NATURE'S GENES IN SPACE AND TIME: USING DNA TO UNCOVER THE HISTORY AND FORMATION OF BIODIVERSITY

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

BAILEY DALE MCKAY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

ADVISED BY

ROBERT ZINK

AUGUST 2011

Acknowledgments

I am grateful to my advisor Bob Zink for creating an environment where I could be a successful graduate student. Bob was a reliable source for good discussion and was always quick with useful written feedback. I especially thank him for giving me the freedom to pursue my own interests. I thank my committee, Sharon Jansa, Ken Kozak, and Scott Lanyon, for all of their work.

My long-time collaborator Herman Mays deserves special thanks. He first introduced me to East Asia and has since provided endless discussion on a great many things. He also provided a lot of help in the lab. Keith Barker, Dominik Halas, Chih-ming Hung, David Porter, Alexis Powell, Andrew Simons, Hernan Vazquez, Susan Weller, Sue Wenker, and Lisa Wiggins provided various additional help during my dissertation.

Many friends and collaborators in Asia have also helped me during these last four years: Cheng-te Yao, Fasheng Zou, Isao Nishiumi, Shou-hsien Li, Fumin Lei, Junhui Huang, Fu-hsiung Hsu, Yasuko, Iwami, Pei-Fen Lee, Hui Li, Ruey-shing Lin, Yi-wen Ping, Sayaka, Someya, Yuchun Wu, and Hsiao-wei Yuan.

I thank the American Museum of Natural History, the University of Kansas Biodiversity Institute, National Taiwan Normal University, the South China Institute for Endangered Animals, the Taiwan Endemic Species Research Institute, and the National Museum of Nature and Science of Japan for providing tissue samples.

Work from this dissertation was supported by grants from the National Science Foundation (NSF) including three awards from the East Asia and Pacific Summer Institutes (EAPSI; to Japan, China, and Taiwan) and an NSF doctoral dissertation

improvement grant (DEB-1011599). Additional support was provided by the EEB department, the American Museum of Natural History, the American Ornithologists' Union, the Wilson Ornithological Society, the Dayton Wilkie Fund, and a Simons Fellowship in Systematic Biology from the Bell Museum of Natural History.

Finally, I owe everything to my wonderful family. My parents Dale and Donna McKay sacrificed a lot to give me every opportunity to succeed in life. They have always encouraged me in everything I have done. They are truly perfect parents. My sister Hollon is a great friend that I am lucky to have shared my childhood with. I am also blessed to have a large extended family, and my grandparents as well as my many aunts, uncles, and cousins have always been a source of support for me. I love them all very much.

Dedication

I dedicate this dissertation to my parents, Dale and Donna McKay. Thank you for always encouraging me to do what I love.

Abstract

Modern science is tasked with explaining biological diversity: where it is, where it comes from, and how it evolves. Molecular tools have revolutionized this endeavor and given us the ability to peer into the evolutionary past like never before. Drawing on examples from the birds and mammals of East Asia, I present a progression of methods for uncovering the history and formation of species using molecular approaches. Because the characterization of diversity must necessarily precede the study of diversity, I begin in Chapter 1 with a species delimitation study aimed at resolving evolutionary lineages. I conclude that fixed character differences are better indicators of lineage limits than statistical approaches and that both morphological and molecular characters may be necessary to reveal lineages when divergence is recent. In Chapter 2 I describe the evolutionary history of a single-species in which I found evidence for a recent range-wide demographic expansion. This particular example demonstrates how phylogeography and ecological niche modeling can reciprocally illuminate aspects of evolutionary history. Finally, in Chapter 3 I discuss a comparative study of the evolutionary histories of multiple co-distributed species. Comparative studies reveal the generalities involved in evolution and allow us to discern the major drivers of biological diversity.

Table of Contents

List of Tables

List of Figures

Chapter 1

Chapter 2

Chapter 3

Introduction

A central challenge in modern science is to explain biological diversity: where it is, where it comes from, and how it evolves. As biologists, we want to understand how one species gave rise to all the diverse forms we see in nature today. In addition, we want to know how species came to occupy their present distributions. Why, for example, are polar bears distributed in the North Pole whereas penguins are distributed in the South Pole? Molecular tools have given us new power to discover the patterns and processes involved in the evolution of life. For example, we can study the distribution of DNA sequences in natural populations to identify cryptic lineages that might not look different phenotypically, yet have still functioned independently in the evolutionary history of a geographic area. Analyses of molecular data shed new light on evolutionary processes. For example, the current distribution of DNA sequence variation can be used to infer past processes such as population size changes and historical migration among populations. With the application of a molecular clock, we can get a general sense of when lineages diverged, providing a timeframe for evolution.

This dissertation is a progression of methods for explaining biological diversity using molecular tools. The East Asian region is used as a study area; birds and mammals are used as study organisms.

Chapter One deals with delimiting population lineages in nature. Before diversity can be explained, it must first be characterized. In this chapter I used a combination of morphological and genetic data to determine the number of lineages contained within a closely related group of birds, the Light-vented/Taiwan Bulbul (*Pycnonotus sinensis*/*P.*

taivanus) complex. I find that it is necessary to use both morphological and mitochondrial DNA characters to delimit all four lineages within this complex. I also employed a suite of 13 nuclear intron sequences in conjunction with a new coalescent-based approach for species delimitation to attempt to validate the four lineages. This produced mixed results. Most lineages were supported by the coalescent-based analysis, but the approach could not distinguish the most morphologically distinct taxon. This is probably the result of recent evolution whereby selection has driven morphological differences to fixation faster than the neutral nuclear markers have coalesced through drift. The coalescent-based species delimitation method was also prone to false positives, likely owing to population substructure within lineages. Therefore, I concluded that fixed character differences, rather than statistical differences, are the most reliable means for delimiting lineages. Further, when lineage divergence is recent, a combination of morphological and molecular markers may be required to uncover all lineages. This study provides the taxonomic framework for my subsequent analyses of evolutionary processes.

Chapter Two is a phylogeographic study of a single species. Before we can get a general sense of the generalities of evolutionary patterns, we must produce the single-species studies that contribute to later comparative studies. In this chapter, I used mitochondrial DNA sequences and nuclear microsatellites to investigate the evolutionary history of Steere's Liocichla (*Liocichla steerii*), a montane bird species endemic to the island of Taiwan. Although many non-volant vertebrates distributed in the Taiwanese mountains show a phylogeographic division between the northern and southern mountains, I found no such pattern in Steere's Liocichla. Although I found some

evidence that gene flow is channeled through altitudinal corridors, populations across the island appear connected by gene flow. There was strong evidence of a recent population expansion. However, I found no genetic evidence of refugial areas from which the current population might have expanded. To further address this issue, I constructed ecological niche models, which predict the distribution of an organism using its affinity for certain abiotic conditions (e.g. temperate and precipitation). The ecological niche model was a good predication of the current distribution of Steere's Liocichla. I then used the abiotic conditions at the Last Glacial Maximum to reconstruct the past distribution of Steere's Liocichla. I found that, while Steere's Liocichla occupied the same general area that it does today, it was probably at a lower population density. Thus, I conclude that the genetic signature of demographic expansion was produced as Steere's Liocichla went from a low population density at the Last Glacial Maximum to its current high population density. This study is a valuable demonstration of how phylogeography and ecological niche modeling can reciprocally illuminate aspects of evolutionary history.

Chapter Three is a comparative study of the phylogeographic histories of multiple co-distributed species. In this chapter, I used mitochondrial sequence data from 24 Japanese land mammals to construct the colonization history of the main Japanese Islands. The Japanese main islands were originally part of the continent until sea-floor spreading made them islands around 15 million years ago. Since then, the islands have been episodically connected to the mainland by landbridges, which formed during periods of low sea-level. Divergence time estimates using a molecular clock indicated mammalian diversification predated the initial isolation of Japan as an island, which is

consistent with the landbridge colonization hypothesis. However, my study implicated a colonization timeframe that is significantly older than the current paradigm based on fossil evidence. Further, half of the divergence time estimates in the Hondo region were clumped around 2.4 million years ago. This suggests a potential dramatic interchange period between Japan and the mainland that is concordant with significant global cooling, a period when the first landbridge connection is likely. This study exemplifies the kinds of generalities that emerge when the evolutionary histories of multiple species are compared. In this case, a period around 2.4 million years ago appears to have been an important time of interchange between Japan and mainland Asia.

Chapter 1

The Challenge of Delimiting Recent Lineages: The Light-vented/Taiwan Bulbul (*Pycnonotus sinensis*/*P. taivanus*) Complex as a Case Study

McKay, B. D., H. L. Mays Jr., Y. Wu, H. Li, C.-T. Yao, I. Nishiumi, and F. Zou (submitted)

Abstract

Species delimitation has recently undergone renewed interest in systematics. Many systematists equate the theoretical concept of species with metapopulation lineages. However, operationally delimiting recently diverged lineages is a challenge because character concordance is expected to be low. We applied morphological and mitochondrial DNA character datasets to develop lineage hypotheses based on fixed character differences in a closely related avian complex, the Light-vented/Taiwan Bulbul (*Pycnonotus sinensis*/*P. taivanus*) complex. Groups identified by the morphological and mitochondrial datasets differed but could be combined to form a reconciled hypothesis of four evolutionary lineages. We attempted to validate lineage hypotheses with sequences from 13 nuclear loci using a new coalescent-based species delimitation method implemented in the program BPP. BPP failed to identify a distinct lineage diagnosable by several morphological characters, possibly as a result of recent selectively driven divergence or gene flow. We also found that BPP can produce high speciation probabilities for arbitrarily defined geographic clusters, suggesting the method may oversplit species if there is population subdivision. We conclude that coalescent-based species delimitation methods can potentially lump morphologically distinct taxa that should be split and can potentially split geographic clusters that should be lumped, presumably due to population substructure within lineages. Therefore, we suggest that these methods be used with caution and, at minimum, interpreted in combination with independent data.

Introduction

Species delimitation, the process of discovering species and determining their boundaries, has recently undergone renewed interest in systematics (Sites and Marshall 2003; Wiens 2007). When delimiting species, it is useful to distinguish between the theoretical concept of species and the operational criteria used for recognizing species (Frost and Kluge 1994; Mayden 1997). De Queiroz (1998; 2007) has argued that all species concepts, either implicitly or explicitly, equate species with metapopulation lineages, and that the major differences among species concepts involve which criteria a lineage must evolve before it is afforded species status. However, the lineage concept of species (i.e. the evolutionary species concept) does not specify operational criteria for delimiting historical lineages (de Queiroz 2007; Mayden 1997).

Many authors have advocated using fixed character differences between populations as evidence the populations do not exchange genes and thus have distinct histories (Cracraft 1983; Nelson and Platnick 1981; Nixon and Wheeler 1990). The boundaries of suitable characters for lineage delimitation should correspond to lineage boundaries. However, evolutionary processes (e.g. selection, introgression, sex-biased gene flow) can create geographic clustering of some characters within a broadly tokogenetic metapopulation. Thus, because single characters can reflect lineage boundaries imperfectly, concordant patterns among multiple independent characters strengthen hypotheses about lineage limits (Avise and Ball 1990). Concordance among characters is expected to increase as lineages diverge, but concordance may be rare when

divergence is recent (Avise 2000). This makes diagnosing recently diverged lineages an expected challenge.

The challenge of delimiting recent lineages is determining which (if any) seemingly fixed characters correctly reflect actual lineage boundaries. Recently, coalescent-based species delimitation methods have been proposed that might help address this problem. Knowles and Carstens (2007) have argued that multiple nuclear loci contain information about lineage limits before fixed nuclear characters evolve. Thus, they suggest that lineage limits suggested from independent data, such as morphology or mtDNA, can be validated using multiple nuclear loci and a coalescent-based framework (Ence and Carstens 2010; Knowles and Carstens 2007). Several different implementations of coalescent-based species delimitation have recently been proposed (Ence and Carstens 2010; O'Meara 2010; Yang and Rannala 2010). However, few empirical studies have applied these methods (see Leaché and Fujita 2010), so it is still unclear how they will perform with real data. Simulations suggests these methods may oversplit lineages when population subdivision is present (O'Meara 2010). This is a potentially serious problem because population subdivision, in various degrees, is likely common in many species. Here we use morphological and mtDNA character datasets to generate lineage hypotheses in a recently evolved avian complex that we then attempt to validate with neutral nuclear loci and a coalescent-based species delimitation method.

The Light-vented/Taiwan Bulbul (*Pycnonotus sinensis*/*P. taivanus*) complex consists of two closely related species distributed in East Asia (Fig. 1). Both species prefer open woodlands and forest edge habitat below 1500 m. The Light-vented Bulbul

(*Pycnonotus sinensis*) is divided into four subspecies (Fishpool and Tobias 2005). *Pycnonotus s. sinensis* is found throughout southeast China. *Pycnonotus s. hainanus* is found on Hainan Island, southern China, and northern Vietnam; it is distinguished from the nominate race by its distinctive all-black nape. *Pycnonotus s. formosae* is distributed in western Taiwan; it is reported to have lighter yellow abdominal streaks and more white in the nape than the nominate race (Yamasaki 2006). *Pycnonotus s. orii* is found in the southern Ryukyu islands from Yonaguni Island to Okinawa Island and was reported in its description to have darker abdominal streaks than the nominate race. The differences distinguishing *P. s. orii* as well as the taxonomic status of this subspecies have been questioned (Yamasaki 2006). The Taiwan Bulbul (*Pycnonotus taivanus*) is monotypic and endemic to southeast Taiwan. It has a restricted range and is considered by the IUCN as a vulnerable and declining species (Butchart et al. 2008), making the assessment of its taxonomic status a conservation priority.

Methods

Sampling and laboratory methods

We collected a total of 87 tissue samples from the Light-vented/Taiwan Bulbul complex (Table 1). Samples are distributed among 18 localities and include all five taxa currently recognized in this complex (Fig. 1). At least two localities were sampled from each taxon.

Whole genomic DNA was extracted using a standard phenol-chloroform protocol followed by ethanol precipitation. The mitochondrial NADH dehydrogenase subunit 2 (ND2) gene (871 bp) was amplified from all samples using the primers L5216 and H6313 (Sorenson et al. 1999). We chose a geographically diverse subset of individuals from each taxon for multi-locus sequencing. This included five individuals from each taxon except *P. s. orii*, which had three individuals. From this subset of 23 individuals, we sequenced thirteen nuclear introns (Table 2) from primers described by Backström et al. (2008) as well as the entire mtDNA ND2 gene (1041 bp) using the primers L5216 and H6313. Two individuals of the Brown-breasted Bulbul (*Pycnonotus xanthorrhous*), the purported sister taxon to the *Pycnonotus sinensis*/*P. taivanus* complex (Oliveros and Moyle 2010), were sequenced for all loci and used as an outgroup.

Polymerase chain reaction (PCR) was performed in 10 µl reactions with Fidelitaq™ Master Mix (Affymetrix/USB) using a thermal profile of 94°C for 4 min followed by 30 cycles of 1 min at 94°C, 1 min at 50°C, and 2 min at 72°C, and then a final extension cycle for 10 min at 72°C. Primers and excess dNTPs were inactivated with ExoSAP-IT (Affymetrix/USB) following the manufacturer's instructions. Sequencing reactions were performed on ExoSAP-IT treated PCR products using BigDye® v. 3.1 Cycle Sequencing Kit (Life Technologies/Applied Biosystems). Unincorporated BigDye® terminators were removed from sequencing reactions using the BigDye Xterminator® kit (Life Technologies/Applied Biosystems) and cleaned sequencing products were separated on a 3130 Genetic Analyzer (Life Technologies/Applied Biosystems). PCR amplicons were sequenced in both directions.

Complementary strands were aligned and edited using Geneious v. 5 (Drummond et al. 2010). Chromatograms were inspected individually and every point mutation was checked for authenticity. The phase of nuclear alleles was determined computationally using the PHASE 2.1 (Stephens et al. 2001) algorithms implemented in DnaSP v. 5 (Librado and Rozas 2009).

We looked for evidence of recombination using the Φ_w-statistic (Bruen et al. 2006) implemented in the program SplitsTree v. 4.10 (Huson and Bryant 2006), and tested for selection using the Hudson-Kreitman-Aguade (HKA) test (Hudson et al. 1987) implemented in DnaSP. We selected the best-fit model of nucleotide substitution for each locus using Akaike information criteria (AIC) in the program jModeltest v. 0.1.1 (Posada 2008). We used PAUP* v. 4.0b10 (Swofford 2003) to generate maximum likelihood gene trees and tested the molecular clock hypothesis for each locus using a likelihood ratio test.

Phylogenetic analysis

Because species-level paraphyly is common in birds (McKay and Zink 2010), we tested the monophyly of the Light-vented and Taiwan Bulbul using two approaches, one that treats individuals as operational taxonomic units (OTUs) and another that treats traditional subspecies as OTUs. We first constructed a maximum-likelihood phylogeny of ND2 sequences using the program PHYML (Guindon and Gascuel 2003). Clade stability was estimated with 100 replicates of nonparametric bootstrapping. We carried out

statistical parsimony analysis of mtDNA ND2 sequences using the program TCS v. 1.21 (Clement et al. 2000). This method partitions haplotypes into independent networks connected by mutations that are non-homoplasious with 95% probability.

We then used the five currently recognized taxa of the Light-vented/Taiwan Bulbul complex as OTUs in a species tree estimation approach implemented in the program *BEAST v. 1.6.1 (Drummond and Rambaut 2007; Heled and Drummond 2010). *BEAST uses a Bayesian Markov chain Monte Carlo (MCMC) method to jointly estimate multiple gene trees embedded in a shared species tree. We ran MCMC chains for 100 million generations (sampling every 10,000 generations and discarding the first 10% as burn-in) and assessed convergence of the MCMC chain in the program Tracer v. 1.5 (Rambaut and Drummond 2007). We conducted two *BEAST analyses, one with all 14 loci and another with only the 13 nuclear loci.

Population aggregation analysis

We used population aggregation analysis (PAA)(Davis and Nixon 1992) to find fixed character differences among populations. PAA begins by summarizing character states for all sampled individuals within a population. Population profiles of character states are compared and populations without fixed differences are combined and the process is iterated until the only remaining population aggregates differ from each other by at least one fixed character. Note that this method assumes samples are representative of populations as a whole (Wiens and Servedio 2000).

We applied PAA to both morphological and molecular character datasets. The morphological dataset included four plumage characters reported to vary within the Light-vented/Taiwan Bulbul complex (Fishpool and Tobias 2005; Yamasaki 2006). The molecular dataset included as characters 871 bp of the mtDNA ND2 gene. As Brower (1999) has shown, the application of PAA to nucleotide bases may produce misleading results in the presence of character homoplasy. Thus, we also applied cladistic haplotype aggregation (CHA) to our molecular dataset (Brower 1999).

Coalescent-based species delimitation

We used the coalescent-based species delimitation method implemented in the program Bayesian Phylogenetics and Phylogeography (BPP v. 2.0) (Rannala and Yang 2003; Yang and Rannala 2010) using our 13 nuclear loci. We preferred BPP over other coalescent-based species delimitation methods because it is currently the only one that incorporates gene tree uncertainly. BPP uses Bayesian Markov chain Monte Carlo (MCMC) algorithms that accommodate the species phylogeny as well as coalescent processes in extant and extinct ancestral species. The method assumes no post-divergence gene flow. A user-specified guide tree is used to reduce the number of possible species delimitations the program must integrate over. At each node on the tree the program assesses whether the sequence data are compatible with a one-species model, or whether a two-species model has to be invoked to explain the data.

BPP incorporates a model that includes the species divergence times (τ), and the population size parameters $\theta = 4N\mu$, where N is the effective population size and μ is the mutation rate per site per generation. For our θ prior, we used a mean of 0.006, which was the average Watterson's θ for all 13 nuclear loci estimated using DnaSP. For our τ_0 prior, we used a mean of 0.0004. This was determined using the equation $\tau = t\mu$, where t is the root age, in generations. For t, we used 100,000, which was the in-group divergence time estimated from a *BEAST analysis that assumed an ND2 mutation rate of 2.76×10^{-8} per site per year (Drovetski et al. 2004) and a generation time of one year. For μ, we used 3.6×10^{-9} per site per year, which was the average μ of all 13 nuclear loci. Individual nuclear locus mutation rates were estimated relative to the ND2 mutation rate in *BEAST.

We used both informative G(10, 1667) and diffuse G(2, 333) gamma priors on the population size parameters (θs) and both informative G(10, 25000) and diffuse G(2, 5000) gamma priors on the age of the root (τ_0), while the other divergence time parameters were assigned the Dirichlet prior (Yang and Rannala 2010). Following the program's recommendations, we ran the species delimitation model using both algorithm 0 (with ε values of 2, 5, 10, and 20) and algorithm 1 (with α values of 1, 1.5, and 2, and m values of 0.5, 1, and 2). We ran the program for 100,000 generations, sampling every five generations, and discarding the first 50,000 generations as burn-in. Each analysis was run at least twice to confirm consistency between runs.

We used the species tree topology estimated from *BEAST (Fig. 2c) as our guide tree. We attempted to validate both a four-lineage hypothesis generated from the

population aggregation analysis as well as a five-lineage hypothesis that assumes all five traditional subspecies are lineages. A simulation study has suggested that coalescent-based species tree delimitation methods may over split species in the presence of population subdivision (O'Meara 2010). Therefore, we tested the effects of artificially splitting peripheral populations within *sinensis* (Chongqing, Miluo, Guilin, and Zhuhai split from Langxi), *hainanus* (Datian split from Fanjia), and *taivanus* (Hwalien split from Taitung). We compared this to splitting these taxa into two equal proportions of random individuals. For all these analyses, we also tested the effects of using fewer loci. In addition to using all 13 nuclear loci, we ran analyses using five random loci (15463, 22187, 25442, 27331, and 27818) and again with just one random locus (27818).

Results

Sequences and phylogenetic analyses

We collected 8,099 bp of nuclear intron sequence and 1041 bp of the mtDNA (ND2) sequence (9,140 bp combined nuclear and mtDNA sequence per individual sample). There were no missing data. No locus exhibited evidence of recombination or selection. The molecular clock was rejected for the ND2 gene, but not for any of the nuclear loci. Individual locus information is available in Table 2.

Phylogenetic analyses rejected monophyly of the Light-vented Bulbul and suggested that the Taiwan Bulbul is embedded within the Light-vented Bulbul (Fig. 2).

Maximum-likelihood phylogenetic analysis of ND2 sequences did not recover any of the five taxa as monophyletic (Fig. 2a). The ND2 haplotype network consisted of two geographic clusters: one comprising individuals from China and Hainan (subspecies *sinensis* and *hainanus*) and the other comprising individuals from Taiwan and the southern Ryukyus (subspecies *formosae*, *orii*, and *taivanus*) (Fig. 2b). The *BEAST species tree analysis recovered *sinensis* and *hainanus* as sister taxa and a clade uniting *formosae* and *orii* as sister to *taivanus* (Fig. 2c). The same topology was produced whether or not the mitochondrial ND2 gene was included in the analysis.

Population aggregation analysis

Population aggregation analysis of morphology and mtDNA character sets produced different population aggregates. However, no individual population showed any fixed differences, and the only groups that did show fixed differences were subspecies or groups of subspecies. That is, PAA did not find distinct groups of populations within any subspecies and no distinct groups identified by PAA spanned two different subspecies (Table 3). The morphological dataset aggregated samples into three groups: one containing *sinensis*, *formosae*, and *orii*; one containing *hainanus*; and one containing *taivanus* (Fig. 3). The mtDNA nucleotide dataset aggregated samples into two groups: one containing *sinensis* and *hainanus*, and one containing *formosae*, *orii*, and *taivanus* (Fig. 3). Population aggregation analysis and cladistic haplotype aggregation both identified the same three fixed differences among aggregates. Thus, there is no evidence

of the kind of homoplasy described by Brower (1999) in our molecular dataset. In addition, there were no fixed base substitutions in the 8,099 bp of nuclear sequence. Thus, collecting nuclear sequence data for all individuals would not have resulted in any fixed differences among subspecies. The group hypotheses generated by the two datasets are not in conflict because the datasets can be combined and reconciled to form a hypothesis of four groups with aggregates than can be diagnosed with characters or combinations of characters from the two datasets (Fig. 3).

Coalescent-based species delimitation

Coalescent-based species delimitation results from BPP are shown in Fig. 4. Analyses using algorithm 0 with various values of ε and analyses using algorithm 1 with various values of α and m produced similar results. Results were also similar whether we used informative or diffuse priors. We report results from runs using algorithm 0 with a ε value of 5 and informative priors.

Analysis of the four-lineage hypotheses derived from population aggregation analysis suggested the model with *sinensis* and *hainanus* split, but *formosae* and *taivanus* lumped, as the most probable (Fig. 4a). In this model, the probability of splitting *formosae* and *taivanus* was low (0.32). Analysis of the five-lineage hypotheses derived from traditional taxa suggested the most probable model was the one with all five taxa split (Fig. 4e). In this model, the probability of splitting *formosae*/*orii* and taivanus was

much higher (0.87) than the probability of splitting *formosae* and *taivanus* in the four-lineage analysis. The probability of splitting *formosae* and *orii* was low (0.55).

In the five-lineage analysis, the most probable model split all five lineages, so this might suggest that *formosae* and *orii* are separate lineages despite not having any known fixed character differences. However, coalescent-based species delimitation may over-split when lineages contain population substructure (O'Meara 2010). Because *orii* could also be considered a peripheral population of *formosae*, we investigated the effects of dividing *sinensis*, *hainanus*, and *taivanus* into geographic clusters (Fig. 4 b-d). Surprisingly, this resulted in high speciation probabilities for some of the geographic clusters. The node probability for splitting the two geographic populations of *hainanus* into two species was quite high (0.92) and exceeded the node probability of splitting *taivanus*, which is very distinct morphologically. The probability of splitting two geographic clusters of *sinensis* was also high (0.82). The probability of splitting the two populations of *taivanus* was low (0.45) and comparable to the probability of splitting *formosae* and *orii*. For comparison, we examined the effects of dividing *sinensis*, *hainanus*, and *taivanus* into two groups with individuals randomly assigned to a group (Fig. 4 f-h). As expected, this resulted in a most probable model with all nodes split except the two random groups. In each case the node probability for splitting the two random groups was very low (0.12-0.30).

Reducing the number of loci resulted in a decrease in speciation probabilities in most cases (Fig. 4). However, reducing the number of loci did not seem to affect the randomly split groups as much as the geographic clusters, and often the speciation

probabilities of the randomly split groups decreased with increasing loci (Fig. 4). Reducing the number of loci had no effect on the speciation probability of the basal split between the *sinensis*/*hainanus* clade and the *formosae*/*orii*/*taivanus* clade, which was 1.0 in every analysis.

Discussion

The objective of this study was to discover identifiable lineages within the Light-vented/Taiwan Bulbul complex. Our operational criteria for diagnosing lineages were fixed character differences. Aggregations of populations with fixed character differences identified by the mtDNA and morphological datasets were different: mtDNA data suggested two groups, whereas morphological data suggested three groups. These results are not necessarily in conflict because group hypotheses generated from each dataset can be combined to form a reconciled hypothesis of four groups. Datasets are expected to differ when divergence is recent (Avise 2000); however, unless one or more characters conflict with the boundaries of the underlying lineage history, it should be possible to reconcile them. Therefore, in contrast to reports of incongruence between species identified with mtDNA and morphology (Wiens and Penkrot 2002), we do not consider our mtDNA and morphological datasets incongruent.

Phylogenetic analyses rejected monophyly of *P. sinensis* as it is currently defined, and suggested that *P. taivanus* is embedded within *P. sinensis*. This is interesting in light of the morphological differences within this complex because it means that *hainanus* and

taivanus have probably recently acquired their distinctive plumage differences whereas *sinensis* and *formosae*/*orii* have retained the ancestral plumage condition. These plumage characters may be under directional selection, which has driven them to fixation before mtDNA haplotypes have sorted though drift. An intriguing possibility is that these plumage characters function in mate recognition or are in some way driven by sexual selection, a process proposed as being an important speciation mechanism in birds (Price 2008; Zink 1996).

Coalescent-based species delimitation

We attempted to validate lineage hypotheses using coalescent-based species delimitation, but this produced mixed results. Splitting *sinensis*/*hainanus* from *formosae*/*orii*/*taivanus* was robustly supported by a speciation probability of 1.0 in all analyses. Splitting *sinensis* from *hainanus* was also strongly supported (≥0.95) by our 13 loci analyses. Leaché and Fujita (2010) suggest speciation probabilities of ≥0.95 as being strong support for species.

In contrast, splitting *taivanus* and *formosae* received low support in the four-lineage analysis (0.32). It received higher support in the five-lineage analysis (0.87), but this still fell short of Leaché and Fujita's 0.95 probability cut-off. There are a couple of reasons why coalescent-based species delimitation may fail to validate real lineages. One reason, as suggested by Knowles and Carstens (2007), is that if lineage divergence is selectively driven, neutral nuclear markers might not contain adequate information about

this divergence, in which case coalescent-based species delimitation methods will tend to be conservative (i.e. fail to recognize real lineages). Another possibility is that there is gene flow between two historical lineages. None of the currently available coalescent-based species delimitation methods, including BPP, account for gene flow, and it has been shown through simulations that gene flow can cause BPP to "lump" lineages (Zhang et al. in press). Yang and Rannala (2010) claim that BPP delimits biological species because it does not include gene flow in its model. However, tree-based species delimitation methods are inherently incompatible with the biological species concept because the ability to interbreed is a retained ancestral trait (Rosen 1979). For example, BPP strongly suggests speciation has occurred between *sinensis* and *formosae*, but these taxa are not considered "good" biological species. Gene flow is known between *taivanus* and *formosae* (Hsu and Lin 1993), so this, in combination with recent (possibly selectively driven) divergence, might explain why BPP assigns such a low speciation probability to these taxa.

Though the speciation probability for splitting *orii* and *formosae* was relatively low (0.55), the most probable model in the five-lineage analysis split all nodes (Fig. 4e). This might suggest that *orii* is a separate lineage that has recently diverged and not yet obtained fixed character differences. However, simulations have suggested that coalescent-based species delimitation methods may falsely split populations if there is population subdivision within lineages (O'Meara 2010), though simulations involving BPP did not oversplit lineages under a stepping-stone model with low migration (Zhang et al. in press). Therefore, we assessed the effects of splitting geographic clusters within

sinensis, *hainanus*, and *taivanus*. Surprisingly, splitting geographic clusters resulted in high speciation probabilities in some cases. Dividing our two sampled populations of *hainanus* on Hainan Island, which are separated by approximately 200 km, resulted in a speciation probability of 0.92 in our 13 loci analysis. This speciation probability is approaching 0.95. The speciation probability of the two *sinensis* clusters was also relatively high (0.82). This is in contrast to the speciation probabilities obtained from dividing individuals randomly between two groups, which were relatively low. The models implemented in BPP are nested statistical hypotheses where the one-species model can be considered the null model and the two-species model is the alternative model. The null model of one species invokes random mating. It might be that coalescent-based species delimitation methods, at least as they are currently implemented, are sensitive to departures from non-random mating, especially when many loci are used. Thus, as the authors warn (Yang and Rannala 2010), caution should be used when interpreting results, which can be integrated with other sources of information, such as morphology, behavior, and ecology. Leaché and Fujita (2010) have formally described new species based only on BPP speciation probabilities under the assumption that these probabilities are conservative (see also Bauer et al. 2011; Fujita and Leaché 2011). However, rather than representing conservative estimates of lineages, coalescent-based species delimitation results may sometimes constitute upper limits to the number of lineages. In other cases, when divergence is recent or when there is gene flow, these methods might not be able to identity lineages.

Lineage delimitation

We conclude that there is congruent evidence for four lineages within the Light-vented/Taiwan Bulbul complex. These are the same four groups identified by the combined morphology + mtDNA population aggregation analyses (Fig. 3). First, the basal split between *sinensis*/*hainanus* and *formosae*/*orii*/*taivanus* is suggested by mtDNA and corroborated by the species tree topology suggested by *BEAST analysis using only nuclear loci as well as by a robustly supported speciation probability of 1.0 from BPP. Splitting *sinensis* and *hainanus* is suggested by a single morphological character (nape color), but this character is corroborated by high (≥0.95) speciation probabilities from BPP, indicating that nape color likely reflects historical isolation. Finally, splitting *taivanus* from *formosae*/*orii* is not strongly supported by molecular data. However, *taivanus* is the most distinctive form, morphologically, and differs from *formosae* in multiple, seemingly independent, plumage characters. Thus, we conclude that *taivanus* and *formosae*/*orii* are separate historical lineages that have separated too recently for their divergence to be detected by the neutral nuclear markers we used.

This study demonstrates the value of including multiple kinds of characters, especially when dealing with recently diverged lineages. Had we only employed mtDNA, we would have concluded that there were two lineages in the Light-vented/Taiwan Bulbul complex. In contrast, had we only looked at morphological characters, we would have concluded that there were only three lineages.

Taxonomic implications

Given that population lineages are fundamental to all species concepts (de Queiroz 1998), we think it is useful to first determine the smallest diagnosable lineages (phylogenetic species *sensu* Cracraft 1983) within a complex and then consider whether those lineages meet the criteria for species status under various species concepts. Diagnostic taxa are important as units in conservation (Cracraft 1997), comparative biology (Eldredge and Cracraft 1980), and biogeography (Cracraft 1989), so if they fail to meet the qualifications of a species under some species concept, they should still be recognized as subspecies under that concept. For example, if evidence suggests diagnostic lineages are reproductively compatible, these lineages could be classified as subspecies under the biological species concept. In fact, redefining the subspecies category to only include diagnostic lineages, as others have suggested (Remsen 2005), would bring a level of rigor to the subspecies concept that is currently lacking (Wilson and Brown 1953; Zink 2004).

The four lineages we identify here meet the qualifications for evolutionary and phylogenetic species. Determining their status as biological species requires considering whether there are intrinsic barriers to reproduction. Although hybridization between *P. s. formosae* and *P. taivanus* is pervasive (Hsu and Lin 1993), *P. taivanus* is distinctive enough in morphology that it is still recognized as a biological species (Clements 2007). *P. s. hainanus* is easily identified by plumage, and IMa analysis indicates little historical gene flow between it and *P. s. sinensis* (B. D. McKay et al. unpublished). Therefore, *P. s. hainanus* could be considered a biological species, though studies of its contact zone with

P. s. sinensis would clarify this determination. Finally, we found no character evidence to support the continued recognition of *P. s. orii* as a taxon. This agrees with a recent quantitative morphological study that found *P. s. orii* indistinctive (Yamasaki 2006). Thus, we suggest *P. s. orii* be subsumed under *P. s. formosae*.

Tables

Table 1.1. Museum voucher information for all samples of the Light-vented/Taiwan Bulbul (*Pycnonotus sinensis*/*P. taivanus*) complex used in this study. Acronyms: South China Institute for Endangered Animals (SCIEA); Taiwan Endemic Species Research Institute (TESRI); American Museum of Natural History (AMNH); National Museum of Nature and Science, Tokyo (NSMT); University of Kansas Natural History Museum (KUNHM).

Catalog number	Taxon	Country	Locality 1	Locality 2
SCIEA S00036	*sinensis*	China	Macao	Macao
SCIEA S00040*	*sinensis*	China	Chongqing	Chongqing
SCIEA S00042	*sinensis*	China	Macao	Macao
SCIEA S00043	*sinensis*	China	Macao	Macao
SCIEA S00045	*sinensis*	China	Macao	Macao
SCIEA S00046	*sinensis*	China	Macao	Macao
SCIEA S00047	*sinensis*	China	Macao	Macao
SCIEA S00067	*sinensis*	China	Macao	Macao
SCIEA S00145*	*sinensis*	China	Anhui	Langxi
SCIEA S00146	*sinensis*	China	Anhui	Langxi
SCIEA S00147	*sinensis*	China	Anhui	Langxi
SCIEA S00148	*sinensis*	China	Anhui	Langxi
SCIEA S00150	*sinensis*	China	Anhui	Langxi
SCIEA S00221	*sinensis*	China	Hunan	Miluo
SCIEA S00227	*sinensis*	China	Hunan	Miluo
SCIEA S00228*	*sinensis*	China	Hunan	Miluo
SCIEA S00229	*sinensis*	China	Hunan	Miluo
SCIEA S00230	*sinensis*	China	Hunan	Miluo
SCIEA S00231	*sinensis*	China	Hunan	Miluo
SCIEA S00232	*sinensis*	China	Hunan	Miluo
SCIEA S00262	*sinensis*	China	Guangxi	Guilin
SCIEA S00264*	*sinensis*	China	Guangxi	Guilin
SCIEA S00266	*sinensis*	China	Guangxi	Guilin
SCIEA S00268	*sinensis*	China	Guangxi	Guilin
SCIEA S00755*	*sinensis*	China	Guangdong	Zhuhai

SCIEA S00759	*sinensis*	China	Guangdong	Zhuhai
SCIEA S00760	*sinensis*	China	Guangdong	Zhuhai
SCIEA S00761	*sinensis*	China	Guangdong	Zhuhai
SCIEA S00763	*sinensis*	China	Guangdong	Zhuhai
SCIEA S01565	*sinensis*	China	Guangdong	Shixing
SCIEA S01566	*sinensis*	China	Guangdong	Shixing
SCIEA S01567	*sinensis*	China	Guangdong	Shixing
SCIEA S01568	*sinensis*	China	Guangdong	Shixing
SCIEA S01878	*sinensis*	China	Guangxi	Jinxiu
SCIEA S01880	*sinensis*	China	Guangxi	Jinxiu
SCIEA S01881	*sinensis*	China	Guangxi	Jinxiu
SCIEA S01882	*sinensis*	China	Guangxi	Jinxiu
SCIEA S01886	*sinensis*	China	Guangxi	Jinxiu
SCIEA S01887	*sinensis*	China	Guangxi	Jinxiu
SCIEA S00012*	*hainanus*	China	Hainan	Datian
SCIEA S00158	*hainanus*	China	Hainan	Datian
SCIEA S00162*	*hainanus*	China	Hainan	Datian
SCIEA S00165*	*hainanus*	China	Hainan	Datian
SCIEA S00167*	*hainanus*	China	Hainan	Datian
SCIEA S00168	*hainanus*	China	Hainan	Datian
SCIEA S00174	*hainanus*	China	Hainan	Fanjia
SCIEA S00184	*hainanus*	China	Hainan	Fanjia
SCIEA S00274*	*hainanus*	China	Hainan	Fanjia
T5454	*formosae*	Taiwan	Tainan	Tainan
T5455	*formosae*	Taiwan	Tainan	Tainan
T5457	*formosae*	Taiwan	Tainan	Tainan
T5458	*formosae*	Taiwan	Tainan	Tainan
T5459	*formosae*	Taiwan	Tainan	Tainan
T5460	*formosae*	Taiwan	Tainan	Tainan
T5462	*formosae*	Taiwan	Tainan	Tainan
T5463	*formosae*	Taiwan	Tainan	Tainan
T5465	*formosae*	Taiwan	Tainan	Tainan
T5466	*formosae*	Taiwan	Tainan	Tainan
T5467	*formosae*	Taiwan	Tainan	Tainan
T5468	*formosae*	Taiwan	Tainan	Tainan
TESRI 3606	*formosae*	Taiwan	Taichung	Taichung
TESRI 3680*	*formosae*	Taiwan	Nantou	Chi-Chi
TESRI 3870	*formosae*	Taiwan	Taichung	Taichung
TESRI 3919	*formosae*	Taiwan	Taichung	Taichung
TESRI 6301*	*formosae*	Taiwan	Taichung	Taichung
TESRI 6324*	*formosae*	Taiwan	Kaohsiung	Kaohsiung
TESRI 6380	*formosae*	Taiwan	Nantou	Chi-Chi
TESRI 6445*	*formosae*	Taiwan	Kaohsiung	Kaohsiung
TESRI 6486	*formosae*	Taiwan	Nantou	Chi-Chi

AMNH DOT5236	*formosae*	Taiwan	Nantou	Chi-Chi
AMNH DOT5237*	*formosae*	Taiwan	Nantou	Chi-Chi
AMNH DOT5238	*formosae*	Taiwan	Nantou	Chi-Chi
AMNH DOT5239	*formosae*	Taiwan	Nantou	Chi-Chi
AMNH DOT5240	*formosae*	Taiwan	Nantou	Chi-Chi
AMNH DOT5241	*formosae*	Taiwan	Nantou	Chi-Chi
TESRI 2261	*taivanus*	Taiwan	Taitung	Taitung
TESRI 2440	*taivanus*	Taiwan	Taitung	Taitung
TESRI 2456*	*taivanus*	Taiwan	Taitung	Taitung
TESRI 2666*	*taivanus*	Taiwan	Taitung	Taitung
TESRI 2667	*taivanus*	Taiwan	Taitung	Taitung
TESRI 4211	*taivanus*	Taiwan	Hwalien	Hwalien
TESRI 4443*	*taivanus*	Taiwan	Hwalien	Hwalien
TESRI 5042*	*taivanus*	Taiwan	Hwalien	Hwalien
TESRI 5169*	*taivanus*	Taiwan	Taitung	Taitung
NSMT YA86*	*orii*	Japan	Okinawa Prefecture	Yonaguni Island
NSMT 2008.95*	*orii*	Japan	Okinawa Prefecture	Okinawa Island
NSMT YA36*	*orii*	Japan	Okinawa Prefecture	Okinawa Island
KUNHM 13691	*xanthorrhous*			
KUNHM 13728	*xanthorrhous*			

* indicates individual selected for multi-locus sequencing

Table 1.2. Descriptive statistics for the 14 molecular loci used in this study.

locus	inheritance	length (bp)	alleles	*s*	PI sites	π (x 10^{-3})	θ (x 10^{-3})
ND2	mtDNA	1041	19	29	14	4.94	7.55
12630	autosome	471	12	15	8	2.82	7.25
13403	autosome	751	17	16	9	3.39	4.85
14572	autosome	722	13	17	13	4.27	5.36
15463	autosome	699	18	20	14	4.38	6.84
17898	autosome	593	16	17	15	5.95	6.52
22187	autosome	514	14	11	6	2.84	4.87
23361	autosome	504	8	7	4	1.81	3.16
24972	autosome	687	23	24	18	6.48	7.95
25442	autosome	641	19	16	11	4.28	6.03
25613	autosome	484	11	10	6	2.21	4.70
27331	autosome	801	20	23	11	3.43	6.53
27818	autosome	632	16	18	12	3.78	6.48
MUSK	Z-linked	600	10	12	10	2.95	4.55

The summary includes the length of the sequences (in base pairs), the number of alleles, the number of segregating sites (*s*), the number of parsimony-informative sites (PI sites), nucleotide diversity (π), and Watterson's theta (θ).

Table 1.3. Population and subspecies profiles representing absence, fixed presence, and nonfixed presence of seven attributes among 18 populations of the Light-vented/Taiwan Bulbul (*Pycnonotus sinensis*/*P. taivanus*) complex. Numbers in the parentheses are sample sizes.

	Biological attributes						
	1	2	3	4	5	6	7
sinensis	1	1	1	0	G	T	C
Guilin (4)	1	1	1	0	G	T	C
Jinxiu (6)	1	1	1	0	G	T	C
Langxi (5)	1	1	1	0	G	T	C
Macao (7)	1	1	1	0	G	T	C
Miluo (7)	1	1	1	0	G	T	C
Shixing (4)	1	1	1	0	G	T	C
Zhuhai (5)	1	1	1	0	G	T	C
Chongqing (1)	1	1	1	0	G	T	C
hainanus	1	1	0	0	G	T	C
Datian (6)	1	1	0	0	G	T	C
Fanjia (3)	1	1	0	0	G	T	C
formosae	1	±	1	0	A	C	T
Kaohsiung (2)	1	1	1	0	A	C	T
Nantou (9)	1	±	1	0	A	C	T
Taichung (4)	1	1	1	0	A	C	T
Tainan (12)	1	±	1	0	A	C	T
orii	1	1	1	0	A	C	T
Okinawa I. (2)	1	1	1	0	A	C	T
Yonaguni I. (1)	1	1	1	0	A	C	T
taivanus	0	0	0	1	A	C	T
Hwalien (4)	0	0	0	1	A	C	T
Taitung (6)	0	0	0	1	A	C	T

Characters: 1=presence of chest band, 2=presence of flank streaks, 3=presence of white in nape, 4=presence of red spot on malar, 5=base position 396 of the mtDNA ND2 gene, 6=base position 649 of the mtDNA ND2 gene, 7=base position 696 of the mtDNA ND2 gene. Note: 1 = fixed presence; 2 = absence; ± = nonfixed presence (polymorphic).

Figures

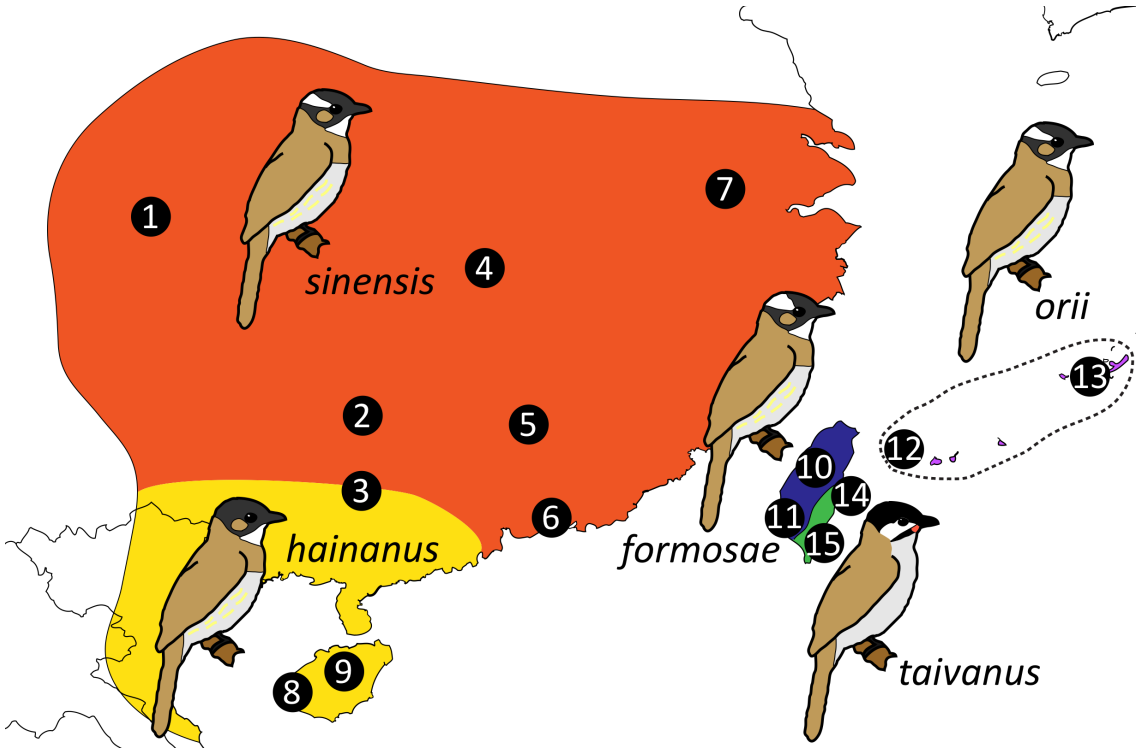


Figure 1.1. Map of East Asia showing the distributions of the five taxa that comprise the Light-vented/Taiwan Bulbul (*Pycnonotus sinensis*/*P. taivanus*) complex. The ranges of the different taxa are colored: *P. s. sinensis*, red; *P. s. hainanus*, yellow; *P. s. formosae*, blue; *P. s. orii*, purple (and outlined by dotted line); *P. taivanus*, green. Black circles represent sampling localities: 1, Chongqing; 2, Guilin; 3, Jinxiu; 4, Miluo; 5, Shixing; 6, Macao & Zhuhai; 7, Langxi; 8, Datian; 9, Fanjia; 10, Taichung & Chi-Chi; 11, Tainan & Kaohsiung; 12, Yonaguni I.; 13, Okinawa I.; 14, Hwalien; 15, Taitung.

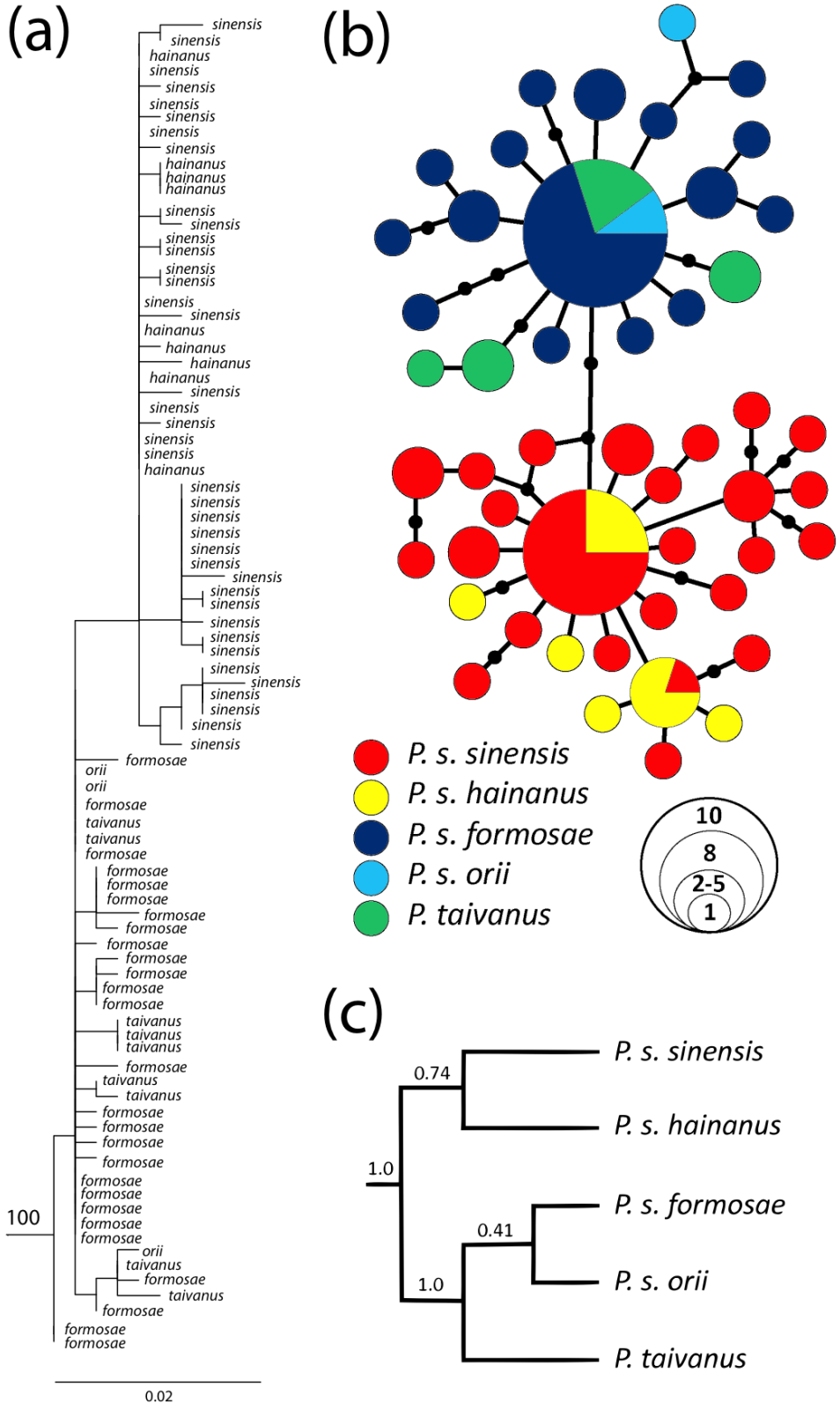

(a)
sinensis
sinensis
hainanus
sinensis
sinensis
sinensis
sinensis
sinensis
sinensis
hainanus
hainanus
hainanus
sinensis
sinensis
sinensis
sinensis
sinensis
sinensis
sinensis
hainanus
hainanus
hainanus
hainanus
sinensis
sinensis
sinensis
sinensis
sinensis
sinensis
sinensis
sinensis
sinensis
sinensis
sinensis
sinensis
sinensis
sinensis
sinensis
sinensis
sinensis
sinensis
formosae
orii
orii
formosae
taivanus
taivanus
formosae
formosae
formosae
formosae
formosae
formosae
formosae
formosae
formosae
formosae
formosae
taivanus
taivanus
taivanus
formosae
taivanus
taivanus
formosae
formosae
formosae
formosae
formosae
formosae
formosae
formosae
formosae
formosae
orii
taivanus
formosae
taivanus
formosae
formosae
formosae
100
0.02
(b)
P. s. sinensis
P. s. hainanus
P. s. formosae
P. s. orii
P. taivanus
10
8
2-5
1
(c)
P. s. sinensis
P. s. hainanus
P. s. formosae
P. s. orii
P. taivanus
0.74
1.0
0.41
1.0

Figure 1.2. (a) Maximum-likelihood analysis of ND2 sequences. No taxa were reciprocally monophyletic, but a clade containing *sinensis* and *hainanus* was embedded within individuals from the other three taxa. Branch lengths are proportional to sequence divergence. Only bootstrap values >75 are shown. (b) Statistical parsimony network of ND2 sequences showing two distinct geographic clusters. Each circle represents a haplotype, and the size of the circle is proportional to that haplotype's frequency. Small black circles represent inferred haplotypes. Colors distinguish taxa. (c) Species tree inferred from *BEAST with posterior probabilities of clade support. Note all phylogenetic analysis support *P. taivanus* as embedded within *P. sinensis*.

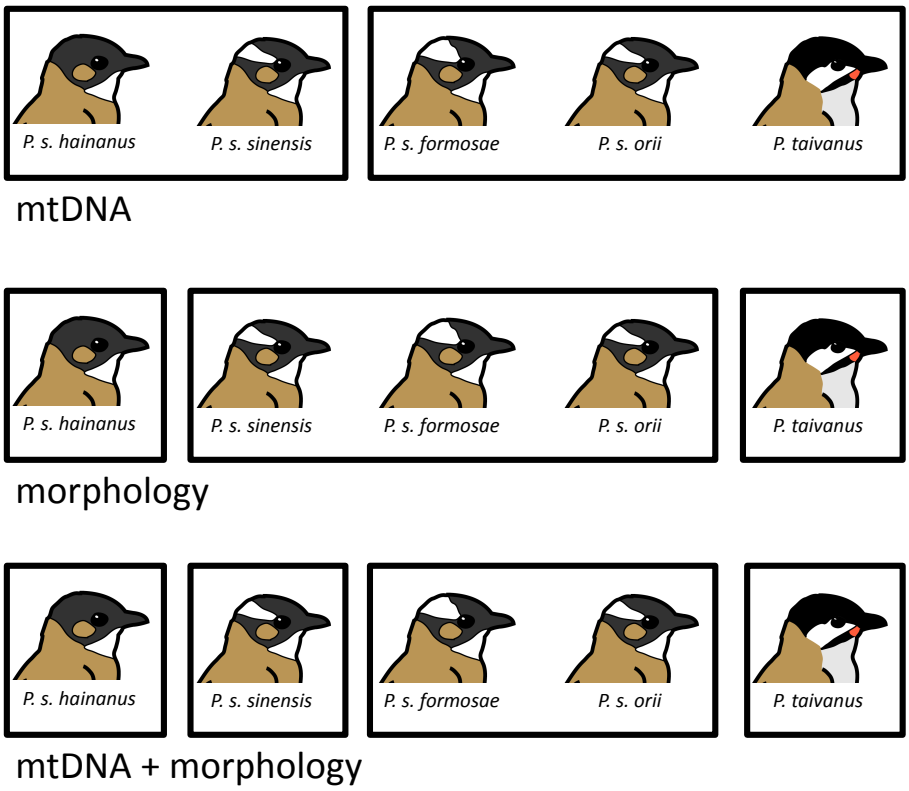
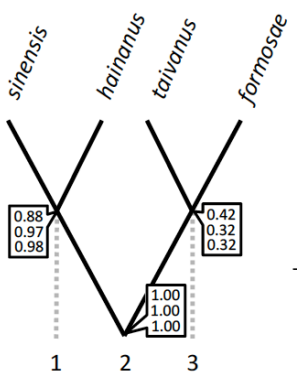


Figure 1.3. Results of population aggregation analysis. The mtDNA dataset suggests two groups. The morphological dataset suggests three groups. The combined dataset suggests a reconciled hypothesis of four groups. Note that of the five traditional taxa recognized in this complex, all were found to show fixed characters (or combinations of characters) except *P. s. orii*, which was indistinguishable from *P. s. formosae*.

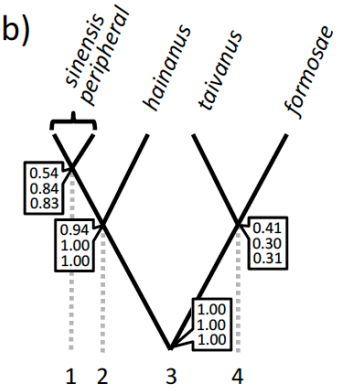
(a)



Tree frequencies

123	Prob.
110	0.66
111	0.32
010	0.01

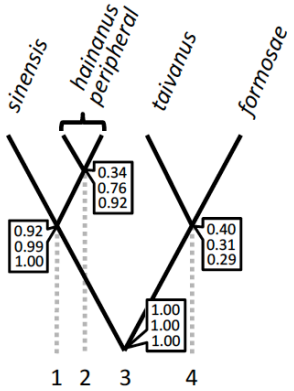
(b)



Tree frequencies

1234	Prob.
1110	0.60
1111	0.25
0110	0.11

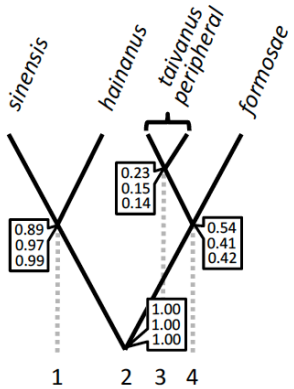
(c)



Tree frequencies

1234	Prob.
1110	0.65
1111	0.27
1010	0.05

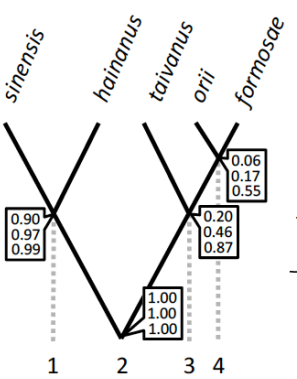
(d)



Tree frequencies

1234	Prob.
1100	0.57
1101	0.27
1111	0.14

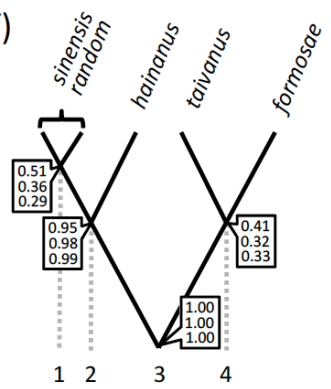
(e)



Tree frequencies

1234	Prob.
1111	0.55
1110	0.31
1100	0.13

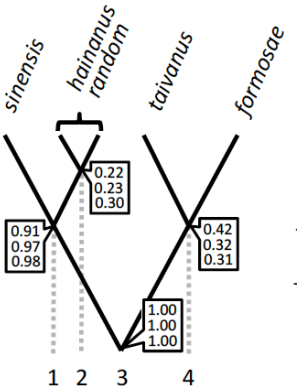
(f)



Tree frequencies

1234	Prob.
0110	0.47
0111	0.23
1110	0.19

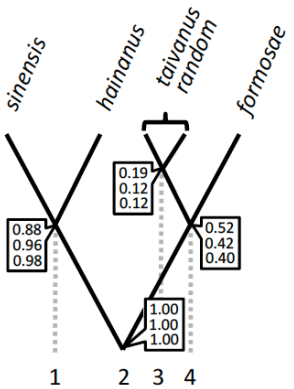
(g)



Tree frequencies

1234	Prob.
1010	0.47
1011	0.21
1111	0.09

(h)



Tree frequencies

1234	Prob.
1100	0.59
1101	0.28
1111	0.12

Figure 1.4. Results from a coalescent-based species delimitation analysis of the Light-vented/Taiwan Bulbul (*Pycnonotus sinensis*/*P. taivanus*) complex. (a) Results from the four lineage hypothesis generated from the combined morphology + molecular character dataset. (e) Results from splitting all five traditional taxa. (b-d) Results from over-splitting the taxa *sinensis*, *hainanus*, and *taivanus* into geographic clusters. (f-h) Results from splitting the *sinensis*, *hainanus*, and *taivanus* into two groups with individuals randomly assigned to groups. The numbers in the boxes associated with each node are the posterior probabilities that the node should be split into two lineages. The top number is from the one-locus analysis, the middle number is from the five-locus analysis, and the bottom number is from the full 13-locus analysis. Posterior probabilities for splitting random groups of individuals are low, but can be quite high for splitting random geographic clusters. The tree frequencies show the top three models suggested from the 13-locus analysis. Note: "1s" and "0s" correspond to the numbered nodes in the tree; "1" represents that node as being resolved (split), whereas "0" represents that node as being collapsed (lumped).

Chapter 2

Recent range-wide demographic expansion in a Taiwan endemic montane bird, Steere's Liocichla (*Liocichla steerii*)

McKay, B. D., H. L. Mays Jr., Y.-W. Peng, K. H. Kozak, C.-T. Yao, H.-W. Yuan, P.-F. Lee, and F.-H. Hsu. 2010. *BMC Evol. Biol.* 10: 71.

Abstract

Climatic cycles have had profound impacts on the genetic diversity of a wide range of taxa. Most studies have focused on temperate areas despite the fact that most biodiversity occurs in tropical areas. The subtropical island of Taiwan is an area of high endemism and a complex topographic environment. We used mitochondrial DNA sequences and microsatellites to investigate the evolutionary history of an endemic montane bird, Steere's Liocichla (*Liocichla steerii*). We found no evidence of the deep genetic divisions reported for other taxa in Taiwan. Our results instead suggest that populations of Steere's Liocichla are connected by gene flow and that this species recently experienced a demographic expansion. Genetic diversity is high and widespread, suggesting expansion from multiple areas rather than a few isolated refugia. Ecological niche models suggested that populations of Steere's Liocichla are connected by climatically suitable habitats and that there was less suitable habitat during the Last Glacial Maximum. Taken together, genetic and ecological niche modeling data corroborate a single history--Steere's Liocichla was at lower density during the Last Glacial Maximum and has subsequently expanded its population. We comment on the conservation implications of our data and urge caution when interpreting results from highly polymorphic markers such as microsatellites.

Introduction

Past climatic cycles have had a profound impact on the levels and distribution of genetic diversity (Hewitt 2000). Despite the fact that most biodiversity occurs in tropical areas, most studies have focused on the temperate regions in North America and Europe where the retreat of ice sheets has left widespread genetic patterns consistent with northward expansion from southern refugia (Hewitt 2004). This imbalanced focus has left us with a much poorer understanding of the impact of climate cycles on the phylogeographical patterns of tropical taxa. However, initial phylogeographic studies indicate that tropical areas are historically complex and contain deep intraspecific lineages (Hewitt 2004; Moritz et al. 2000).

The subtropical island of Taiwan is an area of high endemism as well as topologic, ecologic, and climatic diversity. Taiwan first emerged from the sea following the collision of the Eurasian and Philippine Sea plates approximately four mya (Hsu 1990). Despite periodic connections with the mainland during periods of low sea level, Taiwan's general isolation from mainland Asia has given rise to many endemic species. Within the island, mountains have provided additional opportunities for isolation. This is reflected in the phylogeographic patterns of some lowland taxa. A freshwater crab (*Candidiopotamon rathbunae*, Shih et al. 2006), a frog (*Rana limnocharis*, Toda et al. 1998), and a gecko (*Gekko hokouensis*, Tsai 1999) exhibit a common pattern of genetic division east and west of the Central Mountain Range (CMR), the prominent mountain range that runs almost the entire longitudinal length of the island. In contrast, many

montane species show a pattern of north-south differentiation: Formosan Wood Mouse (*Apodemus semotus*, Hsu et al. 2001), a Ranid frog (*Rana sauteri*, Chen 1994), and Siberian Weasel (*Mustela sibirica*, Wu 2004). The observation of common phylogeographic patterns in such varied taxa suggests vicariance as the cause (Zink et al. 2000).

The east-west pattern of genetic division has been found in lowland birds (Light-vented and Taiwan Bulbul *Pycnonotus sinensis* and *P. taivanus*, Hsu 1999; Ring-necked Pheasant *Phasianus colchicus*, Chen et al. 2004), however, no montane bird has yet been phylogeographically surveyed. Determining whether montane birds were affected by the same isolating events that generated genetic divisions the other montane taxa is important for understanding the impact of mountain vicariance in Taiwan. In a conservation context, uncovering historical lineages is essential for properly guiding management efforts, especially for species confined to single islands. Steere's Liocichla (*Liocichla steerii*) is a montane, resident bird species endemic to Taiwan. It inhabits forest edge and secondary growth from approximately 1000 m to 3000 m. Its dispersal ability was inferred to be low (Huang 2004), and long distance flight has never been observed in this species (Tsai 2005). High song diversity also suggests dispersal may be low (Tsai 2005). Therefore, Steere's Liocichla is an ideal avian candidate to have developed the same genetic structure observed in potentially less vagile vertebrates.

In this paper, we used mitochondrial DNA sequences and nuclear microsatellites to investigate the evolutionary history of Steere's Liocichla. Our primary goal was to determine whether this endemic species showed the same phylogeographic patterns of

deep divergence that characterize other montane vertebrates in Taiwan. As a supplement to our genetic data, we used ecological niche modeling to predict this species' range during the Last Glacial Maximum (LGM). We also considered whether any populations of Steere's Liocichla might be of special management concern for conservation efforts.

Methods

Sampling and laboratory methods

A total of 122 individuals were sampled from 13 populations located throughout the breeding range of the species (Fig. 1; Table 1). Birds were caught during the breeding season (March to September) by mist net and 50µl of blood was collected via brachial venipuncture. Blood samples were stored in lysis buffer (Seutin et al. 1991) at room temperature until they could be transported to the laboratory where they were stored at -20°C.

Whole genomic DNA was extracted using a standard phenol-chloroform protocol followed by ethanol wash. DNA was re-suspended in 1 X TE (0.01M Tris, 0.001M EDTA, pH 8.0) and stored at -20°C. The mitochondrial NADH dehydrogenase subunit 2 (ND2) gene was amplified from a subset of 91 individuals representing all populations (Table 1) using the primers L5216 and H6313 (Sorenson et al. 1999). Polymerase chain reaction (PCR) was performed in 10 µl reactions on a MJ Research PTC-100 thermocycler using a thermal profile of 94°C for 4 min followed by 30 cycles of 1 min at

94°C, 1 min at 55°C, and 2 min at 72°C, and then 10 min at 72°C. Primers and excess dNTPs were removed from the PCR product with ExoSAP-IT (USB Corporation) following the manufacturer's instructions. Sequencing was performed on an ABI 3700 automated sequencer using BigDye kit v. 3.0 according to recommended protocols (Applied Biosystems). All products were sequenced in both directions. Usually, the amplification primers were sufficient to obtain >1,000 base pairs (bp) of unambiguous sequence. However, whenever an ambiguous stretch of sequence was produced, internal sequencing primers (L5758 and H5766; Sorenson et al. 1999) were used to clarify ambiguous bases. Complementary strands were aligned and edited using SEQUENCHER v. 4.6 (GeneCodes Corporation). All sequences were inspected individually using the raw spectrograph data and every point mutation was checked for authenticity. Sequences have been deposited in GenBank (GU560065-560155).

All individuals were also genotyped at seven microsatellite loci. Four loci were developed specifically for Steere's Liocichla (lsgata17, lsgata 21, lsgata 24, lsgata 25; Yeung et al. 2004) and three additional loci (GC-GATA10, GC-GATA 15, GC-GATA 23; Huang et al. 2004) were originally developed for another babbler, the Hwamei (*Garrulax canorus*). Forward primers were fluorescently labeled, and the PCR cycling protocol followed the above for all loci except GC-GATA 15, which used an annealing temperature of 50°C. Fragment analysis of the amplification product was conducted on a Beckman CEQ 8000 sequencer (Beckman Coulter), and genotypes were scored by eye.

Genetic data analysis

For mtDNA sequences, we estimated nucleotide diversity (π), number of haplotypes (nh), and haplotype diversity (h) for each population using the program DnaSP v. 4.9 (Rozas et al. 2003). For each population, we also computed Fu's F_S, a neutrality statistic that is particularly sensitive to population demographic expansion (Fu 1997). Its significance was tested with 10,000 coalescent simulations using the program ARLEQUIN 3.11 (Excoffier et al. 2005). To further test for demographic expansion, we compared mtDNA mismatch distributions with expectations of a sudden-expansion model (Rogers 1995) using DnaSP. Populations that have experienced a sudden demographic expansion are expected to show a unimodel mismatch distribution (Slatkin & Hudson 1991). We tested overall genetic structure of populations with a hierarchical analysis of molecular variance (AMOVA; Excoffier et al. 1992) as implemented in ARLEQUIN. To test for restricted gene flow between the north and south groups identified by previous studies (e.g. Hsu et al. 2001), we assigned populations to either a "north" or "south" group. Based on the study of Hsu et al. (2001), we used the Choushui River as the boundary between groups (see Table 1). To visualize haplotype relationships, we constructed a haplotype network using the parsimony-based algorithm of Templeton et al. (1992) and implemented in the program TCS 1.21 (Clement et al. 2000). Networks are generally better than bifurcating trees at representing the relationships of intraspecific phylogenies (Posada & Crandall 2001).

Microsatellite genotypes were tested for departures from Hardy-Weinberg equilibrium and evidence for linkage disequilibrium using ARLEQUIN. To assess allelic variation within populations, we calculated allele frequencies, observed (Ho) and expected (He) heterozygosities using GENEPOP v. 4.0.9 (Rousset 2008). To correct for sample size differences, we calculated allelic richness using rarefaction with the program FSTAT 2.9.3 (Goudet 2001). A sequential Bonferroni correction was applied to correct for multiple comparisons (Rice 1989). Principal components analysis (PCA) is method of simplifying multivariate data with a minimum loss of information. It can be applied to allele frequency data as a measure of strong multilocus associations among populations. We performed a PCA with arcsin-square root transformed allelic frequencies using SAS (SAS Institute).

To assess genetic structure between populations, we computed pairwise F_{ST} and R_{ST} values for all populations. R_{ST} is analogous to F_{ST} but incorporates a stepwise mutation model of evolution that was specifically designed for microsatellite data (Slatkin 1995). Next, we performed another hierarchical AMOVA to determine the proportions of microsatellite variation among and within populations. Groups were defined in the same way as for the mtDNA AMOVA. Fixation indices and AMOVA were computed in ARLEQUIN. We also inferred genetic population structure using a Bayesian assignment approach implemented in the program BAPS 5.2 (Corander & Marttinen 2006; Corander et al. 2008). BAPS has been shown to perform better than STRUCTURE when F_{ST} is small (Latch et al. 2006). We began with a mixture analysis to determine the most probable number of populations (K) given the data. We performed the

"clustering of groups of individuals" analysis, considering clusters to contain a minimum of three individuals. We estimated the number and composition of clusters considering upper limits of *K* at 2-10, 15, and 20. We ran 10 repetitions for each maximum *K*. Using the groups identified in the mixture analysis, we conducted an admixture analysis using 100 realizations from the posterior of the allele frequencies. We repeated the admixture five times to confirm consistent results. We also performed a "spatial clustering of groups" analysis that incorporates geographical information in the estimation of the genetic population structure (Corander et al. 2008). This approach provides more statistical power to correctly infer structure when molecular data is weak. The spatial mixture analysis followed the above procedure.

Finally, to test for an association between genetic distance (R_{ST}) and geographic distance, we performed a Mantel test using GenAlEx 6 (Peakall & Smouse 2006). We ran two Mantel tests with different geographic distances. The first used "straight-line" distances, which were simply the shortest distances between populations. However, Steere's Liocichla has a restricted altitudinal range and is therefore unlikely to traverse very high or very low altitudes. Thus, the second Mantel test used "least-cost" distances, which were calculated as the shortest distances between populations assuming travel was limited to elevations between 1000m and 2500m. Straight-line and least-cost geographic distances were calculated in ArcGIS 9 (ESRI).

Ecological niche modeling

We used ecological niche modeling (ENM) to independently evaluate our molecular analyses of demographic history and population structure. Briefly, the general ENM approach applied here generates a map of the expected distribution of a species using data on the environmental conditions where it is known to occur and randomly selected background locations in the study area (Kozak et al. 2008). We used Maxent version 3.2 (Phillips et al. 2006; Phillips et al. 2004) to model the potential geographic distribution of Steere's Liochicla. Maxent is a general approach for characterizing probability distributions from incomplete information (Phillips et al. 2006). Maxent computes a probability distribution that describes the relative suitability of each grid cell as a function of the environmental variables at all the known occurrence locations, and when projected into geographic space it produces a map of the species' potential geographic distribution.

To construct the model, we used 19 bioclimatic variables from the WorldClim dataset with 2.5-minute spatial resolution (Hijmans et al. 2005) and 286 georeferenced breeding-season localities for Steere's Liocichla. The climatic variables are derived from weather station data and quantify annual variation in temperature and precipitation. Occurrence records occurring within the same map pixel were removed to avoid pseudoreplication. To calibrate the model, we used quadratic features, and default parameters for the number of background pixels, regularization, the convergence threshold, and the maximum number of iterations. We randomly selected 75% of the

occurrence locations to construct the model; the remaining 25% were set aside to test the model. We calculated the area under the receiving operator characteristic (AUC) to test whether the model could discriminate between the test localities and 10,000 localities randomly selected from across Taiwan. Finally, we selected logistic output, which ranges from 0-1 and quantify the probability of suitable environmental conditions for the species in each grid cell.

To estimate the geographic distribution and extent of suitable climate for Steere's Liocichla at the LGM (21,000 years before present), we applied the contemporary ecological niche model to a set 19 bioclimatic variables generated from CCSM3 paleoclimatic model. In doing so, we used the fade-by-clamp option in Maxent version 3.2 to remove heavily clamped pixels from the final prediction. Failing to remove such pixels from the region might lead to erroneously extensive predictions of suitable climate when transferring a model to a different time period.

We caution that distribution-based niche modeling makes at least three important, biological assumptions. First, it assumes that a species distribution is at equilibrium with the present-day climate conditions. Second, it assumes that the distribution of the species is not strongly influence by biotic factors (e.g. competitors or predators). Third, it assumes that the climatic tolerances of the species are conserved over the timescale for which its distribution is being modeled. If any of these conditions is violated, the extent of climatically suitable habitat could be underestimated.

Results

mtDNA

A total of 1026 bp were obtained from the mitochondrial ND2 gene. There were 48 polymorphic and 20 parsimony informative sites resulting in 36 unique haplotypes. Nuclear copies of mitochondrial genes are known in birds (Sorenson & Quinn 1998). However, the large number of observed haplotypes and an absence of nonsensical stop-codons support a mitochondrial origin for our sequences (Zhang & Hewitt 1996). Overall nucleotide diversity (π) was 0.00272 for all samples combined and ranged from 0.00143 to 0.00446 (Table 1). Overall haplotype diversity was 0.9 and ranged from 0.69 to 1.00 (Table 1). AMOVA indicated non-significant differentiation (1%) between north and south groups and significant population differentiation, with 4.5% ($P < 0.001$) of the molecular variance partitioned among populations. The mtDNA mismatch distribution had a mean of 2.8 nucleotide substitutions, was distinctly unimodel, and did not differ significantly from a model of sudden expansion (Fig. 2). The haplotype network showed no obvious genetic structure (Fig. 3).

Microsatellites

The microsatellite dataset showed no statistically significant deviations from Hardy-Weinberg equilibrium or evidence of lineage disequilibrium. There were a total of 137

alleles, and the average number of alleles per loci was 19.6. The number of alleles per locus ranged from 51 in lsgata17 to 9 in lsgata 21. Allelic richness ranged from 4.00 in Aowanda to 7.57 in Liyuan (Table 1). PCA resulted in relatively little explanatory power with only 15.6% and 12.6% of the variation explained by the first and second principal components, respectively. It took seven principal components to explain 75% of the variation. This suggests weak multilocus associations among populations.

Overall, our analyses of population structure gave conflicting results with different populations highlighted by different analyses. F_{ST} pairwise population comparisons resulted in significant differences between the Beidongyenshan population and every other population (Table 2). R_{ST} pairwise indices were larger that F_{ST} comparisons (Table 2). This is expected because F_{ST} estimated from highly variable markers is known to bias estimated downwards. This is because the proportion of the total variation among populations can never be very high when there is a lot of variation within populations (Hedrick 1999). R_{ST} comparisons did not highlight the Beidongyenshan population as particularly distinct. Rather, there were several significant comparisons that seem haphazard with respect to geography (Table 2). AMOVA of microsatellite alleles indicted non-significant amounts of differentiation (0%) distributed between north and south groups and little but significant population differentiation with 1.8% ($P < 0.001$) of the variance distributed among populations.

The mixture analysis in BAPS indicated the presence of two distinct groups of populations, with the Meifeng population the sole member of one group and all other populations comprising the other group. The admixture analysis correctly assigned most

individual to groups (Fig. 4). In contrast, the spatial analysis that incorporated geographic information about populations suggested only a single group (Fig. 4). The Mantel test indicated a positive but non-significant correlation between genetic (R_{ST}) and geographic distances (km) when straight-line geographic distances were used ($R^2 = 0.01$; $P = 0.23$). However, when least-cost geographic distances were considered, the Mantel test resulted in a positive and significant correlation ($R^2 = 0.07$; $P = 0.04$). This suggests that dispersal for these birds is limited by altitude.

Ecological niche modeling

The predicted geographic distribution of Steere's Liocichla for the present-day and LGM are shown in Figure 5. The area under the receiving operator characteristic demonstrated that the model discriminates strongly between randomly selected locations across Taiwan, and the training (AUC = 0.93) and test localities (AUC = 0.97).

Intriguingly, despite the many assumptions that are made when constructing distribution-based niche models, both the contemporary and LGM predictions are consistent with the findings of the molecular analyses. The model predicts that populations of Steere's Liocichla are connected by climatically suitable habitats (Fig. 5A). This result corroborates the analyses of population structure, which suggest that the degree of gene exchange is related to the geographic distance separating populations and not strong barriers to dispersal. In addition, both the geographic extent and relative suitability of habitats are predicted to have been reduced at the LGM in comparison to the

present (Fig. 5B). The latter result parallels the demographic analyses, which suggest that populations may have increased in size in response to geographic expansion of suitable habitats across Taiwan.

Discussion

Evolutionary history

While we found high levels of genetic variation within populations, these results demonstrate a lack of genetic differentiation among populations. This is in contrast to the deep genetic divisions reported for other montane species in Taiwan. The haplotype network showed no obvious structure with respect to geography. The AMOVA indicated no genetic differentiation in either mtDNA or microsatellites between north and south groups. However, significant amounts of differentiation were partitioned among populations, implying that populations are not panmictic and that dispersal between populations is somewhat constrained. This was true of both mtDNA and microsatellites, though the percent of variation among populations was higher for mtDNA. This is expected because mtDNA has, on average, a four times more rapid time to coalescence than nuclear loci (Moore 1995; Zink & Barrowclough 2008). A lack of genetic structure has been found in other mountain species in Taiwan (e.g. White-bellied Rat *Niviventer culturatus*, Hsu et al. 2000 and Taiwan Alpine Skink *Sphenomorphus taiwanensis*, Guo 2002), and this might result from a recent colonization of Taiwan or historical gene flow

between populations. We favor a scenario of historical gene flow because Steere's Liocichla is an endemic species that's purported closest relative (*L. omeiensis*; Collar & Robson 2007) was estimated to have diverged about 5 million years ago (Luo et al. 2008). This suggests that either Steere's Liocichla has existed on Taiwan for a long time or that it has recently gone extinct elsewhere. The idea that populations of Steere's Liocichla have been historically well-connected by gene flow seems plausible given the dispersal potential of volant birds, and the geographic continuity of climatically suitable habitats over time (Fig. 5).

We also find evidence for a recent demographic expansion in Steere's Liocichla. For example, Fu's F_S, which is particularly sensitive to departures from population equilibrium (Fu 1997), was negative for most populations and significantly negative when all samples were combined. The distinctly unimodal mismatch distribution did not differ significantly from a model of sudden expansion, and the average number of pairwise differences was low, indicating the expansion was relatively recent. High haplotype diversity, low nucleotide diversity, and the star-like haplotype network are also consistent with a recent population expansion (Avise 2000). Ecological niche modeling indicated that there was less suitable habitat for Steere's Liocichla during the LGM. However, the marginal habitat seems to have been distributed across the island. This might suggest that Steere's Liocichla occupied much of its current range during the LGM, but populations were less dense. Thus, as habitat improved with the warming climate, population densities increased leaving a pattern of recent demographic expansion. This scenario is consistent with the genetic data, which show similar levels of

genetic diversity across the island rather than isolated pockets of unusually high diversity that might be interpreted as refugia. It would be interesting to determine whether other montane bird species in Taiwan exhibit similar patterns.

Conservation implications

Taken separately, the results of some of our analyses could be used as evidence that some populations should be of special management concern for conservation efforts. For example, we found significant differences in microsatellite F_{ST} between the Beidongyenshan population and all other populations. This is surprising because there is no reason to expect restricted gene flow to or from the Beidongyenshan population; it is centrally located, only two km from the Meifeng population, and there is adequate habitat and a continuous population of Steere's Liocichla between Beidongyenshan and Meifeng. However, caution must be taken when interpreting F_{ST} distances associated with highly polymorphic loci. Hedrick (1999) has shown that relatively weak bottlenecks can cause statistically significant differences in microsatellite F_{ST} distances. There is evidence that the Beidongyenshan population has recently undergone a population bottleneck and this would explain why it appears as if it had a high genetic distance from the other populations. First, the Beidongyenshan population had the lowest amount of allelic richness. Second, it had a high observed heterozygosity (H_O) relative to its expected heterozygosity (H_E), as indicated by the smaller than zero, and nearly significant, F_{IS} value (Table 1; Note that the AW population has a smaller F_{IS} value, but this likely had a

smaller affect on F_{ST} significance because the sample size is much smaller.). A large H_O relative to H_E is expected shortly after a population bottleneck because rare alleles are quickly lost, resulting in a decrease in the H_E. At the same time, H_O levels are relatively unaffected until a new mutation-drift equilibrium can be reached because H_O is mostly dictated by common alleles. Such a bottleneck may have been caused by anthropogenic disturbance from construction of an agricultural research station at Beidongyenshan, which is where we collected our samples. A change in F_{ST} caused by a bottleneck is likely ephemeral, existing only until a new mutation-drift equilibrium is reached. Thus, individuals sampled from Beidongyenshan in the future may not show significant F_{ST} differentiation from other populations. We note that in such a situation, F_{ST} may actually show biologically relevant differences between populations, but this differentiation might not result from restricted gene flow and might be ephemeral and not of management concern. Clearly, caution should be taken before making conservation recommendations based on the results of F_{ST} alone.

Because it considers information on the length of alleles, R_{ST} is generally thought to perform better than F_{ST} with microsatellite data, especially when there are many alleles. In contrast to F_{ST} results, R_{ST} comparisons did not highlight the Beidongyenshan population as particularly distinct. Instead, the few significant values of R_{ST} do not seem to correlate with geography. For example, there is a significant R_{ST} value between Meifeng and Beidongyenshan, which again are only two km apart, but no significant differences between Meifeng and Moshan or Chilan, which are much farther apart. Overall, significant R_{ST} values seem haphazard with respect to geography. Again, the

statistical power associated with highly variable markers, such as microsatellites, is high, so statistically significant differences should not always be taken to mean biologically significant differences (Hedrick 1999). It should also be noted that statically non-random patterns may result from the stochastic nature of population structure (Ball et al. 1990).

Bayesian analysis of population structure using BAPS found that including the Meifeng population in the pool of remaining populations decreased the fit of a panmictic total, and thus two groups were identified. Again, this is surprising because Meifeng is only two km from Beidongyenshan and there is adequate habitat and a continuous population of Steere's Liocichla between these two populations. However, the spatial analysis in BAPS, which incorporates information about the geographic location of populations, suggested that all populations represented one group. If separating the Meifeng population in the original BAPS analysis was a borderline case, it is understandable that the spatial model suggests only a single population (J. Corander, pers. comm.). This was probably caused by an excess of private alleles in the Meifeng population. Possibly as the result of sampling error, Meifeng had an inordinate number of private alleles and all sampled Meifeng individuals had at least one private allele. This would explain why the admixture analysis was able to accurately assign Meifeng individuals to population. Some authors have suggested private alleles are important for conservation. For example, Funk et al. (2007) defined subspecies for conservation and included the criteria that they contain "unique alleles or haplotypes." We are not sure about the practicality of this particular criterion because with highly polymorphic loci, most, if not all, populations are likely to contain at least some unique alleles. In our study,

all 13 populations we sampled had unique alleles. We question the relevance of private alleles when deciding conservation priority or taxonomic decisions.

Overall, it appears populations of Steere's Liocichla across Taiwan have been historically well-connected by gene flow. Results of least-cost Mantel tests, however, indicate that dispersal has been at least somewhat restricted by altitude. Separately, some of our analyses might seem to indicate biologically relevant differentiation of a population. However, different analyses disagree about which populations are distinct. Therefore, we are cautious in recommending any single population be considered of conservation concern. We urge conservation workers to be very careful when interpreting results from highly polymorphic markers and when recommending populations as special management concern based on such markers.

Tables

Table 2.1. Intrapopulation statistics for each Steere's Liocichla population separately and for all samples combined. The summary of genetic diversity includes sample size (n) nucleotide diversity (π), number of haplotypes (nh), haplotype diversity (h), Fu's F_S (F_S), average number of alleles per locus (A), allelic richness (A_R), observed heterozygosity (H_O), expected heterozygosity (H_E), and inbreeding coefficient (F_{IS}). Microsatellite statistics are based on the average of seven polymorphic loci.

			mtDNA					microsatellites					
Population			n	π	nh	h	F_S	n	A	A_R	H_O	H_E	F_{IS}
Aowanda	AW	South	3	0.00260	3	1.00	-0.34	4	4.00	4.29	0.86	0.73	-0.171
Beidongyanshan	BS	North	9	0.00184	4	0.69	0.27	11	5.43	4.00	0.80	0.72	-0.121
Bilu	BL	North	7	0.00446	7	1.00	-3.23*	11	6.14	4.42	0.70	0.80	0.119
Chilan	CL	North	3	0.00195	3	1.00	-0.69	5	4.00	4.10	0.80	0.75	-0.062
Chitou	CT	South	9	0.00309	5	0.81	0.17	9	5.43	4.48	0.74	0.80	0.073
Danda	DD	South	9	0.00309	6	0.92	-1.03	12	6.29	4.26	0.72	0.77	0.060
Guanwu	GW	North	9	0.00271	6	0.89	-1.35	11	6.00	4.48	0.67	0.79	0.149

Liyuan	LI	South	10	0.00143	5	0.76	-1.32	13	7.57	4.81	0.84	0.82	-0.034
Mingchi	MC	North	7	0.00912	6	0.95	-2.20*	10	5.43	4.47	0.85	0.80	-0.070
Meifeng	MF	North	10	0.00236	7	0.87	-2.67*	10	6.71	4.80	0.81	0.83	0.015
Moshan	MS	South	8	0.00164	5	0.79	-1.54	9	5.43	4.42	0.75	0.79	0.050
Nanxi	NX	South	1	n/a	n/a	n/a	n/a	7	5.14	4.39	0.60	0.80	0.245
Taipingshan	TP	North	6	0.00286	5	0.93	-1.33	10	6.57	4.65	0.74	0.81	0.087
Total			91	0.00272	36	0.90	-26.6**	122	5.70	4.43	0.76	0.79	0.026

* $P < 0.05$; ** $P < 0.001$

Table 2.2. Analyses of molecular variance (AMOVAs) showing the distribution of genetic variation among populations of Steere's Liocichla on Taiwan.

Source of variation	mtDNA				microsatellites			
	df	% total	*F*-statistic	*P*-value	df	% total	*F*-statistic	*P*-value
Among groups (north v. south)	1	0.98	0.010	0.140	1	-0.20	-0.002	0.713
Among populations within groups	11	4.48	0.045	0.011	11	1.92	0.019	<0.001
Within populations	78	94.54	0.055	0.003	231	98.28	0.017	<0.001

Table 2.3. Pairwise estimates of F_{ST} (below diagonal) and R_{ST} (above diagonal). Bolding indicates a significant statistic.

	AW	BS	BL	CL	CT	DD	GW	LI	MC	MF	MS	NX	TP
AW		0.063	0.000	0.028	**0.325**	0.018	0.135	**0.363**	0.000	0.251	0.175	0.210	0.081
BS	**0.062**		0.039	0.000	**0.202**	0.000	0.047	**0.265**	0.045	**0.246**	0.050	0.129	0.000
BL	0.029	**0.044**		0.041	**0.288**	0.055	0.124	**0.370**	0.000	**0.301**	**0.194**	**0.219**	0.076
CL	0.051	**0.097**	0.013		0.057	0.000	0.000	0.157	0.000	0.173	0.018	0.024	0.000
CT	0.050	**0.043**	0.016	0.004		0.063	0.000	0.000	**0.222**	0.181	0.005	0.000	0.034
DD	0.000	**0.036**	0.006	0.018	0.027		0.000	0.159	0.000	**0.240**	0.042	0.010	0.000
GW	0.000	**0.049**	0.017	0.036	0.000	0.008		0.096	0.049	**0.213**	0.012	0.000	0.000
LI	0.035	**0.049**	0.019	0.018	0.000	0.006	0.001		**0.280**	**0.194**	0.053	0.018	0.121
MC	0.031	**0.043**	0.019	0.001	0.015	0.000	0.025	0.020		**0.267**	0.076	0.121	0.021
MF	0.050	**0.055**	0.000	0.008	0.006	0.012	0.019	0.004	0.020		0.193	0.000	0.184
MS	0.035	**0.046**	0.009	0.020	0.013	0.020	0.015	0.011	0.000	0.017		0.000	0.016
NX	0.073	**0.131**	0.003	0.000	0.006	0.027	0.024	0.013	0.000	0.002	0.000		0.000
TP	0.009	**0.039**	0.002	0.027	0.000	0.000	0.000	0.008	0.000	0.002	0.000	0.023	

Figures

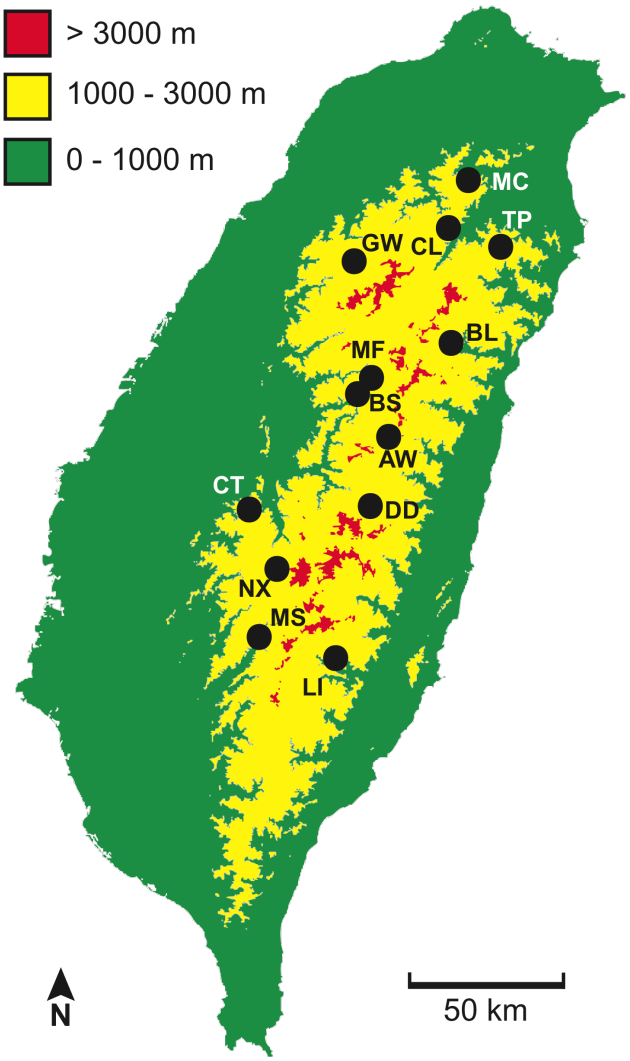


Figure 2.1. Map of Taiwan showing elevation and sampling site locations. Colors denote elevation. Sampling sites are indicated by black circles. The dotted line represents a phylogeographic division reported for many taxa. The yellow area (between 1000 m and 3000 m is the approximate range of Steere's Liocichla). Population abbreviations are from Table 1.

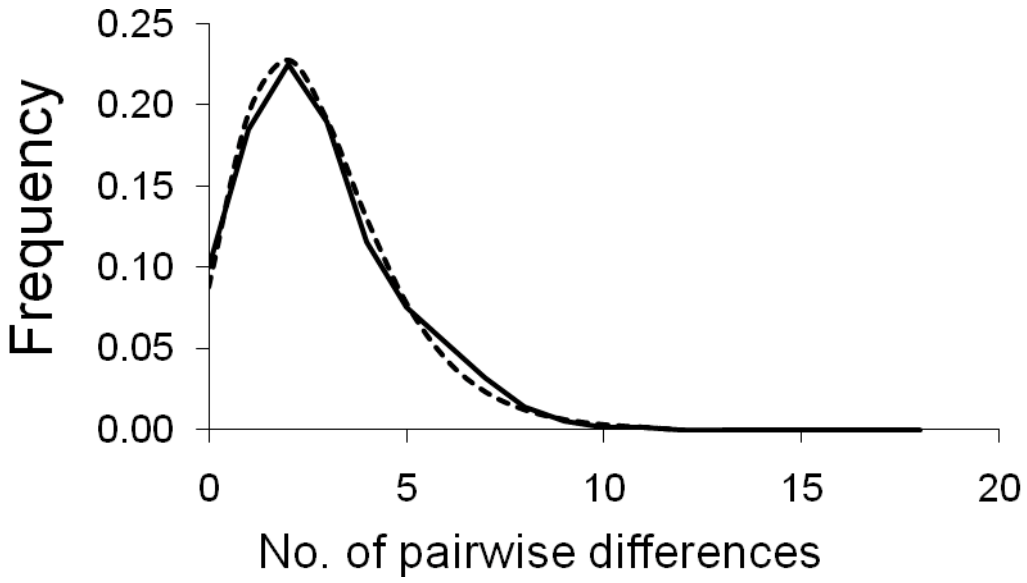


Figure 2.2. Mismatch distribution of Steere's Liocichla mitochondrial sequence data. There is a significant correlation between observed (solid line) and expected frequencies under a model of sudden expansion (dotted line) for the number of pairwise differences.

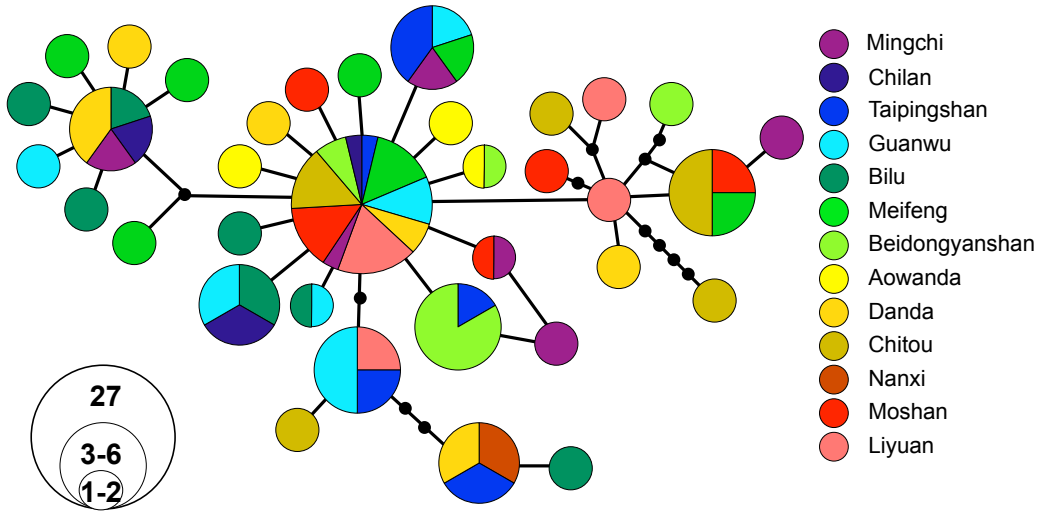


Figure 2.3. Mitochondrial DNA haplotype network for Steere's Liocichla ND2 sequence data**.** Colors indicate the population of origin. Populations from the northern group are represented by cool colors (blues and greens), and populations from the southern group are represented by warm colors (reds and yellows). Each circle represents a haplotype, and the size of the circles is proportional to its frequency. Small black circles represent unsampled haplotypes.

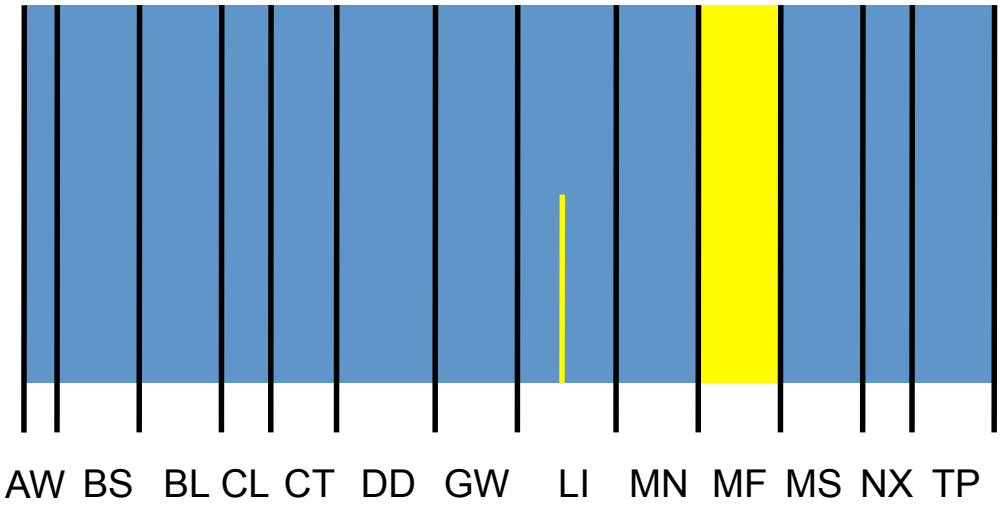


Figure 2.4. Admixture coefficients for Steere's Liocichla estimated using BAPS. Vertical columns correspond to individuals; black lines separate populations. Two groups were identified and they are indicated by different colors. Columns are colored in proportion to the estimated admixture coefficients for each individual. Population abbreviations are from Table 1.

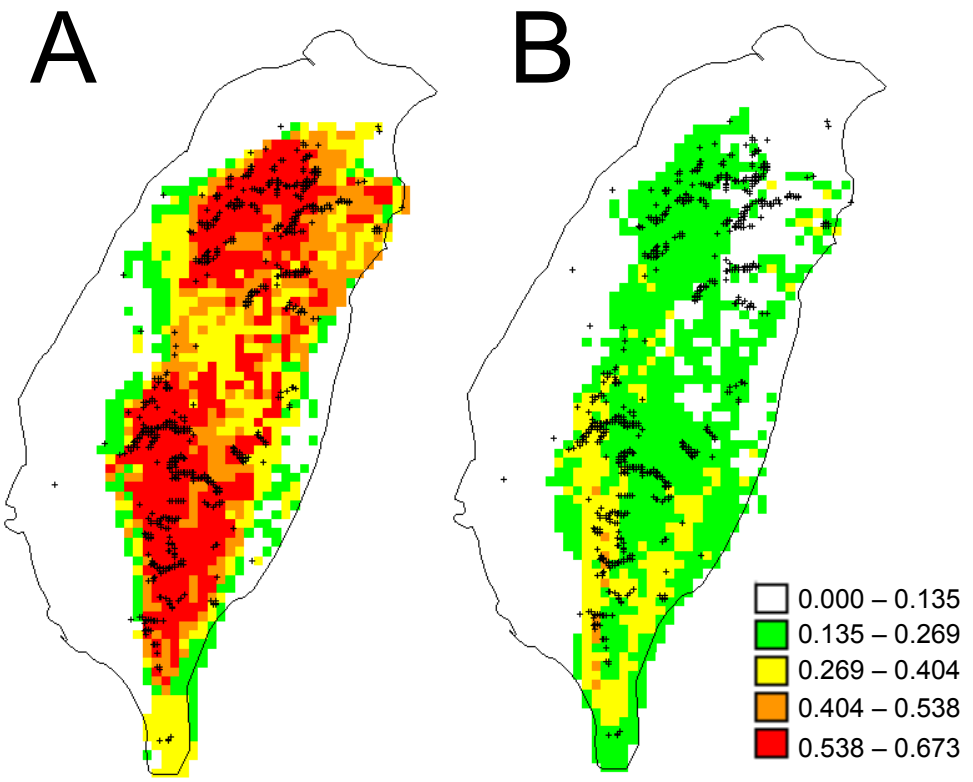


Figure 2.5. Predicted geographic distribution of Steere's Liocichla based on (a) present-day climatic conditions and (b) climatic conditions at the Last Glacial Maximum. Grid cells are classified by predicted suitability with white being least suitable and red being most suitable. Georeferenced occurrence records used to generate the models are indicated by plus signs.

Chapter 3

A new timeframe for the diversification of Japan's mammals

McKay, B. D. (submitted)

Abstract

Landbridge islands, which maintain episodic connections with the mainland, are understudied island systems with the potential to offer fresh insights into the evolution of isolated biotas. The main Japanese Islands are landbridge islands divided by the biogeographic division Blakiston's Line and represent two natural laboratories for studying landbridge diversification. Consistent with the landbridge colonization hypothesis, divergence time estimates from 24 Japanese mammals were found to postdate the initial breakup of Japan from the mainland. However, this study implicates a colonization timeframe that is significantly older than the current paradigm based on fossil evidence. Further, half of the divergence time estimates in the Hondo region were clumped around 2.4 Ma ago. This suggests a potential dramatic interchange period that is concordant with significant global cooling, a period when the first landbridge may have connected Japan to the mainland.

Introduction

Islands have long been regarded as model systems for evolutionary study (Wallace 1881). Isolation makes islands prone to allopatric speciation, and distinct boundaries and simplified biotas make them easier to study than continental systems. To date, most research has focused on oceanic islands that are in complete isolation from continental landmasses, and we have learned much about evolutionary processes such as adaptive radiation (Schluter 2000; Losos and Ricklefs 2009). In contrast, landbridge islands, which are periodically connected to the mainland during periods of low sea level, have received less attention, though their incomplete isolation makes them valuable for studying the affects of reduced isolation on island biotas. In this way, landbridge islands are ideal intermediate systems that can help bridge continental and oceanic island studies. Further, variation in the depths of isolating straits provides opportunities for studying a continuum of landbridge islands, from historically more isolated (deep straits) to historically less isolated (shallow straits).

The Japanese main islands of Honshu, Shikoku, Kyushu, and Hokkaido and the Russian island of Sakhalin comprise an extended series of landbridge islands in East Asia (Fig. 1). These islands were originally part of the mainland until seafloor spreading formed the Sea of Japan, a backarc basin, approximately 15-17 million years (Ma) ago (Taira 2001; Neall and Trewick 2008). The area is divided into two groups of islands by the Tsugaru Strait, which separates Hokkaido and Sakhalin in the north from Honshu, Shikoku, and Kyushu in the south (Fig. 1). This division corresponds to a major

biogeographic break known as Blakiston's Line (Blakiston 1883). The fauna north of Blakiston's Line is taxonomically aligned with the fauna of the northern Palearctic region, whereas the fauna south of Blakiston's Line is highly endemic and thought to derive largely from the fauna of northern China and the Indo-Malayan region (Smith 1983; Dobson 1994). Few non-volant animals are distributed across Blakiston's Line.

Although vicariance would have followed the initial separation of Japan from the mainland 15-17 Ma ago, much of Japan's current fauna is thought to be of a more recent origin, a result of landbridge colonization (Dobson 1994). Two landbridges have historically connected the main Japanese islands to the mainland (Fig. 2). North of Blakiston's line, Hokkaido has been periodically connected to the Russian mainland via Sakhalin. Based on present-day ocean depths, Hokkaido would be connected to the mainland with a 55 m reduction in current sea-level. Thus, Hokkaido would have been connected to the mainland during most glacial periods. South of Blakiston's Line, the islands of Honshu, Shikoku, and Kyushu are divided by marine straits that rarely exceed 50 m, so they were frequently connected as one landmass in the past (Fig. 2) and can be considered a single biogeographic unit that I will hereafter refer to as "Hondo" (following Dobson 1994). Hondo is separated from the Korean Peninsula by a deep strait (130 m) that would have only formed landbridge connections to the mainland at the height of major glacial periods. Thus, Hondo has maintained a more isolated history than Hokkaido.

Determining the timing of Japan landbridge connectivity is complicated by complex tectonic activity in the region (Ota and Machida 1987). However, Dobson and

Kawamura (1998) studied fossil non-volant mammals, which are considered to be poor over-water dispersers and thus their distributions on landbridge islands should largely result from connections to adjacent mainlands. Dobson and Kawamura (1998) suggest four major periods of landbridge connection between Hondo and the Korean Peninsula. Two are dated to the Middle Pleistocene (0.3 and 0.5 Ma ago), one is dated to the Early Pleistocene (1.0 Ma ago), and another is dated to the Pliocene (between 3.0-4.0 Ma ago). They further suggest that the current Hokkaido mammal fauna is the result of periodic colonization from the mainland during the last 0.1 Ma (Dobson and Kawamura 1998). Though fossil records of terrestrial mammals in Hondo are fairly extensive from the Middle Pleistocene to the Holocene, they are scare earlier, whereas in Hokkaido, fossils are scarce throughout the Quaternary and nearly absent earlier (Kawamura 1991). This paucity of early fossils implicates a potential basis in the landbridge hypotheses proposed by Dobson and Kawamura (1998). In this paper, I use molecular approaches to estimate the divergence times between Japanese mammal species and their closest mainland relative to test the landbridge colonization model of Dobson and Kawamura.

Methods

Datasets

I compiled datasets of Japanese mammals and their mainland sister taxa (or mainland conspecifics) using sequences obtained from GenBank. I started with a list of the native

mammals of Japan (Ohdachi et al. 2009). I excluded volant and marine species and species not distributed on the Japanese main islands (i.e. Honshu, Shikoku, Kyushu, and Hokkaido). I also excluded Japanese taxa without a published comprehensive molecular phylogeny that clearly identified their relationships with taxa on the mainland. The species *Glirulus japonicus* was excluded because fossil evidence suggests it is a relict species with a recently extinct relative in mainland Asia (Dobson and Kawamura 1998). To control for rate heterogeneity among loci, I used the same molecular locus for all species. I used the mitochondrial DNA (mtDNA) cytochrome b gene because it was the most widely available on GenBank. I only used species that had at least two individuals from Japan and two individuals from the mainland. I refer to a Japanese taxon and its mainland sister taxon as a taxon-pair.

To accommodate the two possible landbridge routes into Japan (Fig. 2), taxon-pairs were divided into two datasets. The first dataset included taxa from Hokkaido; the second dataset included taxa from Honshu-Shikoku-Kyushu (Hondo). Species found in both Hokkaido and Hondo (i.e. across Blakiston's line) were excluded because it is unclear whether they colonized Japan from the north via Hokkaido or from the south via Hondo. Sequences were aligned independently for each taxon-pair using Clustal W (Larkin et al. 2007).

Divergence time estimates

I estimated divergence times for each taxon-pair using a Bayesian Markov chain Monte Carlo (MCMC) method of species tree estimation implemented in the program *BEAST v. 1.6.1 (Drummond and Rambaut 2007; Heled and Drummond 2010). Analyses were run for each taxon-pair using an uncorrelated lognormal molecular clock (Drummond et al. 2006) and an HKY nucleotide substitution model. I used a UPGMA topology as a starting tree; all priors were set to their default values. I used a cytochrome *b* mutation rate of 1.0×10^{-8} per site per year (divergence rate of 2% per million years)(Brown et al. 1979) to convert divergence time estimates into years. I ran MCMC chains for 100 million generations (sampling every 10,000 generations and discarding the first 10% as burn-in) and assessed convergence of the MCMC chain in the program Tracer v. 1.5 (Rambaut and Drummond 2007). I used Tracer to determine the mean species tree root estimate (the divergence time of the taxon-pairs) and its 95 per cent highest posterior density (HPD).

Controlling for coalescent stochasticity

Divergence time estimates from single markers have large confidence intervals because of the expected stochasticity associated with the coalescent process. This is reflected in the divergence time estimates obtained from *BEAST, and it makes it difficult to pinpoint periods of divergence. However, when multiple taxa are sampled, it is possible

to statistically identify time periods consistent with the simultaneous divergence of a group of co-distributed taxon-pairs separated by a common barrier (Hickerson et al. 2006). Such a method is implemented through a hierarchical approximate Bayesian computation (HABC) model in the program MTML-msBayes (Huang et al. 2011). The msBayes model assumes that ancestral populations are split into two daughter populations, but the ancestral and current effective population sizes and the divergence times are free to vary among taxon-pairs.

I ran two sets of msBayes analyses: one using the Hokkaido taxon-pair dataset and another using the Hondo taxon-pair dataset. I first used msBayes to determine the number of simultaneous divergence events that were consistent with the data. The upper and lower limits of the uniform prior distribution for each taxon-pair's θ ($\theta = N\mu$, where N is the effective population size and μ is the mutation rate) were determined by msBayes using empirical π estimates. The prior for the number of possible divergence events (Ψ) equaled the number of taxon-pairs. Posteriors were constructed using 1000 accepted draws from 10 million simulated draws from the prior. Each analysis was run twice to confirm consistency between runs.

Once the most probable number of simultaneous divergence events was estimated, I conducted additional msBayes analyses to determine the timing of each of the divergence events. I did this by constraining Ψ to equal the most probable number of divergence events suggested by the unconstrained analyses. All other conditions were as left unchanged. Divergence times were converted into years using the formula $t = \theta_{ave}\mu\tau$,

where t is the divergence time in years, θ_{ave} is the mean θ prior, and τ is the divergence time in coalescent units. I used the same mutation rate as above.

Results

I assembled cytochrome *b* sequence datasets for 24 taxon-pairs. This represents 48% of the non-volant terrestrial mammal species native to the Japanese main islands and includes 12 taxa from Hokkaido and 12 taxa from Hondo. The species list I used, including information for excluded species, can be found in the supplementary material. Sample sizes ranged from 2 to 50 individuals and averaged 14.7 individuals per taxon. Sequences ranged from 224 to 1,143 base pairs (bp) and averaged 933 bp per taxon-pair (see Table 1). GenBank numbers of all sequences used can be found in the supplementary material.

Divergence times estimated from the species tree approach in *BEAST were dramatically different between Hokkaido and Hondo. Hokkaido samples were typically not reciprocally monophyletic with respect to their closest relatives on the mainland. Divergence time estimates of Hokkaido samples ranged from 0.00 to 3.37 Ma ago and averaged 0.76 Ma ago (Table 1). In contrast, samples from Hondo were all reciprocally monophyletic with respect to their closest relatives on the mainland. Divergences time estimates of Hondo samples ranged from 0.50 to 8.69 Ma ago and averaged 2.79 Ma ago (Table 2), making them nearly four times older, on average, than the Hokkaido samples. The single locus estimates of divergence time had high levels of uncertainty evidenced by

the large 95 per cent highest posterior densities (Fig. 3). MtDNA gene trees for all taxon-pairs can be found in the supplementary material.

The msBayes analysis suggested five divergence events (Ψ) as most probable for the Hokkaido dataset and four divergence events (Ψ) as most probable for the Hondo dataset (Fig. 4). Subsequent analysis of the Hokkaido dataset with Ψ constrained to be five resulted in divergence events dated at 48,000 (0-192,000), 104,000 (0-456,000), 256,000 (0-896,000), 896,000 (384,000-1,416,000), and 1,360,000 (736,000-1,608,000) years ago (Fig. 5). Analysis of the Hondo dataset with Ψ constrained to be four resulted in divergence events dated at 1.18 (0.25-2.24), 2.35 (1.34-3.39), 3.46 (2.24-4.26), and 4.06 (3.04-4.47) Ma ago (Fig. 5).

Discussion

The main Japanese Islands were originally part of the continental mainland until 15-17 Ma ago when they became isolated through sea-floor spreading (Taira 2001; Neall and Trewick 2008). Vicariance would have followed this initial separation of Japan from the mainland; however, the results presented here, which are based on the molecular clock, can reject divergence times on the order of 15-17 Ma for most species. Divergence between Japanese species and their closest mainland relatives has clearly been more recent. This finding is consistent with the traditional paradigm that implicates a major role for episodic landbridge connections with the mainland in the diversification of the Japanese fauna (Dobson 1994).

The landbridge divergence model is a kind of episodic (or cyclic) vicariance, which can occur if vicariant barriers are eroded and reformed due to geological or climatic processes. Lieberman (2004) describes this as a three-step process. First, an existing barrier to dispersal is eroded and this causes concurrent range expansion (dispersions sensu Platnick 1976) of many different species into the newly accessible area. Second, the barrier reemerges near its original position. Third, populations isolated by the new vicariant barrier differentiate via allopatry. If this process repeats, for example, due to repeated climatic cycles, pairs of sister taxa on either side of the barrier would be expected to have clumped divergence times according to the specific vicariant cycle by which they were isolated. Although the individual divergence time estimates presented here are associated with large uncertainty, the mean divergence times of the Hondo dataset (Fig. 3) do appear clumped, which is consistent with the landbridge divergence hypothesis.

Timeframe of divergence

The timeframe of the landbridge divergence suggested by this study was significantly older than the timeframe held under the current paradigm. This was true for both Hokkaido and Hondo species. For example, Dobson and Kawamura (1998) hypothesized that 10 of the 12 Hondo species in this study colonized Japan during the Middle Pleistocene (0.3-0.5 Ma ago). In contrast, divergence time estimates presented here indicate it is unlikely any of the species sampled colonized Hondo in the Middle

Pleistocene, and, instead, most Hondo species seem to have diverged much earlier (>1.8 Ma ago). Only the lower bound of the most recent simultaneous divergence event suggested by the msBayes analysis is consistent with the Dobson and Kawamura paradigm. There are reasons to believe this study presents a more accurate picture for the timeframe mammalian colonization that the current paradigm. Dobson and Kawamura (1998) allocated species to specific landbridge periods based on the appearance of those species in the fossil record. Fossil evidence represents a minimum age for the existence of a species in a given area, so species will colonize areas before they leave fossil evidence (Marshall 1990). The lag between the time a species colonizes a new area and the time it leaves fossil evidence of its colonization depends on the quality of the fossil record. Mammalian fossils are scare in Hondo before the Middle Pleistocene and scarce in Hokkaido throughout the Quaternary and nearly absent earlier (Kawamura 1991). Thus, given the paucity of early fossils, it is not surprising that the fossil record would underestimate colonization times. In this sense, the timeframe of colonization suggested by this study is consistent with the fossil record.

Individual divergence time estimates were associated with high uncertainty due to stochasticity inherent in the coalescent. The msBayes analysis, which combined all species, is designed to account for the expected stochasticity in single species estimates and determine time periods that are consistent with simultaneous divergence among a group of species (Hickerson et al. 2006). This helped narrow down periods when landbridges were most likely. In Hondo, landbridge connections were suggested at approximately 1.2, 2.4, 3.5, and 4.0 Ma ago. These estimates have associated confidence

intervals and should not be over-interpreted. However, individual divergence time estimates seem to suggest which species are associated with each of the four landbridge connections. For example, the mean divergence estimates of *Meles anakuma*, *Martes melampus*, *Microtus montebelli*, and *Crocidura dsinezumi* are very close to 1.2 Ma ago, so these species likely diverged at the same time. Additionally, the mean divergence estimates of *Mustela itatsi*, *Lutra nippon*, *Chimarrogale platycephalus*, *Sorex shinto*, *Sciurus lis*, and *Ursus thibetanus* are very close to 2.4 Ma ago, so these species likely diverged at the same time.

The mean divergence time estimates of *Pteromys momonga* and *Sorex hosonoi* were unusually large and not obviously clumped with other species. These two species likely account for the landbridge periods indicated by msBayes at approximately 3.5 and 4.0 Ma ago. Clumped divergence time estimates are consistent with periods of landbridge connections; however, idiosyncratic divergence times shown by only one species may have alternative causes. For example, overwater dispersal could be invoked to explain such patterns. This seems particularly unlikely for *Sorex hosonoi*, however, because shrews are highly susceptible to starvation (Hanski 1985) and can not swim distances of over 500 m (Hanski 1986); thus, they are considered to be particularly poor over-water dispersers. Another explanation is that the true sister species for *Pteromys momonga* and for *Sorex hosonoi* was not sampled. This could result from incorrectly sampling the wrong mainland species or in instances where the true mainland sister species has gone extinct. Assuming the wrong sister species would artificially inflate divergence time estimates and cause colonization times to be overestimated and idiosyncratic, which

would explain *Pteromys momonga* and *Sorex hosonoi* as outliers. Future work should focus on obtaining multi-locus nuclear sequences, which will help narrow the uncertainty associated with individual divergence time estimates and allow a more accurate allocation of species to colonization period. Additional species will also help determine if the divergence times of *Pteromys momonga* and *Sorex hosonoi* are truly idiosyncratic.

It is unlikely that any of the species included in this study colonized Hondo during the Last Glacial Maximum (LGM). This is consistent with the idea that sea-level reduction during the Late Quaternary was insufficient to connect Hondo to the mainland, which as been proposed based on palaeoceanographic evidence (Ohshima 1990). The current minimum depth of the Korean Strait is 130 m, and the maximum inferred sea-level drop during the LGM was 120 m (Tushingham and Peltier 1993). This would have reduced the Korean Strait to a width of 20 km and a maximum depth of approximately 10 m (Dobson 1994); this appears to have effectively isolated Hondo's terrestrial mammal fauna.

Because of a poor fossil record in Hokkaido, Dobson and Kawamura (1998) did not make strong statements about the colonization of Hokkaido. However, they did suggest that colonization occurred during a series of periodic landbridges beginning around 0.1 Ma ago. The molecular evidence presented here suggests a more protracted colonization period that extends back until at least 1.4 Ma ago. As with the Hondo dataset, individual divergence time estimates were associated with high uncertainty. However, the msBayes analysis, which combined all Hokkaido species, did not support a single simultaneous divergence event for all Hokkaido samples, and instead suggested

that at least five divergence periods must be invoked to explain the variation in divergence time estimates observed. This is not surprising because, due to the shallow depth of the strait isolating Hokkaido, Hokkaido was likely connected to the mainland during most glacial periods, which occurred in cycles of approximately 100,000 years throughout the Quaternary (Hewitt 2004). The number of divergence events implicated in this study should be considered a minimum number, and future analyses with multi-locus sequences, which will provide tighter confidence intervals for individual divergence time estimates, and additional species may reveal that the divergence time periods described here include multiple divergence events.

Conclusions

Reduced isolation of the Japanese Islands seems to have had a major effect on its mammalian fauna. The original fauna that occupied Japan 15-17 Ma ago are mostly gone (endemic Japanese genera such as *Dymecodon* and *Urotrichus* may be relicts of this initial fauna), and instead Japan is currently home to more recently diverged mammalian species. What happened to the original fauna? Was the transition dramatic, like the Great American Interchange (Marshall et al. 1982)? Unfortunately, the fossil record in Japan is scare before the Middle Pleistocene; however, molecular patterns may provide some clues. For example, half of the Hondo species seem to have diverged from their closest mainland relative approximately 2.4 Ma ago, implicating a dramatic interchange event. This timeframe roughly corresponds to a period of significant global cooling ~2.75 Ma

ago that followed the end of the early Pliocene warm period (Ravelo et al. 2004). This raises the intriguing possibility that this cooling period, which would have been the first major cooling period following Japan's initial isolation from the mainland, initiated the first landbridge connection, which led to a dramatic interchange of Japan's fauna. It would be interesting to see if this pattern holds for other groups of organisms.

Tables

Table 3.1. Summary of the dataset that includes the common and scientific names of Japanese taxa, the mainland sister group used in analyses, sample sizes of both Japanese and mainland samples, the state of reciprocal monophyly, the length, in base pairs, of the cytochrome b sequence alignment, the divergence time estimates from *BEAST analyses in millions of years ago with the highest posterior density in parenthesis, and references for the molecular phylogeny supporting the sister group status of the mainland samples used.

Common Name	Scientific Name	Sister	Sample Size (Japan/ mainland)	Reciprocally Monophyletic	Sequence Length (bp)	Divergence (Ma)	References
Hokkaido							
Eurasian Red Squirrel	*Sciurus vulgaris*	mainland individuals	4/4	no	1040	0.89 (0.00-1.62)	Oshida & Masuda 2000
Siberian Flying Squirrel	*Pteromys volans*	mainland clade	6/23	yes	1140	0.97 (0.32-1.9)	Oshida et al. 2005
Hokkaido Red-backed Vole	*Myodes rex*	*M. rufocanus*	3/16	yes	402	3.37 (1.55-5.48)	Cook et al. 2004

Gray Red-backed Vole	*Myodes rufocanus*	mainland clade	9/16	yes	1140	1.59 (0.58-2.85)	Iwasa et al. 2000
Northern Red-backed Vole	*Myodes rutilus*	mainland clade	8/7	yes	1140	0.80 (0.23-1.49)	Iwasa et al. 2002
Korean Field Mouse	*Apodemus peninsulae*	mainland individuals	3/15	no	1143	0.30 (0.06-0.55)	Serizawa et al. 2002
Laxmann's Shrew	*Sorex caecutiens*	mainland clade	7/39	yes	1140	0.39 (0.20-0.62)	Ohdachi et al. 2001
Slender Shrew	*Sorex gracillimus*	mainland individuals	10/9	no	630	0.26 (0.00-0.50)	Ohdachi et al. 2001
Eurasian Least Shrew	*Sorex minutissimus*	mainland individuals	3/7	no	1140	0.38 (0.00-0.84)	Ohdachi et al. 2001
Long-clawed Shrew	*Sorex unguiculatus*	mainland individuals	26/5	no	1140	0.00 (0.00-0.02)	Ohdachi et al. 2001
Brown Bear	*Ursus arctos*	mainland individuals	9/34	no	1140	0.12 (0.05-0.21)	Matsuhashi et al. 1999; Matsuhashi et al. 2001
Sable	*Martes zibellina*	mainland clade	11/14	yes	720	0.08 (0.00-0.22)	Hosoda et al. 2000

Hondo							
Japanese Squirrel	*Sciurus lis*	*S. vulgaris*	35/8	yes	1040	2.24 (0.16-8.45)	Oshida and Masuda 2000; Oshida et al. 2009
Japanese Dwarf Flying Squirrel	*Pteromys momonga*	*P. volans*	4/55	yes	1068	8.69 (0.67-16.78)	Oshida et al. 2000
Japanese Grass Vole	*Microtus montebelli*	*M. oeconomus*	2/50	yes	930	1.31 (0.08-5.61)	Conroy and Cook 2000; Bannikova et al. 2010
Dsinezumi Shrew	*Crocidura dsinezumi*	*C. lasiura*	21/3	yes	1140	0.50 (0.00-1.12)	Ohdachi et al. 2004
Japanese Water Shrew	*Chimarrogale platycephalus*	*C. himalyica*	50/6	yes	930	2.40 (0.31-4.15)	Ohdachi et al. 2006
Azumi Shrew	*Sorex hosonoi*	*S. minutissimus*	6/10	yes	630	5.84 (0.34-12.08)	Ohdachi et al. 2001
Shinto Shrew	*Sorex shinto*	*S. caecutiens*	13/46	yes	1140	2.28 (0.54-6.07)	Ohdachi et al. 2001; Ohdachi et al. 2006

Asiatic Black Bear	*Ursus thibetanus*	NE Asia clade	7/11	yes	1140	2.09 (0.19-4.08)	Yasukochi et al. 2009
Japanese Weasel	*Mustela itatsi*	*M. sibirica*	5/19	yes	375	2.88 (0.18-6.79)	Hosoda et al. 2000; Marmi et al. 2004
Japanese Martin	*Martes melampus*	*M. zibellina/martes* clade	20/19	yes	1140	1.36 (0.12-2.48)	Hosoda et al. 2000
Japanese Badger	*Meles anakuma*	*M. leucurus*	17/4	yes	629	1.40 (0.16-2.86)	Kurose et al. 2001
Japanese Otter	*Lutra nippon*	*L. lutra*	2/6	yes	224	2.49 (0.00-4.87)	Suzuki et al. 1996

Figures

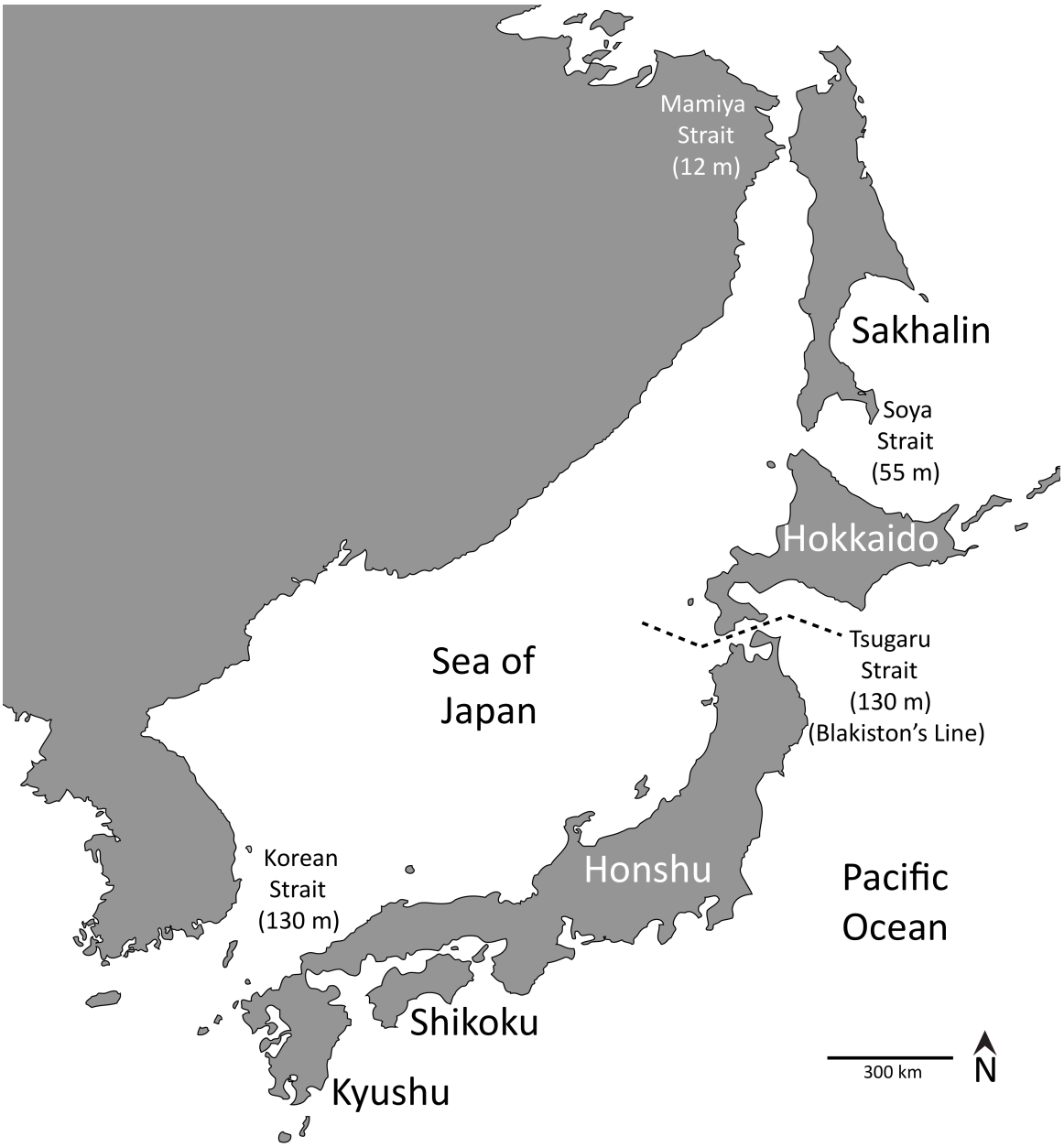


Figure 3.1. Map of the main Japanese Islands (Honshu, Shikoku, Kyushu, and Hokkaido) showing the straits currently isolating these islands from the mainland. The depths of straits are indicated in parenthesis. A biogeographic division, Blakiston's Line, separates Hokkaido in the north from Honshu, Shikoku, and Kyushu in the south.

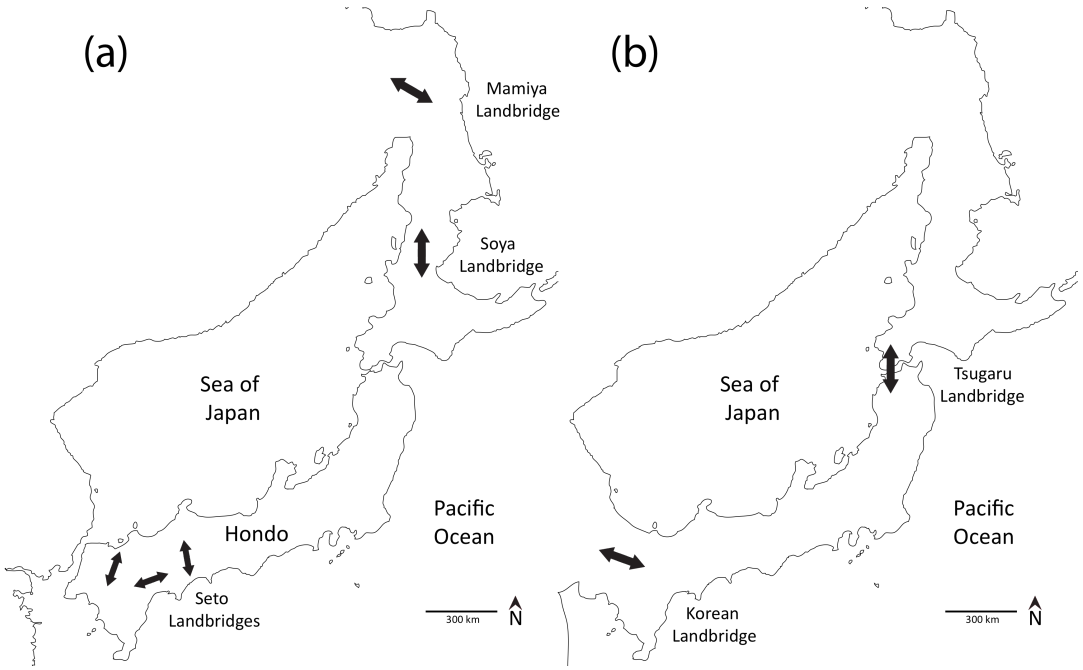


Figure 3.2. Map of the Japan area showing the distribution of land if the sea level fell by (a) 120 m and (b) 150 m (modified, after Dobson 1994). Sea level at the Last Glacial Maximum is estimated to have fallen approximately 120 m. This would have connected Hokkaido to the mainland via Sakhalin and connected Honshu, Shikoku, and Kyushu into a single landmass (Hondo), but it would have been insufficient to connect Hondo to the mainland. However, a sea-level drop of 150 m would have connected Hondo to the Korean Peninsula and connected Hokkaido to Hondo. Arrows indicate open avenues for range expansion (dispersion *sensu* Platnick 1976) during open landbridge connections.

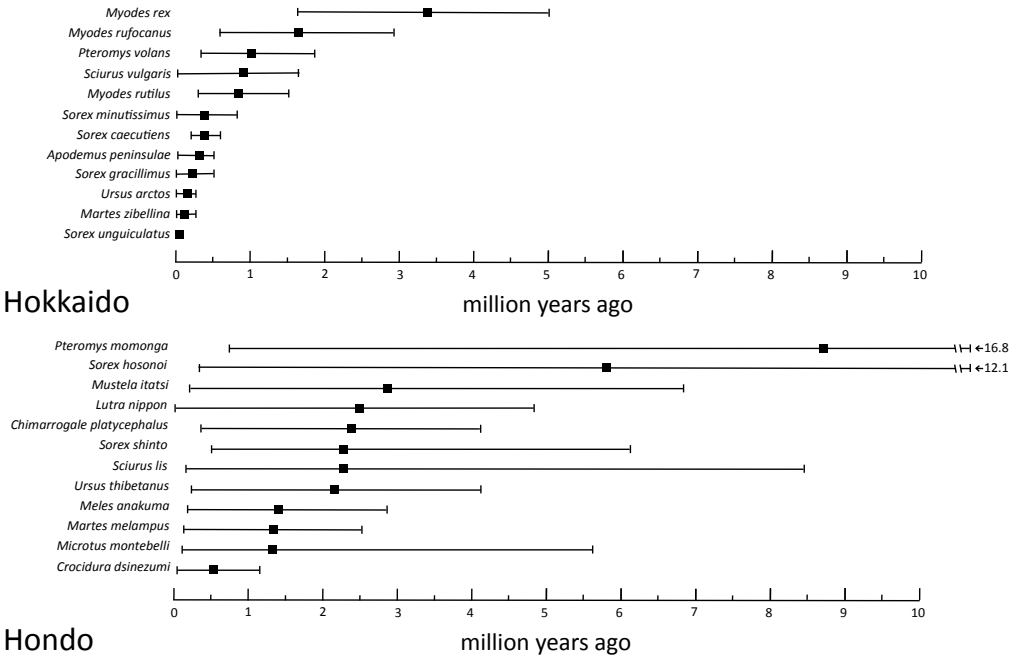


Figure 3.3. Divergence time estimates for each taxon-pair from the species tree approach in *BEAST. Top: results from Hokkaido dataset. Bottom: results from Hondo dataset. Squares are the means and lines are the 95% highest posterior densities (HPD).

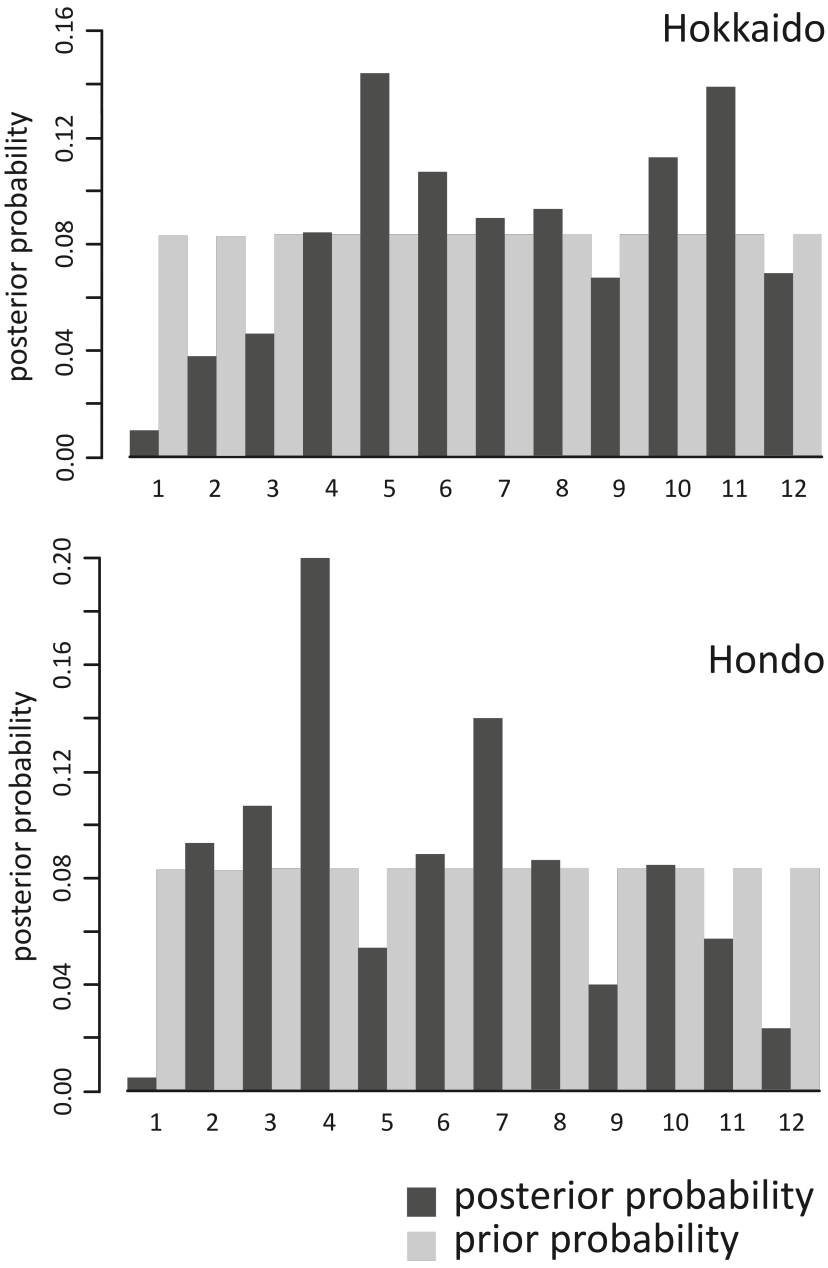


Figure 3.4. Bayesian probability densities for Ψ, the number of divergence events. Top: results from Hokkaido dataset. Bottom: results from Hondo dataset. Prior probabilities are in light gray, posterior probabilities are in dark gray. The most probable Ψ for the Hokkaido dataset was five; the most probable Ψ for the Hondo dataset was four.

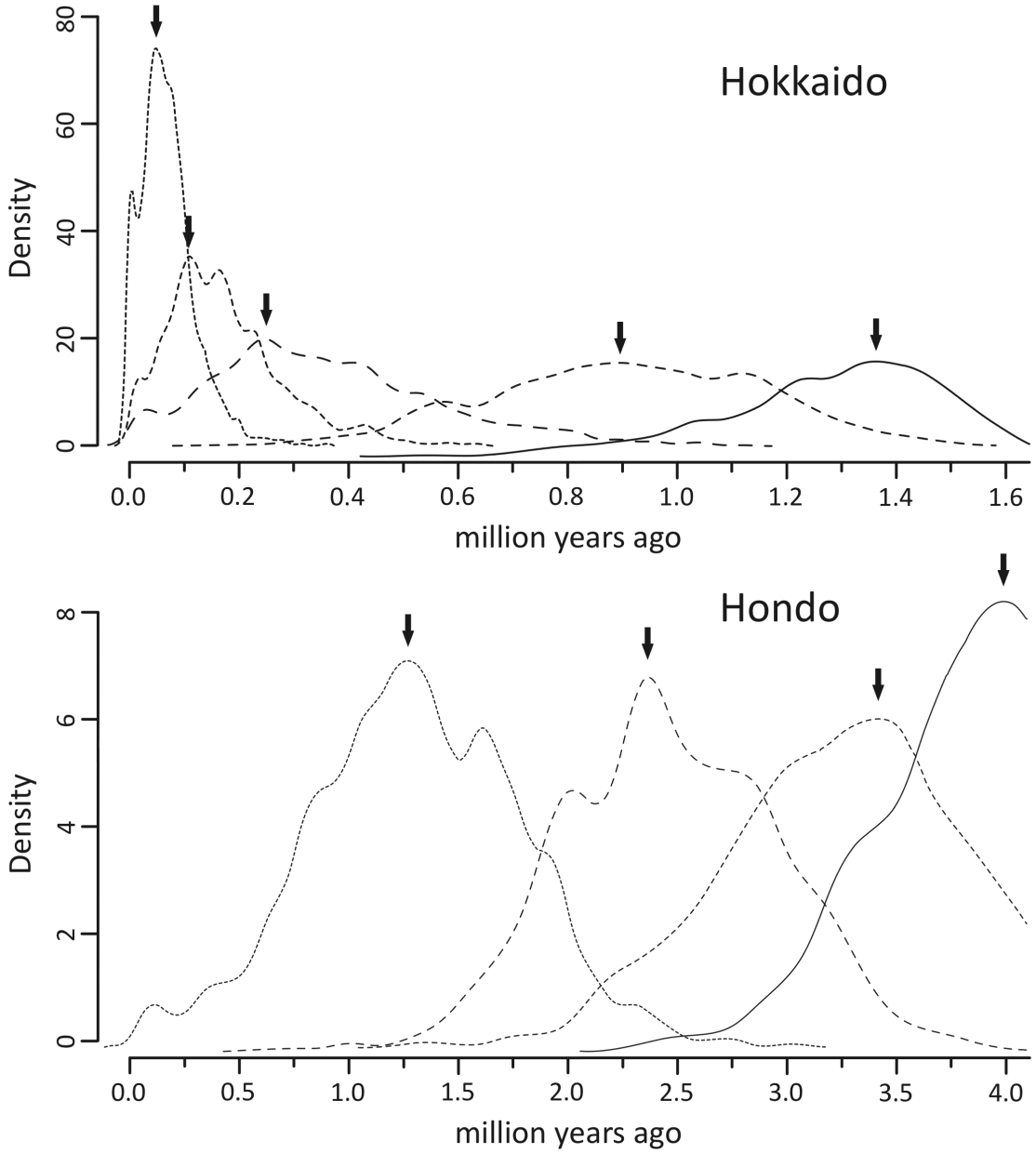


Figure 3.5. Posterior distributions of the timing of divergence events suggested by msBayes. Top: results from Hokkaido dataset when Ψ was constrained to be five. Bottom: results from Hondo dataset when Ψ was constrained to be four. The modes of each divergence event are indicated with arrows.

References

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

Avise, J. C., and R. Ball. 1990. Principles of genealogical concordance in species concepts and biological taxonomy. *Oxf. Surv. Evol. Biol.* 7:45-67.

Backström, N., S. Fagerberg, and H. Ellegren. 2008. Genomics of natural bird populations: a gene-based set of reference markers evenly spread across the avian genome. *Mol. Ecol.* 17:964-980.

Bannikova, A. A., V. S. Lebedev, A. A. Lissovsky, V. Matrosova, N. I. Abramson, E. V. Obolenskaya, and A. S. Tesakov. 2010. Molecular phylogeny and evolution of the Asian lineage of vole genus *Microtus* (Rodentia: Arvicolinae) inferred from mitochondrial cytochrome b sequence. *Biol. J. Linn. Soc.* 99:595-613.

Ball, R. M., J. E. Neigel, and J. C. Avise. 1990. Gene genealogies within the organismal pedigrees of random mating populations. *Evolution* 54:1414-1422.

Blakiston, T. 1883. Zoological indications of ancient connection of the Japan Islands with the Continent. *Transactions of the Asiatic Society of Japan* 11:126-140.

Brower, A. V. Z. 1999. Delimitation of phylogenetic species with DNA sequences: a critique of Davis and Nixon's population aggregation analysis. *Syst. Biol.* 48:199-213.

Brown, W. M., M. George, and A. C. Wilson. 1979. Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* 76:19671971.

Bruen, T. C., H. Philippe, and D. Bryant. 2006. A simple and robust statistical test for detecting the presence of recombination. *Genetics* 172:2665-2681.

Butchart, S., M. Crosby, J. Gilroy, and S. Chan. 2008. Pycnonotus taivanus. IUCN 2010. IUCN Red List of Threatened Species. Version 2010.4. <www.iucnredlist.org>.

Chen, H.-C. 1994 Preliminary study on mitochondrial DNA sequence and population variation of *Rana sauteri*. MS Thesis Taipei: National Taiwan University.

Chen, H.-C., H.-W. Yuan, and Y.-S. Lin. 2004. Genetic diversity and population genetic structure of the Ring-necked Pheasant (*Phasianus colchicus*) in Taiwan. *Quarterly Journal of the Experimental Forest of National Taiwan University* 18:65-75.

Clement, M., D. Posada, and K. A. Crandall. 2000. TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* 9:1657-1169.

Clements, J. F. 2007. The Clements checklist of birds of the world, 6th edn. Cornell University Press, Ithaca, New York.

Collar, N. J., and C. Robson. 2007. Family Timaliidae (Babblers). Pp. 70-291. *in* J. del Hoyo, A. Elliott, and D. Christie, eds. Handbook of the birds of the world: Picarthartes to Tits and Chickadees, vol. 12. Lynx Edicions, Barcelona, Spain.

Conroy, C. J., and J. A. Cook. 2000. Molecular systematics of a holarctic rodent (Microtus: Muridae). *J. Mammal.* 81:344-359.

Cook, J. A., A. M. Runck, and C. J. Conroy. 2004. Historical biogeography at the crossroads of the northern continents: molecular phylogenetics of red-backed voles (Rodentia: Arvicolinae). *Mol. Phylogenet. Evol.* 30:767-777.

Corander, J., and P. Marttinen. 2006. Bayesian identification of admixture events using multilocus molecular markers. *Mol. Ecol.* 15:2833-2843.

Corander, J., P. Marttinen,, J. Siren, and J. Tang. 2008. Enhanced Bayesian modeling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* 9:539

Cracraft, J. 1983. Species concepts and speciation analysis. *Current Ornithology* 1:159-187.

Cracraft, J. 1989. Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. Pp. 28-59 *in* D. Otte, and J. A. Endler, eds. Speciation and its Consequences. Sinauer Associates, Inc., Sunderland, Massachusetts.

Cracraft, J. 1997. Species concepts in systematics and conservation biology-an ornithological viewpoint. Pp. 325-340 *in* M. F. Claridge, H. A. Dawah, and M. R. Wilson, eds. Species: the Units of Biodiversity. Chapman & Hall Ltd., London.

Davis, J. I., and K. C. Nixon. 1992. Populations, genetic variation, and the delimitation of phylogenetic species. *Syst. Biol.* 41:421-435.

de Queiroz, K. 1998. The general lineage concept of species, species criteria, and the process of speciation: A conceptual unification and terminological recommendations *in* D. J. Howard, and S. H. Berlocher, eds. Endless forms: Species and speciation. Oxford University Press, New York.

de Queiroz, K. 2007. Species concepts and species delimitation. *Syst. Biol.* 56:879-886.

Dobson, M. 1994. Patterns of distribution in Japanese land mammals. *Mamm. Rev.* 24:91-111.

Dobson, M., and Y. Kawamura. 1998. Origin of the Japanese land mammal fauna: allocation of extant species to historically-based categories. *The Quaternary Research (Daiyonki-Kenkyu)* 37:385-395.

Drovetski, S. V., R. M. Zink, S. Rohwer, I. V. Fadeev, E. V. Nesterov, I. Karagodin, E. A. Koblik, and Y. A. Red'kin. 2004. Complex biogeographic history of a Holarctic passerine. *Proc. R. Soc. Lond. B.* 271:545-551.

Drummond, A. J., B. Ashton, S. Buxton, M. Cheung, A. Cooper, J. Heled, M. Kearse, R. Moir, S. Stones-Havas, S. Sturrock, T. Thierer, and A. Wilson. 2010. Geneious v5.1, Available from http://www.geneious.com.

Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4:e88.

Eldredge, N., and J. Cracraft. 1980. Phylogenetic patterns and the evolutionary process. Method and theory in comparative biology. Columbia University Press, New York.

Ence, D. D., and B. C. Carstens. 2010. SpedeSTEM: a rapid and accurate method for species delimitation. *Mol. Ecol. Resour.* Online Early.

Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol. Bioinform. Online* 1:47-50.

Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479-91.

Fishpool, L., and J. Tobias. 2005. Family Pycnonotidae (Bulbuls). Pp. 124-250 *in* J. del Hoyo, A. Elliott, and D. Christie, eds. Handbook of the birds of the world: Cuckoo-shrikes to Thrushes, vol. 10. Lynx Edicions, Barcelona, Spain.

Frost, D. R., and A. G. Kluge. 1994. A consideration of epistemology in systematic biology, with special reference to species. *Cladistics* 10:259-294.

Fu, Y. X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147:915-25.

Fujita, M. K., and A. D. Leaché. 2011. A coalescent perspective on delimiting and naming species: a reply to Bauer et al. *Proc. R. Soc. Lond. B* 278:493-495.

Funk, W. C., T. D. Mullins, and S. M. Haig. 2007. Conservation genetics of snowy plovers (*Charadrius alexandrinus*) in the Western Hemisphere: population genetic structure and delineation of subspecies. *Conserv. Genet.* 8:1287-1309.

Goudet, J. 2001. FSTAT Version 2.9.3. a program to estimate and test gene diversities and fixation indices. http://www2.unil.ch/popgen/softwares/fstat.htm

Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52:696-704.

Guo, Q.-H. 2002. The population genetic structure of *Sphenomorphus taiwanensis*. MS Thesis. Taipei: National Taiwan Normal University.

Hanski, I. 1985. What does a shrew do in an energy crisis? Pp. 247-252 *in* R. M. Sibley, and R. M. Smith, eds. Behavioural ecology: ecological consequences of adaptive behavior. Blackwell Scientific Publications, Oxford.

Hanski, I. 1986. Population dynamics of shrews on small islands accord with the equilibrium model. *Biol. J. Linn. Soc.* 28:23-36.

Hedrick, P. W. 1999. Highly variable loci and their interpretation in evolution and conservation. *Evolution* 53:313-318.

Heled, J., and A. J. Drummond. 2010. Bayesian inference of species trees from multilocus data. *Mol. Biol. Evol.* 27:570-580.

Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405:907-13.

Hewitt, G. M. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Phil. Trans. R. Soc. B.* 359:183-95.

Hickerson, M. J., E. A. Stahl, and H. Lessios. 2006. Test for simultaneous divergence using approximate Bayesian computation. *Evolution* 60:2435-2453.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25:1965-1978.

Hosoda, T., H. Suzuki, M. Harada, K. Tsuchiya, S. H. Han, Y. Zhang, A. P. Kryukov, and L. K. Lin. 2000. Evolutionary trends of the mitochondrial lineage

differentiation in species of genera *Martes* and *Mustela*. *Genes Genet. Syst.* 75:259-267.

Hsu, F.-H., F.-J. Lin, and Y.-S. Lin. 2000. Phylogeographic variation in mitochondrial DNA of the Formosan white-bellied rat *Niviventer culturatus*. *Zool. Stud.* 39:38-46.

Hsu, F.-H., F.-J. Lin, and Y.-S. Lin. 2001. Phylogeographic structure of the Formosan wood mouse, *Apodemus semotus* Thomas. *Zool. Stud.* 40:91-102.

Hsu, M. J., and Y. S. Lin. 1993. Morphology, sexual dimorphism and hybridization in Styan's Bulbuls *Pycnontus taivanus* and Chinese Bulbuls *P. sinensis* formosae. *Acta Zool. Taiwanica* 4:103-111.

Hsu, V. 1990. Seismicity and tectonics of a continent-island arc collision zone at the island of Taiwan. *J. Geophys. Res.* 95:4725-4734.

Hsu, Y.-T. 1999. The evolutionary relationship of Chinese bulbul (*Pycnonotus sinensis*) and Taiwan bulbul (*P. taivanus*) revealed by their population genetic structure. MS Thesis. Taipei: National Taiwan University.

Huang, Y.-K. 2004. The island effect on the avifauna of the mountain islands in southern Taiwan. MS Thesis. NeiPu Township: National Pingtung University of Science and Technology.

Huang, Y. J., C. S. Chen, and S. H. Li. 2004. Polymorphic tetranucleotide microsatellite loci in the Hwamei (*Garrulax canorus canorus*) (Timaliidae). *Mol. Ecol. Notes* 4:170-172.

Huang, W., N. Takebayashi, Y. Qi, and M. J. Hickerson. 2011. MTML-msBayes: approximate Bayesian comparative phylogeographic inference from multiple taxa and multiple loci with rate heterogeneity. *BMC Bioinformatics* 12:1.

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

Huson, D. H., and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23:254-267.

Iwasa, M., I. Kartavtseva, A. Dobrotvorsky, V. Panov, and H. Suzuki. 2002. Local differentiation of *Clethrionomys rutilus* in northeastern Asia inferred from mitochondrial gene sequences. *Mamm. Biol.* 67:157-166.

Iwasa, M. A., Y. Utsumi, K. Nakata, I. V. Kartavtseva, I. A. Nevedomskaya, N. Kondoh, and H. Suzuki. 2000. Geographic patterns of cytochrome b and Sry gene lineages in the gray red-backed vole *Clethrionomys rufocanus* from Far East Asia including Sakhalin and Hokkaido. *Zool. Sci.* 17:477-484.

Kawamura, Y. 1991. Quaternary mammalian faunas in the Japanese Islands. *The Quaternary Research (Daiyonki-Kenkyu)* 30:213-220.

Knowles, L. L., and B. C. Carstens. 2007. Delimiting species without monophyletic gene trees. *Syst. Biol.* 56:887-895.

Kozak, K. H., C. H. Graham, and J. J. Wiens. 2008. Integrating GIS-based environmental data into evolutionary biology. *Trends Ecol. Evol.* 23:141-148.

Kurose, N., Y. Kaneko, A. V. Abramov, B. Siriaroonrat, and R. Masuda. 2001. Low genetic diversity in Japanese populations of the Eurasian badger *Meles meles*

(Mustelidae, Carnivora) revealed by mitochondrial cytochrome b gene sequences. *Zool. Sci.* 18:1145-1151.

Larkin, M., G. Blackshields, N. Brown, R. Chenna, P. McGettigan, H. McWilliam, F. Valentin, I. Wallace, A. Wilm, and R. Lopez. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947-2948.

Latch, E. K., G. Dharmarajan, J. C. Glaubitz, and O. E. Rhodes Jr. 2006. Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conserv. Genet.* 7:295-302.

Leaché, A. D., and M. K. Fujita. 2010. Bayesian species delimitation in West African forest geckos (Hemidactylus fasciatus). *Proc. R. Soc. Lond. B* 277:3071-3077.

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

Lieberman, B. S. 2004. Range expansion, extinction, and biogeographic congruence: a deep time perspective. Pp. 111-124 *in* M. V. Lomolino, and L. R. Heaney, eds. Frontiers of biogeography: new directions in the geography of nature. Sinauer Associate Inc., Sunderland, MA.

Losos, J. B., and R. E. Ricklefs. 2009. Adaptation and diversification on islands. *Nature* 457:830-836.

Luo, X., Y.-H. Qu, L.-X. Han, S.-H. Li, and F.-M. Lei. 2008. A phylogenetic analysis of laughingthrushes (Timaliidae: *Garrulax*) and allies based mitochondrial and nuclear DNA sequences. *Zool. Scr.* 38:9-22.

Marmi, J., J. F. López-Giráldez, and X. Domingo-Roura. 2004. Phylogeny, evolutionary history and taxonomy of the Mustelidae based on sequences of the cytochrome b gene and a complex repetitive flanking region. *Zool. Scr.* 33:481-499.

Marshall, C. R. 1990. The fossil record and estimating divergence times between lineages: Maximum divergene times and the importance of reliable phylogenies. *J. Mol. Evol.* 30:400-408.

Marshall, L. G., S. D. Webb, J. J. Sepkoski, and D. M. Raup. 1982. Mammalian evolution and the great American interchange. *Science* 215:1351-1357.

Matsuhashi, T., R. Masuda, T. Mano, K. Murata, and A. Aiurzaniin. 2001. Phylogenetic relationships among worldwide populations of the brown bear *Ursus arctos*. *Zool. Sci.* 18:1137-1143.

Matsuhashi, T., R. Masuda, T. Mano, and M. C. Yoshida. 1999. Microevolution of the mitochondrial DNA control region in the Japanese brown bear (Ursus arctos) population. *Mol. Biol. Evol.* 16:676-684.

Mayden, R. L. 1997. A hierarchy of species concepts: the denouement in the saga of the species problem. Pp. 381-424 *in* M. F. Claridge, H. A. Dawah, and M. R. Wilson, eds. Species: the Units of Biodiversity. Chapman & Hall Ltd., London.

McKay, B. D., and R. M. Zink. 2010. The causes of mitochondrial DNA gene tree paraphyly in birds. *Mol. Phylogenet. Evol.* 54:647-650.

Moore, W. S. 1995. Inferring Phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution* 49:718-726.

Moritz, C., J. L. Patton, C. J. Schneider, and T. B. Smith. 2000. Diversification of rainforest faunas: an integrated molecular approach. *Annu. Rev. Ecol. Syst.* 31:533-563.

Neall, V. E., and S. A. Trewick. 2008. The age and origin of the Pacific islands: a geological overview. *Phil. Trans. R. Soc. B.* 363:3293-3308.

Nelson, G., and N. Platnick. 1981. Systematics and biogeography: cladistics and vicariance. Columbia University Press, New York.

Nixon, K. C., and Q. D. Wheeler. 1990. An amplification of the phylogenetic species concept. *Cladistics* 6:211-223.

O'Meara, B. C. 2010. New heuristic methods for joint species delimitation and species tree inference. *Syst. Biol.* 59:59-73.

Ohdachi, S., N. E. Dokuchaev, M. Hasegawa, and R. Masuda. 2001. Intraspecific phylogeny and geographical variation of six species of northeastern Asiatic Sorex shrews based on the mitochondrial cytochrome b sequences. *Mol. Ecol.* 10:2199-2213.

Ohdachi, S., M. Hasegawa, M. Iwasa, P. Vogel, T. Oshida, L. K. Lin, and H. Abe. 2006. Molecular phylogenetics of soricid shrews (Mammalia) based on mitochondrial cytochrome b gene sequences: with special reference to the Soricinae. *J. Zool.* 270:177-191.

Ohdachi, S. D., Y. Ishibashi, M. A. Iwasa, and T. Saitoh. 2009. The wild mammals of Japan. Shoukadoh Book Sellers, Kyoto, Japan.

Ohdachi, S. D., M. A. Iwasa, V. A. Nesterenko, H. Abe, R. Masuda, and W. Haberl. 2004. Molecular phylogenetics of *Crocidura* shrews (Insectivora) in east and central Asia. *J. Mammal.* 85:396-403.

Ohshima, K. 1990. The history of straits around the Japanese Islands in the late-Quaternary. *The Quaternary Research (Daiyonki-Kenkyu)* 29:193-208.

Oliveros, C. H., and R. G. Moyle. 2010. Origin and diversification of Philippine bulbuls. *Mol. Phylogenet. Evol.* 54:822-832.

Oshida, T., A. Abramov, H. Yanagawa, and R. Masuda. 2005. Phylogeography of the Russian flying squirrel (*Pteromys volans*): implication of refugia theory in arboreal small mammal of Eurasia. *Mol. Ecol.* 14:1191-1196.

Oshida, T., A. Arslan, and M. Noda. 2009. Phylogenetic relationships among the Old World Sciurus squirrels. *Folia Zool.* 58:14-25.

Oshida, T., L. K. Lin, H. Yanagawa, H. Endo, and R. Masuda. 2000. Phylogenetic relationships among six flying squirrel genera, inferred from mitochondrial cytochrome b gene sequences. *Zool. Sci.* 17:485-489.

Oshida, T., and R. Masuda. 2000. Phylogeny and zoogeography of six squirrel species of the genus *Sciurus* (Mammalia, Rodentia), inferred from cytochrome b gene sequences. *Zool. Sci.* 17:405-409.

Ota, Y., and H. Machida. 1987. Quaternary sea-level changes in Japan. Pp. 182-224 *in* M. J. Tooley, and I. Shennan, eds. Sea-level changes. Basil Blackwell, New York.

Peakall, R., and P. E. Smouse. 2006. genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 6:288-295.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Model* 190:231-259.

Phillips, S. J., M. Dudik, and R. E. Schapire. 2004. A maximum entropy approach to species distribution modeling. In *Proceedings of the Twenty-First International Conference on Machine Learning*, pp. 655-662.

Platnick, N. I. 1976. Concepts of dispersal in historical biogeography. *Syst. Biol.* 25:294-295.

Posada, D. 2008. jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* 25:1253-1256.

Posada, D., and K. A. Crandall. 2001. Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol. Evol.* 16:37-45.

Price, T. 2008. Speciation in birds. Roberts and Company, Greenwood Village, Colorado.

Rambaut, A., and A. J. Drummond. 2007. Tracer v1.4, Available from: http;//beast.bio.ed.ac.uk/Tracer.

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

Ravelo, A. C., D. H. Andreasen, M. Lyle, A. O. Lyle, and M. W. Wara. 2004. Regional climate shifts caused by gradual global cooling in the Pliocene epoch. *Nature* 429:263-267.

Remsen, J., JV. 2005. Pattern, process, and rigor meet classification. *Auk* 122:403-413.

Rice, W. R. 1989. Analyzing Tables of Statistical Tests. *Evolution* 43:223-225.

Rogers, A. R. 1995. Genetic Evidence for a Pleistocene Population Explosion. *Evolution* 49:608-615.

Rosen, D. E. 1979. Fishes from the uplands and intermontane basins of Guatemala: Revisionary studies and comparative geography. *Bulletin of the American Museum of Natural History* 162:267-375.

Rousset, F. 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Resour.* 8:103-106.

Rozas, J., J. C. Sanchez-DelBarrio, X. Messeguer, R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496-2497.

Schluter, D. 2000. The ecology of adaptive radiation. Oxford University Press, New York.

Serizawa, K., H. Suzuki, M. A. Iwasa, K. Tsuchiya, M. V. Pavlenko, I. V. Kartavtseva, G. N. Chelomina, N. E. Dokuchaev, and S. H. Han. 2002. A spatial aspect on mitochondrial DNA genealogy in *Apodemus peninsulae* from East Asia. *Biochem. Genet.* 40:149-161.

Seutin, G., B. N. White, and P. T. Boag. 1991. Preservation of avian blood and tissue samples for DNA analyses. *Can. J. Zool.* 69:82-90.

Shih, H.-T., H.-C. Hung, C. D. Schubart, C. A. Chen, and H.-W. Chang. 2006. Intraspecific genetic diversity of the endemic freshwater crab *Candidiopotamon rathbunae* (Decapoda, Brachyura, Potamidae) reflects five million years of the geological history of Taiwan. *J. Biogeogr.* 33:980-989.

Sites, J. W., Jr., and J. C. Marshall. 2003. Delimiting species: a Renaissance issue in systematic biology. *Trends Ecol. Evol.* 18:462-470.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457-62.

Slatkin, M., R. R. Hudson. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129:555-62.

Smith, C. H. 1983. A system of world mammal faunal regions. I. Logical and statistical derivation of the regions. *J. Biogeogr.* 10:455-466.

Sorenson, M. D., and T. W. Quinn. 1998. Numts: A challenge for avian systematics and population biology. *Auk* 115:214-221.

Sorenson, M. D., J. C. Ast, D. E. Dimcheff, T. Yuri, and D. P. Mindell. 1999. Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. *Mol. Phylogenet. Evol.* 12:105-114.

Stephens, M., N. J. Smith, and P. Donnelly. 2001. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* 68:978-989.

Suzuki, T., H. Yuasa, and Y. Machida. 1996. Phylogenetic position of the Japanese river otter *Lutra nippon* inferred from the nucleotide sequence of 224 bp of the mitochondrial cytochrome b gene. *Zool. Sci.* 13:621-626.

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

Taira, A. 2001. Tectonic evolution of the Japanese island arc system. *Annu. Rev. Earth Planet. Sci.* 29:109-134.

Templeton, A. R., K. A. Crandall, and C. F. Sing. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 132:619-33.

Toda, M., M. Nishida, M. Matsui, K.-Y. Lue, and H. Ota. 1998. Genetic variation in the Indian rice frog, *Rana limnocharis* (Amphibia: Anura), in Taiwan, as revealed by allozyme data. *Herpetologica* 54:73-82.

Tsai, C.-Y. 1999. Population phylogeny of *Gekko hokouensis* based on mitochondrial 12S rRNA and cytochrome b sequences. MS Thesis. Kaohsiung: National Sun Yat-sen University.

Tsai, Y.-L. 2005. Geographic song variation of Steere's Liocichla (*Liocichla steerii*). MS Thesis. Taipei: National Taiwan University.

Tushingham, A. M., and W. R. Peltier. 1993. Implications of the radiocarbon timescale for ice-sheet chronology and sea-level change. *Quatern. Res.* 39:125-129.

Wallace, A. R. 1881. Island life. Harper and Brothers, New York.

Wiens, J. J. 2007. Species delimitation: new approaches for discovering diversity. *Syst. Biol.* 56:875-878.

Wiens, J. J., and T. A. Penkrot. 2002. Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (Sceloporus). *Syst. Biol.* 51:69-91.

Wiens, J. J., and M. R. Servedio. 2000. Species delimitation in systematics: inferring diagnostic differences between species. *Proc. R. Soc. Lond. B* 267:631-636.

Wilson, E. O., and W. Brown. 1953. The subspecies concept and its taxonomic application. *Syst. Zool.* 2:97-111.

Wu, M.-C. 2004. Phylogeography variation of Siberian weasel (*Mustela sibirica*) in Taiwan, based on control region sequences of mitochondrial DNA. MS Thesis. Kaohsiung: National Sun Yat-sen University.

Yamasaki, T. 2006. Taxonomic status of populations of the Light-vented Bulbul Pycnonotus sinensis (Gmelin, 1789)(Passeriformes: Pycnonotidae) in Taiwan and the southern Ryukyus. *Zool. Stud.* 45:168-179.

Yang, Z., and B. Rannala. 2010. Bayesian species delimitation using multilocus sequence data. *Proc. Natl. Acad. Sci. USA* 107:9264-9269.

Yasukochi, Y., S. Nishida, S. H. Han, T. Kurosaki, M. Yoneda, and H. Koike. 2009. Genetic structure of the Asiatic black bear in Japan using mitochondrial DNA analysis. *J. Hered.* 100:297-308.

Yeung, C., Y. J. Huang, and S. H. Li. 2004. Development of polymorphic microsatellite markers for the Steere's Liocichla (*Liocichla steerii*). *Mol. Ecol. Notes* 4:420-422.

Zhang, C., D.-X. Zhang, and Y. Ziheng. in press. Evaluation of a Bayesian coalescent method of species delimition. *Syst. Biol.*

Zhang, D.-X., and G. M. Hewitt. 1996. Nuclear integrations: challenges for mitochondrial DNA markers. *Trends Ecol. Evol.* 11:247-251.

Zink, R. M. 1996. Species concepts, speciation, and sexual selection. *J. Avian Biol.* 27:1-6.

Zink, R. M. 2004. The role of subspecies in obscuring avian biological diversity and misleading conservation policy. *Proc. R. Soc. Lond. B* 271:561-564.

Zink, R. M., G. F. Barrowclough. 2008. Mitochondrial DNA under siege in avian phylogeography. *Mol. Ecol.* 17:2107-2121.

Zink, R. M., R. C. Blackwell-Rago, and F. Ronquist. 2000. The shifting roles of dispersal and vicariance in biogeography. *Proc. R. Soc. Lond. B.* 267:497-503.