Phylogeography and diversification of Taiwanese bats

Hao-Chih Kuo

School of Biological & Chemical Sciences
Queen Mary University of London
Mile End Road, London E1 4NS
United Kingdom

A thesis submitted to the
University of London for the
Degree of Doctor of Philosophy
September 2012

Statement of originality

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with standard refereeing practices in biological sciences. I acknowledge the helpful guidance and support of my supervisors and other advisors in sections dedicated to this purpose. Any views expressed in the thesis without reference to others are those of the author.

Hao-Chih Kuo

September 2012

Acknowledgements

I would like to thank my supervisors, Stephen Rossiter (Queen Mary, University of London; QM) and James Cotton (Wellcome Trust Sanger Institute), who guided me through the challenges of my PhD. In particular, I have received much support, not only academically but also mentally, from Stephen Rossiter.

I also wish to express my full appreciation towards people who assisted during my study. Research funding in Taiwan was obtained with the help of Siang-Fan Chen (Providence University, Taiwan) and Yin-Ping Fang (National Chiayi University, Taiwan) with additional help from Stephen Rossiter. Chia-Hong Chen (Shei-Pa National Park Headquarters, Taiwan), Cheng-Han Chou and Chao-Lung Hsu (Bat Association of Taiwan), His-Chi Cheng (Endemic Species Research Institute, Taiwan), Gábor Csorba (Hungarian Natural History Museum), Yin-Ping Fang and Ling-Ling Lee (National Taiwan University), and Judith Eger and Burton Lim (Royal Ontario Museum, Canada) all contributed tissue samples for my studies. I benefited greatly from discussing work with Richard Nichols and Rob Verity (QM), who shared their deep insights into population genetics, and also with Kalina Davies, Steve Le Comber, Silvia Pérez-Espona, Georgia Tsagkogeorga (all QM) and Mark Beaumont (University of Bristol), who all offered advice on analyses. Jon Flanders introduced me to GIS data processing, while Joe Parker (QM) helped with me with running analyses on computer clusters. Many thanks also go to those who helped during fieldwork: Yi-Chun Chang, Chun-Chia Huang, Han-Chun Hou, Chun-Jen Hsiao, Kun-Jung Li, Chung-Hsien Wang and Hsuan-Yu Wu. Special thanks to my uncle, who found a superb vehicle for me and I am very grateful to people who accompanied me for longer or short periods during my study, especially Tzu-Chun, and Mu-Yun.

Little can I achieve without support from my families. I love you.

I am indebted to the Overseas Research Students Awards Scheme of the UK and the Taiwanese Ministry of Education, both of which provided me with support to cover tuition fees and other expenses. I also thank the University of London Central Research Fund and the National Science Council of Taiwan, both of which awarded me funds for research.

Abstract

Gene flow is a central evolutionary force that largely determines the level of differentiation among populations of organisms and thus their potential for divergence from each other. Identifying key factors that influence gene flow among populations or closely related taxa can thus provide valuable insights into how new species arise and are maintained. I undertook a comparative study of the factors that have shaped range-wide intraspecific differentiation in four related and broadly co-distributed Taiwanese bat species of the genera *Murina* and *Kerivoula*. Bats were sampled from sites across Taiwan and sequenced at two mitochondrial genes as well as genotyped at newly developed and/or existing multi-locus microsatellite markers. To improve phylogeographic inference of existing patterns of population genetic structure, I undertook spatial distribution modeling of the focal species at both the present time and at the Last Glacial Maximum. Genetic data were analysed using traditional and new methods, including Bayesian clustering, coalescent-based estimation of gene flow, and haplotype network reconstruction. My findings revealed contrasting signatures of population subdivision and demographic expansion that appear in part to reflect differences in the altitudinal ranges of the focal taxa. Mitochondrial analyses also revealed a putative sister relationship between two of the Taiwanese endemic taxa - *M. gracilis* and *M. recondita*, which - given the fact both are restricted to Taiwan - presents an unusual case of potential non-allopatric divergence. To dissect this divergence process in more detail, I used 454-Pyrosequencing to obtain ten nuclear loci sequences of these two taxa, and a third taxon from mainland Asia, *M. eleryi*. Based on these loci, Bayesian isolation-migration models provided no strong evidence of post-split gene flow and, therefore, did not support speciation within Taiwan. Instead, the divergence process reconstructed from ncDNA loci was found to be incompatible with the mtDNA tree, with *M. recondita* showing a sister relationship with *M. eleryi*. This conflict is best explained by the ancient introgression of mtDNA between the two insular species following their colonization of Taiwan at different times.

Contents

List of figures

List of tables

Abbreviations

AIC - Akaike Information Criterion

AUC - Area under the curve of the receiver-operating characteristic plot

bp - base pair

BSP - Bayesian skyline plot

CCSM - Community Climate System Model

CMR - Central Mountain Range of Taiwan

COI - Cytochrome c oxidase subunit 1 gene

CP_{12} - Protein coding sequence comprising the first and second codon positions

CP_3 - Protein coding sequence comprising the third codon position

Cyt-*b* - Mitochondrial cytochrome b gene

EBSP - Extended Bayesian skyline plot analysis

EGD - Euclidean geographic distance

ESS - Effective sampling size

GTR - Tavaré's (1986) generalized time reversible nucleotide substitution model

HKY - Hasegawa *et al.*'s (1985) nucleotide substitution model

HWE – Hardy-Weinberg equilibrium

IBD – Isolation by distance

IM - Isolation with migration coalescent model

Kyr - Thounsand years

LCPD - Least-cost path distance

LD - Linkage disequilibrium

LGM - Last Glacial Maximum

LRT - Likelihood ratio test

MCC - Maximum clade credibility

MCMC - Markov chain Monte Carlo

MC^3 - Metropolis-coupling Markov chain Monte Carlo

MIROC - the Model for Interdisciplinary Research on Climate

MJ - median-joining algorithm

MP - maximum parsimony

mtDNA - mitochondrial DNA

MTPT - minimum training-presence threshold

Myr - million years

ncDNA - nuclear DNA (microsatellites or their flanking region sequences)

ND1 - mitochondrial NADH dehydrogenase subunit 1 gene

PC - piecewise-constant coalescent model

PCR - polymerase chain reaction

PLCR – 'piecewise-linear and constant root' coalescent model

RD - resistance distance based on the application of electrical circuit theory

RI - reproductive isolation

SD - standard deviation

SDM - species distribution modeling

SH - allelic size homoplasy of microsatellites

TMRCA – time to most recent common ancestor

TN93 - Tamura and Nei's (1993) nucleotide substitution model

5OT - 5% omission threshold

95% CI - 95% confidence interval

Chapter 1: General Introduction

As well as revolutionizing our understanding of the basic processes of living things, molecular genetic techniques have allowed biologists to shed new light on the ecology, evolution and genetics of natural populations. Knowledge of population genetic structure is crucial for understanding multiple aspects of population dynamics, from individual dispersal behaviour (e.g. Sundstrom *et al.* 2003) and reproductive skew (e.g. Rossiter *et al.* 2006), to the impact of barriers to gene flow (e.g. Fairley *et al.* 2002) and the responses of species to changes in both their biotic and abiotic environment (e.g. Albaladejo & Aparicio 2007; Laroche & Durand 2004).

The field of population genetics was developed to offer theory-based explanations for how different evolutionary forces (i.e. natural selection, genetic drift, mutation, and migration) cause changes in the genetic composition of populations (Hamilton 2009). The subsequent introduction of the discipline of phylogeography has focused on attempting to understand how historical and contemporary processes that have shaped the geographical distribution of genetic lineages (Avise *et al.* 1987). To this end, phylogeography has, for many years, relied on extending traditional phylogenetic principles and methods to the intraspecific level, and such intraspecific phylogenies have indeed proven highly useful for evaluating the genetic subdivision of populations (Avise 2000; Soltis *et al.* 2006).

In phylogeography, a hypothesis is typically sought for the relationship between the observed pattern of genetic variation and the geographic distribution of populations, by matching the observed genetic pattern to those expected under various historical processes, such as vicariance (Arbogast & Kenagy 2001; Avise 1992), population bottlenecks (Galtier *et al.* 2000; Garza & Williamson 2001) and population expansions (Pybus *et al.* 2002; Ramos-Onsins & Rozas 2002) (reviewed in Avise 2000). Given that the occurrence of many of these historical processes also require evidence from geology, (paleo-)climatology and (paleo-)ecology, as well as other fields, phylogeography can be considered an interdisciplinary subject (Cruzan & Templeton 2000; Hewitt 1996, 2001a, 2003, 2004).

Most early studies of phylogeography typically made inferences about a given observed genetic pattern on the basis of a single gene (e.g. reviewed in Avise 2000; Hewitt 2004; Soltis *et al.* 2006). Here much importance was typically placed on agreement between

an observed genetic pattern to that expected under a specific historical or biogeographical process (e.g. see fig 1 in Petit (2008)). This rationale was formalized in the so-called Nested Clade Analysis (NCA) in which inferences were made based on a key of potential drivers of genetic patterns (Templeton 1998, 2002a, b, 2004; Templeton *et al.* 1995). Although initially popular, NCA type methods have been criticized for high chance of false-positive inferences (Panchal & Beaumont 2007; Petit & Grivet 2002) as well as for their lack of considerations for competing hypotheses (see 'Statistical phylogeography').

Statistical phylogeography

In recent years, investigations of phylogeography have increasingly incorporated model-based approaches from population genetic theory (e.g. Drummond *et al.* 2005; Lemey *et al.* 2009; Pritchard *et al.* 2000) and, during the past two decades, numerous studies have integrated these two fields to give a more comprehensive picture of how genes in natural populations at different geographic scales can be shaped by a range of historical processes, including demographic and environmental changes (e.g. Bernatchez & Dodson 1991; Encalada *et al.* 1996; Hoffmann & Baker 2003; Randi *et al.* 1994; Rossiter *et al.* 2007). Knowles and Maddison (2002) and Knowles (2004) termed this emerging field 'statistical phylogeography' and stressed that at least two pivotal aspects of inferring phylogeographic history had been overlooked in earlier studies.

The first is the substantial stochastic variance seen in rates of genic coalescence and mutation, which can have large proportional effects when studying processes over short evolutionary timeframes, such as those that characterise phylogeography studies (Hudson 1990; Tajima 1983). When these processes are incorporated into methods for estimating various demographic parameters via population genetic methods, large variances will often result, so highlighting the uncertainty around parameter estimates. It follows that any conclusions drawn upon these estimates may also be flawed. Indeed, it is not unusual for divergence dates (e.g. Arbogast *et al.* 2002; Heled & Drummond 2010) and quantities and timing of demographic growth (e.g. Heled & Drummond 2008; Storz & Beaumont 2002) and migration (e.g. Nichols & Freeman 2004; Vitalis & Couvet 2001) to show broad confidence intervals when Bayesian methods are used to integrate sources of uncertainties in estimations.

Rather than being considered a problem for phylogeographic inference, broad confidence intervals around such parameters reflect the limits of the data (Huelsenbeck *et al.* 2001).

The growing appreciation that stochastic processes - such as the timing and order of the coalescence of alleles, and the rate of occurrence and fate (probability of fixation or extinction) of neutral mutations along a gene genealogy - have large variances, means that it is now well-recognized that any single gene will seldom capture accurately all of the features of these processes, and thus any phylogeographic inference based on a single gene is likely to be imprecise and frequently misleading (Baer *et al.* 2007; Hudson 1990). This realization, coupled with trends in high throughput sequencing, has encouraged more and more studies to include multiple loci in their analyses, which has been shown to be an efficient way of reducing uncertainty in parameter estimation (e.g. Haasl & Payseur 2011; Heled & Drummond 2008; Hey 2010b; Vitalis & Couvet 2001). Indeed, it has been suggested that to assess the influence of variance in these stochastic processes and to improve the reliability of relevant estimates, a sample of multiple unlinked loci should be used (Arbogast *et al.* 2002; Edwards & Beerli 2000).

The second possible weakness of many earlier methods and studies was their failure to consider alternative hypotheses when explaining observed phylogeographic patterns, which was a criticism directed towards NCA. The fact that multiple processes can often lead to very similar patterns of genetic variation, or can act in concert with each other to shape genetic patterns, means that the effects of specific historical processes can be hard to resolve or separate. Fortunately, some advanced phylogeographic approaches have been proposed to address such difficulties, by accommodating these different processes into a generalized model, and separating their respective contribution through the estimation of relevant parameters by maximum likelihood or Bayesian approaches (Beaumont *et al.* 2002; Becquet & Przeworski 2007; Leman *et al.* 2005; Wakeley & Hey 1997).

The potential for multiple processes to lead to similar patterns can be illustrated with reference to two examples, the first of which is migration-drift equilibrium. Here, the widely used equation that describes the relationship between inter-population gene flow and their differentiation, F_{ST}, is valid only under the assumption that the mutation rate is low enough to be ignored (Slatkin 1993; Wright 1965). Such an assumption might be unsafe where genetic markers with high mutation rates - such as microsatellites - are used, thus challenging the

utility of F_{ST} as a measure of gene flow (Balloux & Lugon-Moulin 2002; Neigel 2002; Whitlock 2011). This will be especially problematic when comparing genetic differentiation between different population pairs in cases where different sets of markers have been used. Performing simulations under the finite island model, Nichols and Freeman (2004) showed that markers with different mutation rates can indeed lead to contrasting F_{ST} values at equilibrium. The variance around these F_{ST} values, reflecting their mutation rates, is dependent on the level of inter-population gene flow: that is, smaller variances in F_{ST} are associated with greater inter-population gene flow. To distinguish effects of mutations and gene flow in shaping a given level of population differentiation, Nichols and Freeman (2004) developed a likelihood function containing both factors on which Bayesian analyses were applied.

The second example of different evolutionary processes leading to similar patterns is that of population divergence, which is particularly pertinent to this thesis. The time elapsed since populations started to diverge (the 'split time') has a similar but opposite effect as subsequent gene flow ('post-split gene flow') on the extent of their genic divergence. Wakeley (1996) showed that the exact same level of divergence can arise from combinations of these parameters, ranging from complete isolation with no post-split gene flow to migration-drift equilibrium between these populations. At the same time, however, Wakeley (1996) demonstrated that the split time and post-split gene flow contribute differentially to the variance in intra- and inter-population genic differences. To assess the relative contribution of isolation and migration to a given observed genetic pattern, several methods have been recently developed, termed 'isolation with migration' (IM) models (Nielsen & Wakeley 2001). These models parameterize six main elements to describe the divergence process between populations (or closely related species): the effective population sizes of the ancestral population and two descendant populations, the split time of descendant populations, and the migration rates (to specify the extent of gene flow) in either direction between them. By fitting this model on sampled genetic data from populations under comparison, these parameters can be estimated via optimizations of Bayesian Inference. Based on these estimates, it is possible to determine the relative input of gene flow and split time in determining an observed population subdivision. Since their introduction, IM models have undergone considerable refinement, and likelihood functions have been developed for single locus (Nielsen & Wakeley 2001) and multiple loci (Hey & Nielsen 2004; Hey &

Nielsen 2007), and also can accommodate divergence among multiple populations (Hey 2010b).

Divergence and speciation in a phylogeographic framework

The development of new statistical IM models may provide particularly useful insights into how phylogeographic history has driven the formation of new taxa. The use of IM models to detect non-trivial gene flow - either in one or both directions - during divergence of related taxa could point to non-allopatric speciation. Indeed an increasing number of studies have recently applied these methods to pairs of taxa (divergent populations, sister species, or even non-sibling species) to examine whether gene flow occurred after they split from each other (e.g. partially reviewed by Becquet & Przeworski 2009) and, therefore, to help distinguish between strict allopatric speciation and other speciation modes (Llopart *et al.* 2005; Won & Hey 2005). In a few cases, reported support for gene flow was seen as casting doubt on the paradigm of allopatric speciation in explaining the formation of biodiversity (Becquet & Przeworski 2009).

To dissect the speciation process between a pair of sister species, it is straightforward to imagine a scenario in which an ancestral population splits at some past time to give two descendant populations. In such a system, an inference of a non-allopatric speciation process (i.e. a process accompanied with gene flow) will be made if the gene flow between descendant populations (i.e. after they split from each other) can be detected by examining their contemporary genetic data; otherwise an allopatric speciation process cannot be rejected. Methods for this purpose have been developed using either selectively neutral genes (IM-based methods) or both neutral genes and those potentially associated with the establishment of RI (so-called 'speciation genes') (e.g. Osada & Wu 2005). This latter study is based on Wu's (2001) theory in which he proposes that neutral and speciation genes will leave different evolutionary signals in cases of non-allopatric speciation due to their different behaviours during the establishment of RI, but will leave similar signals when allopatric speciation is involved. However, such methods typically use a single representative genome from each of the taxa under comparison, and, to compensate for the lower power for speciation inference from omitting information of genetic polymorphism in each taxon, a

very large number of genes of the two categories from each taxon are required (Osada & Wu 2005; Wu & Ting 2004). Such a requirement makes it difficult to apply this latter approach to non-model animals - including bats - for which genome data are generally lacking.

A challenge for identifying post-split genetic mixing is ruling out other processes that can give similar signatures. Gene flow tends to diminish the genetic differences between descendant populations by mixing recently coalesced gene lineages within each of them, which can lead to non-reciprocal monophyly of alleles. However, a similar genealogical pattern can result where the time elapsed since the split has not been sufficient for drift to act on descendant populations to remove shared ancestral genes, a phenomenon also known as 'incomplete lineage sorting' (Avise 2000; Avise et al. 1983). Hudson and Coyne (2002: table 1) demonstrated that under an allopatric mode of divergence, it will take around 12 to 15 N generations (where N is the effective population size of each descendant population) for a moderate sample size of 5 to 25 loci to reach reciprocal monophyly with 95% probability. Moreover, as mentioned above (see section 'Statistical phylogeography'), Wakeley (1996) has demonstrated that exactly the same average genetic difference within and between descendant populations can be produced in scenarios of both long isolation with moderate gene flow (the extreme case is migration-drift equilibrium) or recent isolation with low gene flow (the extreme case is pure isolation with no gene flow).

Even where post-split gene flow can be safely inferred from IM estimations (and thus recent divergence and/or incomplete lineage sorting can be ruled out), then such analyses cannot by themselves distinguish between a scenario in which gene flow has occurred continually until the establishment of RI (genuine non-allopatric speciation) and one in which gene flow stems from secondary contact. Unfortunately, the vast majority of studies that have tested for non-allopatric speciation using IM have been focused on animals that are likely to be highly mobile, such as mammals (Becquet & Przeworski 2007; Kotlik et al. 2006), birds (e.g. Balakrishnan & Edwards 2009; Cabanne et al. 2008), and fishes (e.g. Hua et al. 2009; Nielsen & Wakeley 2001), and also have not been undertaken in suitable geographical context (e.g. a small isolated habitat) for ruling out secondary contact. Such studies include a case for the two only known bat species of the genus *Myzopoda* which are endemic to Madagascar (Russell et al. 2008). Madagascar is a very large island (16 times larger than Taiwan) with numerous biogeographical zones, and thus offers good conditions for allopatric

separation between species. Moreover, no gene flow was detected between these sister species after their divergence.

One exception to this trend comes from studies for cichlid fishes in Lake Malawi (Hey et al. 2004; Won et al. 2005) where more than 450 cichlid species (Genner et al. 2004) co-exist and are thought to have radiated from a single ancestral species within a short geological timeframe (570 years to1 million years, Delvaux 1996). Hence all of these species are closely related to each other and restricted to a small area. Some of these aforementioned studies have shown non-trivial gene flow after ancestral populations diverged from each other. These results, along with some theoretical (e.g. Kawata et al. 2007) and empirical studies (e.g. Seehausen et al. 2008) relevant to cichlid fishes in Lake Malawi, as well as other East African lakes, have arguably offered the most rigorous evidence to date of speciation processes in which geographic isolation has not played a major role.

Focal taxa

Murina and *Kerivoula* are the most speciose genera within the bat subfamilies Murininae and Kerivoulinae, respectively, both of which are found within the major family Vespertilionidae (Chiroptera). Phylogenetic studies based on representatives from these subfamilies have established their relationship as sister clades (Hoofer *et al.* 2003; Hoofer & Van den Bussche 2003; Lack *et al.* 2010; Stadelmann *et al.* 2004b). A total of 17 distinct species of *Murina* were recognized by Simmons (2005), all distributed in Oriental and Palearctic ecozones with a western boundary at Pakistan and a northern boundary at southern Siberia; one (or two) species also occurs in archipelagos east of Borneo as well as New Guinea and northeastern Australia. New species of this genus have continued to be reported in recent years (Csorba & Bates 2005; Csorba *et al.* 2011; Csorba *et al.* 2007; Eger & Lim 2011; Furey *et al.* 2009; Kruskop & Eger 2008; Kuo *et al.* 2009; Ruedi *et al.* 2012). Species of the genus *Kerivoula* show remarkably similar patterns; Simmons (2005) recognized 19 distinct species, 11 of which occur in Oriental, Palearctic, and Australasian ecozones with a northern boundary in Central China. Unlike *Murina*, however, the other seven species of this genus occur in sub-Saharan Africa in the Afrotropic ecozone. Numerous new species of *Kerivoula* have been reported from Southeast Asia (Bates *et al.* 2007; Bates *et al.* 2004;

Francis *et al.* 2007), and the taxonomy and systematics of many species of both genera is in need of further revision (Francis *et al.* 2010; Hasan & Abdullah 2011; Khan *et al.* 2010; Kruskop 2005).

The ecology of all species within the Murininae and Kerivoulinae has been understudied and is still poorly known. Rhodes (1995) showed that the Australian species *Phoniscus papuensis* (a member of the Kerivoulinae) has a wing morphology (low wing loading with round wing tip) that is typical of bats specialized for foraging in cluttered environments such as the forest interior (Norberg & Rayner 1987), and this was corroborated by experiments on flight performance. Kingston *et al.* (2003) showed the same was true of several Malaysian species of *Murina*, *Kerivoula* and *Phoniscus*. Furthermore, Kingston *et al.* (1999) showed that these Malaysian species showed highly specialized echolocation calls with very large bandwidths and high frequencies, features also since reported for other members of these groups (Fukui *et al.* 2004; Ma *et al.* 2008; Pennay *et al.* 2004; Sripathi *et al.* 2006). A series of behavioural experiments conducted in a very recent study (Schmieder *et al.* 2012) linked the echolocation call parameters of *Kerivoula* to their superior performance in discriminating prey very close to and in vegetation.

The ecomorphological traits of *Murina* and *Kerivoula* that are highly adapted for living in clutter might also constrain their ability to undertake movements across large patches of unsuitable habitat, such as open spaces and water due to the rapid attenuation of their calls and the energetic inefficiency of their low wing aspect ratio (Aldridge & Rautenbach 1987; Kingston *et al.* 1999; Norberg & Rayner 1987; Schnitzler & Kalko 2001; Schnitzler *et al.* 2003). Surveys corroborate the close association of all of four species with forest habitat, with no or extremely few records from open habitats.

For my research I focused on four bat species: *Murina gracilis*, *M. recondita* , *M. puta* and *Kerivoula* sp., each shown in Figure 1.1. All of these taxa have wing morphologies and echolocation calls that are adapted for foraging in highly cluttered environments of the forest interior. *Murina gracilis* and *M. recondita* are both newly discovered species known only from Taiwan and described by me (Kuo *et al.* 2009). They are similar in size and appearance (ca. 3-5g) and, on the basis of preliminary sequence data appear to be very closely related to each other (3.5-4% uncorrected distance in cytochrome *b* and NADH dehydrogenase subunit 1) (Kuo 2004).

Figure 1.1 Four focal Taiwanese bat species in this research.

Murina gracilis (from Fang YP)	*Murina recondita*
Murina puta	Taiwanese *Kerivoula* sp. (from Chen SF)

Although hitherto only a few individuals of each taxon have been recorded, they appear to show clear differences in habitat use; in particular, *M. gracilis* occurs in mountain regions with elevations from 1400 to 2700 m, while *M. recondita* typically occurs below 1500 m (although a few individuals have been recorded at higher elevations of up to 2200 m). Both species have been found in all of Taiwan's mountain ranges except for the Coast Range, and, in most cases, were caught in the field within or at the edges of forest. Two individuals of *M. gracilis* were found in an abandoned tunnel within a forest, although their roosting habitats are not known. To date, there are only two records of these two species found at the same (or nearby) site (Kuo 2004) .

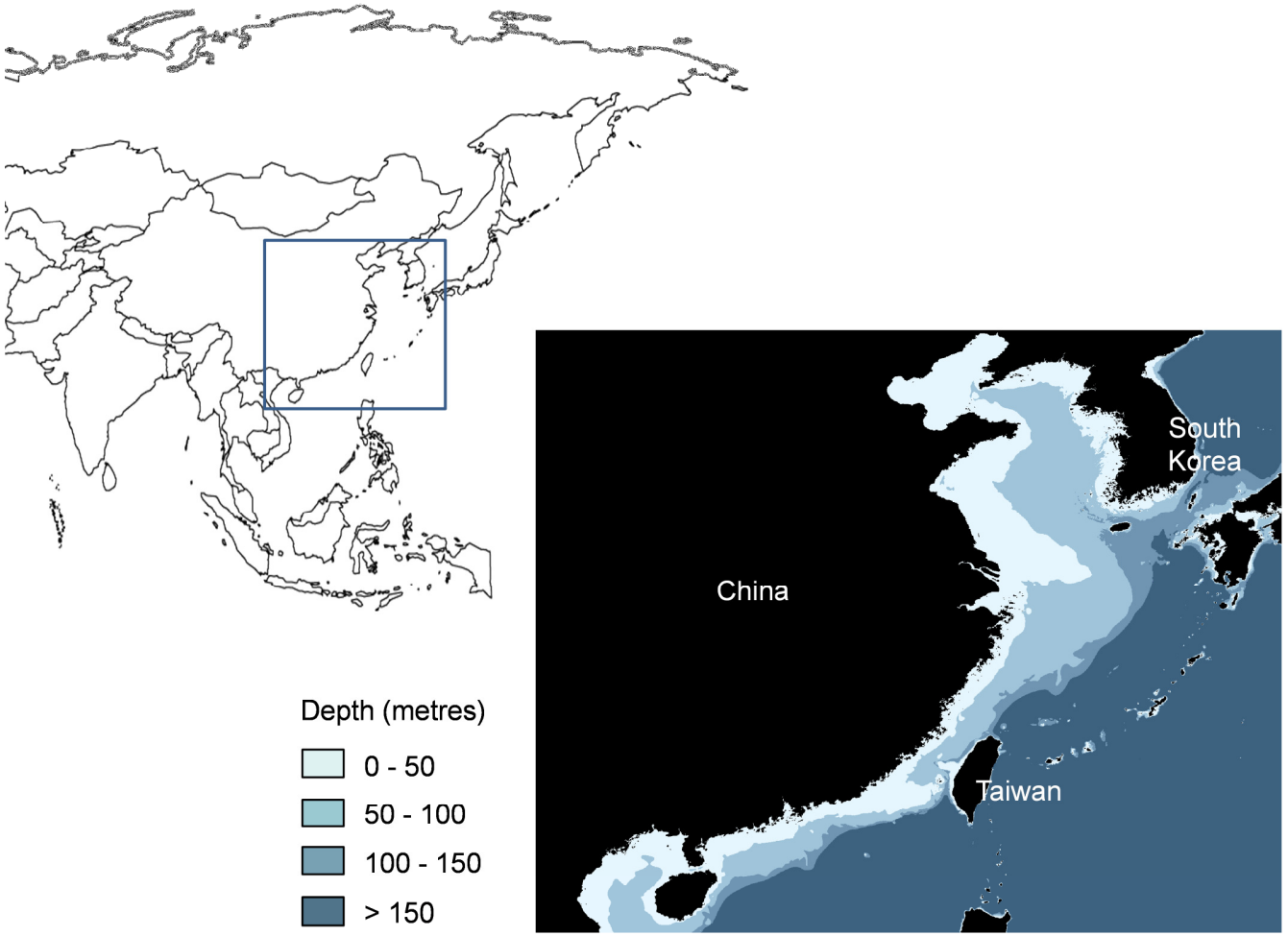
The congener *M. puta* is a slightly larger taxon (ca. 5-10g) that is also endemic to Taiwan. Preliminary genetic evidence suggests this taxon has a sister relationship with a clade comprising *M. ryukyuana* and *M. silvatica* from Japan (11-12% uncorrected distance in ND1) (Kuo 2004), although it might in fact be more closely related to *M. huttoni* (Kuo 2004; Yoshiyuki 1989), which has a wide distribution from the Himalayas to South-East Asia and South-East China. *M. puta* appears to have a wider distribution in Taiwan than the other two focal *Murina* species, and is found in the Coast Range, as well as Yangmingshan and Kenting National Parks (Lee *et al.* 2007; this research) located, respectively in the northernmost and southernmost parts of Taiwan. It is typically found within or near to forest at elevations below 2400 m, and largely co-occurs with the other two *Murina* species. Although its roosting ecology is almost unknown, a small colony of three individuals was found roosting inside a dried, furled leaf of a banana tree (Chou *et al.* 2008).

The fourth focal species, *Kerivoula sp.*, is a small (ca. 4-7g) species of uncertain taxonomic status. It is the only example of its genus in Taiwan and is probably an undescribed endemic species on the basis that it looks dissimilar to other *Kerivoula* from mainland Asia. Its range on Taiwan appears to be similar to that of *M. recondita*, also found at sites within or near to forest. A preliminary phylogeographic study (Yang 2008) revealed two mitochondrial lineages based on cytochrome *b* sequences (ca. 2% Kimura-2-parameters distance) with largely overlapping distributions. However, no divergence was found between these lineages using data from seven microsatellite loci.

Study objectives

I undertook a comparative study of range-wide phylogeographic structure and demographic dynamics in the four focal bat species described in the previous section. All four are considered to be endemic to Taiwan (see Fig. 1.2 for map), and together they include lowland and upland dwellers as well as an altitudinal generalist. My overall aim was to determine the extent to which historical and environmental factors have resulted in similar patterns and modes of intra-specific diversification across these species. Particularly, I hypothesized that the presence of mountain ranges (which represent huge altitudinal changes over small geographic scales on the island) will have had differential impacts on levels of population structure and gene flow in these species, due to their contrasting altitudinal distributions, and thus expected tolerance to colder upland conditions. For this study I established novel microsatellite markers for *Murina* species and I also cross-amplified previously developed markers for the Taiwanese *Kerivoula* sp.; details of these procedures are described in Chapter 2. In Chapter 3, I describe the use of these microsatellite markers as well as two mitochondrial genes to elucidate the population structure, intra-specific phylogenetic structure and historical demography of each of the four focal species. I also characterise gene flow using new approaches in landscape genetics. In Chapter 4, I undertake a more detailed study of the relationship and divergence process between two of the *Murina* species, which were found to be sister taxa based on mitochondrial markers. In particular, I tested the hypothesis that, because these taxa are thought to be restricted to Taiwan, they could have evolved *in situ* via non-allopatric speciation. For this analysis, I examined demographic parameters for divergence and estimated split times and levels of post-split gene flow using IM models and coalescent-based methods based on multiple anonymous nuclear makers. In Chapter 5, I consider how genetic structure in these species might be impacted during future climate change.

Figure 1.2 A map showing the position of Taiwan and the bathymetric profile around this island (modified from source data downloaded from: http://www.gebco.net/).



Chapter 2: General methods

SUMMARY

Microsatellites are regularly used to detect subtle genetic differences between individuals or populations because of their hypervariable nature as well as their assumed neutrality that means they are subject to genetic drift. In order to conduct comparisons of range-wide phylogeographic structure in four Taiwanese bat species (Chapter 3) and also assess gene flow among some of these (Chapter 4), I characterised a set of microsatellite markers for each of my focal taxa. First, markers developed for *Murina gracilis* were used to screen this taxon as well as cross-amplify two Taiwanese congeneric species: *M. recondita* and *M. puta*. The final number of useable markers that did not deviate from Hardy-Weinberg or Linkage equilibrium was 18 markers for *M. gracilis* and *M. recondita,* including one that was X-linked, and nine for *M. puta*. For the Taiwanese *Kerivoula* sp., I tested a panel of markers developed previously for *Kerivoula papillosa*, and identified eight useable markers that met the above criteria, in addition to one *Murina* locus that cross-amplified in *Kerivoula*. Here I describe the characteristics of these microsatellite loci in all four taxa, and describe the range wide sampling for population genetic analyses in following chapters.

2.1 INTRODUCTION

Bat surveys in Taiwan

Bats constitute the largest component of Taiwanese mammalian biodiversity, with 30 distinct species reported in Lin *et al.* (2004) representing nearly 40% of the island's 80 or so recorded terrestrial mammal species (Taiwan Biodiversity Information Facility: http://www.taibif.org.tw/en/node/265). One third of Taiwan's bat species are endemic (Lin *et al.* 2004), although this might be an underestimation because several taxonomic studies of the genera *Myotis* (Chou 2004) and *Pipistrellus* (Wu 2007) have listed additional potentially distinct species, which have not yet been formally described.

Studies of bats in Taiwan have a short history. For example, to date, only a handful of studies have been conducted on the basic ecology of Taiwanese bats, and most of these focused on cave-dwelling taxa whose roosts are relatively easy to locate, plus species that are tolerant to human activities (reviewed in Lin *et al.* 2004). In comparison, there are less data available for tree-roosting species, including *Kerivoula* and *Murina*. There is also a paucity of information on the distributions of nearly all of Taiwan's bats, largely because surveys have seldom been conducted in the montane regions that constitute more than half of the island's area. There are a few exceptions to this rule, although they have focused on restricted areas of mountainous habitat (e.g. Huang & Lee 2007; Lee *et al.* 2007).

Microsatellites

Microsatellites are short simple sequences of one to six nucleotides that are tandemly repeated. They appear to evolve via a mechanism of strand slippage in which repeat motifs are gained or lost following a step-wise model of mutation (Ellegren 2004). Microsatellites are found throughout eukaryotic genomes, although they predominantly occur in non-coding regions. This is particularly true of microsatellites with di- and tetra-nucleotides repeat units, which are not multiples of three base-pairs and thus more likely to cause frame-shifts (conversely, loss or gain of tri- and hexa-nucleotide motifs will not cause frame-shifts and thus these sorts of microsatellites are more tolerated in coding regions) (reviewed in Li *et al.* 2002). Given their scarcity in coding sequences, microsatellites are usually considered to

evolve neutrally (Ellegren 2004; Li *et al.* 2002). Stepwise loss or gain of repeat units means that microsatellites tend to have characteristically high mutation rates when compared to markers with similar modes of inheritance (i.e. where effective population sizes of the marker are not a confounding factor) with rates in the order of 10^{-3} to 10^{-4} mutations per locus per gamete per generation (Ellegren 2000). Consequently, microsatellites typically show much higher levels of allelic diversity than many other markers in a given population (Avise 2004; Wan *et al.* 2004).

Genotyping of microsatellite loci is undertaken by designing primers to attach to the adjacent flanking regions for amplification by polymerase chain reaction (PCR). Amplified allele sizes are then scored on the basis of size separation that corresponds to their number of repeat units. Given that microsatellites can be amplified from very small amounts of starting material, and also show levels of polymorphism, means that they are powerful markers for genetically discriminating among individuals and groups of individuals. As such, since their discovery microsatellites have become the marker of choice in numerous areas of forensics, molecular ecology and evolution. Applications to date include genetic fingerprinting of individuals, parentage inference, detecting population differentiation and gene flow, and estimating demographic parameters (Avise 2004; Selkoe & Toonen 2006).

To date, no microsatellite markers have been developed for any *Murina* species, or other members of the Murininae. For the related subfamily Kerivoulinae, microsatellite markers have been developed for a Southeast Asian species, the papillose woolly bat, *Kerivoula papillosa*, reported in Struebig *et al.* (2008). Previous attempts to cross-amplify these *K. papillosa* markers in other bats suggested some success with congeneric species (see also Rossiter *et al.* 2012 and this study) but not with different genera (Rossiter SJ, personal communication). I therefore tested the same panel of markers in the Taiwanese *Kerivoula* but decided to develop *de novo* a set of microsatellite markers for *Murina* species using *Murina* tissue collected from one of my focal taxa.

2.2 MATERIALS AND METHODS

Bat sampling and DNA extraction

To collect tissue samples from bats of the four focal species, I conducted 145 nights of survey work in Taiwan spanning several months during 2009 (Apr-Aug) and 2010 (May-Sep). In total, 52 sites were visited across the island, although at a few sites traps were deployed at multiple localities; thus 'site' can be defined as the collection of such localities that are no more than 5 km apart from each other, but which are farther than 5km from localities in other sites. To maximize sample sizes of the focal bat species, especially *Murina gracilis* and *M. recondita* that are both relatively rare, I spent variable periods of time at each site, ranging from one to eight nights. For bat capture I used between two and six 'four-bank harp traps' (Appendix 2.1) per night, each placed along trails in the forest interior such as logging roads and other paths. On a small number of occasions I also set up one or two mist-nets (of length 12 m and/or 6 m) to increase sampling, although these proved to be less effective than the harp traps at capturing the focal taxa.

Captured bats were identified to species level based on Kuo *et al.* (2009) for *Murina* bats and based on Lin *et al.* (2004) for the others. For each individual bat I recorded the sex, age (adult versus juvenile) and reproductive status of the female (non-reproductive, pregnant, lactating, post-lactating) following the guidelines of Kingston *et al.* (2006). Additionally I measured the body mass (± 0.25 g; spring balance), forearm length (± 0.1 mm; dial caliper), and tibia length (± 0.1 mm; dial caliper) to aid with species identification. For each bat, two wing biopsies (3-mm diameter) were collected using a dermatological punch (Stiefel Laboratories, Wooburn Green, UK) and transferred to 90% ethanol for storage at -20 ºC. All bats were immediately released after the above procedure. DNA was extracted using commercial kits (Qiagen, UK).

Survey work conducted during this research achieved reasonable sample sizes for *M. gracilis* and *M. recondita*, however both are uncommon. Thus to improve sample sizes I requested additional material from several Taiwanese researchers who have captured and sampled *Murina* incidentally while working on other taxa. These researchers are Chia-Hong Chen (Shei-Pa National Park Headquarters), His-Chi Cheng (Endemic Species Research Institute), Cheng-Han Chou (Bat Association of Taiwan), Yin-Ping Fang (National Chiayi

University), and Ling-Ling Lee (National Taiwan University). Since *M. gracilis* and *M. recondita* were only described as distinct species very recently (Kuo *et al.* 2009), identifications by third-parties were verified by examining any linked voucher specimens deposited in the collections of National Chiayi University and National Taiwan University (Appendix 2.2). Individual bats with wing biopsies collected and provided by above researchers also had their identification confirmed based on molecular methods (see Result and Discussion). Chia-Hong Chen and Chao-Lung Hsu (Bat Association of Taiwan) also contributed for additional wing biopsies of *M. puta* and *Kerivoula* sp. to this and subsequent studies. The distributions of samples analysed in my study, from all four focal species, are summarized in Table 2.1 and Fig. 2.1.

Development of microsatellites for *M. gracilis*

For development of microsatellites, a single individual of *M. gracilis* sampled at site 23 (Fig. 2.1) in 2009 was sacrificed (specimen ID. KHC09081208). A sample of muscle tissue (approx. 5g) was sent to Genetic Identification Services (GIS; Chatsworth, CA, USA) where enriched microsatellite libraries were created, cloned, and a subset were sequenced. In brief, restriction fragments (350-700 base pairs) of genomic DNA of *M. gracilis* were enriched for microsatellite motifs CA, GA, ATG, and TAGA. These fragments were ligated into the *Hind* III restriction site of pUC19 plasmid and were together introduced into *E. coli* strain DH5α (ElectroMax, Invitrogen) by electroporation. Recombinant clones were identified using blue-white screening media, and target plasmids were purified with Millipore Multiscreen MAFB NOB plates (http://www.millipore.com/techpublications/tech1/TN004). Using the DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Biosciences), GIS performed sequencing for 144 positive clones on an ABI Prism 377 DNA Sequencer (Applied Biosystems). Of these clone sequences, 128 were seen to contain microsatellites.

I selected 32 clones that contained suitable microsatellite motifs, and designed primers using the software Primer3 (Rozen & Skaletsky 2000). For multiplexing, the 5′ end of each forward primer was fluorescently labeled (FAM, HEX, or TAMRA). A ‘pig-tail’ was incorporated into the 5′ end of the reverse primer for each locus with CA repeat motifs

(Table 2.2), so to standardize for variation in adenylation introduced by *Taq* during PCR (Brownstein *et al.* 1996).

Characterisation of microsatellite loci in Taiwanese *Murina* spp. and *Kerivoula* sp.

The 32 primer pairs developed for *M. gracilis* were initially tested on eight individual bats of each of the three focal *Murina* species. In such tests, multiplex PCRs were carried out with the Type-it Microsatellite PCR Kit (Qiagen) in a final volume of 7.5µl following the manufacturer's protocol. For each species an annealing temperature gradient of 55 to 65 ºC was performed on an ABI 3730 DNA Analyzer (Applied Biosystems). ROX-400 was used as the size standard and allele sizes were scored using the software GeneMapper 4.0 (Applied Biosystems).

Using the approach described, 24 loci were amplified successfully and also produced interpretable results in at least one species (Table 2.2). Of these I selected 21 markers (14 for *M. puta*) for screening in a larger sample (see Results and Discussion) of 70 individuals of *M. gracilis*, 94 of *M. recondita*, and 96 of *M. puta*, covering 5, 8, and 9 demes (defined by sampling sites), respectively. Tested demes were sites 4, 19, 24, 31 and 54 for *M. gracilis*, sites 2, 7, 10, 11, 16, 28, 40 and 54 for *M. recondita*, and sites 26, 29, 41, 42, 43, 47, 49, 54 and 55 for *M. puta* (see Fig. 2.1 for locations).

For testing the utility of *K. papillosa* primers for cross amplification of loci in the Taiwanese *K.* sp., eight individual bats were used for initial screening. The majority of primer pairs described by Struebig *et al.* (2008) were initially used for testing along with four loci originally developed for *M. gracilis* (A107, A112, A118 and D110). Multiplex PCRs were undertaken using the same conditions and as for *Murina* except that a fixed annealing temperature of 59 ºC was used for all eight samples. From a set of 21 markers, nine produced interpretable results (see Results and Discussion) and were subsequently used for screening 96 individual bats of this species from nine demes (sites 2, 6, 11, 16, 26, 35, 36, 53 and 55). Genotypes were scored with GeneMapper.

For each microsatellite locus for each taxon, I estimated the observed and expected heterozygosity values averaged over individual demes using the program FSTAT 2.9.3

(Goudet 2001). To test for deviations from Hardy-Weinberg equilibrium (HWE), I estimated the F_{IS} value of each locus using Weir and Cockerham's estimator (1984) and compared this value to a null distribution produced by randomization (2000 iterations) of intra-demic alleles among individual bats. Significance was assessed using a one-tailed test (that is, a deficit of heterozygosity than expected). Note that *M. gracilis* and *M. recondita* were typed at an X-linked locus (see Results and Discussion) and thus for these species, only demes with good sample sizes of females were analysed (and only data from females were included in calculations of heterozygosity and HWE tests). Finally I tested each species for potential linkage disequilibrium (LD) between pairs of loci based on the log-likelihood ratio G-statistic, which was compared to its null distribution produced by randomization (iteration number set by FSTAT 2.9.3 for a table-wide error rate of 0.05) of pairs of genotypes between loci. For assessing significance, a one-tailed test was performed. For both *M. gracilis* and *M. recondita*, the X-linked locus was excluded from LD tests. Both HWE and LD tests described above were carried out using FSTAT 2.9.3., and the sequential Bonferroni correction (Holm 1979) was applied.

Table 2.1 Sample sizes of the four Taiwanese bat species of genera *Murina* and *Kerivoula* from 56 sampling sites. The four species are abbreviated as MG for *M. gracilis*, MR for *M. recondita*, MP for *M. puta*, and KSP for *Kerivoula* sp.

No.	Site (County)	MG	MR	MP	KSP
1	Yangmingshan National Park (Taipei)	-	-	2	-
2	Linmei Village (Yilan)	-	15	10	16
3	Fushan Garden (Yilan)	-	-	1	-
4	#100 Logging Road (Yilan)	20	-	2	-
5	Sanguang Village (Taoyuan)	-	1	-	-
6	Da-ai Village (Hsinchu)	-	-	1	5
7	Bailan Tribe (Hsinchu)	-	10	1	6
8	Guanwu Recreation Area-2 (Hsinchu)	3	-	-	-
9	Guanwu Recreation Area-1 (Hsinchu)	2	-	1	-
10	Syuejian Recreation Area (Miaoli)	1	12	19	11
11	Jhongkeng Village (Taichung)	-	14	8	9
12	Dasyueshan Recreation Area (Taichung)	-	-	1	-
13	Huisun Forest Station (Nantou)	-	5	5	-
14	Heping Logging Road (Hualien)	2	-	13	-
15	Auhua Village (Yilan)	-	2	1	4
16	Jinyang Village (Yilan)	-	13	5	10
17	Jinyue Village (Yilan)	-	-	-	2
18	Cueifong Lake (Yilan)	1	-	3	-
19	Jiaping Logging Road (Yilan)	12	1	8	4
20	Sihyuan-yakou (Yilan)	1	-	1	-
21	Wuling Farm (Taichung)	4	-	1	-
22	Lushui Trail (Hualien)	-	-	3	-
23	Dayuling (Hualien)	5	-	1	-
24	Hehuan River Trail (Nantou)	11	-	4	-
25	Meifong (Nantou)	1	1	2	-
26	Wanda Logging Road (Nantou)	-	-	10	11
27	Dili Village (Nantou)	-	-	2	1
28	Sianjin Logging Road (Nantou)	-	9	3	3
29	Shanlinsi Recreation Area (Nantou)	2	-	7	2
30	Shuisheliao (Chiayi)	-	5	2	1
31	Alishan Recreation Area (Chiayi)	19	-	3	-
32	Yushankou (Chiayi)	7	-	-	-
33	Nan-tz-sian-si Logging Road (Chiayi)	1	-	-	-
34	Dongpu Village (Chiayi)	-	1	-	-
35	Chashan Village (Chiayi)	-	5	3	6
36	Jiadong Village (Chiayi)	-	4	-	16
37	Dongshan Fuyou Temple (Chiayi)	-	5	4	4
38	Siaoguanshan Logging Road (Kaohsiung)	-	-	1	-
39	Shihshan Logging Road (Kaohsiung)	-	1	-	-
40	Shanping (Kaohsiung)	-	11	-	-

Table 2.1 Continued.

No.	Site (County)	MG	MR	MP	KSP
41	Taiwu Village (Pingtung)	-	-	6	-
42	Jinshuiying Trail (Pingtung)	-	3	15	-
43	Hsinhua Village (Taitung)	-	-	6	-
44	Yima Logging Road (Taitung)	-	6	3	2
45	Dongsing Village (Taitung)	-	-	1	-
46	Dulanshan Trail (Taitung)	-	-	9	5
47	Yanping Logging Road (Taitung)	-	2	6	1
48	Hongshih Tribe (Taitung)	-	7	8	3
49	Wulu Logging Road (Taitung)	5	1	8	-
50	Siangyang (Taitung)	2	-	-	-
51	Changliang Logging Road (Hualien)	-	1	5	12
52	Yuli Logging Road (Hualien)	-	1	2	-
53	Sanming Logging Road (Hualien)	-	-	1	12
54	Ruisui Logging Road (Hualien)	9	10	14	-
55	Fuyuan Village (Hualien)	-	-	9	11
56	Shuhu Village (Hualien)	-	-	13	7

Figure 2.1 Maps of sampling sites for (a) *Murina gracilis*, (b) *M. recondita*, (c) *M. puta*, and (d) *Kerivoula* sp. Sites are numbered as in Table 2.1 and those with and without GPS coordinates are symbolized as solid circles and open squares, respectively. Areas with altitudes 1500 m a.s.l. or higher are coloured with darker grey.

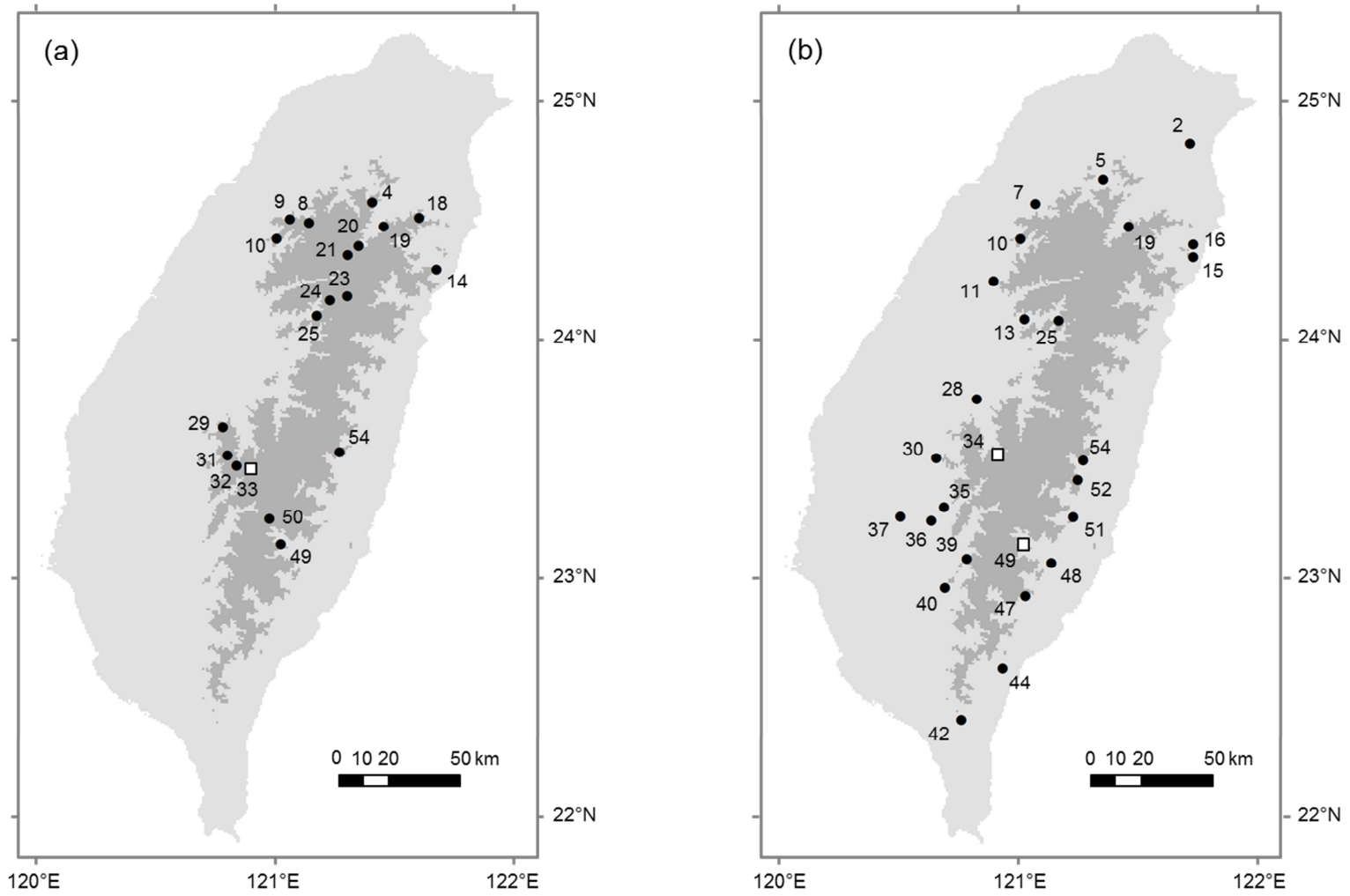


Figure 2.1 Continued.

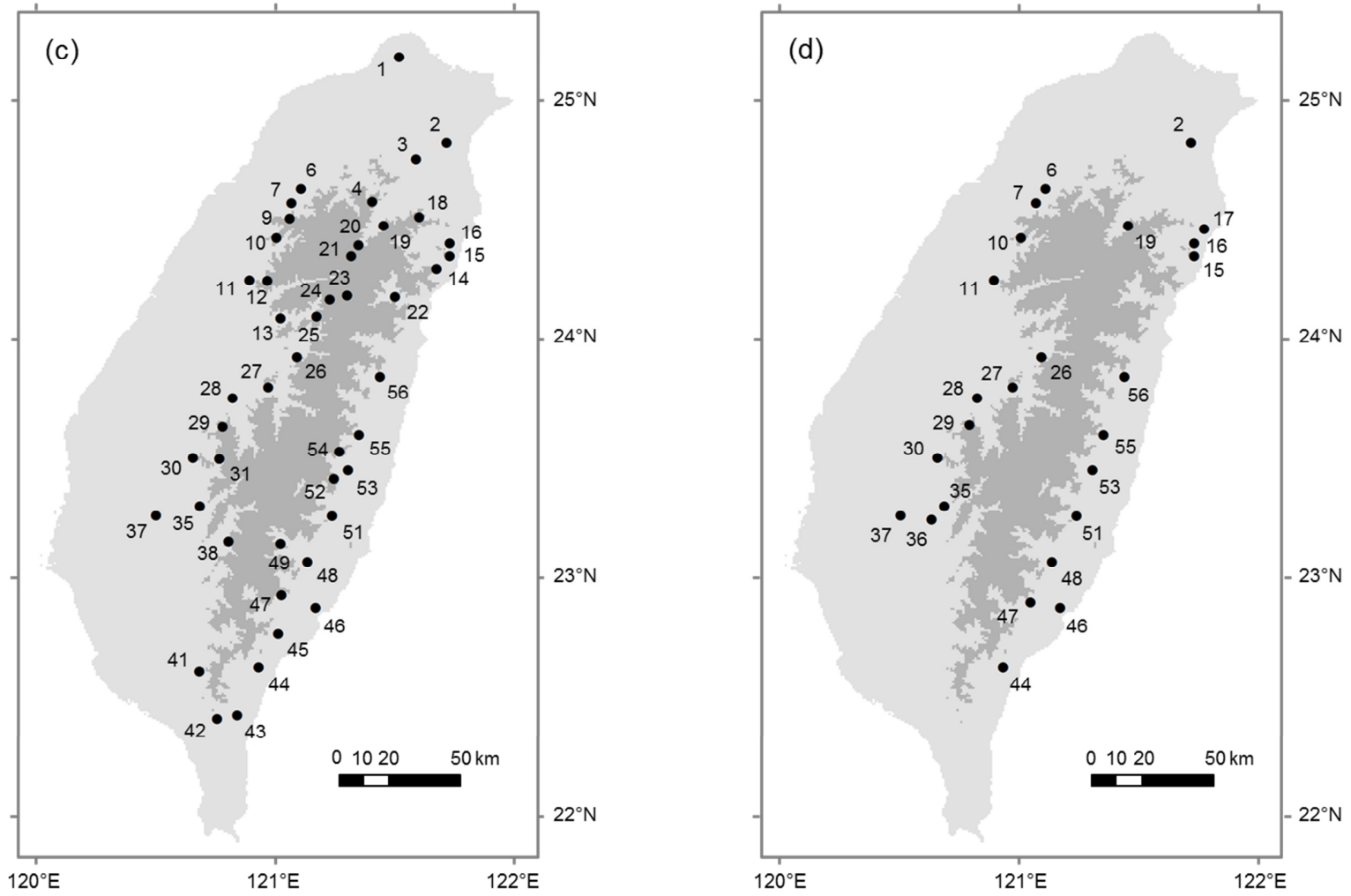


Table 2.2 Primer pairs for 21 microsatellite loci developed for *Murina gracilis*. 'Pig-tails' (before dash) were added to reverse primers for loci with CA repeat motifs. Primer pairs of additional three loci not widely tested are provided below the table[a].

Locus	Repeat motif	Primer sequence (5'-3')	Ta[b] (ºC)
Mugr_A4	$(CA)_{22}$	F: CGGAGCAGGTGAGTAGGAG	60-65
		R: GTTT-GGTGTGAGCAGCAATTCTG	
Mugr_A9	$(CA)_{15}AA(CA)_{11}$	F: GCTCATCAAAGCACTGTTACTC	55-60
		R: GT-TTCATTGTGTCCCTTGTTGT	
Mugr_A10	$(CA)_{19}CG(CA)_3CG$	F: ATTTTTTCCCTTCACTTGCTAG	55-60
	$CACG(CA)_{11}$	R: GTTT-CGGATGAGTGGGTGATTATAC	
Mugr_A103	$(GT)_{21}$	F: CCTTCTCTTGGGATGGCTACT	60-65
		R: GTTT-CCCTTAACCTCCCTGAAAATG	
Mugr_A104	$(CA)_{19}$	F: CTGTTTCCTCATCTGCCAAA	60-65
		R: GTTCTT-TCTGGGTGGTTTATCTGCAA	
Mugr_A107	$(CT)_{13}(CA)_{19}$	F: CCAATTTGTCTGTTAAAACCAC	60
		R: GTTT-AGTCATTCTGGGAAATGTCAG	
Mugr_A109	$(CA)_{18}$	F: GGACCAAATATCAAAGTTCAGA	57-63
		R: GTTT-AGTTGTGTGGGAAGGTATGAT	
Mugr_A112	$(CA)_{16}$	F: GTTTCCATCATCACCACCAG	60-65
		R: GTTT-GCACCACATAGACCCCATC	
Mugr_A118	$(CA)_{15}$	F: AGGAAATCAAGATGCAAGTTC	55-65
		R: GT-TTGGTTGTAATGGTTGCTTAG	
Mugr_A122	$(CA)_{18}$	F: CAGCAACTGTTCATAAAAGACATTC	55
		R: GTTT-CTTCTTGCAATGTGGGGACT	
Mugr_B5	$(CA)_{20}C(GA)_{23}$	F: CCTCTGTTGTCAGGGAAACC	55-60
		R: GTTT-ATTCACATGCCCCTGCTCT	
Mugr_B9	$(GA)_{18}GG(GA)_7$	F: TTGAGGCTGGAATACCAAG	55-60
		R: TTACGGAATGGATGTTTGTG	
Mugr_B12	$(CT)_{23}$	F: CAGGGGCATTTGGTTCTG	57-65
		R: CTGGCCGGACATAAACACTT	
Mugr_B103	$(GT)_{21}GAGG(GA)_{22}$	F: CCCTAAAGATGGAGAAATGAA	55-60
	$GT(GA)_4(GT)_2$	R: GTTT-CCTCTGTCACTGGCTGAAC	
Mugr_B114	$(CT)_{20}$	F: TGACCAGTATGGCTCAGTTG	55-65
		R: GCTCTTCTCATTAGGGCACTTA	
Mugr_B121	$(CT)_{10}TT(CT)_{12}$	F: TCTTCCAGTGCTGCTATGG	55-60
		R: CCAAGGGGTAAAAATGGG	
Mugr_B124	$(CT)_{24}$	F: GCTTTGTACTCAATGGGTGAT	55-60
		R: GCTGGTTGATTAAGCAATCTC	

Table 2.2 Continued.

Locus	Repeat motif	Primer sequence (5'-3')	Ta[b] (ºC)
Mugr_B125	$(GA)_{19}$	F: TAGCACGCATCACCTTCTG	55-63
		R: ATCACCAGTTGGGGAATGT	
Mugr_D9	$(ATCT)_{14}AACT$	F: AAAGGAGATAAACTTTGCACAC	55-63[c]
		R: AACAAAACACTCTGGTCACTG	
Mugr_D110	$(ATCT)_{10}(ACCT)_{12}$	F: GCAGGAGACAGCCAATCAAT	60-65
	ATCT	R: CTAGGTCCTGCGGCTTTTC	
Mugr_D117	$(GATA)_{12}GAT$	F: GTTTGGTTGAGCCTTTAGTTG	55-60
	$(GATA)_2$	R: TTCCTGGGAGATTTTGAAGTA	

[a] Primer pairs of three loci tested on eight individual bats of each of the three *Murina* species: CTC CAC AAG CCA ACT TCT G and GTTT-CTG CCT CCT GAC AGT GTT C for Mugr_A120 (repeat motif: $(CA)_{14}$; successfully amplified in *M. gracilis* (55-65ºC); number of alleles and their size range: 6, 206-218 bps); CGC CTG TAC GTC ACA TCT CTA and CTG GAG CGT CTC CCT ATA CAT for Mugr_B7 (repeat motif: $(GA)_{27}$; successfully amplified in *M. gracilis* (55-63ºC) and *M. recondita*; numbers of alleles and their size ranges: 6, 251-265 bps and 4, 247-273bps in respective species); CGT CGC ATT CTT GGT CTT C and CAG TGG TCA GAG GAG TGT GG for Mugr_B115 (repeat motif: $(GA)_3GG(GA)_{20}\ldots(GA)_4$; amplified in all three species but only scorable in *M. gracilis* (57-63ºC); number of alleles and their size range in *M. gracilis*: 8, 261-293 bps).

[b] Annealing temperature for *M. gracilis*.

[c] Annealing temperature for *M. recondita*.

2.3 RESULTS AND DISCUSSION

Bat sampling

Tissue samples from a total of 652 individual bats of four focal species were obtained for studies in this as well as in the subsequent two chapters. These samples included 108 *M. gracilis*, 146 *M. recondita*, 234 *M. puta*, and 164 *Kerivoula* sp. from a total of 56 sampling sites (Table 2.1; Fig. 2.1; see also Appendix 2.2 for GPS coordinates of these sites). For the majority of bats sampled, I undertook species identification in the field or from examination of voucher specimens (Appendix 2.2). These 'reference samples' accounted for 87 (81%) of *M. gracilis*, 107 (73%) of *M. recondita*, 217 (93%) of *M. puta*, and 159 (97%) of *Kerivoula* sp. For remaining biopsies donated by other researchers, I verified species identities by sequencing of mitochondrial genes (see Chapter 3 for markers used) and with reference to diagnostic alleles at two microsatellite loci (see below).

Microsatellites for *Murina* spp.

Characteristics of the 21 microsatellite loci tested in the three *Murina* species are summarized in Table 2.3. One locus, Mugr_B9 (the prefix 'Mugr_' is omitted hereafter for simplification) was inferred to be X-linked because homozygotes and heterozygotes were observed for females while all males were characterised by a single allele, giving the impression of a homozygote. This pattern was seen in both *M. gracilis* and *M. recondita*, whereas this locus failed to amplify in *M. puta*. All other loci showed both homozygotes and heterozygotes in both sexes, and thus were considered to have an autosomal mode of inheritance.

M. puta had fewest loci successfully amplified, which was expected given its more distant relatedness to the other two taxa. Loci that did not amplify in *M. puta* were A4, A103, A104, B12, B125, D9 and B9. Two loci showed diagnostic differences in amplification or in the allelic profile between *M. gracilis* and *M. recondita*. The first of these was A103, which was polymorphic in *M. gracilis* but fixed in *M. recondita*, and the second one was D9, which failed to be amplified in *M. gracilis*. These loci were thus used to cross-validate species identification based on morphology and sequence data.

A number of loci showed signatures of null amplification in one of the taxa, or otherwise had alleles with shifted peak sizes when visualized using GeneMapper (Table 2.3). These markers were excluded from Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) tests for the corresponding species. A further two loci (A109 and B124) appeared not to amplify in a subset of *M. puta* individuals, and, based on significant deviations from HWE, were removed from downstream analyses of this taxon. For all other markers, no significant deviation from HWE was detected, nor was LD found between pairs of loci in any of the three species following sequential Bonferroni correction. Finally, given that *post hoc* corrections based on many comparisons (here, 136 comparisons for each of *M. gracilis* and *M. recondita*) can lead to overly conservative adjusted alpha values, I also inspected the non-adjusted *P* value of LD tests. Low non-adjusted *P* values were scattered randomly with respect to species and locus, and thus appear to be represent false positives introduced by multiple comparisons rather than genuine LD. Based on these findings I considered that different microsatellite loci used in my studies segregate independently from each other.

After extensive testing, 18 microsatellite loci (all 21 except A107, B103, and D9) were chosen for screening *M. gracilis* while a slightly different set of 18 loci (all 21 except A103, B12, and B103) were chosen for *M. recondita* in Chapter 3. Nine loci were selected for *M. puta*: A9, A10, A112, A122, B5, B114, B121, D110, and D117. All loci in each of these datasets were assumed not to be linked.

Cross-amplification of microsatellites in Taiwanese *Kerivoula* sp.

All nine loci used for screening *Kerivoula* (including eight loci originally developed for *K. papillosa* and one developed for *M. gracilis*) gave scorable genotypes with heterozygotes in both sexes, suggesting autosomal inheritance (Table 2.4). Following sequential Bonferroni correction, none of the nine loci showed a significant deviation from HWE, or evidence of linkage disequilibrium. Therefore, all nine loci were selected for subsequent studies of population genetic structure in the Taiwanese *Kerivoula* sp.

Table 2.3 Characterisation of 21 microsatellite markers for three Taiwanese *Murina* species. A, number of alleles; AS, range of allele sizes (bp); H_O, observed heterozygosity; H_E, expected heterozygosity. Significant deviations from Hardy-Weinberg equilibrium are shown in bold.

	M. gracilis (41 ♂, 29 ♀)			*M. recondita* (57 ♂, 37 ♀)			*M. puta* (65 ♂, 31 ♀)		
Locus	A	AS	H_O/H_E	A	AS	H_O/H_E	A	AS	H_O/H_E
A4[a]	11	174-204	0.84/0.78	12	176-200	0.87/0.80	-	-	-
A9	16	186-218	0.85/0.89	14	184-212	0.77/0.77	4	184-190	0.62/0.67
A10	19	230-274	0.81/0.88	14[e]	216-236[e]	0.70/0.76	19[e]	213-251[e]	0.89/0.88
A103[a]	18	180-218	0.91/0.89	1	160	-	-	-	-
A104[a]	9	217-235	0.76/0.75	11	215-235	0.82/0.81	-	-	-
A107	>35[d]	153-197[d]	-	16	159-191	0.89/0.88	>23[d]	164-201[d]	-
A109	11	248-272	0.73/0.80	10	254-272	0.87/0.84	27	254-305	**0.44/0.93**
A112	18	280-324	0.88/0.90	13[e]	283-303	0.66/0.66	17	276-314	0.93/0.87
A118	11	278-300	0.89/0.89	11	262-288	0.76/0.76	10[d]	275-288[d]	-
A122	16	155-195	0.79/0.82	14	161-189	0.81/0.86	10	147-181	0.61/0.58
B5	15	192-230	0.79/0.86	27[e]	198-234[e]	0.94/0.90	24[e]	203-244[e]	0.87/0.91
B9[ab]	16	248-294	0.69/0.80[a]	19[e]	257-278[e]	0.68/0.82	-	-	-
B12[a]	13	187-215	0.79/0.85	>13[d]	177-209[d]	-	-	-	-
B103	>17[d]	234-266[d]	-	>22[d]	198-249[d]	-	>14[d]	222-271[d]	-
B114	14	188-216	0.92/0.88	16[e]	192-214[e]	0.86/0.86	36[e]	176-257[e]	0.93/0.94
B121	14	190-228	0.81/0.82	7	198-210	0.75/0.78	37[e]	177-291[e]	0.93/0.96
B124	18	164-208	0.82/0.88	16	158-196	0.92/0.89	25[e]	156-212[e]	**0.17/0.89**
B125[a]	10	185-227	0.79/0.80	12	192-222	0.86/0.84	-	-	-
D9[ac]	-	-	-	7	145-169	0.70/0.76	-	-	-
D110	13	190-250	0.87/0.84	9	214-246	0.75/0.76	48[e]	215-315[e]	0.97/0.97
D117	13[e]	281-316[e]	0.78/0.78	7	293-317	0.75/0.76	29[e]	282-391[e]	0.93/0.95

[a] Failed to be amplified in *M. puta* (tested on eight individual bats).
[b] X-linked marker; H_O and H_E are calculated based on females, for demes with five or more females.
[c] Failed to be amplified in *M. gracilis*.
[d] Presence of unscorable alleles due to either faint signals, compared to those with normal signals, or < 1 bp differences in allele sizes from others when visualized in GeneMapper 4.0.
[e] Presence of imperfect mutations.

Table 2.4 Characterisation for nine microsatellite markers (with GenBank accession numbers) on 96 individual bats (56 ♂, 40 ♀) of Taiwanese *Kerivoula* sp. sampled from nine sites. See Table 2.3 for explanations of abbreviations for quantities. None of these nine markers showed significant deviation from the Hardy-Weinberg equilibrium after sequential Bonferroni correction.

Locus[a]	Accession	Repeat unit (bp)	A	AS	H_O/H_E
Kpa02	AM157681	2	6	172-188	0.56/0.61
Kpa05	AM157684	2	7	156-178	0.58/0.64
Kpa08[c]	AM157687	2	26	210-269	0.89/0.88
Kpa24[c]	AM180141	2	24	265-302	0.84/0.88
Kpa25	AM180142	2	15	112-142	0.77/0.84
Kpa26[c]	AM180143	2	14	113-183	0.69/0.70
Kpa32[c]	AM180149	2	11	177-201	0.89/0.85
Kpa46[c]	AM408251	4	11	165-205	0.66/0.59
A122[b]	-	2	7	296-314	0.74/0.72

[a] Markers from Struebig *et al.* (2008).
[b] A marker originally developed for *Murina* spp. in this study.
[c] Presence of imperfect mutations.

Appendix 2.1 A four-bank harp trap (provided by Struebig MJ).

Appendix 2.2 GPS coordinates (in decimal degrees) of sampling sites for the four focal species of this study. Sites are numbered as in Table 2.1. Examined vouchers of *Murina gracilis* (MG) and *M. recondita* (MR) are given.

Site no.	GPS-E	GPS-N	Altitude (m)	Voucher
1	121.524	25.185	855	
	121.542	25.152	437	
2	121.723	24.826	414	
	121.715	24.817	424	
	121.707	24.819	504	
3	121.595	24.757	650	
4	121.412	24.577	1700	MG: NTU 1996.10.3
5	121.360	24.673	607	
6	121.116	24.632	521	
7	121.076	24.573	1268	MR: NCYU FYPB001, FYPB040, KHC040
	121.082	24.587	1350	
8	121.117	24.505	2050	MG: NCYU FYPB168
	121.147	24.493	1989	
9	121.069	24.508	2469	
10	120.988	24.401	1366	
	121.008	24.420	1796	
	121.012	24.426	1862	
11	120.922	24.244	1333	
	120.902	24.249	996	MR: NCYU FYPB102
12	120.976	24.248	1950	
13	121.031	24.090	684	
14	121.680	24.297	1457	MG: NTU 2004.3.1
15	121.735	24.351	173	
16	121.772	24.406	320	
	121.737	24.403	792	
	121.753	24.402	860	
	121.748	24.388	1124	
	121.740	24.385	1145	
17	121.779	24.464	138	
18	121.609	24.513	1987	
19	121.460	24.478	1723	
	121.472	24.460	2145	
20	121.356	24.396	1959	
21	121.325	24.350	1740	
	121.311	24.357	1722	MG: NCYU FYPB092, FYPB169
22	121.508	24.179	420	
23	121.322	24.184	2600	
	121.309	24.184	2594	

Appendix 2.2 Continued.

Site no.	GPS-E	GPS-N	Altitude (m)	Voucher
24	121.235	24.167	2582	
	121.221	24.193	2292	
25	121.181	24.097	2173	
	121.173	24.084	2170	MR: NTU 2005.4.2
26	121.098	23.926	1616	
27	120.978	23.800	688	
28	120.830	23.754	1045	
29	120.794	23.649	1954	
	120.788	23.634	1734	
	120.798	23.642	1754	
30	120.665	23.503	1228	MR: NCYU FYPB068
31	120.775	23.499	1921	
	120.807	23.517	2170	MG: NCYU FYPB050
	120.814	23.519	2270	MG: NCYU FYPB065
32	120.846	23.472	2310	
33	-	-	-	MG: NCYU FYPB048
34	-	-	-	MR: NCYU FYPB039
35	120.692	23.301	942	
36	120.613	23.257	467	MR: NCYU FYPB192
	120.640	23.245	718	
	120.605	23.261	365	
37	120.511	23.262	607	MR: NCYU FYPB103
38	120.814	23.154	1767	
39	120.789	23.085	1650	MR: NCYU FYPB066
40	120.697	22.963	879	
	120.683	22.975	798	
	120.690	22.970	712	
41	120.692	22.608	1058	
42	120.757	22.408	1456	
	120.765	22.409	1460	
43	120.847	22.404	680	
	120.850	22.424	667	
44	120.935	22.633	1292	
	120.938	22.626	1264	MR: NCYU FYPB013
	120.936	22.617	1127	
45	121.020	22.764	212	
46	121.177	22.874	725	
47	121.053	22.898	728	
	121.033	22.929	1605	MR: NCYU FYPB012
48	121.142	23.067	961	MR: NCYU FYPB134

Appendix 2.2 Continued.

Site no.	GPS-E	GPS-N	Altitude (m)	Voucher
49	121.030	23.143	1867	
	121.019	23.141	1940	
50	120.983	23.249	2456	MG: NCYU FYPB057
51	121.245	23.259	825	
	121.234	23.259	904	
52	121.252	23.416	1224	
53	121.312	23.452	635	
54	121.276	23.497	1370	MR: NTU 2003.8.1
	121.281	23.513	1435	
	121.276	23.530	1534	MG: NTU 2003.8.2
55	121.358	23.599	635	
56	121.445	23.844	951	

NTU: Zoology Museum of National Taiwan University; NCYU: National Chiayi University.

Chapter 3: Comparative phylogeography of four Taiwanese bat species from the family Vespertilionidae

SUMMARY

Comparative phylogeographic studies of co-distributed taxa are concerned with identifying shared responses to abiotic factors such as climate history, which have shaped population and community structure. Growing evidence suggests that species responses might also be heavily influenced by biological or ecological attributes. Here I used phylogenetic and population genetic analyses of two different types of genetic markers (mitochondrial DNA and nuclear microsatellites) to investigate range-wide phylogeographic structure and associated demographic processes in four endemic Taiwanese bat species, *Murina gracilis*, *M. recondita*, *M. puta*, and *Kerivoula* sp. For each species, I also modeled its current and former distribution, to provide greater context for interpreting my results. Comparisons of results across the four taxa revealed contrasting population and phylogeographic histories. This variation probably reflects differences in the way each has interacted with extrinsic factors. For example, genetic differentiation among populations separated by mountains was greater in the two lowland species, *M. recondita* and *Kerivoula* sp., than in the highland species, *M. gracilis* and the altitudinal generalist, *M. puta*. The extent to which population genetic structure followed an isolation by distance (IBD) model of gene flow also varied among taxa, with *M. gracilis* and *M. recondita* showing strongest IBD, and *M. puta* showing no detectable trend. Further differences among the taxa might arise from differences in their dispersal behaviour. For example, in *M. gracilis* and *M. recondita*, male-biased gene flow is suggested by detectable nuclear gene flow among populations that show an apparent long-term lack of exchange of mtDNA.

3.1 INTRODUCTION

Phylogeographic studies of individual species or populations seek to identify causal links between observed patterns of genetic variation and extrinsic and/or historical factors affecting these organisms (Avise *et al.* 1987). However, comparative phylogeographic investigations of co-distributed organisms can provide additional important information by revealing extrinsic or historical factors that have shaped multiple species and even entire communities (Bermingham & Moritz 1998). The power of comparative phylogeography is demonstrated in a series of reviews by Hewitt (Hewitt 1996, 1999, 2000, 2001b, 2004) that synthesized data from co-distributed taxa and described shared responses to climatic oscillations and topographic features as inferred from similar patterns of spatial variation in genetic diversity and/or similar distributions of divergent lineages. In particular, paradigms of post-glacial colonization have been identified for temperate European taxa, whereas similar genetic discontinuities associated with topographic features have been detected in a number of eastern North American taxa (reviewed in Soltis *et al.* 2006). Shared patterns across temperate taxa are likely to partially reflect the greater abundance of data available for higher latitudes (see references in Hewitt's reviews), where phylogeographic reconstructions have been conducted at a level of detail not possible elsewhere. These studies have highlighted the importance of climatic factors, geographic barriers and historical refugia, and have suggested different tempos and modes of post-glacial population expansions leading to latitudal clines in genetic diversity.

Glaciations

Since just before the Pliocene-Pleistocene boundary (Gibbard *et al.* 2010), the past 2.75 million years of the history of Earth's history has been characterised by global cooling (Ravelo *et al.* 2004) and for this reason is sometimes referred to as the Quaternary glaciation. During this period, the Earth has experienced a number of cycles of colder episodes called glacials, which have been interrupted with shorter and warmer ones, called interglacials. The consensus of evidence from a number of different studies (Augustin *et al.* 2004 and references therein) suggests that these glacial-interglacial cycles have been remarkably regular with durations of each at around 100 Kyr over the last 500 Kyr (and shorter cycles

during the earlier period). During the glacial episodes, extensive ice sheets formed and grew toward lower latitudes, and thus covered larger parts of Northern hemisphere continents than during warmer periods. As more water became trapped in the ice sheets, glaciation led to a drop in global sea levels. Studies have shown that the sea level during the time when ice sheets reached their maximum in the last glacial cycle - the Last Glacial Maximum (LGM) at around 20-25 Kyr before present (Clark *et al.* 2009) - was more than 100 metres lower than at present (Lambeck *et al.* 2002; Voris 2000). Similar sea level changes are likely to have also occurred during earlier glacial episodes (e.g. Siddall *et al.* 2003).

Genetic data have suggested far-reaching effects of the Quaternary glacial cycles on terrestrial taxa from tropical to arctic regions (Hewitt 2003). For those in northern temperate Europe, rapid expansions of populations northward following the retreat of ice sheets during the interglacial episodes have been proposed to contribute to a common pattern of reduced genetic diversity in northern populations relative to southern ones (Hewitt 2004; Rossiter *et al.* 2007), due to a series of founder events during range expansion (Hewitt 1996; Ibrahim *et al.* 1996). Glacial cycles have also had major impacts on divergence due to the isolation of populations into spatially separate refugia during glacial episodes and admixture of these diverged genomes, often in contact zones following post-glacial expansions (Moritz *et al.* 2009; Swenson & Howard 2005). Indeed allopatric divergence of populations in separate glacial refugia has been suggested as a process facilitating speciation (e.g. Hoskin *et al.* 2005; Knowles 2001; Richardson *et al.* 2001; Vila *et al.* 2005; Weir & Schluter 2004).

Quaternary glacial cycles have had rather different influences on the diversification of terrestrial taxa in tropical regions. However, to date, far fewer phylogeographic studies have been undertaken on tropical taxa (Hewitt 2004; Moritz *et al.* 2000) and, consequently, much less is known about the extent and pattern of genetic diversity in the tropics, or about the processes that have produced this diversity. This is unfortunate given that these population processes must ultimately be responsible for generating the enormous species diversity of tropical taxa (Plana 2004; Solomon *et al.* 2008a), and that understanding the genetic structure of tropical populations is key to conserving the genetic and species diversity of tropical environments.

Lowland tropical regions, such as those of Southeast Asia, were not directly affected by ice sheets during the LGM (or earlier more severe glacial episodes). In fact, recent

evidence has even suggested that humid forests persisted in the Sundaland area of Southeast Asia throughout the LGM, although the extent of past forest coverage varies depending on the data used for habitat reconstruction (Cannon *et al.* 2009; Wang *et al.* 2009; Wurster *et al.* 2010). It is thus likely that tropical taxa survived in this and other low latitude regions over multiple glacial cycles, and that this has contributed to high levels of genetic differentiation and/or divergence and complex histories reported for many tropical fauna (Bermingham & Martin 1998; Garcia-Paris *et al.* 2000; Nater *et al.* 2011; Vidya *et al.* 2009).

One aspect of past climatic oscillations that appears to have had had a major indirect impact on populations of many tropical regions is the associated sea-level fluctuations. Southeast Asia is a region with hundreds of islands, which are connected with each other and with the continent by a shallow continental shelf. This continental shelf was largely exposed during LGM (Voris 2000) due to the drop of sea level, and thus land-bridges were formed that allowed the exchange of terrestrial fauna and flora among formerly isolated landmasses. This periodic isolation and exchange of terrestrial taxa would have brought together genomes that had allopatrically diverged during periods of isolation. Past multiple incursions of taxa into one area via land-bridges can leave complex population genetic signatures; for example, paraphyly can result where some individuals in an island population share close relatives with their parental mainland population. Such patterns have been observed in a number of pairs of taxa in tropical Southeast Asia (e.g. Guicking *et al.* 2011; Ueda *et al.* 2010) including horseshoe bats distributed between the Asian continent and an offshore island separated by a shallow continental shelf (Mao *et al.* 2010).

Mountains

Mountains have been suggested to have a number of possible roles in promoting diversification. First, they can act as physical barriers across which genetic exchange is prevented or reduced (Brown *et al.* 2002; Swenson & Howard 2005; Xu *et al.* 2010). Second, montane regions are also frequently characterised by heterogeneous topography that can interact with the climate to create local microclimates. This means that parts of montane regions might be have served as refugia for taxa wiped out elsewhere during inhospitable climatic episodes. Geographical isolation between such montane refugia could have

subsequently led to accelerated allopatric divergence. Indeed this type of scenarios has been used to explain the observed divergence of lineages of montane taxa in tropical America and Africa (Fjeldså 1994; Fjeldså & Lovett 1997; Fjeldså & Rahbek 2006; Missoup *et al.* 2012; Roy 1997; Taylor *et al.* 2009; Voelker *et al.* 2010). The final way in which mountains could promote diversification is via the impact of temperature zonation along the altitudinal gradient. Such zonation could induce reproductive isolation between taxa showing differential adaptation to clinal environments, so representing a form of parapatric speciation (e.g. Guarnizo *et al.* 2009; Hall 2005). However, few empirical studies have provided compelling evidence for this latter mode of speciation (reviewed in Keller & Seehausen 2012), which reflects either genuine rarity of such speciation events in nature or the inherent difficulties of demonstrating these events (see next chapter for details). Nonetheless it is noteworthy that several investigations have indicated that mountains in tropics harbour greater genetic diversity than surrounded lowland regions, a trend also described from some temperate regions (e.g. Fjeldså & Lovett 1997; Plana 2004).

Ecological considerations in phylogeography

While similar genetic patterns among co-distributed taxa may point to common forces driving diversification (as described above), naïve application of this approach ignores some potential complexities. First, similar genetic patterns can arise from different geological and/or historical processes (e.g. Creer *et al.* 2001; Soltis *et al.* 2006), and second, co-distributed taxa can respond differently to their physical environment (e.g. Marske *et al.* 2012; Rossiter *et al.* 2012). Overemphasis of common genetic patterns may thus obscure intrinsic differences between taxa that could have important implications for their diversification, as exemplified by the following two studies that consider the ecology of focal taxa in a geographic context.

Handley *et al.* (2007) showed in their review that genetic variation among human populations across the world can be very well described by the geographic distance between samples, revealed by an isolation by distance analysis. This highlights that genetic variation of humans is a cline. On the other hand, other studies (e.g. Rosenberg *et al.* 2002), using clustering-based analyses implemented in the program STRUCTURE (Pritchard *et al.* 2000),

showed clear genetic clusters of human populations highly congruent with the presence of major barriers such as oceans and the Himalayas. When specifying a model which describes the level of gene flow as decreasing with geographic distance among samples while prohibiting gene flow directly across major barriers, Handley *et al.* (2007) showed an only minor improvement in model fit (2%) compared to the model considering geographic distance alone. This suggests that human genetic diversity may be distributed as a cline rather than in discrete clusters. These two possibilities would be difficult to distinguish if ecological information - about likely barriers to gene flow for humans - was not added into analyses because both factors (geographic distance and barriers) have similar influences on population differentiation.

Ecological information of extant populations can also be used to reconstruct the likely past distributions of species (e.g. reviewed in Phillips *et al.* 2006b). This information can be then be used to strengthen inferences made from comparative phylogeography while also providing insights into differences among species. Waltari *et al.* (2007) used two species distribution models to hindcast the historical ranges of 20 North American vertebrates during the LGM. For 14 of these species, the authors found strong spatial correlations between the inferred LGM distributions and the inferred locations of refugia as determined from genetic data.

Studies of Taiwanese taxa

Taiwan is a continental island around 150 km off mainland Asia, in the subtropical zone at approximately 22 to 25º north. It is a small island with a land area of 36000 km^2 and is longer than it is wide (400 km in length and 150 km in width). While Taiwan is widely accepted to have formed from the collision between the Luzon arc and the Eurasia plate, the detailed orogeny of Taiwan is still debated (reviewed in Simoes *et al.* 2012). New evidence has suggested that the collision progressed from north to south, with most (though not all) studies suggesting the collision in both the north and central parts of Taiwan started at around 6.5 to 4 Myr before present (reviewed in Kirstein *et al.* 2010). Therefore these parts of Taiwan might have emerged above the sea no earlier than 4 Myr ago (Pliocene), although before this can be more reliably estimated, more data are needed on the extent of temporal

and spatial variation in the pace of deformation of Taiwan (Simoes *et al.* 2012). The topography of Taiwan appears to have formed very rapidly since the relatively recent arc-continental collision, resulting in the complex landscape currently observed and, in support, Taiwan has one the highest long-term exhumation rates in the world (reviewed in Kirstein *et al.* 2010).

Currently one third of Taiwan's land area is at 1000 m or higher above sea level (a.s.l.). Montane areas are mostly concentrated in the centre of Taiwan, forming the Central Mountain Range (CMR) in the east, Hseuhshan Range in the northweast, and the Yushan and Alishan Ranges in the central-southwest, which together account for over 260 peaks with altitudes exceeding 3000 m a.s.l. A fifth mountain range, the Coastal Range, is much lower at < 1000 m a.s.l, and lies east of the CMR, separated by the Longitudinal Valley. Due to their complex topographies and large altitudinal ranges, Taiwan's mountain ranges offer a great range of habitats to its biota (e.g. Yu 1994). In addition, diversification of the Taiwanese terrestrial biota is likely to have been driven by past connections between Taiwan and the continent during glacial maxima, including the LGM (Fairbanks 1989; Voris 2000). Specifically, successive connection and isolation of Taiwan and mainland Asia will have allowed exchanges of terrestrial taxa between the two regions interspersed with periods of allopatric divergence (López-Pujol *et al.* 2011). Together these factors help to account for the island's very high recorded level of endemism (Mittermeier *et al.* 2005).

Phylogeographic studies have been conducted on a number of Taiwanese terrestrial taxa, most of which focused on examining genetic variation in mitochondrial DNA only (e.g. see review of Huang & Lin 2011). Results have revealed multiple well-defined genealogical groups (e.g. Shih *et al.* 2006; Wang *et al.* 2004; Wang *et al.* 2000) suggesting a complex phylogeographic history for each of these species, as also seen in tropical American species (e.g. Garcia-Paris *et al.* 2000). In particular, many have shown divergent lineages with varying degrees of geographic overlap (Creer *et al.* 2001; Hsu *et al.* 2001; Huang & Lin 2010; Jang-Liaw & Lee 2009; Jang-Liaw *et al.* 2008; Oshida *et al.* 2006; Shih *et al.* 2006; Wang *et al.* 2004; Yeh *et al.* 2004; Yuan *et al.* 2006). Vicariance due to isolation of populations into insular montane refugia was proposed for highland taxa showing north-south segregations of distinct lineages (Hsu *et al.* 2001; Yuan *et al.* 2006) although these authors had different opinions on the presence of such refugia during glacial or interglacial episodes. The presence

of glacial refugia has also been proposed for lowland taxa at different locations (Cheng *et al.* 2005; Oshida *et al.* 2006). To date, the population genetics of only one bat species in Taiwan, *Rhinolophus monoceros*, has been reported (Chen *et al.* 2008; Chen *et al.* 2006), which showed isolation by distance at both mitochondrial and nuclear markers.

Reported cases for many lowland taxa (ranges ≤ 2000 m a.s.l.) of reduced or no genetic exchange across the CMR indicates that these mountains serve as an effective barrier to gene flow between populations on its opposite sides (Creer *et al.* 2001; Jang-Liaw *et al.* 2008; Oshida *et al.* 2006; Shih *et al.* 2006; Yeh *et al.* 2004). However, these studies provide little evidence to support a direct role of the CMR in initiating divergence because an alternative explanation - that these divergent lineages represent successive incursions of populations into Taiwan via land-bridges during glacial - appears to be equally or even more likely.

Scenarios of multiple incursions appear to have received best support from empirical studies in which samples of related taxa or populations from mainland China were also analysed (Tzeng *et al.* 2006; Yu 1995). Here some Taiwanese lineages were found to cluster with continental ones on the basis of genetic similarity. However, most such studies of Taiwanese taxa have not used multiple genetic markers to take account of the variance of genetic patterns, nor have they used modern statistical approaches that can rigorously test among different phylogeographic hypotheses. For these reasons, inferences of recolonization events on the basis of non-monophyly alone must be treated with caution due to the stochastic effect of lineage sorting. In fact, for rigorous discrimination among alternative phylogeographic hypotheses, several pieces of information should be considered, including the geographic pattern of lineages with respect to the presence of geographic barriers (Avise 2000), the timing of divergence and presence/absence of migration since then (Becquet & Przeworski 2007), the ecology of the subject taxa including their potential ability to disperse across geographic barriers and over distances (Hewitt 2004), as well as information about the geology, paleoclimatology, and paleoecology of Taiwan (e.g. Hebenstreit *et al.* 2006; Liew *et al.* 1998; Sibuet & Hsu 2004).

Study objectives

I conducted a comparative phylogeographic study of four Taiwanese bat species: *Murina gracilis*, *M. recondita*, *M. puta*, and *Kerivoula* sp. (see Chapter 1 for background information). To elucidate the common and different forces that have shaped range-wide patterns of population genetic structure, for each species I generated sequence data from two mitochondrial genes as well as genotype data from multiple microsatellite loci. I applied a combination of population genetic and phylogenetic methods; in particular, coalescent-based demographic analyses were used to date crucial phylogenetic and demographic events, which were then related to events in the geological history of Taiwan. I also undertook spatial distribution modeling of each of the focal species with the aim of providing explicit links between genetic and geographic patterns. First, I hypothesized that each of these four forest interior species would show evidence of restricted gene flow across Taiwan, related to their low dispersal ability. Second, I hypothesized that the magnitude of the effect of mountain barriers on gene flow would differ among the four taxa, corresponding to their tolerance to high altitude. Specifically I predicted that species that occur at higher elevations would be less affected by mountains and thus would show less population subdivision.

3.2 MATERIALS & METHODS

For comparative phylogeographic analyses, I analysed DNA from four focal bat species (*Murina gracilis*, *M. recondita*, *M. puta* and *Kerivoula sp.*), all collected from across their respective ranges. In addition, I included a sample of a continental species, *M. eleryi*, in the phylogenetic analyses. This sample of *M. eleryi* was requested from Hungarian Natural History Museum (HNHM) and represents a specimen (ID: 2007.28.2) collected from Vietnam (Bac Kan Province: Kim Hy Nature Reserve: 22°14.849'N, 105°58.376'E).

Tissue collection and DNA extraction are both described in Chapter 2. I used two sets of markers that show contrasting modes of inheritance: mitochondrial DNA (mtDNA) and polymorphic microsatellite loci (ncDNA).

Selection of genetic markers

For mtDNA analysis of phylogeography and population structure, I focused on sections of two loci: the 3′-end of the cytochrome b gene (Cyt-*b*) and the barcoding fragment of the cytochrome c oxidase subunit 1 gene (COI). Both of these were screened in *M. gracilis* and *M. recondita*, however, due to problems of amplification, only the former was amplified in *M. puta* and the latter in *Kerivoula* sp. Similarly, sequences of both Cyt-*b* and COI were obtained for *M. eleryi* while only Cyt-*b* sequence was obtained for *M. tiensa*.

The Cyt-*b* fragment was amplified using the primer pair LMu313m (5′- GGA TCC TAC ATA TAT AAA GAG ACR TG -3′; modified from Kuo (2004)) and BSves268H (5′- ATT TCT GGY TTA CAA KAC CRG TGT AA -3′) (Stadelmann *et al.* 2004a). For amplifying the COI fragment, I used the primers BatCO1F (5′- TTT TCA ACC AAY CAY AAA GAY ATY GG -3′) and BatCO1Ra (5′- TAT ACT TCY GGG TGR CCR AAG AAY CA -3′), both of which were based on modified sequences obtained from Ivanova *et al.* (2006). Polymerase chain reactions (PCR) were carried out using the Type-it Microsatellite PCR Kit (Qiagen) with a final volume of 10 μL of the reaction cocktail containing 0.2 μM each primer, 1.6 μL of template, and 5 μL of the provided master mix for each sample. PCRs were performed on a C1000 Thermal Cycler (BIO-RAD Laboratories) with the following

thermal profile: 95 ºC 5 min; 36 cycles of 95 ºC 30 sec, 55 ºC 90 sec, 72 ºC 30 sec; 60 ºC 30 min. After checking band quality on a 2% agarose gel, amplified products were sent to Tri-I Biotech, Inc. (New Taipei City, Taiwan) where sequencing reactions were conducted on an ABI 3730 DNA Analyzer (Applied Biosystems). For resulting sequences for each mitochondrial gene, I first inspected the chromatograms by eye to check for mismatches, and then aligned the sequences using the MUSCLE algorithm (Edgar 2004) implemented in the software MEGA 5 (Tamura *et al.* 2011).

I also investigated population genetic structure of all four focal species based on multi-locus microsatellites. Protocols and procedures for amplification and genotyping of microsatellites for Taiwanese *Murina* and *Kerivoula* bats are described in Chapter 2. Each species required a different panel of markers. For *M. gracilis* and *M. recondita* I used the 20 autosomal markers listed in Table 2.3, however, due to poor amplification results, A107, B103, and D9 were not used for the former species, and A103, B12, and B103 were not used for the latter one. The X-linked marker, B9, was added into corresponding sets of these two species for the Bayesian clustering analyses (see *Genetic clusters* in the section 'Population genetics revealed by microsatellites'). For *M. puta* I was able to obtain good results with nine of the 20 markers, all of which were retained for analysis (A9, A10, A112, A122, B5, B114, B121, D110, and D117), and for *Kerivoula* sp. I used eight markers developed for a related *Kerivoula* species, described in Struebig *et al.* (2008) (Kpa02, Kpa05, Kpa08, Kpa24, Kpa25, Kpa26, Kpa32, and Kpa46), as well as one *Murina* marker described in this study (A112). See Chapter 2 for characterisation of all above microsatellite markers and procedures for marker selection for the four focal species.

Population genetics revealed by mtDNA

Sequence analyses and genetic variability of demes

For each of the three focal *Murina* species, Cyt-*b* sequences were edited to give a final alignment of 768 bp, while for each of *M. gracilis*, *M. recondita*, and *Kerivoula* sp., COI sequences were edited to give a final alignment of 640 bp. To verify the identities of these amplified fragments and ensure that stop codons were not present, indicative of pseudogenized copies, all sequences were translated into amino acids using MEGA 5. In

addition, for each of the six alignments (i.e. each gene for each species) the program DnaSP 5.10.01 (Librado & Rozas 2009) was used to calculate proportions of synonymous versus non-synonymous nucleotide substitutions. For each gene, the McDonald-Kreitman test (McDonald & Kreitman 1991) was applied to the two closely related species, *M. gracilis* and *M. recondita* (see Results), to test for potential positive selection acting on these genes. The rationale behind this test is that under such selection, fixed nucleotide differences between related taxa will show a greater proportion of non-synonymous changes than synonymous ones in either taxon. These tests were implemented with DnaSP 5.10.01.

To assess the genetic variability of demes (defined in this study as sampling sites) within each focal species, conventional measures including the number of segregating nucleotide sites (S), nucleotide diversity (π), and haplotype diversity (Hd) were used. These were estimated for each gene separately, and also concatenated. Considering that the value of the number of segregating sites is expected to depend on the sequence length and the sample size, I also estimated Watterson's (1975) theta (θ_S) to facilitate comparisons across genes. Using DnaSP 5.10.01, all these measures were estimated for individual demes of each focal species with samples of five or more individual bats.

Genetic differentiation among demes and tests for isolation by distance

For each of the four species, global and pairwise genetic differentiation among demes was estimated based on the F_{ST} analogue statistic Φ_{ST} (Excoffier *et al.* 1992). This statistic quantifies differences among populations by taking into account both the frequencies of unique haplotypes and also the nucleotide differences among haplotypes. Significance was evaluated by permutation tests (i.e. by shuffling sequences among demes to produce the appropriate null distributions of Φ_{ST} assuming no differentiation, against which the observed value was compared). For global tests 2000 permutations were used, and for pairwise comparisons 5000 were used followed by a sequential Bonferroni correction (Holm 1979) to correct for multiple tests. The above analyses were implemented with the program Arlequin 3.5.1.2 (Excoffier & Lischer 2010), and only demes with five or more individual bats were included in analyses. For each of *M. gracilis* and *M. recondita*, analyses were separately applied to datasets comprising the Cyt-*b* gene on its own (for which there was better spatial

coverage of sampled demes in *M. recondita* and usually larger sample sizes for demes in both species) and the concatenated sequence of Cyt-*b* and COI genes.

Tests of isolation by distance (IBD) were conducted to determine whether within each species the spatial pattern of genetic differentiation based on Φ_{ST} was consistent with a stepping stone model of gene flow. For these models, model fit was assessed from the coefficient of determination (R^2), and Mantel tests were performed to assess significance. Following initial exploration of the data, I observed that Φ_{ST} was not an appropriate measure of genetic difference among demes in both *M. gracilis* and *M. recondita* because each of these two species was found to be characterised by the presence of highly divergent mitochondrial clades coupled with a large number of haplotypes that were private (i.e. unique) to individual sampling sites (see detailed explanations in the corresponding sections in the Results). Therefore, descriptions of IBD analyses described below based on mtDNA only apply to *M. puta* and *Kerivoula* sp.

For IBD analyses, Φ_{ST} values between pairwise demes were transformed into $\Phi_{ST}/(1-\Phi_{ST})$ and then regressed against corresponding pairwise spatial distances following Rousset (1997). In additional to Euclidean geographic distances (EGD; sometimes referred to as the uncorrected distance in this study) that are conventionally used, I also investigated IBD models using two other measures of spatial distance (sometimes referred to as corrected distances in this study). These modified measures of geographic distance were the 'least-cost path distance' (LCPD) and the 'resistance distance' (RD). Both of these corrected measures are considered to represent effective spatial distances that better account for landscape effects and barriers, such as the mountains that occur in central Taiwan. The application of the LCPD (e.g. Adriaensen *et al.* 2003; Arnaud 2003) is based on the idea that dispersers will take a single route to commute between any pair of localities. More recently, McRae (2006) proposed the RD to take into account the probabilities of dispersal along each of all possible routes connecting localities. This approach uses algorithms based in electrical circuit theory, with the RD representing an 'average' resistance over resistances along each of all possible routes. Specifically, the resistance along each route is taken to be analogous to a series of resistors in an electrical circuit, and the average resistance will be inversely related to the number of parallel routes (an analogue to parallel connections) so that localities connected by a wider patch of permeable habitats will have a lower average resistance than those

connected by a narrower patch of habitats with the same level of resistance for each unit of habitat (McRae 2006; McRae *et al.* 2008).

To calculate all spatial distances described above (EGD, LCPD, and RD), average GPS coordinates were used when individual sampling sites comprised more than one locality. Values of EGD were calculated directly from GPS coordinates using the software Geographic Distance Matrix Generator 1.2.3 (Ersts 2011). For estimating estimates of LCPD and RD, a resistance map was needed that could be superimposed on the landscape. To generate these resistance maps I undertook distribution modeling in the software MaxEnt, which provided estimates of the limits and coverage of suitable (and conversely unsuitable) habitat. Based on the corresponding MaxEnt model, binary predictions of suitable habitats versus unsuitable ones were carried out for each of the two species using specific thresholds. I assumed that each unit of such suitable habitats was permeable with the same cost, while each unit inferred as unsuitable was impermeable to dispersers. For detailed description of MaxEnt model development and selection of thresholds for binary predictions, see the section below on *Species distribution modeling*.

Pairwise LCPDs were estimated using the software PATHMATRIX 1.1 (Ray 2005) and pairwise RDs were estimated using the software Circuitscape 3.5 (McRae & Shah 2009). For the latter measure, dispersal was assumed to only occur between the first-order neighbours (i.e. between each map unit and its four cardinal neighbours). For *Kerivoula* sp., I further replicated all above analyses on a subset of nine demes from western side of the CMR. These were conducted to investigate potentially different IBD patterns in demes of this lowland species from different sides of CMR; only a few well-sampled demes from eastern side of CMR were available, and thus IBD tests were not applied to these demes.

Phylogenetic analyses

Prior to undertaking demographic analyses it was necessary to estimate a molecular clock for Cyt-*b* and COI for each of the focal species. For this I expanded taxonomic sampling in Kuo (2004) and built a mitochondrial gene tree containing representatives of 44 species of subfamilies Kerivoulinae, Murininae and Myotinae (see Appendix 3.1 for materials). Two mitochondrial genes, Cyt-*b* (1140 bp) and NADH dehydrogenase subunit 1

(ND1; 800 bp), were used since sequences of these genes have already been published on GenBank for a wide range of species from the focal groups. Importantly, both gene sequences were nearly always obtained from the same individual bat of each species (see Appendix 3.1 for references) so avoiding problems of taxonomic uncertainty for some of these taxa (see Chapter 1; see also Ruedi & Mayer 2001).

I applied a Drummond *et al.*'s (2006) algorithm implemented in the program BEAST 1.6.1 (Drummond & Rambaut 2007), which uses the Bayesian method to simultaneously estimate the phylogeny together with dates of internal nodes. Drummond *et al.*'s (2006) uncorrelated lognormal relaxed clock algorithm was used to model variation in branch length across the tree. Two splits were used for calibrations: the root of the three analyzed subfamilies (23.3 ± 2.25 Myr) and the time to the most common recent ancestor (TMRCA) of all New World bat species of the genus *Myotis* (9.1 ± 1.3 Myr). Dates of my calibration points were estimates taken from Lack *et al.* (2010), who used concatenated sequences of multiple mitochondrial and nuclear markers to build a phylogeny that included representatives of four bat families as the ingroup (yet with fewer species of Kerivoulinae and Murininae than included in my study). Lack *et al.* (2010) used early fossil records to calibrate their phylogeny.

Prior to Bayesian tree-building, sequence data were partitioned based on codon position: the combination of the first and second codon positions as one partition, CP_{12}, and the third codon position as the other one, CP_3. This type of data partitioning has been shown to result in superior phylogenetic inference when applied to protein coding sequences (Shapiro *et al.* 2006). For each of these partitions, I used the program jModelTest 0.1.1 (Posada 2008) to estimate, via maximum likelihood optimization, the log-likelihood of the data under each of 24 alternative nucleotide substitution models, and I then performed model selection based on the Akaike Information Criterion (AIC). The final selected model for both CP_{12} and CP_3 partitions were the generalized time reversible nucleotide substitution model (GTR, Tavaré 1986) plus rate heterogeneity among sites modeled both by a proportion of invariable sites and by the gamma distribution (GTR+I+G). The branching of the tree was modeled with a constant-branching prior, also called Yule prior (Yule 1924). Two replicated runs of Bayesian MCMC were implemented with BEAST 1.6.1, and each had a total of 50 million iterations (2000 iterations per sample) with the first 10% discarded as burn-in. Since

the two runs gave virtually the same result, their results were combined via the BEAST adjunct software, LogCombiner.

The resulting tree from the above analyses is presented in Fig. 3.4 (see Results for details). Based on this tree, I was also able to examine the intraspecific phylogenies of *M. gracilis* and *M. recondita* by performing further phylogenetic analyses on the pooled sample of these species with *M. eleryi* as an outgroup, as described below. Concatenated sequence of Cyt-*b* and COI was used to improve resolution of these latter analyses. Also based on the tree in Fig. 3.4, another continental *Murina* species, *M. tiensa*, was used as an outgroup for investigating intraspecific phylogeny of *M. puta* using the Cyt-*b* gene (accession number for *M. tiensa*: GQ168913). For the intraspecific phylogeny of Taiwanese *Kerivoula* sp. where the COI gene was used, I queried the nucleotide collection (nr/nt) database of the National Centre for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/) for the closest sequences using the default setting of MEGABLAST. Nine COI sequences of *K. cf. hardwickii* from mainland China showed 99% identity and 100% coverage with the query *Kerivoula* sp. sequence and were thus used as outgroups (accessions HM540676-540679 and HM540681-540685). In addition I included homologous sequences from one *K. cf. hardwickii* from Vietnam (accession number HM540704; 91% identity to the queried sequence) and one *K. kachinensis* from Laos (accession number HM540737; 90% identity to the queried sequence) as sequential outgroups.

'Intraspecific' trees that also included the interspecific relationship of *M. gracilis* and *M. recondita* were built using maximum parsimony (MP) and Bayesian methods. For MP, all nucleotide substitutions were weighted equally, and heuristic searches for the most-parsimonious tree were performed based on ten replicates of random addition of sequences, each followed by tree bisection recombination (TBR) tree search. Standard bootstrapping with 1000 replicates was used to evaluate levels of support for clades; for each bootstrap replicate, tree searching was performed with one random sequence addition followed by TBR. These above analyses were implemented with the program TNT 1.1 (Goloboff *et al.* 2008).

For each Bayesian analysis, sequence data was again partitioned into CP_{12}, and CP_3. I used the following nucleotide substitution models selected with jModelTest 0.1.1 for the pooled data of *M. gracilis* and *M. recondita*: HKY (Hasegawa *et al.* 1985) plus rate heterogeneity among sites modeled both by a proportion of invariable sites and by the

gamma distribution (HKY+I+G) for partition CP_{12}, and GTR plus rate heterogeneity among sites modeled by the gamma distribution (GTR+G) for $CP_{3.}$ For *M. puta* and *Kerivoula* sp., the HKY nucleotide substitution model was used for each partition. Using corresponding models for the two partitions, Bayesian phylogenetic reconstruction was carried out in MrBayes 3.1.2 (Ronquist & Huelsenbeck 2003). For each dataset, two independent runs of a Metropolis-coupling Markov chain Monte Carlo (MC^3) were performed, each with seven heated chains (temperature parameter = 0.1) and with a burn-in of 0.2 million iterations followed by a sampling phase of 0.8 million iterations (100 iterations per sample).

Tree-based phylogenetic methods are expected to be able to represent relationships among sequences more accurately where ancestral sequences have gone extinct. However, this is unlikely to be true of DNA haplotypes in natural populations, where ancestral haplotypes quite often co-exist with their descendants (Posada & Crandall 2001). Therefore, I also investigated phylogenetic connections among mitochondrial haplotypes within each focal species using a network method which relaxes the unrealistic assumption of tree-based methods. I built such networks using the median-joining algorithm (Bandelt *et al.* 1999) implemented with the program Network 4.6.0.0 (Fluxus Technology, www.fluxusengineering.com). The resolution of built networks was examined under sequential numbers of the epsilon parameter. Following extensive testing on all datasets, I built a separate network for each of the four species under an epsilon value of 10, removing unnecessary median vectors and links using the Steiner MP algorithm (Polzin & Daneshmand 2003) in order to provide good network resolution and interpretable connections.

Dynamics of effective population sizes

Two types of approach were used to investigate the dynamics of effective population sizes for demes or groups of demes of each species. First I undertook neutrality tests based on the statistics Fu's F_S (1997) and Ramos-Onsins and Rozas' R_2 (2002). The statistic F_S aims to detect an excess of rare haplotypes with reference to Ewens' (1972) sampling distribution for different haplotypes, whereas R_2 aims to detect an excess of singleton mutations per sequence in relation to pairwise differences between sequences. Simulations by

Ramos-Onsins and Rozas (2002) and Ramirez-Soriano *et al.* (2008) showed these two tests to be among the most powerful for detecting changes in the population size under a range of conditions. In particular Ramirez-Soriano *et al.* (2008) demonstrated power to detect population contraction or a weak bottleneck at the left tail of the null distribution, as well as population expansion or a strong bottleneck at the right tail. Here I applied two-tailed tests to concatenated sequences of Cyt-*b* and COI for individual demes of *M. gracilis* and *M. recondita* (for samples of ≥5) since both taxa showed strong demic genetic structure (see Results). Neutrality tests were also applied to samples of *M. puta* (Cyt-*b*) and *Kerivoula* sp. (COI), as well as to groups of demes of *K.* sp. from the western and eastern side of CMR. All null distributions were produced by simulation of the coalescence while controlling the number of observed segregating sites and the sample size, implemented with the program DnaSP 5.10.01. Note that a number of demes of *M. gracilis* and *M. recondita* contained no intra-demic genetic variability (and thus no coalescent information) and so were excluded from neutrality tests as well as from the following demographic analyses.

To investigate potentially more complex effective population size (Ne) dynamics in the focal demes, I also undertook extended Bayesian skyline plot analyses (EBSP, Heled & Drummond 2008) on the same datasets as described above. Skyline plot methods were first developed from coalescent theory to estimate the dynamics of Ne based on lengths of successive coalescent intervals in an *a priori* estimated dated gene tree for a given population (Pybus *et al.* 2000; Strimmer & Pybus 2001). More specifically, the coalescent time of each such interval is expected to be proportional to Ne, such that the successive lengths of these intervals through time can provide information on the Ne (Pybus *et al.* 2000). Nevertheless, estimating Ne based on every interval would result in an uneven dynamic curve due to stochastic variance of the coalescence process. To avoid such 'noise', Strimmer and Pybus (2001) proposed a method to lump adjacent coalescent intervals into 'coalescent groups', which are bound by changes in Ne. Strimmer and Pybus' (2001) method was extended by Drummond *et al.* (2005) who developed the Bayesian skyline plot (BSP) to incorporate estimation of Ne during each coalescent group and the gene tree (both estimated directly from the sequence data) in a Bayesian framework, so that the uncertainties associated with the tree could also be taken in account. The EBSP builds on the BSP by allowing analysis of multiple sequence loci, and also offers the following two main advantages: (1) allowing linear change in Ne during each coalescent group (piecewise-linear model; versus piecewise-

constant one in old methods) to reflect better the continuous nature of dynamics in the population size, and (2) parameterizing and then estimating the number of historical changes in the population size rather than specifying an arbitrary number as was the case for BSP (Heled & Drummond 2008).

For estimating the tree for each dataset, sequence data were partitioned as for the phylogenetic analyses described above (i.e. CP_{12} and CP_3) and the Hasegawa-Kishino-Yano nucleotide substitution model (HKY, Hasegawa *et al.* 1985) applied to both partitions. To calibrate the tree, a strict molecular clock was assumed. For *M. gracilis* and *M. recondita*, the rates of the two partitions were specified with lognormal priors as follows: ln(Mean) = -5.8, ln(SD) = 0.17 for CP_{12} and ln(Mean) = -2.7, ln(SD) = 0.17 for CP_3. For *M. puta*, the clock rate of CP_{12} was specified with a log-normal prior as ln(Mean) = -5.5, ln(SD) = 0.17 while that of CP_3 was specified with a uniform prior between zero and one. For *Kerivoula* sp., the clock rate of CP_{12} was specified with a log-normal prior as ln(Mean) = -6.3, ln(SD) = 0.19 while that of CP_3 was specified with a uniform prior between zero and one. Rationale and procedures for these prior settings are described in the next paragraph. Two (unless otherwise stated in the Results) independent Markov chain Monte Carlo (MCMC) runs were performed for each dataset via the program BEAST 1.6.1 with the corresponding input file prepared using the adjunct program BEAUti. Each MCMC was run for a total of 50 million iterations (2,000 iterations per sample). Consistency in the marginal distributions of estimated parameters as well as their effective sampling sizes (ESS) was inspected with the program Tracer 1.5 (Rambaut & Drummond 2007) after discarding the first 10% of each run as burn-in (determined by inspecting trends of parameters along the MCMC chain). After confirming consistent estimation in Tracer, results from replicated runs were combined in LogCombiner. A generation time of two years was used for each of the four focal species to convert the re-scaled effective population sizes outputted by EBSP into demographic ones.

Clock rates (R) for the CP_{12} and CP_3 partitions for the four focal species were estimated from divergence times (T) in the tree shown in Fig. 3.4, using the formula R = D/2T where D is the sequence divergence for the corresponding split event. Such levels of sequence divergence (D) for the two partitions were estimated via the web-based software DIVEIN (Deng *et al.* 2010) based on nucleotide substitution models specific for each dataset. From Fig. 3.4, T between *M. eleryi* and the common ancestor of *M. gracilis* and *M. recondita*

was 2.6 (95% CI: 1.9-3.5) million years (Myr) before present. Associated with this T, D in the concatenated sequence of Cyt-*b* and COI were 0.016 for CP_{12} and 0.354 for CP_3 (per nucleotide site per Myr), estimated based on corresponding models in *Phylogenetic analyses* described above for these same species (HKY+I+G for CP_{12} and GTR+G for CP_3). Thus, clock rates (U) for both *M. gracilis* and *M. recondita* were estimated as 0.0031 (95% CI: 0.0023-0.0043) for CP_{12} and 0.0676 (95% CI: 0.0509-0.0952) for CP_3 (per Myr) based on the formula U = D/2T. From Fig. 3.4, T between *M. puta* and *M. tiensa* was 7.1 (95% CI: 5.1-9.3). Between these species, D in the Cyt-*b* were 0.058 for CP_{12} and 0.469 for CP_3 based on the TN93 nucleotide substitution model (Tamura & Nei 1993). Note that, however, D for CP_3 was apparently an underestimate, judged by comparing the ratios of T and D in this dataset to corresponding quantities in previous datasets (D for CP_{12} was 3.5 times larger in the latter dataset, which was roughly proportional to the ratio of T. Yet D for CP_3 was only 1.3 times larger in the later dataset). I considered that substitutions in CP_3 in this latter dataset had been saturated, and thus only the clock rate estimated for CP_{12} was set as an informative prior in the EBSP analysis for *M. puta*. This clock rate of CP_{12} was estimated as 0.0041 (95% CI: 0.0031-0.0057) per Myr. From Fig. 3.4, T between Taiwanese *Kerivoula* sp. and *K. hardwickii* from Borneo was 6.8 (95% CI: 4.5-9.3) Myr. Comparing COI sequences of Taiwanese *K.* sp. with homologous sequences of *K. hardwickii* from Borneo (accession numbers: GU585629-585631 and GU585635), D estimates were 0.024 for CP_{12} and 0.586 for CP_3 based on the TN93 model. Given potential saturation of substitutions in CP_3, reflected by its high D value, I also set a uniform prior for the clock rate of this partition in EBSP analyses for Taiwanese *Kerivoula* sp., as for *M. puta*. The clock rate of CP_{12} for *K.* sp. was estimated as 0.0018 (95% CI: 0.0013-0.0026) per Myr.

Divergence between major mitochondrial clades in M. gracilis and M. recondita

Highly divergent and often allopatrically distributed mitochondrial clades were observed in both *M. gracilis* and *M. recondita*, suggesting vicariance has shaped current matrilineal genetic structure of these taxa (see *Population genetics revealed by mitochondrial DNA* in Results). To investigate the divergence processes of demes containing different mitochondrial lineages, I examined whether mtDNA nucleotide substitutions for the two focal species have evolved in a strict clock-like manner. These tests were applied to the two

data partitions of the concatenated sequence of the studied genes. I specified the maximum clade credibility (MCC) tree from the MrBayes 3.1.2 Bayesian analysis (described in *Phylogenetic analyses*, above) as the tree topology on which the strict molecular clock hypothesis is tested. This MCC tree was generated by TreeAnnotator in the BEAST 1.6.1 package, and the outgroup *M. eleryi* was removed. For each partition, I then used MEGA 5 to estimate, via the Maximum likelihood method, log-likelihood values with and without the strict clock constraint given the specified topology with the nucleotide substitution model for each partition selected earlier (see *Phylogenetic analyses*). These tests suggested violation of the strict clock model for both the CP_{12} (test statistic $2\Delta = 104.99$, $P < 0.001$) and CP_3 ($2\Delta = 119.07$, $P < 0.001$) data partitions.

Second, for each separate species, I used Bayesian methods implemented in BEAST 1.6.1 to estimate the TMRCA of divergent clades while co-estimating the genealogy in a coalescent-based framework. Note that all sequences including replicates were used in these analyses. For each data partition (CP_{12} and CP_3) I specified the HKY model for nucleotide substitutions. I used a less generalized model than those used in the phylogenetic analysis because 1) lower levels of sequence divergence are involved in these comparisons and because 2) complicated models of clock rates and of demographic changes are used. Keeping the collective model as simple as possible should improve mixing of the MCMC. To model variation in the rate of evolution among and/or within clades as suggested by the molecular clock test described above, I applied the uncorrelated lognormal relaxed clock model (Drummond *et al.* 2006) to each partition. The lognormal priors used in *Dynamics of effective population sizes* (i.e. ln(Mean) = -5.8, ln(SD) = 0.17 for CP_{12} and ln(Mean) = -2.7, ln(SD) = 0.17 for CP_3) were used as hyperpriors for the BEAST parameter 'ucld.mean' to calibrate the trees. Given that no coalescent model has been developed in BEAST for a subdivided population (which characterises both focal species) I used a range of available models for each species to assess the robustness of TMRCA values to model selection. These models included the constant-sized population model, the Bayesian Skyline Plot (BSP) model with effective population sizes modeled in a piecewise-constant manner and with a maximum of five demographic changes, and the BSP model with effective population sizes modelled in a piecewise-linear manner and with a maximum of five demographic changes. The protocol for Bayesian simulations under each model is the same as that in EBSP analyses in *Dynamics of effective population sizes*.

Finally, I used 'isolation with migration' (IM) models (Nielsen & Wakeley 2001) to investigate the divergence process between non-allopatric clades. IM models (Nielsen & Wakeley 2001) describe gene genealogies between closely related populations that have undergone a split with potential subsequent gene flow. By parameterizing the split time, the migration rates in both directions, and the effective sizes of both ancestral and descendant populations, Bayesian simulations can be used to disentangle the contributions of recent divergence (i.e. split time) and gene flow (modeled by post-split migration rates) to shared genealogical ancestry between populations (Hey & Nielsen 2004). Using concatenated mitochondrial sequences of the focal populations, I implemented IM analyses in the program IMa2 8.26.11 (Hey 2010b; Hey & Nielsen 2007). These were carried out on a high performance computer at Queen Mary University of London (64-bit Linux cluster with 24 CPUs and total memory of 264 GB). The HKY model was used for nucleotide substitutions. For details of the populations compared, and prior settings (priors are specified as upper bounds of truncated uniform distributions), see Results.

Given the expectation that the use of a single marker offers very limited information about the demography of ancestral populations, I further explored the model, restricting the number of parameters estimated by assuming that the ancestral population size was equal to the larger of the two descendant ones (-j8 option). Under each setting, I performed three replicates of MC^3 runs, each with ten chains specified with geometric heating (heating parameters -ha0.95 -hb0.9) and a sampling of five million iterations (100 iterations per sample) following a burn-in of one million. I applied a per locus mutation rate of 0.0000352 per calendar year, and a generation time of two years. This mutation rate was calculated from comparing concatenated sequences of the two focal species with that of *M. eleryi* (as described in *Dynamics of effective population sizes*, except for the use of the whole locus instead of partitions based on codon positions). The estimated split time between these taxa (2.6 Myr before present) corresponds to an average sequence divergence of 0.13 per base and thus the mutation rate can be calculated as 0.025 per base per million years, or 0.0000352 per locus per calendar year.

To test statistically for the presence of post-split migration between focal populations, I applied the likelihood ratio test (LRT) proposed by Nielsen and Wakeley (2001). Briefly, assuming that the likelihood ratio of presence ($m \neq 0$) against absence of migration ($m = 0$) is

proportional to their marginal posterior ratio when an 'uninformative' prior on migration is given, the LRT can be conducted based on the marginal posterior ratio (Hey & Nielsen 2007). For each migration parameter, the critical value against which the marginal posterior ratio is compared is 0.5*(chi-square, *df* = 0) + 0.5*(chi-square, *df* = 1) because the value of the parameter under the null hypothesis is at the boundary (i.e. zero) of the parameter space. When performing such LRTs, I examined both marginal posterior ratios calculated based on the entry migration parameter of the IM model, 'm', and the population migration rate (i.e. number of migrant gene copies per generation).

Population genetics revealed by multi-locus microsatellites

Genetic variability, genetic differentiation among demes and tests for isolation by distance

Each species was screened at its respective panel of microsatellite markers (see Chapter 2) and the following measures of genetic variability were estimated for all demes of five or more individual bats: allelic richness, observed heterozygosity and expected heterozygosity. To test for deviations from Hardy-Weinberg equilibrium (HWE), I calculated multi-locus F_{IS} according to the Weir & Cockerham estimator θ (1984), and tested for significance by comparing the observed value to a null distribution of 2,000 expected values obtained by permuting alleles within demes. All above analyses were undertaken in FSTAT 2.9.3 (Goudet 2001). For each species, significance values of F_{IS} were adjusted using a sequential Bonferroni procedure.

To investigate genetic differentiation within each species, I repeated the tests previously described (see 'Population genetics revealed by mitochondrial DNA') but used Weir & Cockerham's (1984) F_{ST} instead of Excoffier *et al.*'s (1992) Φ_{ST} as a measure of differentiation. As described in Chapter 1 (see 'Statistical phylogeography'), F_{ST} is influenced by the mutation rate for fast-evolving markers such as microsatellites and, therefore, it is important to disentangle the effects of mutation from gene flow in multi-species comparisons that use different loci. To do this, I plotted locus-wise global F_{ST} against and the level of average intra-demic heterozygosity for each taxon. Nichols and Freeman (2004: fig. 3) showed that under high inter-demic gene flow, loci will consistently show low F_{ST} regardless of the mutation rate, whereas heterozygosity will be consistently high for

rapidly evolving loci, but will vary widely for more slowly evolving loci. In contrast, under low inter-demic gene flow, F_{ST} values will depend on the mutation rate, with low values for rapidly evolving loci and widely varying values for slowly evolving loci. Under this condition, locus-wise heterozygosity values are always high.

Genetic clustering

For investigating discontinuous genetic variation within each species I used the methods implemented in the program Structure 2.3.2 (Pritchard *et al.* 2000), which assume HWE and linkage equilibrium within a number of genetic clusters. In this method, individuals or fractions of individuals are assigned to clusters via Bayesian estimation on the basis of their multi-locus genotypes. I applied clustering analyses to respective sets of autosomal microsatellites of the four focal species, and, for *M. gracilis* and *M. recondita*, I also included an additional X-linked marker (B9). For each dataset I specified a model that allowed admixture from multiple clusters (Pritchard *et al.* 2000) and correlated allele frequencies among clusters (Falush *et al.* 2003). I also explored a model in which sampling sites were specified as a prior in estimating ancestry (Hubisz *et al.* 2009: LOCPRIOR model).

Bayesian simulations were run under sequential numbers of clusters (K), ranging from one to six. Five independent runs of MCMC were conduct under each K, and each such run had a burn-in of 0.75 million iterations followed by an additional sampling of 0.75 million iterations (1000 iterations per sample). K was inferred following Pritchard *et al.* (2000), and Evanno *et al.*'s (2005) method was used where the likelihood of the data increased monotonically until K = 6. To explore potential multimodality revealed by replicate runs, the program CLUMPP (Jakobsson & Rosenberg 2007) was used to calculate the pairwise similarity statistic G′ among runs. The 'FullSearch' (for $K \leq 4$) or 'Greedy' permutation algorithm (for K = 5 or 6) was then used to find matched labels of clusters across runs by maximizing the average pairwise similarity among them. Following this procedure, runs were clumped where G′ values were ≥ 0.9. Finally, the software DISTRUCT (Rosenberg 2004) was used to visualize inferred ancestries of individual bats for each dataset.

Species distribution modeling (SDM)

To explore the possible impact of habitat on observed patterns of phylogeographic, population genetic and demographic structure, I constructed species distribution models (SDM) for each of the four taxa at both the present time and also at the time of the Last Glacial Maximum (LGM). For this, I used the maximum entropy ('maxent') approach developed by Phillips *et al.* (2006b) implemented in the software MaxEnt version 3.3.3 (Phillips *et al.* 2006a). This method aims to estimate across geographic space the probability distribution of a species' presence that, under constraints, is consistent with known presence data. Maximum entropy modeling has been shown to match or outperform other presence-only SDM methods (Elith *et al.* 2006; Phillips *et al.* 2006b) and is more robust to small sample sizes of presence data (Hernandez *et al.* 2006).

Presence data used for building models were based primarily on localities recorded in this study, for which geographical coordinates were obtained using a GPS (see Appendix 2.2). For *M. puta* and *Kerivoula* sp., these data were supplemented with some additional records provided by other Taiwanese bat researchers (Appendix 3.9). SDMs were developed from 20 potential predictors of species presence; these comprised altitude and 19 bioclimatic variables derived from high-resolution monthly temperature and rainfall data (Hijmans *et al.* 2005) that were downloaded as GIS layers from the WorldClim database (http://www.worldclim.org/). Key bioclimatic variables included annual means of precipitation and temperature as well as descriptors of seasonality and extreme conditions, which are together considered pertinent to species distributions, and consequently have been used in a growing number of studies (e.g. Beaumont *et al.* 2009; Holt *et al.* 2009; Waltari *et al.* 2007) (see Appendix 3.10).

For modeling current conditions, GIS layers of all these 20 environmental variables were downloaded at a spatial resolution of 2.5 arc-min (approx. 4.6 km) and were extracted with a mask ranging 119.9-122.1ºE and 21.8-25.5ºN in order to cover Taiwan. Variables that were strongly correlated with other variables (all Pearson correlation coefficient were ≥ 0.89) and thus contained redundant information were pruned (see Appendix 3.11) to give eight final variables: bio1 - mean annual temperature (ºC), bio2 - mean monthly temperature range (ºC), bio3 - isothermality, bio7 - annual temperature range (ºC), bio12 - annual precipitation (mm), bio15 - precipitation seasonality, bio18 - precipitation of warmest quarter (mm), and

bio19 - precipitation of coldest quarter (mm). For the MaxEnt analysis itself, I used a model convergence threshold of 10^{-5} and a maximum number of iterations of 500. A default value of regularization was used.

To evaluate model predictive performance, I applied a five-fold cross-validation procedure. Briefly, presence data were partitioned at random into five equal-sized 'folds', and replicate models were then built, in turn omitting a different fold. Folds served as 'testing data', and allowed model fit to be assessed based on the remaining 'training data'. For this I used the 'area under the curve' (AUC) of the receiver-operating characteristic (ROC) plot, which measures the predictability of a model in relation to a random prediction and has potential values ranging from 0.5 (predictability no better than random prediction) to a maximum depending on the actual prevalence of the target across the focal geographic space (a lower value for a wider distributed target, Phillips *et al.* 2006b). Finally, the present distribution of each focal species was predicted by averaging the results of models built under the five-fold cross-validation procedure.

To reconstruct the probable distribution of each focal species at the time of the LGM, I projected the (five) present-day models for each species onto LGM climatic conditions. LGM climatic layers were obtained from the WorldClim database at a spatial resolution of 2.5 arc-min., and were developed based on two general circulation model (GCM) simulations: the Community Climate System Model (CCSM, Collins *et al.* 2006) and the Model for Interdisciplinary Research on Climate (MIROC, K-1 model developer 2004) (see Waltari *et al.* 2007 for more details in LGM layer development). Since Taiwan was connected to mainland China during LGM due to drop in the sea level (Voris 2000), additional suitable habitats probably existed between these two regions and thus projecting models onto the LGM climate was carried out at a broader geographic scale than just Taiwan, over an area ranging from 111.3-126.3ºE and 19.2-28.4ºN.

Without knowledge of the probability of the presence of each focal species under a set of specific environmental conditions (set as the average feature condition in the MaxEnt program), SDMs of different species are not directly comparable (Elith *et al.* 2011; Phillips & Dudík 2008). This issue can also limit the use of SDM relative probabilities in IBD tests within a species because arbitrary noise will be introduced when scaling the resistance of each map unit with relative probabilities of presence of the target species at that map unit. A

conventional but simple approach to address both of these issues is to make binary predictions of presence and absence of the target at each map unit by applying a specific threshold or thresholds to relative probabilities of presence at the corresponding map unit. Two such thresholds were used in this study. First, I used the minimum training-presence threshold (MTPT), which corresponds to the lowest relative probability of presence of the target associated with its known presence records and is expected to be a stringent threshold if the target had not been recorded at the extremes of its suitable habitats. Second, I selected the 5% omission threshold (5OT), which omitted potential presence in areas within the lowest 5% of cumulative relative probability of presence.

3.3 RESULTS

Population genetics revealed by mitochondrial DNA

Genetic variability

For mtDNA analyses I generated alignments of Cyt-*b* (768 bp) for 97, 130, and 212 individuals of *M. gracilis*, *M. recondita* and *M. puta*, respectively, and alignments of COI (640 bp) for 88, 121, and 143 individuals of *M. gracilis*, *M. recondita* and *Kerivoula* sp., respectively.

For Cyt-*b,* a total of 52, 45, and 75 segregating sites were observed for the three respective alignments (Appendix 3.2), all of which were single nucleotide polymorphisms except for one tri-allelic site in *M. recondita*. Most substitutions (79-87% in each alignment) were synonymous, which together with no detected stop codons, suggested amplification of real mitochondrial Cyt-*b* genes rather than nuclear pseudogene copies. A McDonald-Kreitman test conducted to compare the distribution of synonymous versus non-synonymous substitutions within and between *M. gracilis* and *M. recondita* did not detect significant departure from neutrality (Fisher's exact test, P = 1.0). Species comparisons of genetic variability at Cyt-*b* within demes (here defined as sampling sites) revealed that of the three *Murina spp*., *M. puta* had the highest mean diversity and *M. recondita* had the lowest (Table 3.1; see Appendix 3.2 for raw data for individual demes). Despite this, all three species showed substantial variation among demes in all measures of genetic variability; for example, among the 12 values for each species-measure combination, standard deviations ranged from 22% to above 50% of the corresponding mean (Table 3.1). No clinal change in diversity with latitude was detected for any of the three species based on the four measures of diversity (Pearson correlation tests, $P > 0.05$).

For COI, a total of 31, 33, and 19 segregating sites were found (Appendix 3.2), with one tetra-allelic site seen in *M. gracilis*. As with Cyt-*b*, no stop codons were detected and substitutions were largely or exclusively (74-100%) synonymous, with no evidence of selection between *M. gracilis* and *M. recondita* based on a McDonald-Kreitman test (Fisher's exact test, P = 1.0). *Kerivoula* sp. had the highest mean level of demic genetic variability, with either *M. gracilis* or *M. recondita* second highest, depending on the exact measure

(Table 3.1). Demic levels of genetic diversity varied widely within each species but were most variable in the *Murina* species (standard deviations all >72% compared to 30-47% for *Kerivoula*; Table 3.1). For each species, no clinal change with latitude was detected for any of the four measures (Pearson correlation tests, $P > 0.05$).

Comparing the results from the two markers indicated that demes of both *M. gracilis* and *M. recondita* often showed higher diversity values at Cyt-*b* than at COI. For example, of seven demes of *M. gracilis* that contained polymorphism information, five almost always had higher diversity at Cyt-*b* than at COI gene (and one had the same haplotype diversity at both) (Appendix 3.2). Likewise, apart from two demes of *M. recondita* showing no polymorphism, seven out of 11 remaining demes showed higher diversity at Cyt-*b* than at COI. These differences were significant; from 19 demes of the two *Murina* species, a Wilcoxon signed-rank test suggested more segregating sites and higher nucleotide diversity at Cyt-*b* than at COI gene (test statistic $V = 120$, $P = 0.007$ and $V = 111$, $P = 0.028$, respectively), although there was no difference based on the other two measures ($P > 0.05$). These results highlight a need for caution when comparing genetic variability across demes based on these two mtDNA markers. This is especially true of *M. puta* and *Kerivoula* sp., for which different combinations of markers were used due to difficulties of amplifying both loci in both taxa.

Table 3.1 Levels of intra-demic genetic variability for four Taiwanese bat species based on the mitochondrial markers Cyt-b and COI. Values are given as means ± standard deviations over demes sampled with five or more individual bats for corresponding datasets. Nd, number of analyzed demes; S, number of segregating sites; θ_S, Watterson's (1975) theta; π, nucleotide diversity; Hd, haplotype diversity. Values of θ_S and π are shown as raw values multiplied by 100.

Dataset	Nd	S	θ_S	π	Hd
Murina gracilis × Cyt-*b*	8	6.75 ± 3.01	0.34 ± 0.17	0.29 ± 0.21	0.42 ± 0.32
M. recondita × Cyt-*b*	13	4.46 ± 3.36	0.22 ± 0.15	0.22 ± 0.15	0.37 ± 0.24
M. puta × Cyt-*b*	19	7.89 ± 4.59	0.39 ± 0.22	0.35 ± 0.18	0.77 ± 0.17
M. gracilis × COI	8	3.50 ± 4.78	0.20 ± 0.27	0.14 ± 0.16	0.27 ± 0.27
M. recondita × COI	11	2.55 ± 2.91	0.15 ± 0.16	0.14 ± 0.16	0.38 ± 0.28
Kerivoula sp. × COI	14	3.71 ± 1.73	0.22 ± 0.10	0.20 ± 0.09	0.51 ± 0.15

Genetic differentiation and isolation by distance among demes

Tests of global genetic differentiation among demes undertaken for *M. gracilis* (n = 8 demes) and *M. recondita* (n = 13) revealed that both species were characterised by similarly high levels of structure at Cyt-*b* (global Φ_{ST} = 0.87 and 0.81, respectively, and $P < 0.001$ in both cases). Values of global Φ_{ST} estimated from the concatenated sequences of both loci were almost identical for both *M. gracilis* (n = 8) and *M. recondita* (n = 11) (data not shown). Pairwise Φ_{ST} values among demes were mostly high and significant based on permutation tests with sequential Bonferroni corrections: mean Φ_{ST} values for *M. gracilis* and *M. recondita* were 0.79 and 0.70, respectively, based on 28 (all $P < 0.05$) and 78 pairwise comparisons (73% $P < 0.05$), respectively (see Appendix 3.3). Pairwise values from concatenated sequences were strongly correlated with results from Cyt-*b* for both taxa (correlation coefficient r = 0.96, $P < 0.001$; Mantel test).

Despite these clear results, close inspection of Φ_{ST} suggested two reasons why Cyt-*b* is probably unsuitable as a marker for quantifying genetic differentiation among demes in either putative sister species of *Murina.* First, highly divergent mitochondrial clades were recorded in both species (see *Phylogeography of mitochondrial haplotypes* below). Such accumulation of fixed nucleotide substitutions cannot be satisfactorily quantified by any measure of 'genetic differentiation' (an upper bound of 1 expresses complete differentiation) and so differences among demes containing different clades might be underestimated (Fig. 3.1). Second, the observation that most demes do not share haplotypes indicates that migration rates among demes are expected to be low relative to mutation rate(s) of the focal mitochondrial genes (see *Phylogeography of mitochondrial haplotypes*). Under such circumstances, F_{ST} and any related measures will underestimate true differentiation because of the time lag for migrations to spread new mutations (Balloux & Lugon-Moulin 2002; Whitlock 2011). For these reasons, I expected large biases using Φ_{ST} as a measure of either genetic difference or differentiation for *M. gracilis* and *M. recondita*, and thus did not perform isolation by distance tests for these. Patterns of pairwise Φ_{ST} values with Euclidean geographic distance are provided in Fig. 3.2 for illustrative purposes only.

M. puta and *Kerivoula* sp. also showed significant differentiation among demes (global Φ_{ST} = 0.28 and 0.35, respectively, $P < 0.001$ in both cases). The mean pairwise Φ_{ST} value among demes was 0.25 for both *M. puta* and *Kerivoula* sp., based on 171 (14%

significant following Bonferroni correction) and 91 comparisons (19% significant following Bonferroni correction), respectively (Appendix 3.3). In contrast to *M. gracilis* and *M. recondita*, pairwise Φ_{ST} in *M. puta* and *Kerivoula* sp. did not show 'saturation' toward high values (Fig. 3.1) and therefore tests for isolation by distance (IBD) were undertaken. In the case of *M. puta*, Mantel tests suggested significant IBD when modeled using Euclidean geographic distance (EGD) and also using the distance along the least-cost path (LCPD) that was based on distribution modeling in MaxEnt. However no significant trend was seen between genetic differentiation and the resistance distance (RD) (Table 3.2). All plots of IBD are shown in Fig. 3.3a. Inspection of R^2 values indicated that IBD models with RD provided poorer fit than those based on EGD and LCPD.

For *Kerivoula* sp. Mantel tests did not support IBD for all models except for one modeled by LCPD under the 5% omission threshold (5OT) test (Table 3.2; see also Fig. 3.3b for visualizations of IBD patterns). When five demes were removed to leave only the nine on the western side of CMR (including those from northeastern and western Taiwan), all models showed significant results (Table 3.2) and model fit was much improved with the exception of those based on RD.

Table 3.2 Model fit and significance of IBD models for *Murina puta* and *Kerivoula* sp. P values were obtained using Mantel tests and significant values are shown in bold. Spatial distances among demes comprised the Euclidean geographic distance (EGD), distance along the least-cost path (LCPD), and resistance distance based on circuit theory (RD). Two thresholds for binary predictions of connectivity between demes were used for estimating LCPD and RD: MTPT - minimum training-presence threshold; 5OT - 5% omission threshold (see Materials and Methods section *Species distribution modeling* for explanations of corrected spatial distances and thresholds used).

Taxon	Distance	Threshold	R^2 (%)	P
M. puta:	EGD	-	10.81	**0.002**
all 19 demes	LCPD	5OT	10.40	**0.003**
	LCPD	MTPT	10.43	**0.002**
	RD	5OT	1.67	0.208
	RD	MTPT	1.35	0.227
K. sp.:	EGD	-	2.42	0.127
all 14 demes	LCPD	5OT	7.69	**0.035**
	LCPD	MTPT	2.34	0.079
	RD	5OT	2.70	0.087
	RD	MTPT	2.57	0.091
K. sp.:	EGD	-	44.43	**0.003**
9 demes in NE and	LCPD	5OT	41.90	**0.002**
W Taiwan	LCPD	MTPT	37.90	**0.002**
(sites 2-37)	RD	5OT	29.02	**0.004**
	RD	MTPT	27.30	**0.010**

Figure 3.1 Scatter plots showing two measures of genetic distance for pairwise comparisons of demes for each of the four Taiwanese bat species: *Murina gracilis* (open triangle), *M. recondita* (filled circle), *M. puta* (open circle), and *Kerivoula* sp. (cross). Values of the two measures were estimated from Cyt-*b* for the three *Murina* species and from COI for *K.* sp. Values of pairwise sequence divergence were standardized for sequence length. Comparisons between many demes of *M. gracilis* and *M. recondita* showed accumulations of fixed nucleotide substitutions. Increased differentiation resulting from such accumulation cannot be adequately quantified by $\Phi_{ST,}$ which will 'saturate' for complete differentiation with a value of 1.

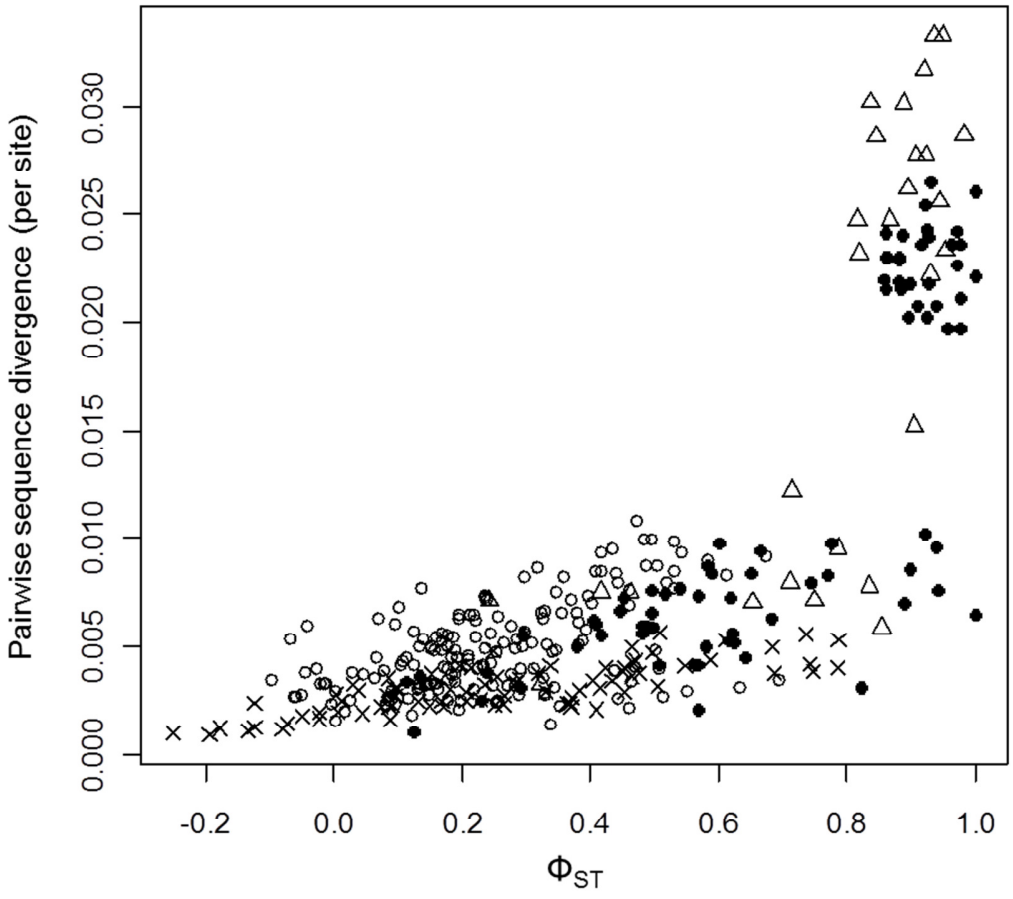


Figure 3.2 Scatter plots showing trends between Φ_{ST} and Euclidean geographic distance (EGD) among pairs of demes for (a) *Murina gracilis* and (b) *M. recondita*. Pairwise Φ_{ST} values were estimated from Cyt-*b*. In (b) symbols depict comparisons among sites from the eastern side of the Central Mountain Range (CMR) (i.e. eastern to southeastern Taiwan; open triangle), from the other side of the CMR (i.e. northeastern and western Taiwan; open circle), and also between sites from these opposite sides (filled circle).

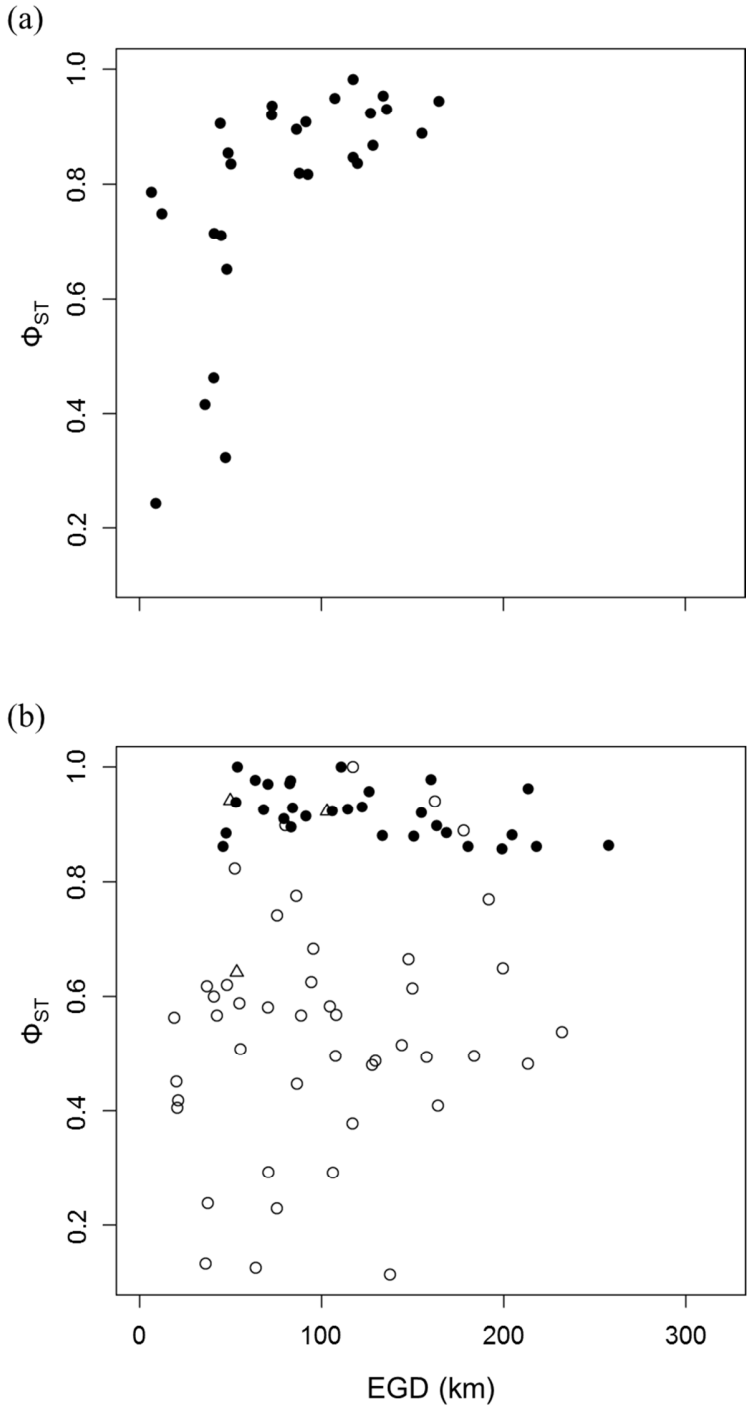
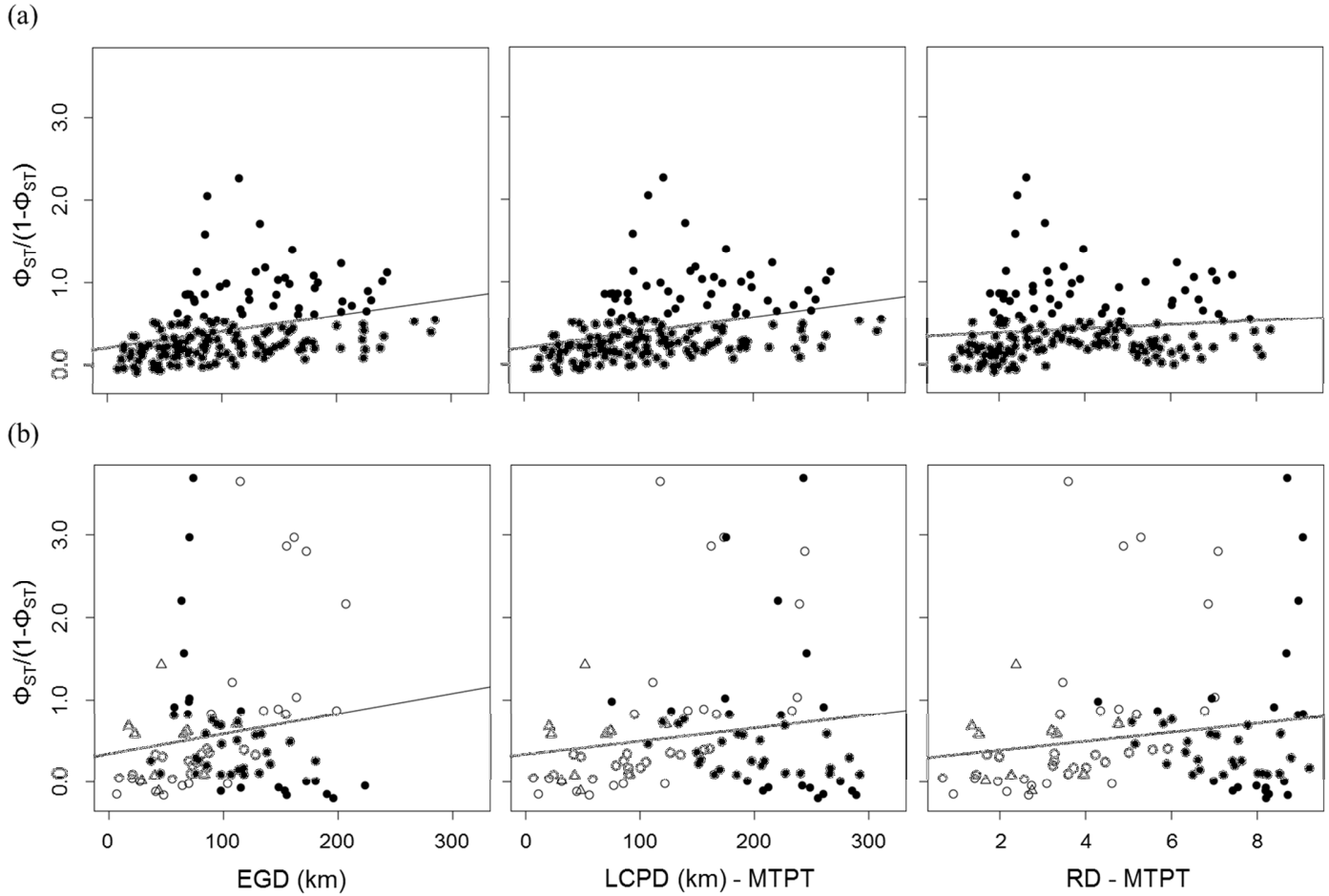


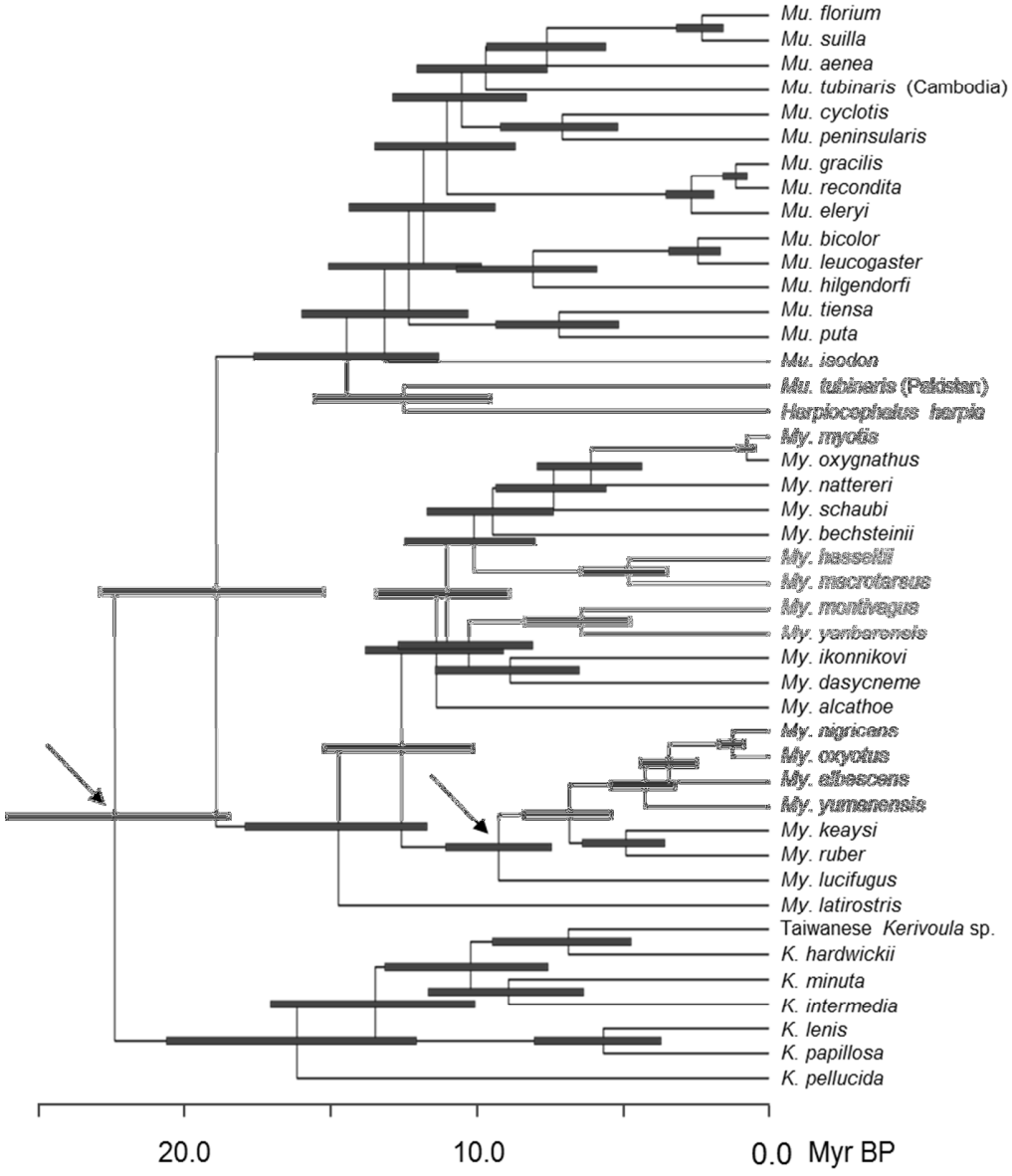
Figure 3.3 Examples of isolation by distance plots modeled using different spatial distances for (a) *Murina puta* and (b) *Kerivoula* sp. Pairwise Φ_{ST} values are estimated based on the Cyt-*b* gene for (a) and based on the COI gene for (b), and are transformed as $\Phi_{ST}/(1-\Phi_{ST})$ according to Rousset (1997). Data symbols in (b) depict comparisons from within the eastern side of the CMR (open triangle), the western side of the CMR (open circle), and from between these two opposite sides (filled circle). Regression lines estimated from the full sample are given under each model. See Table 3.2 for acronyms for various spatial distances and thresholds used.



Mitochondrial phylogeny of Kerivoulinae, Murininae, and Myotinae

A phylogenetic reconstruction revealed a sister relationship between *M. gracilis* and *M. recondita*, the common ancestor of which was observed to be sister to *M. eleryi* (both node posterior probabilities = 1) (see Fig. 3.4). The estimated divergence time between the two Taiwanese species was 1.1 Myr (95% CI: 0.7-1.6) and that between *M. eleryi* and the common ancestor of the two Taiwanese species was 2.6 Myr (95% CI: 1.9-3.5). This latter estimate was used for estimating clock rates of data partitions (CP_{12} and CP_3) of the concatenated sequence of Cyt-b and COI for both *M. gracilis* and *M. recondita* (see *Dynamic of effective population sizes* and *Divergence between major mitochondrial clades in M. gracilis and M. recondita* in Materials and Methods). *M. puta* formed a monophyletic group with *M. tiensa* (posterior probability = 1) and the divergence time between these two species was 7.1 Myr (95% CI: 5.1-9.3). The Taiwanese *Kerivoula* sp. formed a monophyletic grouping with *K. hardwickii* from Borneo (posterior probability = 1) and the divergence time between these two taxa was 6.8 Mya (95% CI: 4.5-9.3). Likewise, these time estimates were used to inform clock rates of data partitions of Cyt-*b* for *M. puta* and of COI for *Kerivoula* sp., respectively.

Figure 3.4 A dated Bayesian maximum clade credibility tree for 44 species of bats of the subfamilies Kerivoulinae, Murininae, and Myotinae (listed in Appendix 3.1) based on two mitochondrial genes, Cyt-*b* and ND1. Arrows point to calibration points while horizontal grey bars mark 95% highest posterior density limits of dates of corresponding splits.



Phylogeography of mitochondrial haplotypes

The concatenated alignment of Cyt-*b* and COI gave 25 and 38 unique haplotypes for *M. gracilis* and *M. recondita*, respectively, with no haplotype shared by these taxa. Maximum parsimony (MP) and Bayesian phylogenetic reconstructions both recovered a reciprocally monophyletic relationship between *M. gracilis* and *M. recondita* to the exclusion of the outgroup *M. eleryi*. Phylogenies also recovered the same major clades within each species (Fig. 3.5a-b).

M. gracilis showed three highly divergent and well-supported clades (Fig. 3.5a; see also Fig. 3.6a): a 'Northern clade' containing bats sampled from sites around Mt. Hehuan of the Central Mountain Range (CMR) as well as from further north (i.e. sites 4-25). The other two clades were from the south, the first ('Southern clade 1') containing bats from the southern part of the CMR as well as from the Alishan Range (sites 29-54), and the second containing bats from the Alishan Range only. The branching order among these three clades was weakly supported (Fig. 3.5a). The second taxon, *M. recondita*, contained two highly divergent and well-supported clades (Fig. 3.5b; see also Fig. 3.6b). An 'Eastern clade' contained bats from sites at the eastern and south-eastern parts of the CMR (sites 44-54) while the 'Western clade' was made up of samples from all other sites (sites 2-42).

Median-joining (MJ) networks, built separately for *M. gracilis* and *M. recondita* recovered the same clades, as did the tree-based methods (Fig. 3.6a-b). Close inspection revealed that haplotypes of both taxa were strongly geographically structured, as visualized in Fig. 3.6a-b. Specifically, only five out of 14 non-singleton haplotypes of *M. gracilis* were shared by two sites while each of the others was exclusive to a single site; weighted by the sample size, each of these non-singletons was shared by an average of 1.4 sampling sites. Likewise, only nine out of 20 non-singletons of *M. recondita* were shared by two or three sites while each of the other 11 was exclusive to a single site; each of these non-singletons were shared by a weighted average of 1.5 sampling sites. In both species, sampling sites that were geographically close to each other shared haplotypes.

For both taxa, geographical structure was less strong when only considering one of the two mitochondrial genes. Eight out of 17 non-singleton haplotypes of *M. gracilis* defined by the Cyt-*b* gene were shared by two or three sites while each of the others were exclusive

to a single site; giving to a weighted average of 1.7 sites per non-singleton. Six out of 12 non-singleton haplotypes of this species defined by the COI gene were shared by two or three sites while each of the others were exclusive to a single site, giving a weighted average of 1.7 sites per non-singleton. Likewise, 11 out of 17 non-singleton haplotypes of *M. recondita* defined by the Cyt-*b* gene were shared by 2-5 sites while each of the others were exclusive to a single site, giving a weighted average of 2.5 sites per non-singleton. Finally, ten out of 17 non-singleton haplotypes of *M. recondita* defined by the COI gene were shared by 2-4 sites while each of the others were exclusive to a single site, leading to a weighted average of 2.7 sites per non-singleton.

For *M. puta*, 61 unique Cyt-*b* haplotypes were observed including 26 non-singletons each shared by a weighted average of 10.2 sampling sites. Overall, 22 haplotypes were shared by more than one site. Tree- and network- based phylogenies of these 61 haplotypes did not recover any well-defined clades (Fig. 3.5c and 3.6c). In the MJ network, most unique haplotypes were linked to one that was much more common and widely distributed across Taiwan (sampled at 20 sites) (Fig. 3.6c), representing a star genealogy that is consistent with a history of population expansion (Slatkin & Hudson 1991). Similarly, several haplotypes (indicated by the ellipse in Fig. 3.6c) were connected to each other and to the main haplotype by several mutations, forming additional star-like genealogies in the network.

For *Kerivoula* sp., 15 unique COI haplotypes were recorded including nine non-singletons of which seven were shared by multiple sites. Each non-singleton was shared by a weighted average of 10.4 sites. The 15 haplotypes of Taiwanese *Kerivoula* sp. were shown to be closely related to those of *K.* cf. *hardwickii* from mainland China in tree- and network-based phylogenies (Fig. 3.5d and 3.6d) although these two taxa did not share haplotypes. The MJ network of Taiwanese *K.* sp., (Fig. 3.6d) indicated a star-like genealogy with an absence of deep structure. Three abundant haplotypes were recorded, all widely distributed across Taiwan (sampled at 8-16 sites) although each appearing to be dominant in a different geographic area of Taiwan (Fig. 3.6d).

Figure 3.5 Bayesian consensus trees reconstructed for (a) *Murina gracilis* and (b) *M. recondita* (Cyt-*b* + COI), (c) *M. puta* (Cyt-*b*), and (d) *K.* sp. (COI) with branches collapsed at posterior probabilities lower than 0.5. In figures (a) and (b), a smaller tree is also provided to show the relationship of the two species, rooted by a continental species, *M. eleryi* (shown in upper left corners). Node support values above branches are shown for highly divergent clades, based on MP bootstrap values (left-hand value) and Bayesian posterior probabilities (right-hand value). In (a) and (b), pie charts are used to present fractions of major clades sampled from different areas. Unlabeled tips in (c) and (d) represent haplotypes of *M. puta* and Taiwanese *K.* sp., respectively. In (d), symbols are given for haplotypes of *K.* cf. *hardwickii* from mainland China.

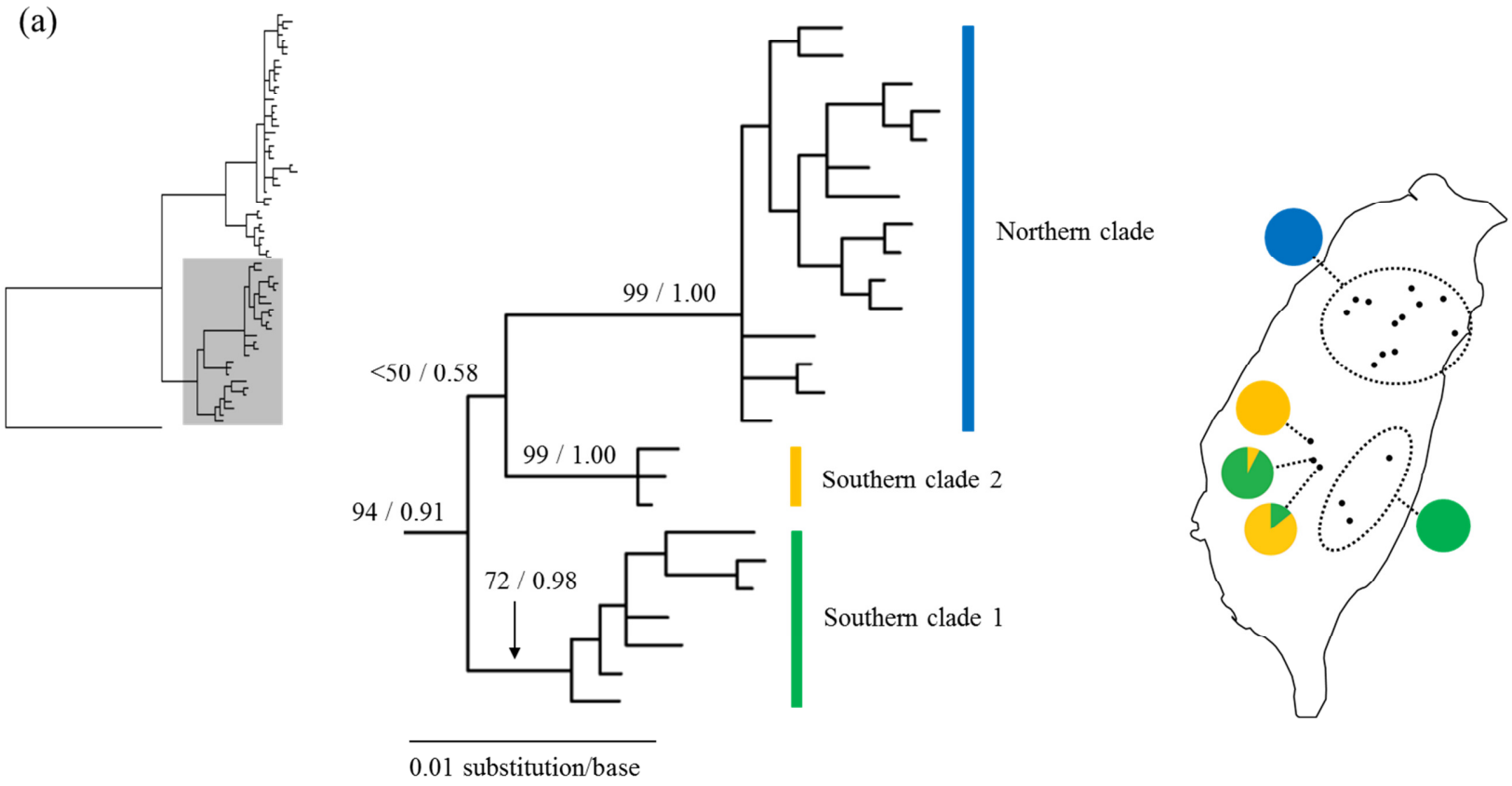


Figure 3.5 Continued.

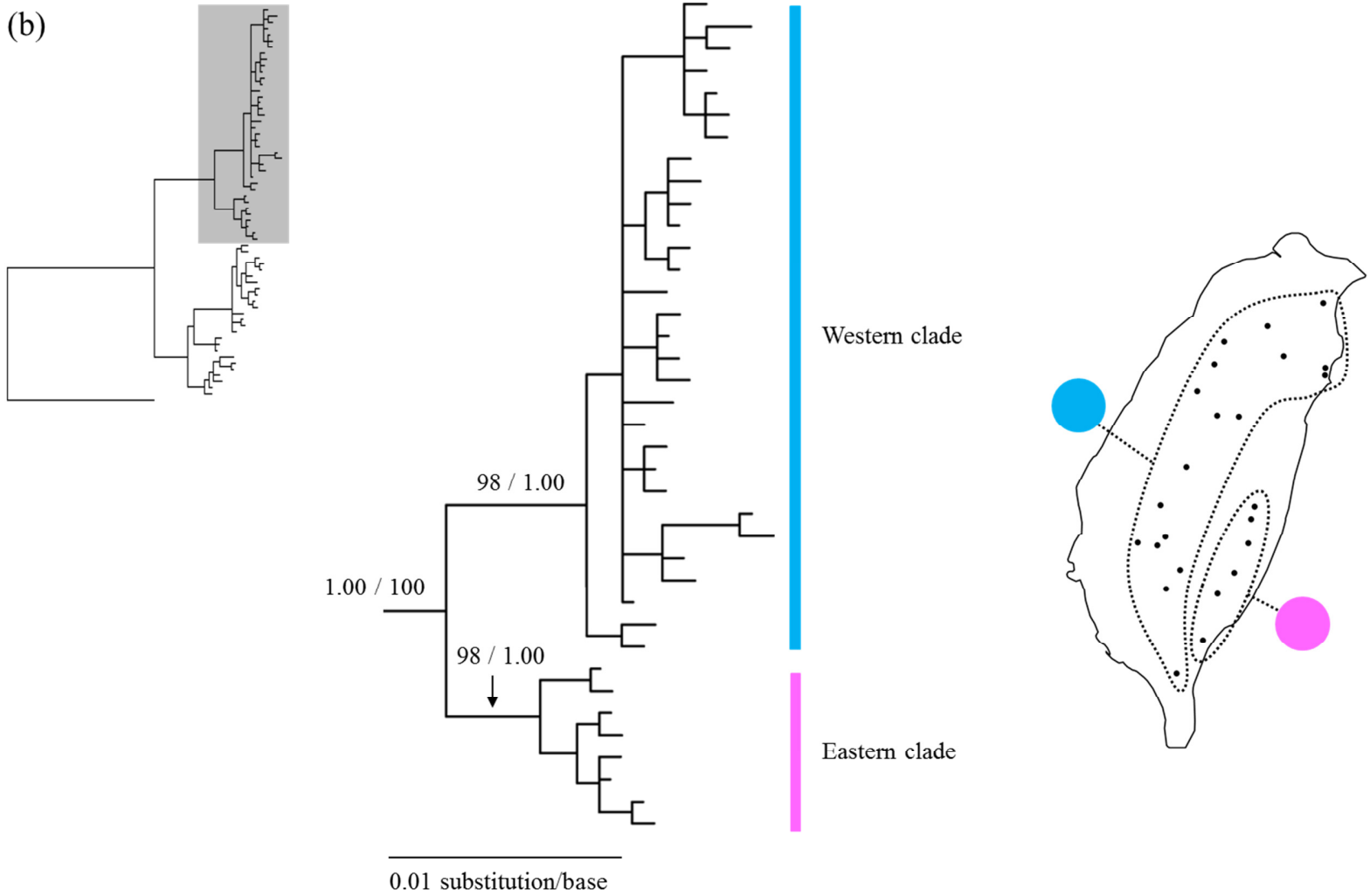


Figure 3.5 Continued.

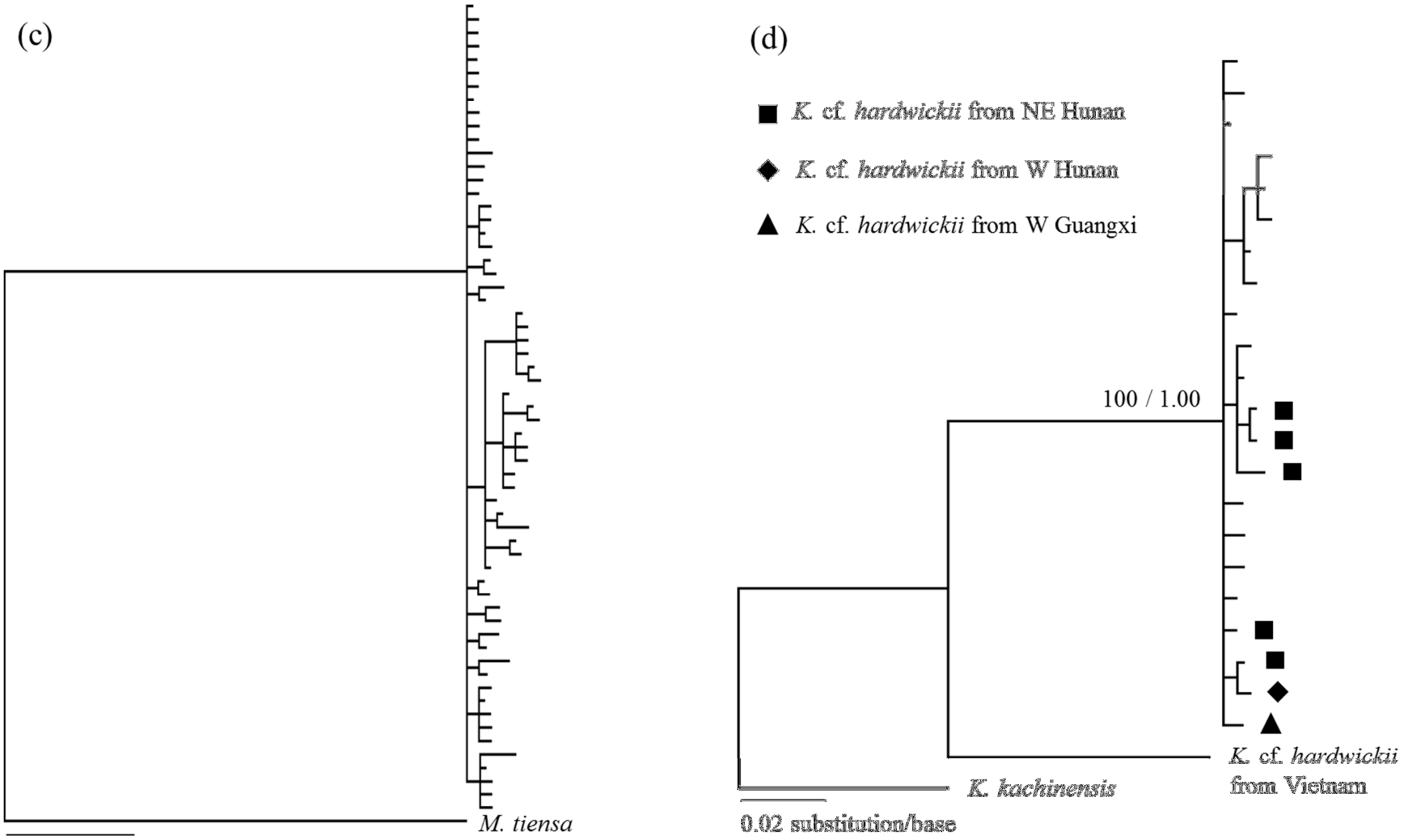
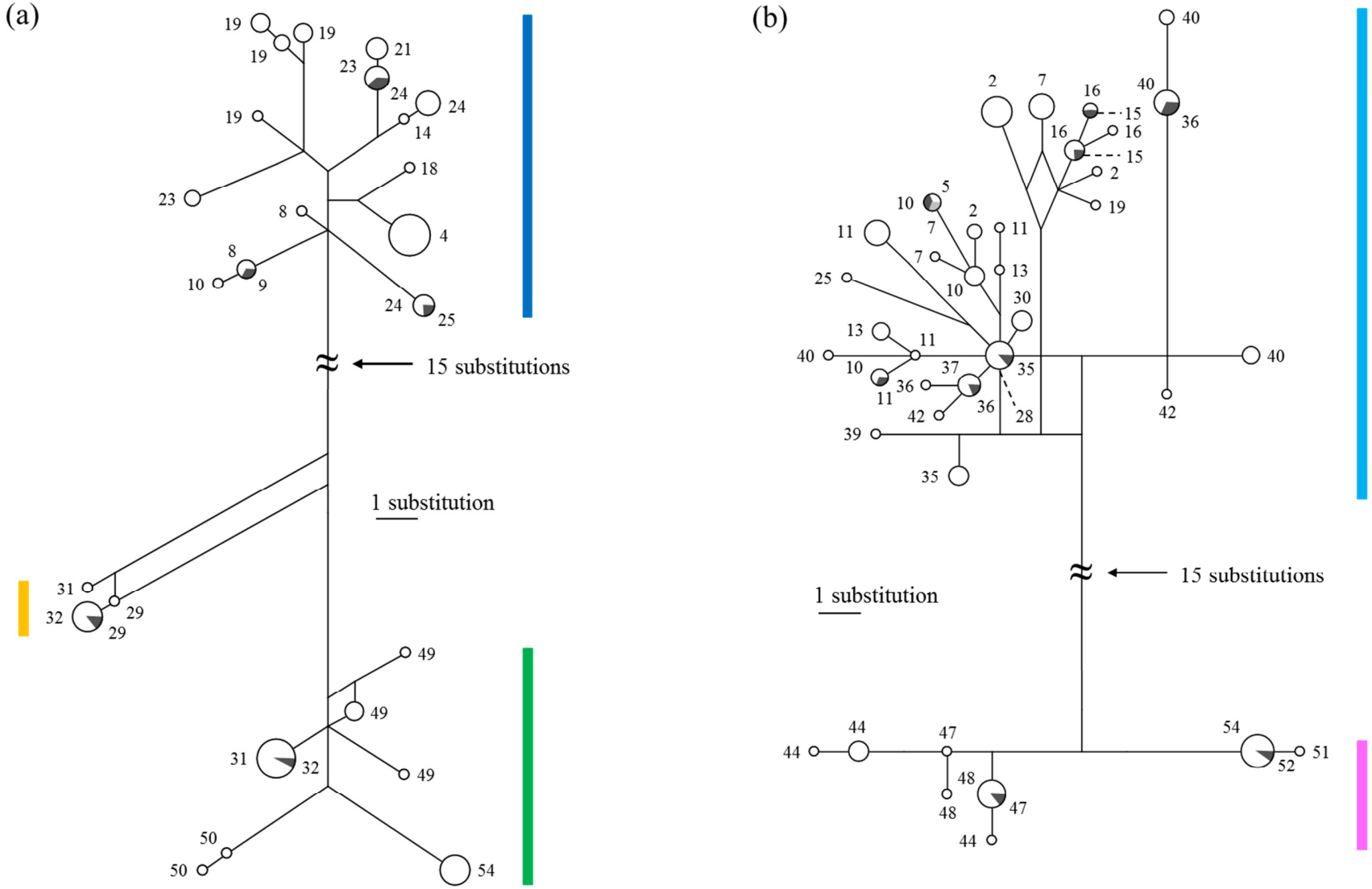
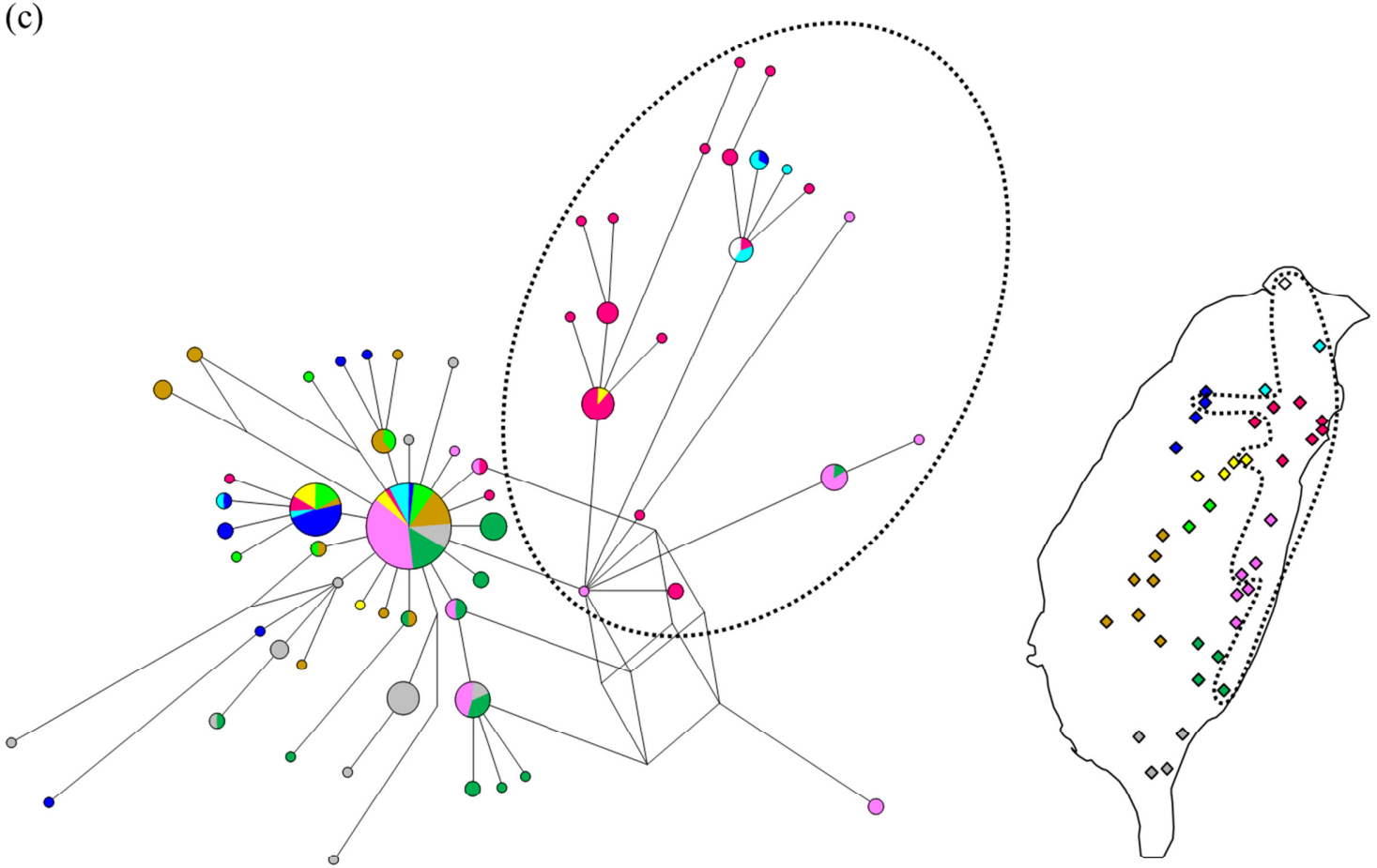
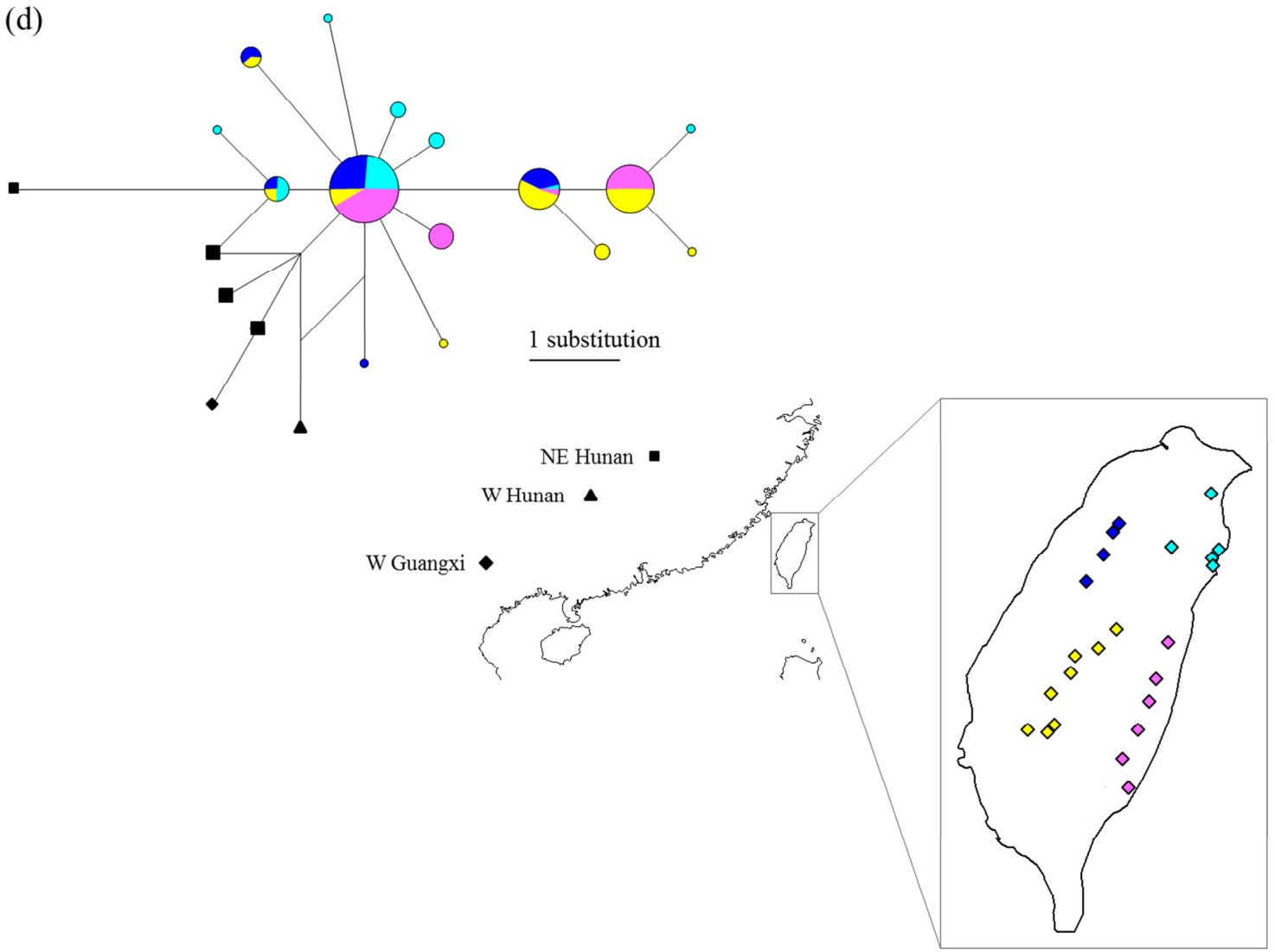


Figure 3.6 (following 3 pages) Median-joining networks reconstructed for (a) *Murina gracilis* (Cyt-*b* + COI), (b) *M. recondita* (Cyt-*b* + COI), (c) *M. puta* (Cyt-*b*), and (d) *Kerivoula* sp. (COI). In all networks, haplotypes of Taiwanese taxa are represented by circles with areas that are proportion to sample size (the smallest circle in each panel represents one sample), while in (d) *K. cf. hardwickii* from mainland China are marked with different symbols. Branches are shown with lengths proportion to the numbers of substitutions between haplotypes. In (a) and (b), circles are labeled with numbers to denote sampling sites of haplotypes. Pie-charts are divided into different shades of grey where a specific haplotype is shared by more than one site, and colour bars depict major clades recovered from the tree-based methods in Fig. 3.3. In (c) and (d), circles are coloured to indicate the geographical source of the haplotype.


(a)
19
19
19
19
21
23
24
24
14
19
18
23
8
4
8
10
9
24
25
15 substitutions
1 substitution
31
32
29
29
49
49
31
32
49
50
50
54
(b)
40
40
36
2
7
16
15
16
16
15
2
19
10
5
7
2
11
11
7
13
25
10
30
13
11
40
40
10
37
35
36
11
36
42
28
42
39
35
15 substitutions
1 substitution
44
44
47
54
52
51
48
48
47
44

(c)




(d)
1 substitution
NE Hunan
W Hunan
W Guangxi

Dynamics of effective population sizes

F_S statistics were significant for four out of six *M. gracilis* demes tested, and one out of nine *M. recondita* demes tested, whereas R_2 tests were only significant for two *M. gracilis* demes (Table 3.3). Many of the demes associated with significant F_S values were located in mountainous areas (the central CMR, the Alishan Range and the northern end of the Hsuehshan Range) and in all such cases, the observed test statistics were above the upper 95% credible limits of corresponding null distributions from simulations, indicating either a history of demographic contraction or a weak bottleneck.

F_S and R_2 neutrality tests applied to the full sample of *M. puta* yielded test statistics that were lower than the 95% credible limits of corresponding null distributions; thus rejecting a history of constant population size in favour of either a population expansion or a strong bottleneck (Table 3.3). For *Kerivoula* sp., neither of the two neutrality tests gave significant results when applied to either the full sample or to separate groups of demes from the western and eastern sides of the CMR (Table 3.3).

Two independent runs of the Extended Bayesian Skyline Plot (EBSP) gave consistent parameter estimates (assessed by visual inspection on the marginal distribution and its summarized statistics) for 14 of the 15 *M. gracilis* and *M. recondita* demes tested. In each case, the effective sampling size (ESS) for individual parameters from the combined two runs was >258. For one deme of *M. gracilis* (site 31) the two EBSP failed to converge due to the presence of a complex multimodal posterior distribution (Appendix 3.4), however, both EBSP runs converged for all other demes and thus in each case their results were combined. For all 14 demes, the estimated number of demographic changes had 95% credible sets that contained zero (Table 3.3) and, therefore, none showed strong support (i.e. with a posterior probability > 0.95) for a non-constant effective population (see Appendix 3.5a-b for the skyline plot for each deme of these two species). Median estimates of current effective population sizes of individual demes of *M. gracilis* varied from 36,500 (site 19) to 82,700 (site 49) and those of *M. recondita* from 17,500 (site 44) to 65,500 (site 40).

Both EBSP runs applied to the full sample of *M. puta* yielded very similar estimates for every parameter, although ESS values were low for tree likelihoods. After addition of a third run and combining across all three runs, the tree likelihood of the CP_{12} data partition gave an ESS value of 151 while all other parameters gave values of >200. The 95% credible

set for the number of demographic changes did not contain zero (Table 3.3), suggesting a non-constant history for this species. The median effective population size appears to have undergone a monotonic 12-fold increase from around 210,000 bats before 300,000 years ago to a current size of 2,860,000 individuals (Fig 3.7). The 95% credible interval for population size dynamics confirms that the onset of this demographic expansion predated 100,000 years ago.

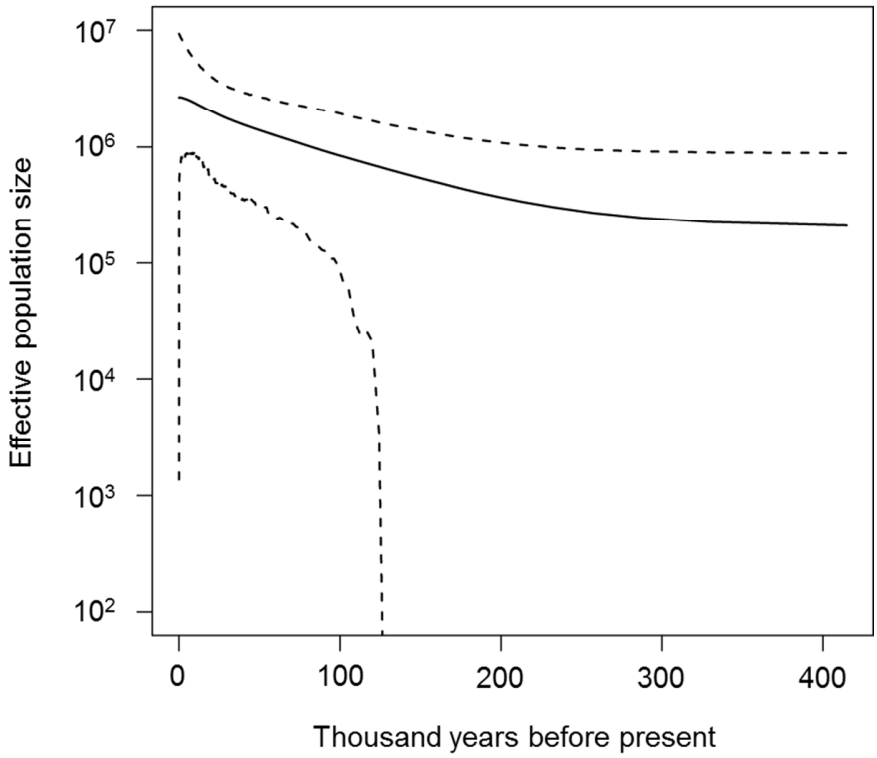
The same analyses applied to the full sample of *Kerivoula* sp., as well as subsamples from either side of the CMR, gave consistent parameter estimates across both EBSP runs. One demographic indicator parameter had a combined ESS of 160 in the full-sample dataset but otherwise parameters had combined ESS values of >200. For all three datasets, the 95% credible sets for numbers of demographic changes contained zero, and thus EBSP did not reject a history of a constant population in *Kerivoula* sp. (see Appendix 3.5c for the skyline plot for each dataset). The median current effective population sizes of the western CMR, eastern CMR, and the range-wide sample were 940,000, 100,000, and 980,000, respectively.

Table 3.3 Neutrality tests results for the four Taiwanese bat species, showing F_S statistics (Fu 1997), R_2 statistics (Ramos-Onsins & Rozas 2002) and Extended Bayesian Skyline Plot estimates of the number of demographic changes (EBSP-r). These analyses are applied to individual demes of *Murina gracilis* and *M. recondita* (marked as in Table 2.1), the full sample of *M. puta* and *Kerivoula* sp. and, in the latter cases, also samples from western (sites 2-37) and eastern (sites 44-56) sides of the CMR. Significant results are shown in bold and arrows indicate the inferred history of demographic expansion (or strong bottleneck; ↑) or contraction (or weak bottleneck; ↓). EBSP-r estimates medians with 95% highest posterior density limits given in parentheses.

Species	Deme/Group	F_S	R_2	EBSP-r
M. gracilis	19	1.02	0.22	0 (0, 2)
	23	**5.04** ↓	0.30	1 (0, 2)
	24	**4.57** ↓	0.23	1 (0, 3)
	31	**6.23** ↓	**0.27** ↓	?[a]
	32	**6.44** ↓	**0.35** ↓	1 (0, 3)
	49	1.22	0.25	1 (0, 3)
M. recondita	2	**4.31** ↓	0.15	1 (0, 3)
	7	3.76	0.17	1 (0, 2)
	10	2.17	0.20	0 (0, 2)
	11	1.78	0.17	0 (0, 3)
	16	-0.83	0.24	1 (0, 3)
	35	2.20	0.40	1 (0, 2)
	40	3.69	0.19	1 (0, 3)
	44	1.51	0.35	1 (0, 2)
	48	1.02	0.30	1 (0, 3)
M. puta	overall	**-57.90** ↑	**0.02** ↑	**1** (1, 3)
K. sp.	overall	-4.12	0.05	1 (0, 3)
	W CMR	0.13	0.05	1 (0, 3)
	E CMR	2.29	0.20	1 (0, 2)

[a] two EBSP runs did not converge.

Figure 3.7 The extended Bayesian skyline plot based on the full sample of *M. puta*. The plot shows the effective population size through time summarized by the median of the posterior probability distribution of the effective population size (solid line) and 95% highest posterior density limits (dashed lines). Note that the effective population size is presented as a $\log_{10}$ scale.



Divergence between major mitochondrial clades in M. gracilis and M. recondita

The main mitochondrial clades of *M. gracilis* and *M. recondita* obtained from traditional phylogenetic analyses (Fig. 3.5a-b and 3.6a-b) were also recovered using BEAST coalescent-based analyses, with strong support values (posterior probabilities > 0.99; trees not shown) regardless of the demographic model used. Bayesian methods also helped to resolve the relationship among the three clades of *M. gracilis* that was unclear from phylogenetic analyses; specifically, all models suggested that the two southern clades were sister groups to the exclusion of the Northern clade (posterior probabilities > 0.96). The median estimate of the TMRCA between the two southern clades was 300 Kyr ago (95% CI: 200-500 Kyr) while the TMRCA between the northern clade and the common ancestor of the two southern clades was 500 Kyr ago (95% CI: 300-800 Kyr) (Table 3.4). Likewise, BEAST analyses based on different models led to essentially the same estimates of TMRCA between the 'Eastern clade' and 'Western clade' of *M. recondita*, which was around 450 Kyr before present (95% CI: 300-800 Kyr) (Table 3.4).

Note that 'Southern clade 1' and 'Southern clade 2' of *M. gracilis* are not allopatrically distributed: samples from the southern CMR are represented by one clade while those from the Alishan Range are represented by both clades (Fig. 3.5a). This pattern could either arise from an ancient population divergence followed by post-split gene flow/secondary contact or by a more recent divergence with incomplete lineage sorting. To test these alternative scenarios, I applied Bayesian analyses based on 'isolation with migration' (IM) models to populations separated by the mountain ranges. The truncated uniform priors for parameters θ (= 4Nu; N is the effective population size and u is the mutation rate), m (the migration rate scaled by mutation rate), and t (a composite parameter of divergence time and mutation rate) were set as 100, 4, and 40, respectively, following some preliminary analyses. A total of six runs were implemented on the computer at Queen Mary University of London, and each took approximately five hours to finish.

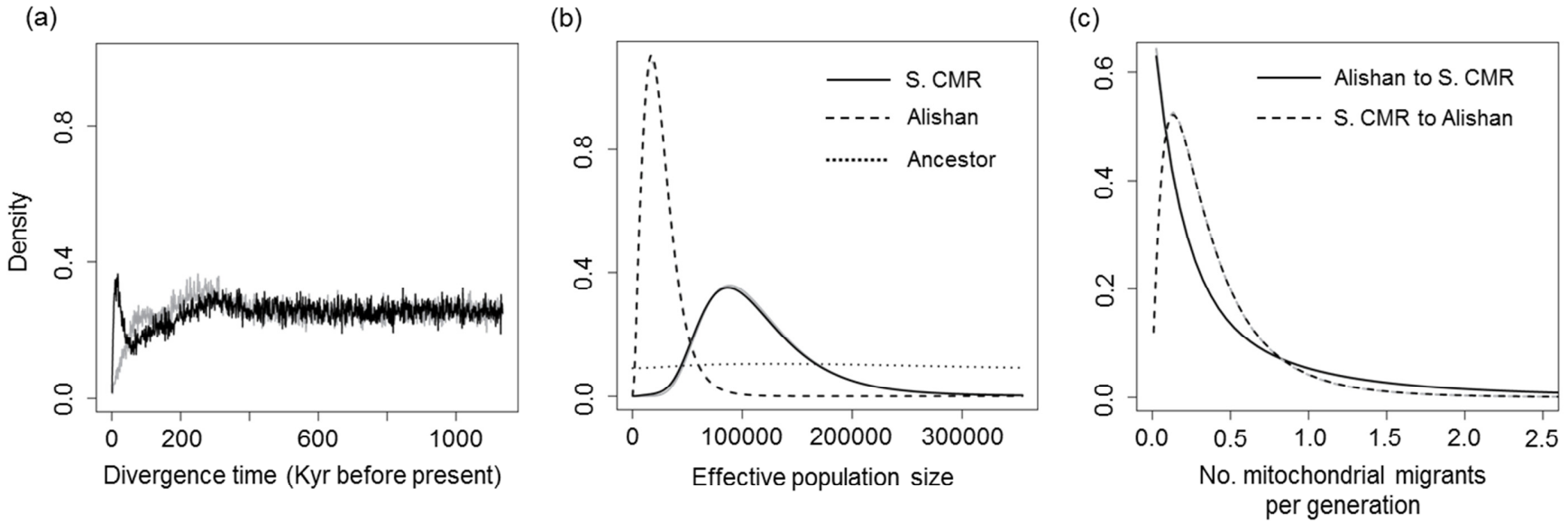
Convergence of chains under both IM models (i.e. with or without -j8 option) was confirmed for good mixing on the basis of stationary trend plots, high ESS values and consistent results across runs. The mode estimates of the divergence time between target populations were 17,611 and 250,533 years before present under the unconstrained and constrained (with -j8) models, respectively. However, marginal distributions of this

parameter were rather flat and/or with multiple peaks (Fig. 3.8a), suggesting little information in the data for this quantity. Chains under both models gave virtually the same estimates of effective population sizes for the two present populations: the mode estimates for the population in southern CMR was around 87,000-88,000 (approx. 95% CI: 33,000-220,000) while that for the population in the Alishan Range was around 18,000 (approx. 95% CI: 2,000-60,000) (Fig. 3.8b). The modal estimate of the effective population size of the ancestral population was around 130,000. Again, this quantity had a flat marginal distribution for these data, and thus the estimate should be interpreted with caution (Fig. 3.8b). Finally, chains under both models gave virtually the same results for post-split migrations between populations in different mountain ranges: the marginal distribution for such migrations from the Alishan Range to southern CMR had its mode at zero (95% CI for the per generation number of mitochondrial migrants: 0-2.42) while that for migrations in the reverse direction had a mode for the per generation number of mitochondrial migrants at 0.13 (95% CI: 0-1.08 or 1.09 under either model) (Fig. 3.8c). The unidirectional gene flow from southern CMR to the Alishan Range was also supported by likelihood ratio tests based on the migration rate parameter 'm' as well as based on the population migration rate ($P < 0.05$ for both cases).

Table 3.4 Times to most recent common ancestor (TMRCA) between major clades in *Murina gracilis* and *M. recondita* estimated using coalescent-based demographic models implemented in BEAST 1.6.1. Demographic models used included the constant-sized model (Model 1), Bayesian Skyline Plot (BSP) model describing population sizes in a piecewise-constant manner and allowing a maximum of five demographic changes (Model 2), and the BSP model describing population sizes in a piecewise-linear manner and allowing a maximum of five demographic changes (Model 3). The three clades of *M. gracilis* have their relationship well-resolved as ((S1,S2),N) in all these BEAST analyses. Acronyms of clades are N, S1, and S2 for the Northern clade, Southern clade1, and Southern clade 2 of *M. gracilis*, and E and W for the Eastern clade and Western clade of *M. recondita*, respectively (See Fig. 3.5a-b for geographic distributions of these clades).

		TMRCA		
Species	Clades pair	Model 1	Model 2	Model 3
M. gracilis	N×(S1+S2)	528	535	528
		312-800	311-810	305-794
	S1×S2	328	334	336
		181-517	186-527	187-525
M. recondita	E×W	452	434	440
		260-706	243-685	245-686

Figure 3.8 Demographic analyses based on isolation with migration models for divergence between populations of *M. gracilis* in the southern CMR and in the Alishan Range. Marginal distributions are shown for (a) divergence time, (b) effective population sizes of either population and of their common ancestor, and (c) numbers of mitochondrial migrants in either direction. Curves are shown in black for estimates under the unconstrained model and in grey for those under the model assuming equal size of the ancestral population to the larger present one (with -j8 option). A mutation rate as 0.0000352 per locus per calendar year as well as a generation time of two years have been used for conversions of raw parameters into these conventional demographic ones.



Population genetic structure revealed by microsatellites

Genetic variability

To further investigate population their population genetic structure, I screened each of the four focal species with between 9 and 17 autosomal microsatellites. These data show that demes of all four species had low F_{IS} values (Appendix 3.6). No deme of any species showed a significant departure from HWE following a sequential Bonferroni correction for multiple testing. Levels of genetic variability were high and consistent across demes within each species, as assessed by allelic richness, observed heterozygosity and expected heterozygosity (as reflected by small standard deviations in all measures for each species; Table 3.5; see also Appendix 3.6 for raw data for individual demes). In *M. puta*, levels of intra-demic allelic richness and expected heterozygosity showed negative correlations with latitude (Pearson's correlation coefficient r = -0.53 and -0.49, respectively, and $P < 0.05$ in both cases), but no other clinal change of intra-demic genetic variability with latitude was detected in the focal species (Pearson's correlation tests, $P > 0.1$ in all cases).

Table 3.5 Levels of intra-demic genetic variability for the four Taiwanese bat species assessed with multi-locus microsatellites. Values are given as means ± standard deviations based on demes consisting of five or more individual bats for each species. Nd, number of analyzed demes; Nl, number of loci; R_S, allelic richness; H_O, observed heterozygosity; H_E, expected heterozygosity. Raw values of R_S of each species (see Appendix 3.6) are standardized to present expected numbers of alleles in four individual bats.

Species	Nd	Nl	R_S	H_O	H_E
Murina gracilis	8	17	5.00 ± 0.22	0.84 ± 0.04	0.85 ± 0.02
M. recondita	14	17	4.46 ± 0.31	0.79 ± 0.04	0.79 ± 0.03
M. puta	18	9	5.49 ± 0.22	0.85 ± 0.03	0.86 ± 0.02
Kerivoula sp.	14	9	4.24 ± 0.14	0.74 ± 0.04	0.74 ± 0.03

Genetic differentiation and isolation by distance among demes

All four species showed significant genetic differentiation between demes (global F_{ST} = 0.05, 0.06, 0.005, and 0.03 for *M. gracilis*, *M. recondita*, *M. puta*, and *Kerivoula* sp. respectively; and P =0.021 for *M. puta* while < 0.001 for the others). The average values of pairwise F_{ST} for these species (see Appendix 3.7 for raw values between pairwise demes) were the same as their respective global F_{ST} values. Thus, *M. gracilis* and *M. recondita* showed, on average, higher levels of differentiation among demes than *K.* sp., and all three had much higher levels of inter-deme differentiation than did *M. puta*. A much larger proportion of pairwise F_{ST} values were significant following sequential Bonferroni correction in *M. gracilis* and *M. recondita* (69-89%) than in the other two species (0-5%) (Appendix 3.7).

Mantel tests suggested that IBD models based on either uncorrected Euclidean geographic distance (EGD) or corrected least-cost path (LCPD) and resistance distance (RD) were significant in all species except for *M. puta* (Table 3.6; see also Fig. 3.9 for visualizations of IBD patterns in the four species). Nonetheless, different distances performed better for different taxa. For *M. gracilis*, corrected distances did not fit the data (as assessed by R^2 values) better than EGD, whereas for *M. recondita*, the model using RD under the MTPT was the best fit (Table 3.6), and models with corrected spatial distances generally performed much better than ones with EGD, although the LCPD under the 5% omission threshold (5OT) had a fit similar to EGD. In this species, model improvement based on corrected spatial distances reflected the major effect of the Central Mountain Range (CMR) on genetic differentiation, as seen in Fig 3.8b where demes from opposite sides of the CMR were more differentiated from each other than those from the same side (data points representing comparisons across the CMR were above the regression line estimated for the full data under the EGD model). These results indicate that direct gene flow between demes on opposite sides of CMR is limited; indeed, when only the 11 demes of *M. recondita* from the western side of CMR were analyzed, all measures of spatial distance gave similar model fit. Finally, for *Kerivoula* sp., models based on corrected spatial distances showed substantially better fit than the EGD model (Table 3.6). Such improvements in model fit when using corrected spatial distances again suggested a major effect of CMR, which was also supported by higher pairwise F_{ST} across CMR (Fig. 3.9d). When only nine demes of

Kerivoula sp. from the western side of CMR were analyzed, models performed similarly with the exception of the RD models, which were still slightly better.

Finally, for each species I investigated the relationship between locus-wise global F_{ST} and intra-demic expected heterozygosity (H_E). In *M. puta*, global F_{ST} values were low with low variance (Fig. 3.10), while in *M. gracilis*, *M. recondita* and *Kerivoula* these values were higher and also showed greater variance among loci (28, 32 and 10 times larger than *M. puta*, respectively). In contrast, H_E values for both *M. puta* and *Kerivoula* sp. showed large variances among loci, while for *M. gracilis* and *M. recondita*, H_E values were consistently high (Fig. 3.10). Based on these above observations, I considered that the lower global F_{ST} values seen in *M. puta* and *Kerivoula* sp., compared to *M. gracilis* and *M. recondita*, were due to higher levels of inter-demic gene flow in the former two species rather than high mutation rates.

Table 3.6 Model fit (assessed using R^2 values) and P values (Mantel tests) for isolation by distance (IBD) in the four Taiwanese bat species as estimated using multi-locus microsatellites. Significant results are shown in bold. See Table 3.2 and the Materials and Methods section *Species distribution modeling* for acronyms and explanations of various spatial distances estimated under different SDM (species distribution modeling) thresholds.

Taxon	Distance	Threshold	R^2 (%)	P
Murina gracilis:	EGD	-	35.27	**0.002**
all 8 demes	LCPD	5OT	34.54	**0.002**
	LCPD	MTPT	33.26	**0.004**
	RD	5OT	28.78	**0.003**
	RD	MTPT	18.47	**0.010**
M. recondita:	EGD	-	6.88	**0.041**
all 14 demes	LCPD	5OT	8.53	**0.021**
	LCPD	MTPT	25.33	**<0.001**
	RD	5OT	22.20	**0.002**
	RD	MTPT	36.01	**<0.001**
M. recondita:	EGD	-	28.66	**<0.001**
11 demes in NE	LCPD	5OT	28.66	**0.002**
and W Taiwan	LCPD	MTPT	29.93	**0.001**
(sites 2-40)	RD	5OT	25.98	**0.003**
	RD	MTPT	27.19	**0.002**
M. puta:	EGD	-	2.69	0.922
all 18 demes	LCPD	5OT	2.70	0.925
	LCPD	MTPT	2.70	0.927
	RD	5OT	2.64	0.860
	RD	MTPT	2.29	0.828
Kerivoula sp.:	EGD	-	9.17	**0.009**
all 14 demes	LCPD	5OT	18.18	**0.001**
	LCPD	MTPT	19.66	**<0.001**
	RD	5OT	25.87	**<0.001**
	RD	MTPT	25.81	**<0.001**
K. sp.:	EGD	-	23.42	**0.017**
9 demes in NE	LCPD	5OT	23.25	**0.011**
and W Taiwan	LCPD	MTPT	27.97	**0.006**
(sites 2-37)	RD	5OT	30.96	**0.002**
	RD	MTPT	30.88	**<0.001**

Figure 3.9 Examples of isolation by distance modeled by different spatial distances for (a) *Murina gracilis*, (b) *M. recondita*, (c) *M. puta*, and (d) *Kerivoula* sp. For each species, pairwise F_{ST} values were estimated based on different sets of 9-17 microsatellites and are transformed as $F_{ST}/(1-F_{ST})$ according to Rousset (1997). Symbols used in (b) and (d) denote comparisons on the western side of the CMR (open circle), the eastern side of CMR (open triangle), and between these two opposite sides (filled circle). The regression line estimated based on the full sample under each model is given. See Table 3.2 for acronyms for various spatial distances and thresholds used.

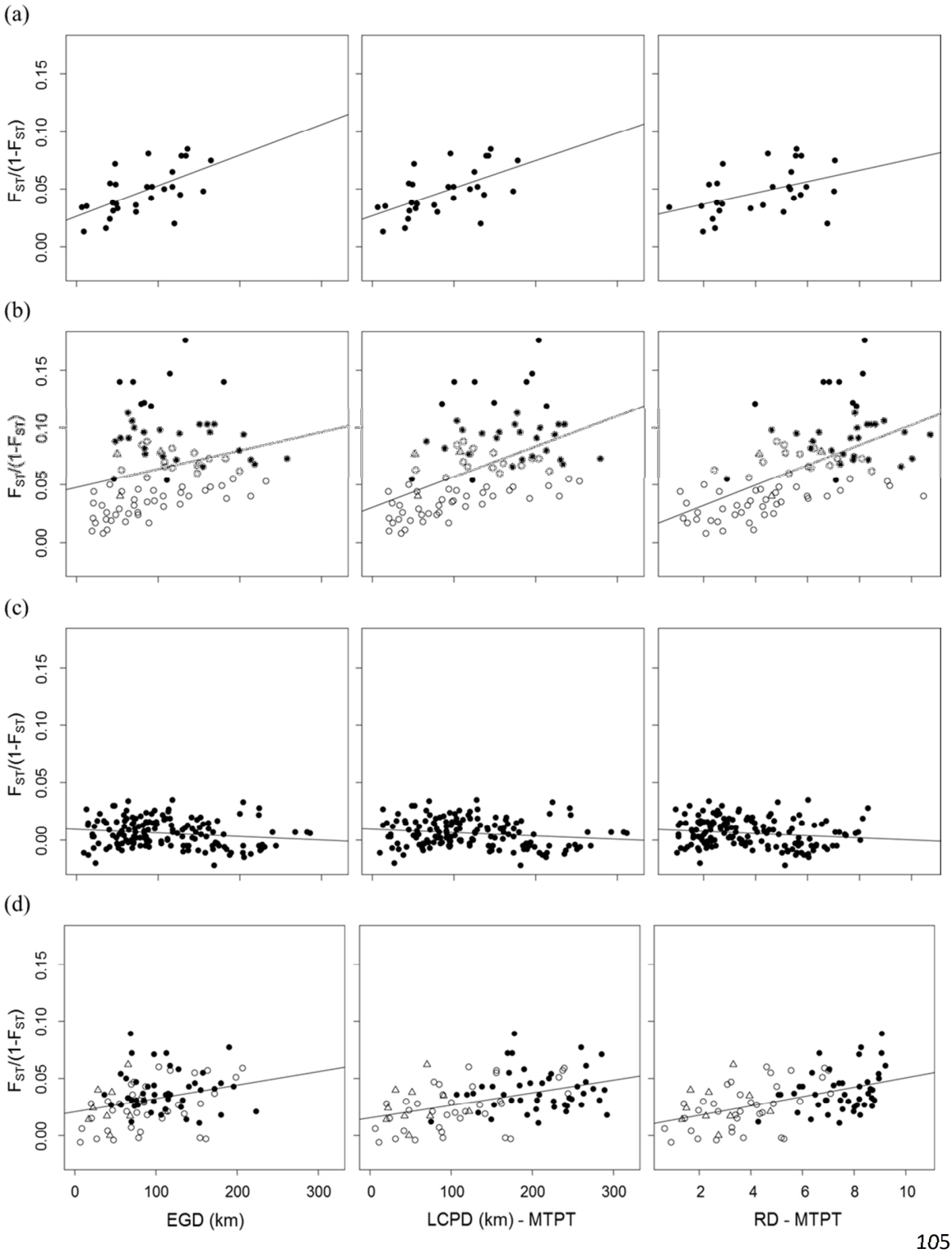
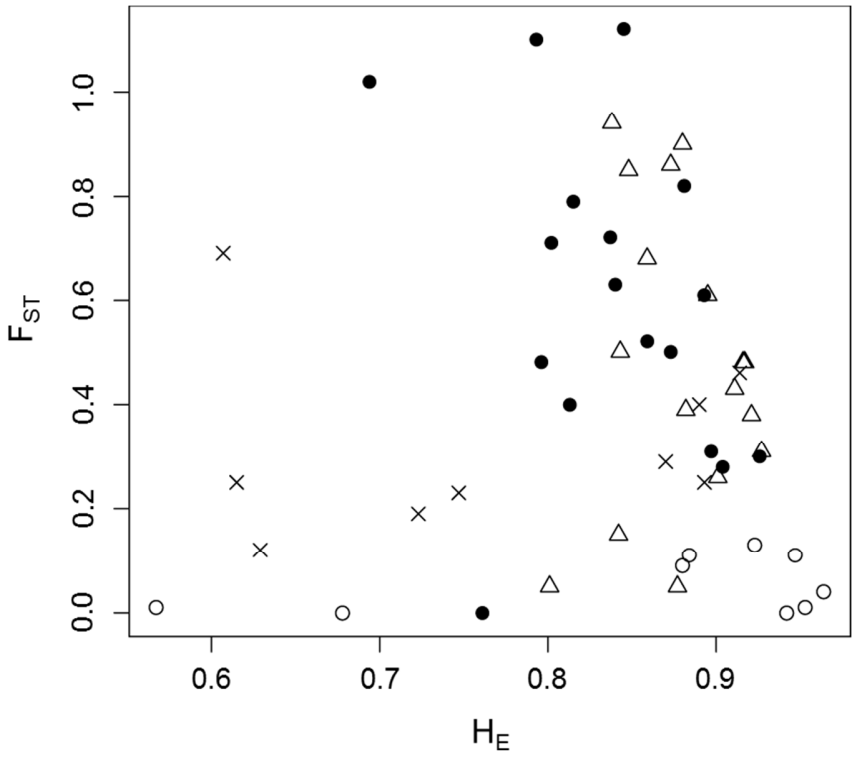


Figure 3.10 Scatter plots showing the relationship between locus-wise global F_{ST} and expected heterozygosity (H_E) for each of the four focal species. Symbols are as follows: open triangle = *M. gracilis*, filled circle = *M. recondita*, open circle = *M. puta* and cross = *Kerivoula*.



Genetic clustering

Bayesian clustering analysis was used to identify discontinuities in genetic variation within each of the four focal species. For all focal species except *M. puta*, the highest log-likelihood values under each specific value of K (Appendix 3.8) were seen when using the LOCPRIOR model (Hubisz *et al.* 2009), which specifies priors to give weight to clustering individuals from the same sites together. For *M. puta,* the LOCPRIOR model introduced larger variances among replicate runs than a model not incorporating sampling sites as priors while this was not the case for the other species (Appendix 3.8). For these reasons, I present results based on the LOCPRIOR model for *M. gracilis*, *M. recondita*, and *Kerivoula* sp. while I present those based on the standard model for *M. puta*.

For *M. gracilis* the likelihood of the data was maximized at K = 4 (Appendix 3.8a). Hierarchical clustering (Fig. 3.11a) suggested that at K = 2, two clear clusters (cyan and light yellow) corresponded to, respectively, all demes in the northern CMR plus Hseuhshan Range (sites 4-25), and three demes in the Alishan Range (sites 29-32). Demes in the southern CMR (sites 49-54) showed admixture of these two clusters. The former of these clusters was then seen to become divided at K = 3 (cyan and red) such that all bats from one deme in the extreme north (site 4) were assigned to the cyan cluster, whereas many bats from the other northern demes and from the southern CMR were assigned to the red cluster, and the others showed admixture of these two clusters. The addition of a fourth cluster (K = 4) led to the recovery of two distinct but similar CLUMPP configurations, each with similar likelihood and patterns of population cluster membership. The fourth cluster (light pink) contained bats from all northern demes except for site 4 and was thus distinct from the red cluster for bats from the southern CMR.

For *M. recondita*, the likelihood curve reached a plateau at K = 4 (Appendix 3.8b) although following Evanno et al. (2005) the most informative K was three, due to larger variances among replicated runs at K = 4. At K = 2, two clusters corresponded broadly to bats from the whole western Taiwan (red) and from all other sites (cyan) (Fig. 3.11b), with many bats sampled at the geographical boundaries of these two groupings also showing cluster admixture. At K = 3, the second of these clusters was subdivided into two (cyan and light yellow), the former of which contained bats from demes in the extreme north (sites 2-5) as well as in northeastern Taiwan (sites 15-19), while the latter contained bats from one deme

in the extreme south (site 42) and most of those east of the CMR (sites 44-51). Bats from the eastern central area (sites 52 and 54) showed admixture of these two clusters. At K = 4, the cluster of bats from western Taiwan (at K =2) subdivided into two (red and light pink), broadly representing bats from the northwest (light pink: sites 7-13 and 25) and southwest Taiwan (sites 28-40), respectively. At this K = 4 configuration, the four genetic clusters broadly corresponded to the four quarters of Taiwan, and demes at boundaries showed admixture between corresponding neighbours.

For *M. puta*, the likelihood of the data was maximized at K = 1, regardless or whether or not the LOCPRIOR model was used (Appendix 3.8c). Fig. 3.11c shows results estimated using the standard correlated allele frequency model for K = 2 and K = 3. All bats were evenly assigned into the specified numbers of clusters and this pattern continued when at higher K values (data not shown). Thus the range-wide sample of *M. puta* appears to be best represented by a single genetic cluster.

For *Kerivoula* sp., the likelihood of data was maximized when K = 3 (Appendix 3.8d). Results for K = 2 and K = 3 are shown in Fig. 3.11d. At K = 2, bats from the whole of western Taiwan (sites 6-11 and 26-37) were almost exclusively assigned to one cluster (red) while most of those from all other sites were assigned to the other cluster (cyan). At K = 3, the latter cluster was subdivided into two, containing bats from northern to northeastern Taiwan (sites 2 and 15-19, cyan) and from eastern to southeastern Taiwan (sites 44-56, light yellow). In this latter configuration, individuals from three demes in northwestern Taiwan (sites 6, 7, and 10) showed admixture between the cyan and red clusters. Likewise, bats from demes at the boundary of the northern and eastern groupings (site 15 and 56) showed admixture between the clusters for the corresponding groupings.

Figure 3.11 Genetic clusters inferred via Bayesian clustering analysis for (a) *Murina gracilis*, (b) *M. recondita*, (c) *M. puta*, and (d) *Kerivoula* sp. based on multi-locus microsatellites. Hubisz *et al.*'s (2009) LOCPRIOR model is used for (a), (b), and (d) but not for (c) (see the text for explanation). Plots show results under sequential numbers of clusters for investigations of hierarchical patterns of clustering in these species. Demes in (a), (b), and (d) are numbered as in Table 2.1 while those in (c) are listed in the order of their coded numbers in Table 2.1.

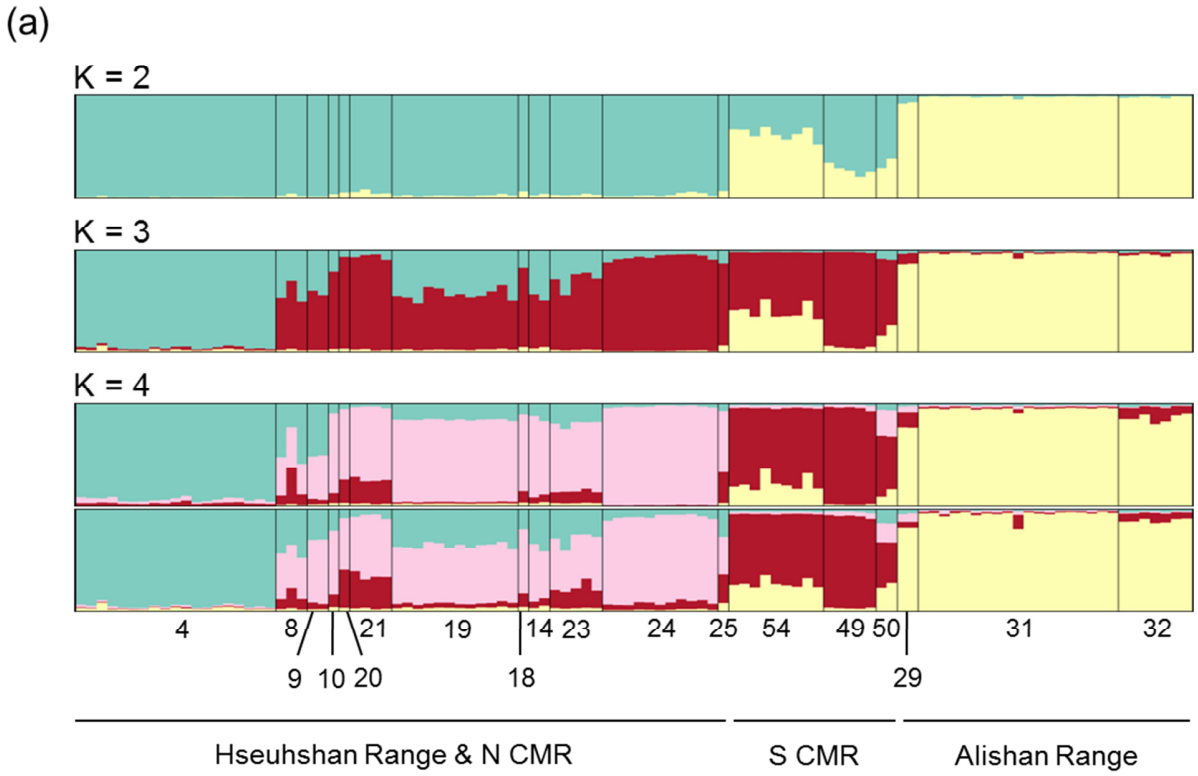


Figure 3.11 Continued.

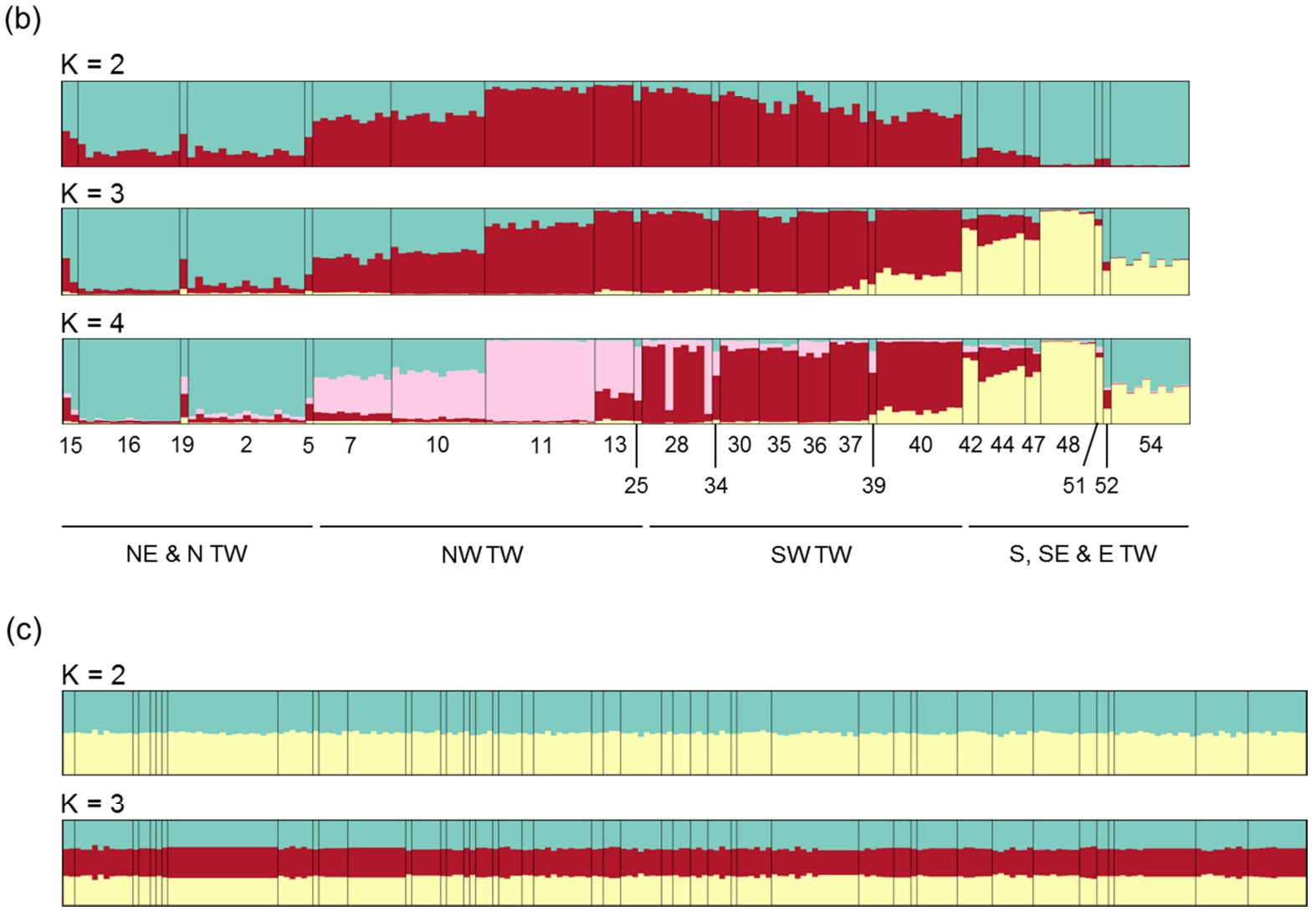
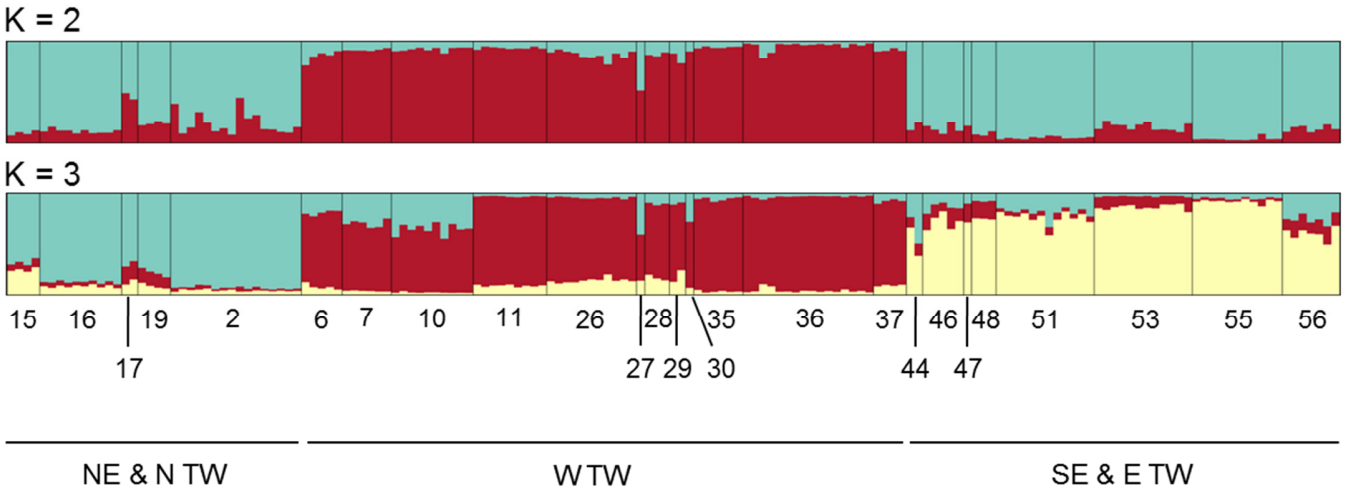


Figure 3.11 Continued.

(d)



Species distribution modeling (SDM)

Averaged over the five replicates of the cross-validation procedure, the AUC values for the Maxent models of each of the four focal species were all substantially higher than 0.5 (mean ± standard deviations: 0.888 ± 0.049 for *M. gracilis*, 0.803 ± 0.021 for *M. recondita*, 0.745 ± 0.022 for *M. puta*, and 0.807 ± 0.016 for *Kerivoula* sp.), indicating they had good predictive power.

M. gracilis

Under the most stringent threshold (minimum training-presence threshold, MTPT), the Maxent model for the present-day distribution predicted suitable habitats in mountainous areas with elevations between 860 and 3150 m a.s.l. (Table 3.7; Fig. 3.12a), although a few areas >2700 m a.s.l. in the Hseuhshan Range and in the northern half of CMR were predicted as unsuitable. Using the less stringent threshold (5% omission threshold, 5OT), suitable habitat was around 20% larger in area, covering a slightly wider altitudinal range (Table 3.7). The addition of some high mountains as suitable habitats under 5OT resulted in a more contiguous block of suitable habitat than that predicted under MTPT (Fig. 3.12a).

Maxent models of suitable habitat for *M. gracilis* during the Last Glacial Maximum (LGM) based on two climatic models - the Community Climate System Model (CCSM) and the Model for Interdisciplinary Research on Climate (MIROC) - gave somewhat different results (Fig. 3.12b-c). The CCSM-based model under both MTPT and 5OT thresholds suggested a wide and continuous past distribution (i.e. defined by suitable habitat) from mainland China to Taiwan that encompassed the land bridge (see Fig. 1.2 for map showing shallow sea areas). In Taiwan, unsuitable areas were only suggested for small parts of the extreme north and south under the MTPT model, and only the extreme south under the 5OT model (Fig. 3.12b). Based on these predictions, it is likely that *M. gracilis* could range down to the current sea level during the LGM (Table 3.7; elevations of habitats during LGM are defined with reference to current sea level) with no major barriers to gene flow, although the highest mountains were probably still inhospitable. Given that *M. gracilis* probably had a much larger historical distribution it seems plausible it will have experienced an extreme range contraction towards the present time. Comparing the areas of suitable habitat within the

area limited by the dashed line shown in Fig 3.10a and b (representing the potential range of *M.* gracilis at present and during the LGM) gives a possible range contraction of 4-4.5fold since the LGM under the two thresholds (Table 3.7), although this probably underestimates the scale of the range contraction since it does not include all of the area *M.* gracilis could have occupied in continental Asia had it been present on both sides of the land bridge.

The MIROC-based MaxEnt model predicted separate blocks of suitable habitat for *M. gracilis* in Taiwan and mainland China, which might have been joined by a narrow corridor under a 5OT model (Fig. 3.12c). Within Taiwan, suitable habitats expanded toward both higher and lower altitudinal areas under MTPT while they expanded toward lower altitudinal areas under 5OT when compared to those in the present model under corresponding thresholds (see Table 3.7 for means and ranges of altitudinal covers of suitable habitats under the two thresholds). Assuming that the two blocks of suitable habitats were not connected in the LGM I estimated that *M. gracilis* has undergone a range contraction of the order of 1.8-2.2 from the LGM to the present (i.e. when comparing suitable habitat at present and during LGM as delimited by dashed lines in Fig. 3.12a and c) under either threshold.

M. recondita

MTPT-based species distribution modelling showed that *M. recondita* is likely to occur in upland regions of between 200 and 2300 m a.s.l. (Table 3.7), areas of which form a ring-shaped block around higher inhospitable parts of the the Hseuhshan Range and the CMR of central Taiwan. Unsuitable habitat in the CMR forms an elongated barrier of around 180km long (Fig. 3.12d). Under a 5OT threshold, the predicted area of suitable habitat was found to be around 20% larger, spanning the highland and lowland extremes of the suitable block predicted under the previous threshold. The additional suitable habitats predicted in high altitude areas might bridge the CMR barrier predicted under the more stringent threshold.

During the LGM, Maxent models based on CCSM and MIROC produced very different predictions about the likely distributions of *M. recondita* (Fig. 3.12e-f). The CCSM-based model predicted that a large landmass including Taiwan and surrounding areas was unsuitable under both MTPT and 5CT thresholds (Fig. 3.12e), implying that this species was

absent from the entire region during the LGM. In contrast, the MIROC-based model predicted large areas in and around Taiwan to offer suitable habitat for *M. recondita* under both MTPT and 5OT criteria. Suitable habitat predicted by the latter model formed two separate blocks in the LGM landmass: one in Taiwan and the other one in the land bridge west of Taiwan (Fig. 3.12f). Within Taiwan, suitable habitats were shifted towards lower altitudes such that the central upland area of inhospitable habitat was larger than at present, probably forming a greater barrier to dispersal (Table 3.7). If I assume that *M. recondita* was never distributed in the disjunct block of suitable habitats west of Taiwan, then the present range of this species is 1.4 to 1.6 times smaller than that during LGM under the two thresholds (estimated as the ratio of areas of suitable habitat within the dashed lines in Fig. 3.12d and f).

M. puta

The present-day Maxent model for *M. puta* gave virtually the same predictions under MTPT and 5OT (showing only a 5% increase in the area predicted as suitable under the latter threshold). In both, suitable habitat forms an approximately contiguous block covering the whole range of Taiwan except for the western lowlands (Fig. 3.12g). Most of the suitable habitat with elevations < 300 m a.s.l. was found to be in northeastern and eastern Taiwan. In high altitudinal areas, patches of unsuitable habitats were predicted, but formed small isolated pockets rather than connecting to form an elongated barrier as observed for *M. recondita*. Apart from the main block of suitable habitat described above, a small patch of isolated habitat was predicted to occur in mountains in the extreme north of Taiwan (Yangmingshan).

As for *M. recondita*, the CCSM- and MIROC-based Maxent models predicted very different former ranges for *M. puta* during the LGM (Fig. 3.12h-i). In the former, the whole of Taiwan and the wider region was predicted as unsuitable under both MTPT and 5OT thresholds (Fig. 3.12h). In contrast, large parts of Taiwan and the former land bridges were predicted as suitable in the MIROC-based model under both thresholds (predictions under the two thresholds were essentially the same; Fig. 3.12i). One such suitable block was located in Taiwan while the other two were located in land bridges western and northeastern to Taiwan, respectively. The suitable block in the western land bridge was disjunct from the Taiwanese

one, whereas a former habitat corridor connected the habitat block in northeastern land bridge to that in Taiwan. Within Taiwan, suitable habitats occurred in lower areas than at present (see Table 3.7 for altitudinal changes of suitable habitats under the two thresholds) and were thus likely separated by a long stretch of higher peaks of the CMR. Within the space delimited by the dashed lines shown in Fig. 3.12g and i, the extent of suitable habitats for *M. puta* at present was about three quarters that during the LGM under both thresholds (Table 3.7). However, such differences will be underestimated if, as seems likely from these predictions, *M. puta* underwent a range expansion to exploit the inferred suitable habitat in the land bridge northeastern to Taiwan during the LGM.

Kerivoula. sp.

Maxent models for current distributions of the endemic *Kerivoula* sp. gave similar predictions under the two thresholds (the area of suitable habitat was 10% larger under 5OT). In both models, this species is predicted to favour lower mountainous areas with elevations of 2200 to 2300 m a.s.l. (Table 3.7; Fig. 3.12j). Consequently, large areas in western Taiwan with elevations < 200 m a.s.l. were predicted to represent unsuitable habitat, although some areas of the same elevation in the eastern part of the island were still predicted as suitable. With the exception of a very small area in the extreme north of the island, areas in the south and north were otherwise considered to be unsuitable. High mountains in the Hseuhshan Range and in CMR appear to form a long and substantial (around 230 km) barrier to this species.

Once again, CCSM- and MIROC-based Maxent models yielded different predictions about the extent of suitable habitat during the LGM. The CCSM-based model indicated a tiny area of suitable habitat persisted in the southernmost part of Taiwan and in the offshore island of Lanyu to the southeast, under both thresholds (Fig. 3.12k). The former patch was < 300 m a.s.l. and represented an area that was 40 to 50 times smaller than its present reconstructed range under the corresponding threshold (Table 3.7). The MIROC-based model predicted three disjunct blocks of suitable habitats in the region during the LGM, one in Taiwan and the other two on the ancient landmass west and northeast (Fig. 3.12l). Within Taiwan, much larger areas of the higher mountains were predicted as unsuitable compared to the present-day model (see Table 3.7 for details of the reduction in elevation of suitable

habitats in the MIROC-based model in relation to those in the present-day model under corresponding thresholds). The central barrier was thus inferred to be even more substantial during this period. Assuming that *Kerivoula* sp. occurs in all habitats predicted to be suitable at the present, and during LGM in all suitable habitats within the space delimited by the dashed lines in Fig. 3.12j and l, then this species will have experienced a range expansion of around 1.2-fold since LGM under the two thresholds (Table 3.7).

Table 3.7 Quantitative comparisons of suitable habitats for the four Taiwanese bat species predicted in Maxent models. These include the mean and range of elevation of suitable habitats (ALT-CUR, ALT-CCSM, and ALT-MIROC for estimates in the present model, LGM model based on CCSM, and LGM model based on MIROC, respectively), and the ratio of the extents of present-day/LGM of suitable habitats (AREA-RATIO-CCSM and AREA-RATIO-MIROC for ratios based on respective climatic models). The 5OT (5% omission threshold) and MTPT (minimum training-presence threshold) are thresholds for binary predictions of suitable versus unsuitable habitats (see Materials and Methods section *Species distribution modeling* for detailed explanations).

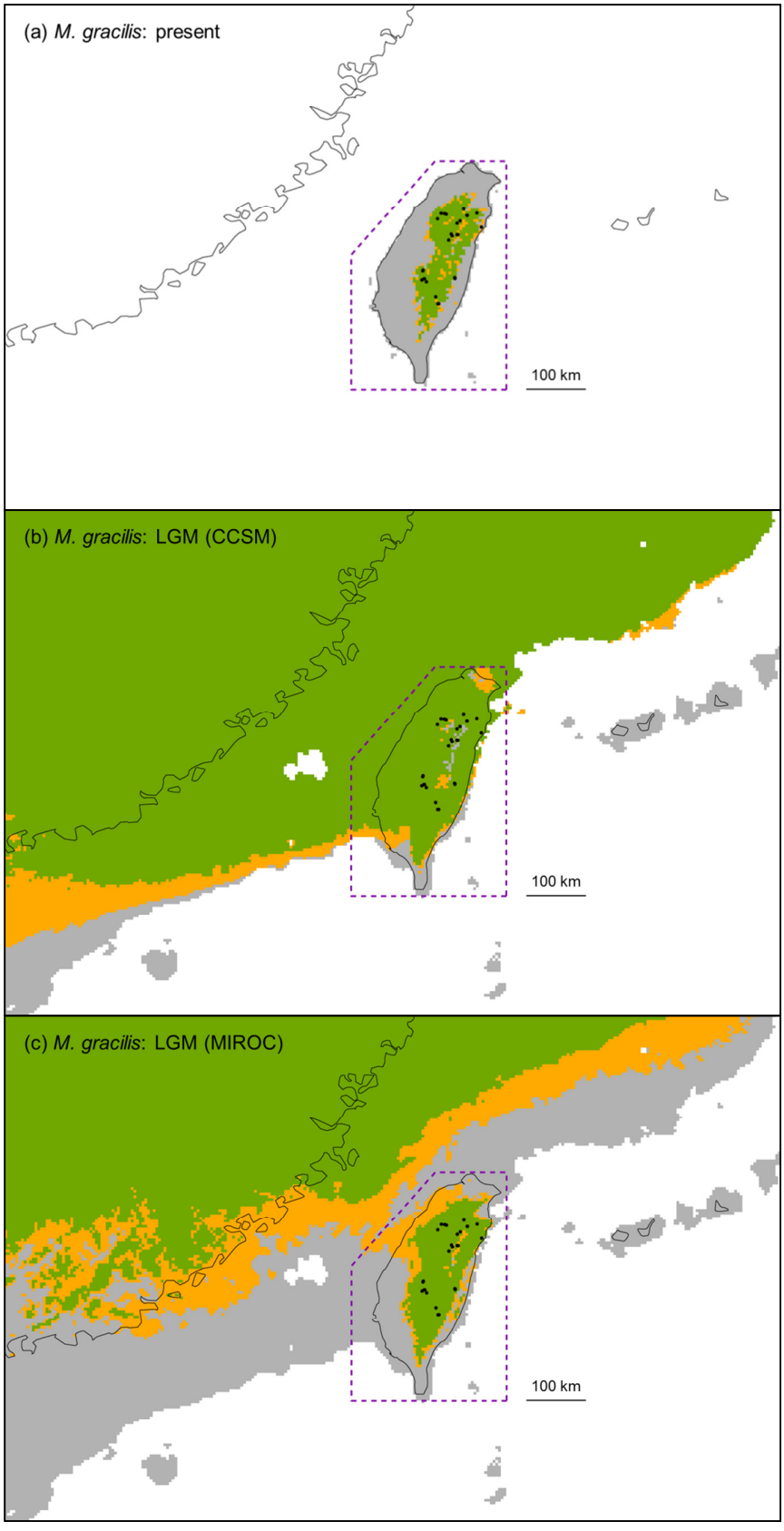
	ALT-CUR		ALT-CCSM[a]		ALT-MIROC[a]		AREA-RATIO-CCSM[b]		AREA-RATIO-MIROC[b]	
Species	5OT	MTPT	5OT	MTPT	5OT	MTPT	5OT	MTPT	5OT	MTPT
Murina gracilis	1743.7	1805.3	742.3	758.8	1083.5	1374.8	0.25	0.22	0.45	0.57
	651-3228	866-3149	-8-3149	-8-3126	-8-3228	172-3228				
M. recondita[c]	1136.8	1095.9	-	-	580.8	539.7	-	-	0.63	0.72
	114-2649	230-2270			-7-2259	-7-1807				
M. puta[c]	1139.7	1147.9	-	-	796.5	819.4	-	-	0.74	0.76
	4-2927	6-2767			-7-2384	1-2329				
Kerivoula sp.	838.3	770.4	128.9	109.2	535.0	526.9	42.00	50.94	1.16	1.23
	2-2321	2-2198	32-289	32-206	2-1733	2-1463				

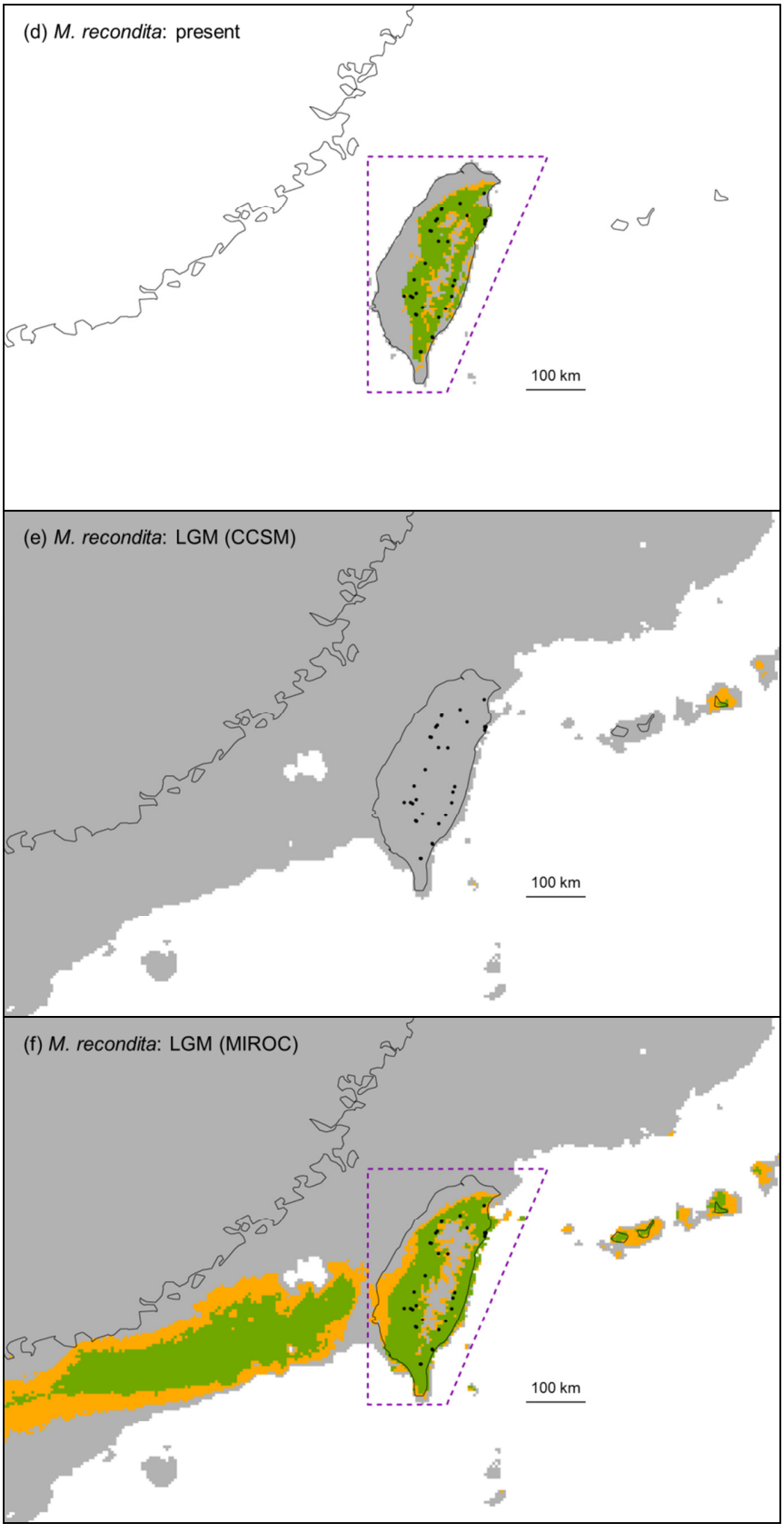
[a] Elevations of suitable habitats during LGM are defined with reference to the current sea level.

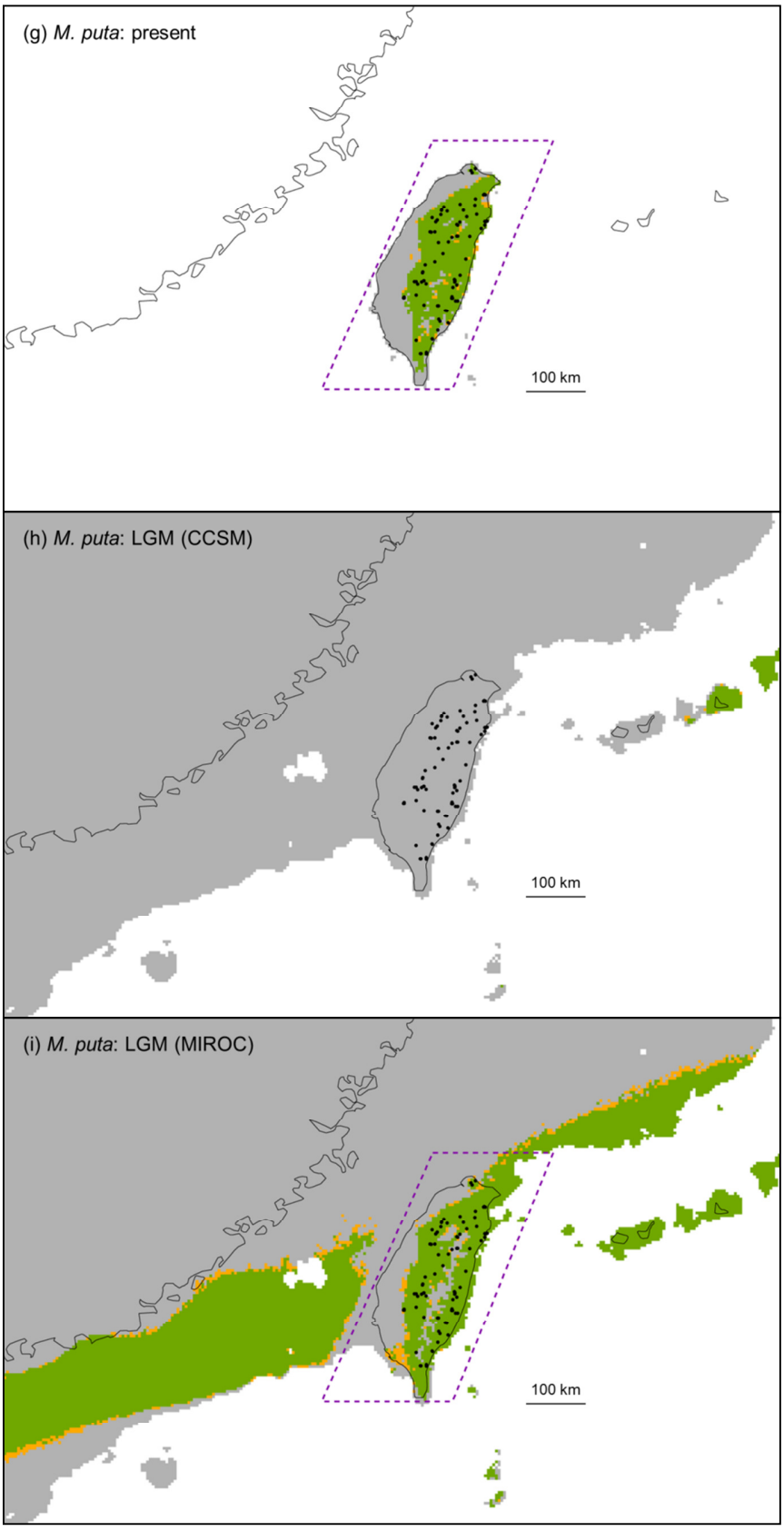
[b] Only suitable habitats in spaces delimited by dashed frames in Fig. 3.12a-l for respective species are used for comparisons.

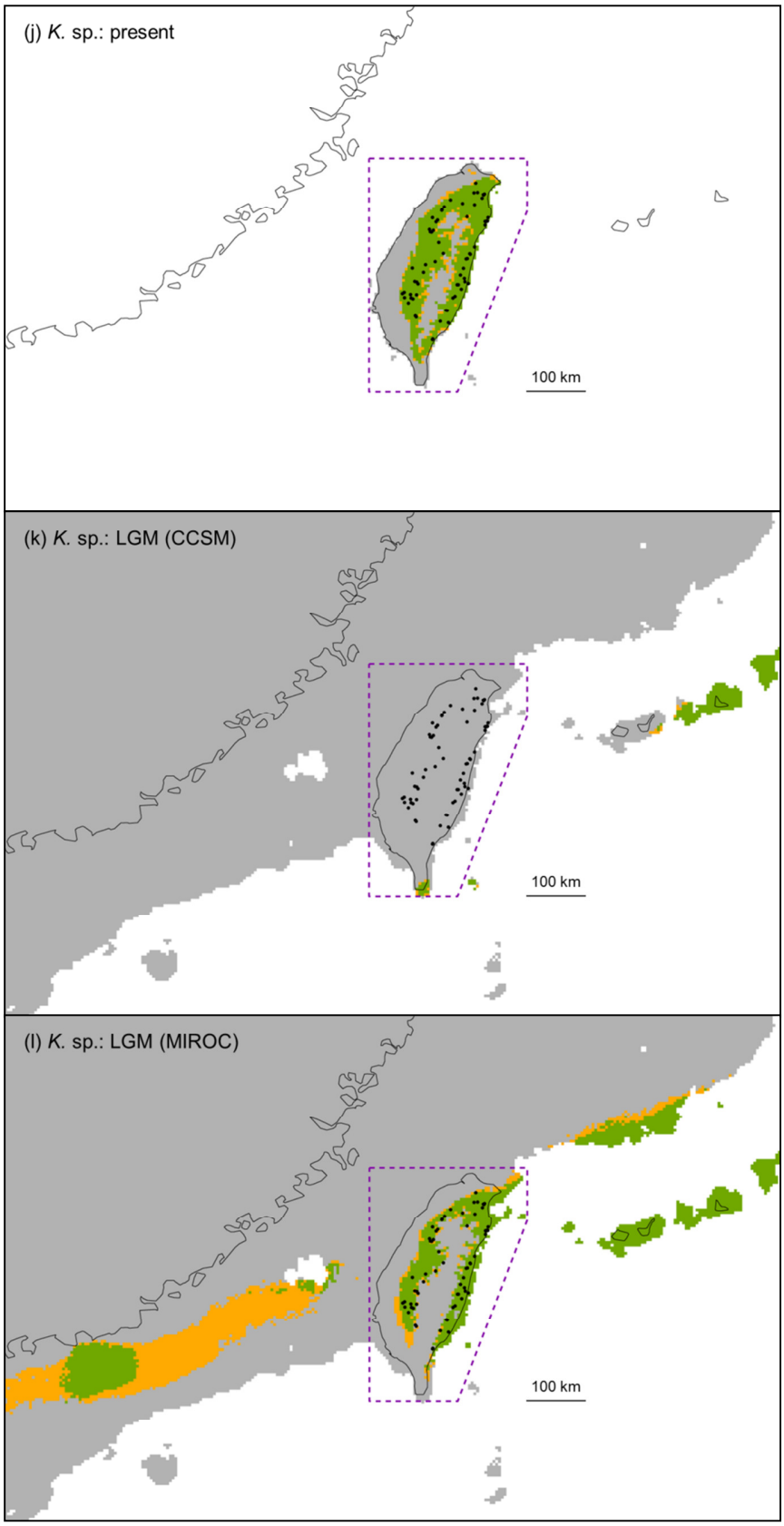
[c] No suitable habitat is predicted to occur in Taiwan in corresponding CCSM-based models for these species.

Figure 3.12 (following 4 pages) Species distribution modeling (SDM) predictions for (a-c) *Murina gracilis*, (d-f) *M. recondita*, (g-i) *M. puta*, and (j-l) *Kerivoula* sp. Panels (a), (d), (g), and (j) show predictions in present models, (b), (e), (h), and (k) show those in LGM models based on CCSM, and (c), (f), (i), and (l) show those in LGM models based on MIROC. In each panel, black dots indicate the presence data used to build the model, grey areas are unsuitable, green areas are suitable habitats predicted under MTPT, and orange areas are additional suitable habitats predicted under 5OT. Purple frames delimit the spaces for respective species where the spreads of suitable habitats at present and during LGM are compared.


(a) M. gracilis: present
100 km
(b) M. gracilis: LGM (CCSM)
100 km
(c) M. gracilis: LGM (MIROC)
100 km


(d) M. recondita: present
100 km
(e) M. recondita: LGM (CCSM)
100 km
(f) M. recondita: LGM (MIROC)
100 km


(g) M. puta: present
100 km
(h) M. puta: LGM (CCSM)
100 km
(i) M. puta: LGM (MIROC)
100 km


(j) K. sp.: present
100 km
(k) K. sp.: LGM (CCSM)
100 km
(l) K. sp.: LGM (MIROC)
100 km

3.4 DISCUSSION

Results from mitochondrial and nuclear markers

I undertook one of the most detailed comparative phylogeographic studies analyses to date and found strikingly contrasting patterns of intraspecific population genetic structure across four Taiwanese endemic bat species.

Based on mtDNA analyses, the upland species *M. gracilis* was characterised by highly divergent clades that were largely confined to different mountainous areas. A Northern clade occurred exclusively in the Hseuhshan Range and the northern part of the CMR, one Southern clade (#1) was found in both the Alishan Range and southern CMR, and a second exclusively in the Alishan Range (#2) (Fig. 3.5a and 3.6a). Paraphyly of sequences from the southern CMR and Alishan Range (Fig. 3.5a) suggested immigration of lineages from the former to the latter following a period of allopatry divergence, and this was supported by isolation with migration (IM) models that revealed unidirectional mitochondrial gene flow (Fig. 3.8c). Coalescent-based gene trees showed good support for the branching order: (Northern clade, (Southern clade 1, Southern clade 2)). Together this evidence suggests that the three divergent clades of *M. gracilis* arose in isolation from each other, first between the north and south with no subsequent admixture, and then between different mountain ranges in southern Taiwan followed by some gene flow. The times to most recent common ancestor (TMRCA) of the three *M. gracilis* clades were all in the middle Pleistocene, 300 to 800 Kyr ago for the north-south split and 200 to 500 Kyr ago for the split within the south (Table 3.4). Divergence times between the ancestral populations for these different clades are expected to be more recent than these TMRCA estimates (Charlesworth 2010), although how recent is uncertain.

Similar phylogeographic patterns of differentiated and/or divergent mitochondrial lineages along the north-south axis of the CMR were reported for two other endemic small mammals of Taiwan, the mole-shrew *Anourosorex yamashinai* (Yuan *et al.* 2006) and the Formosan wood mouse *Apodemus semotus* (Hsu *et al.* 2001). In both taxa, the TMRCA of major lineages, as estimated from a conventional mitochondrial molecular clock rate of 2% per Myr, also corresponded to the middle Pleistocene, although the exact dates did not overlap. Both studies proposed a ‘multiple refugia hypothesis; to explain phylogeographic

patterns, whereby populations survived in separate refugia within Taiwan during unfavorable climatic periods, however, while the refugium for the Formosan wood mouse was inferred to be during a glacial period, for the mole-shrew it was during the interglacial. Given that *M. gracilis* shows both a similar phylogeographic pattern and altitudinal distribution (and presumably occupies a similar ecological niche) as these mammals, the 'multiple refugia hypothesis' may also explain my results. However, the wide confidence intervals around the coalescent-based TMRCA estimates of *M. gracilis* preclude more precise inference of the historical process underlying the phylogeographic pattern, such as whether or not the refugium fell within a glacial maximum. Demographic analyses of the effective population size dynamics of individual demes also cannot address this issue since such analyses can only trace the histories of these demes back to their own TMRCA, which are more recent than would be the TMRCA of the three divergent clades (e.g. see time spans of EBSP results in Appendix 3.5a).

As well as having highly divergent clades, demes of *M. gracilis* were also found to contain many unique and private haplotypes (Fig. 3.6a), leading to strong differentiation even among nearby demes (Appendix 3.3). Conventional Φ_{ST}-based tests for isolation by distance (IBD) were thus not applied to this species, however, the observation that rare cases of shared haplotypes also involved neighboring demes suggests some gene flow, albeit highly restricted over geographic distance. Indeed such strong genetic structure seen in *M. gracilis* appears to present an extreme case when compared to other Taiwanese animals, except for its close relative *M. recondita*. Each non-singleton haplotype of *M. gracilis* was shared by an average (weighted by sampling frequencies of haplotypes) of 1.7 demes while that of *M. recondita* was shared by an average of 2.5-2.7 demes (note that concatenation of the two mitochondrial genes increased the number of private haplotypes). These values are lower than those for *M. puta* and *Kerivoula* sp. (10.2 and 10.4, respectively) as well as those for non-volant animals such as 4.1 for the mole-shrew (Yuan *et al.* 2006), 9.3 for the Formosan wood mouse (Hsu *et al.* 2001), 4.3 for the Formosan white-bellied rat *Niviventer culturatus* (Hsu *et al.* 2000), the Latouch's frog *Hylarana latouchii* (Jang-Liaw *et al.* 2008), the Sauter's frog *Rana sauteri* (Jang-Liaw & Lee 2009), and the Rumsfield stream crab *Candidiopotamon rathbunae* (Shih *et al.* 2006). The very low recorded levels of mtDNA exchange among local demes of *M. gracilis* and also *M. recondita* in this current study suggests extreme female-biased philopatry (and this limited female natal dispersal), which has also been reported for

other bats (Kerth *et al.* 2002; Martins *et al.* 2009; Ramos Pereira *et al.* 2009; Turmelle *et al.* 2011; Wilmer *et al.* 1999)

The strong local genetic structure of *M. gracilis* meant that in order to avoid violating the assumptions of panmixia, demographic analyses were only applied to individual demes. My results suggested that demes around Mt. Hehuan and in the Alishan Range had experienced population contractions or weak bottlenecks based on traditional demographic analyses (Table 3.3), although no shift in population size was detected from the extended Bayesian Skyline Plot (EBSP) analysis (Table 3.3; Appendix 3.5a). It is thus noteworthy that significant results were only found for populations that shared mitochondrial haplotypes with one another (site 23 shares one haplotype with site 24; site 31 shares another haplotype with site 32), and thus might violate a further assumption of these demographic analyses that require the target to be a closed population. Although this violation is common in the literature, in case of demes in the Alishan Range its effect might be severe due to admixture of deeply divergent sequences that bear independent coalescent information. Much of the coalescence history between sequences of two southern clades (1 and 2) reflects the isolation of subdivided populations, and instead of modeling this process, both the neutrality tests and EBSP attempt to model the genetic variation of each of the two demes in the Alishan Range as a coalescence process within a single panmictic population. Inferences based on these analyses are, therefore, likely to be misleading. For example, the genealogy co-estimated in EBSP for site 32 (tree not shown) resembles that expected for a panmictic population experiencing a demographic contraction: most sequences quickly coalesce (backward in time) with each other in very recent history (reflecting coalescence between sequences of the same clade) followed by only one ancient coalescence event (between different clades). Note also that EBSP gave an inflated estimate of the population size of site 32 (the median estimates as 90,000-60,000 from TMRCA to present) in relation to the estimate for the whole population in the Alishan Range from IM analyses (the upper limit of 95% CI as 60,000). When modeled as coalescence within a panmictic population, this inflated population size was introduced to accommodate the long waiting time until the ancient coalescence event. Given the potential influence of such artifacts, significant results in neutrality tests need to be interpreted with caution.

Based on microsatellites, the Bayesian clustering analysis revealed four groups (Fig. 3.11a) that broadly overlapped with the main divergent mtDNA clades. This suggests that genetic structure of *M. gracilis* for these two categories of genetic markers might be shaped by the same or similar historical processes, such as those described by the 'multiple refugia hypothesis'. Nevertheless, the Bayesian clustering analysis showed that there might be nuclear genetic exchange between northern and southern mountain areas, contradicting the pattern revealed by mtDNA. Indeed, microsatellite-based IBD analyses showed that genetic differentiation in this species is clinal (Table 3.6; Fig. 3.9a), suggesting some gene flow between northern and southern populations. Furthermore, IBD models built using resistance distance (RD) (Fig. 3.12a) did not fit the genetic data better than ones using simple Euclidean geographic distance (EGD) (Table 3.6), again revealing no obvious lack of connectivity. Therefore, although the 'multiple refugia hypothesis' might be still explain initial divergence of mitochondrial lineages, the contrast in permeability between mtDNA and ncDNA across central Taiwan might be better explained by behavioural differences between males and females, with male-biased gene flow among demes.

The second of the focal taxa, *M. recondita*, was also found to possess highly divergent mitochondrial clades. These were entirely separated by the CMR (in a strict sense, so that sites in northeastern Taiwan are assigned to the 'western' side of CMR) (Fig. 3.5b and 3.6b). Coalescent-based estimates of TMRCA of these two clades ranged 250 to 700 Kyr, again in the middle Pleistocene (Table 3.4). Thus the CMR has acted as a long-term barrier to east-west gene flow, as suggested for other taxa including a mammal (Oshida *et al.* 2006), reptiles (Creer *et al.* 2001; Lin 2003), amphibians (Jang-Liaw *et al.* 2008; Yeh 1997) and arthropods (Shih *et al.* 2006; Yeh *et al.* 2004). In such studies, although the mountains have probably maintained divergence, the initial differentiation could alternatively have arisen from multiple population incursions from the continent into Taiwan via land bridges across the sea during different glacial periods (see Fig. 1.2 for map showing shallow parts of the Taiwan Strait). In fact, while the 'CMR-initiated-divergence' scenario seems to have been favoured for species endemic to Taiwan (thus intraspecific divergence is supposed to occur *in situ*) (Shih *et al.* 2006; Yeh 1997), in reality there is frequently insufficient sampling from related populations on the continent to allow conclusive inferences in distinguishing the above two scenarios (Creer *et al.* 2001; Jang-Liaw *et al.* 2008; Oshida *et al.* 2006; Yeh *et al.* 2004). A study of two endemic species of *Takydromus* lizards distributed on opposite sides

of CMR attributed their origins to multiple incursions given that one was more closely related to a continental taxon (Lue & Lin 2008: fig. 1, modified from Lin 2003). In the case of *M. recondita*, the fact that both lineages have a sister relationship and are endemic to Taiwan strongly suggests they diverged on the island.

Like *M. gracilis*, *M. recondita* demes were also characterised by many private mtDNA haplotypes (Fig. 3.6b), and in this taxon this led to strong IBD. Demographic analyses show steady population sizes for most individual demes of this species although the F_S-based neutrality test suggested a history of population shrinkage for deme 2 (Table 3.3; Appendix 3.5b). Microsatellite-based analyses recovered four Bayesian clusters, which also showed some correspondence to the two mtDNA clades (Fig. 3.11b). Although little or no gene flow was detected across the CMR, signatures of admixture among clusters suggested considerable genetic mixing through the lowlands of the extreme northern and southern parts of these mountains. Greater mixing among microsatellite clusters than among mtDNA clades would appear to provide indirect evidence for male-biased gene flow. The barrier effect of the CMR was also suggested by IBD analyses, in which models based on spatial distances corrected for this barrier were a better fit of the data than that based on EGD (Table 3.6; Fig. 3.9b).

The third and more widespread of the focal species, *M. puta*, showed no such deep phylogeographic structure based on mitochondrial genes. Nevertheless in a MJ network unique haplotypes grouped into two clades separated by a few substitutions. Contrasting connection patterns within the haplotype network (Fig. 3.6c) could reflect different demographic histories in separate populations. Specifically, one clade containing most bats showed a star-like phylogeny indicative of recent population expansion, perhaps from the east where the putative ancestral haplotype was sampled. The other clade also showed a star-like topology, but contained haplotypes sampled mainly from the northern part of the CMR. In this latter clade, the ancestral haplotype was rare and connected to external haplotypes with more mutational steps, perhaps pointing to a more ancient population expansion or subdivision of an ancient population (Neigel 2002).

Analyses of differentiation among demes rejected island-wide panmixia based on global Φ_{ST}. Significant IBD among demes was detected for geographic distance (EGD) and also least cost path distance (LCPD) based on predicted habitat suitability, but not for

resistance distance (RD) also built from habitat models. Careful examination of the latter IBD patterns (e.g. Fig 3.3a) revealed that non-significance was due to inflated estimates of RD between deme 46 and the others, and to a less extent, between demes 41-43 and the others. Such overestimation of RD was caused by a narrower and more fragmented patch of suitable habitat in southern Taiwan than in the north of the island as revealed by habitat models under both thresholds used (Fig. 3.12g). In fact, predicted maps of suitable habitat did not include some areas with known records of *M. puta* in the extreme south, due to a lack of reliable GPS coordinates for these records. Therefore, I consider the narrow fragmented patch in southern Taiwan could be an artifact that needs to be verified with additional presence records of this species.

Using microsatellites, the Bayesian clustering analysis (Fig. 3.11c) and IBD tests (Table 3.6; Fig. 3.9c) for *M. puta* both failed to detect genetic structure despite a significant global F_{ST} value, although this is unsurprising given the generally low levels of differentiation among demes (see Appendix 3.7c for pairwise F_{ST} values). Indeed, Latch *et al.* (2006: 298), based on a limited set of simulations (where five populations were assumed), showed that the program STRUCTURE consistently failed to detect the presence of genetic structure when F_{ST} is less or equal to 0.01.

The fourth endemic species to be studied was *Kerivoula* sp. COI analyses suggested a close relationship with a taxon from mainland China assigned to *K. hardwickii* (Fig. 3.5d and 3.6d). However, this latter species ranges from India to the Philippines (Simmons 2005) and is in need of taxonomic revision based on recorded sequence diversity at COI (Francis *et al.* 2010).

Analysis of genetic structure based on mitochondrial COI sequences rejected a single island-wide panmictic population, and range-wide IBD was also not detected regardless of the type of distance modeled (Table 3.2). On the other hand, all mtDNA-based IBD models of demes from the western side of CMR showed significant results with good fit. Bayesian clustering analysis of microsatellite loci suggested that this region of the CMR contained two clusters that together were disjunct from the east of Taiwan (Fig. 3.11d) with little or no admixture between populations from opposite sides of the CMR. The role of these mountains in restricting gene flow was also supported by IBD tests in which spatial distances based on habitat suitability resulted in better fit than EGD (Table 3.6; Fig. 3.9d). These patterns were

remarkably similar to those of *M. recondita* which is known to also prefer forest at similar altitudes.

Further exploration of IBD based on microsatellite data showed that, as for the mtDNA analyses described above, a significant trend was found among demes from the western part of the CMR. Interestingly, however, I found no significant correlation between the pairwise values of Φ_{ST} and corresponding pairwise F_{ST} values (r =0.198, P = 0.163, Mantel test, 2000 permutations) suggesting that the IBD pattern revealed by one of these two types of markers might be an artifact. Further inspection of these patterns showed that IBD patterns based on mtDNA data were driven by two demes, 35 and 36 that showed unusually high levels of differentiation from the others, and therefore is less convincing than the IBD based on microsatellites.

Results from species distribution models

MaxEnt-based species distribution models (SDMs) developed for each species showed only slight differences in current distributions when using different thresholds (Fig. 3.12a, d, g, and j), and generally agreed with the known altitudinal ranges (e.g. Appendix 2.2). This congruence is largely due to the strong correlation between mean annual temperatures used in the models and altitude (Appendix 3.11). According to model predictions, high mountains of the CMR and Hseuhshan Range are likely to represent barriers against east-west movements for the two lowland species, *M. recondita* and *Kerivoula* sp., but not for *M. puta* and *M. gracilis* that are known to occupy forest at higher elevations.

Incorporating information from species distribution predictions into microsatellite-based models of gene flow provided biologically sensible explanations for genetic structure. For example, when accounted for the reduced habitat suitability of high altitude areas for lowland taxa, I recorded measurable improvements in the fit of their IBD models; however, no such improvement in fit was seen in the two upland species. These results were supported by findings from phylogenetic and clustering analyses, again revealing the effect of mountains on movements in the two lowland species. In contrast, a range of analyses applied to the upland specialist *M. gracilis*, and the altitudinal generalist, *M. puta*, provided no

evidence of east-west differentiation across the mountain ranges, consistent with their corresponding distribution models (Fig. 3.12a and g).

To date, a handful of other phylogeographic studies have sought to find concordance between genetic or demographic patterns and SDMs of corresponding taxa (e.g. Flanders *et al.* 2011; Jakob *et al.* 2007; Moritz *et al.* 2009; Moussalli *et al.* 2009), however, very few have incorporated SDMs to model gene flow. One exception is Wang *et al.* (2008), who showed that the IBD models had improved fit when using SDM data for an endemic spiny rat in Taiwan, *Niviventer coning*. On the other hand, similar studies of Central American frogs found little value of SDM data (Richards-Zawacki 2009; Wang & Summers 2010).

In recent years, SDMs have become increasingly popular in ecology and thus applications in genetics will also likely increase. In spite of a number of recent advances in SDM techniques, several areas would still benefit from further development. For example, the reliability of these models would be improved if they could incorporate other types of data (e.g. absence of records) (e.g. Lobo *et al.* 2010; Phillips & Dudík 2008), if there were greater balance between the fit of a model and its generality (e.g. Dudik *et al.* 2007; Elith *et al.* 2011), and, perhaps most importantly, if they were able to incorporate ecological processes and interactions relevant to the distributions of target species (Elith *et al.* 2010). In my study, additional problems were faced when assigning sensible resistance values to SDM surfaces for modeling gene flow, which are important for drawing causal links between genetic patterns and landscape features captured in SDM predictions (reviewed in Spear *et al.* 2010).

In my study I also used SDMs to reconstruct the distributions of my focal taxa at the time of the LGM. Contrasting results obtained using the two paleoclimate models (CCSM and MIROC) meant that unlike some previous studies, I was unable to infer the past distribution based on a consensus of these results (e.g. Flanders *et al.* 2011; Waltari & Guralnick 2009; Waltari *et al.* 2007). Observed genetic patterns of the four species suggested that CCSM-based predictions were unrealistic. Indeed these predictions for *M. recondita*, *M. puta*, and *Kerivoula* sp. revealed much smaller past ranges, necessitating dramatic demographic expansions in Taiwan since the past 20 thousand years (Table 3.7). Such population growth should be reliably detected in demographic analyses given this relative short span of time, however, this was not supported by relevant analyses in my study. Instead

for *M. recondita* and *Kerivoula* sp I found evidence of broadly static populations, while for *M. puta*, I detected monotonic population growth since before the LGM (Table 3.3; Fig. 3.7).

In comparison, MIROC-based predictions led to little or no conflict with genetic patterns of corresponding species. In the case of *M. recondita* and *Kerivoula* sp., for example, small differences in range sizes between the LGM and present are broadly compatible with inferred steady demographic histories. For *M. gracilis*, reductions in suitable area from the LGM to the present could be reflected by demographic contractions of a few demes (cf. Table 3.7 and Table 3.3). Finally in the case of *M. puta*, it might appear that the slight reduction of suitable area from the LGM to the present time contradicts the inferred demographic growth in this species. However, the demographic change in *M. puta* seems to be a mild process (growth of the order of 12-fold over a time-span of at least 100 thousand years). Therefore I argue that such a genetic pattern does not strongly contradict the MIROC-based predictions, especially when considering the reported uncertainty that is inherent in extrapolating distributions from the present-day to the past (Elith *et al.* 2010). To summarize, given that MIROC-based MaxEnt predictions are broadly consistent or compatible with genetic patterns in the corresponding species, these predictions might broadly capture the suitability of past habitats. I can therefore conclude that during the LGM, Taiwan provided suitable habitat for all of the four species (Fig. 3.12c, f, i, and l).

General patterns

Evidence from microsatellites provides good evidence that the topology of Taiwan has interacted with the ecology (in particular, the altitudinal ranges) of four focal Taiwanese bat species, and that this has led to some common patterns in genetic structure. As described in the section above, the CMR has prevented recurrent east-west gene flow in two lowland taxa. Consequently, greater gene flow occurs across lowland areas. In contrast, IBD tests based on microsatellites suggest that the genetic structure of *M. gracilis* represents more or less clinal variation, and that of *M. puta* is close to panmixia. These findings are also somewhat supported by data available from a fifth Taiwanese bat species: the lesser horseshoe bat, *Rhinolophus monoceros* (Chen *et al.* 2008; Chen *et al.* 2006). During my own fieldwork I recorded *R. monoceros* at altitudes of 173 to 2726 metres a.s.l. suggesting it to be

an altitudinal generalist like *M. puta*. Population genetic structure in *R. monoceros* also indicated that the CMR did not act as a strong barrier to gene flow (Chen *et al.* 2008; Chen *et al.* 2006).

Nuclear markers also suggested that biological or ecological differences among the four focal species plus *R. monoceros* have led to their different levels of inter-demic differentiation, although it remains unclear exactly what factors may underlie these differences. *R. monoceros* (Chen *et al.* (2008: fig. 2) and *M. puta* show much lower levels of inter-demic differentiation than the other taxa for a given geographic distance, and this holds when only pairwise comparisons between demes on the same side of the CMR are considered to control for any effect of the CMR. Of *M. gracilis*, *M. recondita* and *Kerivoula* sp., the former two have the highest levels of differentiation among demes. By applying Nichols and Freeman's (2004) method, I confirmed that high F_{ST} values reflect low levels of gene flow among demes rather than variation in their mutation rates. Taken together, my evidence suggests different propensities of these five species to differentiate over geographic distance, likely due to different levels of gene flow among demes. Such differences could arise from contrasting ecological traits. For example, in a study of several Malaysian bat species of the genera *Rhinolophus* and *Kerivoula*, Rossiter *et al.* (2012) showed lower levels of genetic connectivity over geographic distance in tree-roosting species than those that roost in caves. Nevertheless, for my focal species, the hierarchy of expected levels of intraspecific gene flow (highest in *M. puta* and *R. monoceros*, followed by *Kerivoula* sp., and lowest in *M. gracilis* and *M. recondita*) seems to correlate poorly with any known ecological or behavioural traits, although more studies are needed.

Appendix 3.1 Representatives (with GenBank accession numbers) of Kerivoulinae, Murininae, and Myotinae used for reconstructing the mitochondrial phylogeny (Fig. 3.4) of these groups.

Taxon	Cyt-*b*	ND1
Subfamily Kerivoulinae		
Kerivoula hardwickii	EU188771[a]	-
K. intermedia	EU188790[a]	-
K. lenis	EU188772[a]	-
K. minuta	EU188774[a]	-
K. papillosa	EU188782[a]	-
K. pellucida	EU188788[a]	-
Taiwanese *Kerivoula* sp.	Unpubl.[b]	Unpubl.[b]
Subfamily Murininae		
Harpiocephalus harpia	Unpubl.[b]	Unpubl.[b]
Murina aenea	GQ168906[c]	FJ743883[c]
Mu. bicolor	Unpubl.[b]	Unpubl.[b]
Mu. cyclotis	GQ168917[c]	FJ743886[c]
Mu. eleryi	GQ168908[c]	FJ743894[c]
Mu. florium	GQ168902[c]	FJ958188[c]
Mu. gracilis	Unpubl.[b]	Unpubl.[b]
Mu. hilgendorfi	Unpubl.[b]	Unpubl.[b]
Mu. isodon	Unpubl.[b]	Unpubl.[b]
Mu. leucogaster	GQ168912[c]	FJ743896[c]
Mu. peninsularis	GQ168911[c]	FJ743891[c]
Mu. puta	Unpubl.[b]	Unpubl.[b]
Mu. recondita	Unpubl.[b]	Unpubl.[b]
Mu. suilla	GQ168905[c]	GQ168899[c]
Mu. tiensa	GQ168913[c]	FJ743901[c]
Mu. tubinaris (Pakistan)	Unpubl.[b]	Unpubl.[b]
Mu. tubinaris (Cambodia)	GQ168915[c]	FJ743889[c]
Subfamily Myotinae		
Myotis albescens	AF376839[d]	AY033952[d]
My. alcathoe	AJ841955[e]	AY027834[f]
My. bechsteinii	AF376843[d]	AY033978[d]
My. dasycneme	AF376846[d]	AY033977[d]
My. hasseltii	AF376850[d]	AY033973[d]
My. ikonnikovi	AB106598[g]	AB106576[g]

Appendix 3.1 Continued.

Taxon	Cyt-*b*	ND1
My. keaysi	AF376852[d]	AY033963[d]
My. latirostris	Unpubl.[b]	Unpubl.[b]
My. lucifugus	AF376854[d]	AY033967[d]
My. macrotarsus	AF376856[d]	AY033951[d]
My. montivagus	AF376858[d]	AY033971[d]
My. myotis	AF376860[d]	AY033986[d]
My. nattereri	AF376863[d]	AY033984[d]
My. nigricans	AF376864[d]	AY033983[d]
My. oxygnathus	AF376841[d]	AY033988[d]
My. oxyotus	AF376865[d]	AY033956[d]
My. ruber	AF376867[d]	AY033981[d]
My. schaubi	AF376868[d]	AY033955[d]
My. yanbarensis	AB106610[g]	AB079828[g]
My. yumanensis	AF376875[d]	AY033979[d]

[a] Khan *et al.* (2010)
[b] Kuo (2004)
[c] Unpublished data on GenBank
[d] Ruedi and Mayer (2001)
[e] Stadelmann *et al.* (2004b)
[f] von Helversen *et al.* (2001)
[g] Kawai *et al.* (2003)

Appendix 3.2 Genetic variability for four Taiwanese bat species based on one or two mitochondrial genes. For each species, values were calculated for the range-wide sample and for demes with five or more sampled individual bats (n, sample size). See Table 3.1 for acronyms of measures. θ_S and π are shown as raw values multiplied by 100. Demes are numbered as in Table 2.1.

(a) *Murina gracilis*

	Cyt-*b* (768 bp)					COI (640 bp)					Cyt-*b* + COI (640 bp)				
Deme	n	S	θ_S	π	Hd	n	S	θ_S	π	Hd	n	S	θ_S	π	Hd
4	16	5	0.20	0.08	0.13	14	0	0	0	0	14	0	0	0	0
19	10	9	0.41	0.32	0.84	9	3	0.17	0.16	0.64	9	8	0.21	0.19	0.81
23	5	8	0.50	0.63	0.60	5	1	0.08	0.09	0.60	5	9	0.31	0.38	0.60
24	10	8	0.37	0.48	0.69	11	2	0.11	0.14	0.44	10	10	0.25	0.33	0.69
31	15	8	0.32	0.14	0.13	13	12	0.60	0.29	0.15	13	20	0.46	0.22	0.15
32	7	9	0.48	0.33	0.29	7	10	0.64	0.45	0.29	7	19	0.55	0.39	0.29
49	5	7	0.44	0.36	0.70	5	0	0	0	0	5	7	0.24	0.20	0.70
54	9	0	0	0	0	8	0	0	0	0	8	0	0	0	0
Overall	97	52	1.32	1.64	0.93	88	31	0.96	1.08	0.91	87	78	1.10	1.39	0.93

Appendix 3.2 Continued.

(b) *M. recondita*

	Cyt-*b* (768 bp)					COI (640 bp)					Cyt-*b* + COI (640 bp)				
Deme	n	S	θ_S	π	Hd	n	S	θ_S	π	Hd	n	S	θ_S	π	Hd
2	13	8	0.34	0.36	0.50	13	4	0.20	0.19	0.41	12	12	0.28	0.26	0.44
7	8	8	0.40	0.42	0.46	10	4	0.22	0.23	0.51	8	12	0.33	0.33	0.46
10	10	10	0.46	0.37	0.71	7	3	0.19	0.19	0.67	7	7	0.20	0.21	0.67
11	12	7	0.30	0.37	0.55	11	3	0.16	0.13	0.60	9	10	0.26	0.26	0.58
13	5	3	0.19	0.23	0.60	4	-	-	-	-	4	-	-	-	-
16	8	0	0	0	0	7	2	0.13	0.09	0.52	5	2	0.07	0.06	0.70
28	9	0	0	0	0	7	0	0	0	0	7	0	0	0	0
35	5	3	0.19	0.16	0.40	5	1	0.08	0.06	0.40	5	4	0.14	0.11	0.40
37	5	4	0.25	0.21	0.40	4	-	-	-	-	4	-	-	-	-
40	10	7	0.32	0.35	0.60	10	10	0.55	0.55	0.78	10	17	0.43	0.44	0.78
44	6	5	0.29	0.22	0.33	6	1	0.07	0.05	0.33	6	6	0.19	0.14	0.60
48	7	3	0.16	0.11	0.29	7	0	0	0	0	7	3	0.09	0.06	0.29
54	10	0	0	0	0	9	0	0	0	0	9	0	0	0	0
Overall	130	45	1.08	1.16	0.94	121	33	0.96	0.76	0.93	113	78	1.05	1.00	0.96

Appendix 3.2 Continued.

(c) *M. puta*: Cyt-*b* (768 bp)

Deme	n	S	θ_S	π	Hd
2	10	8	0.37	0.51	0.73
10	12	16	0.69	0.46	0.89
11	8	1	0.05	0.06	0.43
14	13	16	0.67	0.58	0.95
16	5	13	0.81	0.68	1.00
19	6	11	0.63	0.65	1.00
26	10	3	0.14	0.12	0.71
29	7	5	0.27	0.19	0.95
41	6	8	0.48	0.48	0.73
42	12	10	0.43	0.35	0.86
43	6	4	0.23	0.23	0.73
46	9	6	0.29	0.20	0.56
47	6	7	0.40	0.37	0.93
48	8	3	0.15	0.17	0.61
49	8	9	0.45	0.38	0.96
51	5	3	0.19	0.19	0.80
54	12	15	0.65	0.53	0.67
55	9	7	0.34	0.30	0.69
56	13	5	0.21	0.27	0.50
Overall	212	75	1.65	0.48	0.90

(d) *Kerivoula* sp.: COI (640 bp)

Deme	n	S	θ_S	π	Hd
2	11	7	0.37	0.24	0.67
6	5	2	0.15	0.13	0.40
7	6	3	0.21	0.16	0.60
10	11	7	0.37	0.39	0.76
11	7	2	0.13	0.09	0.29
16	7	5	0.32	0.25	0.71
26	11	4	0.21	0.24	0.69
35	5	5	0.38	0.31	0.40
36	13	2	0.10	0.08	0.50
46	5	2	0.15	0.13	0.40
51	11	3	0.16	0.15	0.33
53	10	4	0.22	0.15	0.38
55	11	3	0.16	0.26	0.55
56	7	3	0.19	0.22	0.48
Overall	143	19	0.54	0.30	0.77

Appendix 3.3 Pairwise estimates of Φ_{ST} for three *Murina* species based on 768 bp of Cyt-*b* gene, and for *Kerivoula* sp. based on 640 bp of COI gene. Comparisons with significant results under sequential Bonferroni corrections are shown in bold phases. Demes are numbered as in Table 2.1.

(a) *M. gracilis*

Deme	19	23	24	31	32	49	54
4	**0.75**	**0.71**	**0.65**	**0.95**	**0.93**	**0.94**	**0.98**
19		**0.42**	**0.46**	**0.92**	**0.87**	**0.89**	**0.95**
23			**0.24**	**0.91**	**0.82**	**0.84**	**0.94**
24				**0.89**	**0.82**	**0.85**	**0.92**
31					**0.79**	**0.32**	**0.85**
32						**0.72**	**0.90**
49							**0.84**

(b) *M. recondita*

Deme	7	10	11	13	16	28	35	37	40	44	48	54
2	0.29	**0.45**	**0.58**	**0.57**	**0.62**	**0.61**	**0.65**	0.48	**0.54**	**0.86**	**0.88**	**0.92**
7		0.45	**0.60**	0.59	**0.58**	**0.68**	0.66	0.49	**0.50**	**0.86**	**0.88**	**0.93**
10			**0.40**	0.13	**0.74**	0.23	**0.48**	0.11	**0.41**	**0.86**	**0.88**	**0.92**
11				0.42	**0.78**	0.51	**0.50**	0.38	**0.52**	**0.86**	**0.88**	**0.91**
13					**0.90**	**0.57**	0.63	0.29	0.49	**0.90**	**0.93**	**0.97**
16						**1.00**	**0.94**	**0.89**	**0.77**	**0.96**	**0.98**	**1**
28							0.82	0.13	**0.57**	**0.96**	**0.98**	**1**
35								0.56	**0.62**	**0.91**	**0.94**	**0.98**
37									0.24	0.89	**0.93**	**0.97**
40										**0.86**	**0.88**	**0.93**
44											0.64	**0.92**
48												**0.94**

Appendix 3.3 Continued.

(c) *M. puta*

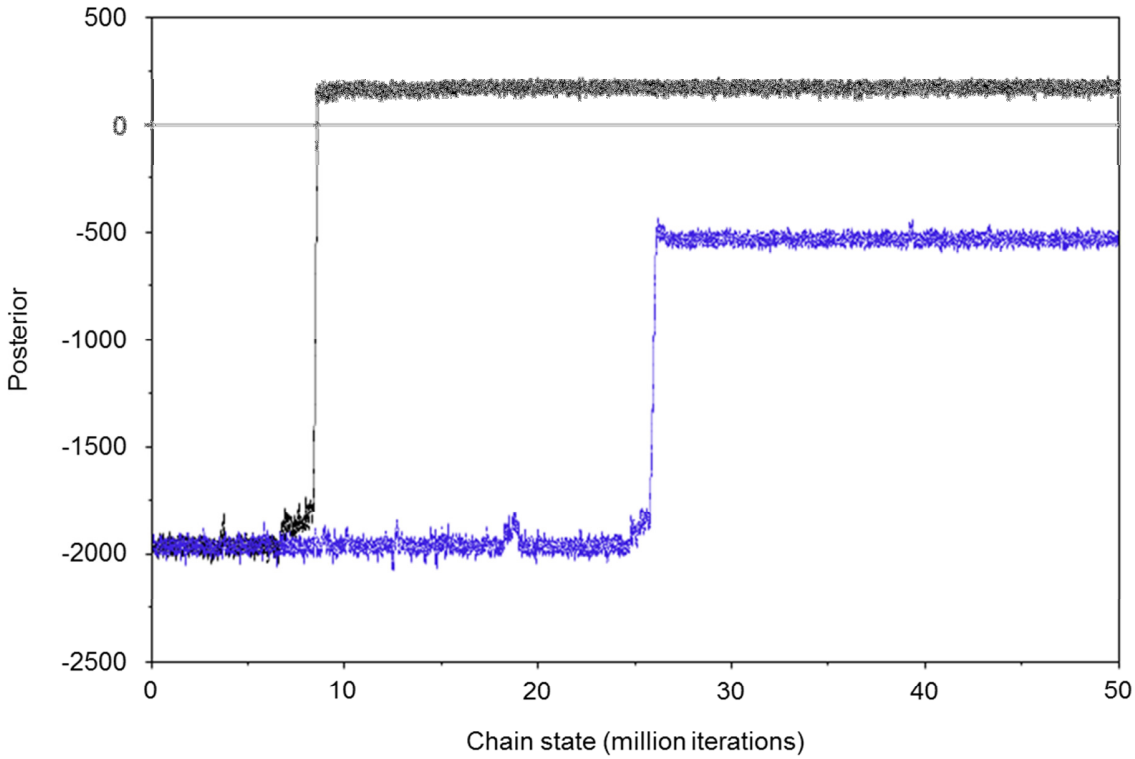
Deme	10	11	14	16	19	26	29	41	42	43	46	47	48	49	51	54	55	56
2	0.22	0.50	0.24	0.30	-0.07	0.38	0.31	0.35	0.36	0.29	0.29	0.33	0.39	0.32	0.38	0.20	0.30	0.27
10		0.13	**0.36**	**0.43**	0.10	0.00	-0.01	0.17	**0.16**	0.08	0.07	0.15	**0.19**	0.15	0.21	0.12	0.11	0.13
11			**0.53**	0.67	0.39	0.34	0.46	**0.50**	**0.44**	0.55	**0.51**	0.51	**0.63**	**0.47**	0.69	**0.35**	**0.46**	**0.46**
14				-0.04	0.10	**0.46**	**0.40**	**0.42**	**0.44**	0.39	**0.38**	**0.41**	**0.46**	**0.42**	**0.44**	**0.24**	**0.37**	**0.33**
16					0.14	0.61	0.53	0.47	0.53	0.50	0.52	0.48	0.58	0.50	0.54	**0.32**	0.49	**0.46**
19						0.25	0.18	0.24	0.26	0.17	0.17	0.21	0.30	0.22	0.28	0.07	0.18	0.14
26							0	**0.31**	0.22	0.19	0.12	0.27	**0.35**	0.25	0.44	0.17	0.18	0.20
29								0.21	0.14	0.08	0.02	0.15	0.25	0.15	0.33	0.11	0.09	0.12
41									-0.03	0.05	**0.25**	0.25	**0.34**	0.22	0.33	0.20	0.24	0.28
42										-0.05	0.16	0.21	0.28	0.19	0.29	0.19	0.18	**0.22**
43											0.10	0.19	0.29	0.18	0.34	0.11	0.13	0.16
46												0.19	**0.26**	0.18	0.33	0.03	0.08	0
47													0.09	-0.10	-0.06	0.14	-0.02	0.23
48														0.16	0.26	0.18	0.15	0.28
49															-0.06	0.16	-0.02	0.23
51																0.19	0.03	0.33
54																	0.07	-0.04
55																		0.13

Appendix 3.3 Continued.

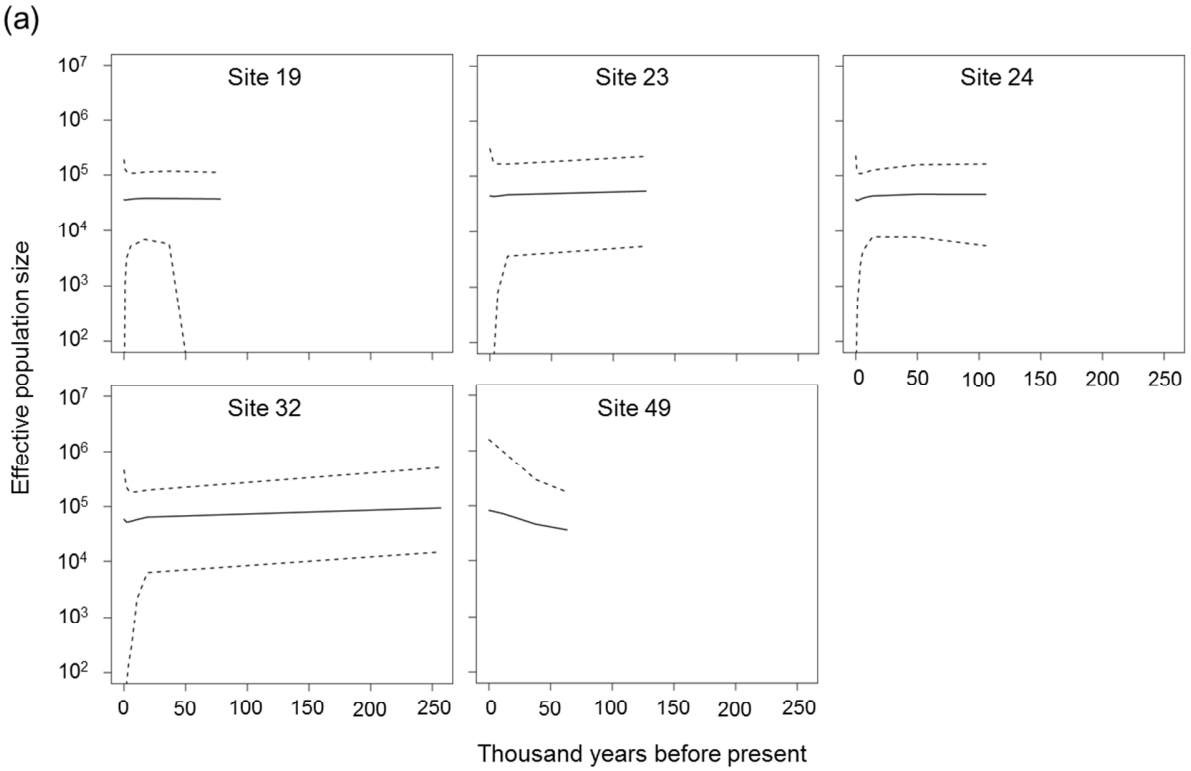
(d) *K.* sp.

Deme	6	7	10	11	16	26	35	36	46	51	53	55	56
2	-0.05	-0.02	0.15	-0.02	0.24	0.28	0.47	**0.68**	-0.05	0.01	**0.33**	0.18	0.42
6		-0.18	0	-0.19	0.21	0.15	0.45	**0.75**	-0.25	-0.12	0.37	0.09	0.42
7			0.04	-0.13	0.09	0.17	0.47	**0.74**	-0.18	-0.07	0.36	0.13	0.43
10				0.08	0.20	0.03	0.25	**0.46**	0	0.10	**0.34**	0.08	0.21
11					0.26	0.25	0.55	**0.79**	-0.19	-0.08	0.41	0.17	0.51
16						0.29	0.51	**0.74**	0.21	0.27	**0.46**	0.32	0.50
26							0.26	**0.45**	0.15	0.23	0.48	0.10	0.21
35								0.04	0.45	0.45	0.61	0.10	-0.12
36									0.75	**0.69**	**0.79**	0.37	0.09
46										-0.12	0.37	0.09	0.42
51											**0.37**	0.07	0.38
53												0.40	0.59
55													0.01

Appendix 3.4 Traces of posteriors for two runs (coloured black and blue) of the Extended Bayesian Skyline Plot analysis applied to one deme (site 31) of *Murina gracilis*. As a standard statistic outputted by BEAST (Drummond & Rambaut 2007) for visualizing the mixing of the MCMC, the posterior is the sum of the log likelihood of the tree, the log prior probability of the tree, and the log prior probability density of all other priors.

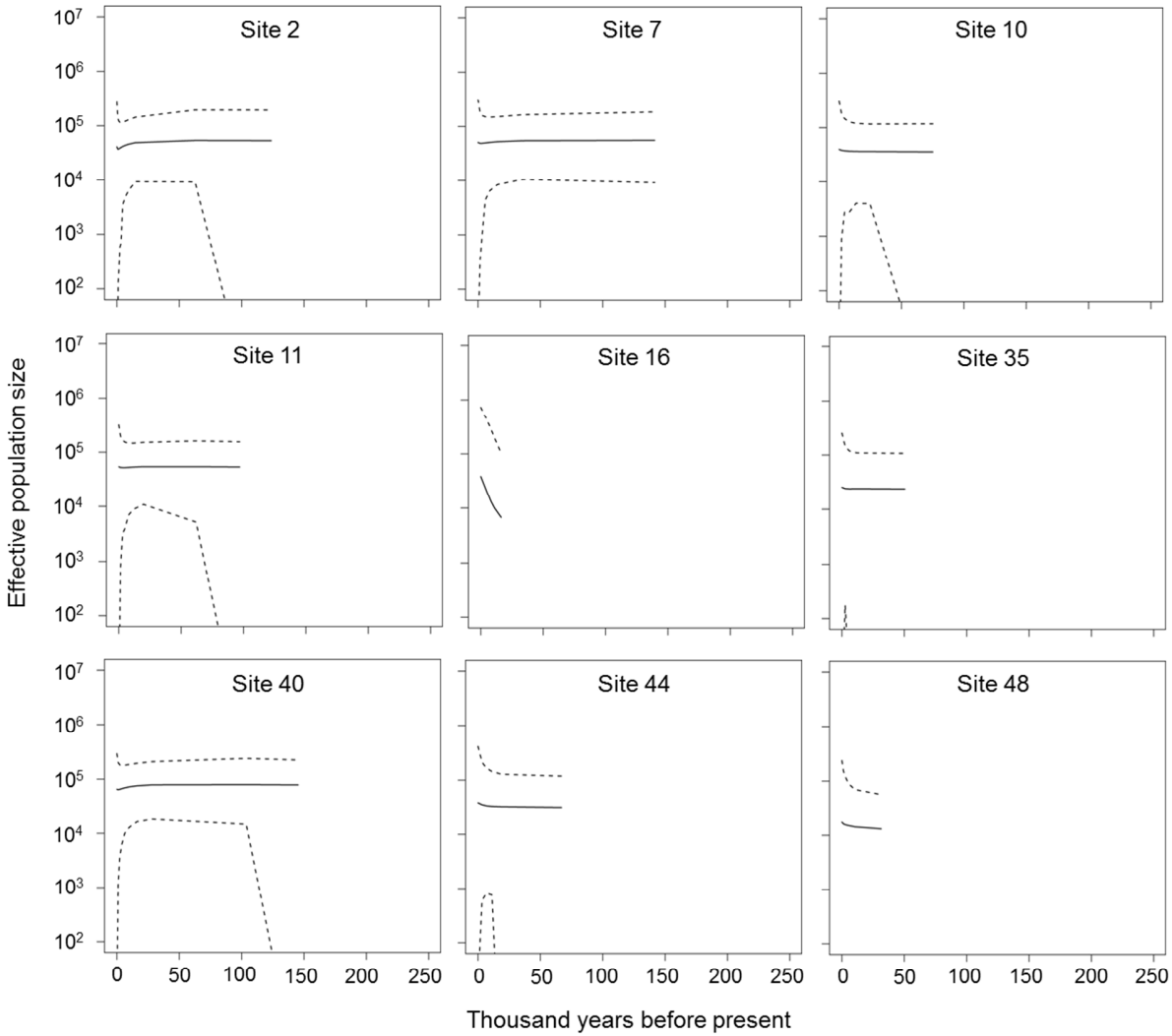


Appendix 3.5 Extended Bayesian skyline plots (solid lines for medians and dashed lines for 95% highest posterior density limits) for (a) individual demes of *Murina gracilis*, (b) individual demes of *M. recondita*, and (c) full sample of *Kerivoula* sp. and its subsets for samples from western (sites 2-37) and eastern (sites 44-56) sides of CMR. Effective population sizes are presented in the $\log_{10}$ scale. Lower credible limits for some demes in (a) and (b) are below the truncated y-axes.



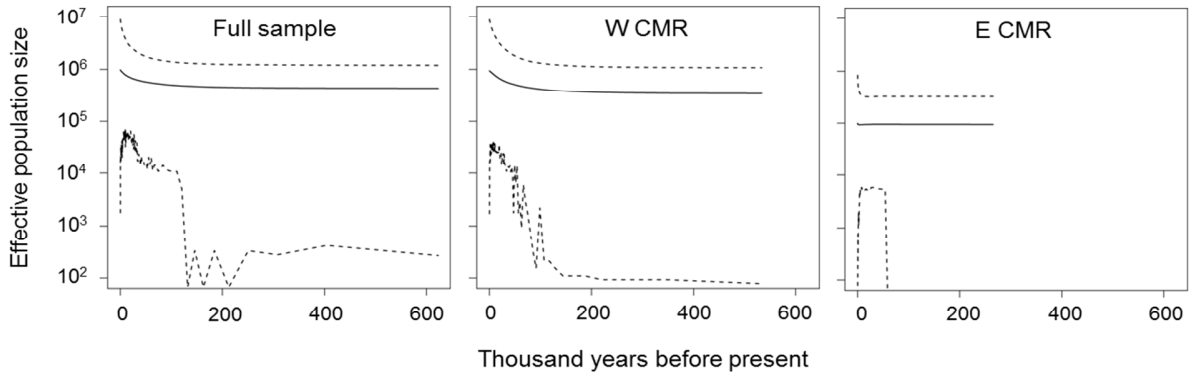
Appendix 3.5 Continued.

(b)



Appendix 3.5 Continued.

(c)



Appendix 3.6 Genetic variability for four Taiwanese bat species based on multi-locus microsatellites. Values are shown for demes sampled with five or more individual bats. n, sample size with the smallest value genotyped at any locus given in parentheses; R_S, allelic richness estimated by rarefaction for a standard sample size of four individuals; H_O, observed heterozygosity; H_E, expected heterozygosity; F_{IS}, inbreeding coefficient. Demes are numbered following Table 2.1.

(a) *Murina gracilis*: 17 loci

Deme	n	R_S	H_O	H_E	F_{IS}
4	19	4.90	0.86	0.83	-0.04
19	12	5.01	0.79	0.84	0.06
23	5	5.38	0.91	0.88	-0.04
24	11(10)	5.05	0.87	0.85	-0.03
31	19(18)	4.88	0.81	0.83	0.02
32	7	4.75	0.83	0.82	-0.01
49	5	4.80	0.86	0.84	-0.02
54	9	5.24	0.80	0.87	0.08

(b) *M. recondita*: 17 loci

Deme	n	R_S	H_O	H_E	F_{IS}
2	15(14)	4.88	0.83	0.83	0.01
7	10(9)	4.54	0.80	0.80	0
10	12(11)	4.74	0.81	0.82	0.01
11	14(13)	4.17	0.77	0.74	-0.04
13	5	3.84	0.78	0.73	-0.07
16	13(11)	4.77	0.78	0.83	0.06
28	9	4.37	0.79	0.78	-0.01
30	5	4.26	0.71	0.77	0.08
35	5	4.41	0.85	0.77	-0.11
37	5	4.53	0.79	0.81	0.03
40	11	4.61	0.83	0.81	-0.03
44	6	4.48	0.84	0.78	-0.08
48	7	4.04	0.71	0.75	0.06
54	10	4.82	0.82	0.82	-0.01

Appendix 3.6 Continued.

(c) *M. puta*: 9 loci

Deme	n	R_S	H_O	H_E	F_{IS}
2	10(9)	5.12	0.85	0.83	-0.02
10	19	5.59	0.88	0.85	-0.03
11	6	5.46	0.85	0.87	0.02
13	5(4)	5.46	0.80	0.86	0.07
14	10	5.37	0.88	0.85	-0.03
16	5	5.09	0.84	0.83	-0.02
26	10(9)	5.41	0.83	0.84	0.01
29	7	5.80	0.84	0.87	0.03
41	6	5.52	0.91	0.86	-0.06
42	15	5.82	0.84	0.88	0.04
43	6(5)	5.76	0.87	0.87	0
46	7(6)	5.73	0.87	0.87	0
47	6	5.38	0.80	0.85	0.06
48	7	5.35	0.84	0.84	0
49	8	5.27	0.85	0.84	-0.01
54	14	5.81	0.82	0.85	0.04
55	9	5.52	0.91	0.87	-0.05
56	10(9)	5.44	0.89	0.86	-0.03

(d) *Kerivoula* sp.: 9 loci

Deme	n	R_S	H_O	H_E	F_{IS}
2	16(14)	4.42	0.70	0.74	0.05
6	5(4)	4.07	0.73	0.78	0.06
7	6	4.39	0.74	0.75	0.01
10	10(9)	4.30	0.73	0.73	0.01
11	9(5)	4.26	0.72	0.75	0.04
16	10(9)	4.26	0.71	0.77	0.08
26	11	4.11	0.74	0.73	-0.01
35	6	4.07	0.70	0.69	-0.02
36	16	4.45	0.81	0.78	-0.04
46	5(4)	4.09	0.78	0.70	-0.12
51	12	4.17	0.70	0.73	0.03
53	12	4.20	0.77	0.74	-0.04
55	11	4.14	0.74	0.73	-0.02
56	7	4.47	0.84	0.79	-0.06

Appendix 3.7 Pairwise F_{ST} for the four Taiwanese bat species estimated based on multi-locus microsatellites. Comparisons with significant results under sequential Bonferroni corrections are shown in bold. Demes are numbered following Table 2.1.

(a) *Murina gracilis*: 17 loci

Deme	19	23	24	31	32	49	54
4	**0.03**	**0.03**	**0.05**	**0.07**	**0.08**	**0.07**	**0.06**
19		0.02	**0.02**	**0.04**	**0.07**	**0.05**	**0.05**
23			0.01	**0.04**	**0.05**	0.02	**0.03**
24				**0.05**	**0.08**	**0.05**	**0.03**
31					**0.03**	**0.07**	**0.04**
32						**0.05**	**0.04**
49							**0.03**

(b) *M. recondita*: 17 loci

Deme	7	10	11	13	16	28	30	35	37	40	44	48	54
2	**0.04**	**0.03**	**0.07**	**0.06**	**0.02**	**0.06**	**0.07**	**0.06**	**0.04**	**0.05**	**0.07**	**0.09**	**0.06**
7		0.02	**0.05**	**0.06**	**0.03**	0.03	0.03	**0.06**	0.04	**0.04**	**0.06**	**0.09**	**0.07**
10			**0.03**	0.02	**0.04**	0.02	0.03	**0.04**	**0.04**	**0.05**	**0.07**	**0.09**	**0.07**
11				0.04	**0.08**	**0.04**	**0.05**	**0.07**	**0.08**	**0.07**	**0.12**	**0.15**	**0.11**
13					**0.08**	0.02	0.03	0.04	0.04	0.05	0.09	**0.13**	**0.09**
16						**0.06**	**0.06**	**0.07**	**0.05**	**0.05**	**0.07**	**0.09**	**0.05**
28							0.01	0.03	0.02	0.02	**0.09**	**0.11**	**0.08**
30								0.02	0.03	0.02	0.09	**0.12**	**0.10**
35									0.01	0.03	0.11	**0.12**	**0.08**
37										0.01	0.08	0.10	**0.09**
40											**0.05**	**0.08**	**0.07**
44												0.04	**0.07**
48													**0.07**

Appendix 3.7 Continued.

(c) *M. puta*: 9 loci

Deme	10	11	13	14	16	26	29	41	42	43	46	47	48	49	54	55	56
2	0.01	0.00	0.02	0.01	0.03	0.01	0.01	0.01	0.01	0.01	0.03	0.02	0.03	0.02	0.02	0.02	0.02
10		-0.01	-0.01	0.00	0.01	0.01	-0.01	-0.01	-0.01	0.00	0.02	0.01	0.01	0.00	0.00	0.02	0.00
11			0.00	-0.01	0.00	-0.01	-0.01	-0.01	-0.02	-0.01	0.01	-0.01	0.01	0.01	0.00	0.02	0.00
13				0.00	0.01	0.01	-0.01	-0.02	-0.01	-0.01	0.01	-0.01	0.02	0.02	0.01	0.01	0.00
14					0.01	0.01	0.01	-0.01	-0.01	-0.01	-0.01	-0.01	0.02	0.00	0.01	0.00	0.00
16						0.02	0.00	0.00	0.00	0.01	0.00	-0.01	0.01	0.00	0.02	0.01	0.01
26							0.00	0.00	0.00	0.00	0.03	0.03	0.03	0.01	0.00	0.03	0.01
29								-0.01	0.00	-0.01	0.02	0.00	0.00	0.00	-0.01	0.02	0.00
41									-0.02	0.00	0.00	0.00	0.02	0.00	0.01	0.01	-0.01
42										-0.01	0.00	0.00	0.01	0.00	-0.01	0.01	-0.01
43											0.02	0.02	0.02	0.01	-0.01	0.01	0.00
46												-0.01	0.01	0.02	0.02	0.00	0.00
47													0.00	0.02	0.01	0.00	0.00
48														0.01	0.00	0.03	0.02
49															0.02	0.03	0.02
54																0.03	0.01
55																	0.02

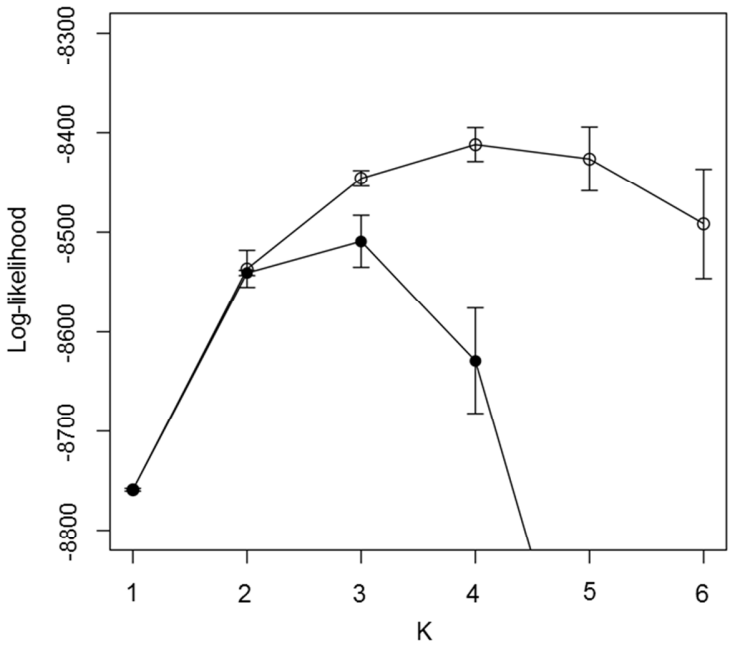
Appendix 3.7 Continued.

(d) *K.* sp.: 9 loci

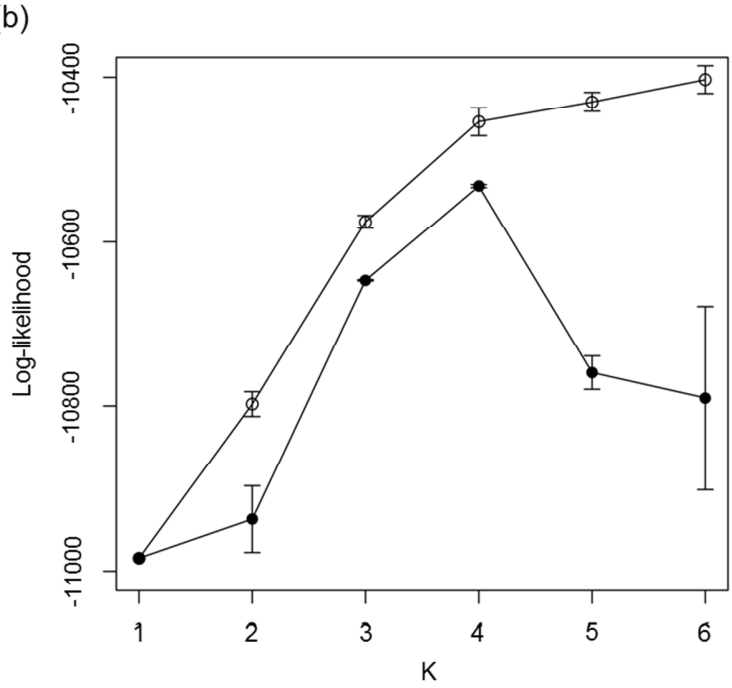
Deme	6	7	10	11	16	26	35	36	46	51	53	55	56
2	0.02	0.04	0.03	0.06	0.02	**0.05**	0.05	**0.06**	0.02	0.02	**0.05**	**0.04**	0.03
6		-0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.04	0.01	0.03	0.03	0.02
7			0.03	0.03	0.03	0.04	0.05	0.03	0.07	0.04	0.05	0.07	0.04
10				0.01	0.02	0.03	0.03	0.02	0.04	0.02	0.02	0.04	0.03
11					0.04	0.00	0.01	0.02	0.04	0.03	0.03	0.03	0.03
16						0.03	0.05	0.04	0.04	0.01	0.04	0.03	0.01
26							0.00	0.02	0.06	0.03	0.03	0.03	0.03
35								0.01	0.08	0.05	0.03	0.04	0.07
36									0.07	**0.05**	0.03	0.04	0.02
46										0.00	0.06	0.03	0.02
51											0.02	0.02	0.02
53												0.01	0.04
55													0.04

Appendix 3.8 Log-likelihood of the data for (a) *Murina gracilis*, (b) *M. recondita*, (c) *M. puta*, and (d) *Kerivoula* sp. estimated in the Bayesian clustering analyses. Plots show means (open and filled circles for models with and without sampling sites set as priors, respectively). Bars denote standard deviations based on five replicate runs. Smaller log-likelihood values on (a) and (d) are truncated.

(a)

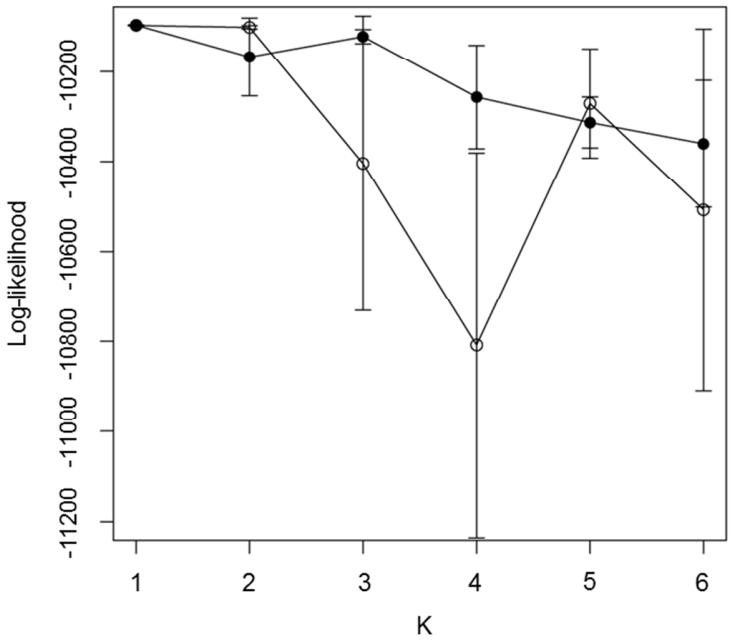


(b)

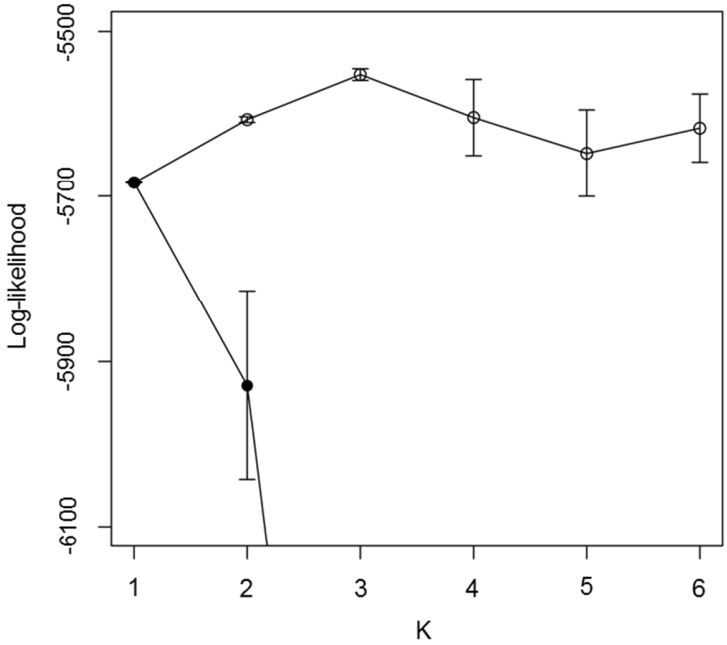


Appendix 3.8 Continued.

(c)



(d)



Appendix 3.9 Records of *Murina puta* and *Kerivoula* sp. individuals for which tissue samples were collected as part of other studies. Collectors were Cheng-Han Chou (Bat Association of Taiwan), Shiang-Fan Chen (Providence University), and Yin-Ping Fang (National Chiayi University). These records were used to improve distribution models.

Species	Locality (County)	GIS-E	GIS-N
M. puta	Chonghe Village (Taipei)	121.5935	25.2141
	Sanguang Village (Taoyuan)	121.3603	24.6725
	Shihlu Trail (Hsinchu)	121.1629	24.5631
	Jinshui Village (Miaoli)	120.9346	24.4647
	Songluo Village (Yilan)	121.5858	24.6585
	Fencihu (Chiayi)	120.6931	23.5037
	Leye Village (Chiayi)	120.7150	23.4689
	Jhutian Village-1 (Hualien)	121.2911	23.2156
	Jhutian Village-2 (Hualien)	121.3226	23.2180
K. sp.	Taipei Zoo (Taipei)	121.5908	24.9919
	Neidong Logging Road (Taipei)	121.5659	24.8451
	Tonghou Logging Road (Taipei)	121.6200	24.8498
	Baling (Taoyuan)	121.3916	24.6812
	Jiale Village (Hsinchu)	121.2079	24.7368
	Jinshui Village (Miaoli)	120.9346	24.4647
	Wushihkeng (Taichung)	120.9494	24.2729
	Sihjiaolin Forest Station (Taichung)	120.8881	24.2855
	Husi Village (Yilan)	121.6907	24.7619
	Songluo Village (Yilan)	121.5755	24.6627
	Hushan Reservoir (Yunlin)	120.6275	23.7061
	near Fencihu (Chiayi)	120.6965	23.4955
	Leye Village (Chiayi)	120.7150	23.4684
	Zengwun Reservoir (Chiayi)	120.6038	23.3088
	Guanzihling-1 (Tainan)	120.5077	23.3387
	Guanzihling-2 (Tainan)	120.4793	23.3231
	Meiling (Tainan)	120.5703	23.1781
	Shanping (Kaohsiung)	120.6877	22.9617
	Changliang Logging Road (Hualien)	121.2778	23.2779
	Chihkeshan (Hualien)	121.3748	23.3838
	Dagangkou (Hualien)	121.4778	23.4718
	Dewu (Hualien)	121.4291	23.4973
	Hegang Village (Hualien)	121.4178	23.5499
	Wanrong Logging Road (Hualien)	121.4049	23.7022
	Shuhu Village (Hualien)	121.4643	23.8440
	Liyushan (Hualien)	121.5101	23.9106
	Chihnan Recreation Area (Hualien)	121.5028	23.9185
	Zuocang Trail (Hualien)	121.5743	24.0200

Appendix 3.10 19 bioclimatic variables used for building species distribution models

BIO1 = Annual Mean Temperature

BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))

BIO3 = Isothermality (BIO2/BIO7) (* 100)

BIO4 = Temperature Seasonality (standard deviation *100)

BIO5 = Maximum Temperature of Warmest Month

BIO6 = Minimum Temperature of Coldest Month

BIO7 = Temperature Annual Range (BIO5-BIO6)

BIO8 = Mean Temperature of Wettest Quarter

BIO9 = Mean Temperature of Driest Quarter

BIO10 = Mean Temperature of Warmest Quarter

BIO11 = Mean Temperature of Coldest Quarter

BIO12 = Annual Precipitation

BIO13 = Precipitation of Wettest Month

BIO14 = Precipitation of Driest Month

BIO15 = Precipitation Seasonality (Coefficient of Variation)

BIO16 = Precipitation of Wettest Quarter

BIO17 = Precipitation of Driest Quarter

BIO18 = Precipitation of Warmest Quarter

BIO19 = Precipitation of Coldest Quarter

Appendix 3.11 Pearson correlation coefficients between pairwise bioclimatic variables from the WorldClim database (http://www.worldclim.org/). Pairwise comparisons were carried out separately between temperature variables (bio1-11) and between precipitation variables (bio12-19) within the geographic space 119.9-122.1ºE, 21.8-25.5ºN (cover the whole land area of Taiwan); the altitude is added into comparisons in the former group.

Variable	bio1	bio2	bio3	bio4	bio5	bio6	bio7	bio8	bio9	bio10	bio11
altitude	-0.98	0.17	0.44	-0.53	-1.00	-0.96	-0.56	-0.97	-0.95	-1.00	-0.94
bio1		-0.02	-0.27	0.37	0.98	0.99	0.42	0.98	0.96	0.98	0.98
bio2			0.82	-0.61	-0.14	-0.03	-0.38	0.00	-0.02	-0.15	0.08
bio3				-0.95	-0.43	-0.20	-0.84	-0.30	-0.29	-0.43	-0.11
bio4					0.53	0.26	0.96	0.42	0.41	0.53	0.20
bio5						0.95	0.57	0.97	0.95	1.00	0.94
bio6							0.30	0.95	0.93	0.96	0.99
bio7								0.49	0.47	0.56	0.26
bio8									0.92	0.97	0.95
bio9										0.95	0.93
bio10											0.94

Variable	bio13	bio14	bio15	bio16	bio17	bio18	bio19
bio12	0.73	0.48	-0.19	0.74	0.49	0.62	0.52
bio13		-0.12	0.49	0.98	-0.13	0.95	-0.13
bio14			-0.85	-0.15	0.99	-0.27	0.89
bio15				0.50	-0.86	0.60	-0.83
bio16					-0.15	0.96	-0.13
bio17						-0.28	0.91
bio18							-0.29

Chapter 4: Divergence and species formation in the *Murina gracilis* complex

SUMMARY

Phylogenetic reconstruction of mitochondrial DNA sequences suggest that the two Taiwanese endemic bats *Murina gracilis* and *M. recondita* are sister species. Such cases of geographically restricted sister taxa are exceptionally rare in nature and could point to potential divergence in the face of ongoing gene flow. To test for non-allopatric divergence and thus gene flow at different temporal scales, I generated sequences using traditional and next generation sequencing methods, and combined these with microsatellite data for coalescent-based analyses. Clustering of microsatellite genotypes revealed several individuals with signatures of species admixture suggesting possible introgression, however, sequencing of microsatellite flankers showed this was likely caused by microsatellite allele size homoplasy rather than recent introgressive hybridisation. Microsatellite flanker sequences also revealed an unexpected sister relationship between *M. recondita* and the continental species *M. eleryi*, to the exclusion of *M. gracilis*, thus conflicting with mtDNA. To dissect this relationship and mode of divergence, I undertook Sanger and 454-Pyrosequencing of ncDNA and mtDNA of larger samples. MtDNA phylogenies unambiguously supported the monophyly of each taxon as well as the sister relationship between *M. gracilis* and *M. recondita*. In contrast, multi-locus ncDNA analyses using *BEAST and isolation with migration (IM) revealed two distinct clades of *M. eleryi*, one of which was sister to *M. recondita*. The best explanation for conflict in the inter-specific relationships between the two categories of markers is that ancestors of *M. gracilis* and *M. recondita* colonized Taiwan at different times, separated by a period of allopatric divergence. *M. recondita* arrived later and captured the mitochondrial genomes of the resident *M. gracilis* population via introgressive hybridisation.

4.1 INTRODUCTION

Evidence for non-allopatric speciation

The question of how new species form is a central theme in evolution. Traditionally, processes of speciation have been categorized in a geographical framework. Most commonly, speciation is considered to be allopatric (meaning 'different place'), where the ancestral populations of new species diverge due to geographical isolation (Mayr 1942, 1963). Here divergence between populations arises from the accumulation of genome-wide genetic differences due to selection and/or neutral drift, eventually leading to reproductive isolation (RI). The establishment of reproductive barriers can prevent mixing between new species in the event that they later undergo secondary contact (Coyne & Orr 2004). Where secondary contact does lead to the formation of hybrids, these tend to be selected against (Servedio 2004). There is some evidence that reduced hybrid fitness can lead to a strengthening of pre-mating barriers, a process termed 'reinforcement'(reviewed in Coyne & Orr 2004). However, regardless of whether geographically isolated species meet again, the expectation that drift and selection will proceed more easily with little or no gene flow has led to the overwhelming view that a period of geographical isolation (allopatry) is essential for initiating speciation (e.g. Barraclough & Vogler 2000; Turelli et al. 2001).

In certain cases, speciation might proceed without isolation. So-called parapatric or sympatric speciation describes situations where the geographical range of a new species partially and completely overlaps with that of its ancestral population (Coyne & Orr 2004). These modes of speciation are considered more difficult because the initial divergence between potential new species would be hindered due to the homogenizing effects of ongoing gene flow between their gene pools at the zone of contact (Coyne & Orr 2004). This is especially true of sympatric speciation, where either one or both parental populations is expected to suffer from constant gene flow across the geographical range, and it follows that there will be a strong constraint on the accumulation of co-adapted gene complexes (and thus the association of particular allele combinations) due to the shuffling effects of recombination (Coyne 2007; Felsenstein 1981). Yet in spite of these difficulties, theoretical models have proposed routes by which parapatric and sympatric speciation (hereafter collectively referred to as 'non-allopatric speciation') can proceed (Dieckmann & Doebeli 1999; Gavrilets & Waxman 2002; Rice 1984; Slatkin 1982). These studies have been coupled

with some intriguing empirical examples that suggest non-allopatric speciation could be more common in nature than was previously assumed (reviewed in Via 2001).

A major problem for demonstrating non-allopatric speciation is that it is virtually impossible to disprove the possibility that species have undergone some period of physical isolation, and this is made especially hard given that species' ranges can shift through time (e.g. Coyne 2007). On the other hand, some researchers have emphasized that it is the gene dynamics of divergence rather than the geographical categorization of speciation processes that is important, and that a focus on geography has oversimplified this field of research. Pivotal questions in speciation biology such as 'can species form in the face of steady gene flow and recombination?', 'what level of gene flow is needed to prevent the formation of new species?' and 'what mechanism is involved in the establishment of RI in the face of gene flow?' all illustrate the importance of understanding gene dynamics (Fitzpatrick et al. 2009; Hey 2006). Indeed among reported examples of non-allopatric speciation, factors such as radical chromosomal rearrangements (e.g. Husband & Sabara 2004) or ecological shifts (e.g. Forbes et al. 2009; Jackson 2008; Kingston & Rossiter 2004) have been causally linked to the establishment of RI. Such studies have thus implied speciation processes in the face of gene flow although they have not explicitly tested for gene flow.

Population genetic methods based on isolation with migration (IM) models (see Chapter 1) offer good opportunities for assessing the extent of gene flow between diverging populations. By applying these methods, a growing number of empirical studies have demonstrated historical gene flow among closely related taxa (see the section 'Divergence and speciation in a phylogeographic framework' in Chapter 1 for examples). These reported cases of divergence with gene flow have sometimes been attributed to non-allopatric speciation (e.g. Hey 2006; Nadachowska & Babik 2009), whereas others have suggested an allopatric period followed by secondary contact (e.g. Geraldes et al. 2006; Llopart et al. 2002). However, such studies have not been able to date the occurrence of gene flow, and recent simulations have shown that IM methods are unable to provide reliable estimates for timing of historical gene flow following the initial divergence of target taxa (Strasburg & Rieseberg 2011). It follows that additional information is required in order to distinguish between cases of species divergence that arise in the face of gene flow (non-allopatric

speciation) versus those that arise as a consequence of a period of geographic isolation (allopatric divergence followed by secondary contact).

A recently proposed approach for detecting sympatric speciation (Coyne and Price (2000) and Coyne (2007)) may complement IM models for detecting historical gene flow. While these two types of methods have not previously been combined, together they offer the power to identify convincing non-allopatric speciation. Coyne and Price (2000) emphasized the value of looking for 'sister species' (species that are each others' closest relatives) that are both mobile yet restricted geographically to the same small area, such as an oceanic island. Under these conditions, they suggested that sympatric speciation rather than allopatric speciation provides the more parsimonious explanation for the origin of the two taxa. Coyne and Price (2000) surveyed 45 small oceanic islands/archipelagos and one continental island (area 0.8 to 3500 km^2, median 78.5 km^2) on which at least one endemic bird species has been recorded. For each taxon, by assessing whether or not sister species also only occurred on the same island/archipelago and thus could have evolved *in situ* via non-allopatric speciation, they found no clear evidence supporting non-allopatric speciation for any of the species pairs studied. Kisel and Barraclough (2010) have extended this approach by surveying pairs of endemic congeneric species on the same islands/ archipelagos (again, suggesting *in situ* speciation) across a much wider range of taxa. The authors found that potential cases of *in situ* speciation in highly mobile animals such as moths (Macrolepidoptera) and mammals (bats and carnivores) were associated with larger island areas (140,000 km^2 and 420,000 km^2, respectively), while much smaller areas were associated with possible examples from plants (ferns and angiosperms) and less mobile animals (snails and lizards). To summarize, there appears to be little evidence for sympatric speciation when applying Coyne and Price's (2000) criteria for island-dwelling taxa, at least for highly mobile animals.

The two newly discovered bat species *Murina gracilis* and *M. recondita* are only known from Taiwan, and have been suggested to represent sister species on the basis of phylogenetic reconstruction using mitochondrial DNA (Fig. 3.4). Such cases of geographically restricted sister species are exceptionally rare in bats; in their review, Kisel and Barraclough (2010) identified two congeneric bat species on New Zealand as the best potential example. Given that Taiwan is a much smaller island (36,000 km^2) with respect to the expected mobility of flying mammals, *M. gracilis* and *M. recondita* match Coyne and

Price's (2000) criteria and thus offer an excellent opportunity to test for non-allopatric divergence. Following the rationale described above, any historical gene flow detected between these two species would help to rule out a period of allopatry and thus lend support to their origin on Taiwan.

Allopatric divergence accompanied with introgressive hybridisation

Unlike in cases of non-allopatric speciation, the establishment of RI is not a necessary step for allopatric divergence to proceed (Turelli *et al.* 2001). As a consequence, RI might sometimes be incomplete, allowing hybridisation on secondary contact. The consequent exchange of genetic material between divergent populations is termed 'introgressive hybridisation'. Introgression between animal species in the wild was long thought to be rare (Dobzhansky 1951; Mayr 1963). Dowling and Secor (1997) suggested that the reason for this viewpoint was twofold. First, it was probably due in part to the former negative attitude of many researchers towards hybridisation, although some researchers did suggest situations where hybridisation might be expected to occur steadily (Arnold & Hodges 1995; Barton 2001; Burke & Arnold 2001; Hubbs 1955). Second, it persisted due to the historical lack of power for detecting introgression. Indeed identifying hybrid taxa is not straightforward since it often requires (i) consideration of multiple sets of characters, plus (ii) the parental taxa need to be sufficiently divergent so that the retention of ancestral polymorphism can be ruled out. In recent years, addressing (i) has become more common given advances in molecular genetic techniques and much greater availability of genetic markers. As a result, numerous examples of probable introgression across taxonomic groups have been documented, many of which have been revealed through conflicts between genetic datasets, or between genetic and other datasets (reviewed in Currat *et al.* 2008; Mallet 2005), and there is now a realization that animal hybridisation occurs more frequently than was earlier recognized. Fewer studies have considered point (ii), however, IM models from population genetics (see section 'Statistical phylogeography' in Chapter 1) provide a more rigorous means to tease apart the contribution of ancestral polymorphism and gene flow.

To explain the establishment and subsequent maintenance of introgressed alleles in a species, some studies have emphasized their potential fitness benefits (e.g. Barton 2001;

Castric *et al.* 2008; Song *et al.* 2011), although others have argued that a selective advantage is not necessary (e.g. Rieseberg *et al.* 2007). Currat *et al*. (2008) performed coalescent simulations for introgression of neutral genes between diverging taxa under different colonization scenarios, and found that introgression can occur with ease given even a modest level of successful interbreeding. These simulations showed that introgressed alleles typically move from the resident taxon into the invading taxon where the latter is out-competed by the former. In instances of no competition, introgressed alleles could be traced in both taxa, although the direction of introgression was biased, from the resident to the invader. A key point for the causality of such patterns is that during initial colonization, the invading taxon will be less abundant and below carrying capacity compared to the resident taxon whose demes will already be at carrying capacity. Consequently, alleles that introgress into the invading taxon will tend to propagate via demographic growth, whereas any that are captured by the resident taxon will not multiply and instead might actually diminish. Overall, Currat *et al*. (2008) found that > 80% empirical studies were characterised by asymmetric introgression in the direction congruent with expectations from simulations

Study objectives

In this Chapter, I assess the divergence process that has led to the formation of two putative sister species *M. gracilis* and *M. recondita*. To achieve this I generated datasets of nuclear markers and combined these with mtDNA markers. I applied a combination of IM-based modeling approaches alongside more conventional phylogenetic and demographic analyses. Since a detailed dissection of the divergence process requires a thorough understanding of the relationship between the two focal taxa, I also generated and analysed sequences for the continental species *M. eleryi*, which was found to have an unexpectedly close relationship with my two focal taxa based on mtDNA phylogenetic reconstruction conducted in Chapter 3. I hypothesized that *M. gracilis* and *M. recondita* are sister species and diverged from each other within Taiwan non-allopatrically.

4.2 MATERIALS AND METHODS

Inclusion of *Murina eleryi*

Murina eleryi is a species from North Vietnam that was very recently described by Furey *et al.* (2009). Since then, Eger and Lim (2011) subsequently reported additional specimens from Southern China, Central Vietnam, and Northern to Central Laos, mostly deposited in the Royal Ontario Museum (ROM). This species has a similar body size to that of *M. gracilis* and *M. recondita* from Taiwan, although specimens from Southern China have smaller forearm and tibia lengths than corresponding measurements of the latter two species (Table 4.1). All three taxa are characterised by a similar sized and shaped skull (Furey *et al.* 2009; Kuo *et al.* 2009), although Furey *et al.* (2009) described minor qualitative differences in their teeth. Furey *et al.* (2009) also emphasized differences between the colour of the dorsal and ventral fur of *M. eleryi* and that of the Taiwanese species, although Eger and Lim (2011) noted colour variation in *M. eleryi* that broadly overlaps with variation in *M. gracilis* and *M. recondita* described by Kuo *et al.* (2009). A diagnostic difference between *M. gracilis* and *M. recondita* based on all specimens in Kuo *et al.* (2009) and those examined during fieldwork in 2009 and 2010 is a pale grey middle portion of the tricolor dorsal fur in *M. recondita*, which is absent in *M. gracilis*. Interestingly this character can also be seen in plates of *M. eleryi* provided by Furey *et al.* (2009: fig. 1C).

Like *M. recondita* and *M. gracilis,* almost nothing is known about the ecology of *M. eleryi*, however, records of *M. eleryi* suggest a restricted altitudinal range with elevations equal to or less than 1000 above sea-level. (Table 4.2; Furey *et al.* 2009). This range fully overlaps with that of *M. recondita* whereas *M. gracilis* is found at higher elevations (see Chapter 2). Therefore, intriguing morphological and ecological variation among these three taxa warrants further taxonomic examination.

Table 4.1 Selected external measurements of *Murina gracilis* and its relatives. These include body mass, forearm length, and tibia length. Values are means ± standard deviations with the range in the next row. Standard deviation values for *M. eleryi* are not given due to small sample sizes of this species.

Sex	Taxa	Body mass (g)	Forearm (mm)	Tibia (mm)	Source
Males	*M. gracilis* (47)	4.4 ± 0.3	29.2 ± 0.6	13.8 ± 0.4	This study[a]
		3.8 - 5.0	27.9 - 30.4	13.1 - 14.8	
	M. recondita (34)	4.3 ± 0.3	28.7 ± 0.6	14.1 ± 0.5	This study[a]
		3.8 - 4.8	27.8 - 30.0	13.3 - 15.2	
	M. eleryi[b] (5)	4.4 (4)	28.3	14	Furey *et al.*
		4.0 - 5.0	27.7 - 29.4	13.0 - 14.7	(2009)
	M. eleryi[c] (5)	3.4	27	12.6	Eger & Lim
		3.0 - 4.0	26.5 - 27.3	12.2 - 13.5	(2011)
	M. eleryi[d] (3)	4	28.7	13.5	Eger & Lim
		?[f]	?[f]	?[f]	(2011)
	M. eleryi[e] (1)	4	28.3	14.2	Eger & Lim (2011)
Females	*M. gracilis* (25)	5.3 ± 0.4	31.0 ± 0.7	14.3 ± 0.4	This study[a]
		4.3 - 6.0	29.2 - 32.0	13.7 - 15.1	
	M. recondita (33)	5.1 ± 0.4	30.6 ± 0.6	14.4 ± 0.4	This study[a]
		4.3 - 6.0	29.1 - 31.7	13.8 - 15.3	
	M. eleryi[b] (6)	4.7	29.9	13.6	Furey *et al.*
		4.0 - 5.5	28.6 - 31.3	12.8 - 14.8	(2009)

[a] From fieldwork 2009-2010; juveniles of both sexes and pregnant females excluded.
[b] From three sites in Northern Vietnam.
[c] From Southern China.
[d] From Central Vietnam.
[e] From Central Laos.
[f] Not provided in the source reference.

Assessment of recurrent gene flow between *M. gracilis* and *M. recondita*

To determine whether there is ongoing gene flow across the species boundary of *M. gracilis* and *M. recondita,* in spite of their unambiguous specific delineations, I applied Bayesian clustering analysis to multi-locus microsatellite genotypes of range-wide samples of these two species. This method, which aims to define distinct genetic clusters based on genetic data without reference to their geographical origin, was also implemented in the program STRUCTURE 2.3.2 (Pritchard *et al.* 2000) (see Chapter 3 for further details).

However, unlike in Chapter 3 (where clustering was examined within each taxon; see Chapter 3, Fig. 3.11), here I pooled the data for both taxa for the following 14 microsatellite loci for which they shared common allele sizes: A4, A9, A10, A104, A109, A118, A122, B5, B9, B114, B121, B124, D110, and D117 (see Chapter 2 for loci details). Given their unambiguous species status, I predicted clear genetic distinction between the two species and thus I used the independent allele frequency model. I repeated this analysis using the admixture model to allow mixed ancestries, such as would arise via introgression between the taxa. I investigated the genetic composition of individual bats under the number of clusters (K) that best fitted the data, and thus conducted Markov chain Monte Carlo (MCMC) simulations with K values ranging from two to ten, each replicated ten times. Each MCMC comprised 750000 iterations for sampling (1000 per sample) and the same number for burn-in. The CLUMPP procedure (Jakobsson & Rosenberg 2007) was used to assess consistency across replicates, and results were visualized using DISTRUCT (Rosenberg 2004), both described in Chapter 3.

Based on the first clustering analysis described above, I calculated for each individual bat the proportion of its genetic composition, as measured by the ancestry coefficient q (Pritchard *et al.* 2000), that could be assigned to its own species. In cases where a species was represented by more than one cluster, then membership of these was summed. A few individual bats of *M. recondita* showed < 0.9 assignment to its own species (see Results), potentially reflecting genetic introgressive hybridisation. Alternatively, such signatures might feasibly also arise through allele size homoplasy (SH), which can occur in microsatellites, due to their hypervariable nature (reviewed in Estoup et al. 2002).

To distinguish between introgression and SH, I selectively amplified and sequenced the flanking regions of any microsatellite markers that showed signatures of moderate to high levels of admixture in STRUCTURE analyses. If such signatures are the result of hybridisation, then this should also be seen in the sequence of the flanking regions due to tight linkage, whereas if signatures are caused by homoplasious allele sizes then the flanking sequences will segregate with species identity. In order to identify loci for flanking region sequencing, I repeated the Bayesian clustering approach under the best-fitting K as described above, but I excluded each locus on at a time. Three replicate runs were conducted for each dataset minus a given locus. Based on the CLUMPP result under each 'pruning scheme', I re-

calculated, for each of the few individual bats of *M. recondita* mentioned above, the ancestry coefficient q assigned to its own species. For each of these, any increase in the proportion of assignment to its species that was seen between the full and reduced datasets revealed microsatellite loci that might have undergone introgression.

Following the methods describe above, I identified two loci that influenced correct species assignment (A9 and A122), for which the flanking regions were amplified and sequenced in selected individual bats of both *M. gracilis* and *M. recondita*. These selected bats included the few *M. recondita* mentioned above (with mixed ancestries across species in the Bayesian clustering analyses) and others that showed no mixed ancestry. In total, ten individual bats from each of the two species were selected (indicated by arrows in Fig. 4.1). I also included a single specimen of *M. eleryi* (HNHM 2007.28.2) from North Vietnam (Chapter 3) for comparative purposes. Primer pairs for flanking regions were designed using software Primer3 (Rozen & Skaletsky 2000) based on clone sequences obtained during microsatellite development, as follows: A9FRL (5′-GCA ATT TCA TTG TGT CCC TTG-3′) paired with A9FRR (5′-GTC ATA GTT CTA GTC TCC CAG ATC C-3′) for A9, and A122FRL (5′-CAT TCT ATC TGC CTA CCT TGA CA-3′) paired with A122FRR (5′-GGC CTT CTC ACT AGG CAC AG-3′) for A122. PCR cocktails of 15 μL, contained 0.2 μM each primer, 2.0 μL of template, and 7.5 μL of the provided master mix of the Type-it Microsatellite PCR Kit (QIAGEN). Reactions were performed on a C1000 Thermal Cycler (BIO-RAD Laboratories) with the following thermal profile: 95 ºC 5 min; 35 cycles of 95 ºC 30 sec, 59 ºC 90 sec, 72 ºC 30 sec; 60 ºC 30 min. Products were checked on the 2% agarose gel, purified with the ExoSAP-IT PCR Clean-up Kit (Affymetrix/USB), and sent to Eurofins MWG Operon (Ebersberg, Germany) where Sanger sequencing reactions were performed on an ABI 3730 DNA Analyzer (Applied Biosystems). Sequencing was performed in one direction using primers A9FRR and A122FRR.

Flanking region sequences were trimmed and the chromatograms were inspected by eye for double peaks that indicate heterozygous sites. For sequences with multiple heterozygous sites, I used the Bayesian statistical method implemented in the program PHASE 2.1.1 (Stephens & Scheet 2005; Stephens *et al.* 2001) to infer 'haplotypes' (i.e. alleles). PHASE analysis was applied separately to *M. gracilis* and *M. recondita*, using models that both allow or disallow intra-genic recombination. To facilitate phasing, known

haplotypes obtained from 454 sequencing were used as references (see next section), conducted with the option -k in PHASE. Under each model, I performed three replicated runs, each with 10000 iterations (10 iterations per sample) following a burn-in of 1000 iterations. For each sample, its inferred pair of haplotypes was accepted and used in downstream analyses when such an inference had a posterior probability of 0.6 or higher. This cut-off value was set because Garrick *et al.* (2010) has shown that PHASE usually has very low false positive rates and that there are minor differences in these rates using a threshold of 0.6 compared to one of 0.9. The inferred pair of haplotypes was also accepted only where inference based on one of the two models passed the cut-off value while a consistent inference was made across models. For each of the alignments of flanking sequences, a median-joining (MJ) haplotype network was reconstructed in the software Network 4.6.0.0 (Fluxus Technology, www.fluxusengineering.com), to provide information in their ancestries with respect to the two focal species. The epsilon parameter was heuristically set as 10 for recovering full connections in each MJ network.

Divergence in the *M. gracilis* complex revealed by mtDNA

MJ networks based on flanking regions of two microsatellites revealed an unexpected clustering of single individual of *M. eleryi* with *M. recondita* to the exclusion of *M. gracilis* (see Results for details), thus conflicting with the mtDNA phylogeny (see Fig. 3.4).

Therefore, to conduct a more comprehensive examination of the respective relationships of these taxa - which is necessary to understand the divergence processes - I obtained tissue samples of an additional 12 individual bats of *M. eleryi* from the Royal Ontario Museum (Toronto, Canada). These individual bats were collected from Southern China (n=7) and Central Vietnam (n=5), around 90 and 800 km away from the type locality of this species, respectively. Available information on voucher specimens is given in Appendix 4.1. Genomic DNA of these bats was extracted using the DNeasy Blood & Tissue Kit (Qiagen).

I used these samples to investigate the divergence histories among the three focal species for their mitochondrial and nuclear genomes. For mtDNA, I expanded the dataset in Chapter 3 by Sanger sequencing the partial cytochrome *b* (Cyt-*b*) and cytochrome c oxidase

subunit 1 (COI) genes of these additional *M. eleryi* samples. In a few cases, sequences were also obtained from GenBank (Table 4.2). Using this extended dataset, I repeated traditional tree-based methods implemented with programs TNT 1.1 (Goloboff *et al.* 2008: performing maximum parsimony analyses) and MrBayes 3.1.2 (Ronquist & Huelsenbeck 2003: performing Bayesian analyses). See Chapter 3 for details of data partition specification for the Bayesian analysis, and for other details (models of nucleotide substitutions, performing tree searching/simulations, and assessing branch supports under each of these methods). To root the trees, I included in these analyses concatenated sequences of the two focal genes from one individual bat of each of *M. puta* and *M. tiensa* (Table 4.2), which were used as outgroups.

Divergence dates (based on mtDNA) among the three taxa *M. gracilis*, *M. recondita* and *M. eleryi*, and among major clades in the former two taxa, were estimated in Chapter 3 (Fig. 3.4 and Table 3.4). In this current chapter I also estimated the TMRCA between major clades within *M. eleryi* using the coalescent-based methods in BEAST 1.7.2 (Drummond & Rambaut 2007). In brief, I partitioned the concatenated alignment for *M. eleryi* into subsets, comprising the first plus second codon positions (CP_{12}) and the third codon position (CP_3). The HKY (Hasegawa *et al.* 1985) and uncorrelated lognormal relaxed clock priors (Drummond *et al.* 2006) were set to model the mode and rates of nucleotide substitutions for these data partitions. Mean values of lognormal priors for the relaxed clock models (i.e. the parameter 'ucld.mean') were set with lognormal hyperpriors ln(Mean) = -5.8, ln(SD) = 0.17 for CP_{12} and ln(Mean) = -2.7, ln(SD) = 0.17 for CP_3 (see Chapter 3 for rationale) while standard deviations of these priors (i.e. the parameter 'ucld.stdev') were left as default (i.e. exponential distribution with mean = 0.33) for both data partitions. Currently no appropriate demographic model is available for describing coalescence in a genetically subdivided population. To assess the robustness of dates obtained from BEAST, I tested on models assuming (1) a constant-sized population, (2) a population with dynamics described by the Bayesian Skyline Plot (BSP) modelled in a piecewise-constant manner and with five groups of coalescent intervals (i.e. four demographic change points), and (3) a population with dynamics described by the BSP modelled in a piecewise-linear manner and with five groups of coalescent intervals. See Chapter 3 for the MCMC setting of Bayesian simulations and procedures for convergence assessments and for log combining.

Table 4.2 Details of loci and sequencing method for all data generated here, together with sources of published data. N = sample size; FR = flanker region sequence obtained for ten microsatellite markers.

	Mitochondrial markers			Nuclear markers	
Taxon	n	Cyt-*b*	COI	n	FR
Murina gracilis	87[a]	Ch. 3	Ch. 3	15[b]	454
M. recondita	113[a]	Ch. 3	Ch. 3	15[b]	454
M. eleryi 1[c]	8	Sanger/ Ch. 3	Sanger /GenBank[e]	7	Sanger
M. eleryi 2[d]	5	Sanger	Sanger /GenBank[e]	5	Sanger
M. puta[f]	1	Ch. 3	Sanger	-	-
M. tiensa[g]	1	GenBank[e]	Sanger	-	-

[a] Range-wide samples described in Chapter 3.

[b] Selected samples from sites 49 and 54 for *M. gracilis* and those from sites 48, 51, 52, and 54 for *M. recondita* (see Chapter 2 for justification for sample selection).

[c] *M. eleryi* 1 refers to samples from Southern China, listed in Appendix 4.1 and a single sample from Vietnam: Bac Kan Province: Kim Hy Nature Reserve (voucher: HNHM 2007.28.2).

[d] *M. eleryi* 1 is refers to samples from Central Vietnam, listed in Appendix 4.1.

[e] Available with accession numbers HM540936-7, JQ601475, JQ601483, JQ601503, and JQ601510 for *M. eleryi* 1, HM540933, HM540938, and JQ601543 for *M. eleryi* 2, and GQ168913 for *M. tiensa*.

[f] Sampled in Taiwan: Taitung County: Dulanshan Trail (site 46; no voucher).

[g] Sampled in Vietnam: Bac Kan Province: Kim Hy Nature Reserve (voucher: HNHM 2007.28.1).

Divergence in the *M. gracilis* complex revealed by ncDNA

Markers

To gain a better resolution of the divergence process in the nuclear genome among the three species, I developed a set of ten anonymous autosomal markers based on microsatellite flanker regions (Table 4.2). These were a subset of the 14 described in the previous section except for A10, B9, B121, and D110. For each locus I re-designed primer pairs based on clone sequences to amplify the flanker but to exclude the microsatellite motif (Appendix 4.2). Primers pairs designed for A9 and A122 amplified fragments that overlap with corresponding ones described in previous section.

For *M. gracilis* and *M. recondita*, I obtained sequences for both alleles of each of the ten nuclear loci via high-throughout 454-Pyrosequencing (Margulies *et al.* 2005). Given that both species show strong genetic substructure associated with geographic divisions within Taiwan (see Chapter 3), which would violate assumptions of coalescent-based analyses, I selected 15 individual bats of each taxon from a limited part of the range. To identify individual bats from their sequences, I incorporated 3-bp tags at the 5′ end of the synthesized forward primer for each locus. For *M. gracilis*, I used the tags ACA, ACG, ACT, AGA, AGC, AGT, ATA, ATC, ATG, CGA, CGC, CGT, CTA, CTC, and for *M. recondita* I used the tags CTG, CAT, GTA, GTC, GTG, GAC, GAG, GAT, GCA, GCG, GCT, TAC, TAG, TAT, TCA and TCG. PCR products of each locus were visualized on a gel and pooled for 454-Pyrosequencing at Mission Biotech (Taipei, Taiwan) on a 454 GS Junior System (Roche/454 Life Sciences).

I used the software CLC Genomics Workbench 4.9 (CLC bio, Denmark) to filter and sort reads generated from 454-Pyrosequencing, as follows. First, original target sequences of corresponding microsatellite clones were used as references to which 454 reads were mapped. I used filtering parameters that specified a sequence length coverage of 0.6 and a similarity of 0.9 of each read to the corresponding reference sequence. For each locus mapped, reads were sorted by individual bat using the 'Process Tagged Sequences' function on the basis of the 3-bp tag plus 2-3 sites of the 5′ end of the forward primer. Subsequently, I used the MUSCLE algorithm (Edgar 2004) in the program MEGA 5 (Tamura *et al.* 2011) to establish an alignment for each locus based on reads sorted into individual bats. A maximum of 50 reads

per individual were randomly selected (the maximum numbers of reads was used if < 50 reads were available) for building each such alignment. Assuming that the rate of error resulting from PCR and 454-Pyrosequencing would be lower per nucleotide site than the correct base (Galan *et al.* 2010), I inferred alleles for each bat. In a small number of individuals, a third haplotype was erroneously inferred, and thus to separate out these assumed artifacts, I adopted the following conservative criteria for acceptance of haplotypes for downstream demographic analyses: (1) for each individual, only inferred haplotypes with frequencies of ≥ 5 were accepted such that the occurrence of two alleles ≥ 5 signified heterozygosity and (2) each individual was considered homozygous for a locus when only one haplotype was detected with a frequency of ≥ 10. Bats that did not meet these criteria, because only one haplotype was detected at a frequency of ≥ 5 but < 10, were considered to be potentially heterozygous but with missing alleles. Here the recorded haplotype was included in the analyses and the second was treated as missing data.

The discovery that *M. eleryi* was closely related to *M. recondita* was made after the 454-Pyrosequencing had been undertaken and, therefore, amplification and sequencing of the ten *M. eleryi* flanker regions was instead completed with Sanger sequencing, in both directions. PCRs were set up with annealing temperatures of 55-60ºC and purification was as outlined in the previous section. Indels (1-9 bps) were observed in four out of these ten loci, shown by double peaks on chromatograms downstream of specific sites. For each individual bat with heterozygous sites, phase of haplotypes was directly inferred by disentangling traces from superimposed sites in combination with complementary information from the reads in both directions (Flot *et al.* 2006). By incorporating phased heterozygous sites as well as those showing no ambiguity (i.e. cases where only one heterozygous site was observed for each individual bat) into known information (option -k), PHASE analyses were carried out to recover pairs of haplotypes of each individual bats at each of the ten loci. This approach was applied separately to *M. eleryi* 1 and *M. eleryi* 2 since they appeared to be divergent (see Results).

Final alignments of each of the ten nuclear markers contained samples of *M. gracilis*, *M. recondita*, *M. eleryi* 1 and *M. eleryi* 2. Prior to downstream demographic analyses, all indels were pruned from alignments, and the web-based software IMgc (Woerner *et al.* 2007) was applied to screen for the largest recombination-free data block in each locus. In brief,

this software first detected potential intra-locus recombination using Hudson and Kaplan's (1985) four-gamete test. Recombination between a pair of segregating sites will result in all four combinations of nucleotide variants of these sites under the assumption that the target sequence has evolved following the infinite site model. Then IMgc purged either individual sequences or the partial alignment containing the segregating site, each of which caused a four-gamete violation, while keeping the largest available data set in terms of both numbers of sequences and segregating sites based on a specified relative weight of one of these factors over the other one. When applying this procedure, I specified equal weights for both factors. The IMgc procedure was not applied to locus A122, due to the detection of a tri-allelic segregating site that suggested this locus did not evolve in an infinite-site manner (and thus violated the assumption of the four-gamete test) and might experience a faster mutation rate (also corroborated by analyses of genetic variability in the four focal taxa; see Results). Throughout this study, the HKY model was used for nucleotide substitutions for A122.

Locus-wise genetic diversity metrics derived for each taxon included the number of segregating sites, the Watterson's (1975) theta value, and the nucleotide diversity, all obtained using the program DnaSP 5.10.01 (Librado & Rozas 2009). Note that theta equals 4Nu (N, effective population size; u, mutation rate) for neutral genetic variation in the diploid organism, and is also the expected value of the nucleotide diversity under the same condition (Tajima 1983). Thus, values of both Watterson's (1975) theta and nucleotide diversity obtained for the ten loci separately can serve as indicators of relative levels of mutation rates of these loci, a rationale that suggested locus A122 evolved faster than the others.

I used the program Network 4.6.0.0 to construct MJ networks for samples of all four taxa in order to gain preliminary information about levels of genetic differences among them. These MJ networks were built based on alignments prior to recombination testing so that changes introduced by IMgc, if any, could be annotated *a posteriori* on the networks. An epsilon value as 10 was used for building each of these MJ networks. Polzin and Daneshmand's (2003) maximum parsimony algorithm was used to remove unnecessary links in the network for the locus A122 where complex connections were seen.

Divergence assuming no post-split gene flow

To infer the phylogeny among the *M. gracilis* complex taxa based on the ten nuclear markers while taking the uncertainty introduced by lineage sorting into account, I used Heled and Drummond's (2010) method *BEAST. When applied to data with multiple representatives from each species, this method models (multi-locus) coalescences along the whole course of divergence among multiple species under the assumption that there is no horizontal transfer/gene flow across species once they have split. Heled and Drummond (2010) have shown that their method outperforms competing ones (include the concatenation method and an another coalescent-based method) in its accuracy for estimating the species tree topology and/or their divergence times. *BEAST analyses were performed with the Jukes and Cantor (1969) nucleotide substitution model of all loci except A122, for which HKY was used. Two demographic models were explored for coalescences in each of the four extant taxa as well as in their ancestral populations; first, the piecewise-constant (PC) model which assumed occurrence of changes of effective population sizes only at split events, and second, the piecewise-linear and constant root (PLCR) model which assumed populations change their sizes linearly through time and the size of each ancestral population was equal to the sum of those of its corresponding pair of descendant populations. The constant-rated speciation model without extinction, the Yule model (Yule 1924), was used for split rates among focal taxa. For calibrations of multi-locus coalescences (and thus to estimate split times between pairs of taxa), I specified for each locus a strict clock with the clock rate sampled from a normal prior which had a mean as 0.0022 and a standard deviation as 0.001 (in a unit of per nucleotide site per Myr) truncated at zero to avoid negative rates. This normal prior was based on Kumar and Subramanian's (2002) study which empirically derived a mean estimate of the clock rate of mammalian genomes as 2.2E-09 substitutions per site per year (p.806). These authors also obtained a bell-shaped and symmetric distribution for levels of sequence divergence between human and mouse based on neutral sites of more than 2,000 genes, which had a mean of 0.466 variance of 0.035 (fig. 3). The standard deviation (SD) of my normal prior was derived by scaling it to 2.2E-09 while keeping the SD/mean ratio as the square root of 0.035 over 0.466. In summary, my rationale for calibrations was to sample clock rates of the ten loci from the same prior while allowing them to vary based on the data. Three replicated runs of MCMC under each demographic model were conducted using the program BEAST 1.7.2, each with 100 million iterations

(10000 iterations per sample) of which first 10% were discarded as burn-in after trend plots were checked. Input files with the above settings were prepared with the help of the adjunct program of BEAST: BEAUti.

To assess mixing of MCMC under each setting, trends and marginal distributions of parameters within and across replicated runs were inspected by eye in the program Tracer 1.5 (Rambaut & Drummond 2007). I ensured each parameter had (when logs were combined over replicate runs) an effective sampling size (ESS) of ≥ 200. It should be noted that parameters for effective population sizes in the BEAST log files can be misleading due to switches of population labels along the MCMC. To assess properly the mixing of these parameters, I repeated the traces using the 'starbeast_demog_log' python script in the biopy-0.1.7 package (available from: http://code.google.com/p/biopy/). Once consistent marginal estimates were observed across replicated runs and high ESS values were obtained for log-combined results, MCMC samples of taxa trees in replicated runs were combined for reports using another adjunct program of BEAST, the LogCombiner. Finally, I calculated the Bayes factor of the PLCR-based model over the PC-based model using Tracer 1.5. The resultant value was compared with Jeffreys' (1961) scales for interpreting the support from the data for the former over the latter model.

Divergence allowing post-split gene flow

The expected level of genetic differentiation between a pair of closely related taxa, will be determined by the size of their ancestral population and the time since they split, both of which were accommodated in the *BEAST analysis above. However, *BEAST ignores the potential presence of gene flow since the split, which is an additional factor able to influence divergence. To assess the effect of such post-split gene flow, together with other factors, in shaping genetic differences among the four taxa in the *M. gracilis* complex, I applied Bayesian analyses based on multi-locus isolation with migration (IM) models (Hey 2010b; Hey & Nielsen 2004) on the dataset of ten loci, implemented with the program IMa2 8.26.11 (Hey 2010b; Hey & Nielsen 2007).

First, I performed IM analyses on each pair of taxa among the four giving size pairwise comparisons. These analyses were conducted because the splitting order of the four

taxa was not known *a priori* from any analysis where post-split gene flow has been taken into account. The HKY model was used for the locus A122 while the infinite site model was used for all other loci. The inheritance scalar was fixed as one for each locus, as each was tightly adjacent to an autosomal microsatellite. Truncated uniform priors for entry demographic parameters of IMa2 were set following guidelines in the software manual (available from: http://genfaculty.rutgers.edu/hey/software); specifically I set upper bounds of priors for the composite parameter of the split time (with the flag -t), the rescaled migration rate in either direction (-m), and the population mutation rate for each of the extant/ancestral taxa (-q) based on the genetic variability of *M. eleryi* 1, as this species had highest genetic variability of the four (see Results). The geometric mean of Watterson's (1975) theta values over loci - an estimate for the population mutation rate - was 0.0055 per nucleotide site, which corresponded to a value of 1.1 per locus given a geometric mean of sequence lengths over the ten loci as 196. Following the IMa2 manual, I set -t2 and -q5 for two and five times of the above estimate, respectively. For post-split migrations, I set -m2 for a moderately high upper bound for each 2NM (number of migrant gene copies per generation), since 2NM = q × m/2 = 1.1 under the set 'm' value. Nevertheless, results under the above setting gave some m estimates with marginal densities that deviated from zero, so I also repeated analyses with -m5. In addition, I explored further m priors, modeled by an exponential distribution with a mean value of 0.5 (with flags -j8 -m0.5). This latter setting, by its nature, circumvents truncation of the marginal estimate by an upper bound. The use of such exponential priors is justified under the expectation of low migration (and thus lower values should have more weight) between diverging taxa (Hey 2010b).

Apart from modeling divergence among pairs of taxa, I also performed IM analyses on all four focal taxa together with the following phylogeny specified as the most likely one based on *BEAST analyses: (((*M. recondita*, *M. eleryi* 1), *M. eleryi* 2), *M. gracilis*). Four-taxon analyses were carried out to complement pairwise analyses with the aim of (1) separately evaluating migration during different episodes, as delimited by successive splits, throughout the whole divergence course of the four focal taxa, and (2) circumventing violations of assumptions of IM analyses potentially incurred in two-taxon analyses. Potential violations include the assumption that the pair of taxa are sister species, and that neither has exchanged genes with any unsampled taxa. Prior settings for demographic parameters for four-taxon analyses were as two-taxon ones, with the exception was that the

setting with -m5 was not applied because Hey (2010b) has shown using simulation that a truncated uniform prior of m with a high upper bound can lead to highly biased estimates of m and other parameters.

Metropolis-coupling MCMC (MC^3) for each of two- and four-taxon analyses was carried out in the software IMa with a geometric heating scheme for 60 chains (heating parameters -ha0.95 -hb0.8), with one million iterations discarded as burn-in followed by five million iterations for sampling (100 iterations per sample). Two (for two-taxon analyses) or three (for four-taxon analyses) replicate runs were performed under each prior setting for assessment of convergence.

Each of these 42 runs was performed on a high performance computer cluster, either at Queen Mary, University of London (see Chapter 3 for specifications) or at the Wellcome Sanger Institute (64-bit Linux computer, 6,800 CPUs, total memory of 18,336 GB). Each two-taxon run took an average of 43 hours to finish while each four-taxon run took an average of 128 hours (total 2380 hours). The results of replicated runs that showed convergence under each setting were combined using the 'L mode' function of IMa2.

The likelihood ratio test (LRT) proposed by Nielsen and Wakeley (2001) was used for statistical evaluation of evidence of post-split gene flow between focal taxa (see Chapter 3 for brief descriptions). LRTs were performed in IMa2 for each of the two- and four-taxon analyses with each prior setting. Hey (2010b) warned that the assumption of this test that the model likelihood is proportional to the posterior density is not met with an exponential prior, although reporting LRT results can still prove useful (e.g. Hey 2010a). Finally, to convert entry demographic parameters into conventional interpretable scales, I applied a mutation rate of 4.3E-07 per locus per calendar year, and a generation time of two years. The former scalar was derived from the mean estimate of the neutral mutation rate of mammalian nuclear genomes as 2.2E-09 per nucleotide site per calendar year (Kumar & Subramanian 2002), together with a geometric mean of sequence lengths of the ten markers used in this study as 196.

4.3 RESULTS

Tests for recent hybridisation between *M. gracilis* and *M. recondita*

When applied to genotypes of all 14 microsatellite loci, clustering implemented in STRUCTURE showed increasing log-likelihood values for K values from two to four (Fig. 4.1). At K = 4 two configurations of clusters were seen in ten replicate runs: in seven cases, there were two clusters within *M. gracilis* and the other two in *M. recondita* (with similarity index G′ between each other > 0.99; the CLUMPP result visualized in Fig. 4.1) whereas in the other three cases there was one cluster for the whole sample of *M. gracilis*, two clusters of *M. recondita* (as above) and the fourth empty cluster containing no substantial fraction of any single individual of either species (not shown). Inferring such empty clusters - sometimes termed 'ghost populations' - in clustering analyses is an issue yet to be solved (reviewed in Guillot 2008). Here I based my subsequent analyses on the former configuration since it showed a better fit to the data under the same K value (Fig. 4.1). Increasing the K value from four to ten consistently introduced more empty clusters besides the same four clusters shown when K = 4 (not shown).

Close inspection of the configuration presented in Fig. 4.1 (K = 4) showed that all but four bats were assigned almost exclusively to a single cluster of their own species, each with a sum of ancestry coefficients q for its own species > 0.9. Four individuals showing >10% 'foreign' genetic composition were all of *M. recondita* (labeled as RA, RB, RC, and RD in Fig. 4.1) sampled from two demes (sampling sites 2 and 16, see Fig. 2.1). Repeated clustering analyses with each locus in turn removed revealed that this signature of genetic admixture was largely attributed to locus A9 and, to a lesser extent, A122 (Appendix 4.3; also visualized in Fig. 4.1). To determine whether alleles of this locus have indeed undergone introgression between the two taxa, I amplified and sequenced the flanking regions of A9 and A122 in selected bats (shown by arrows in Fig. 4.1) including the above four *M. recondita* individuals. A specimen of *M. eleryi* was also included.

In a median-joining (MJ) network of A9 sequences, all individuals of *M. gracilis* and *M. recondita* individuals clustered by species, however, the sequences of the latter taxon were also seen to unexpectedly cluster with that of *M. eleryi* (Fig. 4.2). In a MJ network of A122, sequences, one sequence copy of *M. recondita* did not appear more closely related to

other *M. recondita* sequences copies but instead was shared with *M. eleryi* (this individual of *M. recondita* showed no admixture in the clustering analysis). In summary, flanking regions segregated with the two sister species, and thus provided no evidence that admixture was driven by recent introgression between *M. gracilis* and *M. recondita*. Consequently, erroneous clustering of genetic composition in STRUCTURE is best explained by microsatellite allele size homoplasy. On the other hand, the unexpected close relationship between *M. recondita* and *M. eleryi* could be due to introgression that was not predicted. Based on this finding, I expanded sampling of *M. eleryi* in subsequent analyses for dissecting divergence histories among all these *Murina* taxa.

Figure 4.1 Bayesian clustering analyses implemented in STRUCTURE. The upper panel shows log-likelihood values of ten replicated runs under each of successive numbers of clusters (K) for the full dataset (14 loci). The lower panel shows DISTRUCT plots for analyses based on the full dataset, and for 13 loci with either locus A9 or locus A122 removed. The order of individuals within each species follows Fig. 3.11a-b. Selected individuals for sequencing of flanking regions of loci A9 and A122 are indicated by arrows (see the text for explanation).

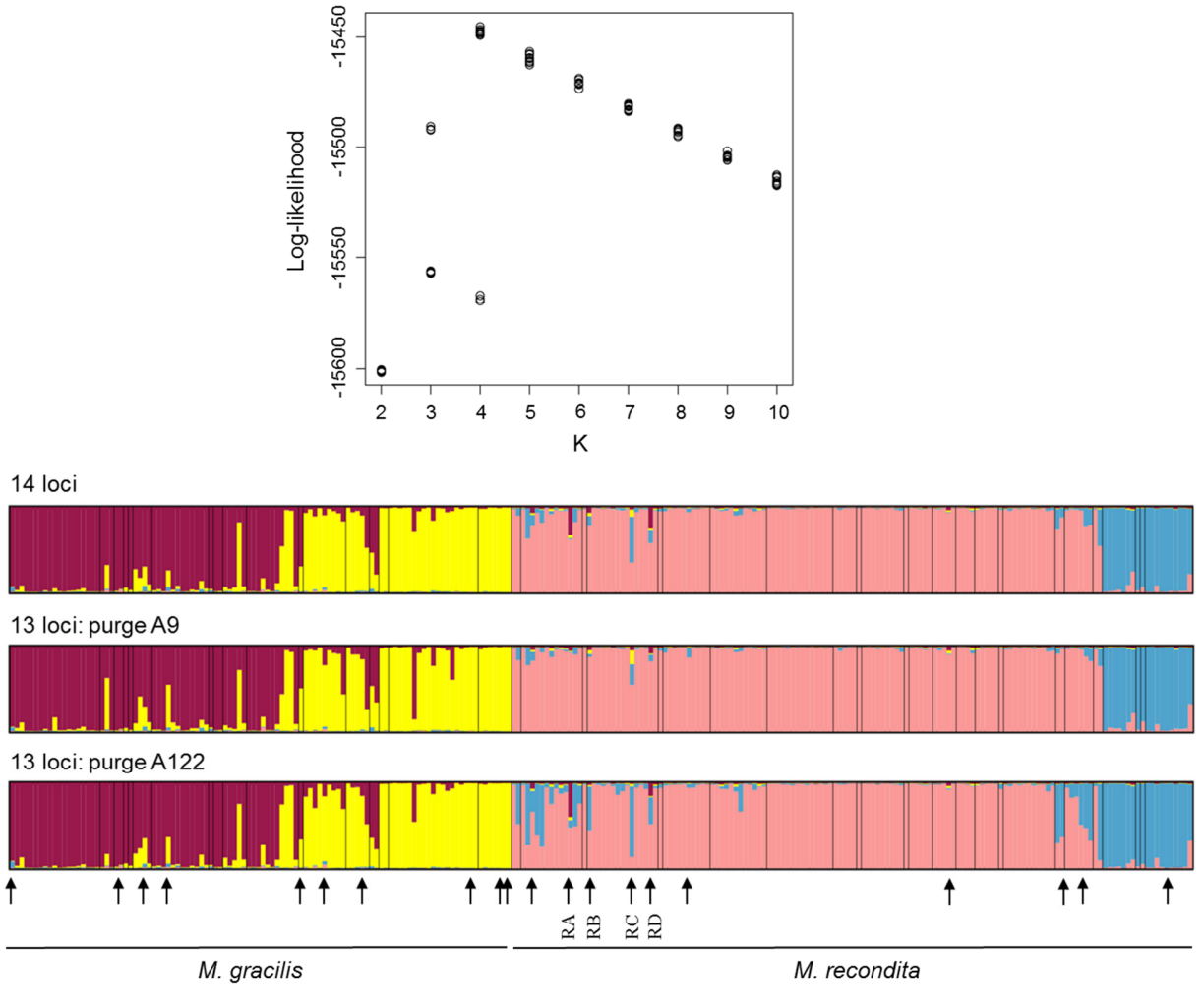
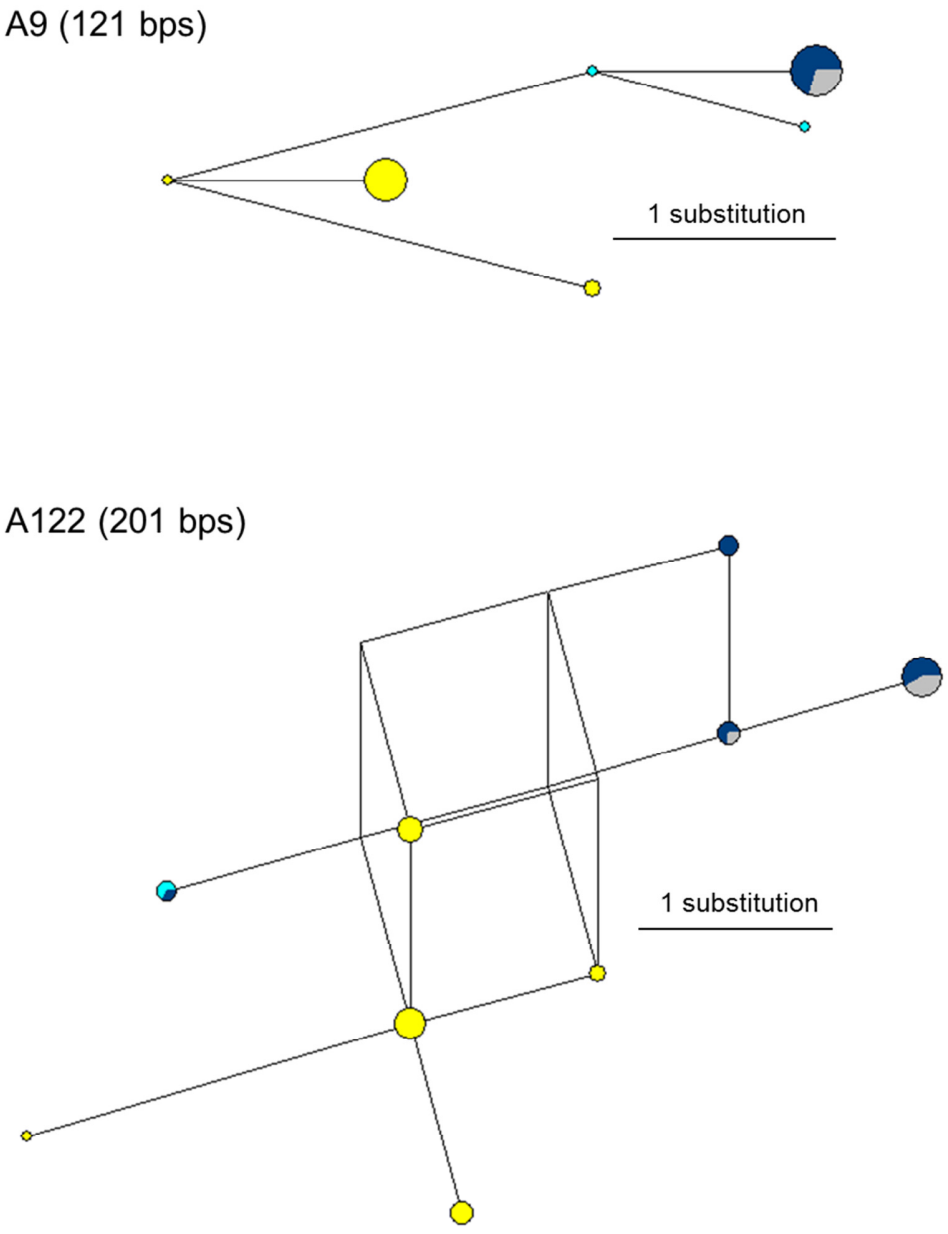


Figure 4.2 Median-joining networks based on flanking region sequences of microsatellite loci A9 and A122. Circles are coloured to represent haplotypes of *Murina gracilis* (yellow), *M. eleryi* (light blue), four individuals of *M. recondita* labeled as RA, RB, RC, and RD in Fig. 4.1 (grey), and remaining individuals of *M. recondita* (dark blue). Sizes of circles are proportional to sample sizes of unique haplotypes.

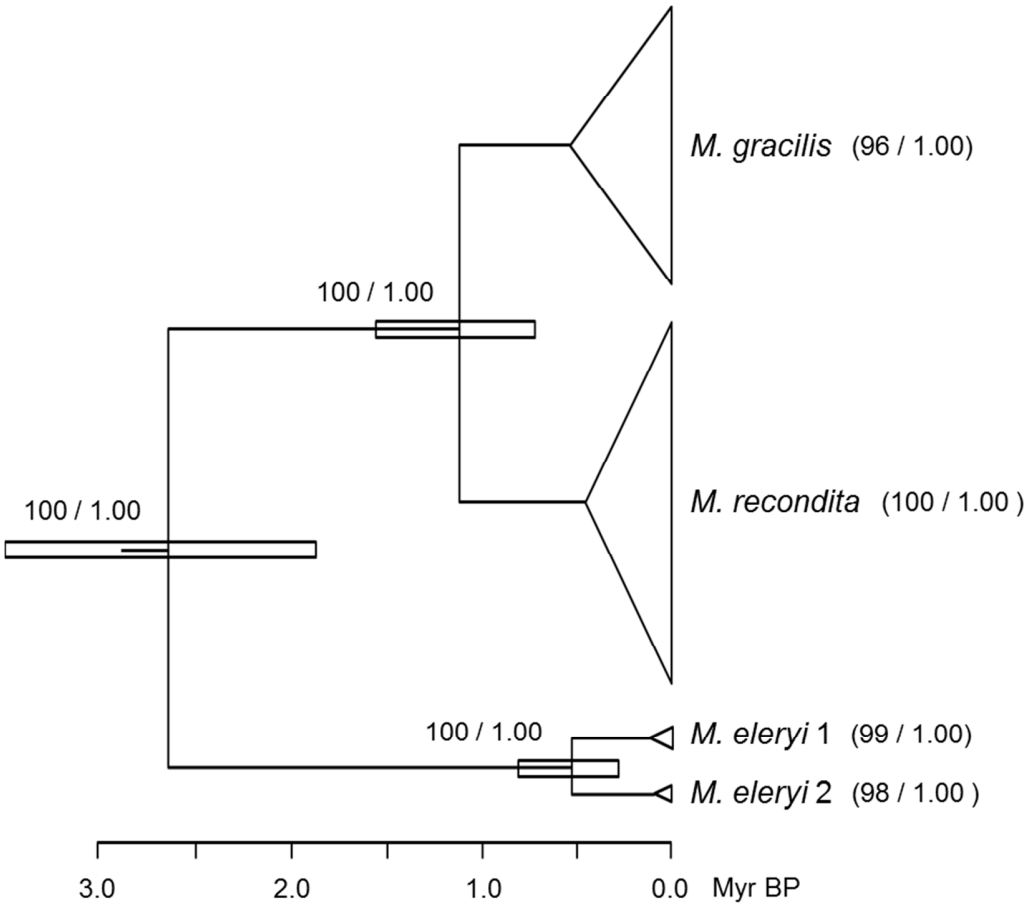


Divergence in the *M. gracilis* complex revealed by mtDNA

Incorporation of additional sequences of *M. eleryi* obtained from museum samples provided more robust estimates of phylogenetic relationships. Concatenated sequences of Cyt-*b* and COI fragments recovered reciprocal monophyletic groupings of *M. gracilis*, *M. recondita*, and *M. eleryi* based on both maximum parsimony (MP) and Bayesian trees (Fig. 4.3). These analyses also provided strong support for the sister relationship of the two Taiwanese taxa, as follows: ((*M. gracilis*, *M. recondita*), *M. eleryi*) (Fig. 4.3). Bayesian estimates of the TMRCA obtained in Chapter 3 are also summarized in Fig. 4.3. From this figure, the TMRCA between *M. eleryi* and the common ancestor of the two Taiwanese species is estimated as 2.62 (95% CI: 1.86-3.48) Myr and that between *M. gracilis* and *M. recondita* is 1.11 (95% CI: 0.72-1.55) Myr.

Two major and well-supported clades were observed within *M. eleryi*, one comprising samples from Northern Vietnam and Southern China and the other comprising samples from Central Vietnam (referred to as *M. eleryi* 1 and *M. eleryi* 2, respectively, in Fig. 4.3 and hereafter). The dating of divergence between these two clades was carried out using coalescent-based analyses implemented by BEAST. Under each of proposed demographic models, the two replicate runs gave high effective sampling sizes for all parameters (>200 for each) and consistent estimates. The combined result gave a median estimate for TMRCA between the two *M. eleryi* clades as 524 (95% CI: 291-815) Kyr before present, when the whole sample of *M. eleryi* was assumed to be from a constant-sized population (Fig. 4.3). Similar median estimates for the same quantity were obtained when other demographic models were used instead: 551 (95% CI: 307-878) and 548 (95% CI: 305-870) for the piecewise-constant and a piecewise-linear Bayesian Skyline Plot models, respectively. In addition, all these analyses also gave approximately the same TMRCA estimates for each of these two clades. Median estimates based on the constant-sized model are shown in Fig. 4.3.

Figure 4.3 A mitochondrial phylogeny for the *Murina gracilis* complex, as recovered from maximum parsimony (MP) and Bayesian (using MrBayes) analyses with outgroups removed. Tips are collapsed into triangles with widths proportional to sample sizes (see Table 4.2) and heights proportional to medians of time to the most recent common ancestor (TMRCA) for the corresponding taxa; TMRCA values were estimated in BEAST analyses conducted in Chapter 3 (for *M. gracilis* and *M. recondita*, separately) and in this chapter (for the whole sample of *M. eleryi*) assuming coalescences in constant-sized populations. Branch support values are based on bootstrap in the MP analysis (before slash) and posterior probabilities in the Bayesian analysis (after slash). Open bars present 95% confidence intervals for nodes; they are estimated in BEAST analyses in Fig. 3.4 and in this chapter.



Divergence in the *M. gracilis* complex revealed by nuclear DNA

Data

For *M. eleryi*, Sanger sequencing followed by phase analysis gave pairs of alleles for over 70% of individuals for each of the ten microsatellite flanking regions. For *M. gracilis* and *M. recondita*, 454-Pyrosequencing generated a total of 73,395 reads across all ten loci. The number of reads successfully mapped to the reference sequences ranged from 1,091 to 9,860 (mean = 4,108) and, after sorting on the basis of individual barcodes, the number of reads per individual bat per locus varied from zero to 613. For *M. gracilis* I obtained both alleles for all ten loci for all individuals with the exception of one bat. For *M. recondita*, both alleles were obtained for around 75% of individual bats (range 40% to 100%), while one allele was obtained for 0% to 20% individuals and no alleles were recovered for 40% individual bats.

Locus-specific median-joining (MJ) networks revealed that *M. recondita* ncDNA sequences were typically more related to those of *M. eleryi* than to those of *M. gracilis* (Appendix 4.4). For three out of ten loci, *M. recondita* shared haplotypes with the two *M. eleryi* clades but not with *M. gracilis*, and in cases where *M. recondita* shared haplotypes with *M. gracilis*, these haplotypes were also shared by the two *M. eleryi* clades. These results suggested a closer relationship between *M. recondita* and the two *M. eleryi* clades than between *M. gracilis* and either of these former species. It was also noteworthy that, at most loci, *M. eleryi* 1 and *M. eleryi* 2 had contrasting haplotype composition, indicative of divergence.

Four-gamete tests suggested potential intra-locus recombination for loci A4, A104, A109, and B124, for which relevant sections of alignments (for A4 and A109) or individual sequences (for A104 and B124) were removed prior to demographic analyses using IMgc (visualized in Appendix 4.4). Locus A122 was also removed from IMgc analyses due to the fact it showed reticulations in the MJ network, probably caused by homoplasious mutations due to its apparently elevated mutation rate (see Materials and Methods). Among the four taxa, *M. recondita* had a lower level of genetic variability than all the others, particularly reflected by its lower θ_S and π values at most of the ten markers (see Table 4.3).

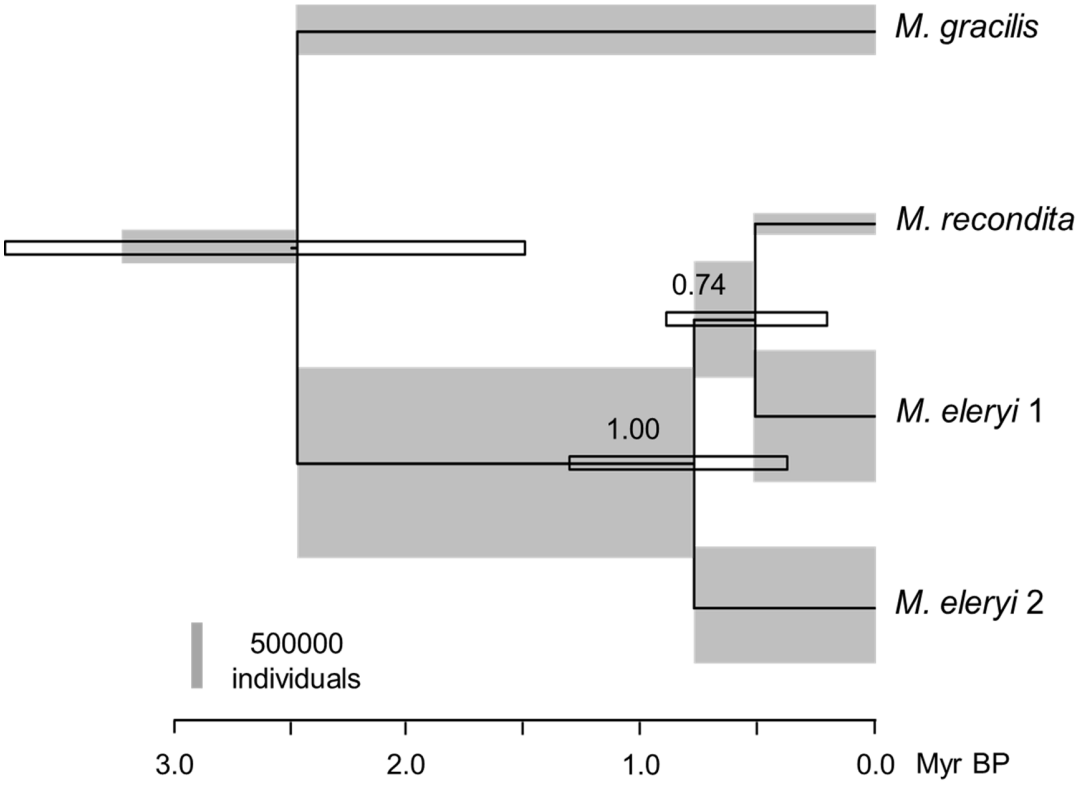
Table 4.3 Genetic variability of four taxa in the *Murina gracilis* complex based on ten microsatellite flanking regions. Values are calculated for each alignment showing no signature of intra-locus recombination except for the locus A122 (see the text for explanation). L, sequence length in number of base pairs; Nseq, sample sizes of sequences; S, number of segregating sites; θ_S, Watterson's (1975) theta; π, nucleotide diversity. Estimates of θ_S and π are presented as corresponding raw values multiplied by 100.

		M. gracilis				*M. recondita*				*M. eleryi* 1				*M. eleryi* 2			
Locus	L	Nseq	S	θ_S	π	Nseq	S	θ_S	π	Nseq	S	θ_S	π	Nseq	S	θ_S	π
A4	216	30	4	0.47	0.48	27	2	0.24	0.22	14	2	0.29	0.19	10	7	1.15	1.12
A9	102	29	3	0.75	1.22	15	0	0	0	14	3	0.93	0.97	10	2	0.69	0.55
A104	156	30	3	0.49	0.93	25	0	0	0	9	4	0.94	0.68	10	1	0.23	0.13
A109	137	30	2	0.37	0.45	30	1	0.18	0.05	14	3	0.69	0.40	6	2	0.64	0.49
A118	180	30	4	0.56	0.59	21	1	0.15	0.24	12	4	0.74	0.37	8	4	0.86	1.03
A122	184	30	5	0.69	0.63	24	7	1.02	0.86	6	10	2.38	2.86	10	7	1.35	1.88
B5	286	30	0	0	0	24	0	0	0	12	1	0.12	0.06	10	1	0.12	0.17
B114	318	30	0	0	0	23	1	0.09	0.16	12	2	0.21	0.27	8	4	0.49	0.60
B124	249	30	0	0	0	20	0	0	0	13	4	0.52	0.45	10	2	0.28	0.27
D117	234	30	4	0.43	0.42	29	0	0	0	12	5	0.71	0.86	8	5	0.82	0.93

Divergence assuming no post-split gene flow

Different coalescent models (PC and PLCR), gave consistent parameter estimates in *BEAST and, after runs were combined, ESS values were >200 for nearly all parameters (see below for exceptions). Results from these two models were almost identical with respect to posteriors of species tree topologies and estimates of divergence times between clades, although a Bayes factor of 0.0037 for the PLCR model over the PC model suggested the latter was a better fit. The maximum clade credibility (MCC) tree showed a well-supported close relationship between *M. recondita* and *M. eleryi* to the exclusion of *M. gracilis* (Fig. 4.4). Divergence between *M. gracilis* and the common ancestor of all the other taxa was dated to 2.42 (95% CI: 1.47-3.65) Myr before present. Within the clade composed of *M. recondita* and *M. eleryi*, the topology with the highest posterior probability corresponded to a grouping of *M. recondita* and *M. eleryi* clade 1 to the exclusion of *M. eleryi* clade 2. Splits within clades 1 and 2 were dated to 0.76 (0.37-1.27) and 0.50 (0.20-0.88) Myr before present, respectively. Lower posterior probabilities were recorded for alternative groupings within the *M. recondita* and *M. eleryi* clade: 0.09 for (*M. recondita*, *M. eleryi* 2) under both coalescent models; 0.17 and 0.12 for (*M. eleryi* 1, *M. eleryi* 2) under PC and PLCR, respectively. Under each model, demographic parameters for ancestral populations associated with these alternative groupings each had an ESS value at around or below 200. These low ESS values of these parameters could, at least partially, reflect low posterior probabilities of corresponding taxa tree topologies rather than indicating poor mixing of MCMC. Effective population sizes of lineages under the PC model ranged from 143,000 (*M. recondita*) to 1,475,000 (common ancestor of *M. recondita* and the two *M. eleryi* taxa), as shown in Figure 4.4.

Figure 4.4 A maximum clade credibility 'species tree' for the four taxa of the *M. gracilis* complex as estimated via a *BEAST analysis based on ten nuclear markers. A Yule model was used for splits between taxa and a piecewise-constant (PC) demographic model was used for coalescences in taxa and in their common ancestors. Numbers above branches represent posterior probabilities of corresponding clades. Branch heights present median estimates of taxa split times with 95% confidence intervals scaled by open bars. Median estimates of effective population sizes of taxa are scaled by grey blocks.



Divergence allowing post-split gene flow

For each pairwise species comparison, isolation with migration (IM) analyses based on different priors for post-split migrations gave highly consistent estimates, especially between those based on truncated uniform priors (Appendix 4.5). Hereafter I use the abbreviations 2PM5, 2PM2, and 2PME0.5 for IM models specifying the migration rate parameter (m) with, respectively, a truncated uniform prior from zero to five, a truncated uniform prior from zero to two, and one with an exponential prior with mean = 0.05. To aid comparisons among datasets, estimates from 2PM2-based analyses are given in Fig. 4.5a and marginal distributions of these demographic parameters from 2PM2 and 2PME0.5-based analyses are given in Fig. 4.6. In this latter figure, the migration rate is presented under the entry-parameter scale of the IMa2 to show truncations due to the prior.

Estimates of split times between each pair of taxa were, in most cases, insensitive to the model used. However, for *M. recondita* and *M. eleryi* 2, the 95% confidence intervals from the 2PME0.5 model analysis were narrower than those from other models because the marginal density reached zero before the set upper bound of the split time (Fig. 4.6a). Note that marginal densities for the split time also did not reach zero within the set upper bound for another two datasets, *M. gracilis* + *M. recondita* and *M. recondita* + *M. eleryi* 1, leading to questionable estimates of credible sets in these cases. From comparing across datasets, mode estimates suggested that the order of splits among taxa was consistent with the results from *BEAST; thus *M. recondita* and *M. eleryi* 1 split most recently (1.1-1.2 Myr ago), then *M. eleryi* 2 and either *M. recondita* or *M. eleryi* 1 (1.4-1.5 Myr ago), and finally the most ancient split was between *M. gracilis* and either of the two *M. eleryi* clades (1.5-1.7 Myr ago) (Fig. 4.5a; see also Appendix 4.5 for raw values). Nontheless, it is important to note that these modal estimates had overlapping marginal distributions (Fig. 4.6a). Finally, the estimated recent split of *M. gracilis* and *M. recondita* appeared to conflict with the splitting order among the four taxa inferred above, given its modal estimates as recent as 1.2-1.3 Myr ago, although these latter marginal estimates had rather flat distributions (Fig. 4.6a), suggesting there is little information contained in the dataset of *M. gracilis* and *M. recondita* for reliably inferring their split time.

Estimates of migration following divergence based on the marginal distribution were lower for the 2PME0.5 model than the other models (Fig. 4.6b), although patterns were

broadly similar in all cases. When *M. gracilis* was paired with either of the other taxa, IM modal estimates of the migration rate parameter were zero for both directions, suggesting no detectable post-split gene flow between these taxa. In contrast, migration mode estimates for the remaining pairwise combinations of taxa were non-zero for both directions. Most of the latter estimates had broad marginal distributions; with the exception of migrations (m) from *M. recondita* to *M. eleryi* 2, all had non-zero marginal densities at the upper bound under 2PM2 (Fig. 4.6b), and estimates of m from *M. eleryi* 1 to *M. recondita* had a nonzero density at the upper bound under 2PM5. Using this parameter as the test statistic, Nielsen and Wakeley's (2001) likelihood ratio tests were significant for migrations from *M. eleryi* 1 to *M. recondita* and from *M. eleryi* 2 to *M. recondita* under all models. Significant results were also obtained for migrations from *M. recondita* to *M. eleryi* 1 under 2PM2.

Marginal distributions of the population migration rate, 2NM, had modes at zero for both directions in pairwise comparisons between *M. gracilis* and other taxa with the exception of detected migration from *M. gracilis* to *M. recondita* (Appendix 4.5). Nonetheless, marginal distributions had high densities when 2NM = 0, and thus likelihood ratio tests (LRTs) following Nielsen and Wakeley's (2001) were not significant for any model comparisons. All other pairwise comparisons gave non-zero modal marginal distributions for 2NM regardless of the direction of migration or the model used. LRTs showed significant results for low levels of migration from *M. eleryi* 1 to *M. recondita* and from *M. eleryi* 2 to *M. recondita* under all models. Higher 2NM estimates were detected in the opposite direction, although LRTs were only significant for migrations from *M. recondita* to *M. eleryi* 1 under 2PM2 and 2PME0.5 and for those in either direction between *M. eleryi* 1 and *M. eleryi* 2 under 2PME0.5 (Appendix 4.5; partially shown in Fig. 4.5a).

For effective population sizes, IM applied to pairwise taxon comparisons commonly showed that *M. eleryi* 1 had a largest effective population size among the four extant taxa, followed by *M. eleryi* 2, then by *M. gracilis*, and then *M. recondita* (Appendix 4.5; partially shown in Fig. 4.6c). For each of these taxa, estimates based on different models on the same datasets showed much less variation than those from different datasets; modal estimates for *M. gracilis*, *M. recondita*, *M. eleryi* 1, and *M. eleryi* 2 were 98000-134000, 21000-40000, 402000-475000, and 278000-374000 individual bats, respectively. For each dataset, the ancestral population of corresponding extant taxa had 95% confidence intervals of its

effective population size containing zero regardless of the model used, suggesting each dataset offered limited information on the ancestral population size.

For IM models applied to all four taxa together, the summary statistics from the marginal estimates of various parameters are given in Appendix 4.6, and those based on 4PM2 are shown in Fig. 4.5b. The actual marginal distributions of these parameters obtained under 4PM2 are given in Fig. 4.7. Under the specified phylogeny, modal estimates for times of splits between *M. recondita* and *M. eleryi* 1, between the common ancestor of these two taxa and *M. eleryi* 2, and between the common ancestor of three taxa and *M. gracilis*, were 1.2-1.4, 1.8-1.9, and around 2.3 Myr, respectively. These values were larger than modal estimates for corresponding taxon pairs in the pairwise analyses. Under both 4PM2 and 4PME05 models, marginal distributions for the most recent split were not smooth unimodal curves, and those for the most ancient split were flat and removed from zero at the set upper bound (see Fig. 4.7a for curves obtained under 4PM2).

Among a total of 18 ‘m’ parameters, seven had marginal estimates with non-zero modes under 4PM2. Of these seven, four corresponded to migrations since the most recent split event: bi-directional between *M. recondita* and *M. eleryi* 1, from *M. recondita* to *M. eleryi* 2, and from *M. eleryi* 2 to *M. eleryi* 1. The three other values were for migrations that predated the most recent split event, from *M. eleryi* 2 to the common ancestor of *M. recondita* and *M. eleryi* 1, from the common ancestor of *M. recondita* and *M. eleryi* 1 to *M. gracilis*, and from the common ancestor of *M. recondita* and both *M. eleryi* taxa to *M. gracilis* (Fig. 4.7b). The former four ‘m’parameters showed less flat marginal distributions than the other three under 4PM2, and also had non-zero modal marginal distributions under 4PME0.5. Using either ‘m’ or 2NM as the test statistic, LRTs were significant for migrations from *M. recondita* to *M. eleryi* 1 and from *M. eleryi* 2 to *M. eleryi* 1 under both models apart from 4PME0.5. Non-significant results were obtained for all other LRTs (Appendix 4.6; partially visualized in Fig. 4.5b).

IM models based on all taxa gave effective population size estimates for *M. gracilis*, *M. recondita*, and *M. eleryi* 2 of 126000-132000, 27000-33000, 319000-349000, respectively, which were broadly consistent with those from two-taxon analyses. In contrast, the estimate for *M. eleryi* 1 of 322000-348000 was lower than obtained from two-taxon analyses (Appendix 4.6; partially visualized in Fig. 4.5b). As with the two-taxon models, four-taxon

analyses gave diffuse marginal distributions for effective population sizes of ancestral populations (Fig. 4.7c), suggesting little information about these parameters was available from the data.

Figure 4.5 Visualization of divergence among taxa of the *Murina gracilis* complex based on isolation with migration (IM) analyses for (a) six pairs of two taxa and (b) four-taxa. Models 2PM2 and 4PM2 (see text for explanation) were used for analyses in (a) and (b), respectively. For (b) the assumed relationship among taxa was (((R, E1), E2), G). Horizontal and vertical dimensions are scaled to represent effective population sizes and split times (present time at the top), respectively. Black boxes and black horizontal lines are scaled to represent mode estimates while grey ones including lines with double arrows are scaled to present 95% confidence intervals of corresponding variables. Black curved arrows mark directions of post-split migrations, each of which was significant using the population migration rate (2NM) based on Nielsen and Wakeley's (2001) likelihood ratio test; modal estimates of 2NM values for these migrations are also given. Extant taxa are labeled as G for *M. gracilis*, R for *M. recondita*, E1 for *M. eleryi* 1, and E2 for *M. eleryi* 2. In (b), ancestral populations are labeled as A1 for that of *M. recondita* and *M. eleryi* 1, A2 for that of *M. recondita* and both *M. eleryi* taxa, and A3 for that of all extant taxa; Successive split times T are also numbered for clarity. A mean mutation rate of 0.43 per locus per Myr and a generation time of two years was used to convert raw parameters to conventional demographic scales.

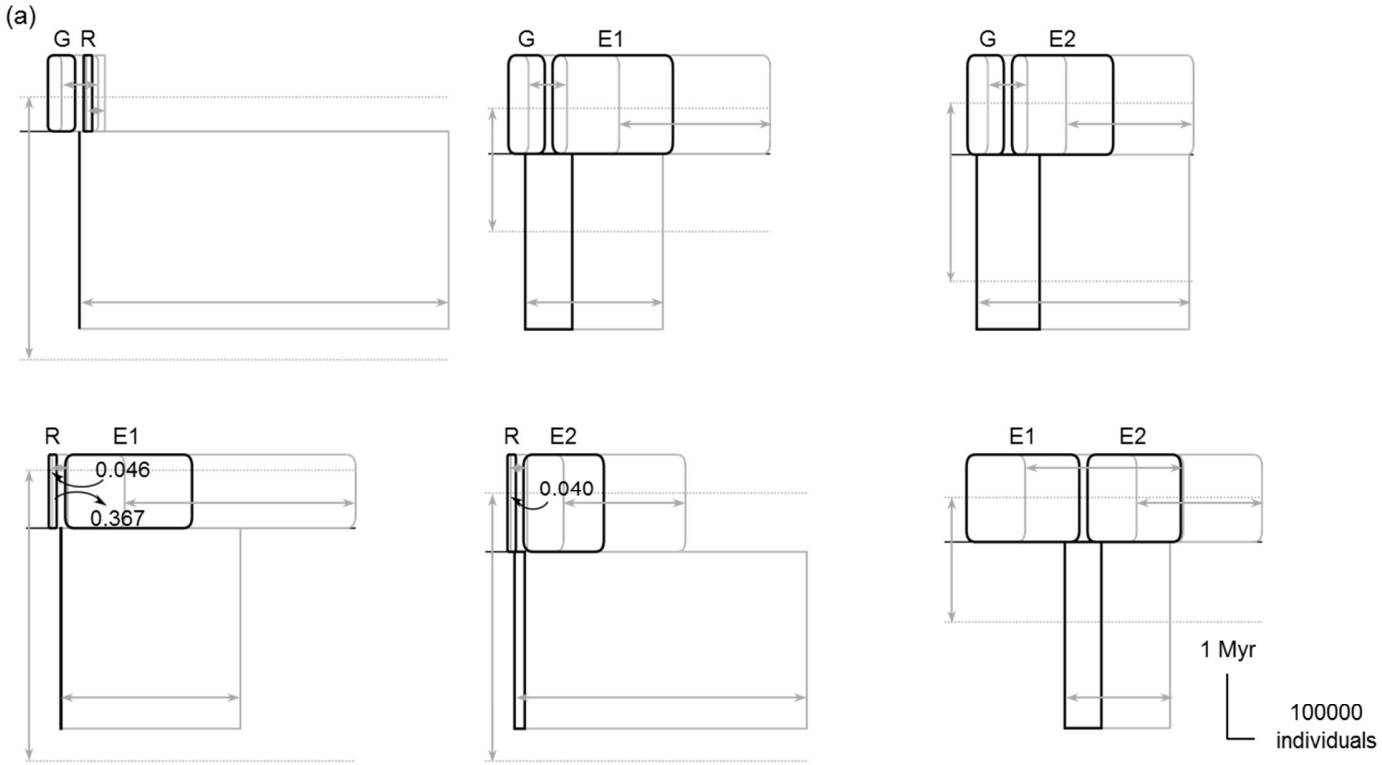


Figure 4.5 Continued.

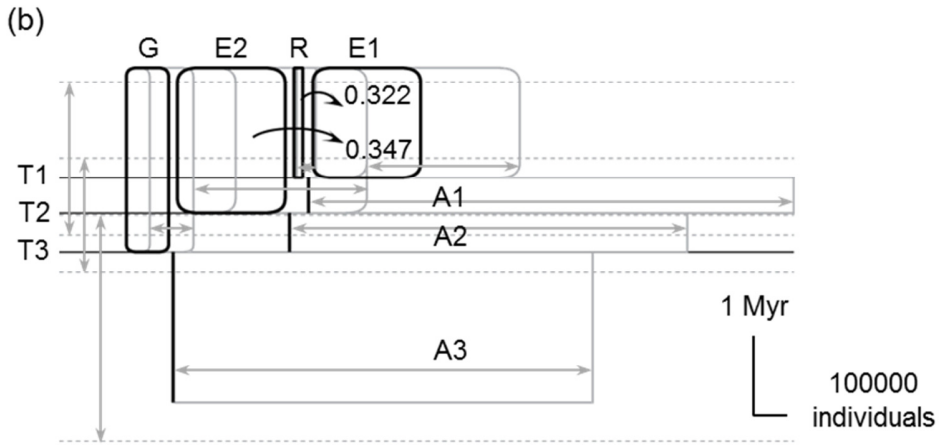


Figure 4.6 Marginal distributions of (a) split times, (b) scaled migration rates, and (c) effective population sizes from IM analyses for the six two-taxon pairs of *Murina gracilis* complex. Curves are shown in black and grey for those based on the 2PM2 and 2PME0.5 model, respectively (see the text for explanation). See Fig. 4.5 for abbreviations for the four extant taxa. In (b), asterisks mark curves showing significant results in Nielsen and Wakeley's (2001) tests under corresponding models ($P < 0.05$ in all such cases).

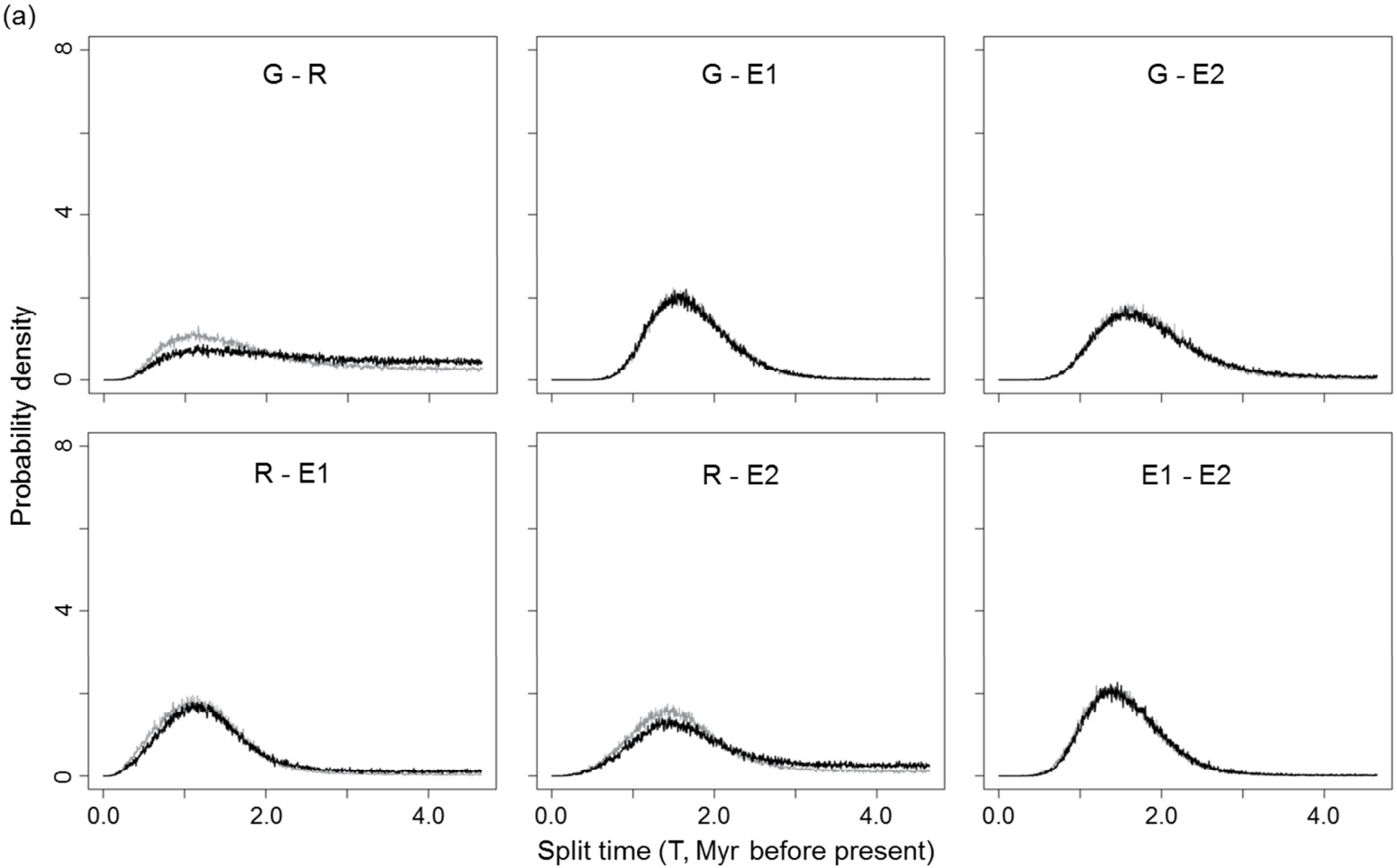


Figure 4.6 Continued.

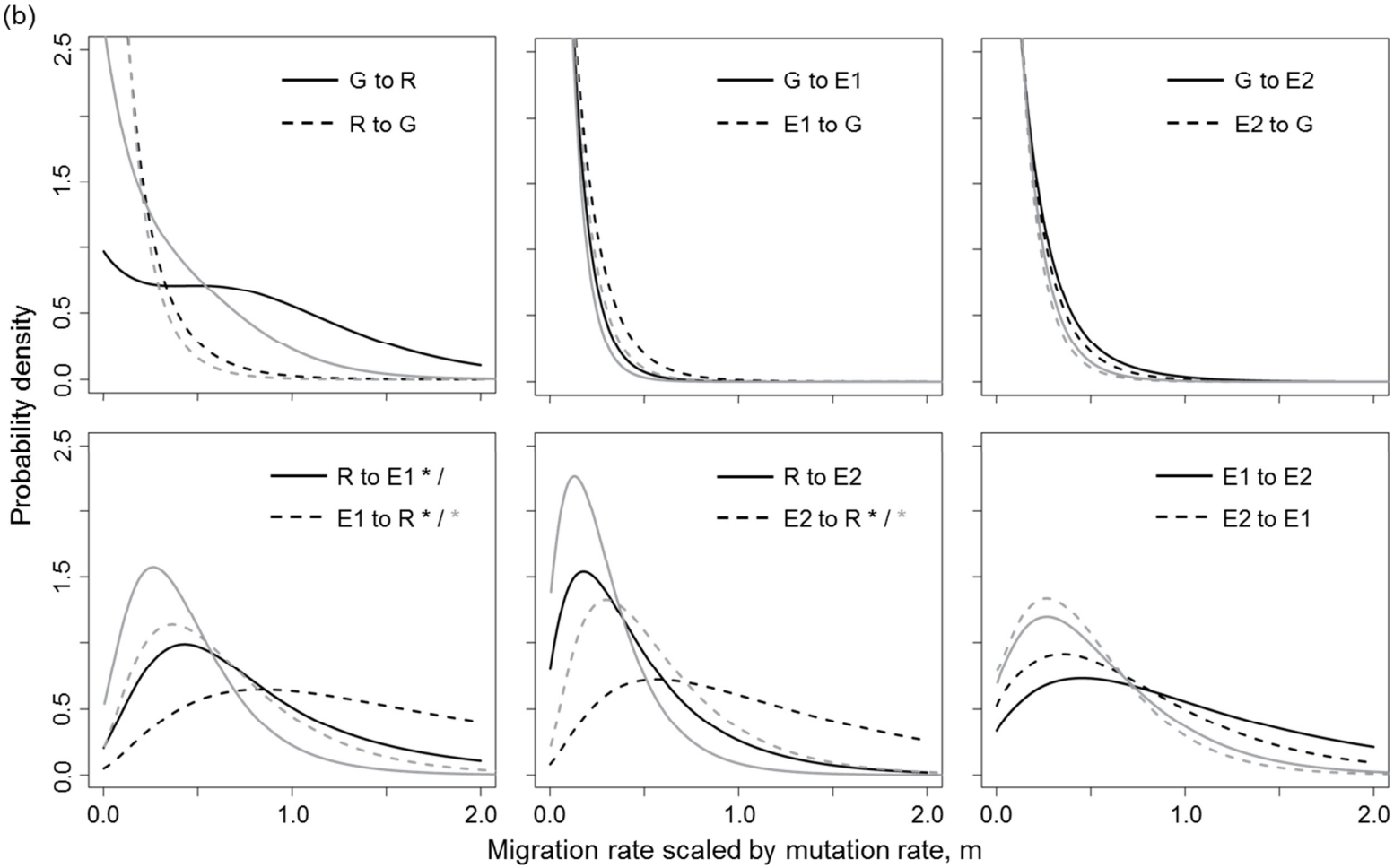


Figure 4.6 Continued.

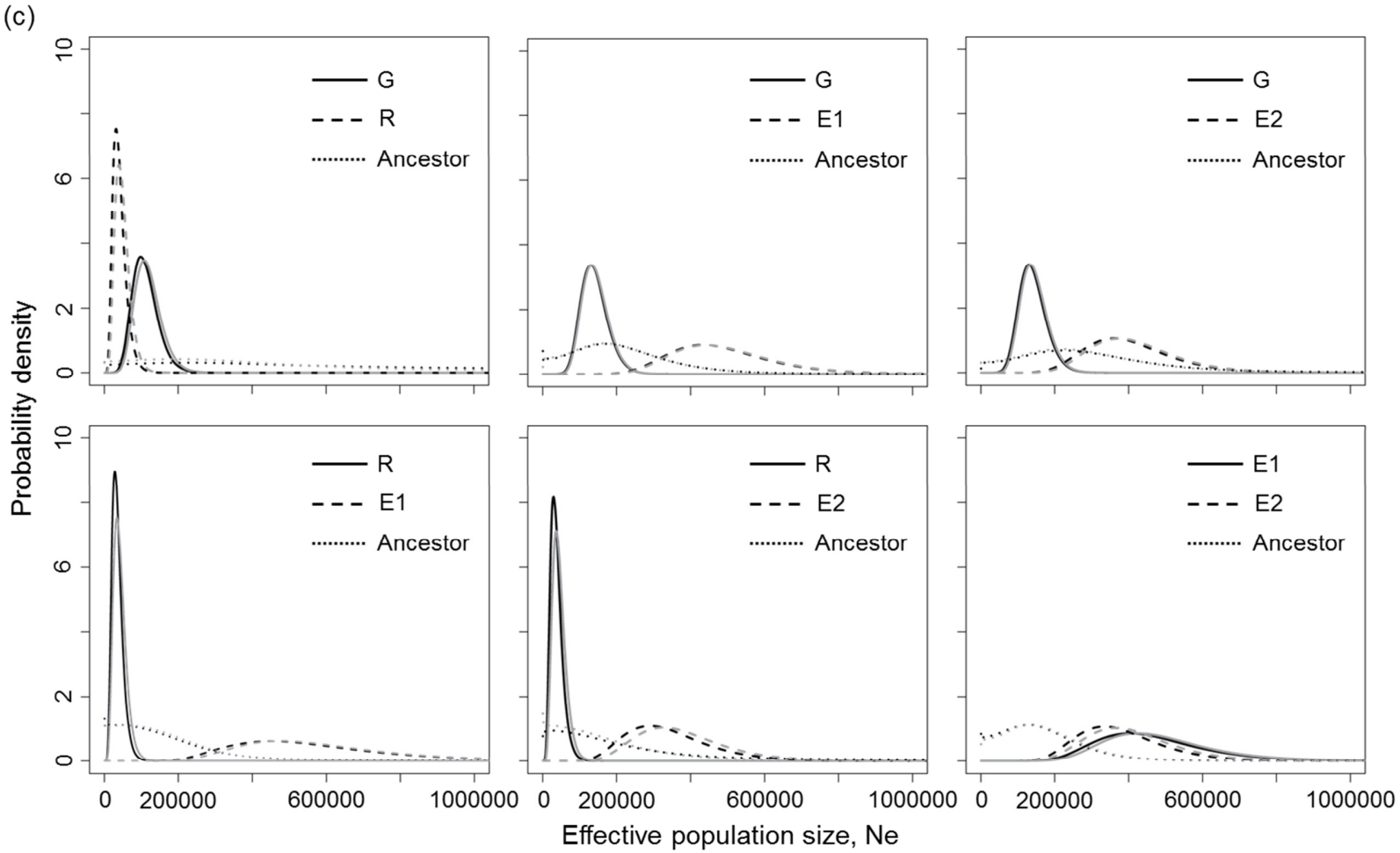
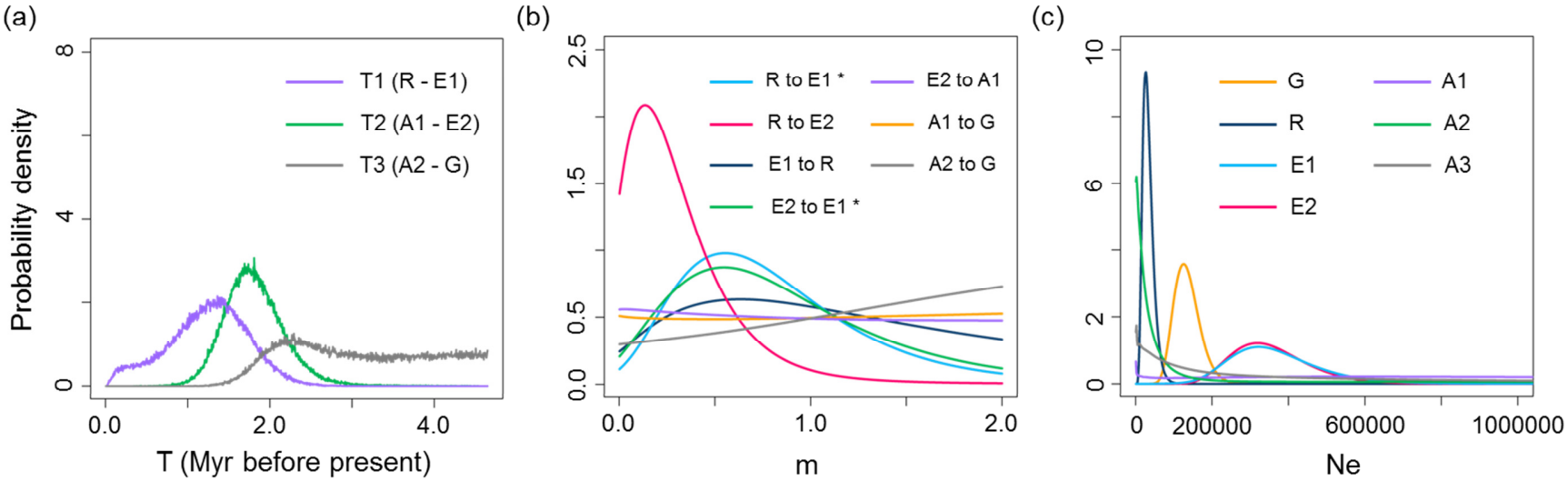


Figure 4.7 Marginal distributions of (a) split times, (b) scaled migration rates and (c) effective population sizes from the four-taxon IM analysis of the *Murina gracilis* complex. These plots are based on the 4PM2 model. See Fig. 4.5 for abbreviations of taxon names and for split times. Only plots for migration rates with nonzero peaks are presented in (b) where asterisks mark significant results obtained using Nielsen and Wakeley's (2001) tests ($P < 0.05$ in both cases).



4.4 DISCUSSION

In this study, phylogenetic reconstructions of three *Murina* species led to conflicting signals between ncDNA and mtDNA with respect to the relationship of the two putative sister species *M. gracilis* and *M. recondita.* Detailed modeling of gene flow suggested this discordance has arisen from introgressive mitochondrial hybridisation between *M. gracilis* and *M. recondita* and therefore, these are not genuine sister species as previously considered. NcDNA sequences also showed two highly divergent evolutionary lineages within the Asian mainland taxon *M. eleryi*, and the divergence process between these two clades, and also between each clade and the other Taiwanese species, are discussed. Clustering of microsatellite genotypes also revealed further conflicts between species identity and cluster assignment, however, these appear not to have resulted from recent hybridisation but instead from allele size homoplasy.

Absence of recent introgressive hybridisation between *M. gracilis* and *M. recondita*

Multi-locus Bayesian clustering analysis of *M. gracilis* and *M. recondita* genotypes using Pritchard et al.'s (2000) admixture model revealed that some individuals were characterised by mixed ancestry from across the species boundary. Such signatures of admixture in STRUCTURE plots are frequently considered to be due to gene flow. Particularly, these signatures are expected to decay rapidly when unlinked loci are used because linkage disequilibrium (LD) resulting from the admixture of populations can last only shortly (Pritchard & Przeworski 2001; Verardi *et al.* 2006) and, therefore, are thought to reflect recent introgression. In my study, from a dataset of 250 bats of two species, each typed at 14 nuclear microsatellite loci (13 autosomal and one X-linked), all but four individuals were seen to cluster with their conspecifics in clearly defined groups (Fig. 4.1). This overwhelming unambiguous and correct assignment to species is consistent with observed maintenance of morphological differences (pelage colour) between *M. gracilis* and *M. recondita*, even at sites where they are sympatric (sites 10, 19, 25, and 54).

Four individual bats of *M. recondita* appeared to possess measurable fractions of the genetic composition of its sister taxon. Surprisingly these animals were sampled from two

demes that were observed to not overlap with the distribution of *M. gracilis* based on field observations and also based on Maxent distribution models of suitable habitat for *M. gracilis* (cf. Fig. 3.12a and d in Chapter 3). These apparently contradictory findings were explored by sequencing of the microsatellite A9 and A122 flanking regions. In addition to the rationale explained above for the expectation that admixture signatures in STRUCTURE plots would only be discovered for recent introgression, the following reason supported the same expectation from another aspect. Rates of stepwise mutation of microsatellite repeat lengths are typically more rapid than those of flanking regions (Wan *et al.* 2004) and thus mutations in the latter would not have occurred since introgression. Consequently, if shared alleles are the result of introgression, then adjacent flanker sequences are expected to be identical or near identical to those of their source species. However, in my study I found discordance between these allelic flanking sequences and number of repeat units, so was able to reject hybridisation as a cause of admixture. Instead admixture is better explained by homoplasious mutations leading to identical allele sizes in some populations of the two species. Similar allele size homoplasy been reported between closely related taxa, including fishes (van Oppen *et al.* 2000), primates (Garza & Freimer 1996) and carnivores (Domingo-Roura *et al.* 2005). Furthermore, such homoplasy can occur within (Anmarkrud *et al.* 2008) and among populations of the same species (Rossiter *et al.* 2007). Rossiter *et al.* (2007) showed that allele size homoplasy can lead to underestimation of population genetic differentiation among distant demes. My findings of homoplasy highlights potential caveats of inferring gene flow from clustering methods such as those developed by Pritchard et al. (2000), where admixture signatures are commonly interpreted as gene flow rather than the potential independent evolution of alleles of the same size.

Evidence of ancient introgression from discordant mitochondrial and nuclear gene genealogies

Further conflicts were observed between sequence datasets of mtDNA and ncDNA based on range-wide samples of the two Taiwanese species, *M. gracilis* and *M. recondita*, and samples of the continental taxon *M. eleryi*. Phylogenetic reconstruction from mitochondrial genes suggested reciprocal monophyly of each of these species, and a sister relationship between *M. gracilis* and *M. recondita* (Fig. 4.3). In contrast, ten gene

genealogies reconstructed for anonymous nuclear fragments together suggested a closer relationship between *M. recondita* and *M. eleryi* to the exclusion of *M. gracilis* (Appendix 4.4). This result was also supported by the results of *BEAST analyses (Fig. 4.4) that simultaneously utilized information from the ten nuclear markers while also taking into account any uncertainty introduced by lineage sorting. Interestingly, *BEAST analyses also revealed that *M. recondita* was a sister taxon of one of two lineages of *M. eleryi* although this was only modestly supported in these analyses (Fig. 4.4).

Although phylogenetic relatedness among taxa is commonly considered to reflect the ordering of their divergence times, a similar level of divergence could arise from different configurations of split time and gene flow. For example, a recent split with little subsequent gene flow might lead to similar levels of divergence as would a more ancient split followed by subsequent genetic mixing (Wakeley 1996). To tease apart these possibilities, analyses based on isolation with migration (IM) models (Hey 2010b; Hey & Nielsen 2004) were performed using the ten nuclear markers. Pairwise IM analyses of the four taxa generally portrayed a splitting order consistent with that suggested by *BEAST in which the first split was between *M. gracilis* and the common ancestor of the others (referred to as T3 in following discussion), then between *M. eleryi* 2 and the common ancestor of *M. recondita* and *M. eleryi* 1 (T2), and most recently between *M. recondita* and *M. eleryi* 1 (T1) (Fig. 4.5a; Appendix 4.5).

However, other aspects of my findings cast doubt on this splitting order. First, the model estimates for T1, T2, and T3 obtained from pairwise IM models were similar to each other (1.1-1.2, 1.4-1.5, and 1.5-1.7 Myr ago, respectively) with overlapping confidence intervals, and variation for each appeared to mainly depend on the pair of taxa under consideration. Differences between these estimates and those obtained from *BEAST analyses (median estimates for T1, T2, and T3 as 0.5, 0.8, and 2.4 Myr ago, respectively) might be due to the inappropriate forcing of no post-split gene flow in the latter analyses (especially for T1 and T2), which would have led to shorter estimated times since divergence than those obtained from the IM analyses in which post-split gene flow is allowed. Indeed IM analyses suggested some bidirectional post-split gene flow between each pairwise combination of the taxa *M. recondita*, *M. eleryi* 1 and *M. eleryi* 2, although statistical support varied depending on prior settings (Fig. 4.5a and 4.6b; Appendix 4.5). The results of IM

analyses also suggested a split time of 1.2-1.3 Myr ago for *M. gracilis* and *M. recondita*, so conflicting with the split order described above. This is, nevertheless, less problematic since marginal distributions for this split time are flat, suggesting little information in the data for an accurate estimation of this quantity.

Parameter estimates obtained from IM models of all four taxa, in which the splitting order was fixed according to *BEAST and pairwise IM analyses, were inconsistent with earlier analyses, giving T1, T2, and T3 values of 1.2-1.4, 1.8-1.9, and 2.3 Myr before present, respectively (Fig. 4.5b and 4.7a; Appendix 4.6). Given the greater parameterization of the four-taxon analyses, it is difficult to ascertain the reason(s) for such discrepanices, although it seems likely that the currently available data did not provide enough information for reliably estimating some of demographic parameters, especially those associated with more ancient events based on flat marginal distributions for all ancestral populations, ancient migrations and T3 (Fig. 4.7). Despite these limitations, the four-taxon IM models did give generally similar estimates of effective population sizes for the four extant taxa as those from pairwise analyses. More importantly, they also suggested the occurrence of uni- or bidirectional migration among *M. recondita* and the two *M. eleryi* taxa since T1 (Fig. 4.5b and 4.7b; Appendix 4.6).

To summarize, evidence from nuclear DNA suggests that the most likely divergence history among the four *M. gracilis* complex taxa is (((*M. recondita*, *M. eleryi* 1), *M. eleryi* 2), *M. gracilis*) although times of these divergence events can rarely be estimated with certainty due to limited information of the data. This history conflicts with that revealed by mitochondrial DNA, showed as ((*M. recondita*, *M. gracilis*), (*M.eleryi* 1, *M. eleryi* 2)). In addition, IM analyses based on nuclear DNA suggest likely post-split gene flow among *M. recondita* and the two *M. eleryi* taxa.

Divergence among *M. gracilis* complex taxa

In light of the split order suggested by multi-locus ncDNA markers, the closer relationship between *M. gracilis* and *M. recondita* based on mtDNA points to introgressive hybridisation in which the mitochondrial genome of one of these taxa has been replaced by that of the other one. Indeed the mitochondrial genealogy reveals a reciprocally

monophyletic relationship between *M. gracilis* and *M. recondita* sampled from across their respective ranges (Fig. 4.3) implying that the hybridisation represents an ancient process. By superimposing the mitochondrial genealogy for the four focal taxa on the hypothesized splitting order depicted by ncDNA, an inference of introgression from *M. gracilis* to *M. recondita* could be made; otherwise (i.e. if introgression occurred from *M. recondita* to *M. gracilis*) the mtDNA clade comprising *M. gracilis* and *M. recondita* should be nested in that of *M. eleryi*. However, my results also suggest possible mitochondrial replacement has occurred between the two *M. eleryi* clades (see below), which has obscured the above inference.

The hypothetical divergence history among *M. gracilis*, *M. recondita*, and *M. eleryi* described above is broadly consistent with expectations based on coalescent simulations conducted by Currat *et al.* (2008). First, the direction of asymmetric introgression appears to have been from the resident species (i.e. the earlier colonizer) to the invading one (i.e. the later colonizer) (see following paragraph). This condition has been documented for a variety of other taxa (reviewed in Currat *et al.* 2008). Second, Currat *et al.* (2008) showed that a pre-existing low level of gene flow among demes of conspecifics provides favorable conditions for introgression into the invading species, for the reason that such newly introgressed genes at the front wave of hybridisation will be less susceptible to being diluted by those arriving from other demes. Moreover, mtDNA is expected to show greater population genetic structure than ncDNA even when migration rates among demes do not vary between different sexes, due to the fact they are haploid and maternally inherited (Birky 2001). Therefore, it can be predicted that the level of introgression into the invading species at a nuclear gene will not exceed that at a mitochondrial gene in the same direction. The case between *M. gracilis* and *M. recondita* fits such a prediction, given that both were found to be characterised by highly reduced mitochondrial gene flow among their corresponding demes (see Chapter 3). Similar situations have been reported for other animal taxa pairs, such as between other bats (Berthier *et al.* 2006), between toads (Sequeira *et al.* 2011) and between hares (Melo-Ferreira *et al.* 2012). Extending this rationale, Du *et al.* (2011) speculated that lower introgression of chloroplast (cp) DNA than of mtDNA between two spruce species can be attributed to the fact that, in gymnosperms, mtDNA gene flow can result from pollen dispersal, whereas cpDNA gene flow can only occur via seed dispersal.

Taking all the evidence into account, I propose a divergence scenario for *M. gracilis* and *M. recondita* as follows. First, the ancestor of *M. gracilis* colonized the area now known as Taiwan from continental Asia via a land bridge during a period of glaciation. With sea-level rises associated with climate warming, the Taiwan Straits later formed, leading to isolation of the colonized population and its subsequent divergence from the common ancestor of *M. recondita* and *M. eleryi*. These proposed colonization and split events are strengthened by our knowledge of the ecomorphological adaptations of *Murina* bats, including their broad wings that are suitable for slow maneuverable flight in vegetation but which are inefficient for long movements, and also their specialized high frequency echolocation that will also constrain flight over wide stretches of unsuitable habitats, including water bodies such as the Taiwan Straits (Kingston *et al.* 1999). Based on mitochondrial genes, this split event occurred at 1.9-3.5 (median: 2.6) Myr before present (Fig. 4.3: TMRCA between *M. eleryi* and the clade made of *M. gracilis* and *M. recondita*) at the boundary between the Pliocene and Pleistocene (Gibbard *et al.* 2010). Divergence times were obtained from the mtDNA tree because methods based on ncDNA did not provide consistent estimates of dates, as described earlier, presumably due to uncertain levels of post-split nuclear gene flow. Taking into account the possible mitochondrial replacement between the ancestor of *M. gracilis* and that of the other taxa, this split event could be even earlier.

The ancestor of *M. recondita* appears to have colonized Taiwan later, undergoing divergence from the ancestral *M. eleryi*. During its colonization, *M. recondita* had its mitochondrial genome completely replaced by that of *M. gracilis* through introgressive hybridisation, which may have ceased at 0.7-1.6 (median: 1.1) Myr before (Fig. 4.3: TMRCA between *M. gracilis* and *M. recondita*). Interestingly, higher levels of post-split nuclear gene flow were inferred to have occurred between *M. recondita* and *M. eleryi*, and also between the two *M. eleryi* lineages, than between *M. gracilis* and *M. recondita* (Fig. 4.5, 4.6b and 4.7b; Appendix 4.5 and 4.6). For this reason, I am unable to rule out the possibility that mitochondrial introgression might have also occurred between *M. recondita* and *M. eleryi*, and between *M. eleryi* 1 and 2. Indeed in the latter case, mitochondrial replacement is suggested by the observation that IM estimates for the split time between *M. eleryi* 1 and *M. eleryi* 2 exceed those for their mitochondrial TMRCA values (medians: 0.5-0.6; upper confidence limits: 0.8-0.9 Myr; Fig. 4.3) regardless of whether these come from pairwise IM models (modes: 1.4-1.5; lower confidence limits: 0.6-0.7 Myr) or from the four-taxon IM

analyses (modes: 1.8-1.9; lower confidence limits: 1.1 Myr). The unidirectional nuclear gene flow from *M. eleryi* 2 to *M. eleryi* 1, which was revealed in the four-taxon analyses, hints at replacement of the mitochondrial genome of *M. eleryi* 1 by *M. eleryi* 2. More information on the demographic histories of these taxa would be helpful before accepting such a scenario, for the reason that simulations suggest it would be more likely had *M. eleryi* 1 undergone population growth (see Currat *et al.*'s (2008)). In spite of the post-split gene flow, I found no evidence of demographic expansion from *BEAST analyses; model comparisons using Bayes factor revealed no greater support for a model that allowed linear demographic changes for each of the four focal taxa than one that specified constant population sizes.

Note that unlike the situation between *M. gracilis* and *M. recondita* where mitochondrial replacement occurred over short geographic distance (e.g. less than half the width of Taiwan, say 70 km), any replacement between *M. eleryi* 1 and *M. eleryi* 2 would have involved much larger geographic distances of up to 900 km, regardless of the direction of introgression. Currat *et al.*'s (2008) simulations show that levels of introgression in the invading species drop in demes away from the front wave of hybridisation. This implies that mitochondrial introgression in my focal bats is very efficient, presumably facilitated by the extremely reduced mitochondrial gene flow among demes of conspecifics, and also raises the possibility of mitochondrial introgression between *M. recondita* and *M. eleryi* 1, especially since that IM analyses indicate nuclear gene flow between these taxa after their initial divergence. Further sampling of *M. eleryi* throughout its range, especially from areas close to Taiwan, would be helpful for resolving this possibility.

Appendix 4.1 Details of 12 voucher specimens of *Murina eleryi* (Royal Ontario Museum) sampled for genetic analysis. GPS coordinates and the altitude are given in decimal degrees and metres above sea-level, respectively.

Catalog No.	Country	Province	Locality	GPS-E	GPS-N	Altitude
111286	Vietnam	Quang Nam	Ngoc Linh Base Camp, 10 km SW Nuoc Xa	15.200	108.033	830
111300	Vietnam	Quang Nam	Ngoc Linh Base Camp, 10 km SW Nuoc Xa	15.200	108.033	830
111308	Vietnam	Quang Nam	Ngoc Linh Base Camp, 10 km SW Nuoc Xa	15.200	108.033	830
111360	Vietnam	Quang Nam	Noc Ong Toan, Tran Don	15.233	108.033	?
111399	Vietnam	Quang Nam	8 km ENE Nuoc Xa	?[a]	?[a]	200
116071	China	Guangxi	Jingxi County Provincial Nature Reserve	23.117	105.967	978
116099	China	Guangxi	Jingxi County Provincial Nature Reserve	23.117	105.967	978
116124	China	Guangxi	Jingxi County Provincial Nature Reserve	23.117	105.967	978
116182	China	Guangxi	Jingxi County Provincial Nature Reserve	23.117	105.967	978
116190	China	Guangxi	Jingxi County Provincial Nature Reserve	23.117	105.967	978
116199	China	Guangxi	Jingxi County Provincial Nature Reserve	23.117	105.967	978
116200	China	Guangxi	Jingxi County Provincial Nature Reserve	23.117	105.967	978

[a] Provided GPS coordinates (1.867ºE, 108.150ºN) are questionable with reference to the recorded locality.

Appendix 4.2 Primer pairs for amplification of flanking regions of ten microsatellite loci.

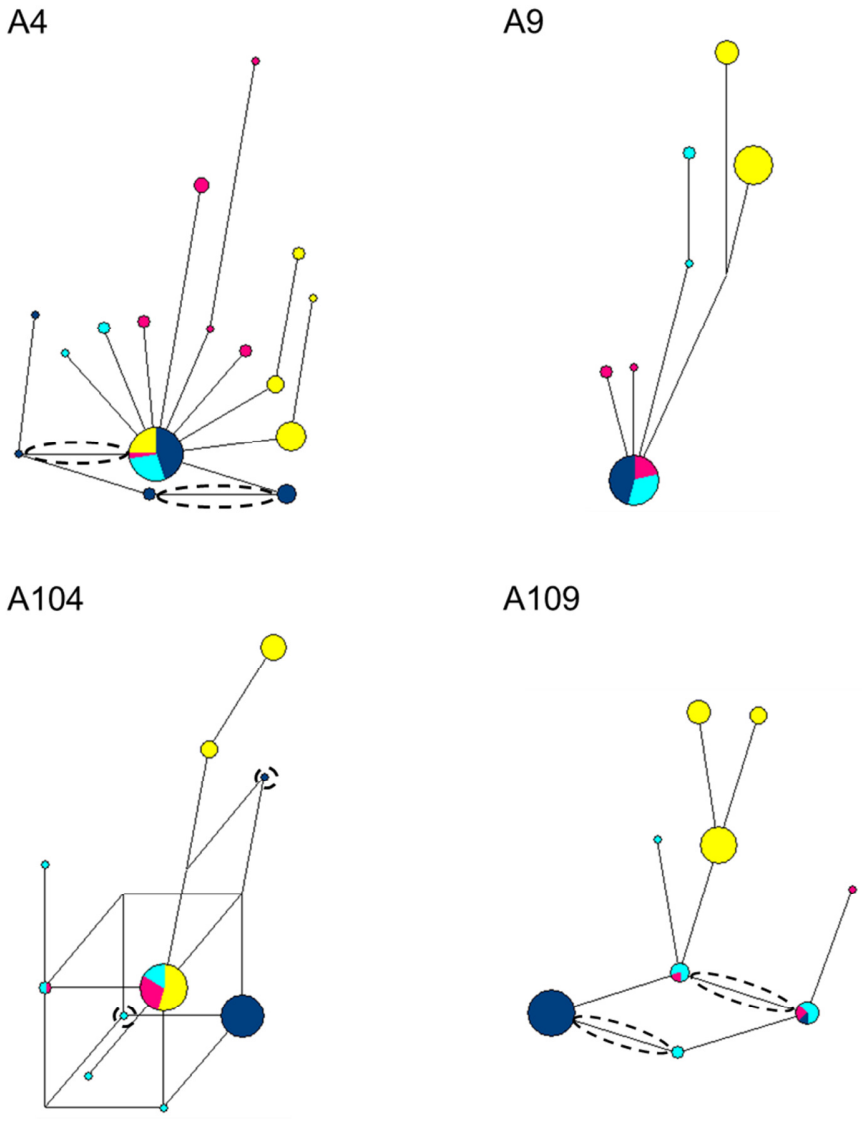
Locus	Pirmer pair	Target length[a] (bps)
A4	F: TCCACTAGCCACATCTCTGT	225
	R: TTTTCAGTAACCACCAGAGG	
A9	F: TTAGGGGAGTCTGAAAAAGG	106
	R: TCCCAGATCCACTTTACAGG	
A104	F: CTCTCAGACCTGGCTTGC	197
	R: GGGCCTAATGCCAAAGTT	
A109	F: TTGAACCTGGGCTTAATCTT	140
	R: CGGTGGACCTTGTTTGTT	
A118	F: TCTCAGAGTCAGCCGGAAT	181
	R: ACATTCTGGATTTGCAACAA	
A122	F: TTCTAAATTCCATCCAACTCC	194
	R: CTCACTAGGCACAGAGTCCA	
B5	F: TCAGTCTAGCGCAGTTCTCA	286
	R: TCCACAGTGGTGTCCCTTA	
B114	F: GGCCTCAGGATTGGATAAG	321
	R: GGAATCGAACCAGTGACCT	
B124	F: TCATTAGCAAAACTCCGACTC	251
	R: TTCCTAATCTGATCACCCATT	
D117	F: GCATCCTTATGGGCATGAT	238
	R: GAAAAGGGGCTGCTGTAAC	

[a] Priming sites removed.

Appendix 4.3 Estimated proportions of 'foreign' genetic composition for four individual bats of *Murina recondita* as inferred from STRUCTURE analyses. Labels follow Fig. 4.1. Values are shown for the analysis based on all 14 microsatellite loci and also for a reduced dataset of 13 loci when each locus was in turn removed.

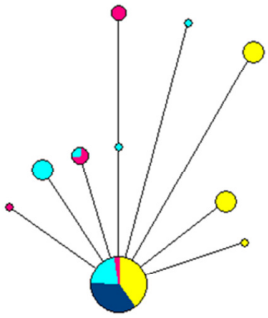
		Locus removed												
Individual	14 loci	A4	A9	A10	A104	A109	A118	A122	B5	B9	B114	B121	B124	D110	D117
RA	0.37	0.23	0.06	0.70	0.48	0.29	0.27	0.35	0.41	0.48	0.34	0.53	0.43	0.42	0.30
RB	0.11	0.07	0.08	0.11	0.07	0.09	0.17	0.05	0.09	0.13	0.08	0.12	0.04	0.13	0.17
RC	0.11	0.13	0.20	0.03	0.11	0.07	0.21	0.04	0.07	0.05	0.08	0.11	0.04	0.12	0.09
RD	0.27	0.24	0.09	0.24	0.27	0.23	0.26	0.32	0.26	0.22	0.16	0.25	0.16	0.19	0.19

Appendix 4.4 Median-joining networks built for the *Murina gracilis* complex based on ten microsatellite flanking regions. Unique haplotypes are colour-coded for *M. gracilis* (yellow), *M. recondita* (dark blue), *M. eleryi* 1 (light blue), and *M. eleryi* 2 (pink). Dashed ellipses indicate haplotypes (for loci A104 and B124) and branches (for loci A4 and A109) that were removed by the IMgc software in order to avoid violations of the four gamete test and thus ensure suitability for downstream analyses (see the text for explanations). The shortest branch in each plot represents one nucleotide substitution.

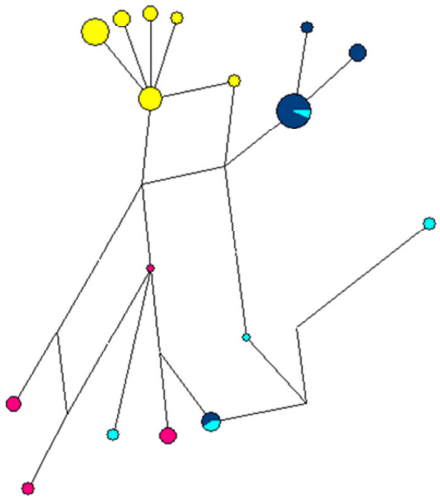


Appendix 4.4 Continued.

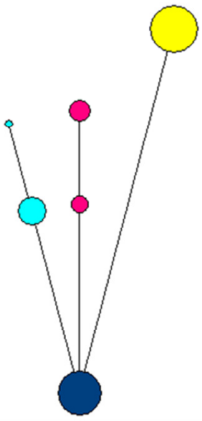
A118



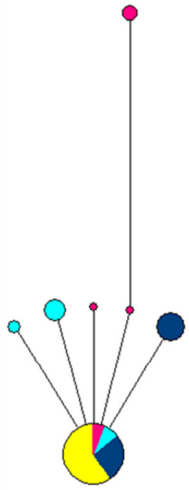
A122



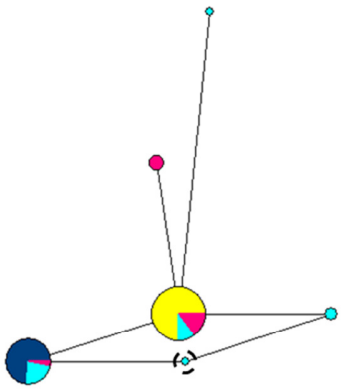
B5



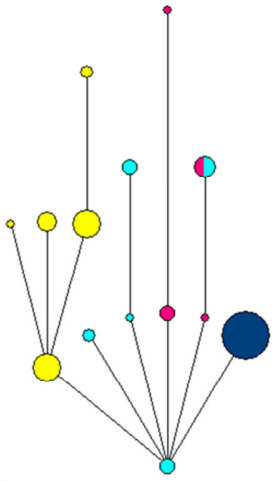
B114



B124



D117



Appendix 4.5 Isolation with migration (IM) estimates for six pairs of two taxa from the *Murina gracilis* complex. Values are modes with 95% confidence intervals given in the row directly below. See Fig. 4.5 for abbreviations of extant taxa ('A' refers to the ancestral population for each taxon pair) and see the text for abbreviations of models. Parameters include the split time (T, Myr), population migration rates (2NM, per generation), and effective population sizes (N, thousand individuals). Asterisks denote significant results from Nielsen and Wakeley's (2001) tests based on 2NM (*, $P < 0.05$; **, $P < 0.01$).

		Model		
Taxon pair	Parameters	2PM5	2PM2	2PME0.5
G - R	T	1.300	1.156	1.160
		0.649-4.644	0.640-4.649	0.444-4.644
	$2NM_{G\rightarrow R}$	0.036	0.035	0
		0-0.106	0-0.098	0-0.072
	$2NM_{R\rightarrow G}$	0	0	0
		0-0.084	0-0.085	0-0.069
	N_G	98	98	108
		47-180	47-180	55-191
	N_R	31	33	40
		11-78	11-79	14-89
	N_A	230	0	191
		0-1341	0-1334	0-1233
G - E1	T	1.607	1.565	1.495
		0.840-2.816	0.840-2.793	0.867-2.733
	$2NM_{G\rightarrow E1}$	0	0	0
		0-0.238	0-0.236	0-0.196
	$2NM_{E1\rightarrow G}$	0	0	0
		0-0.096	0-0.095	0-0.074
	N_G	130	132	133
		75-214	73-213	76-216
	N_{E1}	433	434	441
		242-786	241-786	248-793
	N_A	168	169	166
		0-501	0-497	0-476
G - E2	T	1.681	1.556	1.612
		0.742-3.626	0.747-3.514	0.788-3.142
	$2NM_{G\rightarrow E2}$	0	0	0
		0-0.349	0-0.340	0-0.251
	$2NM_{E2\rightarrow G}$	0	0	0
		0-0.098	0-0.097	0-0.075
	N_G	130	130	134
		72-214	72-214	76-217
	N_{E2}	360	361	374
		194-650	194-649	207-664
	N_A	226	225	219
		0-774	0-761	0-649

Appendix 4.5 Continued.

Taxon pair	Parameters	Model		
		2PM5	2PM2	2PME0.5
R - E1	T	1.151	1.105	1.077
		0.277-4.649	0.226-4.644	0.212-2.560
	$2NM_{R\rightarrow E1}$	0.356	0.367*	0.229*
		0-2.268	0-1.607	0-1.111
	$2NM_{E1\rightarrow R}$	0.057**	0.046**	0.026**
		0.006-0.167	0.006-0.122	0-0.096
	N_R	21	28	34
		7-60	11-68	12-78
	N_{E1}	440	451	475
		206-1046	211-1028	232-1059
	N_A	2	0	2
		0-933	0-636	0-405
R - E2	T	1.477	1.453	1.402
		0.658-4.649	0.588-4.644	0.398-4.649
	$2NM_{R\rightarrow E2}$	0.1	0.114	0.085
		0-0.615	0-0.617	0-0.461
	$2NM_{E2\rightarrow R}$	0.044**	0.040**	0.023**
		0.004-0.135	0.004-0.111	0.001-0.086
	N_R	25	30	37
		8-69	11-72	14-84
	N_{E2}	278	291	326
		140-578	145-588	166-630
	N_A	5	36	0
		0-1149	0-1059	0-687
E1 - E2	T	1.481	1.374	1.398
		0.621-2.965	0.681-2.644	0.663-2.495
	$2NM_{E1\rightarrow E2}$	0.394	0.400	0.214*
		0-1.349	0-1.067	0-0.842
	$2NM_{E2\rightarrow E1}$	0.2805	0.3373	0.239*
		0-1.273	0-1.137	0-0.881
	N_{E1}	402	403	424
		195-793	209-777	227-803
	N_{E2}	318	335	366
		148-616	178-626	194-660
	N_A	127	130	133
		0-418	0-377	0-372

Appendix 4.6 Isolation with migration (IM) estimates based on four-taxa from the *Murina gracilis* complex. Values are modes with corresponding 95% confidence intervals in row directly below. See Fig. 4.5 for the assumed phylogeny of the four extant taxa as well as abbreviations of names and split times, and see text for model abbreviations. See Appendix 4.6 for units of different demographic parameters. Asterisks mark significant results from Nielsen and Wakeley's (2001) tests based on 2NM (*, $P < 0.05$).

Parameters	4PM2	4PME0.5
T1	1.360	1.221
	0.174-2.086	0.142-2.026
T2	1.807	1.863
	1.119-2.542	1.105-2.574
T3	2.291	2.300
	1.83-4.649	1.728-4.649
$2NM_{G\rightarrow R}$	0	0
	0-0.392	0-0.031
$2NM_{R\rightarrow G}$	0	0
	0-0.075	0-0.063
$2NM_{G\rightarrow E1}$	0	0
	0-0.202	0-0.179
$2NM_{E1\rightarrow G}$	0	0
	0-0.086	0-0.072
$2NM_{G\rightarrow E2}$	0	0
	0-0.168	0-0.145
$2NM_{E2\rightarrow G}$	0	0
	0-0.077	0-0.065
$2NM_{R\rightarrow E1}$	0.322*	0.212*
	0.007-1.037	0-0.757
$2NM_{E1\rightarrow R}$	0.036	0.017
	0-0.106	0-0.084
$2NM_{R\rightarrow E2}$	0.080	0.071
	0-0.501	0-0.419
$2NM_{E2\rightarrow R}$	0	0
	0-0.067	0-0.047
$2NM_{E1\rightarrow E2}$	0.063	0.028
	0-0.825	0-0.613
$2NM_{E2\rightarrow E1}$	0.347*	0.245*
	0-0.987	0-0.805

Parameters	4PM2	4PME0.5
$2NM_{G\rightarrow A1}$	0	0
	0-2.911	0-1.384
$2NM_{A1\rightarrow G}$	0	0
	0-0.496	0-0.378
$2NM_{E2\rightarrow A1}$	0	0
	0-3.456	0-2.144
$2NM_{A1\rightarrow E2}$	0	0
	0-1.32	0-0.935
$2NM_{G\rightarrow A2}$	0	0
	0-2.201	0-1.234
$2NM_{A2\rightarrow G}$	0.279	0
	0-0.525	0-0.496
N_G	126	132
	72-203	78-211
N_R	27	33
	9-63	12-76
N_{E1}	322	348
	159-617	171-677
N_{E2}	319	349
	172-563	197-598
N_{A1}	0	483
	0-1453	129-1453
N_{A2}	2	2
	0-1193	0-1265
N_{A3}	2	2
	0-1255	0-1100

Chapter 5: General discussion

Chapter 3 describes one of very few, and arguably the most detailed, comparative range-wide phylogeographic studies undertaken to date. Here patterns of population genetic subdivision within each of four co-distributed taxa were resolved using more than one category of genetic marker, and spatially explicit information on current and former ranges were obtained by species distribution modeling (SDM) techniques. Although a small number of studies of comparative phylogeography of co-distributed taxa have also applied SDM methods (e.g. Carstens & Richards 2007; Marske *et al.* 2012; Solomon *et al.* 2008b) to help infer patterns, each of these studies only employed one type of genetic marker, which was typically mitochondrial DNA (mtDNA) for animals. However, as numerous studies have demonstrated, including my own in Chapters 3 and 4, patterns of mtDNA population genetic structure can frequently differ from those based on nuclear DNA (ncDNA) due to a suite of reasons, including sex-biased dispersal/gene flow (e.g. reviewed in Lawson Handley & Perrin 2007 for mammals) and mtDNA biased genetic introgression (e.g. reviewed in Petit & Excoffier 2009). Consequently perspectives on the historical and demographic processes of the target population inferred from mtDNA must be treated as limited and/or biased.

The potential shortcoming of using mtDNA on its own is pertinent to previous studies of the phylogeography of Taiwanese taxa (reviewed in Chapter 3). Here, inferences from comparative phylogeography of multiple taxa have, in some cases, provided insights into important drivers of diversification, such as the Central Mountain Range (CMR). Nonetheless, in other cases such conclusions are likely to have been spurious. For example, the largely allopatric distributions of highly divergent mitochondrial lineages seen in both *Murina gracilis* and *M. recondita* might have led to an over-importance ascribed on the presence and effect of barriers in space (e.g. mountains) and time (e.g. gaps between fragmented refugia). Yet analyses based on nuclear microsatellites suggested only minor contributions of these putative barriers (except in the case of the CMR for *M. recondita*) relative to inter-demic geographic distance alone in shaping population differentiation in the focal species (see Chapter 3). It is also noteworthy that the vast majority of previous range-wide genetic studies of Taiwanese taxa have not included any quantitative ecological information, although a recent study linked gene flow in a Taiwanese

endemic rat species, *Niviventer coninga*, to fragmentation at very fine spatial scales (10×10 km^2) (Wang *et al.* 2008). To date, these above caveats have restricted the power of comparative phylogeographic inferences for Taiwanese taxa, and thus have limited the implications of such studies.

Compared to continental regions, islands often show high levels of endemism while many of them have only modest levels of species richness (Whittaker & Fernández-Palacios 2007). Using a measure termed 'endemism-scaled richness' (ER) to simultaneously evaluate both endemism and species richness, Kier *et al.* (2009) calculated values of this measure for vascular plants and non-fish vertebrates from 90 biogeographical regions worldwide. These authors showed concentrations of high ER values for all taxonomic groups on islands. In particular, Taiwan was found to consistently show among the very highest ER values for these taxonomic groups (although for birds it was ranked second highest) (fig. 1). Such analyses illustrate the high conservation value of the Taiwanese fauna and flora. Approaches in comparative phylogeography and species distribution modeling might thus provide crucial insights into the conservation value and needs of this largely endemic biota.

Of the many threats to biodiversity, global warming is widely considered to be the most serious (e.g. Hansen *et al.* 2010; Malcolm *et al.* 2006; Thomas *et al.* 2004). In Chapter 3 I reconstructed the likely distribution of my focal taxa at the time of Last Glacial Maximum when the climate was cooler, however, SDM techniques might also be very helpful for understanding the possible responses of Taiwanese species to predicted future climatic trends (Elith *et al.* 2010). To illustrate the potential value of this method, for forecasting distributions of the four focal species, here I repeated these models assuming a rise in the global surface temperature of 2 ºC above present values, which is a rather conservative prediction for this century according to Meehl *et al.* (2007). The results presented in Fig. 5.1 suggest that, of the four species, the upland *M. gracilis* will experience the largest proportional loss of suitable habitat in response to the assumed climatic change (32% loss, based on the minimum training-presence threshold). In comparison, *M. recondita*, *M. puta*, and *Kerivoula* sp., are projected to suffer proportional habitat reductions of 16%, 23%, and 23%, respectively. Perhaps even more importantly, these models suggest that the suitable habitat for *M. gracilis* following a rise of 2 ºC will be more fragmented than at present, especially between the northern and southern parts of Taiwan. In Chapter 3, I

have shown that both *M. gracilis* and *M. recondita* have lower levels of genetic connectivity among demes (i.e. lower levels of gene flow) than the other two species, and thus they are expected to be more sensitive to fragmentation of suitable habitats. Taking account of all of my results, I conclude that *M. gracilis* is more vulnerable than the other three species to global warming and, therefore, should be a priority for conservation. Phylogeographic and ecological data on additional Taiwanese species are now needed to gain a better understanding of the consequences of future climatic conditions.

In Chapter 4 I explored in detail the relationship and process of divergence among two putative sister species, *M. gracilis* and *M. recondita*, and a third taxon from continental Asia, *M. eleryi.* The unexpectedly closer relationship between *M. recondita* and *M. eleryi* revealed by multiple ncDNA sequences contradicted that suggested by mtDNA gene sequences. To reconcile the discordance between two categories of genomes, I hypothesized that while ncDNA describes the correct divergence process among these taxa, the pattern of mtDNA is instead the result of the replacement of mitochondrial genomes of *M. recondita* by those of *M. gracilis* during their secondary contact after a period of allopatric divergence. The direction of this mitochondrial introgression and its high intensity compared to that of the nuclear introgression (for which introgression was barely detected) is compatible with expectations based on simulations of neutral introgression conducted by Currat *et al.* (2008), given the presumably male-mediated inter-demic gene flow within each of these taxa.

Nevertheless, several uncertainties about divergence process hypothesized above need to be addressed in future work before it can be firmly accepted. First, estimates of demographic parameters for such a process that were obtained in Chapter 3 were seen to have very broad confidence intervals, especially divergence times, post-split migration rates and effective population sizes of ancestral populations. Such imprecise estimates hindered reliable comparisons of results obtained based on different IM models (two-taxon vs. four-taxon), and also limited the potential for relating demographic events (e.g. divergence times) to biogeographical factors, such as the formation of land-bridges between Taiwan and the continent during glacial episodes. Second, a key factor determining the direction and intensity of neutral introgression based on Currat *et al.* (2008) is the demographic growth of the latter colonizer (*M. recondita* in this case). Yet, a preliminary analysis in Chapter 4 (*BEAST results) did not

provide evidence for historically demographic changes in the four focal taxa (*M. gracilis*, *M. recondita*, and two *M. eleryi* taxa). Although I used 454-sequencing to screen ten loci, it seems likely that increasing the number of nuclear loci further for these analyses would provide greater power and help to address these issues of statistical power (Heled & Drummond 2010; Hey 2010b).

Discussions above highlight a need for more genetic data, especially additional neutral ncDNA loci, in greater numbers of Taiwanese species for comparative phylogeography and in the *M. gracilis* complex for a better understanding on the divergence process. Next generation sequencing (NGS) techniques appear to be particularly promising for such data collection, and a range of NGS applications have emerged. These include amplicon-sequencing methods (e.g. the 454-Pyrosequening in Chapter 4) that target a relatively small proportion of the target genome, to restriction-digest methods (e.g. RAPD) that can provide time-efficient and cost-effective ways of screening large amounts of anonymous sequence for polymorphisms to address questions in population genetics, phylogeography, and phylogenetics (reviewed in Davey *et al.* 2011; McCormack *et al.* 2012). Recent empirical studies that have applied NGS to non-model organisms with increasingly large numbers of loci are beginning to transform our knowledge in these above fields (Bybee *et al.* 2011; Carstens *et al.* 2012; Emerson *et al.* 2010; Holsinger 2010; Puritz *et al.* 2012). In particular, the emergence of new parallel-amplification techniques promise to reduce costs further (e.g. Smith *et al.* 2011; Tewhey *et al.* 2009).

For future studies of *M. gracilis* and *M. recondita* it would be interesting to test the importance of divergent selection in maintaining the integrity of their nuclear genomes, as well as their parapatric distribution. Indeed it is especially intriguing that *M. gracilis* is a montane species while *M. recondita* tends to be restricted to lower elevations. Although contrasting altitudinal preferences might have played a major role in the reinforcement process during the secondary contact of these taxa, this is not strictly necessary given that divergent selection could also have acted on other trait(s) encoded by genes that differentiated during the allopatric period. In fact, geographic isolation rather than divergent selection in any form is recognized as the primary cause of speciation between these species once the allopatric divergence hypothesis is firmly accepted (Via 2009). Nonetheless, genome-wide studies of the genetic basis of adaptation to high altitude have been conducted on human populations in Tibet (Beall *et al.* 2010; Simonson

et al. 2010; Yi *et al.* 2010) and also on non-model animals such as the deer mouse, *Peromyscus maniculatus* (Storz et al. 2010; Storz et al. 2009; Storz et al. 2007) and the yellow-billed pintail, *Anas georgica* (McCracken *et al.* 2009). These studies have found evidence of molecular adaptation in genes such as hemoglobin genes, which appear to confer selective advantages for living in hypoxic conditions at high altitudes. These and other candidate loci could be compared between *M. gracilis* and *M. recondita*. As this field of research moves forward, NGS methods look set to provide even more efficient ways for identifying large numbers of genes associated with altitudinal or other adaptive differences between my focal species (Stapley *et al.* 2010).

Figure 5.1 MaxEnt predictions of suitable habitats for (a) *Murina gracilis*, (b) *M. recondita*, (c) *M. puta*, and (d) *Kerivoula* sp. In each panel, the left-hand map shows such habitats for corresponding species under current climatic conditions, while the right-hand map shows predicted distributions under the conservative projection of a 2 ºC increase in temperature. See Fig. 3.12 for meanings of areas coloured with green and orange.

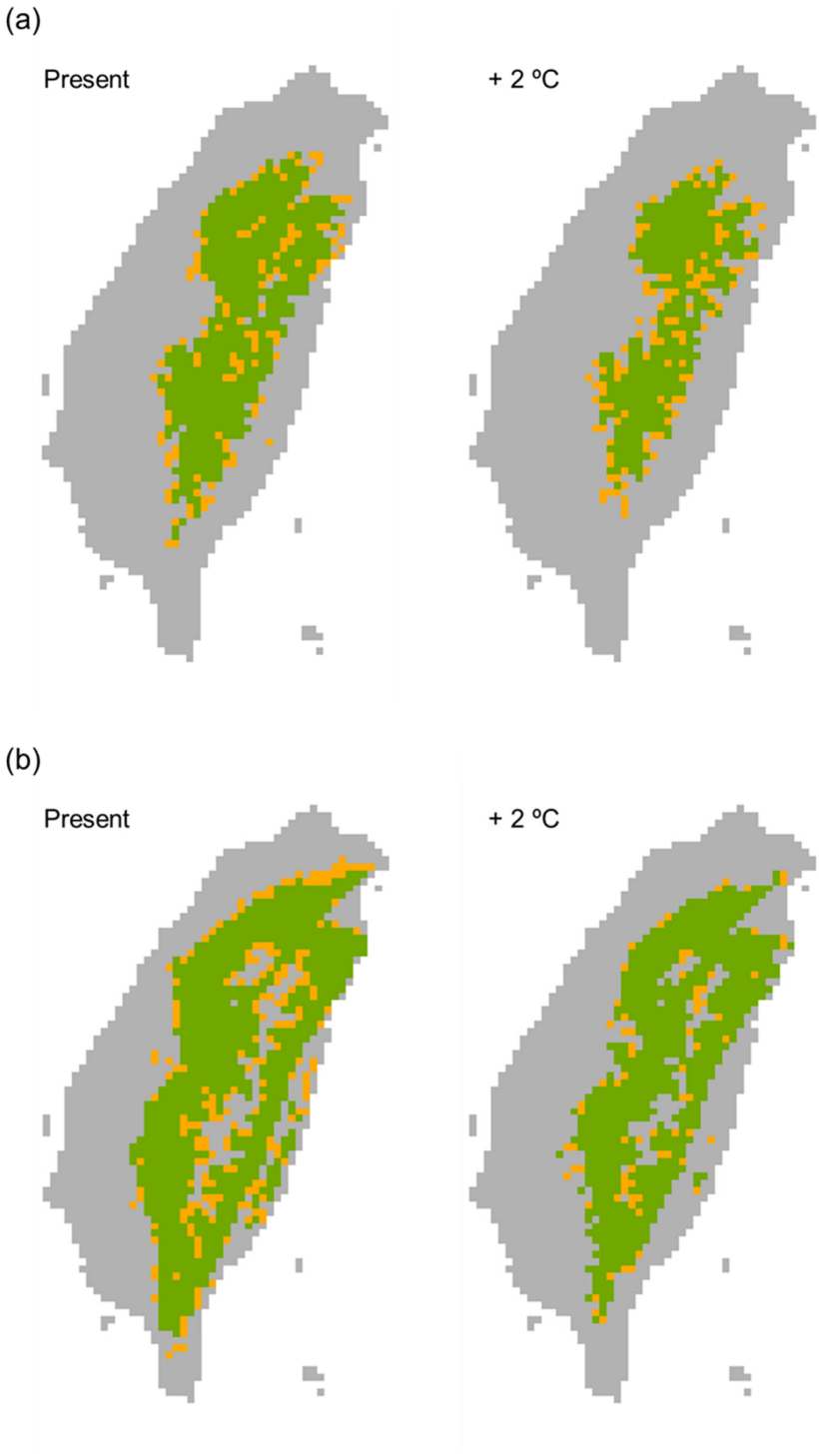
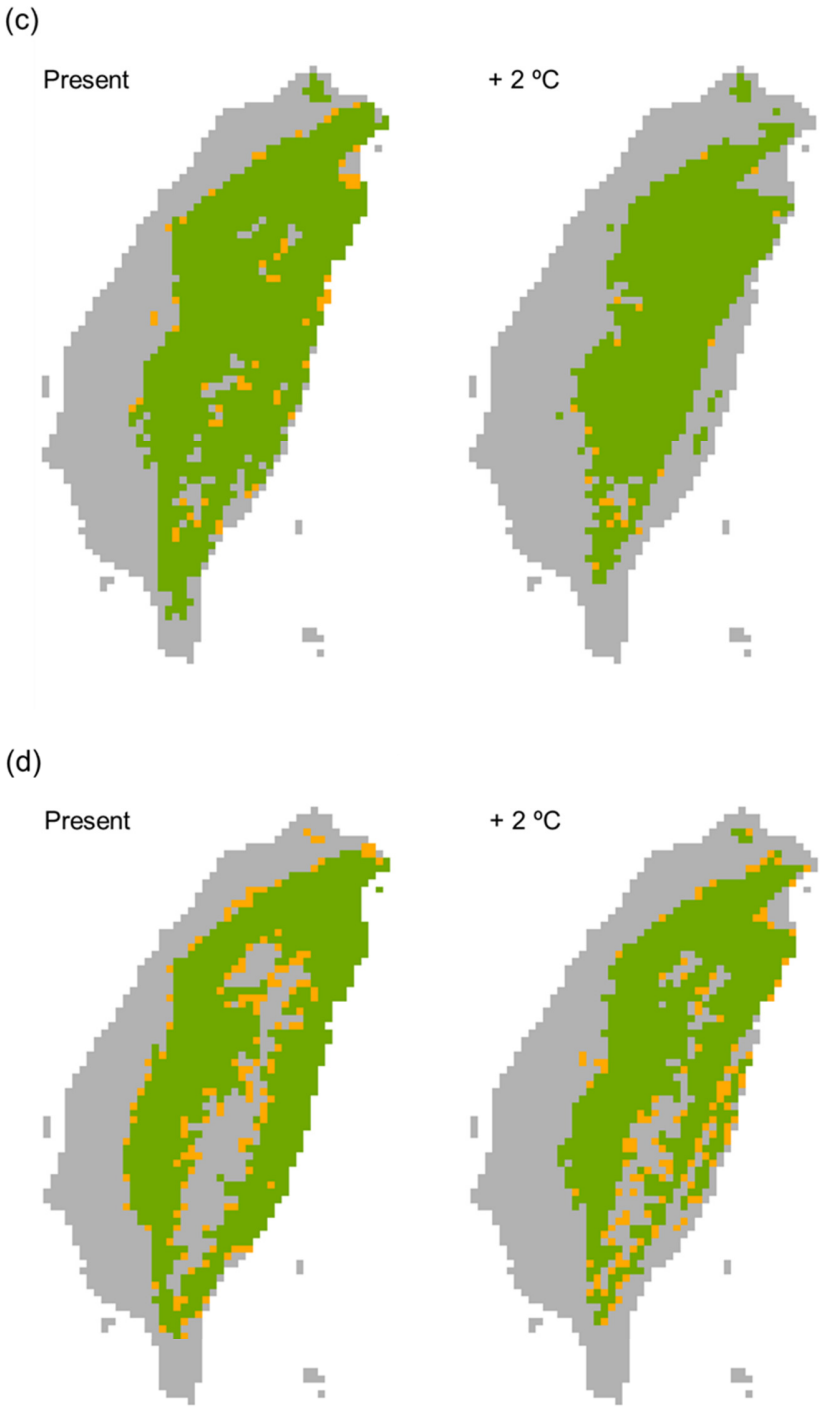


Figure 5.1 Continued.



References

Adriaensen F, Chardon JP, De Blust G*, et al.* (2003) The application of 'least-cost' modelling as a functional landscape model. *Landscape and Urban Planning* **64**, 233-247.

Albaladejo RG, Aparicio A (2007) Population genetic structure and hybridization patterns in the Mediterranean endemics *Phlomis lychnitis* and *P. crinita* (Lamiaceae). *Annals of Botany* **100**, 735-746.

Aldridge HDJN, Rautenbach IL (1987) Morphology, echolocation and resource partitioning in insectivorous bats. *Journal of Animal Ecology* **56**, 763-778.

Anmarkrud JA, Kleven O, Bachmann L, Lifjeld JT (2008) Microsatellite evolution: Mutations, sequence variation, and homoplasy in the hypervariable avian microsatellite locus *HrU10*. *Bmc Evolutionary Biology* **8**.

Arbogast BS, Edwards SV, Wakeley J, Beerli P, Slowinski JB (2002) Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annual Review of Ecology and Systematics* **33**, 707-740.

Arbogast BS, Kenagy GJ (2001) Comparative phylogeography as an integrative approach to historical biogeography. *Journal of Biogeography* **28**, 819-825.

Arnaud JF (2003) Metapopulation genetic structure and migration pathways in the land snail *Helix aspersa*: influence of landscape heterogeneity. *Landscape Ecology* **18**, 333-346.

Arnold ML, Hodges SA (1995) Are natural hybrids fit or unfit relative to their parents? *Trends in Ecology & Evolution* **10**, 67-71.

Augustin L, Barbante C, Barnes PRF*, et al.* (2004) Eight glacial cycles from an Antarctic ice core. *Nature* **429**, 623-628.

Avise JC (1992) Molecular population structure and the biogeographic history of a regional fauna: a case history with lessons for conservation biology. *Oikos* **63**, 62-76.

Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, Massachusetts.

Avise JC (2004) *Molecular Markers, Natural History and Evolution*. Sinauer Associates, Sunderland, Massachusetts.

Avise JC, Arnold J, Ball RM*, et al.* (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* **18**, 489-522.

Avise JC, Shapira JF, Daniel SW, Aquadro CF, Lansman RA (1983) Mitochondrial DNA differentiation during the speciation process in *Peromyscus*. *Molecular Biology and Evolution* **1**, 38-56.

Baer CF, Miyamoto MM, Denver DR (2007) Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nature Reviews Genetics* **8**, 619-631.

Balakrishnan CN, Edwards SV (2009) Nucleotide variation, linkage disequilibrium and founder-facilitated speciation in wild populations of the zebra finch (*Taeniopygia guttata*). *Genetics* **181**, 645-660.

Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. *Molecular Ecology* **11**, 155-165.

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37-48.

Barraclough TG, Vogler AP (2000) Detecting the geographical pattern of speciation from species-level phylogenies. *American Naturalist* **155**, 419-434.

Barton NH (2001) The role of hybridization in evolution. *Molecular Ecology* **10**, 551-568.

Bates PJJ, Struebig MJ, Hayes BD*, et al.* (2007) A new species of *Kerivoula* (Chiroptera : Vespertilionidae) from Southeast Asia. *Acta Chiropterologica* **9**, 323-337.
Bates PJJ, Struebig MJ, Rossiter SJ*, et al.* (2004) A new species of *Kerivoula* (Chiroptera : Vespertilionidae) from Myanmar (Burma). *Acta Chiropterologica* **6**, 219-226.
Beall CM, Cavalleri GL, Deng L*, et al.* (2010) Natural selection on *EPAS1* (*HIF2α*) associated with low hemoglobin concentration in Tibetan highlanders. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 11459-11464.
Beaumont LJ, Gallagher RV, Thuiller W*, et al.* (2009) Different climatic envelopes among invasive populations may lead to underestimations of current and future biological invasions. *Diversity and Distributions* **15**, 409-420.
Beaumont MA, Zhang WY, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics* **162**, 2025-2035.
Becquet C, Przeworski M (2007) A new approach to estimate parameters of speciation models with application to apes. *Genome Research* **17**, 1505-1519.
Becquet C, Przeworski M (2009) Learning about modes of speciation by computational approaches. *Evolution* **63**, 2547-2562.
Bermingham E, Martin AP (1998) Comparative mtDNA phylogeography of neotropical freshwater fishes: testing shared history to infer the evolutionary landscape of lower Central America. *Molecular Ecology* **7**, 499-517.
Bermingham E, Moritz C (1998) Comparative phylogeography: concepts and applications. *Molecular Ecology* **7**, 367-369.
Bernatchez L, Dodson JJ (1991) Phylogeographic structure in mitochondrial DNA of the Lake Whitefish (*Coregonus clupeaformis*) and Its relation to Pleistocene glaciations. *Evolution* **45**, 1016-1035.
Berthier P, Excoffier L, Ruedi M (2006) Recurrent replacement of mtDNA and cryptic hybridization between two sibling bat species *Myotis myotis* and *Myotis blythii*. *Proceedings of the Royal Society of London Series B-Biological Sciences* **273**, 3101-3109.
Birky CW (2001) The inheritance of genes in mitochondria and chloroplasts: Laws, mechanisms, and models. *Annual Review of Genetics* **35**, 125-148.
Brown RP, Suarez NM, Pestano J (2002) The Atlas mountains as a biogeographical divide in North-West Africa: evidence from mtDNA evolution in the Agamid lizard *Agama impalearis*. *Molecular Phylogenetics and Evolution* **24**, 324-332.
Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques* **20**, 1004-1010.
Burke JM, Arnold ML (2001) Genetics and the fitness of hybrids. *Annual Review of Genetics* **35**, 31-52.
Bybee SM, Bracken-Grissom H, Haynes BD*, et al.* (2011) Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. *Genome Biology and Evolution* **3**, 1312-1323.
Cabanne GS, d'Horta FM, Sari EHR, Santos FR, Miyaki CY (2008) Nuclear and mitochondrial phylogeography of the Atlantic forest endemic *Xiphorhynchus fuscus* (Aves: Dendrocolaptidae): Biogeography and systematics implications. *Molecular Phylogenetics and Evolution* **49**, 760-773.
Cannon CH, Morley RJ, Bush ABG (2009) The current refugial rainforests of Sundaland are unrepresentative of their biogeographic past and highly vulnerable to disturbance. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 11188-11193.

Carstens B, Lemmon AR, Lemmon EM (2012) The promises and pitfalls of next-generation sequencing data in phylogeography. *Systematic Biology* **61**, 713-715.
Carstens BC, Richards CL (2007) Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution* **61**, 1439-1454.
Castric V, Bechsgaard J, Schierup MH, Vekemans X (2008) Repeated adaptive introgression at a gene under multiallelic balancing selection. *Plos Genetics* **4**.
Charlesworth D (2010) Don't forget the ancestral polymorphisms. *Heredity* **105**, 509-510.
Chen S-F, Jones G, Rossiter SJ (2008) Sex-biased gene flow and colonization in the Formosan lesser horseshoe bat: inference from nuclear and mitochondrial markers. *Journal of Zoology* **274**, 207-215.
Chen SF, Rossiter SJ, Faulkes CG, Jones G (2006) Population genetic structure and demographic history of the endemic Formosan lesser horseshoe bat (*Rhinolophus monoceros*). *Molecular Ecology* **15**, 1643-1656.
Cheng YP, Hwang SY, Lin TP (2005) Potential refugia in Taiwan revealed by the phylogeographical study of *Castanopsis carlesii* Hayata (Fagaceae). *Molecular Ecology* **14**, 2075-2085.
Chou C-H (2004) *Taxonomic status of the genus Myotis (Chiroptera: Vespertilionidae) in Taiwan*. MS thesis, Tunghai University.
Chou C-H, Hsu C-L, Chuang M-H, Lee L-L (2008) Furled banana leaves as roost for the Formosan tube-nosed bat (*Murina puta*). *Endemic Species Research* **10**, 19-24.
Clark PU, Dyke AS, Shakun JD*, et al.* (2009) The Last Glacial Maximum. *Science* **325**, 710-714.
Collins WD, Bitz CM, Blackmon ML*, et al.* (2006) The Community Climate System Model version 3 (CCSM3). *Journal of Climate* **19**, 2122-2143.
Coyne JA (2007) Sympatric speciation. *Current Biology* **17**, R787-R788.
Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland, Massachusetts.
Coyne JA, Price TD (2000) Little evidence for sympatric speciation in island birds. *Evolution* **54**, 2166-2171.
Creer S, Malhotra A, Thorpe RS, Chou W-H (2001) Multiple causation of phylogeographical pattern as revealed by nested clade analysis of the bamboo viper (*Trimeresurus stejnegeri*) within Taiwan. *Molecular Ecology* **10**, 1967-1981.
Cruzan MB, Templeton AR (2000) Paleoecology and coalescence: phylogeographic analysis of hypotheses from the fossil record. *Trends in Ecology & Evolution* **15**, 491-496.
Csorba G, Bates PJJ (2005) Description of a new species of *Murina* from Cambodia (Chiroptera : Vespertilionidae : Murininae). *Acta Chiropterologica* **7**, 1-7.
Csorba G, Nguyen Truong S, Saveng I, Furey NM (2011) Revealing cryptic bat diversity: three new *Murina* and redescription of *M. tubinaris* from Southeast Asia. *Journal of Mammalogy* **92**, 891-904.
Csorba G, Thong VD, Bates PJJ, Furey NM (2007) Description of a new species of *Murina* from Vietnam (Chiroptera: Vespertilionidae: Murininae). *Occasional Papers, Museum of Texas Tech University* **268**, 1-10.
Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: massive introgression by local genes. *Evolution* **62**, 1908-1920.
Davey JW, Hohenlohe PA, Etter PD*, et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12**, 499-510.
Delvaux D (1996) Age of Lake Malawi (Nyasa) and water level fluctuations. *Musée royal de l'Afrique Centrale (Tervuren), Département de Géologie et Minéralogie, rapport annuel* **1995-1996**, 99-108.

Deng W, Maust BS, Nickle DC, *et al.* (2010) DIVEIN: a web server to analyze phylogenies, sequence divergence, diversity, and informative sites. *Biotechniques* **48**, 405-408.
Dieckmann U, Doebeli M (1999) On the origin of species by sympatric speciation. *Nature* **400**, 354-357.
Dobzhansky T (1951) *Genetics and the Origin of Species*, 3rd edn. Columbia University Press, New York.
Domingo-Roura X, Lopez-Giraldez F, Saeki M, Marmi J (2005) Phylogenetic inference and comparative evolution of a complex microsatellite and its flanking regions in carnivores. *Genetical Research* **85**, 223-233.
Dowling TE, Secor CL (1997) The role of hybridization and introgression in the diversification of animals. *Annual Review of Ecology and Systematics* **28**, 593-619.
Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology* **4**, 699-710.
Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *Bmc Evolutionary Biology* **7**, 214.
Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution* **22**, 1185-1192.
Du FK, Peng XL, Liu JQ, *et al.* (2011) Direction and extent of organelle DNA introgression between two spruce species in the Qinghai-Tibetan Plateau. *New Phytologist* **192**, 1024-1033.
Dudik M, Phillips SJ, Schapire RE (2007) Maximum entropy density estimation with generalized regularization and an application to species distribution modeling. *Journal of Machine Learning Research* **8**, 1217-1260.
Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797.
Edwards SV, Beerli P (2000) Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* **54**, 1839-1854.
Eger JL, Lim BK (2011) Three new species of *Murina* from southern China (Chiroptera: Vespertilionidae). *Acta Chiropterologica* **13**, 227-243.
Elith J, Graham CH, Anderson RP, *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**, 129-151.
Elith J, Kearney M, Phillips S (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution* **1**, 330-342.
Elith J, Phillips SJ, Hastie T, *et al.* (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* **17**, 43-57.
Ellegren H (2000) Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics* **16**, 551-558.
Ellegren H (2004) Microsatellites: Simple sequences with complex evolution. *Nature Reviews Genetics* **5**, 435-445.
Emerson KJ, Merz CR, Catchen JM, *et al.* (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 16196-16200.
Encalada SE, Lahanas PN, Bjorndal KA, *et al.* (1996) Phylogeography and population structure of the Atlantic and Mediterranean green turtle *Chelonia mydas*: a mitochondrial DNA control region sequence assessment. *Molecular Ecology* **5**, 473-483.
Ersts PJ (2011) Geographic Distance Matrix Generator (version 1.2.3). Available from http://biodiversityinformatics.amnh.org/open_source/gdmg.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.
Ewens WJ (1972) The sampling theory of selectively neutral alleles. *Theoretical Population Biology* **3**, 87-&.
Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.
Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479-491.
Fairbanks RG (1989) A 17,000-year glacio-eustatic sea level record: influence of glacial melting rates on the Younger Dryas event and deep-ocean circulation. *Nature* **342**, 637-642.
Fairley TL, Povoa MM, Conn JE (2002) Evaluation of the Amazon River delta as a barrier to gene flow for the regional malaria vector, *Anopheles aquasalis* (Diptera: Culicidae) in northeastern Brazil. *Journal of Medical Entomology* **39**, 861-869.
Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587.
Felsenstein J (1981) Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution* **35**, 124-138.
Fitzpatrick BM, Fordyce JA, Gavrilets S (2009) Pattern, process and geographic modes of speciation. *Journal of Evolutionary Biology* **22**, 2342-2347.
Fjeldså J (1994) Geographical patterns for relict and young species of birds in Africa and South America and implications for conservation priorities. *Biodiversity and Conservation* **3**, 207-226.
Fjeldså J, Lovett JC (1997) Geographical patterns of old and young species in African forest biota: The significance of specific montane areas as evolutionary centres. *Biodiversity and Conservation* **6**, 325-346.
Fjeldså J, Rahbek C (2006) Diversification of tanagers, a species rich bird group, from lowlands to montane regions of South America. *Integrative and Comparative Biology* **46**, 72-81.
Flanders J, Wei L, Rossiter SJ, Zhang S (2011) Identifying the effects of the Pleistocene on the greater horseshoe bat, *Rhinolophus ferrumequinum*, in East Asia using ecological niche modelling and phylogenetic analyses. *Journal of Biogeography* **38**, 439-452.
Flot J-F, Tillier A, Samadi S, Tillier S (2006) Phase determination from direct sequencing of length-variable DNA regions. *Molecular Ecology Notes* **6**, 627-630.
Forbes AA, Powell THQ, Stelinski LL, Smith JJ, Feder JL (2009) Sequential sympatric speciation across trophic levels. *Science* **323**, 776-779.
Francis CM, Borisenko AV, Ivanova NV*, et al.* (2010) The role of DNA barcodes in understanding and conservation of mammal diversity in Southeast Asia. *Plos One* **5**, e12575.
Francis CM, Kingston T, Zubaid A (2007) A new species of *Kerivoula* (Chiroptera : Vespertilionidae) from peninsular Malaysia. *Acta Chiropterologica* **9**, 1-12.
Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915-925.
Fukui D, Agetsuma N, Hill DA (2004) Acoustic identification of eight species of bat (Mammalia: Chiroptera) inhabiting forests of southern Hokkaido, Japan: Potential for conservation monitoring. *Zoological Science* **21**, 947-955.

Furey NM, Thong VD, Bates PJJ, Csorba G (2009) Description of a new species belonging to the *Murina* 'suilla-group' (Chiroptera: Vespertilionidae: Murininae) from north Vietnam. *Acta Chiropterologica* **11**, 225-236.

Galan M, Guivier E, Caraux G, Charbonnel N, Cosson J-F (2010) A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. *Bmc Genomics* **11**.

Galtier N, Depaulis F, Barton NH (2000) Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. *Genetics* **155**, 981-987.

Garcia-Paris M, Good DA, Parra-Olea G, Wake DB (2000) Biodiversity of Costa Rican salamanders: Implications of high levels of genetic differentiation and phylogeographic structure for species formation. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 1640-1647.

Garrick RC, Sunnucks P, Dyer RJ (2010) Nuclear gene phylogeography using PHASE: dealing with unresolved genotypes, lost alleles, and systematic bias in parameter estimation. *Bmc Evolutionary Biology* **10**.

Garza JC, Freimer NB (1996) Homoplasy for size at microsatellite loci in humans and chimpanzees. *Genome Research* **6**, 211-217.

Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology* **10**, 305-318.

Gavrilets S, Waxman D (2002) Sympatric speciation by sexual conflict. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 10533-10538.

Genner MJ, Seehausen O, Cleary DFR, *et al.* (2004) How does the taxonomic status of allopatric populations influence species richness within African cichlid fish assemblages? *Journal of Biogeography* **31**, 93-102.

Geraldes A, Ferrand N, Nachman MW (2006) Contrasting patterns of introgression at X-linked loci across the hybrid zone between subspecies of the European rabbit (*Oryctolagus cuniculus*). *Genetics* **173**, 919-933.

Gibbard PL, Head MJ, Walkers MJC, Subcommission Quaternary S (2010) Formal ratification of the Quaternary System/Period and the Pleistocene Series/Epoch with a base at 2.58 Ma. *Journal of Quaternary Science* **25**, 96-102.

Goloboff PA, Farris JS, Nixon KC (2008) TNT, a free program for phylogenetic analysis. *Cladistics* **24**, 774-786.

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www2.unil.ch/popgen/softwares/fstat.htm. Updated from Goudet (1995).

Guarnizo CE, Amezquita A, Bermingham E (2009) The relative roles of vicariance versus elevational gradients in the genetic differentiation of the high Andean tree frog, *Dendropsophus labialis*. *Molecular Phylogenetics and Evolution* **50**, 84-92.

Guicking D, Fiala B, Blattner FR, *et al.* (2011) Comparative chloroplast DNA phylogeography of two tropical pioneer trees, *Macaranga gigantea* and *Macaranga pearsonii* (Euphorbiaceae). *Tree Genetics & Genomes* **7**, 573-585.

Guillot G (2008) Inference of structure in subdivided populations at low levels of genetic differentiation-the correlated allele frequencies model revisited. *Bioinformatics* **24**, 2222-2228.

Haasl RJ, Payseur BA (2011) Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites. *Heredity* **106**, 158-171.

Hall JPW (2005) Montane speciation patterns in *Ithomiola* butterflies (Lepidoptera : Riodinidae): are they consistently moving up in the world? *Proceedings of the Royal Society B-Biological Sciences* **272**, 2457-2466.

Hamilton MB (2009) *Population genetics*. Wiley-Blackwell, Chichester, West Sussex.
Handley LJL, Manica A, Goudet J, Balloux F (2007) Going the distance: human population genetics in a clinal world. *Trends in Genetics* **23**, 432-439.
Hansen L, Hoffman J, Drews C, Mielbrecht E (2010) Designing climate-smart conservation: guidance and case studies. *Conservation Biology* **24**, 63-69.
Hasan NH, Abdullah MT (2011) A morphological analysis of Malaysian *Kerivoula* (Chiroptera, Vespertilionidae). *Mammal Study* **36**, 87-97.
Hasegawa M, Kishino H, Yano TA (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22**, 160-174.
Hebenstreit R, Bose M, Murray A (2006) Late Pleistocene and early Holocene glaciations in Taiwanese mountains. *Quaternary International* **147**, 76-88.
Heled J, Drummond AJ (2008) Bayesian inference of population size history from multiple loci. *Bmc Evolutionary Biology* **8**.
Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* **27**, 570-580.
Hernandez PA, Graham CH, Master LL, Albert DL (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* **29**, 773-785.
Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**, 247-276.
Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68**, 87-112.
Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907-913.
Hewitt GM (2001a) Speciation, hybrid zones and phylogeography - or seeing genes in space and time. *Molecular Ecology* **10**, 537-549.
Hewitt GM (2001b) Speciation, hybrid zones and phylogeography: or seeing genes in space and time. *Molecular Ecology* **10**, 537-549.
Hewitt GM (2003) Ice ages: their impact on species distributions and evolution. In: *Evolution on Planet Earth* (eds. Rothschild LJ, Lister AM), pp. 339-361. Academic Press, New York.
Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences* **359**, 183-195.
Hey J (2006) Recent advances in assessing gene flow between diverging populations and species. *Current Opinion in Genetics & Development* **16**, 592-596.
Hey J (2010a) The divergence of chimpanzee species and subspecies as revealed in multipopulation isolation-with-migration analyses. *Molecular Biology and Evolution* **27**, 921-933.
Hey J (2010b) Isolation with migration models for more than two populations. *Molecular Biology and Evolution* **27**, 905-920.
Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167**, 747-760.
Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 2785-2790.
Hey J, Won YJ, Sivasundar A, Nielsen R, Markert JA (2004) Using nuclear haplotypes with microsatellites to study gene flow between recently separated Cichlid species. *Molecular Ecology* **13**, 909-919.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**, 1965-1978.
Hoffmann FG, Baker RJ (2003) Comparative phylogeography of short-tailed bats (*Carollia*: Phyllostomidae). *Molecular Ecology* **12**, 3403-3414.
Holm S (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**, 65-70.
Holsinger KE (2010) Next generation population genetics and phylogeography. *Molecular Ecology* **19**, 2361-2363.
Holt AC, Salkeld DJ, Fritz CL, Tucker JR, Gong P (2009) Spatial analysis of plague in California: niche modeling predictions of the current distribution and potential response to climate change. *International Journal of Health Geographics* **8**, 38.
Hoofer SR, Reeder SA, Hansen EW, Van den Bussche RA (2003) Molecular phylogenetics and taxonomic review of noctilionoid and vespertilionoid bats (Chiroptera: Yangochiroptera). *Journal of Mammalogy* **84**, 809-821.
Hoofer SR, Van den Bussche RA (2003) Molecular phylogenetics of the chiropteran family Vespertilionidae. *Acta Chiropterologica* **5**, 1-59.
Hoskin CJ, Higgie M, McDonald KR, Moritz C (2005) Reinforcement drives rapid allopatric speciation. *Nature* **437**, 1353-1356.
Hsu F-H, Lin F-J, Lin Y-S (2001) Phylogeographic structure of the Formosan Wood Mouse, *Apodemus semotus* Thomas. *Zoological Studies* **40**, 91-102.
Hua X, Wang W, Yin W*, et al.* (2009) Phylogeographical analysis of an estuarine fish, *Salanx ariakensis* (Osmeridae: Salanginae) in the north-western Pacific. *Journal of Fish Biology* **75**, 354-367.
Huang CC, Lee L-L (2007) Roost survey of cave-dwelling bats in Yangmingshan National Park, northern Taiwan. *Bat Research News* **48**, 135-135.
Huang J-P, Lin C-P (2010) Diversification in subtropical mountains: Phylogeography, Pleistocene demographic expansion, and evolution of polyphenic mandibles in Taiwanese stag beetle, *Lucanus formosanus*. *Molecular Phylogenetics and Evolution* **57**, 1149-1161.
Huang J-P, Lin C-P (2011) Lineage-specific late pleistocene expansion of an endemic subtropical gossamer-wing damselfly, *Euphaea formosa*, in Taiwan. *Bmc Evolutionary Biology* **11**.
Hubbs CL (1955) Hybridization between fish species in nature. *Systematic Zoology* **4**, 1-20.
Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* **9**, 1322-1332.
Hudson R (1990) Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology* **7**, 1-44.
Hudson RR, Coyne JA (2002) Mathematical consequences of the genealogical species concept. *Evolution* **56**, 1557-1565.
Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**, 147-164.
Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* **294**, 2310-2314.
Husband BC, Sabara HA (2004) Reproductive isolation between autotetraploids and their diploid progenitors in fireweed, *Chamerion angustifolium* (Onagraceae). *New Phytologist* **161**, 703-713.

Ibrahim KM, Nichols RA, Hewitt GM (1996) Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity* **77**, 282-291.
Ivanova NV, Dewaard JR, Hebert PDN (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes* **6**, 998-1002.
Jackson DE (2008) Sympatric speciation: perfume preferences of orchid bee lineages. *Current Biology* **18**, R1092-R1093.
Jakob SS, Ihlow A, Blattner FR (2007) Combined ecological niche modelling and molecular phylogeography revealed the evolutionary history of *Hordeum marinum* (Poaceae) - niche differentiation, loss of genetic diversity, and speciation in Mediterranean Quaternary refugia. *Molecular Ecology* **16**, 1713-1727.
Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801-1806.
Jang-Liaw N-H, Lee T-H (2009) Intraspecific relationships of populations of the brown frog *Rana sauteri* (Ranidae) on Taiwan, inferred from mitochondrial cytochrome *b* sequences. *Zoological Science* **26**, 608-616.
Jang-Liaw N-H, Lee T-H, Chou W-H (2008) Phylogeography of *Sylvirana latouchii* (Anura, Ranidae)in Taiwan. *Zoological Science* **25**, 68-79.
Jeffreys H (1961) *The theory of probability*, 3rd edn. Oxford University Press, Oxford.
Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: *Mammalian protein metabolism* (ed. Munro HH), pp. 21-132. Academic Press, New York.
K-1 model developer (2004) K-1 coupled GCM (MIROC) description. In: *K-1 technical report 1* (eds. Hasumi H, Emori S), pp. 1-34. Center for Climate System Research, University of Tokyo, Tokyo, Japan.
Kawai K, Nikaido M, Harada M*, et al.* (2003) The status of the Japanese and East Asian bats of the genus *Myotis* (Vespertilionidae) based on mitochondrial sequences. *Molecular Phylogenetics and Evolution* **28**, 297-307.
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 Evolutionary Biology* **7**, 22.
Keller I, Seehausen O (2012) Thermal adaptation and ecological speciation. *Molecular Ecology* **21**, 782-799.
Kerth G, Mayer F, Petit E (2002) Extreme sex-biased dispersal in the communally breeding, nonmigratory Bechstein's bat (*Myotis bechsteinii*). *Molecular Ecology* **11**, 1491-1498.
Khan FAA, Solari S, Swier VJ*, et al.* (2010) Systematics of Malaysian woolly bats (Vespertilionidae: *Kerivoula*) inferred from mitochondrial, nuclear, karyotypic, and morphological data. *Journal of Mammalogy* **91**, 1058-1072.
Kier G, Kreft H, Lee TM*, et al.* (2009) A global assessment of endemism and species richness across island and mainland regions. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 9322-9327.
Kingston T, Francis CM, Akbar Z, Kunz TH (2003) Species richness in an insectivorous bat assemblage from Malaysia. *Journal of Tropical Ecology* **19**, 67-79.
Kingston T, Jones G, Akbar Z, Kunz TH (1999) Echolocation signal design in Kerivoulinae and Murininae (Chiroptera: Vespertilionidae) from Malaysia. *Journal of Zoology* **249**, 359-374.
Kingston T, Lim BL, Zubaid A (2006) *Bats of Krau Wildlife Reserve*. Penerbit Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia.
Kingston T, Rossiter SJ (2004) Harmonic-hopping in Wallacea's bats. *Nature* **429**, 654-657.

Kirstein LA, Fellin MG, Willett SD*, et al.* (2010) Pliocene onset of rapid exhumation in Taiwan during arc-continent collision: new insights from detrital thermochronometry. *Basin Research* **22**, 270-285.

Kisel Y, Barraclough TG (2010) Speciation has a spatial scale that depends on levels of gene flow. *American Naturalist* **175**, 316-334.

Knowles LL (2001) Genealogical portraits of speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of the Rocky Mountains. *Proceedings of the Royal Society of London Series B-Biological Sciences* **268**, 319-324.

Knowles LL (2004) The burgeoning field of statistical phylogeography. *Journal of Evolutionary Biology* **17**, 1-10.

Knowles LL, Maddison WP (2002) Statistical phylogeography. *Molecular Ecology* **11**, 2623-2635.

Kotlik P, Deffontaine V, Mascheretti S*, et al.* (2006) A northern glacial refugium for bank voles (*Clethrionomys glareolus*). *Proceedings of the National Academy of Sciences of the United States of America* **103**, 14860-14864.

Kruskop SV (2005) Towards the taxonomy of the Russian *Murina* (Vespertilionidae, Chiroptera). *Russian Journal of Theriology* **4**, 91-99.

Kruskop SV, Eger JL (2008) A new species of tube-nosed bat *Murina* (Vespertilionidae, Chiroptera) from Vietnam. *Acta Chiropterologica* **10**, 213-220.

Kumar S, Subramanian S (2002) Mutation rates in mammalian genomes. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 803-808.

Kuo H-C (2004) *Systematics of bats of genus Murina in Taiwan (Chiroptera: Vespertilionidae)*. MS thesis, National Taiwan University.

Kuo H-C, Fang Y-P, Csorba G, Lee L-L (2009) Three new species of *Murina* (Chiroptera: Vespertilionidae) from Taiwan. *Journal of Mammalogy* **90**, 980-991.

Lack JB, Roehrs ZP, Stanley CE, Jr., Ruedi M, Van den Bussche RA (2010) Molecular phylogenetics of Myotis indicate familial-level divergence for the genus *Cistugo* (Chiroptera). *Journal of Mammalogy* **91**, 976-992.

Lambeck K, Yokoyama Y, Purcell T (2002) Into and out of the Last Glacial Maximum: sea-level change during Oxygen Isotope Stages 3 and 2. *Quaternary Science Reviews* **21**, 343-360.

Laroche J, Durand JD (2004) Genetic structure of fragmented populations of a threatened endemic percid of the Rhône river: *Zingel asper*. *Heredity* **92**, 329-334.

Lawson Handley LJ, Perrin N (2007) Advances in our understanding of mammalian sex-biased dispersal. *Molecular Ecology* **16**, 1559-1578.

Lee Y-F, Kuo Y-M, Chu W-C, Lin Y-H (2007) Chiropteran diversity in different settings of the uplifted coral reef tropical forest of Taiwan. *Journal of Mammalogy* **88**, 1239-1247.

Leman SC, Chen YG, Stajich JE, Noor MAF, Uyenoyama MK (2005) Likelihoods from summary statistics: recent divergence between species. *Genetics* **171**, 1419-1436.

Lemey P, Rambaut A, Drummond AJ, Suchard MA (2009) Bayesian Phylogeography Finds Its Roots. *Plos Computational Biology* **5**.

Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology* **11**, 2453-2465.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.

Liew PM, Kuo CM, Huang SY, Tseng MH (1998) Vegetation change and terrestrial carbon storage in eastern Asia during the Last Glacial Maximum as indicated by a new pollen record from central Taiwan. *Global and Planetary Change* **17**, 85-94.

Lin L-K, Lee L-L, Cheng H-C (2004) *Bats of Taiwan*, 2nd edn. National Museum of Natural Science, Taichung, Taiwan.

Llopart A, Elwyn S, Lachaise D, Coyne JA (2002) Genetics of a difference in pigmentation between *Drosophila yakuba* and *Drosophila santomea*. *Evolution* **56**, 2262-2277.

Llopart A, Lachaise D, Coyne JA (2005) Multilocus analysis of introgression between two sympatric sister species of Drosophila: *Drosophila yakuba* and *D. santomea*. *Genetics* **171**, 197-210.

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

López-Pujol J, Zhang F-M, Sun H-Q, Ying T-S, Ge S (2011) Centres of plant endemism in China: places for survival or for speciation? *Journal of Biogeography* **38**, 1267-1280.

Ma J, Liang B, Zhang S, Metzner W (2008) Dietary composition and echolocation call design of three sympatric insectivorous bat species from China. *Ecological Research* **23**, 113-119.

Malcolm JR, Liu CR, Neilson RP, Hansen L, Hannah L (2006) Global warming and extinctions of endemic species from biodiversity hotspots. *Conservation Biology* **20**, 538-548.

Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* **20**, 229-237.

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

Margulies M, Egholm M, Altman WE*, et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-380.

Marske KA, Leschen RAB, Buckley TR (2012) Concerted versus independent evolution and the search for multiple refugia: comparative phylogeography of four forest beetles. *Evolution* **66**, 1862-1877.

Martins FM, Templeton AR, Pavan ACO, Kohlbach BC, Morgante JS (2009) Phylogeography of the common vampire bat (*Desmodus rotundus*): Marked population structure, Neotropical Pleistocene vicariance and incongruence between nuclear and mtDNA markers. *Bmc Evolutionary Biology* **9**.

Mayr E (1942) *Systematics and the Origin of Species from the Viewpoint of a Zoologist*. Columbia University Press, New York.

Mayr E (1963) *Animal species and evolution*. Harvard University Press, Cambridge, Massachusetts.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2012) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, doi:10.1016/j.ympev.2011.1012.1007.

McCracken KG, Bulgarella M, Johnson KP*, et al.* (2009) Gene flow in the face of countervailing selection: adaptation to high-altitude hypoxia in the βA hemoglobin subunit of yellow-billed pintails in the Andes. *Molecular Biology and Evolution* **26**, 815-827.

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652-654.

McRae BH (2006) Isolation by resistance. *Evolution* **60**, 1551-1561.

McRae BH, Dickson BG, Keitt TH, Shah VB (2008) Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology* **89**, 2712-2724.
McRae BH, Shah VB (2009) Circuitscape (version 3.5). Available from http://www.circuitscape.org.
Meehl GA, Stocker TF, Collins WD*, et al.* (2007) Global climate projections. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds. Solomon S, Qin D, Manning M*, et al.*). Cambridge University Press, Cambridge.
Melo-Ferreira J, Boursot P, Carneiro M*, et al.* (2012) Recurrent introgression of mitochondrial DNA among hares (*Lepus* spp.) revealed by species-tree inference and coalescent simulations. *Systematic Biology* **61**, 367-381.
Missoup AD, Nicolas V, Wendelen W*, et al.* (2012) Systematics and diversification of *Praomys* species (Rodentia: Muridae) endemic to the Cameroon Volcanic Line (West Central Africa). *Zoologica Scripta* **41**, 327-345.
Mittermeier RA, Gil PR, Hoffman M*, et al.* (2005) *Hotspots Revisited: Earth's Biologically Richest and Most Endangered Terrestrial Ecoregions*. Conservation International, Washington D.C.
Moritz C, Hoskin CJ, MacKenzie JB*, et al.* (2009) Identification and dynamics of a cryptic suture zone in tropical rainforest. *Proceedings of the Royal Society B-Biological Sciences* **276**, 1235-1244.
Moritz C, Patton JL, Schneider CJ, Smith TB (2000) Diversification of rainforest faunas: An integrated molecular approach. *Annual Review of Ecology and Systematics* **31**, 533-563.
Moussalli A, Moritz C, Williams SE, Carnaval AC (2009) Variable responses of skinks to a common history of rainforest fluctuation: concordance between phylogeography and palaeo-distribution models. *Molecular Ecology* **18**, 483-499.
Nadachowska K, Babik W (2009) Divergence in the face of gene flow: the case of two newts (Amphibia: Salamandridae). *Molecular Biology and Evolution* **26**, 829-841.
Nater A, Nietlisbach P, Arora N*, et al.* (2011) Sex-biased dispersal and volcanic activities shaped phylogeographic patterns of extant orangutans (genus: *Pongo*). *Molecular Biology and Evolution* **28**, 2275-2288.
Neigel JE (2002) Is F_{ST} obsolete? *Conservation Genetics* **3**, 167-173.
Nichols RA, Freeman KLM (2004) Using molecular markers with high mutation rates to obtain estimates of relative population size and to distinguish the effects of gene flow and mutation: a demonstration using data from endemic Mauritian skinks. *Molecular Ecology* **13**, 775-787.
Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: A Markov chain Monte Carlo approach. *Genetics* **158**, 885-896.
Norberg UM, Rayner JMV (1987) Ecological morphology and flight in bats (Mammalia; Chiroptera): wing adaptations, flight performance, foraging strategy and echolocation. *Proceedings of the Royal Society of London Series B-Biological Sciences* **316**, 337-419.
Osada N, Wu CI (2005) Inferring the mode of speciation from genomic data: A study of the great apes. *Genetics* **169**, 259-264.
Oshida T, Lee JK, Lin LK, Chen YJ (2006) Phylogeography of Pallas's squirrel in Taiwan: geographical isolation in an arboreal small mammal. *Journal of Mammalogy* **87**, 247-254.
Panchal M, Beaumont MA (2007) The automation and evaluation of nested clade phylogeographic analysis. *Evolution* **61**, 1466-1480.

Pennay M, Law B, Reinhold L (2004) *Bat calls of New South Wales: Region-based guide to the echolocation calls of Microchiropteran bats*. NSW Department of Environment and Conservation, Hurstville, New South Wales, Australia.

Petit RJ (2008) The coup de grâce for the nested clade phylogeographic analysis? *Molecular Ecology* **17**, 516-518.

Petit RJ, Excoffier L (2009) Gene flow and species delimitation. *Trends in Ecology & Evolution* **24**, 386-393.

Petit RJ, Grivet D (2002) Optimal randomization strategies when testing the existence of a phylogeographic structure. *Genetics* **161**, 469-471.

Phillips SJ, Anderson RP, Schapire RE (2006a) Maxent v3.3.3. available from: http://www.cs.princeton.edu/~schapire/maxent/.

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

Phillips SJ, Dudík M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**, 161-175.

Plana V (2004) Mechanisms and tempo of evolution in the African Guineo-Congolian rainforest. *Philosophical Transactions of the Royal Society B: Biological Sciences* **359**, 1585-1594.

Polzin T, Daneshmand SV (2003) On Steiner trees and minimum spanning trees in hypergraphs. *Operations Research Letters* **31**, 12-20.

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

Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology & Evolution* **16**, 37-45.

Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: Models and data. *American Journal of Human Genetics* **69**, 1-14.

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

Puritz JB, Addison JA, Toonen RJ (2012) Next-generation phylogeography: a targeted approach for multilocus sequencing of non-model organisms. *Plos One* **7**, e34241-e34241.

Pybus OG, Rambaut A, Harvey PH (2000) An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* **155**, 1429-1437.

Pybus OG, Rambaut A, Holmes EC, Harvey PH (2002) New inferences from tree shape: numbers of missing taxa and population growth rates. *Systematic Biology* **51**, 881-888.

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

Ramirez-Soriano A, Ramos-Onsins SE, Rozas J, Calafell F, Navarro A (2008) Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. *Genetics* **179**, 555-567.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution* **19**, 2092-2100.

Ramos Pereira MJ, Salgueiro P, Rodrigues L, Coelho MM, Palmeirim JM (2009) Population structure of a cave-dwelling bat, *Miniopterus schreibersii*: does it reflect history and social organization? *Journal of Heredity* **100**, 533-544.

Randi E, Gentile L, Boscagli G, Huber D, Roth HU (1994) Mitochondrial DNA sequence divergence among some west European brown bear (Ursus arctos L.) populations. Lessons for conservation. *Heredity* **73**, 480-489.

Ravelo AC, Andreasen DH, Lyle M, Lyle AO, Wara MW (2004) Regional climate shifts caused by gradual global cooling in the Pliocene epoch. *Nature* **429**, 263-267.
Ray N (2005) PATHMATRIX: a geographical information system tool to compute effective distances among samples. *Molecular Ecology Notes* **5**, 177-180.
Rhodes MP (1995) Wing morphology and flight behaviour of the golden-tipped bat, *Phoniscus papuensis* (Dobson) (Chiroptera: Vespertilionidae). *Australian Journal of Zoology* **43**, 657-663.
Rice WR (1984) Disruptive selection on habitat preference and the evolution of reproductive isolation - a simulation study. *Evolution* **38**, 1251-1260.
Richards-Zawacki CL (2009) Effects of slope and riparian habitat connectivity on gene flow in an endangered Panamanian frog, *Atelopus varius*. *Diversity and Distributions* **15**, 796-806.
Richardson JE, Pennington RT, Pennington TD, Hollingsworth PM (2001) Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science* **293**, 2242-2245.
Rieseberg LH, Kim S-C, Randell RA*, et al.* (2007) Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica* **129**, 149-165.
Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574.
Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**, 137-138.
Rosenberg NA, Pritchard JK, Weber JL*, et al.* (2002) Genetic structure of human populations. *Science* **298**, 2381-2385.
Rossiter SJ, Benda P, Dietz C, Zhang S, Jones G (2007) Rangewide phylogeography in the greater horseshoe bat inferred from microsatellites: implications for population history, taxonomy and conservation. *Molecular Ecology* **16**, 4699-4714.
Rossiter SJ, Ransome RD, Faulkes CG, Dawson DA, Jones G (2006) Long-term reproductive skew in male greater horseshoe bats. *Molecular Ecology* **15**, 3035-3043.
Rossiter SJ, Zubaid A, Mohd-Adnan A*, et al.* (2012) Social organization and genetic structure: insights from codistributed bat populations. *Molecular Ecology* **21**, 647-661.
Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**, 1219-1228.
Roy MS (1997) Recent diversification in African greenbuls (Pycnonotidae: *Andropadus*) supports a montane speciation model. *Proceedings of the Royal Society of London Series B-Biological Sciences* **264**, 1337-1344.
Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols* (eds. Krawetz S, Misener S), pp. 365-386. Humana Press Inc., Totowa, New Jersey.
Ruedi M, Biswas J, Csorba G (2012) Bats from the wet: two new species of Tube-nosed bats (Chiroptera: Vespertilionidae) from Meghalaya, India. *Revue Suisse De Zoologie* **119**, 111-135.
Ruedi M, Mayer F (2001) Molecular systematics of bats of the genus *Myotis* (Vespertilionidae) suggests deterministic ecomorphological convergences. *Molecular Phylogenetics and Evolution* **21**, 436-448.
Russell AL, Goodman SM, Fiorentino I, Yoder AD (2008) Population genetic analysis of *Myzopoda* (Chiroptera : Myzopodidae) in Madagascar. *Journal of Mammalogy* **89**, 209-221.

Schmieder DA, Kingston T, Hashim R, Siemers BM (2012) Sensory constraints on prey detection performance in an ensemble of vespertilionid understorey rain forest bats. *Functional Ecology*, doi: 10.1111/j.1365-2435.2012.02024.x.

Schnitzler HU, Kalko EKV (2001) Echolocation by insect-eating bats. *Bioscience* **51**, 557-569.

Schnitzler HU, Moss CF, Denzinger A (2003) From spatial orientation to food acquisition in echolocating bats. *Trends in Ecology & Evolution* **18**, 386-394.

Seehausen O, Terai Y, Magalhaes IS*, et al.* (2008) Speciation through sensory drive in cichlid fish. *Nature* **455**, 620-U623.

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* **9**, 615-629.

Sequeira F, Sodre D, Ferrand N*, et al.* (2011) Hybridization and massive mtDNA unidirectional introgression between the closely related Neotropical toads *Rhinella marina* and *R. schneideri* inferred from mtDNA and nuclear markers. *Bmc Evolutionary Biology* **11**.

Servedio MR (2004) The what and why of research on reinforcement. *PLoS Biology* **2**, e420.

Shapiro B, Rambaut A, Drummond AJ (2006) Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Molecular Biology and Evolution* **23**, 7-9.

Shih HT, Hung HC, Schubart CD, Chen CLA, Chang HW (2006) Intraspecific genetic diversity of the endemic freshwater crab *Candidiopotamon rathbunae* (Decapoda, Brachyura, Potamidae) reflects five million years of the geological history of Taiwan. *Journal of Biogeography* **33**, 980-989.

Sibuet JC, Hsu SK (2004) How was Taiwan created? *Tectonophysics* **379**, 159-181.

Siddall M, Rohling EJ, Almogi-Labin A*, et al.* (2003) Sea-level fluctuations during the last glacial cycle. *Nature* **423**, 853-858.

Simmons NB (2005) Order Chiroptera. In: *Mammal species of the world: A taxonomic and geographic reference* (eds. Wilson DE, Reeder DM), pp. 312-529. Johns Hopkins University Press, Baltimore, Maryland.

Simoes M, Beyssac O, Chen Y-G (2012) Late Cenozoic metamorphism and mountain building in Taiwan: A review. *Journal of Asian Earth Sciences* **46**, 92-119.

Simonson TS, Yang Y, Huff CD*, et al.* (2010) Genetic evidence for high-altitude adaptation in Tibet. *Science* **329**, 72-75.

Slatkin M (1982) Pleiotropy and parapatric speciation. *Evolution* **36**, 263-270.

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**, 264-279.

Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**, 555-562.

Smith MJ, Pascal CE, Grauvogel Z*, et al.* (2011) Multiplex preamplification PCR and microsatellite validation enables accurate single nucleotide polymorphism genotyping of historical fish scales. *Molecular Ecology Resources* **11**, 268-277.

Solomon SE, Bacci M, Jr, Martins J, Jr, Vinha GG, Mueller UG (2008a) Paleodistributions and comparative molecular phylogeography of Leafcutter Ants (*Atta* spp.) provide new insight into the origins of Amazonian diversity. *Plos One* **3**, e2738.

Solomon SE, Bacci M, Jr., Martins J, Jr., Vinha GG, Mueller UG (2008b) Paleodistributions and comparative molecular phylogeography of leafcutter ants (*Atta* spp.) provide new insight into the origins of Amazonian diversity. *Plos One* **3**.

Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology* **15**, 4261-4293.

Song Y, Endepols S, Klemann N*, et al.* (2011) Adaptive introgression of anticoagulant rodent poison resistance by hybridization between Old World mice. *Current Biology* **21**, 1296-1301.

Spear SF, Balkenhol N, Fortin M-J, McRae BH, Scribner K (2010) Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. *Molecular Ecology* **19**, 3576-3591.

Sripathi K, Raghuram H, Nathan PT (2006) Echolocation sounds of the painted bat *Kerivoula picta* (Vespertilionidae). *Current Science* **91**, 1145-1147.

Stadelmann B, Herrera LG, Arroyo-Cabrales J*, et al.* (2004a) Molecular systematics of the fishing bat *Myotis* (*Pizonyx*) *vivesi*. *Journal of Mammalogy* **85**, 133-139.

Stadelmann B, Jacobs DS, Schoeman C, Ruedi M (2004b) Phylogeny of African *Myotis* bats (Chiroptera, Vespertilionidae) inferred from cytochrome *b* sequences. *Acta Chiropterologica* **6**, 177-192.

Stapley J, Reger J, Feulner PGD*, et al.* (2010) Adaptation genomics: the next generation. *Trends in Ecology & Evolution* **25**, 705-712.

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

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

Storz JF, Beaumont MA (2002) Testing for genetic evidence of population expansion and contraction: An empirical analysis of microsatellite DNA variation using a hierarchical Bayesian model. *Evolution* **56**, 154-166.

Storz JF, Runck AM, Moriyama H, Weber RE, Fago A (2010) Genetic differences in hemoglobin function between highland and lowland deer mice. *Journal of Experimental Biology* **213**, 2565-2574.

Storz JF, Runck AM, Sabatino SJ*, et al.* (2009) Evolutionary and functional insights into the mechanism underlying high-altitude adaptation of deer mouse hemoglobin. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 14450-14455.

Storz JF, Sabatino SJ, Hoffmann FG*, et al.* (2007) The molecular basis of high-altitude adaptation in deer mice. *Plos Genetics* **3**, 448-459.

Strasburg JL, Rieseberg LH (2011) Interpreting the estimated timing of migration events between hybridizing species. *Molecular Ecology* **20**, 2353-2366.

Strimmer K, Pybus OG (2001) Exploring the demographic history of DNA sequences using the generalized skyline plot. *Molecular Biology and Evolution* **18**, 2298-2305.

Struebig MJ, Horsburgh GJ, Pandhal J*, et al.* (2008) Isolation and characterisation of microsatellite loci in the papillose woolly bat, *Kerivoula papillosa* (Chiroptera: Vespertilionidae). *Conservation Genetics* **9**, 751-756.

Sundstrom L, Keller L, Chapuisat M (2003) Inbreeding and sex-biased gene flow in the ant *Formica exsecta*. *Evolution* **57**, 1552-1561.

Swenson NG, Howard DJ (2005) Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *American Naturalist* **166**, 581-591.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437-460.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10**, 512-526.

Tamura K, Peterson D, Peterson N*, et al.* (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**, 2731-2739.

Tavaré S (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. In: *Some mathematical questions in biology: DNA sequence analysis* (ed. Miura RM), pp. 57-86. American Mathematical Society, Providence, Rhode Island.

Taylor PJ, Maree S, Van Sandwyk J*, et al.* (2009) Speciation mirrors geomorphology and palaeoclimatic history in African laminate-toothed rats (Muridae: *Otomyini*) of the *Otomys denti* and *Otomys lacustris* species-complexes in the 'Montane Circle' of East Africa. *Biological Journal of the Linnean Society* **96**, 913-941.

Templeton AR (1998) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology* **7**, 381-397.

Templeton AR (2002a) "Optimal" randomization strategies when testing the existence of a phylogeographic structure: A reply to Petit and Grivet. *Genetics* **161**, 473-475.

Templeton AR (2002b) Out of Africa again and again. *Nature* **416**, 45-51.

Templeton AR (2004) Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Ecology* **13**, 789-809.

Templeton AR, Routman E, Phillips CA (1995) Separating population structure from population history: a cladistic analysis of the geographic distribution of mitochondrial DNA haplotypes in the Tiger Salamander, *Ambystoma Tigrinum*. *Genetics* **140**, 767-782.

Tewhey R, Warner JB, Nakano M*, et al.* (2009) Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nature Biotechnology* **27**, 1025-U1094.

Thomas CD, Cameron A, Green RE*, et al.* (2004) Extinction risk from climate change. *Nature* **427**, 145-148.

Turelli M, Barton NH, Coyne JA (2001) Theory and speciation. *Trends in Ecology & Evolution* **16**, 330-343.

Turmelle AS, Kunz TH, Sorenson MD (2011) A tale of two genomes: contrasting patterns of phylogeographic structure in a widely distributed bat. *Molecular Ecology* **20**, 357-375.

Tzeng C-S, Lin Y-S, Lin S-M, Wang T-Y, Wang F-Y (2006) The phylogeography and population demographics of selected freshwater fishes in Taiwan. *Zoological Studies* **45**, 285-297.

Ueda S, Quek S-P, Itioka T, Murase K, Itino T (2010) Phylogeography of the *Coccus* scale insects inhabiting myrmecophytic *Macaranga* plants in Southeast Asia. *Population Ecology* **52**, 137-146.

van Oppen MJH, Rico C, Turner GF, Hewitt GM (2000) Extensive homoplasy, nonstepwise mutations, and shared ancestral polymorphism at a complex microsatellite locus in Lake Malawi cichlids. *Molecular Biology and Evolution* **17**, 489-498.

Verardi A, Lucchini V, Randi E (2006) Detecting introgressive hybridization between free-ranging domestic dogs and wild wolves (*Canis lupus*) by admixture linkage disequilibrium analysis. *Molecular Ecology* **15**, 2845-2855.

Via S (2001) Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology & Evolution* **16**, 381-390.

Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 9939-9946.

Vidya TNC, Sukumar R, Melnick DJ (2009) Range-wide mtDNA phylogeography yields insights into the origins of Asian elephants. *Proceedings of the Royal Society B-Biological Sciences* **276**, 893-902.

Vila M, Vidal-Romani JR, Bjorklund M (2005) The importance of time scale and multiple refugia: Incipient speciation and admixture of lineages in the butterfly *Erebia triaria* (Nymphalidae). *Molecular Phylogenetics and Evolution* **36**, 249-260.

Vitalis R, Couvet D (2001) Estimation of effective population size and migration rate from one- and two-locus identity measures. *Genetics* **157**, 911-925.

Voelker G, Outlaw RK, Bowie RCK (2010) Pliocene forest dynamics as a primary driver of African bird speciation. *Global Ecology and Biogeography* **19**, 111-121.

von Helversen O, Heller KG, Mayer F*, et al.* (2001) Cryptic mammalian species: a new species of whiskered bat (*Myotis alcathoe* n sp) in Europe. *Naturwissenschaften* **88**, 217-223.

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

Wakeley J (1996) Distinguishing migration from isolation using the variance of pairwise differences. *Theoretical Population Biology* **49**, 369-386.

Wakeley J, Hey J (1997) Estimating ancestral population parameters. *Genetics* **145**, 847-855.

Waltari E, Guralnick RP (2009) Ecological niche modelling of montane mammals in the Great Basin, North America: examining past and present connectivity of species across basins and ranges. *Journal of Biogeography* **36**, 148-161.

Waltari E, Hijmans RJ, Peterson AT*, et al.* (2007) Locating Pleistocene refugia: comparing phylogeographic and ecological niche model predictions. *Plos One* **2**, e563.

Wan QH, Wu H, Fujihara T, Fang SG (2004) Which genetic marker for which conservation genetics issue? *Electrophoresis* **25**, 2165-2176.

Wang IJ, Summers K (2010) Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Molecular Ecology* **19**, 447-458.

Wang J-P, Lin H-D, Huang S*, et al.* (2004) Phylogeography of *Varicorhinus barbatulus* (Cyprinidae) in Taiwan based on nucleotide variation of mtDNA and allozymes. *Molecular Phylogenetics and Evolution* **31**, 1143-1156.

Wang JP, Hsu KC, Chiang TY (2000) Mitochondrial DNA phylogeography of *Acrossocheilus paradoxus* (Cyprinidae) in Taiwan. *Molecular Ecology* **9**, 1483-1494.

Wang X, Sun X, Wang P, Stattegger K (2009) Vegetation on the Sunda Shelf, South China Sea, during the Last Glacial Maximum. *Palaeogeography Palaeoclimatology Palaeoecology* **278**, 88-97.

Wang Y-H, Yang K-C, Bridgman CL, Lin L-K (2008) Habitat suitability modelling to correlate gene flow with landscape connectivity. *Landscape Ecology* **23**, 989-1000.

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**, 256-276.

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

Weir JT, Schluter D (2004) Ice sheets promote speciation in boreal birds. *Proceedings of the Royal Society of London Series B-Biological Sciences* **271**, 1881-1887.

Whitlock MC (2011) G $'_{ST}$ and D do not replace F_{ST}. *Molecular Ecology* **20**, 1083-1091.

Whittaker RJ, Fernández-Palacios JM (2007) *Island Biogeography: Ecology, Evolution, and Conservation*, 2nd edn. Oxford University Press, Oxford.

Wilmer JW, Hall L, Barratt E, Moritz C (1999) Genetic structure and male-mediated gene flow in the ghost bat (*Macroderma gigas*). *Evolution* **53**, 1582-1591.

Woerner AE, Cox MP, Hammer MF (2007) Recombination-filtered genomic datasets by information maximization. *Bioinformatics* **23**, 1851-1853.
Won YJ, Hey J (2005) Divergence population genetics of chimpanzees. *Molecular Biology and Evolution* **22**, 297-307.
Won YJ, Sivasundar A, Wang Y, Hey J (2005) On the origin of Lake Malawi cichlid species: A population genetic analysis of divergence. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 6581-6586.
Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* **19**, 395-420.
Wu C-T (2007) *Taxonomic study of the genus Pipistrellus (Chiroptera: Vespertilionidae) in Taiwan*. MS thesis, National Chiayi University.
Wu CI (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology* **14**, 851-865.
Wu CI, Ting CT (2004) Genes and speciation. *Nature Reviews Genetics* **5**, 114-122.
Wurster CM, Bird MI, Bull ID*, et al.* (2010) Forest contraction in north equatorial Southeast Asia during the Last Glacial Period. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 15508-15511.
Xu T, Abbott RJ, Milne RI*, et al.* (2010) Phylogeography and allopatric divergence of cypress species (*Cupressus* L.) in the Qinghai-Tibetan Plateau and adjacent regions. *Bmc Evolutionary Biology* **10**.
Yang C-A (2008) *Population genetic structure and phylogeographic studies of Taiwanese woolly bat (Kerivoula sp.)*. M.S. thesis, National Chiayi University.
Yeh W-B, Chang Y-L, Lin C-H, Wu F-S, Yang J-T (2004) Genetic differentiation of *Loxoblemmus appendicularis* complex (Orthoptera: Gryllidae): speciation through vicariant and glaciation events. *Annals of the Entomological Society of America* **97**, 613-623.
Yi X, Liang Y, Huerta-Sanchez E*, et al.* (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**, 75-78.
Yoshiyuki M (1989) *A Systematic Study of the Japanese Chiroptera*. National Science Museum, Tokyo, Japan.
Yu HT (1994) Distribution and abundance of small mammals along a subtropical elevational gradient in central Taiwan. *Journal of Zoology* **234**, 577-600.
Yu HT (1995) Patterns of diversification and genetic population structure of small mammals in Taiwan. *Biological Journal of the Linnean Society* **55**, 69-89.
Yuan SL, Lin LK, Oshida T (2006) Phylogeography of the mole-shrew (*Anourosorex yamashinai*) in Taiwan: implications of interglacial refugia in a high-elevation small mammal. *Molecular Ecology* **15**, 2119-2130.
Yule GU (1924) A mathematical theory of evolution based on the conclusions of Dr J. C. Willis. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **213**, 21-87.