#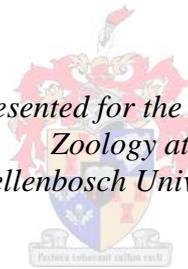 Comparative phylogeography and phylogenetic relationships of the four-striped mouse genus, *Rhabdomys*, and the ectoparasitic sucking louse, *Polyplax arvicanthis*

Nina du Toit

Dissertation presented for the degree of doctor of Zoology at Stellenbosch University

Promoters:
Prof Conrad A. Matthee
Prof Bettine Jansen van Vuuren
Dr Sonja Matthee

Faculty of Science

March 2013

Declaration

By submitting this thesis/dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Some of the contents contained in this thesis (Chapters 2-4) are taken directly from manuscripts submitted or drafted for publication in the primary scientific literature. This resulted in some overlap in content between the chapters.

March 2013

Abstract

Within southern Africa, the widely distributed four-striped mouse *Rhabdomys* is parasitized by, amongst others, the host-specific ectoparasitic sucking louse, *Polyplax arvicanthis*. The present study investigated this parasite-host association from a phylogenetic and phylogeographic perspective utilizing mitochondrial and nuclear DNA markers. The findings support the existence of four species within *Rhabdomys* (three distinct lineages within the previously recognized arid-adapted *R. pumilio* and the mesic-adapted *R. dilectus*). These species have distinct geographic distributions across vegetational biomes with two documented areas of sympatry at biome boundaries. Ecological niche modelling supports a strong correlation between regional biomes and the distribution of distinct evolutionary lineages of *Rhabdomys*. A Bayesian relaxed molecular clock suggests that cladogenesis within the genus coincides with paleoclimatic changes (and the establishment of the biomes) at the Miocene-Pliocene boundary. Strong evidence was also found that the sucking louse *P. arvicanthis* consists of two genetically divergent lineages, which probably represent distinct species. The two lineages have sympatric distributions throughout most of the sampled range across the various host species and also occasionally occur sympatrically on the same host individual. Further, the absence of clear morphological differences among these parasitic lineages suggests cryptic speciation. Limited phylogeographic congruence was observed among the two *P. arvicanthis* lineages and the various *Rhabdomys* species and co-phylogenetic analyses indicated limited co-divergence with several episodes of host-switching, despite the documented host-specificity and several other traits predicted to favour congruence and co-divergence. Also, despite the comparatively smaller effective population sizes and elevated mutational rates found for *P. arvicanthis*, spatial genetic structure was not more pronounced in the parasite lineages compared to the hosts. These findings may be partly attributed to high vagility and social behaviour of *Rhabdomys*, which probably promoted parasite dispersal among hosts through frequent inter-host contact. Further, the complex biogeographic history of *Rhabdomys*, which involved cyclic range contractions and expansions, may have facilitated parasite divergence during periods of host allopatry, and host-switching during periods of host sympatry. Intermittent contact among *Rhabdomys* lineages could also have prevented adaptation of *P. arvicanthis* to specific host lineages, thus explaining the lack of host-specificity observed in areas of host sympatry. It is thus evident that the association between *Polyplax arvicanthis* and *Rhabdomys* has been shaped by the synergistic effects of parasite traits, biogeography, and host-related factors over evolutionary time.

Opsomming

Binne suidelike-Afrika word die wyd-verspreide gestreepte veldmuis, *Rhabdomys*, onder andere deur die gasheer-spesifieke ektoparasitiese luis, *Polyplax arvicanthis*, geparasitiseer. Die huidige studie het hierdie parasiet-gasheer interaksie vanuit 'n filogenetiese en filogeografiese oogpunt ondersoek deur van beide mitokondriale en nukluêre merkers gebruik te maak. Die bevindinge dui op die bestaan van vier spesies binne *Rhabdomys*, waaronder drie nuwe genetiese groepe binne die voorheen erkende *R. pumilio* asook *R. dilectus*. Hierdie spesies het nie-oorvleulende geografiese verspreidings binne spesefieke plantegroei biome met twee geidentifiseerde areas van simpatriese voorkoms by bioom grense. Ekologiese nis modellering ondersteun 'n sterk korrelasie tussen biome en die verspreiding van die evolusionêre groepe binne *Rhabdomys*. 'n Bayesiaanse verslapte molekulêre klok dui daarop dat kladoginese binne die genus gedurende paleoklimatiese veranderinge, wat tot die totstandkoming van die huidige biome gelei het, by die Mioseen-Plioseen grens plaasgevind het. Sterk bewyse is ook gevind dat die parasitiese luis *P. arvicanthis* uit twee geneties verskillende groepe, wat heel moontlik afsonderlike spesies verteenwoordig, bestaan. Hierdie genetiese groepe het simpatriese verspreidings oor meeste van die gebestudeerde geografiese area op die verskeie gasheer spesies en mag ook soms simpatries op dieselfde gasheer individu voorkom. Verder dui die afwesigheid van duidelike morfologiese verskille tusssen die parasiet genetiese groepe op moontlike kriptiese spesiasie. Beperkte filogeografiese ooreenstemming is tussen die *P. arvicanthis* genetiese groepe en die *Rhabdomys* spesies waargeneem en die vergelykende-filogenetiese analises het aangedui dat daar beperkte gesementlike-divergensie plaasgevind het met verskeie episodes van gasheer-wisseling, ten spyte van die gasheer-spesifieke aard van die parasiete asook verskeie ander kenmerke wat veronderstel is om filogeografiese ooreenstemming en gesementlike-divergensie te bevorder. Ten spyte van die vergelykbaar kleiner effektiewe bevolking groottes en verhoogde mutasie tempo wat vir *P. arvicanthis* gevind is, is die geografiese genetiese struktuur nie meer gedifferensieërd in die parasiet groepe as in die gasheer nie. Hierdie bevindinge mag deels verklaar word deur die hoë beweeglikheid asook die sosiale gedrag van *Rhabdomys*, wat waarskynlik parasiet beweging tussen gashere bevorder deur gereelde tussen-gasheer kontak. Die komplekse biogeografiese geskiedenis van *Rhabdomys*, wat sikliese inkrimping en uitsetting van die geografiese verspreiding behels het, het heel moontlik parasiet divergensie tydens tydperke van gasheer allopatrie asook gasheer-wisseling tydens tydperke van gasheer simpatrie, gefasiliteer. Tussentydse kontak tussen *Rhabdomys* genetiese groepe kon aanpassing van *P. arvicanthis* tot sekere gasheer genetiese groepe

verhoed het en verklaar dus die afwesigheid van waargenome gasheer-spesifisiteit in areas van gasheer simpatrie. Dit is dus duidelik dat die assosiasie tussen *P. arvicanthis* en *Rhabdomys* deur die sinergistiese uitwerking van parasiet kenmerke, biogeografie, asook gasheer-verwante faktore oor evolusionêre tyd gevorm is.

Acknowledgements

I would like to express my sincere gratitude to my supervisors for providing invaluable guidance and allowing me creative freedom during this research.

Table of Contents

List of Figures

List of Tables

Chapter 1

General Introduction

1. Host-parasite evolutionary interactions

Co-evolution may be defined as the process whereby interacting species exert selective pressures on each other, resulting in reciprocal evolutionary change (Thompson 1994). In the context of host-parasite interactions, parasite infectivity and host resistance result in such selective pressures (Thompson 1994). Co-speciation is the joint speciation of ecologically associated lineages (Page 2003), and can be seen as co-evolution occurring at a macroevolutionary scale (Brooks & McLennan 1991, 1993; Light & Hafner 2008). If co-speciation repeatedly takes place in a system, a pattern of co-phylogeny (significantly similar branching topologies) may be observed (Clayton *et al.* 2004; Light & Hafner 2008). Erosion of congruence may occur due to host-switching, parasite duplication (parasite speciation without host speciation), sorting events such as parasite extinction or "missing the boat" (absence of parasites on hosts undergoing founder event speciation), and failure of the parasite to speciate when the host does (Clayton & Johnson 2003; Johnson *et al.* 2003).

Most studies on host-parasite associations have been conducted above the species level and focused on macroevolutionary trends (Page 2003). Resulting patterns range from complete and partial congruence between host and parasite phylogenies (Hafner & Nadler 1990; Moran & Baumann 1994; Thomas *et al.* 1996; Haukisalami *et al.* 2001) to incongruence (Page & Hafner 1996; Charleston & Robertson 2002; Weiblen & Bush 2002; Huyse & Volckaert 2005). Further, truly contemporaneous speciation events appear to be uncommon (Ronsted *et al.* 2005; Switzer *et al.* 2005; Light & Hafner 2008).

Macroevolutionary events are ultimately the result of microevolutionary processes such as selection, dispersal, and drift operating at the intraspecific level (Nieberding & Morand 2006), where co-divergence (co-evolution at microevolutionary scales; Brooks & McLennan 1991, 1993) may result in congruent phylogeographic structures (Nadler 1995; Clayton & Johnson 2003; Clayton *et al.* 2004; Criscione *et al.* 2005; Štefka & Hypša 2008). Studies comparing the intraspecific genetic structures of hosts and their associated parasites over intermediate spatial scales has started to gain momentum and have revealed varying levels of congruence (see Clayton *et al*. 2004; Criscione *et al*. 2005; Nieberding & Morand 2006; Stefka & Hypsa 2008; Demastes *et al*. 2012). However, more comparative phylogeographic studies of hosts and parasites are needed to

investigate the microevolutionary processes which reinforce macroevolutionary trends (Criscione *et al.* 2005).

2. Factors affecting parasite-host congruence

The level of congruence observed between host and parasite structures appear to be highly dependent on the intimacy of the interaction between the host and parasite species (Charleston & Perkins 2006) and is determined by several parasite life-history, ecological and demographic traits (Nieberding & Morand 2006).

2.1 Life history traits

Co-evolutionary processes in parasites are underpinned by variation in natural history traits (Whiteman *et al.* 2007). Indeed, parasite population structure and the level of congruence varies according to the specific natural history traits of the parasite species (Whiteman & Parker 2005) and congruence between host and parasite patterns appear to be primarily determined by the duration and intimacy of their association, i.e. long-term host specificity (Johnson *et al.* 2003).

Host specificity may be defined as the range of hosts that can be exploited by the parasite as the result of the evolutionary and biogeographic history of the association (Poulin & Keeney 2007). The degree of host specificity has a significant influence on parasite genetic structures through the creation of gene flow barriers (Nadler 1995; Johnson *et al.* 2002; McCoy *et al.* 2003) which limits host-switching opportunities (Blouin *et al.* 1995). This is illustrated by the fact that congruent patterns are mostly observed among host-parasite pairs with highly specific interactions (Hafner *et al.* 1994; Johnson *et al.* 2002; Clayton & Johnson 2003; Nieberding *et al.* 2004). Generalist parasites usually do not co-evolve with any of their hosts in particular as indicated by their incongruent phylogeographic structures (Brown *et al.* 1997; Althoff & Thompson 1999; Joseph *et al.* 2002). It has been suggested, however, that the apparent host specificity of parasites may simply reflect their incapability to disperse between hosts (Tompkins & Clayton 1999), or the lack of opportunities to do so (Reed & Hafner 1997).

Parasites with a direct life cycle (lack of intermediate hosts) that lack a free-living phase are more likely to have congruent structures with their host, since parasite migration and therefore gene flow is completely dependent on host movements (Blouin *et al.* 1995; Jerome & Ford 2002a; Johnson *et al.* 2003). A direct life cycle thus prevents "parasite release", which is the failure of a parasite to speciate when the host does (Johnson *et al.* 2003). In addition, sexually reproducing parasites will reflect historical differentiation and migration events of the host better than asexual parasites since they will track host movements more closely because they have to meet on a living host in order to reproduce (Nieberding *et al.* 2004; Nieberding & Olivieri 2007).

2.2 Demographic parameters

The effective population size (N_e) can significantly influence the level of parasite-host congruence, the genetic diversity within parasite populations, and the extent of population genetic structure between parasite populations (Nadler 1995; Rannala & Michalakis 2003; Criscione *et al.* 2005; Huyse *et al.* 2005). Small effective population sizes in parasites increase the likelihood of congruence with host structure and leads to more pronounced genetic differences within parasite populations in comparison with the host (Huyse *et al.* 2005). Populations with smaller effective sizes will reach reciprocal monophyly faster than populations with larger sizes (Avise 1994) due to an increased probability and rate of fixation for neutral alleles (Huyse *et al.* 2005).

Not much is known about the factors that control N_e in parasite populations (Criscione *et al.* 2005). However, it is predicted that several parasitic life cycle features may act jointly to reduce the effective population size below that expected for a free-living population of equal census size (Criscione *et al.* 2005). Smaller effective population sizes may be the result of features such as shorter generation times, the highly fragmented nature of populations, strong seasonal fluctuations in population sizes, bottlenecks and frequent extinction and re-colonization events (Huyse *et al.* 2005). Also, N_e is expected to be small if the parasite lacks intermediate hosts and has a short or no free-living phase (Rannala & Michalakis 2003; Criscione & Blouin 2005). Parasites such as lice with a direct life cycle that spend several generations on a single host individual are thus expected to have a small effective population size and will probably show strong diversification and specialization (de Meeûs 2000). N_e is also expected to decrease with an increase in the skew of the

sex ratio (Huyse *et al.* 2005), which is often typical for ectoparasitic insects such as lice (Marshall 1981).

If a parasite displays high prevalence (proportion of hosts examined that are infected with one or more individuals of a particular parasite species) and intensity (number of individuals of a particular parasite species on an individual host) on the host, the probability of it tracking the migration and differentiation patterns of the host is greater (Nieberding & Morand 2006). This in turn reduces the probability of parasite extinction and "missing the boat" (absence of parasite lineage during host founder event speciation) and strengthens the probability of congruent patterns among host and parasite (Rozsa 1993; Paterson *et al.* 1999; Clayton *et al.* 2003).

2.3 Ecological factors

If parasites are transmitted vertically through successive host generation, the genealogical history of the host is more likely to be congruent with that of the parasite (Johnson & Clayton 2004; Whiteman & Parker 2005; Wirth *et al.* 2005). Vertical transmission is usually ensured by strong host specificity and local adaptation of the parasite (Clayton *et al.* 2003; Prugnolle *et al.* 2005) and as expected, congruent host and parasite trees are usually found for parasites that display persistent vertical transmission (Hafner *et al.* 1994; Funk *et al.* 2000; Baumann & Baumann 2005). However, it is important to realize that horizontal transmission does not necessarily preclude co-differentiation between a parasite and its host as long as parasites are transmitted within and not between diverging host lineages (Nieberding & Olivieri 2007).

2.4 Host related factors

In parasites with low dispersal capabilities and lacking a free-living stage, host vagility is an important determinant of population structure since parasite gene flow is dependent on host migration (Blouin *et al.* 1995; McCoy *et al.* 2003; Criscione & Blouin 2004; Criscione *et al.* 2005). Inter-host contact as a result of to social behaviour will thus promote parasite dispersal (Huyse *et al.* 2005) while dispersal will be restricted in solitary hosts (Demastes *et al.* 2012).

2.5 Biogeography

Biogeographic changes can also have significant impact on host-parasite interactions over evolutionary time (Hoberg & Brooks 2008). A shared biogeographic history among closely interacting parasites and their hosts may lead to congruent genetic relationships through similar responses to vicariant events (Hoberg & Klassen 2002; Page 2003; Hoberg & Brooks 2008). Biogeography may also determine the genetic structure of parasites irrespective of host associations, particularly in systems involving multi-host parasites, parasites with free-living phases, and parasites with intermediate hosts (Clayton 1990; Weckstein 2004; Hoberg & Brooks 2008; Nieberding *et al.* 2008). Episodes of environmental change have been suggested as the main drivers for diversification in parasite host systems by inducing cyclical episodes of expansion and contraction in geographical ranges ("Taxon pulse hypothesis"; Halas *et al.* 2005). Biogeographic shifts may thus facilitate co-divergence during periods of refugial isolation and host-switching during periods of expansion (Weckstein 2004; Brooks & Ferrao 2005).

3. Parasites as biological magnifying glasses

An intimate relationship between parasites and their hosts (as a product of the various factors outlined above) may result in co-differentiation (through shared founding, differentiation or migration events) at the intraspecific level (Wickström *et al.* 2003; Reed *et al.* 2004; Nieberding & Olivieri 2007). Parasites may potentially provide better resolution of this common evolutionary history if their genealogical history is similar to and their phylogeographic structure more diversified than that of the host (Nieberding *et al*. 2004; Nieberding & Morand 2006; Nieberding & Olivieri 2007). Parasites may thus be used as biological "magnifying glasses" to illuminate host history (Nieberding & Morand 2006) and the comparison of host and parasite genetic structures can provide additional information about host evolutionary history that could not be obtained from simply studying the host directly (Thomas *et al.* 1996; Nieberding *et al.* 2004). Indeed, comparative studies of host parasite structures have revealed several features of host genealogy and allowed for the generation of new hypotheses about host evolutionary history (Nieberding & Morand 2006). These include historical migration, gene flow patterns, and the location of cryptic refuges in the host (Burban *et al.* 1999; Burban & Petit 2003; Nieberding *et al.* 2004; Nieberding *et al.* 2005) as well as host colonization history and the existence of cryptic lineages (Pellmyr *et al.* 1998; Jerome & Ford 2002a; Jerome & Ford 2002b; Wickström *et al.* 2003; Whiteman *et al.* 2007).

The spatial genetic structure of a parasite is also expected to be more pronounced if parasite populations experience lower levels of gene flow and higher levels of genetic drift due to their more limited dispersal abilities and expected smaller effective population sizes (N_e) (Huyse *et al.* 2005; Criscione *et al.* 2005; Nieberding & Morand 2006). Host evolutionary history will also be amplified by the parasite if the parasite DNA has a higher substitution rate relative to the host, leading to coalescent processes proceeding much more rapidly in the former (Blouin *et al.* 1995; Whiteman & Parker 2005). Many studies on parasite-host systems have shown that parasite genes have a faster rate of molecular evolution than the homologous genes of their hosts (Hafner *et al.* 1994; Page & Hafner 1996; Page *et al.* 1998; Paterson & Banks 2001; Nieberding *et al.* 2004; Page *et al.* 2004; Light & Hafner 2007, 2008) and this could possibly be ascribed to differences in cell division rate, DNA repair efficiency, metabolic rate, body size and generation time (Martin & Palumbi 1993). Parasite generation times are usually shorter than that of their host (Huyse *et al.* 2005; Whiteman & Parker 2005), which could allow for the accumulation of more mutations within a certain time relative to the host.

4. Parasite-host study system

4.1 Host

The genus *Rhabdomys* (Rodentia: Muridae) was first recognized by Thomas in 1916, comprising of a single species, the four-striped mouse, *Rhabdomys pumilio* (Sparrman 1784). *Rhabdomys pumilio* has traditionally been regarded as widely distributed in the southern African subregion and also occurring in isolated areas north of the subregion (Tanzania, Kenya, Uganda, the DRC, Angola, Zambia and Malawi; Skinner & Chimimba 2005). This rodent is a generalist opportunistic omnivore, that occupies a variety of altitudes and habitat types (Skinner & Chimimba 2005). It is abundant overall, well adapted to both natural and urbanized habitats and of economic importance due to the damage it can cause to crops and cultivated land (de Graaff 1981).

The taxonomy of *Rhabdomys* is surrounded by much uncertainty regarding the number of morphologically distinct species and/or subspecies that should be recognized (Musser & Carleton 2005). This is mostly due to extensive variation in pelage colouration across the species' range. Initially, 20 subspecies from southern Africa were listed by Roberts (1951) based on pelage colour patterns and morphological measurements. Meester *et al.* (1986) subsequently retained only seven

subspecies, which subsumed many of those described by Roberts (1951). The exact distribution limits of these subspecies are poorly understood (Skinner & Chimimba 2005). Others have argued that only two species/subspecies (de Graaff 1981) or none (Misonne 1974) should be recognized. Further, the results of both allozyme analysis (Mahida *et al.* 1999) and breeding studies (Pillay 2000) have been ambiguous and inconclusive (Musser & Carleton 2005). Rambau *et al*. (2003) identified two distinct mtDNA clades regarded as separate species, owing to the marked genetic divergence, variable chromosome number, and differences in ecology and sociality (Musser & Carleton 2005). The arid-adapted *R. pumilio* (2n=48) with an arid central and western distribution within South Africa, Namibia, and Botswana form social groups (Schradin & Pillay 2004; Schradin *et al.* 2010), while the mesic-adapted *R. dilectus* is solitary and has an east-central distribution in South Africa, Zimbabwe, Uganda, and Tanzania (Schradin & Pillay 2005). Rambau *et al*. (2003) also retrieved two subgroups within *R. dilectus,* representing the proposed subspecies *R. d. dilectus* (2n=46) and *R. d. chakae* (2n=48). A recent study has shown that *R. dilectus* is even more diverse and consists of at least three distinct mitochondrial haplo-groups (Castiglia *et al.* 2011).

4.2 Parasite

Anoplura (sucking lice; Insecta: Phthiraptera) are true, obligate, permanent parasites of eutherian mammal hosts (Kim 2006). These wingless insects inhabit the pelage of mammals, where they feed from blood vessels with their unique piercing-sucking mouthparts (Lavoipierre 1967). Most species of Anoplura have a simple life cycle including the egg, 3 larval instars and adult stage (Kim 2006). Once established on an individual host, sucking lice will complete several generations on the living host until its death (Kim 2006). It is believed that parasitic lice (Phthiraptera) in general will not leave a living host except under circumstances involving contact between individual hosts, such as seen during copulation, offspring care, and other social interactions (Ledger 1980; Marshall 1981). It is thought that intraspecific dispersal of parasites between individual hosts usually occurs from adult to offspring, although parasite transfer can take place via shared nests or burrows also (Ledger 1980; Marshall 1981) and in rare instances phoresis (dispersal via non-host organism) has also been reported (Durden 1990). The fate of sucking lice is therefore closely tied to their mammalian hosts and this close association provide interesting models for the study of parasite-host co-evolution (Kim 1985, 2006). The parallel evolution of sucking lice and mammalian lineages is supported by molecular studies that indicate a close alignment between placental mammal diversification and Anoplura phylogeny, which probably led to the close association seen today (e.g. Kim 1985, 1988,

2006; Springer *et al*. 2003; Light *et al*. 2010; Smith *et al*. 2011). Molecular dating indicates that the diversification of Anoplura took place during the late Cretaceous (approximately 75 Ma) with radiation occurring after the K-Pg boundary, which is in line with the evolutionary history of mammals (Bininda-Emond *et al*. 2007; Light *et al*. 2010; Smith *et al*. 2011). Widespread incongruence among the subsequent phylogenetic histories of mammals and sucking lice, however, indicate that their interaction has been complex and involved multiple host-switching and extinction events through evolutionary time (Light *et al*. 2010).

The intimate biological relationships among sucking lice and their hosts throughout evolutionary history led to a high incidence of host specificity and monoxeny (one parasite species on one host species), with over 63% of known species being monoxenous (Kim 2006). About 70% of the known species of sucking lice are associated with rodents (Kim 1988), with 62% of these being host specific (Kim 2006). From the genus *Polyplax*, approximately 77 species are associated with the monophyletic Muridae (Anderson & Jones 1984; Kim 1985). Within southern Africa, sucking lice from the genus *Polyplax* parasitize several rodent species (Ledger 1980; Durden & Musser 1994) and currently a single morphologically described species, *Polyplax arvicanthis* (Bedford 1919), has been recorded from *Rhabdomys* and is regarded as host-specific (Ledger 1980; Matthee *et al.* 2007).

5. Aims and objectives

The overarching aim of the current investigation was to explore the evolutionary interactions of a parasite-host association within southern Africa using the four-striped mouse genus, *Rhabdomys*, and the specific ectoparasitic sucking louse, *Polyplax arvicanthis* as model taxa. The main objectives were as follows:

1. To investigate   phylogenetic relationships and taxonomy of the four-striped mouse genus, *Rhabdomys*, particularly focusing on the variation within *R. pumilio* throughout its broad distribution in the arid western regions of South Africa and Namibia

2. To investigate   broad-scale genetic variation within the specific ectoparasitic louse, *Polyplax arvicanthis*, across the distribution of its host genus, *Rhabdomys*

3. To investigate potential congruence between the phylogenetic and phylogeographic patterns of *Polyplax arvicanthis* and *Rhabdomys*

Chapter 2

Biome specificity of distinct genetic lineages within the four-striped mouse *Rhabdomys pumilio* (Rodentia: Muridae) from southern Africa with implications for taxonomy

1. Introduction

It is well established that global paleoclimatic changes have fundamentally influenced speciation processes through altering the habitats and ranges of species (Hewitt 2011). Within the southern African context, the onset of xeric conditions toward the end of the Miocene (6.7 to 6.5 Ma) can be attributed to the glaciation of Antarctica that resulted in rapid cooling of ocean temperatures (Tyson & Partridge 2000) and the associated intensified upwelling of the Benguela current system (Diester-Haass *et al.* 2002). In addition, tectonic uplift along the margins of the Great Escarpment approximately 5 Ma (Partridge 1997; Partridge & Maud 2000), contributed towards an east-to-west sloping topography and an associated rain-shadow effect across the region. In combination these events resulted in significant vegetation changes across southern Africa and the subsequent establishment of the modern biomes (Coetzee 1978; Scott *et al.* 1997). It is thus not surprising that many faunal diversification events within the region date to the Pliocene and Pleistocene (5.3 Ma onwards), and span a diverse range of taxa including reptiles (Matthee & Flemming 2002; Bauer & Lamb 2005; Tolley *et al.* 2006; Makokha *et al.* 2007; Tolley *et al.* 2008; Portik *et al.* 2011), small mammals (Smit *et al.* 2007; Taylor *et al.* 2009; Willows-Munro & Matthee 2009, 2011; Russo *et al.* 2010) , and invertebrates (Prendini *et al.* 2003; Daniels *et al.* 2006; Price *et al.* 2007).

The common African four-striped mouse, genus *Rhabdomys* Thomas 1916, was long regarded as monotypic comprising a single species, *R. pumilio* (Sparrman 1784). The seemingly generalist nature of *Rhabdomys* enables it to maintain a high overall abundance and a wide distribution across a variety of altitudes and habitat types (de Graaff 1981; Skinner & Chimimba 2005). In southern Africa the taxon occurs throughout most of Namibia, Botswana, Zimbabwe, Mozambique, Swaziland, Lesotho, and South Africa, but is also found in Tanzania, Kenya, Uganda, the DRC, Angola, Zambia, and Malawi (Skinner & Chimimba 2005). Extensive variation in pelage colour and morphology resulted in the description of 20 subspecies from southern Africa alone (Roberts 1951), but Meester *et al*. (1986) regarded seven as being valid.

The exact distributional limits of the proposed subspecies are poorly understood (Skinner & Chimimba 2005). Allozyme analysis (Mahida *et al.* 1999) has failed to clearly describe the variation within the genus, and breeding studies (Pillay 2000a; Pillay 2000b) have been inconclusive in ascertaining whether more than one species is present (Musser & Carleton 2005). Based on variable chromosome numbers and the presence of two distinct mtDNA clades within *Rhabdomys,* two

geographically distinct species, *R. pumilio* and *R. dilectus*, are currently recognized (Rambau *et al.* 2003; Musser & Carleton 2005). Within the subregion, the mesic-adapted *R. dilectus* (2n=46 and 2n=48) has an eastern distribution in South Africa, Zimbabwe, Uganda, and Tanzania, and a xeric-adapted *R. pumilio* (2n=48) occurs widely in the arid central and western regions of South Africa, Namibia, and Botswana. Rambau *et al.* (2003) further distinguished two subgroups with different cytotypes within *R. dilectus,* representing what they refer to as the subspecies *R. d. dilectus* (2n=46) and *R. d. chakae* (2n=48). A recent study has shown that *R. dilectus* is even more diverse and consists of at least three distinct mitochondrial lineages (Castiglia *et al.* 2011).

Specific factors driving the diversification within *Rhabdomys* are not well defined. It has been suggested that the arid-adapted *R. pumilio*, with a western distribution, forms social groups in the Succulent Karoo Biome as a result of habitat saturation (Schradin & Pillay 2004; Schradin *et al.* 2010), whereas *R. dilectus* in the east is solitary within the mesic grassland due to the lower abundance and higher dispersion of food resources (Schradin & Pillay 2005). Since large scale changes in the distributions of vegetation have been directly linked to diversification among lineages (Tolley *et al.* 2008; Linder *et al.* 2010; Edwards *et al.* 2011), a prediction can be made that different biomes could harbour distinct evolutionary lineages of the four-striped mouse. Importantly, should this pattern emerge in *Rhabdomys,* it will not be unique. Biomes have previously been found to harbour distinct taxon groups (Chimimba 2001; Russo *et al.* 2010) and there is now extensive evidence of secondary contact among distinct faunal lineages where vegetation types/biomes meet (Tolley *et al.* 2004; Tolley *et al.* 2010; Engelbrecht *et al.* 2011; Willows-Munro & Matthee 2011). Particularly relevant to *Rhabdomys* would be the "Bedfort-gap" (Lawes 1990) which represents a complex region where several biomes meet (Mucina & Rutherford 2006). The region is interspersed within the transitional Albany Thicket Biome (Vlok & Euston-Brown 2002), contains elements of a variety of vegetation types (Mucina & Rutherford 2006) and also provides the interface between the all-year rainfall zone and the summer-rainfall zone (Chase & Meadows 2007b).

To test the hypothesis that changes in vegetation resulted in evolutionary divergences in the four-striped mouse, *Rhabdomys*, we investigated the spatial genetic structure of *R. pumilio*, which has a distribution spanning six different biomes (*sensu* Mucina & Rutherford 2006; Fynbos, Nama-Karoo, Succulent Karoo, Desert, Savanna, and Albany Thicket) across the mainly arid regions of

South Africa and Namibia. *Rhabdomys dilectus* from the mesic regions of the Grassland and Savanna Biomes were included as a reference taxon to provide estimates of interspecific variation within the genus. A Bayesian relaxed molecular clock was used to date divergences among geographic assemblages and ecological niche modelling was applied to better understand the influence of present and past climatic conditions on the potential distribution of *R. pumilio*.

2. Materials and Methods

2.1 Sample collection

Live traps (Sherman-type) baited with a mixture of peanut butter and oats were used to capture the mice. Individuals were euthanized with 2-4 ml of 200 mg/kg sodium pentobarbitone. Tongue or tail tissue was obtained, preserved in 100% ethanol, and deposited in the SUN (Stellenbosch University) tissue database (Table B.1). *Rhabdomys dilectus* specimens were mostly obtained from Rambau *et al*. (2003; Table 2.1, Fig. 2.1). A total of 521 *R. pumilio* specimens from 31 localities and 33 *R. dilectus* specimens from 10 localities, spanning 7 biomes in total, were included in the analyses (Tables 2.1, B.2; Figs. 2.1, 2.2b).

Table 2.1: Localities within the different biomes from which *Rhabdomys* specimens were sampled, with codes corresponding to Fig. 2.1. Localities falling near biome boundaries are indicated as both.

Species	Country/ Province	Locality	Code	Geographic coordinates	Biome
R. pumilio	**Namibia**				
		Otjiamongombe	OR	21°35' S, 16°56' E	Savanna
		Narais	NR	23°07' S, 16°53' E	Savanna
		Windhoek	WH	22°31' S, 17°25' E	Savanna
		Mariental	MT	24°34' S, 18°02' E	Nama-Karoo
		Keetmanshoop	KH	26°21' S, 18°29' E	Nama-Karoo
		Gellap	GR	26°24' S, 18°00' E	Nama-Karoo
		Fish River Canyon	FR	27°41' S, 17°48' E	Nama-Karoo
		Swakopmund	SM	22°41' S, 14°32' E	Dessert
R. pumilio	**South Africa**				
	Northern Cape	Richtersveld	RV	28°12' S, 17°06' E	Succulent Karoo
		Springbok	GP	29°42'S, 18°02' E	Succulent Karoo
		Loeriesfontein	LF	31°04' S, 19°13' E	Nama-Karoo
		Groblershoop	GH	28°37' S, 21°42' E	Nama-Karoo

Table 2.1 continued

Species	Country/ Province	Locality	Code	Geographic coordinates	Biome
R. pumilio	*Northern Cape*	Sutherland	SL	32°24' S, 20'54' E	Nama-Karoo
		Dronfield	DF	28°37' S, 24°48' E	Savanna
		Rooipoort	RP	28°39' S, 24°80' E	Savanna
	Western Cape	Vanrhynsdorp	VR	31°44' S, 18°46' E	Succulent Karoo
		Porterville	PV	32°59' S, 19°01' E	Fynbos
		Rocher Pan	RR	32°36' S, 18°18' E	Fynbos
		Paulshoek	PR	30°23' S, 18°17' E	Succulent Karoo
R. pumilio	*Western Cape*	Stellenbosch	SB	33°55' S, 18°49' E	Fynbos
		De Hoop	DH	34°29' S, 20°24' E	Fynbos
		Oudtshoorn	OH	33°36' S, 22°08' E	Fynbos
		Beaufort West	BW	32°13' S, 22°48' E	Nama-Karoo
		Laingsburg	LB	33°10' S, 20°55' E	Nama-Karoo
		Twee Rivieren	TR	26°30' S, 20°37' E	Savanna
	Eastern Cape	Sneeuberg	MB	31°45' S, 24°46' E	Nama-Karoo
		Fort Beaufort	FB	27°08' S, 20°32' E	Albany Thicket/Mixed
	Free State	Benfontein	BF	28°49' S, 24°49' E	Nama-Karoo
		Gariep Dam	GD	30°33' S, 25°32' E	Nama-Karoo
		Tussen-die-Rivieren	TDR	30°28' S, 26°09' E	Nama-Karoo
		Sandveld	SV	27°40' S, 25°41' E	Savanna/Grassland
R. dilectus	**South Africa**				
	Eastern Cape	Alice	AL	32°47' S, 26°50' E	Albany Thicket/Mixed
		Fort Beaufort	FB	32°51' S, 26°27' E	Albany Thicket/Mixed
	Free State	Willem Pretorius	WP	28°17' S, 27°15' E	Grassland
R. dilectus	**South Africa**	Sandveld	SV	27°40' S, 25°41' E	Savanna/Grassland
		Viljoenskroon	KD	27°00' S, 27°00' E	Grassland
	Gauteng	Suikerbosrand	SR	26°30' S, 28°15' E	Grassland
		Irene	IR	25°53' S, 28°18' E	Savanna/Grassland
	Mpumalanga	Pilgrim's Rest	PS	24°51' S, 30°45' E	Savanna/Grassland
	Zimbabwe				
		Inyanga	IN	18°12' S, 32°40' E	Savanna
		Vumba	VM	19°07' S, 32°47' E	Savanna
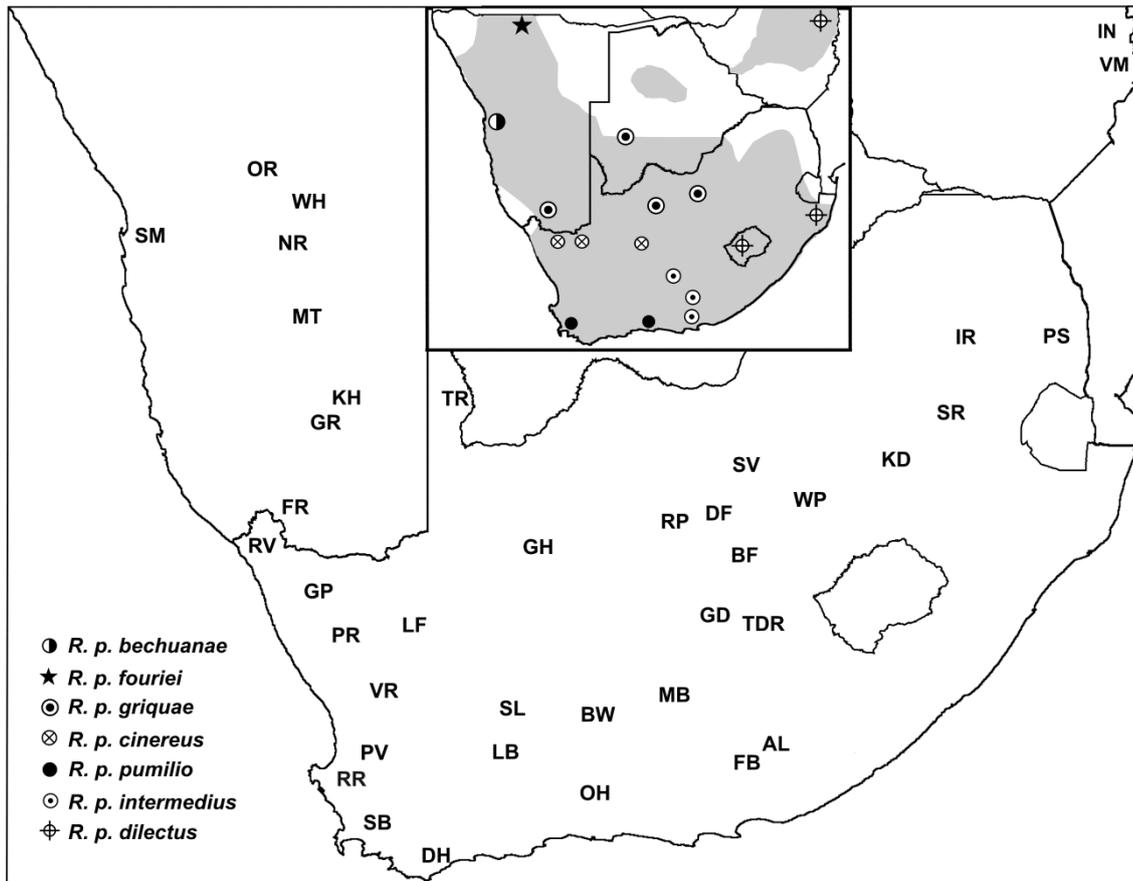
Figure 2.1: Localities from which specimens were analysed in this study with codes as in Table 2.1. The localities from which subspecies (following Meester *et al*. 1986) have been described and the shaded distribution of *Rhabdomys* within southern Africa are indicated in the insert.

2.2 Molecular techniques

Total genomic DNA was extracted with a commercially available kit (Qiagen, DNeasy® Blood and Tissue). PCR and sequencing were performed on the mitochondrial gene Cytochrome Oxidase I (COI), which was generated for all specimens, while the nuclear introns Eef1a1 (eukaryotic translation elongation factor 1 alpha 1), SPTBN1 (beta-spectrin 1 nonerythrocytic), MGF (stem cell factor), and Bfib7 (β-fibrinogen intron 7) were included for a subset of 19 or 20 selected specimens from 19 localities (Tables 2.2, B.2), specifically selected to represent the mtDNA variation observed. Amplification of the gene fragments were performed following standard polymerase chain reaction (PCR) protocols in a GeneAmp® PCR system 2700 thermal cycler (Applied Biosystems). General PCR cycling conditions included an initial denaturation of 3 min at 94°C followed by 30-40 cycles of 30s denaturation at 94°C, 45-60s annealing at the primer-specific temperature (Table B.3), and 45-60s extension at 72°C, followed by a final extension period of 5

min at 72°C. All PCR products were separated by electrophoresis on a 1% agarose gel for visual inspection. If a single clean amplification product was present, purification was performed directly from the PCR product with a commercial kit (Macherey-Nagel, NucleoFast 96 PCR Kit). All other products were purified using a commercial gel purification kit (Promega, Wizard® SV Gel Clean-Up System). All cycle-sequencing reactions were performed using BigDye Chemistry and products were analysed on an automated sequencer (ABI 3730 XL DNA Analyzer, Applied Biosystems).

Table 2.2: The total number of *Rhabdomys* sequences generated (n), the number of alleles used in the phylogenetic analyses (*N*), length in basepairs (bp) after trimming to avoid inclusion of missing data, polymorphic sites (P), parsimony informative sites (PI), the outgroups used for each gene fragment, and the GenBank accession.

Fragment	n	*N*	bp	P	PI	Outgroup	Accession
COI	554	151	900	217	194	*Rattus rattus*	EU273707
						R. norvegicus	AY172581
						Mus musculus	FJ374665
SPTBN1	20	40	847	48	7	*M. musculus*	AL731792.12
MGF	20	40	553	100	28	*M. musculus*	DQ318971
Bfib7	20	40	633	51	15	*M. musculus*	EF605471.1
Eef1a1	19	38	238	22	16	*M. musculus*	NM010106.2

2.3 Data analysis

Sequence alignment and editing was performed in BioEdit Sequence Alignment editor 7.0.5 (Hall 2005). The ends of sequences were trimmed to avoid the inclusion of missing data and gaps were introduced in the intron datasets to allow alignment with the final alignment lengths as indicated in Table 2.2. Collapse 1.21 (Posada 2004) was used to identify the mtDNA haplotypes. Phylogenetic analyses were performed on 151 unique mtDNA haplotypes (GenBank accession numbers JQ 003320 - JQ 003470; Tables 2.2, B.1). For each intron, sequences of individuals were submitted to GenBank (JQ 003241 - JQ 003319; Tables 2.2, B.4). Sequence ambiguities resulting from heterozygous positions were resolved by determining the gametic phase of alleles in PHASE v2.1.1 (Stephens *et al.* 2001; Stephens & Scheet 2005). The algorithm was run for 100,000 generations with a thinning interval of 1 and 10,000 generations were discarded as burn-in. Phases were

considered resolved at a probability threshold of 0.9. All alleles for each intron were included in subsequent analyses.

2.3.1 Phylogenetic reconstructions

Phylogenetic trees were constructed for the complete mtDNA dataset and individual intron datasets, followed by the combined dataset consisting of all nuclear introns and the matching mtDNA subset. For the COI analyses sequences for *Mus musculus*, *Rattus rattus* and *R. norvegicus* were downloaded from GenBank and used as outgroup taxa (Table 2.2). The monophyly of *Rhabdomys* was strongly supported in all analyses and for the combined analyses only *M. musculus* was used as outgroup (Table 2.2). Unweighted parsimony analyses were conducted in PAUP* v4.0b10 (Swofford 2000), using the heuristic search option with random taxon addition (10 replicates) and tree bisection and reconnection (TBR) branch swapping. To reduce computational time, the maximum number of equally parsimonious trees saved during each step was constrained to 100. Nodal support was assessed by 1,000 bootstrap replicates (Felsenstein 1985). The best-fit model of sequence evolution for each of the gene fragments was determined in jModelTest v0.1.1 (Guindon & Gascuel 2003; Posada 2008). The AICc (Burnham & Anderson 2002, 2004), which is a derivation of the original Akaike Information Criterion (AIC; Akaike 1973), was used to choose among alternative models. Bayesian inference (BI) was performed in MrBayes v3.1.2 (Ronquist & Huelsenbeck 2003). Alternative partitioning schemes for the protein-coding COI were evaluated with Bayes factors (Kass & Raftery 1995) as calculated in Tracer v1.5 (Newton & Raftery 1994; Suchard *et al.* 2001; Drummond & Rambaut 2007). For all analyses, the general structure of the models was defined and the default priors used to estimate the parameters (unlinked across all partitions). In each analysis, two parallel Markov Chain Monte Carlo (MCMC) simulations, consisting of 5 chains each, were run for 5-10 million generations with a sampling frequency of 100 generations. Parameter convergence and ESS values were monitored in Tracer v1.5 (Rambaut & Drummond 2007). All independent runs had reached stationarity after 10% of the total number of generations (discarded as burnin). Posterior probabilities for nodal support were obtained by using the sumt command in MrBayes. Due to reticulation among nuclear alleles, networks were implemented to visualize relationships. Individual networks for the nuclear introns were constructed using the NeighborNet method (Bryant & Moulton 2004) implemented in SplitsTree 4.10 (Huson & Bryant 2006). A standardized matrix of multilocus pairwise distances among individuals was generated from pairwise allelic distances of each intron using the POFAD (Phylogeny of Organisms

From Allelic Data) algorithm implemented in the program POFAD 1.03 (Joly & Bruneau 2006). A multilocus network was then constructed from these distances using the NeighborNet method in SplitsTree v4.10 (Huson & Bryant 2006). The average COI HKY-corrected sequence distances among haplotypes, were calculated in PAUP* v4.0b10 (Swofford 2000).

2.3.2 Divergence dating

The nucleotide diversity for each population (excluding those with <5 samples) was calculated in Arlequin 3.5.1.2 (Excoffier & Lischer 2010), using an estimated gamma correction (α=1.71). Divergence dates between clades were estimated from the COI dataset using a Bayesian relaxed molecular clock approach (Drummond *et al.* 2006) as implemented in BEAST v1.6.1 (Drummond & Rambaut 2007). It is widely acknowledged that fossil dates represent good minimum age constraints, but poor maximum age constraints (Donoghue & Benton 2007). Parametric statistical distributions can be used as priors to incorporate uncertainty into calibrations and impose "hard" minimum and "soft" maximum boundaries (Yang & Rannala 2005). Thus, in our analysis, exponential priors were used for fossil calibrations with hard minimum (lower) bounds and soft maximum (upper) bounds, so that 95% of the probability was contained between the two. Two fossil dates were used: the split between *Mus* and *Rattus* approximately 11-12.3 Ma (Benton & Donoghue 2007) and the age of the oldest known *Rhabdomys* fossil, 4-5 Ma (Hendey 1976; Denys 1999). The HKY+I+G model was specified and the data partitioned into $1^{st}+2^{nd}$and 3rd codon positions, with the Yule speciation process as tree prior. The MCMC simulation ran for 80 million generations, sampling every 8,000 generations. Convergence and mixing were assessed in Tracer v1.5 (Rambaut & Drummond 2007) to ensure that all effective sample size (ESS) values were greater than 200, after which the first 2,000 trees were discarded as burn-in and the maximum clade credibility tree produced in TreeAnnotator v1.6.1.

2.3.3 Ecological niche modelling

The MaxEnt (Maximum Entropy) algorithm (Phillips *et al.* 2006; Phillips & Dudik 2008) is a commonly-used method for modelling the potential distributions of species based on their ecological niche requirements. MaxEnt requires only presence data which, despite potential limitations (Elith *et al.* 2011), is advantageous since absence data are often unavailable or unreliable (Anderson *et al.* 2003) and generating pseudo-absences can be problematic (VanDerWal *et al.*

2009; Lobo *et al.* 2010). MaxEnt also performs well with small sample sizes, since its regularization mechanism prevents over-fitting, and it has been shown to outperform other available methods (Elith *et al.* 2006; Elith *et al.* 2011).

The probable current distribution of the mtDNA assemblages within *R. pumilio* was modelled in MaxEnt after which the ecological niche model was projected to climate conditions during the Last Glacial Maximum (LGM; ~21 000 BP). Four bioclimatic variables representing current mean annual trends and seasonality (annual range) in temperature and precipitation (Bio 1: mean annual temperature, Bio 4: temperature seasonality, Bio 12: annual precipitation, Bio 15: precipitation seasonality), as well as altitude data were downloaded from the WORLDCLIM website (version 1.4: http://biogeo.berkeley.edu/worldclim/; Hijmans *et al.* 2005). Climate data for these same bioclimatic variables at the LGM, which has been downscaled using current conditions from the original data of the PIMP2 project (Braconnot *et al.* 2007), were also downloaded from WORLDCLIM. The environmental layers were projected to an equal area map of Africa with a spatial resolution of 5 x 5 km square grids and clipped to fit the extent of the study area. Correlation among variables was assessed with a Pearson correlation coefficient analysis in ENMTools v.1.3 (Warren *et al.* 2010), which indicated that all pairwise values were below 0.7.

The presence records used in our analyses consisted of geo-referenced records with known genotypes from the current study (Table B.5) as well as museum records for *Rhabdomys* published by the following institutions: American Museum of Natural History, California Academy of Sciences, Field Museum of Natural History, Museum of Comparative Zoology (Harvard University), Museum of Vertebrate Zoology (University of California), Los Angeles County Museum of Natural History, Michigan State University Museum, National Museum of Natural History, Zoological Museum Amsterdam (University of Amsterdam), and Yale University Peabody Museum (accessed through GBIF Data Portal, data.gbif.org, 2011-08-22). Geographic co-ordinates for the museum records were obtained from Google Earth (http://earth.google.com; Table B.5). Duplicate records were removed and the remaining filtered to only include records that fell within the distribution of *R. pumilio*. Since the *R. pumilio* clades retrieved from the phylogenetic analyses are geographically structured (see results), museum presence records (unknown genotypes) could be assigned to specific clades based on geographic proximity to localities with known genotypes (Table B.5; Fig. A.1). Areas around the edges of the probable distributions of the clades were excluded, since these could not be assigned to a particular clade with high confidence. This resulted

in a total of 98 presence records of which 31 had known genotypes as determined in this study (values for individual clades are indicated in Table 2.3). The environmental layers together with the presence records were then used to predict the potential distribution of *R. pumilio* assemblages in MaxEnt v3.3.3e (Phillips *et al.* 2006).

Table 2.3: Total number of localities and the proportion with known genotypes within each clade of *Rhabdomys pumilio*, used for the MaxEnt analysis.

Clade	Total	Genotyped
Coastal	27	9
Central	12	6
Northern	57	16
Total	**96**	**31**

The MaxEnt algorithm was run with the following parameter values: regularization multiplier=1, maximum number of background points=10000, maximum iterations=1000, convergence threshold=1×10^{-5}, and the auto features option. Models were evaluated with the threshold-independent AUC (area under the curve) statistic of an ROC (receiver operating characteristic) analysis, which reflects how accurately the model predicts presences. Statistical significance is determined by comparing the model AUC with the null hypothesis that presences are predicted no better than random (AUC=0.5). A 10 fold cross-validation procedure, which utilizes small datasets better than a single training-testing split, was used to generate error margins. The importance of each environmental variable in predicting the distribution of the clades was assessed with the jack-knife procedure, in which several models are constructed, by first removing each variable in turn and then using it in isolation, to determine the effect on model gain.

3. Results

3.1 Phylogenetic reconstructions

As expected, the mtDNA shows a much higher number of polymorphic and parsimony-informative sites compared to the nDNA (Table 2.2). The COI dataset of 554 individuals revealed 151 haplotypes. The parsimony analysis saved the maximum of 100 equally parsimonious trees during each search (Length=901 steps, CI=0.46, RI=0.95). The preferred COI partitioning schemes, as

indicated by Bayes factors (Table B.6), were implemented in the Bayesian analyses. For the independent COI analysis the HKY+I+G (nst=2; rates=invgamma) model was specified for all codon partitions. The combined analysis was partitioned by gene and codon (COI) with the GTR+I (nst=6; rates=inv), JC (nst=1; rates=equal), and GTR (nst=6; rates=equal) models specified for the first, second, and third codon positions of COI, respectively; the HKY+G model (nst=2; rates=gamma) specified for the SPTBN1, MGF and Bfib7; and the GTR+G model (nst=6; rates=gamma) for Eef1a1. Both the parsimony and Bayesian analyses indicated the presence of four well-supported reciprocally monophyletic clades (Fig. 2.2a) with distinct geographic distributions (Fig. 2.2b). The mesic-adapted *R. dilectus* was retrieved, consisting of two subclades which correspond to *R. d. dilectus* and *R. d. chakae* as previously described (Fig. 2.2a; Rambau *et al.* 2003). Three geographically structured *R. pumilio* clades are also present (Coastal, Central and Northern; Fig. 2.2a, b). The Coastal clade consists of individuals originating from the coastal areas of the Western and Northern Cape provinces of South Africa. Two subclades (Coastal A and Coastal B; Fig. 2.2a, b) are present within this clade. Coastal B is represented by individuals from Fort Beaufort (FB) while all other lowland populations in this clade form part of Coastal A. The Central clade consists of individuals from the higher-altitude interior of South Africa (Western and Northern Cape provinces, mainly above the Great Escarpment; Fig. 2.2b) and the Northern clade contains individuals originating from Namibia, the Free State Province, and the northern reaches of the Northern Cape and are mainly distributed north of the Orange River (Fig. 2.2b). Average HKY-corrected sequence divergence values between the clades are comparable to the sequence divergence between the recognized *R. dilectus* and *R. pumilio* species (Table B.7). Also, the divergence between the subclades within the Coastal clade of *R. pumilio* is comparable to divergence between the proposed subspecies within *R. dilectus*. From a phylogenetic perspective, *R. pumilio* is not monophyletic (Fig. 2.2a). Contact zones were found among the Northern clade of *R. pumilio* and *R. d. dilectus* at Sandveld (SV) in the Free State (also see Ganem *et al.* 2012) and between the Coastal and Central *R. pumilio* clades and *R. d. chakae* in the Eastern Cape at Fort Beaufort (FB) (Fig. 2.2b), respectively.
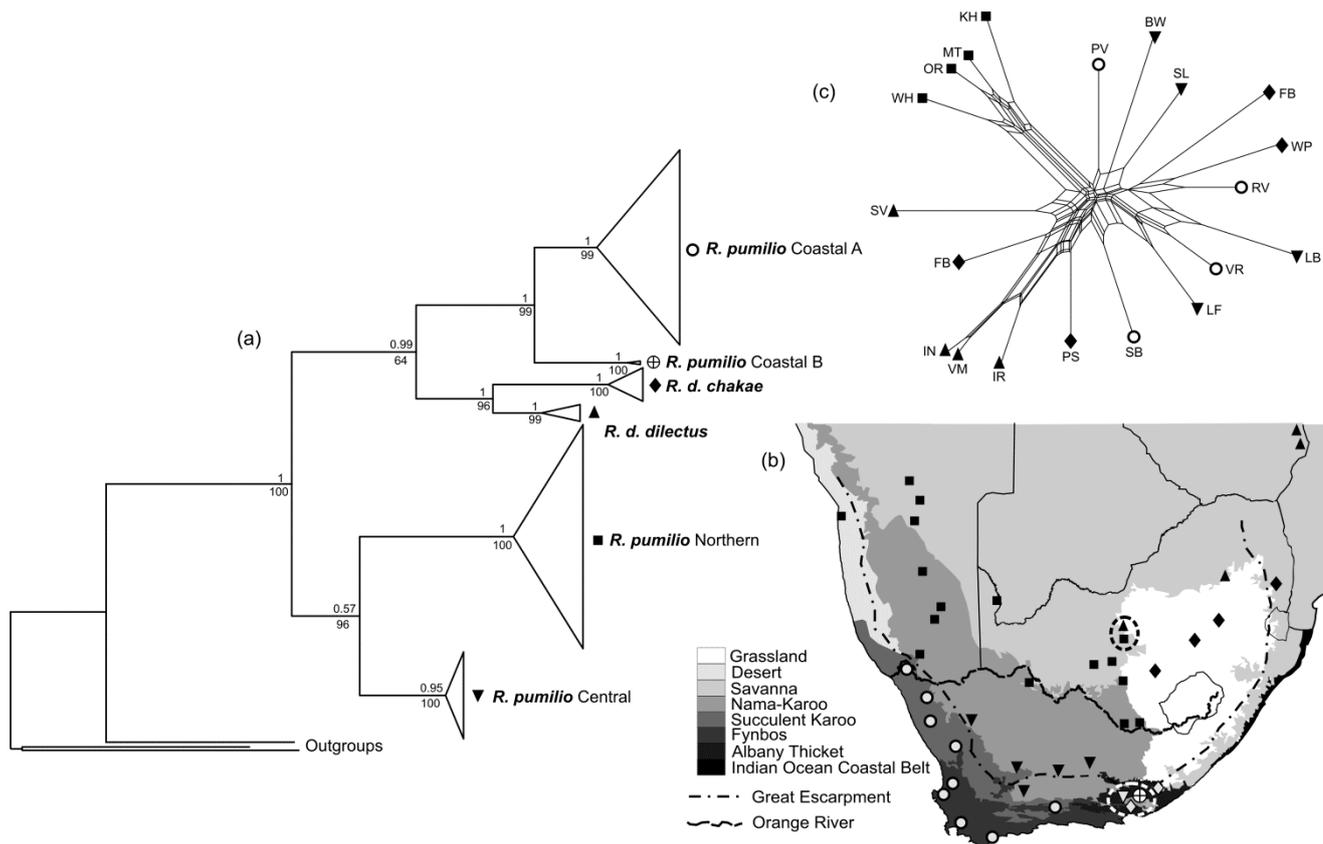
Figure 2.2: (a) The mtDNA parsimony and Bayesian consensus topology with nodal support indicated by posterior probabilities above and bootstrap values below nodes (outgroup OTU's from top to bottom: *Mus musculus*, *Rattus rattus,* and *R. norvegicus*), (b) the distribution of *Rhabdomys* clades across the study area with the biomes of South Africa (following Mucina & Rutherford 2006) and Namibia (following Namibian Atlas Project v.4.02), as well as the locations of the Orange River and the Great Escarpment indicated, and (c) the multilocus nuclear intron network with locality codes as in Fig. 2.1. Clade symbols in (a) and (c) correspond to the sampled localities indicated in (b). Contact zones are encircled, with the black circle indicating Sandveld (SV) and the white circle indicating Fort Beaufort (FB). The position of the white circle also coincides with the location of the "Bedfort-gap" (Lawes 1990).

The combined mtDNA and nDNA datasets yielded a total alignment length of 3,171 base pairs, of which 546 sites were variable and 227 were parsimony-informative. Independent phylogenetic analyses of each nuclear fragment provided little resolution (Figs. A.2, A.3). The parsimony analysis of all data combined (mtDNA + nDNA) resulted in 4 equally-parsimonious trees (Length=998 steps, CI=0.65, RI=0.85) and there is strong overall congruence between the consensus parsimony and Bayesian topology for the combined dataset (Fig. A.4) and the mtDNA topology (Fig. 2.2a). In the combined analyses, the monophyly of all clades are supported by the high parsimony bootstrap values and significant (>0.95) Bayesian posterior probabilities, with the exception of the *R. pumilio* Central clade which had a low posterior probability (Table 2.4). The

individual intron networks provide differential support for the various clades (Table 2.4). The multilocus network (Fig. 2.2c) strongly supports the Northern assemblage within *R. pumilio* and also retrieved the mesic *R. dilectus* and *R. d. chakae* subclade therein. Also, all alleles obtained from individuals sampled in the Northern clade are characterized by a 2bp deletion in the MGF intron. The Central and Coastal assemblages are not differentiated in the individual or multilocus intron networks (Table 2.4; Fig. 2.2c).

Table 2.4: Nuclear support for the mtDNA clades resulting from the combined phylogenetic analyses of all gene fragments (bootstrap/posterior probability), and the clades retrieved from the individual nuclear intron networks (indicated by asterisks).

| | | *R. pumilio* | | *R. dilectus* | *R. d. dilectus* | *R. d. chakae* |
	Coastal	Central	Northern			
All combined	100/1	100/0.66	100/1	95/0.99	100/0.99	100/1
MGF			*	*		
SPTBN1				*	*	*
Bfib7					*	*
Eef1a1			*		*	

3.2 Divergence dating

The overall rate of mtDNA evolution was estimated to be around 1 % ($\pm$0.01) per million years, which is the same as the average rate which has previously been described for COI in mammals (Bininda-Emonds 2007). The BEAST topology (Fig. 2.3) differs slightly from the mtDNA consensus topology with regard to the placement of Central clade (Figs. 2.2, 2.3). In the parsimony and Bayesian analyses, the Central and Northern clades are sister taxa, although the posterior probability values are not significant (Fig. 2.2). In the BEAST topology, the northern assemblage is in a basal position to the exclusion of the rest. Given the uncertain position of the central assemblage, no 95% HPD estimated age could be determined for this node (Fig. 2.3). Divergence of the major lineages within *Rhabdomys* (*R. pumilio* Coastal, Northern, Central, and *R. dilectus*) occurred during the early and middle Pliocene (4.30 to 3.09 Ma). The divergence between the two *R. pumilio* Coastal subclades and the divergence between the two subspecies of *R. dilectus* followed nearly contemporaneously during the late Pliocene (2.34 and 2.17 Ma respectively).
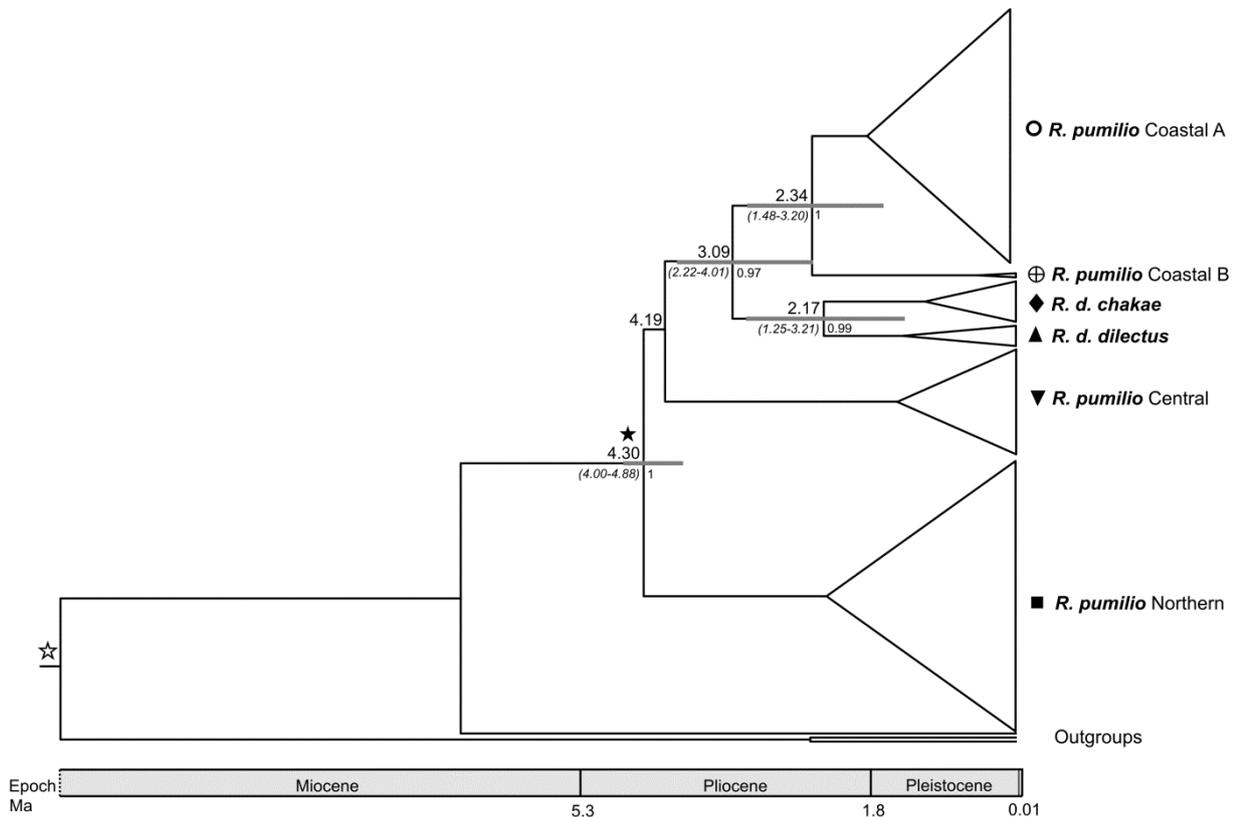
Figure 2.3: Maximum clade probability tree obtained from the fossil-calibrated BEAST analysis (outgroup OTU's from top to bottom: *Mus musculus*, *Rattus rattus,* and *R. norvegicus*). Fossil-calibrated nodes are indicated by the shaded star (oldest *Rhabdomys* fossil; 4-5 Ma) and non-shaded star (split between *Mus* and *Rattus*; 11-12.3 Ma) respectively. Values above nodes indicate the posterior mean divergence dates in millions of years before present. Shaded bars and values in brackets under nodes indicate the 95% HPD credibility intervals. Non-bracketed values under nodes indicate posterior probabilities above 0.95. The epochs spanning the divergence events are also indicated.

When comparing the nucleotide diversities of sampled localities within the different clades (Fig. 2.4), it is evident that the nucleotide diversity of Otjiamongombe (OR) and Narais (NR) is much higher than the other sampled localities within the Northern clade, while Rocher Pan (RR) had the highest diversity in the Coastal clade (Fig. 2.1). All populations within the Central clade had approximately equal nucleotide diversities.

25
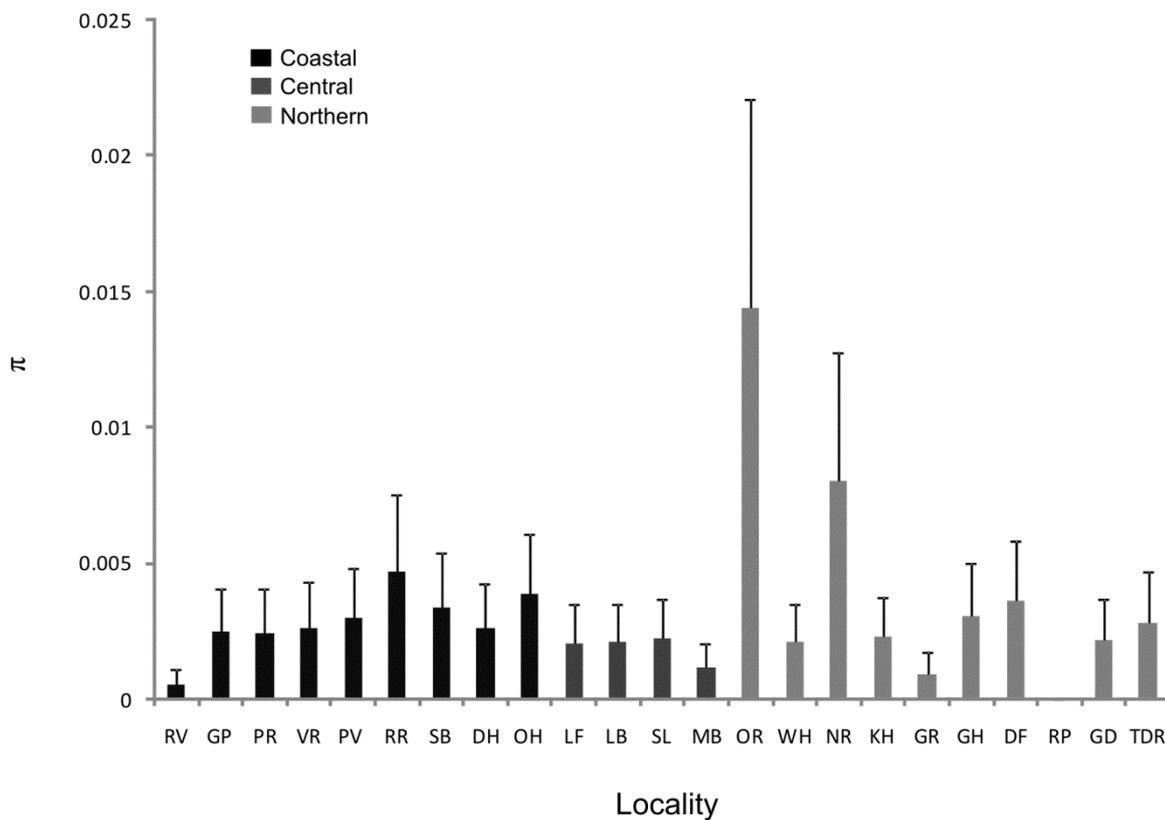
Figure 2.4: Relative COI nucleotide diversity values (π) for the sampled localities within each clade of *Rhabdomys pumilio*.

3.3 Ecological niche modelling

AUC values for the Coastal (0.964, SD=0.024) and Central (0.924, SD=0.075) clades were high, thus indicating very good model performance (Pearce & Ferrier 2000). The AUC value for the Northern clade was lower (0.824, SD=0.081), indicating moderate model performance. The MaxEnt predictions of the current distributions closely resemble the proposed geographic distributions for each of the *R. pumilio* mtDNA clades and lend support to the hypothesis that the genetic assemblages occupy distinct environmental niches (Fig. 2.5a).
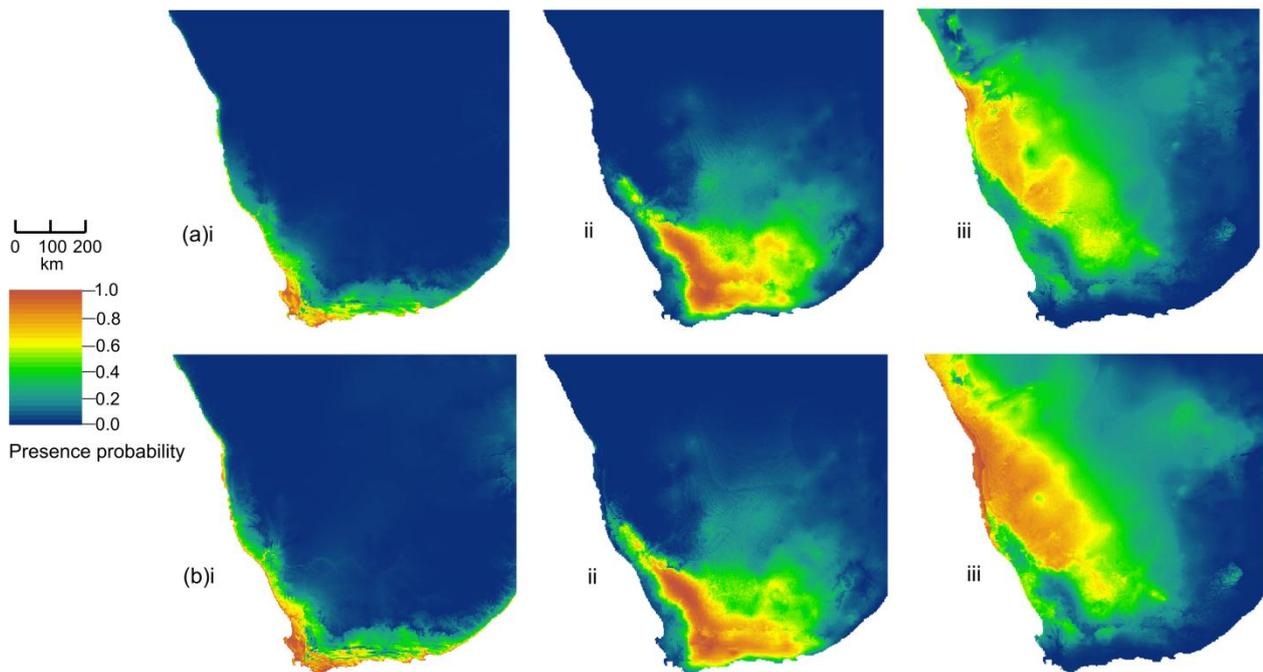
Figure 2.5: Predicted probability of presence from MaxEnt during: (a) current climatic conditions and (b) the Last Glacial Maximum for the Coastal (i), Central (ii), and Northern (iii) clades within *Rhabdomys pumilio*.

Predictions are somewhat broader (although these areas have a lower predicted probability of occurrence) than the recorded presence of the clades, indicating areas of potential contact among the clades. However, distributions predicted by ecological niche modelling will often be wider than the actual distribution of species due to other factors such as physical barriers and biotic interactions constraining the extent of their realized distribution (Phillips *et al.* 2006). It is evident that the projected distribution ranges, particularly the Northern clade (in terms of areas with high suitability), were much more extensive during the LGM than the current clade distributions (Fig. 2.5b).

Altitude was the most important variable for predicting the current distribution of the Coastal clade, while precipitation seasonality (Bio 15) and annual precipitation (Bio 12) had the greatest influence on the distributions of the Central and Northern clades respectively. Altitude, precipitation seasonality, and annual precipitation explained 73%, 66.4%, and 76% of the variation in the respective models. Also, jack-knife tests indicated that these variables resulted in the greatest test gain when used in isolation and the greatest decrease in gain when excluded from the model.

Response curves for all these variables were negative, indicating that probability of occurrence decreases with an increase in altitude, a higher amount of annual precipitation, and an increase in the variance of precipitation seasonality, for the different clades (Fig. 2.6a(i), b, c). For the Coastal clade, mean annual temperature (Bio 1) and precipitation seasonality (Bio 15) also significantly contributed to the model prediction, although to a lesser extent than altitude, as indicated by the resulting increase in test gain when these variables were used in isolation. Response curves indicate an initial positive response for annual mean temperature, after which the relationship becomes negative (Fig. 2.6a(ii)), while there is a negative relationship with probability of occurrence for precipitation seasonality (Fig. 2.6a(iii)).
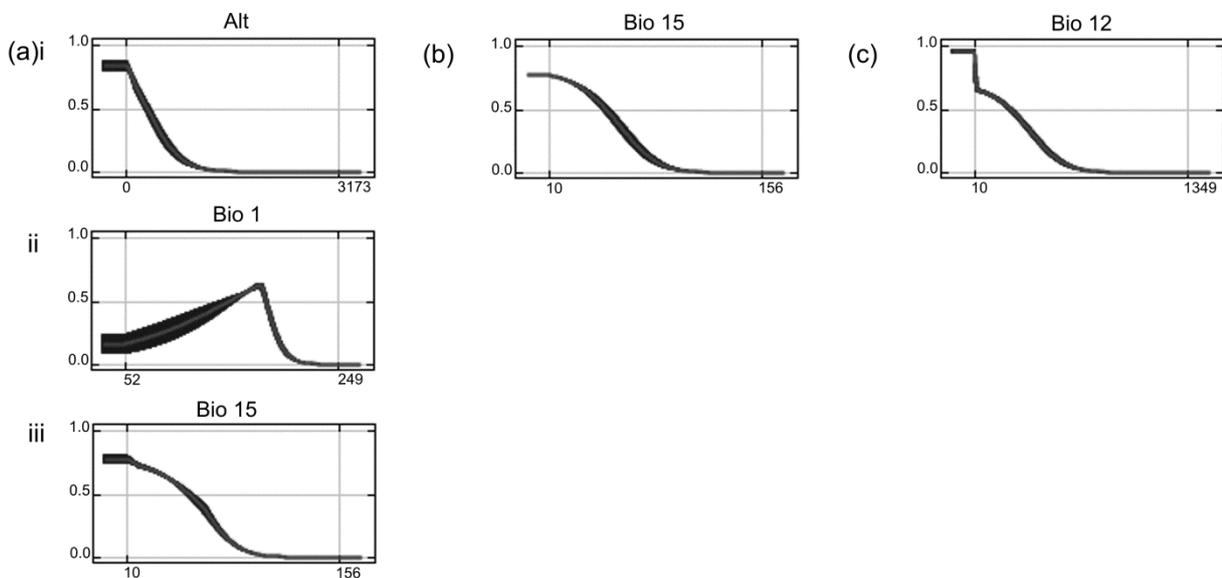
Figure 2.6: Response curves of the most important variables for the (a)i-iii Coastal, (b) Central, and, (c) Northern clades of *Rhabdomys pumilio*, from the MaxEnt analysis.

4. Discussion

Genetic analyses coupled with ecological niche modelling on a large number of *R. pumilio* specimens originating from the southern African subregion clearly support the hypothesis that evolutionary divergences within the four-striped mouse were associated with changes in vegetation. A close correlation between the dates of divergence and environmental changes in the region (specifically the tectonic uplift of the Great Escarpment, the associated establishment of a rain shadow effect across the region, and the establishment of the current biome distributions) supports

the hypothesis that large scale changes in the distributions of vegetation can be linked to the diversification among faunal lineages (Tolley *et al.* 2008; Linder *et al.* 2010; Edwards *et al.* 2011). Apart from the phylogeographic explanations, the non-monophyly of *R. pumilio* strongly suggests that a taxonomic revision is needed.

4.1 Genetic relationships

The existence of four genetically-distinct monophyletic mtDNA clades within *Rhabdomys* (Fig. 2.2), and the non-monophyly of *R. pumilio,* is in sharp contrast to the currently proposed taxonomy for the genus (Musser & Carleton 2005). Although strong genetic evidence exists to support the recognition of the mesic-adapted *R. dilectus* (with the proposed subspecies *R. d. dilectus* and *R. d. chakae*; Rambau *et al*. 2003), the formerly recognized *R. pumilio* consists of at least three evolutionarily distinct geographic mtDNA lineages (Coastal, Central and Northern clades). The partial lack of phylogenetic resolution of the nuclear data (only supporting two assemblages) is probably the result of incomplete lineage sorting (Felsenstein 2004) due to the retention of ancestral polymorphisms (Maddison & Knowles 2006) and the larger effective population size of nDNA compared to mtDNA (Ballard & Whitlock 2004). This is particularly likely in the case of *Rhabdomys* due to the short time between divergence events (supported by the lack of resolution for the phylogenetic position of the Central clade; Figs. 2.2a, 2.3).

Given the high level of mtDNA sequence divergence among the three monophyletic mtDNA clades, the partial support from nuclear DNA markers, and the paraphyletic clustering with *R. dilectus*, we propose that the clades within *R. pumilio* represent distinct species. The recognition of these species is further supported by ecological divergence as indicated by the ecological niche modelling and fairly narrow contact zones in regions where biomes meet (Fort Beaufort (FB) and Sandveld (SV); Fig. 2.2). Although further studies incorporating additional nuclear and more detailed morphometric data are needed to gain additional support for the formal description of species, the non-monophyly of *R. pumilio* provides compelling evidence that a taxonomic revision for *Rhabdomys* is needed. The following recommendations are intended to guide future taxonomic revision (also see Rambau *et al*. 2003). When comparing the geographic distributions of the seven morphologically-defined subspecies (Meester *et al.* 1986) with the genetic patterns (Figs. 2.1, 2.2), some broad concordance is evident. The Coastal clade coincides with the distribution of *R. p. pumilio* (Sparrman 1784),

which was first described from Knysna and the Cape of Good Hope. *Rhabdomys pumilio intermedius* (Wroughton 1905), described from Deelfontein in the Northern Cape as well as Craddock, and Port Elizabeth in the Eastern Cape Province, is mostly represented by the Central clade. The distribution of the Northern clade coincides with *R. p. bechuanae* (Thomas 1893), *R. p. griquae* (Wroughton 1905), and *R. p. fouriei* (Roberts 1946), described from several localities north of the Orange River in Namibia, Botswana, and parts of the Northern Cape and North West Provinces of South Africa. Rambau *et al*. (2003) suggested that the subspecies name *R. p. dilectus* (De Winton 1897) be used to describe the mesic-adapted *R. dilectus* and following this operation we propose the following species names for the clades within the currently recognized *R. pumilio*: *Rhabdomys pumilio* Sparrman 1984 has priority for the Coastal clade, subsuming the subspecies *R. p. pumilio* (Sparrman 1984), *R. p. cinereus* (Thomas & Schwann 1904), and *R. p. orangiae* (Roberts 1946). For the Central clade *R. p. intermedius* (Wroughton 1905) is elevated to *R. intermedius* Wroughton 1905. For the Northern clade *R. bechuanae* Thomas 1893 has priority, subsuming subspecies *R. p. bechuanae* (Thomas, 1893), *R. p. griquae* (Wroughton 1905), and *R. p. fouriei* (Roberts 1946).

There are two well-supported subclades within the greater Coastal clade (*R. pumilio*), with Coastal B composed exclusively of individuals from Fort Beaufort (FB) in the Eastern Cape Province, while Coastal A contains individuals from all other localities within the greater Coastal clade. If taxonomic rank is assigned according to sequence distance alone, these two subclades are different at the subspecific level, since their sequence divergence is similar to that among the subspecies (*R. d. dilectus* and *R. d. chakae*) proposed by Rambau *et al*. (2003). However, additional sampling would be needed to make any firm suggestions regarding the two Coastal subclades.

4.2 Biogeography

The underlying mechanisms leading to speciation are often complex and the pattern in *Rhabdomys* is no exception. The most obvious explanation is that this supposed "habitat generalist" (Skinner & Chimimba 2005) actually consists of several specialized taxa. The clades within *Rhabdomys* are noticeably structured according to biomes (also see Chimimba 2001; Russo *et al.* 2010), with most clades being predominantly associated with a specific vegetation type (Fig. 2.2a, b). It is thus not surprising that the ecological niche modelling strongly support that the clades within the arid-

adapted *R. pumilio* occupy distinct environmental niches (Fig. 2.5). Rainfall patterns, specifically annual amount and seasonality, are important determinants of biome boundaries (Mucina & Rutherford 2006). For the Central clade, the seasonality of rainfall is the most important determinant of the distribution. This clade is associated with the arid Nama-Karoo Biome, which receives mostly late summer rainfall (Mucina & Rutherford 2006). The preference for areas receiving summer rainfall is indicated by the fact that increased variance in rainfall seasonality (thus more rainfall outside of predominant summer rainfall regime) results in a deceased probability of occurrence for this clade (Fig. 2.6b). The Coastal clade is associated with the Fynbos and Succulent Karoo Biome, mostly situated in the winter-rainfall zone of South Africa. According to the ecological niche model, altitude is the most important variable limiting its distribution, with mean annual temperature and rainfall seasonality playing a secondary role. This clade prefers relatively higher mean annual temperatures found in lowland areas, however when temperatures increase beyond a certain point the predicted suitability declines rapidly (Fig. 2.6a(ii)). A winter rainfall regime is preferred as indicated by the highest predicted presence in strictly winter rainfall areas (Fig. 2.5a) and a decline in occurrence probability with increased rainfall outside of the winter season (Fig. 2.6a(iii)). The amount of annual rainfall is the most important variable driving the distribution of the Northern clade. It has the broadest biome distribution, occurring mostly in the Nama-Karoo and arid Savanna Biome (Huntley & Walker 1982; Mucina & Rutherford 2006), but also in the Desert Biome. This clade is adapted to arid conditions as indicated by the fact that probability of occurrence declines rapidly with an increase in the amount of annual rainfall. Its distribution in the east appears to be restricted by an increase in annual rainfall associated with the Grassland Biome and moist vegetation type of the Savanna Biome (Huntley & Walker 1982; Mucina & Rutherford 2006). The mesic adapted *R. dilectus* is restricted to the Grassland Biome and moist vegetation type of the Savanna Biome, with *R. d. dilectus* being predominantly associated with savanna vegetation in South Africa and Zimbabwe, while *R. d. chakae* is mostly restricted to the Grassland Biome within South Africa.

The timing of divergences suggest a radiation within *Rhabdomys* between 4.30 and 3.09 Ma (Fig. 2.3), which given the above, likely occurred in response to rainfall and vegetation changes in southern Africa around the Miocene-Pliocene boundary. This period was characterized by an increase in the east-west rainfall gradient (Partridge 1997; Tyson & Partridge 2000), coupled with an increase in aridity and seasonality of rainfall in the interior (McCarthy & Rubidge 2005), and the establishment of the western winter-rainfall zone (Coetzee 1978; Linder & Hardy 2004). The

overall result was a change from sub-tropical vegetation, which dominated most of the Miocene (Scott *et al.* 1997; Bamford 2000), to the precursors of the modern biomes (Coetzee 1978; Scott *et al.* 1997; Mucina & Rutherford 2006). Divergence within the arid adapted *R. pumilio* probably followed the radiation of the Fynbos, Succulent and Nama-Karoo Biomes within southern Africa at the end of the Miocene (Scott *et al.* 1997). Savanna and grassland vegetation also expanded during this time, with a shift from predominantly C3 to C4 vegetation (Cerling *et al.* 1997; Bonnefille *et al.* 2004; Hopley *et al.* 2007). This probably led to the divergence of *R. dilectus* within the grassland and moist savanna, as also previously suggested by Rambau *et al*. (2003).

Interestingly, two zones of secondary contact among distinct genetic lineages occur where different biomes meet (Fig. 2.2a, b). Specifically, individuals from *R. pumilio* (Coastal clade), *R. intermedius* (Central clade) and the mesic *R. d. chakae* occur sympatrically in the Eastern Cape Province in the region also known as the "Bedfort-gap" (Fig. 2.2b; Lawes 1990). This region is characterized by transitional vegetation of the Albany Thicket Biome (Vlok & Euston-Brown 2002), interspersed by the patches of the Nama-Karoo, Fynbos, Grassland and Savanna Biomes (Mucina & Rutherford 2006). This area is also the cross-over point from aseasonal to summer rainfall (Schulze 1997; Chase & Meadows 2007) and from a mixture of C3 and C4 to predominantly C4 vegetation (Vogel 1978; Vogel *et al.* 1978). In the Free State province at Sandveld (SV), where arid savanna and grassland vegetation types interface, there is a narrow contact zone between *R. bechuanae* and *R. d. dilectus* and these taxa have been shown to have different ecological preferences allowing them to co-exist in this area (Ganem *et al.* 2012).

Given the habitat specificity of *Rhabdomys* lineages, it is expected that their distributions will shift during repeated glacial-interglacial cycles (Zachos 2001). Within southern Africa, climatic oscillations led to drier conditions during glacial and wetter conditions during interglacial phases (Shackleton & Kennet 1975; Partridge 1997; Scott *et al.* 1997). Decreased rainfall during glacial times would have resulted in the expansion of the more arid vegetation types such as savanna, succulent karoo, and nama-karoo at the expense of grassland and fynbos (Ellery *et al.* 1991; Scott & Nyakale 2002; Dupont 2006). The expansion of arid biomes is reflected by the more extensive predicted distribution ranges of the three *R. pumilio* clades during the LGM (Fig. 2.5b). *Rhabdomys bechuanae* (Northern clade) had a much more extensive predicted distribution, while habitat towards the east of South Africa was more favorable for *R. pumilio* (Coastal clade) and *R. intermedius* (Central clade). It can be postulated that during inter-glacial periods, the ranges of these

clades probably contracted to some extent and that populations in more suitable areas of the distribution would have maintained larger population sizes, and would thus have the highest nucleotide diversity. Indeed Otjiamongombe (OR) and Narais (NR) within *R. bechuanae* and Rocher Pan (RR) within *R. pumilio*, which are also located within areas with high predicted suitability under both present and past climates, have significantly higher nucleotide diversities (Figs. 2.4, 2.5a, b).

Apart from the vegetation differences, landscape features may also contribute towards the preservation of the species boundaries by acting as physical barriers to dispersal. For example, altitude was shown to be the most important variable in determining the distribution of *R. pumilio* (Fig. 2.6a(i)). The western Great Escarpment (Fig. 2.2b) presents such an altitudinal barrier, separating *R. pumilio* from *R. intermedius*. Uplift of the Escarpment, which occurred around 5 Ma (Partridge 1997; Partridge & Maud 2000), may have thus also contributed to their isolation. In addition, the Orange River (Fig. 2.2b) appears to form a barrier separating *R. bechuanae* from the remaining species (also see Bauer 1999; Lamb & Bauer 2000; Matthee & Flemming 2002). This contemporary barrier, however, is not absolute as indicated by the inclusion of Groblershoop (GH) below the river within *R. bechuanae*. Nevertheless, both the uplift of the Great Escarpment and the course of the river in the west were established toward the end of the Miocene (Partridge & Maud 2000; Goudie 2005), and there seems to be a correlation between these events and the dates of divergence within *Rhabdomys* (Fig. 2.3).

Chapter 3

The sympatric occurrence of two genetically divergent lineages of sucking louse, *Polyplax arvicanthis* (Phthiraptera: Anoplura), on the four-striped mouse genus, *Rhabdomys* (Rodentia: Muridae)

1. Introduction

Sucking lice (Anoplura) are obligate, permanent ectoparasites of eutherian mammals and are believed to have shared an intimate biological relationship with their hosts through evolutionary time, as evidenced by the high incidence of host specificity and monoxeny (Kim 2006). The probable origin of Anoplura date back to the mid-Cretaceous period and the diversity within the group is frequently correlated to the divergence of placental mammals (Hopkins 1949; Ledger 1980; Smith *et al.* 2011). It is believed that global parasite biodiversity is currently greatly underestimated (Bensch *et al.* 2000; Poulin & Morand 2004; Locke *et al.* 2010). This is also exemplified by the Anoplura since the number of recognized species worldwide (n = 532) has quadrupled within the last century (Durden & Musser 1994) and it has been speculated that the true number of species is probably between 1000 -1500 (Kim *et al*. 1990).

The genus *Polyplax* (Phthiraptera: Anoplura) contains 78 known species, which occur predominantly on members of the rodent family Muridae (Durden & Musser 1994). *Polyplax arvicanthis* (Bedford 1919) has been documented exclusively on the four-striped mouse genus, *Rhabdomys* (Ledger 1980) which is widely distributed throughout the southern African subregion (Skinner & Chimimba 2005). The specificity of *P. arvicanthis* is reflected in the nomenclature (*Rhabdomys* was originally assigned to the genus *Arvicanthis* (Meester *et al.* 1986)), and where this association has been thoroughly studied, *P. arvicanthis* has been recorded to have high prevalence (60%) and abundance (Matthee *et al.* 2007; Matthee *et al.* 2010).

Within *Rhabdomys* (Thomas 1916), a single species (*R. pumilio*) was originally recognized, until molecular evidence (supported by ecological divergence) led to the recognition of two species; the arid-adapted *R. pumilio* (Sparrman 1784) and mesic-adapted *R. dilectus* (De Winton 1897; (Rambau *et al.* 2003; Musser & Carleton 2005). Subsequently, more in depth analyses indicated that both *R. dilectus* and *R. pumilio* contain multiple genetic lineages (Chapter 2; Castiglia *et al*. 2011). Pertinent to the focus of the current study, the arid-adapted *R. pumilio* consists of three geographically structured genetic lineages representing distinct species (Chapter 2). When this diversity is incorporated into *Rhabdomys* taxonomy, it can be argued that at least four species exist within southern Africa. The names *R. dilectus, R. pumilio, R. intermedius* and *R. bechuanae* have been suggested based on the distributions of previously described subspecies and are further used herein (Fig. 3.1; Rambau *et al.* 2003; Chapter 2).

Cryptic species (morphologically similar, but genetically distinct; Andrews *et al.* 1998) have been documented within a wide range of parasitic groups (see de León & Nadler 2010; Nadler & de León 2011; Perkins *et al.* 2011). Parasites are especially prone to harbour cryptic diversity since their reduced bodily features, small size, and morphological stasis often make finding variable morphological characters problematic (Perkins *et al.* 2011), particularly among closely related species (Nadler & de León 2011). Hence, the importance of molecular tools for cataloguing diversity within parasites, preferably using multiple genes, is being increasingly recognized (Nadler 1990; McManus & Bowles 1996; Nadler & de León 2011; Perkins *et al.* 2011). Since the current description of *P. arvicanthis* is based solely on morphological characteristics (Johnson 1960; Ledger 1980) and given that this parasite has a large geographic range spanning four host species within southern Africa (Chapter 2), it is probable that undetected evolutionary divergence may exist within what is currently recognized as a single species.

To test the hypothesis that hidden genetic diversity, potentially associated with *Rhabdomys* divergences (Rambau *et al.* 2003; Chapter 2), may be present within *P. arvicanthis* we embarked on an investigation using multiple molecular markers and included several *P. arvicanthis* specimens sampled throughout southern Africa.

2. Materials and Methods

2.1 Taxon and gene sampling

Polyplax arvicanthis specimens were collected from 16 localities across the distribution of the four *Rhabdomys* species (Table 3.1; Fig. 3.1). Live traps (Sherman-type) baited with a mixture of peanut butter and oats were used to capture host individuals. Mice were euthanized with 0.2-0.4 ml sodium pentobarbitone (200 mg/kg) and placed in individual plastic bags to prevent the loss of ectoparasites post mortem. Host specimens were frozen in the field at -20°C and subsequently thawed in the laboratory, where all lice were removed with forceps under a stereoscopic microscope. For the DNA analyses, *P. arvicanthis* specimens were selected from as many different host individuals as possible per sampled locality (determined by parasite prevalence and abundance; Table 3.1), and placed in a 100% EtOH solution. The remainder of lice collected from *Rhabdomys* at each site were preserved in 70% EtOH for morphological confirmation (identifications provided by L.A. Durden;

Department of Biology, Georgia Southern University, USA). All specimens were identified as belonging to a single species, *P. arvicanthis,* based on gross morphological examination.

Table 3.1: Geo-referenced localities and hosts from which *Polyplax arvicanthis* were collected. The total number of hosts captured, number of host with lice, total number of lice collected, and the sub-sample of lice (specified in Table B.8) used in subsequent analyses are indicated per sampled locality.

Locality	Code	Geographic coordinates	Host species	Total hosts	Hosts with lice	Total lice	Sub-sampled lice
Windhoek	WH	22°31' S, 17°25' E	*R. bechuanae*	20	17	20	2
Keetmanshoop	KH	26°21' S, 18°29' E	*R. bechuanae*	21	12	20	1
Richtersveld	RV	28°12' S, 17°06' E	*R. pumilio*	31	27	27	2
Springbok	GP	29°42'S, 18°02' E	*R. pumilio*	30	28	28	2
Groblershoop	GH	28°37' S, 21°42' E	*R. bechuanae*	14	8	18	1
Sutherland	SL	32°24' S, 20'54' E	*R. intermedius*	13	6	17	1
Rooipoort	RP	28°39' S, 24°8' E	*R. bechuanae*	15	11	28	2
Vanrhynsdorp	VR	31°44´ S, 18°46´ E	*R. pumilio*	30	23	23	2
Porterville	PV	32°59´ S, 19°01´ E	*R. pumilio*	30	18	18	2
Stellenbosch	SB	33°55' S, 18°49' E	*R. pumilio*	31	12	15	1
De Hoop	DH	34°29' S, 20°24' E	*R. pumilio*	19	7	14	2
Oudtshoorn	OH	33°36' S, 22°08' E	*R. pumilio*	31	29	29	2
Beaufort West	BW	32°13' S, 22°48' E	*R. intermedius*	33	15	20	2
Laingsburg	LB	33°10' S, 20°55' E	*R. intermedius*	23	7	10	2
Chelmsford	CH	28°00' S, 29°54' E	*R. dilectus*	7	6	8	2
Alice	AL	32°47' S, 26°50' E	*R. dilectus*	6	2	4	1
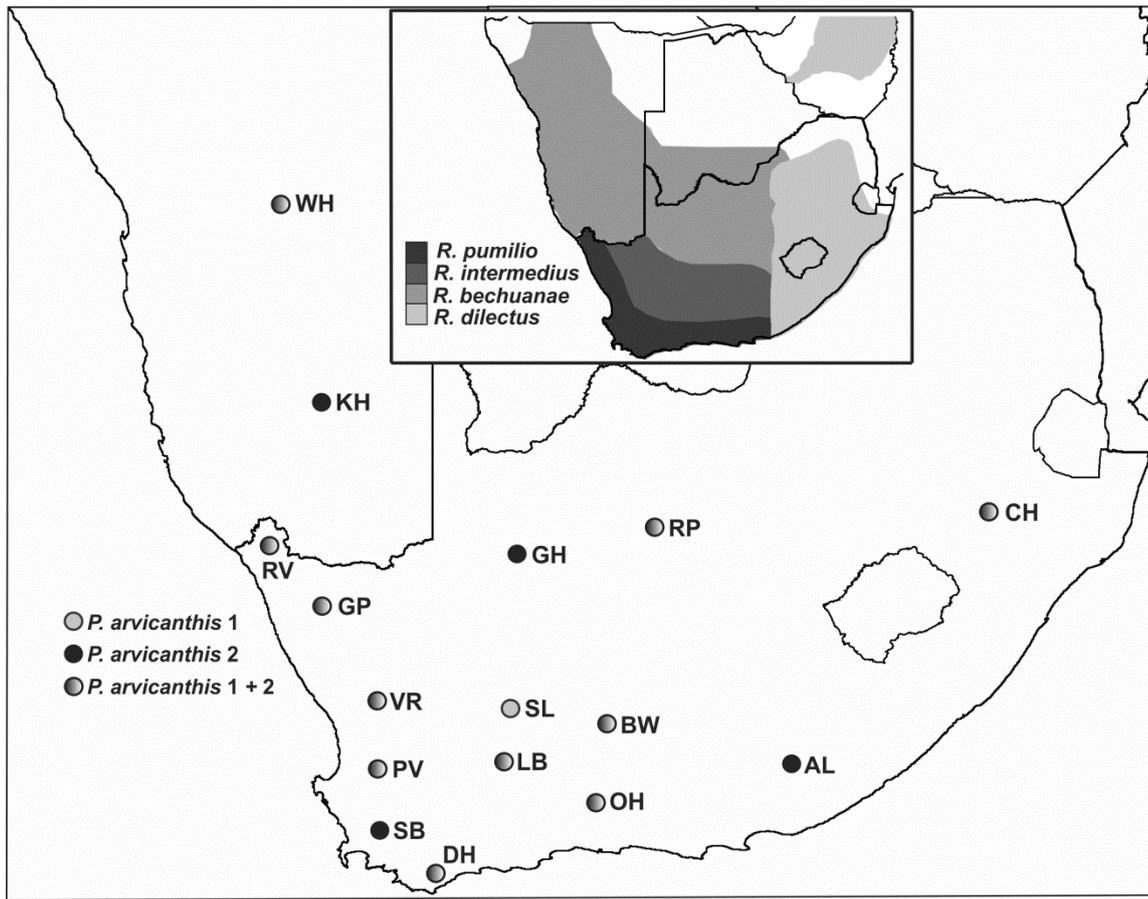
Figure 3.1: Localities from which *Polyplax arvicanthis* were sampled, indicating areas of sympatric and allopatric occurrence of the two clades (*P. arvicanthis* 1 and 2), with locality codes as in Table 3.1. The inset represents the distribution of the different host species following Rambau *et al.* (2003) and Chapter 2.

Mitochondrial COI (Cytochrome Oxidase I) sequence data were generated for 299 *P. arvicanthis* specimens from the 16 sampled localities (Table 3.1) and these sequences were collapsed to haplotypes (Collapse 1.2; Posada 2004). Phylogenetic reconstructions (outlined below) were performed on these haplotypes, which indicated the presence of two highly divergent genetic clades (see results). Based on the need to include multiple markers for accurate phylogenetic inference (Nadler 2002), the COI gene tree was supplemented with data derived from the mitochondrial 12S and 16S rRNA, and nuclear CAD (carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase) genes. The latter was performed for a representative sub-sample of the COI haplotypes, specifically selected to optimally cover the genetic and geographic variation present within *P. arvicanthis* based on the results of the phylogenetic analyses (see below). One haplotype from each COI clade present at each of the 16 sampled localities were included (in most instances two haplotypes per locality, Tables 3.1, B.8, B.9; Fig. 3.1).

2.2 Molecular techniques

Total genomic DNA was extracted from whole individual louse specimens with a commercial kit (Qiagen, DNeasy® Blood and Tissue). PCR and sequencing of all gene fragments were performed following standard polymerase chain reaction (PCR) protocols in a GeneAmp® PCR system 2700 thermal cycler (Applied Biosystems). General PCR cycling conditions for the 12S rRNA, 16S rRNA and CAD genes included an initial denaturation of 3 min at 94°C followed by 30-40 cycles of 30s denaturation at 94°C, 45-60s annealing at the primer-specific temperature (Table 3.2), and 45-60s extension at 72°C, followed by a final extension period of 5 min at 72°C. For the COI gene, the primers LCOIP6625 and HCOIPrev were used (Table 3.2) and the PCR conditions were as follows: an initial denaturation of 1 min at 95° C followed by 10 cycles of 1 min denaturation at 95°C, 1 min annealing at 45°C, and 1 min extension at 72°C after which 30 cycles of 1 min denaturation at 93°C, 1 min annealing at either 58°C or 59°C (depending on amplification success), and 1 min extension at 72°C was performed followed by a final extension of 5 min at 72°C.

Table 3.2: Primers used for PCR amplification of the various gene fragments.

Gene	Primer name	F/R	Sequence (5'-3')	Annealing Temperature (°C)
COI	LCOIP6625[1]	F	CCCGATCCTTTTGGTTTTTTGGGCATCC	45-58/59° C
	COIPRev[2]	R	CCYCCTACNGTAAAYAGRAARATRAARC	
12S	12SAL[3]	F	AAACTGGGATTAGATACCCCACTAT	50°C
	12SBH[3]	R	AGAGTGACGGGGGGTGTGT	
16S	16SF[4]	F	TTAATTCAACATCGAGGTCGCAA	58°C
	Lx16SR[4]	R	GACTGTGCTAAGGTAGCATAAT	
CAD	CADPfor[5]	F	ACGACAACTGCATTACCGTTTGCA	55°C
	CADPrev[6]	R	CCACCGGGGAATTTTGACAAC	

[1] Adapted from L6625, Hafner *et al*. 1994; [2] Designed in this study; [3] Palumbi *et al*. 1991; [4] Shao *et al*. 2009; [5] Adapted from ApCADfor1, Danforth *et al*. 2006; [6] Adapted from Ap835rev1, Danforth *et al*. 2006

Aliquots of PCR products (5µl) were separated by electrophoresis on a 1% agarose gel for visual inspection, after which purification was performed on the remaining PCR product with a commercial kit (Macherey-Nagel, NucleoFast 96 PCR Kit). In some instances fragments were excised and purified using a commercial gel purification kit (Promega, Wizard® SV Gel Clean-Up System). All cycle-sequencing reactions were performed using BigDye Chemistry and products were analysed on an automated sequencer (ABI 3730 XL DNA Analyzer, Applied Biosystems).

2.3 Data analysis

Sequences were edited in BioEdit Sequence Alignment editor 7.0.5 (Hall, 2005) and aligned with multiple alignment mode in Clustal X2 (Larkin *et al.* 2007). The ends of sequences were trimmed to avoid the inclusion of missing data, with the final alignment lengths as indicated in Table 3.3. Within the COI alignment, a 3 bp or 6 bp insert was present within *P. arvicanthis* specimens when compared to outgroup taxa. No double reads were present in the original chromatograms and to further ensure that these sequences represented the functional copy of the protein coding COI gene, all sequences were translated into protein sequences using the online tool EMBOSStranseq (www.ebi.ac.uk/Tools/st/emboss_transeq).

Table 3.3: The number of ingroup samples (N), amplified and final alignment length, polymorphic sites (P), and parsimony informative sites (PI) for each gene fragment.

Gene fragment	N	Amplified length	Alignment length	P	PI
COI	27	300	270*	147	115
12S	26	420	406	185	141
16S	27	350	336	133	92
CAD	27	370	349	80	30

* excluding the two 3 bp indels coded as presence/absence

2.3.1 Phylogenetic reconstructions

A single specimen of *Haematopinus phacochoeri* (sampled from the warthog, *Phacochoerus africanus*) was used as a distant outgroup in all analyses (GenBank accession JX 218028 - 218030; Table 3.4). Several additional species of Anoplura, for which sequence data were available on GenBank, were used as reference taxa in the various phylogenetic analyses (Table 3.4). Parsimony and Bayesian reconstructions were performed on all the COI haplotypes. For the subsampled datasets, parsimony and Bayesian trees were constructed for each gene fragment individually (COI, 12S, 16S, and CAD) followed by combined analyses of the mtDNA fragments as well as a combined mtDNA and nDNA data set. The COI insertions were coded as present/absent data and included in the analyses. Sequence ambiguities resulting from heterozygous positions of the nuclear CAD gene were treated as polymorphisms. To further investigate the divergence among the two *P.*

arvicanthis clades with respect to other recognised *Polyplax* species, a reduced COI Bayesian topology was constructed using data derived from GenBank (Tables 3.4, B.10).

Table 3.4: GenBank accession numbers for outgroup taxa used in the different gene analyses.

Taxon	Gene fragment	Accession
Fahrenholzia pinnata	COI	EF152557.1
	CAD	FJ267404.1
Hoplopleura ferrisi	COI	HM171428.1
Pedicinus badii	CAD	FJ267414.1
Pediculus humanus capitis	12S	AY139881.1
	16S	AY139928.1
	CAD	FJ267404.1
Polyplax serrata	COI	EU162172.1
Haematopinus phacochoeri	COI	JX218028
	12S	JX218029
	16S	JX218030

Unweighted parsimony analyses were conducted in PAUP* v4.0b10 (Swofford 2000). In each analysis, the heuristic search option with random taxon addition (10 replicates) and tree bisection and reconnection (TBR) branch swapping was implemented and a maximum of 100 equally parsimonious trees were saved during each step. Nodal support was assessed with 1000 bootstrap replicates (Felsenstein 1985). Best-fit models of sequence evolution for all gene fragments were determined under the AICc (Akaike 1973; Burnham & Anderson 2002, 2004) in jModelTest v0.1.1 (Guindon & Gascuel 2003; Posada 2008) and implemented in Bayesian tree reconstructions in MrBayes v3.2 (Ronquist *et al.* 2012). The effect of codon partitioning on the marginal likelihoods for the protein coding COI and CAD genes was evaluated with Bayes factors (Kass & Raftery 1995), as calculated in Tracer v1.5 (Newton & Raftery 1994; Suchard *et al.* 2001; Rambaut & Drummond 2007). For all analyses, only the general structure of the model was defined and the default priors were used to estimate parameters. In each analysis, two parallel Markov Chain Monte Carlo (MCMC) simulations, consisting of five chains each, were run for 2-5 million generations depending on when it was estimated that stationarity had been reached. Trees and parameters were sampled every 100 generations, and 25% of the total number of generations sampled were discarded as burn-in after convergence and ESS values were assessed in Tracer v1.5 (Rambaut & Drummond 2007). Posterior probabilities for nodal support were obtained by using the sumt command in

MrBayes. COI GTR-corrected sequence distances were calculated in PAUP* v4.0b10 (Swofford 2000) to compare sequence divergence estimates among species/lineages.

3. Results

The 299 *P. arvicanthis* specimens revealed 93 COI haplotypes (Table B.8; GenBank JX629372 - JX629438, JX198372 - JX198398). A total of 27 COI haplotypes were selected (see above, Table B.8) and sequences were generated for the 16S rRNA and CAD gene fragments (GenBank JX198319 - JX198345, JX198399 - JX198425; Tables 3.3, B.9). Despite numerous attempts, 12S rRNA data for one haplotype (LB_1) could not be obtained, resulting in a total of 26 haplotypes for this dataset (GenBank JX198346 - JX198371; Tables 3.3, B.9).

Bayes factors indicated that partitioning by codon position with a separate model assigned to each partition was preferred over an unpartitioned scheme with a single model, for both the protein coding COI and CAD genes (Table B.11). Thus, the JC model (nst=1, rates=equal) was assigned to the first and second codon positions of COI and all three codon positions of CAD. The HKY (nst=2, rates=equal), GTR+I+G (nst=6; rates=invgamma), and GTR+G (nst=6; rates=gamma) models were specified for the third codon position of COI, the 12S rRNA, and 16S rRNA genes respectively.

Phylogenetic analyses of the 93 COI haplotypes revealed the presence of two well supported genetic clades (Table 3.5, Fig. A.5) differentiated by an average of 25% (±0.02) sequence divergence, which is comparable to values observed among other recognized species of *Polyplax* (Table B.10). Considerable variation is also present within the two *P. arvicanthis* clades, especially within *P. arvicanthis* 2 (Fig. 3.2) which has an average intra-clade divergence of 16% (±7.34) compared to 11% (±7.91) within *P. arvicanthis* 1. For the representative subset of 26/27 haplotypes, the monophyly of *P. arvicanthis* was strongly supported in nearly all individual and combined analyses (parsimony and Bayesian; Table 3.5; Fig. 3.2) of the various genes. The combined analyses (parsimony and Bayesian) of the three mtDNA fragments (Fig. 3.2) provided the highest posterior probability support for the existence of two clades within *P. arvicanthis* (1 and 2) and the same pattern emerged when all the data were combined (Table 3.5; also see (Light & Reed 2009; Light *et al.* 2010). Support for the two clades varied among the individual parsimony and Bayesian

gene trees (Table 3.5), and in instances where the nodes were not obtained they reflect polytomies (unresolved). All of the nodal uncertainty found in the individual gene analyses surrounded the monophyly of *P. arvicanthis* 2. Importantly, however, individuals belonging to the two *Polyplax* clades are consistently differentiated by the presence of the 6 bp (*P. arvicanthis* 1) or 3 bp (*P. arvicanthis* 2) insert within COI when compared to other *Polyplax* species, and this more conservative substitution (Matthee *et al.* 2001) provides further support for the monophyly of each of the two lineages.

Table 3.5: Bootstrap and posterior probability support values for the monophyly of *Polyplax arvicanthis* and the two clades therein resulting from the various single and combined gene analyses.

Clade	Gene fragment						
	COI (full)	**COI (subset)**	**12S**	**16S**	**mtDNA**	**CAD**	**mtDNA+nDNA**
P. arvicanthis	78/1	82/0.99	100/0.8	98/0.99	85/1	100/1	96/1
P. arvicanthis 1	72/0.99	82/0.99	59/0.57	82/0.98	93/0.99	78/0.90	98/1
P. arvicanthis 2	76/0.99	82/0.98	89/0.89	nf/nf	95/0.96	nf/nf	97/0.91

nf = not found
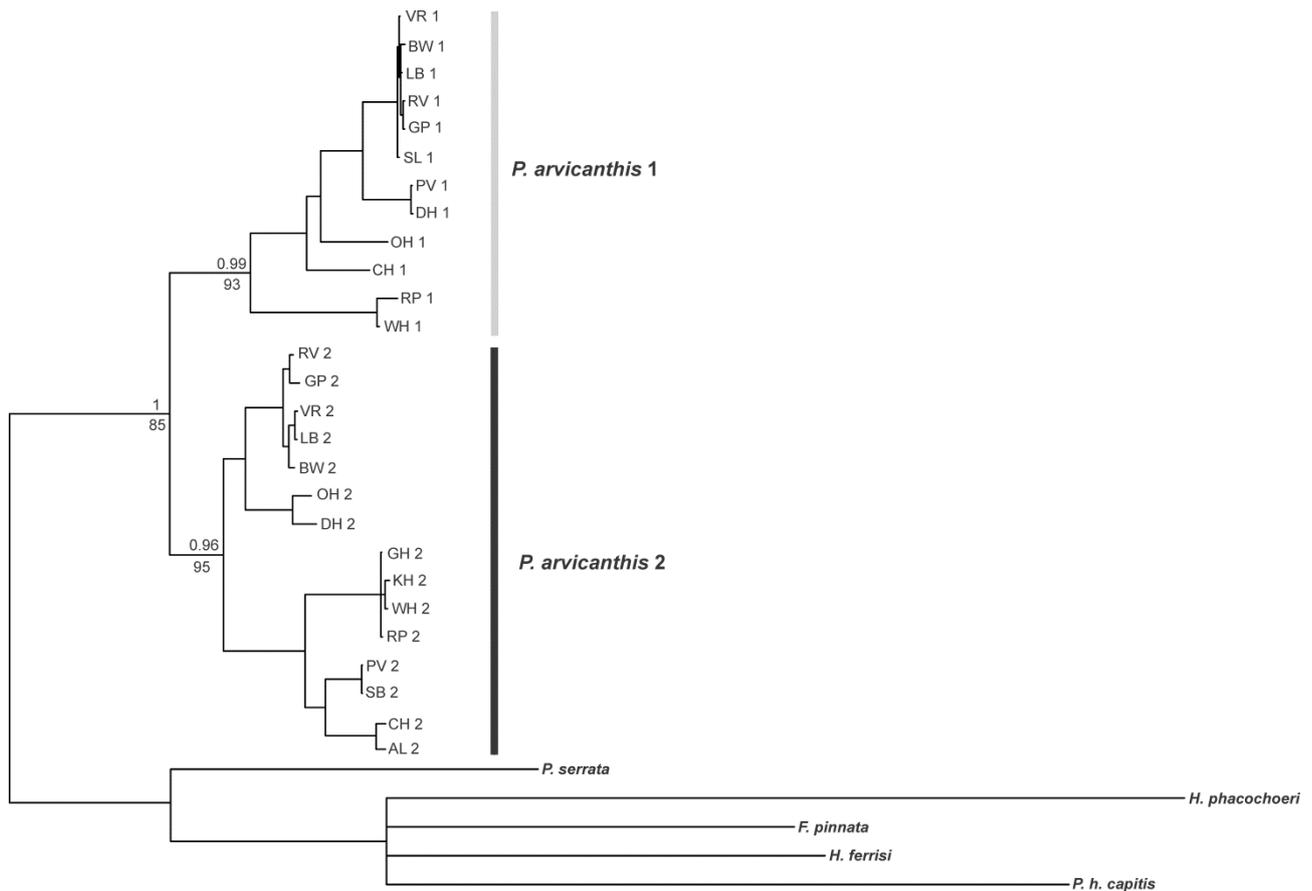
Figure 3.2: Consensus parsimony and Bayesian topology of the combined mtDNA dataset (posterior probabilities above and bootstraps below nodes), indicating the two clades within *Polyplax arvicanthis*.

Analysis of the phylogenetic relationships within *Polyplax* (using COI data obtained from GenBank; Fig. 3.3), place the two *P. arvicanthis* lineages as sister taxa within a clade containing *P. spinulosa*, *P. borealis* and *P. serrata.* Branch lengths separating the *P. arvicanthis* lineages are comparable to those separating the other recognized species. The topology also supports the previously reported non-monophyly of the family *Polyplacidae* (Light *et al.* 2010). It is, however, important to realize that the relationships within the genus as portrayed herein is based on a fraction of the total number of known *Polyplax* species and additional data (species and molecular markers) are needed to obtain more robust support for these preliminary phylogenetic findings.
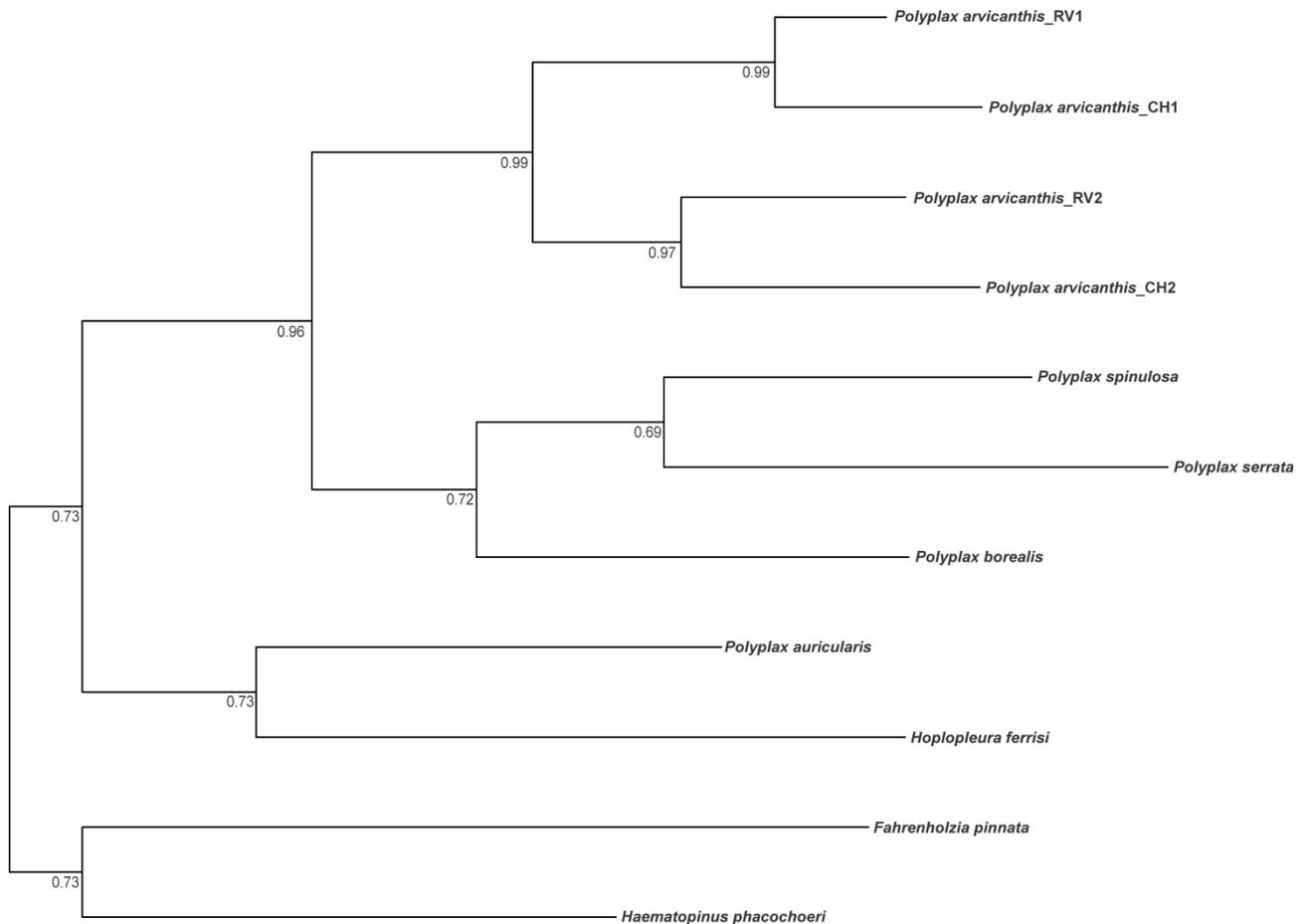
Figure 3.3: Bayesian COI topology indicating the phylogenetic position of the two clades within *Polyplax arvicanthis* with respect to other recognized *Polyplax* species (GenBank accession numbers listed in Tables 3.4, B.10).

4. Discussion

The existence of two genetically distinct, reciprocally monophyletic clades within *P. arvicanthis* is strongly supported in nearly all our analyses (Table 3.5; Fig. 3.2, 3.3). The COI sequence divergence between these clades (Table B.10), also illustrated by branch lengths (Fig. 3.3), is comparable to that found between other *Polyplax* species. Although the exact taxonomic status of these clades is not yet clear and more data from African *Polyplax* is needed to determine whether they represent sister-taxa, this questions the notion that *P. arvicanthis* represents a single species (Sparrman 1784; Ledger 1980).

The apparent absence of morphological differentiation among the two genetically divergent *P. arvicanthis* lineages suggests that they may represent cryptic species. The discovery of cryptic species through the use of molecular tools has been documented widely in several parasitic groups (de León & Nadler 2010) and this may be attributed to the reduced morphological features associated with the parasitic mode of life (Nadler & de León 2011; Perkins *et al.* 2011). It should, however, be noted that further morphological investigation, especially using features that have not previously been used in alpha taxonomy, may reveal differences among the *P. arvicanthis* lineages (Yoshizawa & Johnson 2006; Schlick-Steiner *et al.* 2007), thus providing further evidence for the existence of two species on *Rhabdomys*. Until such investigations have been conducted, the two genetic lineages identified in our study should therefore remain provisionally cryptic (de León & Nadler 2010).

Parasite genetic divergence does not mirror that observed within the host since, for both *P. arvicanthis* lineages, the lice collected from *R. dilectus*, *R. pumilio*, *R. intermedius*, and *R. bechuanae* do not form monophyletic entities (Fig. 3.1, 3.2; Chapter 2). In addition, the two parasite lineages have sympatric distributions throughout most of the sampled range of all host lineages (Fig. 3.1). Our findings are not unique in the sense that closely related lineages of sucking lice occurring sympatrically on a single host species has been documented previously (Reed *et al.* 2004; Štefka & Hypša 2008). It has been postulated that the sympatric occurrence of closely related parasitic groups may arise through colonization (switching from another host) or parasite duplication (shared common ancestor on particular host; Page 2003). A sister relationship between the two lineages would thus favour parasite duplication over colonization due to host-switching. In the present analysis the two divergent *P. arvicanthis* lineages represent sister-taxa (0.99 posterior probability support; Fig. 3.3), which in combination with the presence of unique insertions within both *P. arvicanthis* lineages (when compared to other available *Polyplax* species) supports the idea of a common ancestor on *Rhabdomys* followed by duplication. Unfortunately, the paucity of sequence data for African *Polyplax* precludes a firm statement regarding the sister-taxon status of these lineages and we cannot rule out the possibility that the insertions may also be present in other African *Polyplax*. However, *P. arvicanthis* is specific to *Rhabdomys* as it has not been documented on any other co-occurring rodent taxa within southern Africa (Ledger 1980; Durden & Musser 1994), and is also morphologically distinct from the *Polyplax* species parasitizing these co-occurring hosts (Ledger 1980). Given the specificity to *Rhabdomys* and the apparent lack of

morphological differences among the two *P. arvicanthis* lineages, a recent common ancestor on *Rhabdomys* followed by a parasite duplication event is a more parsimonious evolutionary scenario than morphological convergence of distantly related louse lineages following colonization of one of the lineages from another host (i.e. host-switching).

Parasite duplication may occur via sympatric or allopatric speciation (Johnson & Clayton 2004). Multiple instances of sympatric speciation would be needed to explain the broad sympatric occurrence of the two louse taxa in this study. Thus, parasite duplication through allopatric speciation caused by temporary fragmentation of host populations (Page 2003), followed by mutual secondary colonisations of hosts, appears to be the most parsimonious explanation for the pattern we observed. *Rhabdomys* probably experienced multiple expansion-contraction cycles during glacial-interglacial phases (Zachos 2001) and allopatric divergence of the parasite would be enhanced by the short generation times (Huyse *et al.* 2005; Whiteman & Parker 2005) and faster evolutionary rates (Paterson & Banks 2001; Nieberding *et al.* 2004) generally observed in parasites. It is thus plausible that the two lineages within *P. arvicanthis* diverged in allopatry while isolated on host lineages within different refugia (Chapter 2). If this holds, the current pattern resulted from mutual host colonisations of the two parasite taxa across the region. This dispersal would have been facilitated through the secondary contact observed among hosts (Chapter 2). Importantly, however, host sympatry does not necessarily equate to parasite sympatry (McCoy 2003), especially for *P. arvicanthis* which requires bodily contact to move between hosts (Hopkins 1949; Ledger 1980; Marshall 1981). From the parasite's perspective, host syntopy (occupation of same macro-habitat; Rivas 1964) will therefore be more important than host sympatry in terms of dispersal opportunity between hosts. For *P. arvicanthis*, host syntopy probably does occur within the host contact zone and the co-occurrence of both cryptic lineages on the same host individuals implies that parasite syntopy also occurs in certain areas.

The present study clearly lends support to the suggestions of Kim (2006), in that much of Anoplura diversity is yet to be discovered. Within *P. arvicanthis,* detailed morphological comparsions are needed to fully resolve the taxonomic status of the two genetic lineages and given the extensive variation observed within each of the two *Polyplax* assemblages, more fine-scale comparative phylogeographic analyses is required to unravel the complex patterns observed.

Chapter 4

Limited congruence among the genetic structures of two specific ectoparasitic lice and their rodent hosts: biogeography and host-related factors trumps parasite life-history

1. Introduction

Parasites and hosts interact along an evolutionary continuum across micro- and macroevolutionary scales (Brooks & McLennan 1993; Huyse *et al.* 2005). At the macroevolutionary scale repeated co-speciation (contemporaneous cladogenesis; Brooks 1979; Light & Hafner 2008) within a parasite-host system will lead to congruent phylogenetic relationships (co-phylogeny), while erosion of congruence may occur via host-switching, parasite duplication and sorting events (Johnson *et al.* 2003; Page 2003; Clayton *et al.* 2004). Such macroevolutionary patterns are influenced by microevolutionary processes such as selection, drift, and dispersal, all of which may result in intraspecific co-divergence and congruent phylogeographic structures (Nadler 1995; Clayton & Johnson 2003; Clayton *et al.* 2004; Criscione *et al.* 2005; Nieberding & Morand 2006; Štefka & Hypša 2007). While examples of near-perfect congruence between parasite and host genetic structures exist (Hafner *et al.* 1994; Light & Hafner 2007; Demastes *et al.* 2012), these patterns seem to be the exception rather than the rule (Paterson *et al.* 2000; Hoberg & Klassen 2002; Clayton *et al.* 2004; Weckstein 2004; Huyse *et al.* 2005) and parasite-host associations more often represent a complex mosaic of episodic evolutionary events (Hoberg & Brooks 2008).

Sucking lice (Phthiraptera: Anoplura) are obligate and permanent parasites of eutherian mammals (Kim 2006) and their close association with their hosts provide interesting models for the study of parasite-host coevolution (Kim 1985; Reed *et al.* 2004; Light & Hafner 2008). Within southern Africa, the widely distributed four-striped mouse, *Rhabdomys* (Rodentia: Muridae) is parasitized by the sucking louse *Polyplax arvicanthis* (Ledger 1980; Skinner & Chimimba 2005). Recent phylogenetic evidence based on multiple genes indicated that *P. arvicanthis* consists of two divergent lineages (*P. arvicanthis* 1 and 2) with sympatric distributions across most of the sampled range (Chapter 3). The apparent lack of gross morphological differences between the two *Polyplax* lineages suggests the presence of two cryptic species within *P. arvicanthis* (Chapter 3). A concurrent phylogeographic study on the host genus *Rhabdomys* indicated the existence of at least four distinct species (*R. pumilio*, *R. intermedius*, *R. bechuanae*, and *R. dilectus*) throughout the southern African distribution (Chapter 2) that are restricted to certain vegetational biomes (Rambau *et al.* 2003; Chapter 2).

The extent of parasite-host congruence appears to be primarily determined by the intimacy of the parasite-host interaction as determined by several parasite life-history, ecological and demographic traits (Johnson *et al.* 2003; Huyse *et al.* 2005; Charleston & Perkins 2006; Nieberding & Morand 2006). *Polyplax arvicanthis* possesses several traits expected to lead to strong genetic structuring as well as congruence with the genetic structure of the host, *Rhabdomys*. They are reported to be specific to the host genus, have a direct life cycle with no free-living phase or intermediate hosts and may even complete several generations on a single host individual (Ledger 1980; Durden & Musser 1994; Kim 2006). The dispersal of *P. arvicanthis* is thus completely dependent on that of *Rhabdomys*, and vertical transmission through successive generations is expected (Blouin *et al.* 1995; Jerome & Ford 2002a; Johnson *et al.* 2003; Johnson & Clayton 2004; Whiteman & Parker 2005; Wirth *et al.* 2005). This lifestyle is expected to impose impediments to gene flow and limit opportunities for host-switching to other potential host species (Blouin *et al.* 1995; Nadler 1995; Johnson *et al.* 2002; McCoy *et al.* 2003). The probability that a parasite will track the migration and differentiation patterns of its host is also strengthened by high prevalence and intensity (Clayton *et al.* 2003; Nieberding & Morand 2006), both of which have been found for *P. arvicanthis* on *Rhabdomys* (prevalence = 60%, relative mean infestation intensity = 12.43%; Matthee *et al.* 2007).

Parasites, in general, are also expected to show stronger signatures of genetic structure when compared to their hosts as a result of the combined effects of limited dispersal ability (reduced gene flow), smaller effective population sizes (drift effect particularly in isolated populations), and elevated mutation rates (Nieberding & Morand 2006). It is expected that geographic sampling sites of *P. arvicanthis* will have small effective population sizes since the species is recorded to have a female biased sex ratio (Matthee *et al.* 2007), which is also generally the case for parasites without intermediate hosts and free-living phases (Rannala & Michalakis 2003; Criscione & Blouin 2005). A more pronounced phylogeographic signature is also predicted for *P. arvicanthis* based on the elevated mtDNA mutation rates generally reported for chewing and sucking lice (Hafner *et al.* 1994; Page & Hafner 1996; Page *et al.* 1998; Paterson & Banks 2001; Page *et al.* 2004; Light & Hafner 2007, 2008).

Biogeography (distributions of organisms in space and time) can also significantly shape parasite-host interactions over evolutionary time (Hoberg & Brooks 2008) and in certain parasite-host

systems, biogeography may be equally important, if not more, than host associations in determining the genetic structure of parasites (Clayton 1990; Weckstein 2004; Nieberding *et al.* 2008). A shared biogeographic history with similar responses to vicariant events may also lead to congruent genealogical relationships (Hoberg & Brooks 2008). Importantly, congruence therefore does not necessarily imply contemporaneous parasite and host cladogenesis (Page 2003). In the current study, *Rhabdomys* has a complex biogeographic history and most of the cladogenesis within the genus probably occurred in response to climatic and landscape changes around the Mio-Pliocene boundary (Chapter 2). Several cycles of range expansions and contraction of the various host taxa occurred as a result of climatic oscillations in the region (Zachos 2001; Montgelard & Matthee 2012) and currently three putative species (*R. pumilio*, *R. intermedius*, and *R. dilectus*) have been recorded to occur sympatrically in the Eastern Cape Province at Fort Beaufort (Chapter 2) as well as in the Free State Province (Ganem *et al.* 2012). These areas are thought to represent zones of secondary contact following host range expansion and this host sympatry may have important implications for the parasite's genetic structure by providing opportunities for host-switching (Clayton *et al.* 2004).

The *Polyplax/Rhabdomys* study system provides a unique opportunity to investigate parasite-host associations at the interface between micro- and macroevolutionary scales. Phylogeographic relationships within the two *P. arvicanthis* lineages and the four *Rhabdomys* taxa were compared to test for spatial congruence of genealogical relationships among parasites and hosts, utilizing both mitochondrial and nuclear sequence data. Given the life history of *Polyplax,* we hypothesised that the parasites would have more pronounced spatial genetic structuring than the hosts and that phylogeographic and phylogenetic congruence would be observed among the parasite and host structures. Potential co-phylogeny among the parasite and host genetic groups was evaluated using distance- and topology-based approaches. To determine the relative roles of biogeography and co-divergence in shaping the parasite-host association, putative co-divergence events were assessed for contemporaneity by dating the parasite divergences and comparing these to the associated host divergence times. Contact zones, where host lineages occur sympatrically, were used to evaluate host-specificity by determining whether host-switching occurs among parasites of ecological and genetically divergent host lineages.

2. Materials and Methods

2.1 Taxon and gene sampling

Parasite (lice, n = 327) and host (n = 377) specimens were collected from 17 localities across the southern African distribution of the four putative *Rhabdomys* species (Table 4.1; Fig. 4.1). Live traps (Sherman-type) baited with a mixture of peanut butter and oats were used to capture host individuals. Mice were euthanized with 0.2-0.4 ml sodium pentobarbitone (200 mg/kg) and placed in individual plastic bags to prevent the loss of ectoparasites post mortem. Host specimens were frozen in the field at -20°C and subsequently thawed in the laboratory, where all lice were removed with forceps under a stereoscopic microscope. *Polyplax arvicanthis* specimens included in the genetic analyses were selected from as many different host individuals as possible per sampled locality (determined by parasite prevalence and abundance; Table 4.1), and preserved in 100% ethanol (correct identification of *P. arvicanthis* was morphologically confirmed by L.A. Durden; Department of Biology, Georgia Southern University, USA).

Mitochondrial COI (Cytochrome Oxidase subunit I) sequence data was generated for all host and parasite specimens. These data were augmented by sequencing nuclear introns for as many mtDNA haplotypes as possible. If haplotypes occurred at multiple sampled localities, an attempt was made to sequence at least one individual from each locality. The nuclear Eef1a1 (eukaryotic translation elongation factor 1 alpha 1) and CAD (carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase) were sequenced for *Rhabdomys* and *Polyplax*, respectively (Tables B.12, B.13).

Table 4.1: Localities from which parasite and host specimens were collected indicating host taxon (see Chapter 2), number of hosts captured, parasite prevalence, and the number of specimens for each parasite lineage per locality.

Country/ Province	Locality	Code	Geographic coordinates	Host clade	Hosts (n)	Prevalence (%)	*P. arvicanthis* 1 (n)	*P. arvicanthis* 2 (n)
Namibia								
	Windhoek	WH	22°31' S, 17°25' E	*R. bechuanae*	20	85	18	2
	Keetmanshoop	KH	26°21' S, 18°29' E	*R. bechuanae*	21	57	absent	20
South Africa								
Northern Cape	Groblershoop	GH	28°37' S, 21°42' E	*R. bechuanae*	14	57	absent	18
	Rooipoort	RP	28°39' S, 24°08' E	*R. bechuanae*	15	73	1	27
	Richtersveld	RV	28°12' S, 17°06' E	*R. pumilio*	31	87	15	12
	Springbok	GP	29°42' S, 18°02' E	*R. pumilio*	30	93	21	7
	Sutherland	SL	32°24' S, 20'54' E	*R. intermedius*	13	46	15	2
Western Cape	Vanrhynsdorp	VR	31°44' S, 18°46' E	*R. pumilio*	30	76	16	7
	Porterville	PV	32°59' S, 19°01' E	*R. pumilio*	30	60	12	6
	Stellenbosch	SB	33°55' S, 18°49' E	*R. pumilio*	31	38	absent	15
	De Hoop	DH	34°29' S, 20°24' E	*R. pumilio*	19	36	10	4
	Oudtshoorn	OH	33°36' S, 22°08' E	*R. pumilio*	31	93	21	8
	Laingsburg	LB	33°10' S, 20°55' E	*R. intermedius*	23	30	2	8
	Beaufort West	BW	32°13' S, 22°48' E	*R. intermedius*	33	45	13	7
Eastern Cape	Fort Beaufort_Smutskraal	FB_sk	32°50' S, 26°27' E	Contact zone	14	57	absent	18
	Fort Beaufort_Winterberg	FB_wb	32°47' S, 26°37' E	Contact zone	15	73	3	11
Kwazulu-Natal	Chelmsford	CH	28°00' S, 29°54' E	*R. dilectus chakae*	7	85	4	4
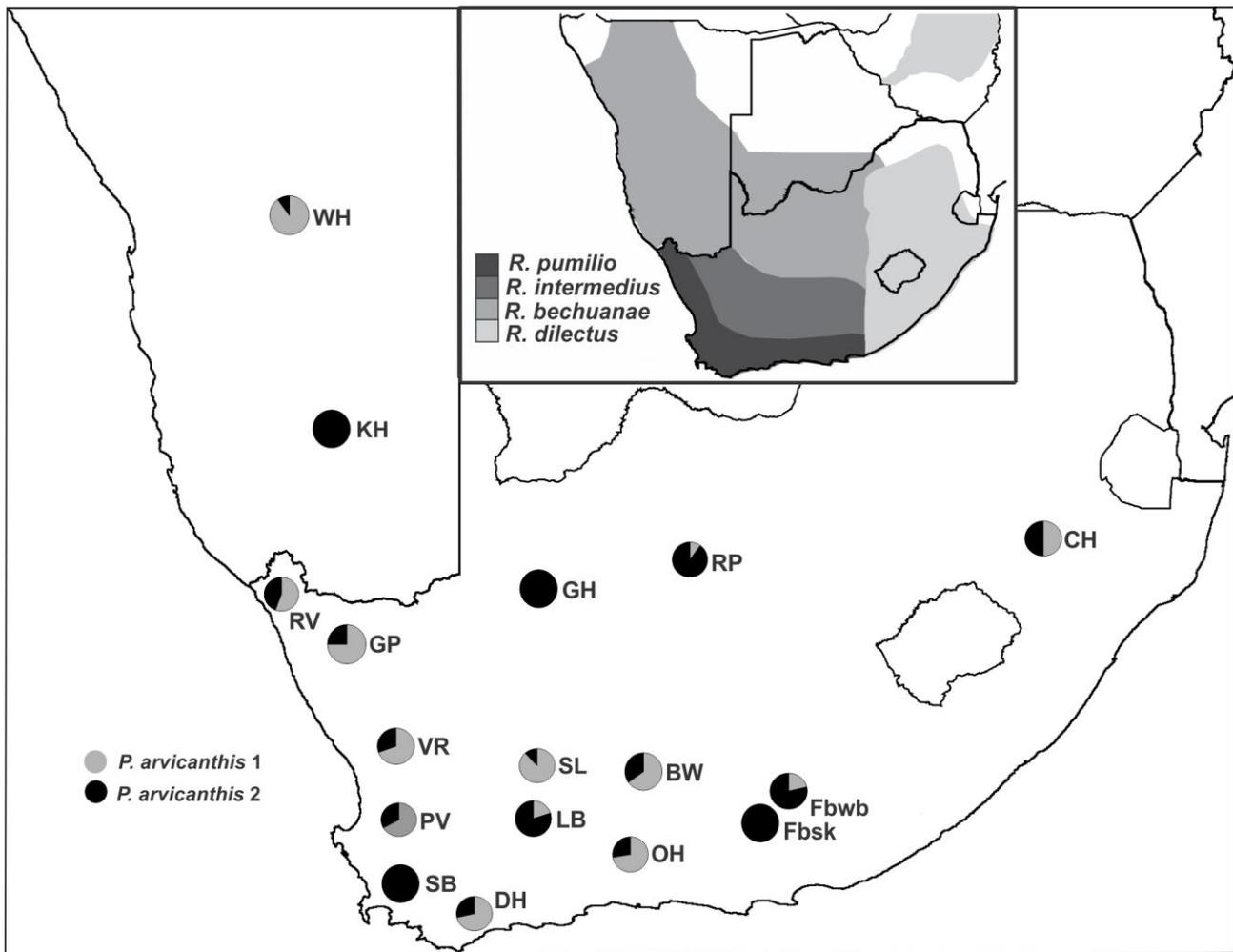
Figure 4.1: Localities from which parasite and host specimens were collected (codes as in Table 4.1), indicating the frequency of the two *Polyplax arvicanthis* lineages (see Chapter 3) at each locality. The distributions of the four putative *Rhabdomys* species (see Chapter 2) are indicated by the dotted lines and accompanying insert. The shaded map area indicates the potential extent of sympatry among the various host lineages.

2.2 Molecular techniques and data analysis

Total genomic DNA was extracted from host tissue and whole individual louse specimens, and PCR and sequencing of the various gene fragments were performed following standard protocols and published primers (Table B.14; see Chapter 2). Sequences were edited in BioEdit Sequence Alignment editor 7.0.5 (Hall 2005) and the alignments were trimmed to avoid missing data (Table 4.2). When compared to other *Polyplax* species, a 6 bp or 3 bp mtDNA insert was present within the *P. arvicanthis* 1 and 2 COI sequences, respectively. Translation into protein sequences using EMBOSStranseq (www.ebi.ac.uk/Tools/st/emboss_transeq) indicated that both variants represent

functional copies. The insertions were coded as presence/absence data in the phylogenetic analyses and not accounted for at the population level analyses (the two parasite lineages were analysed separately). For the nuclear genes, heterozygous positions were resolved in PHASE v2.1.1 (Stephens *et al.* 2001; Stephens & Scheet 2005). The algorithm ran for 100,000 generations with a thinning interval of 1, and 10,000 burn-in generations. Phases with a 0.9 probability or higher were considered resolved. Subsequent nuclear analyses were performed on all alleles (Table 4.2). Haplotypes were identified using Collapse 1.21 (Posada 2004). Standard molecular diversity indices were calculated in Arlequin v3.5.1.2 (Excoffier & Lischer 2010). Relative parasite and host mitochondrial effective population sizes were estimated with theta(π) in Arlequin v3.5.1.2 (Excoffier & Lischer 2010).

2.2.1 Genetic relationships

Statistical parsimony networks were constructed in TCS 1.21 (Clement *et al.* 2000) to depict relationships among the mitochondrial and nuclear haplotypes of *P. arvicanthis* 1, 2, and *Rhabdomys*, separately. The COI haplo-groups of all three datasets could not be connected within the 95% confidence interval and Bayesian phylogenetic trees were therefore constructed from the haplotype data to explore deeper evolutionary relationships. Parasite and host Bayesian COI topologies were constructed in MrBayes v3.2 (Ronquist *et al.* 2012) and best-fit models of sequence evolution, as determined in jModelTest v0.1.1 (Guindon & Gascuel 2003; Posada 2008) using the AICc (Akaike 1973; Burnham & Anderson 2004), were specified in each analysis (Table B.15). All analyses were partitioned by codon, with individual models assigned to each partition and parameters were unlinked across partitions. Each analysis consisted of two parallel Markov Chain Monte Carlo (MCMC) simulations of 5 chains each that were run for 5 million generations with a sampling frequency of 100 generations. Parameter convergence and ESS values were monitored in Tracer v1.5 (Rambaut & Drummond 2007), and 25% of the total number of generations were discarded as burn-in.

Table 4.2: The total number of sequences and alleles after resolving heterozygous positions (nDNA), haplotypes retrieved, total length (bp), polymorphic sites (P), nucleotide diversity (π), haplotype diversity (*h*), and estimated alpha shape parameter of the gamma rate variation distribution for the mitochondrial and nuclear datasets. Theta(π) estimates of mitochondrial effective population size is also indicated.

	Sequences/ Alleles	Total haplotypes	Singleton haplotypes	bp	P	π	*h*	α	Theta(π)
mtDNA									
Rhabdomys (COI)	377	97	56	900	211	0.06 ±0.03	0.96 ±0.003	0.14	59.48 ±28.00
P. arvicanthis 1 (COI)	151	42	29	270*	107	0.09 ±0.04	0.85 ±0.02	0.34	25.27 ±12.34
P. arvicanthis 2 (COI)	176	67	45	270*	114	0.11 ±0.05	0.92 ±0.01	0.34	31.99 ± 15.52
nDNA									
Rhabdomys (Eef1a1)	83/166	92	63	230	30	0.01 ±0.002	0.98 ±0.004	0.90	-
P. arvicanthis 1 (CAD)	46/92	36	22	348	40	0.01 ±0.005	0.93 ±0.01	0.17	-
P. arvicanthis 2 (CAD)	66/132	60	44	348	36	0.01 ±0.006	0.97 ±0.006	0.17	-

*Excluding the two 3bp insertions

Bayesian analysis of population structure (BAPS) was conducted on the mtDNA datasets in the software BAPS v5.3 (Corander *et al.* 2008), using a non-spatial mixture model for linked loci (Corander & Tang 2007). The codon linkage model and a vector of maximum K values (ranging from 1 to the maximum number of populations, each replicated by 5) were specified. For both mitochondrial and nuclear datasets, three-level hierarchical analyses of molecular variance (AMOVA; Excoffier *et al.* 1992) were conducted and Φ statistics were calculated in Arlequin v3.5.1.2 (Excoffier & Lischer 2010) to test for significant differentiation (1) among clusters identified by the mtDNA data, (2) among sampled localities within these clusters, and (3) within sampled localities. Gamma corrections estimated for each dataset in jModelTest were applied (Table 4.2) and significance evaluated with 1,000 permutations. Pairwise Φst values were calculated among localities and sequential Bonferroni corrections were applied within each dataset to account for multiple comparisons (Rice 1989).

2.2.2 Spatial mtDNA structure

Mantel tests (Mantel 1967) for matrix correspondence of genetic relatedness and geographical distance among individuals (Smouse *et al.* 1986; Smouse & Long 1992) were performed in GenAlEx v6.5 (Peakall & Smouse 2006; Peakall & Smouse 2012) to test for isolation-by-distance (IBD). Statistical significance was assessed with 10 000 random permutations. Partial Mantel tests (Legendre & Legendre 1998) were conducted in IBDWS v3.23 (Bohonak 2002; Jensen *et al.* 2005) using 10,000 randomizations to test whether significant differentiation was present among genetic clusters when correcting for IBD (Drummond & Hamilton 2007; Meirmans 2012).

In the presence of hierarchical structuring, IBD may potentially act only at certain distances (Drummond & Hamilton 2007). To further investigate the potential extent of IBD and spatial structure, spatial autocorrelation analyses were performed in GenAlEx v6.5 (Peakall & Smouse 2006; Peakall & Smouse 2012) at the individual level for the two parasite taxa and hosts. The genetic autocorrelation coefficient (r), was estimated for increasing distance size classes and statistical significance was evaluated by generating a 95% confidence interval around r values generated under the null hypothesis of no autocorrelation, using 1,000 random permutations (Smouse & Peakall 1999). Significant autocorrelation is inferred if the observed r value falls outside

this confidence interval. Confidence intervals around the observed r values were calculated with 1,000 bootstrap replicates (Peakall *et al.* 2003).

2.2.3 Co-phylogeny

Co-phylogeny between *Rhabdomys* and *Polyplax* was investigated using distance-based and topology-based approaches on a representative subset of the mtDNA data. Sixteen COI haplotypes representative of all the genetic clusters (parasites = 12 taxa; hosts = 4 taxa) were included. For simplification, the contact zones were excluded. Bayesian phylogenetic trees for the parasites (two taxa combined) and host pruned haplotype datasets were constructed using the Bayesian methods outlined above (Table B.15). The topological relationships retrieved from these pruned datasets were congruent with those obtained using all data. A distance-based co-phylogeny analysis was performed using AxParafit (Stamatakis *et al.* 2007), which is an optimized version of Parafit (Legendre *et al.* 2002) that is implemented in CopyCat v2.02 (Meier-Kolthoff *et al.* 2007). The hypothesis of co-phylogeny between parasite and host is tested through the comparison of parasite and host patristic distances (calculated from the branch length of phylogenetic trees) incorporating the parasite-host associations. Significance was assessed with 10,000 permutations, generated under the null hypothesis of independent parasite and host evolution. Topology-based reconciliation of parasite and host trees were performed in the software Jane v.4 (Conow *et al.* 2010), which is currently the only event-cost method that allows multi-host parasites (Althoff *et al.* 2012; Mendlová *et al.* 2012). A cost is assigned to each potential evolutionary event and the parasite phylogeny mapped onto that of the host using these different events, while attempting to find the solution with the minimum total cost (most parsimonious). The Vertex based cost model was implemented with the following cost scheme: co-speciation/co-divergence = 0, duplication = 1, duplication followed by a host-switch = 2, loss = 1, failure to diverge = 1. The genetic algorithm was set to 500 generations and a population size of 1,500. Statistical significance of the solutions was evaluated with permutations by random tip mating as well as randomization of the parasite tree topology, employing the same genetic algorithm parameter settings as above with a sample size of 1,000.

2.2.4 Evolutionary rates and divergence dates

*BEAST as implemented in BEAST v1.7.4 (Drummond *et al.* 2012) uses the multispecies coalescent to jointly estimate multiple gene trees and a shared species tree from multi-locus data using multiple individuals per species (Heled & Drummond 2010). This approach has been used previously to estimate the relative rates of evolution among hosts and parasites (Štefka *et al.* 2011) and was implemented here to estimate the COI evolutionary rates of the two *P. arvicanthis* lineages relative to that of the host. The two *P. arvicanthis* lineages were analysed separately and the datasets consisted of one representative parasite individual from each sampled population and the complementary host dataset. In each analysis, the rate for *Rhabdomys* was set to 1.0, allowing relative estimation of the parasite rates using an uncorrelated lognormal relaxed clock (Drummond *et al.* 2006). Best-fit models of sequence evolution for whole (unpartitioned) host and parasite datasets were estimated in jModelTest v0.1.1 (Guindon & Gascuel 2003; Posada 2008). Subsequently, the GTR+G and GTR+I models were specified for *Rhabdomys* and for both *P. arvicanthis* 1 and 2, respectively. All datasets were partitioned into 1+2nd and 3rd codon positions, the Yule process was implemented as species tree prior, and the piecewise linear and constant root population size model was used. The analyses ran for 500 million generations, sampling every 50, 000 generations. The output was evaluated in Tracer v1.5 (Rambaut & Drummond 2007) to ensure that all ESS values were greater than 200, ignoring the first 10% of generations as burn-in.

Since timing of divergence events are not incorporated into reconciliation analyses such as Jane (Reed *et al.* 2007; Light & Hafner 2008), inferred co-divergence events may not be truly contemporaneous. To test for contemporaneous parasite and host divergence, the parasite divergences associated with the inferred putative co-divergence events, were estimated in BEAST v1.7.4 (Drummond *et al.* 2012) using a Bayesian relaxed molecular clock (Drummond *et al.* 2006). The *P. arvicanthis* 1 and 2 datasets, consisting of all COI haplotypes, were analysed separately. The estimated relative rate of evolution for each parasite (see results) were converted to substitutions/site/year using the previously estimated rate for *Rhabdomys* (1.0 E-8; Chapter 2) and used to calibrate the respective analyses. The Yule species tree prior and GTR+I model was specified in both analyses, which were unpartitioned to allow estimation of an overall evolutionary rate. The analyses were run for 50 million generations, sampling every 5,000 generations and the output was evaluated employing the same procedure outlined above.

3. Results

3.1 Parasite prevalence and distribution

Polyplax arvicanthis prevalence at most localities was relatively high (mostly > 50%). The two *P. arvicanthis* lineages, however, were found to occur at different frequencies at the various geographic localities and less than 15% of the *Rhabdomys* individuals from which multiple louse specimens were sampled harboured both louse lineages (Table 4.1; Fig. 4.1).

3.2 Genetic relationships among sampling sites

Mitochondrial data was generated for 377 *Rhabdomys*, 151 *P. arvicanthis* 1, and 176 *P. arvicanthis* 2 specimens from 17 localities throughout southern Africa (see Tables B.12, B.13 for GenBank accessions). A total of 83 *Rhabdomys*, 46 *P. arvicanthis* 1, and 66 *P. arvicanthis* 2 nuclear sequences were generated in our attempts to sequence all the mtDNA haplotypes identified (Tables 4.2; see Tables B.12, B.13 for GenBank accessions). *Polyplax arvicanthis* 1 had fewer unique haplotypes and also a lower haplotypic diversity when compared to *P. arvicanthis* 2 for both mtDNA and nDNA datasets (Table 4.2). As expected for species characterized by higher mutation rates (Nieberding & Morand 2006), both parasite lineages had higher mtDNA nucleotide diversities and proportions of polymorphic sites than the host. Estimated theta(π) values generated for the mtDNA data indicated that both parasite taxa had smaller effective population sizes than the host (Table 4.2).

A remarkably high level of genetic divergence was detected for the parasite and host since several distinct, unconnected geographically structured mtDNA haplo-groups were retrieved from the statistical parsimony analyses (Fig. 4.2). The connection limit was 7 and 13 mutational steps for the parasites and hosts, respectively. The distinct genetic clusters retrieved from BAPS (Table B.16, Fig. 4.2) were largely congruent with the TCS haplo-groups found for both parasite and host taxa. *Rhabdomys* consists of five genetic groups (Table B.16, Fig. 4.2), corresponding to four putative species (Chapter 2; Fig. 4.1), with further subdivision of *R. pumilio* into subclades "A" and "B". Subclade "B" consists exclusively of individuals sampled in the contact zones (FB_sk and FB_wb; Fig. 4.1; Chapter 2) where mtDNA haplotypes belonging to *R. pumilio, R. intermedius* and *R. dilectus* were found sympatrically. *Polyplax arvicanthis* 1 and 2 contains 5 and 7 distinct genetic

groups, respectively (Table B.16, Fig. 4.2). No shared haplotypes were detected among the haplogroups of *P. arvicanthis* 1, while some mixture was present at Vanrhysdorp (VR) between haplogroup 1 and 2 of *P. arvicanthis* 2 (Table B.16). While some spatial congruence is evident among the parasite and host genetic groups, parasites collected from *R. pumilio*, *R. dilectus*, and *R. intermedius* within the contact zones (FB_sk and FB_wb) are not genetically differentiated (Fig. 4.3).

The haplo-groups detected by the statistical parsimony analysis and the Bayesian analyses of population structure were also retrieved from the Bayesian phylogenetic analyses (Fig. 4.2), with most haplo-groups forming monophyletic clades. The topologies also indicate the relationships among the haplo-groups, which were well resolved overall (Fig. 4.2). In most instances, haplo-groups in close geographic proximity also cluster as sister assemblages in the topologies (Fig. 4.2). All haplotypes in the nuclear DNA statistical parsimony networks could be connected within the 95% confidence interval (Figs. 4.4, 4.5). Within *Rhabdomys* the highest frequency of shared haplotypes occurs among *R. pumilio* and *R. intermedius*, which is also reflected by the parasites collected from these two hosts (Fig. 4.5). The close relationship between *P. arvicanthis* 2 clusters 3, 4, and 5 at the mtDNA level is reflected by the high incidence of shared haplotypes among these clusters in the nuclear CAD network (Figs. 4.2, 4.4).

There is strong evidence from our data that specimens collected in the northwestern arid region of southern Africa are distinct from the remainder of the individuals sampled within both parasite lineages (*P. arvicanthis* 1 cluster 1 and *P. arvicanthis* 2 cluster 3) and the host (*R. bechuanae*). A similar trend is evident for the mesic-adapted *R. dilectus* and the associated parasite clusters (*P. arvicanthis* 1 cluster 5 and *P. arvicanthis* 2 cluster 4).
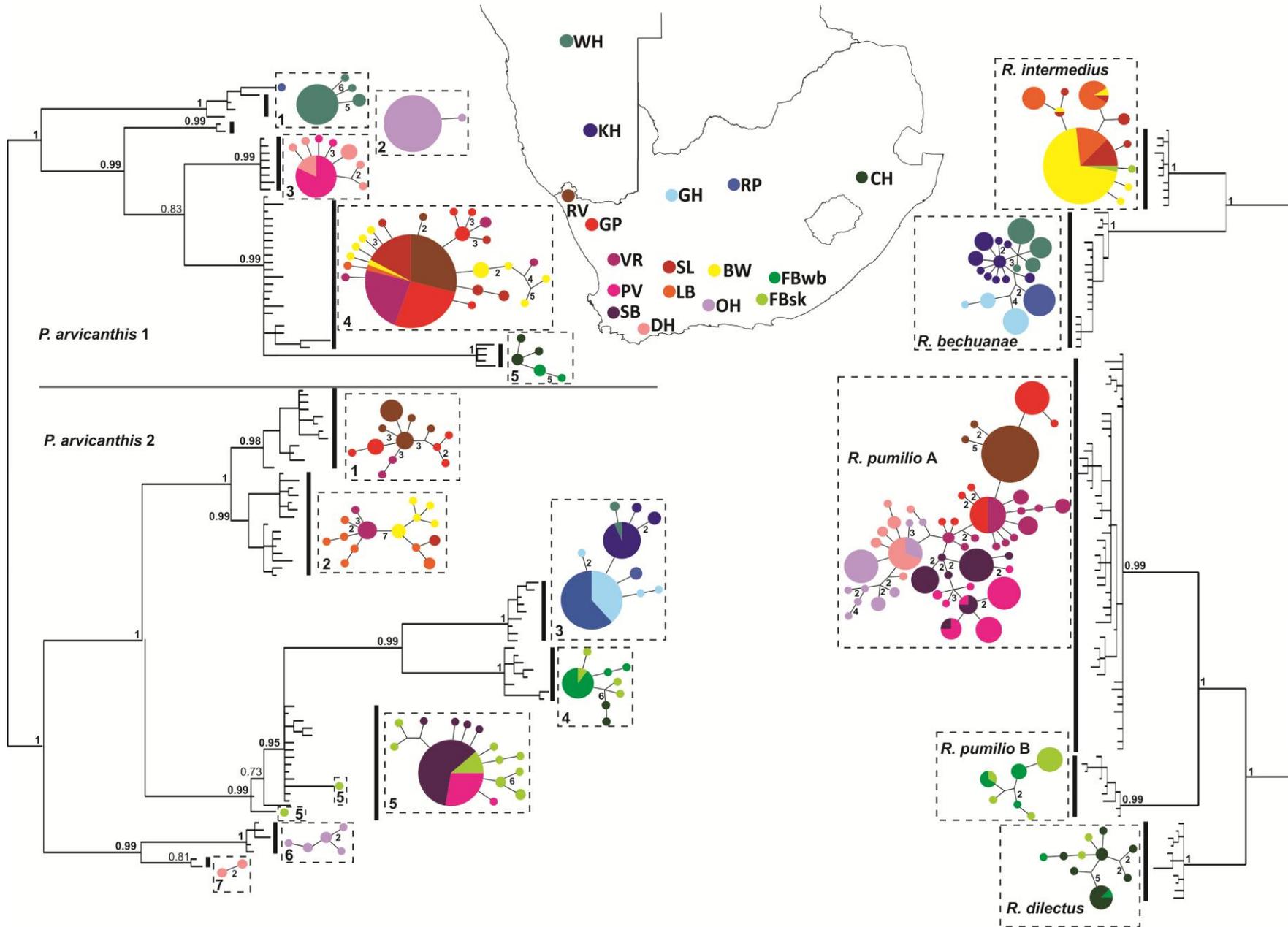
Figure 4.2 (above): Composite of the COI statistical parsimony haplotype networks for both parasite taxa (left) and the host (right) with accompanying Bayesian topologies indicating the relationships among haplogroups. Bayesian posterior probabilities (significant values in bold) are indicated on nodes. Each circle constitutes a particular haplotype with size indicating the relative number of individuals per haplotype. Colours indicate the frequency of each haplotype within the various sampled localities (insert) and each connection constitutes a single mutational step with numbers under lines indicating number of steps if more than one. Dashed-line boxes indicate the genetic clusters retrieved from BAPS, which for *Rhabdomys* coincides with the previously described species (see Chapter 2).
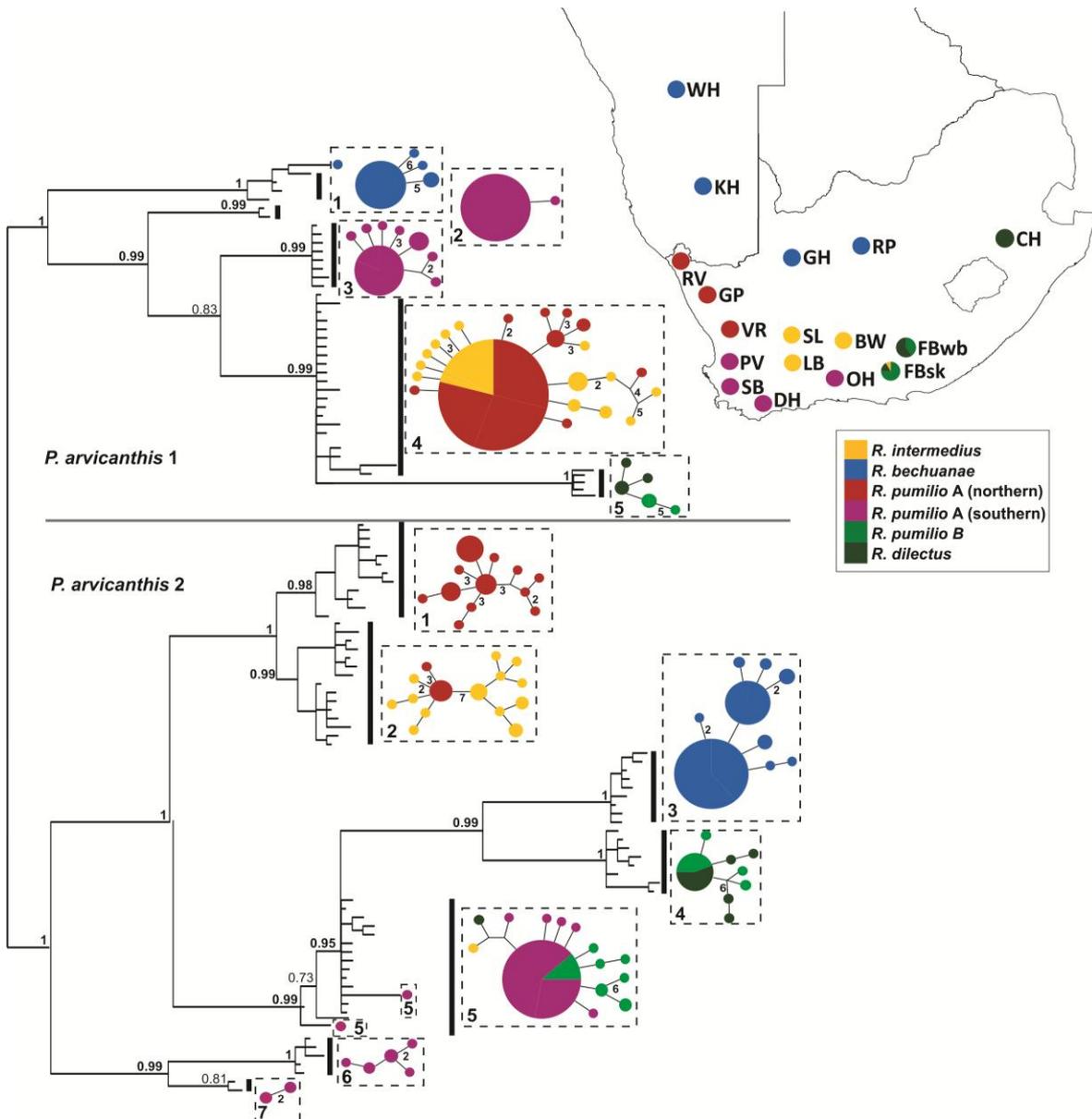
Figure 4.3: *Polyplax arvicanthis* mtDNA haplotype network depicting spatial congruence with the host genetic groups.
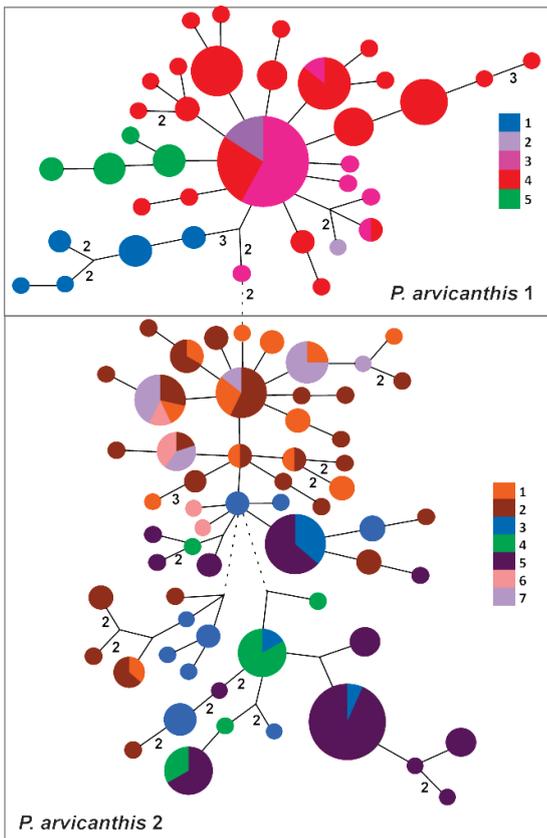
Fig. 4.4: Nuclear statistical parsimony networks (CAD gene) for *P. arvicanthis* 1 and 2 with haplotypes coloured according to mtDNA clusters as in Figs. 4.2 and 4.3. Dotted lines indicate single mutational steps and are used for ease of representation. Clusters with congruent population membership among the two lineages are indicated with corresponding colours.
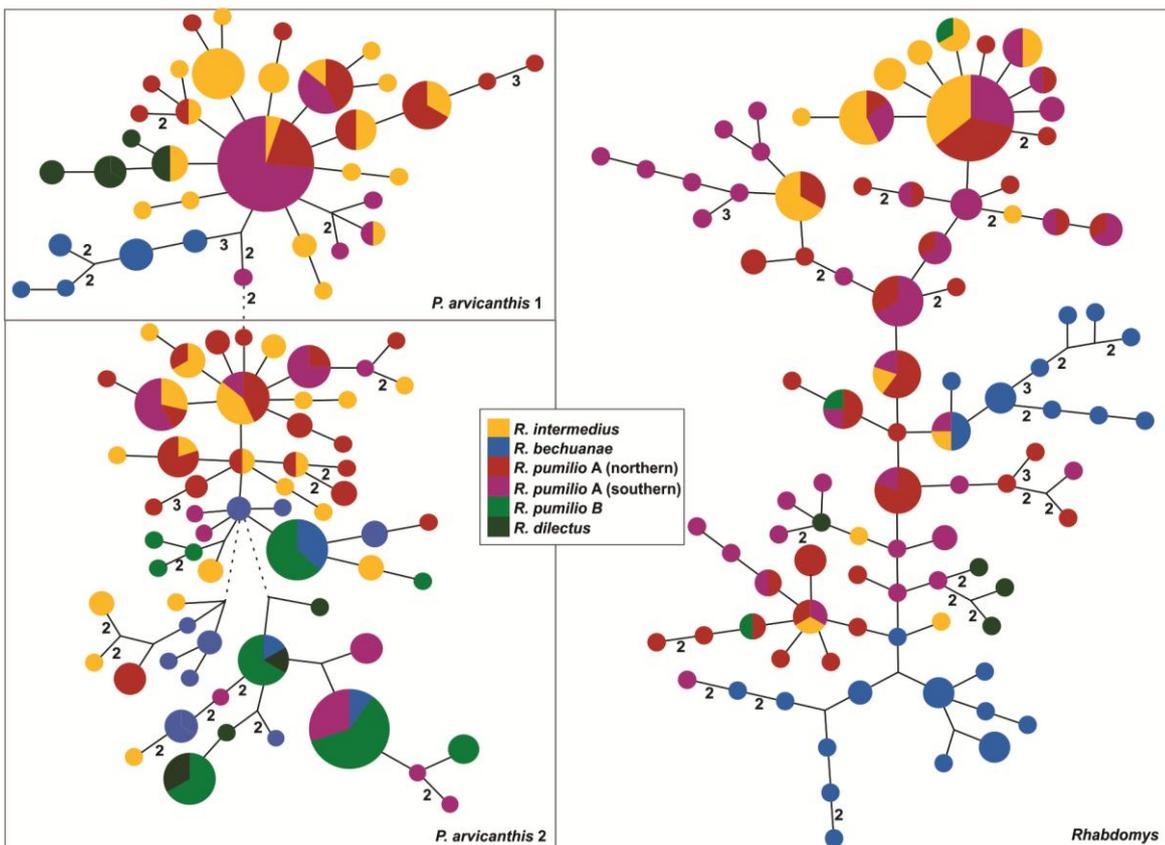
Figure 4.5 (above): Nuclear statistical parsimony networks for the parasite CAD (left) and host Eef1a1 (right) genes. Haplotypes are coloured according to host genetic groups (insert). Dotted lines indicate single mutational steps and are used for ease of representation. Clusters that have congruent population membership within the parasites and the host are indicated in the same colours.

Using the mtDNA haplo-groups as prior, the hierarchical analyses of molecular variance indicated that the genetic groups within both parasite taxa and the host are significantly differentiated at the mtDNA level (Table 4.3). All the respective mitochondrial genetic groups are also significantly differentiated at the nuclear level (stronger in the parasites), although the values suggest lower levels of differentiation for these data. For both parasites and the host, pairwise Φst values among geographic sampling sites indicated significant differentiation among nearly all sites at the mtDNA level (Tables B.17, B.18) and for several pairwise comparisons at the nuclear level (Tables B.17, B.19).

Table 4.3: Results from 3-level hierarchical analyses of molecular variance for the mitochondrial and nuclear datasets of the parasites and hosts.

| | | Fixation index | | | % Variation | | |
		ΦST	ΦCT	ΦSC	Among clusters	Among sites within clusters	Within sites
mtDNA	*Rhabdomys*	**0.97**	**0.94**	**0.60**	94.45	3.37	2.18
	P. arvicanthis 1	**0.96**	**0.96**	**0.19**	95.53	0.87	3.60
	P. arvicanthis 2	**0.90**	**0.87**	**0.31**	86.75	4.18	9.07
nDNA	*Rhabdomys*	**0.29**	**0.21**	**0.10**	20.78	8.50	70.72
	P. arvicanthis 1	**0.43**	**0.37**	**0.11**	36.87	6.95	56.18
	P. arvicanthis 2	**0.40**	**0.28**	**0.17**	27.87	12.56	59.57

*Significant values ($p<0.05$) are indicated in bold

3.3 Spatial mtDNA structure

The Mantel tests indicated significant IBD within *Rhabdomys* and *P. arvicanthis* 2 (Table 4.4). It has been shown that the presence of isolation by distance may lead to the overestimation of distinct genetic clusters by amongst others, BAPS (Frantz *et al.* 2009; Safner *et al.* 2011; Meirmans 2012). In our case this seems unlikely since the clusters identified by BAPS are congruent with the statistical parsimony haplo-groups, which is also supported by the Bayesian topologies.

Furthermore, the results of the partial Mantel tests indicated that the relationship between cluster membership and genetic distance remains significant when controlling for geographic distance (Table 4.4; also see Meirmans 2012).

Table 4.4: Mantel test and partial Mantel test results for the host and parasite mtDNA datasets. Correlation coefficients (r) and statistical significance (p) resulting from 10 000 permutations are indicated.

| | Mantel test[*1] | | Partial Mantel test[*2] | | | |
| | H_1: geography | | H_1: clusters | | H_2: clusters (geography) | |
	r	p	r	p	r	p
Rhabdomys	0.11	0.001	0.7	0.001	0.65	0.001
P. arvicanthis 1	0.02	0.169	0.68	0.0002	0.67	0.0003
P. arvicanthis 2	0.05	0.006	0.76	0.001	0.74	0.001

[*1] H_0: Geographic and genetic distance is independent
   H_1: Genetic distance increases with geographic distance
[*2] H_0: Genetic distance and cluster membership are independent
   H_1: Genetic distance is greater among clusters
   H_2: Genetic distance is greater among clusters, corrected for the effect of geography (distance)

To further evaluate the effects of geographic distance on the genetic signatures obtained in the present study, we examined and found significant positive spatial autocorrelation among sampling sites for both parasite lineages and the hosts combined, with r becoming non-significant at 1200 km (Fig. 4.6). This correlation was strongest at the shortest distances and declined with increasing distance, indicating that isolation by distance is strongest among neighbouring populations and that the positive relationship between genetic and geographic distance deteriorates with increasing distance among populations. When compared among the groups, positive spatial autocorrelation was much stronger in *P. arvicanthis* 2 relative to *P. arvicanthis* 1 and *Rhabdomys*, as indicated by higher r values at shortest distances with the magnitude of difference declining with increasing distance class size. *Polyplax arvicanthis* 1 had the lowest r values, which corresponds to the non-significant Mantel test for overall IBD (Table 4.4).
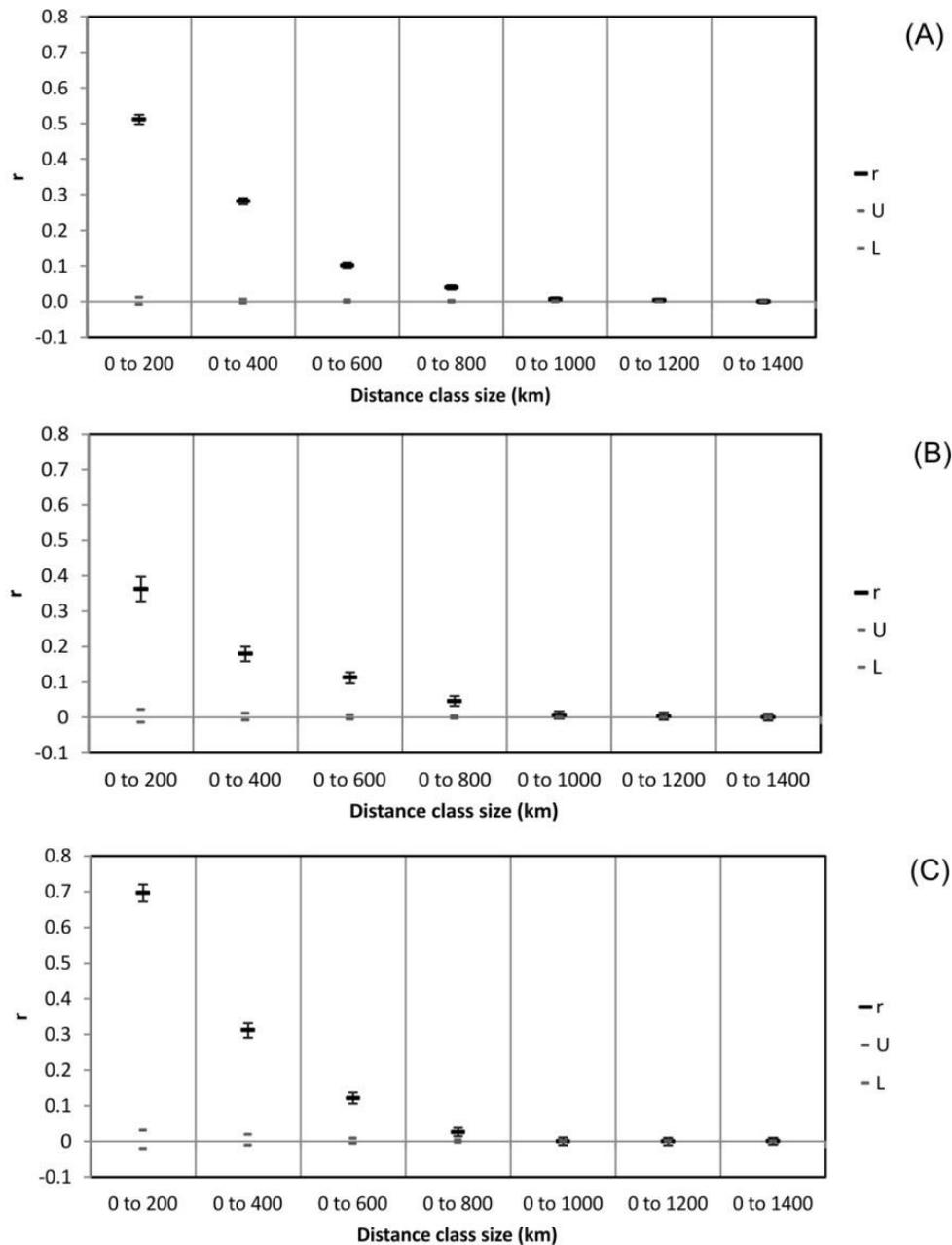
Figure 4.6 (above): Correlograms indicating the average genetic autocorrelation coefficient (r) as a function of increasing distance class size, for (A) *Rhabdomys*, (B) *Polyplax arvicanthis* 1, and (C) *Polyplax arvicanthis* 2. Error bars indicate the 95% confidence interval around the observed r values and grey dash marks indicate the 95% confidence interval surrounding the null hypothesis of no spatial autocorrelation (r = 0).

3.4 Co-phylogeny

Although there is some phylogeographic congruence among parasite and host genetic structures when the haplo-groups are compared, the Bayesian topologies indicate a fair amount of incongruence at the larger scale (Fig. 4.2). Furthermore, within the host contact zones (FB_sk and FB_wb; Fig. 4.2) parasites collected from *R. pumilio*, *R. dilectus*, and *R. intermedius* are not

genetically differentiated (Fig. 4.3). Thus, as expected the AxParafit global test indicated a non-significant association between parasite and host distances (p = 1). The Jane co-phylogeny reconstruction retrieved two solutions, each with a total cost of 19 representing 2 co-divergences, 4 duplications, 5 duplications followed by host-switches, 3 losses, and 2 failures to diverge (Fig. 4.7). The two solutions were identical except for the relative timing of the two co-divergence events. Both permutation procedures indicated non-significant co-phylogeny between parasite and host trees (random tip mating p = 0.92; random parasite tree p = 0.89), since the average total costs resulting from random solutions were lower than that of the observed reconstructions.
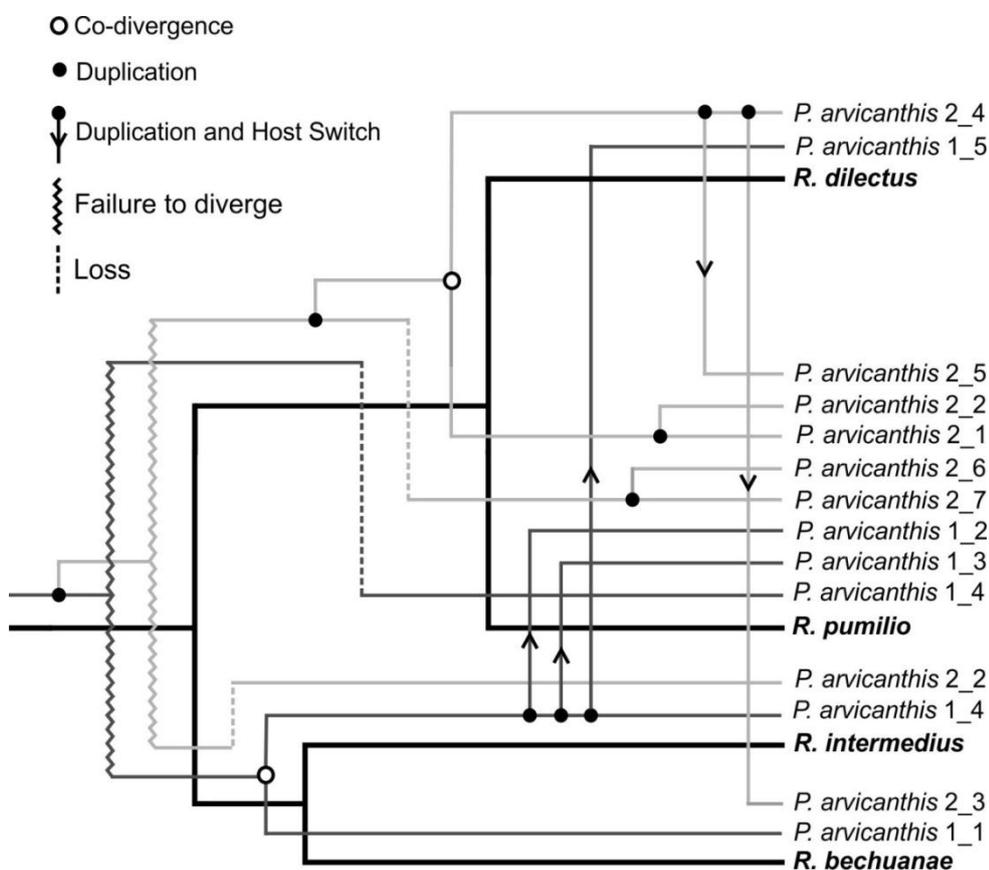
Figure 4.7: Reconciliation of parasite and host phylogenies retrieved from Jane employing the five types of evolutionary events (legend). Numbers following underscores indicate the respective *Polyplax arvicanthis* 1 and 2 clusters.

3.5 Evolutionary rates and divergence dates

The evolutionary rates of *P. arvicanthis* 1 and 2 were approximately 4 (HPD mean = 4.37 ± 0.02) and 5 (HPD mean = 5.18±0.02) times higher than that of *Rhabdomys*, respectively. Standard deviations of the uncorrelated lognormal relaxed clock were less than 1 for all parasite and host

datasets (*P. arvicanthis* 1: 0.67, *P. arvicanthis* 2: 0.62, *Rhabdomys*: 0.57), indicating only moderate deviation from clock-like behaviour (Drummond *et al.* 2006). The estimated parasite divergence dates for the two putative co-divergence events (indicated by Jane; see above) were 3.03 Ma (2.35 - 3.82) and 2.5 Ma (1.91 - 3.14) for *P. arvicanthis* 1 and 2, respectively (Fig. 4.8). These divergences occurred sometime after the two respective host divergences at 4.3 (4.00 - 4.88) and 3.09 (2.22 - 4.01) Ma (Chapter 2).
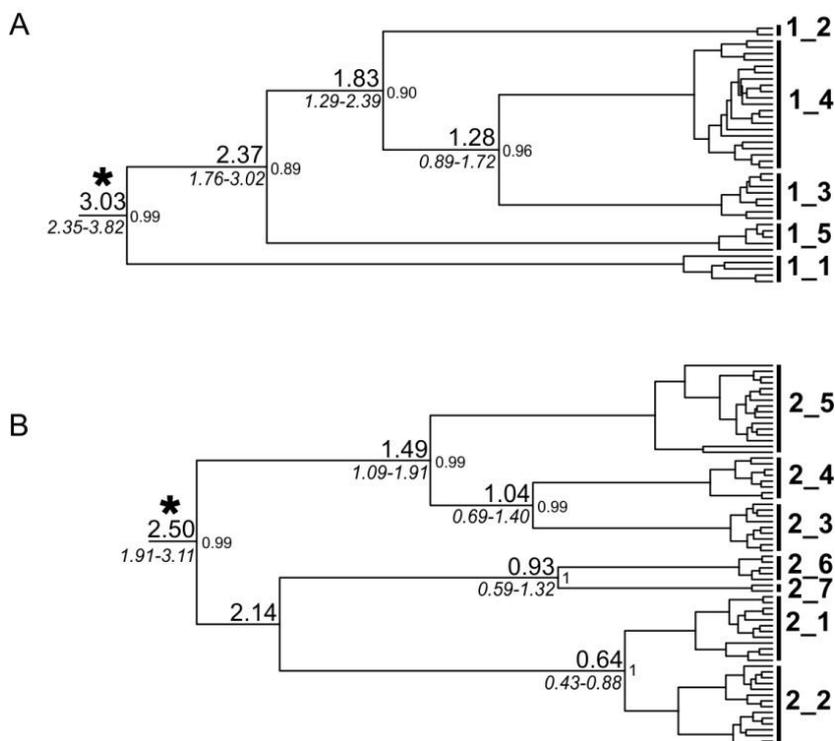
Figure 4.8: Maximum clade probability trees resulting from the rate-calibrated BEAST analyses for (A) *P. arvicanthis* 1 and (B) *P. arvicanthis* 2. Posterior mean divergence dates among the genetic groups (as in Fig. 4.2) are indicated above nodes in millions of years before present. Posterior probabilities and 95% HPD credibility intervals are indicated below nodes to the right and left, respectively. Parasite divergence dates associated with the putative co-divergence events identified by Jane (Fig. 4.7) are indicated by asterisks.

4. Discussion

Polyplax arvicanthis has high specificity to *Rhabdomys*, a direct life cycle (Ledger 1980, Kim 2006), and high prevalence (Table 4.1; Matthee *et al.* 2007; Matthee *et al.* 2010). These factors are predicted to act in concert to enhance vertical transmission and limit host-switching opportunities (Blouin *et al.* 1995; Nadler 1995; Jerome & Ford 2002a; Johnson *et al.* 2003; Johnson & Clayton

2004; Whiteman & Parker 2005; Wirth *et al.* 2005), which will also tend to increase the probability of parasites tracking host movements and thus promote congruence (Clayton *et al.* 2003; Nieberding & Morand 2006). In the present study, only limited congruence is present among the phylogeographic relationships of both *P. arvicanthis* lineages and the four *Rhabdomys* species (Figs. 4.2, 4.3) and this finding is contrary to the expectation that highly host-dependant parasites should show high levels of congruence with the genetic structures of their hosts (Blouin *et al.* 1995; Whiteman & Parker 2005; Nieberding & Morand 2006). This emphasizes the complexity of parasite-host associations and clearly suggests that factors such as biogeography and host life history can be equally or even more important in shaping their outcome (Huyse *et al.* 2005; Barrett *et al.* 2008; Hoberg & Brooks 2008).

Parasites are also expected to have more pronounced spatial structure than hosts due to reduced gene flow and increased genetic drift resulting from the combined effects of limited dispersal abilities, smaller effective population sizes, and elevated mutation rates (Criscione & Blouin 2005; Huyse *et al.* 2005; Nieberding & Morand 2006; Matthee *et al.* 2007). In our study, *Polyplax* has a smaller estimated effective population size than the host (Table 4.2) which is expected for parasites characterized by a female biased sex ratio and a direct life cycle (Rannala & Michalakis 2003; Criscione & Blouin 2005; Huyse *et al.* 2005; Matthee *et al.* 2007). Also, mitochondrial COI evolutionary rates of *P. arvicanthis* 1 and 2 were estimated at approximately 4 and 5 times higher when compared to that of *Rhabdomys*, which is in keeping with previous reported values for chewing and sucking lice and their hosts (Hafner *et al.* 1994; Page & Hafner 1996; Page *et al.* 1998; Paterson & Banks 2001; Page *et al.* 2004; Light & Hafner 2007, 2008). Faster evolutionary rates may be a product of the short generation times of lice which leads to a faster accumulation of mutations (Hafner *et al.* 1994), which is corroborated by the higher levels of sequence diversity observed for *Polyplax* when compared to *Rhabdomys* (Table 4.2). Despite these traits, the *P. arvicanthis* lineages do not fulfil the expectation of more pronounced spatial genetic structure, when compared to *Rhabdomys* (Nieberding & Morand 2006). Geographic genetic structure is actually less pronounced in both parasites when compared to *Rhabdomys*, as indicated by the high incidence of shared haplotypes among sampled localities (Fig. 4.2) and lower haplotypic diversity values (Table 4.2). These findings may potentially be explained by high parasite dispersal (Frankham *et al.* 2002). While the dispersal potential of *P. arvicanthis* is expected to be low due to its life history traits (Hopkins 1949; Kim 2006), the autocorrelation analyses indicated that the inferred levels of

historical gene flow of the parasites (particularly over shorter distances) are more extensive or only slightly more restricted compared to the host (Fig. 4.6).

In sucking lice, dispersal occurs predominantly through inter-host contact from adult to offspring and during social interactions, but may also occur via shared nests or burrows (Ledger 1980; Marshall 1981). Parasite dispersal is therefore dependent on host vagility (Criscione *et al.* 2005) and dispersal will be restricted in solitary hosts (Demastes *et al.* 2012) while movement will be enhanced by social contact (Huyse *et al.* 2005). Some evidence exists to show that the arid-adapted *R. pumilio* is social and group-living while the mesic-adapted *R. dilectus* is more solitary (Schradin & Pillay 2005). Although it has not been explicitly tested, it follows that *R. intermedius* which is also distributed in arid environments will probably also be social in nature. This is mainly based on a higher similarity in vegetation and climatic condition between regions occupied by *R. pumilio* and *R. intermedius* (Mucina & Rutherford 2006). Host social interactions and shared nests are expected to facilitate louse gene flow and it is interesting to note that most haplotype sharing for both *P. arvicanthis* lineages occur among the lice of the social *R. intermedius* and *R. pumilio* (Figs. 4.2, 4.3, 4.4). While the females of *R. pumilio* tend to be philopatric, males may stay in their natal group as helpers or disperse, to either become the territorial breeding male of another group or solitary-living roamers (Solmsen *et al.* 2011). Male-biased infestation of *P. arvicanthis*, which is possibly the result of the depression of male immunity by testosterone, has been reported for *R. pumilio* (Matthee *et al.* 2010) and solitary roamers have been shown to have highest testosterone levels (Schradin *et al.* 2009). Therefore, these roaming males are ideal candidates to facilitate the high levels of *Polyplax* dispersal across the geographic landscape. If these arguments hold, our findings strongly support the premise that host life history characteristics (social behaviour and mobility), are equally as important, if not more than parasite characteristics in shaping parasite population structure and determining the extent of congruence (Huyse *et al.* 2005; Barrett *et al.* 2008).

At the macroevolutionary level, the results obtained from Parafit and Jane indicated limited co-divergence between the four host species and the haplo-groups detected for *P. arvicanthis* 1 and 2 (Fig. 4.7). The Jane reconciliation suggests that the most parsimonious explanation for the existence of the two parasite species on all four *Rhabdomys* species is a parasite duplication event prior to host divergence, which would imply sympatric parasite speciation (Brooks & McLennan 1993;

Huyse *et al.* 2005). During the first host divergence event both parasite lineages failed to diverge, which could possibly be ascribed to intermittent host contact allowing for parasite gene flow among diverging hosts (Banks & Paterson 2005). Subsequently, the evolution of *Polyplax* was characterized by two co-divergences and several duplication, duplication with host-switching, and parasite loss events (Fig. 4.6). Co-divergence events of the two parasite lineages occurred on allopatric hosts (*P. arvicanthis* 1 co-diverged with *R. intermedius* and *R. bechuanae* while *P. arvicanthis* 2 co-diverged with *R. pumilio* and *R. dilectus*). Certain parasitic lineages co-diverging or persisting during host divergence events and others being lost is probably a product of stochasticity and the relative abundance of parasite lineages on the particular host lineage since parasites with higher abundance are more likely to track host divergences (Clayton *et al.* 2003; Clayton *et al.* 2004; Nieberding & Morand 2006). Following the co-divergences, several duplication and host-switching events took place within *P. arvicanthis* 1 and 2. Interestingly, host-switching events only occurred among host taxa that are currently in contact with each other, indicating that this host sympatry was probably also present previously. Switching events among *R. pumilio*, *R. dilectus* and *R. intermedius* could have been facilitated by the contact zones at Fort Beaufort, especially as our data indicates a lack of parasite genetic differentiation among host taxa (FB_sk and FB_wb; Fig. 4.3). A contact zone in the Free State Province (Chapter 2; Ganem *et al.* 2012) probably created opportunities for host-switching among *R. bechuanae* and *R. dilectus*.

The limited genealogical congruence among *P. arvicanthis* and *Rhabdomys* at both micro- and macroevolutionary scales is probably also a direct result of the strong influence of biogeographic history (Chapter 2), coupled to incomplete lineage sorting as a result of recent cladogenesis (Rannala & Michalakis 2003; Nieberding & Olivieri 2007; Demastes *et al.* 2012). Episodes of environmental change have been suggested as the main drivers for diversification in parasite-host systems by inducing cyclical episodes of expansion and contraction in geographical ranges ("Taxon pulse hypothesis"; Halas *et al.* 2005) and such biogeographical change can thus facilitate co-divergence during periods of isolation and host-switching during periods of expansion (Weckstein 2004; Brooks & Ferrao 2005).Importantly, a shared history may produce congruent parasite and host structures in the absence of contemporaneous divergence (Page 2003; Hoberg & Brooks 2008) and inferred co-divergence events may thus not be truly contemporaneous (Reed *et al.* 2007; Light & Hafner 2008). In our case the dated parasite divergences are roughly contemporaneous with the associated host divergences (Fig. 4.8), although parasite divergences lag slightly behind host

divergences. *Rhabdomys bechuanae* and *R. intermedius* diverged approximately 4.3 Ma (4.00 - 4.88) ago with the associated divergence within *P. arvicanthis* 1 at 3.03 Ma (2.35 - 3.82), while *R. pumilio* and *R. dilectus* diverged around 3.09 Ma (2.22 - 4.01) followed by divergence of *P. arvicanthis* 2 around 2.5 Ma (1.91 - 3.14). It is expected that divergence of a parasite will lag behind that of its host if gene flow continues among parasites after hosts divergence (Light & Hafner 2008), a scenario which likely occurred within the current system.

Cladogenesis within *Rhabdomys* occurred as a result of biogeographic changes in response to climatic and landscape changes that occurred within southern Africa around the Mio-Pliocene boundary and led to the establishment of the biomes (Scott *et al.* 1997; Chapter 2). Interestingly, in the arid north-western region and mesic eastern region distinct monophyletic genetic assemblages within both parasite lineages and the host is present at the mtDNA level (see *P. arvicanthis* 1 cluster 1, *P. arvicanthis* 2 cluster 3, and *R. bechuanae* as well as *P. arvicanthis* 1 cluster 5, *P. arvicanthis* 2 cluster 4, and *R. dilectus*; Figs. 4.2, 4.3), and this is also partly supported by the nuclear data (Fig. 4.4). Similar genetic disjunctions for non-related species in the region (Lamb & Bauer 2000; Matthee & Flemming 2002; Redman & Hamer 2003; Tolley *et al.* 2004; Bauer *et al.* 2006; Smit *et al.* 2007; Tolley *et al.* 2010; Engelbrecht *et al.* 2011; Willows-Munro & Matthee 2011; Montgelard & Matthee 2012) support the premise that the pattern observed in both *P. arvicanthis* lineages is the result of vicariance and thus emphasises a strong influence of biogeography. The differentiation between the parasite assemblages, however, breaks down in contact zones at Fort Beaufort (FB_sk and FB_wb; Fig. 4.3) where lice parasitizing *R. pumilio* and *R. dilectus* share haplotypes. There is also another host contact zone in the Free State among *R. dilectus* and *R. bechuanae* (Chapter 2; Ganem *et al.* 2012) which could facilitate gene flow especially within *P. arvicanthis* 2, for which the lice from these hosts are sister taxa and indicate a moderate amount of gene flow at the nuclear level (Fig. 4.4). Reproductive isolation due to both pre- and post-zygotic barriers have been demonstrated between *R. pumilio* and *R. dilectus* as well as the two subspecies within the latter taxon (*R. d. dilectus* and *R. d. chakae*) (Pillay 2000; Lancaster 2001; Pillay *et al.* 2006; Stippel 2009). It is thus unlikely that that the various host taxa will hybridize extensively within contact zones and phylogenetic analyses indicate very narrow contact zones where biomes meet (Chapter 2).

It therefore seems reasonable to suggest that cladogenesis and ecological divergence within *Rhabdomys* occurred in response to biogeographic changes as a result of climatic oscillations in the region which produced cyclic host range contractions and expansions (Zachos 2001). The pattern of genetic divergence within *P. arvicanthis* is most likely the direct result of these cycles which would have facilitated parasite divergence during host allopatry and host-switching during periods of host sympatry (Clayton *et al.* 2004). The latter is supported by the presence of both parasite lineages on all four *Rhabdomys* species and it is evident that despite their presumable permanency and vertical mode of transmission, lice frequently disperse among host taxa in the zones of contact. Further, we hypothesise that it is indeed possible that more extensive geographic sampling within the distribution of *R. dilectus* and *R. bechuanae* may reveal breakdown in the reciprocal monophyly of the parasite mtDNA haplo-groups associated with these species.

The absence of genetic differentiation of parasites according to host lineages within contact zones indicates that *P. arvicanthis* is not adapted to specific hosts, and host-switching frequently occurs where the opportunity arises. The previously documented host-specificity for this taxon is thus merely an index describing the current perceived association between *P. arvicanthis* and *Rhabdomys* (Ledger 1980; Clayton *et al.* 2004). The evolutionary and biogeographic history of a parasite and host determines the range of hosts that can be exploited by the parasite (Poulin & Keeney 2007) and it is expected that the horizontal transmission of the parasite among diverging host lineages will prohibit the development of host-specificity (Nieberding & Olivieri 2007). Intermittent contact among *Rhabdomys* lineages would therefore have prevented the adaptation of *P. arvicanthis* to specific host lineages.

From the current study it is thus evident that biogeographic history coupled to host-related factors, such as vagility and sociality, have important consequences for the genetic outcome of parasite-host associations and can lead to limited congruence despite the presence of parasite traits that are expected to favour congruence.

Summary

The current study explored the evolutionary interactions of the four-striped mouse genus, *Rhabdomys* and its specific ectoparasitic sucking louse, *Polyplax arvicanthis* across southern Africa, employing both phylogenetic and phylogeographic techniques.

The morphology-based taxonomy of *Rhabdomys* is complicated by phenotypic plasticity across the wide distribution within the southern African subregion. Previously, the recognition of two species with distinct distributions (mesic-adapted *R. dilectus* and arid-adapted *R. pumilio*) was proposed based on genetic divergence, chromosomal differences, as well as ecological and behavioural differences. In Chapter 2, we focussed on the spatial genetic structure of the arid-adapted *R. pumilio* (521 specimens from 31 localities) with limited sampling of the mesic-adapted *R. dilectus* (33 specimens from 10 localities) across seven biomes to reappraise the variation within the genus. The mitochondrial COI gene and four nuclear introns (Eef1a1, MGF, SPTBN1, Bfib7) were used for the construction of gene trees. Bayesian and parasimony phylogenetic analyses of the mitochondrial DNA data indicated that *Rhabdomys* consists of four reciprocally monophyletic, geographically structured clades, with three distinct lineages present within the arid-adapted *R. pumilio*. These monophyletic lineages differ by at least 7.9% ($\pm$0.3) GTR-corrected sequence distance and these results are partly confirmed by a multilocus network of the combined nuclear intron dataset. Ecological niche modelling in MaxEnt supports a strong correlation between regional biomes and the distribution of distinct evolutionary lineages of *Rhabdomys*. A Bayesian relaxed molecular clock suggests that the geographic clades diverged between 3.09–4.30 Ma, supporting the hypothesis that the radiation within the genus coincides with paleoclimatic changes (and the establishment of the biomes) characterizing the Miocene-Pliocene boundary. Marked genetic divergence at the mitochondrial DNA level, coupled with strong nuclear and mtDNA signals of non-monophyly of *R. pumilio*, support the notion that a taxonomic revision of the genus is needed.

The molecular systematics of *P. arvicanthis* was investigated in Chapter 3. Representatives of *P. arvicanthis* were sampled from *Rhabdomys* at 16 localities throughout southern Africa. Parsimony and Bayesian gene trees were constructed for the mitochondrial COI, 12S rRNA, 16S rRNA and nuclear CAD genes. Our findings support the existence of two genetic groups within *P. arvicanthis* separated by at least 25% COI GTR-corrected sequence divergence, which is comparable to that observed among recognized *Polyplax* species. We therefore propose that these two genetic lineages probably represent distinct species and that the apparent absence of clear morphological differences

may point to cryptic speciation. The two taxa have sympatric distributions throughout most of the sampled host range and also occasionally occur sympatrically on the same host individual.

In Chapter 4 we investigated the comparative phylogeography and phylogeny of the two *P. arvicanthis* lineages and *Rhabdomys*, utilizing mitochondrial and nuclear sequence data. Parasite (lice, n = 327) and host (n = 377) specimens were collected from 17 localities across the southern African distribution of the four *Rhabdomys* species. Despite the smaller effective population sizes and elevated mutational rates found for *P. arvicanthis*, spatial genetic structure was not more pronounced in the parasite lineages compared to the hosts. This is probably the result of the high vagility and sociality of the hosts promoting parasite dispersal during inter-host contact. Limited phylogeographic congruence and a co-phylogenetic history characterised by several duplications, duplications with host-switching, and extinctions with only two roughly contemporaneous co-divergences were retrieved. These findings are contrary to expectations, since *P. arvicanthis* is highly host-dependant and possesses several traits predicted to favour congruence with host structures at both micro- and macroevolutionary scales. The patterns observed are likely the result of a complex biogeographic history, which involved cyclic range contractions and expansions of host ranges in response to climatic oscillations. This process probably facilitated parasite divergence during host allopatry and subsequent host-switching during periods of host sympatry. Intermittent contact among *Rhabdomys* lineages would also have prevented adaptation of *P. arvicanthis* to specific host lineages, thus explaining the current lack of host-specificity in areas of sympatry.

The outcomes of the present investigation reiterate the premise that parasite-host associations represent a complex mosaic shaped by the effects of parasite traits, host factors, and biogeography over evolutionary time.

References

Akaike H (1973) Information theory and an extention of the maximum likelihood principle. In: *Second International Symposium on Information Theory* eds. Petrov PN, Csaki F), pp. 267-281. Adad. Kiado, Budapest.

Althoff DM, Segraves KA, Smith CI, Leebens-Mack J, Pellmyr O (2012) Geographic isolation trumps coevolution as a driver of yucca and yucca moth diversification. *Molecular Phylogenetics and Evolution* **62**, 898-906.

Althoff DM, Thompson JN (1999) Comparative geographic structures of two parasitoid-host interactions. *Evolution* **53**, 818-825.

Anderson RP, Lew D, Peterson AT (2003) Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling* **62**, 211-232.

Anderson S, Jones JK (1984) *Orders and families of recent mammals of the World* John Wiley, New York.

Andrews RH, Monis PT, Ey PL, Mayrhofer G (1998) Comparison of the levels of intraspecific genetic variation within *Giardia muris* and *Giardia intestinalis*. *International Journal for Parasitology* **28**, 1179-1185.

Avise J (1994) Molecular Markers and Evolution.Chapman & Hall.

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology* **13**, 729-743.

Bamford MK (2000) Cenozoic macro-plants. In: *Cenozoic of Southern Africa* (eds. Partridge TC, Maud RR), pp. 351-356. Oxford University Press, New York.

Banks JC, Paterson AM (2005) Multi-host parasite species in cophylogenetic studies. *International Journal for Parasitology* **35**, 741-746.

Barrett LG, Thrall PH, Burdon JJ, Linde CC (2008) Life history determines genetic structure and evolutionary potential of host-parasite interactions. *Trends in Ecology and Evolution* **23**, 678-685.

Bauer AM (1999) Evolutionary scenarios in the Pachydactylus Group geckos of southern Africa: new hypotheses. *African Journal of Herpetology* **48**, 53-62.

Bauer AM, Lamb T (2005) Phylogenetic relationships of southern African geckos in the Pachydactylus group (Squamata:Gekkonidae). *African Journal of Herpetology* **54**, 105-129.

Bauer AM, Lamb T, Branch WR (2006) A revision of the *Pachydactylus serval* and *P. weberi* groups (Reptilia: Gekkota: Gekkonidae) of Southern Africa, with the description of eight new species. *Procceedings of the California Academy of Sciences* **57**, 595-709.

Baumann L, Baumann P (2005) Cospeciation between the primary endosymbionts of mealybugs and their hosts. *Current Microbiology* **50**, 84-87.

Bensch S, Stjerman M, Hasselquist D*, et al.* (2000) Host specificity in avian blood parasites: a study of *Plasmodium* and *Hemoproteus* mitochondrial DNA amplified from birds. *Proceedings of the Royal Society of London B Series* **267**, 1583-1589.

Benton MJ, Donoghue PCJ (2007) Paleontological Evidence to Date the Tree of Life. *Molecular Biology and Evolution* **24**, 26-53.

Bininda-Emonds ORP (2007) Fast genes and slow clades: Comparative rates of molecular evolution in mammals. *Evolutionary Bioinformatics* **3**, 59-85.

Bininda-Emonds ORP, Cardillo M, Jones KE, *et al*. (2007) The delayed rise of present-day mammals. *Nature* **446**, 507-512.

Blouin MS, Yowell CA, Courtney CH, Dame JB (1995) Host movement and the genetic structure of populations of parasitic nematodes. *Genetics* **141**, 1007-1014.

Bohonak AJ (2002) IBD (isolation by distance): a program of analyses of isolation by distance. *Journal of Heredity* **93**, 153-154.

Bonnefille R, Potts R, Chalie F, Jolly D, Peyron O (2004) High resolution vegetation and climate change associated with Pliocene *Australopithecus afarensis*. *Proceedings of the National Academy of Sciences* **101**, 12125-12129.

Braconnot P, Harrison S, Joussaume S, Peterchmitt J, Crucifix M (2007) Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum – Part 1 : experiments and large-scale features. *Climate Of The Past* **3**, 261-277.

Brooks DR (1979) Testing the context and extent of host-parasite coevolution. *Systematic Zoology* **28**, 299-307.

Brooks DR, Ferrao AL (2005) The historical biogeography of co-evolution: emerging infectious diseases are evolutionary accidents waiting to happen. *Journal of Biogeography* **32**, 1291-1299.

Brooks DR, McLennan DA (1991) *Phylogeny, ecology, and behaviour.* University of Chicago Press, Chicago.

Brooks DR, McLennan DA (1993) *Parascript. Parasites and the language of evolution* Smithsonian Institution Press, Washington, D.C.

Brown JM, Lebens-Mack JH, Thompson JN, Pellmyr O, Harrison RG (1997) Phylogeography and host association in a pollinating seed parasite *Greya politella* (Lepidoptera: Prodoxidae). *Molecular Ecology* **6**, 215-224.

Bryant D, Moulton V (2004) NeighborNet : An agglomerative method for the construction of planar phylogenetic networks. *Molecular Biology and Evolution* **21**, 255-265.

Burban C, Petit RJ (2003) Phylogeography of maratime pine inferred with organelle markers having contrasted inheritance. *Molecular Ecology* **12**, 1487-1495.

Burban C, Petit RJ, Carcreff E, Jactel H (1999) Rangewide variation of the maratime pine bast scale Matsucocus feytaudi Duc. (Homoptera: Matsucoccidae) in relation to the genetic structure of its host. *Molecular Ecology* **8**, 1593-1602.

Burnham KP, Anderson DA (2002) *Model selection and multimodel inference: A practical information-theoretic approach* Springer-Verlag, New York.

Burnham KP, Anderson DA (2004) Multi-model inference: understanding AIC and BIC in model selection. *Sociological Methods and Research* **33**, 261-304.

Castiglia R, Solano E, Makundi RH, *et al.* (2011) Rapid chromosomal evolution in the mesic four-striped grass rat *Rhabdomys dilectus* (Rodentia, Muridae) revealed by mtDNA phylogeographic analysis. *Journal of Zoological Systematics and Evolutionary Research* **50**, 165-172.

Cerling TE, Harris JM, Macfadden BJ, *et al.* (1997) Global vegetation change through the Miocene / Pliocene boundary. *Nature* **389**, 153-158.

Charleston MA, Perkins SL (2006) Traversing the tangle: algorithms and applications for cophylogenetic studies. *Journal of Biomedical information* **39**, 62-71.

Charleston MA, Robertson DL (2002) Preferential host-switching by primate lentiviruses can account for phylogenetic similarity with the primate phylogeny. *Systematic Biology* **51**, 528-535.

Chase BM, Meadows ME (2007) Late Quaternary dynamics of southern Africa's winter rainfall zone. *Earth-Science Reviews* **84**, 103-138.

Chimimba CT (2001) Infraspecifc morphometric variation in *Aethomys namaquensis* (Rodentia : Muridae) from southern Africa. *Journal of Zoology* **253**, 191-210.

Clayton DH (1990) Host specificity of *Strigiphilus* owl lice (Ischnocera: Philopteridae), with the description of new species and host associations. *Journal of Medical Entomology* **27**, 257-265.

Clayton DH, Al-Tamimi S, Johnson KP (2003) The ecological basis of coevolutionary history. In: *Tangled trees: Phylogeny, cospeciation, and coevolution* (ed. Page RDM), pp. 330-350. University of Chicago Press, Chicago.

Clayton DH, Bush SE, Johnson KP (2004) Ecology of congruence: Past meets present. *Systematic Biology* **53**, 165-173.

Clayton DH, Johnson KP (2003) Linking coevolutionary history to ecological process: doves and lice. *Evolution* **57**, 2335-2341.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**, 1657-1659.

Coetzee JA (1978) Climate and biological changes in south-western Africa during the Late Cainozoic. *Paleoecologica Africana* **10**, 13-29.

Conow C, Fielder D, Ovadia Y, Libeskind-Hadas R (2010) Jane: a new tool for the cophylogeny reconstruction problem. *Algorithms for Molecular Biology* **5**: 16.

Corander J, Marttinen P, Sirén J, Tang J (2008) Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* **9**: 539.

Corander J, Tang J (2007) Bayesian analysis of population structure based on linked molecular information. *Mathematical Biosciences* **205**, 19-31.

Criscione CD, Blouin MS (2004) Life cycles shape parasite evolution: comparative population genetics of salmon trematodes. *Evolution* **58**, 198-202.

Criscione CD, Blouin MS (2005) Effective sizes of macroparasite populations: a conceptual model. *Trends in Parasitology* **21**, 212-217.

Criscione CD, Poulin R, Blouin MS (2005) Molecular ecology of parasites: elucidating ecological and microevolutionary processes. *Molecular Ecology* **12**, 2247-2257.

Daniels SR, Gouws G, Crandall KA (2006) Phylogeographic patterning in a freshwater crab species (Decapoda: Potamonautidae: Potamonautes) reveals the signature of historical climatic oscillations. *Journal of Biogeography* **33**, 1538-1549.

de Graaff GD (1981) *The Rodents of southern Africa: Notes on their identification, distribution, ecology and taxonomy* Butterworths, Pretoria.

de León GPP, Nadler SA (2010) What we don't recognize can hurt us: A plea for awareness about cryptic species. *Journal of Parasitology* **96**, 453-464.

de Meeûs T (2000) Adaptive diversity, specialization, habitat preference and parasites. In: *Evolutionary Biology of Host-Parasite Relationships: Theory Meets Reality* (eds. Poulin R, Morand S, Skorping A), pp. 27-42. Elsevier.

Demastes JW, Spradling TA, Hafner MS, *et al.* (2012) Cophylogeny on a fine scale: *Geomydoecus* chewing lice and their pocket gopher hosts, *Pappogeomys bulleri*. *Journal of Parasitology* **98**, 262-270.

Denys C (1999) Of mice and men: evolution in East and South Africa during the Plio-Pleistocene times (eds. Bromage TG, Schrenk F), pp. 226-252. Oxford University Press.

Diester-Haass L, Meyers PA, Vidal L (2002) The late Miocene onset of high productivity in the Benguela Current upwelling system as part of a global pattern. *Marine Geology* **180**, 87-103.

Donoghue PCJ, Benton MJ (2007) Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends in Ecology & Evolution* **22**, 424-431.

Drummond AJ, Ho S, Phillips M, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLOS Biology* **4**: e88.

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

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUTi and the BEAST 1.7. *Molecular Biology and Evolution*, In press.

Drummond CS, Hamilton MB (2007) Hierarchical components of genetic variation at a species boundary: population structure in two sympatric varieties of *Lupinus microcarpus* (Leguminosae). *Molecular Ecology* **16**, 753-769.

Dupont LM (2006) Late Pliocene vegetation and climate in Namibia (southern Africa) derived from palynology of ODP Site 1082. *Geochemistry, Geophysics, Goesystems* **7**, 1-15.

Durden LA (1990) Phoretic relationships between sucking lice (Anoplura) and flies (Diptera) associated with humans and livestock. *The Entomologist* **109**, 191-192.

Durden LA, Musser GG (1994) The sucking lice (Insecta, Anoplura) of the world: A taxonomic checklist with records of Mammalian hosts and geographical distributions. *Bulletin of the American Museum of Natural History* **218**, 1-90.

Edwards S, Claude J, Jansen van Vuuren B, Matthee CA (2011) Evolutionary history of the Karoo bush rat , Myotomys unisulcatus (Rodentia: Muridae): disconcordance between morphology and genetics. *Biological Journal of the Linnean Society* **102**, 510-526.

Elith J, Graham CH, Anderson RP *et al*. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**, 129-151.

Elith J, Phillips SJ, Hastie T, Dudik M (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* **17**, 43-57.

Ellery WN, Scholes RJ, Mentis MT (1991) An initial approach to predicting the sensitivity of the grassland biome to climate change. *South African Journal of Science* **87**, 499-503.

Engelbrecht A, Taylor PJ, Daniels SR, Rambau RV (2011) Cryptic speciation in the southern African vlei rat Otomys irroratus complex: evidence derived from mitochondrial cyt b and niche modelling. *Biological Journal of the Linnean Society* **104**, 192-206.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetic analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479-491.

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

Felsenstein J (2004) *Inferring Phylogenies* Sinauer Associates, Sunderland, Massachusetts.

Frankham R, Ballou JD, Briscoe DA (2002) *Introduction to conservation genetics.* Cambridge University Press, Cambridge.

Frantz AC, Cellina S, Krier A, Schley L, Burke T (2009) Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *Journal of Applied Ecology* **46**, 493-505.

Funk DJ, Helbling L, Wernegreen JJ, Moran NA (2000) Intraspecific phylogenetic congruence among multiple symbiont genomes. *Proceedings of the Royal Society of London B Series* **267**, 2517-2521.

Ganem G, Meynard CN, Perigault M*, et al.* (2012) Environmental correlates of co-occurrence of three mitochondrial lineages of striped mice (*Rhabdomys*) in the Free State Province (South Africa). *Acta Oecologica* **42**, 30-40.

Goudie AS (2005) The drainage of Africa since the Cretaceous. *Geomorphology* **67**, 437-456.

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**, 696-704.

Hafner MS, Nadler SA (1990) Cospeciation in host-parasite assemblages: Comparative analysis of rates of evolution and timing of cospeciaiton events. *Systematic Zoology* **39**, 192-204.

Hafner MS, Sudman PD, Villablanca FX*, et al.* (1994) Disparate rates of molecular evolution in cospeciating hosts and parasites. *Science* **265**, 1087-1090.

Halas D, Zamparo D, Brooks DR (2005) A historical biogeographical protocol for studying diversification by taxon pulses. *Journal of Biogeography* **32**, 249-260.

Hall T (2005) BioEdit, Biological sequence alignment editor for Win95/98/NT/2K/XP. Available from http://www.mbio.ncsu.edu/BioEdit/bioedit.html.

Haukisalami V, Wickström LH, Hantula J, Henttonen H (2001) Taxonomy, genetic differentiation and Holarctic biogeography of *Paranoplocephala* spp. (Cestoda: Anoplocephalidae) in collared lemmings (*Dicrostonyx*; Arvicolinae). *Biological Journal of the Linnean Society* **74**, 171-196.

Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* **27**, 570-280.

Hendey QB (1976) The Pliocene fossil occurences in the 'E' quarry, Langebaanweg, South Africa. *Annals of the South African Museum* **69**, 215-247.

Hewitt GM (2011) Quaternary phylogeography: the roots of hybrid zones. *Genetica* **139**, 617-638.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolatedclimate surfaces for global land areas *International Journal of Climatology* **25**, 1965-1978.

Hoberg EP, Brooks DR (2008) A macroevolutionary mosaic: episodic host-switching, geographical colonization and diversification in complex host-parasite systems. *Journal of Biogeography* **35**, 1533-1550.

Hoberg EP, Klassen GJ (2002) Revealing the faunal tapestry: co-evolution and historical biogeography of hosts and parasites in marine systems. *Parasitology* **124**, S3-S22.

Hopkins GHE (1949) The host association of the lice of mammals. *Proceedings of the Zoological Society, London* **119**, 387-604.

Hopley PJ, Marshall JD, Weedon GP*, et al.* (2007) Orbital forcing and the spread of C4 grasses in the late Neogene: stable isotope evidence from South African speleothems. *Journal of Human Evolution* **53**, 620-634.

Huntley BJ, Walker BH (1982) *Ecology of Tropical Savannas* Springer-Verlag, Berlin

Huson DH, Bryant D (2006) Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution* **23**, 254-267.

Huyse T, Poulin R, Theron A (2005) Speciation in parasites: a population genetics approach. *Trends in Parasitology* **21**, 469-475.

Huyse T, Volckaert FAM (2005) Comparing host and parasite phylogenies: *Gyrodactylus* flatworms jumping form goby to goby. *Systematic Biology* **54**, 710-718.

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. *BMC Genetics* **6**: 13.

Jerome CA, Ford BA (2002a) Comparative population structure and genetic diversity of *Arceuthobium americanum* (Viscaceae) and its *Pinus* host species: insight into host-parasite evolution in parasitic angiosperms. *Molecular Ecology* **11**, 407-420.

Jerome CA, Ford BA (2002b) The discovery of three genetic races of the dwarf mistletoe Arceuthobium americanum (Viscaceae) provides insight into the evolution of parasitic angiosperms. *Molecular Ecology* **11**, 387-405.

Johnson KP, Adams RJ, Page RDM, Clayton DH (2003) When do parasites fail to speciate in response to host speciation? *Systematic Biology* **52**, 37-47.

Johnson KP, Clayton DH (2004) Untangling coevolutionary history. *Systematic Biology* **53**, 92-94.

Johnson KP, Williams BL, Drown DM, Adams RJ, Clayton DH (2002) The population genetics of host specificity: genetic differentiation in dove lice (Insecta: Phthiraptera). *Molecular Ecology* **11**, 25-38.

Johnson PT (1960) The Anoplura of African rodents and insectivores. *Technical Bulletin of the United States Department of Agriculture* **1211**, 1-116.

Joly S, Bruneau A (2006) Incorporating allelic variation for reconstructing the evolutionary history of organisms form multiple genes: an example from Rosa in North America. *Systematic Biology* **55**, 623-626.

Joseph L, Wilke T, Alpers D (2002) Reconciling genetic expectations from host specificity with historical population dynamics in an avian brood parasite, Horsfield's Bronze-Cuckoo *Chalcites basalis* of Australia. *Molecular Ecology* **11**, 829-837.

Kass RE, Raftery AE (1995) Bayes Factors. *Journal of the American Statistical Association.* **90**, 773-795.

Kim KC (1985) Evolution and host associations of Anoplura. In: *Coevolution of parasitic arthropods and mammals* (ed. Kim KC), pp. 197-231. John Wiley, New York.

Kim KC (1988) Evolutionary parallelism in Anoplura and eutherian mammals. In: *Biosystematics of haematophagous insects* (ed. Service MW), pp. 91-114. Oxford University Press.

Kim KC (2006) Blood-sucking lice (Anoplura) of small mammals: True parasites. In: *Micromammals and Macroparasites: From Evolutionary Ecology to Management.* (eds. S.Morand, Krasnov BR, Poulin R), pp. 141-160. Springer-Verlag, Tokyo.

Kim KC, Emerson KC, Traub R (1990) Diversity of parasitic insects: Anoplura, Mallophaga, and Siphonaptera In: *Systematics of the North American insects and arachnids: status and needs* (eds. Kosztarab M, Schaefer CW), pp. 91-103. Virginia Agricultural Experimental Station Information Series, Blacksburg, Virginia.

Lamb T, Bauer AM (2000) Relationships of the Pachydactylus rugosus group of geckos (Reptilia: Squamata: Gekkonidae). *African Zoology* **35**, 55-67.

Lancaster JL (2001) *Reproductive isolation in two populations of the striped mouse Rhabdomys pumilio*, University of Witwatersrand.

Larkin MA, Blackshields, G, Brown NP *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.

Lavoipierre MMJ (1967) Feeding mechanism of *Haematopinus suis* on the transilluminated mouse ear. *Experimental Parasitology* **20**, 303-311.

Lawes MJ (1990) The distribution of the samango monkey (*Cercopithecus mitis erythrarchus* Peters, 1852 and *Cercopithecus mitis labiatus* I. Geoffroy, 1843) and the forest history in southern Africa. *Journal of Biogeography* **17**, 669-680.

Ledger JA (1980) *The arthropod parasites of vertebrates in Africa south of the Sahara. Volume IV. Phthiraptera (Insecta).* Publications of the South African Institute for Medical Research No. 56, South African Institute for Medical Research, Johannesburg, South Africa.

Legendre P, Desdevises Y, Bazin E (2002) A statistical test for host-parasite coevolution. *Systematic Biology* **51**, 217-234.

Legendre P, Legendre L (1998) *Numerical ecology*, 2nd edn. Elsevier, New York.

Light JE, Hafner MS (2007) Disparate rates of evolution in sympatric lineages of chewing lice on pocket gophers. *Molecular Phylogenetics and Evolution* **45**, 997-1013.

Light JE, Hafner MS (2008) Codivergence in Heteromyid rodents (Rodentia: Heteromyidae) and their sucking lice of the genus *Fahrenholzia* (Phthiraptera: Anoplura). *Systematic Biology* **57**, 449-465.

Light JE, Reed DL (2009) Multigene analysis of phylogenetic relationships and divergence times of primate sucking lice (Phthiraptera: Anoplura). *Molecular Phylogenetics and Evolution* **50**, 376-390.

Light JE, Smith SS, Allen JM, Durden LA, Reed DL (2010) Evolutionary history of mammalian sucking lice (Phthiraptera: Anoplura). *BMC Evolutionary Biology* **10**: 292.

Linder HP, Hardy CR (2004) Evolution of the species-rich Cape flora. *Philosophical Transactions Of The Royal Society Of London, B series* **359**, 1623-1632.

Linder HP, Johnson SD, Kuhlmann M, *et al.* (2010) Biotic diversity in the Southern African winter-rainfall region. *Current Opinion in Environmental Sustainability* **2**, 109-116.

Lobo JM, Jimenez-Valverde A, Hortal J (2010) The uncertain nature of absences and their importance in species distribution modelling. *Ecography* **33**, 103-114.

Locke S, Mclaughlin JD, Marcogliese DJ (2010) DNA barcodes show cryptic diversity and potential physiological basis for host specificity among Diplostomoidea (Platyhelminthes:

Digenea) parasitizing freshwater fishes in the St. Lawrence River, Canada. *Molecular Ecology* **19**, 2813-2827.

Maddison WP, Knowles LL (2006) Inferring phylogeny despite incomplete lineage sorting *Systematic Biology* **55**, 21-30.

Mahida H, Campbell GK, Taylor PJ (1999) Genetic variation in *Rhabomys pumilio* (Sparrman, 1784) -an allozyme study. *South African Journal of Zoology* **34**, 91-101.

Makokha JS, Bauer AM, Mayer W, Matthee CA (2007) Nuclear and mtDNA phylogenetic inferences among southern African sand lizards, *Pedioplanis* (Sauria: Lacertidae). *Molecular Phylogenetics and Evolution* **44**, 622-633.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**, 209-220.

Marshall AG (1981) *The ecology of ectoparasitic insects.* Aberdeen University Academic Press, London.

Martin AP, Palumbi SR (1993) Body size, metabolic rate, generation time, and the molecular clock. *Procceedings of the National Academy of Sciences of the United States of America* **90**, 4087-4091.

Matthee CA, Burzlaff JD, Taylor, JF, Davis, SK (2001) Mining the mammalian genome for artiodactyl systematics. *Systematic Biology* **50**, 367-390.

Matthee CA, Flemming AF (2002) Population fragmentation in the southern rock agama, *Agama atra*: more evidence for vicariance in Southern Africa. *Molecular Ecology* **11**, 465-471.

Matthee S, Horak IG, Beaucournu J-C, *et al.* (2007) Epifaunistic arthropod parasites of the four-striped mouse, *Rhabdomys pumilio*, in the Western Cape Province, South Africa. *Journal of Parasitology* **93**, 47-59.

Matthee S, McGeoch MA, Krasnov B (2010) Parasite-specific variation and the extent of male-biased parasitism; an example with a South African rodent and ectoparasitic arthropods. *Parasitology* **137**, 651-660.

McCarthy T, Rubidge B (2005) *The story of earth and life: A southern African perspective on a 4.6-billion-year journey* Struik Publishers, South Africa.

McCoy KD (2003) Sympatric speciation in parasites - what is sympatry? *Trends in Parasitology* **19**, 400-404.

McCoy KD, Boulinier T, Tirard C, Michalakis Y (2003) Host-dependant genetic structure of parasite populations: differential dispersal of seabird tick host races. *Evolution* **57**, 288-296.

McManus DP, Bowles J (1996) Molecular genetic approaches to parasite identification: their value in diagnostic parasitology and systematics. *International Journal for Parasitology* **26**, 687-794.

Meester JAJ, Rautenbach IL, Dippenaar NJ, Baker CM (1986) Classification of southern African Mammals. *Transvaal Museum Monograph* **5**, 275-277.

Meier-Kolthoff JP, Auch AF, Huson DH, Göker M (2007) Copycat: Co-phylogenetic Analysis Tool. *Bioinformatics* **23**, 898-900.

Meirmans PG (2012) The trouble with isolation by distance. *Molecular Ecology* **21**, 2839-2846.

Mendlová M, Desdevises Y, Civáňová K, Pariselle A, Šimková A (2012) Monogeneans of West African cichlid fish: evolution and cophylogenetic interactions. *Plos One* **7**, e37268.

Misonne X (1974) Order Rodentia In: *The mammals of Africa: An Identification Manual* (eds. Meester J, Setzer HW). Smithsonian Institution Press, Washington.

Montgelard C, Matthee CA (2012) Tempo of genetic diversification in southern African rodents: The role of Plio-Pleistocene climatic oscillations as drivers for speciation. *Acta Oecologica* **42**, 50-57.

Moran N, Baumann P (1994) Phylogenetics of cytoplasmically inherited microorganisms of arthropods. *Trends in Ecology and Evolution* **9**, 15-20.

Mucina L, Rutherford MC (2006) *The Vegetation of South Africa, Lesotho and Swaziland.* South African National Biodiversity Institute, Pretoria.

Musser GG, Carleton MD (2005) Order Rodentia. In: *Mammal species of the world. A taxonomic and geographic reference.* (eds. Wilson DE, Reeder DM). The John Hopkins University Press, Baltimore.

Nadler SA (1990) Molecular approaches to studying Helminth population genetics and phylogeny. *International Journal for Parasitology* **20**, 11-29.

Nadler SA (1995) Microevolution and the genetic structure of parasite populations. *Journal of Parasitology* **81**, 395-403.

Nadler SA (2002) Species delimitation and Nematode biodiversity: Phylogenies rules. *Nematology* **4**, 615-625.

Nadler SA, de León GPP (2011) Integrating molecular and morphological approaches for characterizing parasite cryptic species: implications for parasitology. *Parasitology* **138**, 1688-1709.

Newton MA, Raftery AE (1994) Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society, Series B* **56**, 3-48.

Nieberding C, Morand S, Douady CJ, Libois R, Michaux J (2005) Phylogeography of a nematode (Heligmosomoides polygyrus) in the western Palearctic region: Persistence of northern cryptic populations during ice ages? *Molecular Ecology* **14**, 765-779.

Nieberding C, S.Morand, Libois R, Michaux JR (2004) A parasite reveals cryptic phylogeographic history of its host. *Procceedings of the Royal Society of London, B Series* **271**, 2559-2568.

Nieberding CM, Durette-Desset MC, Vanderpoorten A*, et al.* (2008) Geography and host biogeography matter for understanding the phylogeography of a parasite. *Molecular Phylogenetics and Evolution* **47**, 538-554.

Nieberding CM, Morand S (2006) Comparative phylogeography: The use of parasites for insights into host history. In: *Micromammals and macroparasites: From evolutionary ecology to management* (eds. Morand S, Krasnov BR, Poulin R), pp. 277-293. Springer-Verlag, Tokyo.

Nieberding CM, Olivieri I (2007) Parasites: proxies for host genealogy and ecology? *Trends in Ecology & Evolution* **22**, 156-165.

Page DM, Hafner MS (1996) Molecular phylogenies and host-parasite co-speciation: gophers and lice as a model system. In: *New uses for new phylogenies* (eds. Harvey PH, Brown AJL, Smith JM, Nee S), pp. 255-270. Oxford University Press.

Page RDM (2003) *Tangled trees: Phylogeny, cospeciation, and coevolution.* University of Chicago press, Chicago.

Page RDM, Cruickshank RH, Dickens M*, et al.* (2004) Phylogeny of *"Philoceanus* complex" seabird lice (Phthiraptera: Ischnocera) inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution* **30**, 633-652.

Page RDM, Lee PLM, Becher SA, Griffiths R, Clayton DH (1998) A different tempo of mitochondrial DNA evolution in birds and their parasitic lice. *Molecular Phylogenetics and Evolution* **9**, 276-293.

Partridge TC (1997) Evolution of landscapes. In: *Vegetation of southern Africa* (eds. Cowling RM, Richardson DM, Pierce SM), pp. 5-20. Cambridge University Press, Cambridge.

Partridge TC, Maud RR (2000) Macro-scale geomorphic evolution of southern Africa. In: *The Cenozoic of Southern Arica* (eds. Partridge TC, Maud RR), pp. 3-18. Oxford University Press, New York.

Paterson AM, Banks J (2001) Analytical approaches to measuring cospeciation of host and parasites: through a looking glass, darkly. *International Journal for Parasitology* **31**, 1012-1022.

Paterson AM, Palma RL, Gray RD (1999) How frequently do avian lice miss the boat? *Systematic Biology* **48**, 214-223.

Paterson AM, Wallis GP, Wallis LJ, Gray RD (2000) Seabird and louse co-evolution: complex histories revealed by 12S rDNA sequences and reconciliation analyses. *Systematic Biology* **49**, 383-399.

Peakall R, Ruibal M, Lindenmayer DB (2003) Spatial autocorrelation analysis offers new insights into gene flow in the Australian bush rat, *Rattus fuscipes*. *Evolution* **57**, 1182-1195.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research - an update. *Bioinformatics* In press.

Peakall ROD, Smouse PE (2006) GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.

Pearce J, Ferrier S (2000) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling* **133**, 225-245.

Pellmyr O, Leebens-Mack J, Thompson JN (1998) Herbivores and molecular clocks as tools in plant biogeography. *Biological Journal of the Linnean Society* **63**, 367-378.

Perkins L, Martinsen ES, Falk BG (2011) Do molecules matter more than morphology? Promises and pitfalls in parasites. *Parasitology* **138**, 1664–1674.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecography* **190**, 231-259.

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

Pillay N (2000a) Female mate preference and reproductive isolation in populations of the striped mouse, Rhabdomys pumilio. *Behaviour* **137**, 1431-1441.

Pillay N (2000b) Reproductive isolation in three populations of the striped mouse Rhabdomys pumilio: interpopulation breeding studies. *Mammalia* **64**, 461-470.

Pillay N, Eborall J, Ganem G (2006) Divergence of mate recognition in the African striped mouse (*Rhabdomys*). *Behavioural Ecology* **17**, 757-764.

Portik DM, Bauer AM, Jackman TR (2011) Bridging the gap: western rock skinks (Trachylepissulcata) have a short history in South Africa. *Molecular Ecology* **20**, 1744-1758.

Posada D (2004) Collapse 1.2: Describing haplotypes from sequence alignments. Available from http://darwin.uvigo.es/software/collapse.html.

Posada D (2008) jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* **25**, 1253-1256.

Poulin R, Keeney DB (2007) Host specificity under molecular and experimental scrutiny. *Trends in Parasitology* **24**, 24-28.

Poulin R, Morand S (2004) *Parasite Biodiversity* Smithsonian Institution, Washington.

Prendini L, Crowe TM, Wheeler WC (2003) Systematics and biogeography of the family Scorpionidae (Chelicerata: Scorpiones), with a discussion on phylogenetic methods. *Invertebrate Systematics* **17**, 185-259.

Price BW, Barker NP, Villet MH (2007) Patterns and processes underlying evolutionary significant units in the Platypleura stridula L. species complex (Hemiptera: Cicadidae) in the Cape Floristic Region, South Africa. *Molecular Ecology* **16**, 2574-2588.

Prugnolle F, Theron A, Pointer JP, *et al.* (2005) Dispersal in a parasitic worm and its two hosts: consequences for local adaptation. *Evolution* **59**, 296-303.

Rambau RV, Robinson TJ, Stanyon R (2003) Molecular genetics of *Rhabdomys pumilio* subspecies boundaries: mtDNA phylogeography and karyotypic analysis by fluorescence in situ hybridization. *Molecular Phylogenetics and Evolution* **28**, 564-575.

Rambaut A, Drummond AJ (2007) Tracer v1.4, Available from http://beast.bio.ed.ac.uk/Tracer.

Rannala B, Michalakis Y (2003) Population genetics and cospeciation: from process to pattern. In: *Tangled trees. Phylogeny, Cospeciation and Coevolution* (ed. Page RDM), pp. 120-143. Chicago University Press, Chicago.

Redman GT, Hamer ML (2003) The distribution of southern African Harpagophoridae Attems, 1909 (Diplopoda: Spirostreptida). *African Invertebrates* **44**, 213-226.

Reed DL, Hafner MS (1997) Host specificity of chewing lice on pocket gophers: A potential mechanism for cospeciation. *Journal of Mammalogy* **78**, 655-660.

Reed DL, Smith VS, Hammond SL, Rogers AR, Clayton DH (2004) Genetic analysis of lice supports direct contact between modern and archaic humans. *PLOS Biology* **2**, 1972-1983.

Reed LK, Nyboer M, Markow TA (2007) Evolutionary relationships of *Drosophila mojavensis* geographic host races and their sister species *Drosophila arizonae*. *Molecular Ecology* **16**, 1007-1022.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution* **43**, 223-225.

Rivas LR (1964) A reinterpretation of the concepts "sympatric" and "allopatric" with proposal of the additional terms "syntopic" and "allotopic". *Systematic Zoology* **13**, 42-43.

Roberts A (1951) *The mammals of South Africa* (eds. Bigalke R, Fitzzimmons V, Malan DE), pp. 497-505. Trustees of "The mammals of South Africa" book fund, Johannesburg.

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

Ronquist F, Teslenko M, van der Mark P*, et al.* (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* In Press.

Ronsted N, Weiblen GD, Cook JM*, et al.* (2005) 60 million years of co-divergence in the fig-wasp symbiosis. *Procceedings of the Royal Society of London B Series* **272**, 2593-2599.

Rozsa L (1993) Speciation patterns of ectoparasites and "straggling" lice. *International Journal for Parasitology* **23**, 859-864.

Russo I, Chimimba CT, Bloomer P (2010) Bioregion heterogeneity correlates with extensive mitochondrial DNA diversity in the Namaqua rock mouse, *Micaelamys namaquensis* (Rodentia: Muridae) from southern Africa - evidence for a species complex. *BMC Evolutionary Biology* **10**, 307-307.

Safner T, Miller MP, McRae BH, Fortin M-J, Manel S (2011) Comparison of Bayesian clustering and edge detection methods for inferring boundaries in landscape genetics. *International Journal of Molecular Sciences* **12**, 865-889.

Schlick-Steiner BC, Seifert B, Stauffer C*, et al.* (2007) Without morphology, cryptic species stay in taxonomic crypsis following discovery. *Trends in Ecology & Evolution* **22**, 391-392.

Schradin C, Ko B, Pillay N (2010) Reproductive competition favours solitary living while ecological constraints impose group-living in African striped mice. *Journal of Animal Ecology* **79**, 515-521.

Schradin C, Pillay N (2004) The striped mouse (*Rhabdomys pumilio*) from the Succulent Karoo, South Africa: a territorial group-living solitary forager with communal breeding and helpers at the nest. *Journal of Comparative Psychology* **118**, 37-47

Schradin C, Pillay N (2005) Intraspecific variation in the spatial and social organization of the African striped mouse. *Journal of Mammalogy* **86**, 99-107.

Schradin C, Scantlebury M, Pillay N, König B (2009) Testosterone levels in dominant sociable males are lower than in solitary roamers: physiological differences between trhee male reproductive tactis in a socially flexible mammal. *The American Naturalist* **173**, 376-388.

Schulze RE (1997) Climate. In: *Vegetation of southern Africa* (eds. Cowling RM, Richardson DM, Pierce SM), pp. 21-42. Cambridge University Press, Cambridge.

Scott L, Anderson HM, Anderson JM (1997) Vegetation history. In: *Vegetation of Southern Africa* (eds. Cowling RM, Richardson DM, Pierce SM), pp. 62-84. Cambridge University Press, Cambridge.

Scott L, Nyakale M (2002) Pollen indicators of Holocene palaeoenvironments at Florisbad in the central Free State, South Africa. *The Holocene* **12**, 497-503.

Shackleton NJ, Kennet NP (1975) Paleotemperature history of the Cenozoic and the initiation of antarctic glaciation: Oxygen and carbon isotope analyses in DSDP sites 277, 279 and 281. In: *Initial Reports. DSDP, 29* (eds. Kennett JP, Houtz RE), pp. 743-755. U.S. Government Printing Office, Washington.

Skinner JD, Chimimba CT (2005) *The Mammals of the South African subregion.* Cambridge University Press, Cambridge.

Smit HA, Robinson TJ, Jansen van Vuuren B (2007) Coalescence methods reveal the impact of vicariance on the spatial genetic structure of *Elephantulus edwardii* (Afrotheria, Macroscelidea). *Molecular Ecology* **16**, 2680-2692.

Smith VS, Ford T, Johnson KP*, et al.* (2011) Multiple lineages of lice pass through the K-Pg boundary. *Biology letters* **7**, 782-785.

Smouse PE, Long JC (1992) Matrix Correlation-Analysis In Anthropology and Genetics. *Yearbook Of Physical Anthropology* **35**, 187-213.

Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology* **35**, 627-632.

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity* **82**, 561-573.

Solmsen N, Johannesen J, Schradin C (2011) Highly asymmetric fine-scale genetic structure between sexes of African striped mice and indication for condition dependent alternative male dispersal tactics. *Molecular Ecology* **20**, 1624-1634.

Springer MS, Murphy WJ, Eizirk E, O'Brien SJ (2003) Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 1056-1061.

Stamatakis A, Auch AF, Meier-Kolthoff JP, Göker M (2007) AxPcoords & parallel AxParafit: statistical co-phylogenetic analyses on thousands of taxa. *BMC Bioinformatics* **8**, 405.

Štefka J, Hoeck PEA, Keller LF, Smith VS (2011) A hitchhikers guide to the Galápagos: co-phylogeography of Galápagos mockingbirds and their parasites. *BMC Evolutionary Biology* **11**: 284.

Štefka J, Hypša V (2008) Host specificity and genealogy of the louse *Polyplax serrata* on field mice, *Apodemus* species: A case of parasite duplication or colonization? *International Journal for Parasitology* **38**, 731-741.

Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing data imputation. *American Journal of Human Genetics* **76**, 449-462.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* **68**, 978-989.

Stippel BG (2009) *Reproductive isolation in four populations of the striped mouse Rhabdomys*. University of Witwatersrand.

Suchard MA, Weiss RA, Sinsheimer JS (2001) Bayesian selection of continuous-time Markov chain evolutionary models. *Molecular Biology and Evolution* **18**, 1001-1013.

Switzer WM, Salemi M, Shanmugam V, *et al.* (2005) Ancient co-speciation of simian foamy viruses and primates. *Nature* **374**, 376-380.

Swofford DL (2000) PAUP: Phylogenetic Analysis Using Parsimony (and other methods). Sinauer Associates, Sunderland, Massachusetts.

Taylor PJ, Maree S, van Sandwyk J, Baxter R, Rambau RV (2009) When is a species not a species? Uncoupled phenotypic, karyotypic and genotypic divergence in two species of South African laminate-toothed rats (Murinae: Otomyini). *Journal of Zoology, London* **277**, 317-332.

Thomas F, Verneau O, Meeûs TD, Renaud F (1996) Parasites as host evolutionary prints: Insights into host evolution from parasitological data. *International Journal for Parasitology* **26**, 677-686.

Thompson JN (1994) *The coevolutionary process.* University of Chicago Press, Chicago.

Tolley KA, Braae A, Cunningham M (2010) Phylogeography of the Clicking Stream Frog *Strongylopus grayii* (Anura, Pyxicephalidae) reveals cryptic divergence across climatic zones in an abundant and widespread taxon. *African Journal of Herpetology* **59**, 17-32.

Tolley KA, Burger M, Turner AA, Matthee CA (2006) Biogeographic patterns and phylogeography of dwarf chameleons (Bradypodion) in an African biodiversity hotspot. *Molecular Ecology* **15**, 781-793.

Tolley KA, Chase BM, Forest F (2008) Speciation and radiations track climate transitions since the Miocene Climatic Optimum: a case study of southern African chameleons. *Journal of Biogeography* **35**, 1402-1414.

Tolley KA, Tilbury CR, Branch WR, Matthee CA (2004) Phylogenetics of the southern African dwarf chameleons, Bradypodion (Squamata : Chamaeleonidae). *Molecular Phylogenetics and Evolution* **30**, 354-365.

Tompkins DM, Clayton DH (1999) Host resources govern the specificity of swiftlet lice: Size matters. *Journal of Animal Ecology* **68**, 489-500.

Tyson PD, Partridge TC (2000) The evolution of Cenozoic climates. In: *The Cenozoic of Southern Africa* (eds. Partridge TC, Maud RR), pp. 371-387. Oxford University Press, New York.

VanDerWal J, Shoo LP, Graham C, Williams SE (2009) Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? *Ecological Modelling* **220**, 589-594.

Vlok JHJ, Euston-Brown DIW (2002) The patterns within, and the ecological processes that sustain, the subtropical thicket vegetation in the planning domain for the Subtropical Thicket Ecosystem Planning (STEP) project. *TERU Report* **40**, 1-142.

Vogel JC (1978) Isotopic assessment of the dietary habits of ungulates. *South African Journal of Science* **74**, 298-301.

Vogel JC, Fuls A, Ellis RP (1978) The distribution of Kranz grasses in South Africa. *South African Journal of Science* **74**, 209-215.

Warren DL, Glor RE, Turelli M (2010) ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography* **33**, 607–611.

Weckstein JD (2004) Biogeography explains cophylogenetic patterns in toucan chewing lice. *Systematic Biology* **53**, 154-164.

Weiblen GD, Bush GL (2002) Speciation in fig pollinators and parasites. *Molecular Ecology* **11**, 1573-1578.

Whiteman NK, Kimball RT, Parker PG (2007) Co-phylogeography and comparative population genetics of the threatened Galápagos hawk and three ectoparasite species: ecology shapes population histories within parasite communities. *Molecular Ecology*, 1-15.

Whiteman NK, Parker PG (2005) Using parasites to infer host population history: a new rationale for parasite conservation. *Animal Conservation* **8**, 175-181.

Wickström LH, Haukisalmi V, Varis S*, et al.* (2003) Phylogeography of the circumpolar Paranoplocephala arctica species complex (Cestoda: Anoplocephalidae) parasitizing collared lemmings (Dicrostonyx spp.). *Molecular Ecology* **12**, 3359-3371.

Willows-Munro S, Matthee CA (2009) The evolution of the southern African members of the shrew genus *Myosorex*: Understanding the origin and diversification of a morphologically cryptic species. *Molecular Phylogenetics and Evolution* **51**, 394-398.

Willows-Munro S, Matthee CA (2011) Linking lineage diversification to climate and habitat heterogeneity: phylogeography of the southern African shrew *Myosorex varius*. *Journal of Biogeography* **38**, 1976-1991.

Wirth T, Meyer A, Achtman M (2005) Deciphering host migrations and origins by means of their microbes. *Molecular Ecology* **14**, 3289-3306.

Yang Z, Rannala B (2005) Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution* **23**, 212-226.

Yoshizawa K, Johnson KP (2006) Morphology of male genitalia in lice and their relatives and phylogenetic implications. *Systematic Entomology* **31**, 350-361.

Zachos J (2001) Trends , Rhythms , and Aberrations in Global Climate 65 Ma to Present. *Science* **292**, 686-693.

Appendix A

Figures
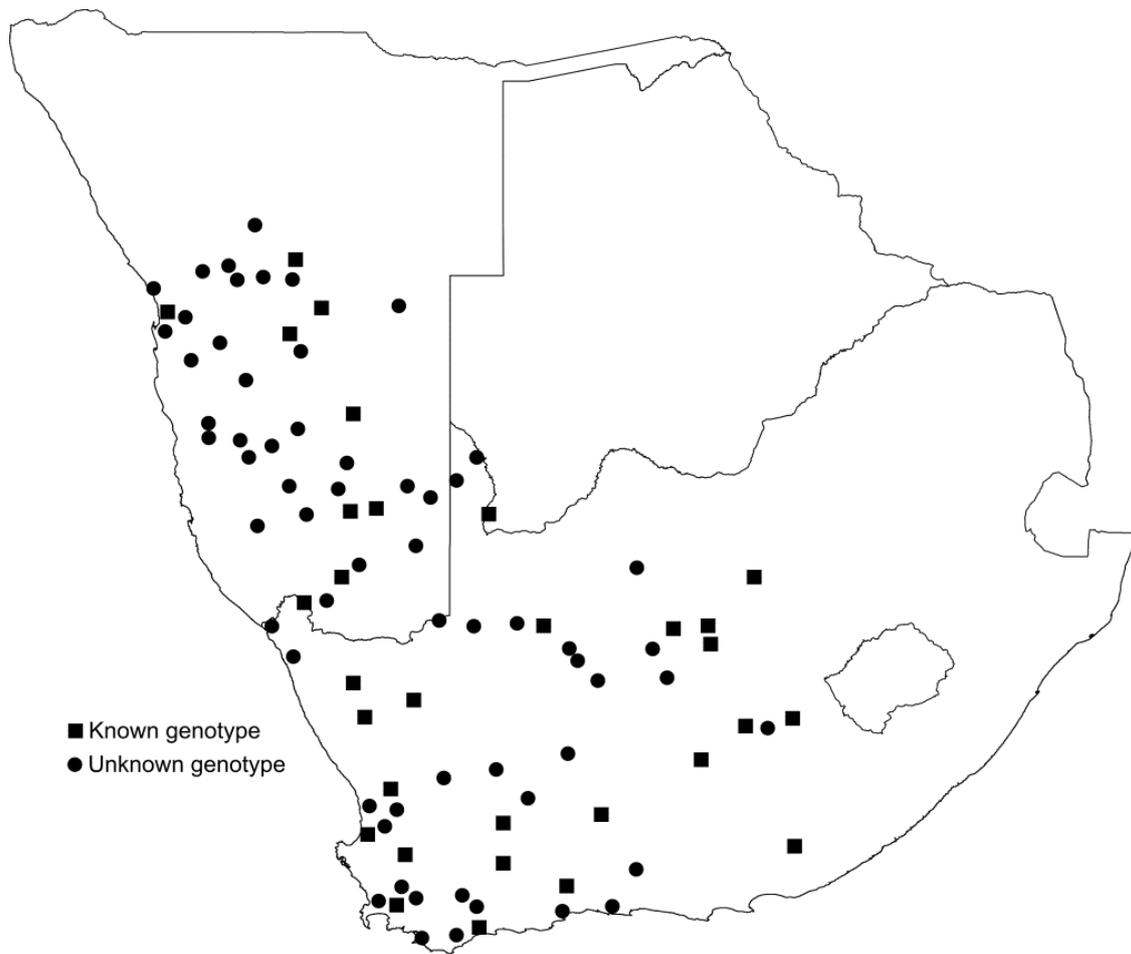
Figure A.1: *Rhabdomys pumilio* presence records used in the MaxEnt analysis, with known genotypes (from this study) and unknown genotypes (museum records).
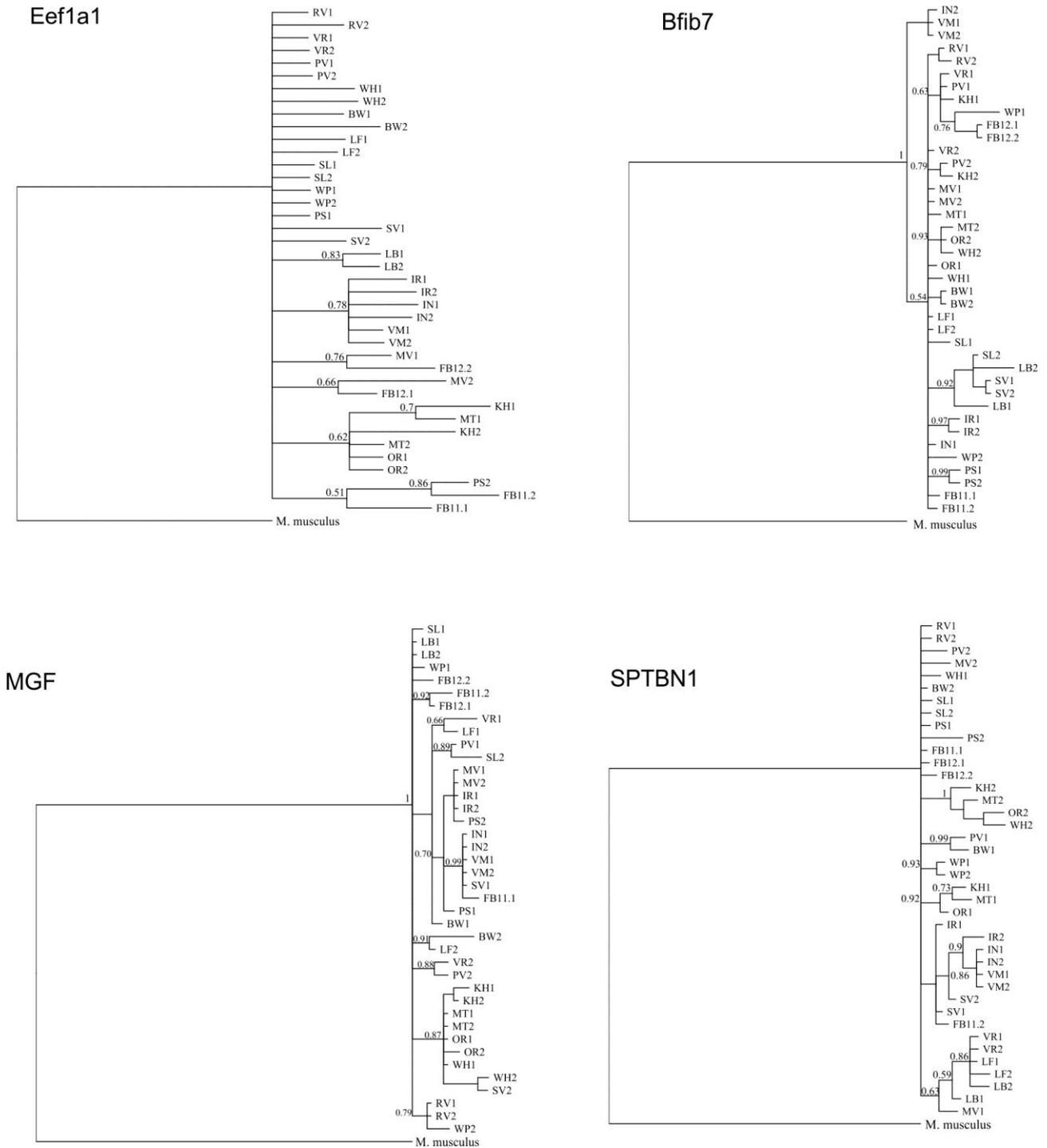
Figure A.2: Bayesian phylograms of the individual nuclear intron analyses with posterior probabilities indicated.
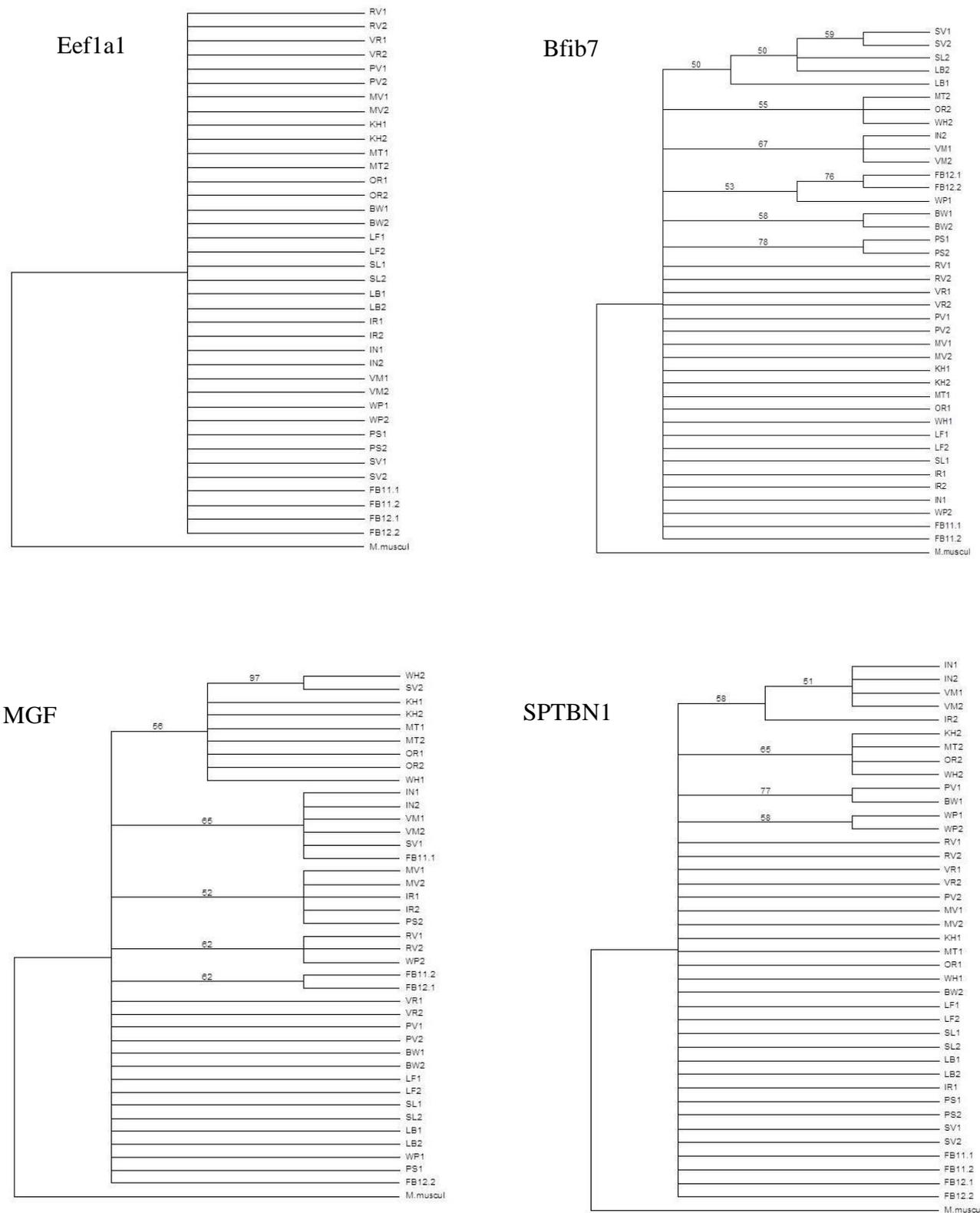
Figure A.3: Parsimony topologies resulting from the individual nuclear intron analyses with bootstrap values indicated.
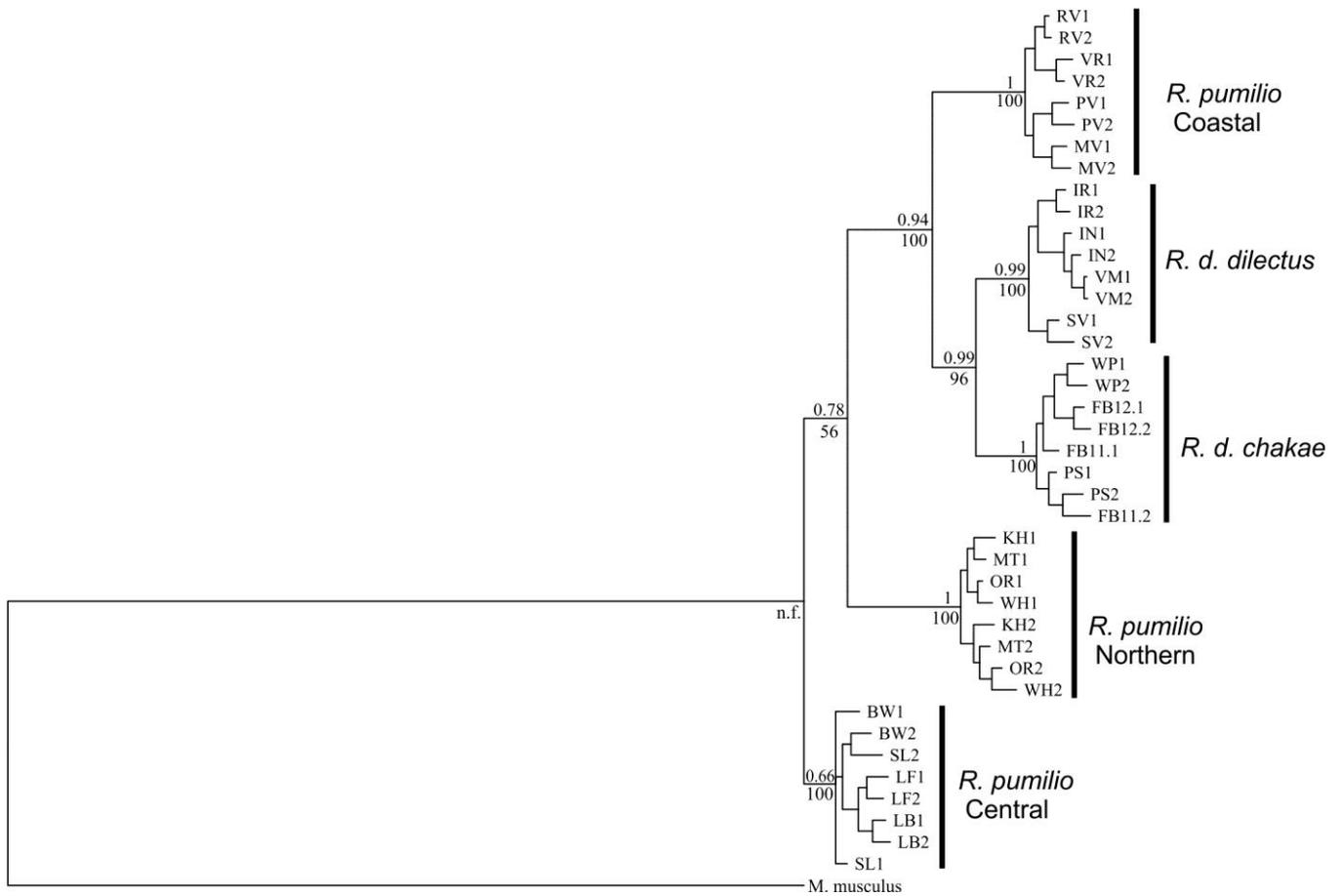
Figure A.4: Consensus Bayesian and parsimony topology for the combined mitochondrial and nuclear dataset with locality codes as in Table 2.1. Posterior probabilities and bootstrap support values are indicated above and below nodes, respectively.
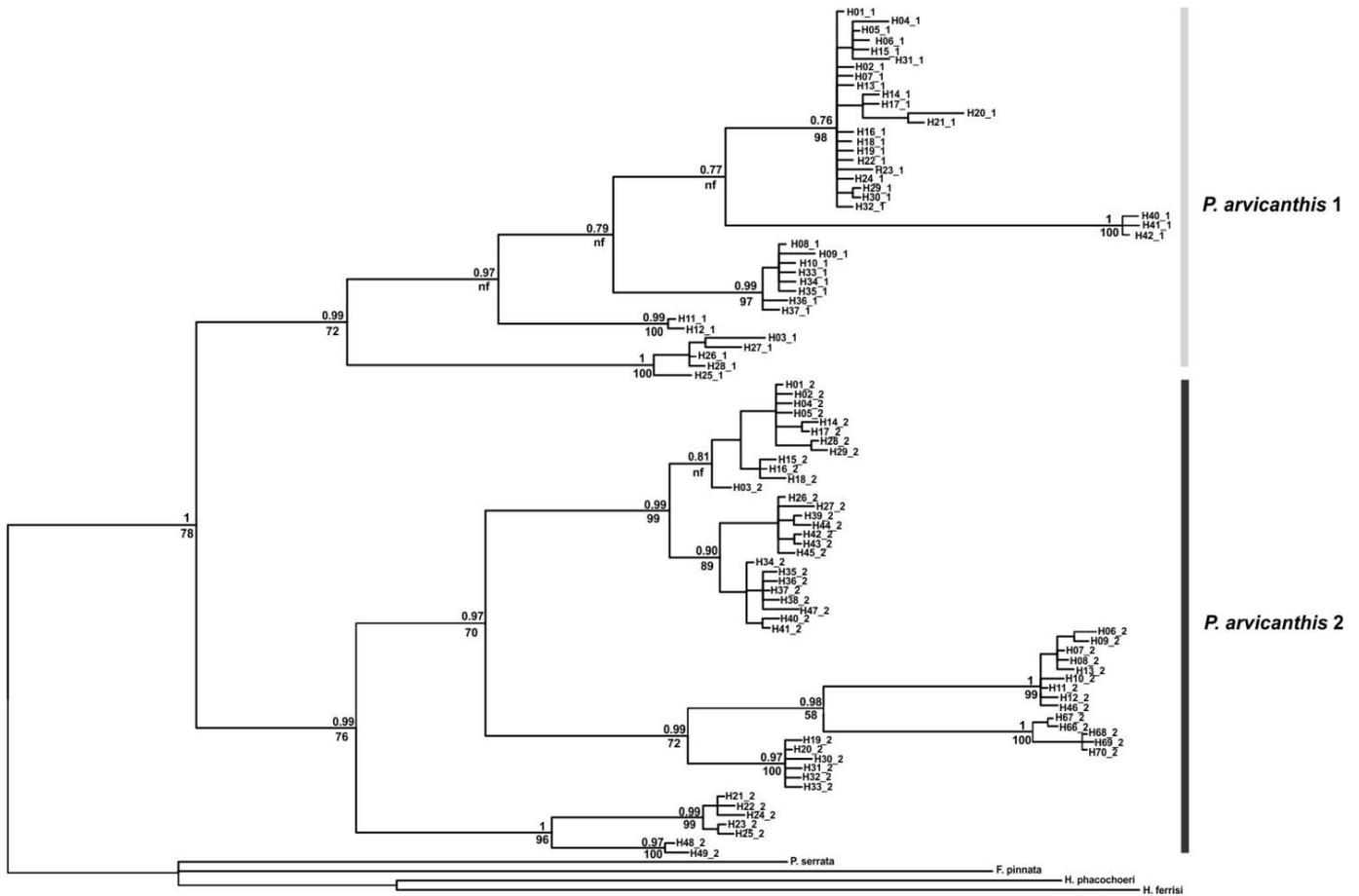
Figure A.5: Consensus Bayesian and Parsimony COI topology indicating the relationships among the 94 haplotypes within *P. arvicanthis* (codes as in Table 3.1). Posterior probabilities and bootstrap support values are indicated above and below nodes, respectively.

Appendix B

Tables

Table B.1: All haplotypes used in the mitochondrial COI analyses with SUN database numbers, GenBank accession numbers, and individual specimens belonging to each haplotype. Locality codes correspond to those in Table 2.1.

Haplotype	Accession	SUN database	Individuals
H001	JQ003320	3838	RV1,2,3,4,5,6,7,8,9,10,11,13,15,14,16,17,18,19,20, 21,22,23,24,25,26,27,28,29,31
H002	JQ003321	3839	RV12
H003	JQ003322	3840	RV30
H004	JQ003323	3841	MT1,3; KH7
H005	JQ003324	3842	MT2
H006	JQ003325	3843	MT4
H007	JQ003326	3844	KH1,2,6,14
H008	JQ003327	3845	KH3,4,15; GR2,3,4,5,8,9,12; FR1,3,4,6,8,10,13; DF15
H009	JQ003328	3846	KH5,12
H010	JQ003329	3847	KH8,9,10,13,17,19
H011	JQ003330	3848	KH11
H012	JQ003331	3849	KH16
H013	JQ003332	3850	KH18
H014	JQ003333	3851	KH20
H015	JQ003334	3852	KH21; GR14
H016	JQ003335	3853	GP1,2,3,5,9,10,11,13,14,15,18,21,25,26,28,30
H017	JQ003336	3854	GP4
H018	JQ003337	3855	GP6,8,17,19,20,27,29; VR1,3,4,8,16,23,24; PR5,6,8,9,17,18
H019	JQ003338	3856	GP7
H020	JQ003339	3857	GP12
H021	JQ003340	3858	GP16,22; PR1,10,12,13
H022	JQ003341	3859	GP23
H023	JQ003342	3860	GP24
H024	JQ003343	3861	OH1
H025	JQ003344	3862	OH2,6,11,16
H026	JQ003345	3863	OH3,4,5; DH2,3,5,7,8,12,13,18,19
H027	JQ003346	3864	OH7,8,17,18,19,20,21,22,24,25,26,27,28,29,32,33
H028	JQ003347	3865	OH9
H029	JQ003348	3866	OH10
H030	JQ003349	3867	OH12
H031	JQ003350	3868	OH14
H032	JQ003351	3869	OH15,23
H033	JQ003352	3870	OH30
H034	JQ003353	3871	PV1,3,9,11,14,17,23,26

Table B.1 continued

Haplotype	Accession	SUN database	Individuals
H035	JQ003354	3872	PV2
H036	JQ003355	3873	PV4,6,12,13,15,18,21,25,27,28,29,30,31
H037	JQ003356	3874	PV5
H038	JQ003357	3875	PV7,20,24; RR13; MV4,19
H039	JQ003358	3876	PV8; MV2,12,29
H040	JQ003359	3877	PV16,22; RR7
H041	JQ003360	3878	PV19
H042	JQ003361	3879	VR2
H043	JQ003362	3880	VR5
H044	JQ003363	3881	VR6,7,15,20
H045	JQ003364	3882	VR9,18,25,26,28,30
H046	JQ003365	3883	VR10
H047	JQ003366	3884	VR11,12,17,27
H048	JQ003367	3885	VR13
H049	JQ003368	3886	VR14
H050	JQ003369	3887	VR19,22
H051	JQ003370	3888	VR21
H052	JQ003371	3889	VR29
H053	JQ003372	3890	FB1
H054	JQ003373	3891	FB2,3,5,6,7,8,10
H056	JQ003375	3893	FB9
H057	JQ003376	3894	FB11
H058	JQ003377	3895	FB12
H059	JQ003378	3896	PR3
H060	JQ003379	3897	PR4,7
H061	JQ003380	3898	PR11
H062	JQ003381	3899	PR14
H063	JQ003382	3900	PR16
H064	JQ003383	3901	NR2,10,11; OR1,6,12,15
H065	JQ003384	3902	NR4
H066	JQ003385	3903	NR6
H067	JQ003386	3904	NR7
H068	JQ003387	3905	NR12
H069	JQ003388	3906	NR15
H070	JQ003389	3907	NR18
H071	JQ003390	3908	GR6
H072	JQ003391	3909	GR7

Table B.1 continued

Haplotype	Accession	SUN database	Individuals
H073	JQ003392	3910	GR13; TR1
H074	JQ003393	3911	GR15
H075	JQ003394	3912	FR5
H076	JQ003395	3913	FR9
H077	JQ003396	3914	FR12
H078	JQ003397	3915	FR14
H079	JQ003398	3916	OR2,9
H080	JQ003399	3917	OR3
H081	JQ003400	3918	OR4
H082	JQ003401	3919	OR5
H083	JQ003402	3920	OR7
H084	JQ003403	3921	OR8
H085	JQ003404	3922	OR10
H086	JQ003405	3923	OR13
H087	JQ003406	3924	OR17
H088	JQ003407	3925	OR18
H089	JQ003408	3926	GD14; SV 2,3; GD1,2,4,6,9,10,11,12,13; TDR1,5,6,7,8,9; BF2; DF2,5,6,8,9,18,20 GD15,5,7,8L TDR3;
H090	JQ003409	3927	DF3,11,12,16,17,23,25,26,27,28
H091	JQ003410	3928	RR1,5,8,10
H092	JQ003411	3929	RR2
H093	JQ003412	3930	RR4,6,9,11,14
H094	JQ003413	3931	RR15
H095	JQ003414	3932	DH1,10
H096	JQ003415	3933	DH4
H097	JQ003416	3934	DH6
H098	JQ003417	3935	DH9,11,15
H099	JQ003418	3936	DH14,16
H100	JQ003419	3937	DH20
H101	JQ003420	3938	MV1,3,15,16,17,18,20,21,22,23,27,28,30,31
H102	JQ003421	3939	MV5
H103	JQ003422	3940	MV6
H104	JQ003423	3941	MV7
H105	JQ003424	3942	MV8,9,10,11,13,14,24,25,26
H106	JQ003425	3943	LF1,2,5,6,7,8,9
H107	JQ003426	3944	LF3,12
H108	JQ003427	3945	LF11,14; BW6; SL11
H109	JQ003428	3946	PS
H110	JQ003429	3947	SB1; AL1,2; KD2
H111	JQ003430	3948	SB2
H112	JQ003431	3949	IR

Table B.1 continued

Haplotype	Accession	SUN database	Individuals
H113	JQ003432	3950	IN
H114	JQ003433	3951	VM1
H115	JQ003434	3952	VM2
H116	JQ003435	3953	WP1
H117	JQ003436	3954	WP2
H118	JQ003437	3955	WP3
H119	JQ003438	3956	BW25; MB3
H120	JQ003439	3957	BW27; LB1,3,7,8,9,10,11,16,22,23; SL10
H121	JQ003440	3958	BW28
H122	JQ003441	3959	SM1
H123	JQ003442	3960	SM2
H124	JQ003443	3961	MB4
H125	JQ003444	3962	MB5
H126	JQ003445	3963	MB7,9,10,12
H127	JQ003446	3964	MB8
H128	JQ003447	3965	MB14
H129	JQ003448	3966	LB2,5,6,18,19,20,21
H130	JQ003449	3967	SL2,5,9
H131	JQ003450	3968	SL4
H132	JQ003451	3969	SL7
H133	JQ003452	3970	SL13
H134	JQ003453	3971	SV1; DF7,10; RP1,2,3,4,5,6,7,8,9,10,11,12,13,14,15
H135	JQ003454	3972	SV4,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20
H136	JQ003455	3973	SV5
H137	JQ003456	3974	GD3
H138	JQ003457	3975	TDR2,4
H139	JQ003458	3976	TDR10
H140	JQ003459	3977	KD1
H141	JQ003460	3978	BF1; DF24
H142	JQ003461	3979	GH1,4,6,8,9,11,13,14
H143	JQ003462	3980	GH2,3,7,10,12
H144	JQ003463	3981	GH5
H145	JQ003464	3982	DF4
H146	JQ003465	3983	DF14,22
H147	JQ003466	3984	WH3,6,9,12,13,17,19,21
H148	JQ003467	3985	WH4,7,14,16
H149	JQ003468	3986	WH5,10,11,15,18,22
H150	JQ003469	3987	WH20
H151	JQ003470	3988	WH23

Table B.2: Number of specimens from each sampled locality (Codes as in Table 2.1 and Fig. 2.1) for which mtDNA and nDNA sequences were generated.

Species	Code	COI	Introns
R. pumilio			
	OR	15	1
	NR	9	
	WH	20	1
	MT	4	1
	KH	21	1
	GR	12	
	FR	11	
	SM	2	
	RV	31	1
	GP	30	
	LF	11	1
	GH	14	
	SL	13	1
	DF	24	
	RP	15	
	VR	30	1
	PV	30	1
	RR	13	
	PR	16	
	SB	31	1
	DH	19	
	OH	31	
	BW	33	1
	LB	23	1
	TR	1	
	MB	22	
	FB	10	
	BF	2	
	GD	15	
	TDR	10	
	SV	3	1
R. dilectus			
	AL	2	
	FB	2	2
	WP	3	1
	SV	17	
	KD	2	

Table B.2 continued

Species	Code	COI	Introns
R. dilectus	SR	2	
	IR	1	1
	PS	1	1
	IN	1	1
	VM	2	1

Table B.3: Primers used for PCR amplification of the different gene fragments.

Target gene	Primer	Sequence (5'-3')	Annealing temperature (°C)
COI	COIL1490[1]	GGTCAACAAATCATAAAGATATTGG	45
	COIH7005[2]	CCGGATCCACNACRTARTANGTRTCRTG	
Eef1a1	EF1-R-For3[3]	GGGGACAAYGTTGGTTTCAACG	55-56
	Cho 10[4]	ACGGCVACKGTYTGHCKCATGTC	
MGF	MGF_A[5]	ATCCATTGATGCCTTCAAGG	62
	MGF_B[5]	CTGTCATTCCTAAGGGAGCTG	
	MGF_R[3]	AGTCGTCTTGTAGAGTGATGG	
SPTBN1	SPTBN1_A[6]	TCTCAAGACTATGGCAAACA	60
	SPTBN1_B[6]	CTGCCATCTCCCAGAAGAA	
Bfib7	BFIBR1[7]	ATTCACAACGGCATGTTCTTCAG	58
	BFIBR2[7]	AANGKCCACCCCAGTAGTATCTC	

[1] Folmer *et al*., 1994; [2]Anderson *et al*. 1981; [3]This study; [4] Danforth & Ji 1998;
[5] Lyons *et al*. 1997; [6] Venta *et al*. 1996; [7] Seddon *et al*. 2001

Table B.4: GenBank accession numbers for the nuclear sequences generated in this study. Haplotype numbers reflect the corresponding mitochondrial COI haplotype for each intron sequence (Table B.1).

Gene	Haplotype	Accession
SPTBN	H001	JQ003241
	H004	JQ003242
	H007	JQ003243
	H018	JQ003244
	H034	JQ003245
	H055	JQ003246
	H057	JQ003247
	H058	JQ003248
	H064	JQ003249
	H101	JQ003250
	H106	JQ003251
	H109	JQ003252
	H112	JQ003253
	H113	JQ003254
	H114	JQ003255
	H116	JQ003256
	H120	JQ003257
	H129	JQ003258
	H134	JQ003259
	H147	JQ003260
MGF	H001	JQ003261
	H004	JQ003262
	H007	JQ003263
	H018	JQ003264
	H034	JQ003265
	H055	JQ003266
	H057	JQ003267
	H058	JQ003268
	H064	JQ003269
	H101	JQ003270
	H106	JQ003271
	H109	JQ003272
	H112	JQ003273
	H113	JQ003274
	H114	JQ003275
	H116	JQ003276
	H120	JQ003277
	H129	JQ003278
	H134	JQ003279

Table B.4 continued

Gene	Haplotype	Accession
MGF	H147	JQ003280
Bfib7	H001	JQ003281
	H007	JQ003282
	H004	JQ003283
	H018	JQ003284
	H034	JQ003285
	H055	JQ003286
	H057	JQ003287
	H058	JQ003288
	H064	JQ003289
	H101	JQ003290
	H106	JQ003291
	H109	JQ003292
	H112	JQ003293
	H113	JQ003294
	H114	JQ003295
	H116	JQ003296
	H120	JQ003297
	H129	JQ003298
	H134	JQ003299
	H147	JQ003300
Eef1a1	H001	JQ003301
	H004	JQ003302
	H007	JQ003303
	H018	JQ003304
	H034	JQ003305
	H055	JQ003306
	H057	JQ003307
	H058	JQ003308
	H064	JQ003309
	H101	JQ003310
	H106	JQ003311
	H109	JQ003312
	H112	JQ003313
	H113	JQ003314
	H114	JQ003315
	H116	JQ003316
	H120	JQ003317
	H129	JQ003318
	H134	JQ003319

Appendix B

Table B.5: All geo-referenced presence records used in the MaxEnt analyses with the origin of each record (genotyped in this study or museum record) indicated.

Clade	Locality	Geographic coordinates	Origin
R. pumilio Coastal	Bredasdorp	34°32'S, 20°02'E	museum
	Clanwilliam	32°11'S, 18°53'E	museum
	De Hoop	34°29'S, 20°24'E	genotyped
	Fort Beaufort	32°51'S, 26°27'E	genotyped
	Franskraal	34°36'S, 19°23'E	museum
	Klipfontein	30°30'S, 17°49'E	museum
	Knysna	34°02'S, 23°03'E	museum
	Lambert's bay	32°06'S, 18°18'E	museum
	Milnerton	33°52'S, 18°32'E	museum
	Modderfontein	30°45'S, 18°19'E	museum
	Montagu	33°47'S, 20°07'E	museum
	Mosselbaai	34°11'S, 22°08E	museum
	Oudtshoorn	33°36'S, 22°08'E	genotyped
	Paulshoek	30°23'S, 18°17'E	genotyped
	Port Nolloth	29°15'S, 16°53'E	museum
	Porterville	32°59'S, 19°01'E	genotyped
	Redelinghuys	32°29'S, 18°32'E	museum
	Richtersveld	28°12'S, 17°06'E	genotyped
	Rocher Pan	32°36'S, 18°18'E	genotyped
	Springbok	29°42'S, 18°02'E	genotyped
	Stellenbosch	33°55'S, 18°49'E	genotyped
	Stettynskloof	33°50'S, 19°16'E	museum
	Swellendam	34°01'S, 20°25'E	museum
	Vanrhynsdorp	31°44'S, 18°46'E	genotyped
	Wellington	33°38'S, 18°59'E	museum
	Wilowmore	33°18'S, 23°28'E	museum
R. pumilio Central	Beaufort West	32°13'S, 22°48'E	genotyped
	Calvinia	31°28'S, 19°46'E	museum
	Carnarvon	30°58'S, 22°08'E	museum
	Fort Beaufort	32°51'S, 26°27'E	genotyped
	Fraserburg	31°53'S, 21°32'E	museum
	Hopetown	29°37'S, 24°04'E	museum
	Laingsburg	33°10'S, 20°55'E	genotyped
	Loeriesfontein	31°04'S, 19°13'E	genotyped
	Sneeuberg	31°45'S, 24°46'E	genotyped
	Sutherland	32°24'S, 20'54'E	genotyped
	Vanwykvlei	30°12'S, 21°49'E	museum
	Williston	31°20'S, 20°55'E	museum

Appendix B

Table B.5 continued

Clade	Locality	Geographic coordinates	Origin
R. pumilio Northern	Alexander bay	28°38'S, 16°30'E	museum
	Aandster	25°23'S, 16° 04'E	museum
	Amichab	23°12'S, 15°31'E	museum
	Asab	25°28'S, 17°57'E	museum
	Augrabies falls	28°36'S, 20°20'E	museum
	Aus	26°40'S, 16°15'E	museum
	Benfontein	28°49'S, 24°49'E	genotyped
	Berseba	25°59'S, 17°46'E	museum
	Bethanien	26°30'S, 17°46'E	museum
	Bethulie	30°31'S, 25°59'E	museum
	Boegoeberg dam	29°02'S, 22°12'E	museum
	Douglas	29°03'S, 23°46'E	museum
	Dronfield	28°37'S, 24°48'E	genotyped
	Erongo	21°40'S, 15°40'E	museum
	Fish River Canyon	27°41'S, 17°48'E	genotyped
	Gababeb	23°31'S, 14°59'E	museum
	Gaibis	26°09;S, 19°32'E	museum
	Gariep Dam	30°33'S, 25°32'E	genotyped
	Gellap	26°24'S, 18°00'E	genotyped
	Goanikontes	22°40'S, 14°50'E	museum
	Gobabis	22°27'S, 18°57'E	museum
	Gorab	25°09'S, 16°30'E	museum
	Great Karas	27°15'S, 18°03'E	museum
	Groblershoop	28°37'E, 21°42'E	genotyped
	Helmeringhausen	25°54'S, 16°49'E	museum
	Henties	22°07'S, 14°16'E	museum
	Kalkfeld	20°54'S, 16°11'E	museum
	Kannabeam	28°07S, 17°33'E	museum
	Karibib	21°57'S, 15°51'E	museum
	Keetmanshoop	26°21'S, 18°29'E	genotyped
	Koegas	29°18'S, 22°21'E	museum
	Koes	25°56'S, 19°07'E	museum
	Kuruman	27°28'S,23°26'E	museum
	Louisville	28°34'S, 21°11'E	museum
	Maltahohe	24°50'S, 16°59'E	museum
	Mariental	24°34'S, 18°02'E	genotyped
	Mata mata	25°49'S, 20°01'E	museum
	Namibrand	25°01'S, 15°55'E	museum
	Narais	23°07'S, 16°53'E	genotyped
	Nossob	25°22'S, 20°25'E	museum
	Okahandja	21°59'S, 16°54'E	museum
	Otjiomongombe	21°35'S, 16°56'E	genotyped

Table B.5 continued

Clade	Locality	Geographic coordinates	Origin
R. pumilio Northern	Prieska	29°40'S, 22°44'E	museum
	Rehoboth	23°20'S, 17°05'E	museum
	Rooipoort	28°39'S, 24°08'E	genotyped
	Sandveld	27°40'S, 25°41'E	genotyped
	Solitaire	23°53'S, 16°00'E	museum
	Sossusvlei	24°44'S, 15°19'E	museum
	Spitzkoppe	21°49'S, 15°11'E	museum
	Stolzenfels	28°30'S, 19°41'E	museum
	Swakopmund	22°41'S, 14°32'E	genotyped
	Tsaris	24°50'S, 16°19'E	museum
	Tussen die Rivieren	30°28'S, 26°09'E	genotyped
	Twee Rivieren	26°30'S, 20°37'E	genotyped
	Walvis bay	22°59'S, 14°29'E	museum
	Warmfontein	27°05'S, 19°15'E	museum
	Wilhelmstal	21°55'S, 16°19'E	museum
	Windhoek	22°31'S, 17°25'E	genotyped

Table B.6: Log_{10} Bayes Factors for alternative partitioning schemes of COI (unpartitioned, partitioned by codon, and partitioned by codon with individual models assigned to each partition) for the individual and combined MrBayes analyses respectively. Standard errors were estimated with 1000 bootstrap replicates.

Model	lnP	S.E.	Log_{10} Bayes Factor		
			Unpartitioned	Partitioned	Partitioned + individual
Individual					
Unpartitioned	-5879	0.288	-	-279.96	-71.227
Partitioned	-5234.4	0.407	**279.96**	-	208.734
Partitioned + individual	-5715	0.273	71.227	-208.734	-
All combined					
Unpartitioned	-9956	0.138	-	-0.799	-193.38
Partitioned	-9954.2	0.151	0.799	-	-192.581
Partitioned + individual	-9510.8	0.114	**193.38**	192.581	-

Table B.7: Average HKY-corrected sequence distances (%), with standard deviation in brackets, among the clades and subclades of *Rhabdomys*.

R. pumilio	*R. pumilio* Coastal	Central	Northern
Coastal	**-**		
Central	11.6 (±0.2)	**-**	
Northern	13.5 (±0.3)	7.9 (±0.3)	**-**
R. dilectus	9.5 (±0.7)	10.7 (±0.4)	12.3 (±0.6)

	R. d. chakae
R. d. dilectus	5.6 (±0.1)

	R. pumilio **Coastal B**
R. pumilio **Coastal A**	4.9 (±0.2)

Table B.8: COI haplotypes identified from the 299 *Polyplax arvicanthis* specimens sampled from 16 localities (codes as in Table 3.1). GenBank accession numbers and haplotypes selected for the representative subset (sequenced for additional markers) are also indicated. Subscripts indicate *P. arvicanthis* clade 1 or 2.

Haplotype	Accession	Frequency	Subset	Individuals
H01_1	JX 198378	52	GP_1	RV1_1, 3_1, 4_1, 6_1, 7_1, 10_1, 11_1, 12_1, 13_1, 18_1, 20_1, 21_1, 23_1, 24_1, 30_1; GP3_1, 4_1, 5_1, 9_1, 10_1,12_1, 13_1, 14_1, 16_1, 17_1, 20_1, 22_1, 24_1, 27_1; VR1_1, 2_1, 4_1, 5_1, 6_1, 8_1, 9_1, 17_1, 21_1, 23_1, 24_1, 27_1; BW10.2_1; LB2_1; SL3.6_1, 4.1_1, 4.2_1, 4.3_1, 4.4_1, 4.5_1, 4.6_1, 6.1_1, 8.2_1;
H02_1	JX 198372	1	RV_1	RV14_1
H03_1	JX 198377	1	RP_1	RP12.2_1
H04_1	JX 629372	1		GP1_1
H05_1	JX 629373	4		GP2_1, 6_1, 8_1, 28_1
H06_1	JX 629374	1		GP7_1
H07_1	JX 629375	1		GP23_1
H08_1	JX 198380	12	PV_1	PV4_1, 5_1, 9_1, 10_1, 14_1, 17_1, 20_1, 23_1, 15.1_1, 29_1, 30_1; DH3.2_1,
H09_1	JX 629376	1		PV18_1
H10_1	JX 629377	1		PV26_1
H11_1	JX 198382	20	OH_1	OH3_1, 5_1, 6_1, 9_1, 10_1, 12_1, 13_1, 15_1, 16_1, 17_1, 18_1, 19_1, 21_1, 22_1, 25_1, 26_1, 27_1, 30_1, 31_1, 32_1
H12_1	JX 629378	1		OH4_1
H13_1	JX 198384	1	VR_1	VR7_1
H14_1	JX 629379	1		VR19_1
H15_1	JX 629380	2		VR22_1 , 26_1
H16_1	JX 629381	1		BW1_1
H17_1	JX 629382	1		BW8_1
H18_1	JX 198387	5	BW_1	BW9_1, 10.1_1, 10.4_1, 18_1, 19_1
H19_1	JX 629383	1		BW17.1_1
H20_1	JX 629384	1		BW21_1
H21_1	JX 629385	1		BW28.1_1
H22_1	JX 629386	1		BW28.2_1
H23_1	JX 629387	1		BW32_1
H24_1	JX 198389	1	LB_1	LB1.3_1
H25_1	JX 629388	1		WH3.4_1
H26_1	JX 629389	12		WH4_1, 5_1, 6_1, 8_1, 9_1, 10_1, 12_1, 13_1, 16_1, 17_1, 18_1, 21_1
H27_1	JX 629390	3		WH7_1, 19_1, 23_1
H28_1	JX 198391	1	WH_1	WH11_1
H29_1	JX 198393	2	SL_1	SL3.1_1, 3.2_1

Table B.8 continued

Haplotype	Accession	Frequency	Subset	Individuals
H30_1	JX 629391	2		SL3.3_1, 7.3_1
H31_1	JX 629392	1		SL8.1_1
H32_1	JX 629393	1		SL8.3_1
H33_1	JX 629394	4		DH2_1, 5.1_1, 5.2_1, 5.3_1
H34_1	JX 198394	1	DH_1	DH3.1_1
H35_1	JX 629395	1		DH3.5_1
H36_1	JX 629396	1		DH5.5_1
H37_1	JX 629397	1		DH5.6_1
H40_1	JX 629398	1		CH6.1_1
H41_1	JX 198397	1	CH_1	CH7_1
H42_1	JX 629399	2		CH14.1_1, 14.2_1
H01_2	JX 629400	3		RV2_2, 19_2, 28_2
H02_2	JX 629401	5		RV8_2, 9_2, 15_2, 16_2, 17_2
H03_2	JX 198373	1	RV_2	RV25_2
H04_2	JX 629402	1		RV26_2
H05_2	JX 629403	1		RV29_2
H06_2	JX 629404	1		GH1.2_2
H07_2	JX 198376	39	RP_2	GH2.1_2, 2.2_2, 2.3_2, 4.2_2, 4.3_2, 4.4_2, 4.6_2, 5.1_2, 5.2_2, 8.1_2, 8.2_2, 9.1_2, 14.1_2, 14.3_2, 14.4_2; RP1.1_2, 1.2_2, 3.2_2, 3.1_2, 4.1_2, 4.2_2, 4.2_2, 5.1_2, 5.2_2, 6.1_2, 7.1_2, 7.2_2, 9.1_2, 9.2_2, 9.3_2, 11.1_2,12.1_2, 12.3_2, 12.5_2, 12.7_2, 12.8_2, 12.9_2, 12.10_2, 14.1_2, 15.1_2
H08_2	JX 198374	1	GH_2	GH4.5_2
H09_2	JX 629405	1		GH10_2
H10_2	JX 198375	4	KH_2	KH1_2, 3_2, 6_2, 14.2_2
H11_2	JX 629406	15		KH4.1_2, 4.2_2, 4.3_2, 7_2, 11_2, 15_2, 16_2, 18.1_2, 18.2_2, 20.1_2, 20.3_2, 21.1_2, 21.2_2, 21.3_2; WH14_2
H12_2	JX 629407	2		KH14.1_2, 20.2_2
H13_2	JX 629408	3		RP11.3_2, 12.4_2, 12.6_2
H14_2	JX 198379	1	GP_2	GP11_2
H15_2	JX 629409	1		GP15_2
H16_2	JX 629410	1		GP18_2
H17_2	JX 629411	3		GP19_2, 21_2, 30_2
H18_2	JX 629412	1		GP26_2
H19_2	JX 629413	1		PV1_2
H20_2	JX 198381	16	PV_2	PV2_2, 7_2, 21_2, 24_2, 27_2; SB2.2_2, 2.4_2, 4.1_2, 4.2_2, 5_2, 8_2, 14_2, 18_2, 22_2, 24_2, 26_2
H21_2	JX 629414	3		OH1_2, 2_2, 20_2
H22_2	JX 629415	1		OH8_2
H23_2	JX 198386	2	OH_2	OH11_2, 33_2

Table B.8 continued

Haplotype	Accession	Frequency	Subset	Individuals
H24_2	JX 629416	1		OH14_2
H25_2	JX 629417	1		OH24_2
H26_2	JX 198385	4	VR_2	VR3_2, 11_2, 16_2, 29_2
H27_2	JX 629418	1		VR12_2
H28_2	JX 629419	1		VR15_2
H29_2	JX 629420	1		VR18_2
H30_2	JX 629421	1		SB2.1_2
H31_2	JX 629422	1		SB12_2
H32_2	JX 198386	1	SB_2	SB25_2
H33_2	JX 629423	1		SB30_2
H34_2	JX 629424	3		BW7_2, 10.5_2, 10.6_2
H35_2	JX 198388	1	BW_2	BW20_2
H36_2	JX 629425	1		BW22_2
H37_2	JX 629426	1		BW26_2
H38_2	JX 629427	1		BW27_2
H39_2	JX 629428	1		LB9.1_2
H40_2	JX 198390	2	LB_2	LB9.3_2 , 14.1_2
H41_2	JX 629429	1		LB10_2
H42_2	JX 629430	1		LB12.1_2
H43_2	JX 629431	1		LB12.2_2
H44_2	JX 629432	1		LB12.4_2
H45_2	JX 629433	1		LB17_2
H46_2	JX 198392	2	WH_2	WH3.1_2, 3.3_2
H47_2	JX 629434	2		SL6.2_2, 13.2_2
H48_2	JX 198395	2	DH_2	DH4.1_2, 8_2
H49_2	JX 629435	2		DH14.2_2, 14.3_2
H66_2	JX 629436	1		CH2_2
H67_2	JX 198396	3	CH_2	CH6.2_2, 12_2, 8_2
H68_2	JX 629437	2		AL 4.2, 5.1
H69_2	JX 629438	1		AL5.2
H70_2	JX 198398	1	AL_2	AL4.1

Table B.9: GenBank accession numbers for the gene sequences generated for the subset of specimens, with locality codes as in Table 3.1 and subscripts indicating *P. arvicanthis* clade 1 or 2.

Specimen	16S	12S	CAD
RV_1	JX198319	JX198346	JX198399
RV_2	JX198320	JX198347	JX198400
GH_2	JX198321	JX198348	JX198401
KH_2	JX198322	JX198349	JX198402
RP_2	JX198323	JX198350	JX198403
RP_1	JX198324	JX198351	JX198404
GP_1	JX198325	JX198352	JX198405
GP_2	JX198326	JX198353	JX198406
PV_1	JX198327	JX198354	JX198407
PV_2	JX198328	JX198355	JX198408
OH_1	JX198329	JX198356	JX198409
OH_2	JX198330	JX198357	JX198410
VR_1	JX198331	JX198358	JX198411
VR_2	JX198332	JX198359	JX198412
SB_2	JX198333	JX198360	JX198413
BW_1	JX198334	JX198361	JX198414
BW_2	JX198335	JX198362	JX198415
LB_1	JX198336	JX198363	JX198416
LB_2	JX198337	no data	JX198417
WH_1	JX198338	JX198364	JX198418
WH_2	JX198339	JX198365	JX198419
SL_1	JX198340	JX198366	JX198420
DH_1	JX198341	JX198367	JX198421
DH_2	JX198342	JX198368	JX198422
CH_2	JX198343	JX198369	JX198423
CH_1	JX198344	JX198370	JX198424
AL_2	JX198345	JX198371	JX198425

Table B.10: GTR-corrected COI sequence distances among recognized *Polyplax* species, with GenBank accession numbers.

	P. spinulosa	*P. auricularis*	*P. borealis*	*P. serrata*
Accession	HQ542196.1	DQ324549.1	DQ324548.1	EU162264.1
P. spinulosa	-			
P. auricularis	0.28	-		
P. borealis	0.23	0.28	-	
P. serrata	0.27	0.29	0.32	-

Table B.11: Log_{10} Bayes Factors for alternative partitioning schemes (unpartitioned and partitioned by codon) of the protein coding COI and CAD genes. Standard errors were estimated with 1000 bootstrap replicates.

Gene	lnP	S.E.	Partitioned	Unpartitioned
COI				
Partitioned	-2442.79	0.132	-	**38.206**
Unpartitioned	-2530.76	0.126	-38.206	-
CAD				
Partitioned	-898.67	0.201	-	**35.869**
Unpartitioned	-981.262	0.296	-35.869	-

Table B.12: Mitochondrial COI haplotypes obtained from the various host taxa sampled (locality codes as in Table 4.1) with GenBank accession numbers. Nuclear Eef1a1 sequences generated for the COI haplotypes are also indicated (including accession numbers) with multiple individuals sequenced per haplotype indicated in bold.

Taxon	Haplotype	COI Accession	Eef1a1 Accession	Frequency	Individuals
R. pumilio	H001	JQ003320	JQ003301	29	RV1,2,3,4,5,6,7,8,9,10,11,13,15,14, 16,17,18,19,20,21,22, 23,24,25,26,27,28,29,31
R. pumilio	H002	JQ003321	KC296593	1	RV12
R. pumilio	H003	JQ003322	no data	1	RV30
R. bechuanae	H004	JQ003323	JQ003302	1	KH7
R. bechuanae	H007	JQ003326	JQ003303	4	KH1,2,6,14
R. bechuanae	H008	JQ003327	KC296594	3	KH3,4,15
R. bechuanae	H009	JQ003328	KC296595	2	KH5, 12
R. bechuanae	H010	JQ003329	KC296596	6	KH8,9,10,13,17,19
R. bechuanae	H011	JQ003330	KC296597	1	KH11
R. bechuanae	H012	JQ003331	KC296598	1	KH16
R. bechuanae	H013	JQ003332	KC296599	1	KH18
R. bechuanae	H014	JQ003333	no data	1	KH20
R. bechuanae	H015	JQ003334	KC296600	1	KH21
R. pumilio	H016	JQ003335	KC296601	16	GP1,2,3,5,9,10,11,13,14,15,18,21,2 5,26,28,30
R. pumilio	H017	JQ003336	KC296602	1	GP4
R. pumilio	H018	JQ003337	JQ003304	14	**GP6**,8,17,19,20,27,29; **VR1**,3,4,8,16,23,24
R. pumilio	H019	JQ003338	KC296603	1	GP7
R. pumilio	H020	JQ003339	KC296604	1	GP12
R. pumilio	H021	JQ003340	KC296605	2	GP16,22
R. pumilio	H022	JQ003341	KC296606	1	GP23
R. pumilio	H023	JQ003342	KC296607	1	GP24
R. pumilio	H024	JQ003343	KC296608	1	OH1
R. pumilio	H025	JQ003344	KC296609	4	OH2,6,11,16
R. pumilio	H026	JQ003345	KC296610	12	**OH3**,4,5; **DH2**,3,5,7,8,12,13,18,19
R. pumilio	H027	JQ003346	KC296611	17	OH7,8,17,18,19,20,21,22,24,25,26, 27,28,29,31, 32,33
R. pumilio	H028	JQ003347	KC296612	1	OH9
R. pumilio	H029	JQ003348	no data	1	OH10
R. pumilio	H030	JQ003349	KC296613	1	OH12
R. pumilio	H031	JQ003350	KC296614	1	OH14
R. pumilio	H032	JQ003351	KC296615	1	OH15,23
R. pumilio	H033	JQ003352	KC296616	1	OH30
R. pumilio	H034	JQ003353	JQ003305	8	PV1,3,9,11,14,17,23,26
R. pumilio	H035	JQ003354	no data	1	PV2

Table B.12 continued

Taxon	Haplotype	COI Accession	Eef1a1 Accession	Frequency	Individuals
R. pumilio	H036	JQ003355	KC296617	13	PV4,6,12,13,15,18,21,25,27,28,29, 30,31
R. pumilio	H037	JQ003356	KC296618	1	PV5
R. pumilio	H038	JQ003357	KC296619	5	**PV**7,**20**,24; **MV4**,19
R. pumilio	H039	JQ003358	KC296620	4	PV8; MV2,12,29
R. pumilio	H040	JQ003359	KC296621	2	PV16,22
R. pumilio	H041	JQ003360	no data	1	PV19
R. pumilio	H042	JQ003361	KC296622	1	VR2
R. pumilio	H043	JQ003362	KC296623	1	VR5
R. pumilio	H044	JQ003363	KC296624	4	VR6,7,15,20
R. pumilio	H045	JQ003364	KC296625	6	VR9,18,25,26,28,30
R. pumilio	H046	JQ003365	KC296626	1	VR10
R. pumilio	H047	JQ003366	KC296627	4	VR11,12,17,27
R. pumilio	H048	JQ003367	KC296628	1	VR13
R. pumilio	H049	JQ003368	KC296629	1	VR14
R. pumilio	H050	JQ003369	KC296630	2	VR19,22
R. pumilio	H051	JQ003370	KC296631	1	VR21
R. pumilio	H052	JQ003371	KC296632	1	VR29
R. intermedius	H053	JQ003372	no data	1	FBsk1
R. pumilio	H054	JQ003373	KC296633	7	FBsk2,3,5,6,7,8,10
R. intermedius	H055	JQ003374	JQ003306	41	**FBsk4**; **BW1**,2,3,4,5,7,8,9,10,11,12,13,14,1 5,16,17,18,19,20,21, 22,23,24,26,29,30,31,32,33; **LB4**,12,13,14,15,17; **SL**1,3,**6**,8,12
R. pumilio	H056	JQ003375	no data	3	FBsk9, FBwb11, 14
R. dilectus	H057	JQ003376	JQ003307	1	FBsk11
R. dilectus	H058	JQ003377	JQ003308	1	FBsk12
R. pumilio	H095	JQ003414	KC296634	2	DH1,10
R. pumilio	H096	JQ003415	KC296635	1	DH4
R. pumilio	H097	JQ003416	KC296636	1	DH6
R. pumilio	H098	JQ003417	KC296637	3	DH9,11,15
R. pumilio	H099	JQ003418	KC296638	2	DH14,16
R. pumilio	H100	JQ003419	KC296639	1	DH20
R. pumilio	H101	JQ003420	JQ003310	14	MV1,3,15,16,17,18,20,21,22,23,27, 28,30,31
R. pumilio	H102	JQ003421	KC296640	1	MV5
R. pumilio	H103	JQ003422	KC296641	1	MV6
R. pumilio	H104	JQ003423	KC296642	1	MV7
R. pumilio	H105	JQ003424	KC296643	9	MV8,9,10,11,13,14,24,25,26
R. intermedius	H108	JQ003427	KC296644	2	BW6; SL11
R. intermedius	H119	JQ003438	KC296645	1	BW25

Table B.12 continued

Taxon	Haplotype	COI Accession	Eef1a1 Accession	Frequency	Individuals
R. intermedius	H120	JQ003439	JQ003317	12	**BW27**;**LB1**,3,7,8,9,10,11,16,22,23; **SL10**
R. intermedius	H121	JQ003440	KC296646	1	BW28
R. intermedius	H129	JQ003448	JQ003318	7	LB2,5,6,18,19,20,21
R. intermedius	H130	JQ003449	KC296647	3	SL2,5,9
R. intermedius	H131	JQ003450	KC296648	1	SL4
R. intermedius	H132	JQ003451	KC296649	1	SL7
R. intermedius	H133	JQ003452	KC296650	1	SL13
R. bechuanae	H134	JQ003453	no data	15	RP1,2,3,4,5,6,7,8,9,10,11,12,13,14,15
R. bechuanae	H142	JQ003461	KC296651	8	GH1,4,6,8,9,11,13,14
R. bechuanae	H143	JQ003462	KC296652	5	GH2,3,7,10,12
R. bechuanae	H144	JQ003463	no data	1	GH5
R. bechuanae	H147	JQ003466	KC296653	8	WH3,6,9,12,13,17,19,21
R. bechuanae	H148	JQ003467	KC296654	4	WH4,7,14,16
R. bechuanae	H149	JQ003468	KC296655	6	WH5,10,11,15,18,22
R. bechuanae	H150	JQ003469	KC296656	1	WH20
R. bechuanae	H151	JQ003470	no data	1	WH23
R. pumilio	H152	KC296581	no data	1	FBsk13
R. pumilio	H153	KC296582	no data	1	FBsk14
R. pumilio	H154	KC296583	no data	1	FBwb1
R. pumilio	H155	KC296584	no data	3	FBwb2, 3, 5
R. dilectus	H156	KC296585	no data	1	FBwb4
R. dilectus	H157	KC296586	no data	8	FBwb6, 7, 9, 10, 12, 13, 16; CH12
R. dilectus	H158	KC296587	no data	1	FBwb15
R. dilectus	H159	KC296588	no data	1	CH1
R. dilectus	H160	KC296589	no data	1	CH2
R. dilectus	H161	KC296590	no data	1	CH6
R. dilectus	H162	KC296591	no data	2	CH7, 8
R. dilectus	H163	KC296592	no data	1	CH14

Table B.13: Mitochondrial COI Haplotypes identified within the two *Polyplax arvicanthis* lineage from the various sampled localities (codes as in Table 4.1) with GenBank accession numbers. Haplotypes sequenced for the nuclear CAD gene are also indicated (including GenBank accessions) and multiple individuals sequenced per haplotype are indicated in bold.

Taxon	Haplotype	COI Accession	CAD Accession	Frequency	Individuals
P. arvicanthis 1	H01_1	JX 198378	JX198405	52	**RV**1_1, **3_1,** 4_1, 6_1, 7_1, 10_1, 11_1, 12_1, 13_1, 18_1, 20_1, 21_1, 23_1, 24_1, 30_1; **GP**3_1, 4_1, 5_1, 9_1, **10_1**, 12_1, 13_1, 14_1, 16_1, 17_1, 20_1, 22_1, 24_1, 27_1;
					VR1_1, 2_1, 4_1, 5_1, 6_1, 8_1, 9_1, 17_1, 21_1, 23_1, 24_1, 27_1; **BW10.2_1**; **LB2_1**; **SL**3.6_1, 4.1_1, 4.2_1, 4.3_1, 4.4_1, 4.5_1, 4.6_1, 6.1_1, **8.2_1**;
	H02_1	JX 198372	JX198399	1	RV14_1
	H03_1	JX 198377	JX198404	1	RP12.2_1
	H04_1	JX 629372	KC296502	1	GP1_1
	H05_1	JX 629373	KC296503	4	GP2_1, 6_1, 8_1, 28_1
	H06_1	JX 629374	KC296504	1	GP7_1
	H07_1	JX 629375	KC296505	1	GP23_1
	H08_1	JX 198380	JX198407	12	**PV4_1**, 5_1, 9_1, 10_1, 14_1, 17_1, 20_1, 23_1, 29_1, 30_1; **DH3.2_1**, 15.1_1
	H09_1	JX 629376	KC296506	1	PV18_1
	H10_1	JX 629377	KC296507	1	PV26_1
	H11_1	JX 198382	JX198409	20	OH3_1, 5_1, 6_1, 9_1, 10_1, 12_1, 13_1, 15_1, 16_1, 17_1, 18_1, 19_1, 21_1, 22_1, 25_1, 26_1, 27_1, 30_1, 31_1, 32_1
	H12_1	JX 629378	KC198410	1	OH4_1
	H13_1	JX 198384	JX198411	1	VR7_1
	H14_1	JX 629379	KC296509	1	VR19_1
	H15_1	JX 629380	KC296510	2	VR22_1 , 26_1

Table B.13 continued

Taxon	Haplotype	COI Accession	CAD Accession	Frequency	Individuals
P. arvicanthis 1	H16_1	JX 629381	KC296511	1	BW1_1
	H17_1	JX 629382	KC296512	1	BW8_1
	H18_1	JX 198387	JX198414	5	BW9_1, 10.1_1, 10.4_1, 18_1, 19_1
	H19_1	JX 629383	KC296513	1	BW17.1_1
	H20_1	JX 629384	KC296514	1	BW21_1
	H21_1	JX 629385	KC296515	1	BW28.1_1
	H22_1	JX 629386	KC296516	1	BW28.2_1
	H23_1	JX 629387	KC296517	1	BW32_1
	H24_1	JX 198389	JX198416	1	LB1.3_1
	H25_1	JX 629388	KC296518	1	WH3.4_1
	H26_1	JX 629389	KC296519	12	WH4_1, 5_1, 6_1, 8_1, 9_1, 10_1, 12_1, 13_1, 16_1, 17_1, 18_1, 21_1
	H27_1	JX 629390	KC296520	3	WH7_1, 19_1, 23_1
	H28_1	JX 198391	JX198418	1	WH11_1
	H29_1	JX 198393	JX198420	2	SL3.1_1, 3.2_1
	H30_1	JX 629391	no data	2	SL3.3_1, 7.3_1
	H31_1	JX 629392	KC296521	1	SL8.1_1
	H32_1	JX 629393	KC296522	1	SL8.3_1
	H33_1	JX 629394	KC296523	4	DH2_1, 5.1_1, 5.2_1, 5.3_1
	H34_1	JX 198394	JX198421	1	DH3.1_1
	H35_1	JX 629395	KC296524	1	DH3.5_1
	H36_1	JX 629396	KC296525	1	DH5.5_1
	H37_1	JX 629397	KC296526	1	DH5.6_1
	H38_1	KC296482	KC296527	2	FBwb9.1_1, 11_1
	H39_1	KC296483	KC296528	1	FBwb16.2_1
	H40_1	JX 629398	no data	1	CH6.1_1
	H41_1	JX 198397	JX198424	1	CH7_1
	H42_1	JX 629399	KC296529	2	CH14.1_1, 14.2_1

Table B.13 continued

Taxon	Haplotype	COI Accession	CAD Accession	Frequency	Individuals
P. arvicanthis 2	H01_2	JX 629400	KC296530	3	RV2_2, 19_2, 28_2
	H02_2	JX 629401	KC296531	5	RV8_2, 9_2, 15_2, 16_2, 17_2
	H03_2	JX 198373	JX198400	1	RV25_2
	H04_2	JX 629402	KC296532	1	RV26_2
	H05_2	JX 629403	KC296533	1	RV29_2
	H06_2	JX 629404	KC296534	1	GH1.2_2
	H07_2	JX 198376	JX198403	39	**GH2.1_2**, 2.2_2, 2.3_2, 4.2_2, 4.3_2, 4.4_2, 4.6_2, 5.1_2, 5.2_2, 8.1_2, 8.2_2, 9.1_2, 14.1_2, 14.3_2, 14.4_2; **RP1.1_2**, 1.2_2, 3.2_2, 3.1_2, 4.1_2, 4.2_2, 5.1_2, 5.2_2, 6.1_2, 7.1_2, 7.2_2, 9.1_2, 9.2_2, 9.3_2, 11.1_2, 12.1_2, 12.3_2, 12.5_2, 12.7_2, 12.8_2, 12.9_2, 12.10_2, 14.1_2, 15.1_2
	H08_2	JX 198374	JX198401	1	GH4.5_2
	H09_2	JX 629405	KC296535	1	GH10_2
	H10_2	JX 198375	JX198402	4	KH1_2, 3_2, 6_2, 14.2_2
	H11_2	JX 629406	KC296536	15	KH4.1_2, 4.2_2, 4.3_2, 7_2, 11_2, 15_2, 16_2, 18.1_2, 18.2_2, 20.1_2, 20.3_2, 21.1_2, 21.2_2, 21.3_2; WH14_2
	H12_2	JX 629407	KC296537	2	KH14.1_2, 20.2_2
	H13_2	JX 629408	KC296538	3	RP11.3_2, 12.4_2, 12.6_2
	H14_2	JX 198379	JX198406	1	GP11_2
	H15_2	JX 629409	KC296539	1	GP15_2
	H16_2	JX 629410	no data	1	GP18_2
	H17_2	JX 629411	KC296540	3	GP19_2, 21_2, 30_2
	H18_2	JX 629412	KC296541	1	GP26_2
	H19_2	JX 629413	KC296542	1	PV1_2
	H20_2	JX 198381	JX198408	18	**PV**2_2, 7_2, 21_2, 24_2, **27_2**; **MV2.2_2**, 2.4_2, 4.1_2, 4.2_2, 5_2, 8_2, 14_2, 18_2, 22_2, 24_2, 26_2; FBsk3.3_2, 8.3_2

Table B.13 continued

Taxon	Haplotype	COI Accession	CAD Accession	Frequency	Individuals
P. arvicanthis 2	H21_2	JX 629414	KC296543	3	OH1_2, 2_2, 20_2
	H22_2	JX 629415	KC296544	1	OH8_2
	H23_2	JX 198386	JX198410	2	OH11_2, 33_2
	H24_2	JX 629416	KC296545	1	OH14_2
	H25_2	JX 629417	KC296546	1	OH24_2
	H26_2	JX 198385	JX198412	4	VR3_2, 11_2, 16_2, 29_2
	H27_2	JX 629418	KC296547	1	VR12_2
	H28_2	JX 629419	KC296548	1	VR15_2
	H29_2	JX 629420	KC296549	1	VR18_2
	H30_2	JX 629421	KC296550	1	MV2.1_2
	H31_2	JX 629422	KC296551	1	MV12_2
	H32_2	JX 198386	JX198413	1	MV25_2
	H33_2	JX 629423	KC296552	1	MV30_2
	H34_2	JX 629424	KC296553	3	BW7_2, 10.5_2, 10.6_2
	H35_2	JX 198388	JX198415	1	BW20_2
	H36_2	JX 629425	KC296554	1	BW22_2
	H37_2	JX 629426	KC296555	1	BW26_2
	H38_2	JX 629427	KC296556	1	BW27_2
	H39_2	JX 629428	KC296557	1	LB9.1_2
	H40_2	JX 198390	JX198417	2	LB9.3_2 , 14.1_2
	H41_2	JX 629429	KC296558	1	LB10_2
	H42_2	JX 629430	KC296559	1	LB12.1_2
	H43_2	JX 629431	KC296560	1	LB12.2_2
	H44_2	JX 629432	KC296561	1	LB12.4_2
	H45_2	JX 629433	KC296562	1	LB17_2
	H46_2	JX 198392	JX198419	2	WH3.1_2, 3.3_2
	H47_2	JX 629434	KC296563	2	SL6.2_2, 13.2_2

Table B.13 continued

Taxon	Haplotype	COI Accession	CAD Accession	Frequency	Individuals
P. arvicanthis 2	H48_2	JX 198395	JX198422	2	DH4.1_2, 8_2
	H49_2	JX 629435	KC296564	2	DH14.2_2, 14.3_2
	H50_2	KC296484	KC296565	1	FBsk1.3_2
	H51_2	KC296485	KC296566	1	FBsk2_2
	H52_2	KC296486	KC296567	10	FBsk3.1_2; FBwb2_2, 3_2, 4.1_2, 4.2_2, 5_2, 7_2, 12.2_2, 13.1_2, 13.2_2
	H53_2	KC296487	KC296568	1	FBsk3.2_2
	H54_2	KC296488	KC296569	2	FBsk3.4_2, 7.3_2
	H55_2	KC296489	no data	1	FBsk7.1_2
	H56_2	KC296490	KC296570	1	FBsk7.2_2
	H57_2	KC296491	KC296571	1	FBsk8.2_2
	H58_2	KC296492	KC296572	1	FBsk8.4_2
	H59_2	KC296493	KC296573	1	FBsk9.2_2
	H60_2	KC296494	KC296574	1	FBsk9.3_2
	H61_2	KC296495	KC296575	1	FBsk10.2_2
	H62_2	KC296496	KC296576	2	FBsk10.3_2, 10.4_2
	H63_2	KC296497	KC296577	1	FBwb 6.1_2
	H64_2	KC296498	KC296578	1	FBwb16.1_2
	H65_2	KC296499	KC296579	1	FBsk13_2
	H66_2	JX 629436	no data	1	CH2_2
	H67_2	JX 198396	KC296580	3	CH6.2_2, 12_2, 8_2

Table B.14: Primers used for PCR amplification of parasite and host mitochondrial and nuclear genes.

Taxon	Gene	Primer	F/R	Sequence (5'-3')	Annealing temperature (^{O}C)
Parasites	COI	LCOIP6625[1]	F	CCCGATCCTTTTGGTTTTTTGGGCATCC	45-58/59^0 C
		COIPRev[2]	R	CCYCCTACNGTAAAYAGRAARATRAARC	
	CAD	CADPfor[3]	F	ACGACAACTGCATTACCGTTTGCA	55^{0}C
		CADPrev[4]	R	CCACCGGGGAATTTTGACAAC	
Host	COI	COIL1490[5]	F	GGTCAACAAATCATAAAGATATTGG	45
		COIH7005[6]	R	CCGGATCCACNACRTARTANGTRTCRTG	
	EF1α1	EF1-R-For3[7]	F	GGGGACAAYGTTGGTTTCAACG	55-56
		Cho 10[8]	R	ACGGCVACKGTYTGHCKCATGTC	

[1]Adapted from L6625, Hafner *et al*. 1994; [2]Designed in this study; [3]Adapted from ApCADfor1, Danforth *et al.* 2006; [4]Adapted from Ap835rev1, Danforth *et al.* 2006; [5]Folmer *et al*. 1994; [6]Anderson *et al*., 1981; [7]Adapted from EF1-For3, Danforth & Ji, 1998; [8]Danforth & Ji, 1998

Table B.15: Models of evolution specified for the various codon positions in the mtDNA Bayesian analyses of all haplotypes and the reduced datasets (clusters) used as input in Jane.

Analysis	Taxon	Codon		
		1	**2**	**3**
All haplotypes	*Rhabdomys*	JC	JC	HKY+G
	Polyplax	GTR+I+G	GTR+I+G	GTR+I
Clusters	*Rhabdomys*	JC	JC	GTR
	Polyplax	GTR+G	JC	GTR

Table B.16: The distinct genetic clusters identified from BAPS within the host and two parasite taxa. Matching superscript numbers indicate mixture detected among the specified clusters.

Clusters	Localities
Rhabdomys	
R. intermedius	SL, LB, BW[2]
R. bechuanae	WH, KH, GH, RP
R. pumilio A	RV, GP, VR, PV, SB, DH, OH
R. pumilio B	FBsk[1,2]
R. dilectus	CH, FBwb[1]
P. arvicanthis 1	
1	RP, WH
2	OH
3	PV, DH
4	RV, GP, VR, BW, LB, SL
5	FBwb, CH
P. arvicanthis 2	
1	RV, GP[3]
2	VR, BW, LB, SL[3]
3	WH, KH, GH, RP
4	FBwb1, CH
5	PV, SB, FBsk1
6	DH
7	OH

Table B.17: Pairwise Φst values among *Rhabdomys* sampled localities with COI above and Eef1a1 below the diagonal. Dash marks indicate localities for which Eef1a1 data was not available. Statistically significant values (p <0.05) are indicated in bold and values remaining significant after Bonferroni correction are indicated with asterisks.

	WH	KH	GH	RP	RV	GP	VR	PV	SB	DH	OH	SL	LB	BW	FB_sk	FB_wb	CH
WH		0.58*	0.78*	0.87*	0.99*	0.98*	0.97*	0.97*	0.97*	0.98*	0.97*	0.97*	0.97*	0.98*	0.79*	0.79*	0.97*
KH	0.24*		0.69*	0.80*	0.98*	0.97*	0.97*	0.97*	0.97*	0.97*	0.97*	0.96*	0.96*	0.98*	0.79*	0.79*	0.97*
GH	0.19	0.07		0.76*	0.98*	0.97*	0.97*	0.97*	0.97*	0.97*	0.96*	0.96*	0.96*	0.98*	0.74*	0.75*	0.96*
RP	-	-	-		0.99*	0.98*	0.98*	0.98*	0.98*	0.98*	0.97*	0.98*	0.98*	0.99*	0.76*	0.77*	0.98*
RV	0.37	0.33*	0.35	0.99*		0.37*	0.44*	0.76*	0.59*	0.73*	0.59*	0.98*	0.98*	0.99*	0.70*	0.74*	0.98*
GP	0.36*	0.31*	0.27*	0.98*	0.06		0.34*	0.70*	0.55*	0.62*	0.54*	0.97*	0.97*	0.98*	0.67*	0.72*	0.96*
VR	0.40*	0.32*	0.31*	0.98*	0.02	0.09*		0.69*	0.55*	0.62*	0.54*	0.97*	0.97*	0.98*	0.66*	0.72*	0.96*
PV	0.35*	0.29*	0.25*	0.98*	0.00	0.03	0.10		0.49*	0.64*	0.60*	0.97*	0.97*	0.98*	0.67*	0.72*	0.96*
SB	0.37*	0.31*	0.27	0.98*	0.00	0.04	0.07	0.00		0.43*	0.41*	0.97*	0.97*	0.98*	0.67*	0.72*	0.95*
DH	0.50*	0.45*	0.48*	0.98*	0.10	0.08	0.03	0.17	0.09		0.25*	0.97*	0.97*	0.98*	0.62*	0.67*	0.96*
OH	0.42*	0.36*	0.35*	0.97*	0.00	0.09*	0.12*	0.00	0.03	0.18*		0.96*	0.97*	0.98*	0.67*	0.72*	0.95*
SL	0.48*	0.38*	0.46*	0.98*	0.08	0.23*	0.30*	0.07	0.17	0.42*	0.10*		0.09	0.17*	0.70*	0.73*	0.96*
LB	0.49*	0.43*	0.55*	0.98*	0.41*	0.21	0.38*	0.14	0.25	0.49*	0.31*	0.31*		0.32*	0.76*	0.78*	0.96*
BW	0.32*	0.27*	0.23*	0.99*	0.00	0.02	0.03	0.00	0.00	0.11	0.00	0.10	0.26*		0.81*	0.83*	0.98*
FB_sk	0.33*	0.24*	0.23*	0.76*	0.00	0.15*	0.10	0.06	0.04	0.27*	0.07	0.23*	0.38*	0.00		0.18	0.59*
FB_wb	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		0.24
CH	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	

Table B.18: Φst values of the pairwise comparisons among parasite localities for COI with *P. arvicanthis* 1 above and *P. arvicanthis* 2 below the diagonal. Dash marks indicate localities where *P. arvicanthis* 1 was not present. Statistically significant values (p <0.05) are indicated in bold and values remaining significant after Bonferroni correction are indicated with asterisks.

	WH	KH	GH	RP	RV	GP	VR	PV	SB	DH	OH	SL	LB	BW	FB_sk	FB_wb	CH
WH		-	-	0.60	**0.96***	**0.96***	**0.95***	**0.95***	-	**0.94**	**0.96***	**0.95***	**0.93**	**0.92***	-	**0.94***	**0.95***
KH	0.31		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GH	**0.71***	**0.60***		-	-	-	-	-	-	-	-	-	-	-	-	-	-
RP	**0.85***	**0.69***	**0.07**		1.00	**0.98**	0.97	0.98	-	0.95	0.99	0.97	0.97	0.91	-	0.93	0.98
RV	**0.79***	**0.89***	**0.88***	**0.91***		**0.11**	0.03	**0.98***	-	**0.97***	**0.99***	**0.11**	0.78	**0.16***	-	**0.99***	**0.99***
GP	**0.91**	**0.95***	**0.96***	**0.97***	0.06		0.01	**0.96***	-	**0.95***	**0.98***	**0.09**	0.09	**0.17***	-	**0.97***	**0.97***
VR	**0.89**	**0.95***	**0.95***	**0.96***	**0.40***	**0.56***		**0.96***	-	**0.95***	**0.98***	**0.05**	0.02	**0.11***	-	**0.96***	**0.97***
PV	**0.98**	**0.97***	**0.98***	**0.99***	**0.80***	**0.93***	**0.89***		-	**0.12***	**0.99***	**0.96***	**0.97**	**0.9***	-	**0.97***	**0.98***
SB	**0.98***	**0.97***	**0.98***	**0.98***	**0.86***	**0.95***	**0.92***	0.00		-	-	-	-	-	-	-	-
DH	**0.97**	**0.98***	**0.98***	**0.99***	**0.80***	**0.92***	**0.88***	**0.98***	**0.98***		**0.97***	**0.94***	**0.93**	**0.88***	-	**0.95***	**0.96***
OH	**0.97***	**0.97***	**0.98***	**0.98***	**0.84***	**0.93***	**0.91***	**0.97***	**0.98***	**0.94***		**0.98***	**0.99**	**0.95***	-	**0.99***	**0.99***
SL	0.99	**0.98***	**0.99***	**0.99***	0.52	0.78	0.53	**0.99**	**0.98***	0.97	**0.97**		0.02	**0.13***	-	**0.96***	**0.97***
LB	**0.90**	**0.95***	**0.95***	**0.96***	**0.52***	**0.66***	0.08	**0.90***	**0.93***	**0.89***	**0.91***	0.45		0.00	-	0.97	0.93
BW	**0.97**	**0.98***	**0.98***	**0.99***	**0.56***	**0.78***	**0.52***	**0.97***	**0.97***	**0.96***	**0.96***	0.68	0.41		-	**0.90***	**0.91***
FB_sk	**0.63***	**0.76***	**0.76***	**0.81***	**0.66***	**0.69***	**0.67***	0.08	0.17	**0.73***	**0.77***	0.66	**0.69***	**0.71***		-	-
FB_wb	**0.98**	**0.97***	**0.98***	**0.99***	**0.88***	**0.95***	**0.93***	**0.98***	**0.98***	**0.98***	**0.98***	0.99	**0.94***	**0.98***	**0.65***		0.36
CH	**0.98**	**0.97***	**0.98***	**0.99***	**0.82***	**0.93***	**0.89***	0.98	**0.98***	**0.98***	**0.97***	0.99	**0.90***	0.97	**0.59***	**0.93***	

Table B.19: Nuclear CAD pairwise Φst values among sampled localities with *P. arvicanthis*1 above and *P. arvicanthis* below the diagonal. Dash marks indicate localities where *P. arvicanthis* 1 was not present. Statistically significant values (p <0.05) are indicated in bold and values remaining significant after Bonferroni correction are indicated with asterisks.

	WH	KH	GH	RP	RV	GP	VR	PV	SB	DH	OH	SL	LB	BW	FB_sk	FB_wb	CH
WH		-		**0.53**	**0.64**	**0.65***	**0.65***	**0.73***	-	**0.60***	**0.66***	**0.63***	**0.65***	**0.68***	-	**0.48***	**0.73***
KH	0.30		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GH	**0.82**	**0.26***		-	-	-	-	-	-	-	-	-	-	-	-	-	-
RP	0.70	0.12	**0.29***		0.73	**0.66**	**0.70**	**0.95**	-	**0.67**	0.83	**0.66**	0.74	**0.75***	-	0.36	0.89
RV	**0.66**	**0.38***	**0.54***	**0.34***		0.09	0.08	0.43	-	0.09	0.16	0.15	0.07	0.09	-	0.08	0.46
GP	**0.55***	**0.33***	**0.47***	**0.31***	**0.11**		0.00	0.09	-	**0.15***	0.14	0.03	0.00	**0.15***	-	**0.25***	**0.40***
VR	**0.56***	**0.28***	**0.45***	0.28	0.07	0.00		0.14	-	**0.13***	0.16	0.04	0.00	**0.11**	-	**0.24***	**0.45***
PV	0.85	**0.40***	**0.54***	0.13	**0.59**	**0.51***	**0.51***		-	0.00	0.07	0.06	0.08	0.05	-	0.25	**0.65***
SB	**0.75**	**0.42***	**0.44***	0.14	**0.59***	**0.56***	**0.55***	0.00		-	-	-	-	-	-	-	-
DH	0.73	**0.30**	**0.61***	0.36	**0.32***	0.20	**0.22**	**0.67**	**0.61***		0.00	**0.11**	0.06	**0.10***	-	0.22	**0.31***
OH	**0.78**	0.55	**0.70***	**0.59***	**0.18***	0.08	**0.18***	**0.73***	**0.71***	0.33		0.07	0.08	0.08	-	0.18	**0.47**
SL	0.85	0.33	**0.56**	**0.43**	0.14	0.14	0.21	0.67	**0.53**	0.58	**0.46**		0.00	**0.12**	-	0.21	**0.36***
LB	**0.52**	**0.25***	**0.40***	**0.25***	**0.06***	0.02	0.02	**0.47***	**0.51***	0.13	**0.11**	0.16		0.01	-	0.14	**0.44**
BW	**0.47**	**0.27***	**0.38***	0.29	0.03	0.04	**0.09**	**0.43***	**0.48***	**0.21**	**0.09**	0.00	**0.06**		-	**0.33***	**0.40***
FB_sk	**0.68***	**0.29***	0.09	0.02	**0.48***	**0.49***	**0.45***	0.18	**0.22***	**0.45***	**0.61***	0.32	**0.43***	**0.44***		-	-
FB_wb	0.64	0.17	0.37	0.15	**0.48**	0.38	**0.38**	0.36	0.35	**0.33***	0.47	0.29	**0.36**	**0.31**	0.03		0.08
CH	0.75	0.17	**0.57**	0.16	**0.48**	**0.35**	**0.35**	0.57	**0.46**	**0.66***	0.66	0.57	**0.32**	**0.30***	0.21	0.50	