for all key nodes is 1 and 95% credible intervals for MRCA estimates were calculated for all key nodes. In Figtee, the cladogram was manually offset by a positive facor of 0.6, which increased the age of the nodes. In Figure 2, one can oberve near the node between *Bunomys* and *Rattus* that one specimen (in orange belonging to the Bunomys population Lanowulo) stands apart from the rest, creating a node of 2.33 between it and the rest of Bunomys. This is the same specimen that caused an unexpected result in the *Structure* histograms in Chapter 2 ("Lanowulo X"). To find out whether Lanowulo X caused any major alteration in node age, BEAST was re-run for *Bunomys* and *Rattus* using the same settings as previously described, yet I manually deleted the aforementioned specimen from my initial Nexus file. The resulting consensus tree (not shown) was calibrated in Figtree so that the node between *Rattus* and *Bunomys* was approximately 3.3. The node between the Lanowulo populations and the Buton Island populations

only underwent minor changes (1.68, as opposed to 1.78) which indicates that “Lanowulo X” specimen did not adversely affect the BEAST tree.

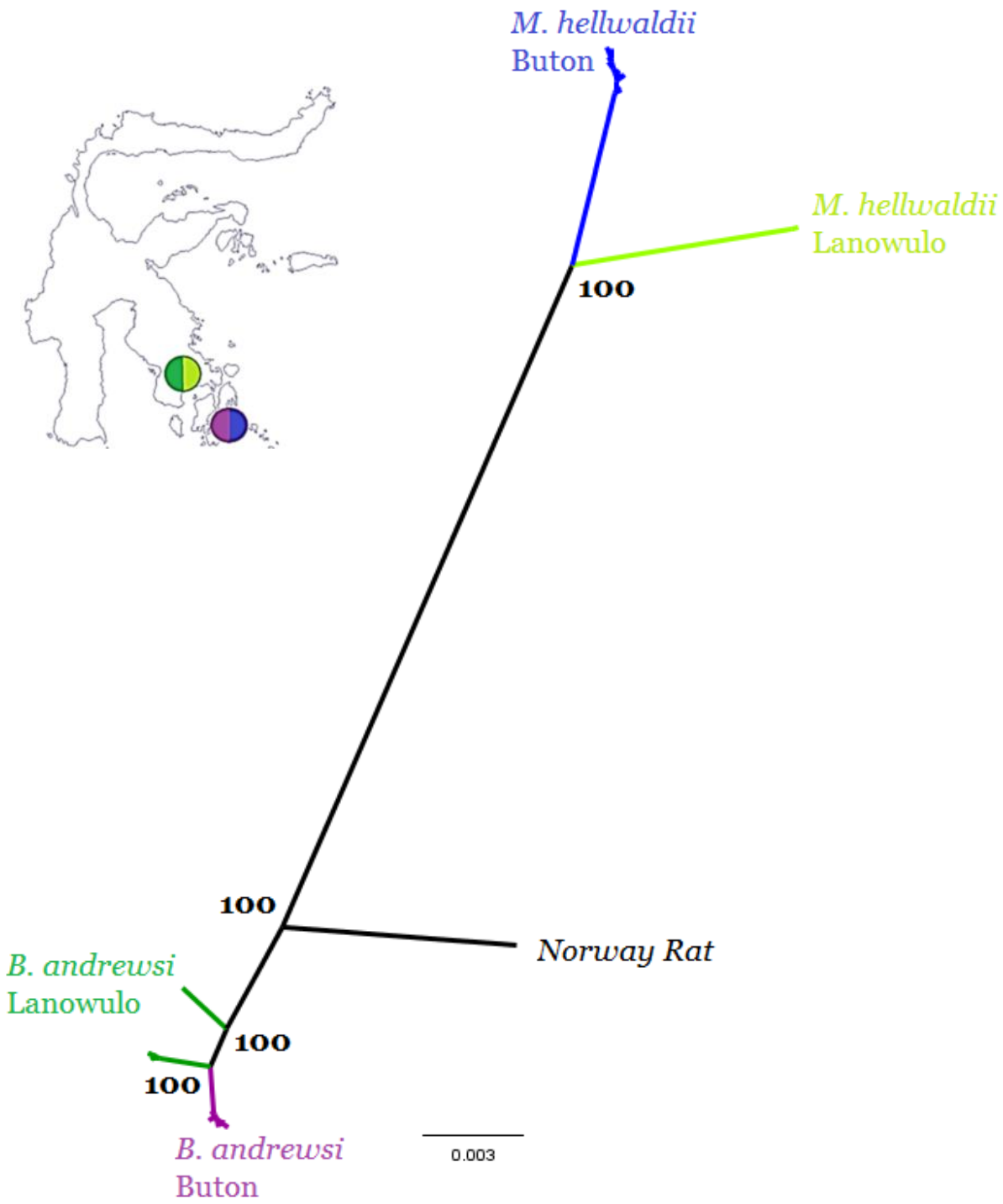


Figure 1. MrBayes unrooted posterior probability tree which ran for 20 Million iterations until achieving a standard deviation of <0.01. Bootstrap values are shown at the nodes and Bremer values are 1 for all key nodes. N=36.

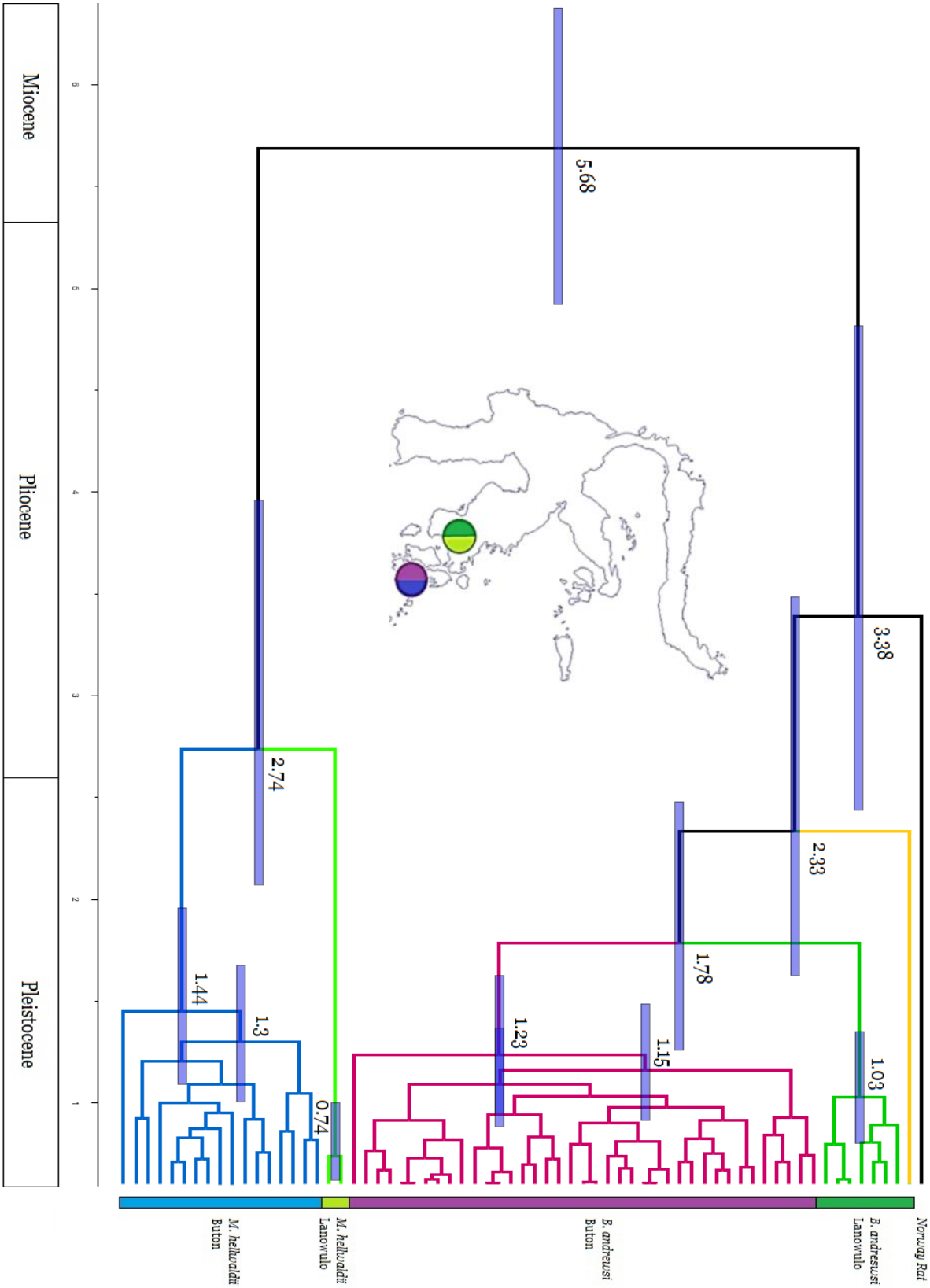


Figure 2. Consensus tree created by BEAST (Drummond *et al.*, 2012) using an uncorrelated lognormal relaxed clock model and the Yule speciation process. Bremer support for key nodes is 1. Bars reprisent 95% credible intervals for MRCA estimates. N=67.

3.10. Discussion

The results show that a single concurrent event occurring during the turn of the Pliocene (between 3.3 and 2.7 Mya) was most likely responsible for the current levels of diversity observed in both species today. Incipient speciation occurred in *B. andrewsi* and *M. hellwaldi* in parallel. The event responsible for the observed pattern is most likely the global rise in temperature and subsequent rise in sea level. It is to be expected that other terrestrial species with habitats encompassing both islands (Buton and mainland Sulawesi) would share a similar evolutionary history. The results confirm the second hypothesis of this study, that a defined geological event equally affected the evolutionary biology of both species of interest. From the trees calculated via MrBayes and BEAST, one specimen within *B. andrewsi* (Lanuwulo X) could represent a deeper lineage within the sampled population.

In a phylogenetic study on South-East Asian murids (not confined to Sulawesi, yet including the genus *Maxomys*) which was conducted by Verneau *et al.* (1998), the author found 5 divergence events which were clustered between 5.5 and 7 Mya, and additional 5 divergence events clustered around 2.7 Mya. For the former, the time was concurrent with the climatic cooling which occurred during the Miocene whereas the more recent divergence is concurrent with the Pliocene to Pleistocene transition. Verneau *et al.* (1998) estimated that *Maxomys* split off from *Rattus* (in the broad sense) sometime between 7.3 and 5.7 Mya, a figure which was consistent with an estimate the authors had on DNA/DNA hybridization data, though Verneau *et al.* (1998) based their study on a single locus without any fossil calibration. I initially expected my BEAST tree to show an early split at around 7 Mya, according to the phylogenetic tree by Fabre *et al.* (2013), yet the split could have occurred more recently. A relatively recent split occurring slightly more than 5Mya can be observed between the genus Rattus (in the broad sense) and the *Melomys* clade in Bryant *et al.* (2011). Although Bryant *et al*. (2011) studied Papuan murids as opposed Indonesian rodents, the geological events would have affected their species of interest in a similar way. Fabre *et al.* (2013) provide

more data-intensive node dating, partly because they used a calibrated tree (including 6 fossil calibrations though none are available for my key nodes). Verneau *et al.* (1998) also hypothesized that *Maxomys* then underwent a period of relatively low speciation until the Early Pleistocene.

The *Bunomys* specimen "Lanowulo X" from mainland Sulawesi could represent a deeper lineage within the population that was by chance only sampled with this one individual. The sample was checked for sequence quality and provenance, and appears to represent a true lineage from the correct geographic area. More samples should be collected to explore its taxonomic position more fully.

A core part of this chapter is centred on comparing the divergence times these two species with geological events. This has been the focus of a number of studies who used divergence time estimates to revise hypotheses which advocated that the current genetic variation found in a given species is attributed to vicariance (e.g. Bartish *et al*., 2011; study on *Sapotaceae;* Schweizer *et al.*, 2010; study on parrots). Crisp *et al*. (2011) made an important point that biogeographical calibrated trees need to be used in conjunction with pre-defined hypotheses, instead of using what they refer to as an "inductive pattern first, process later procedure of fitting geological data on a tree" (Crisp *et al.*, 2011). For this reason I based my calibrations on the data available in Fabre *et al.* (2013). Although stem age estimates are commonly used in biogeographical papers, the approach has been subjected to some criticism by Heads (2011). Heads (2011) argues that age estimates are often underestimates (as opposed to overestimates), yet primarily concerning molecular clocks which were calculated using fossil calibrations (Donoghue & Benton, 2007) which were not directly used to analyse this data. In fact, Heads (2011) recommends relying on well-established geological events rather than fossils when fixing node age. However, Heads (2011) also pointed out that age estimates for geological events can also be imprecise and that researchers sometimes approached their work with an 'a priori acceptance of a vicariance model', which I tried to avoid, partly by testing multiple different TMRCA prior options when creating the BEAST files.

The taxonomic position of *Maxomys* has been controversial. Jansa *et al.* (2006) inferred a sister relationship between *Crunomys* and *Maxomys*, whereas Buzan *et al.* (2011) inferred that *Crunomys* is nested in *Maxomys*. Achmadi *et al.* (2013) found the latter to be the case. There is also a debate going on concerning the phylogeny within the genus *Maxomys* (Ruedas & Kirsch, 1997). For instance, a phylogenetic study by Gorog *et al.* (2004) on two *Maxomys* genera (in South-East Asia, but not exclusively Sulawesi) rejected the idea of Pleistocene migration. With a few exceptions (e.g. Herder *et al.* (2012); Mokodongan & Yamahira, (2015)) it is difficult to find studies conducted in Sulawesi which showed monophyly among groups.

Stelbrink *et al.* (2012) suggested that many species in Sulawesi have undergone a combination of *in situ* diversification and of colonization events. From my results, I was unable to accurately pinpoint any directional movements between Sulawesi and Buton, yet it seems most likely that the ancestral population would have been on the larger island of Sulawesi. If there had been populations on each island during the Pliocene, then the land bridges during the Pleistocene would have promoted mixing. Instead it seems that a colonisation event founded one population from the other once the land bridge was formed (early in the Pleistocene). It is clear that for both *Maxomys* and *Bunomys,* populations on Sulawesi and Buton diverged very early on after it first became possible for a colonisation event to occur (after the extended warm period of the Pliocene) and a subsequent interglacial isolated them again. These populations then remained sufficiently isolated to diverge extensively, forming reciprocally monophyletic incipient species. This is in spite of the fact that connectivity would have been repeatedly possible throughout the Pleistocene during periods of glacial maxima and consequently low sea levels. There must have been some mechanism following the initial colonisation events that for each species kept the new founder and source populations isolated. One possible candidate seems to be altitude, since each island population would have been several hundred meters above sea level, separated by a valley. Apart from inter-island colonization being important for generating the deep evolutionary divergence between populations of *MH* and *BA* (now divisible into two species pairs), the presence of numerous other

species within the genus *Maxomys* emphasises the importance of in situ differentiation accounting for the present levels of diversity. Levels of in situ differentiation and species diversification in Sulawesi can be seen perhaps more obviously in meta-analysis studies such as Stelbrink *et al*. (2012) who studied a range of species such as tarsiers, shrews and macaques.

My species of interest belong to an order which has historically been difficult to study. Jansa *et al.* (2006) showed how the divergence dates of murids can cause significant debate and are influenced by several factors, including the choice of calibration points and which type of data was initially used. Although Jansa *et al.* (2006) did use *Maxomys* as part of their study, only a single specimen was included as part of their analysis (which was not *M. hellwaldi*), hence the study is most informative in relation to murids in general. Gorog *et al*. (2004) recommended primarily using the geological timing of the Sunda shelf fragmentation for any divergence estimation.

As this study has focussed more on the vicariance and isolation of rodent species, a useful subsequent study could, for instance, investigate the effect of adaptive evolution and behavioural ecology on diversification. Webb *et al.* (2002) introduced a novel field of research around 'phylogenetic community ecology', for which this data could have some potential use. The phylogenetic results can also be used to complement the findings of the previous chapter. According to Stevens *et al*. (2012), who studied the phylogenetic community structure of desert rodents, phylogenetic studies can often be useful indicators of the overall ecology of a given area. With data from more species, one would be able to perform a small-scale phylogenetic community ecology study, similar to that of Lanier *et al.* (2013) on Australian vertebrates. This may help reveal the ecological aspects that could have led to persistent isolation early on after colonisation.

A section on the validity of the methods used is given at the beginning of the discussion section of chapter 2. In chapter 1, I introduced the so-called 'Pleistocene Pump' hypothesis (den Tex *et al*., 2010), which endeavours to explain the diversity in Sulawesi via multiple colonization and isolation events during the Pleistocene. More specifically, populations will speciate in allopatry (in

isolation from other populations) during periods of high sea level and expand their population range across different islands during periods of low sea level when the islands are connected (Gorog *et al*., 2003). When this cycle repeats itself several times, there is a strong potential to generate high levels of diversity. The split between populations for both *MH* and *BA* is evidently early on during the Pleistocene, however, even if the precise dating is incorrect, the key point is that the events occurred at a similar time within each species (with node dates having overlapping confidence intervals), and the populations then remained in isolation (as evidenced by the evolution of reciprocal monophyly).

One aspect I could not take into account is how historical vegetation may have affected the evolution of *Bunomys* and *Maxomys*. The vegetation, and therefore the habitat for a murid species, would have changed over the course of the Pliocene and during the Pleistocene. I have often mentioned how sea level changes will affect species diversification, yet the interactions between sea level and climate and their effects on the distribution of an areas biota is both complex and difficult to study (van den Bergh *et al*., 2001). These complications have led to some research on *Maxomys* by Gorog *et al.* (2004) giving broad estimates (such as "mid- to late Pliocene") for the isolation of two Malay Peninsula populations. Although the effect of possible migrations during the mid to late Pleistocene from Southern Philippines via the Sulu Islands has been the topic of much study in Bornean biogeography (Jones & Kennedy, 2008), more research needs to be conducted on migrations to Sulawesi from the Philippines as a possible population source.

Chapter 4 – Conclusion

In order to prevent rapid biodiversity loss, effective allocation of conservation efforts within biodiversity hotspots is required (Pimm *et al*., 2001). The extent of biodiversity in peril can clearly be seen in a study by Sodhi *et al*. (2004), which was bluntly labelled 'Southeast Asian biodiversity: an impending disaster'. These biodiversity hotspots are primarily regions with both high levels of human activity along with a high concentration of endemic species (Myers *et al*., 2000). Regretfully, there seems to be a trend that disproportionately high areas of biodiversity are under disproportionately high threat from adverse human impact. A good case emphasizing the practical need to conserve tropical rainforests in South-East Asia can be found in Woodruff (2010). Wilson *et al.* (2005), in a letter to the journal *Nature* which aimed to prioritize global conservation efforts, has shortlisted Sulawesi as the first and foremost area in South-East Asia which is in most need of conservation. Without such conservation efforts, the remaining tropical diversity in Sulawesi is likely to be lost in near future (Myers *et al*., 2000). With sufficient proof, legislative bodies can declare parts of Sulawesi as protective zones, which would reduce land clearing and logging in such areas (Bruner *et al*., 2001). Admittedly, the effective management and conservation of wildlife is a complex issue, requiring both an understanding of the threats a species faces as well as knowledge of the species itself. Protecting wildlife is essential in protecting ecosystems and ecosystem services, yet equally important for the sake of a species' intrinsic value. Successful conservation efforts should lead to reduced anthropogenic pressures as well as effective legislation, which would only take place when sufficient evidence is presented. I hope my data on diversification can serve as an example of one mechanism underlying the evolution of Sulawesi's rich biota.

The current distribution of protected areas in Sulawesi revolves around a small number of large national parks (such as the Morowali and Lore Lindu National Parks) which are both situated in Central-West Sulawesi. Vast areas in the North-West and Buton remain disproportionately sparsely protected (IUCN, 1992). This study might help give conservation authorities more overview of species range, and promote the designation of more areas of protected forest which would

complement the existing biodiversity in the available reserves. The study of the molecular phylogeography and population structure of a species can further help single out areas with the most genetic endemism (Moritz & Faith, 1998) and provide useful information for conservation management. These results give insight into the ongoing allopatric speciation and the evolutionary history of two species found on separate islands, one on mainland Sulawesi and one on Buton Island. Apart from comparing the phylogeny of 2 species which have received little scientific attention, I trust that my study can serve as an example of the astounding diversity of these Indonesian islands as well as aiding in the resource allocation of future conservation efforts. The results presented here identify two new cryptic species, and identify Buton as a potential source of further unrecognised diversity. These data also identify the importance of the Pliocene to Pleistocene transition towards the evolution of diversity in the region.

The management of conservation strategies has become an increasingly difficult problem due to the continuous novel unearthing of data which changes the way we think about the biogeographical history of Sulawesi (Shekelle *et al*., 2010). More studies which are able to provide a comprehensive review of Sulawesi's biota would greatly aid conservation efforts, as long-term plans could be envisioned which would benefit the majority of the island's biota. This is needed as there will always be exceptions to rules which try to explain general evolutionary processes in Sulawesi and the population structure of most of the island's species. The key point here is that conservation authorities need to find out which evolutionary pathways occurred in the majority of Sulawesi species and base conservation efforts around the population structure which one would then expect, providing transferable inference. At the moment, it seems that population diversity by means of repeated cycles of vicariance and dispersal has first and foremost been responsible for the patterns observed in Sulawesi; authorities should therefore envision conservation strategies with this in mind. Even if this is taken as a fact, it should not in any way discourage future efforts to study the population structure of any other species, but a resolved conservation strategy needs to be

implemented on the basis of the available information, despite gaps in our knowledge of the Sulawesi ecosystem.

Although this paper's data does not challenge any defined trends (Evans *et al*., 2003) that have already been observed in the evolutionary history of Sulawesi species, it can serve as an example of island evolution. Both species have relatively special evolutionary histories: both occurred on separate islands on the paleontological proto-Sulawesi archipelago, which was subjected to periods of connectivity and fragmentation during the Pleistocene leading to a remarkably diverse population structure. Although this evolutionary pattern will be less pronounced with some species, primarily due to their differing ecology and evolutionary processes (such as the Celebes toad studied by Evans *et al.* (2003)), one can expect that future analyses on terrestrial mammals will yield similar results. Although I only studied two different species and am not in the position to make general remarks on the evolution of Sulawesi mammals as a whole, there is increasing evidence that the islands of Sulawesi are subdivided into many regions of endemism, even within the smaller outlying islands such as Buton. These endemic populations are not solely the product of Pleistocene vicariance, as suggested by Evans *et al.* (2003), but have also been influenced by dispersal throughout the history of the proto-Sulawesi archipelago. This view is also taken on by Shekelle *et al.* (2010). In many cases there is likely a pattern of cyclical vicariance by recurrent isolation; where the habitat of a widespread species has been repeatedly fragmented and re-merged due to climate oscillations, generating high levels of diversity (Haffer, 1997). Janssen *et al*. (2009) and Richardson *et al*. (2001) both acknowledge the potential of cyclical vicariance as mechanism for species diversification. It is interesting that the two species investigated here and the two cryptic species discovered seem to present an exception to that broader pattern, instead seeming to generate founder populations that remained isolated.

Another way to obtain more information on the current population structure of *MH* and *BA* would be the use of visible tags (or even genetic markers) which can track the short-term

movements in populations. This would involve a prolonged method which requires the recapture of tagged specimen and suffers from particular limitations (Slatkin, 1985), but would nonetheless provide valuable insight into any directional movement between populations, especially between the Buton populations Kakenauwe and Lambusango. Papers which discuss to what extent genetic tags can be used to track population movement include Palsbøll *et al.* (1997) who used genetic tagging in humpback wales. Non-genetic direct movement studies could also complement this data (Bossart & Prowell, 1998).

In the last 2 decades, there has been a remarkable increase in generating genetic data, and the invention of new sequencing techniques have brought data analysis to the forefront, as opposed to data collection. Coupled with the emergence of new phylogenetic reconstruction methods, researchers have more opportunity to test which processes could have defined the current biodiversity patterns in Sulawesi. The acquired data in this study can help as part of a meta-analysis in biodiversity patterns. More data will still be needed in order to obtain a truly comprehensive overview of Sulawesi's biogeography, including an assessment of the remarkable levels of inter-island diversification which can be seen in this paper's results. It would be beneficial to study unrelated organisms in order to create a small library of biogeographical studies, on which a meta-analysis could be based. Before the advent of molecular studies, biogeographers have come up with many distribution patterns of Wallacea, when debating the origin of Sulawesi's species. Even with molecular methods, a great improvement to this recent study would be additional data on Sulawesi's geology, as opposed to animal biology. My main results are based on the paleogeographic reconstructions of other researchers, which admittedly have their limitations (Hall, 2009), yet at the very least this study is able to add to the extraordinary examples of biodiversity in Sulawesi which have been observed since Wallace (1863), and contribute fresh data to complement modern diversification studies (Kohli *et al*., 2015).

References

Abdo, Z., Crandall, K., Joyce, P. (2004) Evaluating the performance of likelihood methods for detecting population structure and migration. *Molecular Ecology* **13**, 837–851.

Achmadi, A.S., Esselstyn, J.A., Rowe, K.C., Maryanto, I., Abdullah, M.T. (2013) Phylogeny, diversity, and biogeography of Southeast Asian spiny rats (*Maxomys*). *Journal of Mammalogy* **94**(6), 1412–1423.

Akey, J. M. (2004) Population history and natural selection shape patterns of genetic variation in 132 genes*. PLoS Biol.* **2**, e286.

Alpin, K.P., Chesser, R., ten Have, J. (2003) Evolutionary biology of the genus Rattus: Profile of an archetypal rodent pest. *In* 'Rats, mice and people: Rodent biology and management.' Singleton, G.R. (ed.). *Australian Center for International Agricultural Research*. Canberra, Australian Capital Territory, Australia.

Antao, T., Lopes, A., Lopes, R.J., Beja-Pereira, A., Luikart, G. (2008) LOSITAN: A workbench to detect molecular adaptation based on a F_{st}-outlier method. *BMC Bioinformatics* 9, 323.

Aplin, K.P., Helgen, K.M. (2010) Quaternary murid rodents of Timor. Part I: new material of *Coryphomys buehleri* (Schaub, 1937) and description of a second species of the genus. *Bulletin of the American Museum of Natural History* **341,** 1–80.

Amores, A., Catchen, J., Ferrara, A., Fontenot, Q., Postlethwait, J. (2011) Genome evolution and meiotic maps by massively parallel DNA sequencing: Spotted gar, an outgroup for the teleost genome duplication. *Genetics* **188** (4), 799–808.

Amori, G., Gippoliti, S., Helgen, K.M. (2008) Diversity, distribution, and conservation of endemic island rodents. *Quaternary International* **182**, 6-15.

Amori, G., Clout, M. (2003) Rodents on islands: a conservation challenge. *In*: Singleton, G.R., Hinds, L.H., Krebs, C.J., Spratt, D.M. (*eds*.), Rats, Mice and People: Rodent Biology and Management. *ACIAR Monograph* **96**, 63–68.

Avise, J.C. (1994) Molecular Markers, Natural History, and Evolution. Chapman & Hall, New York, USA.

Avise J.C. (2000) Phylogeography: the history and formation of species. Harvard University Press, Cambridge, MA.

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

Avise, J.C., Ball, R.M. (1990) Principles of genealogical concordance in species concepts and biological taxonomy. *Oxford Surveys in Evolutionary Biology* **7**, 45–67.

Avise, J.C., Walker, D. (1998) Pleistocene phylogeographic effects on avian populations and the speciation process. *Proc. R. Soc. Lond. Ser. B.* **265**, 457-463.

Avise, J.C., Walker, D., Johns, G.C. (1998) Speciation durations as Pleistocene effects on vertebrate phyulogeography. *Proc. R. Soc. Lond. Ser. B.* **265**, 1701-1712.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A. (2008) Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers**.** *PLoS One. **3**, e3376.*

Baldauf, S.L. (2003) Phylogeny for the faint of heart: a tutorial. *Trends in Genetics* **19** (6), 345-351.

Bartish, I.V., Antonelli, A. Richardson, J.E., Swenson, U. (2011) Vicariance or long-distance dispersal: historical biogeography of the pantropical subfamily Chrysophylloideae (*Sapotaceae*). *J. Biogeogr*. **38**, 177–190.

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., Bonhomme, F. (2001) GENETIX 4.02, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).

Bergman, C.M., Quesneville, H. (2007) Discovering and detecting transposable elements in genome sequences. *Briefings in Bioinformatics* **8** (6), 382-92.

Bird, M.I., Taylor, D., Hunt, C., (2005) Palaeoenvironments of insular Southeast Asia during the Last Glacial Period: a savanna corridor in Sundaland? *Quaternary Science Reviews* **24**, 2228–2242.

Bossart, J.L., Pashley Prowell, D. (1998) Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends in Evolutionary Ecology* **13** (5), 202-205.

Bowen, B.W., Karl, S.A. (2007) Population genetics and phylogeography of sea turtles. *Mol Ecol.* **16**, 4886–4907.

Bremer, K. (1988) The limits of amino 12 acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* **42**, 795-803.

Bridle, J. R., Pedro, P. M., Butlin, R. K. (2004). Habitat fragmentation and biodiversity: testing for the evolutionary effects of refugia. *Evolution* **58**, 1394–1396.

Brody, J.R., Kern, S.E. (2004) History and principles of conductive media for standard DNA electrophoresis. *Analytical Biochem*istry **333** (1), 1-13.

Brooks, D.R., McLennan, D.A. (1991) Phylogeny. Ecology and Behavior: a Research Program in Comparative Biology. Chicago: University of Chicago Press.

Brown, L.R. (1997). *State of the World 1997: A Worldwatch Institute Report on Progress Toward a Sustainable Society* (14th ed.). W. W. Norton & Company, New York, USA. ISBN 0-393-04008-9.

Brown, R.P., Thorpe, R.S. (1991) Within-island microgeographic variation in body dimensions and scalation of the *skink Chalcides sexlineatus*, with testing of causal hypotheses. *Biological Journal of the Linnean Society* **44**, 47-64.

Brown J.H. (2012) The role of phylogeny in desert rodent community assembly. *J. Anim. Ecol.* **81**, 307–309.

Bruner, A.G., Gullison, R.E. Rice, R.E. and da Fonseca G.A.B. (2001) Effectiveness of parks in protecting tropical biodiversity. *Science* **291**, 125-128.

Bryant, L.M., Donnellan, S.C., Hurwood, D.A., Fuller, S.J. (2011) Phylogenetic relationships and divergence date estimates among Australo-Papuan mosaic-tailed rats from the *Uromys* Division (Rodentia: Muridae). *Zoologica Scripta* **40**, 433–447.

Burton, J.A., Nietsch, A. (2010) Geographical Variation in Duet Songs of Sulawesi Tarsiers: Evidence for New Cryptic Species in South and Southeast Sulawesi. *Int J Primatol* DOI 10.1007/s10764-010-9449-8.

Carleton, M.D., Musser, G.G. (1984) Muroid rodents. *In* S. Anderson and J.K. Jones (eds.) *Orders and Families of Recent Mammals of the World*, pp. 289-379. Wiley, New York, USA.

Catchen, J., Hohenlohe, P., Bassham, S., Amores, A., Cresko, W. (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22** (11), 3124-3140.

Ceballos, G., Brown, J.H. (1995) Global patterns of mammalian diversity, endemism and endangerment. *Conservation Biology* **9**, 559–568.

Chaimanee, Y. (1998) Plio-Pleistocene rodents of Thailand. *Thai Studies in Biodiversity* **3**, 1–303.

Chen, D., Ahlford, A., Schnorrer, F., Kalchhauser, I., Fellner, M. (2008) High-resolution, high-throughput SNP mapping in Drosophila melanogaster. *Nat Methods* **5**, 323–329.

Chevret P., Catzeflis F., Michaux J.R. (2001) Acomyinae: new molecular evidences for a muroid taxon (Rodentia: Muridae). *In*: Denys, C., Granjon, L., Poulet, A. (Eds.), African small mammals. Paris: IRD Editions.

Chuquiyauri, R., Penataro, P., Brouwer, K.C. (2013) Microgeographical differences of Plasmodium vivax relapse and re-infection in the Peruvian Amazon. *American Journal of Tropical Medicine and Hygiene* **89** (2), 326-328.

Curran, L.M., Trigg, S.N., McDonald, A.K., Astiani, D., Hardiono, Y.M., Siregar, P., Caniago, I., Kasischke, E. (2004) Lowland forest loss in protected areas of Indonesian Borneo. *Science* **303**, 1000-1003.

Corbett, G.B., Hill, J.E. (1992) The mammals of the Indonesian region: a systematic review. Oxford University Press, Oxford, UK.

Darlington, P.J. (1957) Zoogeography: the Geographical Distribution of Animals. John Wiley & Sons, Inc., New York.

DeAngelis, M.M., Wang, D.G., Hawkins, T.L. (1995) Solid-phase reversible immobilization for the isolation of PCR products. *Nucleic Acids Research* **23** (22), 4752-4743.

Dehais, C., Eudeline, R., Berrebi, P., Argillier, C. (2010) Microgeographic genetic isolation in chub (Cyprinidae: Squalius cephalus) population of the Durance River: estimating fragmentation by dams. *Ecology of Freshwater Fish* **19** (2), 267-278.

den Tex, R. J., Thorington, R., Maldonado J. E., Leonard, J. A. (2010) Speciation dynamics in the SE Asian tropics: putting a time perspective on the phylogeny and biogeography of Sundaland tree squirrels, Sundasciurus*. Molecular Phylogenetics and Evolution* **55**, 711–720.

den Tex, R. J. (2010) Patterns and Processes of Evolution in Sundaland. Acta Universitatis Upsaliensis. Digital Comprehensive Summaries of Uppsula Dissertations from the Faculty of Science and Technology **835**, 34pp.

Dewi, K., Hasegawa, H. (2010) A new Syphacia Species (Nematoda: Oxyuridae) collected from Bunomys spp. (Rodentia: Muridae) in Central Sulawesi, Indonesia. *The Journal of Parasitology* **96** (1), 125-128.

Donoghue, P. C. J., Benton. M. J. (2007) Rocks and clocks: calibrating the tree of life using fossils and molecules. *Trends Ecol. Evol*. **22**, 424–431.

Drummond, A.J., Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling *trees BMC Evolutionary Biology* **7**, 214.

Durden, L.A., Musser, G.G. (1991) A new species of Sucking Louse (Insecta, Anoplura) from a Montane Forest Rat in Central Sulawesi and a Preliminary Interpretation of the Sucking Louse Fauna of Sulawesi. *American Museum Novitates* **3008**, 10.

Durden, L.A. (1986) The reinfestation of forest rats (*Maxomys musschenboekii*) by epifaunistic arthropods in Sulawesi, Indonesia. *Journal of Tropical Ecology* **2**, 283-286.

Durrett, R. (2008). Probability Models for DNA Sequence Evolution. Springer. ISBN 978-0-387-78168-6.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist*. **7**, 1–26.

Ellerman, J.R. (1961) The fauna of India including Pakistan, Burma and Ceylon. Mammalia (2nd edn.) *Baptist Mission Press, Calcutta*, **3** (2), 1-482.

Esselstyn J.A., Achmadi A.S., Rowe K.C. (2012) Evolutionary novelty in a rat with no molars. *Biol. Lett.* **8**, 990–993.

Evans, B.J., Supriatna, J., Andayani, N., Setiadi, M.I., Cannatella, D.C., Melnick, D.J. (2003). Monkeys and toads define areas of endemism on Sulawesi. *Evolution* **57**, 1436–1443.

Ewens, W.J., Grant R. (2005) Statistical Methods in Bioinformatics. Springer Science and Business Media: New York.

Excoffier, L., Heckel, G. (2006) Computer programs for population genetics data analysis: a survival guide. *Nat Rev Genet* **7** (10), 745-758.

Fabre, P.-H., Pages, M., Musser, G.G., Fitriana, Y.S., Fjeldsa, J., Jennings, A., Jonsson, K.A., Kennedy, J., Michaux, J., Semiadi, G., Supriatna, N., Helgen, K.M. (2013) A new genus of rodent from Wallacea (Rodentia: Muridae: Murinae: Rattini), and its implication for biogeography and Indo-Pacific Rattini systematics. *Zoological Journal of the Linnean Society* **169**, 408–447.

Falush, D., Stephens, M., Pritchard, J. K. (2003). Inference of population structure: Extensions to linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587.

Felsenstein, J. (1985) Confidence limits on phylogenetics: An approach using the 16 bootstrap. *Evolution* **39**, 783-791.

Fernandez, J., Villanueva, B., Pong-Wong, R. & Toro, M. A. (2005) Efficiency of the use of pedigree and molecular marker information in conservation programs. *Genetics* **170**, 1313–1321.

Ferris, S.D., Sage, R.D., Huang, C.-M., Nielsen, J.T., Ritte,U., Wilson, A.C. (1983) Flow of mitochondrial DNA across a species boundary. *Proc. Natl. Acad. Sci. USA* **80**, 2290-2294.

FitzJohn R.G., Maddison W.P., Otto S.P. (2009) Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Syst. Biol.* **58**, 595–611.

Flannery, T. (1995) Mammals of the South-West Pacific and Moluccan Islands. Comstock/Cornell University Press, Ithaca, New York, USA, pp. 1–464.

Foote, A.D., Newton, J., Piertney, S.B., Willerslev, E., Gilbert, M.T.P. (2009) Ecological, morphological and genetic divergence of sympatric North Atlantic killer whale populations. *Molecular Ecology* **18**, 5207–5217.

Francis, C. M. (2008) A Field Guide to the Mammals of South-East Asia: Thailand, Peninsular Malaysia, Singapore, Myanmar, Laos, Vietnam and Cambodia. New Holland Publishers, London. *pp. 1-392.*

Frankham, R., Ballou, J.D., Briscoe, D.A. (2002) *Introduction to Conservation Genetics*. Cambridge University Press, Cambridge, UK.

Gaither, M.R., Toonen, R.J., Robertson, D.R., Planes, S., Bowen, B.W. (2010) Genetic evaluation of marine biogeographical barriers: perspectives from two widespread Indo-Pacific snappers (*Lutjanus kasmira* and *Lutjanus fulvus*) *Journal of Biogeography* **37**, 133–147.

Geffen, E., Rowe, K.C., Yom-Tov,Y. (2011) Reproductive Rates in Australian Rodents are related to phylogeny. *PLoS ONE* **6** (4): e19199.

Gernhard, T., Hartmann, K., Steel, M. (2008) Stochastic properties of generalised Yule models, with biodiversity applications. *J. Math. Biol.* **57**, 713-735.

Gibbard, P., Head, M J. (2009) The Definition of the Quaternary System/Period and the Pleistocene Series/Epoch. *Quaternaire* **20** (2), 125-133.

Glaubitz, J.C. (2004) CONVERT: a user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Mol. Ecol. Notes* **4**, 309–310.

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

Gompert, Z., Forister, M.L., Fordyce, J.A., Nice, C.C., Williamson, R.J., Buerkle, C.A. (2010) Bayesian analysis of molecular variance in pyrosequences quantifies population genetic structure across the genome of *Lycaeides* butterflies. *Molecular Ecology* **19**, 2455–2473.

Goodall, G. (ed.) (1943) *Philips' International Atlas.* pp.91-92. George Philip and Son map 'East Indies' London, UK.

Gorog, A., Sinaga, M., Engstrom, M.D. (2003) Vicariance or dispersal? Historical biogeography of three Sunda shelf murine rodents (Maxomys surifer, Leopoldamys sabanus and Maxomys whiteheadi). *Biological Journal of the Linnean Society* **81**, 91–109.

Graur, D., Li, W-H. (2002) Fundamental of Molecular Evolution. *(2nd ed)*. Sinauer Associates, Massachussets. pp. 1-439.

Groves, C.P. (1976) The origin of mammalian fauna of Sulawesi (Celebes). *Zeitschrift fuer Saeugetierkunde* **41**, 201-216.

Groves, C.P. (2001) Mammals in Sulawesi: where did they come from and when, and what happened to them when they got there? *In*: Metcalfe, I., Smith, J.M.B., Morwood, M., Davidson, I. (eds) 'Faunal and floral migrations and evolution in SE Asia-Australasia'. Balkema, A.A., Netherlands, pp 333–342.

Haffer, J. (1997) Alternative models of vertebrate speciation in Amazonia: An overview. *Biodiversity and Conservation* **6**, 451-476.

Hall, R. (2009) Indonesia, geology. *In:* Gillespie, R., Clague, D., eds. Encyclopedia of islands. Berkeley, CA. *University of California Press*, 454–460.

Hall, R. (2001). Cenozoic reconstructions of SE Asia and the SW Pacific: Changing patterns of land and sea. In I. Metcalf, J. Smith, M. Morwood, & I. Davidson (Eds.), Faunal and floral migrations and evolution in SE Asia-Australia (pp. 35–56). Lisse: Swets and Zeitlinger

Hall, R., Cottam, M., Wilson, M.E.J. (2011) The SE Asian Gateway: history and tectonics of Australia-Asia collision. London: Geological Society of London (Special Publication).

Hanebuth, T.J.J., Stattegger, K., Grootes, P.M. (2000) Depositional sequences on a late Pleistocene-Holocene tropical siliciclastic shelf (Sunda Shelf, southeast Asia). *Journal of Asian Earth Sciences* **23**, 113-126.

Haq, B.U., Berggren W.A., Van Couvering, J.A. (1977) Corrected age of the Pliocene/Pleistocene boundary *Nature* , **269**, 483-488

Harmon-Threatt A.N., Ackerly D.D. (2013) Filtering across spatial scales: phylogeny, biogeography and community structure in bumble bees. *PLoS One*. **8**, e60446.

Hasnawir, Omura H, Kubota T. (2006) Landslide disaster at Mt. BawakaraengnCaldera. South Sulawesi. Indonesia. *Kyushu Journal of Forest Research*. **59**, 269-272.

Heads, M. (2011) Old taxa on young islands: a critique of the use of island age to date island-endemic clades and calibrate phylogenies. *Syst. Biol*. **60**, 204–218.

Heaney, L.R. (1986) Biogeography of mammals in SE Asia: estimates of rates of colonization, extinction and speciation. *Biological Journal of the Linnean Society* **28**, 127–165.

Heaney, L.R., Walsh, J.S. Jr., Peterson, T. (2005) The roles of geological history and colonization abilities in genetic differentiation between mammalian populations in the Philippine archipelago. *Journal of Biogeography* **32**, 229-247.

Hendry, A.P., Wenburg J.K., Bentzen P., Volk E.C., Quinn T.P. (2000) Rapid evolution of reproductive isolation in the wild: evidence from introduced salmon. *Science* **290**, 516-518.

Hennig, W. (1966). Phylogenetic Systematics. Univ. Illinois Press, Urbana.

Herder, F., Hadiaty, R.K., Nolte, A.W. (2012) Pelvic-fin brooding in a new species of riverine ricefish (Atherinomorpha: Beloniformes: Adrianichthyidae) from Tana Toraja, Central Sulawesi, Indonesia. *Raffles Bull. Zool*. **60**, 467–476.

Hickerson, M.J., Carstens, B.C., Cavender-Bares, J. (2000) Phylogeography's past, present, and future: 10 years after Avise. *Mol Phylogenet Evol.* **54**, 291–301.

Hillis, D.M., Bull, J.J. (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analyses. *Syst. Biol.* **42**, 182–192.

Hoelzel, A.R. (ed.) (1998) *Molecular Genetic Analysis: A Practical Approach,* 2nd edn. Oxford University Press, Oxford, UK.

Hoelzel, A.R. (2010) Looking backwards to look forwards: conservation genetics in a changing world. *Conservation Genetics* **11**, 655-660.

Hoelzel, A.R., Dahlheim. M. and Stern, S.J. (1998) Low genetic variation among killer whales (*Orcinus orca*) in the eastern North Pacific, and genetic differentiation between foraging specialists. *J. Hered.* 89:121-128

Hoelzel, A.R., Hey, J., Dahlheim, M.E., Nicholson, C., Burkanov, V., Black, N. (2007) Evolution of population structure in a highly social top predator, the Killer Whale. *Mol Biol Evol.* **24**, 1407-1415.

Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A., Cresko, W.A. (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* **6** (2), e1000862.

Hubert, M, P.J. Rousseeuw, S.V. Aelst. (2005) Multivariate outlier and detection of robustness *In* : C.L. Rao, E.G. Wegman, J.L. Solka. (eds). *Handbook of Statistics* **24**, 263 – 302. Elsevier, New York.

Huelsenbeck, J.P., Ronquist, F. (2001) MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755.

Isaac N.J., Jones K.E., Gittleman J.L., Purvis A. (2005) Correlates of species richness in mammals: body size, life history, and ecology. *Am. Nat.* **165**, 600–607.

Jansa, S.A., Barker, F.K., Heaney, L.R. (2006) The pattern and timing of diversification of Philippine endemic rodents: evidence from mitochondrial and nuclear gene sequences. *Systematic Biology* **55,** 73–88.

Janssens, S.B., Knox, E.B., Huysman, S., Smets, E.G., Merkx, V.S.F.T. (2009) Rapid radiation of Impatiens (Balsaminaceae) during Pliocene and Pleistocene: Result of a global climate change. *Molecular Phylogenetics and Evolution* **52**, 806-824.

Jennings, A.P., Seymour, A.S., Dunstone, N. (2005) Ranging behaviour, spatial organization and activity of the Malay civet (*Viverra tangalunga*) on Buton Island, Sulawesi. *Journal of Zoology* **268**, 63-71.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28** (12), 1647-1649.

Kettlewell, H.B.D. (1973) *The Evolution of Melanism*. Clarendon Press, Oxford, UK.

Kingman, J. F.C. (1982) On the genealogy of large populations. *J. Appl. Probab.* **19A**, 27–43.

Kingston, T., Rossiter, S.J. (2004) Harmonic-hopping in Wallacea's bats. *Nature* **429**, 654-657.

Kitchener, D.J., Maryanto, I. (1997) New species of Melomys (Rodentia, Muridae) from Kai Besar. I. Maluku Tengah, Indonesia. *Record of the Western Australian Museum* **16** (3), 427-436.

Kincka, J., Zink, R.M. (1999) Pleistocene effects on North American songbird evolution. *Proc. R. Soc. Lond. Ser. B.* **266**, 695-700.

Knowles, L.L. (2009) Statistical phylogeography. *Annu Rev Ecol. Evol. S.,* **40**, 593–612.

Kohli, B.A., Fedorov, V.B., Waltari, E., Cook, J.A. (2015) Phylogeography of a Holarctic rodent (Myodes rutilus): testing high-latitude biogeographical hypotheses and the dynamics of range shifts. Journal of Biogeography (J. Biogeogr.) **42**, 377-389.

Kraus, R.H.S., vonHoldt, B., Cocchiararo, B., Harms, V., Bayeri, H. (2015) A single-nucleotide polymorphism-based approach for rapid and cost-effective genetic wolf monitoring in Europe based on noninvasively collected samples. *Molecular Ecology Resources* **15** (2), 296-305.

Kumar, S., Tamura, K and Nei, M. (2004) MEGA 3: Integrated software for Molecular Evolutionary Genetics Analysis And Sequence Alignment. *Briefings in Bioinformatics* **5**, 150-163.

Lambertini, M. (2000) *A Naturalist's Guide to the Tropics, Revised edn.* University of Chicago Press, Illinois, USA.

Lanier H.C., Edwards D.L., Knowles L.L. (2013) Phylogenetic structure of vertebrate communities across the Australian arid zone. *J. Biogeogr*. **40**, 1059–1070.

Larget, B., Simon, D. (1999) Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Mol. Biol. Evol.* **16**, 750–759.

Larson, A., Wake, D.B., Yanev, K.P. (1984) Measuring gene flow among populations having high levels of genetic fragmentation. *Genetics* **106** (2), 293-308.

Lewontin, R.C. (1974) The Genetic Basis of Evolutionary Change. Columbia University Press, New York, USA.

Li, H., Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25**, 2078-9.

Li, H. (2012) Exploring single-sample SNP and INDEL calling with whole-genome de novo assembly. *Bioinformatics* **28**, 1838-1844.

Life Technologies Corporation (2009) PureLink™ PCR Micro Kit - For rapidly purifying and concentrating DNA (Manual). http://www.lifetechnologies.com

Linder, C.R., Warnow, T. (2005) "An overview of phylogeny reconstruction." In the Handbook of Computational Molecular Biology, Chapman and Hall/CRC Computer & Information Science.

Lisher, H.E., Excoffier, L. (2012) PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28** (2), 298-299.

Lohman, D.J., de Bruyn, M., Page, T., von Rintelen, K., Hall, R., Ng P.K.L., Shih, H.-T., Carvalho, G.R., von Rintelen, T. (2011) Biogeography of the indo-Australian Archipelago. *Annual Review of Ecology, Evolution, and Systematics* 42, 205-226.

MacPhee, R.D.E., Fleming, C. (2003) A possible heptaxodontine and other caviidan rodents from the Quaternary of Jamaica. *American Museum Novitates* **3422**, 1–42.

Mayr, E. (1947) Ecological Factors in Speciation. *Evolution* 1 (4), 263-288.

Mayr, E. (1982) Speciation and macroevolution. *Evolution* **36**, 1119-1132.

Majerus, M.E.N. (1998) *Melanism: Evolution in Action*. Oxford University Press, Oxford, UK.

Manel, S., Gaggiotti, O.E., Waples, R. (2005) Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology & Evolution* **20**, 136–142.

Martin, T.E., Blackburn, G.A. (2010) Impacts of Tropical Forest Distrurbance Upon Avifauna on a Small Island with High Endemism: Implications for Conservation. *Conservation and Society* **8** (2), 127-139.

Maryanto, I, Kitchener D.J., Prijono S.N. (2005) Morphological analysis of house mice, *Mus musculus* (Rodentia, Muridae) in Southern and Eastern Indonesia and Western Australia. *Mammal Study* **30**, 53 – 63.

Marrison-Scott, T.C.S., Ellerman, J.R. (1951) Checklist of Palaearctic and Indian mammals, 1759 to 1946. British Museum of Natural History, London, UK, pp. 810.

Mayr, E. (1942) Systematics and the origin of species. Columbia University Press, pp. 334.

McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., Brumfield, R.T. (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution* **66**, 526–538.

McCracken K.G., Bulgarella M., Johnson K.P. (2009) Gene flow in the face of countervailing selection: adaptation to high-altitude hypoxia in the beta A hemoglobin subunit of yellow-billed pintails in the Andes. *Molecular Biology and Evolution* **26**, 815–827.

McNeely, J.A., Miller, K.R., Reid, W.V., Mittermeier, R.A., Werner, T.B. (1990) *Conserving the World's Biological Diversity.* IUCN, World Resources Institute, Conservation International, WWF-US and World Bank the, Washington DC., USA.

Medway, L., Yong H.-S. (1976) Problems in the systematics of the rats (Muridae) of Peninsular Malaysia. *Malaysian Journal of Science* **4,** 1–4.

Meirmans, P.G. (2014) Nonconvergence in Bayesian estimation of migration rates. *Molecular Ecology Resources* **14**, 726-733.

Mercer, J.M., Roth, V.L. (2003) The effects of Cenozoic global change on squirrel phylogeny. *Science* **299**, 1568–1572.

Merila, J., Bjorklund, M., Baker, A.J. (1997) Historical Demography and Present Day Population Structure of the Greenfinch, Carduelis chloris-An Analysis of mtDNA Control-Region Sequences. *Evolution* **51** (3), 946-956.

Merker, S., Groves, C.P. (2006) *Tarsius lariang*: a new primate species from western Central Sulawesi. *International Journal of Primatology* **27**, 465-485.

Michel, C., Hicks, B.J., Stölting, K.N., Clarke, A.C., Stevens, M.I., Tana, R., Meyer, A., van den Heuvel, M.R. (2008). Distinct migratory and non-migratory ecotypes of an endemic New Zealand eleotrid (Gobiomorphus cotidianus) – implications for incipient speciation in island freshwater fish species. *BMC Evolutionary Biology* **8**, 49.

Miller, M.R. Dunham, J.P. Amores, A. (2007) Rapid and cost-effective polymorphism identificationand genotyping using restriction site associated DNA (RAD) markers. *Genome Res.* **17**, 240-248.

Missone, X. (1969) African and Indo-Australian Muridae. *Mus. Roy. l'Afrique Cen., Tervuren, Zool.* **172**, 1-219.

Miyamoto, M.M., Cracraft, J. (1991). Phylogenetic Inference, DNA Sequence Analysis and the Future of Molecular Systematics. *In*: Miyamoyo, M.M. and 145 Cracraft, J. *(eds).* Phylogenetic Analysis of DNA Sequence. Oxford University Press, Oxford. *pp. 1-13.*

Mokodongan, D.F., Yamahira, K. (2015) Origin and intra-island diversification of Sulawesi endemic Adrianichtyidae. *Molecular Phylogeny and Evolution* **93**, 150-160.

Monk, K.A., Fretes, Y., Reksodiharjo-Lilley, G. (1996). *The Ecology of Nusa Tenggara and Maluku*. Periplus Editions Ltd, Hong Kong, China (S.A.R.). ISBN 962-593-076-0.

Moss, S.J., Wilson, E.J. (1998) Biogeographic implications of the Tertiary palaeogeographic evolution of Sulawesi and Borneo. Biogeography and Geological Evolution of SE Asia, pp. 133-163.

Moura, A.E., Natoli, A., Rogan, E., Hoelzel, A.R. (2013a) Evolution of functional genes in cetaceans driven by natural selection on a phylogenetic and population level. *Evolutionary Biology*, **40** (3), 341-354.

Moura, A., Natoli, A., Rogan, E., Hoelzel, A.R. (2013) Atypical panmixia in a European dolphin species (*Delphinus delphis*): implications for the evolution of diversity across oceanic boundaries. *Journal of Evolutionary Biology* **26** (1), 63-75.

Moura, A.E., Kenny, J.G., Chaudhuri, R., Hughes, M.A., Welch, A., Reisinger, R.R., Nico de Bruyn, P. J., Dahlheim, M.E., Hall, N., Hoelzel, A.R. (2014a) Population genomics of the killer whale indicates ecotype evolution in sympatry involving both selection and drift. *Molecular Ecology,* doi: 10.1111/mec.12929.

Moura, A.E., Kenny, J.G., Chaudhuri, R., Hughes, M.A., Reisinger, R.R., de Bruyn, P.J.N., Dahlheim, M.E., Hall, N., Hoelzel, A.R. (2014b) Phylogenomics of the killer whale indicates ecotype divergence in sympatry. *Heredity,* doi:10.1038/hdy.2014.67.

Musser, G.G. (1987) The Mammals of Sulawesi. T.C. Whitmore (ed.) *Biogeographic Evolution of the Malay Archipelago*. Oxford University Press, New York, USA.

Musser, G.G. (1990) Sulawesi Rodents: Species Traits and Chromosomes of Haeromys minahassae and Echiothrix leucura (Muridae: Murinae). *American Museum Novitates*. ISSN 0003-0082.

Musser, G.G. (1991) Sulawesi Rodents: Description of New Species of *Bunomys* and *Maxomys* (Muridae, Murinae) *American Museum Novitates*. ISSN 0003-0082.

Musser, G.G. (2014) A systematic review of Sulawesi *Bunomys* (Muridae, Murinae) with the description of two new species. *Bulletin of the American Museum of Natural History* **392**, pp. 313.

Musser, G.G., Carleton, M.D. (2005) Family muridae. '*Mammal Species of the World: A taxonomic and geographic reference*' (eds. Wilson, D.E., Reeder, D.M.) pp. 894-1531. The John Hopkins University Press, Baltimore, USA. ISBN 978-0-8018-8221-0.

Musser, G.G., Holden, M.E. (1991) Sulawesi rodents (Muridae: Murinae): Morphological and geographical boundaries of species in the *Rattus hoffmanni* group and a new species from Pulau Peleng. Pp. 322-413, *In* Contributions to mammalogy in honour of Karl F. Koopman (T. A. Griffiths and D. Klingener, eds.). *Bulletin of the American Museum of Natural History* **206**, 1-432.

Musser, G.G., Newcomb, C. (1983) Malaysian murids and the giant rat of Sumatra. *Bulletin of the American Museum of Natural History* **174**, 327-598.

Myers, N., Mittermeier, R.A. Mittermeier, C.G. da Fonseca, G.A.B., Kent.J. (2000) Biodiversity hotspots for conservation priorities. *Nature* 403, 853–858.

O'Brien T.G., Kinnaird M.F. (2000) Differential vulnerability of large birds and mammals to hunting in North Sulawesi, and the outlook for the future. *In*: Robinson J.G. and Bennett E.L. (eds) '*Hunting for Sustainability in Tropical Forest*'. Columbia University Press, New York, USA. pp. 199–213.

O'Donovan, G. (2001) Report on the botanical and ecological status of the Kakenauwe and Lambusanga Nature reserves on Buton Island, Sulawesi. Operation Wallacea report.

Orr, H.A. and Presgraves, D.C. (2000) Speciation by postzygotic isolation: Forces, genes and molecules. *Bioessays* **22**, 1085–1094.

Paetkau, D., Calvert, W., Stirling, I., Strobeck, C. (1995) Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology* **4**, 347–357.

Page, R.D.M. (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**, 357-358.

Palsball, P.J., Allen, J., Berube, M., Clapham, P.J., Feddersen, T.P. (1997) Genetic tagging of humpback whales. *Nature* **388**, 767-769.

Partridge, T.C. (1997) Reassessment of the position of the pio-pleistocene boundary: Is there a case for lowering it to the Gauss-Matuyama Paleomagnetic reversal? *Quaternary International* **40**, 5-10.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E. (2012) Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species *PLoS ONE* **7**(5), e37135.

Pimm, S. L., Ayres, M., Balmford, A. Branch, G. ,Brandon, K. (2001) Can we defy nature's end? *Science* **293**, 2207–2208.

Pritchard, J. K., Rosenberg N. A. (1999) Use of unlinked genetic markers to detect population stratification in association chain. *Am. J. Hum. Genet.* **65**, 220–228.

Pritchard, J. K., Stephens, M., Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.

Quicke, D.L.J. (1997) Principles and Techniques of Contemporary Taxonomy. *Systematic Biology* **44**, 120-122.

Tamura, K. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.,* **28**, 2731–2739.

Rambaut, A., Suchard, M.A., Xie, D., Drummond, A.J. (2014) Tracer v1.6, Available from http://beast.bio.ed.ac.uk/Tracer

Randi, E., Lucchini, V., Diong, C.H. (1996) Evolutionary genetics of the suiformes as reconstructed using mtDNA sequences. *J. Mamm. Evol.* **3**, 163–94.

Raup, D.M. (1986) Biological extinctions in earth history. *Science* **231**, 1528-1533.

Rawe, K.C., Reno, M.L., Richmond, D.M., Adkins, R.M., Steppan, S.J. (2008) Pliocene colonization and adaptive radiations in Australia and New Guinea (Sahul): Multilocus systematic of the old endemic rodents (Muroidea: Murinae) *Molecular Phylogenetics and Evolution* **47**, 84-101.

Richardson, J.E., Pennington, R.T., Pennington, T.D., Hollingsworth, P.M. (2001) Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science* **293**, 2242-2245.

Robert, C., Casella, G. (2004). Monte Carlo Statistical Methods. New York: Springer-Verlag, 2nd edition.

Robyt, J.F., White, B.J. (1990) Biochemical Techniques Theory and Practice. *Waveland Press Inc., Illinois, USA*. ISBN 0-88133-556-8.

Röder, B., Frühwirth, K., Vogl, C., Wagner, M., Rossmanith, P. (2010) Impact of Long-Term Storage on Stability of Standard DNA for Nucleic Acid-Based Methods. *J. Clin. Microbiol*. **48** (11), 4260-4262.

Rojas D., Vale Á., Ferrero V., Navarro L. (2012) The role of frugivory in the diversification of bats in the Neotropics*. J. Biogeogr*. **39**, 1948–1960.

Ronquist, F., Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19** (12), 1572-1574.

Rosenbaum B., O'Brien T.G., Kinnaird M.F. and Supriatna J. (1998) Population densities of Sulawesi black macaques (*Macaca nigra*) on Bacan and Sulawesi, Indonesia: effects of habitat disturbance and hunting. *American Journal of Primatology* **44**, 89–106.

Rosenberg, N.A. (2004) Distruct: a program for the graphical display of population structure. *Mol. Ecol. Notes* **4**, 137–138.

Ross, K.S., Haites, N.E., Kelly, K.F. (1990) Repeated freezing and thawing of peripheral blood and DNA in suspension: effects on DNA yield and integrity. *J. Med. Genet*. **27**, 569-570.

Rousset, F. (2008) Genepop'007: a complete re-implementation of the genepop software for Windows and Linux*. Molecular Ecology Resources* **8** (1), 103-106.

Rowe K.C., Aplin K.P., Baverstock P.R., Moritz C. (2011) Recent and rapid speciation with limited morphological disparity in the genus *Rattus*. *Systematic Biology* **60**, 188– 203.

Ruedas, L.A., Kirsch, J.A.W. (1997) Systematics of Maxomys Sody, 1936 (Rodentia: Muridae: Murinae) DNA/DNA hybridization studies of some Borneo-Jarvan species and allied Sundaic and Australia-Papuan genera. *Biological Journal of the Linnean Society* **61**, 385-408.

Ruedas, L.A., Musser, G. (2008) Bunomys andrewsi. *IUCN 2013*. IUCN Red List of Threatened Species. Version 2013.1.

Ruedi, M. (1996) Phylogenetic evolution and biogeography of Southeast Asian shrews (genus Crocidura: Soricidae). *Biol. J. Linnean Soc.* **57**

Ruedi, M., Fumagalli, L. (1996) Genetic structure of Gymnures (genus *Hylomys*; *Erinaceidae*) on continental islands of Southeast Asia: historical effects of fragmentation. *Journal of Zoological Systematics and Evolutionary Research* **34**, 153–162.

Rundle, H.D., Nosil, P. (2005) Ecological speciation. *Ecol. Lett*. **8**, 336–352.

Sathiamurthy, E., Voris, H.K. (2006) Maps of Holocene Sea Level Transgression and Submerged Lakes in the Sunda Shelf. *Nat. Hist. J. Chulalongkorn Univ.* Suppl. 2.

Sandeson, M.J. (1989) Confidence limits on phylogenies: The bootstrap revisited. *Cladistics* **5**, 113–129.

Sawyer, S., Hartl, D (1981). "On the evolution of behavioural reproductive isolation: The Wallace effect." *Theoretical Population Biology*, **19** (2), 261-273.

Schaffner, S.F. (2005) Calibrating a coalescent simulation of human genome sequence variation. *Genome Res*. **15**, 1576–1583.

Schenk J.J., Rowe K.C., Steppan S.J. (2013) Ecological opportunity and incumbency in the diversification of repeated continental colonizations by muroid rodents. *Syst. Biol.* **62**, 837– 864.

Schluter, D. (2001) Ecology and the origin of species. Trends in Ecology & Evolution **16** (7), 372-380.

Schmitt, L.H., Kitchener D.J., How R.A. (1995) A genetic perspective of mammalian variation and evolution in the Indonesian archipelago: biogeographic correlates in the fruit bat genus Cynopterus. *Evolution* **49**, 399–412.

Schweizer, M., Seehausen, O., Guntert, M., Hertwig. S. T. (2010) The evolutionary diversification of parrots supports a taxon pulse model with multiple trans-oceanic dispersal events and local radiations. *Mol. Phylogenet. Evol*. **54**, 984–994.

Seymour, A. (ed.) (2006) Summary of biological and sociological research carried out by Operation Wallacea in the forest of Central Buton. Operation Wallacea, London, UK.

Seymour, A. (ed.) (2004) Monitoring forest degradation and animal populations in the forest of Central Buton: preliminary results from the pilot study. Operation Wallacea, London, UK.

Shekelle, M., Meier, R., Wahyu, I., Ting, N. (2010) Molecular Phylogenetics and Chronometrics of Tarsiidae Based on 12S mtDNA Haplotypes: Evidence for Miocene Origins of Crown Tarsiers and Numerous Species within the Sulawesian *Clade. Int. J. Primat*. DOI 10.1007/s10764-010-9457-8

Shendure, J., Ji, H. (2008) Next-generation DNA sequencing. *Nature Biotechnology*, **26**, 1135–1145.

Simonson T.S., Yang Y.Z., Huff C.D. (2010) Genetic evidence for high-altitude adaptation in Tibet. *Science* **329**, 72–75.

Simpson, G. G. (1953) The major features of evolution. Columbia Univ. Press, New York.

Singer, H.A., E. Purwanto (2006) Misteri kekayaan hayati hutan Lambusango. Program konservasi hutan Lambusango (PKHL). Bau Bau: Operation Wallacea Trust.

Slatkin, M. (1985) Gene flow in natural populations. *Annu. Rev. Ecol. Syst*. **25**, 45-69.

Smith, T.B., Wayne, R.K. (1996) *Molecular Genetic Approaches in Conservation*. Oxford University Press, New York, USA.

Sneath, P.H.A., Sokal, R.R. (1973) Numerical Taxonomy, W.H. Freeman.

Sodhi, N.S., Koh, L.P., Brook, B.W., Ng, P.K.L. (2004) Southeast Asian biodiversity: an impending disaster. *Trends in Ecology and Evolution* **19**, 654-660.

Soltis, P.S., Soltis, D.E. (2003) Applying the Bootstrap in Phylogeny Reconstruction. *Statistical Science* **18** (2), 256–267.

Sorenson, M.D., Franzosa, E.A. (2007) TreeRot, version 3. Boston University, Boston, MA.

Spakman, W., Hall. R. (2010) Surface deformation and slab-mantle interaction during Banda Arc subduction rollback. *Nat. Geosci*. **3**, 562–566.

Sparks, J.S, Smith, W.L. (2004) Phylogeny and biogeography of the Malagasy and Australasian rainbowfishes (Teleostei: Melanotaenioidei): Gondwanan vicariance and evolution in freshwater. *Mol. Phylogenet. Evol.* **33**, 719–34.

Stelbrink, B., Albrecht, C., Hall, R., von Rintelen, T. (2012). The biogeography of Sulawesi revisited: is there evidence for a vicariant origin of taxa on Wallace's 'anomalous island'? *Evolution* **66**, 2252–2271.

Steppan, S.J., Adkins, R.M., Spinks, P.Q., Hole C. (2005) Multigene phylogeny of the Old World mice, Murinae, revelas distinct geographical lineages and the declining utility of mitochondrial genes compared to nuclear genes. *Molecular Phylogenetics and Evolution* **37**, 370-388.

Stevens R.D., Gavilanez M.M., Tello J.S., Ray D.A. (2012) Phylogenetic structure illuminates the mechanistic role of environmental heterogeneity in community organization. *J. Anim. Ecol*. **81**, 455–462.

Stickney, H.L., Schmutz, J., Woods, I.G., Holtzer, C.C., Dickson, M.C. (2002) Rapid mapping of zebrafish mutations with SNPs and oligonucleotide microarrays. *Genome Res* **12**, 1929–1934.

Svennblad, B., Erizon, P., Oxelman, B., Britton, T. (2006) Fundamental Differences Between the Methods of Maximum Likelihood and Maximum Posterior Probability in Phylogenetics. *Syst. Biol.* **55** (1), 116-121.

Swofford, D. L. (2003) PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* **30**, 2725-2729.

Teske, P.R., McLay, C., Sandoval-Castillo. J. (2009) Tri-locus sequence data reject a "Gondwanan origin hypothesis" for the African/South Pacific crab genus Hymenosoma. *Mol Phylogenet Evol.* **53**, 23–33.

United Nations, Department of Economic and Social Affairs, Population Division (2013) *World Population Prospects The 2012 Revision: Highlights and Advance Tables*. Working Paper No. ESA/P/WP.228.

Van den Bergh, G.D., de Vos, J., Sondaar, P.Y. (2001) The Late Quaternary palaeogeography of mammal evolution in the Indonesian Archipelago. *Paleogeography, Palaeoclimatology, Palaeoecology* **171**, 385-408.

Vander Wall, S. B., K. M. Kuhn & M. J. Beck (2005) Seed removal, seed predation, and secondary dispersal. *Ecology* **86**, 801–806.

van Oosterzee, P. 1997. Where Worlds Collide: The Wallace Line. Ithaca, NY: Cornell Univ. Press. 234 pp.

van Orsouw, N.J., Hogers, R.C., Janssen, A., Yalcin, F., Snoeijers, S. (2007) Complexity reduction of polymorphic sequences (CRoPS): a novel approach for large-scale polymorphism discovery in complex genomes. *PLoS ONE* **2**, e1172.

Verneau, O., Catzeflis, F., Furano, A.V. (1998) Determining and dating recent rodent speciation events using L1 (LINE-1) retrotransposons. *Proceedings of the National Academy of Science USA* **95**, 11284-11289.

Vila, C., Savolainen, P., Maldonado, J.E., Amorim, I.R., Rice, J.E., Honeycutt, R.L., Crandall, K.A., Lundeberg, J., Wayne, R.K. (1997) Multiple and ancient origins of the domestic dog. *Science* **276**, 1687-1698.

Vines, T. H., Schluter, D. (2006) Strong assortative mating between allopatric sticklebacks as a by-product of adaptation to different environments*. Proc. R. Soc. Lond. B* **273**, 911–916.

Voris, H.K. (2000) Maps of Peistocene sea levels in Southeast Asia: shorelines, river systems, and time durations. *Journal of Biogeography* **27**, 1153-1167.

Wallace, A. R. (1863) On the physical geography of the Malay Archipelago. *J. Roy. Geogr. Soc.* **33**, 217–234.

Walpole, M.J., Goodwin, H.J. (2001) Local attitudes towards conservation and tourism around Komodo National Park, Indonesia. *Environmental Conservation* **28** (2), 160–166.

Walton, C., Butlin, R. K., & Monk, K. A. (1997). A phylogeny for grasshoppers of the genus *Chitaura* (Orthoptera: Acrididae) from Sulawesi, Indonesia, based on mitochondrial DNA sequence data. *Biological Journal of the Linnean Society* **62**, 365–382.

Waser P.M., Ayers J.M. (2003) Microhabitat use and population decline in banner-tailed kangaroo rats. *J. Mammal.* **84**, 1031–1043.

Webb C.O., Ackerly D.D., McPeek M.A., Donoghue M.J. (2002) Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.* **33**, 475–505.

Weiher E., Freund D., Bunton T., Stefanski A., Lee T., Bentivenga S. (2011) Advances, challenges and a developing synthesis of ecological community assembly theory. *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 2403–2413.

Wheeler, P. (2004). Estimating anoa abundance. *In:* Seymour, A. (ed.). Monitoring forest degradation and animal populations in the forests of Central Buton: Preliminary results from the pilot study. Unpublished report - Operation Wallacea.

Whittaker, R.J., Araújo, M.B., Jepson, P., Ladle, R.J., Watson, J.E.M., Willis, K. (2005) Conservation biogeography: assessment and prospect. *Diversity and Distributions* **11**, 3-23.

Whitten, T., Henderson, G.S., Mustafa, M. (2002) The Ecology of Indonesia Series 4. The Ecology of Sulawesi. Periplus Editions, Hong Kong, China (S.A.R.).

Wicks, S.R., Yeh, R.T., Gish, W.R., Waterston, R.H., Plasterk, R.H. (2001) Rapid gene mapping in *Caenorhabditis elegans* using a high density polymorphism map. *Nat Genet* **28**, 160–164.

Wiens J.J., Parra-Olea G., García-París M., Wake D.B. (2007) Phylogenetic history underlies elevational biodiversity patterns in tropical salamanders. **Proc. R. Soc. B Biol. Sci. 274**, 919–928.

Wilson, G.A., Rannala, B. (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163** (3), 1177-1191.

Wollenberg K.C., Vieites D.R., Glaw F., Vences M. (2011) Speciation in little: the role of range and body size in the diversification of Malagasy mantellid frogs. *BMC Evol. Biol*. **11**, 217.

Woodruff, D.S. (2010) Biogeography and conservation in Southeast Asia: how 2.7 million years of repeated environmental fluctuations affect today's patterns and the future of the remaining refugial-phase biodiversity. *Biodiversity and Conservation* **19**, 919-941.

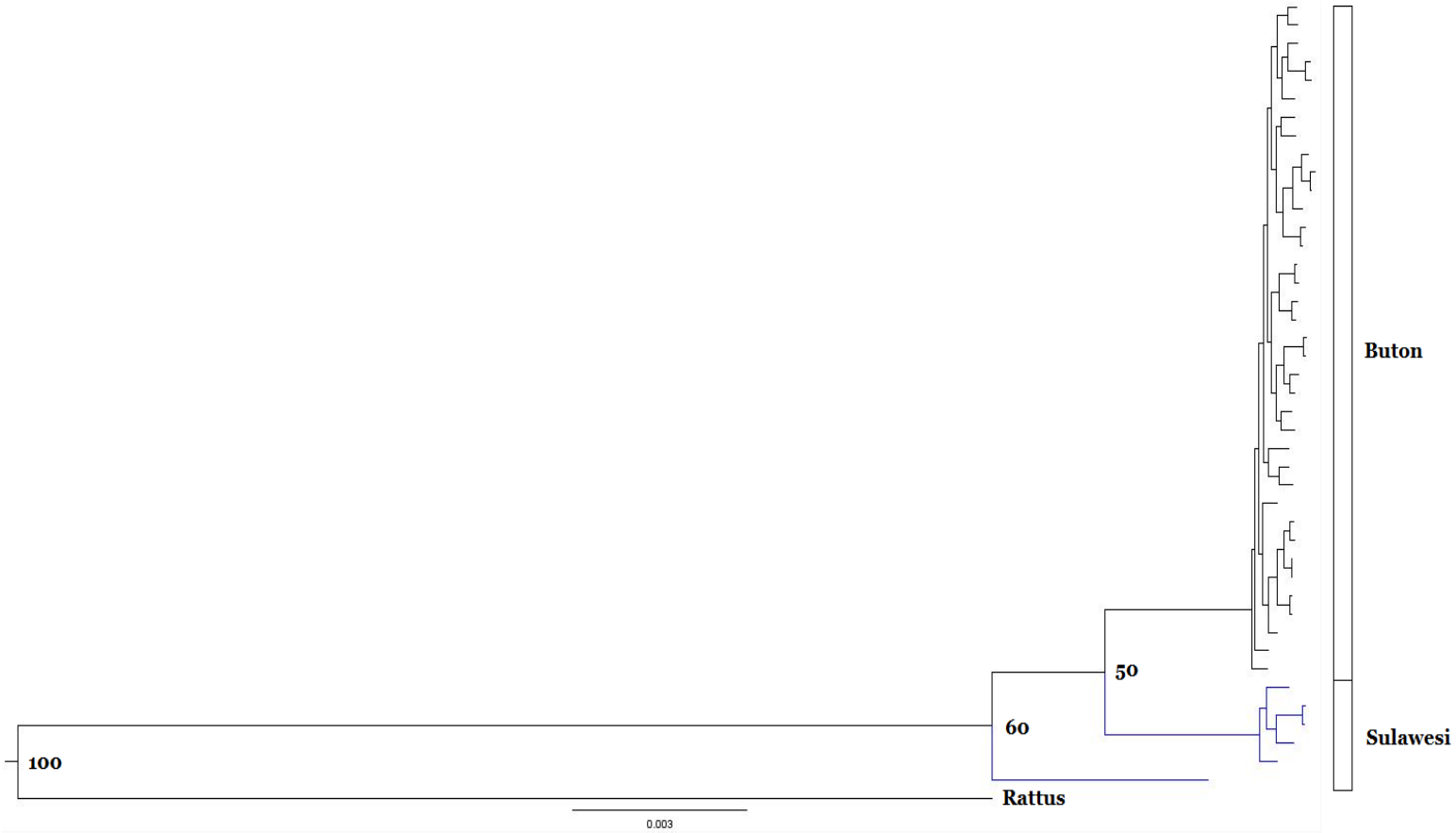
Wright, S. (1931) Evolution in Mendelian Populations. *Genetics* **16**(2), 97-159.

Wright, S. (1943) Isolation by distance. *Genetics* **28**, 114-138.

Yang, Z., Rannala, B. (2012) Molecular phylogenetics: principles and practice. *Nature Reviews* **13**, 303-314.
Zuckerkandl, E., Pauling, L. (1965) Evolutionary divergence and convergence in proteins. *In* Evolving Genes and Proteins (Bryson, V. and Vogel, H.J., eds) pp. 97–166, Academic Press.

Supplementary Information

MrBayes was run for *Bunomys* (without *Maxomys*) along with the *Rattus* reference (N=36). The same settings were used as in Figure 1 in Chapter 3. As the un-shortened Nexus file was too large to be appropriately processed by MrBayes, the standard deviation was never able to drop <0.05, irrespective of the number of iterations. Figure 1 in Chapter 3 was calculated using a manually shortened Nexus file with specimen from both species, whereas all *Bunomys* samples were used in this run. Unlike the Figure 1 in Chapter 3, which is unrooted, this consensus tree is rooted to *Rattus*. The consensus tree was calculated at a standard deviation of 0.18 after 1 Million iterations. Bootstrap values were calculated via MrBayes.



Acknowledgements

I am not young enough to know everything.

- J. M. Barrie

The author wishes to express his sincere thanks to Professor Rus Hoelzel, under whose supervision this work was carried out, for his unfailing guidance and encouragement. The author is indebted to Drs Kim Andrews, Michelle Gaither and David Dolan for their guidance both in the lab and while handling raw data. He is grateful for all the advice given by Ratima Karuwancharoen, Inge Chen, Howard Gray, Andreanna Welch and Catherine Pepper. He is equally grateful for all the reinforcement given to him by his family, from Robert, Christina and Anna.