The molecular basis of high duty-cycle echolocation in bats, and its role in the divergence of populations and species.

Kim Warren

School of Biological and Chemical Sciences

Queen Mary, University of London

Mile End Road, London, E1 4NS

Submitted in partial fulfilment of the requirements of the Degree of Doctor of Philosophy.

Abstract

How populations diverge and form new species in the face of gene flow is a key question in evolutionary biology. Recent research suggests this may be possible where the same traits affect the ecological niche and are involved in assortative mating, and that a small number of genes could be involved in driving speciation in these cases.

Echolocation call frequency in bats has roles in ecology and social communication. Bats using HDC echolocation have hearing tuned to specific frequencies, with frequency shifts impacting ecological niche and mate recognition, meaning this is a good candidate trait to drive speciation. HDC echolocation has evolved independently in two highly divergent groups of bats, providing a unique opportunity to study the molecular basis of a trait potentially driving speciation. I have combined selection testing of specific loci with genomewide divergence scans to test hypotheses concerning the evolution of HDC echolocation.

Members of the yangochiropteran genus *Pteronotus* use low duty-cycle echolocation, except for the subgenus *Phyllodia*. Selection tests on coding sequence data revealed loci associated with hearing under positive selection in *Phyllodia* and in *Pteronotus*, including eleven shared with a yinpterochiropteran HDC echolocator, *Rhinolophus sinicus*.

Three size and acoustic morphs of *Rhinolophus philippinensis* exist in sympatry on Buton Island. Phylogenetic reconstructions revealed population structure between the morphs, though with conflicting topologies based on mitochondrial and nuclear data. Species delimitation identified at least two separate taxa. Genomewide scans of divergence indicated low background F_{ST} between the morphs, punctuated with highly diverged islands featuring an overrepresentation of genes associated with body size and hearing.

This thesis represents the first genome-wide investigation of HDC echolocation, highlighting candidate genes related to this trait. It additionally describes a rarely observed mammalian ecological speciation, providing support for the claim that species designated *R. philippinensis* represent a complex across their range.

Acknowledgements

Behind every thesis stands a list of individuals without whom it would not have been possible, and this thesis is no exception. I've been lucky to have so many people willing to help me get through the last four years with my sanity (more or less) intact.

First thanks go to my supervisor Stephen Rossiter for trusting me with the execution of this project and for his advice, patience and support throughout. Additionally, I would like to thank the rest of my panel, Richard Nichols and Christophe Eizaguirre, for their constructive criticism in panel meetings and for providing guidance on methodological questions. Further thanks go to Elizabeth Clare for providing much appreciated information about the current state of taxonomy and research on the *Pteronotus* genus for Chapter 2, Rodrigo Pranca for advice on data preparation for Chapters 3 and 4, and to Steven LeComber for his help with implementations of statistical tests in Chapter 4.

When I initially arrived at Queen Mary I had very little practical experience working with large-scale genetic data, and profuse thanks go to Kalina Davies, Georgia Tsagkogeorga and Joe Parker for introducing me to the bioinformatics techniques I would need to start analysing my data, and for their willingness to lend their expertise to surmount any hurdles I encountered. In the same vein, I would like to thank Bob Verity for his help with interpreting the output of the MavericK software, and to Philine Feulner for readily providing the R permutation scripts that she had used to carry out the permutation test of F_{ST} islands of divergence in three-spine stickleback.

Many people have been involved in collecting samples and providing sequence data used in this research. Particular thanks go to Seb Bailey, Liliana Dávalos, Tigga Kingston, Felicia Lasmana, Josh Potter, Juliana Senawi, Sigit Wiantoro, Laurel Yohe, and to all

others who were involved in the field over a period of many years. Further thanks go to Burton Lim for providing additional tissue samples of *Pteronotus* and outgroup species for Chapter 2 from the Royal Ontario Museum, and to Kyle Armstrong for providing samples of *Rhinolophus montanus* for Chapters 3 and 4. I would also like to extend thanks to Monika Struebig and Philip Howard for training and supervising me on wet lab techniques, so that I was able to get from these tissue samples to analysis-ready data, and to Jeremy Johnson for providing advanced access to the Broad Institute *Rhinolophus ferrumequinum* assembly that enabled construction of the *Rhinolophus philippinensis* reference for Chapters 3 and 4.

I have been fortunate to be working in a department with a rich PhD student culture. Thanks go to my lab mates Dave Bennett, Rosie Drinkwater, Hernani Oliveira, Tiago Teixeira, Sandra Alvarez-Carretero, Joe Williamson, Ilya Levantis and James Gilbert for sharing lunch times, beers after work and periods of sudden deadline stress. While all the PhD students in the department over the last few years have been great, I also particularly want to call out Sally Faulkner and Liz Archer their support. Away from university, huge thanks go to my housemate Allie Hill, for making sure that I semi-regularly stepped away from the computer to talk to people, occasionally ate a vegetable, and for making sure that I got outside every single day by giving me a puppy in the middle of my PhD!

Finally, I would be remiss if I concluded these acknowledgements without saying a massive thank you to my family, who supported me with their love, unfailing belief that I would succeed, and acceptance that it was not going to be a good idea to ask how things were going with my PhD for the last two years. I would not have got here without you.

Author's declaration

I, Kim Warren, confirm that the research included with this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature:

Date:

Details of collaboration and publications:

Tissue samples used in Chapter 2 were provided by Liliana M. Dávalos and Burton Lim.

The lab work for many of the samples used in Chapter 2, including RNA extraction and library preparation, was assisted by Monika Struebig, Philip Howard and Kalina Davies.

The bioinformatics data preparation of short-read RNASeq data, including read trimming and transcriptome assembly, for all outgroup species presented in Chapter 2 was performed by Kalina Davies.

Tissue for *Rhinolophus montanus* in Chapter 3 and 4 was provided by Kyle Armstrong; Sigit Wiantoro assisted with harvesting *Rhinolophus philippinensis* tissue at the Museum of Bogor and provided mainland Sulawesi samples while Tigga Kingston provided Kabaena and Buton Island samples.

This work was funded as part of an ERC grant.

Table of Contents

List of Figures

List of Tables

CHAPTER ONE

Introduction

The details of how and why novel species arise from existing populations have been an area of hot debate for many decades, particularly where speciation appears to have occurred in the absence of physical barriers. Systems in the early stages of a speciation event offer us unique chances to expand our understanding how these processes progress in natural populations. In this introduction I review current thinking on ecological speciation in wild animal populations with the assistance of some well-studied examples. I discuss the role of 'magic traits' and explore how high duty-cycle (HDC) echolocation using predominantly constant-frequency (CF) calls in bats might qualify as one such trait. Finally, I provide background information on my study systems, the *Phyllodia* subgenus (of the *Pteronotus* genus) and the large-eared horseshoe bat *Rhinolophus philippinensis*, both of which use HDC echolocation. Finally, I summarise the main aims of this study and the organisation of this thesis.

Divergent selection and ecological speciation

During the latter half of the twentieth century, speciation research became focussed on a model in which population divergence was more driven by neutral mechanisms such as genetic drift than by natural selection (as discussed in Via 2001). Consequently, the presence of geographical barriers physically separating diverging populations became viewed as an almost essential prerequisite for the evolution of new species (Mayr 1963). The notion of species diverging in the absence of geographical barriers ('sympatric speciation') became highly controversial and was rebutted by major names in the field of speciation research (e.g. Futuyama and Mayr 1980). A problem for this form of speciation

is that without any physical separation between diverging populations, gene flow will homogenize gene pools and lead to a break up of any genomic divergence via recombination. Yet despite this antagonism between gene flow and divergence, by the turn of the century, there was a build in momentum behind research into modes of 'non-allopatric' speciation, fuelled in part by the revelation that there was genetic differentiation between ecomorphs in *Rhagoletis pomonella*, one suggested example of sympatric speciation (McPheron, Smith and Berlocher 1988). As we have progressed through the first two decades of the 21st century, the nascent field of speciation genomics has been able to leverage the rapid advances in sequencing technology to delve deeper into questions surrounding speciation with gene flow: whether speciation can occur in the face of homogenising gene flow, how common this is, and what mechanisms could enable it to happen (see e.g. Nosil 2008; Feder, Egan and Nosil 2012).

While mathematical demonstrations of how sympatric speciation could take place in the face of gene flow have existed for decades (e.g. Maynard Smith 1966), it has proven difficult to show conclusively that species have diverged in the face of gene flow in the real world. Relatively few good examples of this process have been described (Via 2001), though this number is rapidly increasing to encompass a wide variety of taxonomic groups as research effort in this area increases (e.g. Western Diamondback Rattlesnakes *Crotalus atrox* (Schield et al 2015), whiptail lizards *Cnemidophorus ocellifer* (Oliveira et al 2015), *Heliconius* butterflies (Supple et al 2015), *Myotis* bats (Morales 2017)). The most well-established cases that have been extensively studied over decades include the apple maggot *Rhagoletis pomonella* and the East African crater lake cichlids (Family: Cichlidae).

R. pomonella was proposed early on as a possible example of speciation without the presence of geographical barriers to impede gene flow (Bush 1969), though it was not until later that differences in allele frequencies between the different races were

confirmed (McPheron, Smith and Berlocher 1988). Strong host fidelity for breeding behaviour established a strong pre-mating reproductive barrier between populations preferring apple or hawthorn, with the populations then diverging to adapt to the conditions associated with their preferred host (Filchak, Roethele and Feder 2000). Genome-wide allele frequency shifts have been demonstrated between naturally co-occurring populations of *Rhagoletis* with different hosts, though this particularly marked impact of ecological adaptation has been credited to multiple genomic inversions (Egan et al 2015). Genetic evidence from a range of *Rhagoletis* populations suggests that host shifting, ecological divergence and the initiation of speciation with gene flow to capitalise on ecological opportunity may be common in this group (Powell et al 2014) and has been suggested to play a larger role in explaining the diversity of phytophagous insects.

The *Rhagoletis* system is one example of ecological speciation, defined as "the process by which barriers to gene flow evolve between populations as a result of ecologically based divergent selection between environments" (Nosil 2012). This process has not solely been described in insects, but has also been observed in vertebrates. Species of cichlid fish are found in many crater lakes in East Africa, with the system being remarkable for the fact that the lakes were each thought to have been colonised by single events with a range of assortatively mating descendant species diverging subsequently from that ancestor. This was supported by an apparent monophyly of species with differing trophic and reproductive ecology within single lakes based on molecular data (e.g. Schliewen, Tautz and Pääbo 1994; Barluenga et al 2006). More recently the role of sympatric speciation in this system has been thrown into question, with SNP-based population genetics analyses indicating that these apparent sympatric radiations may have been helped along their way by a mixture of further colonisation events and specialisation to microhabitats within their home lakes (Martin et al 2015). This particular facet of the debate is still very much active, with even more recent research suggesting that secondary

gene flow is likely to have been weak and thus unable to explain the radiation of at least one cichlid species (Richards, Poelstra and Mann 2017). Evidence of complex colonisation histories has shed doubt on other systems proposed as examples of sympatric speciation. Multiple co-occurring pairs of three-spined stickleback *Gasterosteus aculeatus* are now thought to have undergone periods of physical separation in the course of developing reproductive isolation (as reviewed in McKinnon and Rundle 2002). In fact, all the most reliable examples we have of speciation with gene flow in nature involve division of the population over microhabitats with subsequent assortative mating based on locality (Martin et al 2015). Ecological speciation via specialisation to different hosts or microhabitats and subsequent assortative mating provides one way around the antagonistic relationship between genetic divergence and the homogenising effects of gene flow.

Speciation genes

While reproductive isolation happens at the level of the whole genome, adaptation happens at the level of individual genes – so it may be appropriate to look for signatures of speciation happening at this level (Wu 2001). Mathematical models of speciation also concluded that speciation with gene flow would be more likely where fewer loci were implicated (e.g. Maynard Smith 1966, Felsenstein 1981). It follows that genomic divergence between two diverging populations would not be homogenous, and that we could learn more about the biological processes underlying diversity by finding the genetic basis of reproductive isolation (Turner, Hahn and Nuzhdin 2005). Early research in this area used quantitative trait locus mapping, but this had limitations when trying to look for fine-scale genetic differences.

The use of microarray hybridisation methods opened up the possibility to look for genomic differentiation at this finer scale in the mosquito *Anopheles gambiae* (Turner, Hahn and Nuzhdin 2005). This led to the identification of three highly diverged regions, including two (on the 2L and on the X chromosome) that contained multiple fixed differences and no shared polymorphisms, in stark contrast with flanking control regions. While the M and S forms generally appeared to mate assortatively, there was evidence for substantial migration between the two forms based on the existence of gravid females mated by the other form and of F1 hybrids in the wild. These regions of high divergence were dubbed 'speciation islands', potentially containing 'speciation genes' that were responsible for pre- or postzygotic isolation between the two forms. Differentiation in these regions was hypothesised to be as a result of selection against hybrid genotypes, potentially because recombination could break linkage disequilibrium between genes responsible for isolating factors such as assortative mating and postzygotic isolation (Turner, Hahn and Nuzhdin 2005).

Via and West (2008) proposed that these heterogeneous levels of differentiation could be explained by divergence hitchhiking, with regions near quantitative trait loci (QTLs) becoming strongly linked due to low recombination near the trait under selection. They noted a significant relationship from a logistic regression of F_{ST} outlier status against distance to the nearest QTLs associated with traits which cause ecological reproductive isolation in pea aphids. These divergently selected QTLs between incipient pea aphid species generated large divergence 'footprints' on the genome (with an average distance between an F_{ST} outlier and the nearest QTL being 10.6cM), where other loci diverged by 'divergence hitchhiking'. For this reason, the authors suggested that these unexpectedly large hitchhiking regions were protected from recombination due to population subdivision in cases of speciation with gene flow, a phenomenon which they referred to as the 'genetic mosaic of speciation'.

Genomic regions of elevated F_{ST} have come to be known as islands of differentiation (Harr 2006) or islands of divergence (Nosil, Funk and Ortiz-Barrientos 2009). Genome scans, which estimate the relative differentiation across the genome between a pair of species by looking at many loci (Beaumont and Balding 2004, Saetre 2014), have become a key technique in the nascent field of speciation genomics (Feder, Egan and Nosil 2012, Seehausen et al 2014). Genome scan methods have been applied in a wide range of species. Eight autosomal regions of high differentiation were identified between two subspecies of house mice (Harr 2006), some of which were enriched for genes associated with host-pathogen interactions or with olfaction. Variants were identified in genes of interest with regards to adaptations to different migratory strategies that showed moderate to high divergence between two species of willow warbler (Lundberg et al 2013). Genome-wide scans in nearly 1000 *Arabidopsis thaliana* to look for climate associations of around 215,000 SNPs informed a model that could accurately predict the fitness performance of different genotypes in a particular climate, indicating that this kind of approach can pick up true signals (Hancock et al 2011).

Controversy has developed over the suitability of genome scans for identifying islands of diversity and implicating them in driving speciation by divergence. Years after their original paper, Turner and Hahn (2010) conceded that it was hard to demonstrate unambiguously the link between their differentiated islands and loci responsible for reproductive isolation, especially after White et al (2010) demonstrated that the DNA near the centromeres of all three of the mosquito chromosomes was in near-perfect linkage disequilibrium. Cruickshank et al (2014) noted that if absolute rather than relative measures of divergence were used then genomic islands of differentiation disappeared, with these regions showing reduced diversity. They suggested that the explanation that these regions were protected from introgression and so experienced reduced gene flow compared to the surrounding less diverged regions is flawed in light of this finding. As

relative measures of divergence can also be elevated by selection as well as by reduced gene flow, they suggest that selection and linkage may be enough to explain the patterns seen and that further information is needed to conclude the presence of speciation with gene flow. A combinatory approach of genome scans and candidate gene studies may be more suitable when looking for genomic regions related to speciation (Saetre 2014). This approach has been used in the Swainson's thrush *Catharus ustulatus* (Ruegg et al 2014), where candidate genes associated with characters that are divergent between two subspecies were more diverged than would be expected by chance, but lay outside of the islands of divergence identified by genome scans.

Magic traits

One option for defining candidate genes is to identify loci affecting traits with dual roles in ecology and in reproduction or mate choice – so-called 'magic traits' (Servedio et al 2011). Divergent ecological selection acting on these traits leads to positive assortative mating, with individuals selecting mates more similar to themselves, thereby reinforcing the divergence between the different forms or populations. While magic traits are identified at the phenotypic level (Servedio et al 2011), the genes underlying the traits provide a point of entry for looking for molecular evidence of divergent selection acting between putative incipient species in sympatry.

Some examples of 'magic trait' speciation involve traits which directly affect the ability of individuals of the diverging populations to communicate with one another and signal their availability as a mate. Darwin's finches (*Geospiza*) are a classic example of geographical speciation (Lack 1983). Finch species are distributed over several islands, with species co-occurring on single islands differentiated by body size and beak

morphology, which is linked to the trophic niche that they exploit (Grant and Grant 2011). Individuals recognise conspecifics both by a combination of head and body morphology (Ratcliffe and Grant 1983) and by the structure of their song, which functions as a vocal mating signal (Huber et al 2007). Notably, song features also vary with beak morphology (Podos 2001, Huber and Podos 2006). This means that ecological selection on beak morphology to allow exploitation of different feeding niches would affect a mating signal, promoting the development of assortative mating within populations with different beak sizes due to song changes affecting mate recognition and enabling rapid speciation in the group. While there appear to have been a range of forces involved in driving this adaptive radiation (Farrington 2014), 'magic trait' speciation appears to have played a role (Huber et al 2007). Beak morphology has also been proposed as a factor reducing gene-flow between populations of Island Scrub-Jay (*Aphelocoma insularis*) adapted for oak or pine habitats, where longer or shorter bills favouring different foraging habitats both affect rattle calls used for mate recognition and also act as a visual cue (Langin et al 2017). This may be a common process in birds; ecological selection on body size and beak morphology in ovenbirds (family Furnariidae) has a much stronger effect on song features than direct selection on song features imposed by their habitats, suggesting this radiation was also driven by magic trait speciation (Derryberry et al 2018).

In these bird examples, ecological speciation was acting on a trait that affected resource exploitation, but which also had effects on mate signalling. Another way that this dual purpose can be realised is if ecological selection acts on a trait used for prey detection that is also involved in assortative mating. An example of this is the electric organs discharge (EOD) in mormyrid fish (*Campylomormyrus* species) (Feulner et al 2009). Variation in EOD has been implicated as a driving force for diversification in two species flocks of African mormyrids, which use electrolocation to detect food. These fish display phenotype-assortative social preferences and have been shown to make decisions based

on EOD waveform characteristics (Feulner et al 2009a, Feulner et al 2009b). This dual function of the EOD means that disruptive ecological selection on prey recognition leads to assortative mating, making it a candidate to be a driver for speciation.

Echolocation in bats

While echolocation has been documented in multiple mammalian orders, including cetaceans, marsupials and insectivores, it is most widespread and derived in the bats (Order: Chiroptera). Echolocation in bats involves the use of ultrasound for orientation and obstacle avoidance in the dark and for locating food. However, not all bats echolocate, and the evolutionary history of echolocation in bats has not been fully elucidated. Bats are divided into two suborders based on molecular data; the Yinpterochiroptera and the Yangochiroptera (e.g. Teeling et al 2000, Springer et al 2001, Teeling et al 2002, Teeling et al 2005). While laryngeal echolocation is present in both suborders, the Yinpterochiroptera also includes the non-echolocating Old World fruit bats (Family: Pteropodidae) and the tongue-clicking *Rousettus* (Altringham 2011). Analysis of morphological characters of extant and extinct taxa has not yet fully determined whether laryngeal echolocation evolved once ancestrally in bats and was later lost in the Pteropodidae (reviewed in Jones and Teeling 2006), or whether it evolved independently in the Yangochiroptera and the Yinpterochiroptera suborders (e.g. Davies et al 2013). Bats are one of the most diverse mammalian orders, second only to rodents (Altringham 2011), and echolocation is one of the innovations which appears to have allowed them to speciate so rapidly.

Laryngeal echolocation calls can be divided into two forms – frequency-modulated (FM) calls that consist of a broadband sweep down the frequencies, and longer constant-

frequency (CF) calls that are dominated by a single tone (Altringham 2011) (Figure 1.1). Bats using calls composed of predominantly FM elements are sensitive to a broad range of frequencies; given the volume of their emitted ultrasonic calls, they avoid self-deafening by contracting muscles in their middle ear while calling (Henson 1965). This limits them to short, discrete calls with intervening periods long enough to allow the detection of all returning echoes (Holderied and von Helversen 2003). This strategy of separating the emitted echolocation pulse and the received echo in time is known as low duty-cycle (LDC) echolocation (Lazure and Fenton 2011) and is used by approximately 900 species of bat (Fenton, Faure and Ratcliffe 2012). There is a high variability in the frequency use of LDC bats; some species, such as *Lasiurus coiners* (Corcoran and Weller 2018), use calls with even less frequency modulation than a CF call while still using low-duty cycle echolocation. Typically, bats using CF echolocation have highly modified cochleae (Davies, Maryanto and Rossiter 2013) creating an acoustic fovea analogous to the fovea in the visual system (Schuller and Pollack 1979). Bats in the family Rhinolophidae have particularly long relative basilar membrane lengths (Davies, Maryanto and Rossiter 2013), with a disproportionate representation of neurones in the inferior colliculus that respond maximally to a specific echolocation call frequency (Schuller and Pollack 1979). This acoustic fovea provides these bats with a particularly high sensitivity to a particular call frequency (Schuller and Pollack 1979). These bats use Doppler shift compensation to alter their emitted call frequencies in flight to ensure that the returning echoes fall within their most sensitive frequency range (Metzner et al 2002). As their emitted frequency is outside of this high-sensitivity range, these species do not have the same risk of self-deafening and so are less constrained to have discrete calling and listening phases. This enables them to use high duty-cycle (HDC) echolocation, separating the emitted pulse and the received echo by a slight difference in frequency rather than by time (Lazure and Fenton 2011), and allowing them to spend a higher

proportion of time detecting prey. HDC echolocation is employed by about 160 species of bat, most of which are found in the yinpterochiropteran families Rhinolophidae and Hipposideridae (Fenton, Faure and Ratcliffe 2012).

HDC echolocation is thought to be specialised for flutter detection, with strong acoustic 'glints' being produced when the signal hits an insect at the moment that its wings are perpendicular to the call (Schnitzler and Kalko 2001). These 'glints' increase the detection chance of insects against background clutter (Schnitzler and Kalko 2001). The parameters of the very narrowband calls employed in HDC echolocation have a strong impact on the prey space that can be sampled by the bats. The duty cycle of the call, or the percentage of time for which the call is emitted, impacts the probability of detection of insects with different wing-beat rates (Schnitzler and Kalko 2001). The frequency of the call impacts both the distance over which prey species can be detected and the size of insects which can be found. Higher frequency calls are more strongly subject to atmospheric attenuation, so are effective only over shorter distances (Lawrence and Simmons 1982). Lower frequency calls have longer wavelengths and as such are poor at detecting small prey due to the effects of Rayleigh scattering when the prey is smaller than the wavelength of the echolocation call (Houston et al 2004).

Ultrasonic calls also play a role in social communication (Jones and Siemens 2011). While bats have been shown to use a rich repertoire of non-echolocation vocalizations in social communication (Ma et al 2006), they do recognise the echolocation calls of conspecifics (Li et al 2014) and echolocation calls encode a wealth of personal information about the transmitter that other bats can 'eavesdrop' on and use (Gillam and Brock Fenton 2016). There is evidence that they can determine the sex of the caller from an echolocation call (Schuchmann, Puechmaille and Siemers 2012) and use this discriminatory capacity to inform behavioural decisions. The greater sac-winged bat *Saccopteryx bilineata*, which uses LDC echolocation, responds to the echolocation calls

of conspecific males with aggressive vocalisations and to females with courtship vocalisations (Knörnschild et al 2012). Some bats will produce echolocation calls in response to playbacks of conspecific calls when stationary and not foraging despite this being metabolically costly to do, suggesting that they may be actively produced social signals (Dechmann et al 2013). Considering HDC bats specifically, it has been shown that female *Rhinolophus meleyhii* make mate choice decisions based on male call frequency (Puechmaille et al 2014). *Rhinolophus ferrumequinum* in China can distinguish and exhibits a preference for echolocation calls of its local population to those of more distant conspecifics (Lin et al 2016). The ability to recognise gender from echolocation calls has also been demonstrated in *Rhinolophus clivosus*, and even the ability to recognise individuals (Finger 2015) and body condition (Raw 2016).

Given its dual functions in foraging and social communication, and the fine tuning of the cochlea to narrowband calls, HDC echolocation satisfies the necessary criteria of a magic trait and have been recognised as such (Servedio et al 2011). In these bats, the impact of multiple parameters of the call structure on the ability to sample the environment and detect prey means that there is a lot of scope for divergent ecological selection to act on this trait. The tight tuning of the acoustic fovea in HDC echolocators means that there are constraints on perception of sounds used for communication, so disruptive ecological selection acting on the call frequency could drive reproductive isolation (Kingston et al 2001). HDC echolocation may predispose rapid speciation, as in the case of horseshoe bats, where a large number of morphologically highly similar species in close proximity confounding taxonomists for years (Csorba 2003). HDC echolocation is itself thought to have evolved independently twice in bats (Figure 1.2).

High duty-cycle echolocation in the Yinpterochiroptera: family Rhinolophidae

The family Rhinolophidae is one of the echolocating clades of bats within the suborder Yinpterochiroptera, and is more closely related to the non-echolocating fruit bats of the family Pteropodidae than to echolocating bats in the Yangochiroptera (Teeling et al 2005). The Rhinolophidae split from the Pteropodidae about 60 million years ago (Lei and Dong 2016) and consists of 77 listed species (Simmons 2005), although new taxa have since been discovered, with recognised species across its range regularly being reported as actually being complexes of cryptic species (e.g. Volleth et al 2015, Soisook et al 2016, Sun et al 2016, Tan Tu et al 2017, Taylor et al 2018). All horseshoe bats are distributed across temperate and tropical regions in the Old World (Nowak and Paradiso 1999). These bats are closely related to the Hipposideridae, another family using HDC echolocation from which they diverged approximately 42 million years ago (Foley et al 2015). Both of these genera are notable not only for their high species number and cryptic species, but also for their rapid diversification (Foley et al 2015). This rapid diversification has been proposed to be related to their use of HDC echolocation, supported by examples such as the Australian *R. philippinensis*, which has been proposed to represent three separate species separated by size and echolocation call frequency (Cooper et al 2008).

High duty-cycle echolocation in the Yangochiroptera: subgenus Phyllodia

The Yangochiroptera and Yinpterochiroptera have been estimated to have diverged from one another approximately 63 million years ago (Lei and Dong 2016). The former

suborder contains 12 families, all of which use LDC echolocation. The exception to this can be found within the Mormoopidae, a family of neotropical bats that contains the two genera *Mormoops* and *Pteronotus*. Up until the last few years there were thought to be six extant species of *Pteronotus* of which some were divided into multiple subspecies, but more recently these subspecies have been elevated to full species status (Clare et al 2013, Pavan and Marroig 2016). In particular, the species formally known as *Pteronotus parnellii* is now recognised as a complex of nine full species, known collectively as the subgenus *Phyllodia*.

Members of the *Phyllodia* subgenus are unique in the Yangochiroptera in their use of high duty-cycle echolocation. Its congeners use predominantly low duty-cycle echolocation with broadband FM calls featuring short CF tails, while members of its sister genus *Mormoops* use pure FM calls (Mora et al 2013). This marked divergence in echolocation call strategy at a shallow node within the phylogeny presents an opportunity to investigate the genetics underlying the evolution of high duty-cycle echolocation, and the possibility of identifying loci that have contributed to this divergence.

Bats in the *Phyllodia* subgenus show convergent phenotypic adaptations with the high duty-cycle echolocators in the Yinpterochiroptera, including a sharply tuned auditory fovea (Kössl and Vater 1985), the use of Doppler shift compensation (Jen and Kamuda 1982), and a disproportionately large cochlea (Davies, Maryanto and Rossiter 2013). However, some traits commonly associated with high duty-cycle echolocation are present in other species within the genus. There is evidence that species in the *P. personatus* complex, the next most basal group in the *Pteronotus* genus, also exhibit Doppler shift compensation despite not using a long constant-frequency component when echolocating (Smotherman and Guillén-Servent 2008). Additionally, *Pteronotus quadridens* has been shown to use heteroharmonic target-range computation (Mora et al 2013). This raises some questions about where within the genus the transition to HDC echolocation

happened; was it basally present with a reversion after the divergence of the *Phyllodia* subgenus, or did it occur specifically in the *Phyllodia*?

Aims

This thesis breaks down into two major sections, centred around three main objectives. The first section focusses on the *Pteronotus* genus, with the objective of understanding the switch from LDC to HDC echolocation in the *Phyllodia* subgenus. This section of the thesis includes two aims: to identify the point at which the shift to HDC echolocation occurred within the *Pteronotus*, and to identify candidate genes in the evolution of this potential magic trait by looking for common patterns of selection in hearing genes between the *Phyllodia* and *Rhinolophus* bats.

The aim of the second section is to investigate a putative incipient speciation process underway within the horseshoe bats of Buton Island, Indonesia. The first objective here is to investigate population genetic structure on the island using both mitochondrial and genomic data. The second objective is to look for 'islands of divergence' within the genome by performing F_{ST} scans on whole-genome data, and using information about candidate genes implicated in the evolution of constant-frequency echolocation to understand how these islands of speciation have developed and how they may lead to reproductive isolation.

The overarching aim of this thesis is to make a first attempt to dissect the genetics of a putative magic trait. While there has been a lot of research into the genetics of echolocation in general, very little is known about the genes specifically underpinning the evolution of high duty-cycle echolocation. The switch in echolocation call type within

the *Pteronotus* genus and the abrupt call differences that have evolved between the *R. philippinensis* morphs on Buton Island suggests that relatively few loci have been involved, or else that loci involved are linked. These loci represent potential 'speciation genes', and identifying them would allow us to gain further understanding into the links between HDC echolocation and speciation within Chiroptera.

Thesis organisation

Each of the chapters within this thesis addresses one of the key objectives listed above, as follows:

In Chapter 2, I describe the molecular methods used to obtain coding sequence data from members of the Mormoopidae and discuss the results of multiple sequence alignments and selection tests within this family. I identify candidate genes for association with constant-frequency echolocation and seek evidence to allow determination of whether this trait evolved ancestrally to the *Pteronotus* genus or specifically within the *Phyllodia* subgenus.

In Chapter 3, I investigate population structure between the three sympatric morphs of *Rhinolophus philippinensis* using mitochondrial data extracted from whole-genome sequencing reads.

In Chapter 4, I explain the methodology for conducting genome scans to identify islands of speciation and combine the results of this with transcriptome data to investigate the roles of genomic regions that may be involved in driving the development of reproductive isolation.

Each data chapter is structured as a paper, including Abstract, Introduction, Results, Discussion and Methods sections. Chapter 5 is a general discussion in which all results are considered and drawn together to inform overarching conclusions.

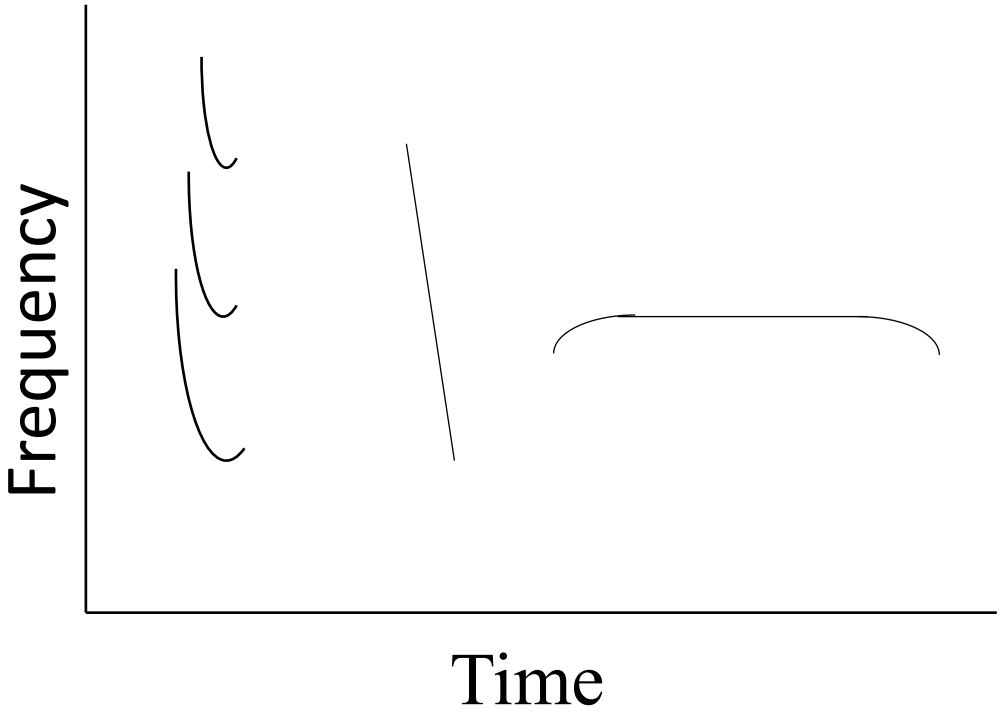


Figure 1.1: Frequency modulated and constant frequency echolocation calls, adapted from Jones and Waters 2000. Left: *Pipistrellus pipstrellus*: Multi-harmonic FM sweep with a narrowband tail. Middle: *Myotis nattereri*: broadband FM call. Right: *Rhinolophus hipposideros*: long CF call with short FM components.

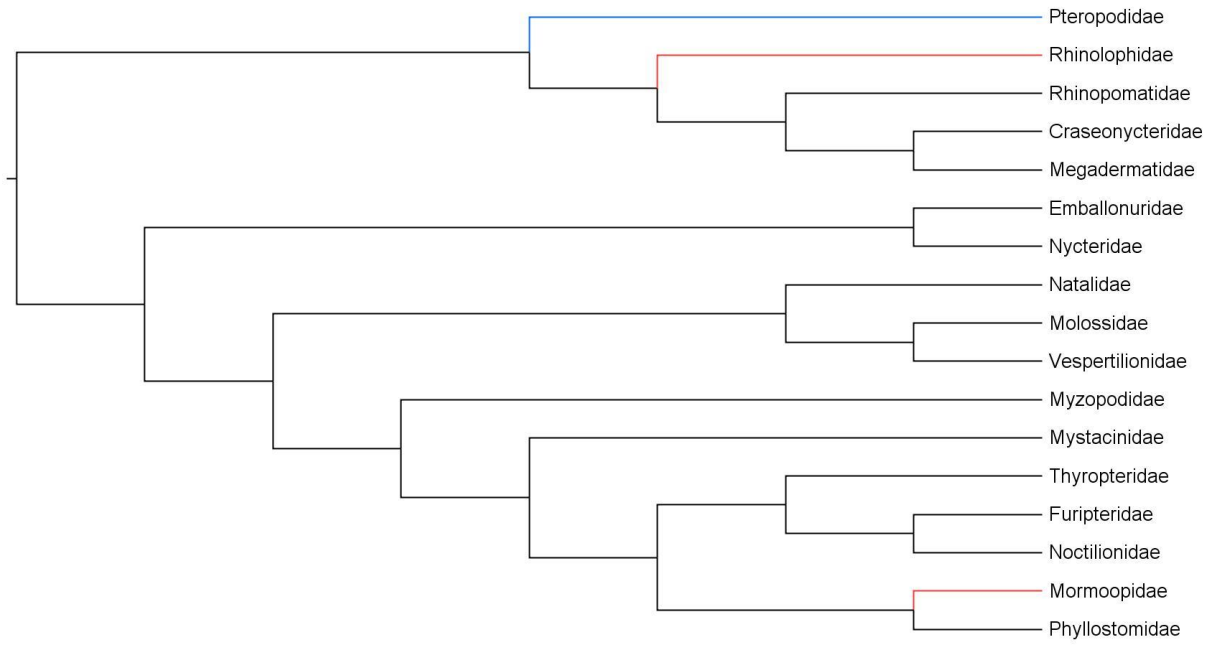


Figure 1.2: The two independent origins of HDC echolocation in bats. Phylogeny derived from Teeling et al 2005. Branches coloured in red indicate the clades in which HDC echolocation has evolved – the Rhinolophidae (here containing the hipposiderids as a subfamily; Hipposideridae is now commonly accepted to be a sister family to the Rhinolophidae and also uses HDC echolocation) and the Mormoopidae. The blue branch indicates the non-echolocating Pteropodidae.

CHAPTER TWO

The evolution of echolocation in mormoopid bats: searching for the genetic basis of divergence in a highly specialised behaviour

Abstract

While echolocation is employed for hunting and obstacle avoidance by most species of bat, there is a large amount of variation in the types of echolocation call used by different species. Echolocating bats can be divided according to duty cycle, or the proportion of echolocating time that they spend emitting sonic pulses. Low duty-cycle echolocation, using short calls separated by long periods, is the most phylogenetically widespread. High duty-cycle echolocation, with long calls separated by short gaps, is associated with a range of specific adaptations. These include the use of narrowband, constant-frequency pulses and changes to the structure of the cochlea. High duty-cycle echolocation has evolved at least twice independently in bats; in the superfamily Rhinolophoidea and in the subgenus *Phyllodia* (family Mormoopidae).

The *Phyllodia* subgenus is notable as the rest of the mormoopid bats use low duty-cycle echolocation. This divergence in call structure between closely-related species allows us a unique opportunity to search for genes associated with specific call types. Previous studies have identified candidate genes implicated in the evolution of echolocation by making comparisons between echolocating and non-echolocating lineages of bats. However, no genes have yet been associated specifically with the evolution of high duty-cycle echolocation. To identify genes potentially associated with the evolution of high

duty-cycle echolocation I have looked for signatures of positive selection in *Phyllodia* in genes linked to hearing using *de novo* assembled transcriptomes from eight species from the family Mormoopidae and a further 23 background species from the suborder Yangochiroptera. Additionally, I have investigated parallelism in selection in hearing genes with a distantly-related high duty-cycle bat, *Rhinolophus sinicus*, using genomic sequence data available on GenBank for it and for six further species. I have identified 11 hearing genes with signatures of positive selection in both of these lineages, including four genes (*Cacna1d*, *Col4a5*, *Eral1*, and *Ush1g*) that each include a parallel substitution between *R. sinicus* and either the subgenus *Phyllodia* or the ancestral *Pteronotus* branch.

Introduction

Even within such a well-researched biological class as Mammalia, there remains much variation which is poorly explored and understood. A range of specialised traits have evolved in only small numbers of mammals, such as poison glands in slow lorises, infrared sensing in vampire bats and echolocation in bats and toothed whales. While early genetic research predominantly focussed on a limited range of model organisms for which complete genomes had been assembled, it was not possible to explore the genetics underlying these kinds of specialist phenotypes. With continued advances in high-throughput sequencing technology meaning that greater quantities of data can be generated at increasingly lower prices, it is becoming feasible to carry out genome-scale molecular evolutionary analyses on groups of non-model organisms to answer existing biological questions concerning the evolution of these traits.

Echolocation involves the use of sound, normally ultrasound, to gain information on the local environment for hunting and for obstacle avoidance. It is best known in toothed whales and in bats. All species of bat outside of the pteropodid fruit bats use laryngeal echolocation. Echolocation in bats can be divided into two major forms, low duty-cycle (LDC) and high duty-cycle (HDC), where the duty cycle refers to the proportion of time that is spent emitting signals (Fenton, Faure and Ratcliffe 2012). These forms differ in the way that the bat separates their emitted sonar pulse and the informative returning echo to avoid self-deafening, and are associated with specific physiological adaptations.

About 80% of bat species use LDC echolocation (Lazure and Fenton 2011), in which the emitted pulse and returning echo are separated temporally. Bats using LDC generally emit short, broadband pulses that sweep down a range of frequencies, known as frequency-modulated (FM) calls (Figure 1.1). These bats contract muscles in their middle ear while

emitting their high-intensity calls to protect their auditory mechanisms (Henson 1965), so are constrained to short call durations so as not to miss echoes returning from nearby objects. The pause between calls is long enough to allow the bat to receive echoes from the edge of its detection range without overlap in echoes returning from different calls, which would create ambiguity as to which call had generated the echo and thus how distant the object was (Holderied and von Helversen 2003).

In contrast, in HDC echolocation the emitted pulse is separated from the returning echo by frequency, allowing bats using this form of echolocation to spend a greater proportion of time emitting calls (Lazure and Fenton 2011) and so build a more constant image of their surroundings. Bats using this form of echolocation concentrate the energy of their call into a single frequency, emitting long, narrow-band pulses known as constant-frequency (CF) calls (Figure 1.1). Bats using HDC echolocation have modified cochleae with an acoustic fovea highly sensitive to a very narrow band of frequencies due to a disproportionately large population of neurons specifically tuned to their reference frequency (Schuller and Pollak 1979, Neumann and Schuller 1991). To ensure that the returning echo falls within this frequency range, these bats adjust their emitted call to leverage Doppler-shift compensation (Pollack, Henson Jr. and Novick 1972, Suga, Neuweiler and Möller 1976). Doppler-shift compensation also allows these bats to extract information about their target's movements relative to their own (Fenton, Faure and Ratcliffe 2012). The long, CF elements of HDC echolocation calls are particularly good at detecting acoustic glints from fluttering insects to distinguish them against complex, cluttered backgrounds (Odendaal, Jacobs and Bishop 2014). Neural processing strategies for target-range compensation also differ between CF and FM bats (Mora et al 2013), as do cochlear morphology, particularly with respect to the basal turn (Davies, Maryanto and Rossiter 2013). While a species of bat may use a mix of FM and CF elements in their calls, all members of a species use the same kind of echolocation call. Most bat species

employ LDC echolocation (approximately 900 species, as compared to around 160 species using HDC) (Fenton, Faure and Ratcliffe 2012).

HDC echolocation with long narrowband CF calls has evolved independently at least twice within bats; once in the yinpterochiropteran Rhinolophoidea and once in the subgenus *Phyllodia* within the yangochiropteran family Mormoopidae. Members of the *Phyllodia* and the Rhinolophoidea share convergent adaptations associated with HDC echolocation, such as the use of Doppler-shift compensation when echolocating in flight (Henson, Schuller and Vater 1985) and the possession of a highly sensitive acoustic fovea tuned to their target frequency range (Kossel and Vater 1985). However, bats of the *Phyllodia* subgenus emit echolocation calls through their mouths while the Rhinolophidae emit them through their nostrils (Jones and Teeling 2006), and there are also differences in the organisation of the auditory cortex between these two groups of HDC echolocators (O'Neill 1995).

Until relatively recently, the Mormoopidae was described as a family consisting of two genera with eight extant species, *Mormoops* and *Pteronotus* (van den Bussche and Weyandt 2003). Of these, a single species, *Pteronotus parnellii*, had been observed to use HDC echolocation. Subsequent research painted a much more evolutionarily complicated picture (Clare et al 2013), elevating subspecies attributed as being part of *P. parnellii* to species level (Gutiérrez and Molinari 2008, Clare et al 2013, Thoisy et al 2014) and reclassifying *P. parnellii* as part of a complex of nine cryptic species (the subgenus *Phyllodia* - Smith 1972, Pavan and Marroig 2016) separated geographically (Pavan and Marroig 2017). The origin of the genus was approximately 16 million years ago, with the most closely related species within the subgenus *Phyllodia* diverging less than 2.6 MYA (Pavan and Marroig 2017) (Figure 2.1).

While a parsimonious evolutionary history for echolocation within the Mormoopidae would have HDC echolocation evolving ancestrally to the *Phyllodia* subgenus after its divergence, there is conflicting phenotypic evidence in the rest of the *Pteronotus* genus. While these species all use LDC echolocation, they all have short CF tails on their FM calls, and traits commonly associated with HDC echolocation have been observed throughout the genus (Figure 2.2). *Pteronotus personatus* employs Doppler-shift compensation in flight (Smotherman and Guillén-Servent 2008). Doppler-shift compensation has been reported in at least one other CF/FM LDC echolocator, *Noctilio albiventris* (Roverud and Grinnell 1985). Similarly, the heteroharmonic strategy for target-range computation has been associated primarily with HDC echolocators, having been describe in *P. parnellii* (Suga et al 1978) and *Rhinolophus rouxii* (Schuller et al 1991). It has also been described in *P. quadridens* (Hechavarría et al 2013) and has been hypothesised to be a basal trait in the *Pteronotus* genus and possibly even of the whole Mormoopidae family (Mora et al 2013). This mix of traits commonly linked with HDC echolocation throughout the phylogeny gives rise to a second possible hypothesis; that HDC echolocation evolved ancestrally in the *Pteronotus* genus, but with a subsequent reversion after *Phyllodia* diverged.

Molecular research in bats and cetaceans has highlighted genes putatively associated with echolocation by identifying genes known to be associated with hearing or vocalisation in other species that shows signals of evolving under a different selection regime in one or both these lineages in comparison to related non-echolocating taxa. Accelerated evolution in the vocalisation gene *Foxp2* has been detected in echolocating bats (Li et al 2007). The high-frequency hearing gene *Prestin* has been shown to create a monophyletic clade of echolocating bats, in contrast to the species phylogeny (Li et al 2008). Evidence of positive selection acting on echolocating bat species and cetaceans has been described in the hearing genes *Tmc1* and *Pjvk*, as well as convergent evolution in these genes between

the echolocating bat lineages and between echolocating bats and dolphins (Davies et al 2012). Parallel evolution was found between the two echolocating bat lineages in the outer hair cell gene *Kcnq4*, including shared fixation of amino acid substitutions by positive selection (Liu et al 2012). Parallel evolution was shown between the two lineages of echolocating bats and between echolocating bats and the dolphin in the auditory genes *Otof*, *Cdh23* and *Pcdh15*, with the latter two also showing signals of positive selection in echolocators (Shen et al 2012). Genome-wide scans have detected signatures of convergence between the two lineages of echolocating bats and between echolocating bats and dolphins in hundreds of genes, including genes linked to hearing and to vision (Parker et al 2013). Relatively little molecular research has looked at genetic underpinnings of different echolocation call structures, and there is currently not a clear picture of whether similar mechanisms were in play in both lineages. For example, there is a parallel substitution in the vocalisation gene *Foxp2* between *Hipposideros* and *P. parnellii*, but not in *Pteronotus quadridens* or *Pteronotus macleayii* (Li et al 2007). However, there is no convergence between *P. parnellii* and the Rhinolophidae in the high-frequency hearing gene *Prestin*, which does not appear to have evolved under positive selection in *Pteronotus* (Shen et al 2011).

The sharp divergence in echolocation call strategy between the *Phyllodia* subgenus and the rest of the *Pteronotus* genus provides an opportunity to study the genetics underlying the evolution of HDC echolocation while attempting to elucidate where this trait evolved in the genus. I started with two hypotheses; that HDC echolocation was the basal condition for the *Pteronotus* genus, with a subsequent reversion after the divergence of the *Phyllodia* subgenus, and that HDC echolocation evolved in the *Phyllodia* subgenus after its divergence. I generated transcriptomes from seven *Pteronotus* species and two *Mormoops* species using high-throughput sequencing. I used these to look for signals of increased rates of positive selection in hearing genes on the branches ancestral to

Pteronotus and to *Phyllodia*, performing independent selection tests on single gene alignments and comparing the results to control branches throughout the Mormoopidae. Finally, I compared genes thus identified as evolving under positive selection on these branches with those evolving under positive selection in a bat that has independently evolved HDC echolocation, *Rhinolophus sinicus*, to look for evidence of parallel substitutions in these genes.

Materials and methods

Tissue collection, RNA extraction and sequencing

Mormoops megalophylla, *Pteronotus davyi* and *Pteronotus mesoamericanus* wing punches were collected from wild-caught individuals in Belize by Stephen Rossiter in 2012 (Belize forest department CD/60/3/12 (11) to M. Brock Fenton). Samples were stored in ethanol in the field and transferred to -80 ^{o}C storage subsequently. *Mormoops blainvillei*, *P. pusillus* and *P. quadridens* wing punches were collected from wild-caught individuals in the Dominican Republic by Stephen Rossiter in 2014 (DR permit 0673). These tissue samples were stabilised in RNALater immediately after collection in the field and subsequently stored at -80^{o}C. *Pteronotus gymnonotus* (voucher ID 117653_F54974, from Suriname), *Pteronotus macleayii* (voucher ID 120832_F53533, from Jamaica) and *Pteronotus personatus* (voucher ID 117617_F54938, from Suriname) tissue samples were provided by the Royal Ontario Museum. RNA was extracted from all mormoopid samples using the Qiagen RNEasy mini kit, with the Illumina TruSeq RNA Sample Prep Kit v2 used to build sequencing libraries with standard Illumina indexing. RNA extraction and library preparation were carried out by Dr Monika Struebig, Dr Kalina Davies and myself, with indexed libraries then being pooled prior to sequencing. *P. mesoamericanus* and *P. davyi* were sequenced by BGI (China) on the Illumina HiSeq 2000 using 90 base pair paired end sequencing. *M. megalophylla*, *P. pusillus* and *P. quadridens* were sequenced by TGAC (Norwich, UK) on the Illumina HiSeq 2500 with 150 base pair paired end sequencing. *M. blainvillei*, *P. gymnonotus*, *P. macleayii* and *P. personatus* were sequenced by Barts and the London Genome Centre

(London, UK) on the Illumina NextSeq using 150bp paired end sequencing. Sequencing information for mormoopid species is summarised in Table 2.1.

P. mesoamericanus and *P. pusillus* are both members of the subgenus *Phyllodia*. As the change in echolocation call frequency occurred at the root of this genus, I have handled them together as *Phyllodia* through the rest of this chapter.

Read processing and assembly

The heart, cochlea and eye raw reads within a species were sequenced and assembled for another study (Sadier A, Davies K, Yohe L, Yun K, Donat P, Hedrick BP, Dumont E, Davalos L, Rossiter S and Sears KE, unpublished data https://www.biorxiv.org/content/early/2018/04/12/300301). mRNA is differentially expressed in different tissues, with many transcripts relevant to hearing only expressed in the cochlea, making this tissue of particular interest to sequence. Transcripts expressed in the eye are also interesting as this is another sensory tissue, with sensory trade-offs hypothesised to exist between vision and hearing (Zhao et al 2009). Finally, including the heart tissue provides a baseline set of sequences from a non-sensory tissue so that I have a point of comparison to see whether sensory genes, particularly those associated with hearing, are experiencing an elevated rate of selection in comparison to other genes. The transcriptomes for these three tissues were pooled by species after sequencing and used assembled in Trinity by Dr Kalina Davies (r20140717, Grabherr et al. 2011) to generate three mixed tissue assemblies, one per species. The *P. mesoamericanus* and *P. davyi* sequences from Belize were assembled independently in Trinity by Dr Kalina Davies. For all other samples, I removed residual adaptor sequences, clipped reads where the average quality of a four-base pair sliding window dropped below five and then removed

any reads below 36 base pairs in length using Trimmomatic (version 0.35, Bolger, Lohse and Usadel 2014). I assembled transcriptomes for each species from the resulting trimmed read files using Trinity (r20140717, Grabherr et al. 2011). In total, this generated 12 assemblies; one for each of *M. megalophylla*, *M. blainvillei*, *P. mesoamericanus*, *P. pusillus*, *P. personatus*, *P. gymnonotus*, *P. davyi*, *P. macleayii* and *P. quadridens* from wing punch tissue, plus additional combined heart, cochlea and eye transcriptome assemblies for each of *M. blainvillei*, *P. pusillus* and *P. quadridens*.

Gene identification and multiple sequence alignment

I used a reciprocal BLAST procedure to identify genes with one-to-one orthology with human genes from the assembled mormoopid transcriptomes and selected outgroups, discussed below. I used Ensembl BioMart to obtain the list of human protein-coding genes with known one-to-one orthologues in at least one of *Myotis* or *Pteropus* to use as query sequences (Ensembl release 78). I performed blastx searches using the bat nucleotide sequences as queries and the human amino acid sequences as a database, then tblastn searches using the bat transcriptomes as the database and the human proteins as queries. I used custom perl scripts and scripts packaged with Trinity to identify and extract the most likely orthologues from the bat transcriptomes into FASTA files.

Transcriptome sequencing captures only the genes that were being expressed in the sequenced tissue at the point it was collected and stored. I had samples from cochlear tissue, which is of particular interest to the present study, for *P. pusillus*, *P. quadridens* and *M. blainvillei*. In order to maximise the value of the generated alignments, I decided to build two separate data sets using *Pteronotus* sequences; one using muscle and wing tissues but featuring a large amount of background context, and one focussed very

precisely on the specialist tissues with a small number of background species chosen based on genetic completeness to maximise the coverage of these genes within the alignments.

The first data set, the sensory tissue data set, used separate heart, eye and cochlea transcriptomes for *M. blainvillei*, *P. pusillus* and *P. quadridens*. I aligned these with cDNAs derived from three genome assemblies in GenBank. I used genome assemblies for outgroups in this data set to increase the chances of genes expressed in the specific tissues sequenced (heart, eye and cochlea) being represented in the outgroups. To maintain a similar taxonomic distribution among the ingroups and outgroups, I downloaded the cDNAs for two *Myotis* species (*Myotis brandtii*; GenBank assembly GCA_000412655.1, and *Myotis davidii*; GenBank assembly GCA_000327345.1) and one other vesper bat (*Eptesicus fuscus*; GenBank Assembly GCA_000308155.1) (Figure 2.3).

The second data set, the Yangochiroptera data set, contained transcriptomes generated from muscle tissue samples for *M. megalophylla*, *M. blainvillei*, *P. mesoamericanus, P. pusillus*, *P. personatus*, *P. davyi*, *P. gymnonotus*, *P. macleayii* and *P. quadridens*. I combined these with 23 unpublished yangochiropteran transcriptomes (Table 2.2) as outgroups. These included 12 members of the family Phyllostomidae (*Macrotus waterhousii*, *Micronycteris microtis*, *Desmodus rotundus*, *Trachops cirrhosis*, *Glossophaga soricina*, *Erophylla sezekorni*, *Anoura geoffroyi*, *Hylonycteris underwoodi*, *Lichonycteris obscura*, *Lonchophylla thomasi*, *Sturnira lilium*, *Uroderma bilobatum*), four Vespertilionidae (*Bauerus dubiaquercus*, *Myotis elegans*, *Myotis ricketti*, *Rhogeessa aeneus*) and representatives from the Emballonuridae (*Peropteryx kappleri*, *Saccopteryx bilineata*, *Rhynchonycteris naso*), Molossidae (*Molossus sinaloae*), Noctilionidae (*Noctilio leporinus*), Nycteridae (*Nycteris tragata*) and Thyropteridae (*Thyroptera tricolor*) (Figure 2.4). For all these species except *M. ricketti*, RNA was extracted from wing punches by either Dr Monika Streubig or Dr Kalina Davies using the Qiagen

RNEasy mini kit, with cDNA libraries then being constructed using the Illumina TruSeq RNA Sample Prep Kit v2 with standard Illumina indexing. *M. waterhousii* was sequenced by TGAC on the Illumina HiSeq 2500 with 150 base pair paired end sequencing; the other libraries were pooled and sequenced by BGI on the Illumina HiSeq 2000 using 90 base pair paired end sequencing. The reads were then assembled into transcriptomes using Trinity by Dr Kalina Davies. The *Myotis ricketti* sequence was provided by Dr Dong Dong with sequencing and assembly details provided in (Dong et al 2013). Additionally, I included the GenBank cDNAs for *Myotis brandtii* (GCA_000412655.1) and *Myotis davidii* (GCA_000327345.1).

I additionally constructed a third data set, the *Rhinolophus* data set, to carry out selection tests in a rhinolophid bat representing the other independent lineage of HDC echolocators. In this case, cDNA sequences were downloaded from GenBank for *Rhinolophus sinicus* (GenBank assembly GCA_001888835.1), the non-echolocating yinpterochiropteran species *Pteropus alecto* (GenBank assembly GCA_000325575.1) and *Rousettus aegypticus* (GenBank assembly GCA_001466805.2) and the Yangochiroptera *Eptesicus fuscus* (GenBank assembly GCA_000308155.1), *Miniopterus natalensis* (GenBank assembly GCA_001595765.1) and *Myotis brandtii* (GenBank assembly GCA_000412655.1). In these cases, no reciprocal blast procedure was required due to the use of downloaded cDNAs. Additionally, I included cDNAs derived from the *Desmodus rotundus* genome (Mendoza et al 2018) by blasting the *E. fuscus* cDNAs against the genome assembly (Figure 2.5).

For all data sets, to enable codon-based multiple sequence alignments I used custom perl scripts to remove sequences with premature stop codons and to add Ns to the tails of any sequence that was not a multiple of three nucleotides in length to make it up to an appropriate length. I discarded any gene within a data set for which we had identified orthologous sequences from fewer than four species. I performed multiple sequence

alignments using the codon-aware alignment software PRANK (version 120626, Löytynoja and Goldman 2008), with Guidance (version 1.5, Penn et al. 2010) to assess the quality of the PRANK alignments. I executed this alignment procedure iteratively, removing any sequences that Guidance marked as low quality and repeating the alignment with only the remaining species until no sequences were indicated for removal. At this point, I retained only the positions in the alignment that Guidance flagged as having been aligned with high confidence. I further pruned alignments using custom perl and python scripts to remove positions where there was data for fewer than half of the included species and to remove sequences that were shorter than 300 base pairs in length. Finally, I retained alignments that had been left with at least four species, including a representative of *Phyllodia* in the sensory tissue and Yangochiroptera data set or *R. sinicus* in the *Rhinolophus* data set.

Selection tests

To identify genes that may be related to the change in echolocation call strategy, I first looked for evidence of positive selection acting at different levels within the Mormoopidae. I used two different models when testing for positive selection; test 2 of the new branch-site model A in PAML codeml (version 4.8, Yang 2007) and the aBSREL model (Smith et al 2015) in HyPhy (version 2.220170207beta, Pond, Frost and Muse 2004). These are both explicit models of positive selection that calculate an ω value – the rate of non-synonymous substitutions divided by the rate of synonymous substitutions – and infer positive selection if this value is greater than one. PAML generates log likelihood values for both an alternate model and for a corresponding null model, which differ only in that the ω value is a free parameter in the alternate model and fixed at one

in the null, and you assess the model fit of the two models by perform a log-likelihood ratio test. HyPhy works similarly, but rather than assuming each branch can be described by three ω classes, it uses AIC_c to determine the optimal number of classes.

I additionally tested for changes in the strength of selection using the RELAX model (Wertheim et al 2014) in HyPhy (version 2.220170207beta, Pond, Frost and Muse 2004). This model generates a p-value to indicate whether there has been a significant change in selection pressure on the foreground branch and a k-value to indicate the direction of the change. A k-value of less than one indicates a relaxation in the strength of all modes of selection acting on the branch, whereas a k-value greater than one shows an increase. In addition to using this model to look for indications of reversion, I used RELAX to confirm whether genes returning a significant result under the positive selection models were experiencing an increase in selection pressure, rather than the apparent signal of positive selection being the result of a relaxation in purifying selection (Zhang, Nielsen and Yang 2005).

In both the sensory tissue and Yangochiroptera data sets I tested two hypotheses by modelling positive selection on specific branches; firstly, that the transition to HDC echolocation occurred specifically in *Phyllodia* and secondly that it was an ancestral trait for the genus with a reversion after *Phyllodia* had diverged. To test the former of these hypotheses, I used the branch ancestral to the subgenus as the foreground and modelled selection on all other *Pteronotus* tips in the phylogeny to provide context to the results. I assigned the ancestral branch of *Pteronotus* as the foreground to test the second hypothesis, with the ancestral *Mormoops* branch providing a point of comparison. In the Yangochiroptera data set I also modelled selection on the internal *Pteronotus* branch after *Phyllodia* had diverged to look for evidence of trait reversion. In the *Rhinolophus* data set, where I was generating a list of genes under positive selection in a separate constant-frequency echolocator as a point of comparison rather than testing a specific hypothesis,

I only ran models with *R. sinicus* in the foreground. I further investigated these hypotheses using the RELAX model, which was applied to the same branches in all data sets with the exception that in the Yangochiroptera data set we did not apply this model to the *P. davyi*, *P. gymnonotus* or *P. macleayii* tips.

To reduce false positives due to gene assembly or alignment errors, I filtered the results based on the clustering of positively selected sites as indicated by the Bayes Empirical Bayes values calculated in PAML. I removed genes from subsequent analysis where the median interval between these sites was less than 10 codons (as described in Tsagkogeorga et al 2015). The remaining p-values from the PAML branch-site models and the HyPhy RELAX models were subjected to correction for multiple testing in SGoF+ (version 3.8, Carvajal-Rodriguez and Uña-Alvarez 2011), a tool that applies several different multiple correction algorithms to a list of p-values, prior to being assessed for significance. The HyPhy aBSREL model performs its own multiple testing correction step when executed on multiple branches within a phylogeny, so no further corrections were applied to the results from this model. Results were considered both prior to and post multiple testing correction.

As the focus of this study is the evolution of echolocation, I concentrated my analysis on hearing genes. I defined hearing genes as those that were associated with ear or hearing disorders in the Rat Genome Database (Shimoyama et al 2015, accessed 03/02/2018). All genes that were identified as experiencing positive selection or a change in the strength of selection on one of the foreground branches were cross-referenced with this database, with the hearing genes identified then being verified by eye to check for evidence of assembly or alignment errors that may suggest that the result was a false positive.

Results

Gene identification and multiple sequence alignment

I constructed three data sets for separate analysis. The first of these, hereafter the sensory tissue data set, included transcriptomes from the hearts, eyes and cochlear of three mormoopid species. These were aligned with CDS data from GenBank for a further three yangochiropteran species as outgroups (Figure 2.3). The second data set, henceforth the Yangochiroptera set, was built by aligning transcriptomes generated from muscle tissue or wing punches of 34 yangochiropteran species (Figure 2.4), including 7 species of *Pteronotus*. The third data set, the *Rhinolophus* set, was derived from multiple sequence alignments of CDSs downloaded from GenBank for seven species, including *R. sinicus* and two species of Pteropodidae from the Yinpterochiroptera (Figure 2.5).

I identified up to 13,806 putative transcripts with single-copy (one-to-one) orthology with human genes in individual species from the family Mormoopidae in the sensory tissue data set and 12,483 single-copy orthologs in the Yangochiroptera set. Only a subset of these genes passed alignment and filtering steps prior to selection testing (Figure 2.6 for *Phyllodia* and ancestral *Pteronotus* branches; detail for all models in Table 2.3). 12,149 genes were successfully aligned in the *Rhinolophus* data set, with 11,881 genes passing through the post-alignment filtering steps for use in selection testing.

Positive selection in Pteronotus

I tested two hypotheses concerning where the transition to HDC echolocation featuring long CF calls took place in the *Pteronotus* phylogeny. The first of these was that HDC echolocation evolved on the branch leading to the *Phyllodia* subgenus; the second that this was an ancestral state in the *Pteronotus* genus with a reversion following the divergence of the *Phyllodia*. I tested these hypotheses by modelling selection rates on specific branches of the phylogeny using two independent models that explicitly test for positive selection - the PAML branch-site model and the HyPhy aBSREL model.

The sensory tissue data set featured a single member of the *Phyllodia*. I modelled selection rates independently on the *Phyllodia* tip and on the node ancestral to the *Pteronotus* genus, as well as *P. quadridens* and *M. blainvillei* (the only representative of the *Mormoops* genus in this data set) tips to provide appropriate comparisons. I found 438 genes under positive selection on the ancestral *Pteronotus* branch using PAML (21 after correction for multiple testing) or 109 using HyPhy. This compared to 367 genes (6 after correction for multiple testing) using PAML or 89 using HyPhy in *M. blainvillei* (Table 2.4). There were 340 genes (33 after correction for multiple testing) that showed signals of positive selection in *Phyllodia* using PAML or 99 using HyPhy, which compared 268 (12 after multiple testing correction) for PAML or 106 with HyPhy respectively in *P. quadridens* (Table 2.4).

I modelled selection rates on the ancestral *Pteronotus* branch with the ancestral *Mormoops* branch as a point of comparison in the Yangochiroptera data set. I also modelled selection on the branch ancestral to *Phyllodia* along with each *Pteronotus* tip present, as well as on the internal *Pteronotus* node after *Phyllodia* diverged from the rest of the genus. In this case, 149 genes (three after multiple testing correction) were found

to be under positive selection in *Pteronotus* by PAML or 27 by HyPhy, compared to 163 (one after multiple testing correction) and 17 respectively on the ancestral *Mormoops* branch. PAML identified 110 genes (three after multiple testing correction) evolving under positive selection in *Phyllodia*, with HyPhy identifying 61. For PAML this was a higher number pre-correction than on any other *Pteronotus* tip, though not post-correction, with the HyPhy result being much higher than in any of the other congeners (Table 2.4).

Across both data sets, most genes identified as being positively selected by the PAML branch-site model following correction for multiple testing were also identified as positively selected by the HyPhy aBSREL model (Figure 2.7). However, there was little concordance in genes that showed signatures of positive selection between the sensory tissue and the Yangochiropteran data sets (Figure 2.8). This remained true when considering only the 6,447 genes that were present in both data sets (Table 2.5). The vesper bats had the most genes found to be evolving under positive selection, with *P. parnellii* falling towards the middle of the range of values in both data sets (Figures 2.3 and 2.4).

Relaxation of selection in Pteronotus

In addition to explicitly testing for positive selection, I looked at changes in selection pressures on target branches in both data sets using the HyPhy RELAX model. A significant p-value under this model indicates that there is a change in selection pressure on the target branch relative to the background, with an associated k-value providing information as to whether that represents an intensification or a relaxation in the strength of any form of selection otherwise operating on that branch.

In the sensory tissue data set, I applied the RELAX model to the same set of branches previously tested for positive selection (Table 2.6). RELAX identified changes in selection pressure in 1,064 genes on the ancestral *Pteronotus* branch (22 after correction for multiple testing). There was an overlap of 121 genes between those detected as experiencing a change in selection pressure on the ancestral *Pteronotus* branch and the genes previously identified as evolving under positive selection on this branch by PAML (of which 108 were experiencing an intensification in selection pressure and 13 a relaxation), with 43 having been detected by the HyPhy aBSREL model (41 intensification, 2 relaxation) (Table 2.7). There was evidence for a change in selection pressure on 866 genes (9 after correction for multiple testing) in *Phyllodia*. This included 103 genes found to be evolving under positive selection by PAML (93 experiencing an intensification of selection, 10 experiencing a relaxation in the strength of selection) or 51 that were detected by HyPhy (48 experiencing an intensification in the strength of selection, 3 a relaxation) (Table 2.7).

In the Yangochiroptera data set, I modelled selection on the same set of branches as with the positive selection models, except for the *P. macleayii*, *P. gymnonotus* and *P. davyi* tips (Table 2.6). I detected 439 genes showing a change in the strength of selection on the ancestral *Pteronotus* branch, though none remained significant after correction for multiple testing. Of these genes, 51 had been identified as evolving under positive selection by the PAML branch-site model (46 showing an intensification in the strength of selection, 5 showing relaxation) or 18 by the HyPhy aBSREL model (all showing an intensification in the strength of selection) (Table 2.7). There were 284 genes (4 after correction for multiple testing) that showed a change in selection strength in *Phyllodia*, of which 39 had been identified by PAML as evolving under positive selection (37 showing an intensification in the strength of selection and 2 showing a relaxation) or 13 by HyPhy aBSREL (12 of which showed an intensification in the strength of selection)

(Table 2.7). On the internal *Pteronotus* branch after the divergence of *Phyllodia*, 120 genes showed evidence of a change in the strength of selection (1 after correction for multiple testing). Of the genes that PAML reported as evolving under positive selection on this branch, 17 were also detected by RELAX (15 showing an intensification of selection), or 3 out of the genes that had been detected as being under positive selection by HyPhy (all showing an intensification of selection) (Table 2.7).

Hearing genes

The Rat Genome Database (Shimoyama et al 2015) identifies 609 genes (as of point of access, 05/02/2018) as being associated with ear and hearing disorders (hereafter hearing genes). I used this database to identify hearing genes within our selection test results.

The sensory tissue dataset included alignments for 388 hearing genes, of which 57 unique genes were identified as being under positive selection by at least one model, with 50 being retained following visual inspection of the alignments (Genes rejected: *Crym, Espn, Gsdme, Prkcb, Tjp2, Tnfrsf1a, Tnfrsf1b*) (Table 2.8). There were 19 hearing genes with signals of evolving under positive selection on the ancestral *Pteronotus* branch, of which three (*Dnmt3a*, *Slc29a3* and *Ush1g*) were detected by both the PAML branch-site and the HyPhy aBSREL models (Figure 2.9). In *Phyllodia* 15 genes were found to be evolving under positive selection, with four (*Col4a5*, *Rela*, *Serpinf1* and *Rela*) being identified by both the models of positive selection (Figure 2.9). Under the HyPhy RELAX model, 26 hearing genes (six after multiple testing correction) showed intensification of selection pressures and five (none after multiple testing correction) showed relaxation of selection on the ancestral *Pteronotus* branch (Table 2.9). This included five genes that had been identified as evolving under positive selection (*Dnmt3a, F2, Fn1, Kdr, Map2*, all showing

an intensification in selection pressure) (Table 2.10). In the *Phyllodia* branch, 30 (eight after multiple testing correction) were experiencing an intensification of selection pressures and 12 (two after multiple testing correction) experiencing a relaxation (Table 2.9). Eight of these genes had been identified as evolving under positive selection (Intensification: *Col4a5, Lars2, Mtor, Opa1, Rela, Tnc*. Relaxation: *Gmnn, Mtr*) (Table 2.10).

There were 225 hearing genes in the Yangochiroptera data set prior to selection testing. 25 of these were found to be evolving under positive selection by at least one model, with three (*Myh14*, *Tnfrsf1a*, *Tnfrsf1b*) being rejected upon visual inspection of the alignments (Table 2.8). The ancestral *Pteronotus* branch carried five hearing genes evolving under positive selection, with *Nefh* being identified as evolving under positive selection by both the PAML branch-site and the HyPhy aBSREL models (Figure 2.10). In the *Phyllodia* subgenus, seven hearing genes were evolving under positive selection, with two (*Rela* and *Serpinf1*) being detected by both models of positive selection (Figure 2.10). On the *Pteronotus* branch, 11 genes were identified as experiencing an intensification in the strength of selection and eight a relaxation in the strength of selection by the HyPhy RELAX model (Table 2.9), though none of these were significant after multiple correction and there was no overlap between these genes and those evolving under positive selection. In *Phyllodia*, eight (three after multiple testing correction) genes were experiencing an intensification in the strength of selection and three (none after multiple testing correction) a relaxation (Table 2.9). All three of the genes found to be experiencing an intensification in the strength of selection after multiple testing correction had also been identified as evolving under positive selection (*Clcn3, Serpinf1* and *Rela*) (Table 2.10).

Three hearing genes, *Lars2*, *Serpinf1* and *Rela*, were found to be evolving under positive selection in *Phyllodia* in both the sensory tissue and the Yangochiroptera data sets. All

three of these genes were also identified as experiencing an intensification of selection pressures by the RELAX model. In all three of these genes, all private substitutions present in *Phyllodia* in the overlapping regions of the two alignments were present in both alignments (*Lars2*: C36H, H64R, R163Q; *Serpinf1*: T352Q, P353H, S364P; *Rela*: Q133A). No hearing genes evolving under positive selection in the ancestral *Pteronotus* branch were shared between both data sets. While there appeared to be a slight enrichment for hearing genes amongst the set of positively selected genes on both the ancestral *Pteronotus* and *Phyllodia* branches, this is also seen at other *Pteronotus* tips (Table 2.11).

Comparison of high duty-cycle echolocators

In the *Rhinolophus* data set I identified 978 genes (155 after correction for multiple testing) as being under positive selection in *Rhinolophus sinicus* using the PAML branch-site model and 228 using the HyPhy aBSREL model. Of these, 163 were found to be evolving under positive selection by both models. Dong et al (2016) had previously found 577 genes evolving under positive selection in the same *R. sinicus* genome assembly, 488 of which were present in my alignments of the *Rhinolophus* data set. I found signatures of positive selection in 44 of the genes found to be under positive selection by Dong et al using the PAML branch-site model on my assemblies, or 12 with the HyPhy aBSREL model. Of 251 hearing genes subjected to selection tests in this data set, I identified 14 as being under positive selection in *R. sinicus* using the HyPhy aBSREL model and 54 with the PAML branch-site model (10 after correction for multiple testing). Five of these genes (*Alb, Fgfr3, Ltf, Nbn, Syp*) were found by both models. In comparison, 29 of the genes identified as being under positive selection by Dong et al were identified as hearing genes

in the rat genome database. Two hearing genes, *Cdh23* and *Slc26a5*, were found to be under positive selection in *R. sinicus* by both my analyses and Dong et al.

Phyllodia and *R. sinicus* shared 34 genes under positive selection in based on my alignments, with 20 genes showing signals of positive selection between *Phyllodia* and Dong et al's results. There were 57 shared genes under positive selection on the ancestral *Pteronotus* branch and my *R. sinicus* analyses and 27 shared with Dong et al's. Of the genes with signatures of positive selection in both *R. sinicus* and *Phyllodia*, five were hearing genes (*Cacna1d, Cat, Col4a5, Gmnn, Map2*), as were six in *R. sinicus* and the ancestral *Pteronotus* branch (*Agt, Eral1, Icam1, Map2, Nefh, Ush1g*) (Table 2.12). *Col4a5* had shared private substitutions between *Phyllodia* and *R. sinicus*, while *Eral1* and *Ush1g* had shared private substitutions between *R. sinicus* and all *Pteronotus* species. *Cat* and *Icam1* showed different substitutions in the same positions in *R. sinicus* and in *Phyllodia* or in the *Pteronotus* species respectively (Figure 2.11).

The proportion of genes in common with *R. sinicus* found to be under positive selection that were also hearing genes was slightly higher in *Phyllodia* and on the ancestral *Pteronotus* branch than in *P. quadridens* and the ancestral *Mormoops* branch, used as control branches (Table 2.13).

Discussion

The origin of HDC echolocation in the Pteronotus *genus*

While most species in the *Pteronotus* genus use LDC echolocation, members of the *Phyllodia* subgenus use HDC echolocation, with individuals separating the long sonar pulses from the returning echo by frequency rather than by time. Other families of bats that use HDC echolocation are distantly related, with *Phyllodia* appearing to have evolved a convergent phenotype independently. However, the evolutionary history of echolocation call structure within the *Pteronotus* genus is not immediately obvious. Traits commonly associated with HDC echolocation have been described in other members of the genus, such as Doppler-shift compensation in the *P. personatus* complex (Smotherman and Guillén-Servent 2008) and the use of the heteroharmonic target range computation strategy in *P. quadridens* (Hechavarría et al 2013). Additionally, all species in the *Pteronotus* genus have short CF tails on their calls, unlike other members of the family Mormoopidae (Mora et al 2013). I have tested two hypotheses concerning the evolution of HDC echolocation within the genus here. The first is that HDC echolocation dominated by long narrowband elements evolved basally in *Pteronotus*, with a reversion to predominantly FM calls after the divergence of the *Phyllodia* subgenus. The second is that the shift in echolocation type occurred specifically on the branch ancestral to the *Phyllodia*, though this does not explain why some features associated with HDC echolocation appear elsewhere in the genus.

The absolute proportion of genes showing signatures of evolving under positive selection in *Phyllodia* is consistent with the rest of the species in the *Pteronotus* genus. This remains true when considering hearing genes in particular. While there may be a slight

enrichment for hearing genes in the set of positively selected genes than in the entire set of genes tested for selection in *Phyllodia* and in the ancestral *Pteronotus* branch the signal is not strong, with some of the other tip species within *Pteronotus* showing a similar enrichment. Similarly, when I tested for relaxation of selection pressures, there was no strong signal within the genus of changes in the strength of selection on hearing genes that could be associated with the change in echolocation call structure. As such, I did not find evidence to support one hypothesis with regards to the evolution of HDC echolocation within the *Pteronotus* over the other.

I was able to highlight some genes of potential interest with relation to the evolution of HDC echolocation. I detected signals of positive selection in 492 genes in *Phyllodia*, including 23 hearing genes, and 587 (of which 27 were hearing genes) in the ancestral *Pteronotus* branch. There was an overlap of 44 that were identified as evolving under positive selection in the *Phyllodia* in both the sensory tissue and the Yangochiroptera data sets, including three hearing genes – *Lars2*, *Rela* and *Serpinf1*. Similarly, 49 were detected on the *Pteronotus* branch in both data sets, though none of these were hearing genes. As the alignments in the sensory tissue and Yangochiroptera data sets were completely independent, derived from separate RNASeq reads sequenced from different tissues, finding the same gene evolving under positive selection in both data sets is a strong signal – especially considering that there were shared private substitutions in all three genes. *Lars2* encodes a tRNA synthetase and is associated with hearing loss via Perrault syndrome (Pierce et al 2013). *Serpinf1* encodes a protein involved in neuronal differentiation and which is an angiogenesis inhibitor; it has been linked to conditions involving retinal degradation, and also to conductive hearing loss via familial otosclerosis (Ziff et al 2016). This gene has previously been observed to be a target of positive selection in the dolphin against a background of non-echolocating taxa (Nery, González and Opazo 2013). *Rela* encodes part of the most abundant NF-kappa-B transcription

factor complex and, while it does not appear to be directly linked to deafness or hearing loss, its localisation in inner ear hair cells has been associated with the effects of drug treatment following ototoxic insult (Layman and Zuo 2015).

Parallelism in the evolution of HDC echolocation between lineages

HDC echolocation has evolved at least twice independently in distantly-related bats; in the yangochiropteran *Phyllodia* subgenus and at a deep node within the Yinpterochiroptera. These two groups of HDC echolocators display shared morphological and behavioural adaptations, including a modified cochlear basilar membrane creating an acoustic fovea (Davies, Maryanto and Rossiter 2013) and the use of Doppler-shift compensation in flight (Henson, Schuller and Vater 1985). It often appears to be the case that where a shared phenotype has involved, there are underlying changes in the same genes (as reviewed in Elmer and Meyer 2011). There are numerous exceptions to this, and even where the same genes are involved there is no guarantee of finding parallel substitutions; a variety of genetic routes can often lead to a similar phenotypic condition (e.g. Steiner et al 2009, Wittkopp et al 2012). I have investigated whether there are any signals of parallelism at the genetic level between two distinct evolutionary origins of HDC echolocation, focussing on genes with known associations with deafness and hearing loss.

I identified 52 genes with signals of positive selection in both the *Phyllodia* genus and in *R. sinicus*, including five hearing-associated genes, and 77 (of which six were hearing genes) under positive selection in both *R. sinicus* and on the ancestral *Pteronotus* branch. Of these genes, four – two in *Phyllodia* (*Cacna1d* and *Col4a5*) and two on the ancestral *Pteronotus* branch (*Eral1* and *Ush1g*) – had shared private substitutions with *R. sinicus*.

Cacna1d encodes a subunit of voltage-dependent calcium channels and has been linked both to forms of congenital deafness (Platzer et al 2000, Baig et al 2011) and of age-related hearing impairment (Chen et al 2013). *Col4a5* encodes one subunit of type IV collagen, which is a major component of basement membranes. It is linked to hearing loss via X-linked Alport syndrome (Barker et al 1990). This gene also been noted to be under accelerated evolution in moles, which was hypothesised to be due to its role in the skin epithelium basilar membrane (Partha et al 2017). *Eral1* encodes a mitochondrial GTPase linked to hearing loss via Perrault syndrome (Chatzispyrou et al 2017). *Ush1g* is linked to hearing loss both via its implication in type 1 Usher syndrome (Mustapha et al 2002), and by associated with non-syndromic hearing loss (Oonk et al 2015). The protein encoded by this gene, SANS, is produced in the sensory hair cells and interacts with harmonin, a protein present in growing hair bundles in inner sensory cells (Weil et al 2003). *Ush1g* has previously been linked to echolocation, being one of the hearing genes displaying codon usage bias in the LDC echolocator *Myotis davidii*, but not in the non-echolocating fruit bat *Pteropus alecto* (Hudson et al 2014).

HDC echolocation is a complex trait which implicates multiple different physiological systems. While I have focussed on auditory genes here, I carried out selection tests on genome-wide transcriptome data and have thus identified a range of other genes under selection on my branches of interest, some of which have the potential to be associated with the change in echolocation call strategy. HDC echolocation involves rearrangement of the auditory cortex, which has not proceeded in the same way in both rhinolophids and the *Phyllodia* subgenus (O'Neill 1995), so genes with roles in brain development may be implicated. Similarly, the sensory trade-off hypothesis postulates that with the evolution of a more energy-demanding echolocation call, other senses, such as vision, may have been reduced, so searching the results for genes associated specifically with vision may also be of interest. One risk with taking this kind of approach is that in such a large data

set there are bound to be positive results, and the challenge is to approach its analysis with specific biological questions and hypotheses to avoid drawing spurious conclusions. Additionally, the use of an RNASeq approach to generate a large amount of coding sequence data limits the power of the investigations that can be carried out. The Yangochiroptera data set was composed of transcriptomes derived from easily-accessible tissues, meaning that genes transcribed in specific brain or sensory tissues will likely be missing entirely from the data set. The sensory tissue data set included transcriptomes derived from cochlear samples and used genomic rather than transcriptomic data for the background species to increase our power to detect hearing genes; the impact of the difference can be seen from the fact that two thirds of the hearing genes listed in the rat genome database (Shimoyama et al 2015) were detected in the sensory tissue data set, compared to only a third in the Yangochiroptera set.

Methodological reflections

There was a substantial difference in the genes found to be under positive selection by the PAML branch-site and the HyPhy aBSREL models when executed on the same input data. As I ran the HyPhy aBSREL model on each tip within the phylogeny, HyPhy applied its own multiple testing correction prior to presenting the results; as such, the number of significant genes discovered by HyPhy was consistently lower than that discovered by PAML prior to correcting the PAML results for multiple testing. Applying correction for multiple testing to the PAML results appeared to disproportionately reduce the number of genes that had been found to be positively selected only by this model, with a smaller effect on the number of genes that had been found to be selected by both models. On the assumption that a gene found to be evolving under positive selection by two separate

models is more likely to be reliable, this suggests that multiple testing correction does bring benefit on whole-genome data sets, even though it does entail a loss of power of detection.

The limited overlap in genes that were found to be under positive selection between the sensory tissue and Yangochiroptera data sets, even among genes that were present in both, was unexpected. There are a few possible explanations for this. The tissues that the sequence data were derived from were different for the two data sets and so may have expressed different isoforms or splice variants of transcripts. The range of species included in the background on the two data sets was vastly different, which also may have affected whether a substitution on a focal branch was likely to be an indication of the action of positive selection on that branch. However, in some cases it was less clear; *Ush1g* was only identified as evolving under positive selection in one data set, despite the phylogeny in both being almost identical and the sequences being similar enough to easily align and identify substitutions in the same positions. PAML and HyPhy both work on maximum likelihood optimisation, and so it is possible that incorrect results can be obtained as a result of local optima in the probability space.

Conclusion

Of 123 genes that I have identified as being under positive selection in both lineages of HDC echolocators in Chiroptera, eleven have known associations with hearing loss or deafness. While HDC echolocation is undoubtedly underlain by a wide variety of molecular adaptations, these eleven genes represent a signal of parallelism between *Phyllodia* and *R. sinicus*, with four genes showing shared private substitutions in the two lineages. These genes provide a possible start point into understanding the genetics

underpinning the evolution of the wide variety of echolocation calls observable across bat species.

Species	Geographic origin	Provenance	Sequencing centre	Sequencing Platform	Reads	Number of raw reads	Number of contigs	Assembly length (bp)
Mormoops blainvillei	Dominican Republic	Feb 2014	Barts and the London Genome Centre	Illumina NextSeq	150bp PE	13,940,556	86,204	84,068,873
Mormoops blainvillei **heart/cochlea/eye**	Dominican Republic	Feb 2014	Barts and the London Genome Centre	Illumina NextSeq	75bp PE	70,658,317	115,489	136,123,629
Mormoops megalophylla	Belize	2012	TGAC	Illumina HiSeq 2500	150bp PE	14,946,982	51,540	53,887,501
Pteronotus davyi	Belize	2012	BGI	Illumina HiSeq 2000	90bp PE	20,589,017	32,581	11,862,757
Pteronotus gymnonotus	Suriname	ROM	Barts and the London Genome Centre	Illumina NextSeq	150bp PE	18,950,896	64,581	54,958,503
Pteronotus macleayii	Jamaica	ROM	Barts and the London Genome Centre	Illumina NextSeq	150bp PE	9,947,590	72,843	51,650,237
Pteronotus pusillus	Dominican Republic	Feb 2014	TGAC	Illumina HiSeq 2500	150bp PE	18,088,264	93,759	93,977,193
Pteronotus pusillus **heart/cochlea/eye**	Dominican Republic	Feb 2014	Barts and the London Genome Centre	Illumina NextSeq	75bp PE	71,281,902	131,574	161,925,039
Pteronotus mesoamericanus	Belize	2012	BGI	Illumina HiSeq 2000	90bp PE	20,062,510	80,676	39,815,755
Pteronotus personatus	Suriname	ROM	Barts and the London	Illumina NextSeq	150bp PE	12,566,333	92,558	70,291,547

			Genome Centre					
Pteronotus quadridens	Dominican Republic	Feb 2014	TGAC	Illumina HiSeq 2500	150bp PE	14,583,821	75,226	75,326,881
Pteronotus quadridens **heart/cochlea/eye**	Dominican Republic	Feb 2014	Barts and the London Genome Centre	Illumina NextSeq	75bp PE	83,150,250	141,274	158,845,843

Table 2.1: Sequencing and assembly information for mormoopid species.

Species	Family	Provenance	Sequencing centre	Sequencing Platform	Reads	Number of raw reads	Number of contigs	Assembly length (bp)
Anoura geoffroyi	Phyllostomidae	Guyana; ROM	BGI	Illumina HiSeq 2000	90bp PE	19,444,106	119,890	58,741,648
Bauerus dubiaquercus	Vespertilionidae	Belize 2012	BGI	Illumina HiSeq 2000	90bp PE	19,746,016	88,758	51,774,594
Desmodus rotundus	Phyllostomidae	Belize 2012	BGI	Illumina HiSeq 2000	90bp PE	20,109,827	89,033	61,997,407
Erophylla sezekorni	Phyllostomidae	Jamaica; ROM	BGI	Illumina HiSeq 2000	90bp PE	20,282,580	96,312	43,897,218
Glossophaga soricina	Phyllostomidae	Belize 2012	BGI	Illumina HiSeq 2000	90bp PE	19,181,759	51,822	23,739,234
Hylonycteris underwoodi	Phyllostomidae	Costa Rica; ROM	BGI	Illumina HiSeq 2000	90bp PE	20,212,773	69,389	31,884,557
Lichonycteris obscura	Phyllostomidae	Ecuador; ROM	BGI	Illumina HiSeq 2000	90bp PE	19,519,969	78,247	49,586,127
Lonchophylla thomasi	Phyllostomidae	Guyana; ROM	BGI	Illumina HiSeq 2000	90bp PE	19,798,642	165,491	90,261,044
Macrotus waterhousii	Phyllostomidae	Dominican Republic 2014	TGAC	Illumina HiSeq 2500	150bp PE		80,185	89,786,186
Micronycteris microtis	Phyllostomidae	Belize 2012	BGI	Illumina HiSeq 2000	90bp PE	20,823,046	71,137	35,628,730
Molossus sinaloae	Molossidae	Belize 2012	BGI	Illumina HiSeq 2000	90bp PE	19,481,973	113,458	67,693,220
Myotis elegans	Vespertilionidae	Belize 2012	BGI	Illumina HiSeq 2000	90bp PE	20,770,231	61,294	29,657,285
Myotis ricketti[1]	Vespertilionidae	Beijing 2009		Illumina Genome Analyzer II	75bp		105,065	65,625,870
Noctilio leporinus	Noctilionidae	Belize 2012	BGI	Illumina HiSeq 2000	90bp PE	20,341,750	109,152	59,655,975
Nycteris tragata	Nycteridae	Malaysia; ROM	BGI	Illumina HiSeq 2000	90bp PE	20,261,468	157,930	100,326,595

Peropteryx kappleri	Emballonuridae	Belize 2012	BGI	Illumina HiSeq 2000	90bp PE	20,485,894	117,342	65,412,313
Rhogeessa aeneus	Vespertilionidae	Belize 2012	BGI	Illumina HiSeq 2000	90bp PE	19,999,149	89,883	49,144,434
Rhynchonycteris naso	Emballonuridae	Belize 2012	BGI	Illumina HiSeq 2000	90bp PE	20,376,163	121,803	72,214,309
Saccopteryx bilineata	Emballonuridae	Belize 2012	BGI	Illumina HiSeq 2000	90bp PE	20,770,231	130,448	84,306,562
Sturnira lilium	Phyllostomidae	Belize 2011	BGI	Illumina HiSeq 2000	90bp PE	20,770,231	106,421	71,328,661
Thyroptera tricolor	Thyropteridae	Suriname; ROM	BGI	Illumina HiSeq 2000	90bp PE	19,199,396	66,540	28,993,309
Trachops cirrhosis	Phyllostomidae	Belize 2011	BGI	Illumina HiSeq 2000	90bp PE	20,527,523	136,832	76,743,986
Uroderma bilobatum	Phyllostomidae	Belize 2012	BGI	Illumina HiSeq 2000	90bp PE	19,726,018	105,134	67,965,240

Table 2.2: Outgroup sample information for the Yangochiroptera data set. ROM: sequence provided by the Royal Ontario Museum.
[1]Sequence published in Dong et al 2013

Species	Sensory tissue data set	Yangochiroptera data set
Mormoops blainvillei	13,604	12,162
Mormoops megalophylla	-	9,920
Pteronotus davyi	-	5,626
Pteronotus gymnonotus	-	10,423
Pteronotus macleayii	-	10,658
Pteronotus pusillus	-	12,483
Pteronotus mesoamericanus	13,751	9,167
Pteronotus personatus	-	11,510
Pteronotus quadridens	13,806	12,194

Table 2.3: Number of putative transcripts with single-copy orthology to human genes identified by reciprocal blast in each mormoopid species.

Data set	Branch	PAML			HyPhy aBSREL		HyPhy Relax		
		Initial	R Sig	C Sig	Initial	Sig[a]	Initial	R Sig	C Sig
ST	*Pteronotus*	11,121	438	21	11,271	109	10,954	1,064	22
	Mormoops	10,846	367	6	12,012	89	10,685	1,152	34
	Phyllodia	11,831	340	33	12,012	99	11,665	866	9
	P. quadridens	11,153	268	12	11,271	106	10,995	826	13
Y	*Pteronotus*	6,699	149	3	6,836	27	6,228	439	0
	Mormoops	6,398	163	1	4,513	17	5,935	395	4
	Phyllodia	6,949	110	3	7,046	61	6,460	284	4
	Internal *Pteronotus*	5,988	38	3	6,104	10	5,542	120	1
	P. personatus	5,343	49	3	5,446	18	4,953	182	3
	P. quadridens	6,143	49	0	6,222	17	5,710	201	4
	P. macleayii	5,043	74	7	5,096	34	-	-	-
	P. gymnonotus	5,060	48	5	5,140	37	-	-	-
	P. davyi	895	8	2	987	14	-	-	-
R	*R. sinicus*	11,497	978	155	11,881	228	11,211	1,315	97

Table 2.4: Number of genes remaining for each model through each stage of the selection testing process. Initial = Number of genes tested; R Sig = Number of significant results before correction for multiple testing; C Sig = Number of significant genes after correction for multiple testing. Data sets: ST = Sensory tissue; Y = Yangochiroptera; R = Rhinolophus.

[a]Using the HyPhy aBSREL model on multiple branches implements its own correction for multiple testing, so these results include that with no further correction steps performed.

Model	Branch	Positive in both	Positive in Sensory Tissue	Positive in Yangochiroptera
HyPhy aBSREL	Ancestral *Pteronotus*	4	42	16
	Ancestral *Mormoops*	3	37	10
	Phyllodia	5	37	51
	P. quadridens	4	67	11
PAML branch-site	Ancestral *Pteronotus*	46 (1)	172 (10)	84 (2)
	Ancestral *Mormoops*	35 (0)	125 (2)	109 (1)
	Phyllodia	42 (1)	101 (16)	49 (1)
	P. quadridens	11 (0)	114 (6)	34 (0)

Table 2.5: Of the genes present in both the sensory tissue and the Yangochiroptera data sets, the number of genes found to be under positive selection in both or in just one. Numbers in brackets shows genes remaining in the category after multiple testing correction.

Data set	Branch	Results before correction		Results after correction	
		Intensification	Relaxation	Intensification	Relaxation
Sensory tissue	Ancestral *Pteronotus*	794	270	19	3
	Ancestral *Mormoops*	843	309	24	10
	Phyllodia	626	239	6	3
	P. quadridens	638	188	10	3
Yangochiroptera	Ancestral *Pteronotus*	315	124	0	0
	Ancestral *Mormoops*	253	142	0	4
	Phyllodia	193	91	2	2
	Internal *Pteronotus*	74	46	0	1
	P. personatus	107	75	1	2
	P. quadridens	122	79	3	1

Table 2.6: Number of genes found to be experiencing an intensification or a relaxation of selection by the RELAX model in the sensory tissue or Yangochiroptera data sets. Results are presented both with and without correction for multiple testing.

		PAML Branch-Site		HyPhy aBSREL	
Data set	**Branch**	**Intensification**	**Relaxation**	**Intensification**	**Relaxation**
Sensory tissue	Ancestral *Pteronotus*	108	13	41	2
	Ancestral *Mormoops*	70	4	34	3
	Phyllodia	93	10	48	3
	P. quadridens	73	3	42	2
Yangochiroptera	Ancestral *Pteronotus*	46	5	18	0
	Ancestral *Mormoops*	46	8	4	1
	Phyllodia	37	2	12	1
	Pteronotus internal	15	2	3	0
	P. personatus	17	4	5	1
	P. quadridens	16	3	8	0

Table 2.7: Number of genes identified as being under positive selection that also returned a significant result under the HyPhy RELAX model (no multiple testing corrections applied).

Branch	Sensory tissue data set	Yangochiroptera data set
Ancestral *Pteronotus*	*A2m, Cd151, Col3a1, Col9a1, Dnmt3a, Eral1, F2, Fdxr, Fgf10, Fn1, Icam1, Kdr, Lars2, Map2, Psap, Ptger4, Slc29a3, Spata17, Ush1g*	*Apoe, Hmox1, Nefh, Myh9, Psap*
Ancestral *Mormoops*	*A2ml1, Cd151, Clpp, Gstp1, Hmox1, P2rx2, Pex1, Plat, Polg, Ppargc1a, Sod1, Tprn, Tyr, Wbp2*	*Cd151, Clpp, Map2, Polg, Wbp2*
Phyllodia	*A2ml1, Cacna1d, Cat, Col4a5, Dcdc2, Dpt, Gmnn, Lars2, Map2, Mtor, Mtr, Opa1, Rela, Serpinf1, Tnc*	*Clcn3, Elavl4, Gsdme, Lars2, Rela, Serpinf1, Stat1*
***Pteronotus* internal**	-	*Il4r*
P. personatus	-	*Myh9*
P. quadridens	*Abcc8, Cat, Ceacam16, Il4r, Met, Nos1ap, Serpinf1, Tpmt, Ush1g*	*Atrx, Ercc2*
P. macleayii	-	*Hsd17b4, Zbtb20*
P. gymnonotus	-	*NA*
P. davyi	-	*NA*

Table 2.8: Hearing genes identified as being under positive selection after visual inspection of alignments. Underlined genes were identified as being under positive selection by both the PAML branch-site and the HyPhy aBSREL model on the same branch.

Data Set	Branch	Intensification	Relaxation
Sensory tissue	Ancestral *Pteronotus*	26 (6)	5 (0)
	Ancestral *Mormoops*	32 (6)	16 (0)
	Phyllodia	30 (8)	12 (2)
	P. quadridens	25 (1)	6 (0)
Yangochiroptera	Ancestral *Pteronotus*	11 (0)	8 (0)
	Ancestral *Mormoops*	12 (1)	6 (0)
	Phyllodia	8 (3)	3 (0)
	Internal *Pteronotus*	5 (0)	3 (0)
	P. personatus	5 (1)	5 (0)
	P. quadridens	5 (0)	3 (0)

Table 2.9: Hearing genes identified as experiencing a change in selection pressure by the HyPhy RELAX model in the sensory tissue and Yangochiroptera data sets following visual inspection of alignments. Numbers in brackets: number of these genes previously found to be evolving under positive selection. Only results prior to multiple testing corrections shown.

Gene Name	Branch positively selected	Gene Function
Clcn3	*Phyllodia*	Voltage-gated chloride channel. Associated with sensorineural hearing loss (ClinVar Accession SCV000323085, Landrum et al 2018)
Col4a5	*Phyllodia*	Collagen – structural component of basement membrane. Associated with hearing loss via X-linked Alport syndrome (Barker et al 1990).
Dnmt3a	Ancestral *Pteronotus*	DNA methyltransferase essential for normal development of the inner ear (Roellig and Bronner 2016).
F2	Ancestral *Pteronotus*	Coagulation factor thrombin potentially associated with sensorineural hearing loss (e.g. Capaccio et al 2009, though mixed evidence cf. Shu et al 2015).
Fn1	Ancestral *Pteronotus*	Glycoprotein involved in cell adhesion and migration processes. Associated with hearing loss through role in cochlear basement membranes (Keithly, Ryan and Woolf 1993); interacts with Ushering in Usher syndrome (Bhattacharya and Cosgrove 2005).
Kdr	Ancestral *Pteronotus*	Receptor of the vascular endothelial growth factor, which has cochlear expression changes after auditory damage and is hypothesised to be associated with recovery and repair (Picciotti 2006).
Lars2*	*Phyllodia*	tRNA synthetase associated with hearing loss through Perrault syndrome (Pierce et al 2013).
Map2	Ancestral *Pteronotus*	Microtubule-associated protein; no clear direct link to hearing loss.
Mtor	*Phyllodia*	Kinase mediating cellular response to stress. Inhibition leads to damage to auditory hair cells (Leitmeyer et al 2015).
Opa1	*Phyllodia*	Regulates mitochondrial stability and energy output. Associated with sensorineural hearing loss (e.g. Leurez et al 2013)
Rela*	*Phyllodia*	NF-kappa-B transcription factor complex; association with drug efficacy in protecting hearing (Layman and Zuo 2015).
Serpinf1*	*Phyllodia*	Angiogenesis inhibitor/neuronal differentiation; associated with hearing loss via familial otosclerosis (Ziff et al 2016).
Tnc	*Phyllodia*	Extracellular matrix protein associated with non-syndromic sensorineural deafness/hearing loss (e.g. Zhao et al 2013).

Table 2.10: Hearing genes signals of positive selection on either the *Phyllodia* or ancestral *Pteronotus* branch which are also supported by a significant intensification of selection by RELAX (Function information from Safran et al 2013; **www.genecards.org**)
***** Indicated genes that were also detected as positively selected independently in both the sensory tissue and Yangochiroptera data sets.

Branch	Tested for positive selection			Signals of positive selection		
	# genes	# hearing genes	% hearing genes	# genes	# hearing genes	% hearing genes
Ancestral *Mormoops*	11,726	377	3.2	585	17	3.2
Ancestral *Pteronotus*	11,914	382	3.2	587	27	4.6
Phyllodia	12,611	410	3.3	492	23	4.7
***Pteronotus* internal[a]**	6,024	181	3.0	39	1	2.6
***P. personatus*[a]**	5,446	150	2.8	59	1	1.7
P. quadridens	11,875	381	3.2	523	20	3.8
***P. macleayii*[a]**	5,096	149	2.9	82	4	4.9
***P. gymnonotus*[a]**	5,140	143	2.8	56	2	3.8
***P. davyi*[a]**	987	28	2.8	11	0	0

Table 2.11: Number and proportion of hearing genes that underwent selection tests and that showed signals of positive selection along branches within the Mormoopidae. Numbers presented combine results from the sensory tissue and Yangochiroptera data sets and do not consider results of visual inspection of genes.
[a]Branch only underwent selection testing in the Yangochiroptera data set.

Gene Name	Branch positively selected	Gene Function
Agt	Ancestral *Pteronotus*	Involved in the maintenance of blood pressure; no clear direct link to hearing loss.
Cacna1d	*Phyllodia*	Voltage-gated calcium channel subunit linked to age-related hearing impairment (Chen et al 2013) and deafness (Platzer et al 2000).
Cat	*Phyllodia*	Antioxidant enzyme associated with noise-induced (Konings et al 2007, Rewerska et al 2013) and cisplatin-induced (Rybak et al 1999) hearing loss.
Col4a5	*Phyllodia*	Collagen – structural component of basement membrane. Associated with hearing loss via X-linked Alport syndrome (Barker et al 1990).
Eral1	Ancestral *Pteronotus*	GTPase linked to hearing loss via Perrault syndrome (Chatzispyrou et al 2017).
Gmnn	*Phyllodia*	Critical role in cell cycle regulation; no clear direct link to hearing loss.
Icam1	Ancestral *Pteronotus*	Cell surface protein; associated with noise-induced hearing loss (Seidman et al 2009).
Map2	*Phyllodia* and Ancestral *Pteronotus*	Microtubule-associated protein; no clear direct link to hearing loss.
Nefh	Ancestral *Pteronotus*	Neurofilament heavy protein linked to hearing loss via Charcot-Marie-Tooth disease (Jacquier et al 2017)
Ush1g	Ancestral *Pteronotus*	Role in development and maintenance of the auditory system (Kazmierczak & Muller 2012); functions in cohesion of hair bundles formed by inner ear sensory cells. Linked to Usher syndrome type 1G (Weil et al 2003).

Table 2.12: Hearing genes with signals of positive selection in either the ancestral *Pteronotus* branch or the *Phyllodia* branch and *Rhinolophus sinicus* (Function information from Safran et al 2013; **www.genecards.org**)

Branch	# shared positively selected genes	# shared positively selected hearing genes	% shared positively selected genes that are hearing genes
Ancestral *Pteronotus*	77	8	0.1
Ancestral *Mormoops*	64	2	0.03
Phyllodia	53	6	0.11
P. quadridens	47	3	0.06

Table 2.13: Proportion of genes with signatures of having evolved under positive selection in both *R. sinicus* and a mormoopid branch of interest that were also hearing genes.

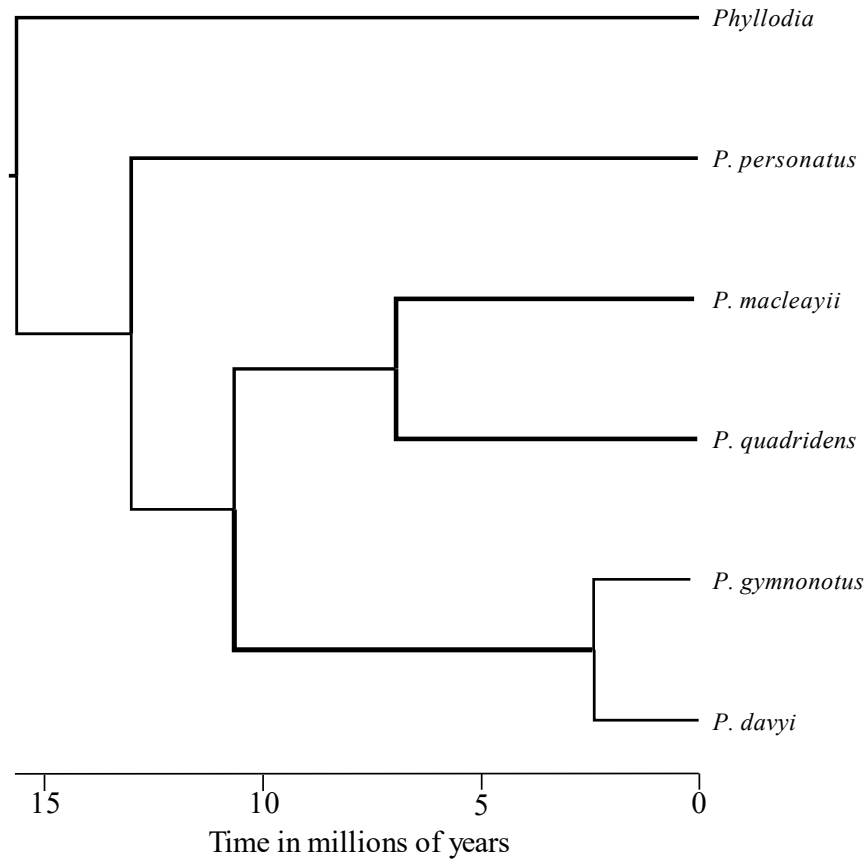


Figure 2.1: *Pteronotus* phylogeny showing divergence times, adapted from Pavan and Marroig 2017 to show only the the species included in my sample. The origin of the genus dates to approximately 16 million years ago.

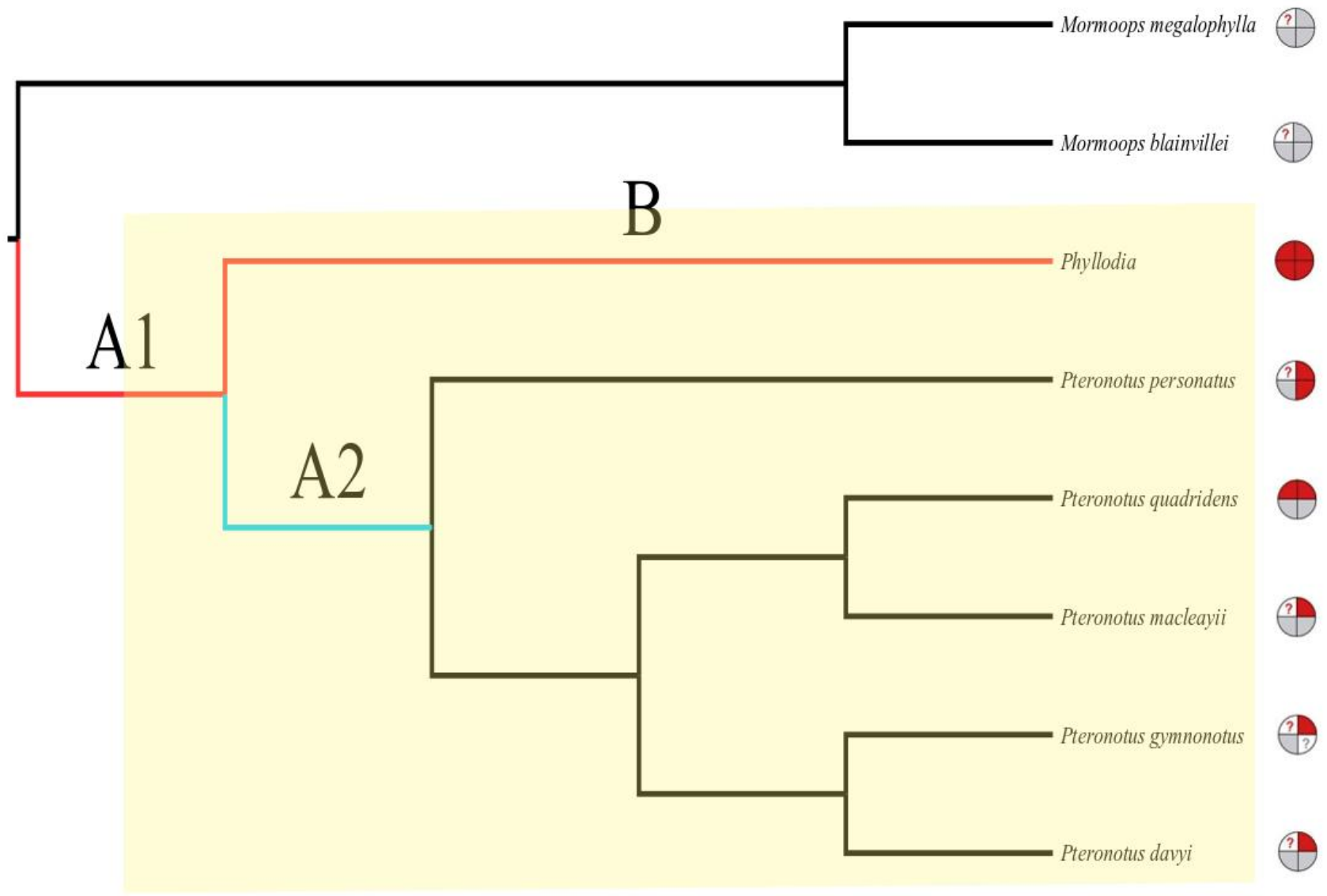

Mormoops megalophylla
Mormoops blainvillei
B
Phyllodia
A1
Pteronotus personatus
A2
Pteronotus quadridens
Pteronotus macleayii
Pteronotus gymnonotus
Pteronotus davyi

Figure 2.2: Hypotheses for the origin of HDC echolocation calls in the *Pteronotus* genus (Topology: van den Bussche and Weyandt 2003). Yellow background indicated the *Pteronotus* genus. Red branches indicate hypothesised gains; blue branches losses. Hypothesis one: HDC echolocation evolved at A1, with a reversion at A2. Hypothesis 2: HDC echolocation evolved at B. Yellow highlight indicates species belonging to the *Pteronotus* genus. Pie Presence/Absence of traits associated with HDC echolocation shown with pie charts after taxon names, taken from Mora et al 2013. Top left: Heteroharmonic computation strategy; top right: CF-FM calls; bottom left: HDC echolocation; bottom right: Doppler-shift compensation. Red = Present; grey = absent; red question mark = probably present; grey question mark = probably absent.

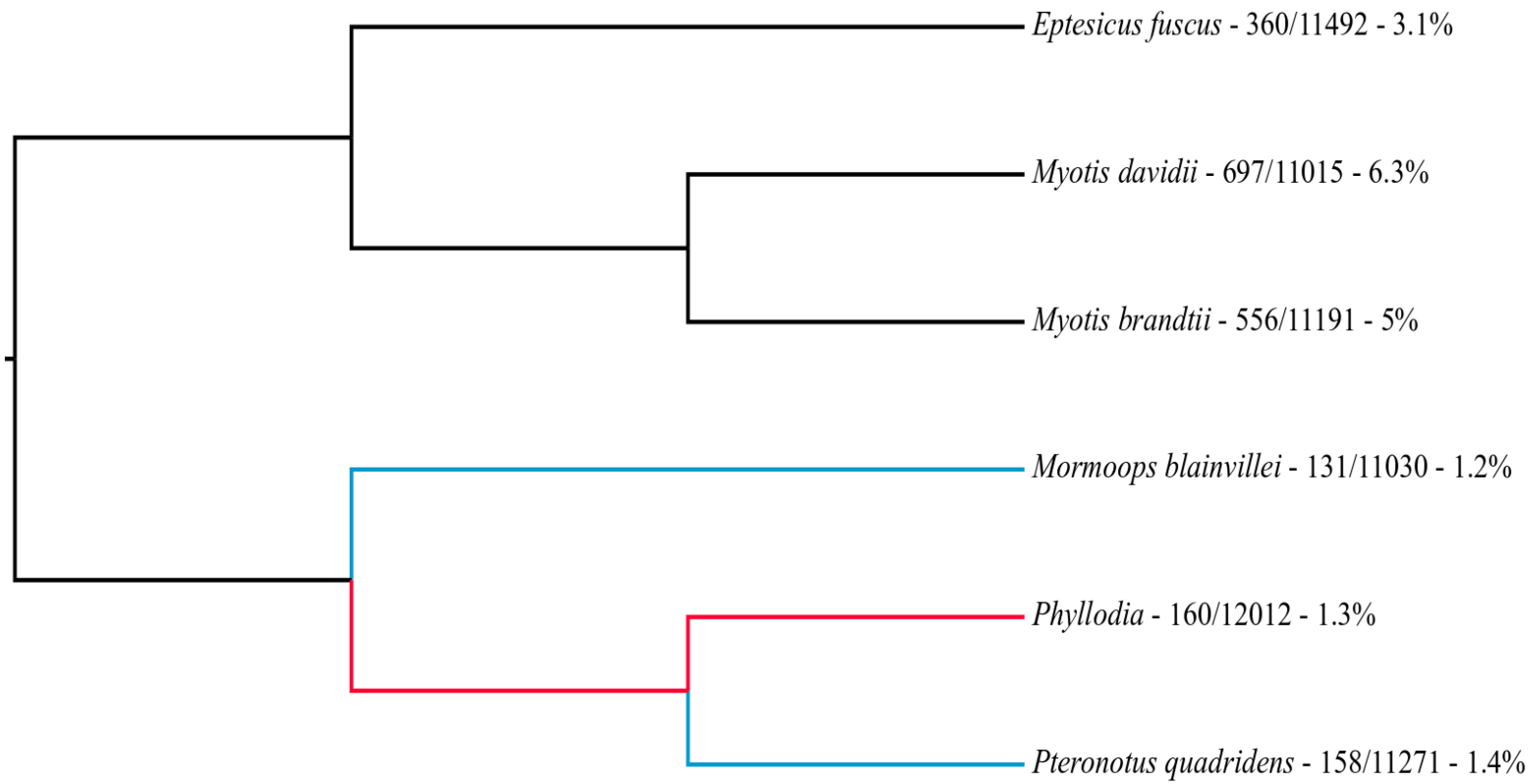


Figure 2.3: Species in the sensory tissue data set. Phylogeny following Potter et al, unpublished. Label format: Species name – genes found under positive selection by HyPhy aBSREL/total genes tested – proportion of genes under positive selection. Red branches are hypothesised points for the evolution of HDC echolocation; blue branches were tested as control branches.

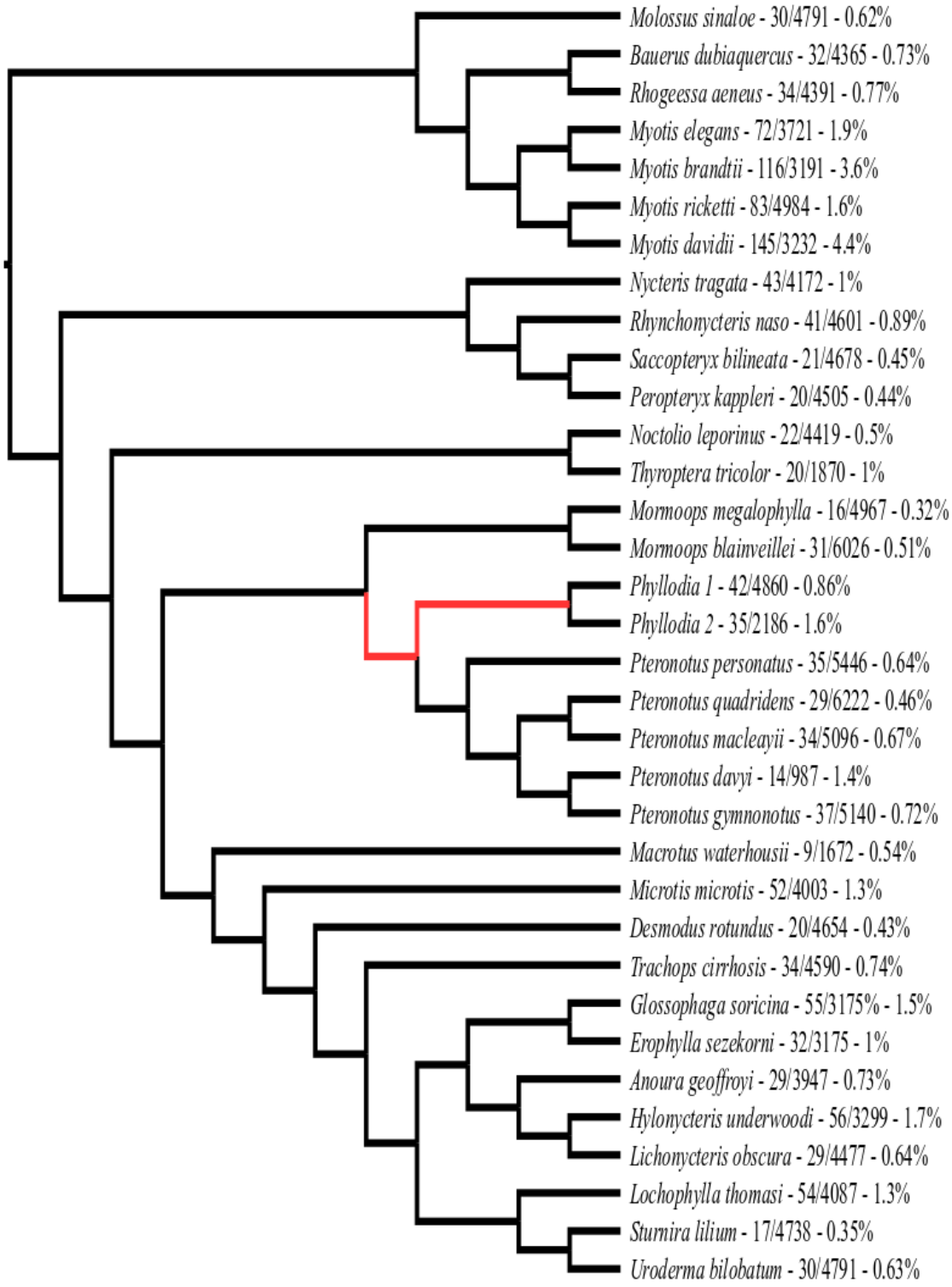


Figure 2.4: Species in the Yangochiroptera data set. Phylogenetic topology following Potter et al, unpublished. Label format: Species name – genes found under positive selection by HyPhy aBSREL/total genes tested – proportion of genes under positive selection. Red branches show the hypothesised points that HDC echolocation evolved.

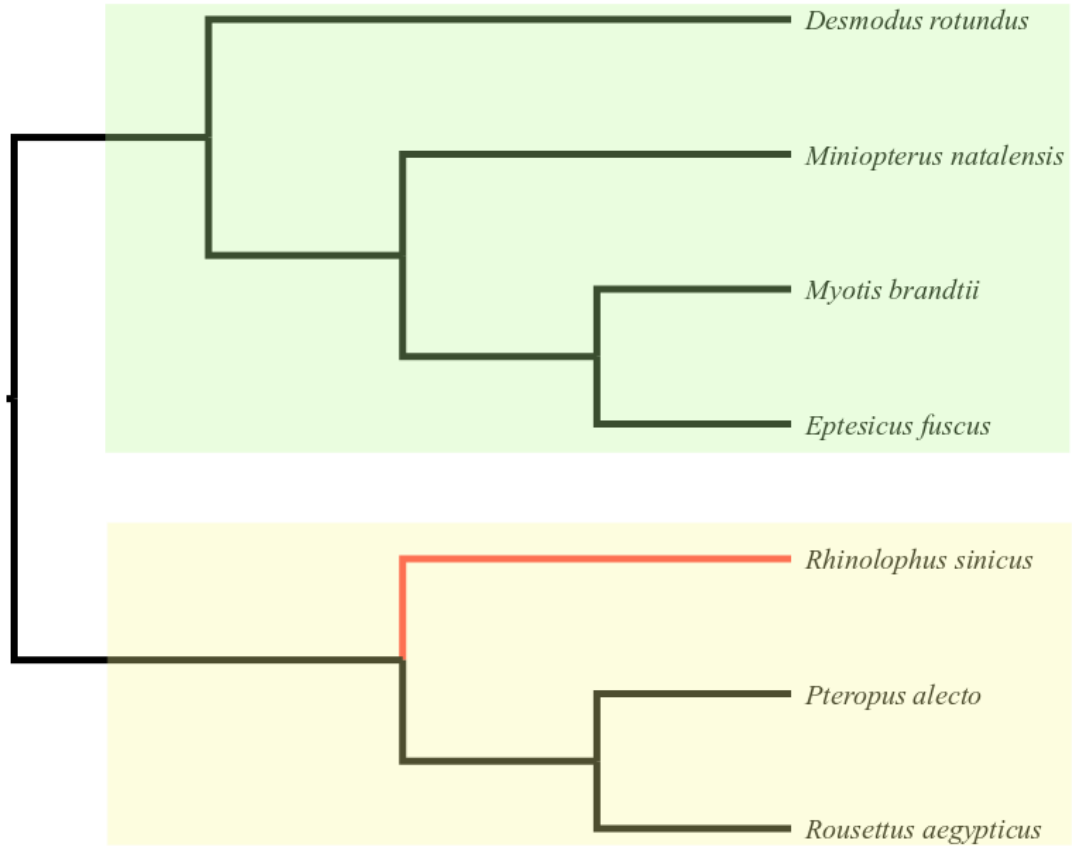


Figure 2.5: Phylogeny of species included in data set three. HDC echolocator branch coloured in red. Yellow highlight indicates members of the Yinpterochiroptera; green highlight members of the Yangochiroptera.

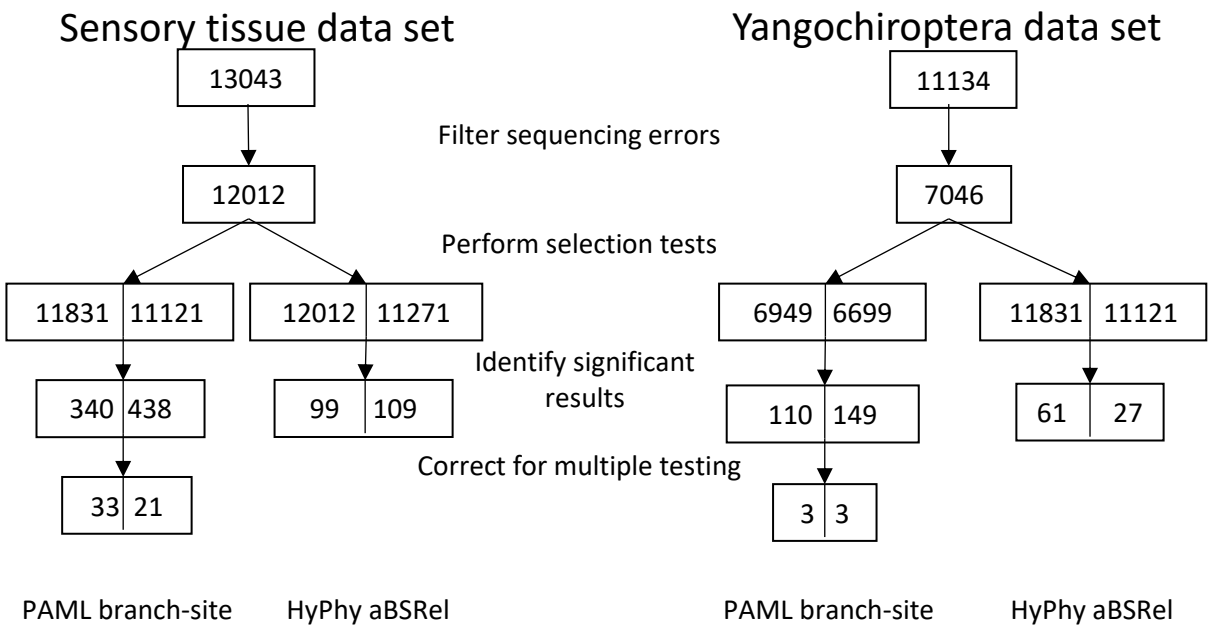


Figure 2.6: Flowcharts showing the number of genes at each step of filtering and selection testing on the *Phyllodia* and ancestral *Pteronotus* branches in the sensory tissue (left-hand chart) and Yangochiroptera (right-hand chart) data sets. Initial boxes show the number of genes successfully aligned; second boxes show the number remaining after alignment filtering steps. Subsequent boxes show number of genes tested for positive selection with the PAML branch-site model (branches to the left) or the HyPhy aBSREL model (branches to the right). Left-hand boxes within a branch show the number of genes tested in *Phyllodia*; right-hand boxes the number tested on the ancestral *Pteronotus* branch. 6,447 genes were present in both the sensory tissue and the Yangochiroptera data sets.

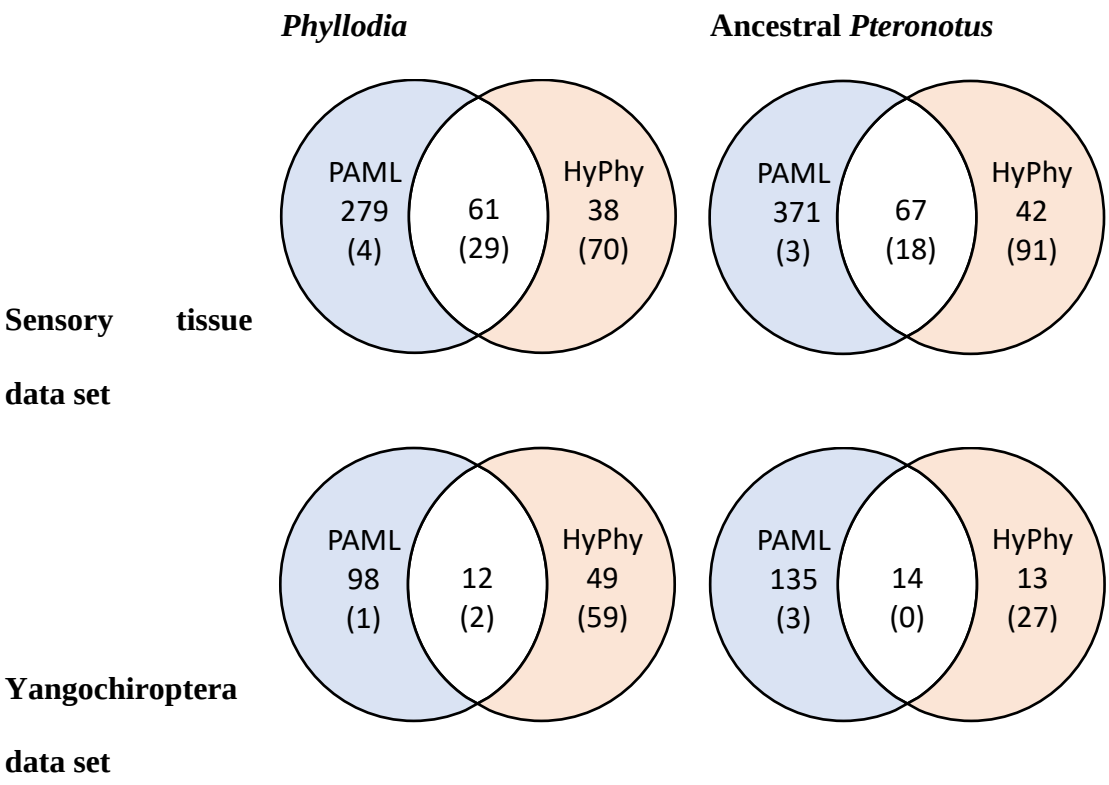


Figure 2.7: Overlap in genes showing signatures of positive selection between the PAML branch-site and HyPhy aBSREL models. Top numbers show uncorrected PAML results; numbers in brackets include PAML results after multiple testing correction.

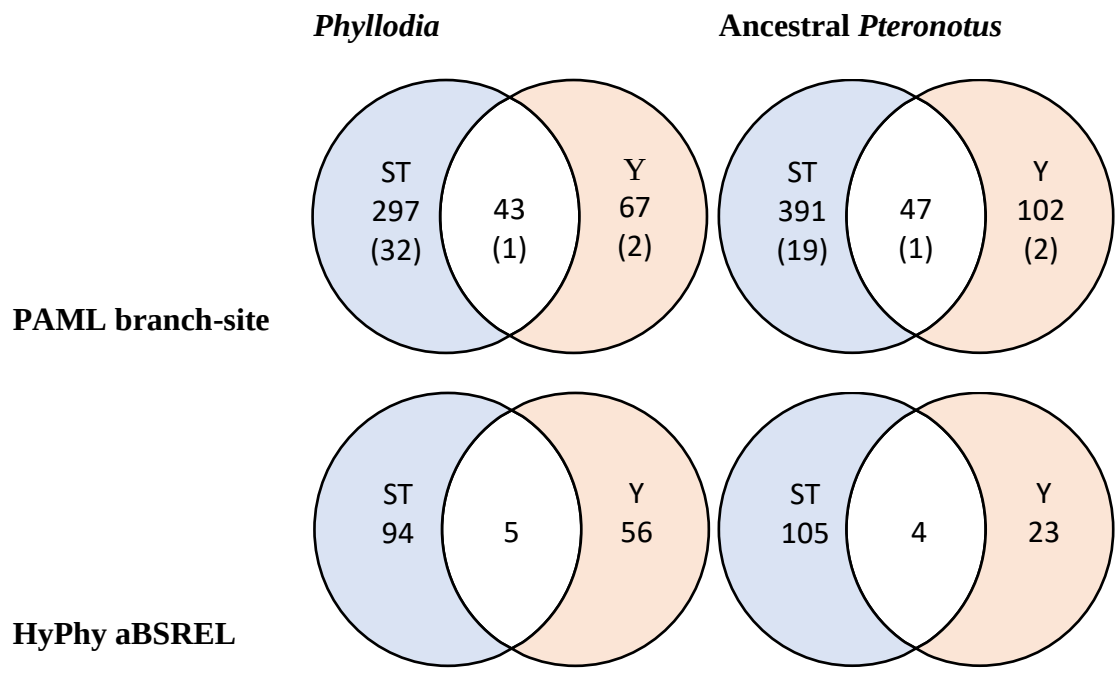


Figure 2.8: Overlap in genes showing signatures of positive selection between the sensory tissue (ST) and Yangochiroptera (Y) data sets. Top numbers show uncorrected PAML results; numbers in brackets include PAML results after multiple testing correction.

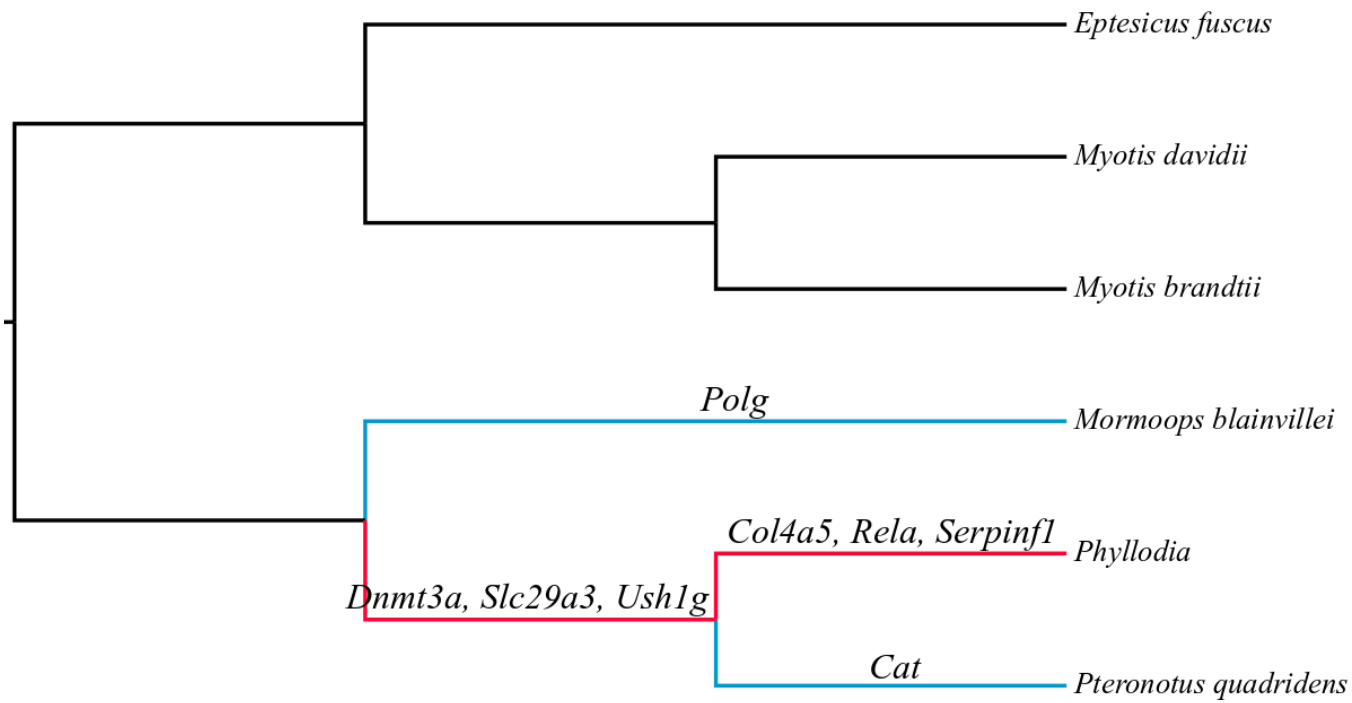


Figure 2.9: Best-supported hearing genes identified as evolving under positive selection in the sensory tissue data set. Genes were found to be evolving under positive selection by both the PAML branch-site model and the HyPhy aBSREL model. Full results in Table 2.8. Red branches are hypothesised points for the evolution of HDC echolocation; blue branches were tested as control branches.

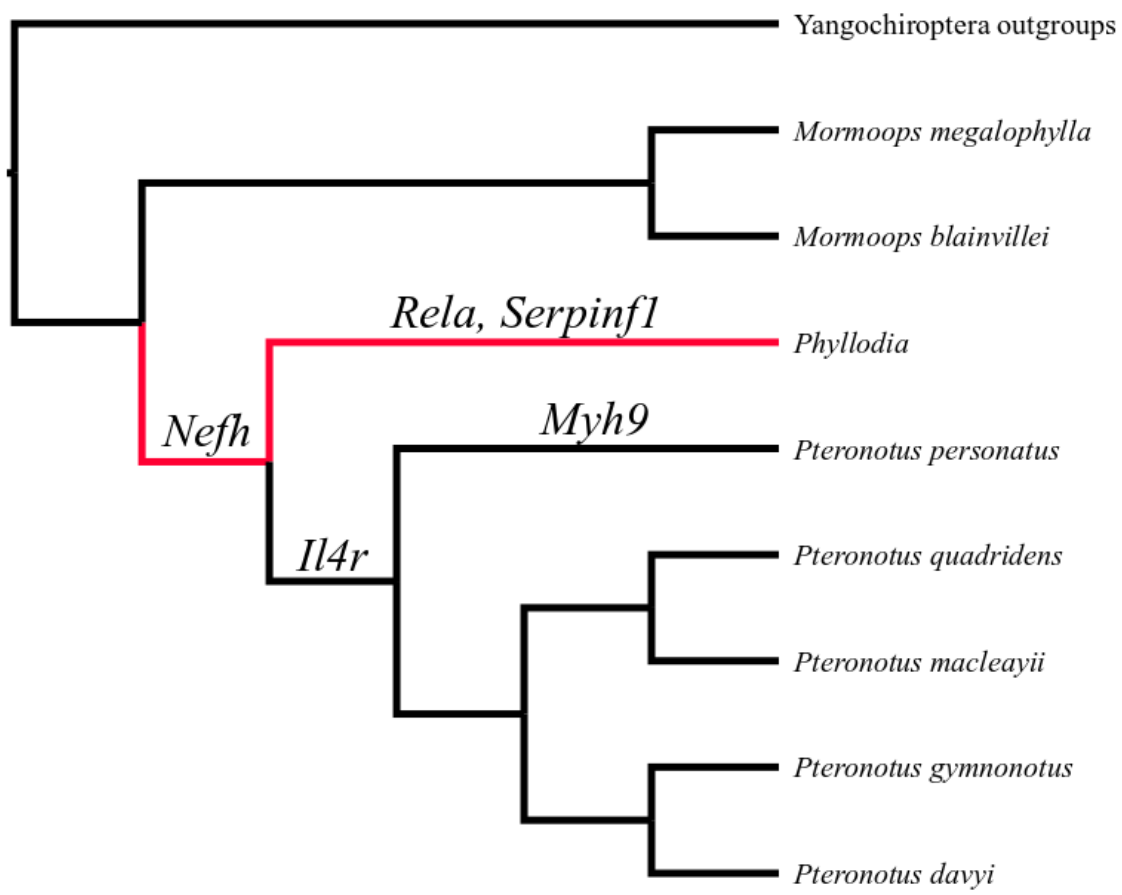


Figure 2.10: Best-supported hearing genes identified as evolving under positive selection in the Yangochiroptera data set. Genes were found to be evolving under positive selection by both the PAML branch-site model and the HyPhy aBSREL model. Full results can be found in Table 2.8. Full topology of outgroups can be found in Figure 2.3. Red branches show the hypothesised points that HDC echolocation evolved.

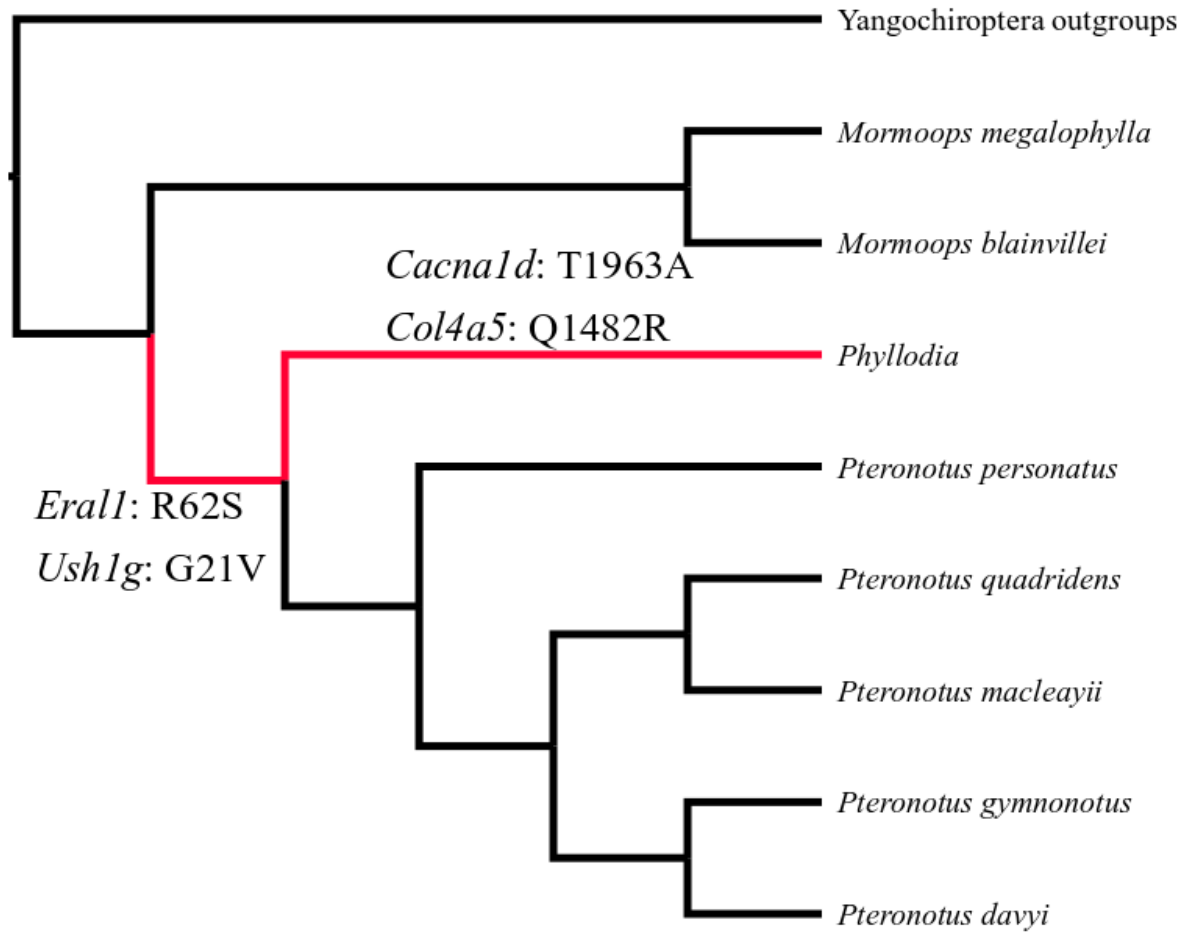


Figure 2.11: Shared private substitutions in hearing genes evolving under positive selection in both *R. sinicus* and either *Phyllodia* or on the ancestral *Pteronotus* branch. Red branches show the hypothesised points that HDC echolocation evolved.

CHAPTER THREE

Mitonuclear discordance among three sympatric morphs of *Rhinolophus philippinensis*.

Abstract

The current gold standard for using genetic data to test population structure and dynamics is to use evidence from a mixture of mitochondrial and nuclear loci. Due to unique features of its inheritance mechanisms, mitochondrial DNA has a shorter coalescence time than nuclear DNA and so can act as an indicator of developing population structure earlier than it can be detected from nuclear markers. However, analysis of a mixture of nuclear and mitochondrial loci frequently reveals conflicting gene tree generated by complex demographic histories featuring migration, hybridisation and asymmetric introgression of genetic markers. As these conflicts can in themselves be informative about population demographics, combining information from multiple sources can give us the most insight.

Three morphs of the large-eared horseshoe bat *Rhinolophus philippinensis* can be found in sympatry on Buton Island, Sulawesi. These morphs are differentiated from one another by body size and by echolocation call frequency, with the Small and Intermediate morphs concentrating the energy of their narrowband call into different harmonics of the Large morph's fundamental frequency. Initial research into this system by Kingston and Rossiter (2004) revealed evidence of genetic differentiation between the morphs based on a limited number of mitochondrial and microsatellite markers, leading to a suggestion that reproductive isolation may be evolving or may have recently evolved between them.

Over a decade on from this original research, there have been substantial advances in high-throughput sequencing and in bioinformatics tools available to analyse the ensuing large quantities of sequence data. I have revisited this system with whole-genome sequencing data for a larger number of individuals to test the hypotheses and findings of the original harmonic hopping paper. Using whole mitochondrial sequences and a panel of genomic SNPs, I have found evidence against panmixia within the Buton *R. philippinensis*, with differentiation between the morphs based on both mitochondrial and nuclear loci. However, there is discordance between these data, with SNPs supporting monophyletic clades for the morphs while the mitochondria indicate paraphyly of the Large and Small morphs. These differences suggest population histories that include either a recent divergence with incomplete lineage sorting, or a degree of gene flow with hybridisation or introgression.

Introduction

While there has been an increasing appreciation over the last decade that the dynamics of speciation on the genetic level can be complex and varied (Campbell, Poelstra and Yoder 2018), speciation in the face of ongoing gene flow remains a controversial topic in evolutionary biology (Foote 2018). The key antagonism at the heart of speciation with gene flow is that any genetic mixing between diverging populations will tend to homogenise gene pools via recombination (Felsenstein 1981, Via 2001), so opposing their divergence and eventual speciation (Kopp et al 2018). One proposed solution to this problem is the action of divergent natural selection on a trait favouring partitioning and divergence of a population. If this trait under ecological selection also has either a direct or an indirect influence on mating behaviour, then the antagonism between divergence and recombination can be overcome, such that divergence will proceed to the formation of novel species (Kopp et al 2018) in a process that is now termed ecological speciation (Nosil 2012).

Traits under natural selection that also have roles in mate recognition and mating decisions have been called 'magic traits' (Servedio et al 2011), and have been implicated in a number of speciation events, particularly in birds (e.g. Huber et al 2006, Huber et al 2007, Derryberry et al 2017, Langin et al 2017). For example, in Darwin's finches *Geospiza fortis*, beak morphology impacts the food sources a bird can exploit and so is a target of natural selection, but also imposes strong constraints upon song parameters (Huber and Podos 2006). Song is commonly used as a mating signal, as are visual cues from beak morphology. Ecological divergent selection on beak morphology to allow exploitation of different resources thus causes divergence in mating signals, reinforcing the differentiation of the populations (Huber et al 2007). A magic trait has also been

implicated in the diversification of the African weakly electric fish (Mormyridae) (Feulner et al 2009a). These species use an electric organ discharge (EOD) in electrolocation to detect prey, but also in intraspecific communication. Crucially, the EOD appears to play a role in mate recognition. EOD playback experiments demonstrated that females were attracted to the EOD of conspecific males, suggesting that it plays a role in assortative mating (Feulner et al 2009b).

Echolocation in bats shows strong parallels with electrolocation in Mormyridae. While the primary functions of echolocation are prey detection and obstacle avoidance, echolocation calls also encode information about the transmitter (Schuchmann, Puechmaille and Siemers 2012, Li et al 2014) that other bats can access and respond to (Gillam and Fenton 2016). Indeed, bats use a range of vocalisations in their social communications (Ma et al 2006), including ultrasonic calls (Jones and Siemers 2011), and there is evidence that they will intentionally emit echolocation calls in response to conspecifics when stationary and not foraging, despite the metabolic cost of doing so (Dechmann et al 2013). In particular, bats respond differently to eavesdropped echolocation calls from male and female conspecifics, indicating that these can function as a mating signal (Knörnschild et al 2012, Puechmaille et al 2014).

Echolocation shows particular potential as a magic trait in bat species that use high duty-cycle (HDC) echolocation (Kingston et al 2001, Servedio et al 2011). These bats have highly derived auditory systems with an acoustic fovea (Schuller and Pollack 1979; Davies, Maryanto and Rossiter 2013), giving them a high sensitivity to a very narrow band of frequencies. Change in echolocation call parameters will affect the prey space that the bats are able to sample, but will also impact the ability for these bats to recognise and respond to each other's calls (Kingston et al 2001). The majority of HDC echolocating bats are found in two highly speciose Old World sister-families from the suborder Yinpterochiroptera, the Rhinolophidae and the Hipposideridae (Foley et al

2015). Divergent ecological selection on call frequency has been proposed as a speciation model in rhinolophid bats (Kingston 2001).

An example where the narrowband echolocation has been implicated in potential ecological speciation is the case of the large-eared horseshoe bat *Rhinolophus philippinensis*. On Buton Island in south-east Sulawesi, Indonesia, three distinct size and acoustic morphs of this taxon have been described, which can be easily distinguished from one another based on these characters (Figure 3.1), but are otherwise identical in phenotype (Kingston and Rossiter 2004). The emitted call frequencies of the intermediate and small morphs correspond to different harmonics of the large morph's fundamental frequency, with modelling suggesting that the frequency difference between the morphs has a substantial impact on the prey space they sample (Kingston and Rossiter 2004). Based on predictions from work on horseshoe bat audiograms (e.g. Long and Schnitzler 1975), these three morphs are thought to not be mutually audible to one another (Kingston and Rossiter 2004). Given the evidence that bats use echolocation calls in mate recognition, this opens a route to premating isolation and ultimately sympatric speciation.

Other members of the *R. philippinensis* clade in Australia and Asia also echolocate at different harmonics of the Buton large morph's fundamental frequency, suggesting that 'harmonic hopping' may have played a role in the rapid radiation of bat species in southeast Asia and the wider region (Kingston and Rossiter 2004). Indeed, two separate size morphs of bats assigned to *R. philippinensis* have also been recorded in Queensland, Australia, though the taxonomic status of these morphs remains unresolved (Cooper et al 1998, Churchill 2008, Pavey and Kutt 2008, Sedlock et al 2008), and also in the Philippines (Flannery 1995).

To date, evidence for potential rapid speciation in *R. philippinensis* is limited. Kingston and Rossiter (2004) applied microsatellite analysis based on small samples and recorded

low genetic differentiation between morphs at a level that suggests that they are not panmictic, but that reproductive isolation is either recently evolved or still incomplete (Kingston and Rossiter 2004). Mitochondrial sequencing of the non-coding control region locus in these individuals also revealed limited divergence among morphs together with a lack of reciprocal monophyly among the three morphs, and specifically some large individuals forming a cluster with the intermediate bats to the exclusion of the other large bats. One explanation for this reported signature is the independent divergence of different size morphs consistent with a history of ecological speciation in sympatry. However, an alternative explanation is that introgression has occurred among these morphs, perhaps with the intermediate representing a hybrid taxon from mating between large and small taxa. Finally, it is plausible that these signatures from mtDNA reflect the retention of ancestral polymorphisms and that incomplete lineage sorting has occurred since divergence.

Unfortunately, the datasets available to Kingston and Rossiter lacked power to tease apart these scenarios, although the authors did note that the observed lower levels of allelic diversity in the small and intermediate morphs compared to the large morph was more consistent with ecological speciation than a hybrid origin for the intermediate morph. In my study I aim to build on this earlier work by increasing the sample size to increase the power of the results, and by leveraging technological developments in both sequencing technology and bioinformatics over the last 15 years. My goals are to identify the genetic differences and relationships between the morphs. To this end, I focus on genetic variation within the morphs, and differentiation among them, as well as their relationships constructed based on ncDNA and mtDNA.

Mitochondrial DNA has long been the most popular marker for investigating molecular diversity in mammals (Galtier et al 2009). Mitochondrial DNA has numerous advantages in this context; it is easily sequenced at great depth due to being present in multiple copies

in cells, it has a high mutation rate so can provide information on population history over a short time frame, and it acts as a single uniparentally inherited non-recombining locus allowing for simple inference from data (Galtier et al 2009). Although there are limitations to using only mitochondrial DNA to reconstruct reliable species' histories, this maker can be informative when examined in conjunction with nuclear markers (Toews and Brelsford 2012). In particular, nuclear and mitochondrial DNA (mitonuclear) discordance can offer insights into population demographics, with differences between phylogenies derived from mitochondrial and nuclear loci giving us clues as to the histories of divergence and contact between closely-related species (Toews and Brelsford 2012). Where there is interspecific gene flow, for example in the case of hybridisation, alleles from one species can introgress into the gene pool of the other (Funk and Omland 2003). Due to the non-recombining nature of the mitochondria, mitochondrial introgression generates a particularly strong phylogenetic signal (Smith 1992) that can lead to polyphyly in mitochondrial trees even where other loci suggest the existence of monophyletic groups (Funk and Omland 2003). In extreme cases, mitochondrial introgression can lead to complete mitochondrial replacement within populations, even in the absence of any signal of introgression in the nuclear genome (e.g. Good et al 2015). Mitochondrial DNA introgression appears to be a common phenomenon in hybridising or migrating taxa, occurring as a result either of chance or of selective pressures (Ballard and Whitlock 2004). Because mitochondrial DNA is for the most part coding and thus not neutral (Dowling, Friberg and Lindell 2008), mitochondrial introgression could be adaptive, or alternatively might exert negative fitness effects due to incompatibilities between the introgressed mitochondria and the native genome (Burton and Barreto 2012; Hoekstra, Siddig and Montooth 2013). This latter effect has been observed in *Urosaurus* lizards, where introgressed mitochondrial DNA persists in an expanding population

despite individuals with heterospecific mitochondria being significantly smaller than those with homospecific mitochondria (Haenel 2017).

Conflicting signals between mitochondrial and nuclear loci can also result from incomplete lineage sorting in recently diverged species, where the partitioning of allele copies between the daughter species results in some loci being more closely related to those in a different species (Funk and Omland 2003). Over time this effect will break down, a process which occurs more quickly in the mitochondrial genome than in the nuclear genome due to its smaller effective population size (Funk and Omland 2003), meaning that the mitochondrial tree may be a more reliable guide than single nuclear gene trees in these cases (Moore 1995). The phylogenetic patterns that can be seen in mitochondrial introgression and in incomplete lineage sorting of mitochondrial DNA are very similar (Ballard and Whitlock 2004).

In this chapter, I first analyse the morphological data that have been collected from *R. philippinensis* individuals in Buton to confirm that the clear separation into three morphs as identified by Kingston and Rossiter (2004) remains for a larger sample size. I then test the hypothesis that the morphs are genetically distinct and form monophyletic groups by constructing phylogenetic trees and haplotype networks base on both mitochondrial and nuclear data. Analysing these two kinds of markers independently will also allow me to identify whether mitonuclear discordance exists in this system; finding dissonance between mitochondrial and nuclear phylogenies could lend support to the hypothesis that these morphs represent a recent speciation event in the case of incomplete lineage sorting, or that there is or has recently been gene flow between the morphs if there has been mitochondrial introgression. Finally, I use coalescent methods (BPP; Rannala and Yang 2003, Rannala and Yang 2013) and MavericK (Verity and Nichols 2016) for population delimitation to test the hypothesis that there is sufficient genetic differentiation between

the morphs that they could be considered separate populations in the process of developing reproductive isolation, rather than being a single panmictic population.

Materials and methods

Sample collection and sequencing

Rhinolophus philippinensis were captured, measured and had wing punch tissue samples taken as described in Kingston and Rossiter (2004) on Buton and Kabaena islands by trained volunteers between 2000 and 2012. I also took additional measurements and tissue samples in spring 2015 from all individuals designated as *R. philippinensis* that were present in the collections of the Bogor Zoology museum, including outgroup samples from individuals captured in Flores and the Moluccas. These collections additionally included 13 individuals that had been collected by Sigit Wiantoro from the Mangolo caves in southern mainland Sulawesi, which had been suggested to represent another *R. philippinensis* morph pair. Final outgroup samples were *Rhinolophus montanus* from East Timor (provided by Kyle Armstrong) and Bornean *R. philippinensis* (sampled under permit). I extracted DNA from all samples but one using the DNEasy Blood and Tissue extraction kit (Qiagen). The final individual was also used in transcriptome sequencing, so I extracted both RNA and DNA from heart tissue using the AllPrep DNA/RNA mini kit (Qiagen). I quantified the extracted DNA with a Qubit 2.0 fluorometer (Life Technologies) and performed quality assessment using a 2100 Bioanalyzer (Agilent).

Given that the majority of these tissue samples had been in storage for many years, I selected six individuals (two small morphs, one intermediate morph, one large morph, the Bornean *R. philippinensis*) to be sequenced in a preliminary run to verify that sufficient DNA remained of sufficient quality to enable sequencing. I prepared 500bp insert libraries for these samples using the NEBNext Ultra DNA Library Prep Kit for Illumina. The libraries were pooled and sequenced by the Barts and the London Genome Centre

(Illumina HiSeq 2500 V4 Chemistry; 2 x 125bp; 5X sequencing depth). I sent the remaining 53 DNA extractions to Novogene, who used the NEBNext Ultra II kit to prepare 350bp insert libraries. Due to variable library quality, they sequenced an initial run of nine samples (Illumina HiSeq; 2 x 150bp; 10X sequencing depth) in order to allow me to make an informed decision on which samples to proceed with. Novogene carried out some initial quality assessments and filtration of raw reads to remove reads containing adapter sequences, reads containing > 10% Ns and reads with a Qscore <= 5 for over 50% of the read. Novogene additionally carries out a cross-species contamination detection summary in which they blast 5000 randomly selected high quality reads against the NT database.

Reference-guided mitochondrial genome assembly

I used two methods for constructing the mitochondrial genomes from the whole genome sequencing reads; by alignment to reference sequences and *de novo* assembly.

In order to create reference-guided mitochondrial genome assemblies, I trimmed my short read sequencing data using Trimmomatic (Bolger, Lohse and Usadel 2014) (Parameters: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36) and aligned it to the *Rhinolophus macrotis* complete mitochondrion available from GenBank (NCBI Reference Sequence: NC_026460.1). I used the mpileup function available in samtools version 1.1 (Li et al 2009) to generate these alignments, then transformed the resulting bam files into fasta sequences using seqtk (https://github.com/lh3/seqtk). I initially generated assemblies for a sample of 12 individuals representing different morphs and collection dates to verify that the reference-guided assemblies looked accurate.

For the purposes of checking that the mitochondrial sequences had been accurately reconstructed this way, I used a quick and light-weight alignment and tree building method in BioEdit version 7.2.5 (Hall 1999). I generated multiple sequence alignments using these sequences using the clustalw algorithm (Thompson, Higgins and Gibson 1994). I created multiple alignments of the 12 sequenced individuals with the *R. macrotis* sequence that was used as the reference for the assembly, as well as with mitochondrial control region sequences downloaded from GenBank for Buton *R. philippinensis* (accessions: AY568637.1 - AY568646.1, Kingston and Rossiter 2004) and Australian bats assigned as *R. philippinensis* (accessions: AF065069.1 – AF065076.1, Cooper and Skilins 1998). Following alignment of all sequences, I selected just the control sequence region to generate a neighbour-joining tree using the DNADIST version 3.5c (c. Felsenstein 1986-1993, http://evolution.genetics.washington.edu/phylip/doc/dnadist.html).

De novo *mitochondrial genome assembly*

For the *de novo* assemblies, I used NOVOplasty (Dierckxsens, Mardulyn and Smits 2017). This software is designed to construct mitochondrial genomes from whole genome sequencing reads from an input seed sequence. It uses the principles that the mitochondrial reads should be highly enriched, and that the mitochondrial genome circularizes to identify and assemble mitochondrial sequences from subsets of the sequencing reads. I selected seed reads for each individual by blasting reads that successfully aligned to the *R. macrotis* mitochondrial genome until I found a match to one of the mitochondrial genes CO1, CYTB, ND1 or ND2. If NOVOplasty failed to generate a circularized assembly using this seed, I ran it again with a new seed from the

same individual, then with an increased sample of reads, and finally using a seed that was used to successfully generate a circularized assembly for another individual. I aligned these sequences manually in BioEdit version 7.2.5, then aligned them with the Kingston and Rossiter (2004) control region sequences to quickly generate a neighbour-joining tree with DNADIST version 3.5c as described above to assess the assemblies.

Whole genome sequencing reference assembly

In order to create a reference-guided assembly of the Bornean *R. philippinensis* individual, I used the method described by Wang et al (2014). This individual was chosen for the reference as it was a closely-related outgroup to the Buton *R. philippinensis* that had been sampled recently and thus had not spent much time in storage. Additionally, it had been collected and stored in optimal conditions, with harvested tissue being put immediately into AllPrep solution and liquid nitrogen.

I first trimmed the raw Bornean *R. philippinensis* sequencing reads to remove low-quality and adaptor sequences using Trimmomatic version 0.36 (Bolger, Lohse and Usadel 2014) (Parameters: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36). I aligned the cleaned reads to a high-quality *Rhinolophus ferrumequinum* genome that had been sequenced and assembled into DISCOVAR contigs (provided by Jeremy Johnson, Broad institute) using BWA version 0.7.8 (BWA-MEM algorithm; Li and Durbin 2009, Li 2013). I also created a *de novo* assembly of the Bornean *R. philippinensis* reads using SOAPdenovo2 version 2.4 (Luo et al 2012) with a k value of 19, as recommended by KmerGenie (Chikhi and Medvedev 2014). I took all contigs over 1000bp in length from this *de novo* assembly and also aligned those to the Broad Institute *R. ferrumequinum* genome with BWA version 0.7.8

(BWA-MEM algorithm). I combined the mapped reads and the mapped contigs using samtools merge version 1.3.1 (Li et al 2009), then extracted the consensus sequences using mpileup from samtools version 0.1.18 (Li et al 2009) and converted them to fasta files using vcfutils from bcftools version 0.1.17 (Li 2011) and seqtk. I used the ensuing fasta file as a reference to map all of the contigs from the *de novo* assembly that were over 100bp in length using BWA version 0.7.8 (BWA-SW algorithm, Li and Durbin 2010). I also remapped the reference used in this step back onto itself in order to generate another bam file that could be combined with the alignment file of all contigs over 100bp in length using samtools merge version 1.3.1. I extracted the consensus sequences from this final merged bam as described previously and discarded any scaffolds fewer than 1000 base pairs in length in order to generate the final alignment. This whole process has been summarised as a flowchart in Figure 3.2.

Whole genome sequencing alignment and SNP calling

I trimmed each set of *R. philippinensis* sequencing reads using Trimmomatic version 0.36 as described previously, then mapped the trimmed reads to the Bornean *R. philippinensis* assembly with BWA version 0.78 (BWA-MEM algorithm). I marked duplicates in the bam files with Picard tools (http://broadinstitute.github.io/picard) prior to following the best practices documentation for SNP calling with GATK version 3 (McKenna et al 2010, DePristo et al 2011, Van der Auwera et al 2013). This involved first calling variants per sample by using HaplotypeCaller in GVCF mode, splitting each sample into subsets of scaffolds in order to reduce the data to a size that GATK was able to process. I then joint-called variants across all samples using GenotypeGVCFs. I applied a set of hard filters to the called variants across the whole dataset to retain only non-singleton biallelic SNPs

that had been genotyped in every individual with a minimum quality of 20 and a maximum depth of 30 using VCFtools (Danecek et al 2011).

In order to generate a set of high-confidence SNPs for downstream analysis, I called SNPs independently using two further methods and retained only variants identified by all three models. In addition to the GATK pipeline described above, I called SNPs using freebayes version 1.1.0 (Garrison and Marth 2012) and mpileup from bcftools version 1.8 (Li 2011) with default parameters. These SNP sets were filtered as described for the GATK set prior to taking the intersection of the three sets of SNP calls using isec from bcftools version 1.8 (Li 2011).

Population genetics

I calculated measures of genetic differences and population genetics statistics within and between the morphs based on the mitochondrial data using DNASP v6 (Rozas et al 2017). In particular, I used this software to calculate haplotype diversity (after Nei 1987 equation 8.4), number of nucleotide differences (according to Tajima 1983 equation A3), nucleotide diversity (from Nei 1987 equation 10.5), the average number of nucleotide substitutions between populations D_{XY} (following Nei 1987 equation 10.21) and F_{ST} (Hudson et al 1992 equation 3).

I calculated measures of genetic diversity across the whole SNP data set (heterozygosity as the inbreeding coefficient F and mean site-wise nucleotide diversity within populations) using VCFtools (Danecek at al 2011). Additionally, used MavericK to infer population structure between the different individuals (Verity and Nichols 2016). This software uses mixture modelling to infer population structure in the same way as the

better-known STRUCTURE software (Pritchard 2000), but uses a new thermodynamic integration method to estimate the appropriate number of demes to partition the data into (*K*) more accurately. I ran MavericK independently four times using subsets of 10,000 SNPs randomly selected from the set of high-confidence SNP calls with a constraint that each SNP from a set came from a different contig to reduce the chance of linkage between SNPs. I used sets of 10,000 SNPs as increasing the data volume beyond this point would not have led to a clearer result (personal communication from Dr Robert Verity). I ran MavericK with *K* values between one and nine, then looked at the model evidence produced using the thermodynamic integration technique to determine the most appropriate value of *K* for each subset. I first executed this process without admixture, then repeated it with admixture.

Phylogeny and network construction

I constructed maximum-likelihood phylogenetic trees based on multiple sequence alignments of the *de novo* assembled mitochondrial genomes using RAxML HPC version 8.2.11 (Stamatakis 2014) with the GTRGAMMA model. I followed the recommended 'Easy and Fast' protocol as described in the RAxML v8.2.X manual (July 20 2016; https://sco.h-its.org/exelixis/resource/download/NewManual.pdf; accessed July 10 2018) to run rapid bootstrap analysis and search for the best-scoring maximum likelihood tree in a single programme run using the bootstrap convergence criterion. I built trees based on the entire mitochondrial alignments, as well as on alignments of just the control regions with the addition of the control region sequences reported in Kingston and Rossiter (2004). The alignments of the mitochondrial sequences revealed that there was a tandem repeat region with variation in the number of repeats between sequences, so I also

generated a phylogeny from these sequences with a standardised number of tandem repeats.

Additionally, I constructed phylogenies using the same sets of 10,000 SNPs generated previously for the MavericK analysis. MavericK takes biallelic calls from each position, whereas RAxML only takes a single nucleotide sequence; as such, each set of SNPs split into two nucleotide sequences for input into RAxML and I ran RAxML eight times in total. I used the same method as for the mitochondrial phylogenies, except that I used the ASC-GTRGAMMA model with the 'lewis' correction. This model corrects for ascertainment bias introduced by using a SNP data set composed only of variable sites. I had chosen to use sets of 10,000 independent SNPs due to MavericK not benefiting from more information than that in generating results; as this not necessarily true of RAxML, I chose to also generate two phylogenies using a larger number of SNPs. In order to ensure that the SNPs in these sets were unlinked I limited them to containing a single SNP from each contig, giving a total size of 26,946 SNPs.

In order to look at the relationship between different mitochondrial haplotypes, I also built mitochondrial haplotype networks using Network 4.6 (www.fluxus-engineering.com) with Median Joining (Bandelt, Forster and Röhl 1999) and default parameters.

Species delimitation using the multi-species coalescent

I carried out a multi-species coalescent analysis based on the aligned mitochondria with a standardised number of tandem repeats using Bayesian Phylogenetics and Phylogeography (BPP v3.4, Rannala and Yang 2003, Rannala and Yang 2013). BPP is a Markov Chain Monte Carlo programme using a reversible-jump algorithm that analyses

DNA sequences under a multispecies coalescent model (Yang 2015). BPP takes a guide tree and population assignments and uses Bayesian inference to generate a species tree and perform species delimitation based on this data (Yang 2015).

In the first instance, I assigned Buton individuals into one of three populations based on their assigned morphs and counted each of the three outgroup species as their own population, for a total of six populations. For the guide tree I used the most commonly recovered phylogenetic arrangement produced by RAxML based on the sets of 10,000 genomic SNPs. In the second instance, I split the Buton individuals into five populations following the major clades identified by RAxML based on the mitochondrial data, and provided this topology as the guide tree.

For each data set I ran model A11 to perform joint species tree estimation and species delimitation using BPP v3.4. In this model the assignment of species to populations remains fixed; while BPP can merge different populations into single species, it cannot split populations (Zhang, Rannala and Yang 2014). I ran this model with species model prior 1, assigning equal probabilities to the rooted trees. I ran each model four times changing only the starting seed to check for convergence of the models and robustness of the inferences. These analyses require inverse gamma priors; I provided diffuse priors with an uninformative shape parameter due to having little *a priori* information about the parameters (theta ~ IG(3,0.004); tau ~ IG(3,0.002)) and ran the MCMC with 100,000 generations (sampling interval of 2) with a burn-in period of 8,000.

Results

Phenotypic distinctiveness between morphs

Forearm measurements were taken on capture from 92 bats, including 47 Buton *R. philippinensis* (26 Small morph individuals, 7 Intermediates and 14 Large), one Intermediate *R. philippinensis* from neighbouring Kabaena island, 20 *R. philippinensis* from two different locations on mainland Sulawesi, 16 *R. philippinensis* from Queensland, Australia (eight Large morphs and eight Intermediates), 5 *R. philippinensis* from Malaysian Borneo, one *R. philippinensis* from the Moluccas, one *R. philippinensis* from Flores and one *R. montanus* from East Timor.

The Buton Small morph had a forearm length of 47.04mm ± 0.82mm (standard deviation of the mean), the Intermediate 50.64mm ± 1.86mm and the Large 56.86mm ± 1.87mm (Figure 3.8). These ranges were consistent with those described based on smaller sample sizes in the original harmonic hopping paper (Kingston and Rossiter 2004). A one-way ANOVA with a Tukey HSD showed that the means were significantly different between each group (ANOVA: $F = 187.8$, $df = 2$, $p < 0.001$; Tukey HSD adjusted p value < 0.001 between each pair of morphs). In comparison, the Queensland Large morph had a forearm length of 55.13mm ± 1.1mm and the Queensland Intermediate 49.24mm ± 2.58mm, also being significantly different (T-test: $t=-5.94$, $df=9.45$, $p < 0.001$). The mean forearm length for all other outgroups was 50.76mm ± 1.99mm. I only had access to forearm measurements for the mainland Sulawesi bats from the Mangolo caves, but these did not support this population representing two distinct morphs, occupying a very similar range of values to those seen for the Buton Island Intermediate morph (Figure 3.8).

Weight measurements were obtained for a subset of 69 of these bats, excluding 20 *R. philippinensis* from mainland Sulawesi and two from Borneo. The Buton Large morph weighed 11.94g ± 1.68g (standard deviation of the mean), the Intermediate 8.33g ± 0.72g and the Small 6.56g ± 0.48g (Figure 9). Again, these pairs were all significantly differentiated by weight (ANOVA: $F = 93.89$, $df = 2$, $p < 0.001$; Tukey HSD adjusted p values: Large-Intermediate < 0.001, Large-Small < 0.001, Small-Intermediate 0.02.) . When considering the Queensland *R. philippinensis*, the Queensland Large morph weighed 11.59g ± 0.52g and the Queensland Intermediate 9.34g ± 1.03g, which were again significantly differentiated (T-test: $t=-5.94$, $df=9.45$, $p < 0.001$). The pooled outgroups had a mean weight of 8.23g ± 1.68g.

I only had access to echolocation call frequency information for the Buton *R. philippinensis* individuals originally reported in the original harmonic hopping paper (2004). New individuals that were being first analysed in this research were assigned to a morph based on their morphological measurement. To confirm the morph allocations of individuals for which I had mitochondrial data, I compared their forearm sizes and weights with the minima and maxima reported by Kingston and Rossiter (2004). When considering forearm length (Figure 3.10), all 19 Small morphs fitted within the range previously described, and the lower bound for the 11 Large morphs was consistent with that described in Kingston and Rossiter (2004) even though the upper bound was higher. The four Intermediates had a greater range of forearm lengths than previously described with two larger individuals present, but they still remained smaller than the Large morphs. Conversely, on weight (Figure 3.11) the four Intermediate morphs fitted inside the previously defined range, with the Small morphs (19) extending into slightly lighter weights, but no heavier. The ten Large individuals spread out both sides of the published weight range, with both heavier and lighter individuals recorded, but remained clearly differentiated from the Intermediates.

DNA sequencing

I initially sent extracted DNA for 53 samples of individuals designated *R. philippinensis* and one of an individual designated *R. montanus* to Novogene. Their internal QC procedure revealed that many of these samples did not meet their quality thresholds, so they carried out library preparation and sequencing for a preliminary run of nine samples to verify that the data quality would be sufficient to pursue this project. Two of these samples failed library preparation, with the remaining seven being successfully sequenced. The two samples that failed were part of a set of 12 samples that had been collected in a single visit to a mainland Sulawesi population of *R. philippinensis*, so I chose not to continue with the sequencing of the remaining ten individuals from that population. Of the remaining samples, three failed library preparation, meaning that I got sequence data back for 38 of these individuals in total.

In combination with the six pilot samples sequenced by Barts and the London Genome Centre, I generated sequence data for 13 large morphs, 21 small morphs and 4 intermediate morphs (one of which was sequenced twice, once by each sequencing center, to check consistency of results) of *R. philippinensis* from Buton island. I additionally had individual outgroup samples sequenced for *R. philippinensis* from the Moluccas, Flores, Kabaena and Borneo and an *R. montanus* individual from East Timor. Novogene's cross-species contamination detection procedure revealed that the sequencing reads for the individual from the Moluccas were heavily contaminated by bacterial sequence (>60% tested reads blasting to *Citrobacter freundii*), so this individual was omitted from further analysis. Details of sequence data volume obtained for all successfully sequenced individuals are in Table 3.1.

Mitochondrial assembly by alignment to a reference

I initially attempted to reconstruct the mitochondrial genomes from the whole genome sequencing reads by aligning them to a reference mitochondrial sequence downloaded from GenBank (*Rhinolophus macrotis*, NC_026460.1). To verify this approach, I initially aligning reads to this reference for a subset of 12 individuals – 10 Buton samples and two outgroup bats – and used these in a multiple sequence alignment including the *R. macrotis* reference and the Buton *R. philippinensis* control region sequences published by Kingston and Rossiter (2004). I extracted a 489 base pair section of the alignment corresponding to the control region and used this to construct a phylogeny (Figure 3.3). The phylogeny generated from this alignment correctly separated the outgroup sequences (*R. montanus*, the Bornean and Australian *R. philippinensis* sequences) from the Buton bats. However, it separated the Kingston and Rossiter (2004) sequences from the newly generated sequences. This result was particularly notable considering that three of the individuals (two small and one large) had been sequenced both in that original paper and in this new research. Given these discrepancies, I chose not to use the mitochondrial sequences generated by alignment to a reference in further analyses.

De novo *mitochondrial read assembly*

NOVOplasty (Dierckxsens, Mardulyn and Smits 2017) successfully generated 40 de novo circularized mitochondrial genomes. An additional three genomes were mostly assembled, missing only small segments, and as such were easily manually aligned with the complete mitochondrial genomes. These initial mitochondrial alignments were 17,091

base pairs in length. When I generated a phylogeny to check these complete mitochondrial alignments, one large morph individual fell outside of the clade defined by the Buton bats, and one outgroup fell within this clade (Figure 3.4). An 11-base tandem repeat region (CATACGCAACG) in the NOVOplasty assemblies was present a variable number of times in the circularized assemblies (range: six to 41). There was no apparent phylogenetic signal in this variable number of repeats (Table 3.2). I trimmed the number of repeats to standardise them at six in each sequence in all downstream analyses, leaving an alignment of 16,706 base pairs in length. This alteration did not change the membership of any of the major groupings visible in the phylogeny, though relationships between individuals within one of these clades appeared slightly rearranged (Figure 3.5). Given that the large morph and outgroup individuals were still placed unexpectedly, and given that these were two individuals were missing data from the alignment as NOVOplasty had failed to circularise these mitogenomes, I decided to remove them from further mitochondrial DNA analyses (Figure 3.6).

When I constructed a phylogeny using a 460 base pair segment corresponding to the control regions extracted from this alignment, along with the control regions from the original harmonic hopping paper (Kingston and Rossiter 2004), the Buton bats no longer separated out based on whether they were represented by the old or the new sequences (Figure 3.7).

Genomic SNP identification

The Bornean *R. philippinensis* reference genome assembly consisted of 1,236,521 contigs assembled into 47,912 scaffolds. The scaffolds were between 1000 and 1,876,277 base pairs in length with a mean length of 43,360 base pairs and an N50 length of 166,896 base

pairs. The total assembly length was 2,077,441,914 base pairs, or just over 2 Gbp. After mapping short reads from all other individuals to this reference assembly, 2,295,016 SNPs were called by GATK, 5,055,388 by samtools and 3,999,838 by Freebayes. 1,319,236 of these SNPs were called using all three methods, with a mean coverage depth per-individual of between 4.2 and 9.3.

Mitochondrial distinctiveness between morphs

Generating a maximum likelihood phylogeny with RAxML (Stamatakis 2014) based on the mitochondrial sequences alone revealed a complex picture (Figure 3.6). While individuals clustered with other individuals of the same morph, only the Intermediate morph formed a monophyletic group. Both the Small and the Large morphs were classified into two polyphyletic groups. The polyphyly of the Large morphs had much stronger bootstrap support, but the reason for the lower support on the groups for the Small morph appears to be because the two groups can switch position with one another, rather than that they break down to a single group. There are no significant difference between the two Large morph groups or between the two Small morph groups on either forearm length (T-test: Large: t=-0.38, df=6.69, p=0.72; Small: t=-0.62, df=14.94, p=0.54) or weight (T-test: Large: t=-1.49, df=6.67, p=0.18; Small: t=-0.25, df=11.61, p=0.81). Sexes are also mixed within the two groups for each morph (Large groups: three males and three females; four males and two females. Small groups: Two males and five females; 10 males, four females and one unreported).

This phylogeny also indicates that the Large morph might be the ancestral state of the Buton *R. philippinensis,* and that the Intermediate morph is more closely related to the

Large morph than it is to the Small morph. The Kabaena individual falls outside of the Buton individuals, with the other outgroups.

A mitochondrial haplotype network paints much the same picture as the maximum likelihood mitochondrial phylogeny (Figure 3.12).

Population genetics

Of 16,706 sites in the mitochondrial alignment, 789 were polymorphic when looking across the whole dataset, of which 561 were singleton variant sites. The Intermediate morphs had the lowest level of within-population diversity (Table 3.3), with the Large and Small morphs being similarly diverse – the Large morphs have a greater level of nucleotide diversity, but the Small morphs have a higher haplotype diversity. There were no fixed differences between the Large and Intermediate mitochondria, one between the Large and Small and eight between the Small and Intermediate (Table 3.4). F_{ST} is highest between the Intermediate and the Small morphs, but similar between the Intermediate and the Small and the Small and the Large morphs (Table 3.4). However, the average nucleotide divergence D_{XY}, showing the average proportion of nucleotide differences between populations, is highest between the Large and Small morphs, and lowest between the Large and Intermediate morphs (Table 3.4). The morphs are significantly differentiated based on their mitochondrial sequences whether we use haplotype-based (Hudson's Hs 0.8; $p < 0.001$) or nucleotide-based (Hudson's Ks 13.7; $p < 0.001$) statistics.

VCFtools calculates the inbreeding coefficient F as a measure of heterozygosity. Based on the full set of 1,319,236 genomic SNPs, the Intermediate morphs had the lowest value of F (0.3), followed by the Large morphs (0.34) with the small morphs having the greatest

value of F (0.41). Positive values of F indicate fewer heterozygotes than expected in line with Hardy-Weinberg principles, so this suggests that the Small morphs have the fewest heterozygotes and the Intermediate morphs the most. Nucleotide diversity within a population, π, was consistent with this result; while the Large and Intermediate morphs had a similar mean genome-wide nucleotide diversity (0.137 and 0.135 respectively), the Small morphs had a slightly lower value (0.121).

Mean Weir and Cockerham F_{ST} values calculated across the whole set of genomic SNPs indicate that the Large and the Intermediate morphs are the most similar ($F_{ST} = 0.12$), with the Small morph being similarly differentiated from each ($F_{ST} = 0.16$). The sex-specific mean F_{ST} values were consistent between males and females for the Large and Small morphs (0.16) and Small and Intermediate morphs (0.13). There was a small difference between the sexes for the Large and Intermediate morphs (male: 0.086, female 0.098).

Population structure

Of the eight maximum likelihood trees generated off 10,000 unlinked genomic SNPs, all found reciprocally monophyletic clades for the three morphs. The most commonly recovered topology (four of the eight trees) placed the Intermediate and Large morphs as sister groups, with all Buton bats forming a monophyletic clade distinct from the outgroup individuals (Figure 13.3a). In the second most common topology (two of the eight trees), the Large morphs split first, with the Small and Intermediate morphs as sister groups and the Kabaena individual in a group with the intermediated (Figure 13.3b). The final two topologies were seen once each; in one the Intermediate and Small morphs were again sister groups, but to the exclusion of the Kabaena individual (Figure 13.3c), while in the other the Intermediate morph was the first to separate, leaving the Small and Large as

sister groups (Figure 13.3d). Increasing the sample size to 26,046 SNPs did not resolve between these topologies, finding the topology seen in Figure 13.3a twice (with the Intermediate morphs as a sister group to the Large morphs), and those from 13.3b and 13.3c once each (with the Intermediate morphs as sister group to the Small morphs, but differing slightly in the placement of the Kabaena individual).

When MavericK was run without admixture, the thermodynamic integration estimator supported a *K* value of three in three runs, while in the final run it supported a *K* of four. At *K*=3, the Buton Small morphs form their own cluster, there is an outgroup cluster containing the Flores and Borneo individuals and *R. montanus*), and the final cluster consists of the Buton Large and Intermediate morphs with the Kabaena individual (Figure 3.14). At *K*=4, each of the Buton morphs separates out into their own cluster, with the Kabaena individual grouping with the Buton Intermediates and the rest of the outgroup individuals forming their own cluster (Figure 3.15).

When admixture is permitted, the thermodynamic integration method still indicates an optimal *K* value of three. At *K*=3 there is very little admixture between the Large and the Small morphs, or with most of the outgroups (Figure 3.16). The Intermediate morphs and the Kabaena individual appear to be an almost even mix of the Small and Large morphs, with the Borneo individual also showing some similarity with the Small Buton morphs.

Species delimitation using the multi-species coalescent

BPP results were consistent over the four runs for each analysis, with Tracer (Rambaut et al 2018) plots of the MCMC demonstrating that the runs had converged.

Joint species tree estimation and species delimitation using model A11 on the six-species data set returned the highest posterior probability for a division of the data into six species (0.71-0.73; second highest posterior probability for five species with the Large and Intermediate morphs combined 0.08-0.14). The differences between the top tree arrangements all concerned the placement of the outgroups (the top two trees, each with a posterior probability of 0.12, only changed between whether the Bornean *R. philippinensis* or the *R. montanus* from East Timor was the most basal taxa), with the relationships between the Buton morphs remaining consistent with that supplied in the guide tree.

The results for the eight-species population assignments differed substantially from those for the six-species assignments, though were consistent across multiple runs. The joint species estimation model in this case found the strongest support for a seven-species tree in which the Intermediates were grouped with one of the Large clades (posterior probability 0.51-0.59), then for all eight clades representing separate species (posterior probability 0.25-0.27). There were multiple tree topologies with similar levels of posterior probability produced by this model for the seven-species tree, representing some rearrangements of the relationships between the Buton morphs. While the guide tree consistently received the highest level of support, rearrangements in which the positions of the two small clades were reversed or in which the two small clades were sister groups received almost as much support (Figure 3.17).

Discussion

Three morphs of *Rhinolophus philippinensis* are found in the same location on Buton island. These morphs differ in size and in the frequency of their echolocation calls. Initial research on this system indicated that these morphs are also genetically differentiated to a level that suggested that they did not represent a single panmictic population (Kingston and Rossiter 2004). The previous research on this system used a small sample size and a very limited number of genetic markers. I have used high-throughput sequencing to generate short-read whole-genome data for 44 Buton bats representing the three morphs and a further four outgroup individuals. Morphological, mitochondrial and genomic results all support the initial conclusions in the Kingston and Rossiter (2004) paper, that the three *R. philippinensis* morphs on Buton island may be in the process of developing reproductive isolation and be in the early stages of speciating.

The Large, Intermediate and Small morphs are significantly differentiated from one another on both weight and on forearm length. The morphological patterns observed in this much larger dataset were consistent with those that were described by Kingston and Rossiter (2004). Adding more individuals increased the ranges of values that were observed in each morph, the means were similar between the two studies and the range of values found in each morph remained non-overlapping. Taken together, the weight and forearm measurements indicate that the three Buton morphs as *R. philippinensis* are morphologically differentiated to a similar degree as the two *R. philippinensis* morphs that have previously been characterised in Queensland morphs (Cooper 1998). While the taxonomic classification of the Queensland *R. philippinensis* is still subject to debate (Sedlock 2012), they have been proposed to represent two separate species.

Genetic analyses based on sets of 10,000 unlinked genomic SNPs indicate divergence between the morphs. Neighbour-joining phylogenetic trees separated the Buton morphs into monophyletic groups, with the Large and Intermediate morphs as sister groups. Mixture modelling using MavericK to infer population structure indicated that the optimal number of demes to divide the population into was either three or four. At K=3, MavericK created a cluster for the Small morphs, a cluster for the outgroups, and a cluster containing the Large and Intermediate morphs as well as the individual from the nearby Kabaena Island. At K=4, MavericK additionally split the Large and Intermediate morphs into their own clusters, though the Kabaena individual remained in a cluster with the Buton Intermediates.

I found the highest inbreeding coefficient and the lowest mean genome-wide nucleotide diversity in the Small morphs. This was consistent with the results of Kingston and Rossiter (2014), who suggested that their discovery of highest heterozygosity in the Large morphs indicated that this was the ancestral form with a founder effect leading to reduced genetic diversity in the other forms. However, I found a lower inbreeding coefficient in the Intermediate morph than the Large, with the two having very similar nucleotide diversity. This result suggests that the Intermediate may be a hybrid of the Large and the Small morphs. This is supported by the results of running MavericK with admixture; at the optimal K value of three there is almost complete separation between the Small and the Large morphs, with the Intermediates appearing to be a mix of the two.

Genetic analyses of mitochondrial data did not provide a clear picture with morphs forming monophyletic groups; while individuals still cluster by morph to the exclusion of individuals of other morphs, the Large and the Small morphs are each split into two distinct clusters, with some Large morphs being more closely related to the Intermediate morphs than they are to the other Large morphs. The relationship between one group of Large morphs and the Intermediate morphs appears even closer than in the trees based on

nuclear SNP data, with one Intermediate morph sitting as an outgroup to the Large morphs in some reconstructions and with the rest of the Intermediate morphs in others. In contrast to the pattern from nuclear variants, the Intermediate morph had the lowest within-population diversity indices based on mitochondrial sequences. Again, though, the Large and the Small morphs were the most differentiated, with the Large and the Intermediate morphs the most similar.

There are a number of possible explanations for mitonuclear discordance, including incomplete lineage sorting of ancestral polymorphisms, sex-biased gene flow, and introgression of mitochondrial DNA (Dong et al 2014). Introgression of mitochondrial DNA between the Intermediate morphs and either the Large or the Small morphs may be favoured by their acoustic system. The Intermediate morph should be able to hear the frequencies used by the Large morph without being audible to the Large morph; similarly, the Intermediate should be able to be heard by the Small morph without being reciprocally receptive (Kingston and Rossiter 2004), which could allow for unbalanced gene flow from one population into the other. Mitochondrial introgression and sex-biased gene flow have both been extensively documented between the *Rhinolophus* species in and around China (Mao et al 2010a, Mao et al 2010b, Mao et al 2013a, Mao et al 2013b). Many of the lines of evidence that can be called upon to differentiate between different causes for mitonuclear discordance are not available in the case of the Buton *R. philippinensis* morphs. There are no clear local source populations where these morphs currently exist in allopatry, removing the option to look at genetic patterns between the morphs in different, non-overlapping parts of their range (e.g. as described in Mao 2013a). In cases of paraphyly generated by incomplete lineage sorting, you would expect to see similar results based on mitochondrial and nuclear markers (Palumbi, Cipriano and Hare 2001); this has been used as evidence against incomplete lineage sorting being the explanation for paraphyly based mitochondrial genes in the presence of monophyly from nuclear loci

in *Rhinolophus sinicus* and *Rhinolophus septentrionalis* (Mao et al 2013a). However, this study differs in that I have not constructed phylogenies based on singular nuclear loci, but rather on a panel of SNPs representing 10,000 independent loci. As such, discordance in a few nuclear genes would be masked by the dominant phylogenetic signal. Under a prediction of male-biased gene flow (as might be expected in bats, due to female philopatry in many species (Dong et al 2014)), you would expect to see greater differentiation based on bi-parentally inherited markers in males than in females (Prugnolle and de Meeus 2002). The only population comparison where there was a difference in F_{ST} based on SNP data was between the Large and the Intermediate morphs, where the males were slightly more similar than the females.

In many cases of mitonuclear discordance one sees a clearer signal of population subdivision using mitochondrial data than using nuclear loci, which is suggested to be a result of faster coalescent times of mitochondrial DNA meaning that nuclear DNA acts as a lagging marker to capture phylogeographic divergence (Barrowclough and Zink 2009). Despite the paraphyly of the Large and Small morphs seen in the maximum-likelihood tree based on the mitochondrial sequences, I found higher values of F_{ST} between the Buton morphs based on the mitochondrial data than based on the whole-genome SNP data. Research in another *Rhinolophus* clade with a complex phylogeny of cryptic species and evidence of mitochondrial introgression cautions against the use of mitochondrial DNA for phylogenetic inference, reporting that using two nuclear introns recovered a better supported species phylogeny than mitochondrial DNA alone, but that combining a small number of nuclear introns with mitochondrial markers would return the mitochondrial tree (Dool et al 2016).

I initially attempted to construct mitochondrial genomes by generating reference alignments to the mitochondrial genomes of a closely related species, *Rhinolophus macrotis*, downloaded from GenBank. However, when I assessed these alignments by

generating phylogenetic trees with the control regions from my sequences and the sequences reported by Kingston and Rossiter (2004), the results were unexpected. The two sets of sequences formed independent clusters in the phylogeny, despite the fact that some individuals had been sequenced in both studies. A plausible explanation for this result is contamination with nuclear mitochondrial pseudogenes (NUMTs). These have been reported in a wide range of species and have been documented as leading to incorrect inferences when DNA was not properly purified prior to PCR amplification (Bensasson et al 2001). NUMTs have been documented as causing problems with high-throughput sequencing, especially where sequencing reads are aligned to mitochondrial genomes, as many aligners will make a random choice in the case of multiple mapping (Ye et al 2014). This poses two possibilities for where the error could have been introduced: either in the original alignment downloaded from GenBank, or in the process of aligning my own reads to the reference genome. NUMTs associated specifically with the mitochondrial control region have been studied in detail in *Rhinolophus*, with at least three independent transpositions of mitochondrial DNA into the nucleus identified (Shi et al 2016). *Rhinolophus ferrumequinum* was found to have a greater total length and total genome proportion of NUMTs than had been documented in many other mammals (though fewer than *Homo sapiens*), with this being suggested to possibly be linked to the rapid radiation of the *Rhinolophus* genus (Shi et al 2016).

Once I generated *de novo* mitochondrial genome assemblies, the separation between the two sets of sequences disappeared in the phylogeny. This method raised a new problem where the mitochondrial genomes for different individuals were different lengths. Visual inspection revealed that this length variation was attributable to an up to seven-fold variation in the number of an 11 base-pair tandem repeat. This variation may be attributable to the mitochondrial assembly method, however high variability in the number of mitochondrial repeats have been widely reported in animals (e.g. Bentzen

1998, Omote 2013). Variation in number of mitochondrial tandem repeats has been used to differentiate between species and subspecies (Hernández et al 2003), but has also been observed within populations of a single species (e.g. Mundy and Woodruff 1996, Padhi 2014, Wang 2015). There was no evidence of phylogenetic signal in the number of tandem repeats present in different individuals in my data, and standardising the number of repeats to six did not change the inferences, so this is the approach that I took in this paper. Heteroplasmy with respect to sequence length has also previously been reported in *Rhinolophus*; in particular, a very similar 11 base pair tandem repeat to that which I found in *R. philippinensis* has been described in the control region of *Rhinolophus sinicus* (Mao et al 2014).

My morphological results support the phenotypic definitions of the three morphs as presented in Kingston and Rossiter (2004). I find the greatest nucleotide diversity in the Large morph and the lowest in the Small morph, consistent with their inference that the Large morph may be the ancestral form with the Small morph losing diversity due to a founder effect, or may be less diverse due to being a younger species. While my phylogeny based on a full mitochondrial alignment provides greater node resolution than in the original control-region phylogeny published by Kingston and Rossiter (2004), the story is the same, with two separate clades of each Small and Large morphs, and some Large morphs being more closely related to the Intermediate morphs than they are to the other Large morphs. However, the addition of a phylogeny based on a large quantity of nuclear loci resolves this paraphyly.

Multiple analyses of independent sets of 10,000 unlinked genomic SNPs support the hypothesis that the *R. philippinensis* morphs of Buton island form three monophyletic groups, albeit with a close enough relationship between the Large and the Intermediate morph that there is conflicting evidence as to whether they should be considered separate demes or not. Results from mitochondrial loci conflict with this finding, with analyses

repeatedly splitting the Large morphs and the Small morphs into two distinct clades each. A finding of mitonuclear discordance in this system could support the hypotheses of gene flow between the morphs or a recent separation, depending on whether it results from introgression or from incomplete lineage sorting. Available data has not allowed me to draw a clear conclusion on the reasons for this mitonuclear discordance, though multiple lines of evidence (MavericK simulations using nuclear SNPs allowing admixture, BPP coalescent simulations, neighbour-joining trees based on mitochondrial data, population statistics) imply gene flow between the Intermediate morphs and some of the Large morphs, with higher measures of heterozygosity in the Intermediate suggesting that it may even have a hybrid origin. Specific analyses of gene flow between the morphs would be needed to confirm this. I have additionally found evidence against panmixia, with results from MavericK and BPP suggesting that at least the Small and the Large morphs are sufficiently genetically distinct to be considered separate populations, though with conflicting evidence as to whether the Intermediate represents a third distinct group.

Individual	Species	Origin	Raw reads
Flores	*R. philippinensis*	Flores, Indonesia	68,324,888
Kabaena	*R. philippinensis*	Kabaena, Indonesia	45,881,400
Montanus	*R. montanus*	East Timor	74,160,166
Mulu	*R. philippinensis*	Borneo	43,111,653
Large 1	*R. philippinensis* Large morph	Buton, Indonesia	59,646,741
Large 2	*R. philippinensis* Large morph	Buton, Indonesia	77,318,322
Large 3	*R. philippinensis* Large morph	Buton, Indonesia	71,176,296
Large 4	*R. philippinensis* Large morph	Buton, Indonesia	68,909,655
Large 5	*R. philippinensis* Large morph	Buton, Indonesia	47,444,574
Large 6	*R. philippinensis* Large morph	Buton, Indonesia	73,659,676
Large 7	*R. philippinensis* Large morph	Buton, Indonesia	76,806,895
Large 8	*R. philippinensis* Large morph	Buton, Indonesia	67,688,372
Large 9	*R. philippinensis* Large morph	Buton, Indonesia	50,373,700
Large 10	*R. philippinensis* Large morph	Buton, Indonesia	87,731,152
Large 11	*R. philippinensis* Large morph	Buton, Indonesia	72,723,432
Large 12	*R. philippinensis* Large morph	Buton, Indonesia	75,286,812
Large 13	*R. philippinensis* Large morph	Buton, Indonesia	73,045,865
Int 1	*R. philippinensis* Intermediate morph	Buton, Indonesia	75,859,527
Int 2	*R. philippinensis* Intermediate morph	Buton, Indonesia	70,389,016
Int 3	*R. philippinensis* Intermediate morph	Buton, Indonesia	57,758,654
Int 3b	*R. philippinensis* Intermediate morph	Buton, Indonesia	68,408,435
Int 4	*R. philippinensis* Intermediate morph	Buton, Indonesia	72,129,157
Small 1	*R. philippinensis* Small morph	Buton, Indonesia	77,883,596
Small 2	*R. philippinensis* Small morph	Buton, Indonesia	71,878,348
Small 3	*R. philippinensis* Small morph	Buton, Indonesia	83,463,703
Small 4	*R. philippinensis* Small morph	Buton, Indonesia	79,530,403

Small 5	*R. philippinensis* Small morph	Buton, Indonesia	68,857,860
Small 6	*R. philippinensis* Small morph	Buton, Indonesia	69,705,605
Small 7	*R. philippinensis* Small morph	Buton, Indonesia	70,949,864
Small 8	*R. philippinensis* Small morph	Buton, Indonesia	70,053,859
Small 9	*R. philippinensis* Small morph	Buton, Indonesia	67,216,697
Small 10	*R. philippinensis* Small morph	Buton, Indonesia	67,133,664
Small 11	*R. philippinensis* Small morph	Buton, Indonesia	70,808,732
Small 12	*R. philippinensis* Small morph	Buton, Indonesia	74,556,135
Small 13	*R. philippinensis* Small morph	Buton, Indonesia	38,134,900
Small 14	*R. philippinensis* Small morph	Buton, Indonesia	40,134,924
Small 15	*R. philippinensis* Small morph	Buton, Indonesia	77,468,812
Small 16	*R. philippinensis* Small morph	Buton, Indonesia	71,090,469
Small 17	*R. philippinensis* Small morph	Buton, Indonesia	83,285,157
Small 18	*R. philippinensis* Small morph	Buton, Indonesia	72,581,494
Small 19	*R. philippinensis* Small morph	Buton, Indonesia	85,797,633
Small 20	*R. philippinensis* Small morph	Buton, Indonesia	75,975,128
Small 21	*R. philippinensis* Small morph	Buton, Indonesia	72,406,167

Table 3.1: Sequence volume obtained for *R. philippinensis* individuals

Number of tandem repeats	Morphs
6	2 Large, 1 Int
7	1 Small
8	1 Int, 4 Small, *R. montanus*
9	1 Int, 2 Small
10	1 Large, 1 Small
11	1 Large, 3 Small
12	1 Large, 1 Int, 1 Kabaena, 2 Small
13	1 Small
14	1 Small
15	1 Large, 1 Borneo
16	1 Large
18	1 Large
20	1 Small
25	1 Flores
26	1 Large
28	2 Small
30	1 Large
34	1 Large
35	1 Small
37	2 Small
38	1 Int
41	1 Large

Table 3.2: Number of tandem repeats in the repetitive mitochondrial region in different individuals

	Large	Intermediate	Small
Number of segregating sites	39	1	37
Number of haplotypes	6	2	9
Haplotype diversity	0.76	0.5	0.83
Average number of differences	18.79	0.4	13.62
Nucleotide diversity	0.0011	0.00003	0.00082
Number of sequences	12	4	21

Table 3.3: Within-population statistics based on mitochondrial data

	Large-Intermediate	Large-Small	Small-Intermediate
Fixed differences between populations	2	1	8
Sites polymorphic in population 1 but not in population 2	39	38	37
Sites polymorphic in population 2 but not in population 1	1	36	1
Average nucleotide differences between populations	19.25	30.8	22.5
F_{ST}	0.47	0.47	0.68
D_{XY}	0.0012	0.0019	0.0014

Table 3.4: Between-population statistics based on mitochondrial data

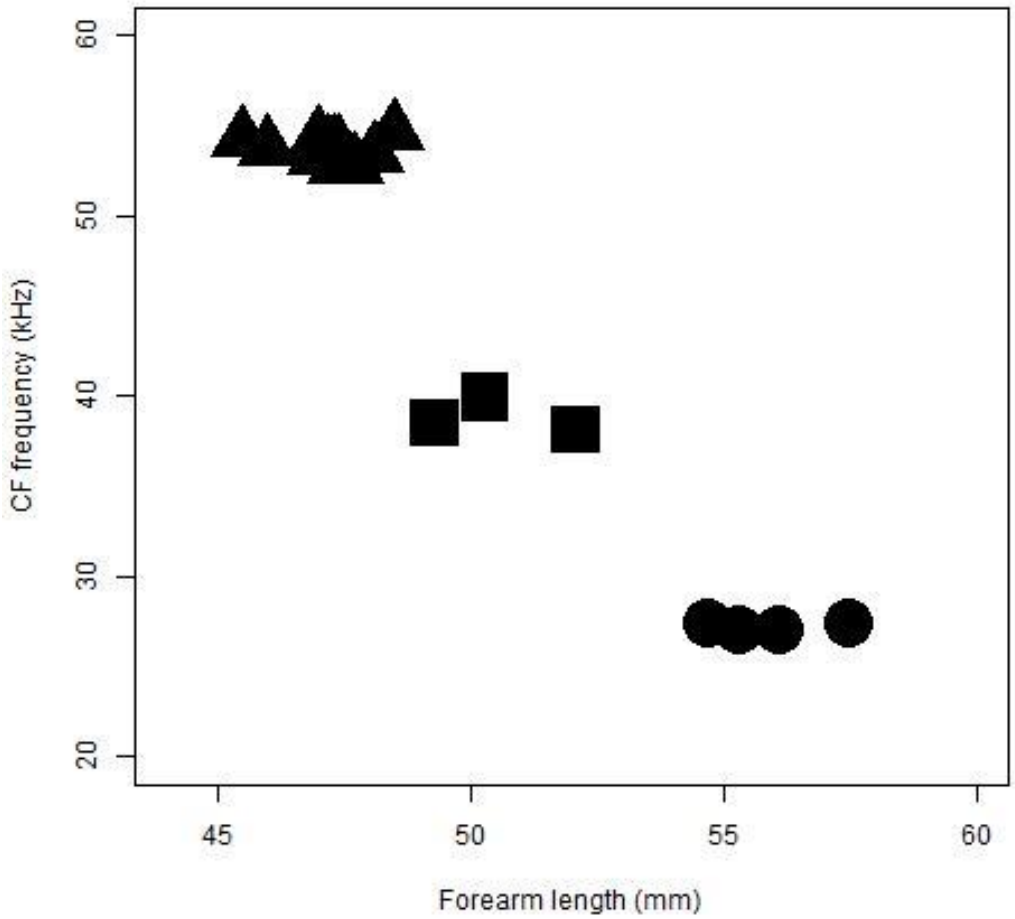


Figure 3.1: Grouping of Buton *R. philippinensis* morphs based on echolocation frequency and forearm length. Filled triangles represent small morphs, filled squares intermediate morphs and filled circles large morphs. Reproduced from Kingston and Rossiter (2004).

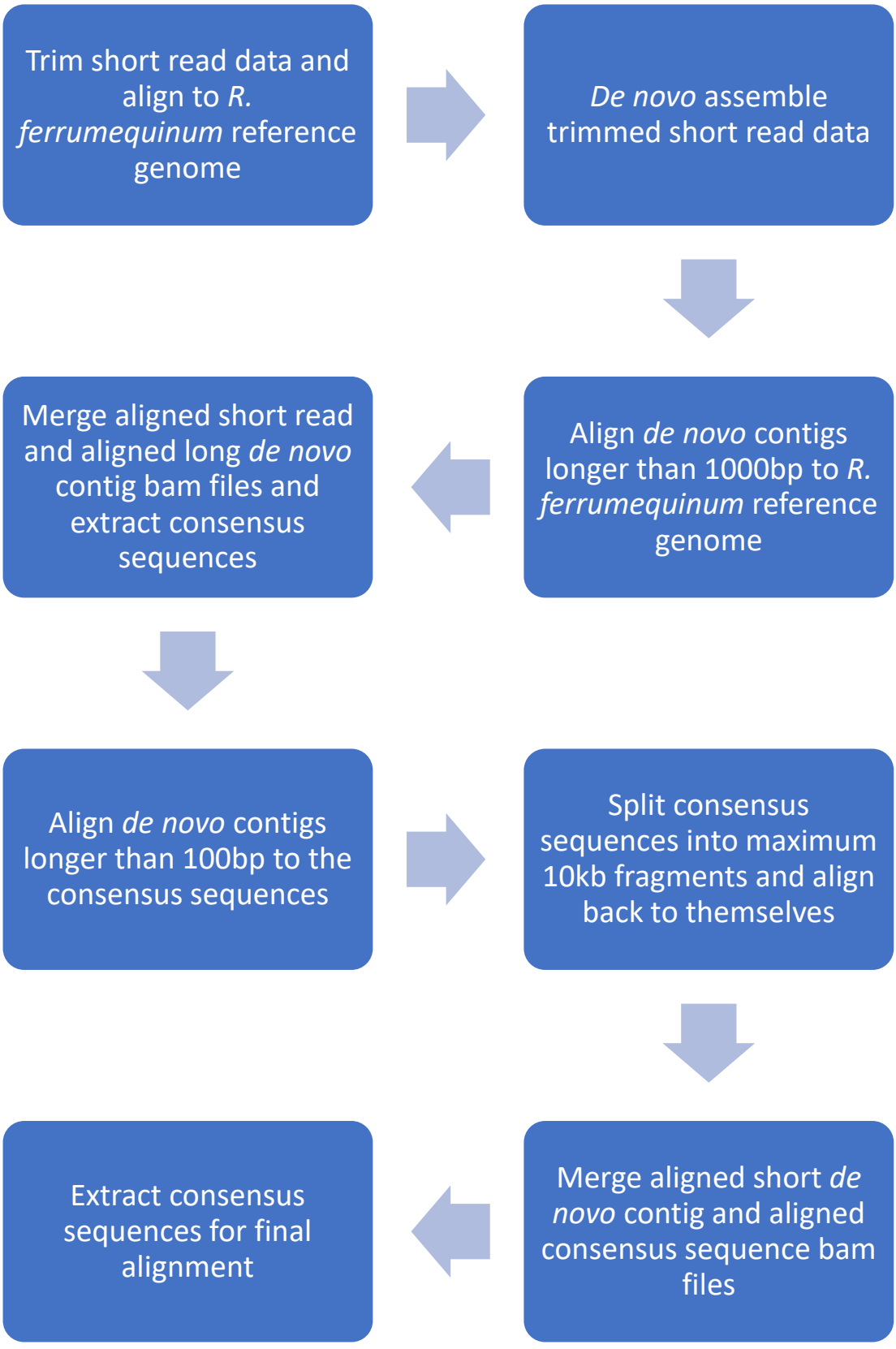


Figure 3.2: Flowchart demonstrating the process used to assemble a *Rhinolophus philippinensis* genome to use as a reference, derived from Wang et al (2014).

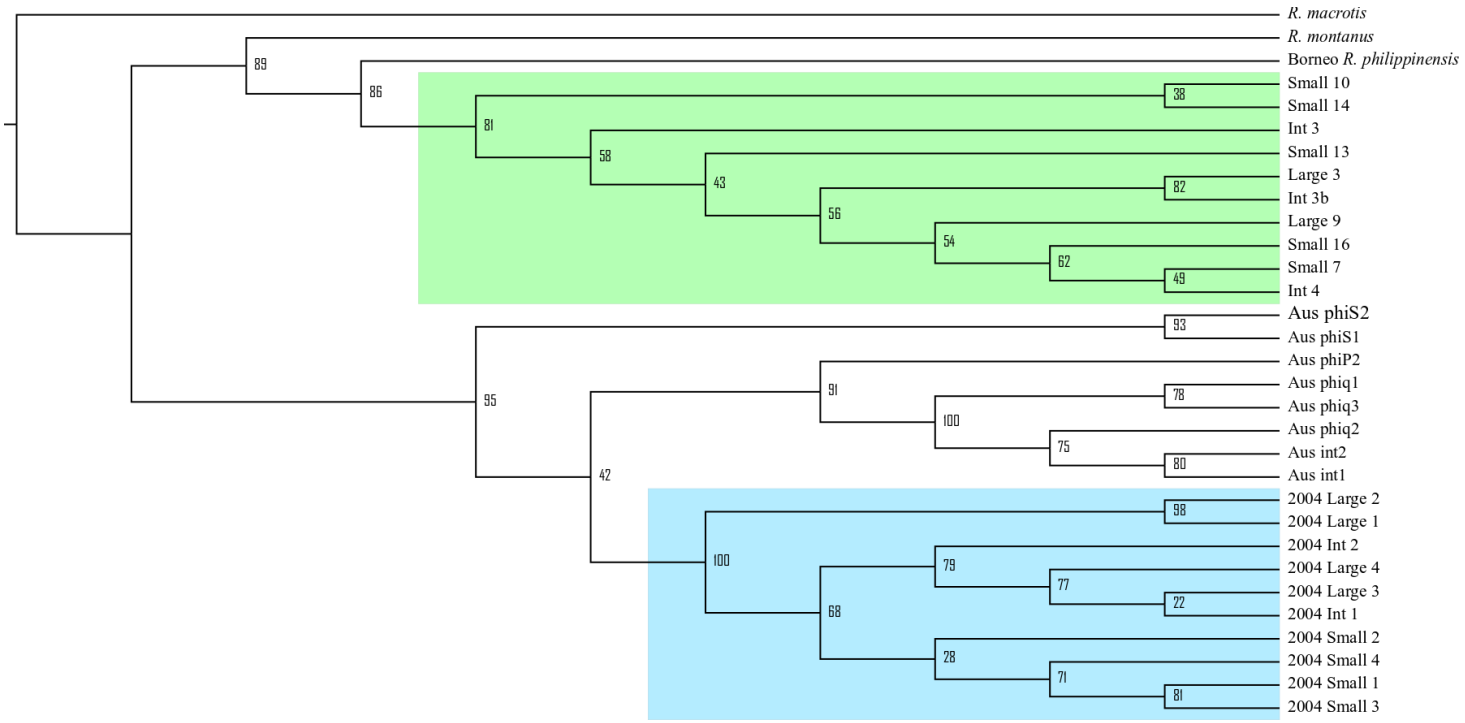
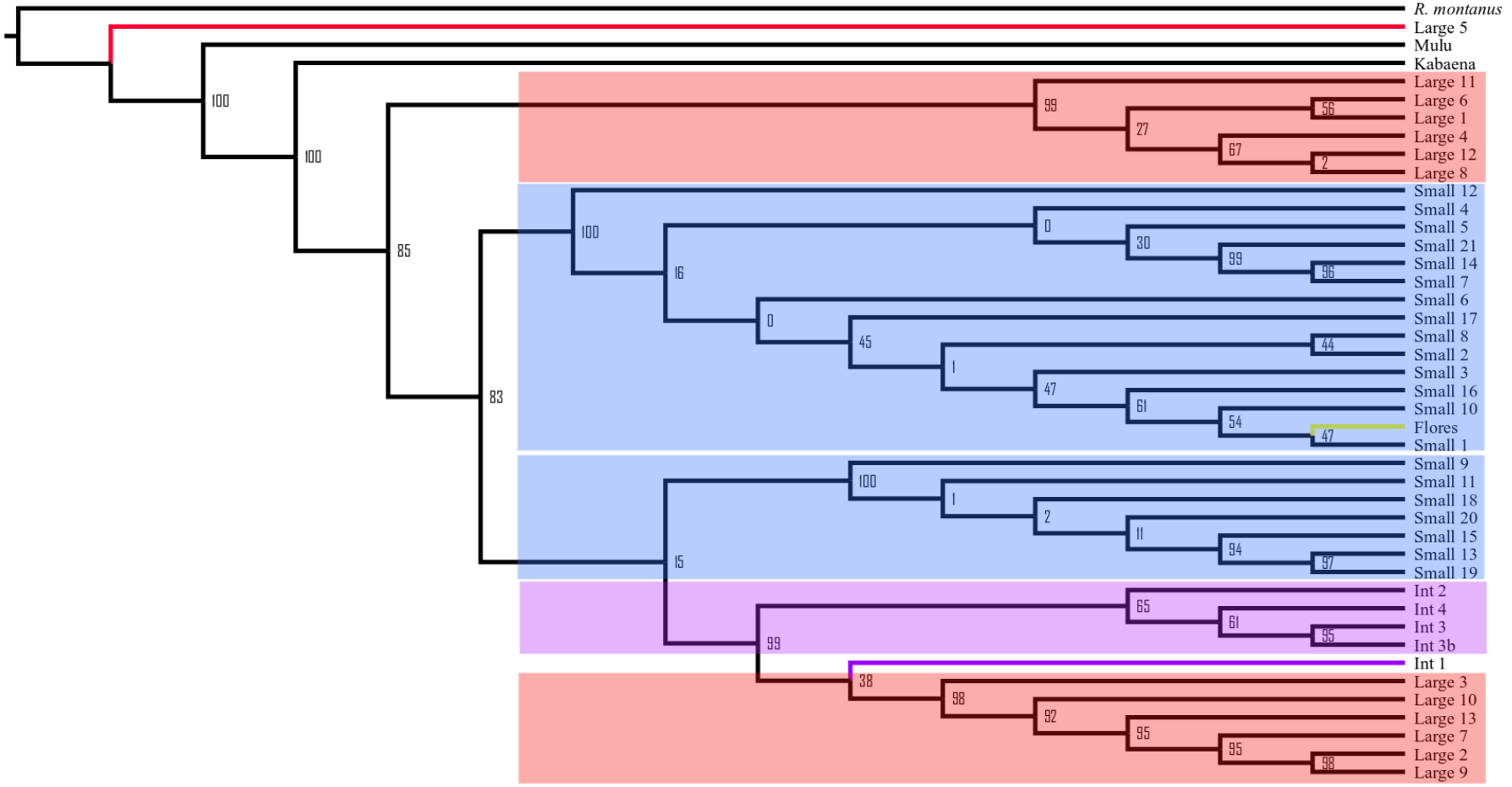
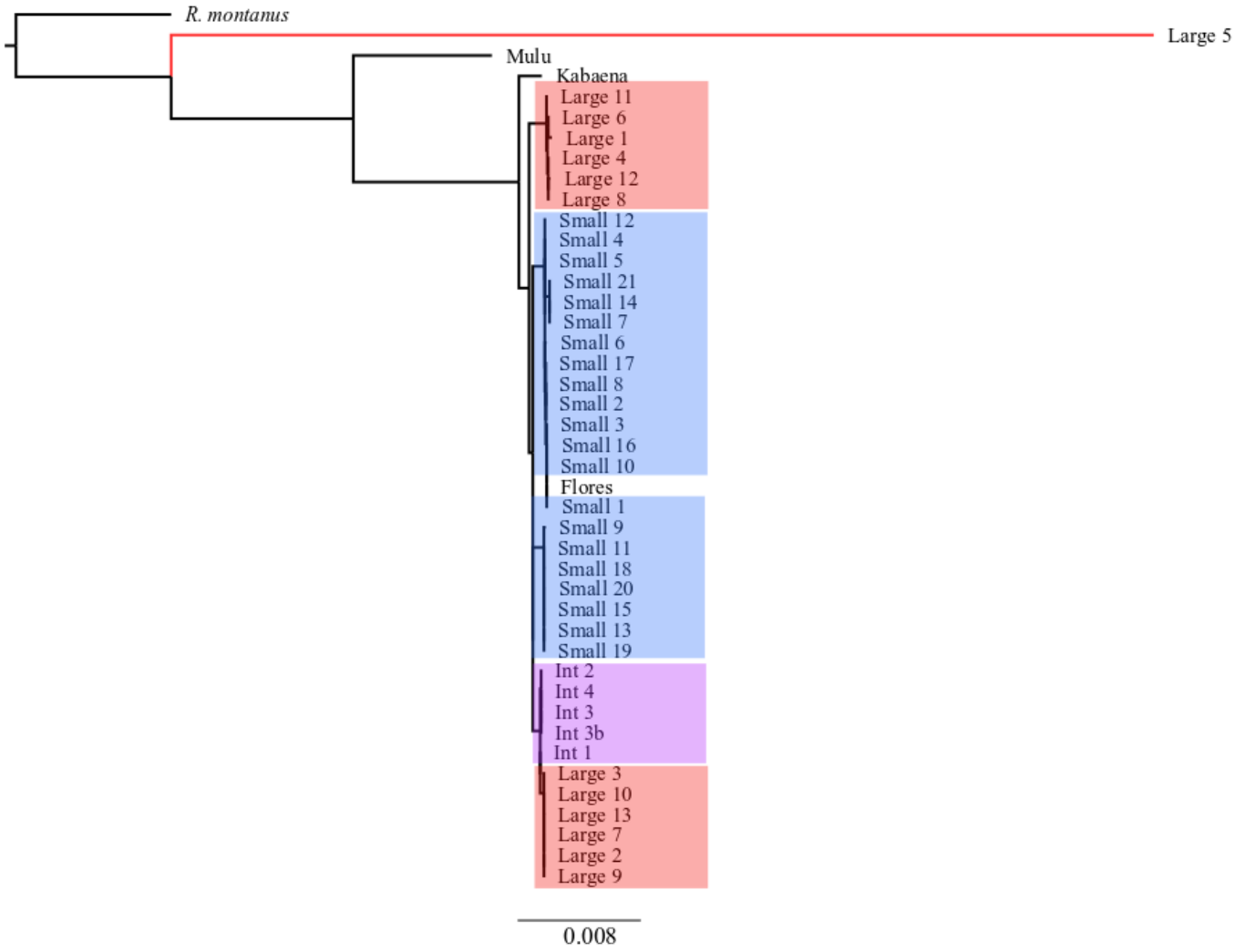


Figure 3.3: Cladogram representation of the neighbour-joining tree of the control region data from *R. philippinensis* as inferred from a reference-guided alignment of my new sequences using *R. macrotis* (489 base pairs). Control region sequences taken from Kingston and Rossiter (2004) (labelled 2004, followed by morph and ID) and Cooper, Reardon and Skilins (1998) (labelled Aus, followed by morph and ID) have also been included in tree construction. Highlighted regions indicate Buton Island *R. philippinensis*; the green highlighted region contains the sequences generated for this study, while the blue highlighted region shows sequences from the Kingston and Rossiter (2004) paper.

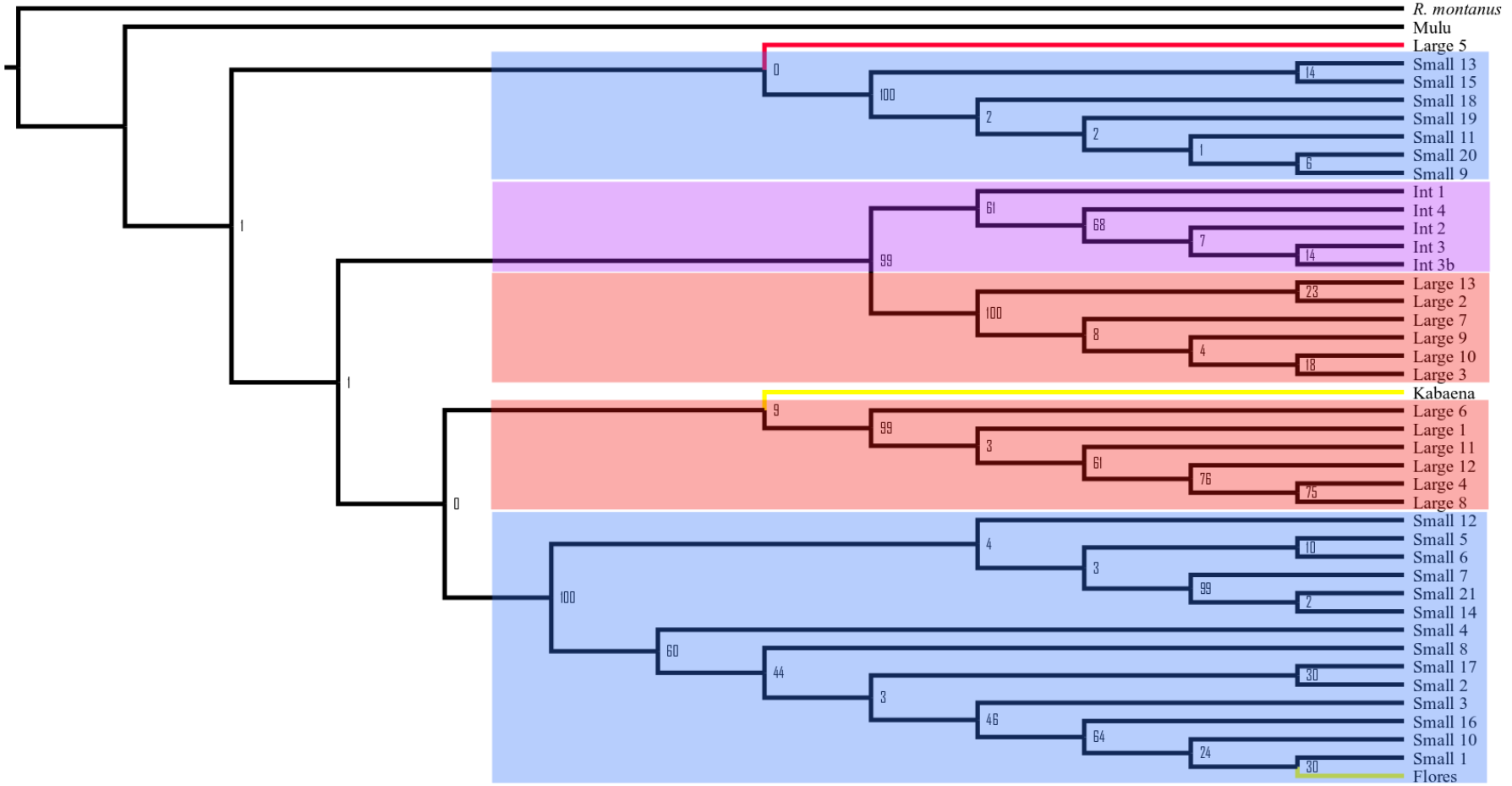

R. montanus
Large 5
Mulu
Kabaena
Large 11
Large 6
Large 1
Large 4
Large 12
Large 8
Small 12
Small 4
Small 5
Small 21
Small 14
Small 7
Small 6
Small 17
Small 8
Small 2
Small 3
Small 16
Small 10
Flores
Small 1
Small 9
Small 11
Small 18
Small 20
Small 15
Small 13
Small 19
Int 2
Int 4
Int 3
Int 3b
Int 1
Large 3
Large 10
Large 13
Large 7
Large 2
Large 9
100
100
85
83
99
56
27
67
2
100
16
0
30
99
96
0
45
1
47
61
54
47
44
100
1
2
11
94
97
15
99
65
61
95
38
98
92
95
95
98

a)

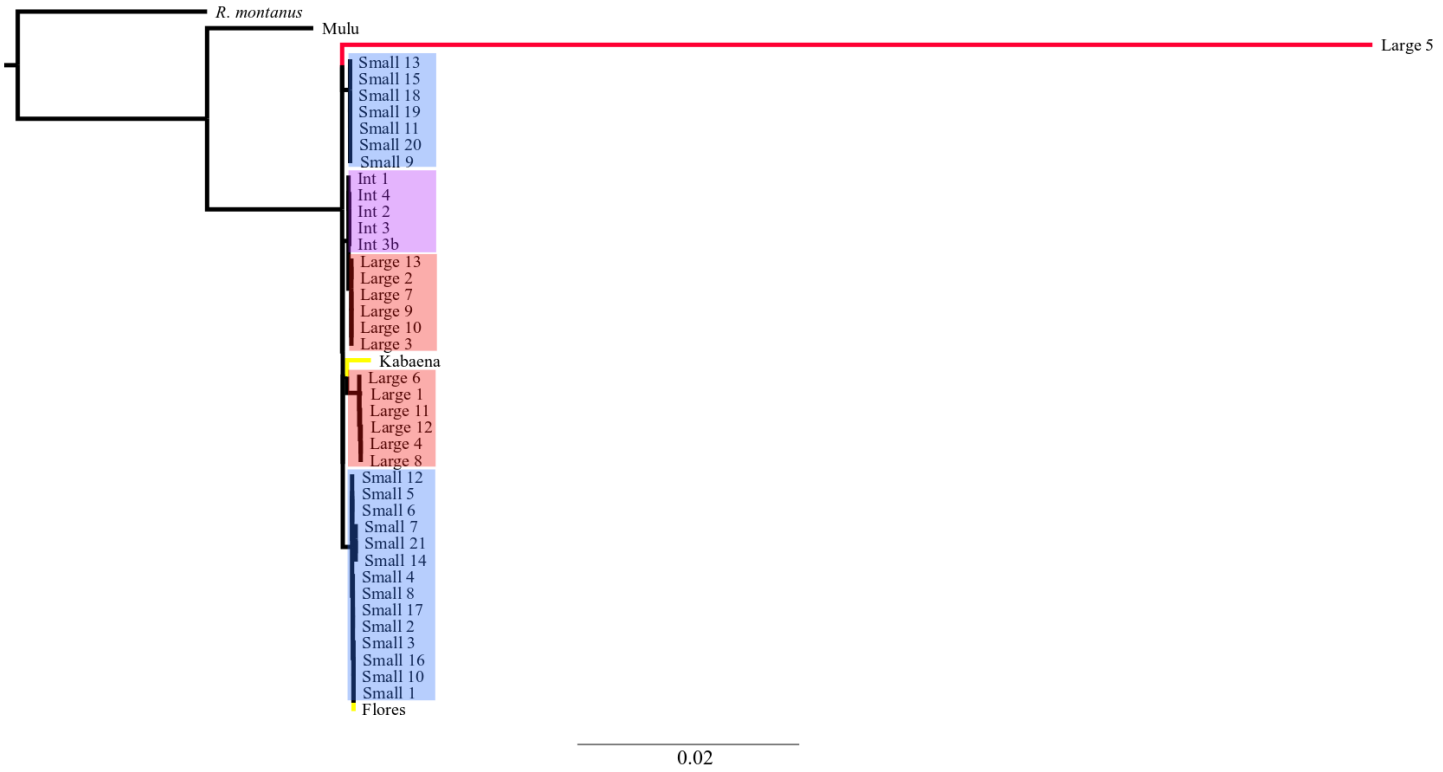


b)

Figure 3.4: Maximum likelihood tree of the full mitochondrial alignment (17,091 base pairs) for all *R. philippinensis* successfully sequenced. Red highlights indicate clades of Large Buton morphs, blue highlights clades of Small Buton morphs and purple highlights clades of Intermediation Buton morphs. Branches falling outside of their clade are highlighted in the appropriate colour: Large 5 in red, Intermediate 4 in purple and the outgroup Flores bat in yellow. A) Cladogram representation. Node labels represent bootstrap support for the node. B) Phylogram representation of the same tree, showing branch lengths.

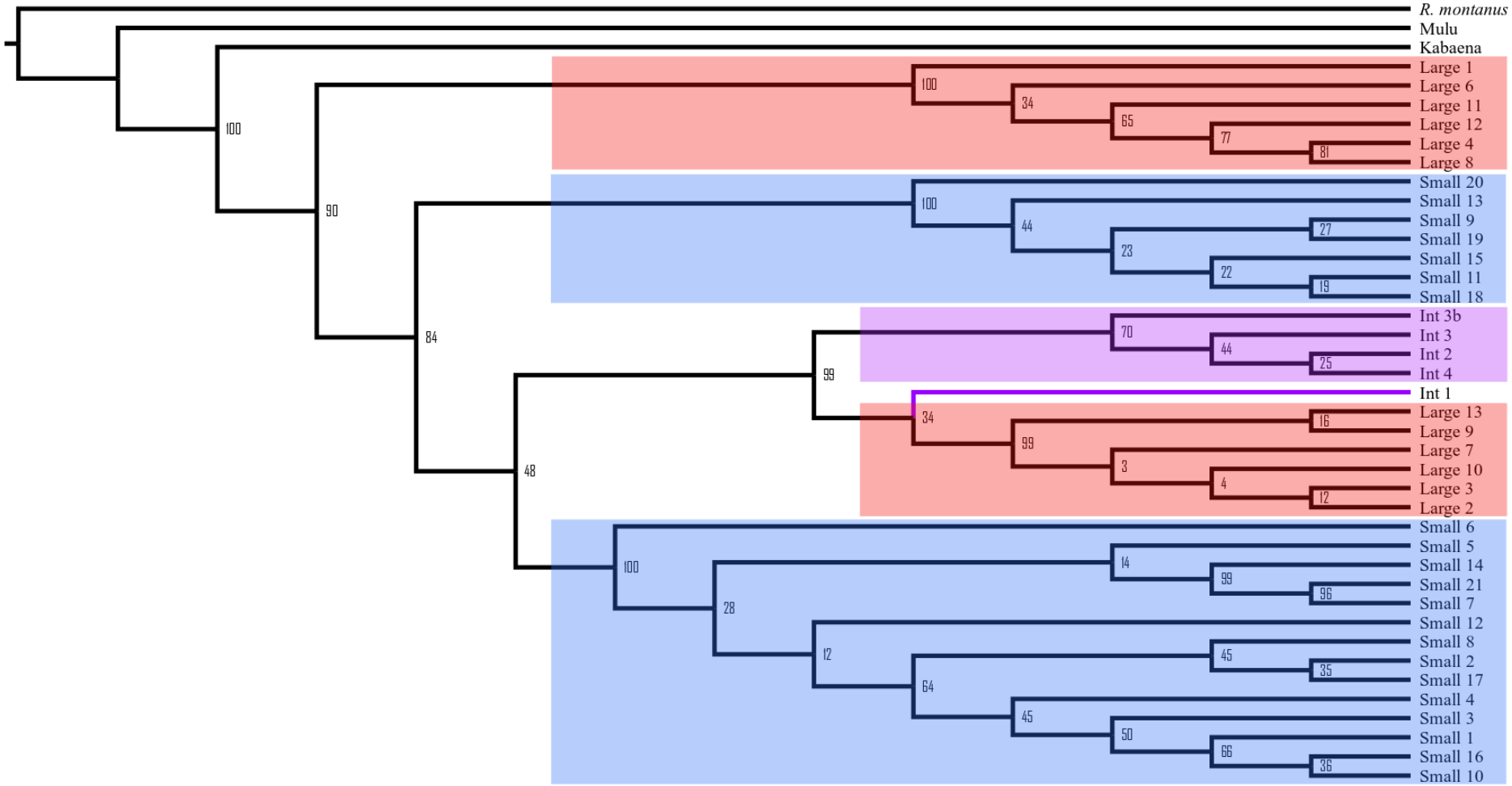

R. montanus
Mulu
Large 5
Small 13
Small 15
Small 18
Small 19
Small 11
Small 20
Small 9
Int 1
Int 4
Int 2
Int 3
Int 3b
Large 13
Large 2
Large 7
Large 9
Large 10
Large 3
Kabaena
Large 6
Large 1
Large 11
Large 12
Large 4
Large 8
Small 12
Small 5
Small 6
Small 7
Small 21
Small 14
Small 4
Small 8
Small 17
Small 2
Small 3
Small 16
Small 10
Small 1
Flores
0
100
2
2
1
6
14
1
99
61
68
7
14
100
8
4
18
23
1
0
9
99
3
61
76
75
100
4
3
99
2
10
60
44
3
46
64
24
30
30

a)

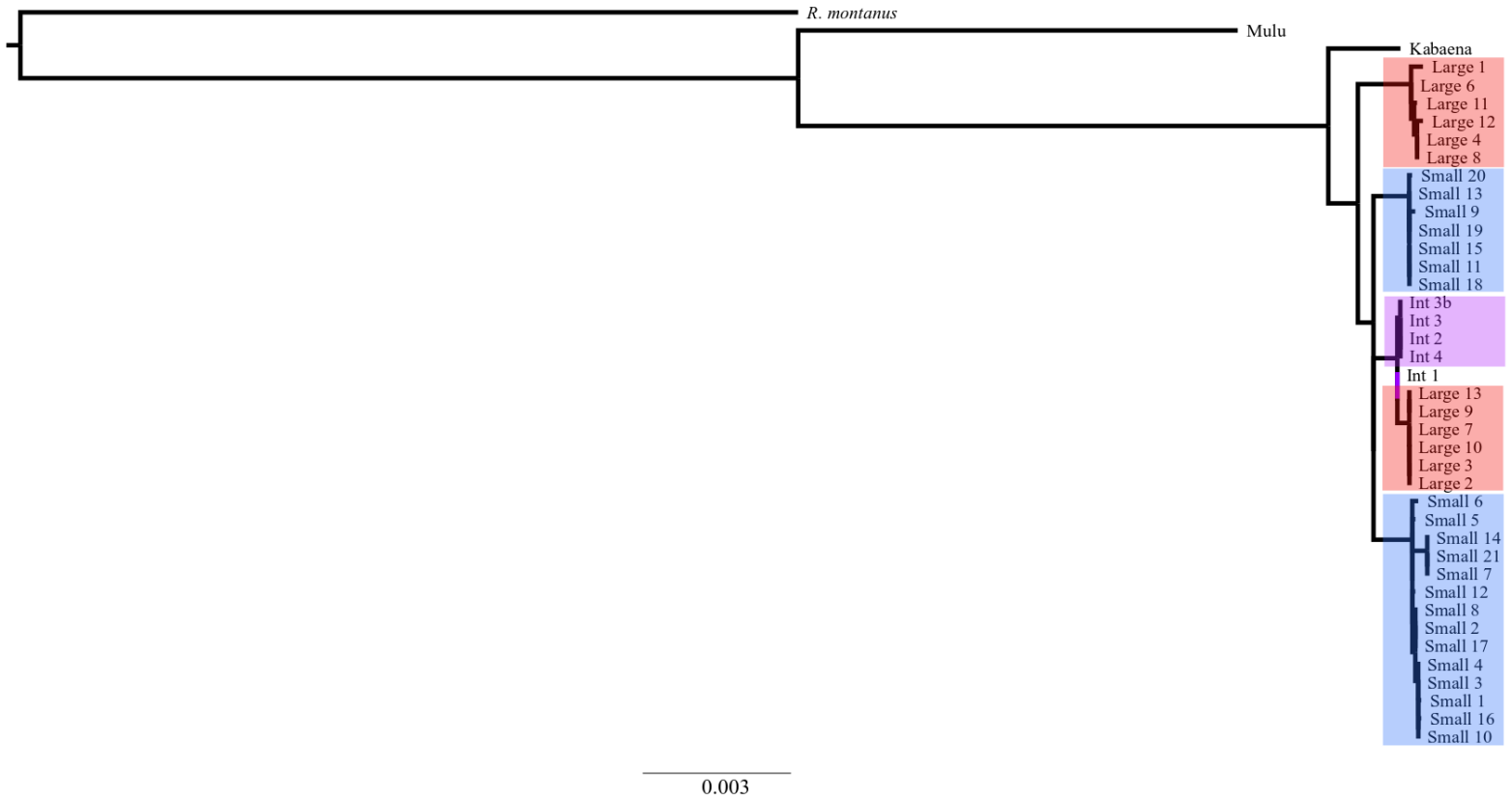


b)

Figure 3.5: Maximum likelihood tree of the mitochondrial alignment for all *R. philippinensis* successfully sequenced after the number of repeats has been standardised to six (16,706 base pairs). Red highlights indicate clades of Large Buton morphs, blue highlights clades of Small Buton morphs and purple highlights clades of Intermediation Buton morphs. Branches falling outside of their clade are highlighted in the appropriate colour: Large 5 in red and the outgroup Flores bat in yellow. A) Cladogram representation. Node labels represent bootstrap support for the node. B) Phylogram representation of the same tree, showing branch lengths.

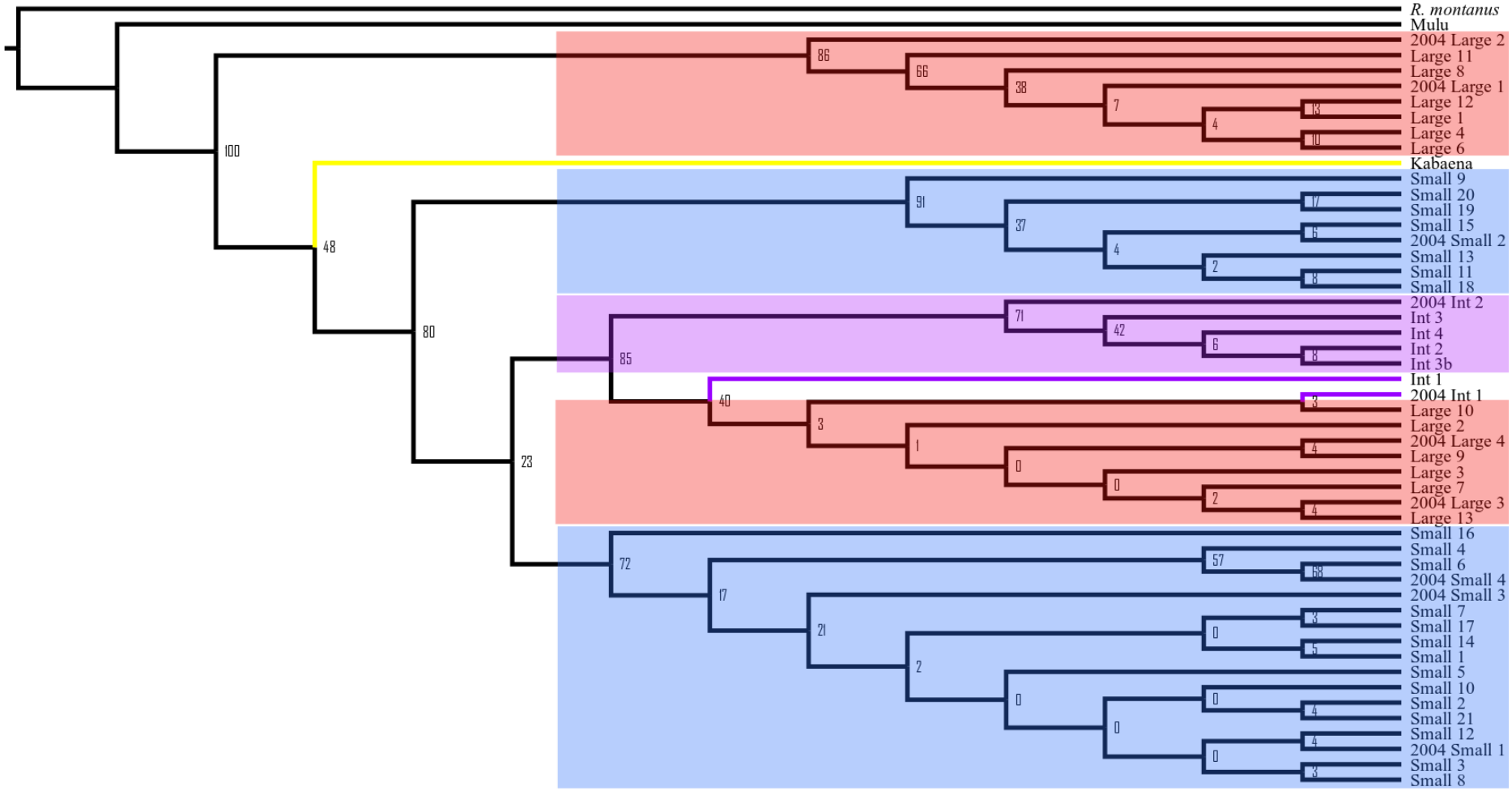

R. montanus
Mulu
Kabaena
Large 1
Large 6
Large 11
Large 12
Large 4
Large 8
Small 20
Small 13
Small 9
Small 19
Small 15
Small 11
Small 18
Int 3b
Int 3
Int 2
Int 4
Int 1
Large 13
Large 9
Large 7
Large 10
Large 3
Large 2
Small 6
Small 5
Small 14
Small 21
Small 7
Small 12
Small 8
Small 2
Small 17
Small 4
Small 3
Small 1
Small 16
Small 10
100
34
65
77
81
100
44
27
23
22
19
70
44
25
99
34
16
99
3
4
12
48
100
14
99
96
28
12
45
35
64
45
50
66
36
90
100
84

a)

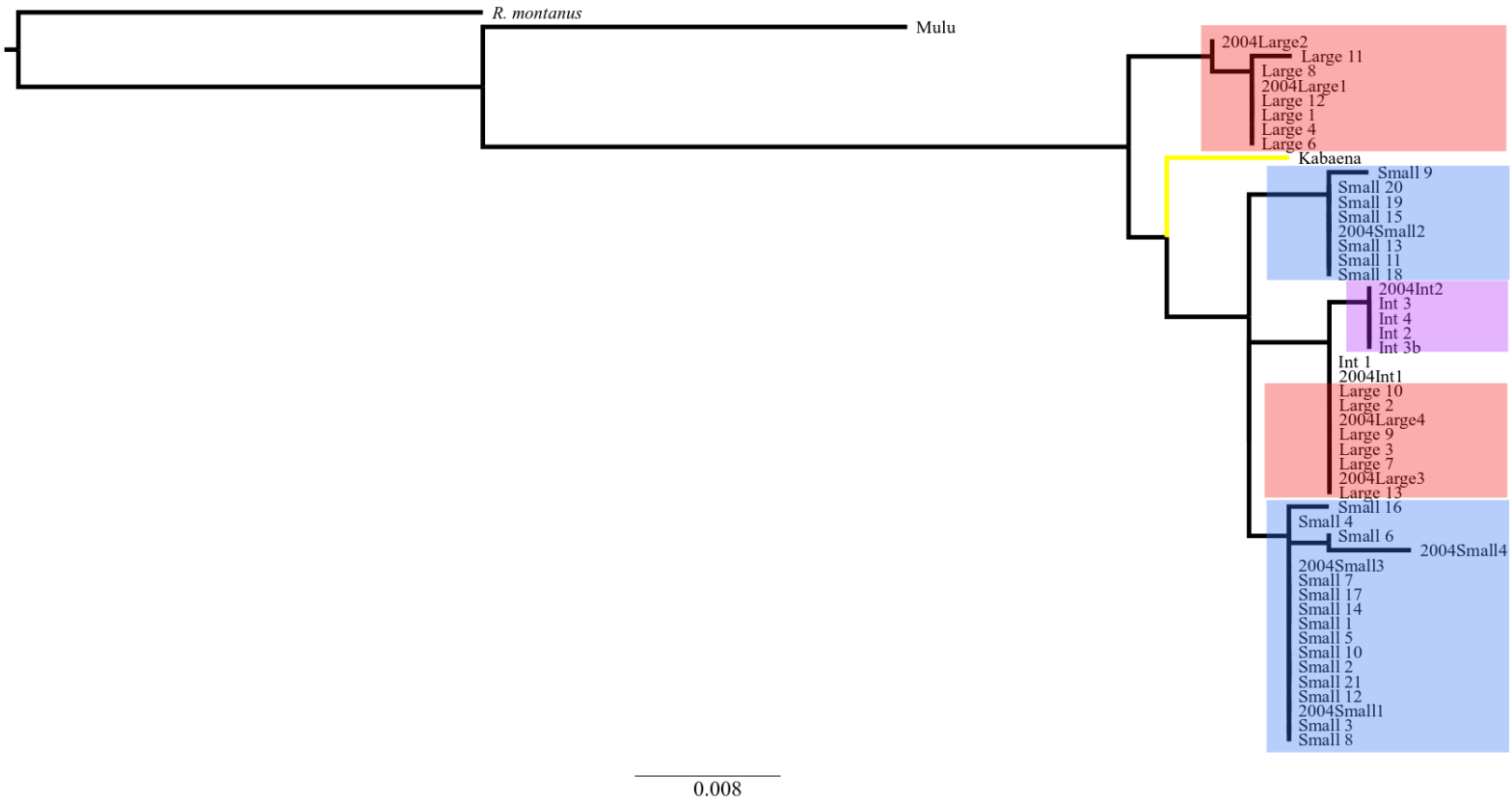


b)

Figure 3.6: Maximum likelihood tree of the *R. philippinensis* mitochondrial alignments after the number of repeats has been standardised to six (16,706 base pairs). The Flores morph and Large 5 have been removed from the data set. Red highlights indicate clades of Large Buton morphs, blue highlights clades of Small Buton morphs and purple highlights clades of Intermediate Buton morphs. A) Cladogram representation. Node labels represent bootstrap support for the node. B) Phylogram representation, with branch legnths.


R. montanus
Mulu
2004 Large 2
Large 11
Large 8
2004 Large 1
Large 12
Large 1
Large 4
Large 6
Kabaena
Small 9
Small 20
Small 19
Small 15
2004 Small 2
Small 13
Small 11
Small 18
2004 Int 2
Int 3
Int 4
Int 2
Int 3b
Int 1
2004 Int 1
Large 10
Large 2
2004 Large 4
Large 9
Large 3
Large 7
2004 Large 3
Large 13
Small 16
Small 4
Small 6
2004 Small 4
2004 Small 3
Small 7
Small 17
Small 14
Small 1
Small 5
Small 10
Small 2
Small 21
Small 12
2004 Small 1
Small 3
Small 8
100
48
80
23
85
40
72
17
21
2
86
66
38
7
4
91
37
4
2
71
42
6
3
1
0
0
2
57
68
3
0
5
0
0
0
4
4
3

a)



b)

Figure 3.7: Maximum likelihood tree of the alignment of *R. philippinensis* control region sequences (460 base pairs). Sequences prefixed with '2004' were included in Kinston and Rossiter (2004) and were downloaded from GenBank. All other sequences were generated for this study. Clades of Large morphs are highlighted in red, clades of Small morphs in blue and clades of Intermediate morphs in purple. A) Cladogram representation. Node labels represent bootstrap support for the node. B) Phylogram representation, with branch legnths.

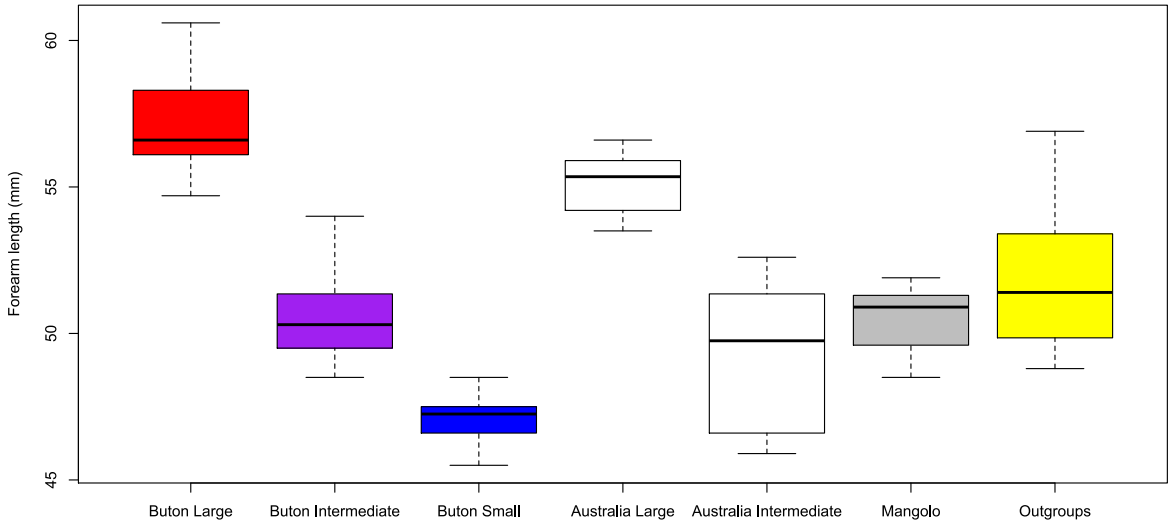


Figure 3.8: Boxplots showing forearm measurements of all *R. philippinensis* caught in the course of this research. Buton Large N = 14; Buton Intermediate N = 7; Buton Small N = 26; Australia Large N = 8; Australia Intermediate N = 8; Mangolo N = 17; Outgroups N = 11.

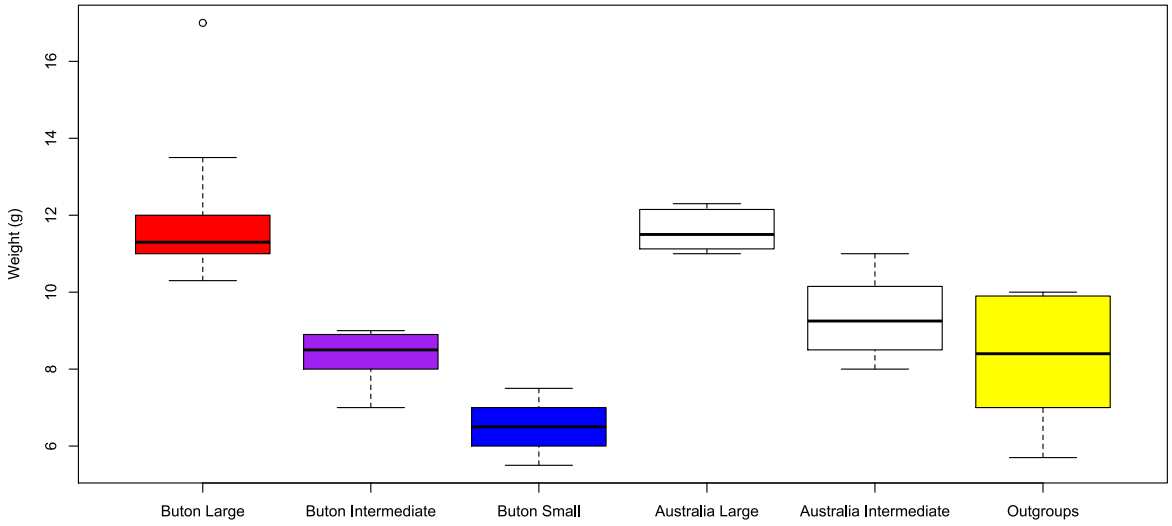


Figure 3.9: Boxplots showing weights of all *R. philippinensis* caught in the course of this research. Buton Large N = 14; Buton Intermediate N = 7; Buton Small N = 26; Australia Large N = 8; Australia Intermediate N = 8; Outgroups N = 6.

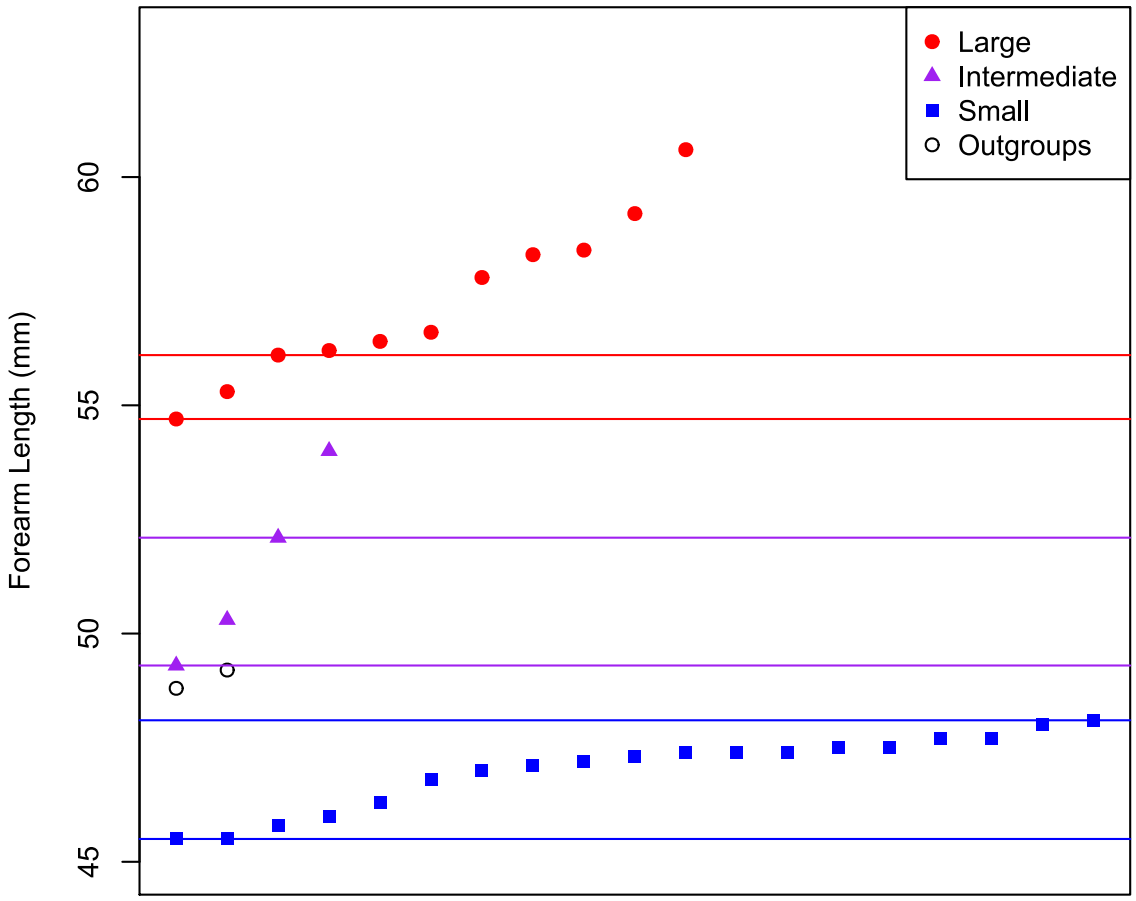


Figure 3.10: Plot of forearm lengths for individuals for which I have mitochondrial data. Coloured lines indicate maximum and minimum forearm lengths described for each morph from individuals assigned to a morph based on their echolocation call frequencies.

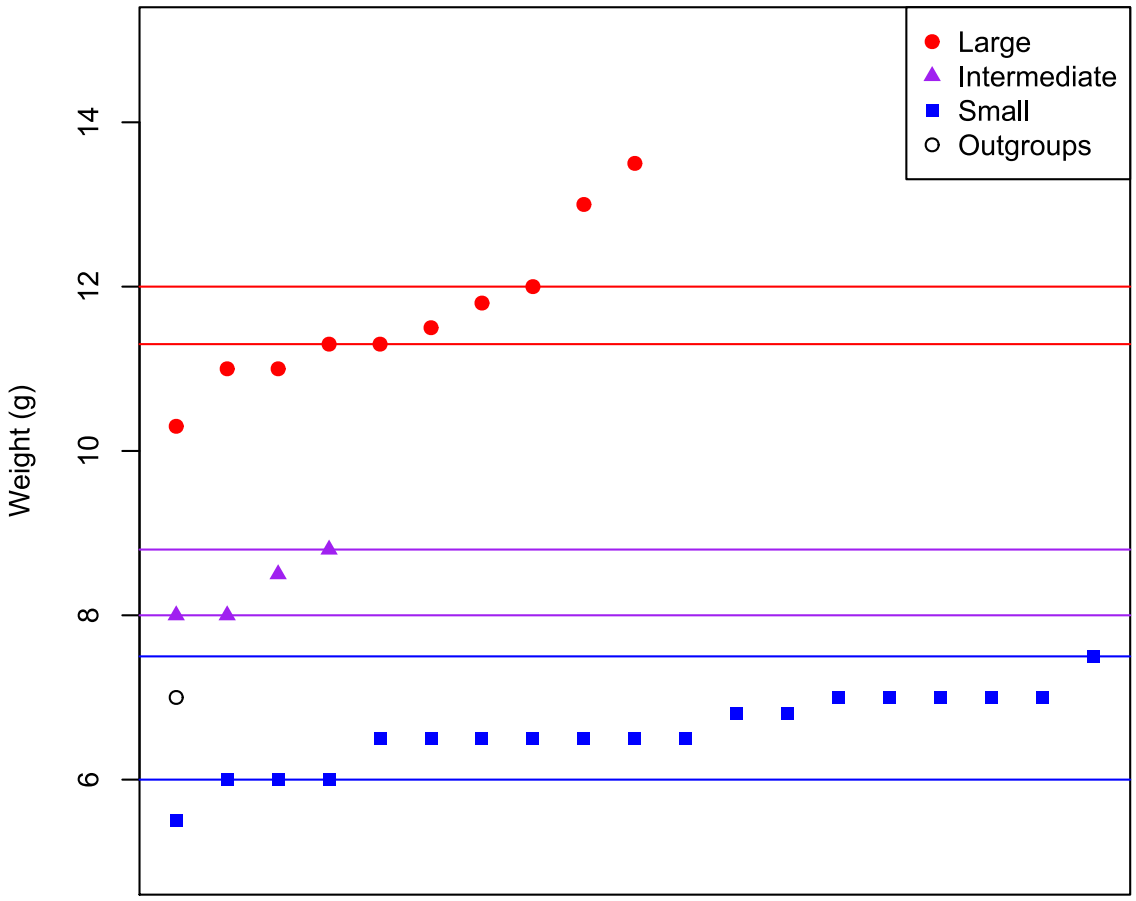


Figure 3.11: Plot of weights for individuals for which I have mitochondrial data. Coloured lines indicate maximum and minimum weights described for each morph based on individuals assigned to a morph by their echolocation call frequencies.

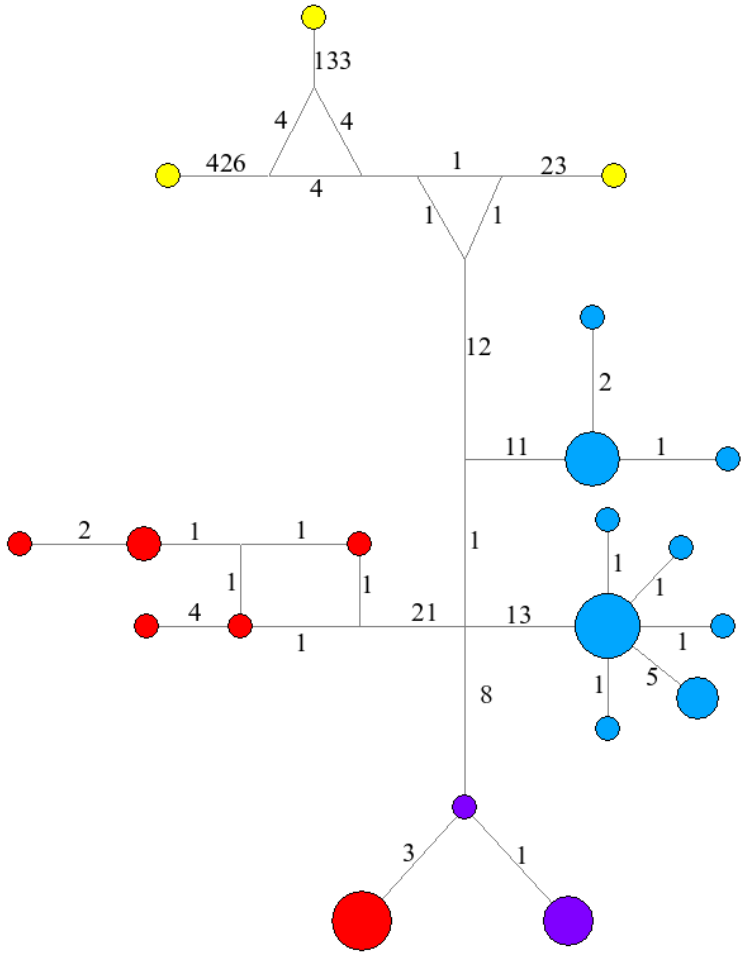
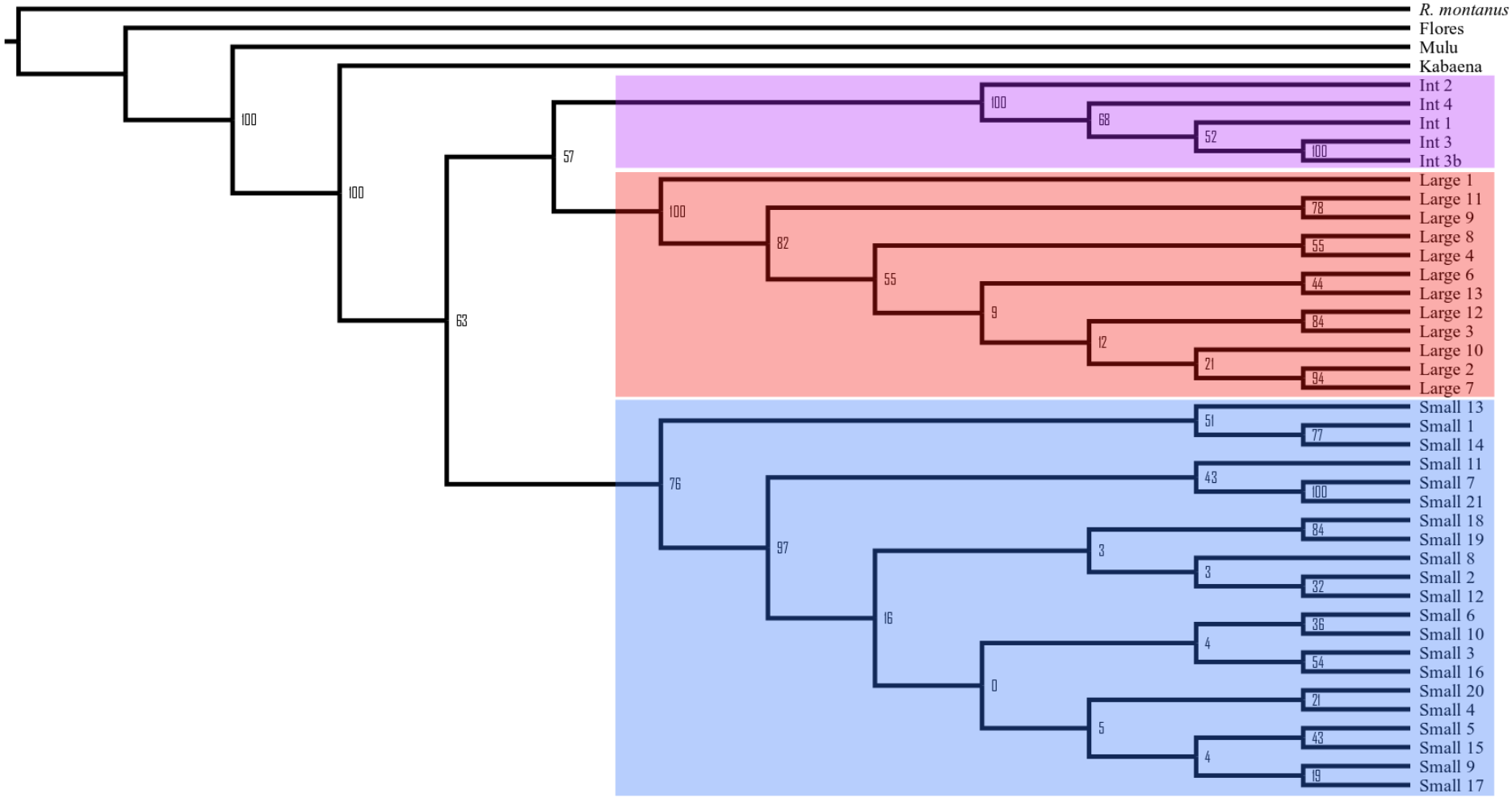
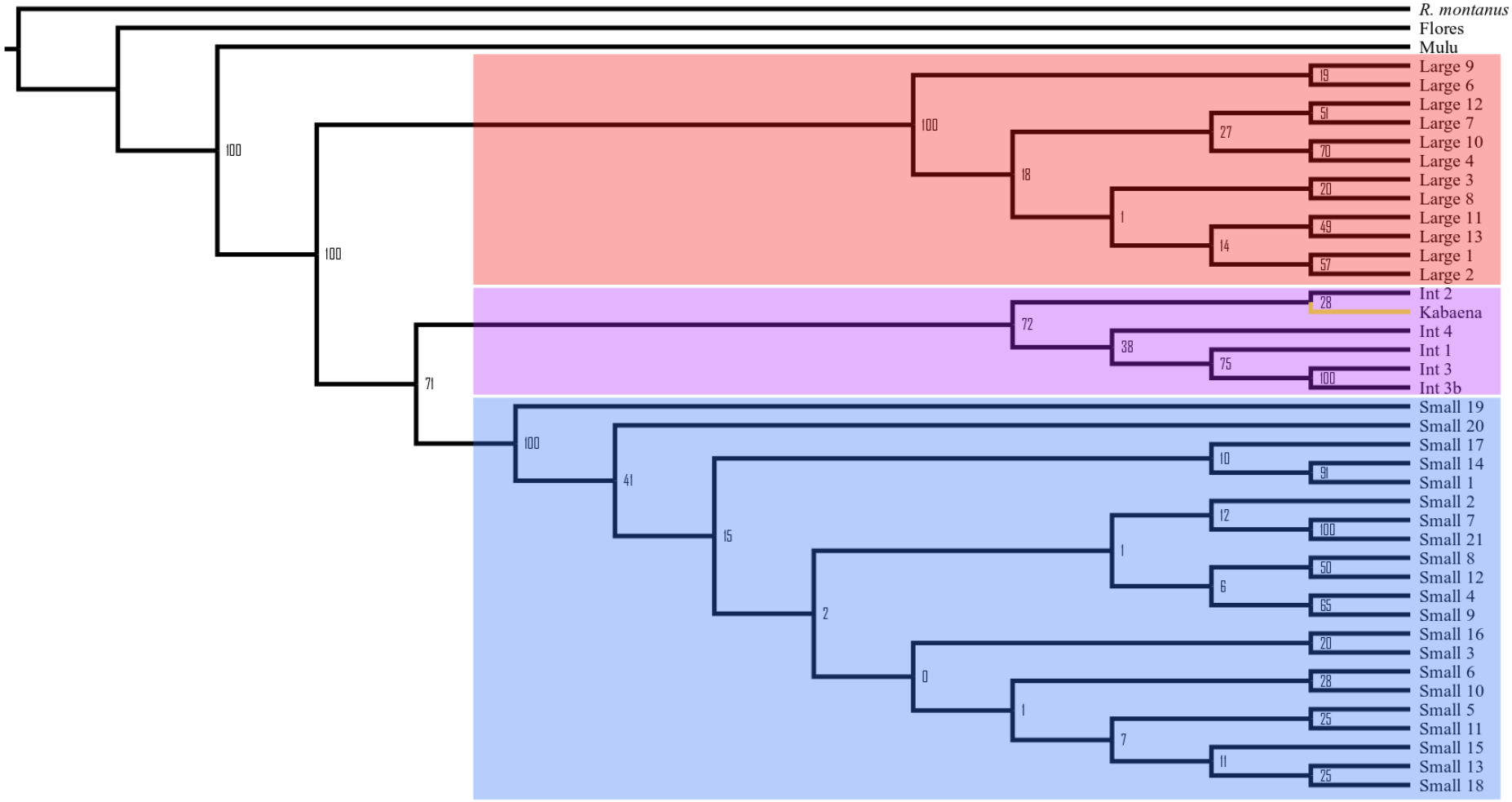


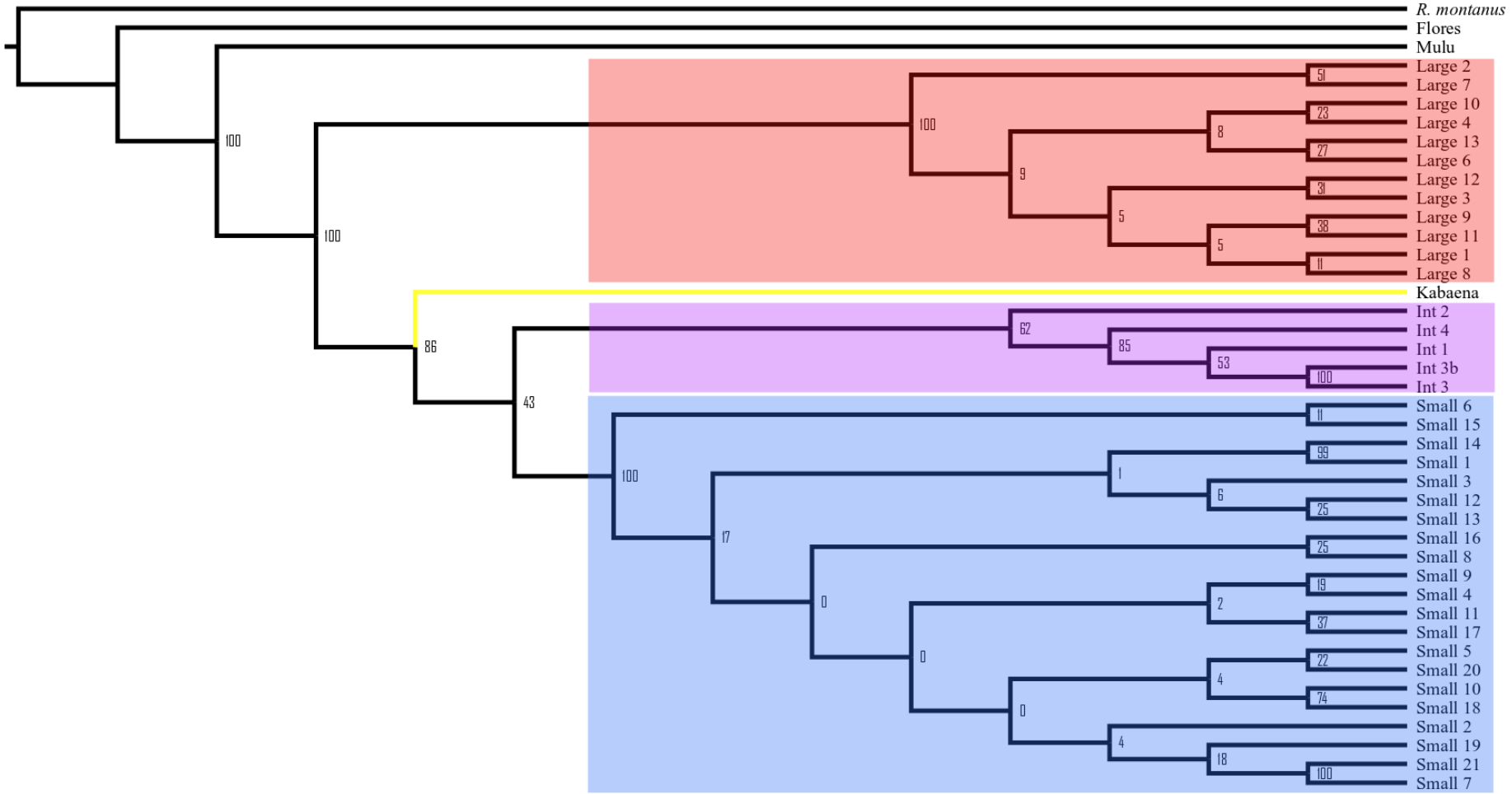
Figure 3.12: Mitochondrial haplotype network using the same data set as Figure 3.6. Nodes are pie charts and proportional to the number of individuals included; red represents large Buton morphs, blue represents small Buton morphs and purple indicates Intermediate Buton morphs. Outgroups are represented in yellow. Numbers on branches indicate the number of mutations that happened along that branch.


R. montanus
Flores
Mulu
Kabaena
Int 2
Int 4
Int 1
Int 3
Int 3b
Large 1
Large 11
Large 9
Large 8
Large 4
Large 6
Large 13
Large 12
Large 3
Large 10
Large 2
Large 7
Small 13
Small 1
Small 14
Small 11
Small 7
Small 21
Small 18
Small 19
Small 8
Small 2
Small 12
Small 6
Small 10
Small 3
Small 16
Small 20
Small 4
Small 5
Small 15
Small 9
Small 17
100
100
63
57
100
68
52
100
100
82
55
9
12
21
94
78
55
44
84
51
77
76
43
100
97
84
3
3
32
16
36
4
54
0
21
5
43
4
19

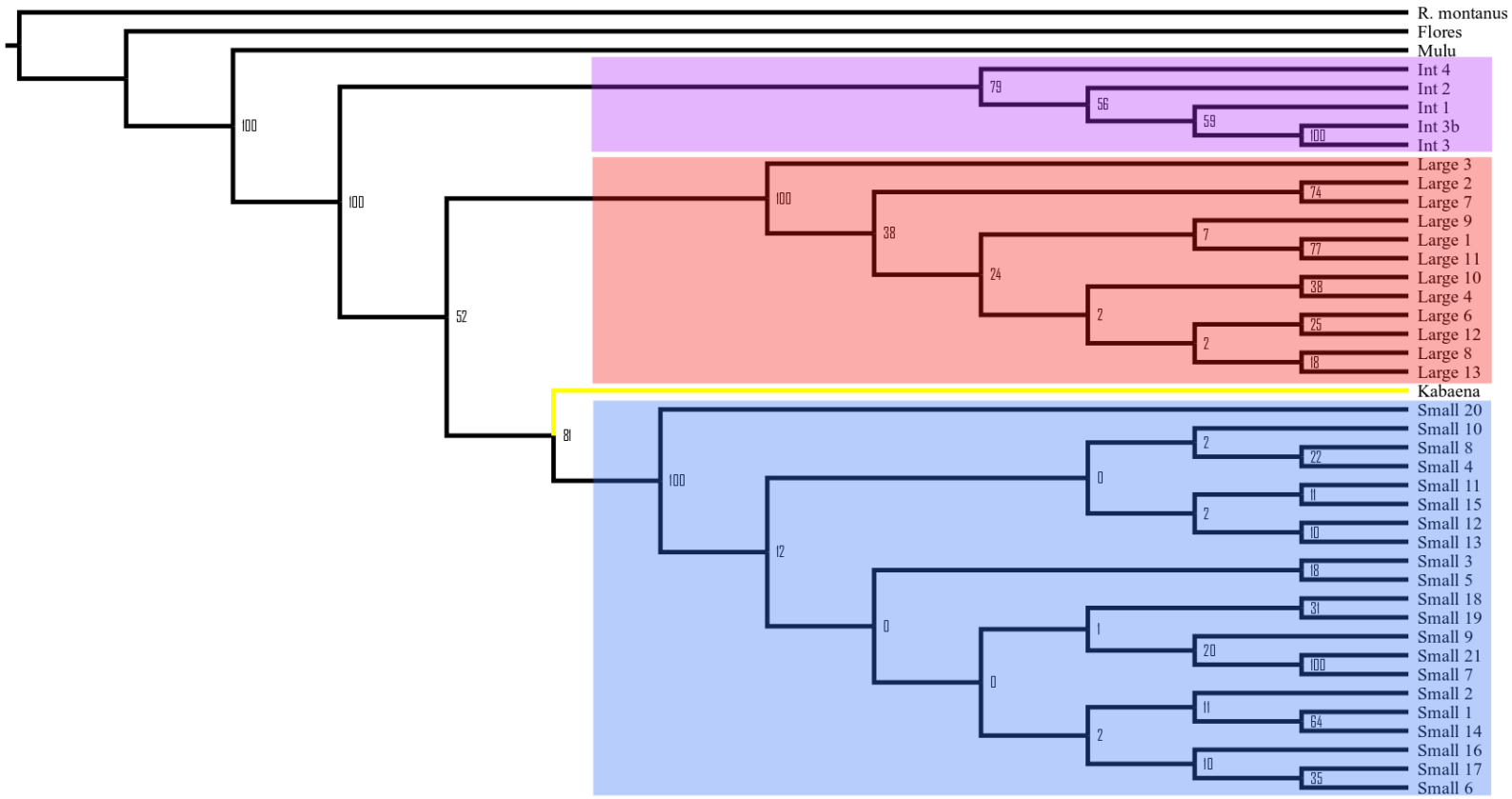
a)


R. montanus
Flores
Mulu
Large 9
Large 6
Large 12
Large 7
Large 10
Large 4
Large 3
Large 8
Large 11
Large 13
Large 1
Large 2
Int 2
Kabaena
Int 4
Int 1
Int 3
Int 3b
Small 19
Small 20
Small 17
Small 14
Small 1
Small 2
Small 7
Small 21
Small 8
Small 12
Small 4
Small 9
Small 16
Small 3
Small 6
Small 10
Small 5
Small 11
Small 15
Small 13
Small 18
100
100
71
100
27
70
18
20
1
49
14
57
72
28
38
75
100
100
41
15
10
91
12
100
1
50
6
65
2
20
0
28
1
25
7
11
25

b)

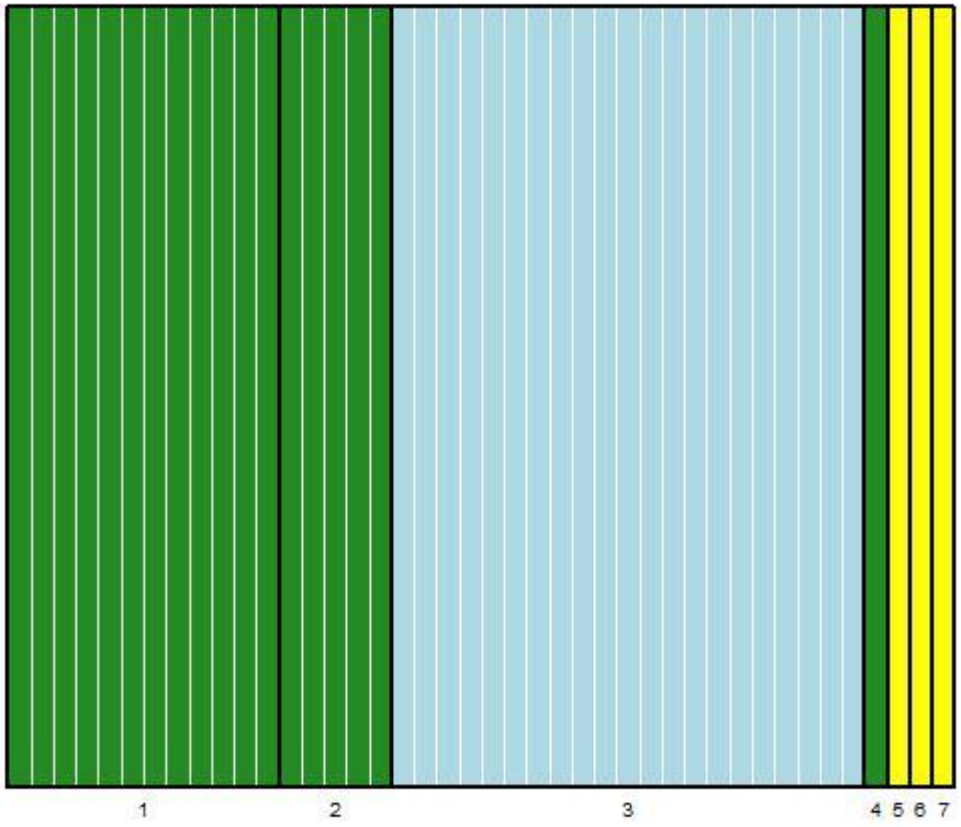

R. montanus
Flores
Mulu
Large 2
Large 7
Large 10
Large 4
Large 13
Large 6
Large 12
Large 3
Large 9
Large 11
Large 1
Large 8
Kabaena
Int 2
Int 4
Int 1
Int 3b
Int 3
Small 6
Small 15
Small 14
Small 1
Small 3
Small 12
Small 13
Small 16
Small 8
Small 9
Small 4
Small 11
Small 17
Small 5
Small 20
Small 10
Small 18
Small 2
Small 19
Small 21
Small 7
100
100
100
9
8
5
5
51
23
27
31
38
11
86
43
62
85
53
100
100
17
0
0
0
1
6
25
25
19
2
37
22
4
74
4
18
100
11
99

c)

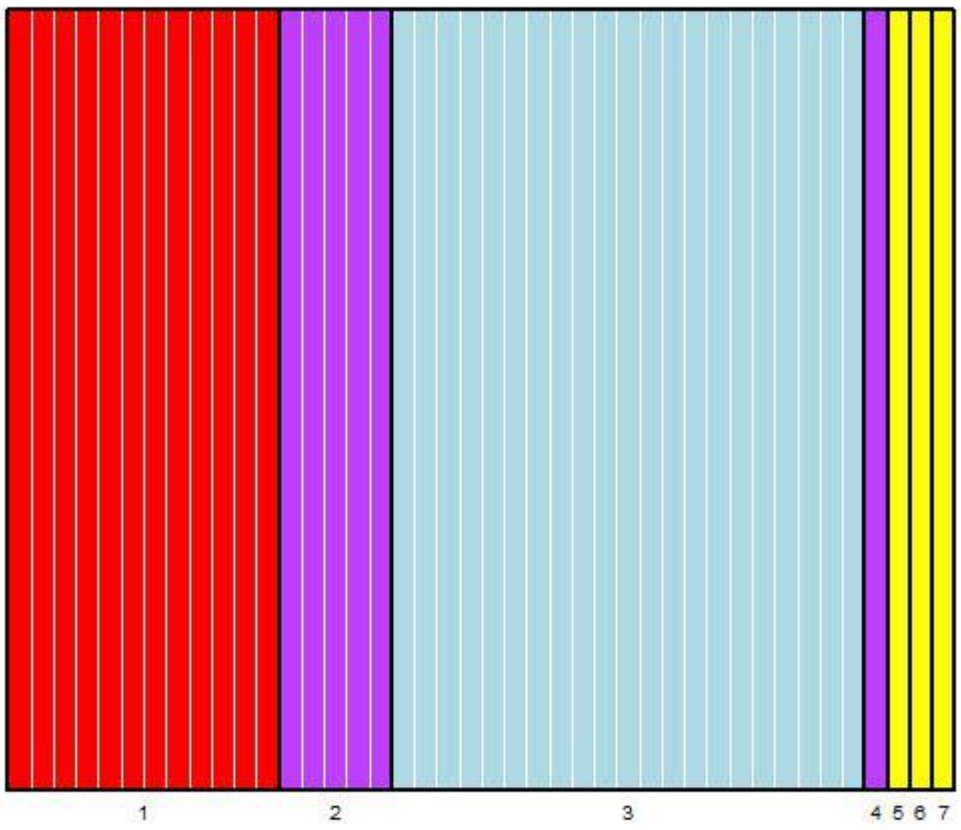


d)

Figure 3.13 a-d: Cladograms constructed from 10,000 unlinked genomic SNPs from *R. philippinensis*, in order of frequency of recovery. Large morph clade is highlighted in red, Small morph clade in blue and Intermediate morph clade in purple. Kabaena individual is coloured yellow when it fails to sit with the other outgroup samples. Topology a was the most commonly recovered, with the other topologies representing less common rearrangements.



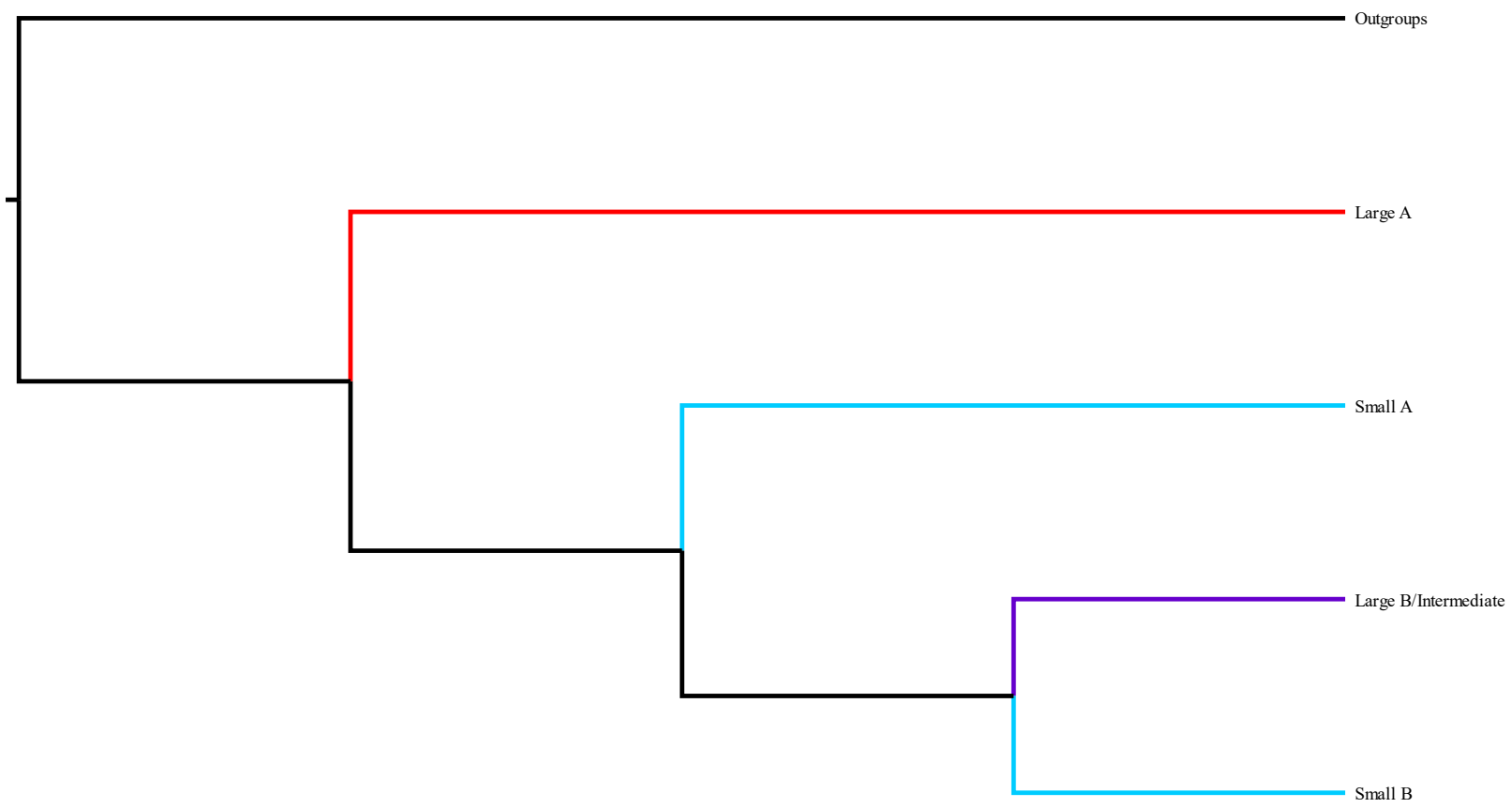
a)



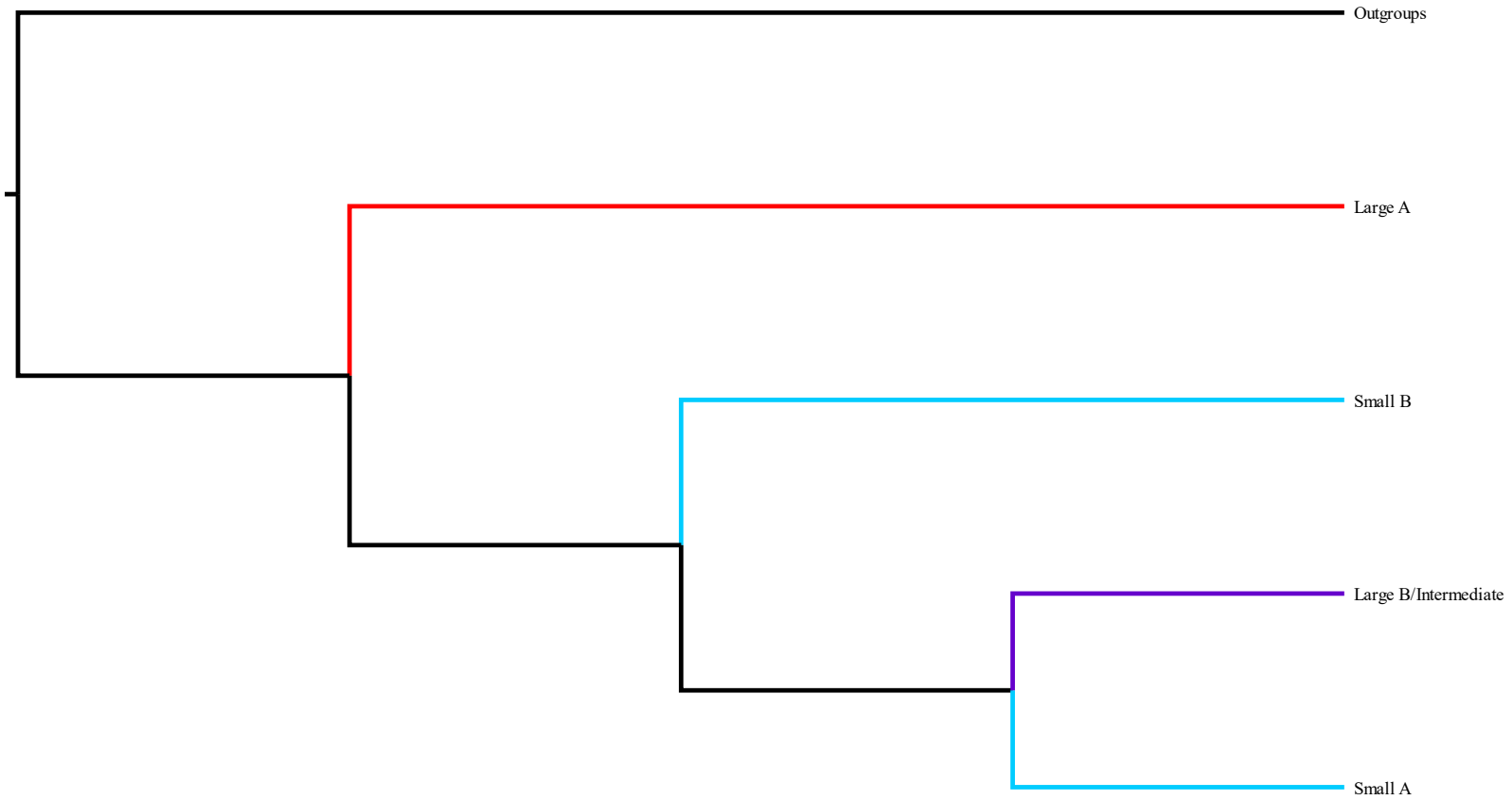
b)

Figure 3.14 a-b: Q-matrix plot showing population structure within *R. philippinensis* based on 10,000 random SNPs. Light bars separate individuals; dark bars separate *a priori* population assignments, indicated by numbers on the x-axis (1 – Large morph; 2 – Intermediate morph; 3 – Small morph; 4 – Kabaena individual; 5 – Flores individual; 6 – Borneo individual; 7 – *R. montanus* individual). Bar colours indicate clusters assigned by MavericK. A) $K = 3$; b) $K = 4$.

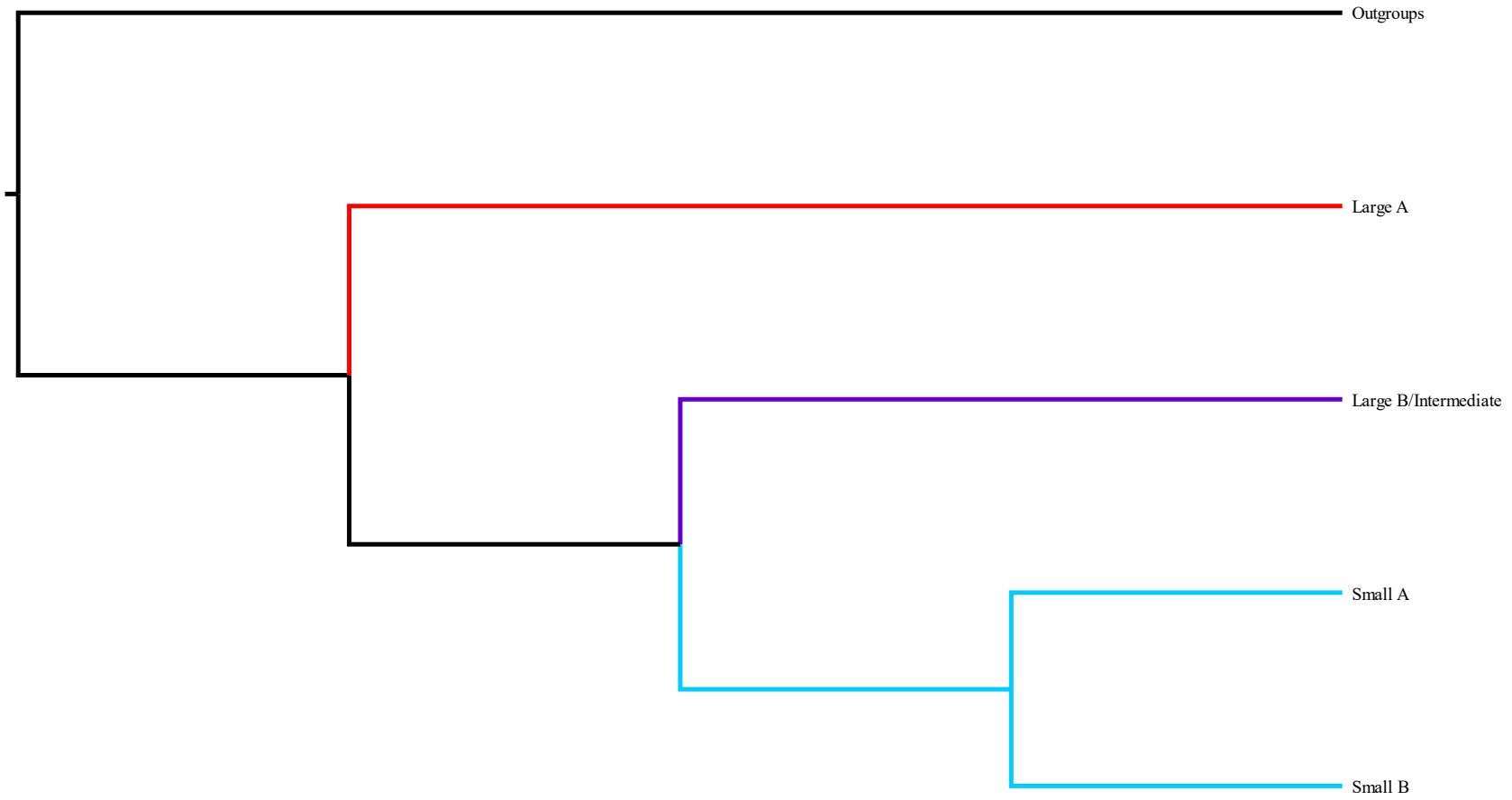
Figure 3.15: Q-matrix plot showing population structure within *R. philippinensis* based on 10,000 random SNPs at *K*=3, where admixture is allowed. Light bars separate individuals; dark bars separate *a priori* population assignments, indicated by numbers on the x-axis (1 – Large morph; 2 – Intermediate morph; 3 – Small morph; 4 – Kabaena individual; 5 – Flores individual; 6 – Borneo individual; 7 – *R. montanus* individual). Bar colours indicate clusters assigned by MavericK.



a)



b)



c)

Figure 3.16: Top three topologies of Buton morphs supported by model A11 provided with the maximum likelihood mitochondrial phylogeny as a guide tree. *R. montanus* and Borneo and Kabaena *R. philippinensis* grouped together as outgroups for the purposes of visualisation. Red branches: large morphs; blue branches: small morphs; purple branches: combined group of Large and Intermediate morphs.

CHAPTER FOUR

Islands of divergence suggest speciation with gene flow among three morphs of *Rhinolophus philippinensis*

Abstract

The genic theory of speciation predicts heterogeneous patterns of genomic differentiation between populations that are speciating in the face of gene flow due to uneven rates of recombination occurring across the genome. In this scenario, most of the genome is expected to exhibit low differentiation alongside stretches of higher divergence around loci that are subject to divergent selection. Genome scans for these so-called 'islands of divergence' can thus be used to demonstrate that putative incipient species are not fully reproductively isolated, or which have only become so recently. Moreover, they can also highlight the genomic regions involved in driving or maintaining the divergence between the populations, allowing us to identify putative speciation genes. Three sympatric acoustic morphs of *R. philippinensis* found on Buton Island, Sulawesi, represent a putative ecological speciation event. I have previously established that these morphs are genetically distinct from one another. To test whether patterns of genetic distinctiveness among morphs are consistent with a genic model of speciation, I performed pairwise genome-wide F_{ST} scans. I also investigated for divergence between the morphs in genes associated with body size and hearing, key characters differentiating the morphs which are also likely to play a role in the development of pre-zygotic reproductive isolation by affecting mate recognition and mate choice. My results show a heterogeneous mosaic of

genomic differentiation existing between the three *R. philippinensis* morphs, with islands of high differentiation between populations showing an overrepresentation of genes associated with body size, and a strong representation of genes associated with hearing that have previously been implicated in the evolution of echolocation or detected as evolving under positive selection in echolocating taxa. These results add support to previous suggestions that these morphs are in the process of undergoing ecological speciation, driven by phenotypic traits that play roles in both ecology and in mate recognition.

Introduction

Under traditional biological species concepts, populations need to be fully reproductively isolated from one another in order to qualify as good species. In the early stages of the 21st century, however, opposition to this view was fully conceptualised in the genic view of the process of speciation (Wu 2001). This proposed the concept of 'speciation genes' that directly affect the process of differential adaptation. These genes would account for only a very small part of the genome, and as long as this part of the genome remained diverged then populations would be able to retain their distinctiveness and continue to differentiate even if there was mixing in the rest of the genome (Wu 2001). The result of this process would be a genetic mosaic of differentiation between the diverging populations. Such instances of these genetic mosaics in nature have been subsequently described in systems suspected to be in the early stages of sympatric divergence, such as fire-bellied toads *Bombina bombina* and *B. variegata* (Vines et al 2003), or, in cases of differential introgression following secondary contact, Mexican red oaks *Quercus affinis* and *Q. laurina* (González-Rodríguez et al 2004) and tiger salamanders *Ambystoma tigrinum mavortium* and *A. californiese* (Fitzpatrick and Shaffer 2004).

Methods first developed over two decades ago showed how calculating F_{ST} at multiple loci across the genome can reveal these islands of differentiation and may thus be a useful first step in identifying candidate genes evolving under selection (Beaumont 2005). These methods have been shown to be robust to vagaries of demographic history and through simulations have been demonstrated to be capable of accurately identifying loci subject to adaptive selection (Beaumont and Balding 2004). Coalescent simulations suggested that regions of elevated divergence found between different forms of *Anopheles gambiae* were not a result of neutral scenarios, supporting the conclusion that these 'speciation

islands' contained genes under adaptive selection that were responsible for reproductive isolation (Turner, Hahn and Nuzhdin 2005).

With recent advances in sequencing technology, it has now become tractable to examine genome-wide patterns of differentiation among multiple individuals from separate populations. Recent studies that have focused on investigating the genes found in diverged genomic islands have reported links to traits involved in reproductive isolation. For example, in two closely related species of nightingale *Luscinia megarhynchos* and *L. luscinia* that hybridise in a secondary contact zone, genomic islands of high differentiation were enriched for genes involved in the oocyte meiosis pathway (Mořkovský et al 2018). In another example, from the three spine stickleback (*Gasterosteus aculeatus* species complex), candidate genes involved in the visual perception of colour were present in moderately differentiated genomic islands between populations and appear to relate to variation in male throat colour between habitats (Marques et al 2017).

This latter example has also been proposed as a case of ecological speciation; male throat colour is a sexual signal under selection to maximise visibility in different light environments, while females also show colour preferences in a controlled light environment (Marques et al 2017, Feller et al 2016). Under ecological speciation, divergence between populations is typically considered to be driven by adaptation to different ecological conditions; this provides a framework to allow for divergence in the face of gene flow (Nosil 2012). This process can be facilitated if the trait subject to divergent natural selection also plays a role in mate choice or mate recognition (Servedio et al 2011). While genome-wide divergence scans can help to identify loci that are highly differentiated between populations and so likely to be driving their divergence this needs to be investigated in the early stages of ecological speciation (Marques et al 2017). Four stages of speciation-with-gene-flow have been proposed (Feder, Egan and Nosil 2012),

starting with direct selection on loci under high divergent selection pressure, moving through divergence hitchhiking and genomic hitchhiking and finishing with post-speciation divergence. In each of these stages, the proportion of the genome which is diverged between the incipient species increases, and thus signals of particular loci driving divergence become harder to detect. Additionally, multiple different pressures imposed by different agents of ecological and sexual selection are likely to have become involved by the later stages of a speciation event, making it difficult to discern which ones were acting at the initial population divergence (Maan and Seehausen 2011). To date, few studies have examined systems that represent potential early stages of speciation-with-gene-flow.

The three *Rhinolophus philippinensis* morphs found on Buton Island appear to be in the early stages of speciating, and low differentiation implies this is occurring in the face of ongoing gene flow (Kingston and Rossiter 2004). In the previous chapter I showed that in addition to the phenotypic differences between the morphs, there were genetic signals of population subdivision in both mitochondrial and genomic DNA and that there was discordance between trees based on nuclear and mitochondrial DNA. I have presented this system as a candidate for the study of ecological speciation, where divergence in echolocation call frequency, a trait closely linked to body size in bats (Jones 1999), has been favoured by natural selection to allow the exploitation of different niches.

Ultrasonic vocalisations as used by bats have been shown to carry information about the transmitter (Gillam and Brock Fenton 2016) and also play roles in social communication (Jones and Siemens 2011). Bats which use high duty-cycle echolocation, such as *R. philippinensis*, focus the main energy of their calls on a single frequency (Altringham 2011; Lazure and Fenton 2011) to which their cochlea is specifically tuned (Schuller and Pollack 1979; Davies, Maryanto and Rossiter 2013). This tuning means that divergence in echolocation call frequency may be potentially more likely to lead to reproductive

isolation in bats using this kind of echolocation, because different call frequencies will be less audible to heterospecifics (Kingston et al 2001, Servedio et al 2011). Moreover, these narrowband calls are thought to play a role in communication, and indeed, echolocation call frequency has been demonstrated to be involved in mate choice in the high duty-cycle echolocating bat, *Rhinolophus mehelyi* (Puechmaille et al 2014).

Relationships between genetic subdivision and echolocation call frequency have been repeatedly observed in bats using high duty-cycle echolocation (e.g. Chen, Jones and Rossiter 2009). Multiple explanations exist for the divergence of echolocation call frequencies among these bats. Partitioning of call frequency bands between Rhinolophidae species, either to reduce competition for resources or to allow for use of private communication channels, has long been proposed to occur (Heller and von Helversen 1989). Twelve syntopic tropical *Rhinolophus* and *Hipposideros* species in Malaysia were initially described as having call frequencies that were more evenly distributed over a frequency range between 40 and 200kHz than would be expected by chance (Heller and von Helversen 1989). While a later reanalysis of the same species assemblage that also included three further syntopic species did not find the same even partitioning of species by frequency (Kingston et al 2000), it did find that the species were non-randomly distributed in multivariate space with trait overdispersion between more similar species (Kingston et al 2000). Where species pairs were identified that used similar or the same echolocation call frequency, such as *Hipposideros larvatus* and *Rhinolophus refulgens*, they would differ greatly along other axes, such as body size.

Character displacement to avoid competition has been studied specifically by comparing island and peninsula populations of rhinolophids (Russo et al 2007). *Rhinolophus hipposideros* and *R. euryale* use higher and lower call frequencies respectively where their ranges overlap with *R. mehelyi* in Sardinia than they do where the latter is absent in peninsula Italy (Russo et al 2007). This strongly indicates a facility for character

displacement of call frequency in rhinolophids to avoid the frequencies used by other species; social selection for clear communication channels has previously been suggested to operate in bats employing high duty-cycle echolocation (Kingston et al 2001). It is worth noting, however, that the call frequency differences between Buton *R. philippinensis* morphs are far greater than the shifts in echolocation call frequency to avoid competition in the Sardinian rhinolophids (*R. euryale*: ~2.5kHz difference between the peninsula and Sardinian populations; *R. hipposideros*: 3-4kHz difference (Russo et al 2007). Buton *R. philippinensis*: Intermediate morph ~12kHz higher than Large morph; Small morph ~15kHz higher than Intermediate morph (Kingston and Rossiter 2004)). Combined with the observation that the Small and Intermediate morphs echolocate at different harmonic frequencies of the Large morph's fundamental frequency, it is unlikely that the divergence in echolocation call frequency between the Buton morphs is a similar case of character displacement in secondary contact to ensure clear communication channels.

Adaptation to local microhabitats has also been implicated as a driver of within-species divergence in echolocation call frequency in *Rhinolophus* species. *Rhinolophus damarensis* can be found over a large geographical range in southern Africa, with a mean resting echolocation call frequency that varies from ~84.4kHz to ~87.6kHz across its range (Maluleke, Jacobs and Winker 2017). These variations in call frequencies correlate with environmental discontinuities across *R. damarensis*' range, particularly with annual mean temperature, and so are proposed to be a result of adaptation to local environments (Maluleke, Jacobs and Winker 2017). The environmental discontinuities also correlate with genetic differences between subpopulations, so this system may represent an incipient speciation driven by isolation by environment (Maluleke, Jacobs and Winker 2017). Divergence in echolocation resting frequency in response to local conditions has also been reported with *Hipposideros ruber* in response to humidity in the Gulf of Guinea

(Guillén, Juste and Ibáñez 2000) and in African *Rhinolophus clivosus*, with differences correlating with temperature and humidity (Jacobs et al 2017). Geographic variation in call frequency in the Taiwanese *Rhinolophus monoceros* can also be partially explained by differences in humidity, although vicariant events and social selection might also have led to abrupt population variation (Chen, Jones and Rossiter 2009).

Examples of divergence related to differences in local environmental conditions are consistent with the Sensory Drive Hypothesis, where the adaptation of sensory systems to local conditions can lead to lineage diversification, reproductive isolation and speciation (Boughman 2002, Kawata et al 2007). The transmission of ultrasonic sound attenuates rapidly as it passes through the atmosphere, with higher frequencies experiencing more extreme attenuation (Lawrence and Simmons 1982). Increasing humidity can lead to further increased atmospheric sound absorption and greater signal attenuation (Hartley 1989). These properties of sound transmission strongly suggest that echolocation call frequency could represent a candidate trait for evolution under the Sensory Drive Hypothesis, especially in high duty-cycle bats. However, in general, the observed divergences in echolocation call frequencies in cases where sensory drive has been implicated as a possible factor are much smaller than the differences between the Buton *R. philippinensis* morphs. The *R. philippinensis* morphs on Buton are found in the same locations, to the point that they can be captured in the same traps, and all three forms have not been reported to occur together in anywhere else. With a surface area of 4,408km^2, Buton Island is unlikely to present a high variety of local environments to drive local adaptation in sensory signals.

In this chapter, I carry out genome-wide divergence scans between the three morphs of Buton *R. philippinensis* looking for evidence of a 'mosaic of divergence'. Finding heterogeneous patterns of genomic divergence would support the hypothesis that reproductive isolation recently developed or is currently in the process of developing

between these morphs, consistent with a scenario of incipient ecological speciation in the face of ongoing gene flow between populations. Given the hypothesis that this ecological speciation event is driven by divergent selection on echolocation call frequency and body size to exploit different niches, supported by assortative mating based on these traits, I hypothesize that genes linked to these phenotypic traits will be found in the islands of divergence. I test this hypothesis both by looking for signs of gene ontology enrichment in diverged islands and by investigating whether there is overrepresentation in these islands of a list of candidate genes associated with hearing or with variation in mammalian body size.

Materials and Methods

Sample collection, DNA sequencing, whole genome sequencing referencing assembly, genome alignment and SNP calling

I used the same samples in these analyses as in Chapter 3. Briefly, individuals of *R. philippinensis* were sampled on Buton Island between 2000 and 2012 as described in Kingston and Rossiter (2004). I collected additional samples from *R. philippinensis* in the collections of the Bogor Zoology museum, and supplemented these further with additional outgroup specimens from East Timor (*Rhinolophus montanus*) (provided by Kyle Armstrong) and Borneo (sampled under permit). This last individual was dissected in the field by Professor Steve Rossiter immediately after capture, with nine individual organs (heart, stomach, intestine, eyes, brain, lung, pancreas, spleen and liver) stored immediately in RNALater.

I extracted DNA from all samples using the Qiagen DNEasy blood and tissue kit, with the exception of the Bornean *R. philippinensis*, where the Qiagen AllPrep kit was used to extract both DNA and RNA from heart tissue. I prepared 500bp insert libraries for six individuals using the NEBNext Ultra DNA Library Prep Kit for Illumina that were sequenced by Barts and the London Genome Centre (Illumina HiSeq 2500 V4 Chemistry, 2 x 125bp, 5X sequencing depth. The rest of the DNA extractions were sent to Novogene (Hong Kong) for construction of 350bp insert libraries using the NEBNext Ultra II kit, and sequencing (Illumina HiSeq; 2 x 150bp, 10X sequencing depth).

I generated a reference genome assembly for the Bornean *R. philippinensis* using a method that combines reference-guided and *de novo* assembly steps, as described in

Wang et al (2014). For the reference-guided steps of this assembly process, I mapped my short-read data to the *Rhinolophus ferrumequinum* DISCOVAR assembly published by the Broad Institute (provided by Jeremy Johnson, Broad institute). Once the *R. philippinensis* reference was assembled, I was able to align all my other *R. philippinensis* short-read data to it using BWA-MEM version 0.78 (Li and Durbin 2009; Li 2013). I identified a set of high-confidence single nucleotide polymorphisms by preforming SNP calling on the mapped bam files using three separate methods (GATK version 3 Haplotype Caller GVCF mode best practices pipeline (McKenna et al 2010, DePristo et al 2011, Van der Auwera et al 2013), bcftools version 1.8 mpileup (Li 2011) and freebayes version 1.1.0 (Garrison and Marth 2012)), applying a set of quality filters with VCFtools (Danecek et al 2011), and then retaining only SNPs that had been called in all three methods (bcftools version 1.8 isec (Li 2011)).

In addition to these methods, I also constructed a second high-confidence SNP dataset by relaxing the requirements around the missing data. Briefly, in the filtering procedure I retained SNPs that had not been genotyped in up to 10% of individuals. This complemented the earlier other dataset described in Chapter 3, in which SNPs were only included if they had been successfully genotyped in all individuals.

RNA extraction and transcriptome assembly

To generate a reference transcriptome, I generated RNASeq data from the nine separate tissues that had been sampled from the Bornean *R. philippinensis* individual. I used the Qiagen AllPrep DNA/RNA mini kit to extract RNA from the heart tissue and the Qiagen RNEasy mini kit to extract RNA from the other eight tissues; stomach, intestine, eyes, brain, lung, pancreas, spleen and liver. Extraction quality control was performed using

the Qubit 2.0 fluorometer (Life Technologies) and the 2100 Bioanalyzer (Agilent), treating all samples showing evidence of DNA contamination with the TURBO DNA-free Kit (Invitrogen). I prepared libraries for each individual tissue using the Illumina TruSeq RNA Sample Prep Kit v2 with standard Illumina indexing. These libraries were then pooled and sequenced on a single lane by Barts and the London Genome Centre (Illumina NextSeq, 2 x 150bp).

All sequencing reads from the nine tissues libraries were used in a single *de novo* transcriptome assembly using Trinity (version 2.4.0, Grabherr et al 2011, Haas et al 2013). To identify coding sequences present in the Bornean *R. philippinensis* transcriptome, I performed a reciprocal blast procedure against human proteins using the same procedure as that described in Chapter 2. Once I had extracted the nucleotide sequences from the *R. philippinensis* transcriptome for genes with apparent single-copy orthology to human genes, I identified their locations in the *R. philippinensis* reference genome using GMAP (version 2018-03-25, Wu and Watanabe 2005, Wu and Nacu 2010).

Genome-wide F_{ST} scans

To calculate pairwise F_{ST} values for each SNP between each pair of morphs (Large-Small, Large-Intermediate, Small-Intermediate), I used the –weir-fst-pop method in VCFtools (Danecek et al 2011). This method generates an F_{ST} estimate based on Weir and Cockerham's paper (1984). Given that I had called SNPs simultaneously across all individuals and I was estimating F_{ST} for subsets, some of the positions were invariant in some pairwise comparisons. In these cases, I elected to remove SNPs that were invariant from the given comparison prior to estimating F_{ST}. I then used custom R scripts based on those used in Feulner et al (2015) to calculate mean F_{ST} over all SNPs within non-

overlapping sliding windows. I executed this procedure twice using different window sizes, with windows of 10,000 and 20,000 base pairs based on Feulner et al (2015).

To identify highly diverged windows, I examined the intersection of two parameters following the approach of Feulner et al (2015). First, I identified the subset of windows that were in the top 1% of the F_{ST} distribution. Additionally, I used a permutation method in which I ran 1,000,000 permutations of loci across the genome to test F_{ST} estimates against permutations containing the same numbers of SNPs. To correct for multiple tests, I applied an FDR correction (Benjamini and Hochberg 1995) in R. Windows that were in the top 1% of the F_{ST} distribution and which returned a significant result from the permutation test after correction for multiple testing were considered to be islands of divergence and used in subsequent analysis.

Association of genes with islands of divergence

To identify putative genes under divergent selection, I screened for all loci located either within anislands of divergence or within one window up- or downstream of an island of divergence among the *R. philippinensis* morphs. For this I compared the genomic locations of windows of elevated F_{ST} with the locations of genes, with the latter determined by mapping the *R. philippinensis* RNA sequences (single-copy orthology with human proteins) to my *R. philippinensis* reference genome. For this, I created custom R scripts to compare the positions of the genomic windows as calculated in R with the genomic regions to which the transcripts had been mapped by GMAP. All genes that fell either wholly or partially within a diverged window, as well as genes that fell within the windows adjacent to it up- or downstream, were retained for further analyses.

Gene ontology analysis

To determine whether genes associated with genomic islands of divergence have particular biological functions, I used a gene ontology approach in which I tested for signatures of functional enrichment. For this I used the PANTHER Overrepresentation Test (released 20171205, Mi et al 2016) as implemented via the Gene Ontology website (www.geneontology.org, Ashburner et al 2000, GO Consortium 2017). I used the "GO biological process complete" analysis (GO Ontology Database Released 2018-06-01), where a biological process is defined as "a recognized series of events or molecular functions" with a discrete beginning and end (Biological Process Ontology Guidelines, www.geneontology.org). In each case, the background set was the list of genes with single-copy orthology to human genes identified in the *R. philippinensis* transcriptome that had been mapped successfully to the *R. philippinensis* genome. The foreground for each test was the list of genes that had been identified in association with diverged windows in any given pairwise F_{ST} scan. This implementation of gene ontology uses a Fisher's exact test with FDR multiple test correction by default.

Candidate gene identification

The three Buton *R. philippinensis* morphs are phenotypically differentiated by size and by echolocation call frequency. In order to establish whether the divergence in these characters was linked to the genomic islands of divergence that I had identified, I collated lists of candidate genes associated with these traits. For my candidate hearing gene list, I downloaded all genes listed as being associated with hearing or with ear diseases in the

rat genome database (Shimoyama et al 2015; downloaded 18/06/2018) for a total of 573 candidate genes. I manually generated a candidate list of size genes by looking at publications investigating size variation in mammals (Sutter et al 2007, Kemper, Visscher and Goddard 2012, Makvandi-Nejad et al 2012, Rimbault et al 2013, Hayward et al 2016, Bouwman et al 2018), giving a list of 73 candidate genes.

To assess whether sets of genes in or near to windows of elevated divergence were associated with size variation or hearing, I simulated 10,000 random gene sets from the background gene list of equal size to the number of genes found in windows of divergence between the three pairs of morphs. From these random gene lists I established the expected distributions for the number of candidate (i) hearing genes and (ii) body size genes. For each pair of morphs, I compared the observed numbers of genes to its random distribution, and obtained the p-value as the chance of obtaining at least this number in a sample of equivalent size. I then performed false discovery rate correction on the obtained lists of p-values in R. This procedure allowed me to determine whether the candidate size or hearing genes were overrepresented in or around windows of divergence.

Finally, I identified whether any of the SNPs within the window of elevated divergence fell within the coding region of a candidate gene and, if so, whether it was a non-synonymous change in the amino acid sequence. I manually compared the detailed GMAP output providing information on the exact mapping positions of the transcriptome to the genome with the positions of the SNPs called within the windows of elevated divergence. This allowed me to define likely introns in the genome and from here to predict whether a SNP within these introns would be synonymous or non-synonymous. The presence of non-synonymous changes within significant islands of divergence would add support to a functional consequence of SNPs among the morphs, consistent with divergence driven by ecological speciation.

Results

Genomic SNP identification

Summary data regarding the DNA sequencing datasets, the Bornean *R. philippinensis* reference genome assembly, and the alignment of short read data from other *R. philippinensis* individuals to this reference, are provided in Chapter 3. In summary, short read sequence data were successfully generated for 42 individuals (13 Large morphs, 21 Small morphs, 4 Intermediate morphs and 4 outgroup individuals). The Bornean *R. philippinensis* reference genome consisted of 47,912 scaffolds between 1,000 and 1,876,277 base pairs in length (mean length: 43,360; N50: 166,896), with a total assembly length of 2,077,441,914 base pairs, or just over 2 Gbp.

To identify SNPs, I used three approaches and two levels of stringency (conservative and relaxed). Briefly, restricting my search to SNPs genotyped in every individual (conservative), I recorded 2,295,016 SNPs called by GATK, 5,055,388 by samtools and 3,999,838 by Freebayes. Of these, 1,319,236 SNPs were called by all three methods and so were retained for use in these analyses. When genotype data were permitted to be missing in up to 10% of individuals (relaxed), 6,903,311 SNPs were called by GATK, 8,237,963 by samtools and 6,255,833 by Freebayes. In this case, 3,463,691 SNPs were identified by all three SNP calling methods and were included in downstream analyses.

Transcriptome assembly and gene identification

I carried out RNAseq of nine separate tissues from one individual (Table 4.1), then used the reads in a single Trinity assembly. Trinity assembled the transcriptome reads into 402,962 contigs of between 201 and 33,877 base pairs in length, with a mean contig size of 1,011 and a total length of 407,342,612 base pairs. 14,997 transcripts with one-to-one orthology to human proteins were identified in this transcriptome with a reciprocal blast procedure, of which 14,962 mapped to the Bornean *R. philippinensis* reference assembly using GMAP. This list of genes formed our background gene set in downstream analyses.

Genome-wide F_{ST} scans

Overall, I detected a slightly lower level of divergence between the Large and Intermediate morphs than between the other morph pairs. Mean whole-genome F_{ST} values were 0.11 and 0.12 based on conservative and relaxed genotyping criteria, respectively, between the Large and Intermediate morphs. In contrast, I recorded a whole-genome F_{ST} of 0.14 based on both criteria for the Large versus Small morph comparison, and also the Large versus Intermediate morph comparison (Table 4.2).

The mean F_{ST} averaged across all diverged windows, regardless of pair of morphs being compared, ranged between 0.73-0.85 (Table 4.2). The mean F_{ST} values in diverged windows followed the same pattern as the genome wise mean F_{ST} in that the values are lower between the Large and Intermediate morphs than they are in either of the other comparisons (Table 4.1). Mean F_{ST} between diverged windows was slightly higher for 10kb windows than for 20kb windows. In terms of the number of significantly diverged

windows, I recorded most between the Large and the Small morphs (Table 4.2), with similar numbers of significantly diverged windows found in the other two comparisons. When calculating F_{ST} over 10kb windows, the comparison between Large and Intermediate morphs had the second greatest number of diverged windows, while when considering 20kb windows this was the comparison between Small and Intermediate morphs.

Diverged genes

By comparing the genomic coordinates of windows and loci, I found ~160 genes in divergent windows in the Large versus Small comparison and ~140 for the Small versus Intermediate comparison with both of these values being robust to window size and to the threshold used for SNP genotyping (Table 4.3). Window size had an impact on the number of genes found in diverged windows between the Large and Intermediate morphs; I found 128 genes with a window size of 10kb regardless of the threshold for SNP genotyping, and 157 and 170 for 20kb windows (conservative and relaxed genotyping threshold, respectively) (Table 4.3). A further 54-69 genes can be identified within one window up or downstream of a 10kb window, or 103-112 genes one window up or downstream of a 20kb window.

I carried out gene ontology analyses on each pairwise comparison, using the set of genes with one-to-one orthology to human genes that were present in the Bornean *R. philippinensis* reference genome as the background. There were no significantly enriched biological process categories after FDR correction in any of the pairwise comparisons.

Candidate genes

For each pairwise comparison of morphs, I compared the observed set of genes associated with diverged windows to 10,000 randomly generated gene sets of equivalent size. Within these observed and random gene sets I counted the numbers of candidate genes linked to hearing (n=456) and mammalian body size (n=65). By comparing empirical and expected counts of genes I found limited evidence of an overrepresentation of genes associated with hearing in the windows directly adjacent to diverged 10kb windows between the Small and the Intermediate morphs (p = 0.02) (Table 4.3) according to the strict criteria for SNP calling. This association was no longer significant when following the relaxed criteria (p = 0.08), and was no longer significant after FDR correction (p = 0.46). None of the other comparisons showed a significant overrepresentation of hearing genes.

I also detected a stronger pattern of overrepresentation of genes associated with size variation in mammals falling in diverged windows between Large and Small morphs (p = 0.01-0.03 depending on window size and proportion missing data allowed; no longer significant after FDR correction) and between Small and Intermediate morphs (p = 0.02 when calculated on 10kb windows; association not significant with 20kb windows. No longer significant after FDR correction) (Table 4.3).

No hearing or size gene was common to a diverged window in all three pairwise comparisons. However, one gene, *Gja1*, was present within one window of a diverged window in each test (Table 4.4). A further 12 genes fell in a diverged window or within one window up- or downstream or a diverged window in two of the pairwise comparisons (Table 4.4).

Screens of SNPs among morphs revealed that 16 candidate genes that overlapped with diverged windows contained SNPs in putative exons (two examples in Figure 4.1). Out of these genes, 11 contained non-synonymous SNPs (Table 4.5).

Discussion

In this study I carried out genome-wide scans of divergence between three sympatric size and acoustic morphs of *R. philippinensis* found on Buton Island. Based on previously published papers (Kingston and Rossiter 2004) and my own research (see Chapter 3), I hypothesised that I would find a genomic mosaic of divergence between these morphs, indicating current or recent gene flow between them with recombination happening unevenly across the genome (Via 2001). Further, I expected to find that 'islands of divergence' identified between the morphs would be associated with putative 'speciation genes' (Wu 2001). These are loci that maintain a high level of genomic differentiation in the face of gene flow and which thus may play a role in driving the ecological speciation. In my study organisms, such loci are expected to be linked to the phenotypic divergence between the morphs, which in this case relates to body size and hearing.

Windowed calculations of F_{ST} across the genome between each pair of morphs revealed that the overall level of genetic divergence is low. Specifically, mean genome-wide F_{ST} between the Intermediate and Large morphs was 0.11-0.12, and between either the Small morph and any other morph was 0.14. These F_{ST} values are close to those initially reported between these three morphs based on variation in twelve microsatellite loci (Kingston and Rossiter 2004) and show the same pattern: there is a similar level of divergence between the Small and Large and the Small and Intermediate morphs, with the Large and Intermediate morphs being less differentiated. This result is consistent with the results of phylogenetic analyses in Chapter 3, with the Large and Intermediate morphs appearing to be less differentiated from each other, and thus more closely related, than the Small morph is to either of the other morphs. As all of these bats are located in the same forest, this pattern is not easily explained by current introgression and so is suggestive of the

Intermediate morph either arising from a past hybridisation event with backcrossing to the Large morph, or from a recent split of the Intermediate and the Large morph.

This level of F_{ST} is commonly detected among island bat populations (Kingston and Rossiter 2004), though is lower than the differentiation seen between the *R. philippinensis* morphs and other *Rhinolophus* species that were caught in the same locality, supporting the hypothesis that the morphs are either not yet fully reproductively isolated or that they only recently became so (Kingston and Rossiter 2004). The background F_{ST} between the Buton *R. philippinensis* morphs is lower than that reported in cases of species hybridising in secondary contact (e.g. *Ficedula* flycatchers, Ellegren et al 2012), but higher than that observed between ecotypes of a single species (e.g. sockeye salmon *Oncorhynchus nerka*, Larson et al 2017) or cases of local adaptation of populations of a species (e.g. North American swamp sparrow *Melospiza georgiana*, Deane-Coe et al 2018).

I identified distinct regions of dramatically elevated F_{ST}, in which the mean F_{ST} values in diverged windows were seven to eight times the background F_{ST} across the rest of the genome. Dependent on morphs and window size, I found between 465 and 1004 diverged windows in the genome. This number is a greater than that reported in three-spine sticklebacks *Gasterosteus aculeatus* using very similar methods (128-192 dependent on population comparison using non-overlapping 10kb windows; Feulner et al 2015). However, the stickleback genome is estimated to be around 460MB in length (Ensembl release 93- July 2018), while bat genomes are estimated to be ~2GB (Fang et al 2015), so this difference in the number of windows is proportionate.

Using randomisations revealed that candidate genes associated with body size variation in mammals were overrepresented within diverged windows in comparison with the genome as a whole between the Small and Intermediate or Small and Large morphs. I found 28 hearing genes that overlapped with regions of F_{ST} divergence. The candidate list

of hearing genes that I used was broad, containing all genes known to have an association with any kind of hearing or ear disorder in the Rat Genome Database (Shimoyama et al 2015). Of these 28 genes, 12 have previously appeared in other publications looking at molecular evolution in echolocating taxa, with four of them being directly highlighted when considering the evolution of echolocation. *Col11a1* has GO annotations concerning the sensory perception of sound and the detection of mechanical stimulus involved in the sensory perception of sound, and has been detected to be convergent between bats and dolphins (Chabrol et al 2017). *Myo7a* was one gene contributing to a functional enrichment of the same GO terms in *Myotis davidii* but not in *Pteropus alecto,* with other genes in the functional cluster leading to the suggestion that it was related predominantly to inner-ear development (Hudson et al 2014). *Strc* was highlighted as possibly containing convergent amino acid sites between different lineages of laryngeal echolocating bats, with all echolocating bats forming a monophyletic clade for this gene with the amino acid variants present having important functional effects (Dong et al 2016). This monophyly of echolocating bats from *Strc* gene trees was in direct contradiction to the findings of an earlier publication (Kirwan et al 2013). Another known hearing gene detected within a divergent window is *Slc26a4*, which encodes pendrin, the closest paralog to *Prestin* (*Slc26a5*) with 40% sequence identity (He et al 2014). Prestin is one of the key genes associated with the evolution of high-frequency hearing in echolocating bats (Li et al 2008, Shen et al 2011).

A single gene typically has roles in a range of different biological pathways or processes, and so genes that I have classified as hearing genes may have been considered from the perspective of some other function in previous research into molecular evolution in echolocating taxa. *Clcn3*, *Lmbrd1* and *S100a9* have been found to be evolving under positive selection in the bottlenose dolphin (Yim et al 2014; McGowen, Grossman and Wildman 2012). *Gja1* has previously been found associated with an island of F_{ST}

divergence in a comparison of great leaf-nosed bat *Hipposideros armiger* populations that live at different altitudes (Dong et al 2016).; while the authors of this study were interested in its role in hypoxia response due to its links with the cardiovascular system, it also facilitates potassium recycling from cochlear hair cells during auditory transduction (Wang et al 2009). In relation to humans, *Myotis brandtii* has been found to have gained an additional copy of *Spata5* (Seim et al 2013). Cetacean-specific mutations predicted to have a functional effect on the gene were discovered in *Fgf10* (Nam et al 2017).

The *Fgf10* gene, along with *Fgfr2* – that later of which I found in association with islands of divergence between Buton *R. philippinensis* morphs and which has been linked to both body size and hearing - form part of a signalling pathway that has been directly connected to the development of bat wings (Tokita, Abe and Suzuki 2012). Other size genes that I found in islands of divergence have also been linked to bat wings; *Plag1* has been implicated specifically in their embryonic development (Booker et al 2016), while *Col10a1* is associated more generally with the process of digit formation (St-Jacques et al. 1999; Yoshida et al. 2004)and *Col11a1* additionally playing a critical role in skeletal morphogenesis and particularly of limb development (Li et al 1995). In bats, these two collagen genes could also be expected to impact on wing morphology, with.

Examining the results of all of the candidate genes that I found to be associated with islands of divergence, it is noteworthy that a large number have already been highlighted as evolutionarily interesting in echolocating taxa, or developmentally interesting in limb development or even in bat wings specifically. This finding lends further support to the hypothesis that the genes associated with these islands of differentiation are involved in driving the phenotypic divergence between the morphs.

Inferring divergence or speciation with gene flow between populations or species on the basis of heterogeneous differentiation between their genomes is not always

straightforward. For example, in *Helianthus* sunflowers, islands of divergence are more closely associated with reduced recombination rates than they are with interspecific gene flow, which tend to co-map with breakpoints of major chromosomal rearrangements (Renaut et al 2013). This has led to the suggestion that genomic architecture plays a larger role in patterns of genomic divergence than speciation geography (Renaut et al 2013). Comparing alleles differentiating benthic and limnetic pairs of stickleback *Gasterosteus aculeatus* in 34 global populations with those found in nearby solitary populations suggested that these adaptive variants evolved in allopatry, rather than the species pairs diverging in sympatry in the face of gene flow (Jones et al 2012). A caveat to this last is that subsequent analyses of specific populations since this study that explicitly test alternate hypotheses have supported genetic differentiation arising and persisting between population pairs in sympatry where genes relevant to adaptation and mate choice are co-localized in the genome (e.g. Marques et al 2016).

More generally, reanalysis of five studies showing islands of elevated F_{ST} between species pairs revealed that regions with high relative measures of divergence also showed reduced diversity and so did not have elevated measures of absolute sequence divergence (e.g. D_{XY}) (Cruickshank and Hahn 2014). They suggest that regions of elevated F_{ST} without elevated absolute divergence are more likely to be driven by linked selection than by differential migration between loci, potentially as a result of local adaptation, and caution against cases of gene flow on secondary contact between species that diverged in allopatry being mistaken for 'speciation-with-gene-flow' (Cruickshank and Hahn 2014). In the case of the Buton Island morphs under study here, one argument against divergence in allopatry followed by secondary contact for explaining the observed genomic patterns is the lack of obvious source populations for the different morphs. While there was only a small representation of outgroup species from surrounding islands designated as *R. philippinensis* in this study, their phenotypic characteristics are most similar to those seen

in the Buton Intermediates (see Chapter 3). Given that the morphs occupy the same areas of Buton Island to the extent that they can be caught in the same traps, local adaptation is also unlikely as an explanation.

The array of methods available for calculating genome-wide F_{ST} and other divergence statistics has exploded in recent years. While we still face challenges in interpreting the biological significance between the patterns of genomic differentiation that we observe between populations, novel approaches to answer questions of hybridization and speciation are continuously being developed (e.g. looking at ancestry junctions to infer history and identify loci driving reproductive barriers in hybrid zones; Hvala, Frayer and Payseur 2018). The methods that I have applied here reveal that there is a genomic mosaic of differentiation between the Buton Island *R. philippinensis*, with a low background F_{ST} punctuated by islands of divergence that are consistent with the hypothesis of speciation with gene flow. Also, in line with a hypothesis of ecological speciation, I expected to find that these islands of divergence were enriched in genes associated with echolocation and body size. These phenotypic traits are closely linked, highly divergent between the morphs, and likely to be acted upon by natural selection while also playing a role in pre-mating reproductive isolation by assortative mating. I found a strong signal for overrepresentation of genes associated with body size in these regions, with a high number of the candidate genes linked to hearing or body size that appeared in these regions having documented links to bat wing development, echolocation, or more generally to positive selection in echolocating taxa. These results provide strong support for the divergence of the three *R. philippinensis* morphs of Buton Island representing an incipient ecological speciation event.

Tissue	# Raw Reads
Brain	20,919,895
Eye	22,667,637
Heart	19,802,219
Intestine	20,420,494
Liver	22,580,443
Lung	22,754,142
Pancreas	22,295,708
Spleen	22,821,602
Stomach	20,142,887

Table 4.1: Volume of RNAseq reads generated from different tissues collected from the Bornean *R. philippinensis*.

	Large-Intermediate				Large - Small				Small - Intermediate			
Window size	10kb		20kb		10kb		20kb		10kb		20kb	
Max % missing	0	10	0	10	0	10	0	10	0	10	0	10
Mean genome-wide F_{ST}	0.12		0.11		0.14		0.14		0.14		0.14	
Mean variants/genome-wide window	6.4		11.5		7.2		13.2		6.6		12	
# diverged windows	685	716	572	602	979	1004	660	665	704	727	465	485
Mean diverged window F_{ST}	0.81	0.8	0.73	0.73	0.84	0.83	0.77	0.77	0.85	0.85	0.8	0.79
Mean variants/diverged window	5.6	5.8	10.4	10.3	6.7	7.1	12.5	12.7	5.2	5.3	8.5	8.9

Table 4.2: Windowed F_{ST} divergence between population pairs.

Comparison	Max % Missing Genotypes	Window Size	Gene Location	All Genes	Hearing Gene #	*p*	FDR *p*	Size Gene #	*p*	FDR *p*
Large vs. Intermediate	0	10kb	In Wins	128	7	0.095	0.46	1	0.43	0.43
			Near Wins	67	5	0.055	0.46	1	0.25	0.39
		20kb	In Wins	157	7	0.20	0.6	2	0.14	0.28
			Near Wins	103	3	0.62	0.64	1	0.36	0.42
	10	10kb	In Wins	128	7	0.095	0.46	1	0.43	0.43
			Near Wins	69	4	0.16	0.55	1	0.26	0.39
		20kb	In Wins	170	7	0.27	0.63	2	0.17	0.31
			Near Wins	105	3	0.62	0.64	1	0.37	0.42
Large vs. Small	0	10kb	In Wins	161	5	0.55	0.64	3	0.03	0.09
			Near Wins	65	3	0.33	0.63	0	NA	NA
		20kb	In Wins	159	5	0.52	0.64	3	0.03	0.09
			Near Wins	107	3	0.64	0.64	1	0.37	0.42
	10	10kb	In Wins	167	6	0.41	0.63	4	0.01	0.09
			Near Wins	68	3	0.34	0.63	0	NA	NA
		20kb	In Wins	159	5	0.52	0.64	3	0.03	0.09
			Near Wins	107	3	0.64	0.64	1	0.37	0.42
Small vs. Intermediate	0	10kb	In Wins	135	6	0.23	0.61	3	0.02	0.09
			Near Wins	55	5	0.02	0.46	0	NA	NA
		20kb	In Wins	140	5	0.42	0.63	2	0.12	0.27
			Near Wins	107	4	0.41	0.63	0	NA	NA
	10	10kb	In Wins	143	7	0.15	0.55	3	0.02	0.09
			Near Wins	54	4	0.08	0.46	0	NA	NA
		20kb	In Wins	140	5	0.42	0.63	2	0.12	0.27
			Near Wins	112	4	0.45	0.64	0	NA	NA

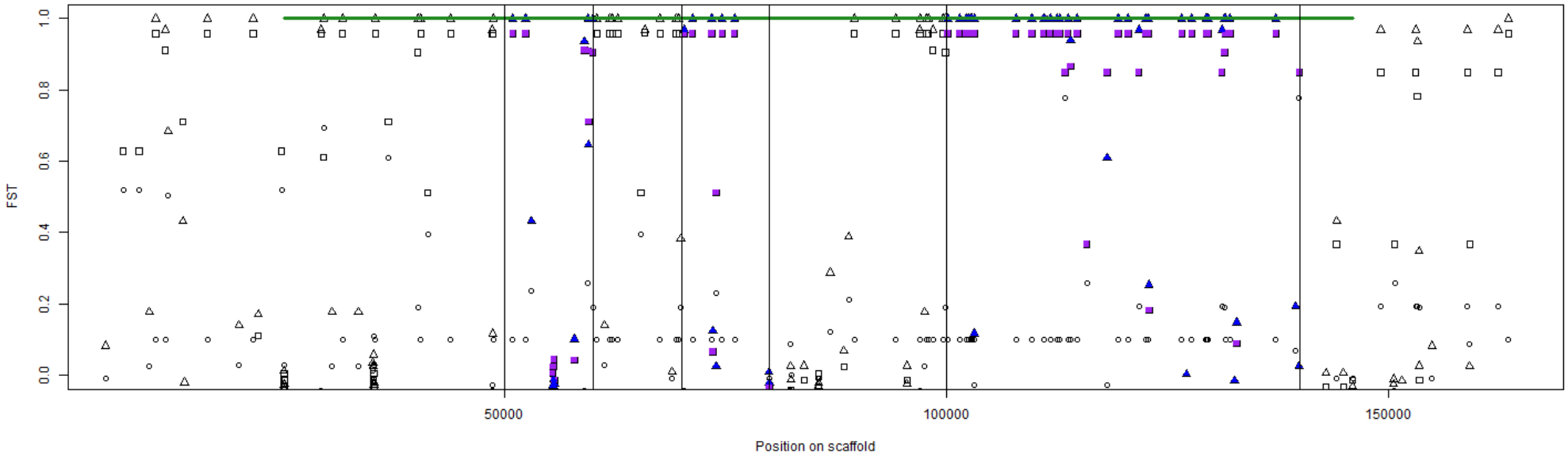
Table 4.3: Number of overall genes, hearing genes and size genes found in diverged windows or within one window of a diverged window.

Gene	Size/Hearing	Large-Int	Large-Small	Small-Int
Alb	Hearing		Near window	
Cav1	Hearing			In window
Ccnd2	Size	In window		
Cdc14a	Hearing	In window		
Clcn3	Hearing	In window		In window
Clic5	Hearing	In window		
Col10a1	Size		In window	In window
Col11a1	Size/Hearing		In window	In window
Cyp2e1	Hearing	Near window		Near window
Elmod3	Hearing	In window		
Fgf10	Hearing	Near window		Near window
Fgfr2	Size/Hearing	In window	In window	
Fndc3b	Size		In window	In window
Gja1	Hearing	Near window	Near window	Near window
Gjb1	Hearing			In window
Lmbrd1	Hearing			In window
Mmp9	Hearing	Near window		
Msrb3	Hearing		In window	
Myo7a	Hearing			Near window
Piezo2	Hearing			In window
Plag1	Size	Near window	Near window	
Pogz	Hearing			In window
Prkcb	Hearing	In window		
Prkg1	Hearing			Near window
Rrp15	Hearing	In window	In window	
S100a9	Hearing			Near window
Slc26a4	Hearing		In window	
Spata5	Hearing	In window		
Strc	Hearing			Near window
Tcfa	Hearing	Near window	Near window	
Tmtc2	Hearing		In window	In window
Vps13b	Hearing		In window	In window

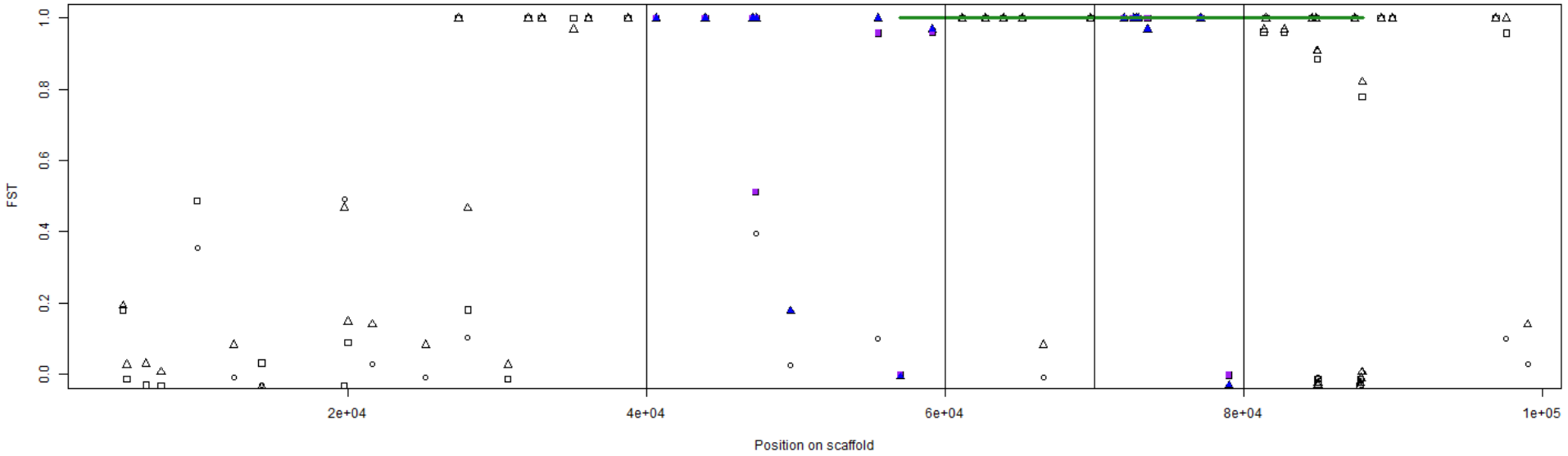
Table 4.4: Candidate size and hearing genes found within diverged windows ('In window') or within one window of a diverged window ('Near window') in pairwise divergence scans of the three Buton morphs.

Gene	Size/Hearing	Alignment length (bp)	Synonymous SNPs	Non-synonymous SNPs
Alb	Hearing	6,812	1	0
Cav1	Hearing	250,051	6	3
Clcn3	Hearing	61,700	0	2
Clic5*	Hearing	96,099	0	1
Col11a1[1]	Size/Hearing	31,012	16	1
Col11a1[2]	Size/Hearing	11,270	8	10
Cyp2e1*	Hearing	11,829	0	1
Fgfr2	Size/Hearing	95,574	20	7
Fndc3b	Size	250,051	3	0
Gja1*	Hearing	641	12	4
Lmbrd1*	Hearing	48,918	1	0
Myo7a	Hearing	30,360	15	2
Piezo2	Hearing	184,515	31	9
Plag1	Size	2,019	9	0
Slc26a4*	Hearing	35,451	2	0
Spata5*	Hearing	192,142	2	1
Strc	Hearing	5,216	4	3
Tgfa*	Hearing	53,941	0	1
Tmtc2	Hearing	172,538	2	0
Vps13b[1]	Hearing	120,946	21	8
Vps13b[2]	Hearing	212,096	71	38

Table 4.5: Candidate size and hearing genes that overlap with diverged windows and contain SNPs within apparent introns.
Superscript numbers: two separate alignments were produced for *Col11a1* and *Vps13b*, so both are presented.
*: SNPs were called on the non-coding strand and so have had to be translated.


FST
0.0
0.2
0.4
0.6
0.8
1.0
50000
100000
150000
Position on scaffold

a)



b)

Figure 4.1: Examples of F_{ST} divergence between morph pairs. Image shows the full length of a single scaffold which contains an island of divergence in the candidate gene. Points indicate F_{ST} values for a SNP between a morph pair; circles = Large vs. Intermediate, squares = Small vs. Intermediate, triangles = Large vs. Small. Filled coloured shapes indicate SNPs within a significantly diverged island. The location of the gene on the scaffold is indicated by a horizontal green line; the vertical lines indicate the start and end of the diverged windows. A) *Vps13b*; b) *Col11a1*.

CHAPTER FIVE

General discussion

This goal of this thesis has been to investigate the genetic basis of echolocation in bats, with a particular focus on high duty-cycle echolocation and its role in divergence and speciation. I have studied two systems at different levels of divergence in the pursuit of this goal. Firstly, I have generated new transcriptomic data and combined these with published genome sequences to look at the Mormoopidae family, in which a complex of sister species within the *Pteronotus* genus, currently assigned to the subgenus *Phyllodia*, are distinct from their congeners in their evolution and use of high duty-cycle echolocation. Secondly, I have applied whole-genome sequencing data to investigate the patterns and order of divergence, and molecular targets of selection, among three acoustic and size morphs of *Rhinolophus philippinensis* that occur in sympatry on Buton Island in Southeast Indonesia.

I have aimed to answer an array of questions in the course of this work. By investigating the selection pressures operating on genes within *Phyllodia* in a phylogenetic framework that included a wide array of other low duty-cycle echolocators in the Yangochiroptera, I addressed the question of whether candidate hearing genes play a role in the change of echolocation call strategy. Second, by comparing genes evolving under positive selection in this subgenus to those detected in members of the distantly related horseshoe bats from the Yinpterochiroptera that have independently evolved high duty-cycle echolocation, I asked whether selection has acted on the same sets of loci. This aim then allowed me to assess where in the *Pteronotus* genus the transition to high duty-cycle echolocation occurred. I also asked whether candidate hearing genes, in addition to other functional genes, might be implicated in the marked changes in the echolocation call frequencies of narrowband calls, and whether these loci would be associated with regions of high

divergence between acoustic morphs of *Rhinolophus philippinensis*. Finally, could I find evidence of population structuring and patterns of genomic differentiation between these acoustic morphs consistent with ecological speciation, driven by 'speciation genes' associated with echolocation and high-frequency hearing?

The evolution of high duty-cycle echolocation in Pteronotus *and* Phyllodia

In Chapter 2, I identified 492 genes that showed signatures of evolving under positive selection specifically in the *Phyllodia* subgenus including 23 hearing genes, three of which contained private substitutions in *Phyllodia* and which were supported by multiple lines of evidence. Given that there were also 27 hearing genes evolving under positive selection on the ancestral *Pteronotus* branch, this approach did not conclusively answer the question as to whether high duty-cycle echolocation was a trait which had evolved ancestrally in the *Pteronotus* with a subsequent reversion following the split between *Phyllodia* and its congeners, or whether it the transition to high duty-cycle echolocation had occurred specifically in the *Phyllodia*. However, given that the other extant *Pteronotus* exhibit partial characteristics typically associated with high duty-cycle echolocators – including Doppler-shift compensation in *Pteronotus personatus* (Smotherman and Guillén-Servent 2008), heteroharmonic target-range compensation in *Pteronotus quadridens* (Hechavarría et al 2013) and echolocation calls that feature short constant-frequency elements on the tails of their broadband sweeps (Mora et al 2013) that are absent in the sister genus *Mormoops* - it is possible that these separate elements evolved independently throughout the history of the *Pteronotus* genus. In this scenario, 'intermediate' echolocation forms are assumed to represent stable strategies that are able to persist. Indeed, some other bat species that use low duty-cycle echolocation, or a

mixture of broadband and narrowband elements in their echolocation calls, also possess some of these traits; for example, *Noctilio albiventris* has been described as using Doppler shift compensation (Roverud and Grinnell 1985).

High duty-cycle echolocation is a complex trait entailing phenotypic adaptations that encompass numerous physiological systems; these include modifications of the cochlear basilar membrane that produces an acoustic fovea (Schuller and Pollack 1979, Neumann and Schuller 1991, Davies, Maryanto and Rossiter 2013) as well as auditory cortex organisation (O'Neill 1995) and neural processing pathways (Mora et al 2013). There is also evidence of sensory trade-offs in bats between vision and echolocation (e.g. Liu et al 2015, Thiagavel et al 2018), particularly in bats using the more derived high duty-cycle forms of echolocation (Zhao et al 2009a, Shen et al 2010, Dong et al 2017, Gutierrez et al 2018a, Gutierrez et al 2018b). As such, looking at patterns of selection in genes associated with other biological functions and pathways, particularly brain development or vision, might help to inform the evolutionary history of this trait within members of the genus *Pteronotus*.

Population structure and islands of genomic divergence between acoustic morphs of Buton Island R. philippinensis

The results in Chapter 3 reveal genetic population structure among the three sympatric *R. philippinensis* morphs of Buton Island based on both mitochondrial and nuclear loci, providing strong evidence against panmixia. Based on nuclear genome-wide SNP data, the morphs form three separate monophyletic groups, with the Intermediate morphs appearing as a sister group to the Large morphs. Population inference based on Bayesian mixture modelling (Verity and Nichols 2016) without admixture provides the most

support for splitting the Small morphs from the Large and Intermediate morphs, then for all three morphs representing separate populations. Allowing admixture in the model still divides the Large and the Small morphs from one another at the optimal number of demes, with the Intermediate morphs containing an almost even mix of ancestry proportions from both of the other morphs.

The results based on mitochondrial data conflict with these findings. While they still form distinct clades in a maximum likelihood tree, the Large and Small morphs both form two groups and appear to be paraphyletic, with some Large individuals being more closely related to the Intermediate morphs than to the rest of the Large morphs. These results are consistent with those reported by Kingston and Rossiter (2004) from sequencing a fragment of the mitochondrial control region in a smaller number of individuals. Bayesian MCMC species delimitation using the multispecies coalescent (Rannala and Yang 2003, Rannala and Yang 2013) on the mitochondrial alignments was highly dependent on the guide tree provided. When I provided the maximum likelihood tree from the SNP data (with one clade per morph and reciprocal monophyly between the morphs), the guide tree topology with the three morphs as separate species had the greatest posterior probability. In this topology, the Small split from the other morphs first, with the Large and Intermediate morphs as sister groups. However, when using the topology generated from the mitochondrial data (with paraphyly of both the Large and the Intermediate morphs), the greatest posterior probability was for a species delimitation in which the Intermediate morphs and a subset of the Large morphs formed a single taxon. The inferred topology starting from the mitochondrial guide tree was also less certain, finding similar posterior probabilities for the maximum-likelihood mitochondrial topology provided as a guide tree and for two rearrangements – switching the positions of the two groups of Small morphs, or else switching the position of one of the groups of Small morphs with the combined Large and Intermediate morph group to position the two Small morph groups as sister

taxa. It is important to note that the algorithm underlying this Bayesian MCMC species delimitation makes an assumption that there is no admixture between the groups, and that violation of this assumption can have an impact on inferences (Yang and Rannala 2010). Despite this, Yang (2018) has stated that in the cases where gene flow is between sister species, then this should facilitate the inference of the correct species tree, with incorrect trees only being produced if the gene flow is between non-sister species.

Mitonuclear discordance has previously been reported in other *Rhinolophus* species (Mao et al 2010a, Mao et al 2010b, Mao et al 2013a, Mao et al 2013b), with multiple possible explanations that include incomplete lineage sorting of ancestral polymorphisms into recently diverged populations, sex-biased gene flow, and introgression of mitochondrial DNA (see also Dong et al 2014). In the Buton *R. philippinensis* it is likely that the genomic SNP trees represent the correct species phylogeny, as four consistent trees were constructed using four different sets of 10,000 unlinked SNPs. This represents a large number of loci, meaning that any discordant signal would be masked in the phylogeny by the dominant signal. Nuclear loci have been demonstrated to be a more reliable method of reconstructing the species phylogeny in *Rhinolophus* than mitochondrial DNA, even based on as few as two introns (Dool et al 2016).

Whole-genome scans described in Chapter 4 revealed low background levels of divergence between pairs of morphs, punctuated by regions with mean F_{ST} values seven times higher than the genomic background. The genic process of speciation (Wu 2001) proposes the concept of 'speciation genes' directly involved in the process of differential adaptation; these would account for a very small part of the genome, and, as long as these remained diverged, then populations could remain distinct even in the presence of recombination between them. Islands of genomic divergence are predicted in cases where genetic differentiation between populations or species is maintained in the absence of complete reproductive isolation (Via and West 2008), whether that is from speciation in

the face of gene flow, local adaptation with assortative mating, or hybridisation following secondary contact (Noor and Bennett 2009, Turner and Hahn 2010, Cruickshank and Hahn 2014). The islands of divergence between the Buton Large and Small morphs, and between the Small and Intermediate morphs, were enriched in genes known to be linked to variations in mammalian body size, and there was weaker evidence of enrichment in hearing genes between the Large and Intermediate, and also the Small and Intermediate morphs. Non-synonymous substitutions arising from the SNPs were also present in the coding regions of candidate hearing genes and size genes in these regions. Despite this, the low background genomic values of F_{ST} indicate that reproductive isolation is possibly not complete between the morphs, or that it has only recently developed (see also Kingston and Rossiter 2004). The overrepresentation of genes associated with the diverged phenotypes between the morphs in islands of divergence suggests some form of selection against recombination in these regions, maintaining their differentiation even if there is some level of genetic exchange (Via 2001). Echolocation call frequency is considered to be a magic trait in high duty-cycle echolocating bats (Servedio et al 2011); it has both an ecological role in that it affects prey species that can be detected (e.g. Houston et al 2004) and a role in assortative mating (e.g. Puechmaille et al 2014). Magic traits are considered to be strong candidates to drive ecological speciation (Servedio et al 2011). Body size and echolocation call frequency are closely linked and inversely correlated in *Rhinolophus* bats (Heller and von Helversen 1989), so theoretically a shift in body size could lead to a change in call or vice versa. For the time being it is unclear whether body size or echolocation call frequency would have changed first in these bats. The inner ear structures of the Buton Island *R. philippinensis* morphs are smaller than those of their relatives (Davies et al 2013), which has been proposed to be due to a recent origin of the morphs with a rapid evolution via shifts in call frequency. A signature of accelerated basilar membrane evolution has also been detected in these morphs (Davies,

Maryanto and Rossiter 2013). In general, the echolocation call frequencies of all three Buton Island morphs of *R. philippinensis* are lower than would be expected given their forearm lengths, body mass and basilar membrane lengths (Davies, Maryanto and Rossiter 2013). From these papers, one could infer that the shift in echolocation call frequency occurred first, with overall body size and size of the inner ear structures in particular still in the process of catching up. The results presented in these two chapters represent evidence that the Buton *R. philippinensis* are in the early stages of ecological speciation.

Signatures of parallel evolution in high duty-cycle echolocation between lineages that evolved it independently

While there has been a lot of research into genes associated with echolocation (e.g. Li et al 2007, Li et al 2008, Davies et al 2012, Liu et al 2012, Shen et al 2012, Parker et al 2013), less is known about the genetic basis of high duty-cycle echolocation specifically. The existence of two lineages of bats that have independently evolved high duty-cycle echolocation provides us with an opportunity to look for genes showing signatures of parallel or convergent signals between them. The results of Chapter 2 highlighted a set of eleven hearing genes that showed signatures of evolving under positive selection in both independent lineages of high duty-cycle echolocators, as elucidated by comparing genes identified in the yinpterochiropteran *Rhinolophus sinicus* (Dong et al 2017) and either ancestrally in the yangochiropteran *Pteronotus* genus or more specifically within its subgenus *Phyllodia*. These genes represent strong candidates to play a role in the evolution of high duty-cycle echolocation. However, none of these genes found to be under positive selection within both the *Pteronotus* genus and *R. sinicus* also showed up

as being associated with islands of divergence between different morphs of *Rhinolophus philippinensis*.

Of the genes found in association with islands of divergence between different morphs of *R. philippinensis*, I also found that 22 were detected as evolving under positive selection in *Phyllodia* and 33 in the ancestral *Pteronotus* branch (in comparison to 18 in *Pteronotus quadridens* and 19 in the ancestral *Mormoops* branch). There was one hearing gene found shared between diverged islands in *R. philippinensis* and each of *Phyllodia* (*Clcn3*), *P. quadridens* (*Prkcb*) and the ancestral *Pteronotus* branch (*Fgf10*). Two of these genes, *Clcn3* and *Fgf10*, have previously been found to be evolving under positive selection in echolocating cetaceans (Yim et al 2014, Nam et al 2017) and represent very interesting targets for further research into the evolution of high duty-cycle echolocation in bats. *Fgf10* plays a major role in mammalian inner ear morphogenesis during embryonic development (Pauley et al 2003), while *Clcn3* has inner ear expression specific to hair cells (Hertzano 2004) and has been associated with sensorineural hearing loss (Yoshikawa et al 2002).

While there are phenotypic similarities between the two lineages of high duty-cycle echolocators (Henson, Schuller and Vater 1985; Kossel and Vater 1985), there are also many phenotypic differences in how they have independently evolved this trait (O'Neill 1995, Jones and Teeling 2006). Previous research has indicated that these differences extend to the genetic level, with different molecular adaptations underlying the two independent origins of this trait (Li et al 2007, Shen et al 2011).

Caveats

The reliability of conclusions drawn from assemblies of whole-genome short-read data are limited by data completeness and quality. Regions of the genome which are not covered by sequencing reads cannot be assembled, leading to gaps in gene-based studies. Short-read assemblers are also still prone to errors, generating chimeric sequences (Yang and Smith 2013) that can lead to incorrect inferences of selection (Mallick et al 2009).

For Chapter 2, I generated sequence data for most members of the Mormoopidae by sequencing RNA from muscle tissue obtained from museum of university collections. While this approach generates very high coverage of the coding fraction of the genome, RNA expression is tissue specific and thus I was only able to obtain transcripts from the tissues available. Many of the hearing genes that may be associated with the evolution of echolocation will only be expressed in the cochlea or in the brain, and I only had access to cochlea tissue for a limited number of the species from the Mormoopidae (*Mormoops blainvillei*, *Pteronotus parnellii*, *Pteronotus quadridens*); this will have reduced my power to detect the candidate hearing genes in the data set and thus my ability to detect evolutionary signal. Indeed, of 609 candidate hearing genes, almost two-thirds (388 genes) were detected within the data set that included transcripts derived from cochlear tissues, while only just over one-third (225 genes) were present in the alignments of transcripts sequenced from muscle tissue.

In Chapters 3 and 4, I sequenced ~50 individual *R. philippinensis* at a low sequencing depth (5X – 10X). Some of these individuals had been in room-temperature storage in a museum in Indonesia for up to 15 years, raising the possibility of DNA degradation. Indeed, Bioanalyzer analyses of DNA quality returned many low DIN scores, with some individuals subsequently being omitted from my research due to either failing library

construction or failing to sequence. Low quality DNA increases the chances of sequencing errors, with low sequencing depth also introducing sequencing errors that can be propagated through downstream analyses (Sims et al 2014). I introduced measures to counter these, taking a conservative approach to SNP calling that may have reduced my power to detect phylogenetic and evolutionary signal by incorrectly filtering out genuine variants.

Future work

In order to further our understanding about the links between high duty-cycle echolocation and speciation, the *Rhinolophus philippinensis* morphs of Buton Island provide a promising system for further study, with multiple avenues of potential interest.

Short-read sequencing data are not highly suitable to detecting large complex genomic elements, such as copy number alterations and structural variations (Goodwin, McPherson and McCombie 2016). These genomic features have been shown to have effects on mammalian phenotypes. Copy number variants having known roles in human disease (McCarroll and Altschuler 2007), with DNA repeats even in non-coding regions altering gene expression in human cells leading to nearly 30 hereditary disorders (Mirkin 2007) and copy number variants even being linked to complex conditions such as autism and schizophrenia (Stankiewicz and Lupski 2010). Long-read sequencing technologies offer a solution for these kinds of complex genome assembly problems, with novel assembly methods being developed that enable us to best leverage their potential e.g. (Huddleston et al 2014, Chaisson et al 2015, Madoui et al 2015). These methods have been successfully applied to mammalian genomes, identifying novel differences between human and gorilla sequences likely to affect gene regulation up to thousands of bases in

length (Gordon et al 2016). Long-read and optical mapping approaches have been used in combination with short-read sequencing in the European crow *Corvus corax*, an avian speciation model, to increase understanding of genetic diversity and differentiation along the genome (Weissensteiner 2017), leading to the suggestion that a combination of approaches can give us access to more information than is accessible via each approach independently. Application of long-read sequencing methods to fresh *Rhinolophus philippinensis* samples would be a valuable next step in understanding the genetic processes underlying this divergence. Separately or in combination with these methods it would be valuable to carry out RNASeq analyses to look for evidence of difference in gene expression between morphs. Gene expression analysis has revealed that gene expression plays a role in humans in the susceptibility of complex traits and diseases (Mancuso et al 2017). Gene dosage has been linked to body size in *Drosophila*, where a gene located on the X-chromosome provides a double dose in females and explains sexual size dimorphism (Mathews, Cavegn and Zwicky 2017). Any of these genetic approaches would require fresh *R. philippinensis* issue collected and stored under optimal conditions to allow extraction of RNA and high-quality DNA. Creating high-quality, chromosome-level assemblies also facilitates a range of other molecular approaches, allowing research into regulatory elements such as epigenetics, small RNAs, control regions and enhancers to see if changes in putative promoter regions near genes led to alterations in their coding or transcription. These novel sequencing technologies represent the most promising directions for future research into this system.

Assortative mating in *Rhinolophus philippinensis* has for the time being only been inferred indirectly from biology in other *Rhinolophus* species and patterns of genetic divergence between morphs. Theoretically, next steps in determining whether this system is in the early stages of an ecological speciation would be to directly study paternity and behaviour in order to integrate behavioural, ecological and evolutionary evidence (Maan

and Seehausen 2011). There are two different approaches that could be taken in this direction. One would be to carry out preference experiments (e.g. as in Puechmaille et al 2014), to determine whether morphs react differently to calls from their own morph and from different morphs. The other is to carry out a long-term mark-release study, taking wing punches from bats and uniquely identifying them (for example by ringing) to determine paternity and relationships within the population (such as has been carried out in horseshoe bats *Rhinolophus ferrumequinum* in the UK; Rossiter et al 2000, Rossiter et al 2006). While these would be promising avenues for research, there are practical barriers to these suggestions. *R. philippinensis* are rare on Buton Island and do not roost in large colonies – in fact, no roost has yet been recorded. This means that they are not easily available for behaviour experiments, and additionally that relationships between individuals become more difficult to determine as we cannot get the mothers and pups attached to one another as in the long-term UK *R. ferrumequinum* study.

The outgroup samples included in the genetic analyses of this system were relatively haphazard, taking advantage of samples which were already available either at the Museum of Bogor or at Queen Mary University of London. To better understand the evolutionary history of *R. philippinensis* in the region, particularly whether the Buton morphs potentially have different source populations, a more systematic sampling process of outgroup populations needs to be undertaken on mainland Sulawesi and on other surrounding islands in Indonesia. *R. philippinensis* is listed as an uncommon species across its range (Sedlock et al 2008), meaning that collecting large sample sizes is challenging even if the necessary research permits are in place.

While there has been a large amount of research into the genetic basis of the evolution of echolocation in bats and toothed whales, this is the first study to attempt to identify genes associated specifically with the evolution of high duty-cycle echolocation in a genome-wide fashion. I have highlighted novel candidate echolocation genes using high-throughput sequencing of coding sequence data in Chapter 2 that establish routes for further research into the molecular basis of this complex phenotypic trait.

Additionally, I have used molecular methods in Chapter 3 and Chapter 4 to investigate a rarely reported case of an incipient ecological speciation event in a mammal. This has opened up future avenues for research into the divergence of *Rhinolophus philippinensis* on Buton Island, which can it turn teach us more about how ecological speciation can progress even with the presence of reproductive isolation.

R. philippinensis is currently categorised as a species of 'Least Concern' in the IUCN red list (Sedlock et al 2008), though with a note that it is a rare to uncommon species across its range and that it represents a species complex that requires taxonomic research to resolve. Research into the taxonomic status of the Buton Island morphs represents a step in this direction. When taken with evidence of distinct morphs of *R. philippinensis* existing in sympatry in Australia and in the Philippines (Flannery 1995, Cooper 1998), this research provides further support that this designated species is in fact a complex of multiple species across its range. Recognising multiple taxonomic divisions within *R. philippinensis* may have implications for its conservation status and actions needed to protect it.

CHAPTER SIX

References

Altringham JD: *Bats: From Evolution to Conservation* (2nd ed). 2011, Oxford Biology.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GB, Sherlock G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25: 25-29.

Baig SM, Koschak A, Lieb A, Gebhart M, Dafinger C, Nürnberg G, Ali A, Ahmad I, Sinnegger-Brauns MJ, Brandt N, Engel J, Mangoni ME, Farooq M, Khan HU, Nürnberg P, Striessnig J, Bolz HJ. 2011. Loss of $Ca_v1.3$ (CACNA1D) function in a human channelopathy with bradycardia and congenital deafness. Nature Neurosci. 14: 77-84.

Ballard JO, Whitlock MC. 2004. The incomplete natural history of mitochondria. Mol. Ecol. 13: 729-744.

Barker DF, Hostikka SL, Zhou J, Chow LT, Oliphant AR, Gerken SC, Gregory MC, Skrolnick MH, Atkin CL, Tryggvason K. 1990. Identification of mutations of the COL4A5 collagen gene in Alport syndrome. Science 248: 1224-1227.

Bandelt H-J, Forster P, Röhl A. 1999. Median-joining networks for inferring specific phylogenies. Mol. Biol. Evol. 16: 37-48.

Barluenga M, Stölting KN, Salzburger W, Muschick M, Meyer A. 2006. Sympatric speciation in Nicaraguan crater lake cichlid fish. Nature 439, 719–23.

Barrowclough GF, Zink RM. 2009. Funds enough, and time: mtDNA, nuDNA and the discovery of divergence. Mol. Ecol. 18: 2934-2936.

Beaumont MA, Balding DJ. 2004. Identifying adaptive genetic divergence among populations from genome scans. Mol. Ecol. 13: 969-980.

Beaumont MA. 2005. Adaptation and speciation: what can F_{ST} tell us? TREE 20: 435-440.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Royal Stat. Soc. B 57: 289-300.

Bensasson D, Zhang D-X, Hartl DL, Hewitt GM. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. TREE 16: 314-321.

Bentzen P, Wright JM, Bryden LT, Sargent M, Zwanenburg KCT. 1998. Tandem repeat polymorphism and heteroplasmy in the mitochondrial control region of redfishes (*Sebastes*: Scorpaenidae). J. Hered. 89: 1-7.

Bhattacharya G, Cosgrove D. 2005. Evidence for functional importance of usherin/fibronectin interactions in retinal basement membranes. Biochemistry 44: 11518-11524.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120.

Booker BM, Friedrich T, Mason MK, VanderMeer JE, Zhao H, Eckalbar WL, Logan M, Illing N, Pollard KS, Ahituv N. 2016. Bat accelerated regions identify a bat forelimb

specific enhancer in the *HoxD* locus. PLoS Genet. https://doi.org/10.1371/journal.pgen.1005738

Boughman JW. 2002. How sensory drive can promote speciation. TREE 17: 571-577.

Bouwman AC, Daetwyler HD, Chamberlain AJ, Ponce CH, Sargolzaei M, Schenkel FS, Sahana G, Govignon-Gion A, Boitard S, Dolezal M, Pausch H, Brøndum RF, Bowman PJ, Thomsen B, Guldbrandsten B, Lund MS, Servin B, Garrick DJ, Reecy J, Vilkki J, Bagnato A, Wang M, Hoff JL, Schnabel RD, Taylor JF, Vinkhuyzen AAE, Panitz F, Bendixen C, Holm L-E, Gredler B, Hozé C, Boussaha M, Sanchez M-P, Rocha D, Capitan A, Tribout T, Barbat A, Croiseau P, Drögemüller C, Jagannathan V, Jagt CV, Crowley JJ, Bieber A, Purfield DC, Berry DP, Emmerling R, Götz K-U, Frischknecht M, Russ I, Sölkner J, Van Tassell CP, Fries R, Stothard P, Veerkamp RF, Boichard D, Goddard ME, Hayes BJ. 2018. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. Nat. Genet. 50: 362-367.

Burton RS, Barreto FS. 2012. A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? Mol. Ecol. 21: 4942-4957.

Bush GL. 1969. Sympatric host race formation and speciation in frugivorous flies of the genus *Rhagoletis* (Diptera: Tephritidae). Evolution 23: 237-251.

Campbell CR, Poelstra JW, Yoder AD. 2018. What is speciation genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. Bio. J. Linn. Soc. Bly063, https://doi.org/10.1093/biolinnean/bly063

Capaccio P, Cuccarini V, Ottaviani F, Fracchiolla NS, Bossi A, Pignataro L. 2009. Prothrombotic gene mutations in patients with sudden sensorineural hearing loss and cardiovascular thrombotic disease. Ann. Otol. Rhinol. Laryngol. 118: 205-210.

Carvajal-Rodriguez A, Uña-Alvarez Jd. 2011. Assessing significance in high-throughput experiments by sequential goodness of fit and q-value estimation. PLoS One 6(9): e24700.

Chabrol O, Royer-Carenzi M, Pontarotti P, Didier G. 2017. Detecting molecular basis of phenotypic convergence. bioRxiv 137174; doi: https://doi.org/10.1101/137174

Chaisson MJP, Huddleston J, Dennis MY, Sudmant PH, Malig M, Hormozdiari F, Antonacci F, Surti U, Sandstrom R, Boitano M, Landolin JM, Stamatoyannopoulos JA, Hunkapiller MW, Korlach J, Eichler EE. 2014. Resolving the complexity of the human genome using single-molecule sequencing. Nature 517: 608-611.

Chatzispyrou IA, Alders M, Guerrero-Castillo S, Zapata Perez R, Haagmans MA, Mouchiroud L, Koster J, Ofman R, Baas F, Waterham HR, Spelbrink JN, Auwerx J, Mannens MM, Houtkooper RH, Plomp AS. 2017. A homozygous missense mutation in ERAL1, encoding a mitochondrial rRNA chaperone, causes Perrault syndrome. Hum. Mol. Gen. 26: 2541-2550.

Chen J, Chu H, Xiong H, Yu Y, Huang X, Zhou L, Chen Q, Bing D, Liu Y, Wang S, Cui Y. 2013. Downregulation of Cav1.3 calcium channel expression in the cochlea is associated with age-realted hearing loss in C57BL/6J mice. Neuroreport 24: 13-17.

Chen S-F, Jones G, Rossiter S. 2009. Determinants of echolocation call frequency variation in the Formosan lesser horseshoe bat (*Rhinolophus monoceros*). Proc. R. Soc. B. 276: 3901-3909.

Chikhi R, Medvedev P. 2013. Informed and automated *k*-mer size selection for genome assembly. Bioinformatics 30: 31-37.

Churchill SK. 2009. Australian bats. Second Edition. Allen and Unwin, Sydney.

Clare EL, Adams AM, Maya-Simoes AZ, Eger JL, Hebert PDN, Fenton MB. 2013. Diversification and reproductive isolation: cryptic species in the only New World high duty-cycle bat, *Pteronotus parnellii*. BMC Evol. Biol. 13:26.

Cooper SJB, Reardon TB, Skilins J. 1998. Molecular systematics of Australian rhinolophid bats (Chiroptera: Rhinolophidae). Aust. J. Zool. 46:203-20.

Corcoran A, Weller TJ. 2018. Inconspicuous echolocation in hoary bats (*Lasiurus cinereus*). Proc. R. Soc. B. 285. 10.1098/rspb.2018.0441.

Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. 23:3133-57.

Csorba G, Ujhelyi P, Thomas N: *Horseshoe Bats of the World (Chiroptera: Rhinolophidae).* 2003, Alana Books.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker R, Lunter G, Marth G, Sherry ST, McVean G, Durbin R, 1000 Genome Project Analysis Group. 2011. Bioinformatics 27: 2156-2158.

Davies KTJ, Cotton JA, Kirwan JD, Teeling EC, Rossiter SJ. 2012. Parallel signatures of sequence evolution among hearing genes in echolocating mammals: an emerging model of genetic convergence. Heredity. 108: 480-489.

Davies KTJ, Bates PJJ, Maryanto I, Cotton JA, Rossiter SJ. 2013a. The evolution of bat vestibular systems in the face of potential antagonistic selection pressures for flight and echolocation. PLoS One 8:e61998.

Davies KTJ, Maryanto I, Rossiter SJ. 2013b. Evolutionary origins of ultrasonic hearing and laryngeal echolocation in bats inferred from morphological analyses of the inner ear. Front Zool. 10: 2.

Deane-Coe P, Butcher BG, Greenberg R, Lovette IJ. 2018. Whole genome scan reveals the multigenic basis of recent tidal marsh adaptation in a sparrow. bioRxiv 360008; doi: https://doi.org/10.1101/360008.

Dechmann DKN, Wikelski M, van Noordwijk HJ, Voigt CC, Voigt-Heucke SL. 2013. Metabolic costs of bat echolocation in a non-foraging context support a role in communication. Front. Physiol. 4: 66.

Derryberry EP, Seddon N, Derryberry GE, Claramunt S, Seeholzer GF, Brumfield RT, Tobias JA. 2018. Ecological drivers of song evolution in birds: disentangling the effects of habitat and morphology. Ecol. Evol. 8: 1890-1905.

DePristo M, Banks E, Poplin R, Garimella K, Maguire J, Hartl C, Philippakis A, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell T, Kernytsky A, Sivachenko A, Cibulskis K, Gabriel S, Altchuler D, Daly M. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43: 491-498.

Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. Nuc. Acid. Res. 45: e18.

Dong D, Lei M, Liu Y, Zhang S. 2013. Comparative inner ear transcriptome analysis between the Rickett's big-footed bats (*Myotis ricketti*) and the greater short-nosed fruit bats (*Cynopterus sphinx*). BMC Genomics 14: 916.

Dong D, Lei M, Hua P, Pan Y-H, Mu S, Zheng G, Pang E, Lin K, Zhang S. 2017. The genomes of two bat species with long constant frequency echolocation calls. Mol Biol Evol. 34: 20-34.

Dong J, Mao X, Sun H, Irwin DM, Zhang S, Hua P. 2014. Introgression of mitochondrial DNA promoted by natural selection in the Japanese pipistrelle bat (*Pipistrellus abramus*). Genetica 142: 484-494.

Dool SE, Puechmaille SJ, Foley NM, Allegrini B, Bastian A, Mutumi GL, Maluleke TG, Odendaal LJ, Teeling EC, Jacobs DS. 2016. Nuclear introns outperform mitochondrial DNA in interspecific phylogenetic reconstruction: Lessons from horseshoe bats (Rhinolophidae: Chiroptera). Mol. Phylo. Evol. 97: 196-212.

Dowling DK, Friberg U, Lindell J. 2008. Evolutionary implication of non-neutral mitochondrial genetic variation. TREE 23: 546-554.

Egan SP, Ragland GJ, Assour L, Powell THQ, Hood GR, Emrich S, Nosil P, Feder JL. 2015. Experimental evidence of genome-wide impact of ecological speciation during early stages of speciation-with-gene-flow. Ecol. Lett. 18: 817-825.

Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, Künsten A, Mäkinen H, Nadachowska-Brzyska K, Qvarnström A, Uebbing S, Wolf JBW. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. Nature 491: 756-760.

Elmer KR, Meyer A. 2011. Adaptation in the age of ecological genetics: insights from parallelism and convergence. TREE. 26: 298-306.

Fang J, Wang X, Mu S, Zhang S, Dong D. 2015. BGD: A database of bat genomes. PLoS One 10(6): e0131296. doi:10.1371/journal. pone.0131296

Farrington HL, Lawson LP, Clark CM, Petren K. 2014. The evolutionary history of Darwin's finches: speciation, gene flow and introgression in a fragmented landscape. Evolution 68: 2932-2944.

Feder JL, Egan SP, Nosil P. 2012. The genomics of speciation-with-gene-flow. TIG 28: 342-350.

Felsenstein J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals? Evolution 35: 124-138.

Fenton MB, Faure PA, Ratcliffe JM. 2012. Evolution of high duty cycle echolocation in bats. J. Exp. Biol. 215: 2935-2944.

Feller AF, Seehausen O, Lucek K, Marques DA. 2016. Habitat choice and female preference in a polymorphic stickleback population. Evol. Ecol. Res. 17: 419-435.

Feulner PGD, Plath M, Engelmann J, Kirschbaum F, Tiedemann R. 2009a. Electrifying love: electric fish use species-specific discharge for mate recognition. Biol. Let. 5: 225-228.

Feulner PGD, Plath M, Engelmann J, Kirschbaum F, Tiedemann R. 2009b. Magic trait electric organ discharge (EOD). Commun. Integr. Biol. 2: 329-331.

Feulner PGD, Chain FJJ, Panchal M, Huang Y, Eizaguirre C, Kalbe M, Lenz TL, Samonte IE, Stoll M, Bornberg-Bauer E, Reusch TBH, Milinski M. 2015. Genomics of divergence along a continuum of parapatric population differentiation. PLoS Genet. 11(7): e1005414.

Filchak KE, Roethele JB, and Feder JL 2000. Natural selection and sympatric divergence in the apple maggot *Rhagoletis pomonella*. Nature 407: 739–42.

Finger NM. 2015. Behavioural evidence for the perception of individual identity and gender via the echolocation calls of a high duty cycle bat, *Rhinolophus clivosus*. MSc Biological Sciences, University of Cape Town, Cape Town.

Fitzpatrick BM, Shaffer HB. 2004. Environment-dependent admixture dynamics in a tiger salamander hybrid zone. Evolution 58: 1282-1293.

Flannery T. 1995. Mammals of the South-West Pacific & Moluccan Islands. Australian Museum/Reed Books, pp 361.

Foley NM, Thong VD, Soisook P, Goodman SM, Armstrong KN, Jacobs DS, Peuchmaille SJ, Teeling EC. 2015. How and why overcome the impediments to resolution: lessons from rhinolophid and hipposiderid bats. Mol. Biol. Evol. 32: 313-333.

Foote AD. 2018. Sympatric speciation in the genomic era. TREE 33: 85-95.

Funk DJ, Omland KE. 2003. Species-level paraphyly and polyphyly: frequency, causes and consequences, with insights from animal mitochondrial DNA. Annu. Rev. Ecol. Evol. Syst. 34: 397-423.

Futuyama DJ, Mayer GC. 1980. Non-allopatric speciation in animals. Syst. Biol. 29: 245-271.

Galtier N, Nabholz B, Glémin S, Hurst GDD. 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Mol. Ecol. 18: 4541-4550.

Garrison E, Gabor M. 2012. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv: 1207.3907 [q-bio.GN]

Gillam E, Brock Fenton M. 2016. Roles of acoustic social communication in the lives of bats. In: Bat bioacoustics. Eds Brock Fenton M, Grinnell AD, Popper AN, Fay RR. 117-139.

González-Rodríguez A, Aria DM, Valencia S, Oyama K. 2004. Morphological and RAPD analysis of hybridization between *Quercus affinis* and *Q. laurina* (Fagaceae), two Mexican red oaks. Am. J. Bot. 91: 401-409.

Good JM, Vanderpool D, Keeble S, Bi K. 2015. Negligible nuclear introgression despite complete mitochondrial capture between two species of chipmunks. Evolution 69: 1961-1972.

Goodwin S, McPherson JD, McCombie WR. 2016. Coming of age: ten years of next-generation sequencing technologies. Nat. Rev. Genet. 17: 333-351.

Gordon D, Huddleston J, Chaisson MJP, Hill CM, Kronenberg ZN, Munson KM, Malig M, Raja A, Fiddes I, Hillier LW, Dunn C, Baker C, Armstrong J, Diekhans M, Paten B, Shendure J, Wilson RK, Haussler D, Chin C-S, Eichler EE. 2016. Long-read sequence assembly of the gorilla genome. Science 352: DOI: 10.1126/science.aae0344

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-

length transcriptome assembly from RNA-seq data without a reference genome. Nat. Biotechnol. 28: 644-652.

Grant PR, Grant BR: How and why species multiply: the radiation of Darwin's finches. Third printing. 2011. Princeton University Press.

Guillén A, Juste BJ, Ibáñez C. 2000. Variation in the frequency of the echolocation calls of *Hipposideros ruber* in the Gulf of Guinea: an exploration of the adaptive meaning of the constant frequency value in rhinolophoid CF bats. J. Evol. Biol. 13: 70-80.

Gutiérrez EE, Molinari J. 2008. Morphometrics and taxonomy of bats in the genus *Pteronotus* (subgenus *Phyllodia*) in Venezuela. J. Mammal. 89: 292-305.

Gutierrez EA, Castiglione GM, Morrow JM, Schott RK, Loureiro LO, Lim BK, Chang BSW. 2018a. Functional shifts in bat dim-light visual pigment are associated with differing echolocation abilities and reveal molecular adaptation to photic-limited environments. Mol. Biol. Evol. msy140, https://doi.org/10.1093/molbev/msy140

Gutierrez EA, Schott RK, Preston MW, Loureiro LO, Lim BK, Chang BSW. 2018b. The role of ecological factors in shaping bat cone opsin evolution. Proc. Royal. Soc. B. 285: 20172835.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8: 1494-1512.

Haenel GJ. 2017. Introgression of mtDNA in *Urosaurus* lizards: historical and ecological processes. Mol. Ecol. 26: 606-623.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41: 95-98.

Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F, Bergelson J. 2011. Adaptation to climate across the *Arabidopsis thaliana* genome. Nature 334:83-6.

Harr B. 2006. Genomic islands of differentiation between house mouse subspecies. Genome Res. 16: 730-737.

Hartley DJ. 1989. The effect of atmospheric sound absorption on signal bandwidth and energy and some consequences for bat echolocation. J. Acoust. Soc. Am. 85: 1338-1347.

Hayward JJ, Castelhano MG, Oliveira KC, Corey E, Balkman C, Baxter TL, Casal ML, Center SA, Fang M, Garrison SJ, Kall SE, Korniliev P, Kotlikoff MI, Moise NS, Shannon LM, Simpson KW, Sutter NB, Todhunter RJ, Boyko AR. 2016. Complex disease and phenotype mapping in the domestic dog. Nat. Commun. 7: 10460.

He DZZ, Lovas S, Ai Y, Li Y, Beisel KW. 2014. Prestin at year 14: Progress and prospect. Hear. Res. 311: 25-35.

Hechavarría JC, Marcías S, Vater M, Mora EC, Kössel M. 2013. Evolution of neuronal mechanisms for echolocation: Specializations for target-range computation in bats of the genus *Pteronotus*. J Acoust Soc Am. 133: 570-578.

Heller K-G, v. Helversen O. 1989. Resource partitioning of sonar frequency in bands in rhinolophid bats. Oecologia 80: 178-186.

Henson Jr OW. 1965. The activity and function of the middle-ear muscles in echo-locating bats. J. Physiol. 180: 871-887.

Henson Jr OW, Schuller G, Vater M. 1985. A comparative study of the physiological properties of the inner ear in Doppler shift compensation bats (*Rhinolophus rouxi* and *Pteronotus parnellii*). J Comp Physiol. 157: 587-597.

Hernández MA, Campos F, Gutiérrez-Corchero F, Amezcua A. 2003. Identification of *Lanius* species and subspecies using tandem repeats in the mitochondrial DNA control region. Ibis 146: 227-230.

Hertzano R, Montcouquiol M, Rashi-Elkeles S, Elkon R, Yücel R, Frankel WN, Rechavi G, Möröy T, Friedman TB, Kelley MW, Avraham KB. 2004. Transcription profiling of inner ears from *Pou4f3*$^{ddl/ddl}$ identifies Gfi1 as a target for the Pou4f3 deafness gene. Hum. Mol. Genet. 13: 2143-2153.

Hoekstra LA, Siddig MA, Montooth KL. 2013. Pleiotropic effects of a mitochondrial-nuclear incompatibility depend on the accelerating effect of temperature in *Drosophila*. Genetics 195: 1129-1139.

Holderied MW, von Helversen O. 2003. Echolocation range and wingbeat period match in aerial-hawking bats. Proc. R. Soc. Lond. B. 270: 2293–2299.

Houston RD, Boorman AM, Jones G. Do echolocation signal parameters restrict bats' choice of prey? In *Echolocation in Bats and Dolphins* (eds Thomas JA, Moss CF, Vater M) pp. 339-44. 2004, Univ. of Chicago Press.

Huber SK, Podos J. 2006. Beak morphology and song features covary in a population of Darwin's finches (*Geospiza fortis*). Bio. J. Linn. Soc. 88: 489-498.

Huber SK, De León LF, Hendry AP, Bermingham E, Podos J. 2007. Reproductive isolation of sympatric morphs in a population of Darwin's finches. Proc. Royal Soc. Lond. B. 274: 1709-1714.

Huddleston J, Ranade S, Malig M, Antonacci F, Chaisson M, Hon L, Sudmant PH, Graves TA, Alkan C, Dennis MY, Wilson RK, Turner SW, Korlach J, Eichler EE. 2014. Reconstructing complex regions of genomes using long-read sequencing technology. Genome Res. 24: 688-696.

Hudson NJ, Baker ML, Hart NS, Wynne JW, Gu Q, Huang Z, Zhang G, Ingham AB, Wang L, Reverter A. 2014. Sensory rewiring in an echolocator: genome-wide modification of retinogenic and auditory genes in the bat *Myotis davidii*. G3 4: 1825-1835.

Hudson RR, Kreitman M, Aguade M. 1987. A test of molecular evolution based on nucleotide data. Genetics 116: 153-159.

Huihua Z, Shuyi Z, Mingxue Z, Jiang Z. 2006. Correlations between call frequency and ear length in bats belonging to the families Rhinolophidae and Hipposideridae. J. Zool. 259: 189-195.

Hvala JA, Frayer ME, Payseur BA. 2018. Signatures of hybridization and speciation in genomic patters of ancestry. Evolution doi:10.1111/evo.13509.

Jacobs DS, Catto S, Mutumi GL, Finger N, Webala PW. 2017. Testing the Sensory Drive hypothesis: Geographic variation in echolocation frequencies of Geoffroy's horseshoe bat (Rhinolophidae: *Rhinolophus clivosus*). PLoS One: 12: e0187769.

Jacquier A, Delorme C, Belotti E, Juntas-Morales R, Solé G, Dubourg O, Giroux M, Maurage CA, Castellani V, Rebelo A, Abrams A, Züchner S, Stojkovic T, schaeffer L, Latour P. 2017. Cryptic amyloidogenic elemnts in mutant NEFH causing Charcot-Marie-Tooth 2 trigger aggresome formation and neuronal death. Acta Neuropathol. Commun. 5: 55.

Jen PH-S, Kamuda T. 1982. Doppler-shift compensation behavior of *Pteronotus parnellii parnellii* during avoidance of stationary and moving obstacles. J. Acoust. Soc. Am. 71:S50.

Jones FC, Chan YF, Schmutz J, Grimwood J, Brady SD, Southwick AM, Absher DM, Myers RM, Reimchen TE, Deagle BE, Schluter D, Kingsley DM. 2012. A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. Curr. Biol. 22: 83-90.

Jones G. 1999. Scaling of echolocation call parameters in bats. Exp. Biol. 202: 3359-3367.

Jones G, Water DA. 2000. Moth hearing in response to bat echolocation calls manipulated independently in time and frequency. Proc. R. Soc. Lond. B. 267: 1627-1632.

Jones G, Teeling EC. 2006. The evolution of echolocation in bats. Trends Ecol. Evol. 21:149–56.

Jones G, Siemers BM. 2011. The communicative potential of bat echolocating pulses. J. Comp. Physiol. A 197: 447-57.

Kawata M, Shoji A, Kawamura S, Seehausen O. 2007. A genetically explicit model of speciation by sensory drive within a continuous population in aquatic environments. BMC Evol. Biol. 7: 99.

Kazmierczak P, Muller U. 2012. Sensing sound: molecules that orchestrate mecahnotransduction by hair cells. Trends Neuroci. 35: 220-229.

Keithley EM, Ryan AF, Woolf NK. 1993. Fibronectin-like immunoreactivity of the basilar membrane of young and aged rats. J. Comp. Neurol. 327: 612-617.

Kemper KE, Visscher PM, Goddard ME. 2012. Genetic architecture of body size in animals. Genome Biol. 13: 244.

Kingston T, Jones G, Zubaid A, Kunz TH. 2000. Resource partitioning in rhinolophid bats revisited. Oecologica 124: 332-342.

Kingston T, Lara MC, Jones G, Akbar Z, Kunz TH, Schneider CJ. 2001. Acoustic divergence in two cryptic Hipposideros species: a role for social selection? Proc. Biol. Sci. 268: 1381-1386.

Kingston T, Rossiter SJ. 2004. Harmonic-hopping in Wallacea's bats. Nature 429:654–7.

Kirwan JD, Bekaert M, Commins JM, Davies KT, Rossier SJ, Teeling EC. 2013. A phylomedicine approach to understanding the evolution of auditory sensory perception and disease in mammals. Evol. Appl. 6: 412-422.

Knörnschild M, Jung K, Nagy M, Metz M, Kalko E. 2012. Bat echolocation calls facilitate social communication. Proc. Biol. Sci. 279: 4827-4835.

Konings A, Van Laer L, Pawelczyk M, Carlsson PI, Bondeson ML, Rajkowska E, Dudarewicz A, Vandevelde A, Fransen E, Huyghe J, Borg E, Sliwinska-Kowalska M,

Van Camp G. 2007. Association between variation in CAT and noise-induced hearing loss in two independent noise-exposed populations. Hum. Mol. Genet. 16: 1872-1883.

Kopp M, Servedio MR, Mendelson TC, Safran RJ, Rodríguez RL, Hauber ME, Scordato EC, Symes LB, Balakrishnan CN, Zonana DM, van Doorn GS. 2018. Mechanisms of assortative mating in speciation with gene flow: connecting theory and empirical research. Am. Nat. 191: 1-20.

Kössl M, Vater M. 1985. The cochlear frequency map of the mustache bat, *Pteronotus parnellii*. J. Comp. Physiol. A. 157:687–97.

Lack D. Darwin's Finches, 1983 reissue. 1983. Cambridge University Press.

Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipralla S, Gu B, Hart J, Hoffman D, Jang W, Karapetyan K, Katz K, Liu C, Maddipalta Z, Malheiro A, McDaniel K, Ovetsky M, Riley G, Zhou G, Holmes JB, Kattman BL, Maglott DR. 2018. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 46: D1062-D1067.

Langin KM, Sillett TS, Morrison SA, Ghalambor CK. 2017. Bill morphology and neutral genetic structure both predict variation in acoustic signals within a bird population. Behav. Eco. 28: 866-873.

Larson WA, Limborg MT, McKinney GJ, Schindler DE, Seeb JE, Seeb LW. 2017. Genomic islands of divergence linked to ecotypic variation in sockeye salmon. Mol. Ecol. 26: 554-470.

Lawrence BD, Simmons JA. 1982. Measurements of atmospheric attenuation at ultrasonic frequencies and the significance for echolocation by bats. J. Acoust. Soc. Am. 71:585–90.

Layman WS, Zuo J. 2015. Preventing ototoxic hearing loss by inhibiting histone deacetylases. Cell Death Dis. 6, e1882; doi: 10.1038/cddis.2015.252.

Lazure L, Fenton MB. 2011. High duty cycle echolocation and prey detection in bats. J. Exp. Biol. 214:1131-7.

Lei M, Dong D. 2016. Phylogenomic analyses of bat subordinal relationships based on transcriptome data. Sci. Rep. 6: 27726.

Leitmeyer K, Glutz A, Radojevic V, Setz C, Huerzeler N, Bumann H, Bodmer D, Brand Y. 2015. Inhibition of mTOR by rapamycin results in auditory hair cell damage and decreased spinal ganglion neuron outgrowth and neurite formation *in vitro*. BioMed Res. Int. Article ID 924890.

Leurez S, Milea D, Defoort-Dhellemmes S, Colin E, Crochet M, Procaccio V, Ferré M, Lamblin J, Drouin V, Vincent-Delorme C, Lenaers G, Hamel C, Blanchet C, Juul G, Larsen M, Verny C, Reynier P, Amati-Bonneau P, Bonneau D. 2013. Sensorineural hearing loss in OPA1-linked disorders. Brain. 136: e236.

Li G, Wang J, Rossiter SJ, Jones G, Zhang S. 2007. Accelerated FoxP2 evolution in echolocating bats. PLoS One 2: e900.

Li G, Wang J, Rossiter SJ, Jones G, Cotton JA, Zhang S. 2008. The hearing gene Prestin reunites echolocating bats. Proc Natl Acad Sci. 105: 13959–13964.

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

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009b. The Sequence/Alignment Map format and SAMtools. Bioinformatics 25: 2078-2079.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26: 589-595.

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27: 2987-1993.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv: 1303.3997v1 [q-bio.GN]

Li Y, Lacerda DA, Warman ML, Beier DR, Yoshioka H, Ninomiya Y, Oxford JT, Morris NP, Andrikopoulos K, Ramirez F, Wardell BB, Lifferth GD, Teuscher C, Woodward SR, Taylor BA, Seegmiller RE, Olsen BR. 1995. A fibrillar collagen gene, *Col11a1*, is essential for skeletal morphogenesis. Cell 80: 423-430.

Li Y, Wang J, Metzner W, Luo B, Jiang T, Yang S, Shi L, Huang X, Yue X, Feng J. 2014. Behavioural responses to echolocation calls from sympatric heterospecific bats: Implications for interspecific competition. Behav. Ecol. Sociobiol. 68:657-67.

Lin A, Liu H, Chang Y, Lu G, Feng J. 2016. Behavioural response of the greater horseshoe bat to geographical variation in echolocation calls. Behav. Ecol. Sociobiol. 70: 1765-1776.

Liu H-Q, Wei J-K, Li B, Wang M-S, Wu R-Q, Rizak JD, Zhong L, Wang L, Xu F-Q, Shen Y-Y, Hu X-T, Zhang Y-P. 2015. Divergence of dim-light vision among bats (order: Chiroptera) as estimated by molecular and electrophysical methods. Sci. Rep. 5: 11531.

Liu Y, Han N, Franchini LF, Xu H, Pisciottano F, Elgoyhen AB, Rajan KE, Zhang S. 2012. The voltage-gated potassium channel subfamily KQT member 4 (KCNQ4) displays parallel evolution in echolocating bats. Mol. Biol. Evol. 29: 1441-1450.

Long GR, Schnitzler H-U. 1975. Behavioural audiograms from the bat, *Rhinolophus ferrumequinum*. J. Comp. Physiol. 100: 211-219.

Löytynoja A, Goldman N. 2008. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. Science 320: 1632-1635.

Lundberg M, Boss J, Canback B, Liedvogel M, Larson KW, Grahn M, Akesson S, Bensch S, Wright A: Characterisation of a transcriptome to find sequence differences between two differentially migrating subspecies of the willow warbler *Phylloscopus trochilus*. *BMC Genomics* 2013, 14:330.

Luo R, Liu B, Xio Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhange H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu S-M, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam T-W, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. Gigascience 1: 18.

Ma J, Kobayasi K, Zhang S, Metzner W. 2006. Vocal communication in adult greater horseshoe bats, *Rhinolophus ferrumequinum*. J. Comp. Physiol. A. 192:535-50.

Maan ME, Seehausen O. 2011. Ecology, sexual selection and speciation. Ecol. Lett. 14: 591-602.

Madoui M-A, Engelen S, Cruaud C, Belser C, Bertrand L, Alberti A, Lemainque A, Winckler P, Aury J-M. 2015. Genome assembly using Nanopore-guided long and error-free DNA reads. BMC Genomics 16: 327.

Maharadatunkamsi S, Hisheh, Kitchener DJ, Schmitt LH. 2000. Genetic and morphometric diversity in Wallacea: geographical pattering in the horseshoe bat, *Rhinolophus affinis*. J. Biogeogr. 27: 193-201.

Makvandi-Nejad S, Hoffman GE, Allen JJ, Chu E, Gu E, Chandler AM, Loredo AI, Bellone RR, Mezey JG, Brooks SA, Sutter NB. 2012. Four loci explain 83% of size variation in the horse. PLoS One 7(7): e39929.

Mallick S, Gnerre S, Muller P, Reich D. 2009. The difficulty of avoiding false positives in genome scans for natural selection. Genome Res. 19: 922-933.

Malueke T, Jacobs DS, Winker H. 2017. Environmental correlates of geographic divergence in a phenotypic trait: A case study using bat echolocation. Ecol. Evol. 7: 7347-7361.

Mancuso N, Shi H, Goddard P, Kichaev G, Gusev A, Pasaniuc B. 2017. Integrating gene expression with summary association statistics to identify genes associated with 30 complex traits. AJHG 100: 473-487.

Mao X, Zhang J, Zhang S, Rossiter SJ. 2010a. Historical male-mediated introgression in horseshoe bats revealed by multilocus DNA sequence data. Mol. Ecol. 19: 1352-1366.

Mao X, Zhu GJ, Zhang S, Rossiter SJ. 2010b. Pleistocene climatic cycling drives intra-specific diversification in the intermediate horseshoe bat (*Rhinolophus affinis*) in Southern China. Mol. Ecol. 19: 2754-2769.

Mao X, He G, Zhang J, Rossiter SJ, Zhang S. 2013a. Lineage divergence and historical gene flow in the Chinese horseshoe bat (*Rhinolophus sinicus*). PLoS One: https://doi.org/10.1371/journal.pone.0056786

Mao X, Thong VD, Bates PJJ, Jones G, Zhang S, Rossiter SJ. 2013b. Multiple cases of asymmetric introgression among horseshoe bats detected by phylogenetic conflicts across loci. Bio. J. Linn. Soc. 110: 346-361.

Mao X, Dong J, Hua P, He G, Zhang S, Rossiter S. 2014. Heteroplasmy and ancient translocation of mitochondrial DNA to the nucleus in the Chinese horseshoe bat (*Rhinolophus sinicus*) complex. PLoS One: https://doi.org/10.1371/journal.pone.0098035.

Marques DA, Lucek K, Meier JI, Mwaiko S, Wagner CE, Excoffier L, Seehausen O. 2016. Genomics of rapid incipient speciation in sympatric treespine stickleback. PLoS Genet. https://doi.org/10.1371/journal.pgen.1005887

Marques DA, Lucek K, Haesler MP, Feller AF, Meier JI, Wagner CE, Excoffier L, Seehausen O. 2017. Genomic landscape of early ecological speciation initiated by selection on nuptial colour. Mol. Ecol. 26: 7-24.

Martin CH, Cutler JS, Friel JP, Dening Touokong C, Coop G, Wainwright PC. 2015. Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation. Evolution 69: 1406–1422.

Mathews KW, Cavegn M, Zwicky M. 2017. Sexual dimorphism of body size is controlled by dosage of the X-chromosomal gene *Myc* and by the sex-determining gene *tra* in *Drosophila*. Genetics 205: 1215-1228.

Maynard Smith J. 1966. Sympatric speciation. Am. Nat. 100: 637-650.

Mayr E. *Animal species and evolution*. 1963. Belknap Press.

McCarroll SA, Altshuler DM. 2007. Copy-number variation and association studies of human disease. Nat. Genet. 39: S37-S42.

McGowen MR, Grossman LI, Wildman DE. 2012. Dolphin genome provides evidence for adaptive evolution of nervous system genes and a molecular rate slowdown. *Proc. R. Soc. B.* 279: 3643-3651.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altschuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analysing next-generation DNA sequencing data. Genome Res. 20: 1297-1303.

McKinnon JS, Rundle HD. 2002, Speciation in nature: the threespine stickleback model systems. Trends Ecol. Evol. 17: 480–488.

McPheron BA, Smith DC, Berlocher SH. 1988. Genetic differences between host races of *Rhangoletis pomonella*. Nature 336: 64-66.

Mendoza MLZ, Xiong Z, Escalera-Zamudio M, Runge AK, Thézé J, Streicker D, Frank HK, Loza-Rubio E, Liu S, Ryder OA, Castruita JAS, Katzourakis A, Pacheco G, Taboada B, Löber U, Pybus OG, Li Y, Rojas-Anaya E, Bohmann K, Baez AC, Arias CF, Liu S, Greenwood AD, Bertelsen MF, White NE, Bunce M, Zhang G, Sicheritz-Pontén T, Gilbert MPT. 2018. Hologenomic adaptations underlying the evolution of sanguivory in the common vampire bat. Nat. Ecol. Evol. doi:10.1038/s41559-018-0476-8

Metzner W, Zhang S, Smotherman M. 2002. Doppler-shift compensation behavior in horseshoe bats revisited: auditory feedback controls both a decrease and an increase in call frequency. J. Exp. Biol. 205:1607–16.

Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD. 2016. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucl. Acids Res. 45: D183-D189.

Mirkin SM. 2007. Expandable DNA repeats and human disease. Nature 447: 932-940.

Moore WS. 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution 49: 718-726.

Mora EC, Macías S, Hechavarría J, Vater M, Kössl M. 2013. Evolution of the heteroharmonic strategy for target-range computation in the echolocation of Mormoopidae. Front. Physiol. 4:141.

Morales AE, Jackson ND, Dewey TA, O'Meara BC, Carstens BC. 2017. Speciation with gene flow in North American *Myotis* bats. Syst. Biol. 66: 440-452.

Mořkovský L, Janoušek V, Reif J, Rídl J, Pačes J, Choleva L, Janko K, Nachman MW, Reifová R. 2018. Genomic islands of differentiation in two songbird species reveal candidate genes for hybrid female sterility. Mol. Ecol. 27: 949-958.

Mundy NI, Winchell CS, Woodruff DS. 1996. Tandem repeats and heteroplasmy in the mitochondrial DNA control region of the loggerhead shrike (*Lanius ludovicianus*). J. Hered. 87: 21-26.

Mustapha M, Chouery É, Torchard-Pagnez D, Nouaille S, Khrais A, Sayegh FN, Mégarbané A, Loiselet J, Lathrop M, Petit C, Weil D. 2002. A novel locus for Usher syndrome type I, USH1G, maps to chromosome 17q24-25. Hum Genet. 110: 348-350.

Nam K, Lee KW, Chung O, Yim H-S, Cha S-S, Lee S-W, Jun J, Cho Y-S, Bhak J, de Magalhães JP, Lee J-H, Jeong J-Y. 2017. Analysis of the FGF gene family provides insights into aquatic adaptation in cetaceans. Sci. Rep. 7: 40233.

Nei M. 1987. Molecular evolutionary genetics. Columbia Univ. Press, New York.

Neumann I, Schuller G. 1991. Spectral and temporal gating mechanisms enhance the clutter rejection in the echolocating bat, *Rhinolophus rouxi*. J. Comp. Physiol. A. 169: 106-116.

Nery MF, González DJ, Opazo JC. 2013. How to make a dolphin: molecular signature of positive selection in cetacean genome. PLoS One: https://doi.org/10.1371/journal.pone.0065491

Noor MAF, Bennett SM. 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintain species. Heredity 103: 439 - 444.

Nosil P. 2008. Speciation with gene flow could be common. Mol. Ecol. 17: 2103-2106.

Nosil P, Funk DJ, Ortiz-Barrientos D. 2009. Divergent selection and heterogenous genomic divergence. Mol. Ecol. 18: 375-402.

Nosil P. *Ecological Speciation*. 2012, Oxford University Press.

Nowak RM, Paradiso JL. 1999. Walker's mammals of the world. Baltimore and London: John Hopkins University Press.

Odendaal LJ, Jacobs DS, Bishop JM. 2014. Sensory trait variation in an echolocating bat suggests roles for both selection and plasticity. BMC Evol Biol. 14: 60.

Oliveira EF, Gehara M, São-Pedro VA, Chen X, Myers EA, Burbrink FT, Mesquita DO, Garda AA, Colli GR, Rodrigues MT, Arias FJ, Zaher H, Santos RML, Costa GC. 2015. Speciation with gene flow in whiptail lizards from a Neotropical xeric biome. Mol. Ecol. 24: 5957-5975.

Omote K, Nishida C, Dick MH, Masuda R. 2013. Limited phylogenetic distribution of a long tandem-repeat cluster in the mitochondrial region in *Bubo* (Aves, Strigidae) and cluster variation in Blakiston's fish owl (*Bubo blakistoni*). Mol. Phylogenetics Evol. 66: 889-897.

O'Neill WE. 1995. The bat auditory cortex. In: Popper AN, Fay RR (eds) Hearing by bats. Springer Handbook of Auditory Research, vol 5. Springer, New York, NY.

Oonk AMM, van Huet RAC, Leijendeckers JM, Oostrik J, Venselaar H, van Wijk E, Beyon A, Kunst HPM, Hoyng CB, Kremer H, Schraders M, Pennings RJE. 2015. Nonsyndromic hearing loss caused by USH1G mutations: widening the USH1G disease spectrum. Ear and Hearing. 36: 205-211.

Padhi A. 2014. Geographic variation within a tandemly repeated mitochondrial DNA D-loop region of a North American freshwater fish, *Pylodictis olivaris*. Gene 538: 63-68.

Palumbi SR, Cipriano F, Hare MP. 2007. Predicting nuclear gene coalescence from mitochondrial data: the three-times rule. Evolution 55: 859-868.

Parker J, Tsagkogeorga G, Cotton JA, Liu Y, Provero P, Stupka E, Rossiter SJ. 2013. Genome-wide signatures of convergent evolution in echolocating mammals. Nature 502: 228–231.

Partha R, Chauhan BK, Ferreira Z, Robinson JD, Lathrop K, Nischal KK, Chikina M, Clark NL. 2017. Subterranean mammals show convergent regression in ocular genes and enhancers, along with adaptation to tunnelling. eLife 6: e25884.

Pauley S, Wright TJ, Pirvola U, Ornitz D, Beisel K, Fritzsch B. 2003. Expression and function of FGF10 in mammalian inner ear development. Dev. Dyn. 227: 203-215.

Pavan AC, Marroig G. 2016. Integrating multiple evidences in taxonomy: species diversity and phylogeny of mustached bats (Mormoopidae: *Pteronotus*). Mol. Phylogenet. Evol. 103: 184-198.

Pavan AC, Marroig G. 2017. Timing and patterns of diversification in the Neotropical bat genus *Pteronotus* (Mormoopidae). Mol. Phylogenet. Evol. 108: 61-69.

Pavey CR, Kutt A. 2008. Large-eared horseshoe bat *Rhinolophus philippinensis*. In *The mammals of Australia.* 3rd edition (Eds S. Van Dyck and R. Strahan). Reed New Holland. Sydney. Pp 454-456.

Penn O, Privman E, Landan G, Graur D, Pupko T. 2010. An alignment confidence score capturing robustness to guide tree uncertainty. Mol Biol Evol. 27: 1759-1767.

Picciotti PM, Fetoni AR, Paludetti G, Wolf FI, Torsello A, Troiani D, Ferraresi A, Pola R, Sergi B. 2006. Vascular endothelial growth factor (VEGF) expression in noise-induced hearing loss. Hear Res. 214: 76-83.

Pierce SB, Gersak K, Michaelson-Cohen R, Walsh T, Lee MK, Malach D, Klevit RE, King M-C, Levy-Lahad E. 2013. Mutations in *LARS2*, encoding mitochondrial leucyl-tRNA synthetase, lead to premature ovarian failure and hearing loss in Perrault syndrome. Am. J. Hum. Genet. 92: 614-620.

Platzer J, Engel J, Schrott-Fischer A, Stephan K, Bova S, Chen H, Zheng H, Striessnig J. 2000. Congenital deafness and sinoatrial node dysfunction in mice lacking class D L-type Ca2+ channels. *Cell* 102: 89-97.

Podos J. 2001. Correlated evolution of morphology and vocal signal structure in Darwin's finces. Nature 409: 185-188.

Pollack G, Henson Jr OW, Novick A. 1972. Cochlear microphonic audiograms in the "pure tone" bat *Chilonyteris parnellii parnellii*. Science. 176: 66-68.

Pond SLK, Frost SDW, Muse SV. 2004. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21: 676-679.

Powell THQ, Forbes AA, Hood GR, Feder JL. 2014. Ecological adaptation and reproductive isolation in sympatry: genetic and phenotypic evidence for native host races of *Rhagoletis pomonella*. Mol. Ecol. 23: 688–704.

Pritchard JK, Stephens MW, Donnelly P. 2000. Inference of population structure using multilocus genome data. Genetics 155: 945-959.

Prugnolle F, de Meeus T. 2002. Inferring sex-biased dispersal from population genetic tools: a review. Heredity 88: 161-165.

Puechmaille SJ, Borissov IM, Zsebok S, Allegrini B, Hizem M, Kuenzel S, Schuchmann M, Teeling EC, Siemers BM. 2014. Female mate choice can drive the evolution of high

frequency echolocation in bats: a case study with *Rhinolophus mehelyi*. PLoS One 9:e103452.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. syy032, https://doi.org/10.1093/sysbio/syy032.

Rannala B, Yang Z. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics 164: 1645-1656.

Rannala B, Yang Z. 2013. Improved reversible jump algorithms for Bayesian species delimitation. Genetics 194: 245-253.

Ratcliffe LM, Grant PR. 1983. Species recognition in Darwin's finches (*Geospiza*, Gould) I. Discrimination by morphological cues. Animal Behaviour 31: 1139-1153.

Renaut S, Grassa CJ, Yeaman S, Moyers BT, Lai Z, Kane NC, Bowers JE, Burke JM, Rieseberg LH. 2013. Genomic islands of divergence are not affected by geography of speciation in sunflowers. Nat. Commun. 4: 1827.

Rewerska A, Pawelczyk M, Rajkowska E, Politanski P, Sliwinska-Kowalska M. 2013. Evaluating D-methionine dose to attenuate oxidative stress-mediated hearing loss following overexposure to noise. Otorhinolaryngol. 4: 1513-1520.

Richards E, Poelstra J, Mann C. 2017. Don't throw out the sympatric species with the crater lake water: fine-scale investigation of introgression provides weak support for functional role of secondary gene flow in one of the clearest examples of sympatric speciation. bioRxiv: https://doi.org/10.1101/217984.

Rimbault M, Beale HC, Schoenebeck JJ, Hoopes BC, Allen JJ, Kilroy-Glynn P, Wayne RK, Sutter NB, Ostrander EA. 2013. Derived variants at size genes explain nearly half of size reduction in dog breeds. Genome Res. 23: 1985-1995.

Robert R. 2016. The role of echolocation in communication in a high duty cycle echolocating bat, *Rhinolophs clivosus* (Chiroptera: Rhinolophidae): an experimental approach. PhD Biological Sciences, University of Cape Town, Cape Town.

Roellig D, Bronner ME. 2016. The epigenetic modifier DNMT3A is necessary for proper otic placode formation. Dev. Biol. 15: 294-300.

Rossiter SJ, Jones G, Ransome RD, Barratt EM. 2000. Parentage, reproductive success and breeding behaviour in the greater horseshow bat (*Rhinolophus ferrumequinum*). Proc. R. Soc. B. 267: 545-551.

Rossiter SJ, Ransom RD, Faulkes CG, Dawson DA, Jones G. 2006. Long-term paternity skew and the opportunity for selection in a mammal with reversed sexual size dimorphism. Mol. Ecol. 15: 3035-3043.

Roverud RC, Grinnell AD. 1985. Frequency tracking and Doppler shift compensation in response to an artificial CF/FM echolocation sound in the CF/FM bat, *Noctilio albiventris*. J Comp Physiol A. 156: 471-475.

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large datasets. Mol. Biol. Evol. 34: 3299-3302.

Ruegg K, Anderson EC, Boone J, Pouls J, Smith TB. 2014. A role for migration-linked genes and genomic islands in divergence of a songbird. Mol. Ecol. 23:4757-4769.

Russo D, Mucedda M, Bello M, Biscardi S, Pidinchedda E, Jones G. 2007. Divergent echolocation call frequencies in insular rhinolophids (Chiroptera): a case of character displacement? J. Biogeogr. 34: 2129-2138.

Rybak LP, Hussain K, Whitworth C, Somani SM. 1999. Dose dependent protection by lipoic acid against cisplatin-induce ototoxicity in rats: an antioxidant defense system. Toxicol. Sci. 47: 195-202.

Saetre G-P. 2014. Genome scans and elusive candidate genes: detecting the variation that matters for speciation. Mol. Ecol. 23:4677-4678.

Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M, Nativ N, Bahir I, Doniger T, Krug H, Sirota-Madi A, Olender T, Golan Y, Stelzer G, Harel A, Lancet D. 2010. GeneCards Version 3: the human gene integrator. Database 2010. www.genecards.org

Schield DR, Card DC, Adams RH, Jezkova T, Reyes-Velasco J, Proctor FN, Spencer CL, Herrmann H-W, Mackessy SP, Castoe TA. 2015. Incipient speciation with biased gene flow between two lineages of the Western Diamondback Rattlesnake (*Crotalus atrox*). Mol. Phlylogenetics Evol. 83: 213-223.

Schield DR, Adams RH, Card DC, Perry BW, Pasquesi GM, Jezkova T, Portik DM, Andrew AL, Spencer CL, Sanchez EE, Fujita MK, Mackessy SP, Castoe TA. 2017. Insight into the roles of selection in speciation from genomic patterns of divergence and introgression in secondary contact in venomous rattlesnakes. Ecol. Evol. 7: 3951-3966.

Schliewen UK, Tautz D, Pääbo S. 1994. Sympatric speciation suggested by monophyly of crater lake cichlids. Nature 368: 629–632.

Schnitzler H-U, Kalko EKV. 2001. Echolocation by insect-eating bats. Bioscience 51: 557-569.

Schuchmann M, Puechmaille SJ, Siemers BM. 2012. Horseshoe bats recognise the sex of conspecifics from their echolocation calls. Acta Chiropterol. 14: 161-166.

Schuller G, Pollak G. 1979. Disproportionate frequency representation in the inferior colliculus of doppler-compensating Greater Horseshoe bats: Evidence for an acoustic fovea. J. Comp. Physiol. A. 132: 47–54.

Schuller G, O'Neill WE, Radtke-Schuller S. 1991. Facilitation and delay sensitivity of auditory cortex neurons in CF-FM bats, *Rhinolophus rouxi* and *Pteronotus p. parnellii*. Eur. J. Neurosci. 3: 1165-1181.

Sedlock J, Francis C, Heaney L, Suyanto I. 2008. *Rhinolophus philippinensis*. The IUCN Red List of Threatened Species. Version 2014.2. <www.iucnredlist.org>. Downloaded on 10 September 2014.

Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenloe PA, Peichel CL, Saetre G-P, Bank C, Brannstrom A, Brelsford A, Clarkson CS, Eroukhmanoff F, Feder JL, Fischer MC, Foote AD, Franchini P, Jiggins CD, Jones FC, Lindholm AK, Lucek K, Maan ME, Marques DA, Martin SH, Matthews B, Meier JI, Most M, Nachman MW, Nonaka E, Rennison DJ, Schwarzer J, Watson ET, Westram AM, Widmer A. 2014. Genomics and the origin of species. Nat. Rev. Genet. 15:176-192.

Seidman MD, Tang W, Shirwany N, Bai U, Rubin CJ, Henig JP, Quirk WS. 2009. Anti-intracellular adhesion molecule-1 antibody's effect on noise damage. Laryngoscope. 119: 707-712.

Seim I, Fang X, Xiong Z, Lobanov AV, Huang Z, Ma S, Feng Y, Turanov AA, Zhu Y, Lenz TL, Gerashchenko MV, Fan D, Yim SH, Yao X, Jordan D, Xiong Y, Ma Y, Lyapunov AN, Chen G, Kulakova OI, Sun Y, Lee S-G, Bronson RT, Moskalev AA, Sunyaev SR, Zhang G, Krogh A, Wang J, Gladyshev VN. 2013. Genome analysis reveals insights into physiology and longevity of the Brandt's bat *Myotis brandtii*. Nat. Commun. 4: 2212.

Servedio MR, Van Doorn GS, Kopp M, Frame AM, Nosil P. 2011. Magic traits in speciation: "magic" but not rare? Trends Ecol. Evol. 26: 389–397.

Shen B, Avila-Flores R, Liu Y, Rossiter SJ, Zhang S. 2011. Prestin shows divergent evolution between constant frequency echolocating bats. J. Mol. Evol. 73: 109-115.

Shen B, Fang T, Dai M, Jones G, Zhang S. 2013. Independent losses of visual perception genes Gja10 and Rbp3 in echolocating bats (Order: Chiroptera). PLoS One 8(7): e68867.

Shen Y-Y, Liu J, Irwin DM, Zhang Y-P. 2010. Parallel and convergent evolution of the dim-light vision gene *RH1* in bats (Order: Chiroptera). PLoS One 5(1): e8838.

Shen Y-Y, Liang L, Li G-S, Murphy RW, Zhang Y-P. 2012. Parallel evolution of auditory genes for echolocation in bats and toothed whales. PLoS Genet. 8(6): e1002788.

Shi H, Dong J, Irwin DM, Zhang S, Mao X. 2016. Repetitive transpositions of mitochondrial DNA sequences to the nucleus during the radiation of horseshoe bats (*Rhinolophus*, Chiroptera). Gene 581: 161-169.

Shimoyama M, De Pons J, Hayman GT, Laulederkind SJ, Liu W, Nigam R, Petri V, Smith JR, Tutaj M, Wang SJ, Worthey E, Dwinell M, Jacob H. 2015. The Rat Genome Database 2015: genomic, phenotypic and environmental variations and disease. Nucleic Acid Res. 28: D743-750.

Shu J, Yin S, Tan AZ, He M. 2015. Association between the prothrombin G20210A mutation and sudden sensorineural hearing loss in European populations: a meta-analysis. Thromb. Res. 135: 73-77.

Simmons NB. 2005. Order Chiroptera. In: Wilson DE, Reeder DM eds. Mammal species of the world: a taxonomic and geographic reference. Washington (DC): Smithsonian Institution Press. 312-529.

Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. 2014. Sequencing depth and coverage: key considerations in genomic analyses. Nat. Rev. Genet. 15: 121-132.

Smith GR. 1992. Introgression in fishes: significance for paleontology, cladistics and evolutionary rates. Syst. Biol. 41: 41-57.

Smith JD. 1972. Systematics of the Chiropteran family Mormoopidae, vol. 56. Miscellaneous publication. Mus. Nat. Hist, Univ. Kansas: 1-132.

Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Pond SLK. 2015. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol Biol Evol. 32: 1342-1353.

Smotherman M, Guillén-Servent A. 2008. Doppler-shift compensation behavior by Wagner's mustached bat, *Pteronotus personatus*. J. Acoust. Soc. Am. 123: 4331–4339.

Soisook P, Karapan S, Srikrachang M, Dejtaradol A, Nualcharoen K, Bumrungsri S, Lin Oo SS, Aung MM, Bates PJJ, Harutyunyan M, Buś MM, Bogdanowicz W. 2016. Hill

forest dweller: a new cryptic species of *Rhinolophus* in the '*pusillus* group' (Chiroptera: Rhinolophidae) from Thailand and Lao PDR. Acta Chiropterologica 18: 117-139.

Springer MS, Teeling EC, Madsen O, Stanhope MJ, de Jong WW. 2001. Integrated fossil and molecular data reconstruct bat echolocation. Proc. Natl. Acad. Sci. U. S. A. 98: 6241–6246.

St-Jacques B, Hammerschmidt M, McMahon AP. 1999. Indian hedgehog signalling regulates proliferation and differentiation of chondrocytes and is essential for bone formation. Genes Dev. 13: 2072-2086.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312-1313.

Stankiewicz P, Lupski JR. 2010. Structural variation in the human genome and its role in disease. Ann. Rev. Med. 61: 437-455.

Steiner CC, Römpler H, Boettger LM, Schönberg T, Hoekstra HE. 2009. The genetic basis of phenotypic convergence in beach mice: similar pigment patterns but different genes. Mol Biol Evol. 26: 35-45.

Stoffberg S, Jacobs DS, Matthee CA. 2011. The divergence of echolocation frequency in horseshoe bats: moth hearing, body size or habitat. J. Mammal. Evol. 18L 117-129.

Suga N, Neuweiler G, Möller J. 1976. Peripheral auditory tuning for fine frequency analysis by the CF-FM bat, *Rhinolophus ferrumequinum*. J. comp. Physiol. 106: 111-125.

Suga N, O'Neill WE, Manabe T. 1978. Cortical neurons sensitive to combinations of information-bearing elements of biosonar signals in the Mustache bat. Science. 200: 778-781.

Sun K, Kimball RT, Liu T, Wei X, Jin L, Jiang T, Lin A, Feng J. 2016. The complex evolutionary history of big-eared horseshoe bats (*Rhinolophus macrotis* complex): insights from genetic, morphological and acoustic data. Sci. Rep. 6: 35417.

Supple MA, Papa R, Hines HM, McMillan WO, Counterman BA. 2015. Divergence with gene flow across a speciation continuum of *Heliconius* butterflies. BMC Evol. Biol. 15: 204.

Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K, Zhu L, Padhukasahasram B, Karlins E, Davis S, Jones PG, Quignon P, Johnson GS, Parker HG, Fretwell N, Mosher DS, Lawler DF, Satyaraj E, Nordborg M, Lark KG, Wayne RK, Ostrander EA. 2007. A single IGF1 allele is a major determinant of small size in dogs. Science 316: 112-115.

Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105: 437-460.

Tu VT, Hassanin A, Görföl T, Arai S, Fukui D, Thanh HT, Son NT, Furey NM, Csorba G. 2017. Integrative taxonomy of the *Rhinolophus macrotis* complex (Chiroptera, Rhinolophidae) in Vietnam and nearby regions. J. Zool. Syst. Evol. Res. 55: 177-198.

Taylor PJ, Macdonald A, Goodman SM, Kerney T, Cotterill FPD, Stoffberg S, Monadjem A, Schoeman MC, Guyton J, Naskrecki P, Richards LR. 2018. Integrative taxonomy resolves three new cryptic species of small southern African horseshoe bats (*Rhinolophus*). Zool. J. Linn. Soc. Zly024.

Teeling EC, Scally M, Kao DJ, Romagnoli ML, Springer MS, Stanhope MJ. 2000. Nature 403: 188-192.

Teeling EC, Madsen O, Van Den Bussche RA, de Jong WW, Stanhope MJ, Springer MS. 2002. Microbat paraphyly and the convergent evolution of a key innovation in Old World rhinolophoid microbats. PNAS 99: 1431-1436.

Teeling EC, Springer MS, Madsen O, Bates P, O'Brien SJ, Murphy WJ. 2005. A molecular phylogeny for bats illuminates biogeography and the fossil record. Science. 307: 580-584.

The Gene Ontology Consortium. 2017. Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Res. 45: D331-D338.

Thiagavel J, Cechetto C, Santana SE, Jakobsen L, Warrant EJ, Ratcliffe JM. 2018. Auditory opportunity and visual constraint enabled the evolution of echolocation in bats. Nat. Commun. 9: 98.

Thoisy BD, Pavan AC, Delaval M, Lavergne A, Luglia T, Pineau K, Ruedi M, Rufray V, Catzeflis F. 2014. Cryptic diversity in common Mustached bats *Pteronotus* cf. *parnellii* (Mormoopidae) in French Guiana and Brazilian Amapa. Acta Chiropt. 16: 1-13.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22: 4673-4680.

Toews DP, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. Mol. Ecol. 21: 3907-3930.

Tokita M, Abe T, Suzuki K. 2012. The developmental basis of bat wing muscle. Nat. Commun. 3: 1302.

Tsagkogeorga G, McGowen MR, Davies KTJ, Jarman S, Polanowski A, Bertelsen MF, Rossiter SJ. 2015. A phylogenomic analysis of the role and timing of molecular adaptation in the aquatic transition of cetartiodactyl mammals. Royal Soc. Open Sci. 2: 150156.

Turner TL, Hahn MW, Nuzhdin SV. 2005. Genomic islands of speciation in *Anopheles gambiae*. PLoS Biol. 3(9):e285.

Turner TL, Hahn MW. 2010. Genomic islands *of* speciation or genomic islands *and* speciation? Mol. Ecol. 19:848-50.

Van der Auwera GA, Carneiro M, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella K, Altschuler D, Gabriel S, DePristo M. 2013. From FastQ to data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinformatics 43: 11.10.1-11.10.33.

Van Den Bussche RA, Weyandt SE. 2003. Mitochondrial and nuclear DNA sequence data provide resolution to sister-group relationships within Pteronotus (Chiroptera: Mormoopidae). Acta Chiropterologica 5: 1–13.

Verity R, Nichols RA. 2016. Estimating the number of subpopulations (K) in structured populations. Genetics 203: 1827-1839.

Via S. 2001. Sympatric speciation in animals: the ugly duckling grows up. TREE 16:381-90.

Via S, West J. 2008. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. Mol. Ecol. 17:4334-45.

Vines TH, Köhler SC, Thiel M, Ghira I, Sands TR, MacCallum CJ, Barton NH. 2003. The maintenance of reproductive isolation in a mosaic hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*. Evolution 57: 1876-1888.

Volleth M, Loidl J, Mayer F, Yon H-S, Müller S, Heller K-G. 2015. Surprising genetic diversity in *Rhinolophus luctus* (Chiroptera: Rhinolophidae) from Peninsular Malaysia: description of a new species based on genetic and morphological characters. Acta Chiropterologica 17: 1-20.

Wang B, Ekblom R, Ignas B, Siitari H, Högland J. 2014. Whole genome sequencing of the black grouse (*Tetrao tetrix*): reference guided assembly suggests faster-Z and MHC evolution. BMC Genomics 15: 180.

Wang L, Li G, Wang J, Ye S, Jones G, Zhang S. 2009. Molecular cloning and evolutionary analysis of the GJA1 (connexin43) gene from bats (Chiroptera). Genet. Res. Camb. 91: 101-109.

Wang X, Liu N, Zhang H, Yang X-J, Huang Y, Lei F. 2015. Extreme variation in patterns of tandem repeats in mitochondrial control region of yellow-browed tits (*Sylviparus modestus*, Paridae). Sci. Rep. 5: 13227.

Weil D, El-Amraoui A, Masmoudi S, Mustapha M, Kikkawa Y, Lainé S, Delmaghani S, Adato A, Nadifi S, Ben Zina Z, Hammel C, Gal A, Ayadi H, Yonekawa H, Petit C. 2003. Usher syndrome type 1 G (USH1G) is caused by mutations in the gene encoding SANS, a protein that associates with the USH1C protein, harmonin. Hum. Mol. Gen. 12: 463-471.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38: 1358-1370.

Weissensteiner MH, Pang AWC, Bunikis I, Höijer I, Vinnere-Pettersson O, Suh A, Wolf JBW. 2017. Combination of short-read, long-read and optical mapping assemblies reveals large-scale tandem repeat arrays with population genetic implications. Genome Res. 27: 697-708.

Wertheim JO, Murrell B, Smith MS, Pond SLK, Scheffler K. 2014. RELAX: Detecting relaxed selection in a phylogenetic framework. Mol. Biol. Evol. 32: 820-832.

Wilson DE, Reeder DM (editors). 2005. Mammal species of the world: a taxonomic and geographic reference (3rd ed). John Hopkins University Press.

Wittkopp PJ, Williams BL, Selegue JE, Carroll SB. 2012. *Drosophila* pigmentation evolution: Divergent genotypes underlying convergent phenotypes. PNAS. 100: 1808-1813.

White BJ, Cheng C, Simard F, Costantini C, Besansky NJ. 2010. Genetic association of physically unlinked islands of genomic divergence in incipient species of *Anopheles gambiae*. Mol. Ecol. 19:925-39.

Wu C-I. 2001. The genic view of the process of speciation. J. Evol. Biol. 14: 851-865.

Wu TD, Watanabe CK. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21: 1859-1875.

Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26: 873-881.

Yang Y, Smith SA. 2013. Optimizing *de novo* assembly of short-read RNA-seq data for phylogenomics. BMC Genomics 14: 328.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24: 1586-1591.

Yang Z, Rannala B. 2010. Bayesian species delimitation using multilocus sequence data. PNAS 107: 9264-9269.

Yang Z. 2015. The BPP programme for species tree estimation and species delimitation. Curr. Zool. 61: 854-865.

Yang Z. 2018. BPP discussion group May 2018; https://groups.google.com/forum/#!forum/bpp-discussion-group

Ye F, Samuels DC, Clark T, Guo Y. 2014. High-throughput sequencing in mitochondrial DNA research. Mitochondrion 17: 157-163.

Yim H-S, Cho YS, Guang X, Kang SG, Jeong J-Y, Cha S-S, Oh H-M, Lee J-H, Yang EC, Kwon KK, Kim YJ, Kim TW, Kim W, Jeon JH, Kim S-J, Choi DH, Jho S, Kim H-M, Ko J, Kim H, Shin Y-A, Jung H-J, Zheng Y, Wang Z, Chen Y, Chen M, Jiang A, Li E, Zhang S, Hou H, Kim TH, Yu L, Liu S, Ahn K, Cooper J, Park S-G, Hong CP, Jin W, Kim H-S, Park C, Lee K, Chun S, Morin PA, O'Brien SJ, Lee H, Kimura J, Moon DY, Manica A, Edwards J, Kim BC, Kim S, Wang J, Bhak J, Lee HS, Lee J-H. 2014. Minke whale genome and aquatic adaptation in cetaceans. Nat. Genet. 46: 88-92.

Yoshida CA, Yamamoto H, Fujita T, Ito K, Inoue KI, Yamana K, Zanma A, Takada K, Ito Y, Komori T. 2004. Runx2 and Runx3 are essential for chondrocyte mutation, and Runx2 regulates limb growth through induction of Indian hedgehog. Genes Dev. 18: 952-963.

Yoshikawa M, Uchida S, Ezaki J, Rai T, Hayama A, Kobayashi K, Kida Y, Noda M, Koike M, Uchiyama Y, Marumo F, Kominami E, Sasaki S. Genes Cells 7: 597-605.

Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol. Biol. Evol. 22: 2472-2479.

Zhang C, Rannala B, Yang Z. 2014. Bayesian species delimitation can be robust to guide-tree inference errors. Sys. Biol. 63: 993-1004.

Zhao H, Rossiter SJ, Teeling EC, Li C, Cotton JA, Zhang S. 2009. The evolution of colour vision in nocturnal mammals. PNAS 106: 8980-8985.

Zhao Y, Zhao F, Zong L, Zhang P, Guan L, Zhang J, Wang D, Wang J, Chai W, Lan L, Han B, Yang L, Jin X, Yang W, Hu X, Wang X, Li N, Li Y, Petit C, Wang J, Wang HY, Wang Q. 2013. Exome sequencing and linkage analysis identified tenascin-C (TNC) as a novel causative gene in nonsyndromic hearing loss. PLoS One 8: e69549.

Ziff JL, Crompton M, Powell HRF, Lavy JA, Aldren CP, Steel KP, Saeed SR, Dawson SJ. 2016. Mutations and altered expression of *SERPINF1* in patients with familial otosclerosis.